Phylogenetic history and global diversity patterns of plants

Dissertation

to attain the doctoral degree (Dr. rer. nat.)

of the Faculty of Forest Sciences and Forest Ecology

Georg-August-Universität Göttingen

Submitted by

Lirong Cai

born in Zhejiang, China

Göttingen, 2023

Thesis Committee:

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

Prof. Dr. Hermann Behling, Department of Palynology and Climate Dynamics, Georg-August-Universität Göttingen.

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

Referees:

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

Prof. Dr. Hermann Behling, Department of Palynology and Climate Dynamics, Georg-August-Universität Göttingen.

Prof. Dr. Wolf Eiserhardt, Department of Biology - Ecoinformatics and Biodiversity, Aarhus University.

Date of oral examination: 25th September 2023

Biodiversity is the greatest treasure we have.

Thomas Eisner

Abstract

Understanding where and how biodiversity originates and how it is maintained is one of the central questions in biogeography and macroecology. Phylogenies capture the evolutionary history of how lineages have diversified over evolutionary time. Integrating information on phylogenetic positions and evolutionary uniqueness of species into biodiversity assessments thus provides insights into biogeographic and evolutionary mechanisms underlying biodiversity, and is of paramount importance for biodiversity conservation. Plants are key elements of terrestrial ecosystems and are essential for biodiversity and humanity in terms of controlling ecosystem functioning and providing essential ecosystem services. Despite their crucial importance, knowledge of plant diversity on a global scale, accounting for evolutionary history, remains limited.

In this thesis, I fill this important gap in our understanding of plant diversity by integrating a comprehensive global dataset of regional plant inventories across different geographic regions comprising up to 320,000 plant species with broad plant phylogenies. I explored global patterns and drivers of three key aspects of plant diversity accounting for evolutionary history in particular: (i) species and phylogenetic richness (Chapter 1); (ii) phylogenetic endemism that accounts for the phylogenetic uniqueness of range-restricted species (Chapter 2); (iii) phylogenetic turnover that quantifies dissimilarities in the evolutionary relatedness of assemblages (Chapter 3).

In Chapter 1, integrating current knowledge of regional vascular plant diversity with past and present environmental variables, I tested environment-related hypotheses of broad-scale vascular plant diversity gradients, and modeled and predicted global species and phylogenetic richness using advanced machine learning techniques. Global patterns of plant diversity are shaped by a range of past and present environmental variables that interact in complex ways. While current climate and environmental heterogeneity emerged as the most important drivers, past environmental conditions left discernible legacies on current diversity patterns. The updated global maps produced as a result of the models at multiple grain sizes provide accurate estimates of vascular plant diversity, which will be a foundation for large-scale biodiversity monitoring, research, and conservation.

In Chapter 2, I uncovered patterns and determinants of phylogenetic endemism, and distinguished the drivers and centers of evolutionarily young (neoendemism) and evolutionarily old endemism (paleoendemism) for seed plants worldwide. Phylogenetic endemism was predominantly driven by environmental heterogeneity. Warm and wet climates, geographic isolation, and long-term climatic stability were also important drivers of phylogenetic endemism. Long-term climatic stability promoted the persistence of paleoendemics, while isolation promoted neoendemism, leading to islands and mountain regions in the tropics and subtropics as global centers. These findings highlight the key role of climatic and geological history in diversification and maintenance of biodiversity, and reinforce the urgency of conserving areas occupied by narrow-ranged species with unique evolutionary histories.

In Chapter 3, I tested hypotheses of environmental filtering and dispersal history on global patterns of phylogenetic and species turnover in seed plants, and assessed the contributions of these processes to phylogenetic turnover along the phylogenetic timescale. Past and present dispersal limitations promoted compositional dissimilarity among regions, but its effect was smaller for phylogenetic turnover than for species turnover, and further diminished when moving back along the phylogenetic timescale. In contrast, environmental filtering strongly

promoted both species turnover and phylogenetic turnover at different phylogenetic timescales. The findings highlight the significant influence of environmental constraints on the distribution of major seed plant lineages and the important impact of dispersal limitation on the younger lineages towards the tips of the phylogeny.

In conclusion, the integration of unprecedented plant distribution and phylogenetic information allows to reveal global patterns and drivers of plant diversity and compositions in terms of evolutionary history. The thesis uncovers global distributions of plant species and phylogenetic richness and phylogenetic endemism, and disentangles the complex effects of past and present environmental drivers. Global patterns of regional seed plant composition result from complex dispersal history related to past and present dispersal limitations and phylogenetically conserved environmental constraints, and further the relative impacts of the processes vary along the phylogenetic timescale. Notably, the findings highlight the importance of past climate change and geological history (e.g. past plate tectonics) on regional plant diversity and composition via altering key evolutionary and ecological processes of diversity generation and maintenance. Consequently, these findings enhance our understanding of biogeographical and evolutionary mechanisms structuring biodiversity and provide essential information for future biodiversity science and conservation.

Zusammenfassung

Wo und wie die biologische Vielfalt entsteht und wie sie erhalten wird, ist eine der zentralen Fragen der Biogeografie und Makroökologie. Phylogenien erfassen die Evolutionsgeschichte, wie sich die Abstammungslinien im Laufe der Evolution diversifiziert haben. Die Einbeziehung von Informationen über die phylogenetische Position und die evolutionäre Einzigartigkeit von Arten in die Bewertung der biologischen Vielfalt ermöglicht somit Einblicke in die biogeografischen und evolutionären Mechanismen, die der biologischen Vielfalt zugrunde liegen, und ist von größter Bedeutung für die Erhaltung der biologischen Vielfalt. Pflanzen sind Schlüsselelemente terrestrischer Ökosysteme und für die biologische Vielfalt und die Menschheit von grundlegender Bedeutung, da sie das Funktionieren von Ökosystemen kontrollieren und wichtige Ökosystemleistungen erbringen. Trotz entscheidenden Bedeutung von Pflanzen ist das Wissen über die Pflanzenvielfalt auf globaler Ebene und unter Berücksichtigung der Evolutionsgeschichte nach wie vor begrenzt.

In dieser Arbeit schließe ich diese wichtige Lücke in unserem Verständnis über Pflanzenvielfalt durch die Integration eines umfassenden globalen Datensatzes regionaler Pflanzeninventuren in verschiedenen geografischen Regionen, die bis zu 320.000 Pflanzenarten mit breiter Pflanzenphylogenie umfassen. Ich untersuchte die globalen Muster und Treiber von drei Schlüsselaspekten der Pflanzenvielfalt, die insbesondere die Evolutionsgeschichte berücksichtigen: (i) taxonomische und phylogenetische Diversität (Kapitel 1), (ii) phylogenetischer Endemismus, der die phylogenetische Einzigartigkeit von Arten mit eingeschränktem Verbreitungsgebiet berücksichtigt (Kapitel 2) und (iii) phylogenetischer *Turnover*, der die Unterschiede in der evolutionären Verwandtschaft von Pflanzengruppen quantifiziert (Kapitel 3).

In Kapitel 1 habe ich – unter Einbeziehung des aktuellen Kenntnisstandes über regionale Gefäßpflanzenvielfalt und sowohl historische als auch aktuelle Umweltvariablen – umweltbezogene Hypothesen zu weiträumigen Gradienten von Gefäßpflanzendiversität getestet und die globale taxonomische und phylogenetische Diversität mit Hilfe moderner maschineller Lernverfahren modelliert und vorhergesagt. Die globalen Muster der Pflanzenvielfalt werden durch eine Reihe historischer und aktueller Umweltfaktoren geprägt, die auf komplexe Weise zusammenwirken. Während sich das aktuelle Klima und die Umweltheterogenität als die wichtigsten Einflussfaktoren herausstellten, hinterließen vergangene Umweltbedingungen deutliche Spuren in den aktuellen Diversitätsmustern. Die aktualisierten globalen Karten, erstellt auf Grundlage der Modelle in verschiedenen Korngrößen, liefern genaue Schätzungen der Gefäßpflanzenvielfalt und somit eine Grundlage für die globales Biodiversitätsmonitoring, und Naturschutz.

In Kapitel 2 habe ich Muster und Determinanten von phylogenetischem Endemismus aufgedeckt und die Triebkräfte und Zentren des evolutionär jungen (Neoendemismus) und evolutionär alten Endemismus (Paläoendemismus) für Samenpflanzen weltweit unterschieden. Der phylogenetische Endemismus wurde in erster Linie durch Umweltheterogenität bestimmt. Warmes und feuchtes Klima, geografische Isolation und langfristige Klimastabilität waren ebenfalls wichtige Faktoren für den phylogenetischen Endemismus. Langfristige Klimastabilität förderte das Fortbestehen von Paläoendemiten, während Isolation Neoendemismus begünstigte, was zu Inseln und Bergregionen in den Tropen und Subtropen als globale Zentren führte. Diese Ergebnisse verdeutlichen die Schlüsselrolle der klimatischen und geologischen Geschichte für die Diversifizierung und das

Fortbestehen der biologischen Vielfalt und unterstreichen die Dringlichkeit des Schutzes von Gebieten, in denen Arten mit geringer Verbreitung und einzigartiger Evolutionsgeschichte leben.

In Kapitel 3 überprüfte ich die Hypothesen zur Umweltfilterung und zur Ausbreitungsgeschichte in Bezug auf globale Muster des phylogenetischen und des taxonomischen *Turnover* bei Samenpflanzen und bewertete die Beiträge dieser Prozesse zum phylogenetischen *Turnover* entlang der phylogenetischen Zeitskala. Historische und gegenwärtige Ausbreitungsbeschränkungen förderten die floristische Verschiedenheit zwischen den Regionen, aber ihr Effekt war kleiner für den phylogenetischen als für den taxonomischen *Turnover* und nahm weiter ab, wenn man sich entlang der phylogenetischen Zeitskala zurückbewegte. Im Gegensatz dazu förderte die Umweltfilterung sowohl taxonomischen als auch phylogenetischen *Turnover* auf verschiedenen phylogenetischen Zeitskalen stark. Die Ergebnisse unterstreichen den signifikanten Einfluss von Umwelteinflüssen auf die Verbreitung der wichtigsten Samenpflanzengruppen und den wichtigen Einfluss der Ausbreitungsbeschränkung auf die jüngeren Entwicklungslinien an den feinen Verästelungen der Phylogenie.

Zusammenfassend lässt sich sagen, dass die Integration von noch nie dagewesenen Informationen über die Verbreitung von Pflanzen und phylogenetischen Daten es ermöglicht, globale Muster und Triebkräfte der Pflanzenvielfalt und -zusammensetzung im Hinblick auf die Evolutionsgeschichte aufzudecken. Die Arbeit deckt die globale Verteilung von Pflanzenarten und phylogenetische Diversitätsowie phylogenetischem Endemismus auf und entschlüsselt die komplexen Auswirkungen vergangener und gegenwärtiger Umweltfaktoren. Globale Muster der regionalen Zusammensetzung von Samenpflanzen sind das Ergebnis einer komplexen Ausbreitungsgeschichte, die mit früheren und heutigen Ausbreitungsbeschränkungen und phylogenetisch konservierten umweltbedingten Zwängen zusammenhängt, und die relativen Auswirkungen der Prozesse variieren entlang der phylogenetischen Zeitskala. Die Ergebnisse unterstreichen insbesondere die Bedeutung des vergangenen Klimawandels und der geologischen Geschichte (z. B. der Plattentektonik) für die regionale Pflanzenvielfalt und -zusammensetzung, da sie wichtige evolutionäre und ökologische Prozesse der Entstehung und Erhaltung der Vielfalt verändern. Folglich erweitern diese Ergebnisse unser Verständnis der biogeografischen und evolutionären Mechanismen, die die biologische Vielfalt strukturieren, und liefern wichtige Informationen für die künftige Biodiversitätsforschung und den Artenschutz.

Table of contents

List of figures

List of tables

Author Contributions

Chapter 1: Global models and predictions of plant diversity based on advanced machine learning techniques

Lirong Cai[1], Holger Kreft[1,2,3], Amanda Taylor[1], Pierre Denelle[1], Julian Schrader[1,4], Franz Essl[5], Mark van Kleunen[6,7], Jan Pergl[8], Petr Pyšek[8,9], Anke Stein[6], Marten Winter[10], Julie F. Barcelona[11], Nicol Fuentes[12], Inderjit[13], Dirk Nikolaus Karger[14], John Kartesz[15], Andreij Kuprijanov[16], Misako Nishino[15], Daniel Nickrent[17], Arkadiusz Nowak[18,19], Annette Patzelt[20], Pieter B. Pelser[11], Paramjit Singh[21], Jan J. Wieringa[22] & Patrick Weigelt[1,2,3].

L.C., H.K., and P.W. conceived of the idea and developed the conceptual framework of the study. All authors were involved in collecting the data. L.C. performed the statistical analyses and wrote the first draft of the manuscript with the help of H.K., A.T. and P.W. All authors contributed to the writing and interpretation of the results.

This chapter is published in New Phytologist.

Chapter 2: Climatic stability and geological history shape global centers of neo- and paleoendemism in seed plants

Lirong Cai[1], Holger Kreft[1,2,3], Amanda Taylor[1], Julian Schrader[1,4], Wayne Dawson[23], Franz Essl[5], Mark van Kleunen[6,7], Jan Pergl[8], Petr Pyšek[8,9], Marten Winter[10], Patrick Weigelt[1,2,3]

L.C., H.K., and P.W. designed research. L.C. performed research. L.C., H.K., A.T., J.S., W.D., F.E., M.v.K., J.P., P.P., M.W., and P.W. collected the data. L.C., H.K., A.T., J.S., W.D., F.E., M.v.K., J.P., P.P., M.W., and P.W. wrote the paper.

This chapter is published in The Proceedings of the National Academy of Sciences (PNAS).

Chapter 3: Environmental filtering, not dispersal history, drives global patterns of phylogenetic turnover in seed plants at deep phylogenetic timescales

Lirong Cai[1], Holger Kreft[1,2,3], Pierre Denelle[1], Amanda Taylor[1], Dylan Craven[24,25], Wayne Dawson[23], Franz Essl[5], Mark van Kleunen[6,7], Jan Pergl[8], Petr Pyšek[8,9], Marten Winter[10], Francisco J. Cabezas[26], Viktoria Wagner[27], Pieter B. Pelser[11], Jan Wieringa[22], Patrick Weigelt[1,2,3]

L.C., H.K., and P.W. conceived the idea and developed the conceptual framework of the study. L.C., H.K., P.D., A.T., W.D., F.E., M.v.K., J.P., P.P., M.W., F. J. C., V. W., P.B.P., J.W., and P.W. collected the data. L.C. and P.W. performed the statistical analyses. P.W. developed the graph framework for calculating the cost distances across the hexagon grids. L.C. wrote the first draft of the manuscript. L.C., H.K., P.D., A.T., D.C., W.D., F.E., M.v.K., J.P., P.P., M.W., F. J. C., V. W., P.B.P., J.W., and P.W. contributed to the writing and interpretation of the results.

This chapter is an unpublished manuscript.

Author affiliations

[1] Biodiversity, Macroecology and Biogeography, University of Göttingen, 37077 Göttingen, Germany.

[2] Centre of Biodiversity and Sustainable Land Use, University of Göttingen, 37077 Göttingen, Germany.

[3] Campus-Institute Data Science, 37077 Göttingen, Germany.

[4] School of Natural Sciences, Macquarie University, Sydney, 2109 New South Wales, Australia.

[5] Division of Bioinvasions, Global Change & Macroecology, University Vienna, 1030 Vienna, Austria.

[6] Ecology, Department of Biology, University of Konstanz, 78464 Konstanz, Germany.

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

[8] Czech Academy of Sciences, Institute of Botany, Department of Invasion Ecology, 252 43 Průhonice, Czech Republic.

[9] Department of Ecology, Faculty of Science, Charles University, 128 44 Prague, Czech Republic.

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

[11] School of Biological Sciences, University of Canterbury, 8140 Christchurch, New Zealand.

[12] Departamento de Botánica, Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción, Chile.

[13] Department of Environmental Studies and Centre for Environmental Management of Degraded Ecosystems (CEMDE), University of Delhi, 110007 Delhi, India.

[14] Swiss Federal Institute for Forest, Snow and Landscape Research WSL, 8903 Birmensdorf, Switzerland.

[15] Biota of North America Program (BONAP), Chapel Hill, 27516 NC, USA.

[16] 650065 Kemerovo, Russia.

[17] Plant Biology Section, School of Integrative Plant Science, College of Agriculture and Life Science, Cornell University, Ithaca, 14853 NY, USA.

[18] Department of Botany and Nature Protection, University of Warmia and Mazury in Olsztyn, 10-728 Olsztyn, Poland.

[19] PAS Botanical Garden, 02-973 Warszawa, Poland.

[20] Hochschule Weihenstephan-Triesdorf, University of Applied Sciences, Vegetation Ecology, 85354 Freising, Germany.

[21] Mohali, 140301 Punjab, India.

[22] Naturalis Biodiversity Center, 2333 CR Leiden, The Netherlands.

[23] Department of Biosciences, Durham University, Durham, DH1 3LE, UK.

[24] Centro de Modelación y Monitoreo de Ecosistemas, Facultad de Ciencias, Universidad Mayor, Santiago, Chile.

[25] Data Observatory Foundation, DO, Eliodoro Yáñez 2990, oficina A5, Providencia, Santiago, Chile.

[26] Deparment of Biodiversity, Ecology and Evolution, Faculty of Biological Sciences, Complutense University of Madrid. Jose Antonio Novais 12, 28040. Madrid. Spain.

[27] Department of Biological Sciences, University of Alberta, Edmonton, T6G 2E9 Alberta, Canada.

General Introduction

Historical developments of biodiversity patterns

Earth is home to a remarkable diversity of vascular plants that comprise well over 340,000 species (Govaerts *et al.*, 2021), derived from evolutionary radiation over more than 400 million years (Morris *et al.*, 2018; Nie *et al.*, 2020). How this biological diversity varies across the Earth's surface and what determines these patterns is one of the main questions in biogeography and macroecology (Lomolino *et al.*, 2016).

Many insights into the geographic variation of biological diversity were gained during the 1700s and 1800s. Carl Linnaeus (1707-1778) remarked that species were adapted to the environmental conditions of particular habitats, and then Georges-Louis, Leclerc, Comte de Buffon (1707-1788) observed that distant areas with the same environmental conditions were often inhabited by different species (Lomolino *et al.*, 2016). Natural scientists such as Johann Reinhold Forster (1729 - 1798) and Alexander von Humboldt (1769-1859) developed a more thorough understanding of biodiversity patterns, including descriptions of elevational and latitudinal diversity gradients, more integrative explanations of the links between abiotic environmental conditions and plants, and visualization of diversity patterns (Lomolino *et al.*, 2016). Particularly, using the data collected by Humboldt, the first map of the known number of plant species at that time for different regions on Earth was published (Berghaus, 1837; Bromme, 1851). In contrast to the prevailing view of that time that "species are static", Charles Darwin (1809-1882) and Alfred Russel Wallace (1823-1913) independently developed the theory of evolution through natural selection, laying the foundations for understanding adaptations and distributions of organisms across space and time (Lomolino *et al.*, 2016).

Moving on to the 20th century, many vital contributions shed light on patterns of biological diversity. There were three striking revolutions: (i) the formulation of theories of plate tectonics, (ii) the development of molecular systematics allowing us to gain information about evolutionary relationships among groups of organisms, and (iii) the theories of ecological biogeography related to biogeographical processes of immigration, extinction, and evolution (Lomolino *et al.*, 2016). These revolutions linked patterns to processes and offered new approaches to studying how and why biodiversity varies across space and time. Simultaneously, the detailed maps of vascular plant species numbers were developed with a worldwide coverage based on consecutively expanded databases and the application of statistical models (Figure 0.1) (Wulff, 1935; Malyshev, 1975; Barthlott *et al.*, 2005; Mutke & Barthlott, 2005; Kreft & Jetz, 2007).

Thus, all of these efforts provide a comprehensive and reliable overview of the distribution of plant diversity across the globe, and deepen our understanding of drivers of biological diversity. However, previous studies have largely focused on species that are only one of the diverse components of biodiversity, and biogeographical and evolutionary mechanisms, including the origin, spread and diversification of biodiversity, remain poorly understood. Nowadays, as biodiversity is under increasing pressure from human impacts (Ceballos *et al.*, 2015; Cowie *et al.*, 2022), understanding biological diversity has taken on a new urgency.
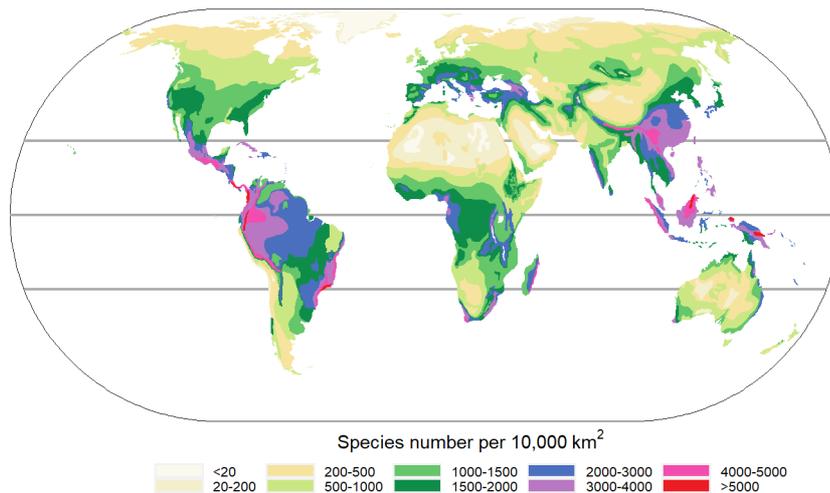
Figure 0.1 Species number of vascular plants from Barthlott et al. (2005). The map use Eckert IV projection.

Biodiversity multidimensionality and its phylogenetic perspective

It is a consensus that biodiversity is fundamentally multidimensional (Magurran, 2021), encompassing the variety of life on Earth at all levels of organization (Gaston & Spicer, 2004), beginning with genes (genetic variation within populations and across populations of the same species), extending to organisms (variety of different species and lower or higher classifications of taxa present in a given area), and ecosystem (variety of different habitats, communities and ecological processes between ecosystems) and beyond, organized at different spatial scales from niches and local habitats, through landscapes and biomes, on up to the entire planet (Díaz & Malhi, 2022). Diversity at the level of organisms can be explored through different dimensions, including taxonomy (Kreft & Jetz, 2007; Kier *et al.*, 2009), phylogenies (Faith, 1992; Mace *et al.*, 2003), and functional traits (Petchey & Gaston, 2006; Cadotte *et al.*, 2011). For each dimension, different measures are designed to capture the state and trends of biodiversity (Petchey & Gaston, 2006; Chiarucci *et al.*, 2011; Tucker *et al.*, 2017).

Species are the most common choice of units when measuring the taxonomic diversity of organisms. Despite considerable disagreement about what precisely constitutes a species, species are broadly considered to be a group of organisms that share a series of common characteristics derived from their independent evolutionary trajectories and unique histories (Gaston & Spicer, 2004). Species, therefore, offer a clear and measurable way of distinguishing one group of organisms from another, allowing us to assess the variety of different life forms in a particular area. For these reasons, coupled with more readily available data at the species level, species-centric metrics have become one of the most widely applied metrics by ecologists (GBIF, 2020; Weigelt *et al.*, 2020; Govaerts *et al.*, 2021), especially species richness (i.e. the number of species in a site, habitat and beyond), which is more intuitive compared to other measures (Purvis & Hector, 2000). However, it has been argued that species are nothing special on the Tree of Life; studies of lineages below or above the species level can also be important and promote a comprehensive understanding of biodiversity patterns and processes (Mishler, 2023). A key way is to integrate phylogenies into biodiversity assessments, accounting for the full set of nested lineages on the Tree of Life as it is currently known, thus providing key information on the evolutionary history of lineages that is missing from species lists or taxonomies.

Phylogenies quantify the evolutionary history of any set of organisms, at least on a rough timescale, as a branching tree (Smith & Brown, 2018). In other words, they detail the patterns of nested relationships among these organisms,

2

which reflect the sequence of all the events of biological evolution that have led to the diversification and adaptation of different lineages over time. As a consequence, these phylogenies represent a new window to look back into the historical ecological and evolutionary processes that have shaped current biodiversity patterns (Mace *et al.*, 2003; Graham & Fine, 2008). Beyond its application to understanding the mechanisms underlying biodiversity, biodiversity from a phylogenetic perspective also plays a prominent role in nature conservation (Winter *et al.*, 2013; Gumbs *et al.*, 2023). While evolutionary history needs to be conserved as a fundamental and crucial component of biodiversity with its intrinsic value (Mace *et al.*, 2003; Rosauer & Mooers, 2013), the conservation of evolutionary history could potentially encompass other desirable components of biodiversity, such as functional diversity (Tucker *et al.*, 2018, 2019; Mazel *et al.*, 2018). Also, evolutionary history can be associated with the conservation of evolutionary potential (i.e., the ability of lineages to evolve in response to environmental change), which is important in the face of ongoing global change (Forest *et al.*, 2007; Winter *et al.*, 2013).

The potential value of phylogenies in addressing ecological questions and in conservation, coupled with the increasing availability of large phylogenies for different taxonomic groups (Prum *et al.*, 2015; Smith & Brown, 2018), has led to a rapid increase in phylogenetic metrics. Previous studies have provided unifying frameworks to elucidate the conceptual relationships between these phylogenetic metrics and assist in their correct application and interpretation (Pavoine & Bonsall, 2011; Tucker *et al.*, 2017). One widely used scheme for grouping phylogenetic metrics is based on whether the metric utilizes the information about a single set of tips of a phylogeny within an assemblage, such as a community or region (i.e., phylogenetic alpha diversity), or multiple sets of tips (i.e., phylogenetic beta diversity) (Tucker *et al.*, 2017). Specifically, phylogenetic alpha diversity is aimed to answer the question of how much evolutionary diversity is present in an assemblage, while phylogenetic beta diversity measures the extent to which different assemblages differ in their phylogenetic composition (Graham & Fine, 2008).

Phylogenetic diversity is a simple and widely used metric of phylogenetic alpha diversity, measured as the sum of the phylogenetic branch lengths on the minimum path linking all terminals coexisting in a region (Faith, 1992). Thus, phylogenetic diversity incorporates the number of taxa and their branch lengths in a phylogeny, and quantifies the combined contribution of the taxa in an assemblage to the overall evolutionary diversity of the group. However, not all lineages are equally important for understanding and conserving biodiversity. Range-restricted species (endemics), for example, have irreplaceable ecological and evolutionary characteristics, leading to endemism as another key concept in biodiversity assessment (Grenyer *et al.*, 2006; Kier *et al.*, 2009). By combining phylogenetic diversity and weighted endemism measures, phylogenetic endemism is a measure that accounts for the phylogenetic uniqueness of range-restricted lineages (Rosauer *et al.*, 2009). Specifically, this metric is fundamentally a phylogenetic diversity-based measure, but is calculated based on a phylogeny where each branch length is divided by the total range size of the terminal taxa that descended from the branch (Rosauer *et al.*, 2009). Therefore, regions with high phylogenetic endemism harbor evolutionarily unique lineages that have restricted geographic distributions. A loss of phylogenetic endemism would cause a disproportionately large loss of evolutionary history, making phylogenetic endemism a key metric of biodiversity in conservation prioritization (Rosauer *et al.*, 2009; Mishler *et al.*, 2014). Additionally, measures of phylogenetic endemism allow for the quantitative distinction between neoendemism (recently evolved endemism) and paleoendemism (evolutionarily old endemism) (Mishler *et al.*, 2014), providing new insights into understanding evolutionary underpinnings related to the origin and maintenance of biodiversity.

Metrics of phylogenetic beta diversity are analogous measures to beta diversity of species (i.e., variation of the species composition of assemblages), such as Jaccard and Sørenson dissimilarity indices (Swenson, 2011; Leprieur *et al.*, 2012; Baselga, 2012). Different from species dissimilarity indices that consider species composition (Anderson *et al.*, 2011), phylogenetic dissimilarity indices take into account branches of a phylogeny and measure the length of branches that are shared or unique between two assemblages (Swenson, 2011; Leprieur *et al.*, 2012). Therefore, studies of phylogenetic beta diversity can help us to understand the biogeographical and evolutionary processes, such as dispersal limitation and environmental filtering, that cause assemblages of lineages to be more or less similar at different places and times (Eiserhardt *et al.*, 2013; König *et al.*, 2017).

Overall, biodiversity from a phylogenetic perspective is a key component of biodiversity, and exploring biodiversity using these complementary phylogenetic metrics could provide new approaches to testing ecological and evolutionary hypotheses of current biodiversity. However, global patterns of diversity from a phylogenetic perspective are largely unstudied in plants, a taxonomic group of fundamental ecological importance.

Drivers of biodiversity

Biological diversity is not evenly distributed across the planet, and this fundamental feature is stated in multiple trends, such as latitudinal diversity gradients (Hillebrand, 2004; Barthlott *et al.*, 2005; Buckley & Jetz, 2007) and the relatively poor diversity of Africa compared to Southeast Asia and the Neotropics (Raven *et al.*, 2020). The underlying mechanism of this geographic discrepancy remains elusive and highly controversial. A large number of hypotheses have been proposed (here, focused mainly on the biodiversity of organisms) (Currie *et al.*, 2004; Wiens & Donoghue, 2004; Mittelbach *et al.*, 2007; Sandel *et al.*, 2011; Stein *et al.*, 2014), and many of these hypotheses are not mutually exclusive. Most importantly, in any comprehensive explanation for the geographic discrepancy of biodiversity, at least one of the three processes that directly shape the diversity in a region, i.e., speciation, extinction, and dispersal, must be included (Ricklefs, 1987; Mittelbach *et al.*, 2007). Thus, environmental factors can influence biodiversity by modulating the rates and patterns of speciation, extinction, and dispersal.

Generally, diversity has been linked to limitations on dispersal and restricted gene flow among regions, which in turn foster speciation and limit range expansion (Flantua *et al.*, 2020). Dispersal limitation is promoted by geographic isolation due to physical and ecological dispersal barriers, including oceans (Kier *et al.*, 2009; Veron *et al.*, 2021), mountain ranges (Hughes & Atchison, 2015), and climatic gradients (Thompson *et al.*, 2005). For example, Mediterranean-type climate regions show high diversity and endemism of plants due to local speciation events driven by their peculiar climate (Cowling *et al.*, 1996; Thompson *et al.*, 2005; Valente *et al.*, 2010), and endemism on isolated islands exceeds that of mainland areas (Kier *et al.*, 2009). However, dispersal limitation is not static over geological time. Past geological processes like plate tectonics have altered past migration routes and biotic isolation and exchange between regions, and potentially leave an imprint on current biodiversity (Lomolino *et al.*, 2016; Couvreur *et al.*, 2021).

Beyond isolation, several environmental factors have been hypothesized to influence the probability of speciation events in a region. Climate, for example, possibly affects speciation rates, either because of the faster speed of molecular evolution at high temperatures (Rohde, 1992; Brown *et al.*, 2004), or because of relatively stronger biotic interactions in warm and humid climates (i.e. the humid tropics) (Currie *et al.*, 2004). Speciation may also be facilitated by environmental heterogeneity (i.e., environmental gradients, the amount of habitat types, resource diversity, and structural complexity) due to geographic isolation or adaptation to diverse environmental conditions

4

(Stein *et al.*, 2014). Additionally, long-term climatic stability may increase opportunities for speciation by allowing the evolution of narrow physiological tolerances and ecological specialization (Dynesius & Jansson, 2000; Jansson & Dynesius, 2002).

Diversity is facilitated by the long-term survival of lineages (i.e., low extinction) and their accumulation over long timescales. Past climatic changes, such as Quaternary glacial cycles, have caused large range shifts of species and increased extinction probabilities (Dynesius & Jansson, 2000; Svenning *et al.*, 2015). Regions with stable climates have suffered less severe environmental change across space and may have acted as refugia where plants could persist over time during the period of climate change (Jansson, 2003; Enquist *et al.*, 2019). Beyond long-term stable climates, environmentally heterogeneous regions can also act as refugia (Stein & Kreft, 2015; McFadden *et al.*, 2019), such as mountains that allow the species to move over only relatively short altitudinal distances in response to climate change, reducing their extinction risk (Jump *et al.*, 2009). Moreover, warm and humid climates are hypothesized to support larger population sizes by offering sufficient resources, promoting the persistence of lineages.

In summary, current patterns of biodiversity are the result of the joint interplay of speciation, extinction, and dispersal. Environmental factors can change biodiversity by affecting the rates of speciation and extinction, limiting species from dispersing into new adaptive areas. However, the relative importance of these processes as well as these environmental factors in modern plant distributions on a global scale remains poorly understood.

Study outline

Given the importance of evolutionary history in understanding mechanisms of biodiversity and conserving biodiversity, and the current lack of comprehensive global investigates of plant diversity from a phylogenetic perspective, I aimed to uncover the global patterns of plant diversity and their potential drivers from three different aspects in terms of evolutionary history in particular: species and phylogenetic richness (Chapter 1), phylogenetic endemism (Chapter 2), and phylogenetic beta diversity (Chapter 3). To this end, I used a global dataset of regional plant inventories from the Global Inventory of Floras and Traits (GIFT) (Weigelt *et al.*, 2020) in Chapter 1. Moreover, I integrated GIFT with the World Checklist of Vascular Plants (WCVP) (Govaerts *et al.*, 2021) to obtain a nearly complete global coverage dataset of seed plants for Chapters 2 and 3.

In Chapter 1, I tested environment-related hypotheses of broad-scale vascular plant diversity gradients and modeled and predicted species and phylogenetic richness at the global scale. I compared the performance of advanced statistical modeling techniques including machine learning in revealing complex diversity–environment relationships, providing improved models of global plant diversity. I tested hypotheses about plant diversity gradients related to geography, environmental heterogeneity, current climate, and past environmental conditions, and quantified their relative effects on plant species and phylogenetic richness. Finally, I produced ensemble predictions of global plant diversity patterns at multiple resolutions and revealed global diversity centers (i.e., regions with predicted richness values higher than the 90th quantile).

In Chapter 2, I uncovered global patterns and determinants of phylogenetic endemism for seed plants, and identify centers and drivers of evolutionarily young (neoendemism) and evolutionarily old endemism (paleoendemism). Specifically, I revealed how phylogenetic endemism of seed plants varies across 912 geographical regions worldwide. I then quantified the extent to which global patterns of plant phylogenetic endemism are driven by isolation, climate, environmental heterogeneity, and past climate change. I distinguished between centers of

neoendemism and paleoendemism and centers of both types of endemism across the globe using a categorical analysis of neo- and paleoendemism (CANAPE) (Mishler *et al.*, 2014). Finally, I assessed the effects of past climate change and geological history, including insularity and the presence of mountain ranges, on the spatial distribution of neo- and paleoendemism, by modeling the standardized effect size of relative phylogenetic endemism in response to past climatic and geological factors.

In Chapter 3, I assessed the effects of environmental filtering and dispersal history on phylogenetic and species turnover for seed plants and tested how these effects vary at different phylogenetic timescales. To do this, I calculated phylogenetic and species turnover of seed plants between all pairwise regions worldwide, and then measured the contribution of each region to global variation in plant phylogenetic and species composition using local contribution to beta diversity (LCBD) (Legendre & De Cáceres, 2013). To account for dispersal history, I calculated past and present geographical linear distances and cost distances across dispersal barriers defined as water, mountain ranges, or unsuitable climates using reconstructions of past tectonic plate arrangements, elevation, and climate. I then used generalized dissimilarity modeling (Ferrier *et al.*, 2007) to test for differences in the extent to which environmental dissimilarity related to environmental filtering process and past and current geographical distance (i.e., geographical linear distances and cost distances) shape global patterns of phylogenetic and species turnover. Furthermore, I used beta diversity through time framework (BDTT) (Groussin *et al.*, 2017; Mazel *et al.*, 2017) to compute phylogenetic turnover at different time periods along the phylogenetic timescale. I quantified how the effects of environmental dissimilarity and geographical distances on phylogenetic turnover vary at different phylogenetic timescales.

Chapter 1 Global models and predictions of plant diversity based on advanced machine learning techniques

Lirong Cai, Holger Kreft, Amanda Taylor, Pierre Denelle, Julian Schrader, Franz Essl, Mark van Kleunen, Jan Pergl, Petr Pyšek, Anke Stein, Marten Winter, Julie F. Barcelona, Nicol Fuentes, Inderjit, Dirk Nikolaus Karger, John Kartesz, Andreij Kuprijanov, Misako Nishino, Daniel Nickrent, Arkadiusz Nowak, Annette Patzelt, Pieter B. Pelser, Paramjit Singh, Jan J. Wieringa & Patrick Weigelt.

Abstract

Despite the paramount role of plant diversity for ecosystem functioning, biogeochemical cycles and human welfare, knowledge of its global distribution is still incomplete, hampering basic research and biodiversity conservation. Here, we used machine learning (random forests, extreme gradient boosting, neural networks) and conventional statistical methods (generalized linear models, generalized additive models) to test environment-related hypotheses of broad-scale vascular plant diversity gradients, and to model and predict species richness and phylogenetic richness worldwide. To this end, we used 830 regional plant inventories including c. 300,000 species and predictors of past and present environmental conditions. Machine learning showed a superior performance, explaining up to 80.9% of species richness and 83.3% of phylogenetic richness, illustrating the great potential of such techniques for disentangling complex and interacting associations between the environment and plant diversity. Current climate and environmental heterogeneity emerged as the primary drivers, while past environmental conditions left only small but detectable imprints on plant diversity. Finally, we combined predictions from multiple modeling techniques (ensemble predictions) to reveal global patterns and centers of plant diversity at multiple resolutions down to 7,774 km^2. Our predictive maps provide accurate estimates of global plant diversity available at grain sizes relevant for conservation and macroecology.

Introduction

Vascular plants comprise well over 340,000 species (Govaerts *et al.*, 2021) and are fundamental to terrestrial ecosystems maintaining ecosystem functioning (Tilman *et al.*, 2014) and providing ecosystem services (Isbell *et al.*, 2011; Cardinale *et al.*, 2012). To preserve and manage this important part of global biodiversity, knowledge of its spatial distribution and location of biodiversity centers is critical. Mapping plant distributions and diversity has a long and rich tradition starting in the 19th century, with the collation of regional plant species numbers and expert-drawn isolines of species richness (Wulff, 1935; Barthlott *et al.*, 2005; reviewed in Mutke & Barthlott, 2005). These maps have since then been refined and scaled to different resolutions (e.g. c. 12,100 km^2 in Kreft & Jetz (2007)) by modelling diversity patterns in response to environmental and spatial variables (Keil & Chase, 2019; Sabatini *et al.*, 2022), allowing for continuous predictions worldwide. The accuracy of such predictive maps depends on the quality and representativeness of available plant diversity data, environmental predictors, and models applied. Recent developments in the availability of both data and modeling techniques allows for models of plant diversity of hitherto unprecedented resolution and accuracy.

Knowledge of plant distributions worldwide has increased in recent years, thanks to international efforts to mobilize and collate species occurrence records (Enquist *et al.*, 2016; GBIF, 2020), vegetation plots (Sabatini *et*

al., 2021) along with regional checklists and floras (Weigelt *et al.*, 2020; Govaerts *et al.*, 2021). However, these data differ in precision, completeness, and scope (König *et al.*, 2019). Specifically, fine-grained data such as occurrence records and vegetation plots are often geographically biased and only partially cover regional floras (Meyer *et al.*, 2016; Qian *et al.*, 2022). Despite being coarse-grained and often delimited by artificial, administrative borders, checklists and floras reflect the most complete and authoritative accounts of regional floristic composition to date, and are available with near-complete global coverage (Weigelt *et al.*, 2020; Govaerts *et al.*, 2021). As such, checklists and floras are useful resources for global-scale modeling of plant diversity–environment relationships (Kreft & Jetz, 2007), and for predicting plant diversity across different grain sizes (Keil & Chase, 2019). Including species identities further allows for the integration of species-level phylogenetic and trait information, offering a unique opportunity to study multiple facets of biodiversity.

Although it is widely accepted that plant diversity reflects the complex interplay of evolutionary, geological, and ecological processes, disentangling the drivers of global plant diversity remains an important topic of modern macroecology (Kreft & Jetz, 2007; Tietje *et al.*, 2022). Several hypotheses related to geography, past and present climate, and environmental heterogeneity of a region have been proposed to explain plant diversity patterns (Currie *et al.*, 2004; Mittelbach *et al.*, 2007; Fine, 2015) (SI Appendix, Table S1.1). Large and heterogeneous areas, for example, are hypothesized to support more species by offering a greater diversity of resources and habitats, thus promoting species coexistence (Connor & McCoy, 1979) and offering refugia during environmental fluctuations (Stein *et al.*, 2014). Also, areas with warm, wet, and relatively stable climates such as humid tropical forests should support more species owing to high speciation (Rohde, 1992; Mittelbach *et al.*, 2007; Brown, 2014) and low extinction rates (Gillooly & Allen, 2007; Eiserhardt *et al.*, 2015). Geographic isolation could simultaneously promote extinction (Brown & Kodric-Brown, 1977; Ouborg, 1993) and speciation (Kisel & Barraclough, 2010), by making populations less well-connected. Finally, historical processes like past plate tectonics and climatic change have influenced diversity patterns through altered biotic isolation and exchange or species range shifts (Dynesius & Jansson, 2000; Svenning *et al.*, 2015; Couvreur *et al.*, 2021). However, past environmental conditions remain underrepresented in global models of plant diversity and their legacies in modern plant distributions are still poorly understood (Kissling *et al.*, 2012; Hagen *et al.*, 2021).

Diversity–environment relationships are often complex, non-linear, and scale-dependent (Francis & Currie, 2003; Keil & Chase, 2019). Many environmental predictors interact and show high levels of collinearity, thus presenting major challenges for conventional statistical models such as generalized linear models (GLMs) and generalized additive models (GAMs). Machine learning approaches represent powerful modeling tools that can effectively deal with multidimensional and correlated data and can reveal non-linear relationships and interactions of predictors without *a priori* specification (Olden *et al.*, 2008; Crisci *et al.*, 2012). Therefore, machine learning has become a promising alternative to conventional techniques in ecology (Hengl *et al.*, 2017; Park *et al.*, 2020; Sabatini *et al.*, 2022). However, its performance in modeling global plant diversity has yet to be explored. In addition to relying on one particular model type, combining predictions based on multiple modeling techniques (i.e. ensemble predictions) might decrease prediction uncertainties (Araújo & New, 2007) and can thereby further improve predictions of global plant diversity patterns.

Here, we present improved models and predictions of two key facets of vascular plant diversity, i.e. species richness and phylogenetic richness, at a global extent using advanced statistical modeling techniques. In addition to non-spatial and spatial GLMs and GAMs, we systematically assess the predictive performance of machine learning methods, including random forests, extreme gradient boosting (XGBoost) and neural networks.

8

Specifically, our aims are: (1) to compare the performance of different modeling techniques in revealing complex diversity–environment relationships and to improve global geo-statistical plant diversity models; (2) to test hypotheses on plant diversity gradients related to geography, environmental heterogeneity, current climate and past environmental conditions and to quantify their relative importance for plant species and phylogenetic richness; and (3) to predict both facets of plant diversity at multiple grain sizes across the globe. Our study is based on c. 300,000 species from checklists and floras for 830 regions across the globe (SI Appendix, Figure S1.1) collated in the Global Inventory of Floras and Traits (Weigelt *et al.*, 2020) (GIFT; SI Appendix, References S1.1), and a large, dated mega-phylogeny of vascular plants (Jin & Qian, 2019).

Materials and Methods

Species distribution data and species richness

To calculate species and phylogenetic richness, we used the species composition of native vascular plants in regional checklists and floras from GIFT (Weigelt *et al.*, 2020) (version 2.1: http://gift.uni-goettingen.de). In GIFT, all non-hybrid species names are standardized and validated based on taxonomic information provided by The Plant List (version 1.1, http:// www.theplantlist.org) and additional resources available via iPlant's Taxonomic Name Resolution Service (TNRS) (Boyle *et al.*, 2013; Weigelt *et al.*, 2020). The original database contains > 3000 geographic regions representing islands, protected areas, biogeographical regions and administrative units (e.g. countries, provinces). We excluded regions with incomplete native vascular plant checklists, incomplete data for predictor variables, or an area of less than 100 km^2. Furthermore, we coped with overlapping regions in two steps. First, for overlapping regions from one individual literature source, we only kept non-overlapping regions preferring smaller over larger regions (e.g. the individual states of Brazil instead of the country). Second, for overlapping regions from different literature sources, we retained both smaller and larger regions if smaller regions covered only parts of the larger regions. Otherwise we removed the larger regions. A total of 298,087 vascular plant species from 775 mainland regions and 55 islands or island groups was used to proceed with the calculation of species richness (i.e. taxonomic richness) and phylogenetic richness. The geographic regions in the dataset were distributed representatively across the entire globe, covering all major biomes (SI Appendix, Figure S1.1).

Phylogeny reconstruction and phylogenetic richness

We used a large, dated megatree of vascular plants, GBOTB_extended (Jin & Qian, 2019), as a backbone to generate a phylogeny for all species in the dataset. The megatree was derived from the GBOTB tree for seed plants by Smith & Brown (2018) and the phylogeny for pteridophytes in Zanne *et al.* (2014). We excluded taxa not identified to the species level for calculating phylogenetic richness, leading to a dataset including 295,417 species in 466 families of vascular plants. All families and 10,128 out of 14,962 genera (67.7%) in the dataset were included in the megatree. We bound the remaining genera and species into their respective families and genera using "Scenario 3" in the R package *V.PhyloMaker* (Jin & Qian, 2019). In "Scenario 3", the weighted positioning of the additional taxa depends on the length and amount of already existing tips per taxon. 91.95% out of the 295,417 species in the dataset were from genera already present in the backbone. It is suggested that patterns of phylogenetic richness would be similar regardless of whether the phylogeny used is resolved at the genus or species level (Qian & Jin, 2021). Additionally, we carried out a sensitivity analysis to test for the effect of adding missing genera to the phylogeny on phylogenetic richness and found consistent patterns, indicating that our method is robust (See SI Appendix, Methods S1.1 for details).

Several indices exist for capturing different dimensions of phylogenetic diversity including richness, divergence and regularity (Tucker *et al.*, 2017). Here, we focus on phylogenetic richness, which represents the amount of unique phylogenetic history present in an assemblage (Tucker *et al.*, 2017). We chose Faith's PD, a common measure of phylogenetic richness, calculated as the sum of the branch lengths of all species coexisting in a region (Faith, 1992), which is directly comparable to species richness. Even though highly correlated to species richness (Pearson's r = 0.98), we did not standardize phylogenetic richness (i.e. assessing the deviation of phylogenetic richness from expectations based on species richness) in our main analyses as we were not interested in whether the phylogenetic structure of a region is overdispersed or clustered, but rather aimed to capture both taxonomic and phylogenetic aspects of plant diversity. However, we presented an analysis on the drivers of deviations in phylogenetic richness from species richness in SI Appendix, Table S1.2.

Predictor variables

We identified a set of candidate predictor variables hypothesized to affect plant distributions and diversity and classified them into four categories: geography, current climate, environmental heterogeneity and past environmental conditions. Twenty-five predictors were considered in the original dataset (SI Appendix, Table S1.1). These have been shown or hypothesized to contribute to geographic patterns of plant diversity in previous studies (Kreft & Jetz, 2007; Kissling *et al.*, 2012; Stein *et al.*, 2014; Keil & Chase, 2019). Geographic variables were region area (km^2) and the summed proportion of landmass area in the surrounding area of the target region within buffer distances of 100 km, 1000 km, 10,000 km, serving as a measure of geographic isolation (Weigelt & Kreft, 2013). Current climatic variables included 13 biologically relevant temperature and precipitation variables. These variables represent annual averages, seasonality and limiting climatic factors (e.g. length of growing season), capturing the main aspects of climate important for plant diversity (Karger *et al.*, 2017). Furthermore, gross primary productivity (Zhao & Running, 2010) was included as a measure of potential plant productivity based on available solar energy and water. Climatic variables were extracted as mean values across the input raster layers per region. The number of soil types (Hengl *et al.*, 2017) and elevational range (Danielson & Gesch, 2011) were calculated for each region as proxies for environmental heterogeneity within regions.

To determine the contribution of past environmental conditions to modern diversity patterns, we calculated biome area variation since the Pliocene and the Middle Miocene, temperature anomaly since the mid-Pliocene warm period, temperature stability since the last glacial maximum (LGM) and velocity of temperature change since the LGM. Terrestrial biomes are affected by multiple drivers containing atmospheric circulation, precipitation and temperature patterns, and thus changes in biome distributions represent major environmental changes through geological time. To calculate biome area variation, we used biome distribution maps at present (Olson *et al.*, 2001), the LGM (c. 25 – 15 ka) (Ray & Adams, 2001), the mid-Pliocene warm period (mid-Piacenzian, c. 3.264 – 3.025 Ma) (Dowsett *et al.*, 2016) and the Middle Miocene (c. 17 – 15 Ma) (Henrot *et al.*, 2010). The three paleo-time periods represented particularly different climates compared to present-day conditions, and showed distinct biome distributions which are hypothesized to have left imprints on current plant diversity (Svenning *et al.*, 2015; Sandel *et al.*, 2020). Since biome definitions differed across the four datasets, we regrouped biomes to match across datasets and then calculated biome area changes (See SI Appendix, Methods S1.2 for details; SI Appendix, Table S1.3). We acknowledge potential drawbacks of this approach due to the coarse resolution and uncertainty of the original past biome maps. Because of the coarse resolution of the Middle Miocene map and absent data for some geographic regions, we only used biome area variation since the Pliocene and excluded Miocene biome variation from further analyses.

In addition, we calculated temperature stability from two paleo-time periods until present, i.e. the LGM and the mid-Pliocene warm period, representing cooler and warmer climates compared to the current climate, respectively. Temperature stability since the LGM was calculated using the *climateStablity* R package (Owens & Guralnick, 2019). It takes temperature differences between 1000 year time slices expressed as standard deviation and averages the results across all time slices. The stability is then calculated as inverse of the mean standard deviation rescaled to [0,1]. Temperature anomaly since the mid-Pliocene was calculated as the difference in mean annual temperature between the mid-Pliocene warm period and present-day. The velocity of temperature change since the LGM was calculated as the ratio between temporal change and contemporary spatial change in temperature, representing the speed with which a species would have to move its range to track analogous climatic conditions (Sandel *et al.*, 2011). For details on paleoclimate estimates see SI Appendix, Methods S1.2.

An alternative way to evaluate effects of biogeographic history on plant diversity is to account for predefined discrete geographic regions influencing diversity via differences in diversification history and dispersal barriers. We therefore included floristic kingdoms (Takhtajan, 1986) as an additional categorical variable in the models and compared the performance of models with and without floristic kingdoms to assess if we managed to model the effect of biogeographic history properly by only including the variables that directly quantify past environmental change.

Statistical models

Predictor variable selection To quantify diversity–environment relationships, we fitted five different types of models with species richness and phylogenetic richness as response variables: GLMs, GAMs, random forests, XGBoost and neural networks. To compare model performance across model types, we used the same set of predictors across models. Since there was significant collinearity between the 22 predictors in the initial dataset, we removed variables with low contribution to predictions until the variance inflation factors (VIF) of all remaining variables was below a threshold of five. It has been suggested that a VIF value that exceeds five indicates a problematic amount of collinearity (James *et al.*, 2013). The contribution to predictions was based on a preliminary ranking of predictor variables using random forests and a stepwise forward strategy for variable introduction (Genuer *et al.*, 2015). Like this, we selected a subset of 15 predictor variables minimizing redundancy and maximizing model performance to fit models (bold in SI Appendix, Table S1.1; SI Appendix, Figure S1.2). The predictors retained represented all aspects (geography, current climate, environmental heterogeneity and past environment) that are hypothesized to affect plant diversity patterns.

Modeling To perform GLMs and GAMs, we used a negative binomial error distribution with a log-link function for species richness to cope with over-dispersion of the response variables, and a Gaussian error distribution with log-link function for phylogenetic richness. For the GLMs, some predictors were log-transformed owing to their skewed distribution (i.e. area, temperature seasonality, number of wet days, precipitation seasonality, precipitation of warmest quarter, gross primary productivity, elevational range, number of soil types and velocity in temperature since the LGM). After log-transformation, all continuous predictor variables were standardized to zero mean and unit variance to aid model fitting and make their parameter estimates comparable. Although fitting GLMs with 15 predictors might seem excessive, it is suggested not to exclude predictors hypothesized to be important when collinearity is minimized and not a hindrance to analysis (Morrissey & Ruxton, 2018). Thus, in our GLMs, we built the full model including 15 predictors and then simplified the model using Akaike's information criterion (AIC). Predictors were tested in turn, and removed if AIC values were larger in the complex models compared to the simpler ones (Phillips *et al.*, 2019) (SI Appendix, Table S1.4). To account for interactive effects of

environmental predictors on diversity patterns, we fitted GLMs including energy-water, energy-environmental heterogeneity and area-environment interactions, as suggested by previous studies (Kreft & Jetz, 2007; Stein *et al.*, 2014; Keil & Chase, 2019). Models including interactions were simplified based on AIC values. First, all interactions were tested, and then, any main effects (i.e. individual predictors) that were not included in the retained interactions were tested (Phillips *et al.*, 2019). In GAMs, we used penalized regression smoothers (with nine spline bases for species richness and 10 spline bases for phylogenetic richness) for each predictor to estimate the smooth terms. The number of spline bases were selected from values between two and 10 using random cross-validation to optimize model performance (i.e. minimizing root mean square error). Additionally, we used a gamma value of 1.4 to reduce overfitting without compromising model fit (Wood, 2006) and also included a double penalty to variable coefficients. We used the R packages *MASS* (Venables & Ripley, 2002) to fit negative binomial generalized linear models and *mgcv* (Wood, 2006) to fit GAMs.

In addition, we applied machine learning techniques, i.e. random forests, XGBoost and neural networks, to fit global models of plant diversity. Random forests are an ensemble learning method that builds a large collection of decision trees and outputs average predictions of the individual regression trees, while XGBoost is an ensemble model of decision trees trained sequentially fitting the residual errors in each iteration. Several innovations make XGBoost highly effective, including a novel tree learning algorithm for handling sparse data, and a theoretically justified weighted quantile sketch procedure enabling handling instance weights in approximate tree learning (Chen & Guestrin, 2016). Neural networks are a machine learning method that comprises a collection of connected units (neurons) and their connections (edges). For these machine learning methods, species and phylogenetic richness were log-transformed prior to modeling to reduce skewness of their distributions. A set of tuning parameters (i.e. hyperparameters), which cannot directly be estimated from the data, needs to be set beforehand. These hyperparameters determine the training strategy and related efficiency of the algorithms. It is commonly suggested to tune hyperparameters to maximize model performance before running models for a certain problem (Bergstra & Bengio, 2012). We used the *train* function from the R package *caret* to optimize the model tuning parameters for the three machine learning models used here (Kuhn, 2008). We used repeated random cross-validation and selected the hyperparameters that produced the lowest root mean squared error. We then refitted the final models using these optimal hyperparameters. The R package *ranger* was used to fit random forests (Wright & Ziegler, 2017), *xgboost* to fit XGBoost (Chen & Guestrin, 2016) and *neuralnet* to fit neural networks (Günther & Fritsch, 2010). Unlike GLMs and GAMs, machine learning can detect and model interactions of predictors without *a priori* specification, and we visualized interactions in machine learning models using partial dependence plots. For details on tuning parameters, model fitting using machine learning techniques and visualization of interactions, see SI Appendix, Methods S1.3.

Spatial terms Species distribution data and environmental predictors are often spatially autocorrelated. On the one hand, this might lead to biased parameter estimates which need to be accounted for (Dormann *et al.*, 2007). On the other hand, including spatial information in models could increase their predictive power (Keil & Chase, 2019). Because of this, we generated spatial models using different modeling techniques. To account for spatial autocorrelation in GLM residuals, we used simultaneous autoregressive (SAR) models of the spatial error type, which is recommended for use when dealing with spatially autocorrelated species distribution data (Kissling & Carl, 2008). We evaluated SAR models with different neighborhood structures and spatial weights (lag distances between 200 and 3000 km, weighted and binary coding). As the final SAR model, we chose a model with weighted neighborhood structure and 800 km lag distance for both species and phylogenetic richness, which had the minimal

AIC and the best reduction of spatial autocorrelation in the residuals. Species and phylogenetic richness were log-transformed prior to modeling. In GAMs, we added a two-dimensional spline on geographical coordinates, which accounts for spatial autocorrelation in model residuals (Dormann *et al.*, 2007; Keil & Chase, 2019). To cope with spatial autocorrelation in machine learning models, we included cubic polynomial trend surfaces (i.e. latitude (Y), centered longitude (X) as well as X^2, XY, Y^2, X^3, X^2Y, XY^2 and Y^3) (Bjorholm *et al.*, 2005; Li, 2019). Overall, the spatial models successfully removed spatial autocorrelation from model residuals (SI Appendix, Figure S1.3).

Comparison to established models To compare our models to published global models of plant species richness, we rebuilt these models for the data set analyzed here. First, we fitted the best model as in Kreft & Jetz (2007), a combined six-predictor model using GLMs; and second, we built a GAM using the same model structure as Keil and Chase's smooth model (Keil & Chase, 2019), which contained a two-dimensional spline on geographical coordinates, 15 single predictors and interactions between each individual predictor and area. We ran models including the same 15 predictor variables and floristic kingdom using random forests and XGBoost, and compared them with the models without floristic kingdom. Adding floristic kingdom increased collinearity between predictors. However, the two tree-based models are able to handle multicollinearity when they are used for prediction. Random forests in the *ranger* R package can handle categorical variables automatically; however, XGBoost only works with numeric vectors. We therefore converted all other forms of data into numeric vectors. Here we used one-hot encoding (0,1) to convert the floristic kingdom into dummy variables for the XGBoost model.

Variable importance To estimate the relative importance of each environmental predictor, we used a consistent method across model types. We randomly reshuffled values of the predictor of interest in the dataset, predicted the response variables based on the modified dataset and calculated the Spearman rank correlation coefficient between those predictions and the predictions using the original dataset. The relative importance of the predictor of interest was calculated as one minus the correlation coefficient divided by the sum of one minus the correlation coefficients of all predictors (Thuiller *et al.*, 2009). Likewise, to compare the relative importance of different categories of predictor variables (categories in SI Appendix, Table S1.1), we permuted values of a subset of predictors belonging to one category, correlated the predictions of the model using the modified dataset and predictions using the original dataset, and estimated the importance of each category as one minus the Spearman rank correlation coefficient divided by the sum of one minus the correlation coefficients of all predictor categories. Relationships between diversity metrics and predictor variables were visualized as partial dependence plots (see SI Appendix, Methods S1.3 for details).

Cross-validation

To assess the accuracy of model predictions across all different model types, we used random 10-fold cross-validation and spatial 68-fold cross-validation following Ploton *et al.* (2020) (for details see SI Appendix, Methods S1.4). To quantify model predictive performance, we summarized the cross-validation results using root mean squared error and two different pseudo-coefficients of determination to quantify the amount of variation explained by the model based on out-of-bag samples. R^2_CORR is the coefficient of determination of a linear model of the predicted and observed values from all repetitions of the cross-validation. R^2_Accuracy is the amount of variation explained by the model, calculated as R^2_Accuracy $=[1-SSE/SST]$ (Hengl *et al.*, 2017), where SSE is the sum of the squared error between observation and prediction and SST is the total sum of squares. The model with the lowest root mean squared error and highest R^2_CORR/ R^2_Accuracy was identified as the best predictive model. For all models, we calculated cross-validation results for log-transformed observed and predicted species and

phylogenetic richness, because species and phylogenetic richness were log-transformed prior to modeling for machine learning models and fitted with log link functions in GLMs and GAMs.

Variation explained according to spatial cross-validation was consistently lower than variation explained according to random cross-validation, likely because the former offers biased and pessimistic estimates (Wadoux *et al.*, 2021). Spatial cross-validation excludes entire portions of regions with specific combinations of environmental characteristics and biogeographic histories from the training data and is therefore less representative of the globe and its environmental spectrum, likely causing predictions outside of covariate space within the models. In contrast, random cross-validation is almost unbiased when the sampling design is systematic or random (Wadoux *et al.*, 2021). Because the geographic regions in our dataset were distributed representatively across the entire globe, covering all major biomes (SI Appendix, Figure S1.1), we argue that random cross-validation offers relatively unbiased assessments of model performance.

Predictions

We used the resulting models to predict vascular plant species and phylogenetic richness across global grids of four different resolutions (i.e. 7,774; 23,322; 69,967 and 209,903 km^2 hexagon size). We used the *dggridR* R package (Barnes & Sahr, 2017) to produce a grid of equal-area and equidistant hexagons across the Earth's surface clipped for global coastlines. Islands smaller than 1.5 times the gridcell size were treated as entire entities instead of subdividing them into several partial grid cells. For each hexagon, we calculated the same predictor variables as for the geographic regions used for fitting the models. We then used the models to predict vascular plant species and phylogenetic richness, and mapped the predictions across the hexagon grid. Due to missing values in some predictor variables, a few values had to be interpolated for predicting (see SI Appendix, Methods S1.5 for details).

Besides predictions based on individual models, we used an ensemble prediction procedure, which averages the predictions based on the models fitted by different techniques weighted by model accuracy (the inverse of the model squared error) from the random cross-validation process (Marmion *et al.*, 2009). Because spatial cross-validation was biased (Wadoux *et al.*, 2021), we used model accuracy from random cross-validation. In addition to the hexagon grids, we generated plant diversity maps in raster format at a resolution of 30 arc seconds based on predictions for the 7,774 km^2 hexagons (see SI Appendix, Methods S1.5; SI Appendix, Figure S1.4). As centers of plant diversity based on the ensemble predictions, we defined regions with predicted richness values higher than the 90th quantile, i.e. containing at least 1,765 plant species and 41,866 Ma of phylogenetic richness at a resolution of 7,774 km^2.

Uncertainty

To assess variation of the predictions across models, we calculated the coefficient of variation of predicted values for each hexagon grid cell. The coefficient of variation is defined as the ratio of the standard deviation to the mean, which accounts for the differences in diversity between regions and thereby avoids artificially high uncertainty of high diversity regions. Additionally, we calculated standard errors of predictions for GLMs, GAMs and random forests. For XGBoost and neural networks, we modelled the relationship between model residuals and environmental predictors from the raw data, and used this model to predict uncertainty across the hexagon grids.

Results and Discussion

Performance of plant diversity models

Our results reveal a great potential of machine learning, particularly decision-tree methods, for modeling plant diversity–environment relationships and for accurately predicting plant diversity across various scales. Overall, the predictive power of the models was high (Table 1.1). Machine learning models and GAMs outperformed GLMs, and spatial models (i.e. models containing spatial terms to account for the spatial non-independence of regions) (Dormann *et al.*, 2007) showed an overall better performance than non-spatial models (except GLMs for species richness). Extreme gradient boosting, an ensemble of sequentially trained decision trees, produced the most accurate predictions for both species richness (70.3% variation explained based on spatial cross-validation, 80.9% based on random cross-validation) and phylogenetic richness (73.7 and 83.3%, respectively), which was consistent across spatial and non-spatial models.

Table 1.1 Performance of global models of vascular plant diversity based on cross-validation.

Models	Species richness				Phylogenetic richness (Faith's PD)			
	Random cross-validation		Spatial cross-validation		Random cross-validation		Spatial cross-validation	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
Non-spatial models								
Full GLM	0.525	0.636	0.582	0.561	0.514	0.452	0.552	0.359
Minimum adequate GLM	0.520	0.643	0.548	0.608	0.513	0.454	0.548	0.369
GLM with interaction terms	0.471	0.704	0.502	0.664	0.412	0.635	0.453	0.559
GAM	0.437	0.742	0.507	0.658	0.359	0.723	0.430	0.604
Random forests	0.415	0.761	0.511	0.639	0.317	0.784	0.395	0.667
Extreme gradient boosting	0.389	0.791	0.487	0.673	0.295	0.813	0.384	0.685
Neural networks	0.451	0.718	0.604	0.496	0.328	0.769	0.419	0.628
Spatial models								
SAR	0.537	0.600	0.548	0.584	0.416	0.629	0.426	0.611
GAM	0.413	0.769	0.499	0.667	0.340	0.751	0.416	0.633
Random forests	0.398	0.780	0.502	0.653	0.303	0.803	0.379	0.694
Extreme gradient boosting	0.371	0.809	0.463	0.703	0.279	0.833	0.351	0.737
Neural networks	0.422	0.753	0.587	0.522	0.314	0.789	0.433	0.597

Each model was evaluated for its predictive performance using both random 10-fold and spatial 68-fold cross-validation. Non-spatial models were fitted with 15 predictors representing geography, current climate, environmental heterogeneity and past environment conditions (SI Appendix, Table S1.1) except for the minimum adequate generalized linear model (GLM) and the GLM with interaction terms. Spatial models in addition contained spatial terms (i.e. simultaneous autoregressive (SAR) models, generalized additive models (GAMs) including splines of geographic coordinates and machine learning methods including cubic polynomial trend surfaces). The minimum adequate GLM was obtained by simplifying the full GLM based on Akaike's information criterion (AIC). The GLM with interaction terms was fitted including all predictors of the full GLM and interactions of energy-water, energy-heterogeneity and area-environment related variables, and was then simplified based on AIC. Because the response variables (i.e. species and phylogenetic richness) were log-transformed in models, the accuracy statistics are provided on log-scale. Based on all out-of-bag samples, values shown are: root mean squared error (RMSE); the amount of variation explained by the model calculated as one minus the ratio of the sum of the squared error between observation and prediction to the total sum of squares (R^2). For more detailed cross-validation results, see SI Appendix, Table S1.4.

The good predictive performance of machine learning models can be attributed to their ability to uncover complex, non-linear diversity–environment relationships (SI Appendix, Figures S1.5 and S1.6) and interactive effects (SI Appendix, Figures S1.7-S1.18). We found strong interactions between spatial terms and environmental variables (SI Appendix, Figures S1.7-S1.18). This indicates regional differences in plant diversity and diversity–environment relationships and shows that different combinations of environmental variables are important when predicting diversity across geographic regions (Keil & Chase, 2019). Moreover, machine learning models revealed strong interactions between energy and water availability, energy and environmental heterogeneity, as well as area and environmental variables (SI Appendix, Figures S1.7-S1.18). Also, the accuracy of GLMs increased when

including the interactions that turned out to be important in machine learning models (70.4% vs. 63.6% in species richness based on random cross-validation; 63.5% vs. 45.2% in phylogenetic richness), highlighting the role of complex interactive effects among biotic and abiotic factors in shaping global plant diversity patterns (Francis & Currie, 2003; Kreft & Jetz, 2007; Keil & Chase, 2019). By implicitly accounting for grain dependence and complex interactions among spatial and environmental variables, our machine learning models outperform previous models of plant diversity (Kreft & Jetz, 2007; Keil & Chase, 2019) (SI Appendix, Table S1.4), improving our understanding of diversity–environment relationships and allowing for improved predictions of plant diversity across scales.

Drivers of global patterns of vascular plant diversity

Current climatic variables emerged as the most important drivers of plant diversity, accounting for 34.4-48.1% of the variation in species richness and 39.7-58.2% in phylogenetic richness across models (Figure 1.1; SI Appendix, Table S1.1). High energy and water availability and low seasonality promoted species and phylogenetic richness (SI Appendix, Figures S1.5 and S1.6), supporting other large-scale studies that report strong effects of current climate on plant diversity (Francis & Currie, 2003; Hawkins *et al.*, 2003; Kreft & Jetz, 2007). Environmental heterogeneity (measured here as elevational range and number of soil types within a region) explained 21.0-40.9% of the variation in species richness and 16.3-27.2% in phylogenetic richness, with increasing heterogeneity leading to higher plant diversity as expected (Stein *et al.*, 2014). Even though species and phylogenetic richness were highly correlated (Pearson's r = 0.98), some differences emerged in diversity–environment relationships. For example, environmental heterogeneity explained less variation in phylogenetic richness than in species richness. This potentially reflects a signal of *in-situ* speciation that is promoted by high environmental heterogeneity, creating clusters of closely related species resulting in relatively low phylogenetic richness compared to species richness (Forest *et al.*, 2007). This notion was also supported by a negative effect of number of soil types on the residual variation of phylogenetic richness after accounting for species richness (SI Appendix, Table S1.2).

Geographic variables (area and geographic isolation) explained 9.8-23.1% of the variation in species richness and 18.0-24.6% in phylogenetic richness. Larger regions tend to have higher *in-situ* speciation rates owing to more opportunities for geographic isolation within a region, and lower extinction rates due to larger populations (Terborgh, 1973; Kisel & Barraclough, 2010). These effects should be most pronounced in self-contained, isolated regions like islands, mountains, or other isolated habitats, and less so in regions that are similar to their surroundings (Rosenzweig, 2003; Testolin *et al.*, 2021). Additionally, larger regions often provide a greater variety of habitats, offering more environmental niches to be occupied by species. Geographic isolation, measured here as the proportion of surrounding landmass, did not explain much variation (0.0-3.9% in species richness; 0.5-3.5% in phylogenetic diversity; SI Appendix, Figure S1.19) for both diversity facets, possibly because our dataset consisted mainly of mainland regions (93.4% of all regions). While geographic isolation is a main driver of insular plant diversity (Weigelt & Kreft, 2013), isolation and peninsular effects seem to play only a minor role on the mainland, where geographic isolation can be expected to be more important for compositional uniqueness of regions and endemism, rather than for richness (Sandel *et al.*, 2020).

We hypothesized that higher plant diversity would accumulate in regions with long-term climate stability because of low extinction and high speciation rates (Fine, 2015; Svenning *et al.*, 2015). We therefore assessed the effects of temperature stability and biome variation as proxies for past climatic change for two paleo-time periods, i.e. the last glacial maximum (LGM) and the mid-Pliocene warming period. In contrast to the expected legacy effects of historical variables on modern plant diversity, past environmental conditions only contributed 0.8-5.5% to

16

explaining species richness in most of our models, but up to 23.8% in neural networks. Likewise, past environmental conditions showed higher explanatory power (15.0%) for phylogenetic richness in neural networks than in other models (4.0%-8.5%). Models including spatial trend surfaces or discrete biogeographic regions (i.e. floristic kingdoms) to account for regional idiosyncrasies (after statistically controlling for current and past environments) further improved model fits (Table 1.1 and SI Appendix, Table S1.4). This suggests that in addition to climate stability since the LGM or mid-Pliocene warm period, biogeographic history pre-dating the Pliocene or regional idiosyncrasies other than climatic changes affected modern plant diversity. These historical regional effects are possibly due to dispersal barriers and idiosyncratic colonization and diversification histories (Qian & Ricklefs, 2004; Ricklefs & He, 2016).
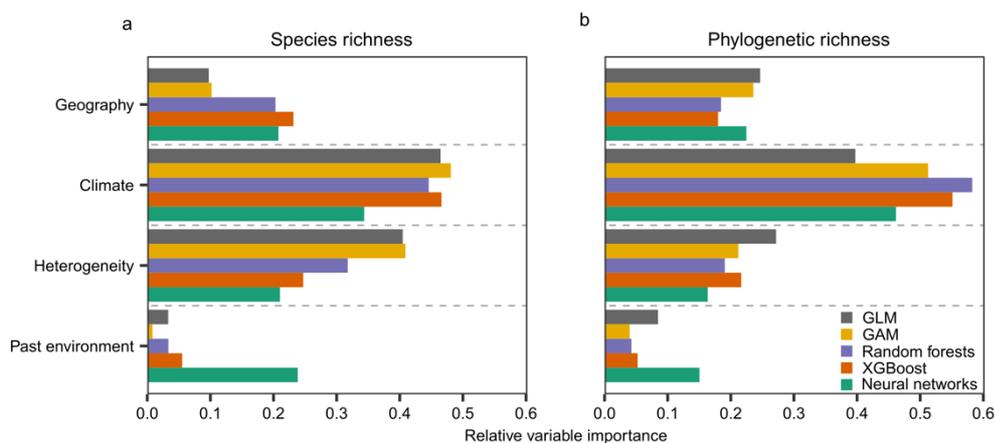
Figure 1.1 Relative importance of environmental variable categories for explaining global patterns of vascular plant diversity across five non-spatial models. a, species richness; b, phylogenetic richness (Faith's PD). Relative importance for different variable categories (scaled to sum up to one) was calculated as one minus the Spearman rank correlation coefficient between predictions of the model using a dataset where the values of the predictors of interest were permuted and predictions using the original dataset. Environmental variables falling into each category are shown in SI Appendix, Table S1.1. For the importance of individual environmental variables, see SI Appendix, Figure S1.19. GLM, generalized linear model; GAM, generalized additive model; XGBoost, extreme gradient boosting.

Improved global plant diversity maps

We produced global diversity maps for species and phylogenetic richness of vascular plants, based on individual well-performing models and model ensembles. Because of its outstanding predictive power and ability to handle missing data, we consider XGBoost (including geographic coordinates) the most powerful single model for predicting plant diversity (SI Appendix, Figures S1.20d and S1.21d). In addition, we present ensemble predictions which reduce the uncertainty introduced by the choice of one particular modeling technique and therefore improve prediction accuracy (Marmion *et al.*, 2009). Including region area and its interactions with other predictor variables allowed us to predict plant diversity across global grids of equal area and equidistant hexagons of different grain sizes (i.e. 7,774; 23,322; 69,967 and 209,903 km^2; SI Appendix, Figures S1.22 and S1.23). All model predictions and their uncertainties are accessible at https://gift.uni-goettingen.de/shiny/predictions/.
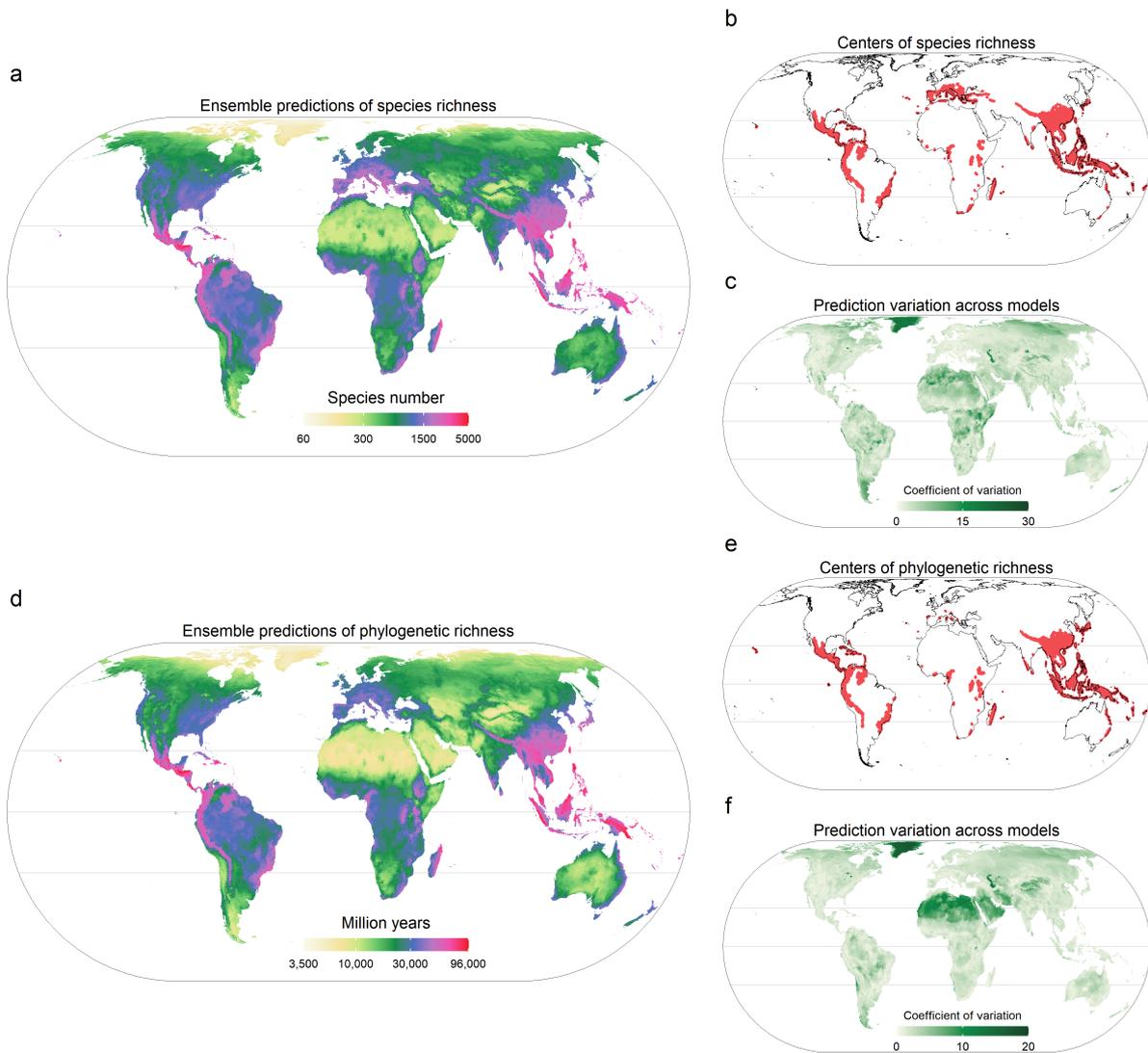
Figure 1.2 Global patterns of vascular plant diversity predicted across an equal area hexagon grid of 7,774 km² resolution. Species richness (a) and phylogenetic richness (Faith's PD, d) based on an ensemble of five different models (i.e. three spatial models using machine learning methods, a spatial generalized additive model, and a non-spatial generalized linear model with interactions) weighted by model accuracy; Species richness (b) and phylogenetic richness (e) centers defined as regions with predicted richness values higher than the 90th quantile of the predictions (i.e. containing at least 1,765 plant species and 41,866 Ma of phylogenetic richness per 7,774 km²).; Variation of predictions across models used for the ensemble predictions calculated as coefficient of variation of predicted values for species richness (c) and phylogenetic richness (f). Horizontal lines depict the equator and borders of the tropics. In a, b, d, e, $\log_{10}$ scale is used and all maps use Eckert IV projection. For maps of all different models and resolutions and data download, see https://gift.uni-goettingen.de/shiny/predictions/.

Our ensemble predictions (Figure 1.2a, d) describe the global patterns of species and phylogenetic richness with unprecedented detail and accuracy. The maps capture how diversity varies along environmental gradients and identify global centers of plant diversity (Figure 1.2b, e). The highest concentrations of plant species and phylogenetic richness are predicted in Central America, southern Mexico, Andes-Amazonia, the Caribbean, south-eastern Brazil, the Cape region of Southern Africa, Madagascar, Malay Archipelago, Indochina and southern China (Figure 1.2b, e), which is in line with empirical observations and previous studies (Myers *et al.*, 2000; Barthlott *et al.*, 2005; Kreft & Jetz, 2007). While patterns of phylogenetic richness closely resembled species richness (Pearson's r = 0.97), discrepancies occurred, for example, around the Mediterranean, in Central America, the Caucasus and Himalayas (SI Appendix, Figure S1.24). Differences might result from unequal taxonomic efforts (e.g. many closely related species described separately in Europe) or the uneven distribution of

evolutionarily old or young clades across the globe (Thorne, 1999; Endress, 2001). The former suggests that predictions of phylogenetic diversity provide a taxonomically less biased representation of global plant diversity patterns.

Thanks to the high-resolution environmental data and modeling techniques that account for complex interactions, regions with steep elevational gradients show finer tuned variation in predicted effects presented here than in previous studies (Barthlott *et al.*, 2005; Kreft & Jetz, 2007). For example, the eastern slopes of the Andes, southern Himalayan slopes, or the northern Kunlun Mountains in China show a finer differentiation from adjacent dryer and less diverse regions than in Kreft & Jetz (2007). At the same time, our ensemble predictions show relatively high values in species-poor regions like non-glaciated parts of Greenland or the Sahara. Here, and in other regions with extreme values of plant diversity, individual models perform better than the ensemble model (SI Appendix, Figures S1.20 and S1.21), which tends to attenuate extreme values. Besides the important differences just outlined, the ensemble predictions presented here were strongly correlated with model predictions in Kreft & Jetz (2007) (Pearson's r = 0.872; SI Appendix, Figure S1.25). Apart from the different modeling techniques used and how they account for complex and interactive diversity–environment relationships, differences with previous maps could derive from the accumulation of knowledge on plant diversity worldwide and the continuously updated species distribution data in GIFT used for modeling.

Regions with high species and phylogenetic richness were found to be distributed mostly in mountainous regions (SI Appendix, Figure S1.26). Specifically, tropical mountain ranges, including the tropical Andes, eastern African highlands and various Asian mountains (e.g. in southern China and the Malay Archipelago), are the global centers of plant diversity. The high diversity of tropical mountain ranges, as also found in previous studies (Testolin *et al.*, 2021), is linked to warm and wet climates and heterogeneous environments (Antonelli *et al.*, 2018). Multiple biogeographical and evolutionary processes, including speciation, dispersal, and persistence that are driven by long-term orogenic and climatic dynamics in mountains have led to outstanding regional plant diversity (Antonelli *et al.*, 2018; Rahbek *et al.*, 2019). Orogenic processes constantly change soil composition, nutrient levels and local climate of mountainous regions, thus creating novel and heterogeneous habitats where plant lineages diversify and colonize from neighboring areas (Antonelli *et al.*, 2018). Moreover, climatic fluctuations stimulate diversification by driving dynamic shifts in habitat connectivity within mountains (Rahbek *et al.*, 2019). Due to their steep environmental gradients and heterogeneous nature, mountain regions provide refugia in times of unfavorable climate (Bennett *et al.*, 1991; Rahbek *et al.*, 2019).

Differences among models (measured as coefficient of variation) were greatest in regions with extreme environments, such as deserts and Arctic regions (Figure 1.2c, f). Arctic regions also consistently showed the highest prediction uncertainty across models (SI Appendix, Figures S1.27 and S1.28). The uncertainties in regions with extreme environments probably stem from two sources. First, extremely species-poor regions might be less well represented in published diversity data. Regions with extreme environments are often part of artificially delimited regions instead of being sampled individually (e.g. Chad and Libya sampled instead of the Sahara). Those artificially delimited regions are more environmentally heterogeneous, which attenuates the extreme values of environmental factors as well as plant diversity. Machine learning models are known to not extrapolate well under such conditions (Elith *et al.*, 2010). Second, even for regions with relatively homogeneous environments, checklists and floras do not only include information on predominant but also azonal vegetation, making them richer than expected from their prevailing conditions and observed at a more local scale (comparing to alpha diversity predictions in Sabatini *et al.* (2022)).

Conclusions

We present the most accurate and comprehensive predictive global maps of regional vascular plant species and phylogenetic richness available to date. They are based on significantly improved global models using comprehensive global inventory-based plant distribution data, high resolution past and current environmental information, and advanced machine learning models. Our findings illustrate that machine learning methods applied to large distribution and environmental datasets help to disentangle underlying complex and interacting associations between the environment and plant diversity. Machine learning methods therefore help to improve both the fundamental understanding and quantitative knowledge in biogeography and macroecology. The updated global diversity maps of vascular plant diversity at multiple grain sizes (available at https://gift.uni-goettingen.de/shiny/predictions/) provide a solid foundation for large-scale biodiversity monitoring and research on the origin of plant diversity, and subsequently support future global biodiversity assessments and environmental policies.

Chapter 2 Climatic stability and geological history shape global centers of neo- and paleoendemism in seed plants

Lirong Cai, Holger Kreft, Amanda Taylor, Julian Schrader, Wayne Dawson, Franz Essl, Mark van Kleunen, Jan Pergl, Petr Pyšek, Marten Winter, Patrick Weigelt

Abstract

Assessing the distribution of geographically restricted and evolutionarily unique species and their underlying drivers is key to understanding biogeographical processes and critical for global conservation prioritization. Here, we quantified the geographic distribution and drivers of phylogenetic endemism for ~320,000 seed plants worldwide, and identified centers and drivers of evolutionarily young (neoendemism) and evolutionarily old endemism (paleoendemism). Tropical and subtropical islands as well as tropical mountain regions displayed the world's highest phylogenetic endemism. Most tropical rainforest regions emerged as centers of paleoendemism, while most Mediterranean-climate regions showed high neoendemism. Centers where high neo- and paleoendemism coincide emerged on some oceanic and continental fragment islands, in Mediterranean-climate regions and parts of the Irano-Turanian floristic region. Global variation in phylogenetic endemism was well explained by a combination of past and present environmental factors (79.8% – 87.7% of variance explained) and most strongly related to environmental heterogeneity. Also, warm and wet climates, geographic isolation, and long-term climatic stability emerged as key drivers of phylogenetic endemism. Neo- and paleoendemism were jointly explained by climatic and geological history. Long-term climatic stability promoted the persistence of paleoendemics, while the isolation of oceanic islands and their unique geological histories promoted neoendemism. Mountainous regions promoted both neo- and paleoendemism, reflecting both diversification and persistence over time. Our study provides insights into the evolutionary underpinnings of biogeographical patterns in seed plants and identifies the areas on Earth with the highest evolutionary and biogeographical uniqueness – key information for setting global conservation priorities.

Significance Statement

Range-restricted and evolutionarily unique species are a crucial yet often overlooked facet of biodiversity. Understanding the distribution of neo- and paleoendemism, i.e., identifying centers of evolutionarily young or old endemics, helps to understand the processes that shaped today's distribution of biodiversity. Here, we uncovered global patterns and determinants of phylogenetic endemism and neo- versus paleoendemism for seed plants. Environmental heterogeneity, climate and geographic isolation emerged as key drivers of phylogenetic endemism. Long-term climatic stability promotes paleoendemism, while isolation promotes neoendemism, jointly leading to oceanic and large continental islands, and mountain regions in the tropics and subtropics being global endemism centers. These results highlight the importance of climatic and geological history on diversification and persistence of biodiversity and aid conservation prioritization.

Introduction

Plant species range sizes vary widely from being nearly cosmopolitan to extremely small, for example, being restricted to a single mountain or island (Sheth *et al.*, 2020). Understanding the global distribution of range-restricted or endemic species and the mechanisms that create centers of high endemism is a central question in biogeography (Enquist *et al.*, 2019), and is crucial for the preservation of biodiversity (Kier *et al.*, 2009). Due to their restricted geographic ranges, endemic species are more vulnerable to extinction (Myers *et al.*, 2000; Pitman & Jørgensen, 2002), and, if simultaneously evolutionarily unique, their extinction may result in significant losses of evolutionary history (Purvis *et al.*, 2000; Mace *et al.*, 2003; Gumbs *et al.*, 2023). Evolutionarily unique endemics are also likely to be associated with irreplaceable ecological and functional characteristics (Faith, 1992; Veron *et al.*, 2021). It is therefore essential to account for the phylogenetic relatedness and evolutionary uniqueness of species when assessing endemism. Measures of phylogenetic endemism (PE) account for the phylogenetic uniqueness of range-restricted species (Rosauer *et al.*, 2009) and allow for the differentiation between neo- and paleoendemism (Mishler *et al.*, 2014). Regions with high PE or paleoendemism harbor more evolutionarily unique lineages with restricted geographic distributions than regions with low PE. Assessing PE for seed plants, the different types of endemism, and their past and present environmental drivers is thus crucial for setting conservation priorities and for understanding the biogeographical mechanisms underpinning plant diversity.

Endemism can originate from multiple biogeographical and evolutionary processes, which promote the formation or persistence of range-restricted species (see SI Appendix, Table S2.1 for main hypotheses of PE determinants). For one, endemism may result from speciation and limited range expansion due to dispersal limitation, which is promoted by physical or ecological barriers such as oceans, mountain ranges or climatic gradients (Thompson *et al.*, 2005; Hughes & Atchison, 2015; Fernández-Mazuecos *et al.*, 2020). Isolated regions, like oceanic islands, are renowned for their high levels of endemism (Kier *et al.*, 2009). Relatively recent speciation events on oceanic islands (Losos & Ricklefs, 2009; Weigelt *et al.*, 2015) may have resulted in an accumulation of recently evolved lineages ("cradles of diversity") that are still restricted to their area of origin, leading to so-called neoendemism (Mishler *et al.*, 2014). Alternatively, endemism can be facilitated by the long-term persistence of range-restricted species and their accumulation over long timescales, leading to paleoendemism ("museums of diversity") (Cronk, 1992; Fjeldså & Lovett, 1997). During periods of pronounced climate change (e.g., Quaternary glacial cycles), plant distributions shifted greatly, resulting in repeated range contractions followed by range expansions in more favorable periods (Dynesius & Jansson, 2000; Davis & Shaw, 2001). Therefore, regions that were climatically stable over long time periods might have served as refugia (Jansson, 2003; Enquist *et al.*, 2019). Particularly, topographically heterogeneous regions allowed species to track climate change over only relatively short altitudinal distances reducing their extinction risk (Bennett *et al.*, 1991; Jump *et al.*, 2009). After periods of climatically unfavorable conditions, not all plants could reoccupy their former ranges (i.e., relictualization) (Crisp *et al.*, 2001; Gillespie & Roderick, 2002). The resulting paleoendemics that were once widespread and are now restricted to former refugia often represent evolutionarily old lineages (Cronk, 1997; Fjeldså & Lovett, 1997).

Factors favoring the formation or persistence of endemic species do not need to be mutually exclusive. However, the influence of these processes may vary across space and over geological time, leading to regional assemblages of more recently evolved endemics or those that diverged long ago, or both (Lu *et al.*, 2018). Floras with high levels of neo- or paleoendemism have likely been shaped by different processes affecting species diversification and persistence, which jointly lead to high endemism. Assessing patterns and drivers of PE accounting for the evolutionary history of range-restricted species and distinguishing between neo- and paleoendemism thus provides

insights into past and present determinants, including geological history, climatic changes, and evolutionary processes, that structure biodiversity (Thornhill *et al.*, 2016) (see SI Appendix, Table S2.2 for main hypotheses of neo- and paleoendemism determinants). However, global tests of how plant PE and neo- versus paleoendemism are driven by climatic and biogeographical history are still lacking.

Here, we reveal global patterns and drivers of PE and neo- versus paleoendemism (Mishler *et al.*, 2014) for ~320,000 seed plant species by integrating the most comprehensive regional plant inventories across 912 geographic regions worldwide (Weigelt *et al.*, 2020; Govaerts *et al.*, 2021) (SI Appendix, Figure S2.1) with a broad seed plant phylogeny (Smith & Brown, 2018). Specifically, our aims are: (i) to reveal geographic patterns of PE for seed plants at the global scale; (ii) to test hypotheses related to isolation, environmental heterogeneity, climate, and past climate change on global patterns of seed plant PE (SI Appendix, Table S2.1); (iii) to identify centers of neoendemism and paleoendemism across the world; (iv) and to assess how past climate change and geological history shaped the centers of neo- and paleoendemism (SI Appendix, Table S2.2).

Results

Global patterns and drivers of phylogenetic endemism

Phylogenetic endemism of seed plants varied greatly among regions, being highest on islands and in topographically heterogeneous tropical mainland regions (Figure 2.1). These and all other main results are based on the global distribution of 212,525 seed plants excluding all species from 293 genera that contain apomictic species (Hojsgaard *et al.*, 2014), to avoid biases introduced by the multitude of apomictic taxa in the temperate Northern Hemisphere (see Materials and Methods for more details and SI Appendix for results based on the datasets including apomictic taxa), and only retaining species that were originally included in the phylogeny (Smith & Brown, 2018; Mishler, 2023), if not stated otherwise. To test for potential biases introduced by incomplete phylogenetic knowledge (i.e., taxa missing from the phylogeny) (Rudbeck *et al.*, 2022), we repeated all analyses based on a phylogeny with unplaced species added to their congeners (including 267,105 species when excluding apomictic taxa; see SI Appendix for the results). Because PE is scale-dependent (Daru *et al.*, 2020) and depends on reliable range size estimates, we calculated PE based on two different calculations of species range sizes: (i) the total area (PE.area) of regions a species occurs in and (ii) the number of these regions (PE.count).

We found that PE.area was almost 17-fold higher on islands than in mainland regions (mean PE.area of islands and mainland regions: 0.50 versus 0.03 $Myr \cdot km^{-2}$). PE.area peaked on subtropical islands located in the Southern Hemisphere, with Lord Howe Island having the highest PE.area overall (30.80 $Myr \cdot km^{-2}$), while the province of Pichincha in Ecuador showed the highest value among mainland regions (0.57 $Myr \cdot km^{-2}$; Figure 2.1a, b and SI Appendix, Table S2.3). In contrast, PE.count peaked in the tropics both for islands (Madagascar: 91,364 Myr) and mainland regions (Peru: 82,911 Myr; Figure 2.1c, d and SI Appendix, Table S2.3).
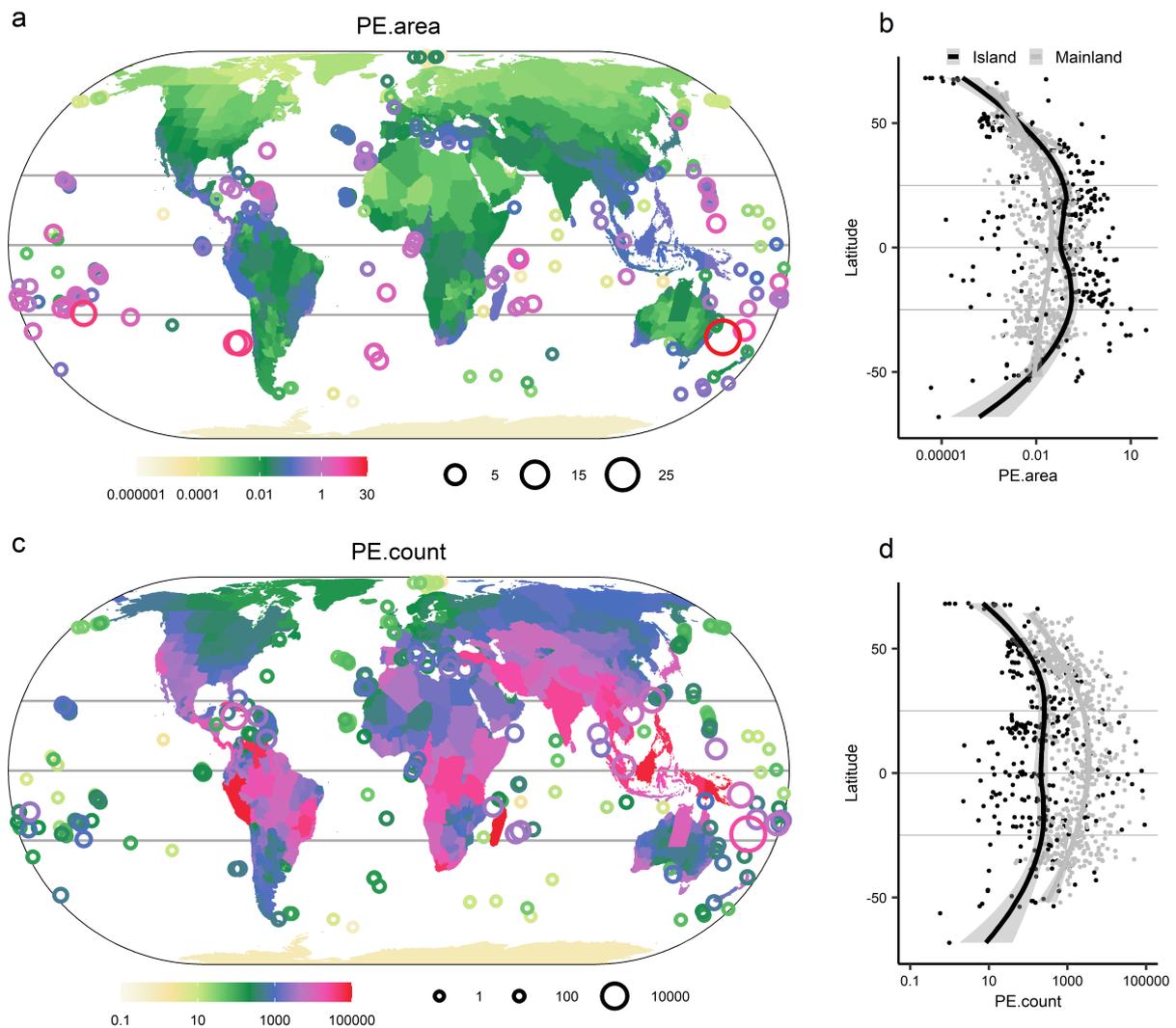
Figure 2.1 Global patterns of phylogenetic endemism of seed plants and its distribution along latitude. In a and b, phylogenetic endemism is calculated based on species range size measured as the total area of regions where a species occurs (PE.area); In c and d, phylogenetic endemism is calculated based on species range size measured as the count of regions where a species occurs (PE.count). In b and d, the fitted lines are lowess regressions, separately fitted for islands and mainland regions. Log$_{10}$ scale is used for phylogenetic endemism in all panels and maps are shown in Eckert IV projection.

The environmental factors we hypothesized to affect PE (i.e., isolation, environmental heterogeneity, climate, and past climate change; SI Appendix, Table S2.1) explained 79.8% of the variance in PE.area and 87.7% in PE.count (Figure 2.2 and SI Appendix, Table S2.4). The effects of environmental factors on PE were largely similar regardless of how range size was quantified (differing most prominently for region area which had a positive effect on PE.count and a negative effect on PE.area) (Figure 2.2a). PE was most strongly associated with environmental heterogeneity, increasing, as expected, with elevational range and number of soil types (Figure 2.2a). Surrounding landmass proportion, a proxy for isolation, which is lowest for remote islands and highest for regions located in the centers of large continents (Weigelt & Kreft, 2013), was negatively related to PE.area. This indicates that high PE occurred on islands and in mainland regions that are partly surrounded by water bodies such as coastal regions or peninsulas. When unplaced species were added to the phylogeny, surrounding landmass proportion turned out to be the most important driver of PE.area and also showed a significant negative effect on PE.count (SI Appendix, Figure S2.2), which may be explained by many species from islands missing from the original phylogeny (SI Appendix, Figure S2.3). Among climatic factors, energy and water availability had strong associations with PE,

with increasing length of the growing season and mean annual temperature leading to higher PE (Figure 2.2a). Temperature and precipitation seasonality, however, had no or only weak positive effects on PE. Relatively recent past climate change left prominent traces in PE, but this was not detectable for climatic changes in deeper time. PE increased with temperature stability since the Last Glacial Maximum (LGM; 21 Ka), while velocity of temperature change since the LGM had a negative effect. However, we found no significant relationship between PE and temperature anomaly since the mid-Pliocene warm period (~3.264 – 3.025 Ma). To test if the effects of environmental predictors on PE varied between isolated regions (e.g., islands) and less isolated regions (e.g., mainland regions), we included interactions between each predictor and surrounding landmass proportion in the models. We found that the positive effect of mean annual temperature on PE increased with decreasing surrounding landmass proportion (Figure 2.2b). Nearly identical PE patterns and drivers were found across all datasets, regardless of the exclusion or inclusion of unplaced species (SI Appendix, Figure S2.2 and S2.4 and Table S2.5) and apomictic taxa (SI Appendix, Figure S2.5 – S2.7, Table S2.6 and Table S2.7).
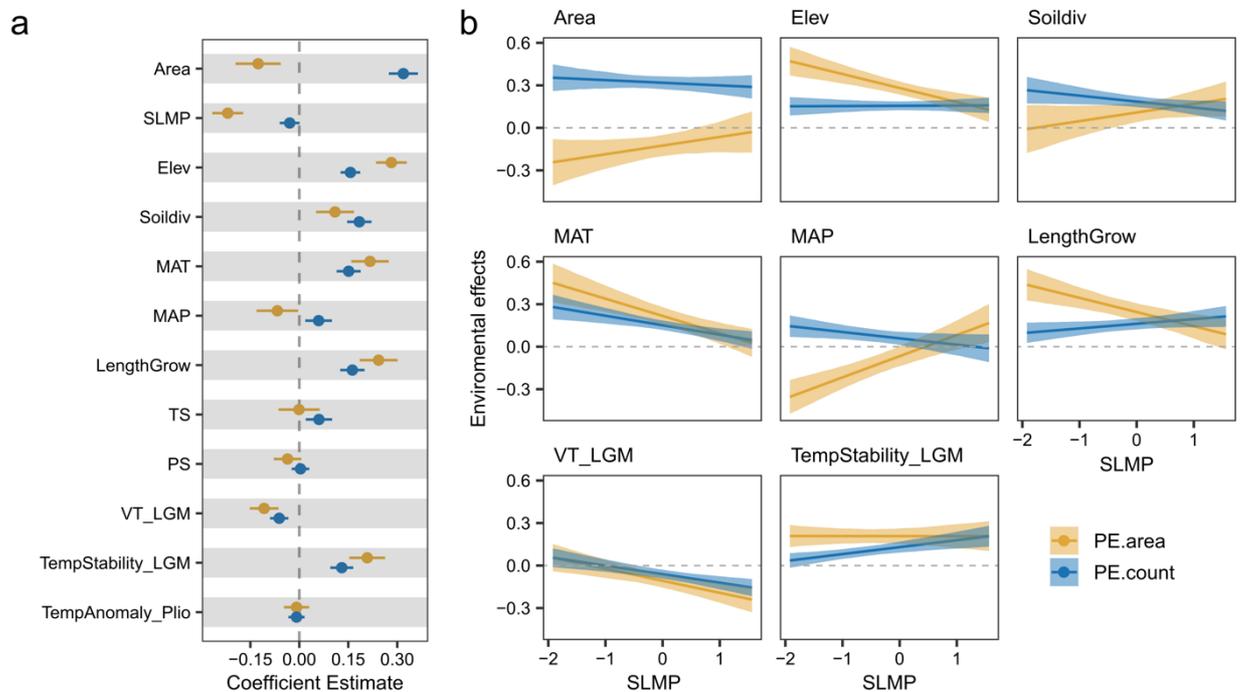
Figure 2.2 Determinants of phylogenetic endemism in seed plants based on spatial models including environmental factors and interactions between each environmental factor and surrounding landmass proportion. a, standardized regression coefficients of individual environmental factors. Bars around each point show the standard error of the coefficient estimate. b, significant interaction terms in the models visualized as effects of environmental factors on phylogenetic endemism (model coefficients on the y-axis) with varying surrounding landmass proportion (x-axis). Lines and shadings represent 95% confidence intervals. Results are shown for phylogenetic endemism based on two competing ways of measuring the range size of species. PE.area (yellow) indicates phylogenetic endemism calculated based on the range size of species as the area of regions where a species occurs, while PE.count (blue) is calculated based on the range size of species as the count of these regions. Area = region area; SLMP = surrounding landmass proportion; Elev = elevational range; Soildiv = number of soil types; MAT = mean annual temperature; MAP= mean annual precipitation; LengthGrow = length of the growing season; TS = temperature seasonality; PS = precipitation seasonality; VT_LGM = velocity of temperature change since the Last Glacial Maximum; TempStability_LGM = temperature stability since the Last Glacial Maximum; TempAnomaly_Plio = temperature anomaly between the mid-Pliocene warm period and present-day.

Figure 2.3 Global centers of neo- and paleoendemism for seed plants. In (A) species unplaced in the phylogeny are missing while they are added in (B). Colored regions present different types of endemism centers according to a categorical analysis of neo- and paleoendemism (CANAPE): violet, neoendemism; green, paleoendemism; yellow, mixed-endemism (i.e., neo- and paleoendemism); and brown indicating super-endemism (i.e., centers with both extremely high neo- and paleoendemism); beige, not significant. Patterns of neo- and paleoendemism have been calculated based on two competing ways of measuring species range size (i.e., as the area of regions where a species occurs versus as the count of these regions). Combinations of colors (hashed in mainland regions and dashed circles for islands) represent different types of endemism for a region based on these two metrics, while solid colors represent consistent endemism types. Islands that were not significant for both two metrics are represented by small and beige circles. See SI Appendix, Figure S2.8 for endemism centers based on each metric separately. Maps are shown in Eckert IV projection.

Global centers and determinants of neo- and paleoendemism

We uncovered centers of evolutionarily old and range-restricted species, centers of evolutionarily young and range-restricted species as well as centers of both using a categorical analysis of neo- and paleoendemism (CANAPE) (Mishler *et al.*, 2014). Regions identified as centers of neo- or paleoendemism occupied 27.4% (PE.area) and 31.4% (PE.count) of the global landmass area including mainland regions and islands, while regions that harbored both high neo- and paleoendemism (i.e., centers of mixed or super-endemism) only occupied 5.1% and 4.4%, respectively (Figure 2.3a and SI Appendix, Figure S2.8a, b). Many remote islands (e.g., Mauritius, Juan Fernández Islands and New Caledonia) emerged as centers of both neo- and paleoendemism (Figure 2.3a). In contrast, some continental fragment islands, such as Madagascar, Cuba and Hispaniola, and large continental islands in southeast Asia (e.g., New Guinea, Sumatra, and Java), were identified as centers of paleoendemism. When adding unplaced species to the phylogeny, Madagascar and Hispaniola turned out to be centers of super-endemism, harboring both unusually high neo- and paleoendemism (Figure 2.3b and SI Appendix, Figure S2.8e, f). Mainland regions characterized by tropical rainforests, such as Amazonia, Peru, western Colombia, central Africa, and large parts of Indochina, showed high paleoendemism (Figure 2.3a). Mediterranean-climate regions and large parts of the Irano-Turanian floristic region (Takhtajan, 1986) stood out as extra-tropical hotspots of seed-

26

plant PE among mainland regions. For example, south-western Australia, the Tibetan plateau, Afghanistan, Tajikistan and Turkey were characterized by both high neo- and paleoendemism, while the Cape of South Africa, central Chile, California, and mainland Spain were centers of neoendemism (Figure 2.3a). When adding unplaced species to the phylogeny, regions from the Irano-Turanian floristic region tended to be centers of neoendemism (Figure 2.3b). Moreover, some differences emerged depending on the measurement of species range size. For example, the Himalayas were a center of neo- and paleoendemism based on PE.count, while it did not emerge as an endemism center based on PE.area (SI Appendix, Figure S2.8a, b). In contrast, Iran tended to be a center of neoendemism based on PE.count, while it was a center of neo- and paleoendemism based on PE.area (SI Appendix, Figure S2.8a, b). Comparing patterns including and excluding apomictic taxa, the most prominent differences occurred in European countries that were identified as endemism centers when including apomictic species, due to high numbers of apomictic range-restricted species in genera like *Rubus* and *Hieracium* (SI Appendix, Figure S2.8 and S2.9).
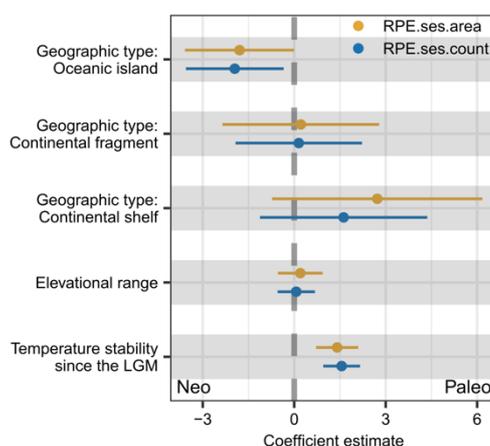
Figure 2.4 Determinants of neo- and paleoendemism. Standardized regression coefficients of environmental factors are shown from spatial models of the standardized effect size of relative phylogenetic endemism of seed plants for regions with significantly high phylogenetic endemism. A positive effect of environmental factors represents higher paleoendemism at higher values of the environmental factor, while a negative effect represents higher neoendemism. RPE.ses.area (yellow) indicates the standardized effect size of relative phylogenetic endemism calculated based on the range size of species as the area of regions where a species occurs, while RPE.ses.count (blue) is calculated based on the range size of species as the count of these regions. The reference level of geographic type is mainland regions. LGM = Last Glacial Maximum.

We assessed the impacts of geological history and past climate change on neo- and paleoendemism by modeling the standardized effect size of relative PE (see Materials and Methods for details) for regions that showed significantly high PE in response to past climatic and geological factors (SI Appendix, Table S2.2). We included the geographic type of each region (distinguishing between mainland regions and continental shelf islands, continental fragments, and oceanic islands) and elevational range (distinguishing between mountainous and non-mountainous regions) to represent geological history. Oceanic islands showed significantly higher neoendemism than mainland endemism centers identified based on PE.count (Figure 2.4 and SI Appendix, Table S2.8), and also based on PE.area when unplaced species were added to the phylogeny (SI Appendix, Figure S2.10a). However, continental islands did not show any significant difference. Elevational range had no significant effect on neo-versus paleoendemism without unplaced species added to the phylogeny (Figure 2.4) and a positive effect on neoendemism when unplaced species were added (SI Appendix, Figure S2.10a). However, when testing for

differences of environmental factors among endemism types, centers of neo- and paleoendemism both consistently had significantly higher elevational ranges than regions with low PE (SI Appendix, Figure S2.11a, b and Figure S2.12a, b). Past climate change was a major driver of neo- versus paleoendemism, with increasing temperature stability since the LGM increasing paleoendemism. Besides climate since the LGM, we also found significant relationships between temperature anomaly since the mid-Pliocene warm period and neo- and paleoendemism. Specifically, regions of super-endemism showed a significantly lower temperature anomaly since the mid-Pliocene than regions with low PE.area (SI Appendix, Figure S2.11g). When unplaced species were added to the phylogeny, a significant difference also emerged between super-endemism regions and other types of regions for both PE.area and PE.count (SI Appendix, Figure S2.12g, h). Comparing the results including and excluding apomictic taxa, the relationships between environmental variables and neo- versus paleoendemism were almost identical when unplaced species were added to the phylogenies (SI Appendix, Figure S2.10b), while differences emerged without unplaced species added (SI Appendix, Figure S2.10c). We found no difference or weakly increased neoendemism with increasing temperature stability since the LGM, while the significant difference in neo- versus paleoendemism for oceanic islands disappeared.

Discussion

Our study reveals islands and mountain regions in the tropics and subtropics as global centers of phylogenetic as well as neo- and paleoendemism of seed plants. Variation in the distribution of neo- and paleoendemism across the globe reflects the varied processes related to the generation and persistence of range-restricted species within a region, which jointly lead to high numbers of geographically restricted and evolutionarily unique lineages. We show that past climate change and geological history help to explain how diversification and relictualization (i.e., the persistence of species that went extinct elsewhere) shape the distribution of neo- and paleoendemism and simultaneously phylogenetic endemism worldwide. Understanding the drivers of different kinds of endemism and knowing particularly those regions with both high neo- and paleoendemism that act simultaneously as "museums" and "cradles" of biodiversity is of great importance for conservation prioritization, especially since global threat assessments for endemic plants are still incomplete (Gallagher *et al.*, 2023).

Geographic isolation resulted in high neo- and paleoendemism and PE on islands, which may stem from *in-situ* speciation in isolation and relictualization (Gillespie & Roderick, 2002). While speciation events require time for island species to evolve into phylogenetically distinct species, endemic species may accumulate over shorter times through relictualization (Cronk, 1987, 1992), resulting from species extinctions on the mainland and other islands. Species from lineages that diversified on islands are often young and closely related, while relict lineages on islands may be old and only distantly related to other species on the same island. Sometimes ancient endemic species are even older than the formation of the island, such as the only member of the genus *Hillebrandia sandwicensis* on the Hawaiian Islands (Clement *et al.*, 2004) and the only member of the oldest known angiosperm family (Amborellaceae) *Amborella trichopoda* on New Caledonia (Zhang *et al.*, 2020). High PE on islands may thus be a result of a combination of diversification leading to neoendemism and relictualization leading to paleoendemism. Furthermore, the diversification of island species is constrained by available resources and niches. For example, the probability of *in-situ* speciation scales positively with island size (Kisel & Barraclough, 2010). This may explain the stronger effects of some environmental factors, such as energy availability and elevational range, on island PE than on mainland PE.

When comparing islands of different geological origins, we found that oceanic islands are characterized by higher neoendemism than continental islands, which may be explained by their unique geological history (Gillespie & Roderick, 2002). Oceanic islands have not been connected with continental landmasses in the past but emerged from the oceans due to volcanic or tectonic activity. Untapped resources and the lack of enemies and competitors allowed plant species that colonized oceanic islands to diversify (Losos & Ricklefs, 2009; Fernández-Mazuecos *et al.*, 2020). Considering the relatively short geological lifespan of oceanic islands, the speciation on these islands happened comparatively recently, leading to neoendemism. However, some oceanic islands were identified as centers of super-endemism (e.g. New Caledonia), where relictualization and diversification happened in concert. Continental fragments and continental shelf islands, in contrast, were once part of continents that became separated by tectonic drift or sea-level rise. These islands were originally inhabited by floras comparable to those of the continents they were connected to. The prolonged isolation (tens of millions of years) of continental fragments allowed for the accumulation of relict lineages as well as *in-situ* speciation, which led to high neo- and paleoendemism on some of these islands. However, the origins of endemism on some large continental fragments are still debated (Antonelli *et al.*, 2022). Apart from more recent colonization events, evolution after vicariance or early long-distance dispersal events may have led to particularly old endemic species. For example, the majority of endemics on Madagascar evolved from lineages that originated from Cenozoic dispersal events (Yoder & Nowak, 2006), while few groups (e.g., the genus *Takhtajania*) date back to a potential Gondwanan vicariance (Buerki *et al.*, 2013). Also, islands located in southeast Asia showed high paleoendemism, which is due to numerous relict lineages that have survived the last two mass extinctions (Procheş *et al.*, 2015). Consequently, our results reinforce the conservation urgency for islands which are often occupied by both neoendemics and paleoendemics that represent millions of years of unique evolutionary history (Fernández-Palacios *et al.*, 2021).

Tropical mountain regions are well-known centers of taxonomic and phylogenetic plant diversity (Cai *et al.*, 2023). Due to their complex topography and geological and climatic histories, they also hold exceptionally range-restricted species (Dagallier *et al.*, 2020). In our study, mountain regions, especially in tropical regions, emerged not only as centers of PE, but also of both neo- and paleoendemism in particular. On the one hand, mountain regions show remarkable diversification of their plant lineages and therefore foster high neoendemism, acting as "cradles" of biodiversity (Merckx *et al.*, 2015; Xing & Ree, 2017). This diversification is the consequence of multiple mechanisms, including adaptation to diverse niches during long-term orogeny (Xing & Ree, 2017; Antonelli *et al.*, 2018), or divergence resulting from dynamic connectivity between habitats related to climatic fluctuation (Rahbek *et al.*, 2019; Flantua *et al.*, 2020). On the other hand, mountain regions support the persistence of ancient lineages over time, acting as "museums" for paleoendemics (Sandel *et al.*, 2011; Dagallier *et al.*, 2020). This results from steep environmental gradients with diverse microclimates in mountain regions, allowing species to track their climate niche through altitudinal range shifts during climate change periods (Jump *et al.*, 2009).

Our results show how past climate has affected present patterns of neo- and paleoendemism, with climate stability since the Last Glacial Maximum promoting the accumulation of paleoendemics as well as PE in general. Cooler temperatures during glacial periods may have caused range contractions and selective extinctions of range-restricted species, and thus likely removed or reduced their ranges in less stable regions (Jansson, 2003). In contrast, some regions such as islands, coastal or mountain regions have suffered less from past climate change because of the buffering effect of the oceans against climatic changes (Cronk, 1997) and the topographically diverse microclimates in mountain regions (Jump *et al.*, 2009). Also, the high concentration of both neo- and paleoendemism in regions with less climatic changes since the mid-Pliocene warm period emphasizes the vital

role of long-term climatic stability on speciation and persistence of range-restricted species. Different from the climate during the LGM, the mid-Pliocene warm period, however, represented warmer climates compared to today. The positive effect of climatic stability on neo- and paleoendemism is therefore also in line with the general positive effect of energy and water availability on PE of plants. This may be linked to lower extinction risks for range-restricted plants under warm and wet climates by offering favorable environments and sufficient resources for larger populations in smaller areas (Evans *et al.*, 2005). In addition, we found that Mediterranean-climate regions acted as extra-tropical hotspots of plant endemism, especially with high neoendemism. This may be attributed to the recent and rapid speciation in these regions, triggered by the unique climatic regime characterized by high seasonality and summer drought (Thompson *et al.*, 2005; Valente *et al.*, 2010).

Generally, larger regions host more endemics as well as wide-ranged species because of their overall higher plant diversity (Storch *et al.*, 2012; Cai *et al.*, 2023). Here, we observed a negative association between region area and PE when species range sizes were measured as the total area of the occupied regions. Specifically, PE.area peaked on some small islands (e.g., Lord Howe Island) and showed lower values in large mainland regions. However, PE.area of large mainland regions was possibly underestimated because the range sizes of endemics that only occur in small suitable habitats within large regions were overestimated. In contrast, there was a positive association between region area and PE when we measured species range sizes as the count of occupied regions. However, this method ignores the variation of area across regions and disregards that endemics in small regions likely have smaller ranges than endemics in larger regions leading to an underestimation of PE for small regions. Area, therefore, acted as a covariate to control for biases in the estimation of range size in our two metrics of PE and the scale-dependency of endemism (Daru *et al.*, 2020) rather than as an environmental predictor. The largely similar effects of environmental drivers on PE and neo- versus paleoendemism for the two ways in which range size was quantified demonstrates the robustness of our results. Similarly, the results were largely the same with and without unplaced species added to the original phylogeny. Differences that did emerge, however, call for rigorous sensitivity analyses when dealing with species without phylogenetic information (Thornhill *et al.*, 2017; Sandel *et al.*, 2020; Mishler, 2023). For example, the emergence of Madagascar as a super-endemism center when unplaced species were added may result from many species from genera endemic to Madagascar missing from the phylogeny, e.g., *Astiella* (38). Similarly, the lack of an effect of elevational range on neo-endemism when unplaced species were not added suggests that particularly endemic species from recent diversification events may be missing phylogenetic data (Rudbeck *et al.*, 2022).

In conclusion, our study uncovers global patterns of phylogenetic endemism for seed plants and disentangles the complex joint effects of isolation, heterogeneity, climate and long-term climatic stability on phylogenetic endemism. Integration of unprecedented phylogenetic information allowed us to distinguish global centers of neo- and paleoendemism, highlighting tropical mountains, oceanic and large continental islands as well as Mediterranean-climate regions as hotspots of evolutionarily distinct endemic species. These regions have experienced unique climatic and geological histories, which have driven the interplay of important evolutionary and ecological processes of diversity generation and maintenance. Consequently, these regions are of crucial conservation value and need to be protected.

Materials and Methods

Species distribution data

We used regional species composition data for native seed plants from the Global Inventory of Floras and Traits (GIFT version 3.0: http://gift.uni-goettingen.de) (Weigelt *et al.*, 2020) and the World Checklist of Vascular Plants (WCVP, http://wcvp.science.kew.org/) (Govaerts *et al.*, 2021). GIFT contains regional plant inventories from published floras and checklists for ~ 3400 geographic regions worldwide representing islands, protected areas, biogeographical regions (e.g., botanical countries) and political units (e.g., countries, provinces). WCVP is a comprehensive taxonomic compilation of vascular plants and offers distribution information of species in WGSRPD Level-3 units (i.e., 369 botanical countries). We downloaded information for each non-synonym species in WCVP (accessed 18 February 2022) using the function *pow_lookup* in the R package *taxize* (Chamberlain & Szöcs, 2013) and extracted their distribution and biogeographical status across all botanical countries. We then combined all native seed plant occurrences from WCVP with all native seed plant checklists from GIFT available for the same regions. To obtain finer-grain distribution information for some large regions, we replaced botanical countries with smaller regions from GIFT where available (e.g., the individual departments of Bolivia instead of the entire country). We removed the larger regions only when smaller regions were nested within the larger regions and all nested regions completely covered the larger regions for mainland regions, and replaced archipelagos with individual islands if the individual islands made up most of the archipelago. Because all non-hybrid species names in GIFT 3.0 were standardized and validated based on taxonomic information provided by WCVP, we were able to directly combine WCVP and GIFT data. We retained taxonomically unmatched species names because of the low percentage of these species per region (i.e., 99.7% of all species names were taxonomically matched on average across regions). We excluded regions with areas not permanently covered by ice smaller than 10 km^2. All small regions excluded were islands and only a few of them host endemic species (49 endemic species on 112 islands < 10 km^2 in GIFT). The final dataset included 317,985 seed plant species for 912 geographic regions covering all landmass worldwide with varying area sizes ranging from 10 to 3,069,766 km^2 (median: 23,192 km^2), consisting of 597 mainland regions and 315 islands or island groups (SI Appendix, Figure S2.1).

Apomictic taxa

Apomixis is a special case of uniparental reproduction via asexually formed seeds (Majeský *et al.*, 2017). Apomixis is tightly associated with hybridization and polyploidization, and may promote reticulate evolution and the formation of a multitude of novel lineages (Majeský *et al.*, 2017). European brambles (*Rubus* subgen. *Rubus*, Rosaceae), for example, consist mostly of apomictic taxa (only 4 out of 748 accepted species are sexual) owing to speciation via reticulation and apomixis (Sochor *et al.*, 2015). However, taxonomic treatment of these complex groups of apomictic taxa and underlying species concepts are contentious. Additionally, regional floras and checklists differ in the level of detail at which these groups are included and taxonomically resolved. Consequently, the global distribution of apomictic taxa is geographically biased (particularly towards the well-sampled European flora), affecting the assessment of endemism, especially for regions with a high proportion of apomictic taxa. To account for the bias introduced by apomictic taxa, we repeated all analyses including and excluding all the species from 293 genera that contain apomictic species according to the Apomixis Database (http://www.apomixis.uni-goettingen.de) (Hojsgaard *et al.*, 2014). The distribution dataset excluding species from apomictic genera included 273,838 species and was used for the main analyses. Results including apomictic species can be found in SI Appendix. The Apomixis Database has been constructed only for angiosperms. It has however been shown that apomixis is very rare in gymnosperms (Majeský *et al.*, 2017).

Phylogeny

To measure phylogenetic endemism, we linked the species from the distribution dataset to a large, dated species-level phylogeny of seed plants with 353,185 tips (Smith & Brown, 2018). A total of 212,525 species from the distribution dataset excluding apomictic taxa (77.6 % of the species), and 244,206 species from the dataset including apomictic taxa (76.8%), could be directly matched to the phylogeny. Species not present in the distribution dataset were excluded from the phylogeny (hereafter called matched phylogeny). Different ways to deal with species missing from phylogenies in biogeographic and macroecological analyses exist (Thornhill *et al.*, 2017; Tietje *et al.*, 2022; Mishler, 2023). Furthermore, it has been shown that range-restricted species are significantly less likely to have phylogenetic data (Rudbeck *et al.*, 2022), suggesting that excluding all species missing from the original phylogeny might systematically underestimate PE and neo- and paleoendemism. Therefore, to test whether removing unplaced species from the distribution dataset or adding them into the phylogeny affects patterns of PE and neo- and paleoendemism, we built an additional phylogeny and repeated all analyses for comparison (Tietje *et al.*, 2022). We bound the missing species to their congeners in the original phylogeny by replacing all species of a given genus by a polytomy using the function *congeneric.merge* in the R package *pez* (Pearse *et al.*, 2015). We then excluded species not present in the distribution dataset from the phylogeny (hereafter called merged phylogeny). The merged phylogeny included 267,105 out of 273,838 species (97.5 %) in the dataset excluding apomictic taxa and 311,250 species (97.9%) including apomictic taxa. Adding species as polytomies may introduce additional uncertainties when working with large phylogenies (Qian *et al.*, 2021; Cai *et al.*, 2023). However, phylogenetic metrics based on phylogenies with higher numbers of polytomies have been shown to be highly correlated with metrics based on trees without or with fewer polytomies (Qian & Jin, 2021). We repeated all analyses using both the matched and merged phylogenies. PE derived from the matched phylogeny was highly correlated to PE based on the merged phylogeny (Pearson's r: 0.98 for PE.area and 0.99 for PE.count based on the dataset excluding species from apomictic genera). We, therefore, present results based on the matched phylogeny in the main text if not stated otherwise, and discuss discrepancies between the different approaches critically. Results based on the merged phylogeny and excluding and including apomictic taxa can be found in SI Appendix.

Phylogenetic endemism

To investigate the distribution of seed plant endemism worldwide, we calculated phylogenetic endemism for each region following Rosauer *et al.* (2009), as the sum of branch lengths connecting all species coexisting in a region, based on a phylogeny where each branch length is divided by the global range size of the species that descended from the branch. Because PE depends on reliable range size estimates and its pattern is sensitive to differences in grain size (Daru *et al.*, 2020), we measured the range size of each species and of each branch in two different ways: (i) as the number of regions a species occurs in (PE.count) and (ii) as the total area (not permanently covered by ice) of these regions (PE.area). PE.count overestimates PE particularly for large regions, since it disregards that the ranges of species endemic to small regions are likely smaller than the ranges of species endemic to larger regions. In contrast, PE.area accounts for the varied areas of regions in our dataset, but likely underestimates PE for large regions because their areas may be larger than the actual ranges of the species occurring inside. Despite the potential biases of both methods, the actual ranges and hence endemism fall within the range that is estimated by the two methods (SI Appendix, Figure S2.13; see Discussion for more details). We, therefore, repeated all analyses based on both metrics, considered those results particularly robust that emerged for both metrics and discussed differences critically.

Neo- and paleoendemism

We used the categorical analysis of neo- and paleoendemism (CANAPE) (Mishler *et al.*, 2014) to distinguish between centers of neoendemism and paleoendemism. CANAPE is based on the assessment of the statistical significance of PE and relative phylogenetic endemism (RPE). RPE is the ratio of PE measured on the actual phylogenetic tree divided by PE measured on a comparison tree that retains the actual tree topology but with all branches having the same length (Mishler *et al.*, 2014). Therefore, RPE allowed us to examine the degree to which branch lengths and hence clade ages matter for the observed patterns of PE. We carried out the CANAPE analysis for PE.count and PE.area, respectively. To test the significance of the metrics, we ran 1000 null model randomizations. In the null models, species occurrences across regions were randomly reassigned without replacement, keeping the species number in each region and the total number of regions occupied by each species constant (Mishler *et al.*, 2014). Distributions of null model values for each region were then used for non-parametric tests for the significance of the observed values of the tested metrics and for calculating the standardized effect size of RPE. If the observed value of the tested metric fell into the highest 2.5% or lowest 2.5% of the null distribution for a region, it was identified as statistically significantly high or low, respectively (two-tailed test, $\alpha=0.05$). This randomization-based significance test was carried out for PE measured on the actual tree (numerator of RPE), PE measured on the comparison tree with equal branch lengths (denominator of RPE), and RPE.

We then followed a two-step process to distinguish different centers of endemism following Mishler *et al.* (2014). First, we identified regions with significantly high PE by testing whether PE measured on the actual tree (numerator of RPE), PE measured on the comparison tree with equal branch length (denominator of RPE), or both were significantly higher than expected (observed value > 95% of the randomization values; one-tailed test, $\alpha= 0.05$). Second, we divided regions with significantly high PE into four categories of centers of endemism (paleo-, neo-, mixed, and super-endemism). If the RPE of a region was significantly high or low (two-tailed test, $\alpha= 0.05$), the region was defined as a center of paleoendemism or neoendemism, respectively. If the RPE was not significantly high or low, but both the numerator and denominator of RPE were significantly high ($\alpha= 0.05$), the region was defined as a center of mixed endemism. If a mixed endemism region had both a significantly high numerator and denominator of RPE at the $\alpha= 0.01$ level, the region was identified as a center of super-endemism.

We also calculated the standardized effect size of relative phylogenetic endemism (RPE.ses) based on the null distributions of RPE obtained from the null model. RPE.ses was calculated as the difference between the observed values and the mean of the null distribution divided by the standard deviation of the null distribution. In contrast to the non-parametric test in CANAPE, RPE.ses quantifies, for each region, the degree to which disproportionately young or old lineages (i.e., shorter or longer branches) are spatially restricted. When only considering regions with significantly high PE, lower values of RPE.ses represent more young or younger lineages that are spatially restricted, while higher values represent more old or older lineages than expected by chance. Thus, this metric offers an opportunity to model and explore the relationship between the historical drivers and the spatial patterns of neo- versus paleoendemism as a continuous variable. However, it should be noted that a region with both high neo- and paleoendemism may show a value close to zero.

Predictors of phylogenetic endemism

We hypothesized that phylogenetic endemism is shaped by biogeographical and evolutionary processes that promote the origin and maintenance of range-restricted lineages. We, therefore, identified a set of candidate predictor variables representing these processes and classified them into five categories: isolation, environmental heterogeneity, energy and water availability, climatic seasonality and long-term climatic stability (SI Appendix, Table S2.1). These factors have been shown or hypothesized to contribute to geographic patterns of plant

endemism in previous studies (Jansson, 2003; Kier *et al.*, 2009; Sandel *et al.*, 2011, 2020). We measured geographic isolation as the sum of the proportions of landmass area in the surrounding of the target regions within buffer distances of 100 km, 1000 km, and 10,000 km (Weigelt & Kreft, 2013). Its value is lowest for remote islands and highest for regions located in the centers of large continents. We considered number of soil types (Hengl *et al.*, 2017) and elevational range (Danielson & Gesch, 2011) for each region as proxies for environmental heterogeneity. We also included five ecologically relevant climatic variables representing the main aspects of climate hypothesized to be important for plant endemism, namely mean annual temperature, mean annual precipitation, length of the growing season (i.e., number of days with temperatures exceeding a threshold of 0.9 °C, without snow cover, and with sufficient soil water), temperature seasonality (i.e., standard deviation of mean monthly temperature × 100) and precipitation seasonality (i.e., coefficient of variation in monthly precipitation) (Karger *et al.*, 2017). Climatic variables were extracted as mean values per region from the input raster layers.

To determine the contribution of long-term climatic stability to PE, we calculated temperature stability since the LGM (21 Ka), velocity of temperature change since the LGM, and temperature anomaly since the mid-Pliocene warm period (~3.264 – 3.025 Ma). The LGM and the mid-Pliocene warm period represent cooler and warmer climates compared to the current climates, respectively. Temperature stability since the LGM was calculated using the *climateStablity* R package (Owens & Guralnick, 2019). It calculates temperature differences between 1000-year time slices expressed as standard deviation and averages the results across all time slices. Temperature stability is then calculated as the inverse of the mean standard deviation rescaled to [0,1] (Owens & Guralnick, 2019). In addition, we calculated the velocity of temperature change since the LGM as the ratio between temporal change and contemporary spatial change in temperature, representing the speed with which a species would have to move its range to track analogous climatic conditions (Sandel *et al.*, 2011). Temperature anomaly since the mid-Pliocene warm period was calculated as the absolute difference in mean annual temperature between the mid-Pliocene warm period and present-day (Brown *et al.*, 2018).

Predictors of neo- and paleoendemism

Neo- and paleoendemism are hypothesized to be driven primarily by the geological history of a region and by past climate change or stability (SI Appendix, Table S2.2). We, therefore, included the geographic type of each region (distinguishing between mainland regions and continental shelf islands, continental fragments and oceanic islands) instead of surrounding landmass proportion, and elevational range (to distinguish between mountainous and non-mountainous regions) to represent geological history (Weigelt *et al.*, 2020). We removed three islands with heterogeneous geological origin from further analyses on neo- and paleoendemism. To test for the impacts of past climate change on neo- and paleoendemism, we included the variables of long-term climatic stability introduced above.

Models of phylogenetic endemism

To assess the relationships between PE and environmental predictor variables, we fitted linear models with PE as a response variable. Beyond all predictor variables hypothesized to be important to PE (SI Appendix, Table S2.1), we included area size (km^2) to control for the over- and underestimation of PE in large regions for PE.count and PE.area, respectively. We excluded regions with incomplete coverage of predictor variables, leading to a dataset including 818 regions (incl. 236 islands and 582 mainland regions; see SI Appendix, Figure S2.14 for correlations between predictors). PE was log_{10}-transformed before modeling. Some predictor variables (i.e., region area, elevational range, number of soil types, mean annual precipitation, temperature seasonality, precipitation

seasonality, velocity in temperature change since the LGM, temperature stability since the LGM, and temperature anomaly since the mid-Pliocene warm period) were also $\log_{10}$-transformed to reduce the skewness of their distributions. All continuous predictor variables were then standardized to zero mean and unit variance to aid model fitting and make their parameter estimates comparable. To test whether the effects of environmental predictors on PE differ for isolated islands compared to less isolated mainland regions, we included the interaction between each predictor and surrounding landmass proportion. To test if including surrounding landmass proportion correctly encapsulated the effect of insularity, we updated the model by replacing surrounding landmass proportion with a categorical variable indicating whether a region is an island. Since these models performed worse than models including surrounding landmass proportion (SI Appendix, Table S2.9), we retained surrounding landmass proportion for all further analyses. We visualized the change in the coefficient of one variable in the interactions in dependence on the value of the other variable included using the function *interplot* in the R package *interplot* (Solt & Hu, 2015).

Species distributions, environmental predictors and model residuals are often spatially autocorrelated, which may lead to biased parameter estimates and the violation of statistical assumptions (Dormann *et al.*, 2007). As spatial autocorrelation was detected in the model residuals (SI Appendix, Figure S2.15a, b), we included a spatial autocovariate that represents the spatial autocorrelation in the residuals of non-spatial models (residual autocovariate models, RAC) (Crase *et al.*, 2012). This autocovariate term was implemented as a spatial weight matrix of non-spatial model residuals based on an optimized neighborhood structure. Because most of our regions are political units with varying geometry and size, we used a sphere of influence to identify neighbors for each region (Lim *et al.*, 2020). The sphere of influence for each focal region was defined as a circle around the centroid of a focal region within a radius equal to the distance to the centroid of the nearest neighboring region. When the sphere of influence of two regions overlapped, the two regions were considered neighbors. Overall, the RAC models successfully removed spatial autocorrelation from model residuals (SI Appendix, Figure S2.15c).

Statistical analyses of neo- and paleoendemism

To explore the potential drivers of spatial patterns of neo- and paleoendemism, we fitted ordinary linear models to explain the variation in RPE.ses only for regions that showed significantly high PE based on CANAPE (for CANAPE categories of each region, see doi: 10.6084/m9.figshare.21909822; SI Appendix, Figure S2.8). We removed velocity of temperature change since the LGM and temperature anomaly since the mid-Pliocene warm period because of their low explanatory power for RPE.ses based on AIC values. Predictors retained for modeling contained all three aspects (i.e., islands, mountains and past climate change) hypothesized to affect neo- and paleoendemism. Likewise, we fitted spatial models by including a spatial autocovariate to remove spatial autocorrelation present in the residuals of the non-spatial models (SI Appendix, Figure S2.16).

In addition, we compared the distribution of environmental factors for all regions (912 regions) across all CANAPE categories (i.e., neo, paleo, mixed, super-endemic and non-significant; SI Appendix, Figures S2.11 and S2.12). Because the environmental factors were not normally distributed for each category separately, we used a non-parametric Kruskal–Wallis test followed by Wilcoxon pairwise comparisons (two-tailed tests with Holm's correction) to identify which categories were different from each other (Dagallier *et al.*, 2020). We repeated all modeling procedures for two PE metrics (PE.area and PE.count) and the datasets with and without unplaced species added to the original phylogeny, and excluding and including apomictic taxa.

Chapter 3 Environmental filtering, not dispersal history, drives global patterns of phylogenetic turnover in seed plants at deep phylogenetic timescales

Lirong Cai, Holger Kreft, Pierre Denelle, Amanda Taylor, Dylan Craven, Wayne Dawson, Franz Essl, Mark van Kleunen, Jan Pergl, Petr Pyšek, Marten Winter, Francisco J. Cabezas, Viktoria Wagner, Pieter B. Pelser, Jan Wieringa, Patrick Weigelt

Abstract

Phylogenetic beta diversity quantifies dissimilarities in the evolutionary relatedness among assemblages and is important for understanding the underlying mechanisms structuring biodiversity. Environmental filtering and dispersal history are two main processes that limit plant distributions and determine biogeographical patterns, and their relative importance might vary across evolutionary timescales. Here, we examined the effects of environmental dissimilarity and past and current geographical distances on the turnover component of phylogenetic and species beta diversity of seed plants globally and across phylogenetic timescales. To calculate species and phylogenetic turnover, we used a global dataset of regional plant inventories across 675 geographic regions comprising ~ 320,000 species matched to a mega-phylogeny of seed plants. To account for past and present dispersal opportunities, we used historical reconstructions of tectonic plate arrangements and calculated geographical linear distances and cost distances accounting for the cost of crossing water bodies, mountains, or unsuitable climates. Geographical distances and environmental dissimilarity together explained species turnover better (up to 86.4% of deviance explained) than phylogenetic turnover (up to 65.7%). The effect of geographical distances diminished when moving back in evolutionary time, while environmental dissimilarity always showed strong effects on phylogenetic turnover. Past cost distances across barriers explained a comparatively low amount of variation across all timescales, peaking slightly at intermediate phylogenetic timescales (20 – 50 Ma BP). Our results suggest that old lineages had enough time to disperse widely, but the fingerprints of environmental limitations on spatial patterns of plant diversity persist, providing insights into the biogeographic and evolutionary processes underlying global biodiversity patterns.

Introduction

Biodiversity varies across the globe (Pianka, 1966; Cai *et al.*, 2023), and understanding how this variation is generated is of particular interest to macroecologists and biogeographers (Mittelbach *et al.*, 2007). Beta diversity, which quantifies variation in species composition among sites, provides a key to exploring the drivers underlying biodiversity patterns (Tuomisto, 2010; Anderson *et al.*, 2011). While classical measures of beta diversity focus on species composition, adding phylogenetic information into analyses of beta diversity can quantify how evolutionary relationships of members from different assemblages change (i.e., phylogenetic beta diversity) (Graham & Fine, 2008), and helps to address how these drivers changed across different phylogenetic timescales (Mazel *et al.*, 2017). Therefore, exploring large-scale phylogenetic beta diversity using an approach that incorporates compositional dissimilarities along the phylogenetic timescale could enhance our understanding of

contemporary ecological and past evolutionary mechanisms structuring biodiversity (Mazel *et al.*, 2017; Liu *et al.*, 2023).

Two major processes are hypothesized to restrict the distribution of species: environmental filtering and dispersal history (Tuomisto, 2010). Niche-assembly theory states that environmental factors determine the establishment or persistence of species in a particular area (Whittaker, 1956; Kraft *et al.*, 2015). As a result, regions with more similar environments are expected to have a higher similarity in species composition as well as in phylogenetic composition. However, different species and phylogenetic compositions can be observed between regions with similar bioclimates (Holt *et al.*, 2013), which may be explained by dispersal history of lineages (Rodrigues & Diniz-Filho, 2017; Daru *et al.*, 2017). If a lineage persists and diversifies in an isolated region, its descendants might still be restricted to their areas of origin owing to the failure of individuals to reach environmentally suitable areas elsewhere (Linder *et al.*, 2013). Consequently, this region may share few species with others. Similarities of phylogenetic lineages may also be low, but when ancient lineages are shared between regions, phylogenetic beta diversity possibly detects relatively high assemblage similarity (Graham & Fine, 2008). Such limited dispersal has frequently been addressed by quantifying geographical linear distances between two focal regions (here, we focus on characteristics of the external environment instead of characteristics of an organism related to dispersal) (Nekola & White, 1999). This measure, however, ignores that rates of dispersal of a taxon might be constrained by landscape attributes, such as the presence of barriers like water bodies, mountains or deserts, or other unsuitable environments (Jürgens, 1997; Garcillán & Ezcurra, 2003). These dispersal barriers can promote geographic isolation, leading to high beta diversity over relatively short distances. Importantly, environmental filtering and dispersal history are not mutually exclusive, and may jointly shape variation in species and phylogenetic composition across space (Kubota *et al.*, 2011; Eiserhardt *et al.*, 2013; König *et al.*, 2017). However, the relative contribution of these two processes to phylogenetic beta diversity is largely unstudied and might vary among taxonomic groups (Eiserhardt *et al.*, 2013; König *et al.*, 2017).

The relative importance of environmental filtering and dispersal history on beta diversity is thought to vary among ancient and more recently diverged lineages (Duarte *et al.*, 2014; Mazel *et al.*, 2017). One hypothesis postulates that ancient lineages originate from adaptations to past environmental conditions and have more time to disperse to other environmentally suitable habitats (Paul *et al.*, 2009), leading to a prediction that current dissimilarities of ancient lineages may be primarily driven by environmental factors. Then, within each regime of environments, speciation coupled with phylogenetic niche conservatism may have prompted the formation of recently evolved lineages. These recently evolved lineages have less time to disperse and might be largely restricted to ancestral areas. Consequently, we expect that dissimilarities of recently diverged lineages may be largely shaped by dispersal limitation. Moreover, dispersal limitation is not stable along evolutionary time due to past geological events (Ficetola *et al.*, 2017). Past plate tectonics, for example, force the movements of continents, possibly altering dispersal routes and affecting dissimilarities of ancient lineages bewteen regions (Condamine *et al.*, 2012). While some movements of continents may create barriers such as oceans and mountains, other movements may promote dispersal among regions. However, it is unknown whether there is still an imprint of these past settings in present-day phylogenetic compositions, and at which depths along the phylogenetic timescale their effects can be perceived. Incorporating compositional dissimilarities along phylogenetic timescales, i.e., separating more recent and ancient lineages in a phylogeny (Mazel *et al.*, 2017), could allow us to distinguish the varied effects of environmental filtering and dispersal history on the different evolutionary-age phylogenetic composition of regional assemblages.

Here, we examine the effects of environmental filtering and dispersal history on phylogenetic turnover, i.e., a component of phylogenetic beta diversity which is insensitive to phylogenetic diversity (Leprieur *et al.*, 2012), at different phylogenetic timescales (i.e., for lineages of different evolutionary age) for seed plants. Our study integrates a global dataset of regional plant inventories (Weigelt *et al.*, 2020; Govaerts *et al.*, 2021) across 675 geographic regions comprising ~ 320,000 species (SI Appendix, Figure S3.1) with a comprehensive mega-phylogeny (Smith & Brown, 2018). Specifically, our aims are to (i) reveal global geographic patterns of phylogenetic turnover for seed plants; (ii) test for differences in the extent to which environmental filtering versus dispersal history, accounting for past tectonic plate arrangements and dispersal barriers, shapes species and phylogenetic turnover; (iii) uncover the current distribution of phylogenetic turnover at different phylogenetic timescales; and (iv) assess how the effects of environmental filtering and dispersal history on phylogenetic turnover vary along the phylogenetic timescale.

Materials and Methods

Species distribution data

We used regional species composition data for native seed plants by merging the Global Inventory of Floras and Traits (GIFT version 3.0: https://gift.uni-goettingen.de) (Weigelt *et al.*, 2020) and the World Checklist of Vascular Plants (WCVP, https://wcvp.science.kew.org/) (Govaerts *et al.*, 2021). Both databases offer comprehensive distribution information of species from a variety of resources, including published floras and checklists. While GIFT contains regional plant inventories for ~3,400 geographic regions globally representing islands, protected areas, biogeographical regions and political units (e.g. countries or states), WCVP provides distribution information of plant species across WGSRPD Level-3 units (i.e. 369 botanical countries). We downloaded distribution data for each non-synonym species in WCVP (accessed 18 February 2022) using the function *pow_lookup* in the *taxize* package in R (Chamberlain & Szöcs, 2013). We then merged records of all native seed plants in WCVP with native seed plant checklists in GIFT available for the same regions. To obtain fine-resolution distribution information, we replaced botanical countries with smaller regions from GIFT where available (e.g. the individual departments of Bolivia instead of the entire country). Specifically, we removed the larger regions only when smaller regions were nested within and completely covered the larger regions for mainland regions, and replaced archipelagos by individual islands if most of the archipelago represented by individual island checklists. Due to the taxonomic standardization and validation of all non-hybrid species names in GIFT 3.0 using taxonomic information from WCVP, we can directly merge data from WCVP and GIFT. We excluded regions < 1000 km^2 and regions with fewer than 50 species to limit the ranges of area and species richness and allow for shared species across all region pairs. The final dataset included 316,348 seed plant species for 675 geographic regions with a nearly complete coverage of the globe. These regions had varying area sizes from 1,010 to 3,069,766 km^2 (median: 64,964 km^2) and consisted of 587 mainland regions and 88 islands or island groups (SI Appendix, Figure S3.1).

Apomictic taxa

Apomixis is a mode of reproduction via asexual seed formation without syngamy of female and male gametes; it is typically linked to both hybridization and polyploidization, and may enhance reticulate evolution and lead to the formation of many novel lineages (Majeský *et al.*, 2017). However, taxonomic treatment of apomictic taxa, particularly the underlying species concept, is contentious. Moreover, the level of detail at which these taxa are included and taxonomically resolved varies among regional floras and checklists. This may result in geographical

biases in the global distribution of apomictic taxa (particularly towards the taxonomically well studied floras of Europe), influencing the estimates of turnover, especially for regions with a high proportion of apomictic taxa (SI Appendix, Figures S3.2, S3.3 and S3.4). For example, if apomictic taxa are recognized as species, half of the species of the British flora would consist of such taxa (Richards, 2003). To account for the bias introduced by apomictic taxa, we repeated all analyses for the dataset including and excluding all the species from 293 genera that contain apomictic species based on the Apomixis Database (Hojsgaard et al., 2014). We used the distribution dataset excluding apomictic species (272,394 species) for the main analyses. Although the Apomixis Database only included occurrence of apomixis in angiosperms, it has been shown that apomixis is very rare in gymnosperms (Majeský et al., 2017).

Phylogeny

To calculate phylogenetic turnover, we linked species from the distribution dataset to a dated species-level mega-phylogeny of seed plants comprising 353,185 terminal taxa (Smith & Brown, 2018). A total of 211,418 species from the distribution dataset (77.6 %) was included in the original phylogeny. Remaining species were conservatively added to their congeners in the phylogeny by replacing all species of the genus with a polytomy, using the function *congeneric.merge* in the R package *pez* (Pearse et al., 2015). We then removed species not present in the distribution dataset from the phylogeny (hereafter called merged phylogeny), leading to 265,693 out of 272,394 species in the distribution dataset (97.5 %) included in the merged phylogeny. Although additional phylogenetic uncertainties may arise from adding polytomies, it has been showed that phylogenetic metrics based on phylogenetic trees with higher numbers of polytomies correlated strongly with metrics based on trees with no or fewer polytomies (Qian & Jin, 2021). As a sensitivity analysis to test whether adding species and replacing their genera with polytomies affects patterns of phylogenetic turnover, we calculated phylogenetic turnover based on the phylogeny only including 211,418 species that was contained in the original phylogeny and without unplaced species added to the genus (hereafter called matched phylogeny). We found that phylogenetic turnover calculated from the merged and matched phylogenies was highly correlated (Pearson's r = 0.99), suggesting that results based on the merged phylogeny are robust. We therefore used the merged phylogeny for all analyses presented in the main text. The same procedure was carried out for the dataset including apomictic species (Pearson's r = 0.99).

Beta diversity

Beta diversity, i.e., variation in species composition between assemblages, consists of two different components: turnover involving species replacement between regions, and nestedness resulting from differences in species numbers (Baselga, 2010). Likewise, phylogenetic beta diversity, i.e. evolutionary dissimilarity between assemblages, can be decomposed into the two separate components accounting for phylogenetic turnover and differences in phylogenetic diversity (Leprieur et al., 2012). Here, we used the Simpson dissimilarity index to measure both species and phylogenetic turnover between all pairs of regions (Baselga, 2010; Leprieur et al., 2012). This index is insensitive to differences in species richness and phylogenetic diversity and thus allows us to compare regions with varying area size, which differ strongly in species richness (Cai et al., 2023). The index is defined as: $\beta_{sim}= min(b,c)/[min(b,c) + a]$, where a is the number of species shared by two regions and b and c is the number of species unique to the first and second region respectively, when species turnover is concerned. When applied to phylogenetic turnover, species numbers are replaced with branch lengths in the phylogeny (Leprieur et al., 2012).

Decomposition of beta diversity through time

To disentangle the relative influence of past and present factors on beta diversity at different phylogenetic timescales, we used the beta diversity through time framework (BDTT) (Mazel *et al.*, 2017) to compute phylogenetic turnover between regions at different timescales along the phylogeny. From the tips to the root, the phylogenetic tree was repeatedly cut into 10 Ma steps from present back to 100 Ma BP by collapsing all nodes younger than the threshold into ancestral branches. For each cutoff, the geographic distribution of new clades within the pruned phylogenetic tree was calculated as the union of the distributions of their descendent tips. The resulting clade × region matrices could then be used to calculate pairwise phylogenetic turnover for each cutoff. The BDTT offered a profile of beta diversity through time that allowed us to assess the contribution of deep and shallow branches to beta diversity patterns as well as their past and present drivers through time (Mazel *et al.*, 2017).

Predictors of species and phylogenetic turnover

To test the effects of environmental filtering and dispersal history on compositional dissimilarities between regions, we identified a set of predictor variables representing environmental dissimilarity and geographical distances, respectively. To capture major ecologically relevant axes of the environmental space, we considered climate, edaphic conditions and topography, which are hypothesized to be the main environmental drivers of plant composition (Eiserhardt *et al.*, 2013; König *et al.*, 2017). We included five ecologically relevant climatic variables representing the main aspects of climate hypothesized to be important for plant distributions (Cai *et al.*, 2023), namely mean annual temperature, mean annual precipitation, temperature seasonality (i.e., standard deviation of mean monthly temperature × 100), precipitation seasonality (i.e. coefficient of variation in monthly precipitation) (Karger *et al.*, 2017) and aridity (i.e., the ratio of precipitation to potential evapotranspiration) (Zomer *et al.*, 2022). We also included cation-exchange capacity (cmol +/kg) for each region as a proxy for soil fertility (Hengl *et al.*, 2017) and elevational range for topographical heterogeneity (Danielson & Gesch, 2011; Stein *et al.*, 2014). All climatic and edaphic variables were extracted as mean values per region from the input raster layers.

To quantify the effects of dispersal history on turnover, we considered current geographical linear distances, past geographical linear distances derived from plate tectonic models, as well as physical and ecological dispersal barriers (e.g., oceans, mountains and unsuitable climate). In order to calculate distances of each factor between pairs of regions that differ in area sizes, we used the *dggridR* R package (Barnes & Sahr, 2017) to produce a grid of equal-area and equidistant hexagons (grid area: 23,323 km^2) across the Earth's surface and calculated distances between each pair of grid cells. We then measured distances between all pairs of regions as the minimum distance between the grid cells that were overlapping with the two focal regions.

To quantify the effects of the past position of major landmasses (continents) on regional plant species composition, we used the GPLATES software (Boyden *et al.*, 2011; Williams *et al.*, 2012) with global plate tectonic models (Müller *et al.*, 2016; Matthews *et al.*, 2016) to obtain the past geographic coordinates of each grid cells' mass centroid in 10 Ma time steps between the present and 100 Ma BP. We calculated past and current geographical linear distances between each pair of grid cells as the shortest distance on an ellipsoid (here the WGS 84 reference ellipsoid).

To account for physical and ecological dispersal barriers when calculating pairwise geographical distances, we measured the least cost distance between regions, i.e., the minimum cost of moving from one region to another across cost surfaces, using network analysis. Network analysis can find routes that minimize costs between entities in a network. A network consists of a set of connected vertices (here representing our hexagon grid cells) and

edges representing the relationship between the entities (McNulty, 2022). Edges can be either unweighted or weighted (here weighted to reflect geographical linear distances combined with costs for crossing a given surface). A path between two vertices is defined as a series of edges from the first vertex to the second, and therefore the length of the path between two vertices is the sum of the weights of the edges traversed in the path (McNulty, 2022). Here, we created a network for the equal-area and equidistant hexagon grids using the *igraph* R package (Csardi & Nepusz, 2006). All grid cells were connected to their five direct neighbors by edges. We calculated the weights of edges as the geographical linear distances between two connected vertices multiplied by the mean of the values extracted from a given barrier cost surface for the two vertices. The barrier cost of a grid cell or vertex was respectively defined as whether a region is covered by water (1: water; 0: not water), mean elevation, annual mean temperature and aridity index to weigh in costs of crossing water bodies, mountains, cold and dry deserts, respectively. To calculate a barrier cost surface, we scaled and transformed mean elevation, annual mean temperature and aridity index to a range of [0,1] where values closer to 1 indicate higher costs for moving from one cell to the other (SI Appendix, Figure S3.5). To consider all barriers together, we created an additional barrier cost surface as the maximum values among whether a region is covered by water, mean elevation, annual mean temperature, and aridity index. Then, we calculated the least cost distance between each pair of grid cells as the length of the shortest path between pairwise grids for each barrier cost surface. The least cost distance between grid cells was then aggregated for all pairs of geographic regions as the minimum distance between grid cells overlapping with the first and second region, respectively. Beyond current barriers, we calculated the past least cost distance accounting for dispersal barriers (i.e., considering both mountains and water, and only accounting for water) for each 10 Ma from present back to 100 Ma BP based on paleo-estimates of mean elevation and whether a region was covered by water (Scotese & Wright, 2018).

Additionally, we calculated the least cost distance across a surface of climate dissimilarities between the focal regions and all other regions to quantify dispersal limitation due to climatically unsuitable habitat to cross. We characterized the contemporary climate of each grid cell using the first five axes of a principal component analysis (PCA) applied to 19 bioclimatic variables extracted from CHELSA (Karger *et al.*, 2017). Before the PCA, eight bioclimatic variables related to precipitation were log-transformed to meet normality assumptions. The five axes represented 93.5% of the total variability of the bioclimatic variables. Costs were then defined individually for each grid cell as the Euclidean distance of the five climatic axes between each grid cell and the focal grid cell divided by the maximum value so that the range was linearly scaled to [0,1] (SI Appendix, Figure S3.6). We then calculated the length of the shortest path from the focal grid cell to all other grid cells across the climate dissimilarity cost surface. We repeated the calculation for each grid cell and cost surface, leading to a grid cell × grid cell matrix with values of each row representing the length of the shortest path of one focal grid cell to all other grid cells across its individual climate dissimilarity surface. The distance between two grid cells was then calculated as the mean between the two distances obtained when considering the one or the other grid cell as the focal cell. We then again extracted the least cost distance across climate dissimilarity surface between pairwise regions from the distance of grid cells.

Beta diversity analysis

To compare and map the contribution of each region to the global variation in plant composition, we calculated the local contribution to beta diversity (LCBD) for species and phylogenetic turnover, which quantifies the degree of uniqueness of the focal region in terms of species and phylogenetic composition (Legendre & De Cáceres, 2013). Large LCBD values indicate that regions have strongly different species and phylogenetic compositions.

42

We applied generalized dissimilarity modeling (GDM) to explore how species and phylogenetic turnover between regions is driven by dispersal history and environmental filtering (Ferrier *et al.*, 2007). GDM is a powerful method for analyzing and predicting spatial patterns of beta diversity between pairs of regions with respect to explanatory variables. A key assumption of GDM is that dissimilarity can only increase as two regions become more different in terms of the values of a predictor variable (Mokany *et al.*, 2022). The approach is specifically designed to accommodate (1) the curvilinear relationship between compositional turnover and increasing ecological and environmental distance between regions, and (2) the variation in the rate of compositional turnover at different positions along an environmental gradient (Ferrier *et al.*, 2007). We fitted GDM for species and phylogenetic turnover, respectively, using the R package *gdm* (Fitzpatrick *et al.*, 2022) with the region × region matrices of geographical distances (i.e., past and current geographical linear distances, the past and current least cost distances across barriers and the least cost distance across climate dissimilarity surface), and the untransformed vectors of environmental variables (i.e., mean annual temperature, mean annual precipitation, temperature seasonality, precipitation seasonality, aridity index, cation-exchange capacity and elevational range) as predictors. To capture the impact of current and past dispersal limitation on turnover, we initially considered two matrices of past geographical linear distances (i.e. 100 and 150 Ma BP), five matrices of the current least cost distance across barriers (i.e., all barriers, mountains, water, cold and dry deserts) and four matrices of the past least cost distance across barriers (i.e. all barriers and water of 100 and 150 Ma BP). As models with different predictors differed only to a small degree, we retained the current least cost distance across all barriers, geographical linear distances 100 Ma BP and the least cost distance across all barriers 100 Ma BP in the final models. Further, we replaced current climatic variables (i.e. mean annual temperature, mean annual precipitation, temperature seasonality, precipitation seasonality) with climate in the Last Glacial Maximum (LGM; 21 Ka BP) and in the mid-Pliocene warm period (~3.264 – 3.025 Ma BP) to test if past climate better explained turnover than current climate.

Beyond fitting models including all regions (> 1000 km²), we also fitted GDMs for regions with minimum area sizes of 10,000 and 100,000 km^2 to test if predictor effects differ due to the inclusion of small regions that are several orders of magnitude smaller than the largest ones. We used the default of three I-spline basis functions per predictor. Predictor significance tests were performed using matrix permutation as implemented in the function *gdm.varImp*. Following König *et al.* (2017), we applied two different methods to quantify the importance of predictors for present day models and along the phylogenetic timescale. First, we quantified the importance of each individual predictor in explaining dissimilarity using the maximum height of fitted spline functions in the GDM (Fitzpatrick *et al.*, 2013). We scaled the heights so that their sums equal the proportion of deviance explained by the model. Second, we used partitioning of deviance, an approach similar to variance partitioning to estimate the relative effect of environmental filtering (i.e. environmental variables) versus dispersal history (i.e., geographical distances) (Borcard *et al.*, 1992). Specifically, we partitioned the explained variation in turnover into a fraction only explained by environmental filtering, a fraction only explained by dispersal limitation, overlapping (i.e. a fraction explained by either dispersal limitation or environment filtering but not possible to disentangle) and unexplained variation using the function *gdm.partition.deviance*.

To test how the effects of dispersal limitation and environment filtering on phylogenetic turnover vary along the phylogenetic timescale, we used GDMs to model current patterns of phylogenetic turnover at different phylogenetic timescales (i.e., for each ten million years from present back to 100 Ma BP) as functions of geographical (past) and environmental (current) distances. Here, we fitted models only for regions with a minimum area of 10,000 km^2 since models including regions with this area threshold showed a relatively high model

performance for current phylogenetic turnover, and still have a near-complete coverage of the globe. For each time cutoff, a GDM was fitted using the matrices of geographical linear distances and the least cost distance across barriers that were computed for the given time period, and the untransformed vectors of mean annual temperature, mean annual precipitation, temperature seasonality, precipitation seasonality and cation-exchange capacity. Elevational range was removed from these models because it was insignificant in the model for current patterns of turnover. To make models comparable across the phylogenetic timescale, the current least cost distance across barriers also included only mountains and water bodies as barriers like the past least cost distance across barriers. Like above, we used two different methods to quantify the importance of predictors.

Results

The local contribution of each region to global beta diversity (LCBD) varied strongly across the globe for both phylogenetic and species turnover (Figure 3.1). The LCBD of phylogenetic and species turnover both peaked in southwest Australia, parts of Colombia and South Africa as well as some tropical and subtropical islands (e.g., New Caledonia). European regions showed lower species turnover in particular and phylogenetic turnover than parts of North America and Central and East Asia. However, some differences emerged when separately comparing LCBD for phylogenetic and species turnover (Pearson's r = 0.46; SI Appendix, Figure S3.7). For example, high LCBD values for species turnover were observed on some remote islands (e.g., Hawaiian Islands) and large continental islands in Southeast Asia, while these regions did not stand out in terms of phylogenetic turnover. When comparing patterns including and excluding apomictic taxa, notable differences were observed primarily in European regions, which showed greater unique compositions when apomictic species were included, owing to high numbers of narrowly distributed apomictic species in genera such as *Rubus* and *Hieracium* (SI Appendix, Figures S3.2 and S3.3). Nearly identical LCBD patterns were found for the dataset retaining only species that were included in the original phylogeny (SI Appendix, Figure S3.8a).
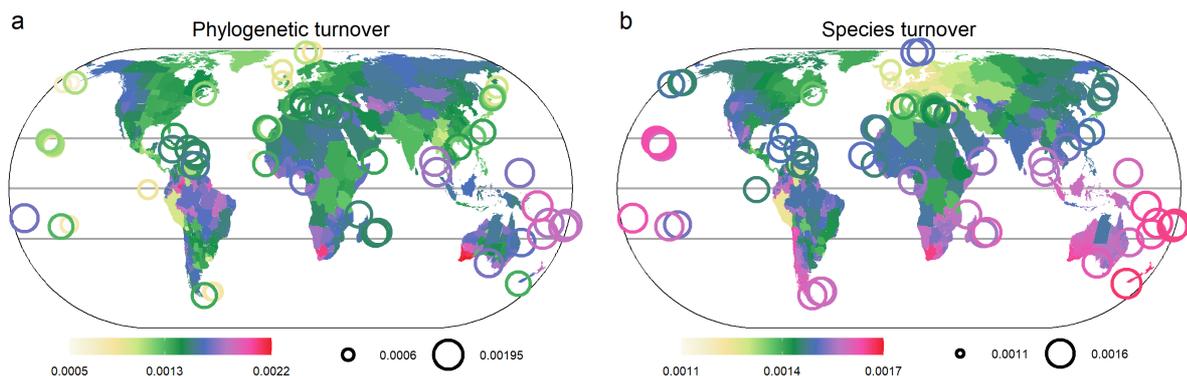
Figure 3.1 Local contribution to global patterns of beta diversity of 675 geographic regions for phylogenetic (a) and species turnover (b) in seed plants. Large values indicate that regions have a strongly different phylogenetic and species composition. Maps use Eckert IV projection.

Based on generalized dissimilarity modeling, environmental dissimilarity and geographical distances (i.e., geographical linear and cost distances) together explained 56.9 – 65.7 % of the variation in current patterns of pairwise phylogenetic turnover and 82.5 – 86.4 % in species turnover, respectively (Figure 3.2; Pearson's r = 0.8 between phylogenetic and species turnover). Models including regions with a minimum area size of 100,000 km² showed the highest explained deviance. According to the deviance-based approach, current and past geographical

44

distances (uniquely explaining 24.4 – 27.5 % in phylogenetic turnover and 42.5 – 44.0% in species turnover) showed a greater effect on both phylogenetic and species turnovers than environmental dissimilarity (17.6 – 20.3% in phylogenetic turnover and 9.2 – 12.8% in species turnover) regardless of the minimum area size of the regions included (Figure 3.2b). The effect of geographical distances was much stronger for species turnover than for phylogenetic turnover, while environmental dissimilarity was slightly more important to phylogenetic turnover than to species turnover. However, the relative importance of environmental dissimilarity was slightly higher when using the curve-based approach (Figure 3.2a) possibly arising from a greater contribution of environmental dissimilarity to the shared fraction (König *et al.*, 2017).
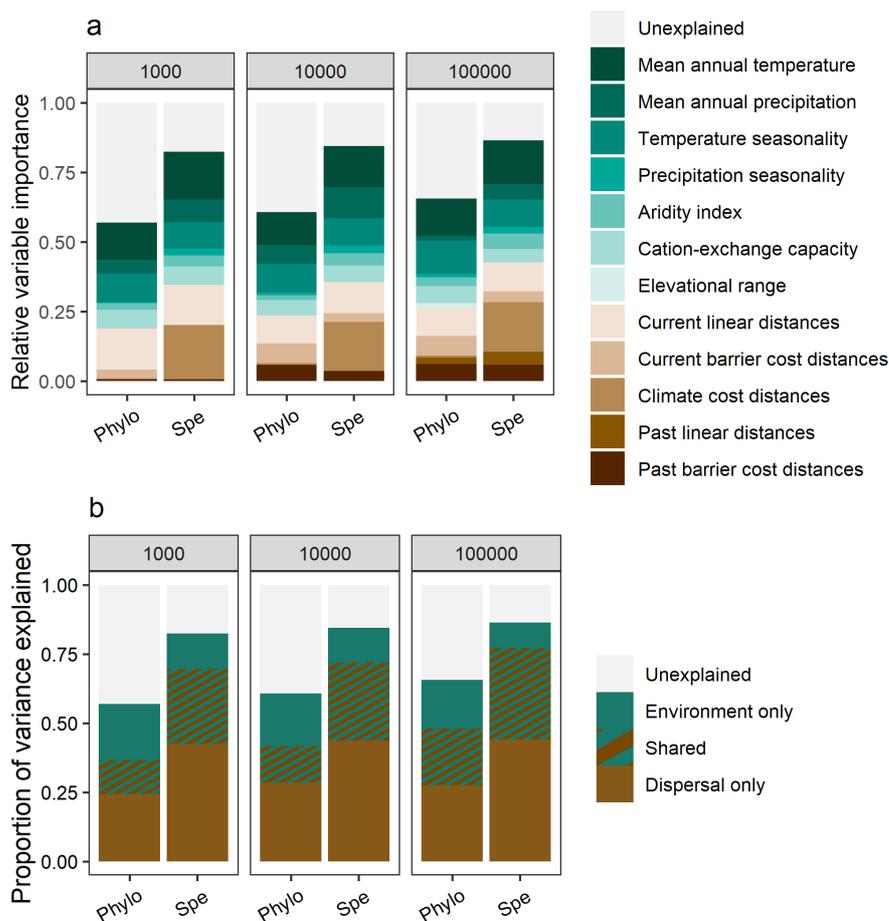
Figure 3.2 Relative importance of predictor variables for plant phylogenetic (Phylo) and species (Spe) turnover in seed plants at the global scale. Variable importance is shown for datasets including regions with a different minimum area size (i.e., 1000; 10,000; 100,000 km^2). In a, relative importance is shown for individual predictor variables based on the height of generalized dissimilarity modeling transformation curves, which is scaled so that their sums equal the proportion of deviance explained by the model; In b, relative importance is shown for thematic groups of predictor variables based on deviance partitioning.

Among the individual predictors, the least cost distance across climate dissimilarity surface most strongly promoted species turnover (17.5 –19.5%), indicating that high species dissimilarity occurred among regions separated by unsuitable climatic conditions. Meanwhile, dissimilarity in mean annual temperature showed the strongest effect on phylogenetic turnover (11.8 and 13.4% for regions > 10,000 and 100,000 km^2, respectively), except for the model including regions > 1000 km^2 where current geographical linear distances were the main driver (14.8%) (Figure 3.2a; SI Appendix, Figure S3.9). Past geographical distances also left small but detectable imprints on current patterns of plant turnover. We found that the past least cost distance across barriers (100 Ma BP) explained 5.8 and 6.3% of the variation in phylogenetic turnover when regions included were > 10,000 and

100,000 km², respectively, and 3.7% and 5.9% of the variation in species turnover. Comparing the effects of past and current climate on plant turnover, the models including climate in the LGM explained less deviance (55.06 – 64.2% in phylogenetic turnover and 80.1 – 85.3% in species turnover; SI Appendix, Figure S3.10 a, c) than the models that included current climate, while the models with climate in the mid-Pliocene warm period showed nearly identical performance (56.6 – 65.5% in phylogenetic turnover and 82.2 – 86.9% in species turnover; SI Appendix, Figure S3.10 b, d).
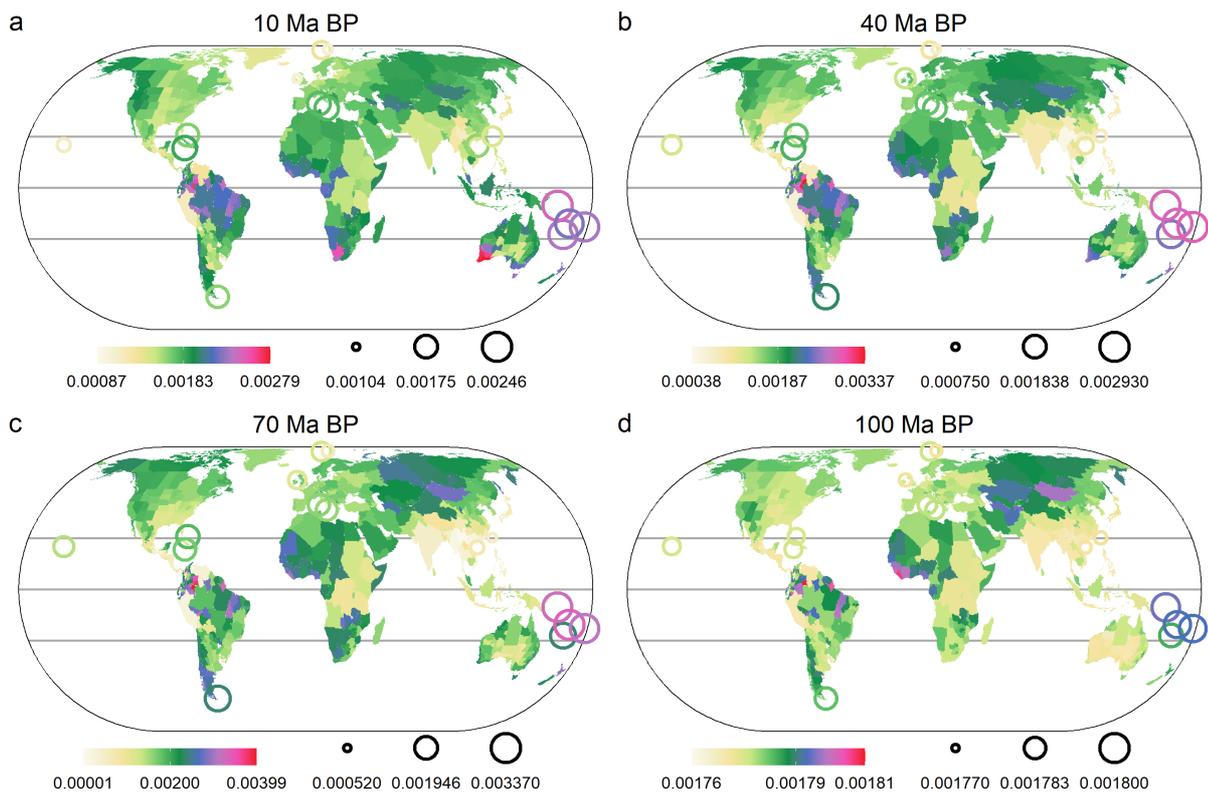
Figure 3.3 Local contribution to global patterns of beta diversity (LCBD) of 562 geographic regions for phylogenetic turnover at different phylogenetic timescales in seed plants. LCBD patterns are calculated for 10 Ma BP (a), 40 Ma BP (b), 70 Ma BP (c), and 100 Ma BP (d), based on the given phylogenies which are obtained by cutting the original phylogeny at a specified time period and collapsing all descendent leaves into ancestral branches. Large LCBD values indicate that regions have a strongly different phylogenetic composition. Regions included are > 10,000 km² and maps use Eckert IV projection.

Along the phylogenetic timescale, the LCBD for phylogenetic turnover, and the contributions of environmental dissimilarity and geographical distances to plant phylogenetic turnover varied considerably. While southwest Australia and parts of South Africa showed lower LCBD at deeper phylogenetic timescales, LCBD always peaked in parts of Colombia and Ecuador (Figure 3.3). The variation in phylogenetic turnover explained by the GDMs decreased from 60.2% (present) to 23.4% (100 Ma BP) along the phylogenetic timescale (Figure 3.4). According to the deviance-based approach, environmental dissimilarity consistently showed strong effects on phylogenetic turnover for both shallow (e.g., uniquely explaining 19.8% for present) and deep phylogenetic timescales (e.g. 17.3% for 100 Ma BP), and peaked at the time slice of 40 Ma BP (26%). In contrast, the effect of geographical distances monotonically decreased from 27.7% to 2.3% when time went back from the present to 100 Ma BP (Figure 3.4b). In other words, phylogenetic turnover of deep phylogenetic time (i.e., time slices before 20 Ma BP) was better explained by current environmental dissimilarity than geographical distances of the same time slices,

while shallow phylogenetic time showed an opposite trend (i.e., 0 and 10 Ma BP; Figure 3.4b). Among predictor variables accounting for dispersal opportunities, the least cost distance across barriers for specific times showed stronger effects on the time slices of 20 – 50 Ma BP (8.6 – 11.6%), while geographical linear distances greatly promoted phylogenetic turnover on shallow phylogenetic time (i.e., 0 and 10 Ma BP; Figure 3.4a). Among environmental predictors, mean annual temperature, temperature seasonality and cation-exchange capacity largely increased phylogenetic turnover regardless of the timescale.
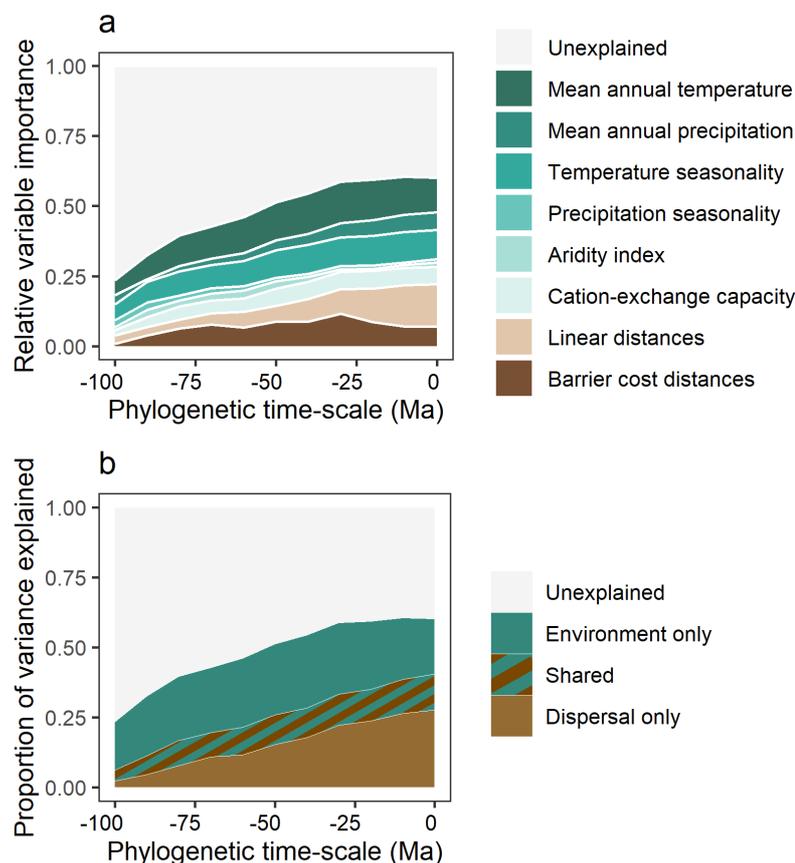
Figure 3.4 Relative importance of predictor variables for phylogenetic turnover along phylogenetic timescales in seed plants at the global scale. In a, relative importance is shown for individual predictor variables based on the height of generalized dissimilarity modeling transformation curves, which is scaled so that their sums equal the proportion of deviance explained by the model; In b, relative importance is shown for thematic groups of predictor variables based on deviance partitioning.

Discussion

Global turnover in seed plants is jointly shaped by environmental filtering and dispersal history. We found that the relative contributions of these two processes vary across species and phylogenetic turnover and at different phylogenetic timescales. While dispersal history had an overall larger effect, it was smaller for phylogenetic than for species turnover and further diminished with phylogenetic time, whereas environmental filtering remained important throughout. This indicates a major role of environmental constraints for the distribution of major seed plant lineages and of dispersal limitation for the younger lineages towards the tips of the phylogeny (species).

Due to dispersal limitation, lineages are restricted geographically, causing a similar distribution of more closely related species and simultaneously increasing evolutionary separation between geographic regions (Linder *et al.*, 2013). This hypothesis is supported by our finding that greater similarity of species and phylogenetic composition occurs between regions that are closer to each other and have fewer barriers to cross. Meanwhile, environmental

filtering determines species distributions (Whittaker, 1956; Currie *et al.*, 2004), and thus has a strong influence on both current patterns of species and phylogenetic turnover as found here. As expected, we found that higher dissimilarity of climatic and edaphic conditions, especially variables related to temperature, strongly promote species and phylogenetic turnover in plants among regions. These findings support the hypothesis that environmental conditions determine the occurrence of species and phylogenetic lineages in a specific area due to inherent differences in environmental tolerance among plants (Wiens & Graham, 2005). Although both showed strong impacts, geographical distances explained current species turnover better than environmental dissimilarity. This disparity was relatively small for phylogenetic turnover, suggesting that dispersal history has a greater effect on regional species composition than environmental filtering at the global scale. This may reflect the high dissimilarity of plant species composition among continents because of prolonged isolation and relatively limited dispersal events (Szövényi *et al.*, 2008; Arakaki *et al.*, 2011).

We found that the relative importance of dispersal history and environmental filtering on current patterns of turnover varied among phylogenetic and species turnover and along the phylogenetic timescale. Whereas environmental dissimilarity explained a high share of both species turnover and phylogenetic turnover in ancient and young lineages, geographical distances showed greater effects on species turnover and on young lineages than ancient lineages. This corroborates our results of the contribution of remote islands to global beta diversity being greater for species turnover than for phylogenetic turnover. Variation in the relative importance of these variables likely indicates the underlying biogeographical and evolutionary processes driving current patterns of diversity (Graham & Fine, 2008). On the one hand, this may stem from "time-for-dispersal" effects, i.e., the time available for dispersal could limit the extent of ranges of taxa (Svenning & Skov, 2007; Paul *et al.*, 2009). Along evolutionary timescales, major ancient lineages of plants have had more time to spread over the world and occupy large parts of their potential range, thereby generating the homogeneity of phylogenetic composition along geographical distances at deep phylogenetic timescales (Hardy *et al.*, 2012). In contrast, recently derived lineages might be still restricted to their area of origins due to the short time available for dispersal, colonization, and range expansion. On the other hand, persistent strong effects of environmental filtering possibly reveal a tendency of lineages to retain characteristics of ancestral environmental tolerances (Burns & Strauss, 2011). Under this scenario, ancient lineages may have originated from adaptations to past environmental conditions. Within a particular environmental regime, allopatric speciation and phylogenetic niche conservatism may have then promoted diversification without major niche shifts and produced recently evolved lineages that share heritable traits through descents, leading to consistently large environmental restrictions for both ancient and recently diverged lineages (Donoghue, 2008). Further, the important role of environmental filtering in structuring turnover of ancient lineages may reflect that major environmental constraints might be phylogenetically conserved in large lineages. For example, palms (Arecaceae), a pantropical plant family, are largely restricted to megathermal climates, possibly due to their lack of physiological mechanisms for frost tolerance (Kissling *et al.*, 2012).

Dispersal limitation is more difficult to detect within continents than among continents and islands, because dispersal barriers such as climate dissimilarity are not always obvious. Current geographical linear distances are most frequently chosen as a proxy of geographical connectivity (Nekola & White, 1999) and also here it showed a strong influence on phylogenetic and species turnover at a global scale. However, this method assumes that individuals spread across a uniformly suitable landscape, which is not always the case. We gained additional insights by including novel distance measures that account for barrier costs such as oceans, mountains, deserts, and climatic gradients, all of which produced high levels of phylogenetic turnover among regions. Meanwhile,

high species turnover was found among regions separated by unsuitable climatic conditions, which could be explained by a lack of potential corridors with suitable climate and consequently constrained species dispersal. These results highlight that complex measures of cost distances across barriers appropriately reflect dispersal barriers and demonstrate that the medium to cross plays an important role in the redistribution of plants. These results also suggest that the success of dispersal for species is not only affected by the direct geographical distances but rather by the ability to establish populations between two regions.

We demonstrate that past plate tectonics have left signatures in the current distribution of plant species and major lineages. Tectonic movements changed geographic proximity of landmass and altered dispersal barriers among regions, including orogenic uplift and expansion or shrinkage of oceans, affecting dispersal history of plant species and, therefore, modifying their distributions (Raven & Axelrod, 1974; Pellissier *et al.*, 2018). While past geographical distances explained little variation in phylogenetic turnover at deeper phylogenetic timescales, we found that past cost distances across barriers, including the emergence of mountains and water bodies, showed relatively strong effects on phylogenetic turnover at intermediate phylogenetic timescales (20 – 50 Ma BP, mainly the Oligocene and Eocene). During this period, continents underwent ongoing divergence and sometimes collided with previously isolated landmasses. The latter process was exemplified by the rapid northward drifting of India and its subsequent collision with Asia, leading to the uplift of the Himalayas (Lomolino *et al.*, 2016), which might have emerged as a key dispersal barrier within continents and thus left an impressive imprint on plant distributions. Additionally, we found that climate in the LGM had less effects on current phylogenetic and species turnover compared to the current climate, possibly derived from postglacial species migration to track climate change (Svenning & Skov, 2007). Such postglacial range expansion has led to large geographic ranges of plant species in Europe (Svenning & Skov, 2007), coupled with high numbers of plant species extinctions due to extreme climatic conditions (Qian & Ricklefs, 2000; Svenning, 2003), possibly leading to the high homogeneity of the European flora found here.

In conclusion, our study reveals that global patterns of regional seed plant composition result from complex dispersal filtering processes related to both past and present dispersal limitations and phylogenetically conserved environmental constraints. Our results show that the relative importance of the two processes on phylogenetic turnover varies along the phylogenetic timescale, i.e., different processes operate at the species versus at the higher phylogenetic level. Using paleogeographical reconstructions and a framework to account for dispersal barriers, we demonstrate the contribution of past plate tectonics to explaining phylogenetic turnover by altering geographical proximity and dispersal barriers among regions. Consequently, our study helps to unfold the complex biogeographical and evolutionary mechanisms that shape global plant biogeography.

General Discussion

Studying biodiversity from a phylogenetic perspective provides new insights into understanding the underlying mechanisms of biodiversity patterns (Donoghue, 2008; Mishler, 2023) and is important for biodiversity conservation (Mace *et al.*, 2003; Gumbs *et al.*, 2023). With newly compiled global datasets of plant distributions, better resolved species-level phylogenies of plants, high-resolution past and present environmental datasets, and the modeling techniques developed, plant diversity can now be studied on a global scale while simultaneously accounting for phylogenetic relationships among species. Specifically, I uncovered global patterns and drivers of three key aspects of plant diversity in terms of evolutionary history in particular: species and phylogenetic richness (Chapter 1), phylogenetic endemism (Chapter 2), and species and phylogenetic turnover (Chapter 3), which together improve both fundamental understanding and knowledge of global biodiversity. In the following sections, I discuss and synthesize the contribution of my thesis and elaborate on challenges and new directions for biodiversity research.

Contribution of this thesis

The thesis contributes to biodiversity research on three different levels: biodiversity patterns, biodiversity mechanisms, and methods of studying biodiversity.

Biodiversity patterns

The thesis provides comprehensive investigations of plant diversity distributions accounting for evolutionary history, which have been so far missing at the global scale. I filled this critical gap in biodiversity knowledge based on regional plant inventory data that have been compiled over a decade (GIFT; Weigelt et al., 2020) combined with a curated global list of all known scientifically described plant species (WCVP; Govaerts et al., 2021), which currently represent the most authoritative and comprehensive distribution datasets of vascular plants with complete global coverage. I provided global maps of species and phylogenetic richness with unprecedented detail and accuracy by combining predictions from multiple modeling techniques, and identified global centers of plant diversity (Chapter 1). Also, I uncovered global patterns of evolutionarily unique and range-restricted lineages (phylogenetic endemism, Chapter 2). I identified regions with significantly unique compositions of species and phylogenetic lineages (species and phylogenetic turnover, Chapter 3). Generally, tropical and subtropical islands consistently stood out for these multiple aspects of plant diversity, such as Madagascar, the Malay Archipelago, and New Caledonia. Meanwhile, mountain regions in the tropics and subtropics were identified as global centers of species and phylogenetic richness and phylogenetic endemism, e.g. the tropical Andes. Furthermore, I identified regions with both high neo- and paleoendemism that act simultaneously as "museums" and "cradles" of biodiversity, including some oceanic and continental fragment islands, Mediterranean-climate regions, and parts of the Irano-Turanian floristic region (Chapter 2). With an explicit consideration of these multiple aspects, the results of my PhD work identify areas with high evolutionary diversity and uniqueness, experiencing important processes related to the origination and persistence of biodiversity, and could be of great importance for biodiversity conservation. Simultaneously, the high-resolution predictive maps of species and phylogenetic richness will offer essential information for future biodiversity science and global environmental impact assessments.

Biodiversity mechanisms

I make empirical contributions to a better understanding of the underlying mechanism that shapes biodiversity. Lineage divergence, persistence, and dispersal are fundamental processes that contribute to why particular Earth areas harbor more lineages than others (Ricklefs, 1987; Mittelbach *et al.*, 2007). Many environmental factors have been hypothesized to determine species' geographic ranges, assemblages of species in a particular area, and variation of diversity across space by changing patterns of these fundamental processes (Condamine *et al.*, 2012). In the thesis, I provided the most comprehensive tests of how this set of past and current environmental factors acts simultaneously on regional species and phylogenetic richness and phylogenetic endemism across the globe for plants (Chapters 1 and 2). Importantly, I also tested the hypotheses of past climate change and geological history on paleo- and neoendemism of plants, providing novel insights into the understanding of processes of lineage divergence and persistence over long-term timescales (Chapter 2). I found that warm and humid climates and environmental heterogeneity strongly promoted both species and phylogenetic richness and phylogenetic endemism, corroborating other large-scale studies that report strong effects of the current climate and environmental heterogeneity on plant diversity (Hawkins *et al.*, 2003; Kreft & Jetz, 2007; Stein *et al.*, 2014; Sandel *et al.*, 2020). Geographic isolation emerged as an important driver of phylogenetic endemism, as expected (Sandel *et al.*, 2020), resulting in high phylogenetic endemism on islands and in mainland regions that are partly surrounded by water bodies like coastal regions or peninsulas.

Notably, the findings of the thesis revealed the critical legacies of past climate change and past geological events in current plant diversity. Earth has experienced strong climatic and geological changes over the last millions of years (Sepulchre *et al.*, 2006; Lomolino *et al.*, 2016). Given the evolution and accumulation of biodiversity under these circumstances, it is hypothesized that current biodiversity is strongly determined by these past dynamics (Hoorn *et al.*, 2010; Svenning *et al.*, 2015; Hagen *et al.*, 2021). The results of the thesis showed that past climate change left a small but detectable impact on current patterns of plant species and phylogenetic richness, while long-term climatic stability significantly promoted the accumulation of evolutionarily unique and narrow-ranged lineages, especially for paleoendemics, as far back as the mid-Pliocene (~3.264 Ma BP). These findings demonstrate the strong and long-lasting effects of past climate change on current plant diversity, and highlight the potential for loss of plant diversity due to future climate change. Also, geological history, including insularity and the presence of mountains, played an important role in plant diversity. On the one hand, I found variations in distributions of neo- and paleoendemism between islands with different geographic origins. The findings could be explained by different evolutionary trajectories of plants determined by geological history of islands, including differences in the timing of colonization, subsequent time for speciation, persistence and accumulation of lineages (Cronk, 1987; Yoder & Nowak, 2006; Losos & Ricklefs, 2009), which together lead to the remarkable plant diversity on large tropical islands as found in Chapter 1. On the other hand, I showed that mountains, especially in tropical regions, were identified as global centers of both diversification and persistence of range-restricted lineages, harboring high species and phylogenetic richness. This possibly reflects the interplay of multiple evolutionary mechanisms in mountain regions: enhanced speciation rates and more opportunities for persistence and coexistence of lineages, resulting from heterogeneous and dynamic landscapes with diverse niches, and dynamic connectivity between habitats within the mountain linked to complex climatic histories (Antonelli *et al.*, 2018; Rahbek *et al.*, 2019).

Patterns and drivers of diversity potentially vary across spatial scales, e.g., from regional scales to local scales (Keil & Chase, 2019). The ratio of regional diversity to local diversity indicates species turnover between sites,

defined as a calculation of beta diversity (Whittaker, 1960). Thus, beta diversity establishes a direct link between biodiversity at large scale (e.g., regional scales) and fine scales (e.g., local scales), illuminating our understanding of the mechanisms underlying species assemblages and variation in biodiversity across regions (Anderson *et al.*, 2011). In Chapter 3, I examined the impacts of environmental filtering and dispersal history on phylogenetic and species turnover of seed plants and along phylogenetic timescales. I found that despite dispersal history exerted the overall larger influence, its effect was smaller for phylogenetic than for species turnover, and further decreased as phylogenetic time was traced back, as far back as 100 Ma BP. In contrast, environmental filtering remained important across phylogenetic timescales. These findings indicate a major role of environmental constraints for the distribution of major seed plant lineages and strong effects of dispersal limitation for the younger lineages owing to relatively short time available for dispersal (Svenning & Skov, 2007; Paul *et al.*, 2009). Integrated with palaeogeographical reconstructions and a framework to account for dispersal barriers, I demonstrated the contribution of past plate tectonics to phylogenetic turnover by altering geographic proximity of landmass and changing dispersal barriers among regions, including orogenic uplift and expansion or shrinkage of oceans (Raven & Axelrod, 1974). These findings add to phylogenetic lineage assembly and enhance our understanding of contemporary ecological and historical evolutionary mechanisms structuring biodiversity.

Methods of studying biodiversity

I provide important practical advances in the methods of studying biodiversity. First, the thesis provides a solid modeling pipeline for the application of machine learning techniques with freely available codes (Chapter 1). The findings of the thesis illustrated the great predictive performance of machine learning models applied to large distribution and environmental datasets, because of the ability of machine learning models to disentangle underlying complex and interacting diversity–environment relationships. Further, with more available techniques to interpret and visualize machine learning models including understanding the global characteristics of the modeled system and unraveling the fitted relationships between independent and dependent variables (e.g., partial dependence plots as used in Chapter 1) (Lucas, 2020), machine learning methods could be used not only for prediction but also for hypothesis testing and inference (Phillips *et al.*, 2019; Větrovský *et al.*, 2019; Liang *et al.*, 2022). Therefore, machine learning coupled with recent increases in plant-ecological information at large scales (Weigelt *et al.*, 2020; Sabatini *et al.*, 2021; Govaerts *et al.*, 2021), provide exciting new opportunities for ecologists to gain an improved understanding and quantitative knowledge in biogeography and macroecology. Chapter 1 could serve as a useful template for various future biogeographical and macroecological research. Second, I developed a new phylogenetic measure, i.e., standardized effect size of relative phylogenetic endemism (Chapter 2), which quantifies the degree to which disproportionately young or old lineages are spatially restricted in a particular area. This metric therefore presents a novel avenue for testing evolutionary and ecological processes that may be responsible for current biodiversity patterns. Third, the thesis provided new environmental measures, including the metric quantifying the degree to which areas of different biomes vary across geological times (Chapter 1) and the framework for calculating cost distance across barriers (Chapter 3). These measures capture important environmental conditions that are hypothesized to affect current diversity patterns and could potentially be applied to future biodiversity research.

Challenges and future perspectives in biodiversity research

Scales

Biodiversity patterns and underlying processes depend on scale (Ricklefs, 1987; Willis & Whittaker, 2002). Patterns and processes found at one spatial scale may not necessarily emerge at other scales. For example, evolutionary processes (speciation, extinction and migration) and geological and climatic history are key drivers of regional diversity (Antonelli *et al.*, 2015; Jiménez-Alfaro *et al.*, 2018), while local diversity tends to be shaped mainly by processes related to community assemblage (dispersal, environmental filtering and biotic interactions) (Cavender-Bares *et al.*, 2009). Scale dependence has been shown to be potentially ubiquitous in patterns and drivers of species richness (Willis & Whittaker, 2002; Rahbek, 2004; Keil & Chase, 2019) and endemism (Daru *et al.*, 2020). Therefore, the patterns and drivers of regional plant diversity found in the thesis cannot be extrapolated to other possible spatial scales, such as plot and landscape scales. For example, Puglielli and Pärtel (2023) compared the results of the thesis (Chapter 1) with a study that predicted global patterns of plant species richness at the plot scale and uncovered their drivers based on a vegetation-plot dataset (Sabatini *et al.*, 2022), and found differences in diversity drivers. Specifically, while climate consistently showed a strong influence on plant species richness at the regional and plot scale, environmental heterogeneity only emerged as a key driver of regional plant diversity (Sabatini *et al.*, 2022). Such difference highlights another important issue of biodiversity research is to identify various drivers of biodiversity across scales. Moreover, global patterns of plant diversity from a phylogenetic perspective at the plot scale remain to be explored, which could be the next step for understanding local biodiversity based on the global dataset of vegetation plots (Sabatini *et al.*, 2021).

Furthermore, knowledge of landscape-scale plant diversity is still lacking, especially when evolutionary history is taken into account. To fill this important gap, additional landscape-scale data are required. Currently, the species composition data in GIFT and WCVP are coarse-grained. Also, there is a lack of data covering some relevant parts of the globe in GIFT. Once GIFT reach the goal of a complete coverage of the globe in coarse grains, GIFT will focus more on the collection of distribution data in fine resolutions, i.e., species checklists and floras for small geographic regions. In this way, regions covered by the same landscapes may be sampled individually. For example, extreme habitats (i.e., deserts and arctic regions) remain under-sampled in the current data, which are often part of artificially delimited regions (e.g., Chad and Libya sampled instead of the Sahara). These efforts in GIFT might provide new opportunities for exploring landscape-scale plant diversity, which shed lights on understanding of biodiversity across various spatial scales and promote basic research and conservation initiatives.

Multifaceted biodiversity

Biodiversity is multifaceted and its three key facets include taxonomy, phylogeny, and traits. The present thesis focuses on plant diversity in terms of evolutionary history in particular and taxonomic identity, but did not consider functional diversity. Functional diversity refers to the variation in functional traits (including morphological, physiological and phenological attributes of organisms) that influence an organism's performance or fitness (Petchey & Gaston, 2002), which in turn can help to explain multiple assembly processes (Spasojevic & Suding, 2012). Because biodiversity effects on ecosystem functioning ultimately result from interactions between organisms and their environment relative to the organism's traits, knowing functional traits and summary them in a diversity measure provide a good way to assess the role of biodiversity in ecosystem functioning (e.g., primary production and ecosystem stability) (Tilman, 2001; Díaz & Cabido, 2001; Petchey & Gaston, 2006). Consequently, functional diversity can be used to predict the functional consequence of biodiversity loss due to human disturbances (Petchey & Gaston, 2006; Cardinale *et al.*, 2012). Despite the importance of assessing functional diversity, the lack of large-scale assessments of functional traits for large taxonomic groups (e.g., vascular plants) restricts our ability to conduct global-scale studies on functional diversity. Efforts are being made to compile

species-level traits for plants at a global scale (Weigelt *et al.*, 2020; Kattge *et al.*, 2020; Govaerts *et al.*, 2021; Díaz *et al.*, 2022), which may provide exciting opportunities to close the gap in the future.

Beyond considering each facet of biodiversity independently, a more integrated approach by considering all three facets jointly should be used for better understanding of biodiversity mechanisms (Guo *et al.*, 2020; Xu *et al.*, 2023). For example, Xu *et al.* (2023) assessed the influence of Quaternary climate change on global beta diversity in trees across the three facets, and their results reflected phylogenetically and functionally selective processes in species replacement, extinction and colonization during climate change periods. Furthermore, given the importance of each facet to biodiversity and ecosystem functioning (Pollock *et al.*, 2017), coupled with the fact that any one diversity facet cannot be reliably captured by other facets (Devictor *et al.*, 2010; Mazel *et al.*, 2018), considering multiple facets of diversity together could provide a biologically comprehensive foundation for future conservation planning efforts. For example, Brum *et al.*, (2017) identified areas of high conservation priority among and across the three facets of mammal diversity, which could serve as an important resource for global protected area network.

Conclusion

The thesis uncovered global patterns of regional diversity and composition in plants accounting for evolutionary history based on the integration of unprecedented plant distribution and phylogenetic information. It thus provides essential information for future biodiversity science and biodiversity conservation priorities. Also, the thesis provided the most comprehensive tests of how past and current environmental factors jointly contribute to plant diversity. With the developments of phylogenetic approaches and paleoenvironmental reconstructions, the thesis emphasized the significance of past climate change and geological history on regional plant diversity, and uncovered the varied effects of dispersal history linked to past plate tectonics and phylogenetically conserved environmental constraints on regional seed plant composition along the phylogenetic timescale. Consequently, the thesis sheds lights on understanding of biogeographical and evolutionary mechanisms structuring biodiversity.

References

Anderson MJ, Crist TO, Chase JM, Vellend M, Inouye BD, Freestone AL, Sanders NJ, Cornell HV, Comita LS, Davies KF, *et al.* 2011. Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist: Roadmap for beta diversity. *Ecology Letters* 14: 19–28.

Antonelli A, Kissling WD, Flantua SGA, Bermúdez MA, Mulch A, Muellner-Riehl AN, Kreft H, Linder HP, Badgley C, Fjeldså J, *et al.* 2018. Geological and climatic influences on mountain biodiversity. *Nature Geoscience* 11: 718–725.

Antonelli A, Smith RJ, Perrigo AL, Crottini A, Hackel J, Testo W, Farooq H, Torres Jiménez MF, Andela N, Andermann T, *et al.* 2022. Madagascar's extraordinary biodiversity: Evolution, distribution, and use. *Science* 378: eabf0869.

Antonelli A, Zizka A, Silvestro D, Scharn R, Cascales-MiÃ±ana B, Bacon CD. 2015. An engine for global plant diversity: highest evolutionary turnover and emigration in the American tropics. *Frontiers in Genetics* 6.

Arakaki M, Christin P-A, Nyffeler R, Lendel A, Eggli U, Ogburn RM, Spriggs E, Moore MJ, Edwards EJ. 2011. Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proceedings of the National Academy of Sciences* 108: 8379–8384.

Araújo MB, New M. 2007. Ensemble forecasting of species distributions. *Trends in Ecology & Evolution* 22: 42–47.

Barnes R, Sahr K. 2017. dggridR: Discrete Global Grids for R.

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

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

Baselga A. 2012. The relationship between species replacement, dissimilarity derived from nestedness, and nestedness: Species replacement and nestedness. *Global Ecology and Biogeography* 21: 1223–1232.

Bennett KD, Tzedakis PC, Willis KJ. 1991. Quaternary refugia of north European trees. *Journal of Biogeography* 18: 103–115.

Berghaus HKW. 1837. *Berghaus physikalischer Atlas.* Gotha: Justus Perthes.

Bergstra J, Bengio Y. 2012. Random search for hyper-parameter optimization. *Journal of Machine learning research* 13: 281–305.

Bjorholm S, Svenning J-C, Skov F, Balslev H. 2005. Environmental and spatial controls of palm (Arecaceae) species richness across the Americas. *Global Ecology and Biogeography* 14: 423–429.

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

Boyden JA, Müller RD, Gurnis M, Torsvik TH, Clark JA, Turner M, Ivey-Law H, Watson RJ, Cannon JS. 2011. Next-generation plate-tectonic reconstructions using GPlates. In: Keller GR, Baru C, eds. Geoinformatics: Cyberinfrastructure for the solid Earth Science. Cambridge: Cambridge University Press, 95–114.

Boyle B, Hopkins N, Lu Z, Raygoza Garay JA, Mozzherin D, Rees T, Matasci N, Narro ML, Piel WH, Mckay SJ, *et al.* 2013. The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinformatics* 14: 16.

Bromme T. 1851. *Atlas zu Alex. v. Humboldt's Kosmos in zweiundvierzig Tafeln mit erläuterndem Texte.* Stuttgart: Krais & Hoffmann.

Brown JH. 2014. Why are there so many species in the tropics? *Journal of Biogeography* 41: 8–22.

Brown JH, Gillooly JF, Allen AP, Savage VM, West GB. 2004. Toward a metabolic theory of ecology. *Ecology* 85: 1771–1789.

Brown JL, Hill DJ, Dolan AM, Carnaval AC, Haywood AM. 2018. PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Scientific Data* 5: 1–9.

Brown JH, Kodric-Brown A. 1977. Turnover rates in insular biogeography: effect of immigration on extinction. *Ecology* 58: 445–449.

Brum FT, Graham CH, Costa GC, Hedges SB, Penone C, Radeloff VC, Rondinini C, Loyola R, Davidson AD. 2017. Global priorities for conservation across multiple dimensions of mammalian diversity. *Proceedings of the National Academy of Sciences* 114: 7641–7646.

Buckley LB, Jetz W. 2007. Environmental and historical constraints on global patterns of amphibian richness. *Proceedings of the Royal Society B: Biological Sciences* 274: 1167–1173.

Buerki S, Devey DS, Callmander MW, Phillipson PB, Forest F. 2013. Spatio-temporal history of the endemic genera of Madagascar. *Botanical Journal of the Linnean Society* 171: 304–329.

Burns JH, Strauss SY. 2011. More closely related species are more ecologically similar in an experimental test. *Proceedings of the National Academy of Sciences* 108: 5302–5307.

Cadotte MW, Carscadden K, Mirotchnick N. 2011. Beyond species: functional diversity and the maintenance of ecological processes and services: Functional diversity in ecology and conservation. *Journal of Applied Ecology* 48: 1079–1087.

Cai L, Kreft H, Taylor A, Denelle P, Schrader J, Essl F, van Kleunen M, Pergl J, Pyšek P, Stein A, *et al.* 2023. Global models and predictions of plant diversity based on advanced machine learning techniques. *New Phytologist* 237: 1432–1445.

Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings C, Venail P, Narwani A, Mace GM, Tilman D, Wardle DA, *et al.* 2012. Biodiversity loss and its impact on humanity. *Nature* 486: 59–67.

Cavender-Bares J, Kozak KH, Fine PVA, Kembel SW. 2009. The merging of community ecology and phylogenetic biology. *Ecology Letters* 12: 693–715.

Ceballos G, Ehrlich PR, Barnosky AD, García A, Pringle RM, Palmer TM. 2015. Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Science Advances* 1: e1400253.

Chamberlain SA, Szöcs E. 2013. taxize: taxonomic search and retrieval in R. *F1000Research* 2: 191.

Chen T, Guestrin C. 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA: ACM, 785–794.

Chiarucci A, Bacaro G, Scheiner SM. 2011. Old and new challenges in using species diversity for assessing biodiversity. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366: 2426–2437.

Clement WL, Tebbitt MC, Forrest LL, Blair JE, Brouillet L, Eriksson T, Swensen SM. 2004. Phylogenetic position and biogeography of *Hillebrandia sandwicensis* (Begoniaceae): a rare Hawaiian relict. *American Journal of Botany* 91: 905–917.

Condamine FL, Sperling FAH, Wahlberg N, Rasplus J-Y, Kergoat GJ. 2012. What causes latitudinal gradients in species diversity? Evolutionary processes and ecological constraints on swallowtail biodiversity: Phylogeny and latitudinal diversity gradient. *Ecology Letters* 15: 267–277.

Connor EF, McCoy ED. 1979. The statistics and biology of the species-area relationship. *The American Naturalist* 113: 791–833.

Couvreur TLP, Dauby G, Blach-Overgaard A, Deblauwe V, Dessein S, Droissart V, Hardy OJ, Harris DJ, Janssens SB, Ley AC, *et al.* 2021. Tectonics, climate and the diversification of the tropical African terrestrial flora and fauna. *Biological Reviews* 96: 16–51.

Cowie RH, Bouchet P, Fontaine B. 2022. The Sixth Mass Extinction: fact, fiction or speculation? *Biological Reviews* 97: 640–663.

Cowling RM, Rundel PW, Lamont BB, Kalin Arroyo M, Arianoutsou M. 1996. Plant diversity in mediterranean-climate regions. *Trends in Ecology & Evolution* 11: 362–366.

Crase B, Liedloff AC, Wintle BA. 2012. A new method for dealing with residual spatial autocorrelation in species distribution models. *Ecography* 35: 879–888.

Crisci C, Ghattas B, Perera G. 2012. A review of supervised machine learning algorithms and their applications to ecological data. *Ecological Modelling* 240: 113–122.

Crisp MD, Laffan S, Linder HP, Monro A. 2001. Endemism in the Australian flora. *Journal of Biogeography* 28: 183–198.

Cronk QCB. 1987. The history of endemic flora of St Helena: a relictual series. *New Phytologist* 105: 509–520.

Cronk QCB. 1992. Relict floras of Atlantic islands: patterns assessed. *Biological Journal of the Linnean Society* 46: 91–103.

Cronk QCB. 1997. Islands: stability, diversity, conservation. *Biodiversity and Conservation* 6: 477–493.

Csardi G, Nepusz T. 2006. The igraph software package for complex network research. *InterJournal, Complex Systems* 1695: 1–9.

Currie DJ, Mittelbach GG, Cornell HV, Field R, Guegan J-F, Hawkins BA, Kaufman DM, Kerr JT, Oberdorff T, O'Brien E, *et al.* 2004. Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters* 7: 1121–1134.

Dagallier LMJ, Janssens SB, Dauby G, Blach-Overgaard A, Mackinder BA, Droissart V, Svenning J, Sosef MSM, Stévart T, Harris DJ, *et al.* 2020. Cradles and museums of generic plant diversity across tropical Africa. *New Phytologist* 225: 2196–2213.

Danielson JJ, Gesch DB. 2011. *Global multi-resolution terrain elevation data 2010 (GMTED2010).*

Daru BH, Elliott TL, Park DS, Davies TJ. 2017. Understanding the Processes Underpinning Patterns of Phylogenetic Regionalization. *Trends in Ecology & Evolution* 32: 845–860.

Daru BH, Farooq H, Antonelli A, Faurby S. 2020. Endemism patterns are scale dependent. *Nature Communications* 11: 2115.

Davis MB, Shaw RG. 2001. Range shifts and adaptive responses to Quaternary climate change. *Science* 292: 673–679.

Devictor V, Mouillot D, Meynard C, Jiguet F, Thuiller W, Mouquet N. 2010. Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world: Spatial mismatch between diversity facets. *Ecology Letters* 13: 1030–1040.

Díaz S, Kattge J, Cornelissen JHC, Wright IJ, Lavorel S, Dray S, Reu B, Kleyer M, Wirth C, Prentice IC, *et al.* 2022. The global spectrum of plant form and function: enhanced species-level trait dataset. *Scientific Data* 9: 755.

Díaz S, Malhi Y. 2022. Biodiversity: Concepts, Patterns, Trends, and Perspectives. *Annual Review of Environment and Resources* 47: 31–63.

Díaz S, Cabido M. 2001. Vive la différence: plant functional diversity matters to ecosystem processes. *Trends in Ecology & Evolution* 16: 646–655.

Donoghue MJ. 2008. A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences* 105: 11549–11555.

Dormann C, M. McPherson J, B. Araújo M, Bivand R, Bolliger J, Carl G, G. Davies R, Hirzel A, Jetz W, Daniel Kissling W, *et al.* 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30: 609–628.

Dowsett H, Dolan A, Rowley D, Moucha R, Forte AM, Mitrovica JX, Pound M, Salzmann U, Robinson M, Chandler M, *et al.* 2016. The PRISM4 (mid-Piacenzian) paleoenvironmental reconstruction. *Climate of the Past* 12: 1519–1538.

Duarte LDS, Both C, Debastiani VJ, Carlucci MB, Gonçalves LO, Cappelatti L, Seger GDS, Bastazini VAG, Brum FT, Salengue EV, *et al.* 2014. Climate effects on amphibian distributions depend on phylogenetic resolution and the biogeographical history of taxa: Historical and climatic correlates of amphibian distributions. *Global Ecology and Biogeography* 23: 213–222.

Dynesius M, Jansson R. 2000. Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the National Academy of Sciences* 97: 9115–9120.

Eiserhardt WL, Borchsenius F, Sandel B, Kissling WD, Svenning J-C. 2015. Late Cenozoic climate and the phylogenetic structure of regional conifer floras world-wide. *Global Ecology and Biogeography* 24: 1136–1148.

Eiserhardt WL, Svenning J-C, Baker WJ, Couvreur TLP, Balslev H. 2013. Dispersal and niche evolution jointly shape the geographic turnover of phylogenetic clades across continents. *Scientific Reports* 3: 1164.

Elith J, Kearney M, Phillips S. 2010. The art of modelling range-shifting species. *Methods in Ecology and Evolution* 1: 330–342.

Endress PK. 2001. The flowers in extant basal angiosperms and inferences on ancestral flowers. *International Journal of Plant Sciences* 162: 1111–1140.

Enquist BJ, Condit R, Peet RK, Schildhauer M, Thiers BM. 2016. *Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity*. PeerJ Preprints.

Enquist BJ, Feng X, Boyle B, Maitner B, Newman EA, Jørgensen PM, Roehrdanz PR, Thiers BM, Burger JR, Corlett RT, *et al.* 2019. The commonness of rarity: Global and future distribution of rarity across land plants. *Science Advances* 5: eaaz0414.

Evans KL, Warren PH, Gaston KJ. 2005. Species–energy relationships at the macroecological scale: a review of the mechanisms. *Biological Reviews* 80: 1–25.

Faith DP. 1992. Conservation evaluation and phylogenetic diversity. *Biological Conservation* 61: 1–10.

Fernández-Mazuecos M, Vargas P, McCauley RA, Monjas D, Otero A, Chaves JA, Guevara Andino JE, Rivas-Torres G. 2020. The radiation of Darwin's giant daisies in the Galápagos Islands. *Current Biology* 30: 4989–4998.

Fernández-Palacios JM, Kreft H, Irl SDH, Norder S, Ah-Peng C, Borges PAV, Burns KC, de Nascimento L, Meyer J-Y, Montes E, *et al.* 2021. Scientists' warning – The outstanding biodiversity of islands is in peril. *Global Ecology and Conservation* 31: e01847.

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.

Ficetola GF, Mazel F, Thuiller W. 2017. Global determinants of zoogeographical boundaries. *Nature Ecology & Evolution* 1: 0089.

Fine PVA. 2015. Ecological and evolutionary drivers of geographic variation in species diversity. *Annual Review of Ecology, Evolution, and Systematics* 46: 369–392.

Fitzpatrick M, Mokany K, Manion G, Nieto-Lugilde D, Ferrier S. 2022. gdm: Generalized Dissimilarity Modeling. R package version 1.5.0-3.

Fitzpatrick MC, Sanders NJ, Normand S, Svenning J-C, Ferrier S, Gove AD, Dunn RR. 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.

Fjeldså J, Lovett JC. 1997. Geographical patterns of old and young species in African forest biota: the significance of specific montane areas as evolutionary centres. *Biodiversity and Conservation* 6: 325–346.

Flantua SGA, Payne D, Borregaard MK, Beierkuhnlein C, Steinbauer MJ, Dullinger S, Essl F, Irl SDH, Kienle D, Kreft H, *et al.* 2020. Snapshot isolation and isolation history challenge the analogy between mountains and islands used to understand endemism (D Storch, Ed.). *Global Ecology and Biogeography* 29: 1651–1673.

Forest F, Grenyer R, Rouget M, Davies TJ, Cowling RM, Faith DP, Balmford A, Manning JC, Proche ş Ş, van der Bank M, *et al.* 2007. Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* 445: 757–760.

Francis AP, Currie DJ. 2003. A globally consistent richness-climate relationship for angiosperms. *The American Naturalist* 161: 523–536.

Gallagher RV, Allen SP, Govaerts R, Rivers MC, Allen AP, Keith DA, Merow C, Maitner B, Butt N, Auld TD, *et al.* 2023. Global shortfalls in threat assessments for endemic flora by country. *PLANTS, PEOPLE, PLANET*: ppp3.10369.

Garcillán PP, Ezcurra E. 2003. Biogeographic regions and β-diversity of woody dryland legumes in the Baja California peninsula. *Journal of Vegetation Science* 14: 859–868.

Gaston KJ, Spicer JI. 2004. *Biodiversity: an introduction*. Malden, MA: Blackwell Pub.

GBIF. 2020. GBIF: The Global Biodiversity Information Facility (year) What is GBIF?. Available from https://www.gbif.org/what-is-gbif [13 January 2020].

Genuer R, Poggi J-M, Tuleau-Malot C. 2015. VSURF: an R package for variable selection using random forests. *The R Journal* 7: 19–33.

Gillespie RG, Roderick GK. 2002. Arthropods on islands: colonization, speciation, and conservation. *Annual Review of Entomology* 47: 595–632.

Gillooly JF, Allen AP. 2007. Linking global patterns in biodiversity to evolutionary dynamics using metabolic theory. *Ecology* 88: 1890–1894.

Govaerts R, Nic Lughadha E, Black N, Turner R, Paton A. 2021. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Scientific Data* 8: 215.

Graham CH, Fine PVA. 2008. Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. *Ecology Letters* 11: 1265–1277.

Grenyer R, Orme CDL, Jackson SF, Thomas GH, Davies RG, Davies TJ, Jones KE, Olson VA, Ridgely RS, Rasmussen PC, *et al.* 2006. Global distribution and conservation of rare and threatened vertebrates. *Nature* 444: 93–96.

Groussin M, Mazel F, Sanders JG, Smillie CS, Lavergne S, Thuiller W, Alm EJ. 2017. Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nature Communications* 8: 14319.

Gumbs R, Gray CL, Böhm M, Burfield IJ, Couchman OR, Faith DP, Forest F, Hoffmann M, Isaac NJB, Jetz W, *et al.* 2023. The EDGE2 protocol: Advancing the prioritisation of Evolutionarily Distinct and Globally Endangered species for practical conservation action (U Ramakrishnan, Ed.). *PLOS Biology* 21: e3001991.

Günther F, Fritsch S. 2010. Neuralnet: training of neural networks. *The R Journal* 2: 30–38.

Guo W-Y, Serra-Diaz JM, Schrodt F, Eiserhardt WL, Maitner BS, Merow C, Violle C, Blach-Overgaard A, Zhang J, Anand M, *et al.* 2020. *Paleoclimate and current climate collectively shape the phylogenetic and functional diversity of trees worldwide*. Ecology.

Hagen O, Skeels A, Onstein RE, Jetz W, Pellissier L. 2021. Earth history events shaped the evolution of uneven biodiversity across tropical moist forests. *Proceedings of the National Academy of Sciences* 118: e2026347118.

Hardy OJ, Couteron P, Munoz F, Ramesh BR, Pélissier R. 2012. Phylogenetic turnover in tropical tree communities: impact of environmental filtering, biogeography and mesoclimatic niche conservatism: Phylogenetic turnover in tropical trees. *Global Ecology and Biogeography* 21: 1007–1016.

Hawkins BA, Field R, Cornell HV, Currie DJ, Guégan J-F, Kaufman DM, Kerr JT, Mittelbach GG, Oberdorff T, O'Brien EM, *et al.* 2003. Energy, water, and broad-scale geographic patterns of species richness. *Ecology* 84: 3105–3117.

Hengl T, Mendes de Jesus J, Heuvelink GBM, Ruiperez Gonzalez M, Kilibarda M, Blagotić A, Shangguan W, Wright MN, Geng X, Bauer-Marschallinger B, *et al.* 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* 12: e0169748.

Henrot A-J, François L, Favre E, Butzin M, Ouberdous M, Munhoven G. 2010. Effects of CO_2, continental distribution, topography and vegetation changes on the climate at the Middle Miocene: a model study. *Climate of the Past* 6: 675–694.

Hillebrand H. 2004. On the Generality of the Latitudinal Diversity Gradient. *The American Naturalist* 163: 192–211.

Hojsgaard D, Klatt S, Baier R, Carman JG, Hörandl E. 2014. Taxonomy and biogeography of apomixis in angiosperms and associated biodiversity characteristics. *Critical Reviews in Plant Sciences* 33: 414–427.

Holt BG, Lessard J-P, Borregaard MK, Fritz SA, Araújo MB, Dimitrov D, Fabre P-H, Graham CH, Graves GR, Jønsson KA, *et al.* 2013. An Update of Wallace's Zoogeographic Regions of the World. *Science* 339: 74–78.

Hoorn C, Wesselingh FP, ter Steege H, Bermudez MA, Mora A, Sevink J, Sanmartin I, Sanchez-Meseguer A, Anderson CL, Figueiredo JP, *et al.* 2010. Amazonia Through Time: Andean Uplift, Climate Change, Landscape Evolution, and Biodiversity. *Science* 330: 927–931.

Hughes CE, Atchison GW. 2015. The ubiquity of alpine plant radiations: from the Andes to the Hengduan Mountains. *New Phytologist* 207: 275–282.

Isbell F, Calcagno V, Hector A, Connolly J, Harpole WS, Reich PB, Scherer-Lorenzen M, Schmid B, Tilman D, van Ruijven J, *et al.* 2011. High plant diversity is needed to maintain ecosystem services. *Nature* 477: 199–202.

James G, Witten D, Hastie T, Tibshirani R (Eds.). 2013. *An introduction to statistical learning: with applications in r.* New York, US: Springer.

Jansson R. 2003. Global patterns in endemism explained by past climatic change. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270: 583–590.

Jansson R, Dynesius M. 2002. The fate of clades in a world of recurrent climatic change: Milankovitch oscillations and evolution. *Annual Review of Ecology and Systematics* 33: 741–777.

Jiménez-Alfaro B, Girardello M, Chytrý M, Svenning J-C, Willner W, Gégout J-C, Agrillo E, Campos JA, Jandt U, Kącki Z, *et al.* 2018. History and environment shape species pools and community diversity in European beech forests. *Nature Ecology & Evolution* 2: 483–490.

Jin Y, Qian H. 2019. V. PhyloMaker: an R package that can generate very large phylogenies for vascular plants. *Ecography* 42: 1353–1359.

Jump AS, Mátyás C, Peñuelas J. 2009. The altitude-for-latitude disparity in the range retractions of woody species. *Trends in Ecology & Evolution* 24: 694–701.

Jürgens N. 1997. Floristic biodiversity and history of African arid regions. *Biodiversity and Conservation* 6: 495–514.

Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder HP, Kessler M. 2017. Climatologies at high resolution for the earth's land surface areas. *Scientific Data* 4: 170122.

Kattge J, Bönisch G, Díaz S, Lavorel S, Prentice IC, Leadley P, Susanne Tautenhahn, Gijsbert D. A. Werner, Tuomas Aakala, Mehdi Abedi, *et al.* 2020. TRY plant trait database – enhanced coverage and open access. *Global Change Biology* 26: 119–188.

Keil P, Chase JM. 2019. Global patterns and drivers of tree diversity integrated across a continuum of spatial grains. *Nature Ecology & Evolution* 3: 390–399.

Kier G, Kreft H, Lee TM, Jetz W, Ibisch PL, 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* 106: 9322–9327.

Kisel Y, Barraclough TG. 2010. Speciation has a spatial scale that depends on levels of gene flow. *The American Naturalist* 175: 316–334.

Kissling WD, Baker WJ, Balslev H, Barfod AS, Borchsenius F, Dransfield J, Govaerts R, Svenning J-C. 2012. Quaternary and pre-Quaternary historical legacies in the global distribution of a major tropical plant lineage. *Global Ecology and Biogeography* 21: 909–921.

Kissling WD, Carl G. 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography* 17: 59–71.

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

König C, Weigelt P, Schrader J, Taylor A, Kattge J, Kreft H. 2019. Biodiversity data integration—the significance of data resolution and domain. *PLOS Biology* 17: e3000183.

Kraft NJB, Adler PB, Godoy O, James EC, Fuller S, Levine JM. 2015. Community assembly, coexistence and the environmental filtering metaphor (J Fox, Ed.). *Functional Ecology* 29: 592–599.

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

Kubota Y, Hirao T, Fujii S, Murakami M. 2011. Phylogenetic beta diversity reveals historical effects in the assemblage of the tree floras of the Ryukyu Archipelago: Correspondence. *Journal of Biogeography* 38: 1006–1008.

Kuhn M. 2008. Building predictive models in R using the caret package. *Journal of Statistical Software* 28: 1–26.

Legendre P, De Cáceres M. 2013. Beta diversity as the variance of community data: dissimilarity coefficients and partitioning (H Morlon, Ed.). *Ecology Letters* 16: 951–963.

Leprieur F, Albouy C, De Bortoli J, Cowman PF, Bellwood DR, Mouillot D. 2012. Quantifying Phylogenetic Beta Diversity: Distinguishing between 'True' Turnover of Lineages and Phylogenetic Diversity Gradients (M Shawkey, Ed.). *PLoS ONE* 7: e42760.

Li L. 2019. Geographically weighted machine learning and downscaling for high-resolution spatiotemporal estimations of wind speed. *Remote Sensing* 11: 1378.

Liang J, Gamarra JGP, Picard N, Zhou M. 2022. Co-limitation towards lower latitudes shapes global forest diversity gradients. *Nature Ecology & Evolution*: 1–15.

Lim JY, Svenning J-C, Göldel B, Faurby S, Kissling WD. 2020. Frugivore-fruit size relationships between palms and mammals reveal past and future defaunation impacts. *Nature Communications* 11: 4904.

Linder HP, Antonelli A, Humphreys AM, Pirie MD, Wüest RO. 2013. What determines biogeographical ranges? Historical wanderings and ecological constraints in the danthonioid grasses (R Ladle, Ed.). *Journal of Biogeography* 40: 821–834.

Liu Y, Xu X, Dimitrov D, Pellissier L, Borregaard MK, Shrestha N, Su X, Luo A, Zimmermann NE, Rahbek C, *et al.* 2023. An updated floristic map of the world. *Nature Communications* 14: 2990.

Lomolino MV, Riddle BR, Whittaker RJ. 2016. *Biogeography: Biological Diversity Across Space and Time.* Sunderland, Massachusetts, U.S.A.: Sinauer Associates.

Losos JB, Ricklefs RE. 2009. Adaptation and diversification on islands. *Nature* 457: 830–836.

Lu L-M, Mao L-F, Yang T, Ye J-F, Liu B, Li H-L, Sun M, Miller JT, Mathews S, Hu H-H, *et al.* 2018. Evolutionary history of the angiosperm flora of China. *Nature* 554: 234–238.

Lucas TCD. 2020. A translucent box: interpretable machine learning in ecology. *Ecological Monographs* 90: e01422.

Mace GM, Gittleman JL, Purvis A. 2003. Preserving the tree of life. *Science* 300: 1707–1709.

Magurran AE. 2021. Measuring biological diversity. *Current Biology* 31: R1174–R1177.

Majeský L, Krahulec F, Vašut RJ. 2017. How apomictic taxa are treated in current taxonomy: A review. *Taxon* 66: 1017–1040.

Malyshev LI. 1975. The quantitative analysis of flora: spatial diversity, level of specific richness, and representativity of sampling areas. *Bot. Zhurn. (Moscow & Leningrad)* 60: 1537–1550.

Marmion M, Parviainen M, Luoto M, Heikkinen RK, Thuiller W. 2009. Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions* 15: 59–69.

Matthews KJ, Maloney KT, Zahirovic S, Williams SE, Seton M, Müller RD. 2016. Global plate boundary evolution and kinematics since the late Paleozoic. *Global and Planetary Change* 146: 226–250.

Mazel F, Pennell MW, Cadotte MW, Diaz S, Dalla Riva GV, Grenyer R, Leprieur F, Mooers AO, Mouillot D, Tucker CM, *et al.* 2018. Prioritizing phylogenetic diversity captures functional diversity unreliably. *Nature Communications* 9: 2888.

Mazel F, Wüest RO, Lessard J, Renaud J, Ficetola GF, Lavergne S, Thuiller W, Lyons K. 2017. Global patterns of β-diversity along the phylogenetic time-scale: The role of climate and plate tectonics. *Global Ecology and Biogeography* 26: 1211–1221.

McFadden IR, Sandel B, Tsirogiannis C, Morueta-Holme N, Svenning J, Enquist BJ, Kraft NJB. 2019. Temperature shapes opposing latitudinal gradients of plant taxonomic and phylogenetic β diversity (M Anderson, Ed.). *Ecology Letters* 22: 1126–1135.

McNulty K. 2022. *Handbook of Graphs and Networks in People Analytics: With Examples in R and Python.* Boca Raton: Chapman and Hall/CRC.

Merckx VSFT, Hendriks KP, Beentjes KK, Mennes CB, Becking LE, Peijnenburg KTCA, Afendy A, Arumugam N, de Boer H, Biun A, *et al.* 2015. Evolution of endemism on a young tropical mountain. *Nature* 524: 347–350.

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

Mishler BD. 2023. Spatial phylogenetics. *Journal of Biogeography*: jbi.14618.

Mishler BD, Knerr N, González-Orozco CE, Thornhill AH, Laffan SW, Miller JT. 2014. Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian Acacia. *Nature Communications* 5: 4473.

Mittelbach GG, Schemske DW, Cornell HV, Allen AP, Brown JM, Bush MB, Harrison SP, Hurlbert AH, Knowlton N, Lessios HA, *et al.* 2007. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters* 10: 315–331.

Mokany K, Ware C, Woolley SNC, Ferrier S, Fitzpatrick MC. 2022. A working guide to harnessing generalized dissimilarity modelling for biodiversity analysis and conservation assessment (V Bahn, Ed.). *Global Ecology and Biogeography* 31: 802–821.

Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, Wellman CH, Yang Z, Schneider H, Donoghue PCJ. 2018. The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences* 115.

Morrissey MB, Ruxton GD. 2018. Multiple regression is not multiple regressions: the meaning of multiple regression and the non-problem of collinearity. *Philosophy, Theory, and Practice in Biology* 10.

Müller RD, Seton M, Zahirovic S, Williams SE, Matthews KJ, Wright NM, Shephard GE, Maloney KT, Barnett-Moore N, Hosseinpour M, *et al.* 2016. Ocean Basin Evolution and Global-Scale Plate Reorganization Events Since Pangea Breakup. *Annual Review of Earth and Planetary Sciences* 44: 107–138.

Mutke J, Barthlott W. 2005. Patterns of vascular plant diversity at continental to global scales. *Biologiske Skrifter* 55: 521–538.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858.

Nekola JC, White PS. 1999. The distance decay of similarity in biogeography and ecology. *Journal of Biogeography* 26: 867–878.

Nie Y, Foster CSP, Zhu T, Yao R, Duchêne DA, Ho SYW, Zhong B. 2020. Accounting for Uncertainty in the Evolutionary Timescale of Green Plants Through Clock-Partitioning and Fossil Calibration Strategies (L Jermiin, Ed.). *Systematic Biology* 69: 1–16.

Olden JD, Lawler JJ, Poff NL. 2008. Machine learning methods without tears: a primer for ecologists. *The Quarterly Review of Biology* 83: 171–193.

Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GVN, Underwood EC, D'amico JA, Itoua I, Strand HE, Morrison JC, *et al.* 2001. Terrestrial ecoregions of the world: a new map of life on earth. *BioScience* 51: 933–938.

Ouborg NJ. 1993. Isolation, population size and extinction: the classical and metapopulation approaches applied to vascular plants along the dutch rhine-system. *Oikos* 66: 298–308.

Owens HL, Guralnick R. 2019. climateStability: an R package to estimate climate stability from time-slice climatologies. *Biodiversity Informatics* 14: 8–13.

Park DS, Willis CG, Xi Z, Kartesz JT, Davis CC, Worthington S. 2020. Machine learning predicts large scale declines in native plant phylogenetic diversity. *New Phytologist* 227: 1544–1556.

Paul JR, Morton C, Taylor CM, Tonsor SJ. 2009. Evolutionary Time for Dispersal Limits the Extent but Not the Occupancy of Species' Potential Ranges in the Tropical Plant Genus *Psychotria* (Rubiaceae). *The American Naturalist* 173: 188–199.

Pavoine S, Bonsall MB. 2011. Measuring biodiversity to explain community assembly: a unified approach. *Biological Reviews* 86: 792–812.

Pearse WD, Cadotte MW, Cavender-Bares J, Ives AR, Tucker CM, Walker SC, Helmus MR. 2015. *pez*: phylogenetics for the environmental sciences. *Bioinformatics* 31: 2888–2890.

Pellissier L, Heine C, Rosauer DF, Albouy C. 2018. Are global hotspots of endemic richness shaped by plate tectonics? *Biological Journal of the Linnean Society* 123: 247–261.

Petchey OL, Gaston KJ. 2002. Extinction and the loss of functional diversity. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 269: 1721–1727.

Petchey OL, Gaston KJ. 2006. Functional diversity: back to basics and looking forward. *Ecology Letters* 9: 741–758.

Phillips HRP, Guerra CA, Bartz MLC, Briones MJI, Brown G, Crowther TW, Ferlian O, Gongalsky KB, van den Hoogen J, Krebs J, *et al.* 2019. Global distribution of earthworm diversity. *Science* 366: 480–485.

Pianka ER. 1966. Latitudinal Gradients in Species Diversity: A Review of Concepts. *The American Naturalist* 100: 33–46.

Pitman NCA, Jørgensen PM. 2002. Estimating the size of the world's threatened flora. *Science* 298: 989–989.

Ploton P, Mortier F, Réjou-Méchain M, Barbier N, Picard N, Rossi V, Dormann C, Cornu G, Viennois G, Bayol N, *et al.* 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications* 11: 4540.

Pollock LJ, Thuiller W, Jetz W. 2017. Large conservation gains possible for global biodiversity facets. *Nature* 546: 141–144.

Procheş Ş, Ramdhani S, Perera SJ, Ali JR, Gairola S. 2015. Global hotspots in the present-day distribution of ancient animal and plant lineages. *Scientific Reports* 5: 15457.

Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, Lemmon AR. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* 526: 569–573.

Puglielli G, Pärtel M. 2023. Macroecology of plant diversity across spatial scales. *New Phytologist* 237: 1074–1077.

Purvis A, Agapow P-M, Gittleman JL, Mace GM. 2000. Nonrandom extinction and the loss of evolutionary history. *Science* 288: 328–330.

Purvis A, Hector A. 2000. Getting the measure of biodiversity. *Nature* 405: 212–219.

Qian H, Jin Y. 2021. Are phylogenies resolved at the genus level appropriate for studies on phylogenetic structure of species assemblages? *Plant Diversity* 43: 255–263.

Qian H, Ricklefs RE. 2000. Large-scale processes and the Asian bias in species diversity of temperate plants. *Nature* 407: 180–182.

Qian H, Ricklefs RE. 2004. Taxon richness and climate in angiosperms: is there a globally consistent relationship that precludes region effects? *The American Naturalist* 163: 773–779.

Qian H, Ricklefs RE, Thuiller W. 2021. Evolutionary assembly of flowering plants into sky islands. *Nature Ecology & Evolution* 5: 640–646.

Qian H, Zhang J, Jiang M-C. 2022. Global patterns of fern species diversity: An evaluation of fern data in GBIF. *Plant Diversity* 44: 135–140.

Rahbek C. 2004. The role of spatial scale and the perception of large-scale species-richness patterns: Scale and species-richness patterns. *Ecology Letters* 8: 224–239.

Rahbek C, Borregaard MK, Antonelli A, Colwell RK, Holt BG, Nogues-Bravo D, Rasmussen CMØ, Richardson K, Rosing MT, Whittaker RJ, *et al.* 2019. Building mountain biodiversity: Geological and evolutionary processes. *Science* 365: 1114–1119.

Raven PH, Axelrod DI. 1974. Angiosperm Biogeography and Past Continental Movements. *Annals of the Missouri Botanical Garden* 61: 539.

Raven PH, Gereau RE, Phillipson PB, Chatelain C, Jenkins CN, Ulloa Ulloa C. 2020. The distribution of biodiversity richness in the tropics. *Science Advances* 6: eabc6228.

Ray N, Adams JM. 2001. A GIS-based vegetation map of the world at the last glacial maximum (25,000-15,000 BP). *Internet Archaeology* 11.

Richards AJ. 2003. Apomixis in flowering plants: an overview (HG Dickinson, SJ Hiscock, and PR Crane, Eds.). *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358: 1085–1093.

Ricklefs RE. 1987. Community Diversity: Relative Roles of Local and Regional Processes. *Science* 235: 167–171.

Ricklefs RE, He F. 2016. Region effects influence local tree species diversity. *Proceedings of the National Academy of Sciences* 113: 674–679.

Rodrigues JFM, Diniz-Filho JAF. 2017. Dispersal is more important than climate in structuring turtle communities across different biogeographical realms. *Journal of Biogeography* 44: 2109–2120.

Rohde K. 1992. Latitudinal gradients in species diversity: the search for the primary cause. *Oikos* 65: 514–527.

Rosauer D, Laffan SW, Crisp MD, Donnellan SC, Cook LG. 2009. Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. *Molecular Ecology* 18: 4061–4072.

Rosauer DF, Mooers AO. 2013. Nurturing the use of evolutionary diversity in nature conservation. *Trends in Ecology & Evolution* 28: 322–323.

Rosenzweig ML. 2003. Reconciliation ecology and the future of species diversity. *Oryx* 37: 194–205.

Rudbeck AV, Sun M, Tietje M, Gallagher RV, Govaerts R, Smith SA, Svenning J, Eiserhardt WL. 2022. The Darwinian shortfall in plants: phylogenetic knowledge is driven by range size. *Ecography* 2022: e06142.

Sabatini FM, Jiménez-Alfaro B, Jandt U, Chytrý M, Field R, Kessler M, Lenoir J, Schrodt F, Wiser SK, Arfin Khan MAS, *et al.* 2022. Global patterns of vascular plant alpha diversity. *Nature Communications* 13: 4683.

Sabatini FM, Lenoir J, Hattab T, Arnst EA, Chytrý M, Dengler J, Ruffray PD, Hennekens SM, Jandt U, Jansen F, *et al.* 2021. sPlotOpen–An environmentally balanced, open-access, global dataset of vegetation plots. *Global Ecology and Biogeography* 30: 1740–1764.

Sandel B, Arge L, Dalsgaard B, Davies RG, Gaston KJ, Sutherland WJ, Svenning J-C. 2011. The influence of Late Quaternary climate-change velocity on species endemism. *Science* 334: 660–664.

Sandel B, Weigelt P, Kreft H, Keppel G, van der Sande MT, Levin S, Smith S, Craven D, Knight TM. 2020. Current climate, isolation and history drive global patterns of tree phylogenetic endemism (R Kelly, Ed.). *Global Ecology and Biogeography* 29: 4–15.

Scotese CR, Wright N. 2018. *PALEOMAP Paleodigital Elevation Models (PaleoDEMS) for the Phanerozoic PALEOMAP Project,*.

Sepulchre P, Ramstein G, Fluteau F, Schuster M, Tiercelin J-J, Brunet M. 2006. Tectonic Uplift and Eastern Africa Aridification. *Science* 313: 1419–1423.

Sheth SN, Morueta-Holme N, Angert AL. 2020. Determinants of geographic range size in plants. *New Phytologist* 226: 650–665.

Smith SA, Brown JW. 2018. Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany* 105: 302–314.

Sochor M, Vašut RJ, Sharbel TF, Trávníček B. 2015. How just a few makes a lot: Speciation via reticulation and apomixis on example of European brambles (Rubus subgen. Rubus, Rosaceae). *Molecular Phylogenetics and Evolution* 89: 13–27.

Solt F, Hu Y. 2015. interplot: Plot the Effects of Variables in Interaction Terms.

Spasojevic MJ, Suding KN. 2012. Inferring community assembly mechanisms from functional diversity patterns: the importance of multiple assembly processes: *Functional diversity along gradients*. *Journal of Ecology* 100: 652–661.

Stein A, Gerstner K, Kreft H. 2014. Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales (H Arita, Ed.). *Ecology Letters* 17: 866–880.

Stein A, Kreft H. 2015. Terminology and quantification of environmental heterogeneity in species-richness research: Environmental heterogeneity and species richness. *Biological Reviews* 90: 815–836.

Storch D, Keil P, Jetz W. 2012. Universal species–area and endemics–area relationships at continental scales. *Nature* 488: 78–81.

Svenning J-C. 2003. Deterministic Plio-Pleistocene extinctions in the European cool-temperate tree flora: Deterministic Plio-Pleistocene extinctions. *Ecology Letters* 6: 646–653.

Svenning J-C, Eiserhardt WL, Normand S, Ordonez A, Sandel B. 2015. The influence of paleoclimate on present-day patterns in biodiversity and ecosystems. *Annual Review of Ecology, Evolution, and Systematics* 46: 551–572.

Svenning J-C, Skov F. 2007. Could the tree diversity pattern in Europe be generated by postglacial dispersal limitation? *Ecology Letters* 10: 453–460.

Swenson NG. 2011. Phylogenetic Beta Diversity Metrics, Trait Evolution and Inferring the Functional Beta Diversity of Communities (A Hector, Ed.). *PLoS ONE* 6: e21264.

Szövényi P, Terracciano S, Ricca M, Giordano S, Shaw AJ. 2008. Recent divergence, intercontinental dispersal and shared polymorphism are shaping the genetic structure of amphi-Atlantic peatmoss populations. *Molecular Ecology* 17: 5364–5377.

Takhtajan AL. 1986. *Floristic regions of the world*. Berkeley, CA: University of California Press.

Terborgh J. 1973. On the notion of favorableness in plant ecology. *The American Naturalist* 107: 481–501.

Testolin R, Attorre F, Borchardt P, Brand RF, Bruelheide H, Chytrý M, De Sanctis M, Dolezal J, Finckh M, Haider S, *et al.* 2021. Global patterns and drivers of alpine plant species richness (A Bjorkman, Ed.). *Global Ecology and Biogeography* 30: 1218–1231.

Thompson JD, Lavergne S, Affre L, Gaudeul M, Debussche M. 2005. Ecological differentiation of Mediterranean endemic plants. *Taxon* 54: 967–976.

Thorne RF. 1999. Eastern Asia as a living museum for archaic angiosperms and other seed plants. *Taiwania* 44: 413–422.

Thornhill AH, Baldwin BG, Freyman WA, Nosratinia S, Kling MM, Morueta-Holme N, Madsen TP, Ackerly DD, Mishler BD. 2017. Spatial phylogenetics of the native California flora. *BMC Biology* 15: 96.

Thornhill AH, Mishler BD, Knerr NJ, González-Orozco CE, Costion CM, Crayn DM, Laffan SW, Miller JT. 2016. Continental-scale spatial phylogenetics of Australian angiosperms provides insights into ecology, evolution and conservation. *Journal of Biogeography* 43: 2085–2098.

Thuiller W, Lafourcade B, Engler R, Araújo MB. 2009. BIOMOD–a platform for ensemble forecasting of species distributions. *Ecography* 32: 369–373.

Tietje M, Antonelli A, Baker WJ, Govaerts R, Smith SA, Eiserhardt WL. 2022. Global variation in diversification rate and species richness are unlinked in plants. *Proceedings of the National Academy of Sciences* 119: e2120662119.

Tilman D. 2001. Functional Diversity. In: Encyclopedia of Biodiversity. Elsevier, 109–120.

Tilman D, Isbell F, Cowles JM. 2014. Biodiversity and ecosystem functioning. *Annual Review of Ecology, Evolution, and Systematics* 45: 471–493.

Tucker CM, Aze T, Cadotte MW, Cantalapiedra JL, Chisholm C, Díaz S, Grenyer R, Huang D, Mazel F, Pearse WD, *et al*. 2019. Assessing the utility of conserving evolutionary history. *Biological Reviews* 94: 1740–1760.

Tucker CM, Cadotte MW, Carvalho SB, Davies TJ, Ferrier S, Fritz SA, Grenyer R, Helmus MR, Jin LS, Mooers AO, *et al*. 2017. A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biological Reviews* 92: 698–715.

Tucker CM, Davies TJ, Cadotte MW, Pearse WD. 2018. On the relationship between phylogenetic diversity and trait diversity. *Ecology* 99: 1473–1479.

Tuomisto H. 2010. A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography* 33: 2–22.

Valente LM, Savolainen V, Vargas P. 2010. Unparalleled rates of species diversification in Europe. *Proceedings of the Royal Society B: Biological Sciences* 277: 1489–1496.

Venables WN, Ripley BD. 2002. *Modern applied statistics with s*. New York, US: Springer.

Veron S, Kondratyeva A, Robuchon M, Grandcolas P, Govaerts R, Haevermans T, Pellens R, Mouchet M. 2021. High evolutionary and functional distinctiveness of endemic monocots in world islands. *Biodiversity and Conservation* 30: 3697–3715.

Větrovský T, Kohout P, Kopecký M, Machac A, Man M, Bahnmann BD, Brabcová V, Choi J, Meszárošová L, Human ZR, *et al*. 2019. A meta-analysis of global fungal distribution reveals climate-driven patterns. *Nature Communications* 10: 5142.

Wadoux AMJ-C, Heuvelink GBM, de Bruin S, Brus DJ. 2021. Spatial cross-validation is not the right way to evaluate map accuracy. *Ecological Modelling* 457: 109692.

Weigelt P, Daniel Kissling W, Kisel Y, Fritz SA, Karger DN, 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. 2020. GIFT – A global inventory of floras and traits for macroecology and biogeography. *Journal of Biogeography* 47: 16–43.

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

Whittaker RH. 1956. Vegetation of the Great Smoky Mountains. *Ecological Monographs* 26: 1–80.

Whittaker RH. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs* 30: 279–338.

Wiens JJ, Donoghue MJ. 2004. Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution* 19: 639–644.

Wiens JJ, Graham CH. 2005. Niche Conservatism: Integrating Evolution, Ecology, and Conservation Biology. *Annual Review of Ecology, Evolution, and Systematics* 36: 519–539.

Williams SE, Müller RD, Landgrebe TCW, Whittaker JM. 2012. An open-source software environment for visualizing and refining plate tectonic reconstructions using high-resolution geological and geophysical data sets. *GSA Today*: 4–9.

Willis KJ, Whittaker RJ. 2002. Species Diversity--Scale Matters. *Science* 295: 1245–1248.

Winter M, Devictor V, Schweiger O. 2013. Phylogenetic diversity and nature conservation: where are we? *Trends in Ecology & Evolution* 28: 199–204.

Wood SN. 2006. *Generalized additive models: an introduction with r*. Boca Raton, Florida: Chapman & Hall/CRC.

Wright MN, Ziegler A. 2017. ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software* 77: 1–17.

Wulff EW. 1935. Versuch einer Einteilung der Vegetation der Erde in pflanzengeographische Gebiete auf Grund der Artenzahl. *Repertorium Specierum Novarum Regni Vegetabilis* 12: 57–83.

Xing Y, Ree RH. 2017. Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. *Proceedings of the National Academy of Sciences* 114: E3444–E3451.

Xu W-B, Guo W-Y, Serra-Diaz JM, Schrodt F, Eiserhardt WL, Enquist BJ, Maitner BS, Merow C, Violle C, Anand M, *et al.* 2023. Global beta-diversity of angiosperm trees is shaped by Quaternary climate change. *Science Advances* 9: eadd8553.

Yoder AD, Nowak MD. 2006. Has vicariance or dispersal been the predominant biogeographic force in Madagascar? Only time will tell. *Annual Review of Ecology, Evolution, and Systematics* 37: 405–431.

Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG, McGlinn DJ, O'Meara BC, Moles AT, Reich PB, *et al.* 2014. Three keys to the radiation of angiosperms into freezing environments. *Nature* 506: 89–92.

Zhang L, Chen F, Zhang X, Li Z, Zhao Y, Lohaus R, Chang X, Dong W, Ho SYW, Liu X, *et al.* 2020. The water lily genome and the early evolution of flowering plants. *Nature* 577: 79–84.

Zhao M, Running SW. 2010. Drought-induced reduction in global terrestrial net primary production from 2000 through 2009. *Science* 329: 940–943.

Zomer RJ, Xu J, Trabucco A. 2022. Version 3 of the Global Aridity Index and Potential Evapotranspiration Database. *Scientific Data* 9: 409.

Acknowledgements

Appendix

Supporting Information for Chapter 1

Table S1.1 List of environmental predictor variables hypothesized to affect plant diversity patterns. Abbreviations in bold highlight predictor variables that were retained after the initial selection process to limit collinearity and used to fit different models of species richness and phylogenetic richness of vascular plants. ↑ represents effects on plant diversity hypothesized to be positive, while ↓ represents negative effects.

Category	Abbreviation	Variable description	Hypoth esized effect	Unit	Resolutio n	Reference
Geography	**Area**	Area of geographic regions	↑	km^2	-	(Weigelt *et al.*, 2020)
Geography	**SLMP**	Surrounding landmass proportion: Summed proportions of landmass area surrounding the target region within buffer distances of 100 km, 1000 km and 10,000 km	↑	-	-	(Weigelt *et al.*, 2013)
Climate	MAT	Mean annual temperature	↑	°C	30 arc-seconds	(Karger *et al.*, 2017)
Climate	PET	Potential evapotranspiration	↑	mm	30 arc-seconds	(Zomer *et al.*, 2008)
Climate	**LengthGrow**	Length of growing season (days): Number of days with temperatures exceeding a threshold of 0.9°C, without snow cover, and with sufficient soil water	↑	n	30 arc-seconds	(Karger *et al.*, 2019)
Climate	**MeanTempGr ow**	Mean temperature of growing season	↑	°C	30 arc-seconds	(Karger *et al.*, 2019)
Climate	MeanT_WetQ	Mean temperature of wettest quarter	↑	°C	30 arc-seconds	(Karger *et al.*, 2017)
Climate	IS	Isothermality, quantifying how large the diurnal variation in temperature is proportional to the annual variation: The ratio of the mean diurnal range divided by the annual temperature range, and then multiplied by 100	↑	-	30 arc-seconds	(Karger *et al.*, 2017)
Climate	**TS**	Temperature seasonality: standard deviation of mean monthly temperature *100	↓	°C	30 arc-seconds	(Karger *et al.*, 2017)
Climate	TAR	Temperature annual range	↓	°C	30 arc-seconds	(Karger *et al.*, 2017)
Climate	AP	Annual precipitation	↑	mm	30 arc-seconds	(Karger *et al.*, 2017)
Climate	**Wetdays**	Number of days per year with precipitation >0.1mm	↑	n	10 arc - minute	(New *et al.*, 2002)
Climate	**PrecipWarmQ uarter**	Precipitation of warmest quarter	↑	mm	30 arc-seconds	(Karger *et al.*, 2017)
Climate	PrecipGrow	Precipitation of growing season	↑	mm	30 arc-seconds	(Karger *et al.*, 2019)
Climate	**PS**	Precipitation seasonality: Ratio of the standard deviation of the monthly total precipitation to the mean monthly total precipitation (coefficient of variation in monthly precipitation)	↓	-	30 arc-seconds	(Karger *et al.*, 2017)
climate	**GPP**	Gross primary productivity	↑	gcarb on/m²	30 arc-seconds	(Zhao & Running, 2010)
Heterogeneity	**Elev_range**	Elevational range: absolute difference between highest and lowest elevation within a given area	↑	m	30 arc-seconds	(Danielson & Gesch, 2011)
Heterogeneity	**Soildiv**	Number of different soil types based on World Reference Base classification system	↑	n	30 arc-seconds	(Hengl *et al.*, 2017)
Heterogeneity	Homogeneity	Homogeneity: similarity of MODIS enhanced vegetation index between adjacent pixels	↓	-	30 arc-seconds	(Tuanmu & Jetz, 2015)
Past environments	**VelocityTemp_ LGM**	Climate change velocity in temperature for the last 21,000 years (since the Last Glacial Maximum) calculated as the ratio between temporal change and contemporary spatial change in temperature	↓	m/yr	30 arc-seconds	(Hijmans *et al.*, 2005; Weigelt *et al.*, 2013)
Past environments	**TempStability _LGM**	Temperature stability for the last 21,000 years (since the Last Glacial Maximum): Mean standard deviation in mean annual temperature between time slices of 1000 years each over the last 21,000 years; "stability" is defined as the inverse of this deviation re-scaled here between 0 and 1.	↑	-	2.5 degrees	(Owens & Guralnick, 2019)
Past environments	**TempAnomaly _midPliocene**	Difference for mean annual temperature between the mid-Pliocene warm period (c. 3.264-3.025 Ma) and the present-day, i.e. the value of past minus present.	↓	°C	2.5 arc-minutes	(Hill, 2015; Brown *et al.*, 2018)
Past environments	Biome_Miocen e	Euclidean distance of biome area changes across four time periods (present, Last Glacial Maximum, Pliocene and Middle Miocene)	↓	-	-	(Olson *et al.*, 2001; Ray & Adams, 2001; Henrot *et al.*, 2010; Dowsett *et al.*, 2016)
Past environments	**Biome_Pliocen e**	Euclidean distance of biome area changes across three time periods (present, Last Glacial Maximum and Pliocene)	↓	-	-	(Olson *et al.*, 2001; Ray & Adams, 2001; Dowsett *et al.*, 2016)
Past environments	Kingdom	Floristic kingdoms: Antarctic kingdom, Australis kingdom, Cape kingdom, Holarctic kingdom, Neotropic kingdom, Paleotropic kingdom.	-	factor	-	(Takhtajan, 1986)

Table S1.2 Coefficients of a linear model between the residuals (deviation) from the linear regression between species richness and phylogenetic richness, and the fifteen predictor variables identified to best explain plant diversity. Residuals obtained from the model: phylogenetic richness = 16.87 * species richness, R^2= 0.96, p < 0.0001. Negative residuals indicate lower phylogenetic richness than expected based on species richness.

Predictors	Estimates	Confidence intervals	P value
(Intercept)	-0.14	0.62 – 0.34	0.573
Area	0.01	0.03 – 0.06	0.571
Surrounding landmass proportion	-0.11	-0.17 – -0.05	0.001
Length of growing season	0.14	0.05 – 0.22	0.002
Mean temperature of growing season	0.19	0.08 – 0.30	0.001
Temperature seasonality	-0.13	-0.25 – -0.01	0.028
Number of wet days	0.07	0.02 – 0.16	0.132
Precipitation of warmest quarter	0.17	0.09 – 0.26	<0.001
Precipitation seasonality	0.07	0.01 – 0.14	0.076
Gross primary productivity	0.02	0.07 – 0.11	0.739
Number of soil types	-0.29	-0.39 – -0.19	<0.001
Elevational range	0.11	0.01 – 0.20	0.026
Temperature change velocity since the LGM	0.17	0.08 – 0.26	<0.001
Temperature stability since the LGM	0.18	0.08 – 0.28	0.001
Temperature anomaly since the mid-Pliocene	0.03	0.04 – 0.09	0.413
Biome area changes	0.04	-0.03 – 0.11	0.302
Observations	830		
R^2 / R^2 adjusted	0.385 / 0.373		

Table S1.3 Homogenization of biome classifications for current maps, Last Glacial Maximum (LGM), Pliocene (mid-Piacenzian) and Middle Miocene.

New biome type	Current	LGM	Pliocene (mid-Piacenzian)	Middle Miocene
Tropical forest	Tropical & subtropical moist broadleaf forests	Tropical rainforest	Tropical forest	Tropical rainforest
	Tropical & subtropical dry broadleaf forests	Tropical woodland		Sub-tropical forest
		Monsoon or dry forest		Tropical seasonal forest
		Tropical thorn scrub and scrub woodland		
	Tropical & subtropical coniferous forests	Montane tropical forest		
	Mangroves			
Temperate forest	Temperate broadleaf & mixed forests	Broadleaved temperate evergreen forest	Warm temperate forest	Temperate broadleaved deciduous forest
	Mediterranean forests, woodlands & scrub	Semi-arid temperate woodland or scrub		Warm temperate open woodland
				Warm temperate mixed forest
				Warm temperate broadleaved evergreen forest
	Temperate conifer forests		Temperate forest	Cool temperate conifer forest
				Cool temperate mixed forest
Boreal forest	Boreal forests/taiga	Open boreal woodlands	Boreal forest	Boreal/montane forest
		Main Taiga		Cold temperate/boreal open woodland
Savanna and grassland	Tropical & subtropical grasslands, savannas & shrublands	Tropical grassland	Savanna &dry woodland	Tropical savanna
		Savanna	Grassland &dry scrubland	Tropical grassland
	Temperate grasslands, savannas & shrublands	Temperate steppe grassland		Temperate grassland
		Forest steppe		
		Dry steppe		
	Flooded grasslands & savannas			
	Montane grasslands & shrublands	Alpine tundra		
		Montane Mosaic		
		Subalpine parkland		
Tundra	Tundra	Tundra	Tundra	Tundra
		Steppe-tundra	Dry tundra	
		Polar and alpine desert		
Deserts	Deserts & xeric shrublands	Tropical semi-desert	Desert	Desert
		Tropical extreme desert		Semi-desert
		Temperate desert		
		Temperate semi-desert		
Not vegetated	Lake	Lakes and open water		
	Rock and ice	Ice sheet and other permanent ice	Ice	Ice

Table S1.4 Model assessment results. Each model was assessed for its fit based on training data, and for its predictive performance based on all out-of-bag samples using both random 10-fold cross-validation and spatial 68-fold cross-validation. Accuracy statistics are provided on log-scale. Values shown are: root mean squared error (RMSE); the coefficient of determination of a linear model of predicted vs. observed richness (R2_CORR); the amount of variation explained by the model calculated as one minus the ratio of the sum of the squared error between observation and prediction to the total sum of squares (R2_Accuracy). Values in gray are showed in Table 1. Non-spatial models were fitted with 15 predictors except Minimum adequate generalized linear model (GLM), Full GLM with interaction terms and GLM simplified with interaction terms, while spatial models also contained spatial effects [i.e. simultaneous autoregressive (SAR) models, generalized additive models (GAMs) included the spline of geographic coordinates and machine learning methods included cubic polynomial trend surfaces]. Minimum adequate GLM for species richness contained 9 predictors (i.e. Area, LengthGrow, MeanTempGrow, GPP, TS, PS, Elev_range, Soildiv, Biome_Pliocene); Minimum adequate GLM for phylogenetic richness contained 13 predictors (PS and wetdays removed). Full GLM with interaction terms was fitted with all 15 environmental predictors and interactions between area and each individual predictor, and between energy and environmental heterogeneity (i.e. MeanTempGrow : Soildiv; MeanTempGrow : Elev_range), as well as energy and water (i.e. MeanTempGrow : PrecipWarmQuarter; MeanTempGrow : Wetdays). Simplified GLM with interaction terms for species richness contained 13 individual predictors (PS and TempStability_LGM removed) and 9 interaction terms between area and other predictors, and between energy and environmental heterogeneity; simplified GLM with interaction terms for phylogenetic richness contained 14 main predictors (PS removed) and 13 interaction terms between area and other predictors, between energy and environmental heterogeneity, as well as energy and water. GLM (Kreft and Jetz, 2007) was a combined six-predictor model as in Kreft & Jetz (2007). SMOOTH Model (Keil & Chase, 2019) was fitted using the same model structure as Keil and Chase's smooth model (Keil & Chase, 2019), which contained a two-dimensional spline on geographical coordination, 15 individual predictors and the interactions between each individual predictor and area. Likewise, REALM Model was fitted using the same model structure as Keil and Chase's realm model (Keil & Chase, 2019), which contained floristic kingdom, 15 individual predictors and the interactions between each individual predictor and area. Extreme gradient boosting (XGBoost) with kingdom and random forests with kingdom were fitted with the same 15 predictors and floristic kingdom. Area = area of the region; SLMP = Surrounding landmass proportion; LengthGrow = Length of growing season; MeanTempGrow = Mean temperature of growing season; TS = Temperature seasonality; Wetdays = Number of wet days; PrecipWarmQuarter = Precipitation of warmest quarter; PS = Precipitation seasonality; GPP = Gross primary productivity; Soildiv = Number of soil types; Elev_range = Elevational range; VelocityTemp_LGM = Temperature change velocity since the LGM; TempStability_LGM = Temperature stability since the LGM; TempAnomaly_midPliocene = Temperature anomaly since the mid-Pliocene; Biome_Pliocene = Biome area changes.

Models	Species richness									Phylogenetic richness								
	Model fits			Random cross-validation			Spatial cross-validation			Model fits			Random cross-validation			Spatial cross-validation		
	RMSE	R²_CORR	R²_Accuracy	RMSE	R²_CORR	R²_Accuracy	RMSE	R²_CORR	R²_Accuracy	RMSE	R²_CORR	R²_Accuracy	RMSE	R²_CORR	R²_Accuracy	RMSE	R²_CORR	R²_Accuracy
Non-spatial models																		
Full GLM	0.510	0.657	0.657	0.525	0.636	0.636	0.582	0.568	0.561	0.506	0.639	0.471	0.514	0.630	0.452	0.552	0.594	0.359
Minimum adequate GLM	0.511	0.657	0.657	0.520	0.643	0.643	0.548	0.609	0.608	0.506	0.640	0.470	0.513	0.631	0.454	0.548	0.600	0.369
Full GLM with interaction terms	0.457	0.722	0.722	0.478	0.695	0.694	0.535	0.627	0.620	0.397	0.697	0.661	0.419	0.668	0.623	0.467	0.616	0.531
Simplified GLM with interaction terms	0.457	0.722	0.722	0.471	0.704	0.704	0.502	0.665	0.664	0.396	0.697	0.663	0.412	0.676	0.635	0.453	0.630	0.559

Model																	
GAM	0.410	0.775	0.775	0.437	0.743	0.742	0.507	0.664	0.658	0.328	0.778	0.769	0.359	0.735	0.723	0.430	0.647	0.604
Random forests	0.403	0.786	0.775	0.415	0.774	0.761	0.511	0.642	0.639	0.309	0.807	0.795	0.317	0.796	0.784	0.395	0.669	0.667
XGBoost	0.107	0.985	0.984	0.389	0.791	0.791	0.487	0.673	0.673	0.084	0.986	0.985	0.295	0.813	0.813	0.384	0.687	0.685
Neural networks	0.285	0.887	0.887	0.451	0.725	0.718	0.604	0.552	0.496	0.244	0.872	0.872	0.328	0.774	0.769	0.419	0.650	0.628
Spatial models																		
SAR	0.385	0.795	0.794	0.537	0.601	0.600	0.548	0.586	0.584	0.295	0.813	0.813	0.416	0.631	0.629	0.426	0.614	0.611
GAM	0.383	0.802	0.802	0.413	0.769	0.769	0.499	0.672	0.667	0.312	0.794	0.791	0.340	0.757	0.751	0.416	0.663	0.633
Random forests	0.384	0.810	0.795	0.398	0.796	0.780	0.502	0.660	0.653	0.292	0.828	0.816	0.303	0.815	0.803	0.379	0.697	0.694
XGBoost	0.096	0.988	0.987	0.371	0.809	0.809	0.463	0.703	0.703	0.081	0.986	0.986	0.279	0.833	0.833	0.351	0.737	0.737
Neural networks	0.253	0.911	0.911	0.422	0.756	0.753	0.587	0.590	0.522	0.203	0.911	0.911	0.314	0.792	0.789	0.433	0.645	0.597
Other models																		
GLM (Kreft and Jezt, 2007)	0.452	0.716	0.716	0.457	0.708	0.708	0.494	0.663	0.659									
SMOOTH Model (Keil & Chase, 2019)	0.394	0.791	0.791	0.421	0.761	0.761	0.487	0.686	0.683									
REALM Model (Keil & Chase, 2019)	0.415	0.770	0.770	0.445	0.734	0.733	0.494	0.679	0.673									
XGBoost with kingdom	0.107	0.985	0.984	0.388	0.791	0.791	0.472	0.693	0.692	0.138	0.960	0.959	0.290	0.820	0.820	0.364	0.720	0.718
Random forests with kingdom	0.398	0.792	0.780	0.414	0.775	0.763	0.527	0.632	0.626	0.305	0.809	0.800	0.316	0.795	0.785	0.407	0.654	0.654
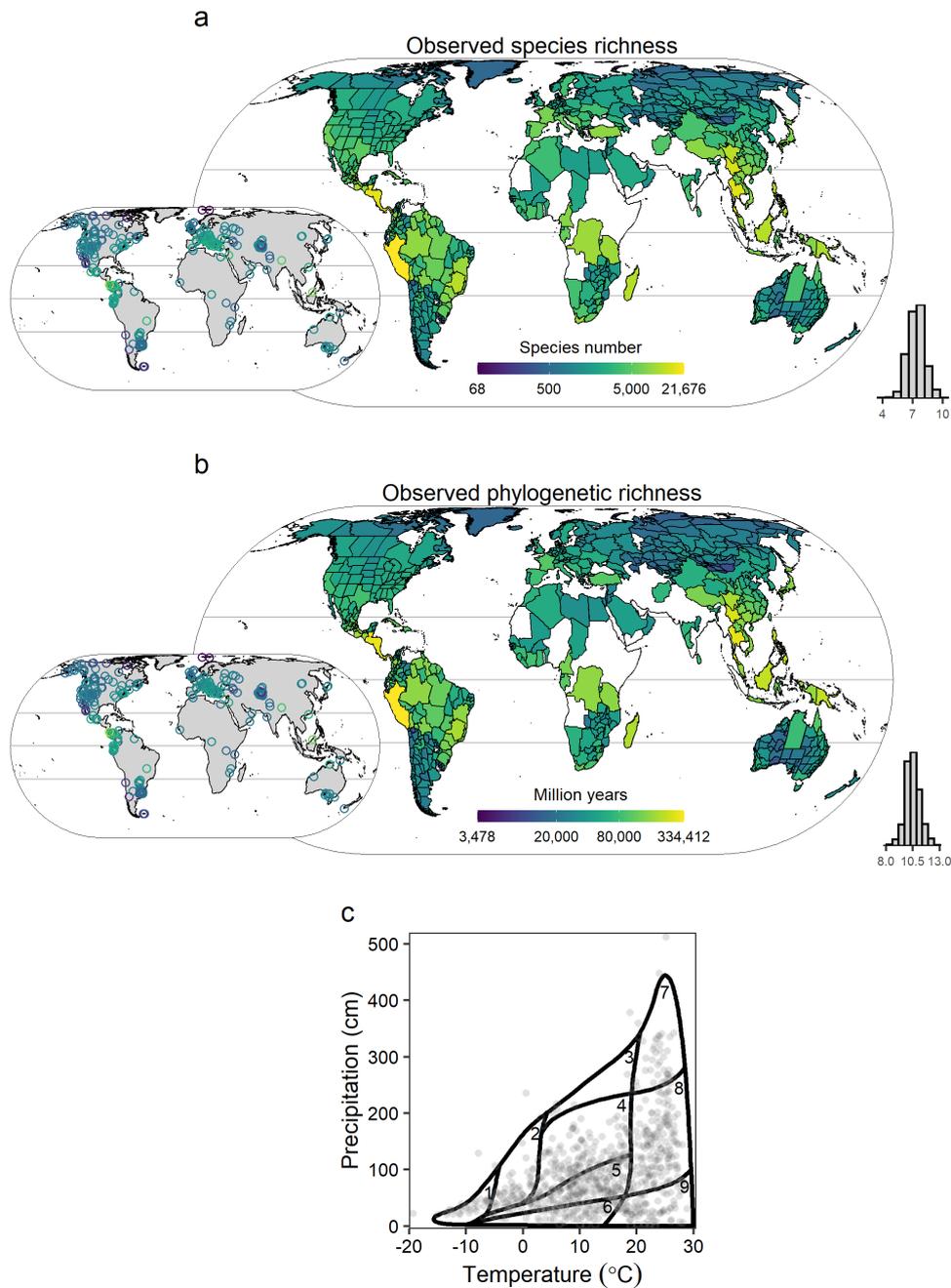
Figure S1.1 Observed species and phylogenetic richness of vascular plants for 830 geographic regions used to train the models. a, species richness. b, phylogenetic richness. Embedded maps show observed species richness (a) and phylogenetic richness (b) for overlapping small regions and regions <10000 km^2. The histograms show the frequency distribution of log species richness (a) and log phylogenetic richness (b). c, Regions plotted onto Whittaker's scheme of biomes delineated based on mean annual temperature and annual precipitation (Whittaker, 1975). Whittaker biomes are numbered as follows: 1 = tundra , 2 = boreal forest, 3 = temperate rainforest , 4 = temperate seasonal forest , 5 = woodland/shrubland , 6 = temperate grassland, 7 = tropical rainforest, 8 = tropical seasonal forest, 9 = subtropical desert.
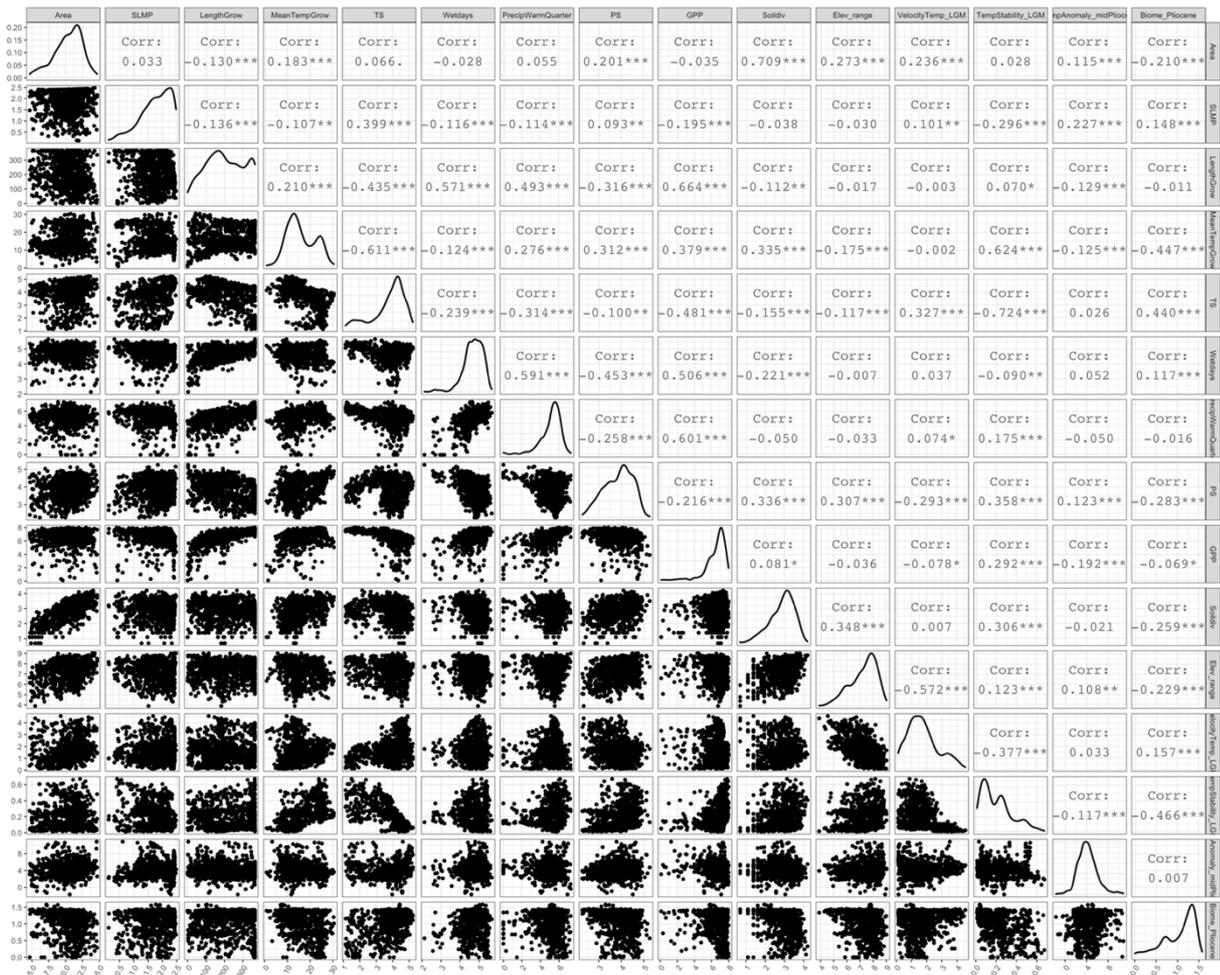
Figure S1.2 Correlations among all predictors and their density distributions. Numbers are Pearson correlation coefficients. Some predictors (i.e. Area, TS, Wetdays, PrecipWarmQuarter, PS, GPP, Soildiv, Elev_range, VelocityTemp_LGM) are shown in log-scale as they were log-transformed for generalized linear models owing to their skewed distributions. Area = area of the region; SLMP = Surrounding landmass proportion; LengthGrow = Length of growing season; MeanTempGrow = Mean temperature of growing season; TS = Temperature seasonality; Wetdays = Number of wet days; PrecipWarmQuarter = Precipitation of warmest quarter; PS = Precipitation seasonality; GPP = Gross primary productivity; Soildiv = Number of soil types; Elev_range = Elevational range; VelocityTemp_LGM = Temperature change velocity since the LGM; TempStability_LGM = Temperature stability since the LGM; TempAnomaly_midPliocene = Temperature anomaly since the mid-Pliocene; Biome_Pliocene = Biome area changes.
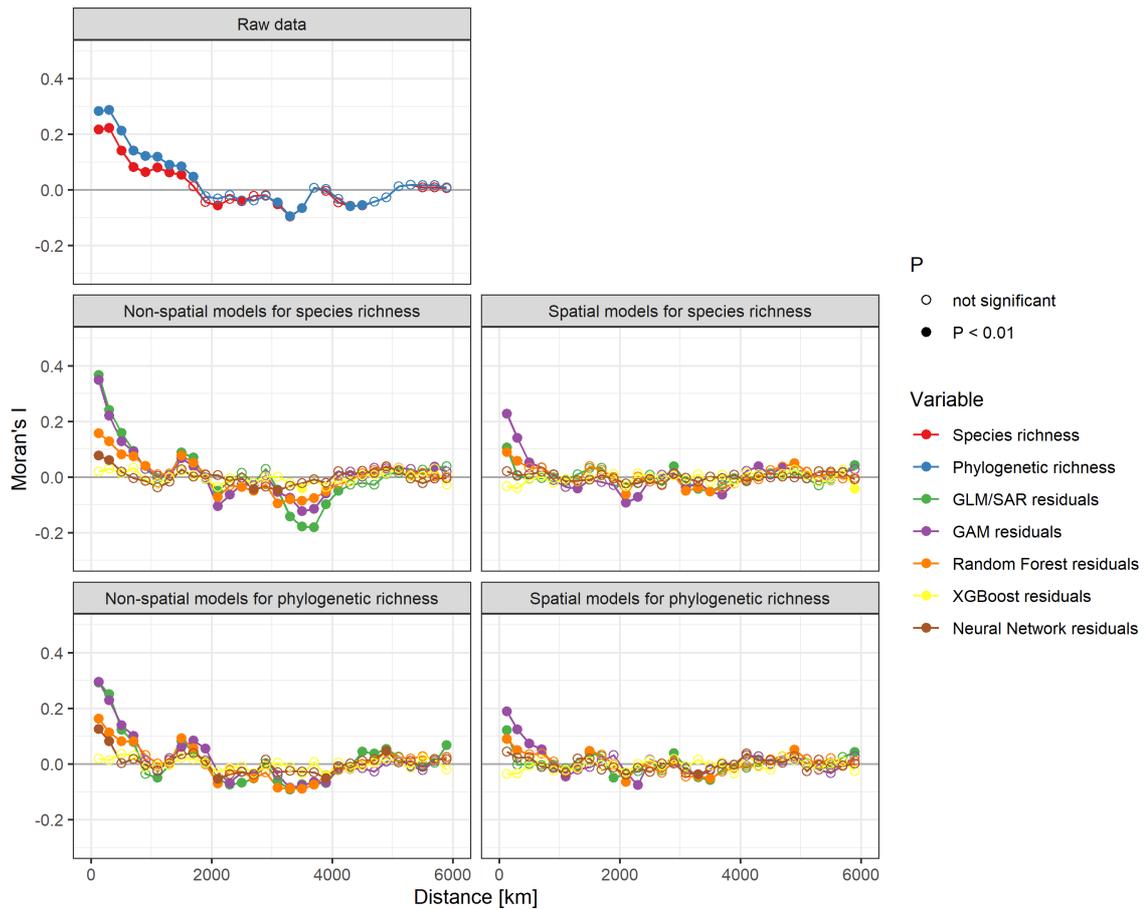
Figure S1.3 Spatial correlograms of raw diversity data, and residuals from non-spatial and spatial models, respectively, fitted for species richness and phylogenetic richness. Full symbols indicate a significant Moran's I correlation at a given lag distance (P < 0.01).
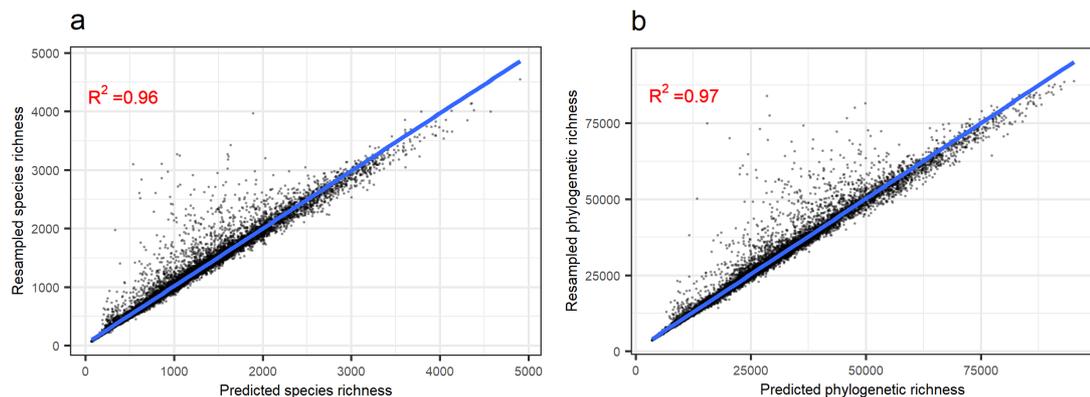
Figure S1.4 Comparison between ensemble predictions of vascular plant diversity across an equal area grid of 7,774 km² hexagons and a raster layer of resampled ensemble predictions at 30 arc second resolution. See Methods S1.5 for details on the resampling procedure. For comparison, the resampled ensemble predictions of (a) species richness and (b) phylogenetic richness were extracted and averaged for each hexagon polygon grid cell.
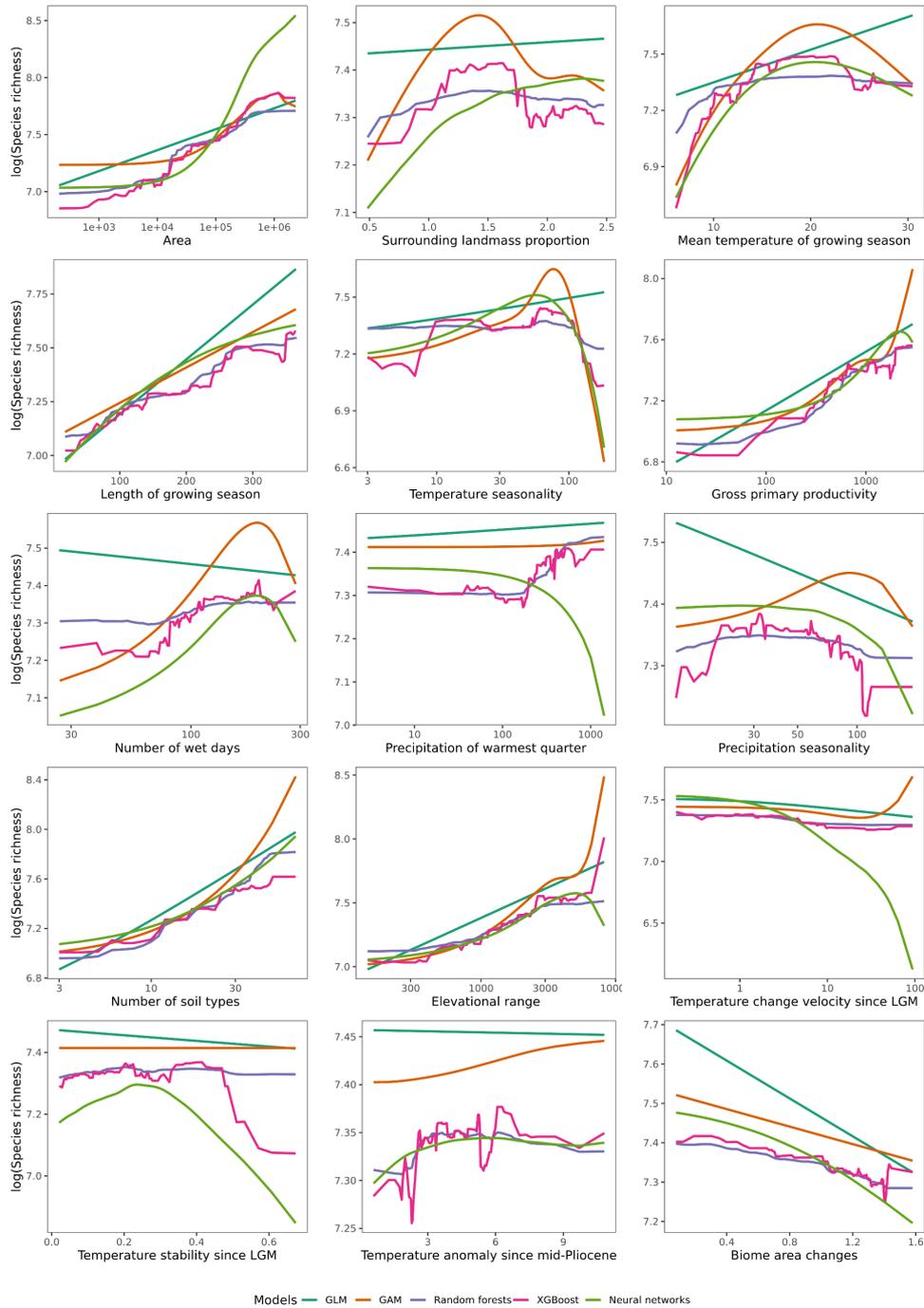
Figure S1.5 Estimated effects of predictor variables on species richness of vascular plants across five non-spatial models (partial dependence plots). Five non-spatial models were fitted with 15 predictors. Predictors shown here were used for both non-spatial and spatial models, and were selected to represent the major hypotheses related to plant diversity–environment relationships and filtered based on their contribution to model performance and collinearity.
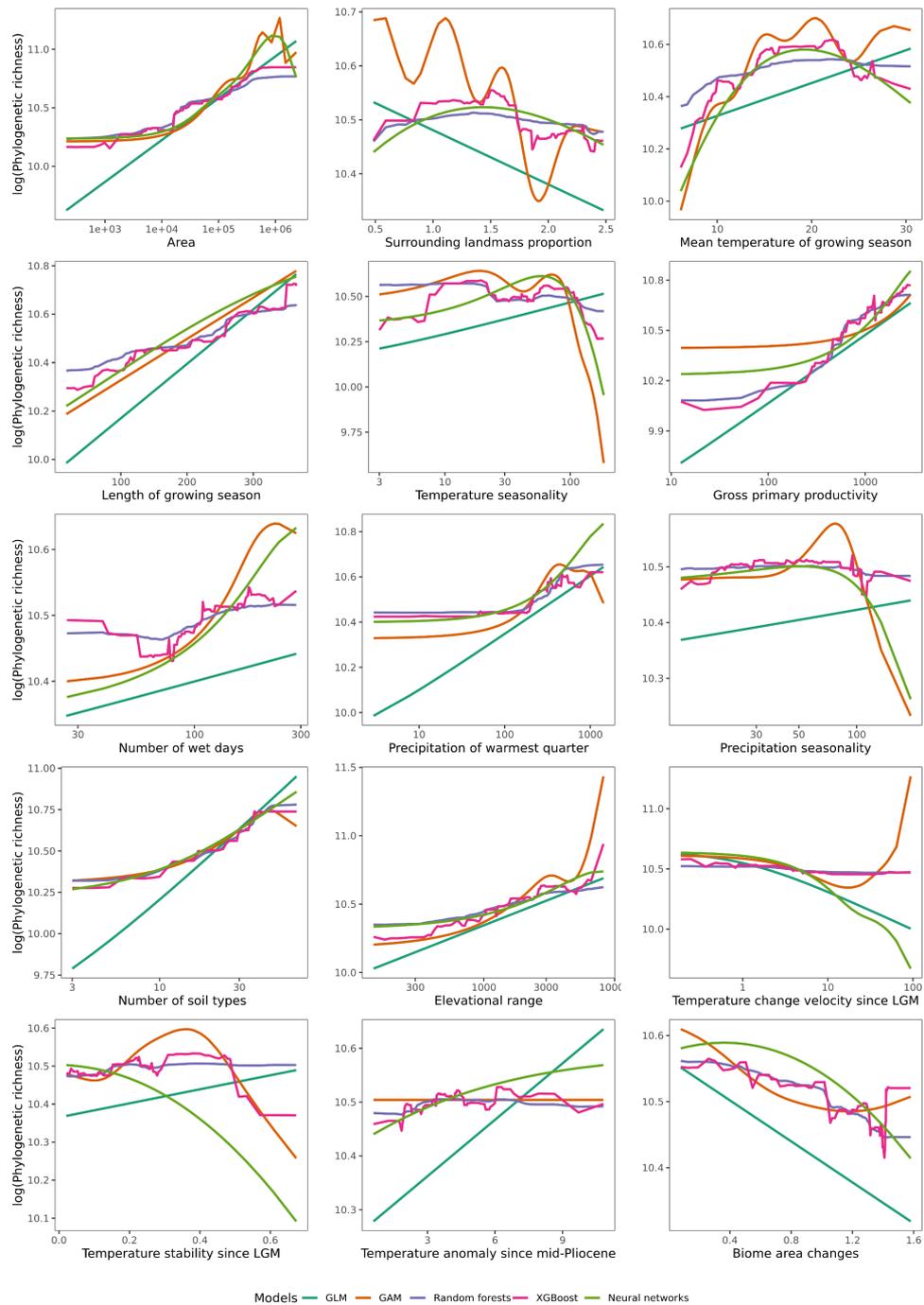
Figure S1.6 Estimated effects of predictor variables on phylogenetic richness of vascular plants across five non-spatial models (the partial dependence plots). Five non-spatial models were fitted with 15 predictors. Predictors shown here were used for both non-spatial and spatial models, and were selected to represent the major hypotheses related to plant diversity–environment relationships and filtered based on their contribution to model performance and collinearity.
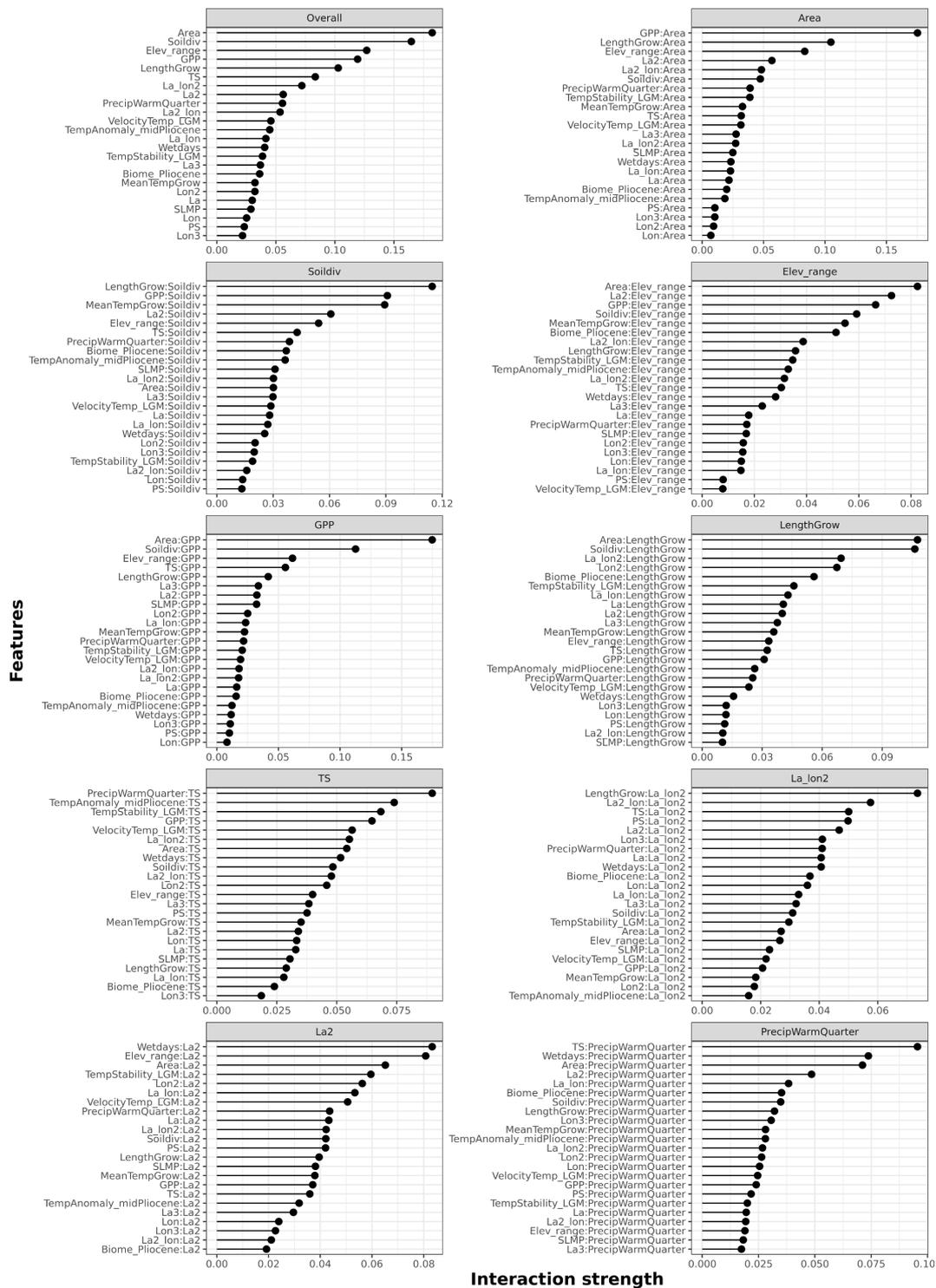
Figure S1.7 Interaction strength of each predictor variable for explaining species richness (Overall) in the spatial random forest model and two-way interaction strengths between the nine top-ranked covariates and all other covariates. Terms of cubic polynomial trend surfaces [i.e. latitude (Y), centered longitude (X) as well as X^2, XY, Y^2, X^3, X^2Y, XY^2 and Y^3] are "La", "Lon", "La2", "La_lon", "Lon2", "La3", "La2_lon", "La_lon2", and "Lon3", respectively. Other predictors are abbreviated as follows: Area = area of the region; SLMP = Surrounding landmass proportion; LengthGrow = Length of growing season; MeanTempGrow = Mean temperature of growing season; TS = Temperature seasonality; Wetdays = Number of wet days; PrecipWarmQuarter = Precipitation of warmest quarter; PS = Precipitation seasonality; GPP = Gross primary productivity; Soildiv = Number of soil types; Elev_range = Elevational range; VelocityTemp_LGM = Temperature change velocity since the LGM; TempStability_LGM = Temperature stability since the LGM; TempAnomaly_midPliocene = Temperature anomaly since the mid-Pliocene; Biome_Pliocene = Biome area changes.

Figure S1.8 Estimated effects of the nine two-way interactions (two-predictors partial dependence plots) in the spatial random forest model for species richness.
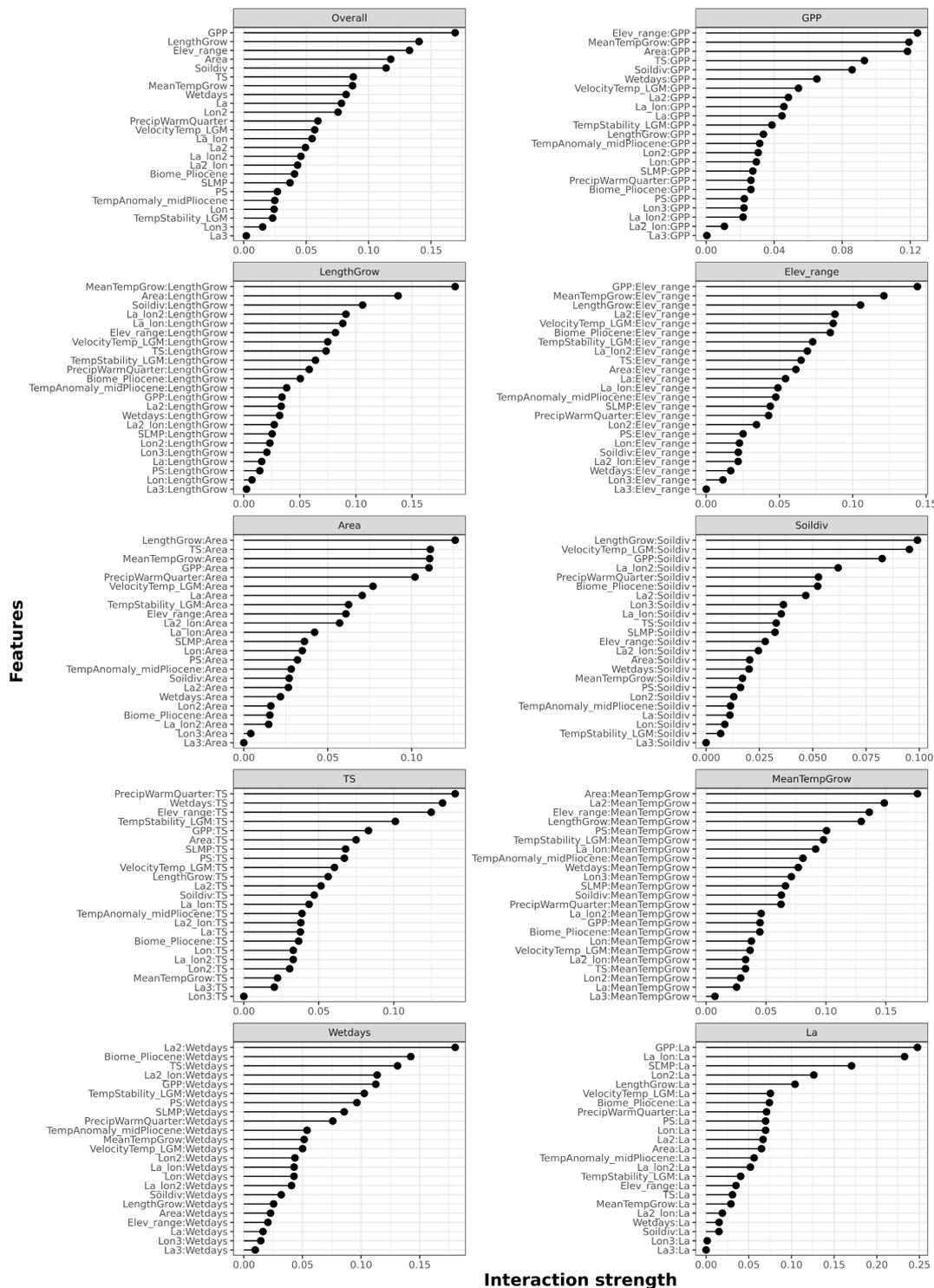
Figure S1.9 Interaction strength of each predictor variable for explaining species richness (Overall) in the spatial extreme gradient boosting model and two-way interaction strengths between the nine top-ranked covariates and all other covariates. Terms of cubic polynomial trend surfaces [i.e. latitude (Y), centered longitude (X) as well as X^2, XY, Y^2, X^3, X^2Y, XY^2 and Y^3] are "La", "Lon", "La2", "La_lon", "Lon2", "La3", "La2_lon", "La_lon2", and "Lon3", respectively. Other predictors are abbreviated as follows: Area = area of the region; SLMP = Surrounding landmass proportion; LengthGrow = Length of growing season; MeanTempGrow = Mean temperature of growing season; TS = Temperature seasonality; Wetdays = Number of wet days; PrecipWarmQuarter = Precipitation of warmest quarter; PS = Precipitation seasonality; GPP = Gross primary productivity; Soildiv = Number of soil types; Elev_range = Elevational range; VelocityTemp_LGM = Temperature change velocity since the LGM; TempStability_LGM = Temperature stability since the LGM; TempAnomaly_midPliocene = Temperature anomaly since the mid-Pliocene; Biome_Pliocene = Biome area changes.

Figure S1.10 Estimated effects of the nine two-way interactions (two-predictors partial dependence plots) in the spatial extreme gradient boosting model for species richness.
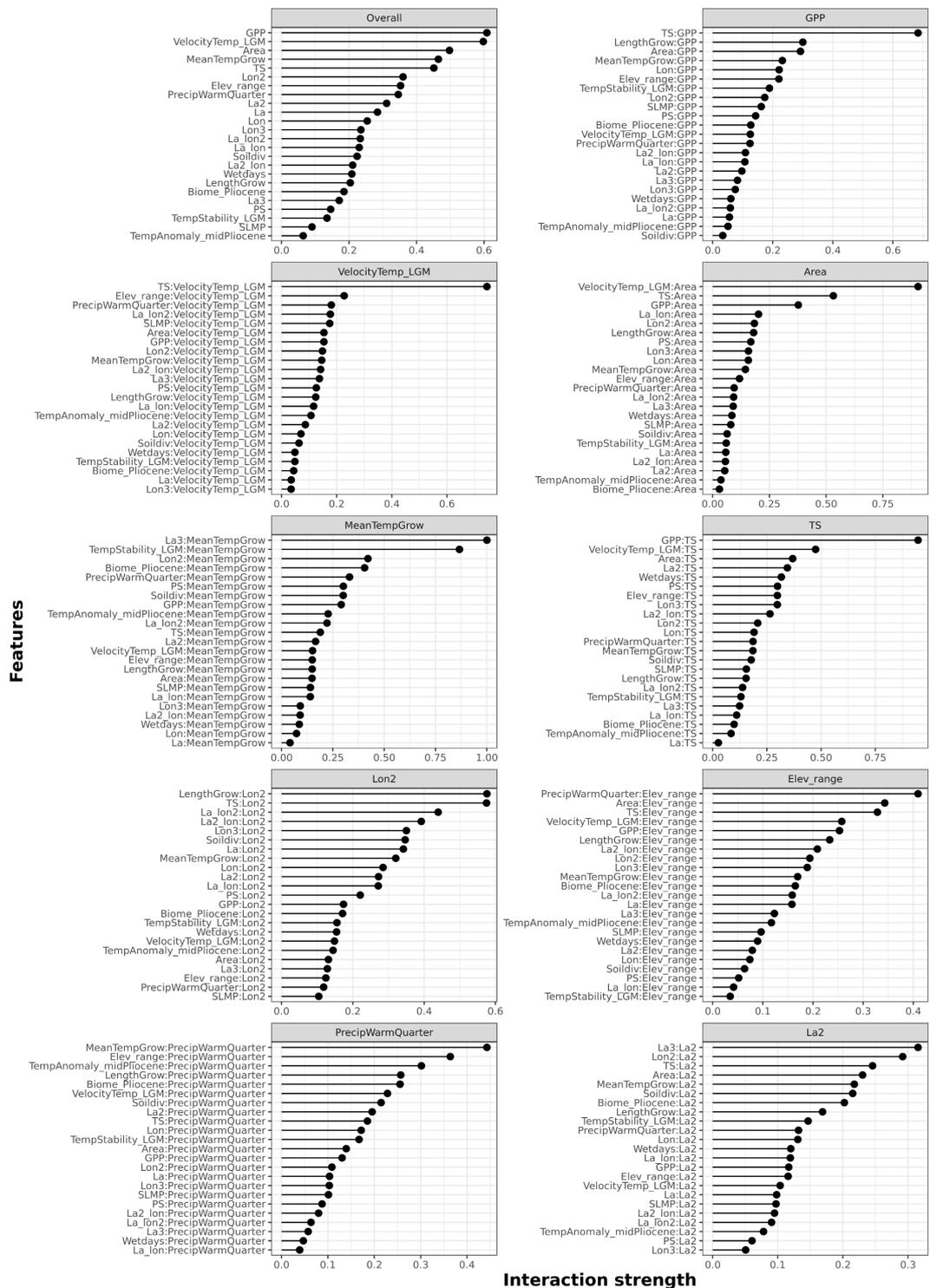
Figure S1.11 Interaction strength of each predictor variable for explaining species richness (Overall) in the spatial neural network model and two-way interaction strengths between the nine top-ranked covariates and all other covariates. Terms of cubic polynomial trend surfaces [i.e. latitude (Y), centered longitude (X) as well as X^2, XY, Y^2, X^3, X^2Y, XY^2 and Y^3] are "La", "Lon", "La2", "La_lon", "Lon2", "La3", "La2_lon", "La_lon2", and "Lon3", respectively. Other predictors are abbreviated as follows: Area = area of the region; SLMP = Surrounding landmass proportion; LengthGrow = Length of growing season; MeanTempGrow = Mean temperature of growing season; TS = Temperature seasonality; Wetdays = Number of wet days; PrecipWarmQuarter = Precipitation of warmest quarter; PS = Precipitation seasonality; GPP = Gross primary productivity; Soildiv = Number of soil types; Elev_range = Elevational range; VelocityTemp_LGM = Temperature change velocity since the LGM; TempStability_LGM = Temperature stability since the LGM; TempAnomaly_midPliocene = Temperature anomaly since the mid-Pliocene; Biome_Pliocene = Biome area changes.

Figure S1.12 Estimated effects of the nine two-way interactions (two-predictors partial dependence plots) in the spatial neural network model for species richness.
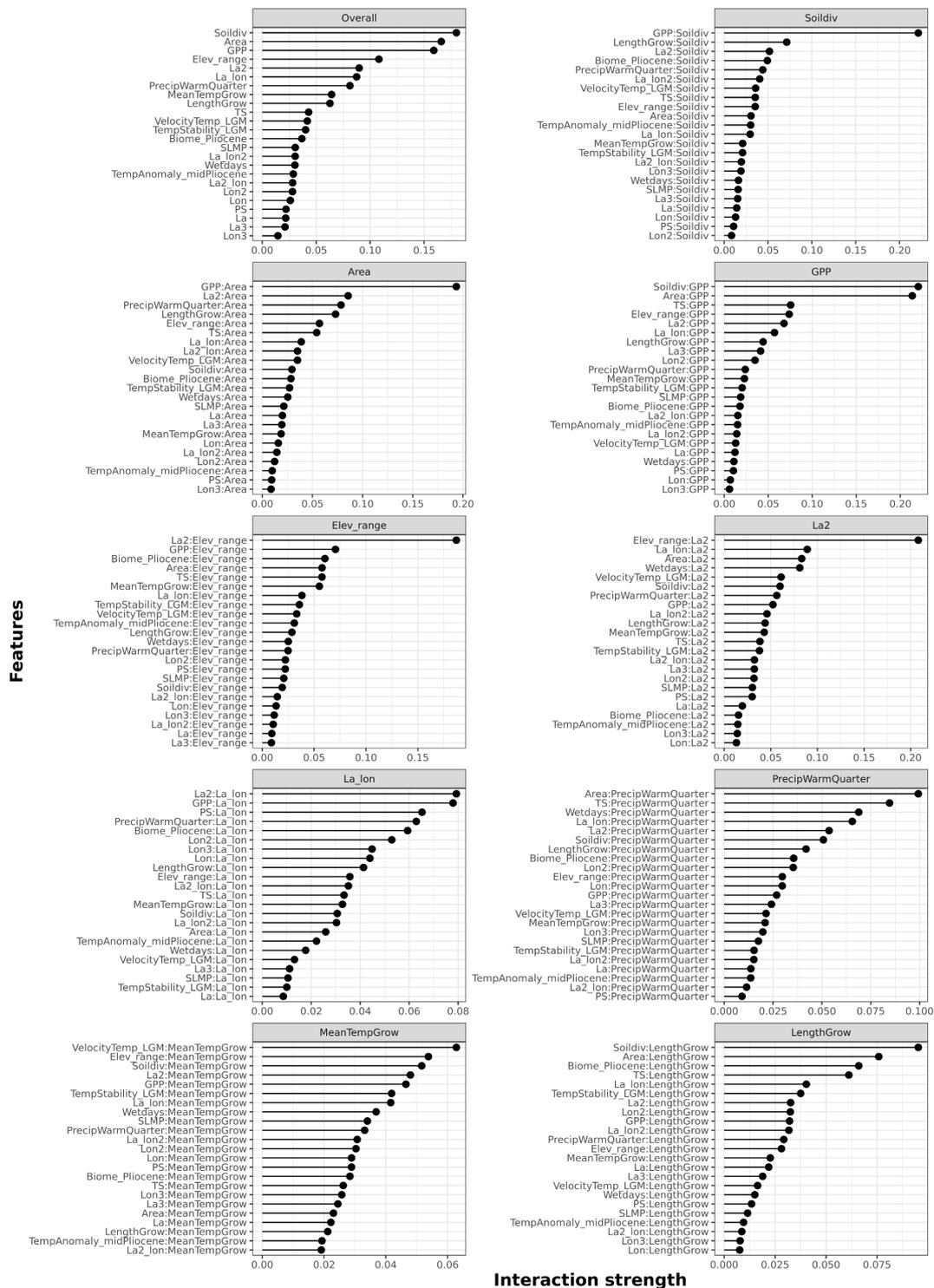
Figure S1.13 Interaction strength of each predictor variable for explaining phylogenetic richness (Overall) in the spatial random forest model and two-way interaction strengths between the nine top-ranked covariates and all other covariates. Terms of cubic polynomial trend surfaces [i.e. latitude (Y), centered longitude (X) as well as X^2, XY, Y^2, X^3, X^2Y, XY^2 and Y^3] are "La", "Lon", "La2", "La_lon", "Lon2", "La3", "La2_lon", "La_lon2", and "Lon3", respectively. Other predictors are abbreviated as follows: Area = area of the region; SLMP = Surrounding landmass proportion; LengthGrow = Length of growing season; MeanTempGrow = Mean temperature of growing season; TS = Temperature seasonality; Wetdays = Number of wet days; PrecipWarmQuarter = Precipitation of warmest quarter; PS = Precipitation seasonality; GPP = Gross primary productivity; Soildiv = Number of soil types; Elev_range = Elevational range; VelocityTemp_LGM = Temperature change velocity since the LGM; TempStability_LGM = Temperature stability since the LGM; TempAnomaly_midPliocene = Temperature anomaly since the mid-Pliocene; Biome_Pliocene = Biome area changes.

Figure S1.14 Estimated effects of the nine two-way interactions (two-predictors partial dependence plots) in the spatial random forest model for phylogenetic richness.
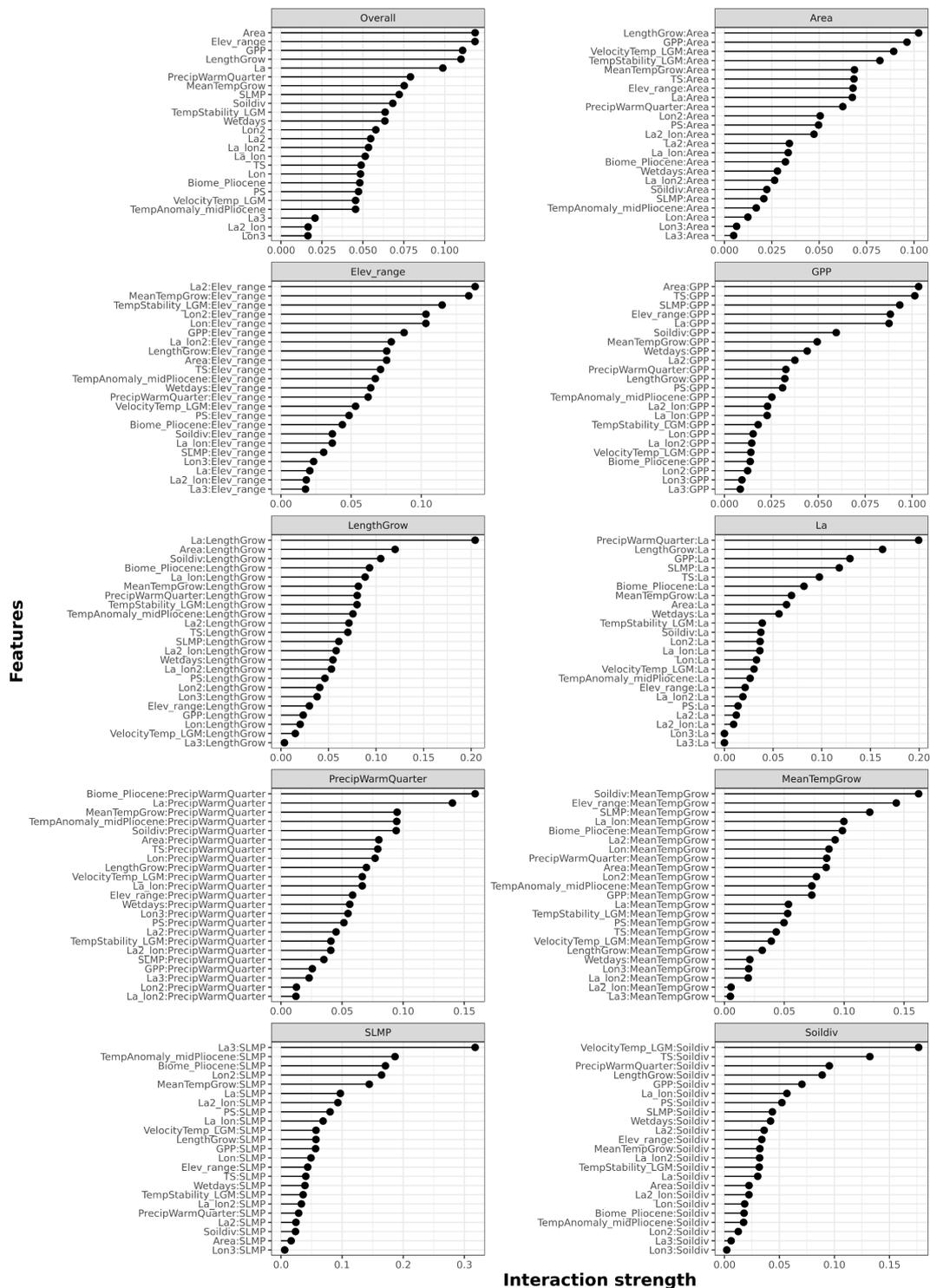
Figure S1.15 Interaction strength of each predictor variable for explaining phylogenetic richness (Overall) in the spatial extreme gradient boosting model and two-way interaction strengths between the nine top-ranked covariates and all other covariates. Terms of cubic polynomial trend surfaces [i.e. latitude (Y), centered longitude (X) as well as X^2, XY, Y^2, X^3, X^2Y, XY^2 and Y^3] are "La", "Lon", "La2", "La_lon", "Lon2", "La3", "La2_lon", "La_lon2", and "Lon3", respectively. Other predictors are abbreviated as follows: Area = area of the region; SLMP = Surrounding landmass proportion; LengthGrow = Length of growing season; MeanTempGrow = Mean temperature of growing season; TS = Temperature seasonality; Wetdays = Number of wet days; PrecipWarmQuarter = Precipitation of warmest quarter; PS = Precipitation seasonality; GPP = Gross primary productivity; Soildiv = Number of soil types; Elev_range = Elevational range; VelocityTemp_LGM = Temperature change velocity since the LGM; TempStability_LGM = Temperature stability since the LGM; TempAnomaly_midPliocene = Temperature anomaly since the mid-Pliocene; Biome_Pliocene = Biome area changes.

Figure S1.16 Estimated effects of the nine two-way interactions (two-predictors partial dependence plots) in the spatial extreme gradient boosting model for phylogenetic richness.
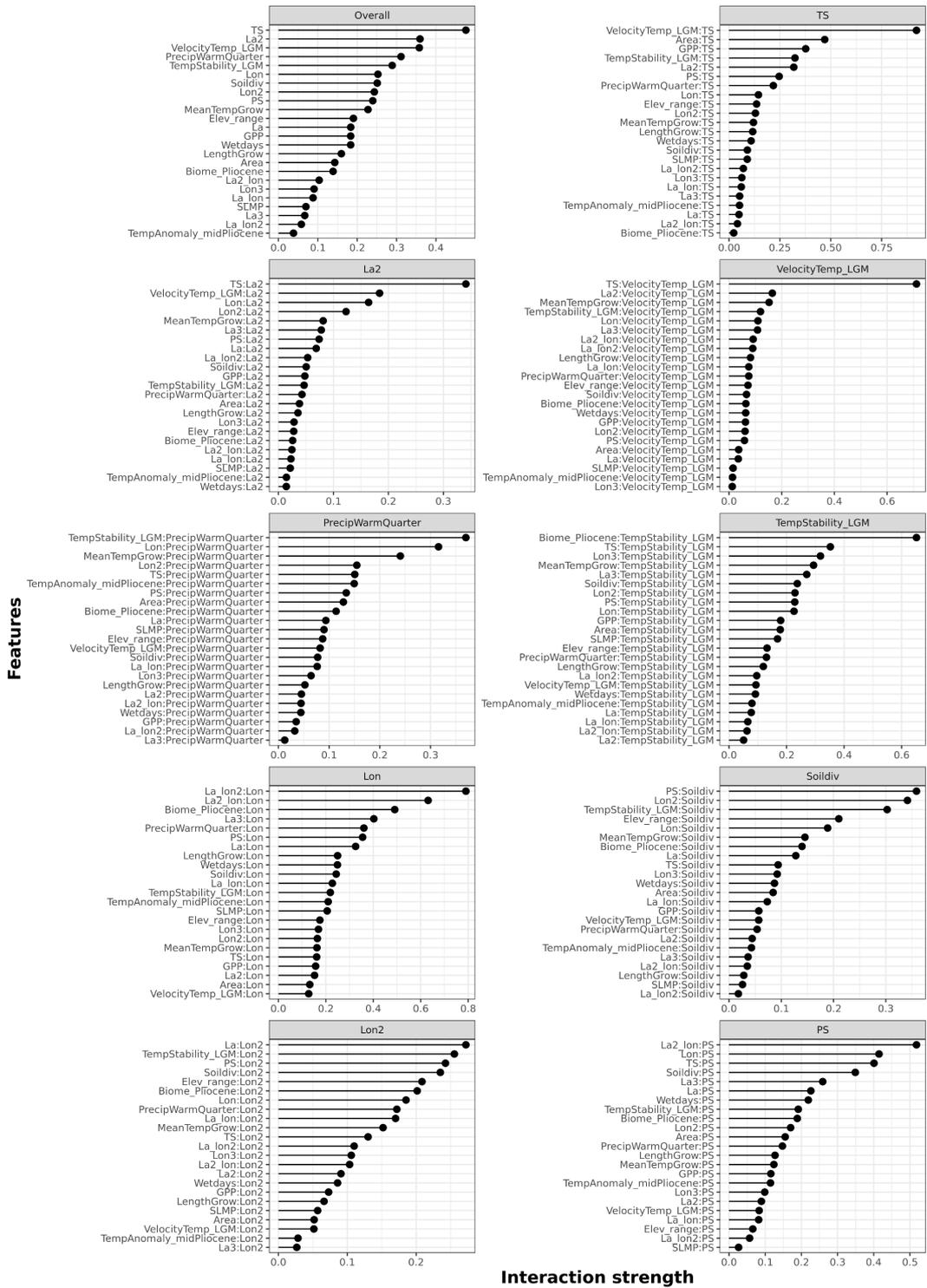
Figure S1.17 Interaction strength of each predictor variable for explaining phylogenetic richness (Overall) in the spatial neural network model and two-way interaction strengths between the nine top-ranked covariates and all other covariates. Terms of cubic polynomial trend surfaces [i.e. latitude (Y), centered longitude (X) as well as X^2, XY, Y^2, X^3, X^2Y, XY^2 and Y^3] are "La", "Lon", "La2", "La_lon", "Lon2", "La3", "La2_lon", "La_lon2", and "Lon3", respectively. Other predictors are abbreviated as follows: Area = area of the region; SLMP = Surrounding landmass proportion; LengthGrow = Length of growing season; MeanTempGrow = Mean temperature of growing season; TS = Temperature seasonality; Wetdays = Number of wet days; PrecipWarmQuarter = Precipitation of warmest quarter; PS = Precipitation seasonality; GPP = Gross primary productivity; Soildiv = Number of soil types; Elev_range = Elevational range; VelocityTemp_LGM = Temperature change velocity since the LGM; TempStability_LGM = Temperature stability since the LGM; TempAnomaly_midPliocene = Temperature anomaly since the mid-Pliocene; Biome_Pliocene = Biome area changes.

Figure S1.18 Estimated effects of the nine two-way interactions (two-predictors partial dependence plots) in the spatial neural network model for phylogenetic richness.
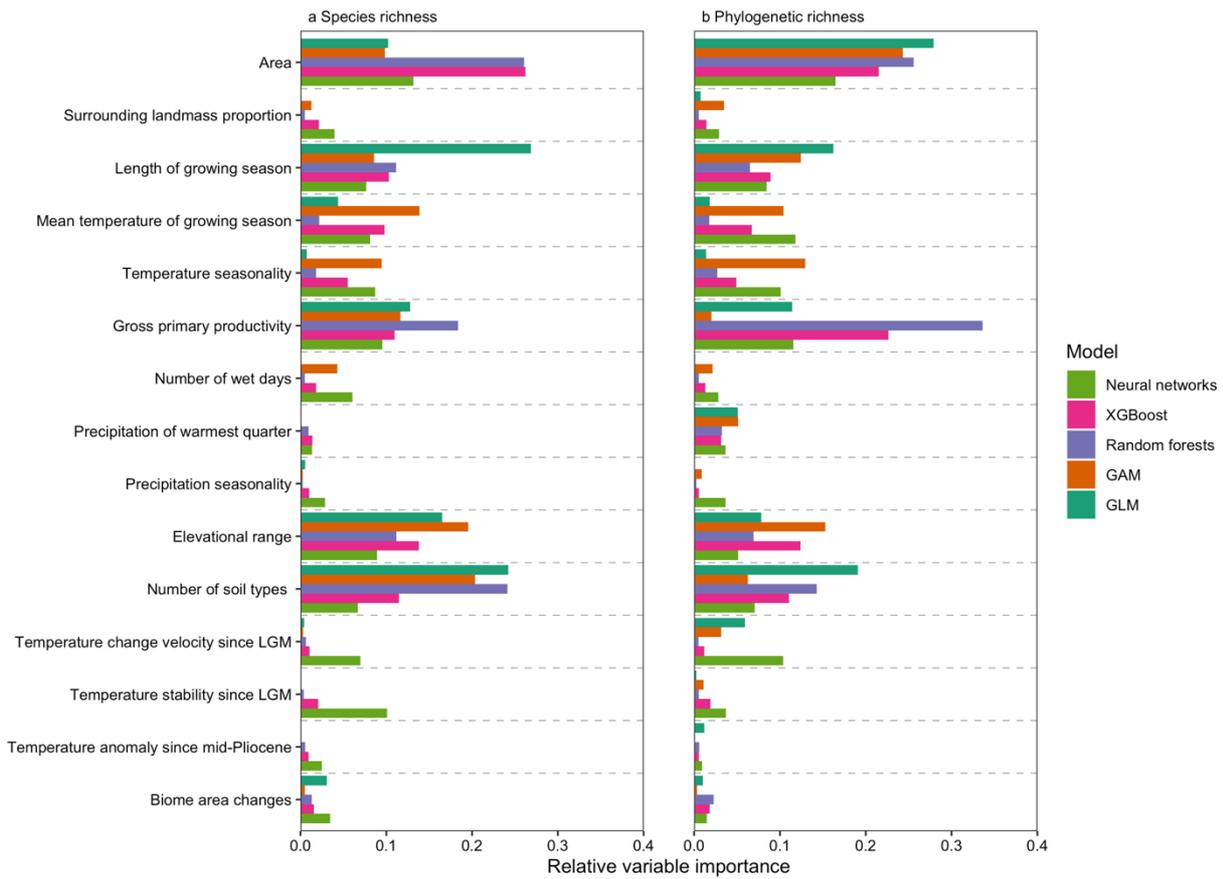
Figure S1.19 Relative importance of environmental variables explaining global pattern of vascular plant diversity across five non-spatial models. The models were fitted including 15 predictors representing geography, climate, environmental heterogeneity and past environmental conditions. a, Importance of predictor variables for species richness; b, Importance of predictor variables for phylogenetic richness.
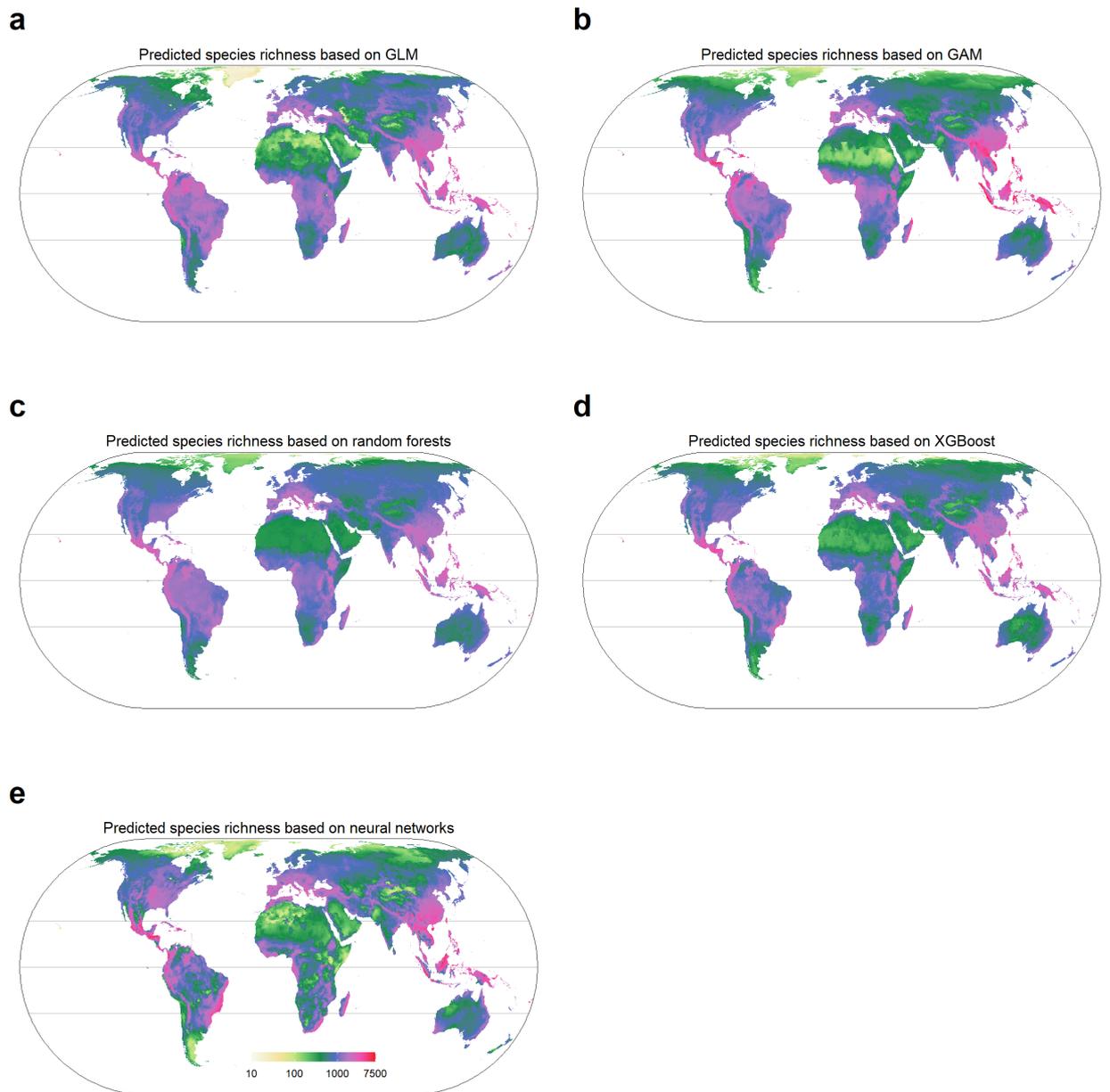
Figure S1.20 Species richness of vascular plants predicted across an equal area grid of 7,774 km^2 hexagons based on different models (i.e. spatial models using machine learning methods and generalized additive models, and a non-spatial generalized linear model with interactions). The same log$_{10}$ scale color gradient is used in all maps. For comparisons across all spatial and non-spatial models and data download, see https://gift.uni-goettingen.de/shiny/predictions/. Projection: Eckert IV. GLM = generalized linear model, GAM = generalized additive model, XGBoost = extreme gradient boosting.
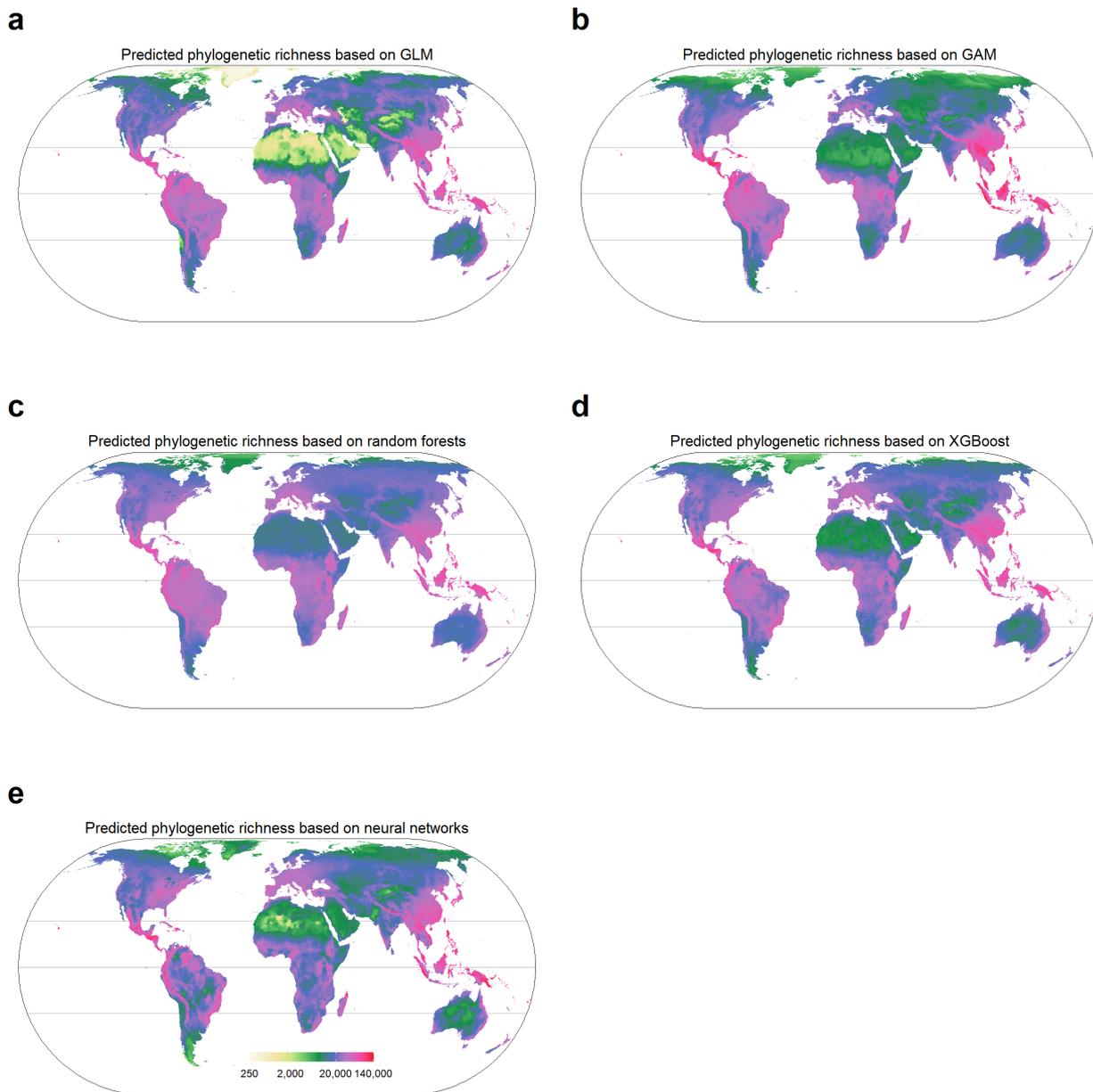
Figure S1.21 Phylogenetic richness of vascular plants predicted across an equal area grid of 7,774 km^2 hexagons based on different models (i.e. spatial models using machine learning methods and generalized additive models, and a non-spatial generalized linear model with interactions). The same log$_{10}$ scale color gradient is used in all maps. For comparisons across all spatial and non-spatial models and data download, see https://gift.uni-goettingen.de/shiny/predictions/. Projection: Eckert IV. GLM = generalized linear model, GAM = generalized additive model, XGBoost = extreme gradient boosting.
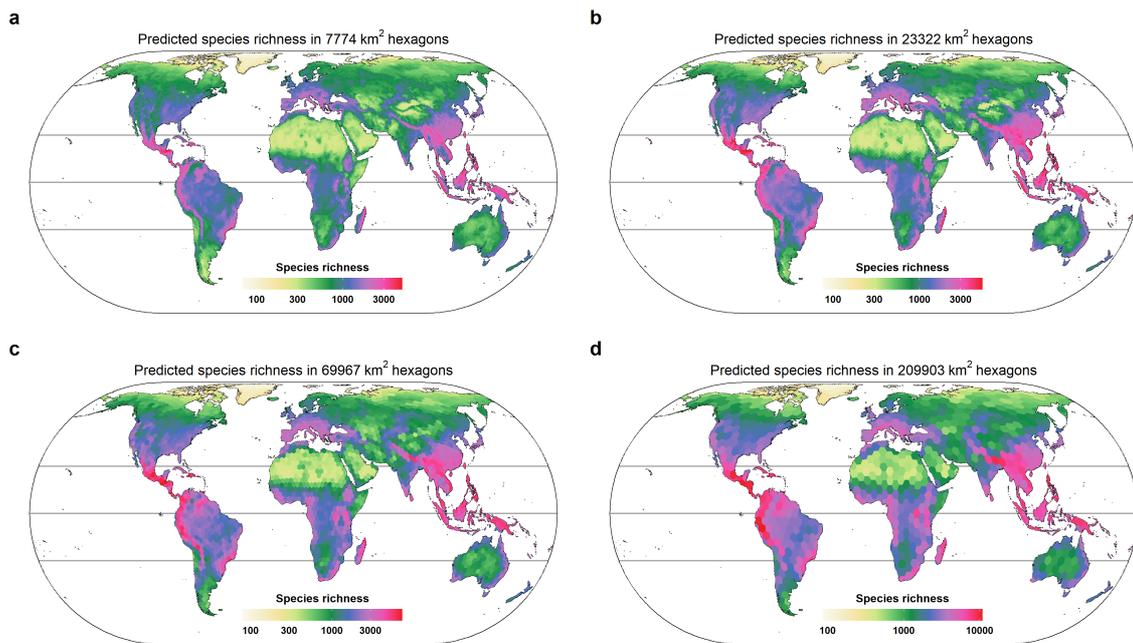
Figure S1.22 Species richness of vascular plants based on ensemble predictions across different grid sizes (i.e. spatial models using machine learning methods and generalized additive models, and a non-spatial generalized linear model with interactions). Grid sizes used for maps are: a, 7774 km²; b, 23322 km²; c, 69967 km²; d, 209903 km². For comparisons across all spatial and non-spatial models and data download, see https://gift.uni-goettingen.de/shiny/predictions/. Projection: Eckert IV.
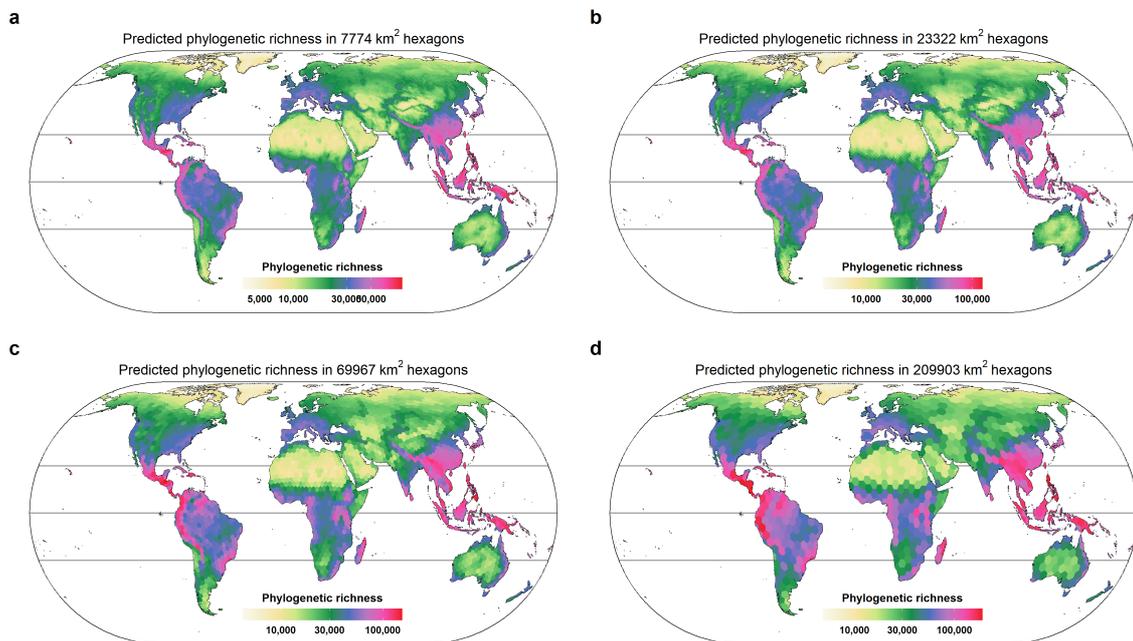
Figure S1.23 Phylogenetic richness of vascular plants based on ensemble predictions across different grid sizes (i.e. spatial models using machine learning methods and generalized additive models, and a non-spatial generalized linear model with interactions). Grid sizes used for maps are: a, 7774 km²; b, 23322 km²; c, 69967 km²; d, 209903 km². For comparisons across all spatial and non-spatial models and data download, see https://gift.uni-goettingen.de/shiny/predictions/. Projection: Eckert IV.
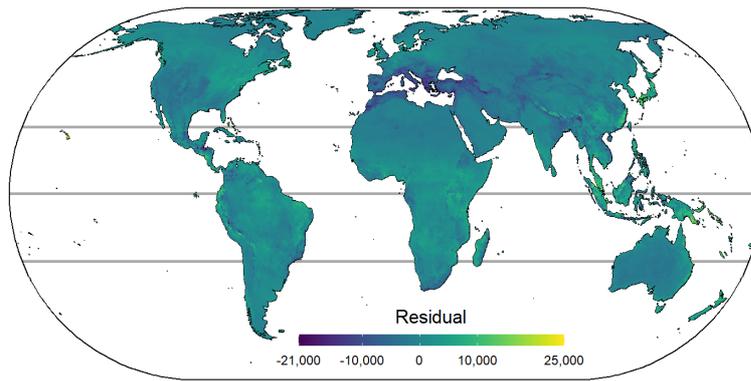
Figure S1.24 Residuals (deviation) from the linear regression between species richness and phylogenetic richness based on Ensemble predictions (phylogenetic richness = 22.1 * species richness, $R^2 = 0.947$, $p < 0.0001$). Negative residuals indicate lower phylogenetic richness than expected based on species richness. Projection: Eckert IV.
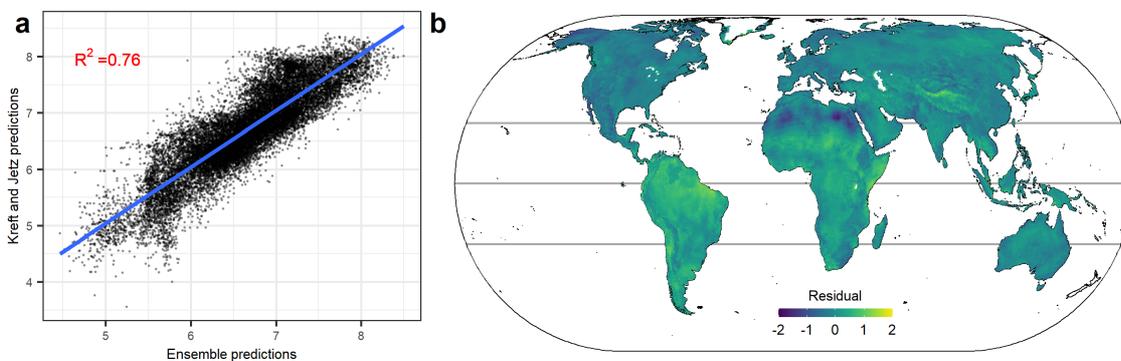
Figure S1.25 Comparison between vascular plant species richness based on Ensemble predictions produced in the scope of this paper (SR.Ensemble) and species richness extracted from Kreft and Jetz's predictions (Kreft & Jetz, 2007) (SR.Kreft) (a, SR.Kreft = 1.01 SR.Ensemble, $R^2 = 0.76$, $p < 0.0001$). Species richness was log-transformed. b, global patterns of residuals from the linear regression between species richness based on the ensemble predictions and species richness extracted from Kreft and Jetz's predictions (Kreft & Jetz, 2007). Positive residuals indicate higher values of SR.Kreft. Projection: Eckert IV.
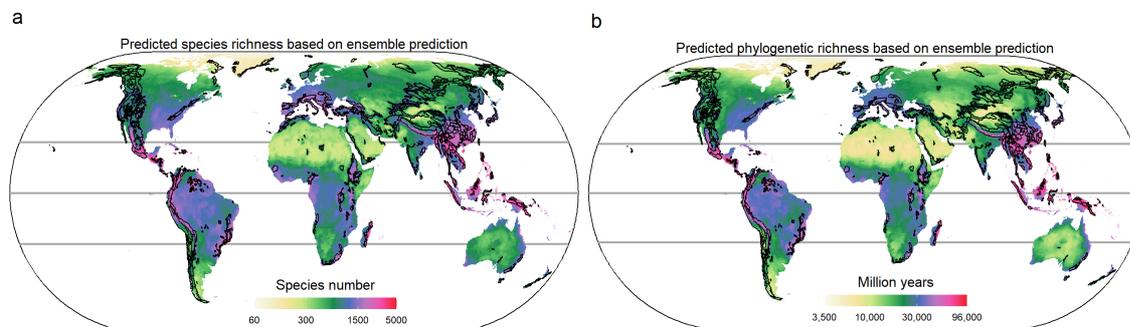
Figure S1.26 Vascular plant diversity based on ensemble predictions across an equal area grid of 7774 km² hexagons and mountain regions. Black lines delineate mountainous regions worldwide based on Körner *et al.* (2017). Projection: Eckert IV.
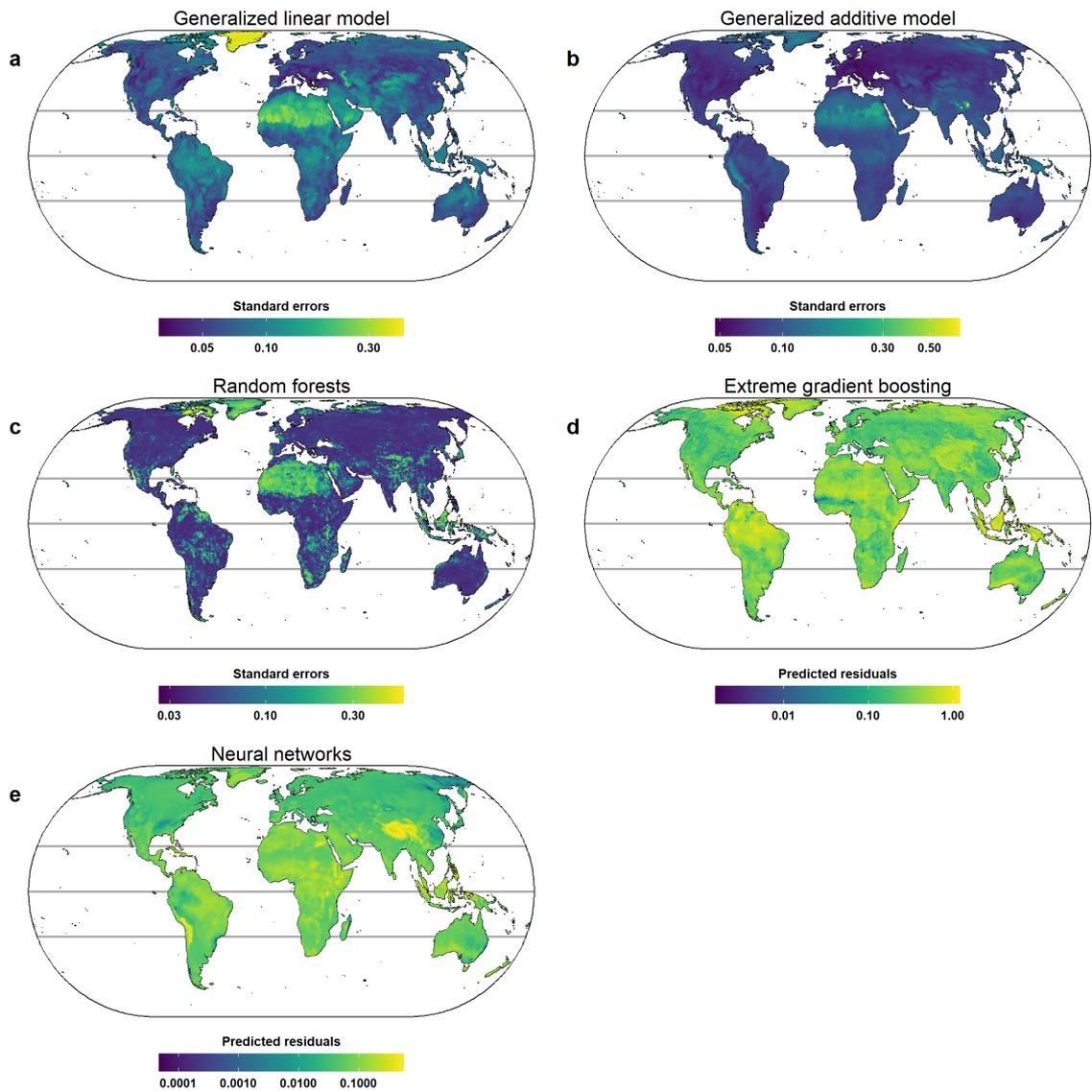
Figure S1.27 Uncertainty in predicted species richness from the five models used for the ensemble predictions (i.e. spatial models using machine learning methods and generalized additive models, and a non-spatial generalized linear model with interactions). Prediction variation is measured as standard errors of predicted values in generalized additive models, generalized linear models and random forests, and predicted residuals in extreme gradient boosting and neural networks based on models fitting the relationship between residuals of trained models and predictors from the raw datasets. For comparisons across all spatial and non-spatial models and data download, see https://gift.uni-goettingen.de/shiny/predictions/. Projection: Eckert IV.
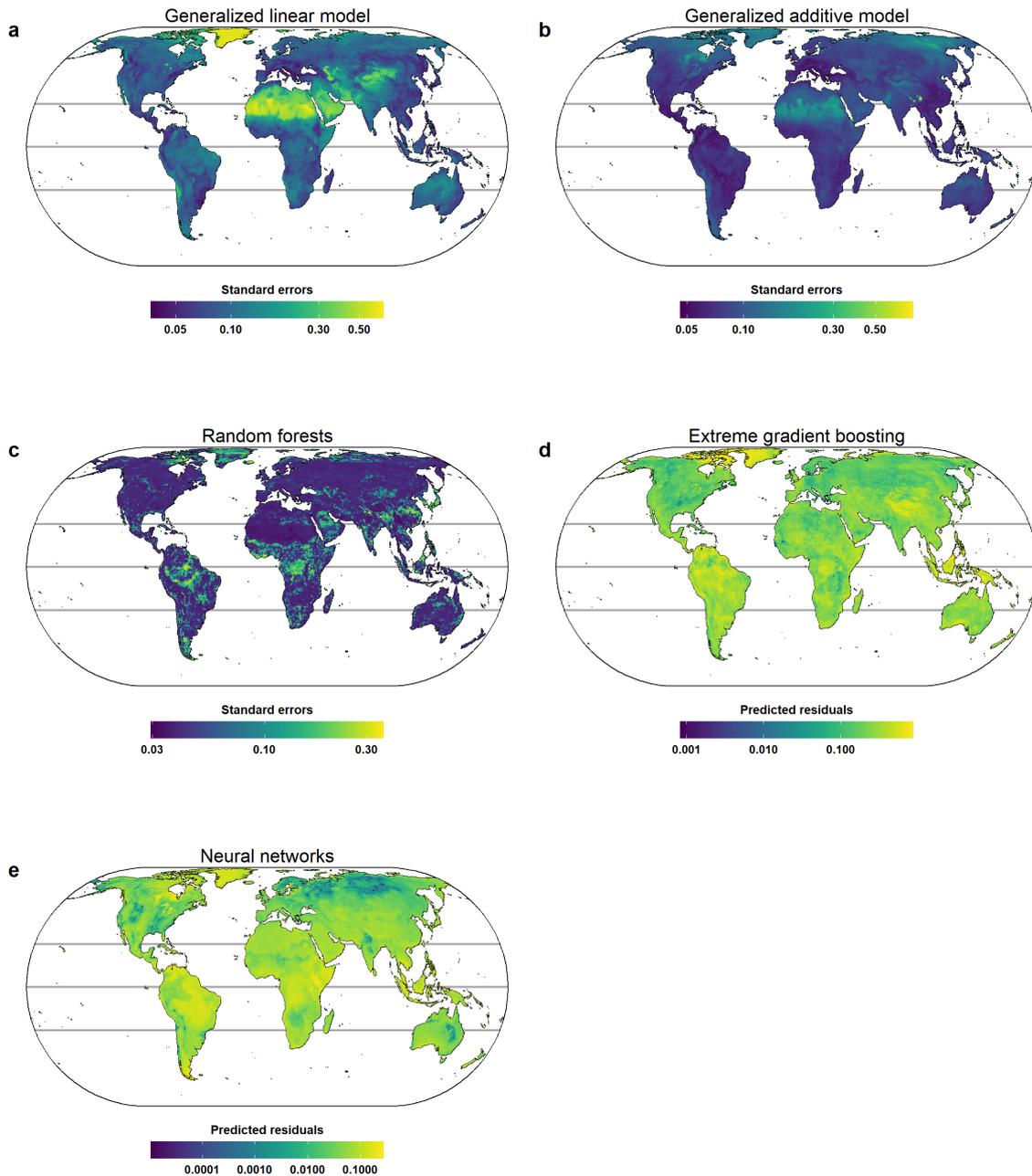
Figure S1.28 Uncertainty in predicted phylogenetic richness from the five models used for the ensemble predictions (i.e. spatial models using machine learning methods and generalized additive models, and a non-spatial generalized linear model with interactions). Prediction variation is measured as standard errors of predicted values in generalized additive models, generalized linear models and random forests, and predicted residuals in extreme gradient boosting and neural networks based on models fitting the relationship between the residuals of trained models and predictors from the raw datasets. For comparisons across all spatial and non-spatial models and data download, see https://gift.uni-goettingen.de/shiny/predictions/. Projection: Eckert IV.

Methods S1.1 Sensitivity analyses of phylogenetic richness

A recent study (Qian & Jin, 2021) showed that phylogenetic richness derived from a phylogeny resolved only at the genus level was nearly perfectly correlated with phylogenetic richness derived from a phylogeny resolved fully at the species level (Pearson's r: 0.997-1). This suggests that patterns of phylogenetic richness would be similar regardless of whether the phylogeny used to calculate phylogenetic richness is resolved at the genus or species level. As a sensitivity analysis assessing a potential effect of adding missing genera to the phylogeny on phylogenetic richness, we disregarded species from genera that were absent from the phylogenetic backbone and constructed a tree only for the remaining species. We found that phylogenetic richness calculated from the tree resolved at the genus level was nearly perfectly correlated to phylogenetic richness based on the tree with missing genera added (Pearson's r = 0.998).

Methods S1.2 Past environmental variables

To determine the contribution of past environmental conditions to modern diversity patterns, we calculated biome area variation since the Pliocene and the Middle Miocene, temperature anomaly since the mid-Pliocene warm period, temperature stability since the last glacial maximum (LGM) and velocity of temperature change since the LGM. Terrestrial biomes are affected by multiple drivers containing atmospheric circulation, precipitation and temperature patterns, and thus changes in biome distributions represent major environmental changes through geological time. To calculate biome area variation, we used biome distribution maps at present (Olson *et al.*, 2001), the LGM (c. 25 – 15 ka) (Ray & Adams, 2001), the mid-Pliocene warm period (mid-Piacenzian, c. 3.264 – 3.025 Ma) (Dowsett *et al.*, 2016) and the Middle Miocene (c. 17 – 15 Ma) (Henrot *et al.*, 2010). The three past periods represent particularly different climates compared to present-day conditions, and showed distinct biome distributions which are hypothesized to have left imprints on current plant diversity (Svenning *et al.*, 2015; Sandel *et al.*, 2020). The LGM represents a cooler period compared to present-day conditions, characterized by large glaciated areas, expanded dry deserts and reduced forest biome areas. In contrast, the mid-Pliocene and the Middle Miocene were two relatively warm periods compared to present-day climate, characterized by decreased ice loadings of the continents and expanded forest biomes at the expense of deserts (Henrot *et al.*, 2010; Dowsett *et al.*, 2016). Since biome definitions differed across the four datasets, we defined six broad biome categories, namely tropical forest, temperate forest, boreal forest, savanna and grassland, tundra and deserts, and regrouped the biomes of each dataset to match the new biome classification (Table S1.3). We extracted each biome's area inside each region at each time slice and standardized it by dividing it by the area of the region. Then we calculated Euclidean distances of biome area change between every two adjacent time-slices for each region and averaged the distance across all time-slice periods.

In addition, we calculated temperature stability from two paleo-time periods until present, i.e. the LGM and the mid-Pliocene warm period. Paleoclimate estimates for calculating temperature stability since the LGM were derived from the TRaCE21ka experiment, based on the Community Climate System Model version 3 (CCSM3) (Owens & Guralnick, 2019). CCSM3 is a global, coupled ocean-atmosphere-sea ice-land surface climate model without flux adjustment (Collins *et al.*, 2006). To calculate temperature anomaly since the mid-Pliocene, we took mean annual temperature of the mid-Pliocene warm period from PaleoClim, simulated following the PlioMIP protocols (Brown et al., 2018). Additionally, we took mean annual temperature for the LGM for calculating the velocity of temperature change since the LGM from the Paleoclimate Modeling Intercomparison Project Phase II, which was averaged from hindcasts of two past climate models (CCSM3 and MIROC3.2) (Braconnot *et al.*, 2007; Weigelt *et al.*, 2013).

Methods S1.3 Statistical models

For each random forest model, we fitted 500 regression trees. After tuning the hyperparameters, we used 8 variables randomly sampled as candidates at each split and each node contained at least 4 samples (minimal node size) for species richness, while 6 variables were randomly sampled at each split and each node contained at least 4 samples for phylogenetic richness.

In Extreme Gradient Boosting (XGBoost), three types of parameters have to be set: general parameters, task parameters and booster parameters. General parameters are related to which booster is used, and here we selected tree-based models. For task parameters, we chose regression with squared loss for ranking tasks. Booster parameters define how to build the tree models. In our resulting XGBoost model for species richness, we used the booster parameters max_depth = 4, eta = 0.1, gamma = 0, colsample_bytree = 0.8, min_child_weight = 1 and subsample = 0.7. In the XGBoost model for phylogenetic richness, we used the booster parameters max_depth = 4, eta = 0.1, gamma = 0, colsample_bytree = 0.7, min_child_weight = 1 and subsample = 0.8. Further, we set the maximum number of boosting iterations to 200. Both random forests and XGBoost control for overfitting by utilizing ensemble strategies, i.e. bagging and gradient boosting. While column (feature) sampling is used by both methods to further prevent overfitting, XGBoost utilizes two additional techniques, regularization and shrinkage (Chen & Guestrin, 2016).

We applied feed-forward neural networks with three hidden layers between the input and output layer. We used resilient backpropagation with the weight backtracking algorithm to compute the neural networks. Compared with

traditional backpropagation algorithms, this algorithm applies a separate learning rate, which can be changed during the training process (Günther & Fritsch, 2010). Moreover, we applied logistic activation. We scaled the response after log-transformation and all predictors to fall within a range of 0 to 1 to improve neural network stability and modeling performance. For every hidden layer, we trained the number of neurons, and the architecture of the final networks for prediction was (8,4,8) in the species richness model and (8,4,10) in the phylogenetic richness model.

We used partial dependence plots from the R package *pdp* to visualize the relationships between diversity metrics and the single predictors across models. Partial dependence plots visualize the relationship between a subset of the predictors (typically 1-3) and the response while accounting for the average effect of all other predictors in the model (Greenwell, 2017). To identify and visualize important two-way interactions between predictor variables, we followed Lucas (Lucas, 2020). We first calculated the interaction importance for each covariate by decomposition of the prediction models (Friedman & Popescu, 2008). Then we calculated the two-way interaction strengths between the covariates of interest and all other covariates. Finally, we visualized important two-way interactions using two-predictors partial dependence plots.

Methods S1.4 Cross-validation

To assess the accuracy of model predictions across all different model types, we used random 10-fold cross-validation and spatial 68-fold cross-validation. In random cross-validation, the observations in the dataset were randomly partitioned into 10 nearly equally-sized sets. Nine sets were used to train the model, which was then used to make predictions for the remaining set. The predictions were then compared to observed values. This process was repeated until all 10 sets were predicted. For spatial cross-validations, observations were grouped into spatially homogeneous clusters if their pairwise geographic distances were smaller than the threshold of spatial autocorrelation to remove potentially spatial dependence between training and test data (Ploton *et al.*, 2020). Spatial clusters were generated using a hierarchical cluster analysis (complete linkage method) of the distance matrix of observed geographical coordinates and a clustering height (i.e. the threshold of spatial autocorrelation). The threshold of spatial autocorrelation (i.e. the maximum distance between regions within each cluster) was defined here as 2,000 km based on the spatial correlograms of raw observed species and phylogenetic richness, where the Moran's I almost reached zero (Figure S1.3).

Methods S1.5 Handling of missing values in predictor variables for predicting and calculating predictions in raster format

Seven predictors (i.e. surrounding landmass proportion, gross primary productivity, mean temperature of growing season, biome area variation since Pliocene, temperature stability since LGM and temperature anomaly since the mid-Pliocene warm period in all resolutions; precipitation seasonality in 7,774 and 23,322 km^2 resolution) had missing values in a small number of hexagons at northern Africa, Greenland and the margins of continents. XGBoost can handle missing data when it is used for predicting. An optimal default direction in each tree node is learned from the trained data in the model constructing process. When there are missing values in predictor data, the observation is classified into the default branch. Random forests, neural networks, GLMs and GAMs cannot deal with missing values. We therefore interpolated predictor values for hexagons from their neighbors. To interpolate biome area change, we defined the biome type of each hexagon where missing values occurred in the Pliocene map according to their neighboring cells.

To generate plant diversity maps in raster format based on the 7,774 km^2 resolution hexagon polygon grids, we applied two different approaches. First, we rasterized the values of the hexagon polygon grids using the *fasterize* R package, with a CHELSA raster layer at a resolution of 30 arc seconds (c. 1 km^2 at the equator) (Karger *et al.*, 2017) as a template. These layers include "rasterized" at the end of their filenames. Second, we aggregated the rasterized data to a resolution of 20 arc minutes (c.1600 km^2 at the equator) as mean values, and then resampled the aggregated layers back to 30 arc seconds resolution using cubic interpolation. Values of these interpolated raster layers still represent species richness and phylogenetic richness per 7,774 km^2 but do not follow the borders of the initial hexagon grid cells any more. These layers include "interpolated" at the end of their filenames. When extracted and averaged per hexagon grid cell, the interpolated values were highly correlated to the original predicted values of the hexagon polygon grids, except for some small fractions of grid cells located on the coasts of continents deviating more strongly from their neighboring cells due to low environmental heterogeneity (Figure S1.4). Predictions as raster layers and hexagon polygon layers can be downloaded at https://gift.uni-goettingen.de/shiny/predictions/ .

References S1.1 References of Checklists and Floras from the Global Inventory of Floras and Traits (GIFT) used to compile the regional species composition data.

3D Environmental. 2008. Vegetation Communities and Regional Ecosystems of The Torres Strait Islands, Queensland, Australia. Queensland, Australia.

3D Environmental. 2013. Profile for management of the habitats and related ecological and cultural resource values of Mua Island. Queensland, Australia.

Acevedo-Rodríguez P, Strong MT. 2007. *Catalogue of the seed plants of the West Indies Website*. [WWW document] URL http://botany.si.edu/antilles/WestIndies/catalog.htm [accessed 1 March 2011].

Backer CA, Bakhuizen van den Brink, RC. 1963. Flora of Java. (Spermatophytes only). Groningen: Noordhoff.

Baker ML, Duretto MF. 2011. A census of the vascular plants of Tasmania. Hobart, Australia: Tasmanian Herbarium, Tasmanian Museum and Art Gallery.

Barker WR, Barker RM, Jessop JP, Vonow HP. 2005. Census of South Australian vascular plants. *Journal of the Adelaide Botanic Gardens Supplement* 1: 1–396.

Belhacene L. 2010. Catalogue 2010 des plantes vasculaires du département de la Haute-Garonne. *Supplément à Isaatis* 10: 1–145.

Bernal R, Gradstein SR, Celis M. 2015. *Catálogo de plantas y líquenes de Colombia*. [WWW document] URL http://catalogoplantasdecolombia.unal.edu.co/ [accessed 15 January 2016].

BGCI. 2021. *GlobalTreeSearch online database (version 1.4)*. [WWW document] URL https://tools.bgci.org/global_tree_search.php [accessed 19 February 2021].

Bingham MG, Willemen A, Wursten BT, Ballings P, Hyde MA. 2016. *Flora of Zambia*. [WWW document] URL http://www.zambiaflora.com/ [accessed 14 November 2016].

BioScripts. 2014. *Flora Vascular*. [WWW document] URL http://www.floravascular.com/ [accessed 25 May 2014].

Botanical Garden Tel Aviv. 2015. Israel Flora. unpublished. Tel Aviv, Israel: Tel Aviv University: Tel Aviv University.

Breckle S-W, Hedge IC, Rafiqpoor MD. 2013. Vascular plants of Afghanistan. An augmented checklist. Bonn: Scientia Bonnensis.

Brennan K. 1996. An annotated checklist of the vascular plants of the Alligator Rivers Region, Northern Territory, Australia. Barton, Australia: Supervising Scientist.

Broughton DA, McAdam JH. 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.

Brundu G, Camarda I. 2013. The Flora of Chad: a checklist and brief analysis. *Phytokeys* 23: 1–17.

Buchwald E, Wind P, Bruun HH, Møller PF, Ejrnæs R, Svart HE. 2013. Hvilke planter er hjemmehørende i Danmark? *Jydsk Naturhistorisk Forening* 118: 72–96.

Bundesamt für Naturschutz. 2016. *Floraweb*. [WWW document] URL www.floraweb.de [accessed 15 September 2016].

Cámara-Leret R, Frodin DG, Adema F, Anderson C, Appelhans MS, Argent G, Arias Guerrero S, Ashton P, Baker WJ, Barfod AS *et al.* 2020. New Guinea has the world's richest island flora. *Nature* 584: 579–583.

Case TJ, Cody ML, Ezcurra E. 2002. A new island biogeography of the Sea of Cortés. New York, NY: Oxford University Press.

Catarino L, Martins ES, Basto MF, Diniz MA. 2008. An annotated checklist of the vascular flora of Guinea-Bissau (West Africa). *Blumea-Biodiversity, Evolution and Biogeography of Plants* 53: 1–222.

Chang C-S, Kim H, Chang K. Provisional Checklist of the Vascular Plants for the Korea Peninsular Flora (KPF). Version 1.0. Korea.

Cheffings CM, Farrell L, eds. 2005. The vascular plant red data list for Great Britain. Peterborough, UK: Joint Nature Conservation Committee.

Chiapella J, Ezcurra C. 1999. La flora del parque provincial Tromen, provincia de Neuquén, Argentina. *Multequina* 8: 51–60.

Chinese Virtual Herbarium. 2016. *The Flora of China v. 5.0*. [WWW document] URL http://www.cvh.org.cn/ [accessed 15 January 2016].

Clark JL, Neill DA, Asanza M. 2006. Floristic checklist of the Mache-Chindul mountains of Northwestern Ecuador. *Contributions from the United States National Herbarium* 54: 1–180.

Cochard R, Bloesch U. 2007. *Electronic plant species database of the Saadani National Park, coastal Tanzania*. [WWW document] URL http://www.wildlife-baldus.com/saadani.html [accessed 14 November 2016].

CONABIO. 2016. *Sistema Nacional de Información sobre Biodiversidad*. [WWW document] URL https://www.gob.mx/conabio [accessed 11 April 2016].

Conti F, Abbate G, Alessandrini A, Blasi C. 2005. Annotated Checklist of the Italian Vascular Flora. Rome, Italy: Palombi Editori.

Conti F, Bartolucci F. 2015. The Vascular Flora of the National Park of Abruzzo, Lazio and Molise (Central Italy). Barisciano, Italy: Springer.

Danihelka J, Chrtek J, Kaplan Z. 2012. Checklist of vascular plants of the Czech Republic. *Preslia* 84: 647–811.

Dauby G, Zaiss R, Blach-Overgaard A, Catarino L, Damen T, Deblauwe V, Dessein S, Dransfield J, Droissart V, Duarte MC *et al.* 2016. RAINBIO. A mega-database of tropical African vascular plants distributions. *PhytoKeys* 74: 1–18.

Dimopoulos P, Raus T, Bergmeier E, Constantinidis T, Iatrou G, Kokkini S, Strid A, Tzanoudakis D. 2013. Vascular plants of Greece. An annotated checklist. Berlin, Athens: Botanischer Garten und Botanisches Museum Berlin-Dahlem, Freie Universität Berlin; Hellenic Botanical Society.

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.

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.

Elizabeth Joyce, Kevin Thiele, Ferry Slik, Darren Crayn, Joyce E, Thiele K, Slik F, Crayn D. 2020. Checklist of the vascular flora of the Sunda-Sahul Convergence Zone. *Pensoft Publishers* 8: e51094.

Euro+Med. 2006. *Euro+Med PlantBase - the information resource for Euro-Mediterranean plant diversity.* [WWW document] URL http://ww2.bgbm.org/EuroPlusMed/ [accessed 1 June 2019].

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.

Fischer MA, 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. Linz, Austria: Oberösereichisches Landesmuseum.

Funk VA, Hollowell T, Berry P, Kelloff C, Alexander SN. 2007. Checklist of the plants of the Guiana Shield (Venezuela: Amazonas, Bolivar, Delta Amacuro; Guyana, Surinam, French Guiana). Washington, DC: Department of Botany, National Museum of Natural History.

Harris DJ. 2002. The vascular plants of the Dzanga-Sangha Reserve, Central African Republic. Edinburgh, Scotland, UK: Royal Botanic Garden Edinburgh.

Hassler M. 2020. *World Ferns. Synonymic Checklist and Distribution of Ferns and Lycophytes of the World.* [WWW document] URL www.worldplants.de/ferns/ [accessed accessed via Species 2000 & ITIS Catalogue of Life 22 September 2020].

Hutchinson J, Dalziel JM, Keay RWJ, Hepper N. 2014. Flora of West Tropical Africa. London, UK: Crown agents for overseas governments.

Hyde MA, Wursten BT, Ballings P, Coates Palgrave M. 2016a. *Flora of Botswana.* [WWW document] URL http://www.botswanaflora.com/ [accessed 14 November 2016].

Hyde MA, Wursten BT, Ballings P, Coates Palgrave M. 2016b. *Flora of Malawi.* [WWW document] URL http://www.malawiflora.com/index.php [accessed 14 November 2016].

Hyde MA, Wursten BT, Ballings P, Coates Palgrave M. 2016c. *Flora of Mozambique.* [WWW document] URL http://www.mozambiqueflora.com/ [accessed 14 November 2016].

Hyde MA, Wursten BT, Ballings P, Coates Palgrave M. 2016d. *Flora of Zimbabwe* [accessed 14 November 2016].

INBIO. 2000. *Lista de planta de Costa Rica. With updates by Eduardo Chacón.* [WWW document] URL http://www.inbio.ac.cr/papers/manual_plantas/index.html [accessed 5 February 2016].

Inderjit. 2017. State level vascular plant checklists for India.

Jardim Botânico do Rio de Janeiro. 2016. *Flora do Brasil 2020 em construção.* [WWW document] URL http://floradobrasil.jbrj.gov.br/ [accessed 9 May 2016].

Junak S, Philbrick R, Chaney S, Clark R. 1997. A checklist of vascular plants of Channel Islands National Park. Tucson, Arizona: Southwest Parks and Monuments Association.

Karlsson T, Agestam M. 2014. *Checklist of Nordic vascular plants. Sweden Checklist.* [WWW document] URL http://www.euphrasia.nu/checklista/index.eng.html [accessed 11 February 2016].

Kartesz JT. 2014. *The Biota of North America Program (BONAP). North American Plant Atlas.* [WWW document] URL http://www.bonap.org/napa.html [accessed 16 February 2016].

Kelloff CL, Funk VA. 1998. Preliminary checklist of the plants of Kaieteur National Park, Guyana. Washington: National Museum of Natural History, Smithsonian Institution.

Kluge J, Worm S, Lange S, Long D, Böhner J, Yangzom R, Miehe G. 2017. Elevational seed plants richness patterns in Bhutan, Eastern Himalaya. *Journal of Biogeography* 44: 1711–1722.

Krestov PV. 2016. Flora of Siberia. unpublished.

Kristinsson H. 2008. Checklist of the vascular plants of Iceland. Reykjavík, Iceland: Náttúrufræðistofnun Íslands.

Kupriyanov AN, Kupriyanov OA. 2014. The study of the flora (by the example of Kemerovo region). unpublished. Kemerovo.

Kuzmenkova SM. 2015. *Plants of Belarus.* [WWW document] URL http://hbc.bas-net.by/plantae/eng/default.php [accessed 15 February 2016].

Landcare Research New Zealand. *Ecological Traits of New Zealand Flora.* [WWW document] URL https://ecotraits.landcareresearch.co.nz/ [accessed Nov/2018].

Lazkov GA, Sultanova BA. 2011. Checklist of vascular plants of Kyrgyzstan. Helsinki: Botanical Museum, Finnish Museum of Natural History.

Le Houerou HN. 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.

Lipkin R. 2005. Aniakchak National Monument and Preserve, vascular plant inventory: final technical report. Anchorage, USA: National Park Service, Southwest Alaska Network Inventory & Monitroing Program.

Lorite J. 2016. An updated checklist of the vascular flora of Sierra Nevade (SE Spain). *Phytotaxa* 261: 1–57.

Marticorena C, Squeo FA, Arancio G, Muñoz M. 2008. Catálogo de la flora vascular de la IV Región de Coquimbo. In: Squeo FA, Arancio G, Gutiérrez JR, eds. *Libro rojo de la flora nativa y de los sitios prioritarios para su conservación: Región de Atacama*. Ediciones Universidad de La Serena La Serena, 105–142.

Masharabu T. 2011. *Flore et végétation du Parc National de la Ruvubu au Burundi: diversité, structure et implications pour la conservation*. PhD Thesis, Université Libre de Bruxelles, Bruxelles, Belgium.

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.

Ministry of Environment. 2012. The national red list 2012 of Sri Lanka. Conservation status of the fauna and flora. Colombo, Sri Lanka.

Nationalpark Eifel. 2015. *Artenliste Farne und Blütenpflanzen*. [WWW document] URL http://www.nationalpark-eifel.de/go/artenliste.html [accessed 11 March 2015].

Newman M. 2007. A checklist of the vascular plants of Lao PDR. Edinburgh, Scotland, UK: Royal Botanic Garden Edinburgh.

Nikolić T. 2016. *Flora Croatica Database*. [WWW document] URL http://hirc.botanic.hr/fcd [accessed 12 February 2016].

Norton J, Majid SA, Allan D, Al Safran M, Böer B, Richer RA. 2009. An illustrated checklist of the flora of Qatar. Gosport, UK: Browndown Publications Gosport.

Notov AA. 2010. National park "Zavidovo": vascular plants, bryophyte, lichens. Moscow: Delovoj mir.

Nowak A. 2018. Vascular plants of Tajikistan. unpublished.

NPS. 2015. *NPSpecies. Information on Species in National Parks*. [WWW document] URL https://irma.nps.gov/NPSpecies/ [accessed 9 April 2015].

Oggero AJ, Arana MD. 2012. Inventario de las plantas vasculares del sur de la zona serrana de Córdoba, Argentina. *Hoehnea* 39: 171–199.

Pal D, Kumar A, Dutt B. 2014. Floristic diversity of Theog Forest Division, Himachal Pradesh, Western Himalaya. *Check list* 10: 1083–1103.

Pandža M. 2010. Flora parka prirode Papuk (Slavonija, Hrvatska). *Šumarski list* 134: 25–43.

Parks Canada. 2015. *Biotics Web Explorer*. [WWW document] URL http://www.pc.gc.ca/apps/bos/BOSIntro_e.asp [accessed 13 April 2015].

Patzelt A. 2020. Plant checklist Oman.

Pelser PB, Barcelona JF, Nickrent DL. 2011. *Co's Digital Flora of the Philippines*. [WWW document] URL https://www.philippineplants.org/ [accessed 13 January 2016].

Peña-Chocarro MdC. 2010. Updated checklist of vascular plants of the Mbaracayú Forest Nature Reserve (Reserva Natural del Bosque Mbaracayú), Paraguay. Auckland, N.Z.: Magnolia Press.

Pôle Flore Habitats. 2015. *Catalogue de la flore vasculaire de Rhône-Alpes*. [WWW document] URL http://www.pifh.fr/pifhcms/index.php [accessed 12 January 2015].

Queensland Government. 2014. *Census of the Queensland flora 2014*. [WWW document] URL https://data.qld.gov.au/dataset/census-of-the-queensland-flora-2014 [accessed 5 February 2015].

Rakov NS, Saksonov S.V., Senator S.A., Vasjukov V.M. 2014. Vascular plants of Ulyanovsk Region. Togliatti, Russia: Russian Academy of Sciences.

Revushkin A. 2014. Key to the flora of the Tomsk Region [Opredelitel rasteniy Tomskoy oblasty]. with minor corrections by Alexandr Ebel. Tomsk, Russia: Tomsk State University Press.

Rodriguez R, Marticorena C, Alarcón D, Baeza C, Cavieres L, Finot VL, Fuentes N, Kiessling A, Mihoc M, Pauchard A *et al.* 2018. Catálogo de las plantas vasculares de Chile. *Gayana. Botánica* 75: 1–430.

Roux JP. 2009. Synopsis of the Lycopodiophyta and Pteridophyta of Africa, Madagascar and neighbouring islands. Cape Town, South Africa: South African National Biodiversity Institute.

Royal Botanic Gardens and Domain Trust. 2017. *PlantNET - The NSW Plant Information Network System*. [WWW document] URL http://plantnet.rbgsyd.nsw.gov.au [accessed 16 September 2016].

Rundel PW, Dillon MO, Palma B. 1996. Flora and Vegetation of Pan de Azúcur National Park in the Atacama desert of Northern Chile. *Gayana Bot* 53: 295–315.

SANBI. 2014. *Plants of Southern Africa. An online checklist*. [WWW document] URL http://posa.sanbi.org [accessed 9 March 2015].

Sandbakk BE, Alsos IG, Arnesen G, Elven R. 1996. *The flora of Svalbard*. [WWW document] URL http://svalbardflora.no/ [accessed 16 March 2011].

Selvi F. 2010. A critical checklist of the vascular flora of Tuscan Maremma (Grosseto province, Italy). *Fl. Medit* 20: 47–139.

Shaw JD, Spear D, Greve M, Chown SL. 2010. Taxonomic homogenization and differentiation across Southern Ocean Islands differ among insects and vascular plants. *Journal of Biogeography* 37: 217–228.

Short PS, Albrecht DE, Cowie ID, Lewis DL, Stuckey BM. 2011. Checklist of the vascular plants of the Northern Territory. Darwin: Department of Natural Resources, Environment, The Arts and Sport.

Silaeva TB, Chugunov GG, Kiryukhin IV, Ageeva AM, Vargot EV, Grishutkina GA, Khapugin AA. 2011. Flora of the national park "Smolny". Mosses and vascular plants: annotated list of species. Moscow, Russia: Commission of RAS for the Conservation of Biological Diversity [Комиссия РАН по сохранению биологического разнообразия].

Silantyeva M. 2013. Synopsis of the flora of Altayskyi Krai [Konspekt flory Altayskogo kraya]. with minor corrections by Alexandr Ebel. Barnaul, Russia: Altay State University Press.

SLUFG. 2015. Rote Liste und Artenliste Sachsen. Farn- und Samenpflanzen. Dresden, Germany: Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie.

Stace CA, Ellis RG, Kent DH, McCosh DJ. 2003. Vice-county Census Catalogue of the vascular plants of Great Britain, the Isle of Man and the Channel Islands. London, UK: Botanical Society of the British Isles.

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.

Tamis WLM, van der Meijden R, Runhaar J, Bekker RM, Ozinga WA, Odé B, Hoste I. 2004. Standard List of the Flora of the Netherlands 2003. *Gorteria* 30: 101–195.

Tatewaki M. 1957. Geobotanical studies on the Kurile Islands. *Acta Horti Gotoburgensis* 21: 43–123.

Tela Botanica. 2015. *Base de données des Trachéophytes de France métropolitaine (bdtfx)*. [WWW document] URL http://www.tela-botanica.org/bdtfx.

Thomas J. 2011. *Plant diversity of Saudi Arabia. Flora checklist*. [WWW document] URL http://plantdiversityofsaudiarabia.info/biodiversity-saudi-arabia/flora/Checklist/Cheklist.htm [accessed 14 January 2016].

Tropicos. 2015a. *Catálogo de las Plantas Vasculares de Bolivia*. [WWW document] URL http://www.tropicos.org/Project/BC [accessed 10 September 2015].

Tropicos. 2015b. *Catalogue of the Vascular Plants of Ecuador*. [WWW document] URL http://www.tropicos.org/Project/CE [accessed 22 October 2015].

Tropicos. 2015c. *Panama Checklist*. [WWW document] URL http://www.tropicos.org/Project/PAC [accessed 22 October 2015].

Tropicos. 2015d. *Peru Checklist*. [WWW document] URL http://www.tropicos.org/Project/PEC [accessed 22 October 2015].

Tropicos. 2020. *Catalogue of the Vascular Plants of Madagascar*. [WWW document] URL http://www.tropicos.org/Project/MADA [accessed 01/22/2020].

Turner IM. 1995. A catalogue of the vascular plants of Malaya. Singapore: Gardens' Bulletin.

Ulloa Ulloa C, Acevedo-Rodríguez P, Beck S, Belgrano MJ, Bernal R, Berry PE, Brako L, Celis M, Davidse G, Forzza RC *et al.* 2017. An integrated assessment of the vascular plant species of the Americas. *Science (New York, N.Y.)* 358: 1614–1617.

USDA, NRCS. 2015. *The PLANTS Database*. [WWW document] URL http://plants.usda.gov [accessed 24 April 2015].

van Vreeswyk AME, Payne AL, Leighton KA, Hennig P. 2004. An inventory and condition survey of the Pilbara region, Western Australia. Perth, Australia: Department of Agriculture.

Vanderplank SE. 2010. The Vascular Flora of Greater San Quintín, Baja California, Mexico. *CGU Theses & Dissertations* Paper 2.

Velayos Rodríguez M. 2016. *Flora de Guinea. Base de datos de Flora de Guinea*. [WWW document] URL http://www.floradeguinea.com/herbario/ [accessed 2 October 2016].

VicFlora. 2016. *Flora of Victoria*. [WWW document] URL http://vicflora.rbg.vic.gov.au [accessed 30 November 2016].

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.

Vogt C. 2011. Composición de la Flora Vascular del Chaco Boreal, Paraguay. I. Pteridophyta y Monocotiledoneae. *Steviana* 3: 13–47.

WCSP. 2014. *World Checklist of Selected Plant Families*. [WWW document] URL http://apps.kew.org/wcsp/home.do [accessed 1 December 2014].

Western Australian Herbarium. 2017. *FloraBase - the Western Australian Flora*. [WWW document] URL https://florabase.dpaw.wa.gov.au/ [accessed 16 September 2016].

Wieringa JJ. 2016. Flora of Gabon. unpublished.

Yena AV. 2012. Prirodnaja flora krymskogo poluostrova. Spontaneous flora of the Crimean Peninsula. Simferopol': Orianda.

Yonekura K, Kajita T. 2013. *BG Plants Japanese name and Scientific name Index (YList)*. [WWW document] URL http://ylist.info [accessed 6 February 2016].

ZDSF, SKEW. 2014. *Info Flora. Artenliste Schweiz 5x5 km*. [WWW document] URL https://www.infoflora.ch/de/daten-beziehen/artenliste-5x5-km.html [accessed 13 February 2015].

Zuloaga FO, Morrone O, Belgrano M. 2014. *Catálogo de las Plantas Vasculares del Cono Sur*. [WWW document] URL http://www.darwin.edu.ar/Proyectos/FloraArgentina/fa.htm [accessed 16 March 2015].

Zvyagintseva KO. 2015. An annotated checklist of the urban flora of Kharkiv. Kharkiv, Ukraine: Kharkiv National University.

Евстигнеев ОИ, Федотов ЮП. 2007. Флора сосудистых растений заповедника "Брянский лес". Брянск: Гос. природ. биосфер. заповедник Брян. лес.

Хапугин АА. 2013. Сосудистые растения Ромодановского района Республики Мордовия (конспект флоры). Saransk.

References of Supporting Information for Chapter 1

Braconnot P, Otto-Bliesner B, Harrison S, Joussaume S, Peterchmitt J-Y, Abe-Ouchi A, Crucifix M, Driesschaert E, Fichefet Th, Hewitt CD, *et al.* 2007. Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum – Part 1: experiments and large-scale features. *Climate of the Past* 3: 261–277.

Brown JL, Hill DJ, Dolan AM, Carnaval AC, Haywood AM. 2018. PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Scientific Data* 5: 1–9.

Chen T, Guestrin C. 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA: ACM, 785–794.

Collins WD, Bitz CM, Blackmon ML, Bonan GB, Bretherton CS, Carton JA, Chang P, Doney SC, Hack JJ, Henderson TB, *et al.* 2006. The community climate system model version 3 (CCSM3). *Journal of Climate* 19: 2122–2143.

Danielson JJ, Gesch DB. 2011. *Global multi-resolution terrain elevation data 2010 (GMTED2010).* URL http://pubs.er.usgs.gov/publication/ofr20111073

Dowsett H, Dolan A, Rowley D, Moucha R, Forte AM, Mitrovica JX, Pound M, Salzmann U, Robinson M, Chandler M, *et al.* 2016. The PRISM4 (mid-Piacenzian) paleoenvironmental reconstruction. *Climate of the Past* 12: 1519–1538.

Friedman JH, Popescu BE. 2008. Predictive learning via rule ensembles. *The Annals of Applied Statistics* 2: 916–954.

Greenwell B M. 2017. pdp: an R Package for constructing partial dependence plots. *The R Journal* 9: 421–436.

Günther F, Fritsch S. 2010. Neuralnet: training of neural networks. *The R Journal* 2: 30–38.

Hengl T, Mendes de Jesus J, Heuvelink GBM, Ruiperez Gonzalez M, Kilibarda M, Blagotić A, Shangguan W, Wright MN, Geng X, Bauer-Marschallinger B, *et al.* 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* 12: e0169748.

Henrot A-J, François L, Favre E, Butzin M, Ouberdous M, Munhoven G. 2010. Effects of CO^2, continental distribution, topography and vegetation changes on the climate at the Middle Miocene: a model study. *Climate of the Past* 6: 675–694.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965–1978.

Hill DJ. 2015. The non-analogue nature of Pliocene temperature gradients. *Earth and Planetary Science Letters* 425: 232–241.

Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder HP, Kessler M. 2017. Climatologies at high resolution for the earth's land surface areas. *Scientific Data* 4: 170122.

Karger DN, Kessler M, Conrad O, Weigelt P, Kreft H, König C, Zimmermann NE. 2019. Why tree lines are lower on islands—Climatic and biogeographic effects hold the answer. *Global Ecology and Biogeography* 28: 839–850.

Keil P, Chase JM. 2019. Global patterns and drivers of tree diversity integrated across a continuum of spatial grains. *Nature Ecology & Evolution* 3: 390–399.

Körner C, Jetz W, Paulsen J, Payne D, Rudmann-Maurer K, M. Spehn E. 2017. A global inventory of mountains for bio-geographical applications. *Alpine Botany* 127: 1–15.

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

Lucas TCD. 2020. A translucent box: interpretable machine learning in ecology. *Ecological Monographs* 90: e01422.

New M, Lister D, Hulme M, Makin I. 2002. A high-resolution data set of surface climate over global land areas. *Climate Research* 21: 1–25.

Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GVN, Underwood EC, D'amico JA, Itoua I, Strand HE, Morrison JC, *et al.* 2001. Terrestrial ecoregions of the world: a new map of life on earth. *BioScience* 51: 933–938.

Owens HL, Guralnick R. 2019. climateStability: an R package to estimate climate stability from time-slice climatologies. *Biodiversity Informatics* 14: 8–13.

Ploton P, Mortier F, Réjou-Méchain M, Barbier N, Picard N, Rossi V, Dormann C, Cornu G, Viennois G, Bayol N, *et al.* 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications* 11: 4540.

Qian H, Jin Y. 2021. Are phylogenies resolved at the genus level appropriate for studies on phylogenetic structure of species assemblages? *Plant Diversity* 43: 255–263.

Ray N, Adams JM. 2001. A GIS-based vegetation map of the world at the last glacial maximum (25,000-15,000 BP). *Internet Archaeology* 11.

Sandel B, Weigelt P, Kreft H, Keppel G, van der Sande MT, Levin S, Smith S, Craven D, Knight TM. 2020. Current climate, isolation and history drive global patterns of tree phylogenetic endemism. *Global Ecology and Biogeography* 29: 4–15.

Svenning J-C, Eiserhardt WL, Normand S, Ordonez A, Sandel B. 2015. The influence of paleoclimate on present-day patterns in biodiversity and ecosystems. *Annual Review of Ecology, Evolution, and Systematics* 46: 551–572.

Takhtajan AL. 1986. *Floristic regions of the world.* Berkeley, CA: University of California Press.

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.

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.

Weigelt P, König C, Kreft H. 2020. GIFT – A global inventory of floras and traits for macroecology and biogeography. *Journal of Biogeography* 47: 16–43.

Whittaker RH. 1975. *Communities and ecosystems*. New York, US: Macmillan.

Zhao M, Running SW. 2010. Drought-induced reduction in global terrestrial net primary production from 2000 through 2009. *Science* 329: 940–943.

Zomer RJ, Trabucco A, Bossio DA, Verchot LV. 2008. Climate change mitigation: A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agriculture, Ecosystems & Environment* 126: 67–80.

Supporting Information for Chapter 2
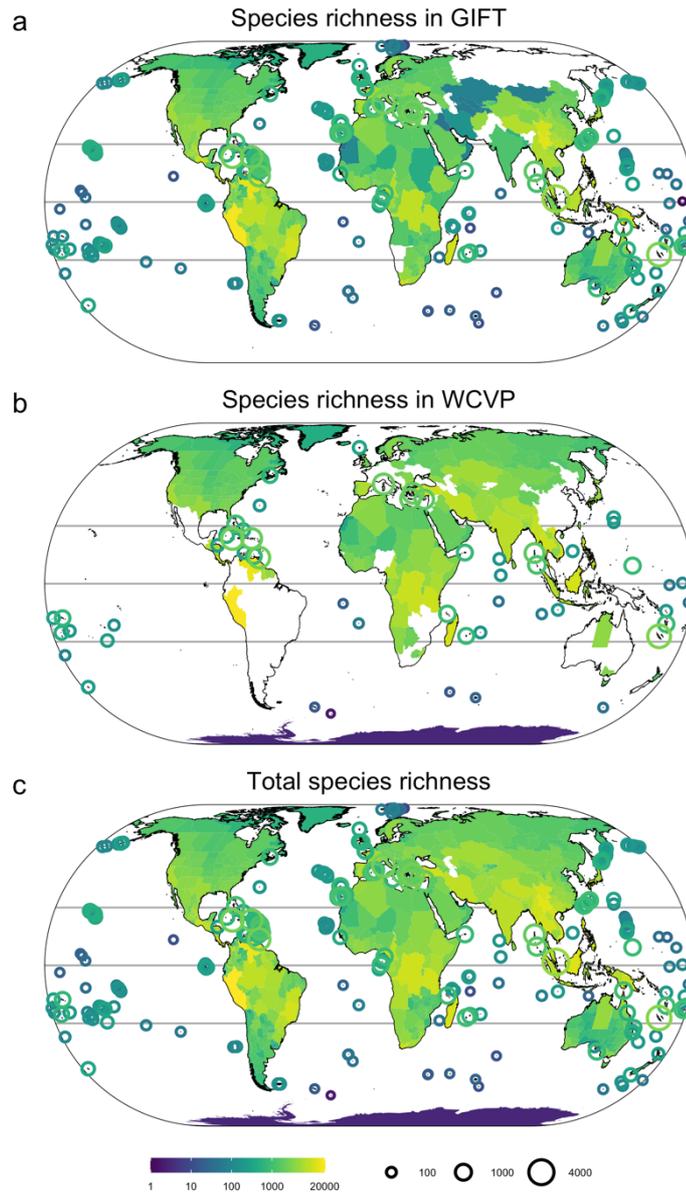
Figure S2.1 Observed species richness of seed plants for 912 geographic regions selected from (a) the Global Inventory of Floras and Traits (GIFT; 879 regions) (Weigelt *et al.*, 2020) and (b) the World Checklist of Vascular Plants (WCVP; 261 regions) (Govaerts *et al.*, 2021), combined in (c) to estimate phylogenetic endemism at the global scale and at a spatial grain as fine as possible. Log$_{10}$ scale is used for species richness and maps are shown in Eckert IV projection.
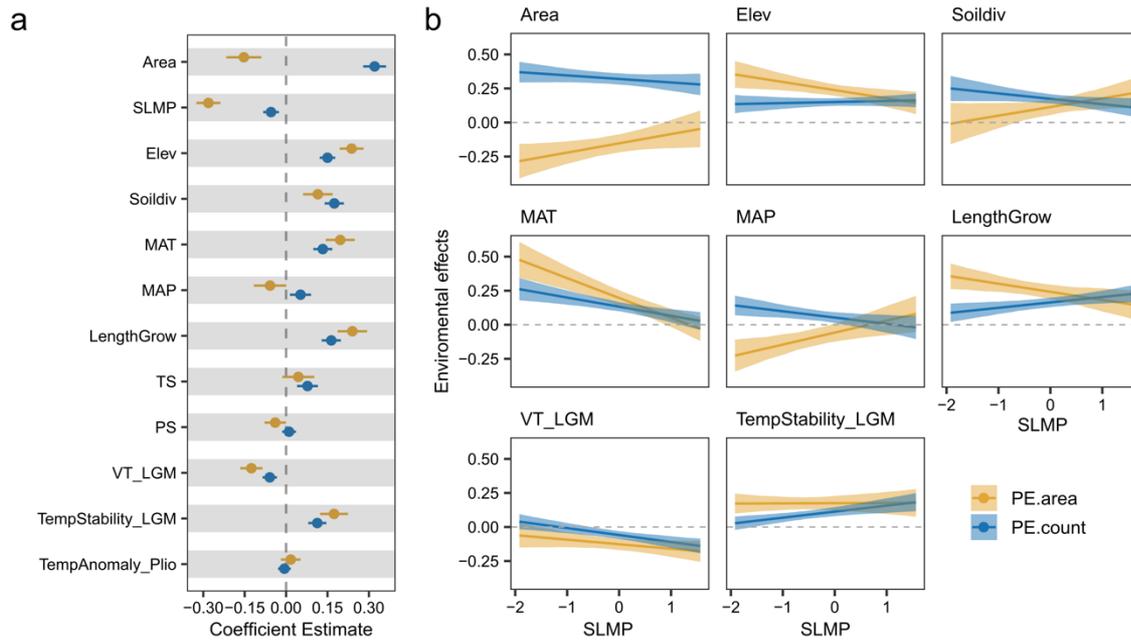
Figure S2.2 Determinants of phylogenetic endemism in seed plants based on the dataset with unplaced species added to the phylogeny and excluding apomictic taxa. Results are obtained from spatial models including environmental factors and interactions between each environmental factor and surrounding landmass proportion. a, standardized regression coefficients of individual environmental factors. Bars around each point show the standard error of the coefficient estimate. b, significant interaction terms in the models visualized as effects of environmental factors on phylogenetic endemism (model coefficients on y-axis) with varying surrounding landmass proportion (x-axis). Lines and shadings represent 95% confidence intervals. Results are shown for phylogenetic endemism based on two competing ways of measuring range size of species. PE.area (yellow) indicates phylogenetic endemism calculated based on range size of species as the area of regions where a species occurs, while PE.count (blue) is calculated based on range size of species as the count of these regions. Area = region area; SLMP = surrounding landmass proportion; Elev = elevational range; Soildiv = number of soil types; MAT = mean annual temperature; MAP= mean annual precipitation; LengthGrow = length of the growing season; TS = temperature seasonality; PS = precipitation seasonality; VT_LGM = velocity of temperature change since the Last Glacial Maximum; TempStability_LGM = temperature stability since the Last Glacial Maximum; TempAnomaly_Plio = temperature anomaly between the mid-Pliocene warm period and present-day.

Figure S2.3 Number (a) and proportion (b) of seed plant species in the dataset missing from the phylogeny (Smith & Brown, 2018) that were either added to the phylogeny replacing their genera by polytomies ("merged phylogeny") or excluded from analyses ("matched phylogeny").
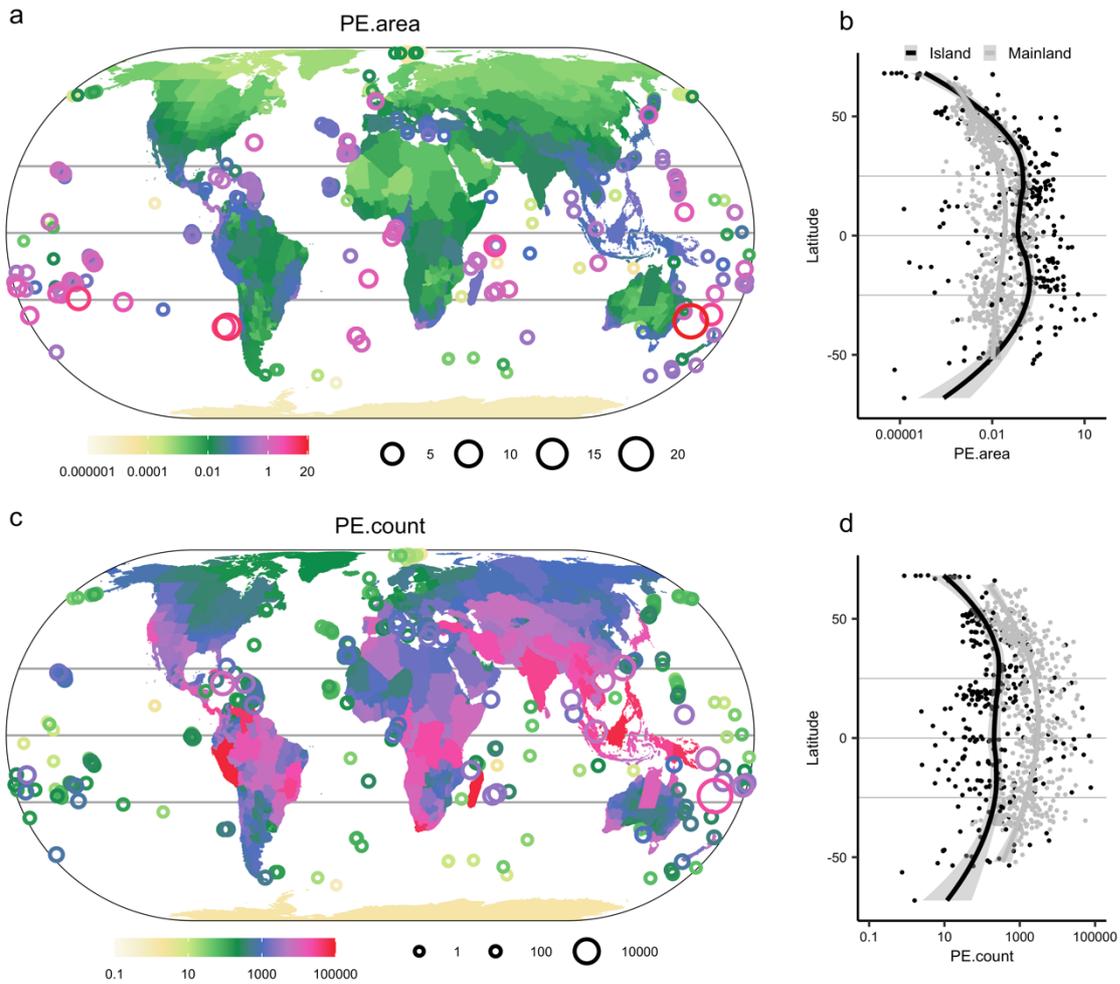
Figure S2.4 Global patterns of phylogenetic endemism of seed plants and its distribution along latitude based on the dataset with unplaced species added to the phylogeny and excluding apomictic taxa. In a and b, phylogenetic endemism is calculated based on species range size measured as the area of regions where a species occurs (PE.area); In c and d, phylogenetic endemism is calculated based on species range size measured as the count of regions where a species occurs (PE.count). In b and d, the fitted lines are lowess regressions, separately fitted for islands and mainland regions. Log$_{10}$ scale is used for phylogenetic endemism in all panels and maps are shown in Eckert IV projection.

Figure S2.5 Global patterns of phylogenetic endemism of seed plants and its distribution along latitude based on the dataset only retaining species that were included in the original phylogeny and including apomictic taxa. In a and b, phylogenetic endemism is calculated based on species range size measured as the area of regions where a species occurs (PE.area); In c and d, phylogenetic endemism is calculated based on species range size measured as the count of regions where a species occurs (PE.count). In b and d, the fitted lines are lowess regressions, separately fitted for islands and mainland regions. Log_{10} scale is used for phylogenetic endemism in all panels and maps are shown in Eckert IV projection.

Figure S2.6 Global patterns of phylogenetic endemism of seed plants and its distribution along latitude based on the dataset with unplaced species added to the phylogeny and including apomictic taxa. In a and b, phylogenetic endemism is calculated based on species range size measured as the area of regions where a species occurs (PE.area); In c and d, phylogenetic endemism is calculated based on species range size measured as the count of regions where a species occurs (PE.count). In b and d, the fitted lines are lowess regressions, separately fitted for islands and mainland regions. Log_{10} scale is used for phylogenetic endemism in all panels and maps are shown in Eckert IV projection.

Figure S2.7 Determinants of phylogenetic endemism in seed plants based on the datasets including apomictic taxa. Results are obtained from spatial models including environmental factors and interactions between each environmental factor and surrounding landmass proportion, based on the phylogenies without (a, b) and with unplaced species added to the original phylogeny (c, d). a and c, standardized regression coefficients of individual environmental factors. Bars around each point show the standard error of the coefficient estimate. b and d, significant interaction terms in the models visualized as effects of environmental factors on phylogenetic endemism (model coefficients on y-axis) with varying surrounding landmass proportion (x-axis). Lines and shadings represent 95% confidence intervals. Results are shown for phylogenetic endemism based on two competing ways of measuring range size of species. PE.area (yellow) indicates phylogenetic endemism calculated based on range size of species as the area of regions where a species occurs, while PE.count (blue) is calculated based on range size of species as the count of these regions. Area = region area; SLMP = surrounding landmass proportion; Elev = elevational range; Soildiv = number of soil types; MAT = mean annual temperature; MAP= mean annual precipitation; LengthGrow = length of the growing season; TS = temperature seasonality; PS = precipitation seasonality; VT_LGM = velocity of temperature change since the Last Glacial Maximum; TempStability_LGM = temperature stability since the Last Glacial Maximum; TempAnomaly_Plio = temperature anomaly between the mid-Pliocene warm period and present-day.

Figure S2.8 Global centers of neo- and paleoendemism for seed plants. Endemism centers are identified based on the datasets without and with unplaced species added to the phylogeny and excluding and including apomictic taxa. Colored regions present different types of endemism centers according to a categorical analysis of neo- and paleoendemism (CANAPE): violet, neoendemism; green, paleoendemism; yellow, mixed-endemism (i.e., neo- and paleoendemism); and brown indicating super-endemism (i.e., centers with both extremely high neo- and paleoendemism); beige, not significant. Patterns of neo- and paleoendemism are distinguished based on phylogenetic endemism with two competing ways of measuring species range size. PE.area indicates phylogenetic endemism calculated based on range size of species as the area of regions where a species occurs, while PE.count is calculated based on range size of species as the count of these regions. Maps are shown in Eckert IV projection.
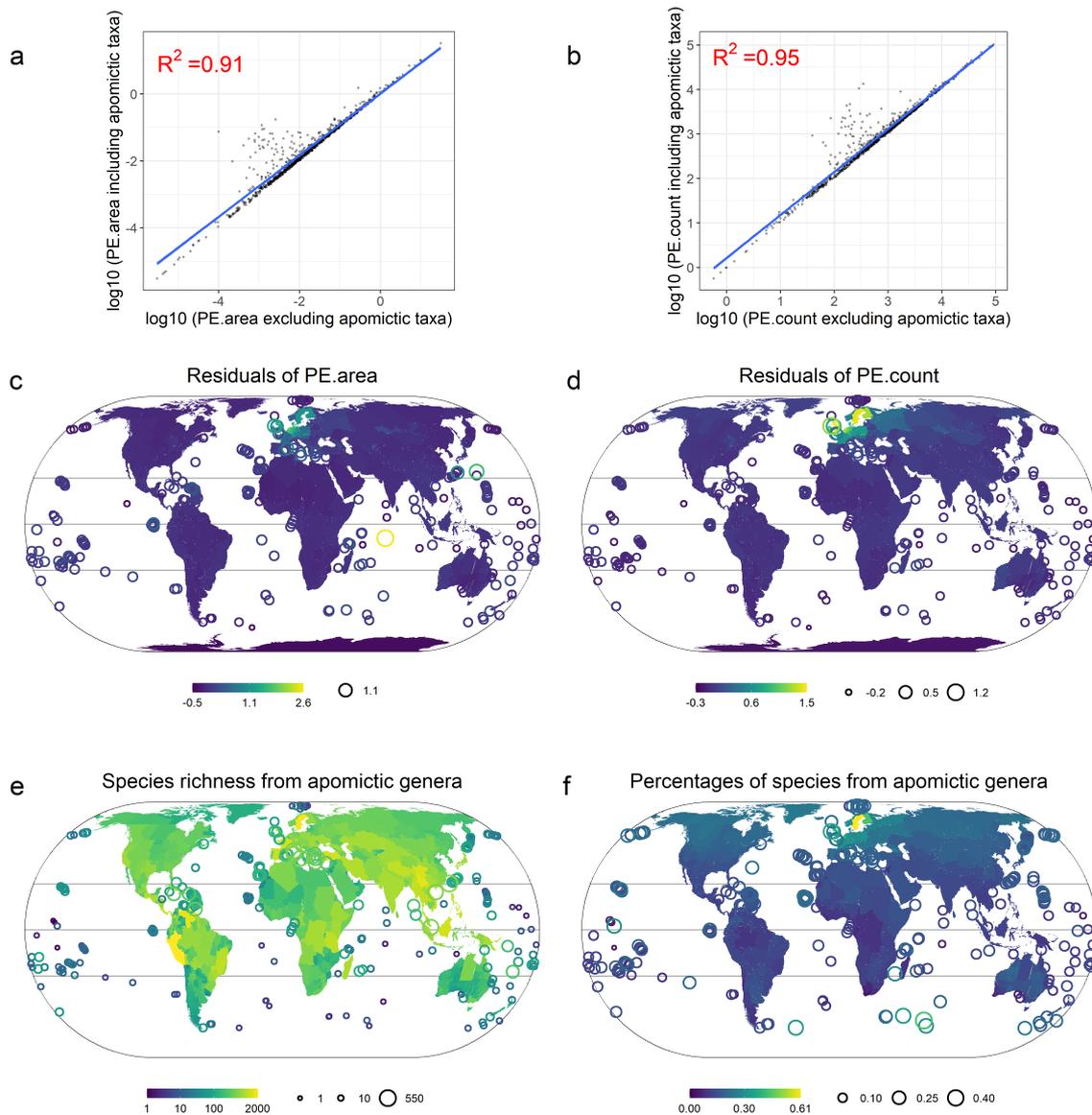
Figure S2.9 Comparison of phylogenetic endemism for seed plants based on species distribution data including and excluding apomictic taxa. a and b, the linear regression between phylogenetic endemism including and excluding apomictic taxa; c and d, residuals from the linear regression. Phylogenetic endemism is calculated using the phylogeny retaining only species originally included and based on two different ways to measure range size of each species: in a and c, as the total area of regions a species occurs in (PE.area); and in b and d, as the number of these regions (PE.count). Positive residuals in c and d indicate higher values of phylogenetic endemism based on the data including apomictic taxa than expected based on the data excluding apomictic taxa. e, species richness of seed plants from genera known to include apomictic species. f, percentages of species from genera known to include apomictic species. $\log_{10}$ scale is used in e and all maps are shown in Eckert IV projection.
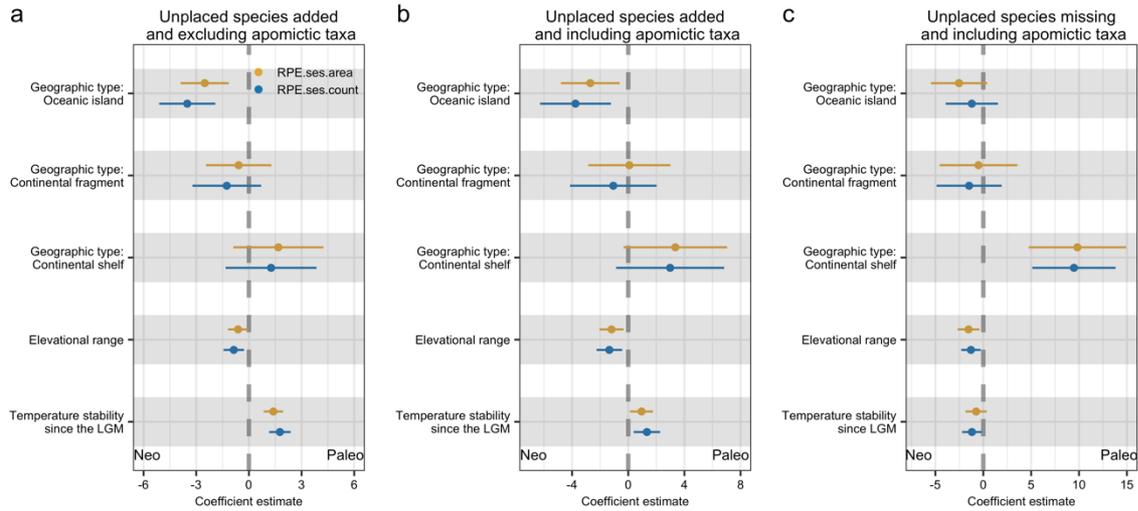
Figure S2.10 Determinants of neo- and paleoendemism. Standardized regression coefficients of environmental factors are shown from spatial models of the standardized effect size of relative phylogenetic endemism of seed plants for regions with significantly high phylogenetic endemism. A positive effect of environmental factors represents higher paleoendemism at higher values of the environmental factor, while a negative effect represents higher neoendemism. Results are shown for the datasets without and with unplaced species added to the original phylogeny and excluding and including apomictic taxa. RPE.ses.area (yellow) indicates the standardized effect size of relative phylogenetic endemism calculated based on range size of species as the area of regions where a species occurs, while RPE.ses.count (blue) is calculated based on range size of species as the count of these regions. The reference level of geographic type is mainland regions. LGM = Last Glacial Maximum.
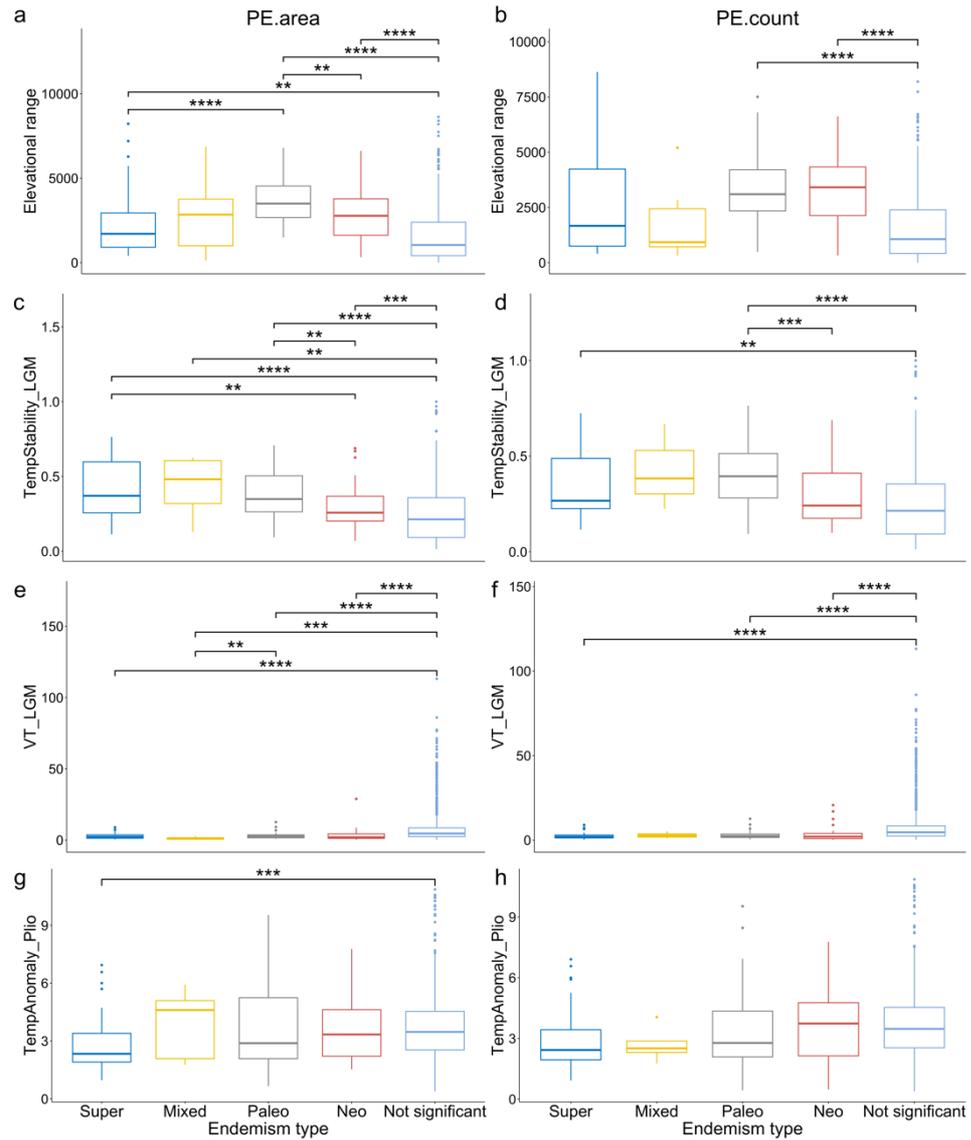
Figure S2.11 Distributions of environmental variables across regions of different endemism types based on the dataset only retaining species that were included in the original phylogeny and excluding apomictic taxa. Endemism types are identified using a categorical analysis of neo- and paleoendemism (CANAPE) based on phylogenetic endemism with two competing ways of measuring species range size. PE.area indicates phylogenetic endemism calculated based on range size of species as the area of regions where a species occurs (a, c, e and g), while PE.count is calculated based on range size of species as the count of these regions (b, d, f and h). Distributions of environmental variables of the sampling regions are compared using pairwise Wilcoxon tests. Asterisks indicate statistical significance (**: $p \leq 0.01$; ***: $p \leq 0.001$; ****: $p \leq 0.0001$). For each box, the middle horizontal line corresponds to the median; the lower and upper bounds of the box correspond to first and third quartiles, respectively. The upper whisker extends from the upper bound of the box to the highest value of the distribution, no further than $1.5 \times$ interquartile range (i.e., distance between the first and third quartiles). The lower whisker extends from the lower bound of the box to the lowest value of the distribution, no further than $1.5 \times$ interquartile range. Dots are values beyond the end of the whiskers ("outliers"). TempStability_LGM = temperature stability since the Last Glacial Maximum; VT_LGM = velocity of temperature change since the Last Glacial Maximum; TempAnomaly_Plio = temperature anomaly between the mid-Pliocene warm period and present-day.

Figure S2.12 Distributions of environmental variables across regions of different endemism types based on the dataset with unplaced species added to the phylogeny and excluding apomictic taxa. Endemism types are identified using a categorical analysis of neo- and paleoendemism (CANAPE) based on phylogenetic endemism with two competing ways of measuring species range size. PE.area indicates phylogenetic endemism calculated based on range size of species as the area of regions where a species occurs (a, c, e and g), while PE.count is calculated based on range size of species as the count of these regions (b, d, f and h). Distributions of environmental variables of the sampling regions are compared using pairwise Wilcoxon tests. Asterisks indicate statistical significance (**: $p <= 0.01$; ***: $p <= 0.001$; ****: $p <= 0.0001$). For each box, the middle horizontal line corresponds to the median; the lower and upper bounds of the box correspond to first and third quartiles, respectively. The upper whisker extends from the upper bound of the box to the highest value of the distribution, no further than $1.5 \times$ interquartile range (i.e., distance between the first and third quartiles). The lower whisker extends from the lower bound of the box to the lowest value of the distribution, no further than $1.5 \times$ interquartile range. Dots are values beyond the end of the whiskers ("outliers"). TempStability_LGM = temperature stability since the Last Glacial Maximum; VT_LGM = velocity of temperature change since the Last Glacial Maximum; TempAnomaly_Plio = temperature anomaly between the mid-Pliocene warm period and present-day.

Figure S2.13 Comparison of phylogenetic endemism for seed plants with two competing ways of measuring species range size. PE.area indicates phylogenetic endemism calculated based on range size of species as the area of regions where a species occurs, while PE.count is calculated based on range size of species as the count of these regions. Phylogenetic endemism is calculated based on the datasets without (matched phylogeny) and with (merged phylogeny) unplaced species added to the phylogeny, and excluding and including apomictic taxa.
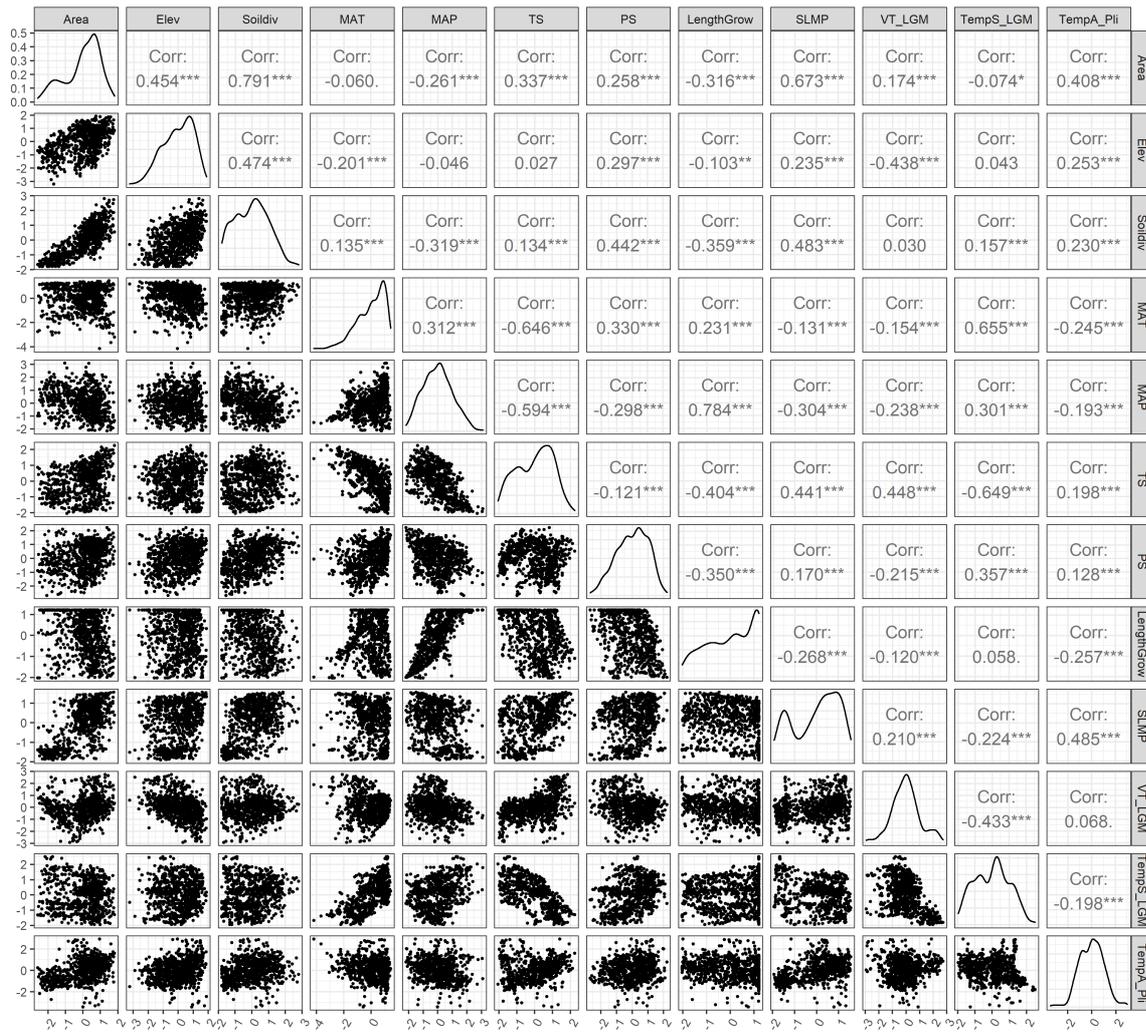
Figure S2.14 Correlations among all predictors and their density distributions after transformation. Numbers are Pearson correlation coefficients. Some predictors (i.e., Area, Elev, Soildiv, MAP, TS, PS, VT_LGM, TempS_LGM; TempA_Pli) are shown in log-scale as they were log-transformed for models owing to their skewed distributions. Area = region area; SLMP = surrounding landmass proportion; Elev = elevational range; Soildiv = number of soil types; MAT = mean annual temperature; MAP = mean annual precipitation; LengthGrow = length of the growing season; TS = temperature seasonality; PS = precipitation seasonality; VT_LGM = velocity of temperature change since the Last Glacial Maximum; TempS_LGM = temperature stability since the Last Glacial Maximum; TempA_Pli = temperature anomaly since the mid-Pliocene.
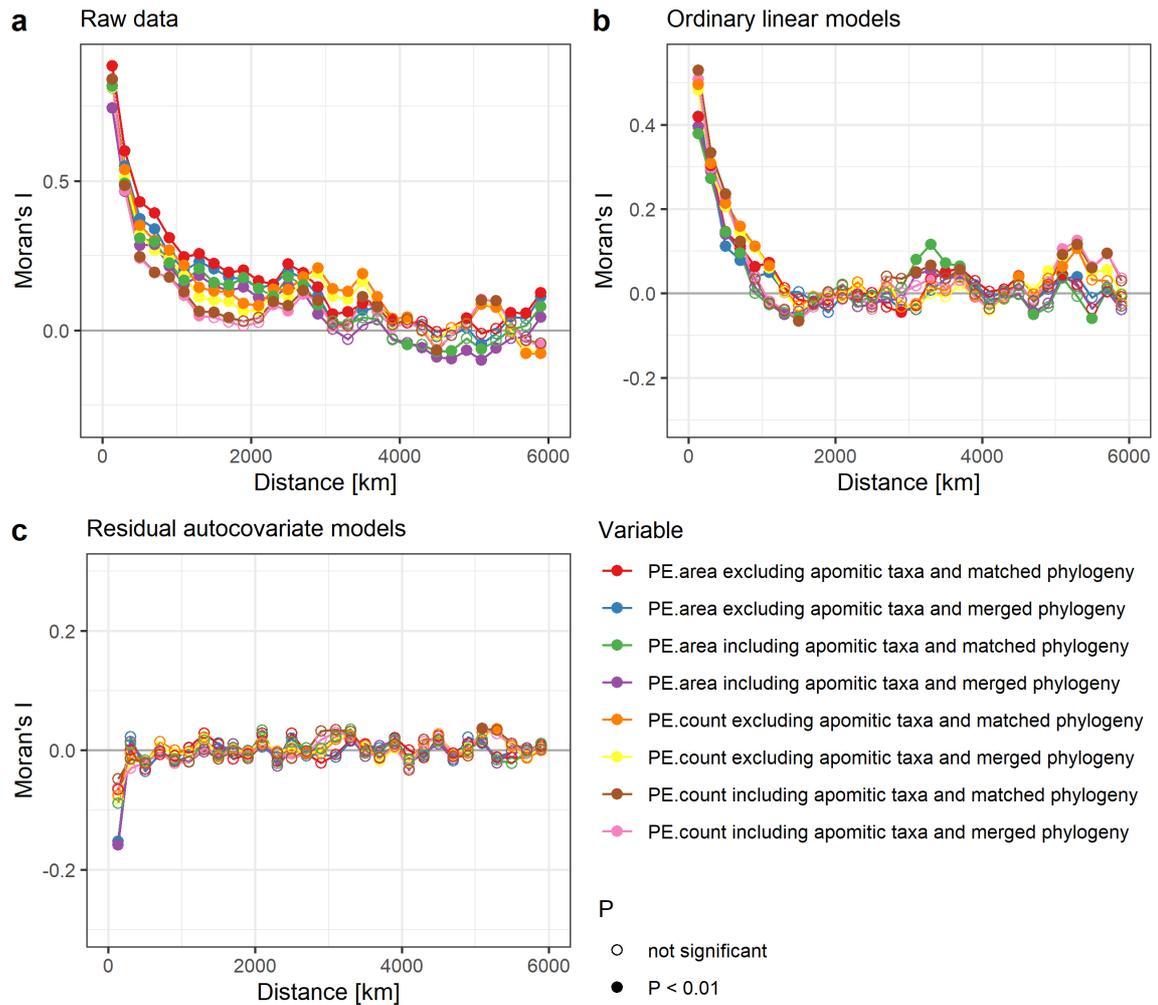
Figure S2.15 Spatial correlograms of raw phylogenetic endemism data (a), residuals from linear regression models for plant phylogenetic endemism (b) and from residual autocovariate models (c). Full symbols indicate a significant Moran's I at a given lag distance ($P < 0.01$). Phylogenetic endemism is calculated based on the datasets without (matched phylogeny) and with (merged phylogeny) unplaced species added to the phylogeny, and excluding and including apomictic taxa. PE.area indicates phylogenetic endemism calculated based on species range size as the area of regions where a species occurs, while PE.count is calculated based on species range size as the count of these regions.
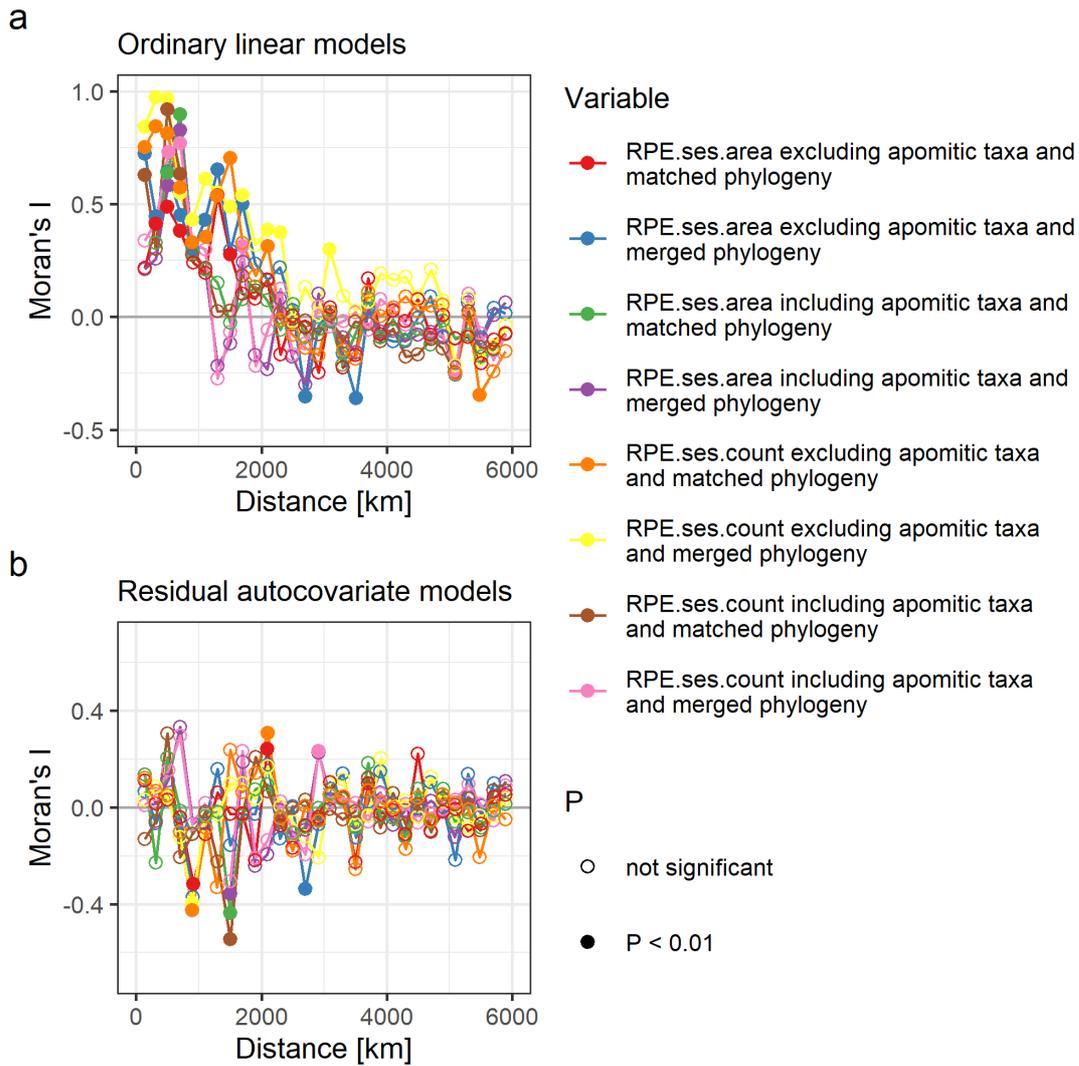
Figure S2.16 Spatial correlograms of residuals from linear regression models of the standardized effect size of relative phylogenetic endemism (a) and from residual autocovariate models (b). Full symbols indicate a significant Moran's I at a given lag distance (P < 0.01). Standardized effect size of relative phylogenetic endemism is calculated based on the datasets without (matched phylogeny) and with (merged phylogeny) unplaced species added to the phylogeny, and excluding and including apomictic taxa. RPE.ses.area indicates standardized effect size of relative phylogenetic endemism calculated based on range size of species as the area of regions where a species occurs, while RPE.ses.count is calculated based on range size of species as the count of these regions.

Table S2.1 Hypotheses and related predictors of phylogenetic endemism in plants. ↑ represents effects on phylogenetic endemism hypothesized to be positive, while ↓ represents negative effects.

Factor category	Variables	Hypothesis
Isolation	Surrounding landmass proportion (↓). A higher surrounding landmass proportion indicates a lower degree of isolation (Weigelt & Kreft, 2013). Islands are characterized by a low surrounding landmass proportion.	Isolation fosters allopatric speciation and limits range expansion due to limited gene flow and dispersal, which in turn promotes endemism (Kier *et al.*, 2009; Sandel *et al.*, 2020).
Environmental heterogeneity	Elevational range (↑); number of soil types (↑)	Heterogeneous regions include small habitats supporting more narrow-ranged species, allowing for geographic isolation promoting specialization, and serving as refugia during unfavorable climate change periods (Stein *et al.*, 2014; McFadden *et al.*, 2019).
Energy and water availability	Mean annual temperature (↑); mean annual precipitation (↑); length of the growing season (↑)	Warm and humid climates are hypothesized to support larger populations in small regions by offering sufficient resources, which promotes long-term survival of spatially restricted species and their accumulation over long timescales (Jetz *et al.*, 2004). Additionally, high energy availability may increase the opportunity for speciation, which promotes endemism (Rohde, 1992; Mittelbach *et al.*, 2007).
Climatic seasonality	Temperature seasonality (↓); precipitation seasonality (↓)	High climatic seasonality selects for species with broader climatic tolerances and larger ranges (Stevens, 1989).
Long-term climatic stability	Temperature stability since the Last Glacial Maximum (LGM) (↑); velocity of temperature change since the LGM (↓); temperature anomaly since the mid-Pliocene warm period (↓)	Long-term climatic stability allows for the evolution of narrow physiological tolerances and specialization, and reduces extinction risk of small-ranged species (Jansson, 2003; Sandel *et al.*, 2011; Enquist *et al.*, 2019)

Table S2.2 Hypotheses and related predictors of neo- and paleoendemism in plants.

Factor category	Variables	Hypotheses
Geologic origin	Geographic type of a region based on its geologic origins, i.e., continental shelf islands, continental fragments, oceanic islands, or mainland regions.	Oceanic islands host higher neoendemism, while continental fragments are centers of paleoendemism (Gillespie & Roderick, 2002).
Mountain regions	Elevational range	On the one hand, mountain regions are centers of neoendemism, due to high speciation rates that are driven by long-term orogenic and climatic dynamics in mountains (Antonelli *et al.*, 2018; Rahbek *et al.*, 2019); on the other hand, mountain regions foster paleoendemism, because mountain regions promote the persistence of ancient lineages during unfavorable climate change periods (Bennett *et al.*, 1991). The two processes may be not mutually exclusive, leading to mountain regions as centers of both neo- and paleoendemism (Dagallier *et al.*, 2020).
Past climate change	Temperature stability since the Last Glacial Maximum (LGM); velocity of temperature change since the LGM; temperature anomaly since the mid-Pliocene warm period	Regions with stable climates have suffered less severe environmental changes across space and may have acted as refugia where plants could persist over time, and host higher paleoendemism (Jump *et al.*, 2009).

Table S2.3 Top ten regions of phylogenetic endemism for seed plants based on the dataset only retaining species that were included in the original phylogeny and excluding apomictic taxa. Phylogenetic endemism is calculated based on two different ways to measure range size of each species: (i) as the total area of regions a species occurs in (PE.area) and (ii) as the number of these regions (PE.count).

PE.area								PE.count			
Mainland	PE	Island >10 km²	PE	Island >100 km²	PE	Island >1000 km²	PE	Mainland	PE	Island	PE
Pichincha, Ecuador	0.57	Lord Howe Island	30.80	Mahé	3.03	Caroline Islands	2.70	Peru	82911	Madagascar	91364
Carchi, Ecuador	0.46	Tubuai Island	9.97	Caroline Islands	2.70	Mauritius	1.37	Venezuela	55927	Papua New Guinea	80038
Costa Rica	0.46	Masatierra	9.73	St. Helena	2.02	New Caledonia	1.26	Western Cape, South Africa	49783	Borneo	76619
Panama	0.38	Masafuera	7.12	Raiatea	1.59	Tahiti	1.08	Minas Gerais	37719	Philippines	54356
Western Cape, South Africa	0.36	Norfolk Island	5.06	Rodrigues	1.56	La Réunion	0.97	Vietnam	35438	Indonesian New Guinea	44115
Antioquia, Colombia	0.31	Silhouette	3.56	Mauritius	1.37	Samoa	0.80	Thailand	34041	Sumatra	33862
Imbabura, Ecuador	0.31	Mahé	3.03	New Caledonia	1.26	Kaua'i	0.74	Yunnan	30380	Cuba	27411
Valle del Cauca, Colombia	0.28	Caroline Islands	2.70	Hiva Oa	1.20	Jamaica	0.71	Turkey asiatic	29485	New Caledonia	24061
Quindío, Colombia	0.28	Rarotonga	2.54	Tahiti	1.08	Comoros	0.68	Tanzania	28554	Hispaniola	20486
Rio de Janeiro, Brazil	0.27	Tristan da Cunha	2.15	Principe	1.00	Tenerife	0.62	India	28054	Sulawesi	19821

Table S2.4 Summary statistics of spatial models showing the effects of environmental factors on phylogenetic endemism of seed plants. Models are fitted for phylogenetic endemism based on two competing ways of measuring range size of species and the datasets without (matched phylogeny) and with (merged phylogeny) unplaced species added to the phylogeny and excluding and including apomictic taxa. PE.area indicates phylogenetic endemism calculated based on range size of species as the total area of regions where a species occurs, while PE.count is calculated based on range size of species as the count of these regions. Area = region area; SLMP = surrounding landmass proportion; Elev = elevational range; Soildiv = number of soil types; MAT = mean annual temperature; MAP = mean annual precipitation; LengthGrow = length of the growing season; TS = temperature seasonality; PS = precipitation seasonality; VT_LGM = velocity of temperature change since the Last Glacial Maximum; TempStability_LGM = temperature stability since the Last Glacial Maximum; TempAnomaly_Plio = temperature anomaly between the mid-Pliocene warm period and present-day; RAC= spatial autocovariate. Combinations of predictors separated by ":" indicate interactions between each predictor and SLMP. std. Error = Standard error. Sig = statistical significance (NS: P>0.05; *: p <= 0.05; **: p <= 0.01; ***: p <= 0.001).

| Models | Excluding apomictic taxa and matched phylogeny | | | | | | | | | | | | Including apomictic taxa and matched phylogeny | | | | | | | | | | | | Excluding apomictic taxa and merged phylogeny | | | | | | | | | | | | Including apomictic taxa and merged phylogeny | | | | | | |
| | PE.area | | | PE.count | | | PE.area | | | PE.count | | | PE.area | | | PE.count | | | PE.area | | | PE.count | | | PE.area | | | PE.count | | | PE.area | | | PE.count | | |
	Estimate	std. Error	Sig	Estimate	std. Error	Sig	Estimate	std. Error	Sig	Estimate	std. Error	Sig	Estimate	std. Error	Sig	Estimate	std. Error	Sig	Estimate	std.	Sig	Estimate	std.	Sig
Intercept	-1.85	0.02	***	2.98	0.01	***	-1.67	0.02	***	3.11	0.01	***	-1.80	0.02	***	3.00	0.01	***	-1.65	0.02	***	3.11	0.01	***
Area	-0.13	0.04	***	0.32	0.02	***	-0.14	0.04	***	0.33	0.02	***	-0.15	0.03	***	0.32	0.02	***	-0.19	0.03	***	0.33	0.02	***
SLMP	-0.22	0.02	***	-0.03	0.02	NS	-0.20	0.02	***	-0.03	0.02	NS	-0.28	0.02	***	-0.05	0.01	***	-0.27	0.02	***	-0.06	0.01	***
Elev	0.28	0.02	***	0.16	0.02	***	0.26	0.02	***	0.15	0.02	***	0.24	0.02	***	0.15	0.01	***	0.23	0.02	***	0.15	0.01	***
Soildiv	0.11	0.03	***	0.18	0.02	***	0.13	0.03	***	0.20	0.02	***	0.12	0.03	***	0.17	0.02	***	0.14	0.03	***	0.19	0.02	***
MAT	0.22	0.03	***	0.15	0.02	***	0.19	0.03	***	0.12	0.02	***	0.20	0.03	***	0.13	0.02	***	0.16	0.03	***	0.10	0.02	***
MAP	-0.07	0.03	*	0.06	0.02	**	-0.16	0.03	***	-0.00	0.02	NS	-0.06	0.03	NS	0.05	0.02	**	-0.12	0.03	***	0.00	0.02	NS
LengthGrow	0.24	0.03	***	0.16	0.02	***	0.32	0.03	***	0.21	0.02	***	0.24	0.03	***	0.16	0.02	***	0.29	0.03	***	0.21	0.02	***
TS	-0.00	0.03	NS	0.06	0.02	**	-0.07	0.03	*	-0.00	0.02	NS	0.04	0.03	NS	0.08	0.02	***	-0.01	0.03	NS	0.03	0.02	NS
PS	-0.04	0.02	NS	0.00	0.01	NS	-0.08	0.02	***	-0.02	0.01	NS	-0.04	0.02	*	0.01	0.01	NS	-0.06	0.02	**	-0.00	0.01	NS
VT_LGM	-0.11	0.02	***	-0.06	0.01	***	-0.09	0.02	***	-0.04	0.01	**	-0.13	0.02	***	-0.06	0.01	***	-0.11	0.02	***	-0.04	0.01	**
TempStability_LGM	0.21	0.03	***	0.13	0.02	***	0.09	0.03	**	0.05	0.02	**	0.17	0.03	***	0.11	0.02	***	0.07	0.03	**	0.05	0.02	**
TempAnomaly_Plio	-0.01	0.02	NS	-0.01	0.01	NS	-0.02	0.02	NS	-0.01	0.01	NS	0.02	0.02	NS	-0.01	0.01	NS	0.01	0.02	NS	-0.01	0.01	NS

RAC	0.81	0.03	***	0.88	0.04	***	0.79	0.04	***	0.86	0.04	***	0.81	0.03	***	0.90	0.04	***	0.82	0.03	***	0.88	0.04	***
Area:SLMP	0.06	0.03	NS	-0.02	0.02	NS	0.01	0.03	NS	-0.03	0.02	NS	0.07	0.03	*	-0.03	0.02	NS	0.04	0.03	NS	-0.04	0.02	NS
Elev:SLMP	-0.10	0.02	***	0.00	0.02	NS	-0.11	0.02	***	-0.01	0.02	NS	-0.06	0.02	**	0.01	0.01	NS	-0.07	0.02	**	0.00	0.01	NS
Soildiv:SLMP	0.06	0.03	NS	-0.04	0.02	NS	0.10	0.03	**	-0.03	0.02	NS	0.06	0.03	*	-0.04	0.02	NS	0.08	0.03	*	-0.03	0.02	NS
MAT:SLMP	-0.12	0.03	***	-0.07	0.02	***	-0.14	0.03	***	-0.08	0.02	***	-0.15	0.03	***	-0.07	0.02	***	-0.16	0.03	***	-0.08	0.02	***
MAP:SLMP	0.15	0.03	***	-0.04	0.02	*	0.10	0.03	***	-0.08	0.02	***	0.09	0.03	***	-0.05	0.02	**	0.03	0.03	NS	-0.07	0.02	***
LengthGrow:SLMP	-0.10	0.03	***	0.03	0.02	NS	-0.09	0.03	***	0.04	0.02	*	-0.06	0.03	*	0.04	0.02	*	-0.03	0.03	NS	0.05	0.02	***
TS:SLMP	0.05	0.03	NS	-0.05	0.02	*	-0.02	0.03	NS	-0.08	0.02	***	-0.03	0.03	NS	-0.05	0.02	**	-0.08	0.03	**	-0.08	0.02	***
PS:SLMP	-0.07	0.03	**	-0.01	0.02	NS	-0.09	0.03	***	-0.03	0.02	NS	-0.06	0.02	**	-0.01	0.02	NS	-0.08	0.02	***	-0.02	0.01	NS
VT_LGM:SLMP	-0.08	0.02	***	-0.06	0.01	***	-0.11	0.02	***	-0.08	0.01	***	-0.03	0.02	NS	-0.05	0.01	***	-0.05	0.02	**	-0.06	0.01	***
TempStability LGM:SLMP	0.00	0.02	NS	0.05	0.02	***	-0.07	0.02	**	0.01	0.02	NS	0.00	0.02	NS	0.04	0.01	**	-0.04	0.02	NS	0.02	0.01	NS
TempAnomaly Plio:SLMP	0.05	0.02	*	0.00	0.01	NS	0.04	0.02	NS	0.00	0.01	NS	0.02	0.02	NS	0.00	0.01	NS	0.02	0.02	NS	0.00	0.01	NS
Observations	818			818			818			818			818			818			818			818		
R^2 / R^2	0.804 / 0.798			0.880 / 0.877			0.781 / 0.774			0.877 / 0.874			0.809/ 0.803			0.885 / 0.881			0.794 / 0.788			0.884 / 0.880		

Table S2.5 Top ten regions of phylogenetic endemism for seed plants based on the dataset with unplaced species added to the phylogeny and excluding apomictic taxa. Phylogenetic endemism is calculated based on two different ways to measure range size of each species: (i) as the total area of regions a species occurs in (PE.area) and (ii) as the number of these regions (PE.count).

PE.area								PE.count			
Mainland	PE	Island> 10 km²	PE	Island >100 km²	PE	Island >1000 km²	PE	Mainland	PE	Island	PE
Costa Rica	0.44	Lord Howe Island	21.65	Mahé	2.92	Caroline Islands	1.93	Peru	77729	Madagascar	76964
Pichincha, Ecuador	0.44	Masatierra	9.37	St. Helena	2.06	New Caledonia	1.27	Western Cape	55197	Borneo	69591
Western Cape	0.40	Tubuai Island	6.16	Caroline Islands	1.93	Mauritius	1.10	Venezuela	47547	Papua New Guinea	49859
Panama	0.40	Masafuera	5.34	Raiatea	1.37	Kaua'i	0.96	Vietnam	34535	Philippines	45383
Carchi, Ecuador	0.30	Norfolk Island	4.78	Rodrigues	1.27	Tahiti	0.83	Thailand	32192	Indonesian New Guinea	29119
Valle del Cauca, Colombia	0.24	Silhouette	3.92	New Caledonia	1.27	La Réunion	0.77	Minas Gerais	30592	Sumatra	27187
Antioquia, Colombia	0.24	Mahé	2.92	Mauritius	1.10	O'ahu Island	0.66	India	30181	New Caledonia	24092
Rio de Janeiro, Brazil	0.23	Henderson	2.24	Príncipe	1.08	Samoa	0.60	Panama	29108	Cuba	23675
Imbabura, Ecuador	0.22	Fulanga	2.09	Sao Tomé	1.02	Tenerife	0.56	Yunnan	28127	Japan	16511
Cotopaxi, Ecuador	0.22	Rarotonga	2.09	Madeira	0.96	Jamaica	0.56	Asiatic Turkey	26746	Hispaniola	14738

Table S2.6 Top ten regions of phylogenetic endemism for seed plants based on the dataset only retaining species that were included in the original phylogeny and including apomictic taxa. Phylogenetic endemism is calculated based on two different ways to measure range size of each species: (i) as the total area of regions a species occurs in (PE.area) and (ii) as the number of these regions (PE.count).

PE.area									PE.count			
Mainland	PE	Island >10 km²	PE	Island >100 km²	PE	Island >1000 km²	PE		Mainland	PE	Island	PE
Pichincha, Ecuador	0.61	Lord Howe Island	32.79	Mahé	3.23	Caroline Islands	2.97		Peru	93046	Madagascar	103183
Costa Rica	0.54	Masatierra	10.30	Caroline Islands	2.97	Mauritius	1.83		Venezuela	61270	Borneo	91674
Carchi, Ecuador	0.48	Tubuai Island	10.02	St. Helena	2.36	La Réunion	1.50		Western Cape	50550	Papua New Guinea	89316
Panama	0.47	Masafuera	7.17	Mauritius	1.83	New Caledonia	1.45		Vietnam	45382	Philippines	68921
Western Cape	0.36	Norfolk Island	6.96	Rodrigues	1.68	Tahiti	1.09		Thailand	40724	Indonesian New Guinea	49085
Antioquia, Colombia	0.33	Silhouette	3.77	Raiatea	1.63	Samoa	1.07		Minas Gerais	40136	Sumatra	39141
Valle del Cauca, Colombia	0.33	Mahé	3.23	La Réunion	1.50	Comoros	0.89		Turkey asiatic	37637	Cuba	33527
Imbabura, Ecuador	0.33	Rarotonga	3.00	New Caledonia	1.45	Jamaica	0.88		Yunnan	37157	New Caledonia	27592
Rio de Janeiro, Brazil	0.31	Caroline Islands	2.97	Hiva Oa	1.23	Kaua'i	0.85		Panama	33685	Hispaniola	24531
Loja, Ecuador	0.29	St. Helena	2.36	Christmas Island	1.14	Viti Levu	0.67		India	33420	Sulawesi	21988

Table S2.7 Top ten regions of phylogenetic endemism for seed plants based on the dataset with unplaced species added to the phylogeny and including apomictic taxa. Phylogenetic endemism is calculated based on two different ways to measure range size of each species: (i) as the total area of regions a species occurs in (PE.area) and (ii) as the number of these regions (PE.count).

PE.area								PE.count			
Mainland	PE	Island >10 km²	PE	Island >100 km²	PE	Island >1000 km²	PE	Mainland	PE	Island	PE
Costa Rica	0.49	Lord Howe Island	26.56	Mahé	3.03	Caroline Islands	2.10	Peru	86902	Madagascar	90040
Pichincha, Ecuador	0.47	Masatierra	10.13	St. Helena	2.33	Mauritius	1.44	Western Cape	56603	Borneo	76810
Panama	0.43	Tubuai Island	6.21	Caroline Islands	2.10	New Caledonia	1.43	Venezuela	51928	Papua New Guinea	55351
Western Cape	0.41	Norfolk Island	6.01	Mauritius	1.44	La Réunion	1.40	Vietnam	37975	Philippines	51794
Carchi, Ecuador	0.32	Masafuera	5.64	New Caledonia	1.43	Kaua'i	1.05	Thailand	34753	Indonesian New Guinea	31787
Valle del Cauca, Colombia	0.26	Silhouette	4.04	Raiatea	1.40	Tahiti	0.84	India	34545	Sumatra	29958
Rio de Janeiro, Brazil	0.26	Mahé	3.03	La Réunion	1.40	Samoa	0.75	Asiatic Turkey	34509	Cuba	27660
Antioquia, Colombia	0.25	Rarotonga	2.54	Rodrigues	1.36	O'ahu Island	0.75	Yunnan	33386	New Caledonia	27282
Quindío, Colombia	0.25	St. Helena	2.33	Madeira	1.13	Comoros	0.73	Minas Gerais	32888	Japan	21096
Loja, Ecuador	0.24	Henderson	2.30	Príncipe	1.08	Tenerife	0.67	Panama	31249	Hispaniola	17330

Table S2.8 Summary statistics of spatial models showing the effects of environmental factors on the standardized effect size of relative phylogenetic endemism (RPE.ses) of seed plants for regions with significantly high phylogenetic endemism. Models are fitted for RPE.ses based on two competing ways of measuring range size of species and the datasets without (matched phylogeny) and with (merged phylogeny) unplaced species added to the phylogeny and excluding and including apomictic taxa. RPE.ses.area indicates the standardized effect size of relative phylogenetic endemism calculated based on range size of species as the area of regions where a species occurs, while RPE.ses.count is calculated based on range size of species as the count of these regions. Geographic type of each region distinguishes between mainland regions, continental shelf islands, continental fragments and oceanic islands. The reference level of geographic type is mainland regions. TempStability_LGM = temperature stability since the Last Glacial Maximum; RAC= spatial autocovariate. std. E = Standard error. Sig = statistical significance (NS: P>0.05; *: p <= 0.05; **: p <= 0.01; ***: p <= 0.001).

Models	Excluding apomictic taxa and matched phylogeny						Including apomictic taxa and matched phylogeny						Excluding apomictic taxa and merged phylogeny						Including apomictic taxa and merged phylogeny					
	PE.area			PE.count			PE.area			PE.count			PE.area			PE.count			PE.area			PE.count		
	Estimate	std. E	Sig	Estimate	std. E	Sig	Estimate	std. E	Sig	Estimate	std. E	Sig	Estimate	std. E	Sig	Estimate	std. E	Sig	Estimate	std. E	Sig	Estimate	std. E	Sig
Intercept	-0.07	0.43	NS	1.25	0.36	***	0.92	0.67	NS	1.60	0.60	**	0.84	0.35	*	1.37	0.35	***	0.61	0.53	NS	1.20	0.55	*
Oceanic	-1.79	0.91	NS	-1.95	0.82	*	-2.53	1.50	NS	-1.20	1.39	NS	-2.52	0.70	***	-3.51	0.82	***	-2.70	1.07	*	-3.76	1.29	**
Continental	0.22	1.31	NS	0.15	1.06	NS	-0.50	2.07	NS	-1.48	1.73	NS	-0.59	0.95	NS	-1.26	1.00	NS	0.08	1.49	NS	-1.06	1.57	NS
Continental	2.72	1.76	NS	1.62	1.40	NS	9.83	2.60	***	9.47	2.22	***	1.68	1.31	NS	1.26	1.32	NS	3.36	1.88	NS	2.98	1.96	NS
Elevational	0.20	0.37	NS	0.07	0.31	NS	-1.54	0.58	**	-1.29	0.51	*	-0.62	0.29	*	-0.86	0.30	**	-1.19	0.44	**	-1.34	0.46	**
TempStabilit	1.41	0.35	***	1.56	0.31	***	-0.75	0.57	NS	-1.19	0.52	*	1.39	0.28	***	1.76	0.31	***	0.95	0.42	*	1.33	0.48	**
RAC	0.26	0.03	***	0.31	0.02	***	0.26	0.02	***	0.34	0.03	***	0.28	0.03	***	0.28	0.03	***	0.24	0.03	***	0.25	0.04	***
Observations	154			141			163			153			138			135			145			140		
R^2 / R^2	0.482 / 0.461			0.631 / 0.615			0.532 / 0.515			0.600 / 0.583			0.570 / 0.550			0.615 / 0.597			0.341 / 0.312			0.324 / 0.293		

Table S2.9 Linear model results for phylogenetic endemism of seed plants including either surrounding landmass proportion or whether a region is an island or not as a covariable. Models are fitted for phylogenetic endemism based on two different methods used to quantify species range size and the datasets without (matched phylogeny) and with (merged phylogeny) unplaced species added to the phylogeny and excluding and including apomictic taxa. The model performances are quantified by Akaike's information criterion (AIC) and adjusted coefficient of determination (R^2). PE.area indicates phylogenetic endemism calculated based on range size of species as the area of regions where a species occurs, while PE.count is calculated based on range size of species as the count of these regions. Models include region area, elevational range, number of soil types, mean annual temperature, mean annual precipitation, length of the growing season, temperature seasonality, precipitation seasonality, temperature stability since the Last Glacial Maximum, velocity of temperature change since the Last Glacial Maximum, temperature anomaly between the mid-Pliocene warm period and present-day, and either surrounding landmass proportion (SLMP) or whether or not a region is an island (Geo_class) as predictor variables as well as all interaction terms between SLMP or Geo_class and all other predictors.

Models	Excluding apomictic taxa and matched phylogeny				Including apomictic taxa and matched phylogeny				Excluding apomictic taxa and merged phylogeny				Including apomictic taxa and merged phylogeny			
	PE.area		PE.count		PE.area		PE.count		PE.area		PE.count		PE.area		PE.count	
	AIC	R^2	AIC	R^2	AIC	R^2	AIC	R^2	AIC	R^2	AIC	R^2	AIC	R^2	AIC	R^2
Including SLMP	1325.7	0.652	618.4	0.783	1277.3	0.639	572.3	0.784	1197.4	0.658	524.5	0.786	1169.5	0.635	460.6	0.791
Including Geo_class	1337.2	0.647	636.3	0.778	1290.2	0.633	601.2	0.777	1226.7	0.645	551.2	0.779	1194.4	0.624	490.2	0.784

References S2.1 References of checklists and floras from the Global Inventory of Floras and Traits (GIFT) used to compile the regional species composition data.

3D Environmental. 2008. Vegetation Communities and Regional Ecosystems of The Torres Strait Islands, Queensland, Australia. Queensland, Australia.

Abe T. 2006. Threatened pollination systems in native flora of the Ogasawara (Bonin) Islands. *Annals of Botany (London)* 98: 317.

Acevedo-Rodríguez P, Strong MT. 2007. *Catalogue of the seed plants of the West Indies Website*. [WWW document] URL http://botany.si.edu/antilles/WestIndies/catalog.htm [accessed 1 March 2011].

Al Khulaidi AWA. 2000. Flora of Yemen. Yemen.

Alsos IG, Ehrich D, Eidesen PB, Solstad H, Westergaard KB, 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: plv036.

Arechavaleta M, Rodríguez S, Zurita N, García A. 2009. Lista de especies silvestres de Canarias. Hongos, plantas y animales terrestres. Santa Cruz de Tenerife, Spain: Consejería de Medio Ambiente y Ordenación Territorial, Gobierno de Canarias.

Arechavaleta M, Zurita N, Marrero MC, Martín JL. 2005. Lista preliminar de especies silvestres de Cabo Verde (hongos, plantas y animales terrestres). Santa Cruz de Tenerife, Spain: Consejería de Medio Ambiente y Ordenación Territorial, Gobierno de Canarias.

Ashmole P, Ashmole M. 2000. St Helena and Ascension Island. A natural history. Oswestry, Shropshire, UK: Anthony Nelson Ltd.

Backer CA, Bakhuizen van den Brink, RC. 1963. Flora of Java. (Spermatophytes only). Groningen: Noordhoff.

Baker ML, Duretto MF. 2011. A census of the vascular plants of Tasmania. Hobart, Australia: Tasmanian Herbarium, Tasmanian Museum and Art Gallery.

Barker WR, Barker RM, Jessop JP, Vonow HP. 2005. Census of South Australian vascular plants. *Journal of the Adelaide Botanic Gardens Supplement* 1: 1–396.

Bernal R, Gradstein SR, Celis M. 2015. *Catálogo de plantas y líquenes de Colombia*. [WWW document] URL http://catalogoplantasdecolombia.unal.edu.co/ [accessed 15 January 2016].

BGCI. 2021. *GlobalTreeSearch online database (version 1.4)*. [WWW document] URL https://tools.bgci.org/global_tree_search.php [accessed 19 February 2021].

Bingham MG, Willemen A, Wursten BT, Ballings P, Hyde MA. 2016. *Flora of Zambia*. [WWW document] URL http://www.zambiaflora.com/ [accessed 14 November 2016].

BioScripts. 2014. *Flora Vascular*. [WWW document] URL http://www.floravascular.com/ [accessed 25 May 2014].

Borges PAV, Abreu C, Aguiar AMF, Carvalho P, Jardim R, Melo I, Oliveira P, Sérgio C, Serrano ARM, Vieira P. 2008. Listagem dos fungos, flora e fauna terrestres dos arquipélagos da Madeira e Selvagens. Funchal and Angra do Heroísmo, Portugal: Direcção Regional do Ambiente da Madeira and Universidade dos Açores.

Borges PAV, Costa A, Cunha R, Gabriel R, Gonçalves V, Martins AF, Melo I, Parente M, Raposeiro P, Rodrigues P *et al.* 2010. A list of the terrestrial and marine biota from the Azores. Cascais: Princípia.

Boullet V, Hivert J, Gigord LDB. 2018. An Updated Account of the Vascular Flora of the Iles Eparses (Southwest Indian Ocean). *Atoll Research Bulletin* 614: 1–64.

Breckle S-W, Hedge IC, Rafiqpoor MD. 2013. Vascular plants of Afghanistan. An augmented checklist. Bonn: Scientia Bonnensis.

Broughton DA, McAdam JH. 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.

Brown EA. 1979. Vegetation and flora of Ponui Island, Hauraki Gulf, New Zealand. *TANE* 25: 5–16.

Brundu G, Camarda I. 2013. The Flora of Chad: a checklist and brief analysis. *Phytokeys* 23: 1–17.

Buchwald E, Wind P, Bruun HH, Møller PF, Ejrnæs R, Svart HE. 2013. Hvilke planter er hjemmehørende i Danmark? *Jydsk Naturhistorisk Forening* 118: 72–96.

Bundesamt für Naturschutz. 2016. *Floraweb*. [WWW document] URL www.floraweb.de [accessed 15 September 2016].

Burton FJ. 2007. Red list assessment of Cayman Islands' native flora for legislation and conservation planning.

Byrd GV. 1984. Vascular vegetation of Buldir Island, Aleutian Islands, Alaska, compared to another Aleutian Island. *Arctic* 37: 37–48.

Cámara-Leret R, Frodin DG, Adema F, Anderson C, Appelhans MS, Argent G, Arias Guerrero S, Ashton P, Baker WJ, Barfod AS *et al.* 2020. New Guinea has the world's richest island flora. *Nature* 584: 579–583.

Cameron EK, Lange PJ de, McCallum J, Taylor GA, Bellingham PJ. 2007. Vascular flora and some fauna for a chain of six Vascular flora and some fauna for a chain of six Hauraki Gulf islands east and southeast of Waiheke Island. *Auckland Botanical Society Journal* 62: 136–156.

CARMABI. 2009. *Dutch Caribbean Biodiversity Explorer*. [WWW document] URL http://www.dcbiodata.net/explorer/home [accessed 24 June 2011].

Carrington CMS, Edwards RD, Krupnick GA. 2018. Assessment of the Distribution of Seed Plants Endemic to the Lesser Antilles in Terms of Habitat, Elevation, and Conservation Status. *Caribbean Naturalist* Special Issue No. 2: 30–47.

Catarino L, Martins ES, Basto MF, Diniz MA. 2008. An annotated checklist of the vascular flora of Guinea-Bissau (West Africa). *Blumea-Biodiversity, Evolution and Biogeography of Plants* 53: 1–222.

Chang C-S, Kim H, Chang K. Provisional Checklist of the Vascular Plants for the Korea Peninsular Flora (KPF). Version 1.0. Korea.

Charters M. 2007. *Flora of Bermuda*. [WWW document] URL http://www.calflora.net/floraofbermuda/ [accessed 30 July 2011].

Cheffings CM, Farrell L, eds. 2005. The vascular plant red data list for Great Britain. Peterborough, UK: Joint Nature Conservation Committee.

Chernyaeva AM. 1973. Flora of Onekotan Island. *Bulletin of Main Botanical Garden* 87: 21–29.

Chinese Virtual Herbarium. 2016. *The Flora of China v. 5.0*. [WWW document] URL http://www.cvh.org.cn/ [accessed 15 January 2016].

Chong KY, Tan TWH, Corlett RT. 2009. A checklist of the total vascular plant flora of Singapore. Native, Naturalised and Cultivated Species. Singapore: Raffles Museum of Biodiversity Research.

Christmas Island National Park. 2002. Third Christmas Island national park management plan. Christmas Island, Australia: Parks Australia North.

CONABIO. 2016. *Sistema Nacional de Información sobre Biodiversidad*. [WWW document] URL https://www.gob.mx/conabio [accessed 11 April 2016].

Conti F, Abbate G, Alessandrini A, Blasi C. 2005. Annotated Checklist of the Italian Vascular Flora. Rome, Italy: Palombi Editori.

Costion C, Lorence D. 2012. The endemic plants of Micronesia. A geographical checklist and commentary. *Micronesica* 43: 51–100.

Cronk QCB. 1989. The past and present vegetation of St Helena. *Journal of Biogeography* 16: 47–64.

Danihelka J, Chrtek J, Kaplan Z. 2012. Checklist of vascular plants of the Czech Republic. *Preslia* 84: 647–811.

D'Arcy WG. 1971. The island of Anegada and its flora. *Atoll Research Bulletin* 139: 1–21.

Dauby G, Zaiss R, Blach-Overgaard A, Catarino L, Damen T, Deblauwe V, Dessein S, Dransfield J, Droissart V, Duarte MC *et al.* 2016. RAINBIO. A mega-database of tropical African vascular plants distributions. *PhytoKeys* 74: 1–18.

Dimopoulos P, Raus T, Bergmeier E, Constantinidis T, Iatrou G, Kokkini S, Strid A, Tzanoudakis D. 2013. Vascular plants of Greece. An annotated checklist. Berlin, Athens: Botanischer Garten und Botanisches Museum Berlin-Dahlem, Freie Universität Berlin; Hellenic Botanical Society.

Du Puy DJ. 1993. *Christmas Island. species lists*. [WWW document] URL http://www.anbg.gov.au/abrs/online-resources/flora/ [accessed 6 April 2011].

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.

Egorova EM. 1964. Flora of Shiashkotan Island. *Bulletin of the Main Botanical Garden* 54: 114–120.

Elizabeth Joyce, Kevin Thiele, Ferry Slik, Darren Crayn, Joyce E, Thiele K, Slik F, Crayn D. 2020. Checklist of the vascular flora of the Sunda-Sahul Convergence Zone. *Pensoft Publishers* 8: e51094.

Euro+Med. 2006. *Euro+Med PlantBase - the information resource for Euro-Mediterranean plant diversity*. [WWW document] URL http://ww2.bgbm.org/EuroPlusMed/ [accessed 1 June 2019].

Exell AW. 1944. Catalogue of the vascular plants of S. Tome (with Principe and Annobon). London, UK: Trustees of the British Museum.

Fall PL, Drezner TD. 2020. Vascular plant species of the Kingdom of Tonga by vegetation type, species origin, growth form, and dispersal mechanism. *Ecology* 101: e02902.

Figueiredo E, Paiva J, Stevart T, Oliveira F, Smith GF. 2011. Annotated catalogue of the flowering plants of São Tomé and Príncipe. *Bothalia* 41: 41–82.

Fischer MA, 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. Linz, Austria: Oberöstereichisches Landesmuseum.

Florence J, Chevillotte H, Ollier C, Meyer J-Y. 2007. *Base de données botaniques Nadeaud de l'Herbier de la Polynésie française (PAP)*. [WWW document] URL http://www.herbier-tahiti.pf/ [accessed 1 July 2011].

Fosberg FR, Renvoize SA, Townsend CC. 1980. The flora of Aldabra and neighbouring islands. London, UK: HMSO.

Fosberg FR, Sachet MH. 1987. Flora of Maupiti, Society Islands. *Atoll Research Bulletin* 294: 1–70.

Francisco-Ortega J, Wang F-G, Wang Z-S, Xing F-W, Liu H, Xu H, Xu W-X, Luo Y-B, Song X-Q, Gale S *et al.* 2010. Endemic Seed Plant Species from Hainan Island: A Checklist. *The Botanical Review* 76: 295–345.

Freid E, Francisco-Ortega J, Jestrow B. 2014. Endemic Seed Plants in the Bahamian Archipelago. *The Botanical Review*: 1–27.

Funk VA, Hollowell T, Berry P, Kelloff C, Alexander SN. 2007. Checklist of the plants of the Guiana Shield (Venezuela: Amazonas, Bolivar, Delta Amacuro; Guyana, Surinam, French Guiana). Washington, DC: Department of Botany, National Museum of Natural History.

Gage S, Joneson SL, Barkalov VY, Eremenko NA, 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.

Garroutte MD. 2016. Species Richness, Community Compositon and Species Distribution Patterns in Aleutian Plants. Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science. University of Alaska Fairbanks.

Gerlach J. 2003. The biodiversity of the granitic islands of Seychelles. *Phelsuma* 11 (Supplement A): 1–47.

Govaerts R, Nic Lughadha E, Black N, Turner R, Paton A. 2021. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Scientific Data* 8: 1.

Graveson R. 2019. *Plants of Saint Lucia. A Pictorial Flora of Wild and Cultivated Vascular Plants.* [WWW document] URL http://www.saintlucianplants.com/index.html [accessed 5 October 2020].

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.

Green PS. 1994a. *Lord Howe Island. species lists.* [WWW document] URL http://www.environment.gov.au/biodiversity/abrs/online-resources/flora/49/index.html [accessed 6 April 2011].

Green PS. 1994b. *Norfolk Island. Species lists.* [WWW document] URL http://www.environment.gov.au/biodiversity/abrs/online-resources/flora/49/index.html [accessed 6 April 2011].

Greene SW, Walton DWH. 1975. An annotated check list of the sub-Antarctic and Antarctic vascular flora. *Polar Record* 17: 473–484.

Hnatiuk RJ. 1993. *Subantarctic Islands. species lists.* [WWW document] URL http://www.environment.gov.au/biodiversity/abrs/online-resources/flora/50/index.html [accessed 7 April 2011].

Huang T-C. 1994. Flora of Taiwan, vol. I–VI. Taipei, Taiwan: National Taiwan University.

Hutchinson J, Dalziel JM, Keay RWJ, Hepper N. 2014. Flora of West Tropical Africa. London, UK: Crown agents for overseas governments.

Hyde MA, Wursten BT, Ballings P, Coates Palgrave M. 2016a. *Flora of Botswana.* [WWW document] URL http://www.botswanaflora.com/ [accessed 14 November 2016].

Hyde MA, Wursten BT, Ballings P, Coates Palgrave M. 2016b. *Flora of Malawi.* [WWW document] URL http://www.malawiflora.com/index.php [accessed 14 November 2016].

Hyde MA, Wursten BT, Ballings P, Coates Palgrave M. 2016c. *Flora of Mozambique.* [WWW document] URL http://www.mozambiqueflora.com/ [accessed 14 November 2016].

Hyde MA, Wursten BT, Ballings P, Coates Palgrave M. 2016d. *Flora of Zimbabwe* [accessed 14 November 2016].

Imada CT, ed. 2012. Hawaiian Native and Naturalized Vascular Plants Checklist. Honolulu, Hawai'i: Hawaii Biological Survey, Bishop Museum.

INBIO. 2000. *Lista de planta de Costa Rica. With updates by Eduardo Chacón.* [WWW document] URL http://www.inbio.ac.cr/papers/manual_plantas/index.html [accessed 5 February 2016].

Jackes BR. 2010. Plants of Magnetic Island. Townsville, Australia: James Cook University.

Jahn R, Schönfelder P. 1995. Exkursionsflora für Kreta. Stuttgart, Germany: Ulmer (Eugen).

Jaramillo Díaz P, Guézou A, Mauchamp A, Tye A. 2018. CDF Checklist of Galapagos Flowering Plants. FCD Lista de especies de Plantas con flores Galápagos. Puerto Ayora, Galapagos: Charles.

Jardim Botânico do Rio de Janeiro. 2016. *Flora do Brasil 2020 em construção.* [WWW document] URL http://floradobrasil.jbrj.gov.br/ [accessed 9 May 2016].

Johnson PN, Campbell DJ. 1975. Vascular plants of the Auckland Islands. *New Zealand journal of botany* 13: 665–720.

Jongbloed M, Feulner GR, Böer B, Western AR. 2003. The comprehensive guide to the wild flowers of the United Arab Emirates. Abu Dhabi: Environmental Research and Wildlife Development Agency.

Karlsson T, Agestam M. 2014. *Checklist of Nordic vascular plants. Sweden Checklist.* [WWW document] URL http://www.euphrasia.nu/checklista/index.eng.html [accessed 11 February 2016].

Kenneally KF. 1993. *Ashmore Reef and Cartier Island. Species lists.* [WWW document] URL http://www.environment.gov.au/biodiversity/abrs/online-resources/flora/50/index.html [accessed 6 April 2011].

Keppel G, Gillespie TW, Ormerod P, Fricker GA. 2016. Habitat diversity predicts orchid diversity in the tropical south-west Pacific. *Journal of Biogeography* 43: 2332–2342.

Kerguelen M. 2005. *Base de Données Nomenclaturales de la Flore de France.* [WWW document] URL http://www.tela-botanica.org/page:284# [accessed 13 January 2012].

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.

Kirchner F, Picot F, Merceron E, Gigot G. 2010. Flore vasculaire de La Réunion. Réunion, France: Conservatoire Botanique National de Mascarin.

Komarov VL, Shishkin BK, Bobrov EG. 1933-1964. Flora of the U.S.S.R. Translated and published 1968. Springfield, US: Israel Program for Scientific Translations.

Kristinsson H. 2008. Checklist of the vascular plants of Iceland. Reykjavík, Iceland: Náttúrufræðistofnun Íslands.

Kuzmenkova SM. 2015. *Plants of Belarus.* [WWW document] URL http://hbc.bas-net.by/plantae/eng/default.php [accessed 15 February 2016].

Landcare Research New Zealand. *Ecological Traits of New Zealand Flora.* [WWW document] URL https://ecotraits.landcareresearch.co.nz/ [accessed Nov/2018].

Lange PJ de, Heenan PB, Rolfe JR. 2011. Checklist of vascular plants recorded from Chatham Islands.

Lazkov GA, Sultanova BA. 2011. Checklist of vascular plants of Kyrgyzstan. Helsinki: Botanical Museum, Finnish Museum of Natural History.

Lester-Garland LV. 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. London, UK: West, Newman & Co.

Lord JM. 2015. Patterns in floral traits and plant breeding systems on Southern Ocean Islands. *AoB Plants* 7: plv095.

Marquand ED. 1901. Flora of Guernsey and the lesser Channel Islands. namely Alderney, Sark, Herm, Jethou, and the adjacent islets. London, UK: Dulau & Co.

Marticorena C, Squeo FA, Arancio G, Muñoz M. 2008. Catálogo de la flora vascular de la IV Región de Coquimbo. In: Squeo FA, Arancio G, Gutiérrez JR, eds. *Libro rojo de la flora nativa y de los sitios prioritarios para su conservación: Región de Atacama*. Ediciones Universidad de La Serena La Serena, 105–142.

Marticorena C, Stuessy TF, Baeza CM. 1998. Catalogue of the vascular flora of the Robinson Crusoe or Juan Fernández islands, Chile. *Gayana Botánica* 55: 187–211.

McClatchey W, Thaman R, Vodonaivalu S. 2000. A preliminary checklist of the flora of Rotuma with Rotuman names. *Pacific Science* 54: 345–363.

Meades SJ, Hay SG, Brouillet L. 2005. Annotated checklist of the vascular plants of Newfoundland and Labrador. Sault Ste. Marie, Ontario: The Provincial Museum of Newfoundland and Labrador, St. John's, NF, Canada.

Mifsud S. 2020. *MaltaWildPlants.com - an online flora of Malta*. [WWW document] URL http://www.maltawildplants.com/ [accessed 2 October 2020].

Miller AG, Morris M. 2004. Ethnoflora of the Soqotra Archipelago. Edinburgh, UK: Royal Botanic Garden.

Ministry of Environment. 2012. The national red list 2012 of Sri Lanka. Conservation status of the fauna and flora. Colombo, Sri Lanka.

Morat P, Jaffré T, Tronchet F, Munziger J, Pillon Y, Veillon J-M, Chalopin M, Birnbaum P, Rigault F, Dagostini G *et al.* 2012. The taxonomic reference base "Florical" and characteristics of the native vascular flora of New Caledonia. *Adansonia* 34: 179–221.

Muer T, Sauerbier H, Cabrera Calixto F. 2016. Die Farn- und Blütenpflanzen der Kanarischen Inseln. Über 2.000 Pflanzenarten, mehr als 2.600 Fotos. Weikersheim: Margraf Publishers.

Nakamura K, Suwa R, Denda T, Yokota M. 2009. Geohistorical and current environmental influences on floristic differentiation in the Ryukyu Archipelago, Japan. *Journal of Biogeography* 36: 919–928.

Newman M. 2007. A checklist of the vascular plants of Lao PDR. Edinburgh, Scotland, UK: Royal Botanic Garden Edinburgh.

Nikolić T. 2016. *Flora Croatica Database*. [WWW document] URL http://hirc.botanic.hr/fcd [accessed 12 February 2016].

Norton J, Majid SA, Allan D, Al Safran M, Böer B, Richer RA. 2009. An illustrated checklist of the flora of Qatar. Gosport, UK: Browndown Publications Gosport.

Pandey RP, Diwakar PG. 2008. An integrated checklist flora of Andaman and Nicobar Islands, India. *J. Econ. Taxon. Bot* 32: 403–500.

Penneckamp Furniel D. 2018. Flora vascular silvestre del Archipiélago Juan Fernández. Valparaíso, Chile: Planeta de Papel Ediciones.

Press JR, Short MJ, Turland NJ. 2016. Flora of Madeira. Exeter: Pelagic Publishing.

Price JP, Wagner WL. 2011. A phylogenetic basis for species–area relationships among three Pacific Island floras. *American Journal of Botany* 98: 449–459.

Queensland Government. 2014. *Census of the Queensland flora 2014*. [WWW document] URL https://data.qld.gov.au/dataset/census-of-the-queensland-flora-2014 [accessed 5 February 2015].

Raulerson L. 2006. Checklist of plants of the Mariana Islands. *University of Guam Herbarium Contribution* 37: 1–69.

Releford JS, Stevens J, Bridges KW, McClatchey WC. 2009. Flora of Rongelap and Ailinginae Atolls, Republic of the Marshall Islands. *Atoll Research Bulletin* 572: 1–13.

Renvoize SA. 1975. A floristic analysis of the western Indian Ocean coral islands. *Kew Bulletin* 30: 133–152.

Rhind M. 2020. *Terrestrial Biozones*. *Endemic Floras*. [WWW document] URL http://www.terrestrial-biozones.net/Endemic%20Floras/Endemic%20Floras.html [accessed 2 February 2021].

Robinson AC, Canty PD, Fotheringham D. 2008. Investigator group expedition 2006. flora and vegetation. *Transactions of the Royal Society of South Australia* 132: 173–220.

Rodriguez R, Marticorena C, Alarcón D, Baeza C, Cavieres L, Finot VL, Fuentes N, Kiessling A, Mihoc M, Pauchard A *et al.* 2018. Catálogo de las plantas vasculares de Chile. *Gayana. Botánica* 75: 1–430.

Royal Botanic Gardens and Domain Trust. 2017. *PlantNET - The NSW Plant Information Network System*. [WWW document] URL http://plantnet.rbgsyd.nsw.gov.au [accessed 16 September 2016].

Sachet M-H. 1962. Flora and vegetation of Clipperton Island. *Proceedings of the California Academy of Sciences* 31: 249–307.

Sáez L, Rosselló JA. 2001. Llibre vermell de la flora vascular de les Illes Balears. Palma de Mallorca, Spain: Direcció General de Biodiversitat, Conselleria de Medi Ambient, Govern de les Illes Balears.

SANBI. 2014. *Plants of Southern Africa. An online checklist*. [WWW document] URL http://posa.sanbi.org [accessed 9 March 2015].

Sandbakk BE, Alsos IG, Arnesen G, Elven R. 1996. *The flora of Svalbard*. [WWW document] URL http://svalbardflora.no/ [accessed 16 March 2011].

Searle J, Madden S. 2006. Flora assessment of South Stradbroke Island. Gold Coast City, Australia: Gold Coast City Council.

Shaw JD, Spear D, Greve M, Chown SL. 2010. Taxonomic homogenization and differentiation across Southern Ocean Islands differ among insects and vascular plants. *Journal of Biogeography* 37: 217–228.

Short PS, Albrecht DE, Cowie ID, Lewis DL, Stuckey BM. 2011. Checklist of the vascular plants of the Northern Territory. Darwin: Department of Natural Resources, Environment, The Arts and Sport.

SLUFG. 2015. Rote Liste und Artenliste Sachsen. Farn- und Samenpflanzen. Dresden, Germany: Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie.

Smith AC. 1979-1996. Flora Vitiensis nova. a new Flora of Fiji (spermatophytes only). Lawaii, Hawaii: Pacific Tropical Botanical Garden.

St John H. 1960. Flora of Eniwetok Atoll. *Pacific Science* 14: 313–336.

Stace CA, Ellis RG, Kent DH, McCosh DJ. 2003. Vice-county Census Catalogue of the vascular plants of Great Britain, the Isle of Man and the Channel Islands. London, UK: Botanical Society of the British Isles.

Stephens K. 2011. Comparative floristic analysis of vegetation on the Dune Islands of South-East Queensland. *The Proceedings of the Royal Society of Queensland* 117: 141–180.

Stoddart DR, Fosberg FR. 1994. Flora of the Phoenix Islands, central Pacific. *Atoll Research Bulletin* 393: 1–60.

Strahm WA. *The conservation and restoration of the flora of Mauritius and Rodrigues,* University of Reading, Reading, UK.

Sykes WR. 1970. Contributions to the flora of Niue. *Bulletin. Department of Scientific and Industrial Research, New Zealand* 200: 321.

Sykes WR. 2016. Flora of the Cook Islands. Kalaheo (Kaua`i, USA): National Tropical Botanical Garden.

Sykes WR, West CJ, Beever JE, Fife AJ. 2000. Kermadec Islands flora-special edition. a compilation of modern material about the flora of the Kermadec Islands. Lincoln, New Zealand: Manaaki Whenua Press, Landcare Research.

Takahashi H, Barkalov VY, Gage S, Semsrott B, Ilushko M, Zhuravlev YN. 2006. A floristic study of the vascular plants of Kharimkotan, Kuril Islands. *Bulletin of the Hokkaido University Museum* 3: 41–66.

Takahashi H, Barkalov VY, Gage S, Zhuravlev YN. 1997. A preliminary study of the flora of Chirpoi, Kuril Islands. *Acta Phytotaxonomica et Geobotanica* 48: 31–42.

Tamis WLM, van der Meijden R, Runhaar J, Bekker RM, Ozinga WA, Odé B, Hoste I. 2004. Standard List of the Flora of the Netherlands 2003. *Gorteria* 30: 101–195.

Tatewaki M. 1957. Geobotanical studies on the Kurile Islands. *Acta Horti Gotoburgensis* 21: 43–123.

Taylor R. 2006. Straight through from London. the Antipodes and Bounty Islands, New Zealand. Christchurch, New Zealand: Heritage Expeditions New Zealand.

Tela Botanica. 2015. *Base de données des Trachéophytes de France métropolitaine (bdtfx)*. [WWW document] URL http://www.tela-botanica.org/bdtfx.

Thaman RR, Fosberg FR, Manner HI, Hassall DC. 1994. The flora of Nauru. *Atoll Research Bulletin* 392: 1–233.

Thaman RR, Tye A. 2015. Flora of Kiritimati (Christmas) Atoll, Northern Line Islands, Republic of Kiribati. *Atoll Research Bulletin* 608: 1–73.

Thomas J. 2011. *Plant diversity of Saudi Arabia. Flora checklist*. [WWW document] URL http://plantdiversityofsaudiarabia.info/biodiversity-saudi-arabia/flora/Checklist/Cheklist.htm [accessed 14 January 2016].

Tropicos. 2015a. *Catálogo de las Plantas Vasculares de Bolivia*. [WWW document] URL http://www.tropicos.org/Project/BC [accessed 10 September 2015].

Tropicos. 2015b. *Catalogue of the Vascular Plants of Ecuador*. [WWW document] URL http://www.tropicos.org/Project/CE [accessed 22 October 2015].

Tropicos. 2015c. *Flora de Nicaragua*. [WWW document] URL http://www.tropicos.org/Project/FN [accessed 16 July 2015].

Tropicos. 2015d. *Panama Checklist*. [WWW document] URL http://www.tropicos.org/Project/PAC [accessed 22 October 2015].

Tropicos. 2015e. *Peru Checklist*. [WWW document] URL http://www.tropicos.org/Project/PEC [accessed 22 October 2015].

Tropicos. 2020. *Catalogue of the Vascular Plants of Madagascar*. [WWW document] URL http://www.tropicos.org/Project/MADA [accessed 01/22/2020].

Turner IM. 1995. A catalogue of the vascular plants of Malaya. Singapore: Gardens' Bulletin.

UIB. 2007. *Herbario virtual del Mediterráneo Occidental*. [WWW document] URL http://herbarivirtual.uib.es/cas-med/ [accessed 7 August 2012].

Ulloa Ulloa C, Acevedo-Rodríguez P, Beck S, Belgrano MJ, Bernal R, Berry PE, Brako L, Celis M, Davidse G, Forzza RC *et al*. 2017. An integrated assessment of the vascular plant species of the Americas. *Science (New York, N.Y.)* 358: 1614–1617.

University of Kent. 2012. *Cook Islands Biodiversity and Ethnobiology Database*. [WWW document] URL http://cookislands.pacificbiodiversity.net/cibed/dbs/search.html [accessed 12 April 2012].

USDA, NRCS. 2015. *The PLANTS Database*. [WWW document] URL http://plants.usda.gov [accessed 24 April 2015].

Vander Velde N. 2003. The vascular plants of Majuro atoll, Republic of the Marshall Islands. *Atoll Research Bulletin* 503: 1–141.

Velayos M, Barberá P, Cabezas FJ, La Estrella Md, Fero M, Aedo C. 2014. Checklist of the Vascular Plants of Annobón (Equatorial Guinea). *Phytotaxa* 171: 1–78.

VicFlora. 2016. *Flora of Victoria*. [WWW document] URL http://vicflora.rbg.vic.gov.au [accessed 30 November 2016].

Virtual Biodiversity Museum of Cyprus. *Flora of Cyprus - Endemic plants of Cyprus*. [WWW document] URL http://www.naturemuseum.org.cy/lang1/endemic_plants.html [accessed 02.20.2020].

Vogt C. 2011. Composición de la Flora Vascular del Chaco Boreal, Paraguay. I. Pteridophyta y Monocotiledoneae. *Steviana* 3: 13–47.

Wace NM. 1961. The vegetation of Gough Island. *Ecological Monographs* 31: 337–367.

Wace NM, Dickson JH. 1965. The terrestrial botany of the Tristan da Cunha Islands. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 249: 273–360.

Wagner WL, Herbst DR, Lorence DH. 2005. *Flora of the Hawaiian Islands website*. [WWW document] URL http://botany.si.edu/pacificislandbiodiversity/hawaiianflora/ [accessed 16 October 2010].

Wagner WL, Lorence DH. 2002. *Flora of the Marquesas Islands website*. [WWW document] URL http://botany.si.edu/pacificislandbiodiversity/marquesasflora/index.htm [accessed 28 November 2014].

Wester L. 1985. Checklist of the vascular plants of the northern Line Islands. *Atoll Research Bulletin* 187: 1–38.

Western Australian Herbarium. 2017. *FloraBase - the Western Australian Flora*. [WWW document] URL https://florabase.dpaw.wa.gov.au/ [accessed 16 September 2016].

Yena AV. 2012. Prirodnaja flora krymskogo poluostrova. Spontaneous flora of the Crimean Peninsula. Simferopol': Orianda.

Yonekura K, Kajita T. 2013. *BG Plants Japanese name and Scientific name Index (YList)*. [WWW document] URL http://ylist.info [accessed 6 February 2016].

Yuncker TG. 1959. Plants of Tonga. *Bishop Museum Bulletin* 220: 1–283.

ZDSF, SKEW. 2014. *Info Flora. Artenliste Schweiz 5x5 km*. [WWW document] URL https://www.infoflora.ch/de/daten-beziehen/artenliste-5x5-km.html [accessed 13 February 2015].

Zizka G. 1991. Flowering plants of Easter Island. *Palmarum hortus francofurtensis* 3: 3–108.

Zuloaga FO, Morrone O, Belgrano M. 2014. *Catálogo de las Plantas Vasculares del Cono Sur*. [WWW document] URL http://www.darwin.edu.ar/Proyectos/FloraArgentina/fa.htm [accessed 16 March 2015].

References of Supporting Information for Chapter 2

Antonelli A, Kissling WD, Flantua SGA, Bermúdez MA, Mulch A, Muellner-Riehl AN, Kreft H, Linder HP, Badgley C, Fjeldså J, *et al.* 2018. Geological and climatic influences on mountain biodiversity. *Nature Geoscience* 11: 718–725.

Bennett KD, Tzedakis PC, Willis KJ. 1991. Quaternary refugia of north European trees. *Journal of Biogeography* 18: 103–115.

Dagallier LMJ, Janssens SB, Dauby G, Blach-Overgaard A, Mackinder BA, Droissart V, Svenning J, Sosef MSM, Stévart T, Harris DJ, *et al.* 2020. Cradles and museums of generic plant diversity across tropical Africa. *New Phytologist* 225: 2196–2213.

Enquist BJ, Feng X, Boyle B, Maitner B, Newman EA, Jørgensen PM, Roehrdanz PR, Thiers BM, Burger JR, Corlett RT, *et al.* 2019. The commonness of rarity: Global and future distribution of rarity across land plants. *Science Advances* 5: eaaz0414.

Gillespie RG, Roderick GK. 2002. Arthropods on islands: colonization, speciation, and conservation. *Annual Review of Entomology* 47: 595–632.

Govaerts R, Nic Lughadha E, Black N, Turner R, Paton A. 2021. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Scientific Data* 8: 215.

Jansson R. 2003. Global patterns in endemism explained by past climatic change. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270: 583–590.

Jetz W, Rahbek C, Colwell RK. 2004. The coincidence of rarity and richness and the potential signature of history in centres of endemism. *Ecology Letters* 7: 1180–1191.

Jump AS, Mátyás C, Peñuelas J. 2009. The altitude-for-latitude disparity in the range retractions of woody species. *Trends in Ecology & Evolution* 24: 694–701.

Kier G, Kreft H, Lee TM, Jetz W, Ibisch PL, 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* 106: 9322–9327.

McFadden IR, Sandel B, Tsirogiannis C, Morueta-Holme N, Svenning J, Enquist BJ, Kraft NJB. 2019. Temperature shapes opposing latitudinal gradients of plant taxonomic and phylogenetic β diversity (M Anderson, Ed.). *Ecology Letters* 22: 1126–1135.

Mittelbach GG, Schemske DW, Cornell HV, Allen AP, Brown JM, Bush MB, Harrison SP, Hurlbert AH, Knowlton N, Lessios HA, *et al.* 2007. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters* 10: 315–331.

Rahbek C, Borregaard MK, Antonelli A, Colwell RK, Holt BG, Nogues-Bravo D, Rasmussen CMØ, Richardson K, Rosing MT, Whittaker RJ, *et al.* 2019. Building mountain biodiversity: Geological and evolutionary processes. *Science* 365: 1114–1119.

Rohde K. 1992. Latitudinal gradients in species diversity: the search for the primary cause. *Oikos* 65: 514–527.

Sandel B, Arge L, Dalsgaard B, Davies RG, Gaston KJ, Sutherland WJ, Svenning J-C. 2011. The influence of Late Quaternary climate-change velocity on species endemism. *Science* 334: 660–664.

Sandel B, Weigelt P, Kreft H, Keppel G, van der Sande MT, Levin S, Smith S, Craven D, Knight TM. 2020. Current climate, isolation and history drive global patterns of tree phylogenetic endemism (R Kelly, Ed.). *Global Ecology and Biogeography* 29: 4–15.

Smith SA, Brown JW. 2018. Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany* 105: 302–314.

Stein A, Gerstner K, Kreft H. 2014. Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales (H Arita, Ed.). *Ecology Letters* 17: 866–880.

Stevens GC. 1989. The latitudinal gradient in geographical range: how so many species coexist in the tropics. *The American Naturalist* 133: 240–256.

Weigelt P, König C, Kreft H. 2020. GIFT – A global inventory of floras and traits for macroecology and biogeography. *Journal of Biogeography* 47: 16–43.

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

Supporting Information for Chapter 3
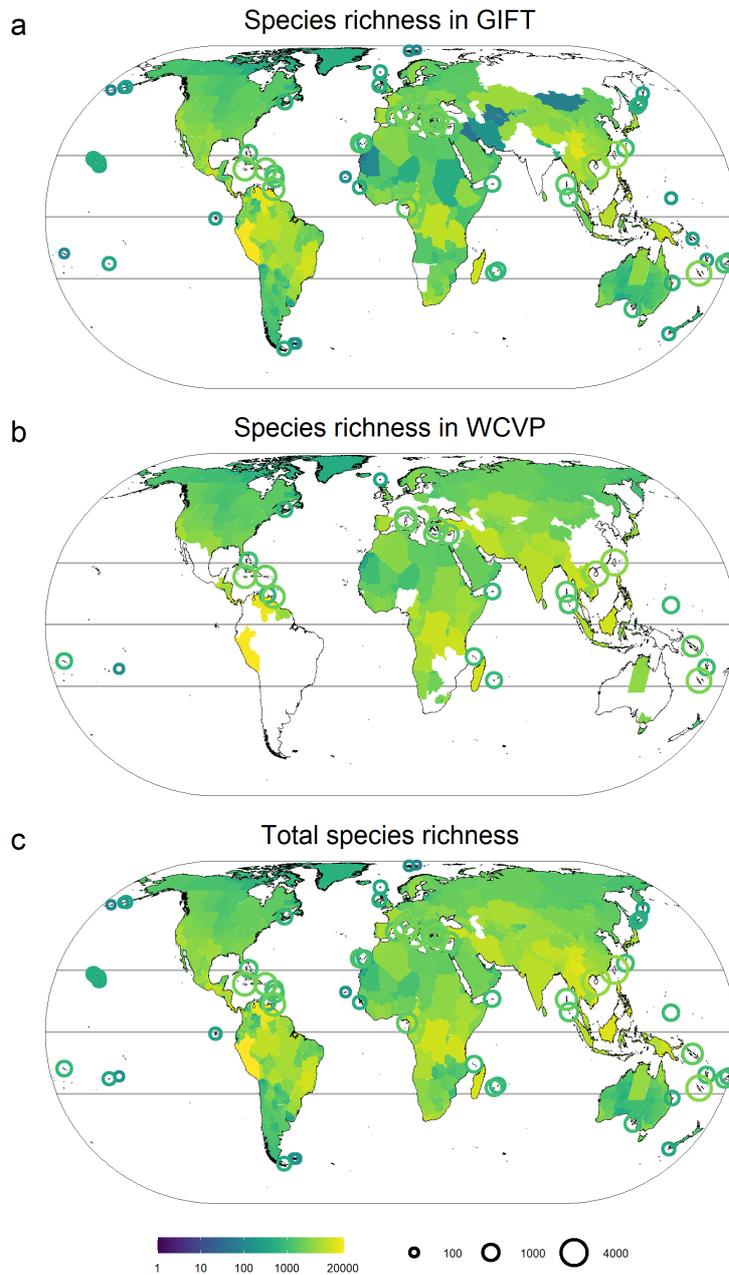
Figure S3.1 Observed species richness of seed plants for 675 geographic regions selected from (a) the Global Inventory of Floras and Traits (GIFT; 650 regions) and (b) the World Checklist of Vascular Plants (WCVP; 228 regions, combined in (c) to estimate phylogenetic and species turnover at the global scale. Log_{10} scale is used for species richness and maps are shown in Eckert IV projection.
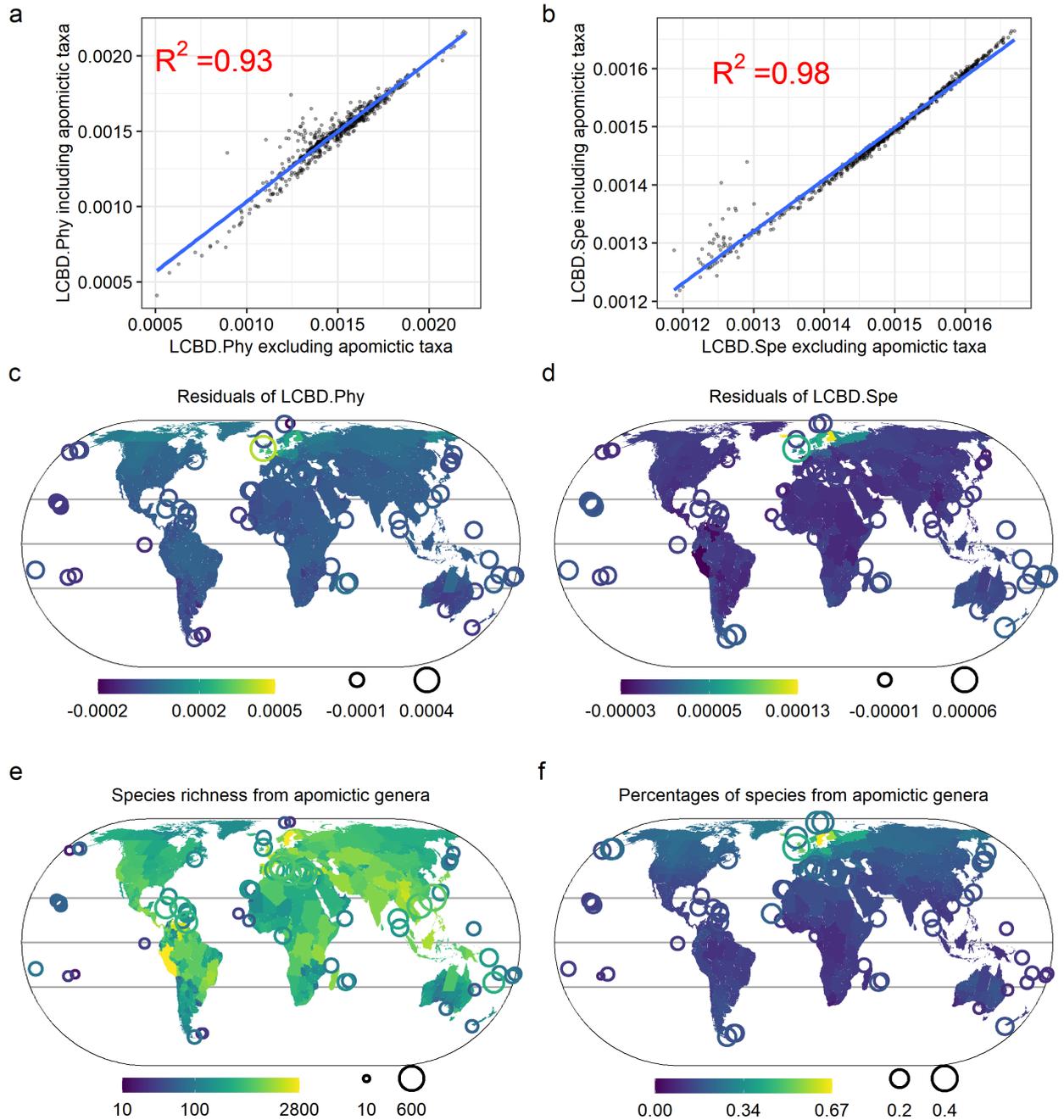
Figure S3.2 Comparison of local contribution to beta diversity (LCBD) for seed plants based on species distribution data including and excluding apomictic taxa. a and b, the linear regression between LCBD including and excluding apomictic taxa; c and d, residuals from the linear regression. LCBD is calculated using the dataset with missing species added to the phylogeny and for two dimensions of diversity: in a and c, phylogenetic turnover (LCBD.Phy); and in b and d, species turnover (LCBD.Spe). Positive residuals in c and d indicate higher values of LCBD based on the data including apomictic taxa than expected based on the data excluding apomictic taxa. e, species richness of seed plants from genera known to include apomictic species. f, percentages of species from genera known to include apomictic species. Log$_{10}$ scale is used in e and all maps are shown in Eckert IV projection.
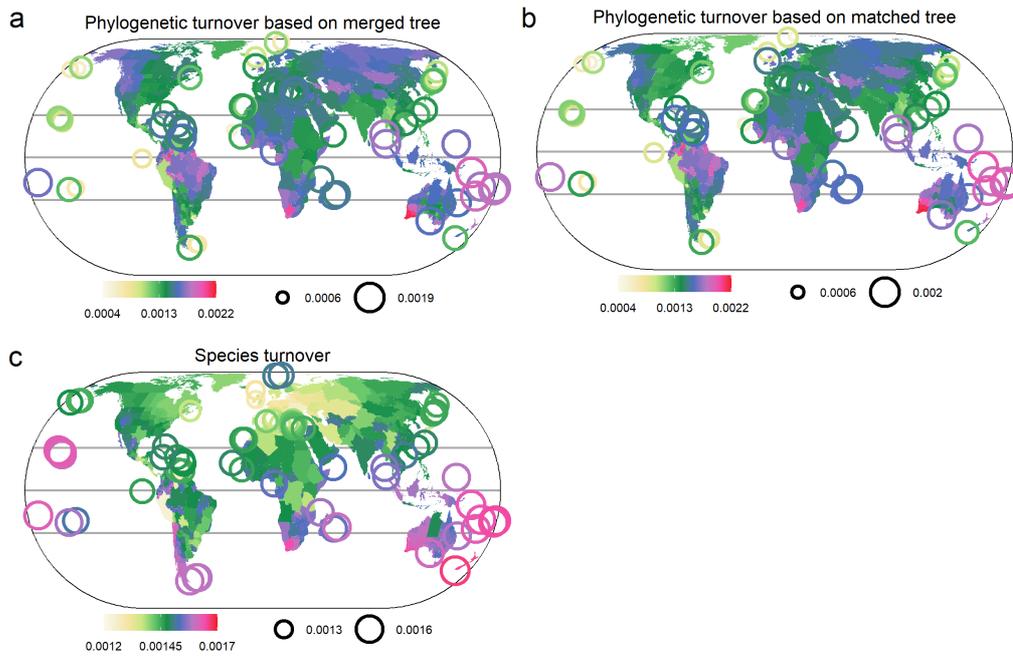
Figure S3.3 Local contribution to global patterns of beta diversity in seed plants based on the dataset including apomictic taxa. Local contribution to beta diversity is calculated respectively for phylogenetic turnover based on the dataset with (i.e. merged tree, a) and without (i.e. matched tree, b) missing species added to the phylogeny, and species turnover (c). Maps use Eckert IV projection.
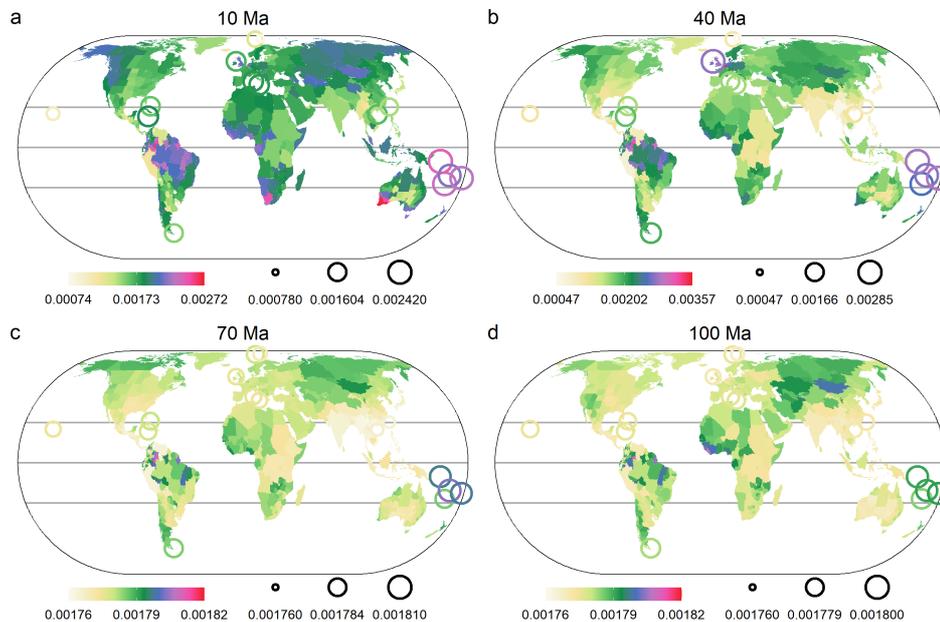
Figure S3.4 Local contribution of each region to global patterns of beta diversity (LCBD) for phylogenetic turnover at different phylogenetic timescales in seed plants based on the dataset including apomictic taxa and with missing species added to the phylogeny. LCBD patterns are calculated for 10 Ma BP (a), 40 Ma BP (b), 70 Ma BP (c), and 100 Ma BP (d), based on the given phylogenies which are obtained by cutting the original phylogeny at a specified time period and collapsing all descendent leaves into ancestral branches. Regions included are > 10,000 km2 and maps use Eckert IV projection.
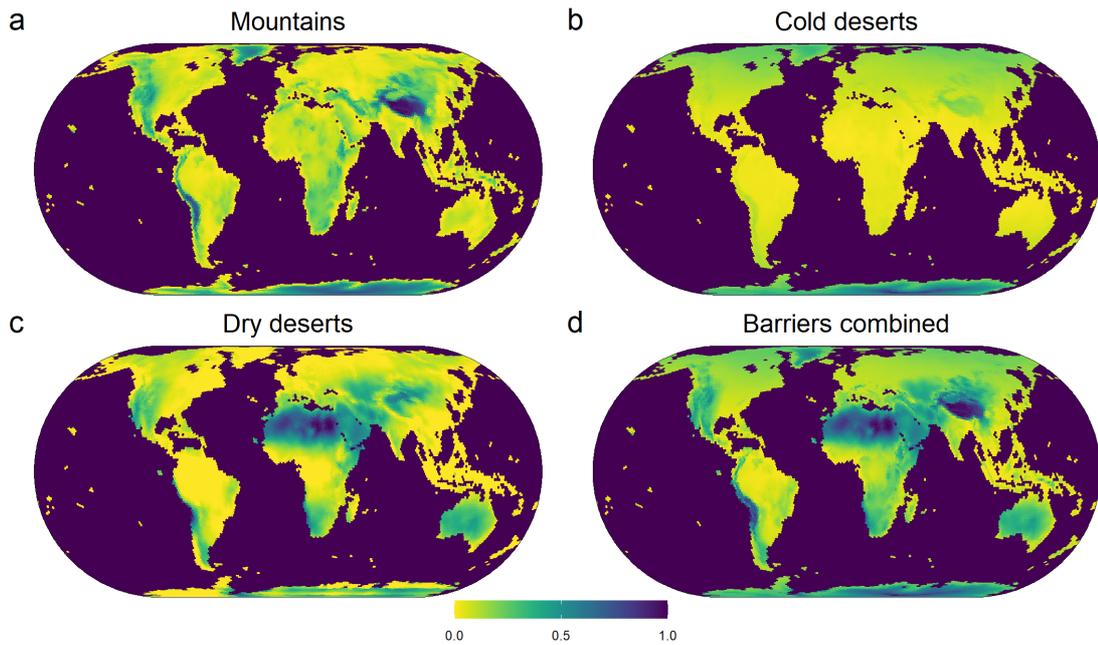
Figure S3.5 Barrier cost across an equal area grid of 23,323 km^2 hexagons. The barrier cost of a grid cell was respectively defined as mean elevation (a), annual mean temperature (b), aridity index (c), and the maximum values among whether a region is covered by water, mean elevation, annual mean temperature, and aridity index (d) to weigh in costs of crossing mountains, cold and dry deserts, and all barriers. High values indicated more cost needed for plants to move from one region to the other.
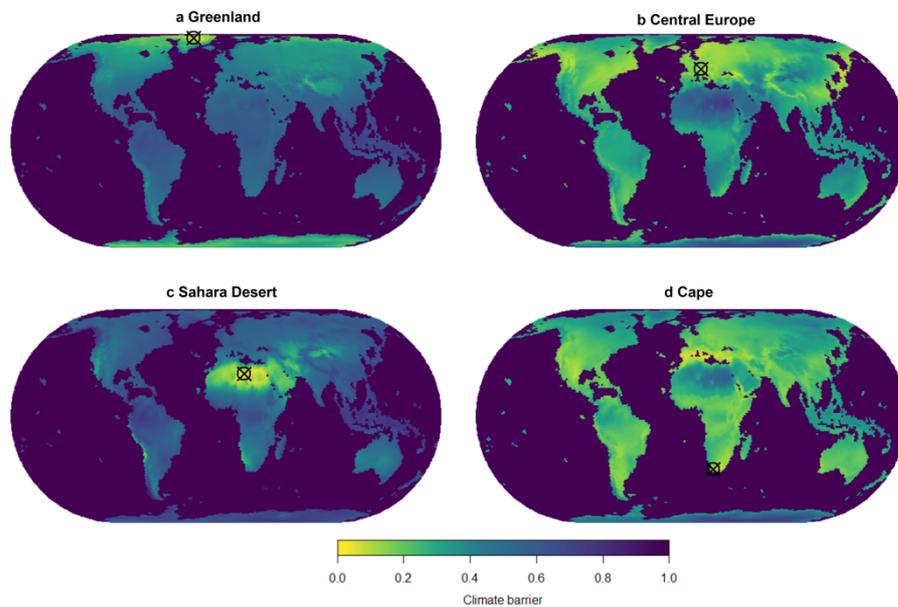
Figure S3.6 Climate cost across an equal area grid of 23,323 km^2 hexagons between a focal grid cell and all other grid cells. For each focal grid cell, the climate cost of a grid cell was defined as the Euclidean distance of the five climatic axes of a principal component analysis applied to 19 bioclimatic variables between the grid cell and the focal grid cell, divided by the maximum value so that the range was linearly scaled to [0,1]. High values indicated more cost needed for plants to move from one region to the other due to unsuitable climates. The focal grid cells of climate cost showed are respectively located in Greenland (a), central Europe (b), Sahara Desert (c) and Cape (d).
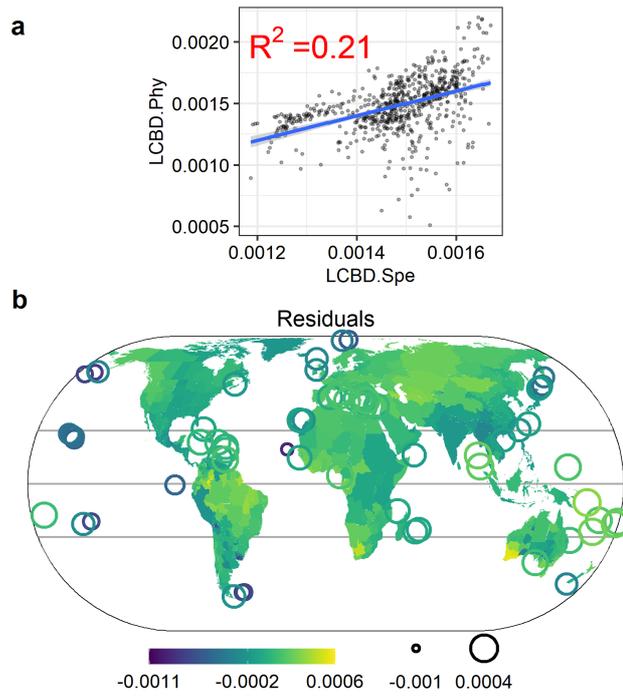
Figure S3.7 Comparison between local contribution to beta diversity for phylogenetic turnover (LCBD.Phy) and for species turnover (LCBD.Spe). a, the linear regression between LCBD.Phy and LCBD.Spe; b, residuals from the linear regression. LCBD is calculated using the dataset with missing species added to the phylogeny and removing apomictic taxa. Positive residuals in b indicate higher values of LCBD for phylogenetic turnover than expected based on species turnover.
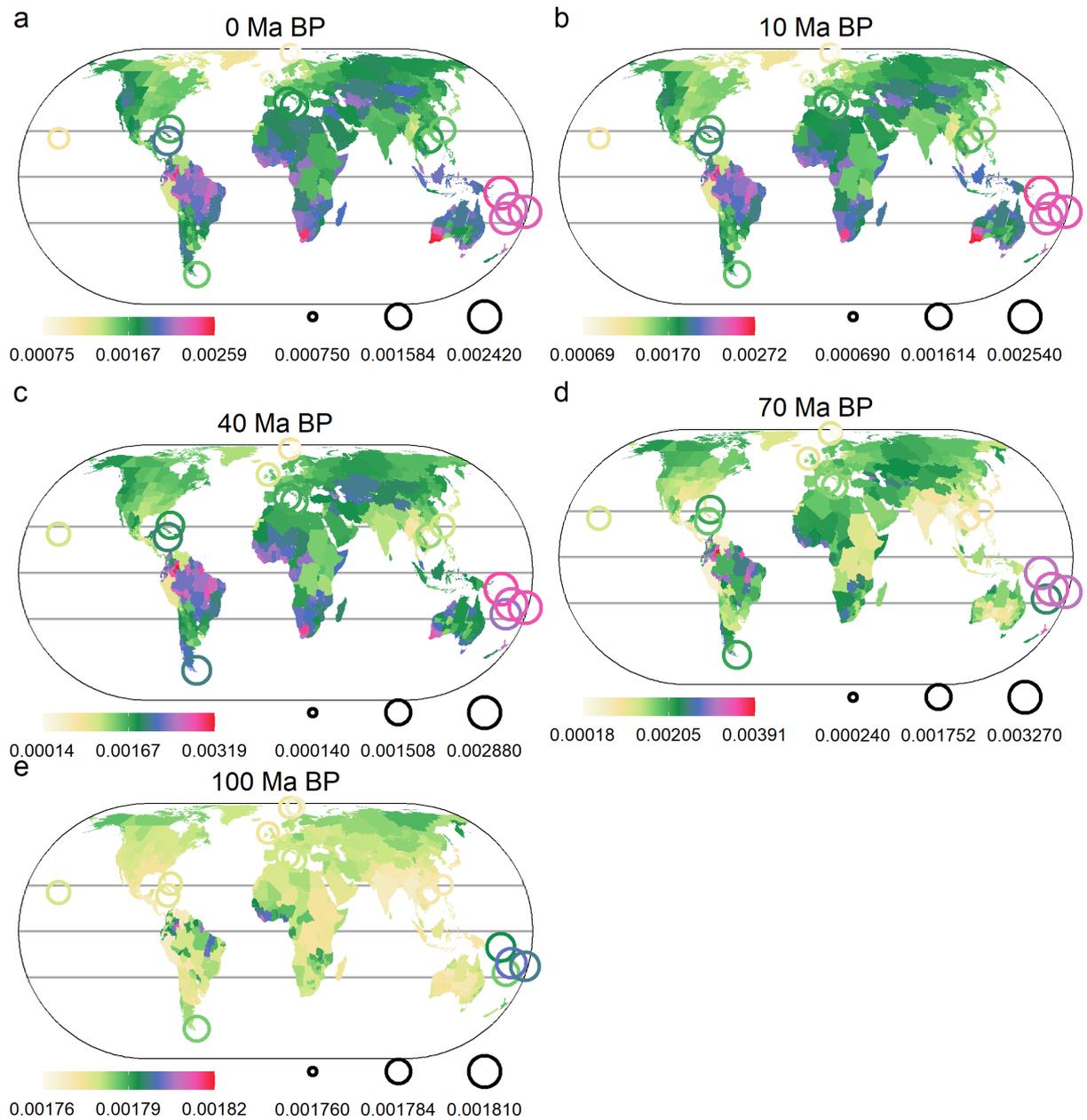
Figure S3.8 Local contribution to global patterns of beta diversity (LCBD) for phylogenetic turnover at different phylogenetic timescales in seed plants based on the dataset excluding apomictic taxa and without missing species added to the phylogeny. LCBD patterns are showed for: a, 0 Ma BP; b, 10 Ma BP; c, 40 Ma BP; d, 70 Ma BP and 100 Ma BP. Regions included are > 10,000 km^2 and maps use Eckert IV projection.
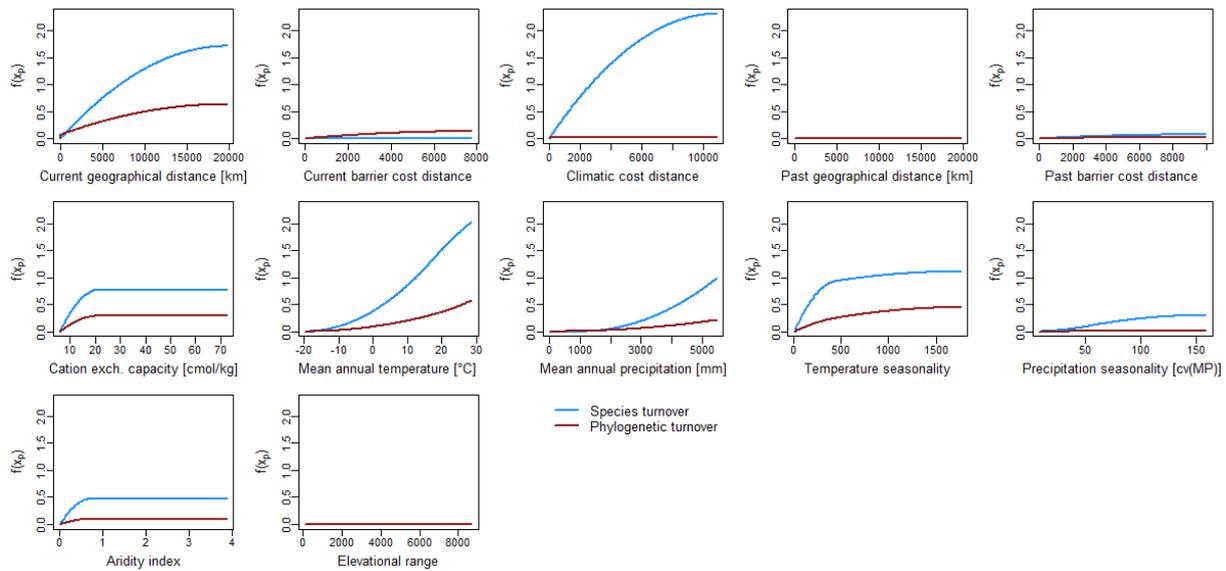
Figure S3.9 Generalized dissimilarity modelling spline functions for each predictor variable, respectively, for phylogenetic and species turnover in seed plants among regions across the world. The maximum height of the spline function indicates the importance of the predictor for explaining dissimilarities.
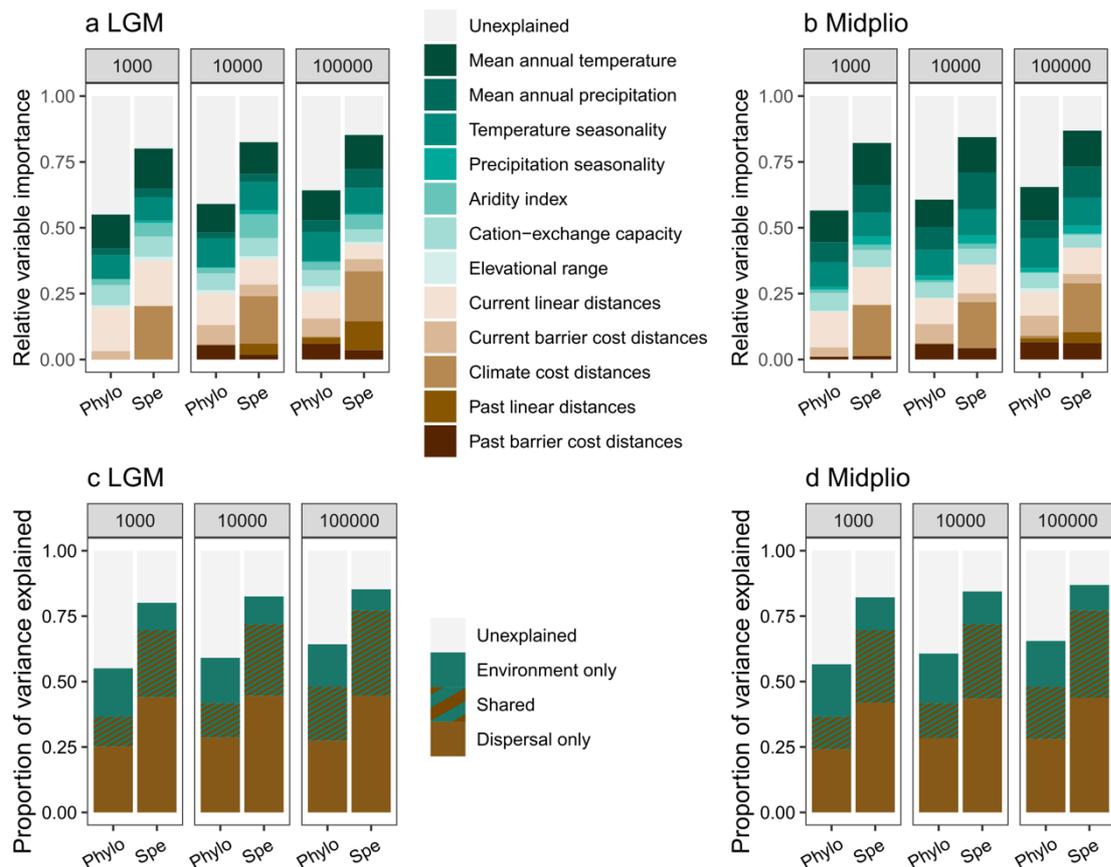
Figure S3.10 Relative importance of predictor variables for global phylogenetic and species turnover in seed plants based on generalized dissimilarity modelling including past climate instead of contemporary climate. Models includes: a and c, climate in the Last Glacial Maximum (LGM); b and d, climate in the mid-Pliocene warm period (midPlio). Two methods are used to quantify variable importance to plant phylogenetic (Phylo) and species turnover (Spe) based on different subsets of regions with a different minimum area size (i.e., 1000; 10,000; 100,000 km^2): a and b, the height of generalized dissimilarity modelling transformation functions; c and d, deviance partitioning.

References S3.1 References of checklists and floras from the Global Inventory of Floras and Traits (GIFT) used to compile the regional species composition data.

Acevedo-Rodríguez P, Strong MT. 2007. *Catalogue of the seed plants of the West Indies Website.* [WWW document] URL http://botany.si.edu/antilles/WestIndies/catalog.htm [accessed 1 March 2011].

Al Khulaidi AWA. 2000. Flora of Yemen. Yemen.

Alsos IG, Ehrich D, Eidesen PB, Solstad H, Westergaard KB, 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: plv036.

Arechavaleta M, Rodríguez S, Zurita N, García A. 2009. Lista de especies silvestres de Canarias. Hongos, plantas y animales terrestres. Santa Cruz de Tenerife, Spain: Consejería de Medio Ambiente y Ordenación Territorial, Gobierno de Canarias.

Arechavaleta M, Zurita N, Marrero MC, Martín JL. 2005. Lista preliminar de especies silvestres de Cabo Verde (hongos, plantas y animales terrestres). Santa Cruz de Tenerife, Spain: Consejería de Medio Ambiente y Ordenación Territorial, Gobierno de Canarias.

Backer CA, Bakhuizen van den Brink, RC. 1963. Flora of Java. (Spermatophytes only). Groningen: Noordhoff.

Baker ML, Duretto MF. 2011. A census of the vascular plants of Tasmania. Hobart, Australia: Tasmanian Herbarium, Tasmanian Museum and Art Gallery.

Barker WR, Barker RM, Jessop JP, Vonow HP. 2005. Census of South Australian vascular plants. *Journal of the Adelaide Botanic Gardens Supplement* 1: 1–396.

Bernal R, Gradstein SR, Celis M. 2015. *Catálogo de plantas y líquenes de Colombia.* [WWW document] URL http://catalogoplantasdecolombia.unal.edu.co/ [accessed 15 January 2016].

BGCI. 2021. *GlobalTreeSearch online database (version 1.4).* [WWW document] URL https://tools.bgci.org/global_tree_search.php [accessed 19 February 2021].

Bingham MG, Willemen A, Wursten BT, Ballings P, Hyde MA. 2016. *Flora of Zambia.* [WWW document] URL http://www.zambiaflora.com/ [accessed 14 November 2016].

BioScripts. 2014. *Flora Vascular.* [WWW document] URL http://www.floravascular.com/ [accessed 25 May 2014].

Breckle S-W, Hedge IC, Rafiqpoor MD. 2013. Vascular plants of Afghanistan. An augmented checklist. Bonn: Scientia Bonnensis.

Broughton DA, McAdam JH. 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.

Brundu G, Camarda I. 2013. The Flora of Chad: a checklist and brief analysis. *Phytokeys* 23: 1–17.

Buchwald E, Wind P, Bruun HH, Møller PF, Ejrnæs R, Svart HE. 2013. Hvilke planter er hjemmehørende i Danmark? *Jydsk Naturhistorisk Forening* 118: 72–96.

Bundesamt für Naturschutz. 2016. *Floraweb.* [WWW document] URL www.floraweb.de [accessed 15 September 2016].

Cámara-Leret R, Frodin DG, Adema F, Anderson C, Appelhans MS, Argent G, Arias Guerrero S, Ashton P, Baker WJ, Barfod AS *et al.* 2020. New Guinea has the world's richest island flora. *Nature* 584: 579–583.

Carrington CMS, Edwards RD, Krupnick GA. 2018. Assessment of the Distribution of Seed Plants Endemic to the Lesser Antilles in Terms of Habitat, Elevation, and Conservation Status. *Caribbean Naturalist* Special Issue No. 2: 30–47.

Catarino L, Martins ES, Basto MF, Diniz MA. 2008. An annotated checklist of the vascular flora of Guinea-Bissau (West Africa). *Blumea-Biodiversity, Evolution and Biogeography of Plants* 53: 1–222.

Chang C-S, Kim H, Chang K. Provisional Checklist of the Vascular Plants for the Korea Peninsular Flora (KPF). Version 1.0. Korea.

Cheffings CM, Farrell L, eds. 2005. The vascular plant red data list for Great Britain. Peterborough, UK: Joint Nature Conservation Committee.

Chinese Virtual Herbarium. 2016. *The Flora of China v. 5.0.* [WWW document] URL http://www.cvh.org.cn/ [accessed 15 January 2016].

CONABIO. 2016. *Sistema Nacional de Información sobre Biodiversidad.* [WWW document] URL https://www.gob.mx/conabio [accessed 11 April 2016].

Conti F, Abbate G, Alessandrini A, Blasi C. 2005. Annotated Checklist of the Italian Vascular Flora. Rome, Italy: Palombi Editori.

Costion C, Lorence D. 2012. The endemic plants of Micronesia. A geographical checklist and commentary. *Micronesica* 43: 51–100.

Danihelka J, Chrtek J, Kaplan Z. 2012. Checklist of vascular plants of the Czech Republic. *Preslia* 84: 647–811.

Dauby G, Zaiss R, Blach-Overgaard A, Catarino L, Damen T, Deblauwe V, Dessein S, Dransfield J, Droissart V, Duarte MC *et al.* 2016. RAINBIO. A mega-database of tropical African vascular plants distributions. *PhytoKeys* 74: 1–18.

Dimopoulos P, Raus T, Bergmeier E, Constantinidis T, Iatrou G, Kokkini S, Strid A, Tzanoudakis D. 2013. Vascular plants of Greece. An annotated checklist. Berlin, Athens: Botanischer Garten und Botanisches Museum Berlin-Dahlem, Freie Universität Berlin; Hellenic Botanical Society.

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.

Elizabeth Joyce, Kevin Thiele, Ferry Slik, Darren Crayn, Joyce E, Thiele K, Slik F, Crayn D. 2020. Checklist of the vascular flora of the Sunda-Sahul Convergence Zone. *Pensoft Publishers* 8: e51094.

Euro+Med. 2006. *Euro+Med PlantBase - the information resource for Euro-Mediterranean plant diversity*. [WWW document] URL http://ww2.bgbm.org/EuroPlusMed/ [accessed 1 June 2019].

Fischer MA, 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. Linz, Austria: Oberösterreichisches Landesmuseum.

Florence J, Chevillotte H, Ollier C, Meyer J-Y. 2007. *Base de données botaniques Nadeaud de l'Herbier de la Polynésie française (PAP)*. [WWW document] URL http://www.herbier-tahiti.pf/ [accessed 1 July 2011].

Francisco-Ortega J, Wang F-G, Wang Z-S, Xing F-W, Liu H, Xu H, Xu W-X, Luo Y-B, Song X-Q, Gale S *et al.* 2010. Endemic Seed Plant Species from Hainan Island: A Checklist. *The Botanical Review* 76: 295–345.

Freid E, Francisco-Ortega J, Jestrow B. 2014. Endemic Seed Plants in the Bahamian Archipelago. *The Botanical Review*: 1–27.

Funk VA, Hollowell T, Berry P, Kelloff C, Alexander SN. 2007. Checklist of the plants of the Guiana Shield (Venezuela: Amazonas, Bolivar, Delta Amacuro; Guyana, Surinam, French Guiana). Washington, DC: Department of Botany, National Museum of Natural History.

Garroutte MD. 2016. Species Richness, Community Compositon and Species Distribution Patterns in Aleutian Plants. Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science. University of Alaska Fairbanks.

Govaerts R, Nic Lughadha E, Black N, Turner R, Paton A. 2021. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Scientific Data* 8: 1.

Huang T-C. 1994. Flora of Taiwan, vol. I–VI. Taipei, Taiwan: National Taiwan University.

Hutchinson J, Dalziel JM, Keay RWJ, Hepper N. 2014. Flora of West Tropical Africa. London, UK: Crown agents for overseas governments.

Hyde MA, Wursten BT, Ballings P, Coates Palgrave M. 2016a. *Flora of Botswana*. [WWW document] URL http://www.botswanaflora.com/ [accessed 14 November 2016].

Hyde MA, Wursten BT, Ballings P, Coates Palgrave M. 2016b. *Flora of Malawi*. [WWW document] URL http://www.malawiflora.com/index.php [accessed 14 November 2016].

Hyde MA, Wursten BT, Ballings P, Coates Palgrave M. 2016c. *Flora of Mozambique*. [WWW document] URL http://www.mozambiqueflora.com/ [accessed 14 November 2016].

Hyde MA, Wursten BT, Ballings P, Coates Palgrave M. 2016d. *Flora of Zimbabwe* [accessed 14 November 2016].

Imada CT, ed. 2012. Hawaiian Native and Naturalized Vascular Plants Checklist. Honolulu, Hawaiʻi: Hawaii Biological Survey, Bishop Museum.

INBIO. 2000. *Lista de planta de Costa Rica. With updates by Eduardo Chacón*. [WWW document] URL http://www.inbio.ac.cr/papers/manual_plantas/index.html [accessed 5 February 2016].

Jahn R, Schönfelder P. 1995. Exkursionsflora für Kreta. Stuttgart, Germany: Ulmer (Eugen).

Jaramillo Díaz P, Guézou A, Mauchamp A, Tye A. 2018. CDF Checklist of Galapagos Flowering Plants. FCD Lista de especies de Plantas con flores Galápagos. Puerto Ayora, Galapagos: Charles.

Jardim Botânico do Rio de Janeiro. 2016. *Flora do Brasil 2020 em construção*. [WWW document] URL http://floradobrasil.jbrj.gov.br/ [accessed 9 May 2016].

Jongbloed M, Feulner GR, Böer B, Western AR. 2003. The comprehensive guide to the wild flowers of the United Arab Emirates. Abu Dhabi: Environmental Research and Wildlife Development Agency.

Karlsson T, Agestam M. 2014. *Checklist of Nordic vascular plants. Sweden Checklist*. [WWW document] URL http://www.euphrasia.nu/checklista/index.eng.html [accessed 11 February 2016].

Keppel G, Gillespie TW, Ormerod P, Fricker GA. 2016. Habitat diversity predicts orchid diversity in the tropical south-west Pacific. *Journal of Biogeography* 43: 2332–2342.

Kerguelen M. 2005. *Base de Données Nomenclaturales de la Flore de France*. [WWW document] URL http://www.tela-botanica.org/page:284# [accessed 13 January 2012].

Kirchner F, Picot F, Merceron E, Gigot G. 2010. Flore vasculaire de La Réunion. Réunion, France: Conservatoire Botanique National de Mascarin.

Komarov VL, Shishkin BK, Bobrov EG. 1933-1964. Flora of the U.S.S.R. Translated and published 1968. Springfield, US: Israel Program for Scientific Translations.

Kristinsson H. 2008. Checklist of the vascular plants of Iceland. Reykjavík, Iceland: Náttúrufræðistofnun Íslands.

Kuzmenkova SM. 2015. *Plants of Belarus*. [WWW document] URL http://hbc.bas-net.by/plantae/eng/default.php [accessed 15 February 2016].

Landcare Research New Zealand. *Ecological Traits of New Zealand Flora*. [WWW document] URL https://ecotraits.landcareresearch.co.nz/ [accessed Nov/2018].

Lazkov GA, Sultanova BA. 2011. Checklist of vascular plants of Kyrgyzstan. Helsinki: Botanical Museum, Finnish Museum of Natural History.

Marticorena C, Squeo FA, Arancio G, Muñoz M. 2008. Catálogo de la flora vascular de la IV Región de Coquimbo. In: Squeo FA, Arancio G, Gutiérrez JR, eds. *Libro rojo de la flora nativa y de los sitios prioritarios para su conservación: Región de Atacama*. Ediciones Universidad de La Serena La Serena, 105–142.

Meades SJ, Hay SG, Brouillet L. 2005. Annotated checklist of the vascular plants of Newfoundland and Labrador. Sault Ste. Marie, Ontario: The Provincial Museum of Newfoundland and Labrador, St. John's, NF, Canada.

Miller AG, Morris M. 2004. Ethnoflora of the Soqotra Archipelago. Edinburgh, UK: Royal Botanic Garden.

Ministry of Environment. 2012. The national red list 2012 of Sri Lanka. Conservation status of the fauna and flora. Colombo, Sri Lanka.

Morat P, Jaffré T, Tronchet F, Munziger J, Pillon Y, Veillon J-M, Chalopin M, Birnbaum P, Rigault F, Dagostini G *et al.* 2012. The taxonomic reference base "Florical" and characteristics of the native vascular flora of New Caledonia. *Adansonia* 34: 179–221.

Muer T, Sauerbier H, Cabrera Calixto F. 2016. Die Farn- und Blütenpflanzen der Kanarischen Inseln. Über 2.000 Pflanzenarten, mehr als 2.600 Fotos. Weikersheim: Margraf Publishers.

Nakamura K, Suwa R, Denda T, Yokota M. 2009. Geohistorical and current environmental influences on floristic differentiation in the Ryukyu Archipelago, Japan. *Journal of Biogeography* 36: 919–928.

Newman M. 2007. A checklist of the vascular plants of Lao PDR. Edinburgh, Scotland, UK: Royal Botanic Garden Edinburgh.

Nikolić T. 2016. *Flora Croatica Database*. [WWW document] URL http://hirc.botanic.hr/fcd [accessed 12 February 2016].

Norton J, Majid SA, Allan D, Al Safran M, Böer B, Richer RA. 2009. An illustrated checklist of the flora of Qatar. Gosport, UK: Browndown Publications Gosport.

Ovczinnikov PN. 1957-1991. Flora Tajikistana. Leningrad, Russia: Nauka.

Pandey RP, Diwakar PG. 2008. An integrated checklist flora of Andaman and Nicobar Islands, India. *J. Econ. Taxon. Bot* 32: 403–500.

Patzelt A. 2020. Plant checklist Oman.

Pavlov NV. 1954-1966. Flora Kazakhstana. Tom I-IX. Alma-Ata, Kazakhstan: Nauka Kazakhskoy SSR.

Pelser PB, Barcelona JF, Nickrent DL. 2011. *Co's Digital Flora of the Philippines*. [WWW document] URL https://www.philippineplants.org/ [accessed 13 January 2016].

Price JP, Wagner WL. 2011. A phylogenetic basis for species–area relationships among three Pacific Island floras. *American Journal of Botany* 98: 449–459.

Queensland Government. 2014. *Census of the Queensland flora 2014*. [WWW document] URL https://data.qld.gov.au/dataset/census-of-the-queensland-flora-2014 [accessed 5 February 2015].

Rodriguez R, Marticorena C, Alarcón D, Baeza C, Cavieres L, Finot VL, Fuentes N, Kiessling A, Mihoc M, Pauchard A *et al.* 2018. Catálogo de las plantas vasculares de Chile. *Gayana. Botánica* 75: 1–430.

Royal Botanic Gardens and Domain Trust. 2017. *PlantNET - The NSW Plant Information Network System*. [WWW document] URL http://plantnet.rbgsyd.nsw.gov.au [accessed 16 September 2016].

Sáez L, Rosselló JA. 2001. Llibre vermell de la flora vascular de les Illes Balears. Palma de Mallorca, Spain: Direcció General de Biodiversitat, Conselleria de Medi Ambient, Govern de les Illes Balears.

SANBI. 2014. *Plants of Southern Africa. An online checklist*. [WWW document] URL http://posa.sanbi.org [accessed 9 March 2015].

Sandbakk BE, Alsos IG, Arnesen G, Elven R. 1996. *The flora of Svalbard*. [WWW document] URL http://svalbardflora.no/ [accessed 16 March 2011].

Shaw JD, Spear D, Greve M, Chown SL. 2010. Taxonomic homogenization and differentiation across Southern Ocean Islands differ among insects and vascular plants. *Journal of Biogeography* 37: 217–228.

Short PS, Albrecht DE, Cowie ID, Lewis DL, Stuckey BM. 2011. Checklist of the vascular plants of the Northern Territory. Darwin: Department of Natural Resources, Environment, The Arts and Sport.

SLUFG. 2015. Rote Liste und Artenliste Sachsen. Farn- und Samenpflanzen. Dresden, Germany: Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie.

Smith AC. 1979-1996. Flora Vitiensis nova. a new Flora of Fiji (spermatophytes only). Lawaii, Hawaii: Pacific Tropical Botanical Garden.

Stephens K. 2011. Comparative floristic analysis of vegetation on the Dune Islands of South-East Queensland. *The Proceedings of the Royal Society of Queensland* 117: 141–180.

Strahm WA. *The conservation and restoration of the flora of Mauritius and Rodrigues,* University of Reading, Reading, UK.

Tamis WLM, van der Meijden R, Runhaar J, Bekker RM, Ozinga WA, Odé B, Hoste I. 2004. Standard List of the Flora of the Netherlands 2003. *Gorteria* 30: 101–195.

Tatewaki M. 1957. Geobotanical studies on the Kurile Islands. *Acta Horti Gotoburgensis* 21: 43–123.

Tela Botanica. 2015. *Base de données des Trachéophytes de France métropolitaine (bdtfx)*. [WWW document] URL http://www.tela-botanica.org/bdtfx.

Thomas J. 2011. *Plant diversity of Saudi Arabia. Flora checklist*. [WWW document] URL http://plantdiversityofsaudiarabia.info/biodiversity-saudi-arabia/flora/Checklist/Cheklist.htm [accessed 14 January 2016].

Tropicos. 2015a. *Catálogo de las Plantas Vasculares de Bolivia*. [WWW document] URL http://www.tropicos.org/Project/BC [accessed 10 September 2015].

Tropicos. 2015b. *Catalogue of the Vascular Plants of Ecuador*. [WWW document] URL http://www.tropicos.org/Project/CE [accessed 22 October 2015].

Tropicos. 2015c. *Flora de Nicaragua*. [WWW document] URL http://www.tropicos.org/Project/FN [accessed 16 July 2015].

Tropicos. 2015d. *Panama Checklist*. [WWW document] URL http://www.tropicos.org/Project/PAC [accessed 22 October 2015].

Tropicos. 2015e. *Peru Checklist*. [WWW document] URL http://www.tropicos.org/Project/PEC [accessed 22 October 2015].

Tropicos. 2020. *Catalogue of the Vascular Plants of Madagascar*. [WWW document] URL http://www.tropicos.org/Project/MADA [accessed 01/22/2020].

Turner IM. 1995. A catalogue of the vascular plants of Malaya. Singapore: Gardens' Bulletin.

UIB. 2007. *Herbario virtual del Mediterráneo Occidental*. [WWW document] URL http://herbarivirtual.uib.es/cas-med/ [accessed 7 August 2012].

Ulloa Ulloa C, Acevedo-Rodríguez P, Beck S, Belgrano MJ, Bernal R, Berry PE, Brako L, Celis M, Davidse G, Forzza RC *et al.* 2017. An integrated assessment of the vascular plant species of the Americas. *Science (New York, N.Y.)* 358: 1614–1617.

USDA, NRCS. 2015. *The PLANTS Database*. [WWW document] URL http://plants.usda.gov [accessed 24 April 2015].

Velayos Rodríguez M, Aedo Pérez C, Cabezas Fuentes FJ, de la Estrella González, Manuel, Fero Meñe MS, Sánchez PB, Quintanar Sánchez A, Medina Domingo L. 2021. *Flora de Guinea. Base de datos de Flora de Guinea*. [WWW document] URL http://www.floradeguinea.com/herbario/ [accessed 2 March 2021].

VicFlora. 2016. *Flora of Victoria*. [WWW document] URL http://vicflora.rbg.vic.gov.au [accessed 30 November 2016].

Virtual Biodiversity Museum of Cyprus. *Flora of Cyprus - Endemic plants of Cyprus*. [WWW document] URL http://www.naturemuseum.org.cy/lang1/endemic_plants.html [accessed 02.20.2020].

Vogt C. 2011. Composición de la Flora Vascular del Chaco Boreal, Paraguay. I. Pteridophyta y Monocotiledoneae. *Steviana* 3: 13–47.

Wagner WL, Herbst DR, Lorence DH. 2005. *Flora of the Hawaiian Islands website*. [WWW document] URL http://botany.si.edu/pacificislandbiodiversity/hawaiianflora/ [accessed 16 October 2010].

Western Australian Herbarium. 2017. *FloraBase - the Western Australian Flora*. [WWW document] URL https://florabase.dpaw.wa.gov.au/ [accessed 16 September 2016].

Wieringa JJ. 2016. Flora of Gabon. unpublished.

Yena AV. 2012. Prirodnaja flora krymskogo poluostrova. Spontaneous flora of the Crimean Peninsula. Simferopol': Orianda.

Yonekura K, Kajita T. 2013. *BG Plants Japanese name and Scientific name Index (YList)*. [WWW document] URL http://ylist.info [accessed 6 February 2016].

ZDSF, SKEW. 2014. *Info Flora. Artenliste Schweiz 5x5 km*. [WWW document] URL https://www.infoflora.ch/de/daten-beziehen/artenliste-5x5-km.html [accessed 13 February 2015].

Zuloaga FO, Morrone O, Belgrano M. 2014. *Catálogo de las Plantas Vasculares del Cono Sur*. [WWW document] URL http://www.darwin.edu.ar/Proyectos/FloraArgentina/fa.htm [accessed 16 March 2015].