IMPERIAL

Climate and the biogeography of plant function and diversity

Jiaze Li

Supervision: Professor Iain Colin Prentice



Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Life Sciences Research and
the Diploma of Imperial College London
September 2025

Declaration of Originality

I declare that the research I present in this thesis is my original work, conducted under the supervision of Professor Iain Colin Prentice. Any material which could be construed as other people's work has been referenced and cited appropriately. All parts of this thesis which been conducted in collaboration with other researchers has been indicated in detail.

Jiaze Li

September 2025

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Funding

This research was funded by a PhD scholarship granted by the China Scholarship Council (CSC No. 202108060263).

Acknowledgements

If this life is the only one we can be sure of having, then I believe it is worth dedicating it to the things and people we truly love. I am lucky not only to find a research direction that excites me, but also to have been guided, supported, encouraged, and accompanied by teachers, family, and friends who have enabled me to grow both academically and personally.

First and foremost, I would like to express my deepest gratitude to my supervisor, Professor Colin Prentice, who has been my longest-standing mentor for almost six years now and will continue to be. You are also the person I most admire. You once told me: "*I've always done what I like. I would recommend doing what you like.*" You gave me the freedom to pursue projects of my own interest, trained me in the research skills needed to complete my work, and shared your knowledge and experience generously. You inspired me to ask, consider, and answer questions from multiple perspectives. Your guidance has been central to my development as an independent researcher. Even when I went through an entire year without progress, you continually supported me, reminding me that this is a normal part of research and that I would eventually find a way. Your encouragement enabled me to overcome difficulties, strengthen my resilience, and maintain confidence in my work.

My deepest thanks also go to my parents, Kaige Wang and Weijie Li. You are my best friends and my first (two) soulmates! Your unwavering support, trust, and protection have allowed me to pursue what I love in health and happiness. Your respect for nature and curiosity about the world have influenced me profoundly. Because of your constant love, I have felt able to explore every path that interested me. My thanks also go to my grandparents, aunts, uncles, and cousins, for your continued support, encouragement, and care, and for respecting my choices in life.

I would like to thank all my friends. Your friendship has been my source of warmth and reassurance, much like family. You shared and celebrated joyful moments with me, always stood by me and offered me advice when I faced challenges. We are also collaborators in our work, offering each other research advice and inspiration, and generously and patiently sharing our expertise. Our personal and academic interactions have been invaluable.

Special thanks to Silwood Park. The sound and scent of the wind passing through the trees, and the sunlight and raindrops dancing on the leaves, have always been my greatest sources of comfort and inspiration. I have witnessed the changing of the seasons at Silwood: from budding plants to colouring forests, flowers blooming and fading in succession, and young rabbits and deer growing into new families. All the life in Silwood has accompanied me, witnessed my growth, and listened to my joys, sorrows, and wishes.

Most importantly, I am grateful for this PhD journey. To me, ecology is the study of the many stories that nature holds. It has been a privilege to collect and share some of these wonderful stories with others through ecological research. I also wish to thank my younger self – the child who delighted in gathering fallen leaves, flowers, and seeds, and who has now grown into a researcher continuing to study and tell the stories of plants. I hope to always preserve the same curiosity and love for exploring the vastness of the world while cherishing the life of a single plant.

Abstract

Climate shapes terrestrial ecosystems by determining which species and functions can persist under given environmental conditions. Recent climate governs the large-scale distribution of contemporary ecosystems and their function; long-term climate dynamics explain why species and functions with distinct evolutionary histories are found in today's environments. However, there is currently no framework that integrates these spatial and temporal dimensions to explain how ongoing climate change shapes terrestrial ecosystems.

Plants form the structural and trophic basis of ecosystems on land. They regulate community composition, biogeochemical cycles, and productivity by responding to climate. This thesis investigates how climate influences plant functional traits and diversity across space and time, using an analytical framework that integrates geographic and climate-defined multidimensional spaces. Firstly, I quantify the relationships among key plant functional traits and between these traits and contemporary climate across herbaceous, deciduous and evergreen woody plants. This generates the most comprehensive global trait maps to date based on a statistical upscaling approach. Secondly, by tracing the emergence, persistence, expansion, contraction, and movement of climatic conditions across the Earth's surface over approximately the past 50 million years, I show that historical persistence and shifts in the geographic extent of climates are crucial for explaining contemporary patterns of global vascular plant diversity. Thirdly, I use C4 plants as a model functional group to reconstruct their biogeographic dynamics over the past 50 million years, based on C4 diversity data and eco-evolutionary optimality models. This analysis demonstrates how Cenozoic climate change drove the origin and spread of C4 photosynthesis.

Taken together, these findings imply that the functional and diversity patterns of modern ecosystems are the result of spatio-temporal variations in climate. This integrated explanatory framework provides an underpinning for better understanding of the resilience and adaptive capacity of the biosphere in an ever-changing world.

Contents

Acknowledgements 2

Abstract 4

Chapter 1: General introduction 10

1.1 Climate: the spatio-temporal driver of ecological patterns 10

1.2 Plant function and diversity 12

1.3 Geographic space and climatic space 18

1.4 Thesis outline 20

Chapter 2: Global patterns of plant functional traits and their relationships to climate 27

2.1 Abstract 27

2.2 Main 27

2.3 Results 33

2.4 Discussion 45

2.5 Methods 51

Chapter 3: Deep-time geographic dynamics of climate shape global vascular plant diversity 59

3.1 Abstract 59

3.2 Main 59

3.3 Results 62

3.4 Discussion 79

3.5 Methods 82

3.6 Extended data figures and tables 91

Chapter 4: Global biogeography of C4 plants through evolutionary time 106

4.1 Abstract 106

4.2 Main 106

4.3 Results 109

4.4 Discussion 124

4.5 Methods 129

4.6 Extended data figures and tables 137
Chapter 5: Conclusion 150
5.1 Key findings 150
5.2 Limitations and future directions 151
5.3 Implications 154
Bibliography 158
Appendices 183
A1 Supplementary information for Chapter 2 183
A2 Supplementary information for Chapter 3 241
A3 Supplementary information for Chapter 4 321
Appendix bibliography 361

List of Figures and Tables

Fig. 1.1: Climate changes over geological timescales. 11

Fig. 1.2: Hutchinson's duality. 19

Fig. 1.3: Representative figure for Chapter 2 – Plant functional trait distributions in climatic and geographic spaces. 21

Fig. 1.4: Representative figure for Chapter 3 – Present-day plant diversity and the age and persistence of recent climatic conditions. 23

Fig. 1.5: Representative figure for Chapter 4 – Global distribution and temporal dynamics of C4 plant diversity. 25

Table 2.1 | Plant functional traits used in this study. 31

Fig. 2.1: Trait-trait and trait-climate correlations. 34

Fig. 2.2: Relative importance of three bioclimatic variables. 36

Fig. 2.3: Examples of climate space diagrams. 37

Fig. 2.4: Examples of global trait maps. 42

Fig. 2.5: Comparison between predicted trait values and the iNaturalist estimated trait values. 44

Fig. 3.1: Mapping climates in geographic and climatic spaces. 64

Fig. 3.2: Global patterns of vascular plant diversity across geographic and climatic spaces. 66

Fig. 3.3: Age and persistence of climate and their associations with recent species richness. 68

Fig. 3.4: Relationships between recent plant species richness and three explanatory variables across selected geological periods. 70

Fig. 3.5: Influence of climate and its deep-time geographic dynamics on plant diversity. 77

Fig. 3.6: Relative importance of climate and its deep-time geographic dynamics in explaining plant diversity. 78

Extended Data Fig. 3.1 Historical geography of recent climates. 91

Extended Data Fig. 3.2 Relative importance of climate and its historical geographic extent in explaining plant species richness. 93
Extended Data Fig. 3.3 Relationships between recent plant diversity and age or persistence of climate. 95
Extended Data Fig. 3.4 Influence of climatic conditions first occurring in the Eocene (*ca*. 50 Ma) and their deep-time geographic dynamics on plant diversity. 97
Extended Data Fig. 3.5 Relative importance of climate climatic conditions first occurring in the Eocene (*ca*. 50 Ma) and their deep-time geographic dynamics in explaining plant diversity. 99
Extended Data Fig. 3.6 Influence of climatic conditions first occurring in the Mid-Pliocene (3.3–3.0 Ma) and their deep-time geographic dynamics on plant diversity. 100
Extended Data Fig. 3.7 Relative importance of climate climatic conditions first occurring in the Mid-Pliocene (3.3–3.0 Ma) and their deep-time geographic dynamics in explaining plant diversity. 102
Extended Data Fig. 3.8 Influence of climatic conditions first occurring in the Last Glacial Maximum (LGM, 22–18 ka) and their deep-time geographic dynamics on plant diversity. 103
Extended Data Fig. 3.9 Relative importance of climate climatic conditions first occurring in the Last Glacial Maximum (LGM, 22–18 ka) and their deep-time geographic dynamics in explaining plant diversity. 105

Fig. 4.1: Contemporary global patterns of C4 plant diversity in geographic space. 110
Fig. 4.2: Contemporary global patterns of C4 plant diversity in climatic space. 112
Fig. 4.3: Simulated historical biogeography of C4 plants in geographic space. 115
Fig. 4.4: Simulated historical biogeography of C4 plants in climatic space. 118
Fig. 4.5: Dominant environmental drivers of simulated C4 plant biogeography in geographic space. 121
Fig. 4.6: Dominant environmental drivers of simulated C4 plant biogeography in climatic space. 123
Extended Data Fig. 4.1 Contemporary global patterns of C4 Poales diversity in geographic space. 137
Extended Data Fig. 4.2 Contemporary global patterns of C4 Poales diversity in climatic space. ... 138

Extended Data Fig. 4.3 Contemporary global patterns of non-Poales C4 diversity in geographic space. 140
Extended Data Fig. 4.4 Contemporary global patterns of non-Poales C4 diversity in climatic space. 141
Extended Data Fig. 4.5 Simulating potential tree cover based on net carbon profit of trees. 143
Extended Data Fig. 4.6 Simulated net carbon profit for trees and tree cover across geological time slices. 144
Extended Data Fig. 4.7 Second dominant environmental drivers of simulated C4 plant biogeography in geographic space. 146
Extended Data Fig. 4.8 Second dominant environmental drivers of simulated C4 plant biogeography in climatic space. 148

Chapter 1: General introduction

1.1 Climate: the spatio-temporal driver of ecological patterns

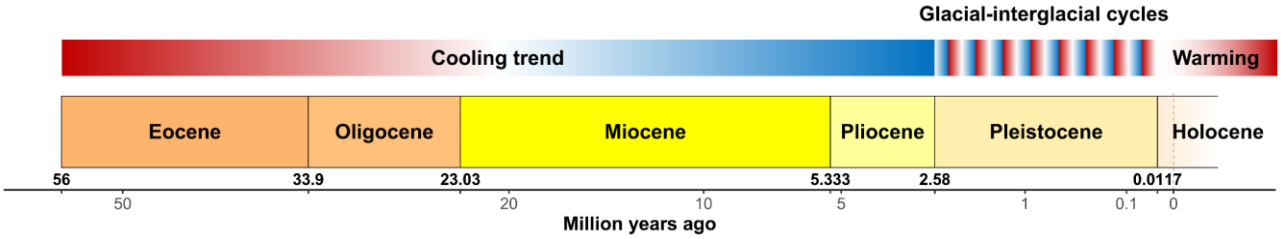
Among all environmental factors, climate is the principal driver shaping the structure (e.g. community composition)[1–5] and function (e.g. ecosystem productivity and biogeochemical cycles) of terrestrial ecosystems[4,6–9]. Climatic variables, principally the availability of energy and water, impose fundamental physiological constraints on organisms, thereby functioning as a large-scale environmental filter that determines which species can persist, thrive, and assemble into a community[10–12]. In turn, the aggregate functional traits and metabolic rates of this assembled community, which are themselves co-limited by climatic conditions, directly govern the ecosystem-level fluxes of energy and matter, such as primary productivity and water and nutrient cycling[13–15]. The consequences of this climatic filtering are evident both spatially, in the large-scale geographic patterns of contemporary biomes, and temporally, in the deep evolutionary and migratory histories of lineages responding to long-term climate change.

The Earth's terrestrial surface forms a mosaic of large-scale ecosystems, from tropical rainforests to subtropical deserts and to Arctic tundra, whose geographic distributions are shaped by macroclimatic gradients of energy and water availability[5]. In equatorial regions, high energy input and abundant water support tropical forests, characterised by complex, high-biomass vegetation that sustains the highest terrestrial species richness and facilitates exceptional rates of net primary productivity and nutrient cycling[16–21]. In contrast, ecosystems constrained by either water or energy display markedly different structures and functions. In subtropical deserts, where water availability is typically the primary limiting factor, biodiversity is generally low, and species are often highly specialised to cope with drought and temperature extremes[22]. In the Arctic, low solar energy input and long freezing periods largely restrict vegetation to tundra ecosystems, characterised by low species diversity dominated by stress-tolerant and low-stature plants, and among the lowest productivity and decomposition rates recorded for all terrestrial ecosystems[23,24]. Similarly, elevational gradients create analogous climatic transitions, with temperature generally declining and

precipitation patterns shifting with increasing altitude, resulting in a compressed succession of ecosystem types – from montane forests to alpine tundra – over short geographical distances[1,25].

However, the spatial patterns of terrestrial ecosystems observed today are far from static. Over geological timescales, Earth's climate has undergone substantial (and sometimes abrupt) fluctuations[26–28] (Fig. 1.1). A long-term global cooling trend began around 55 million years ago after the Paleocene-Eocene Thermal Maximum (PETM). This trend was accompanied by a multi-million-year decline in atmospheric CO_2, thought to be caused by a long-term imbalance between volcanic sources and the chemical weathering sink. This decline continued into the Quaternary period[29]. The Quaternary is characterized by repeated glacial-interglacial cycles paced by variations in Earth's orbit. The glacial periods were further punctuated by abrupt climate changes, including Dansgaard-Oeschger events – during which Greenland temperatures rose by 10–15 °C within decades – and the rapid final warming that terminated the last glacial period[29].

Fig. 1.1: Climate changes over geological timescales.



Geological periods are coloured according to the *International Chronostratigraphic Chart* (https://stratigraphy.org/chart)[30]. Time is shown on a square-root scale.

These historical climatic shifts have shaped ecosystem structure and function by driving habitat creation, alteration of connectivity, and habitat loss[31]. From a structural perspective, these deep-time climatic dynamics drove continuous reorganisation of communities. They caused large-scale species migrations, triggered extinctions, and promoted speciation[26,32–34]. Repeated contraction of populations into climatic refugia during cold glacials, followed by expansion and secondary contact in warmer interglacials, promoted genetic divergence[35]. These processes have directly influenced the species composition and genetic structure of modern ecosystems. From a functional perspective,

long-term climatic trends have exerted selective pressures leading to major evolutionary innovations. For example, the decline in atmospheric CO_2 during the Cenozoic was a key driver for the evolution and expansion of C4 photosynthesis[36–39]. This photosynthetic pathway enhances photosynthetic efficiency under conditions of low CO_2, high temperature, and drought[40,41]. Such adaptations fundamentally altered productivity, fire regimes, and biogeochemical cycling across extensive tropical and subtropical grasslands and savannahs[42–47].

While a spatial analysis can reveal how contemporary climate governs the global patterns of ecosystem structure and function, temporal perspectives are essential to explain why particular species and functions, with their unique evolutionary histories, occur in a given location and assemble under present climatic conditions. Therefore, a comprehensive understanding of how climate shapes the living world requires integrating static geographic patterns with dynamic geohistorical processes. Achieving this integration remains a major challenge in ecology and biogeography, and forms the theme of this thesis.

1.2 Plant function and diversity

Land plants originated during the middle Cambrian-Early Ordovocian interval (*ca*. 510–470 million years ago)[48]. This evolutionary event transformed barren continents into diverse, biologically complex ecosystems[49]. Constituting over 80% of all terrestrial biomass, plants form the structural and trophic basis of terrestrial ecosystems, creating habitats that sustain almost all other life on land[50]. As primary producers, plants are the main pathway in which key elements such as carbon and nitrogen enter the terrestrial biosphere[51,52]. Their evolutionary rise reshaped global biogeochemical cycles and influenced Earth's climate system. By enhancing phosphorus weathering and increasing the burial of organic carbon, early land plants drove a major mid-Palaeozoic oxygenation event (~400 Ma), raising atmospheric oxygen to near-modern levels[53]. The rise of plants also altered long-term atmospheric CO_2 concentrations by increasing the production and burial of organic matter in sediments and by accelerating the chemical weathering of silicate rocks[54]. Through albedo, evapotranspiration and canopy roughness, plants further influenced planetary energetics, the hydrologic cycle and atmospheric composition[55], and regulated regional precipitation

patterns and temperature regimes[56]. Over the past two centuries, following the Industrial Revolution (~1750), terrestrial ecosystems have taken up a significant fraction of anthropogenic CO_2 emissions from fossil fuel burning, with vegetation, soils and permafrost currently sequestering approximately one-third of these anthropogenic CO_2 emissions[57].

Because plants directly mediate the exchange of energy and water between land and the atmosphere, their performance is tightly coupled to climatic conditions[5]. This coupling exerts selective pressure on the physiological and morphological traits of plants, determining whether a species can persist and function in a given environment over time[11,58]. Given their foundational role in the biosphere and their close interactions with climate, plants serve as a lens through which to examine the spatial and temporal impacts of climate on terrestrial ecosystems.

1.2.1 Plant functional traits

Plant functional traits (FTs) are the morphological, physiological and phenological characteristics of individual plants[58]. They determine plant growth, reproduction and survival strategies by influencing life-history processes (e.g., seed traits), resource uptake and utilisation (e.g., leaf nutrient contents), and responses to environmental change and disturbances (e.g., leaf traits and plant height)[58–60].

These traits covary along a limited number of dimensions that reflect fundamental trade-offs in ecological strategies[61–63]. The global spectrum of plant form and function[63] describes two major orthogonal dimensions of trait covariation: one represents the size of the plant and its organs (e.g., plant height and seed mass), while the other is the well-known 'leaf economics spectrum' (LES). The LES describes a universal trade-off between rapid resource acquisition (e.g. low leaf mass per area) and resource conservation (e.g. high leaf mass per area). This spectrum is expressed across nearly all plant communities worldwide[62]. While the LES itself is a fundamental axis of trait variation, environmental conditions determine the relative representation of different strategies along this spectrum[59,64–67]. For instance, warm and resource-rich environments tend to favour acquisitive species with fast return on investment, whereas cold or nutrient-poor environments select for conservative species with long-lived, resource-efficient leaves[59,64–67]. By influencing species-level dynamics such as migration and local extinction, trait covariation shapes the assembly

of plant communities and, in turn, determines the macro-scale geographic patterns and long-term persistence of vegetation types[68].

A quantitative understanding of relationships among traits, and between traits and climate, is essential for predicting how the global distribution of vegetation will respond to continuing climate change[65,68]. Dynamic Global Vegetation Models (DGVMs), the tools most widely used for such simulations, rely on global maps of key traits as inputs[69,70]. Traditionally, such global trait maps have been estimated by classifying plants into a small number of Plant Functional Types (PFTs), each assigned a fixed trait value[69–72]. However, assigning a single trait value to all members within the same PFT ignores the observed variation of a trait among plant individuals[71,72].

Beyond PFT-based modelling, a variety of approaches have been developed to estimate global trait distributions. Satellite remote sensing enables direct mapping of certain leaf and canopy traits at large scales, but is restricted to a limited set of variables and typically achieves only moderate accuracy[73–75]. Statistical upscaling and machine learning methods, which relate traits to environmental predictors, have been applied to generate global maps of several traits (mainly leaf traits)[65,72,74,76–80]. More recently, optimality-based modelling has been used to simulate global patterns of multiple leaf traits from eco-evolutionary principles[81]. Complementary data-driven approaches have also advanced, including the use of large plot-based compilations such as sPlotOpen[82,83], species-level trait databases such as TRY[84,85], and even the global citizen science project such as iNaturalist[86], which provide valuable, spatially explicit observations. Despite this diversity of methods, existing global trait maps are often inconsistent, reflecting differences in input data, modelling frameworks, and spatial coverage[81,87]. Moreover, fundamental trait-trait and trait-climate relationships are not universal – they vary among major plant groups such as woody versus herbaceous[88], and deciduous versus evergreen species[89–91]. To generate more realistic and comprehensive global trait maps, we need a quantitative understanding of how these relationships differ across plant groups. This limitation motivates the central objective of **Chapter 2** in this thesis: to account for important functional differences among evergreen, deciduous, and herbaceous plants, and to evaluate how these differences alter both trait-trait and trait-climate relationships, with the goal of creating a comprehensive set of high-resolution global plant trait maps.

1.2.2 Plant diversity

Plant functional traits influence how species respond to environmental conditions, interact with one another, and withstand disturbance. Traits that allow species to tolerate local abiotic stresses determine which species can establish and persist, while differences in traits promote niche partitioning and coexistence, and similarities can intensify competition and exclusion[92–94]. Traits linked to recovery and resilience also affect how communities reassemble after disturbance[59,95,96]. Together, these processes govern species establishment, dominance, and coexistence, shaping the diversity patterns observed across regions.

Plant diversity encompasses multiple complementary dimensions that capture different aspects of variation within and among plant communities. Taxonomic diversity reflects the number of taxa present and their relative abundances[97]. Functional diversity describes the range, distribution, and divergence of plant functional traits within a community, linking species diversity to ecosystem processes including productivity, nutrient cycling, and resilience[98,99]. Phylogenetic diversity quantifies the evolutionary distances among co-occurring species, representing the accumulated evolutionary history of a community[100]. These dimensions can be examined at multiple spatial scales[101,102]: α diversity captures local species richness and composition; β diversity measures the turnover or nestedness of species and traits between sites and along environmental gradients; and γ diversity represents the total diversity across a region. Integrating these dimensions provides a more comprehensive understanding of the mechanisms shaping global community patterns under environmental change[103].

The latitudinal diversity gradient (LDG) is a fundamental biodiversity pattern on Earth, describing the observation that tropical regions harbour more species than higher-latitude areas[104,105]. This pattern was first recognised by Johann Reinhold Forster and Alexander von Humboldt in the late eighteenth and early nineteenth centuries[1,106–108]. Multiple, often competing, hypotheses have been proposed to explain how climate shapes this gradient[106,109,110]. The ‘energy-water hypothesis’ posits that the warm, humid, and productive conditions in the tropics support larger populations, thereby reducing extinction risk, while elevated temperatures in the tropics may also accelerate speciation by increasing metabolic and mutation rates[111–113]. The ‘time-stability hypothesis’ argues that the tropics, being older and less affected by past glaciations,

have allowed lineages more time to accumulate in a relatively stable environment, promoting speciation and lowering extinction[27,35,109,110,114–117]. The 'area-isolation hypothesis' suggests that larger land area and complex structure of tropical ecosystems provide more niches, reducing competition and allowing more species to coexist, while geographic barriers in the tropics may promote population isolation and further speciation[118–121].

All these hypotheses tend to emphasise single aspects of climate, focusing either on its magnitude, its temporal stability, or its spatial distribution, without integrating these dimensions into a unified spatio-temporal framework. This limitation has contributed to ongoing debate over the dominant climatic drivers of the LDG. To address this gap, **Chapter 3** of this thesis quantifies when and where climatic conditions have appeared, persisted and shifted across the Earth's surface over the past 50 million years, and investigates how such spatio-temporal climate dynamics have shaped contemporary patterns of plant diversity across multiple dimensions.

1.2.3 Diversity of a functional group

The interaction between climate, function, and diversity can be illustrated by the evolution of a major plant functional group. Photosynthesis is a fundamental physiological process underpinning plant growth, reproduction, and survival. Terrestrial plants assimilate atmospheric CO_2 via distinct photosynthetic pathways, primarily C3 and C4, which differ in their biochemical mechanisms and ecological adaptations[38,41,122,123]. The evolution of photosynthetic pathways provides a case study for how climatic conditions select for specific functional strategies, which in turn influence species diversification, distribution, and dominance.

The ancestral C3 photosynthetic pathway is sensitive to atmospheric composition because the carboxylating enzyme of photosynthesis, Rubisco, catalyses both the fixation of CO_2 (carboxylation, contributing to carbon gain) and the fixation of O_2 (oxygenation)[38,40,41,124]. The latter initiates photorespiration, a process that does not generate energy but instead consumes ATP and reducing power produced during the light reactions of photosynthesis. In addition, photorespiration releases previously fixed carbon back into the atmosphere as CO_2, resulting in a loss of carbon and a reduction in photosynthetic efficiency[38,40,41,124]. Photorespiration is primarily governed by the availability of CO_2 within the leaf[41]. At lower ambient CO_2 concentrations, the competition between

CO_2 and O_2 at the Rubisco active site is intensified. Higher temperatures exacerbate this effect by reducing Rubisco's relative affinity for CO_2 relative to O_2, while water-limited conditions can indirectly enhance photorespiration through stomatal closure, which lowers internal CO_2 availability[41].

During the mid- to late Cenozoic, a decline in global atmospheric CO_2 concentrations, coupled with increasing aridity, imposed increasing physiological stress on C3 plants[38,125,126]. Rather than evolving a more efficient Rubisco, natural selection favoured innovations in cellular systems that concentrate CO_2 around Rubisco[41]. The C4 photosynthetic pathway represents one such innovation, employing a CO_2-concentrating mechanism (CCM) to elevate CO_2 levels around the Rubisco[38,122,123]. In C4 plants, atmospheric CO_2 is initially fixed by phosphoenolpyruvate carboxylase (PEPc) into four-carbon acids (giving the C4 photosynthesis its name) in leaf mesophyll cells[38,40,41,122]. These acids are then transported to bundle sheath cells, where they are decarboxylated to release CO_2 in a compartment with limited gas exchange with intercellular air spaces. This spatial separation effectively reduces photorespiration[38,40,41,122]. The C4 pathway enhances carbon-fixation efficiency in environments that restrict CO_2 supply to Rubisco, allowing for higher CO_2 assimilation rates alongside greater nitrogen- and water-use efficiencies[41]. By decoupling the physiological trade-off between fast, acquisitive growth and high resource-use efficiency in C3 plants, C4 photosynthesis enables a 'fast-efficient' phenotype capable of thriving in hot, arid, saline, or nutrient-poor environments[41].

The ecological success of this C4 photosynthesis innovation has reshaped terrestrial ecosystems. C4 plants, particularly grasses and sedges, dominate most grasslands and savannas in tropical and subtropical regions[38,42,127]. Despite comprising only ~3% of land plant species[41,128], C4 plants are responsible for about 20% of global gross primary production[43–47]. This ecological dominance drove one of the most significant global vegetation changes through the Miocene/Pliocene boundary, leading to the expansion of grasslands, which in turn influenced the co-evolution of herbivorous fauna specialising on C4 plants[36,42].

This example illustrates how a functional innovation, arising in response to global climatic change, can drive the assembly of new biomes and fundamentally reshape the functioning of terrestrial ecosystems. However, our understanding of the diversification of this functional group

and its climatic drivers in a spatio-temporal context remains constrained by two major challenges: (1) incomplete records of C4 species and their present-day geographic distributions, and (2) geographically sparse and temporally fragmented palaeoecological evidence. These limitations provide the motivation for **Chapter 4** of this thesis: applying modelling approaches that integrate plant physiology, evolutionary history, and palaeoclimate with comprehensive C4 diversity data to reconstruct the biogeographic dynamics of C4 plants across geological intervals spanning its evolutionary history.

1.3 Geographic space and climatic space

The structure and function of ecosystems are ultimately determined by the distribution of species and their functional traits. However, the presence of a species at a given location and time depends not only on the suitability of the local climate, but also on its ability to persist within the local biotic community, and on whether it has been able to reach that location[129,130]. As a result, analysing the relationships between climate, function, and diversity solely in geographic space tends to mix the effects of environmental filtering with those of geographic processes, such as dispersal limitation and long-term isolation[121,131]. This confounding can make it harder to identify the general principles by which climate shapes ecological patterns.

A useful framework for disentangling these processes is the ecological niche. Grinnell (1917)[132] proposed that each species could persist only where a particular set of environmental conditions occurred in a specific place, which he termed the species' 'niche'. Hutchinson (1957)[133] transformed this concept by defining the 'fundamental niche' as an n-dimensional hypervolume of abiotic conditions within which a species can survive and thrive, thereby separating the niche concept from any specific geographic location. Following this logic, the range of multiple climatic variables that a species can tolerate defines its 'climatic niche'. This concept allows the construction of a measurable, multi-dimensional 'climatic space' in which each point represents a unique climatic condition.

This framework reveals a fundamental duality: a single point in climatic space (a specific climatic condition) can correspond to multiple, geographically disjunct regions on Earth[121,131]. This

idea, known as 'Hutchinson's duality'[134] (Fig. 1.2), enables the mapping of information between climatic and geographic spaces, providing complementary perspectives. Analyses in geographic space capture realised ecological and evolutionary patterns in their physical locations, while analyses in climatic space reveal the general climatic drivers of those patterns[121,131]. Climatic space also provides a stable reference framework for tracking climate change through deep time. A specific climatic condition can emerge, persist, expand, contract, or shift its geographic distribution over millions of years – a dynamic that is difficult to compare across different geological periods with changing continental configurations.

Fig. 1.2: Hutchinson's duality.

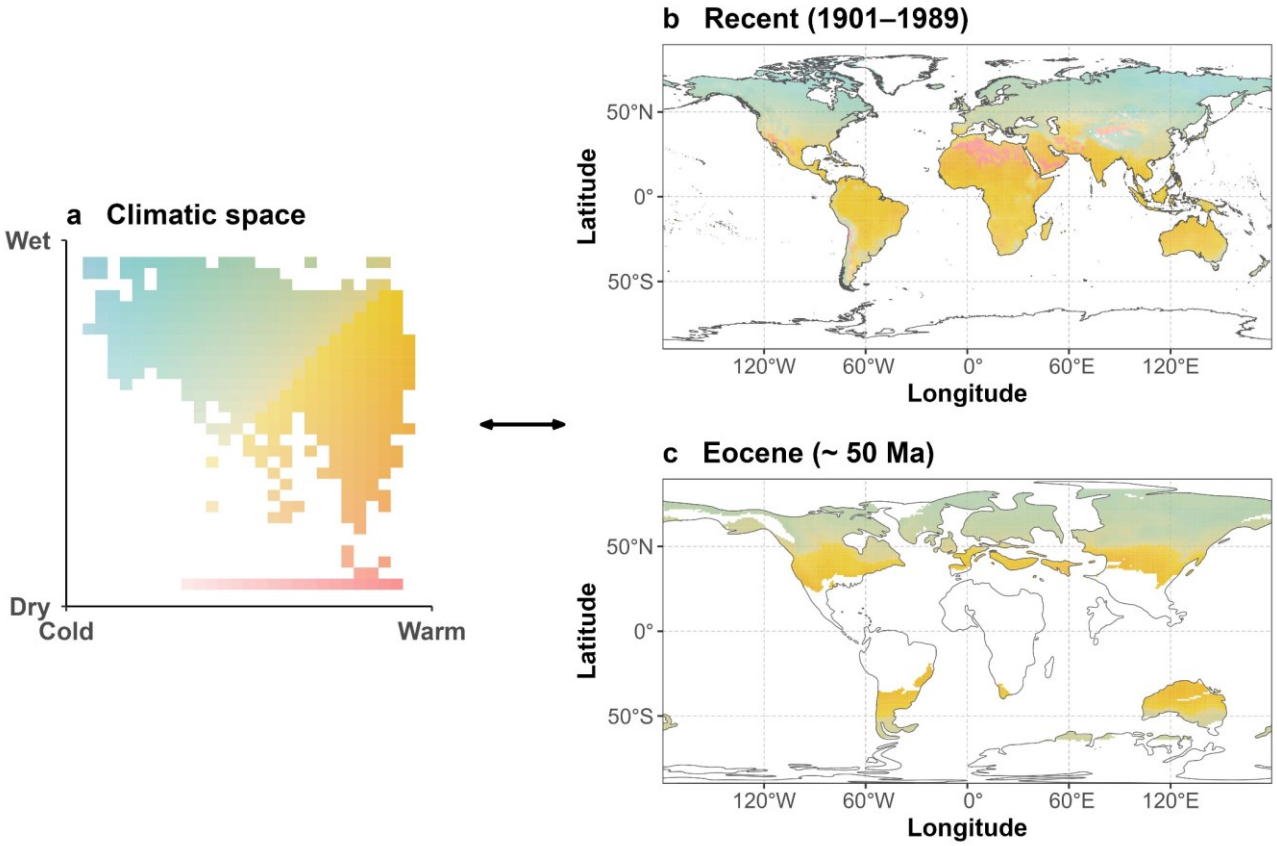


Illustration of how a single climatic condition can occur in multiple geographic locations, whereas different geographic locations may correspond to the same climatic condition. **a**, Two-dimensional climatic space defined by temperature and moisture, with each coloured grid cell ('climatic bin') representing a unique climatic condition. **b**, Geographic locations of all climatic bins (defined in **a**) during the recent period (1901–1989). **c**, Geographic locations of the same climatic conditions

during the Eocene (*ca.* 50 Ma); uncoloured areas indicate climatic conditions absent today. Colours in **b** and **c** correspond to those in **a**, with identical colours indicating identical climatic conditions.

This thesis investigates the relationships between climate, plant function, and plant diversity from perspectives of both geographic and climatic spaces. **Chapter 2** maps the distributions of key functional traits in both spaces to test whether covarying traits show similar relationships with climate. **Chapter 3** uses the climatic space framework to quantify the persistence and shift in geographic extent of climates over the past 50 million years and to assess how these spatio-temporal dynamics have shaped contemporary plant diversity. **Chapter 4** applies this dual-space approach to reconstruct the biogeographic history and climatic niche evolution of C4 plants through deep-time climate change.

1.4 Thesis outline

This thesis investigates how climate shapes the global biogeography of plant function and diversity across space and time. Using an analytical framework that integrates perspectives from both geographic and climatic spaces, it aims to characterise global patterns of plant functional traits in relation to contemporary climate, evaluate how deep-time climate dynamics have influenced global plant diversity, and reconstruct the biogeographic history of a key functional group through evolutionary time. The thesis is organised into five chapters:

Chapter 1: General introduction

This chapter introduces the core concepts, outlines the theoretical background, and presents the central research questions and overall structure of the thesis. I am the sole contributor to this chapter.

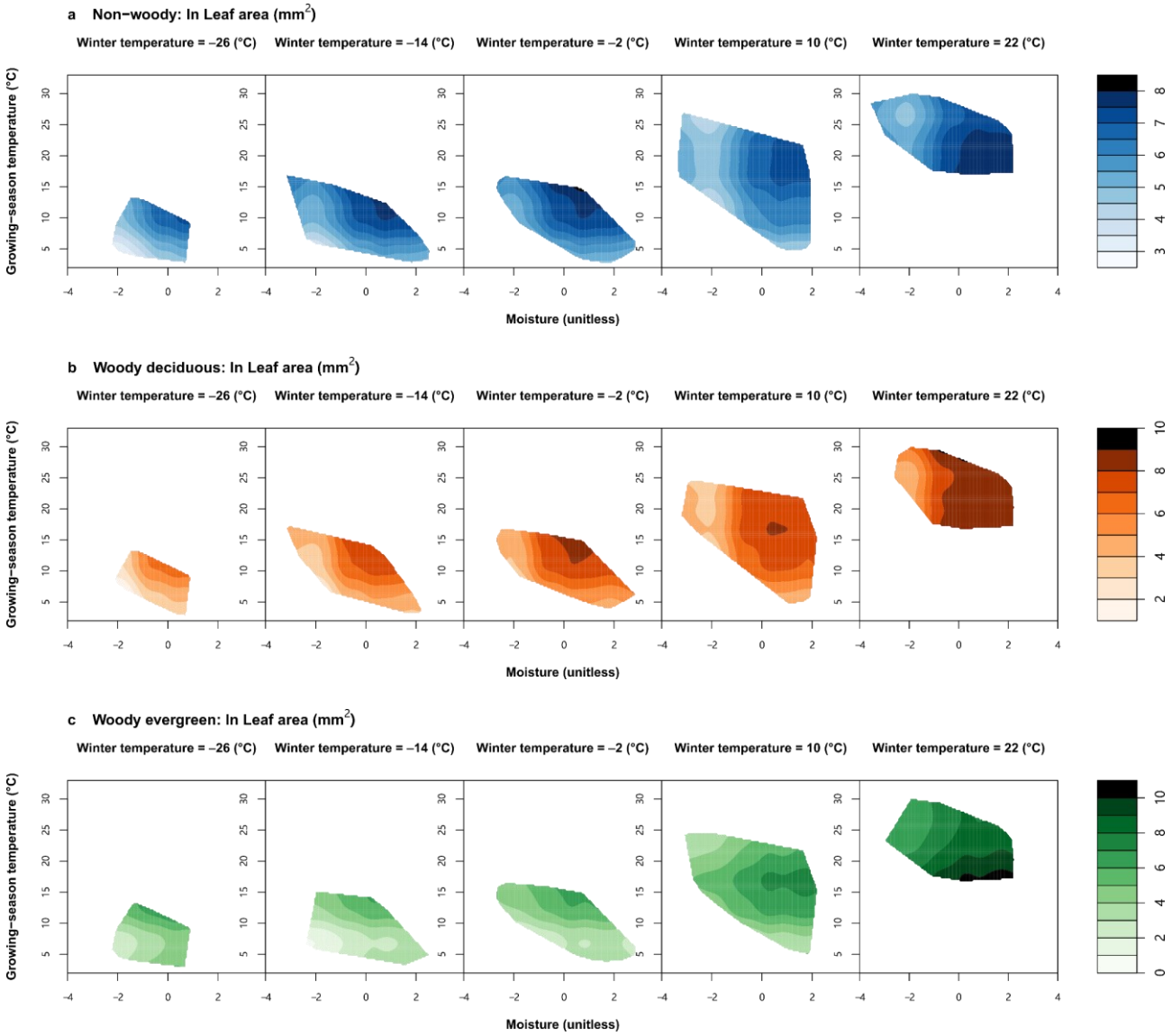
Chapter 2: Global patterns of plant functional traits and their relationships to climate

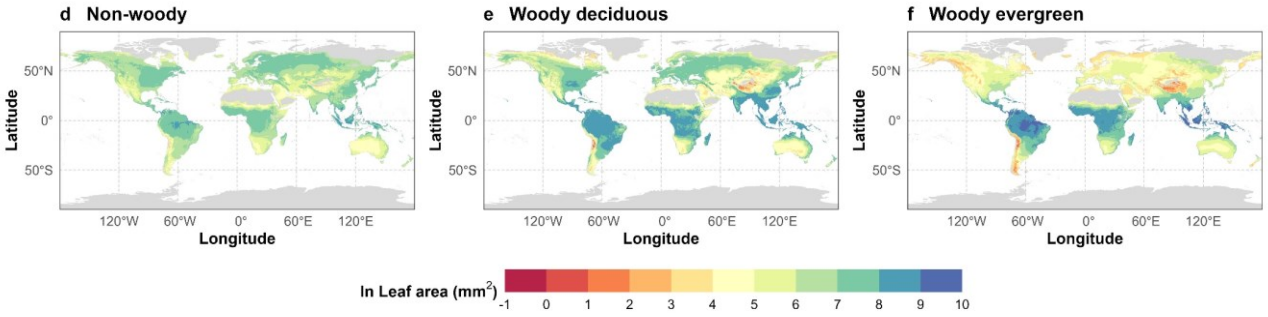
This chapter maps global patterns of key plant functional traits based on the relationships between traits and recent climate. By compiling and analysing a comprehensive global dataset of 16 functional traits for over 42,000 plant taxa, this chapter quantifies how trait-trait and trait-climate relationships differ among non-woody, woody deciduous, and woody evergreen plants, and develops high-resolution global maps for these key traits.

This chapter was conceived and developed by me and my supervisor, Iain Colin Prentice. I compiled the data, designed and carried out all analyses and visualisations.

In September 2024, an adapted version of this Chapter 2 was published in *Communications Biology*.

Fig. 1.3: Representative figure for Chapter 2 – Plant functional trait distributions in climatic and geographic spaces.





a–**c**, Distributions of natural-log transformed leaf area (mm^2) for **a**, non-woody plants; **b**, woody deciduous plants; and **c**, woody evergreen plants within a three-dimensional climatic space defined by three bioclimatic variables: moisture, winter temperature and growing-season temperature. Fitted trait values are presented as contours, with darker colours in the right colour bar representing higher trait values. Values of moisture and growing-season temperature vary continuously along the horizontal and vertical axes, respectively. Each slice is created at an exact winter temperature value. **d**–**f**, Global patterns of natural-log transformed leaf area (mm^2) predicted by Generalised Additive Models (GAMs) based on bioclimatic variables for **d**, non-woody plants; **e**, woody deciduous plants; and **f**, woody evergreen plants at a 0.1° spatial resolution.

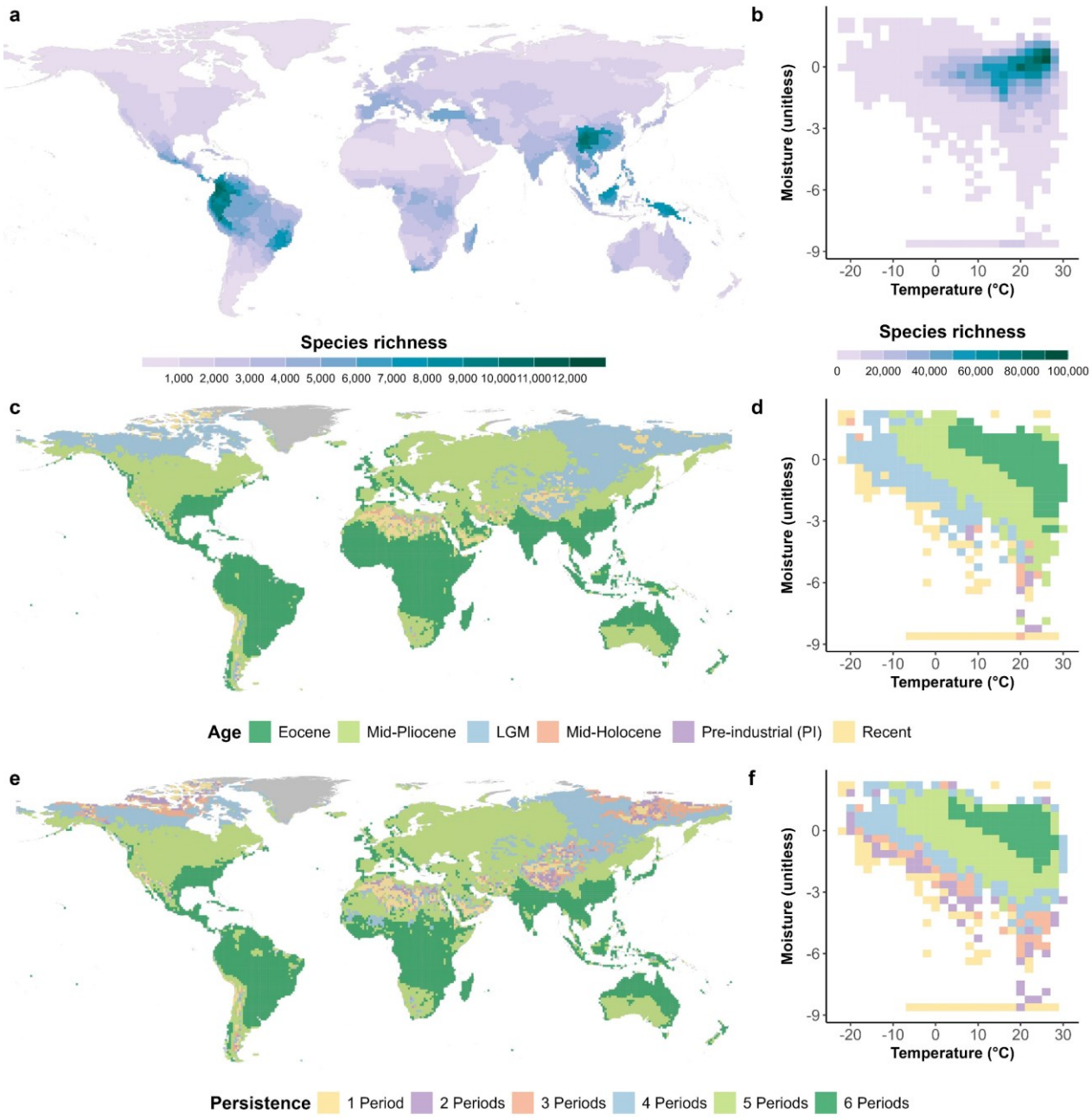
Chapter 3: Deep-time geographic dynamics of climate shape global vascular plant diversity

This chapter investigates how the deep-time geographic dynamics of climate have shaped contemporary global patterns of vascular plant diversity. By compiling occurrence records for over 350,000 species and tracing the geographic distributions of contemporary climatic conditions over the past 50 million years, this chapter quantifies the historical persistence and shifts in the geographic extent of climatic conditions and evaluates their influence on contemporary patterns of plant diversity. This analysis is enabled by palaeoclimate simulations for multiple geological periods, the major features of which are supported by fossil plant evidence. The chapter then assesses the relative contributions of these spatio-temporal climate dynamics in explaining present-day plant diversity patterns.

This chapter was conceived and developed by me and my supervisor, Iain Colin Prentice. I designed the methods, performed data collection and manipulation, carried out all analyses and visualisations.

Chapter 3 is prepared with the aim of submission to a *Nature* journal.

Fig. 1.4: Representative figure for Chapter 3 – Present-day plant diversity and the age and persistence of recent climatic conditions.



a, Global species richness of vascular plants at 1° spatial resolution, based on native distributions derived from occurrence records of 350,864 species. **b**, Species richness in a two-dimensional climatic space defined by temperature and moisture. **c**–**d**, Age of recent climatic conditions mapped

in geographic (**c**) and climatic (**d**) spaces, indicating when each climatic bin first emerged across the selected geological periods (Eocene to recent). **e**–**f**, Persistence of climates, defined as the number of periods in which a recent climatic condition was present, shown in geographic (**e**) and climatic (**f**) spaces.

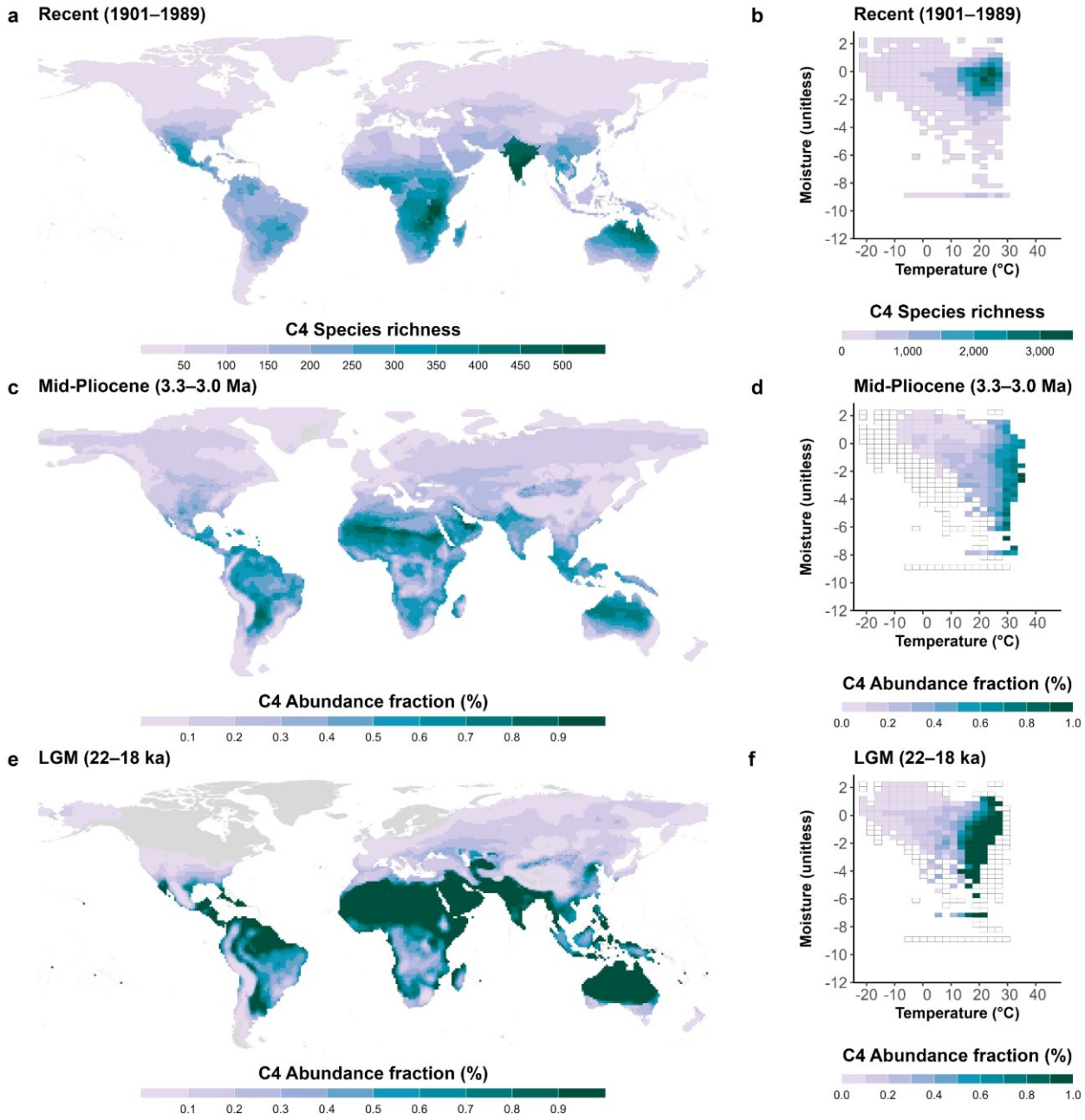
Chapter 4: Global biogeography of C4 plants through evolutionary time

This chapter simulates the global biogeography of C4 plants over the past 50 million years. By assembling a global database of C4 diversity and combining it with eco-evolutionary optimality models, this work reconstructs the global biogeographic dynamics of C4 plants across major intervals of their evolutionary history and quantifies how the relative influence of environmental drivers on their distribution and dominance has shifted over time.

This chapter was conceived and developed by me and my supervisor, Iain Colin Prentice. Boya Zhou, Ruijie Ding and Jierong Zhao provided codes of eco-evolutionary optimality models and assisted with model implementation. Rowan F. Sage contributed data resources on C4 plants. I collected and processed the data, carried out all analyses and visualisations.

Chapter 4 is prepared with the aim of submission to a *Nature* journal.

Fig. 1.5: Representative figure for Chapter 4 – Global distribution and temporal dynamics of C4 plant diversity.



a–**b**, Present-day C4 plant species richness, based on native distributions derived from occurrence records of 7,156 C4 species, shown in geographic space (**a**) and a two-dimensional climatic space (**b**) defined by temperature and moisture. **c**–**d**, C4 abundance fraction during the Mid-Pliocene (3.3–3.0 Ma) in geographic (**c**) and climatic (**d**) space, simulated using models based on eco-evolutionary optimality (EEO) theory. **e**–**f**, Modelled C4 abundance fraction during the Last Glacial Maximum (LGM, 22–18 ka) in geographic (**e**) and climatic (**f**) space.

Chapter 5: Conclusion

This chapter synthesises the findings from the preceding chapters, acknowledges the limitations of the work, outlines potential directions for future research, and discusses their broader implications for ecology, conservation, and Earth system science. I am the sole contributor to this chapter.

Chapter 2: Global patterns of plant functional traits and their relationships to climate

2.1 Abstract

Plant functional traits (FTs) determine growth, reproduction and survival strategies of plants adapted to their growth environment. Exploring global geographic patterns of FTs, their covariation and their relationships to climate are necessary steps towards better-founded predictions of how global environmental change will affect ecosystem composition. We compile an extensive global dataset for 16 FTs and characterise trait-trait and trait-climate relationships separately within non-woody, woody deciduous and woody evergreen plant groups, using multivariate analysis and Generalised Additive Models (GAMs). Among the six major FTs considered, two dominant trait dimensions – representing plant size and the leaf economics spectrum (LES) respectively – are identified within all three groups. Size traits (plant height, diaspore mass) however are generally higher in warmer climates, while LES traits (leaf mass and nitrogen per area) are higher in drier climates. Larger leaves are associated principally with warmer winters in woody evergreens, but with wetter climates in non-woody plants. GAM-simulated global patterns for all 16 FTs explain up to three-quarters of global trait variation. Global maps obtained by upscaling GAMs are broadly in agreement with iNaturalist citizen-science FT data. This analysis contributes to the foundations for global trait-based ecosystem modelling by demonstrating universal relationships between FTs and climate.

2.2 Main

As primary producers, land plants regulate carbon and nitrogen cycling in ecosystems and determine their productivity[65,79,135]. Functional traits (FTs) – morphological, physiological and phenological characteristics of individual plants[58] – determine plant growth, reproduction and survival strategies by influencing life-history processes (e.g., seed traits), resource uptake and

utilization (e.g., leaf nutrient contents), and responses to environmental change and disturbances (e.g., leaf traits and plant height)[58–60]. FTs thus reflect the diverse ecological strategies of plants, which influence plant community assembly and thereby biodiversity and function[59,136–138].

FTs have been found to covary along a limited number of dimensions[58,61]. FTs that vary together at species level might in principle share genetic controls, or depend on trade-offs representing alternative ecological strategies[61]. The leaf economics spectrum (LES) is one well-known axis of trait covariation that embodies a trade-off between resource acquisition and conservation[62]. Díaz and colleagues[63] summarised the global spectrum of plant form and function based on six major FTs that are critical for the growth, survival and reproduction of species. They identified two orthogonal dimensions of trait covariation: one corresponding to the size of whole plants and plant organs, the other to the LES[63]. These two trait dimensions are universal, existing even in ecosystems with extreme climates, such as tundra[139]; at both species and community scales[67,138]; and within plant groups, e.g. trees[140]. Different ecological strategies can facilitate the coexistence of species (and their FTs) within communities[138]. At the community level, covarying FTs tend to perform interrelated functions[65,93,141,142]. Wright *et al.*[141] observed that wood density and leaf size are negatively correlated in neotropical forests. Lower wood density related with higher water conductivity per unit sapwood area and larger leaf area associated with potentially higher photosynthetic rate but also higher evapotranspiration. These two traits are thus interconnected through the hydraulic system governing water transport and its trade-off with photosynthetic efficiency[141]. Quantifying the covariation of FTs and their controls is necessary for a full understanding of how the diversity of plants translates into community composition, productivity and adaptations to environment[61].

Although the two dimensions of trait variation identified in ref. [63] are universal, the prevalence of specific FTs varies along environmental gradients[64,65]: different combinations of FTs are selected for by environmental conditions[59,66,67], so climate change is expected to bring about shifts in species composition determined by their FTs, such as migration and extinction of needle-leaved trees in part of North America since the last glacial maximum[68]. In order to predict such shifts, it is essential to understand global trait-climate relationships[65,68]. But there is no comprehensive global map of trait measurements[70]. The models used most widely to investigate climate change effects on ecosystems

are Dynamic Global Vegetation Models (DGVMs). These typically compress functional diversity into a small (about 5–15) number of plant functional types (PFTs)[69,70], which are assigned fixed trait values. Substantial, observed variability within PFTs is thereby lost[71,72]. Moreover, as past and future climates may have no analogues in present climate conditions, it is likely that this approach misses potential trait variation and trait combinations within PFTs[71,143,144]. A better approach to estimating trait values on a global scale is needed to overcome the limitations of PFT-based modelling. Although satellite remote sensing offers potential to directly map plant traits at large scales[73–75], this approach can only retrieve a limited set of traits from leaves and upper canopies, with moderate accuracy[73]. In addition, statistical upscaling and machine learning approaches based on relationships between FTs and environmental factors have recently been used to produce global maps for a few (mainly leaf) FTs[65,72,74,76–80]. A recently published work simulated global maps of three leaf traits via optimality models based on eco-evolutionary optimality theory[81]. However, the various published global trait maps do not always show consistent global patterns, reflecting differences in data sources and upscaling methods[81,87].

Newly published data and computational resources should facilitate global-scale research on FTs. The sPlotOpen dataset[82,83], a curated subset of a large globally distributed set of vegetation plots, provides plot-level trait data (as community-weighted means, CWMs) for 18 FTs. These CWMs were calculated from trait data in the TRY database[84] and the relative abundances of each taxon within each plot[83]. Díaz *et al*.[85] recently published an enhanced species-level trait dataset containing the species-mean values of the six major FTs defining the primary axes of trait covariation in the global spectrum of plant form and function as defined in ref. [63]. Global trait patterns have also been mapped by complementing plant observations from the global citizen science project iNaturalist with measurements from TRY[86].

A study of the geographical patterns of six FTs across North and South America found that trait-climate relationships differ overall between woody and herbaceous plants, including different climate predictors and different response shapes of FTs along climatic gradients[88]. Previous studies have also found that woody evergreen plants tend to have thicker and longer-lived leaves poleward, while woody deciduous plants show the opposite pattern[89,90]. A recently developed theory explains this divergence of latitudinal trends[91]. According to the theory, leaves maximize the average net

carbon gain over their life cycle. The different relationships between leaf mass per area (LMA) and environmental factors in deciduous and evergreen woody plants then arise because the life cycle of deciduous leaves, unlike that of evergreen leaves, is tied to the annual cycle[91]. It is therefore appropriate to analyse non-woody plants, woody deciduous plants and woody evergreen plants as separate groups, with potentially distinct trait-environment relationships.

We compiled a global dataset for 16 FTs (Table 2.1), including 42,676 plant taxa from 77,074 natural vegetation plots based on the sPlotOpen dataset[83] and the enhanced species-level trait dataset of ref. [85]. All FTs in our dataset have plot-level means for each plot; six also have species-level means. We refer to these six as 'major FTs' as they also form the global spectrum of plant form and function. Plant taxa were grouped into non-woody, woody deciduous and woody evergreen categories according to their life form and leaf phenology. We analysed trait covariation and characterised trait-climate correlations within each category for the six major FTs. For all FTs, we fitted Generalised Additive Models (GAMs) for their relationships to three bioclimatic variables. Then we simulated global patterns for each FT at 0.1° spatial resolution, based on climate data and remotely-sensed global fractional cover of the three plant groups. Finally, we compared our global maps for the 16 FTs against the iNaturalist maps[86] at 2° spatial resolution.

Table 2.1 | Plant functional traits used in this study.

Trait[a]	Abbreviation[b]	Unit[c]	Functions[d]	Availability[e]
Leaf area	LA	mm^2	Leaf size; related to leaf energy and water balance[61]	Plot-level trait means[83] and species-level trait means[85]
Leaf mass per unit area	LMA	kg/m^2	Key traits in the leaf economic spectrum (LES); reflecting investment strategy[62]	
Leaf nitrogen content per unit area	N_{area}	g/m^2		
Stem specific density	SSD	g/cm^3	Trait in wood economics spectrum; reflecting drought-tolerance[145]	
Plant height	H	m	Whole plant size; reflecting ability to compete for light[63,146]	
Diaspore mass	DM	mg	Seed size; reflecting dispersal and establishment ability[63,147]	
Leaf fresh mass	LFM	g	Related to energy transfer, plant growth and productivity[138]	Plot-level trait means[83]
Leaf phosphorus content per unit area	P_{area}	g/m^2		
Leaf carbon content	C_{mass}	mg/g		
Leaf dry matter content	LDMC	g/g		
Stem conduit (vessel and tracheid) element length	WVL	µm	Related to water transport: safety versus efficiency[138]	
Stem conduit density	SCD	mm^{-2}		
Seed number per reproductive unit	SN	NA	Reflecting dispersal, regeneration and reproduction ability[138]. Here, SL is the length of a seed, while DUL is the length of the dispersal unit,	
Seed length	SL	mm		
Dispersal unit length	DUL	mm		

			which may include many seeds.
Leaf nitrogen isotope ratio	$\delta^{15}N$	per million	Reflecting nitrogen source, loss pathways, and root symbionts[138,148]

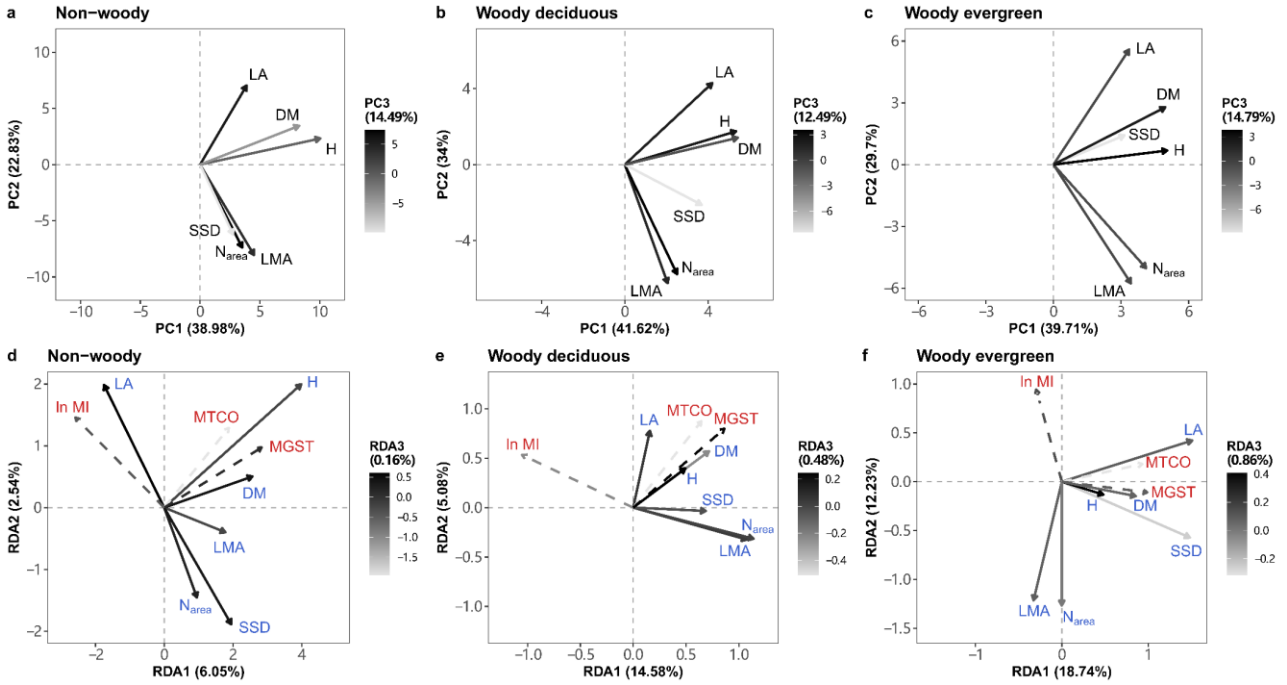
[a] Plant functional traits used in this study. [b] Abbreviation and [c] unit of each of the 16 traits. [d] Brief description of functions for each of the 16 traits. [e] Data availability for each trait. All 16 plant functional traits have plot-level trait means (community-weighted means, CWMs)[83], while the first six traits also have species-level mean trait values[85].

2.3 Results

2.3.1 Trait combinations

We calculated community-weighted means (CWMs) of each of the three plant groups within each vegetation plot for the six major FTs. CWMs were natural-log transformed prior to further analysis. We then used the CWMs to evaluate covariation among FTs within the non-woody, woody deciduous and woody evergreen groups. Patterns of covariation were summarised by Principal Component Analysis (PCA). The first three PCA axes together captured 76% of trait variation in non-woody taxa, 88% in woody deciduous taxa and 84% in woody evergreen taxa (Fig. 2.1a–c, Table A1.6). Two orthogonal dimensions emerged for all groups. One dimension, represented by height (H) and diaspore mass (DM), represents overall plant size – from short plants with light diaspores to tall plants with heavy diaspores. The other, represented by LMA and leaf nitrogen per unit area (N_{area}), reflects the LES[62] and the predominant influence of LMA on N_{area}[149,150]. Leaf area (LA) tended to covary with plant size, while being approximately orthogonal to the LES. Stem specific density (SSD) showed the opposite direction to the other traits along the axis 3 of PCA. SSD in non-woody plants was strongly correlated with LMA and N_{area} (Fig. 2.1a). SSD in woody evergreen plants was more correlated with plant size (Fig. 2.1c), while SSD in woody deciduous plants showed an intermediate pattern (Fig. 2.1b).

Fig. 2.1: Trait-trait and trait-climate correlations.



Principal component analysis (PCA) (**a**–**c**) and redundancy analysis (RDA) (**d**–**f**) for six major plant functional traits of non-woody (**a**, **d**), woody deciduous (**b**, **e**) and woody evergreen (**c**, **f**) plants. Six major traits are leaf area (LA, mm^2), leaf mass per unit area (LMA, kg/m^2), leaf nitrogen content per unit area (N_{area}, g/m^2), stem specific density (SSD, g/cm^3), plant height (H, m) and diaspore mass (DM, mg). The orientation of axes has been rotated according to the Figure 2 in ref. [63]. Gray scales indicate the loadings on the third axis. Solid arrows (with blue labels in RDA) represent traits and dashed arrows (with red labels in RDA) represent bioclimatic variables. All six traits were natural-log transformed before analysis. Both log-transformed traits and bioclimatic variables were rescaled to a mean of 0 and a standard deviation of 1 before analysis. ln MI, log-transformed moisture index; MTCO, mean temperature of the coldest month; MGST, mean growing-season temperature (see Methods for definition). All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[151].

2.3.2 Relationships between traits and climate

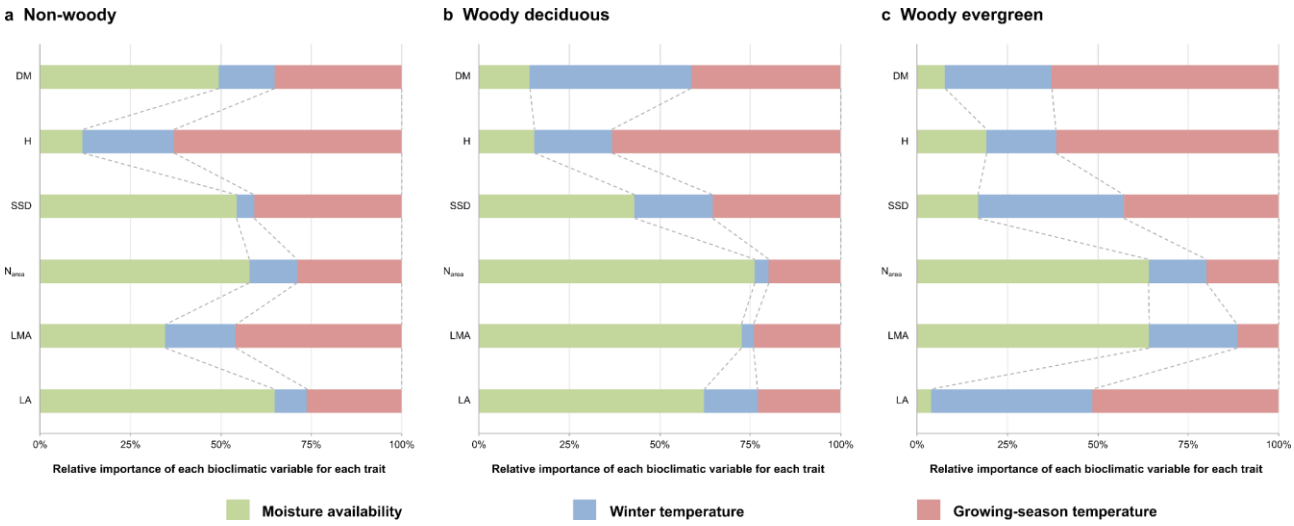
We calculated three bioclimatic variables for each plot (Fig. A1.2), which represent the three major independent climatic controls on the global geographic distribution of vegetation

physiognomy[11,14,152] and therefore likely also the global distribution of key plant traits[152] (see Methods). The three bioclimatic variables are: mean growing-season temperature (MGST, °C, representing summer warmth), mean temperature of the coldest month (MTCO, °C, representing winter cold), and the log-transformed moisture index (ln MI, unitless, representing plant-available moisture: see Methods). We subsequently conducted a Redundancy Analysis (RDA) to describe the extent to which the six major FTs covary with these bioclimatic variables. The first three constrained RDA axes accounted for 9%, 20% and 32% of total trait variation in non-woody, woody deciduous and woody evergreen plants, respectively (Fig. 2.1d–f, Table A1.7). These three axes revealed substantially the same combinations of plant size and LES traits as were shown in the PCA, but further illustrated their relationships with bioclimate. Along the three axes (Fig. 2.1d–f), plant height and diaspore mass increased with the growing-season temperature (MGST) within all groups, and LMA and N_{area} decreased with increasing moisture (ln MI) within all groups. Plant size traits was positively correlated with winter cold (MTCO) on the first two axes, but they were negatively correlated on the third axis. However, LA behaved differently, covarying with ln MI in non-woody plants but with temperature variables in woody plants. The residual or unconstrained RDA axes (Fig. A1.6) approximately replicate the PCA axes, indicating that much trait variation along the main axes identified in ref. [63] persists even after climatic influences have been taken into account. But the strong correlation between SSD and plant size traits in woody evergreen plants disappeared (Fig. A1.6c).

We used Generalised Additive Models (GAMs) to quantify relationships between each major FT and bioclimatic variables in more detail, which allowed us to fit more complex surfaces and to distinguish growing-season from cold-month temperature effects (Figs. 2.2–2.3). We calculated relative importance values of explanatory variables (Fig. 2.2) in GAMs to quantify the individual contribution of three bioclimatic variables to the major FTs. These values measure the average percentage contribution of each variable in turn to the fit of the models based on all three variables. Across all plant groups H variation was dominated by growing-season temperature, and LMA and N_{area} variation by moisture availability (with an additional influence of growing-season temperature on LMA in non-woody plants). DM variation was dominated by moisture availability in non-woody plants but by growing-season temperature (with an additional influence of winter temperature) in

woody deciduous plants. By contrast, variation in SSD and LA showed similar controls in non-woody and woody deciduous plants (dominated by moisture availability and growing-season temperature for SSD and moisture availability for LA), but a different type of control – dominated by growing-season and winter temperatures – for woody evergreen plants.

Fig. 2.2: Relative importance of three bioclimatic variables.



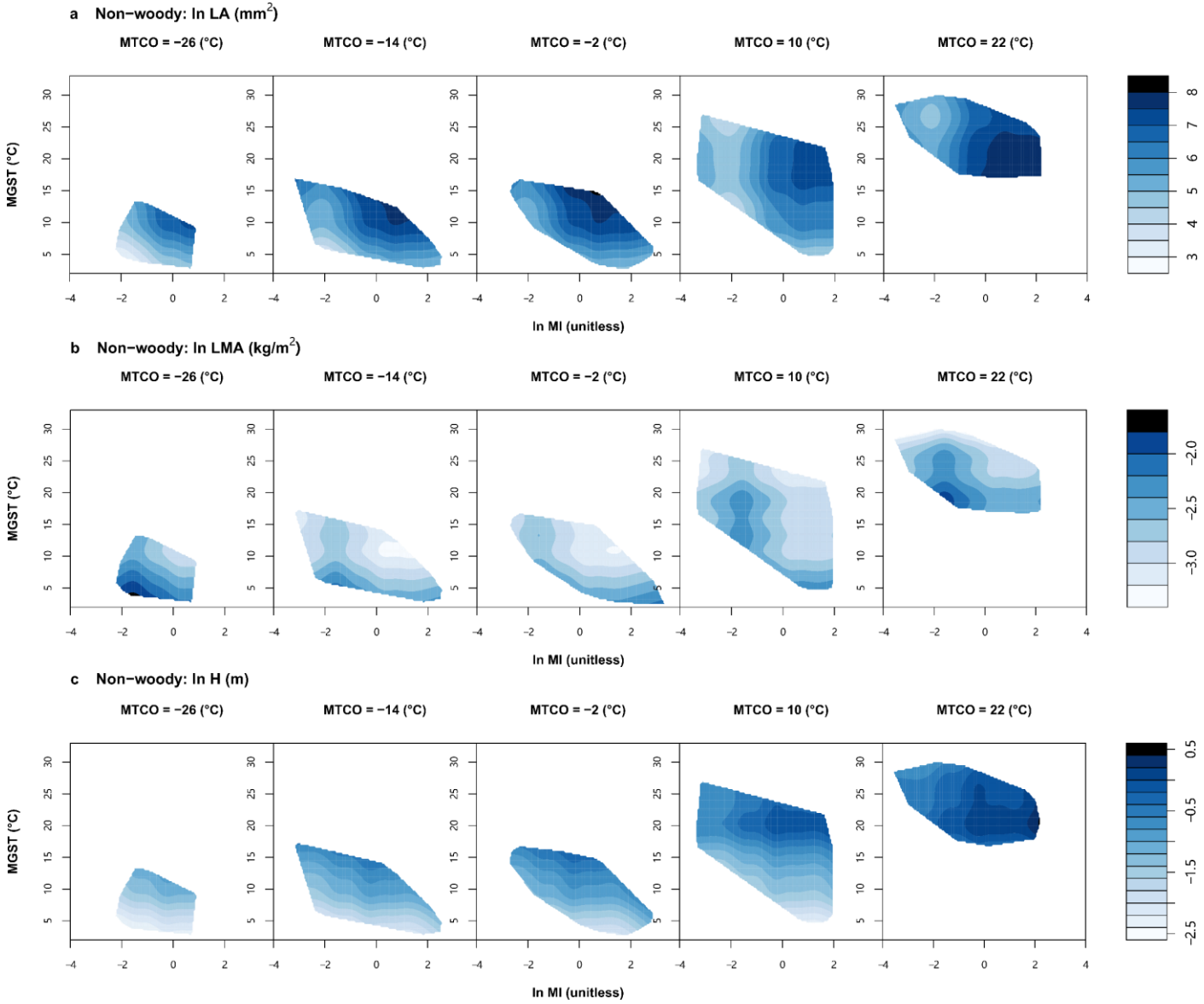
Relative importance of each bioclimatic variable in predicting six major plant functional traits of non-woody (**a**), woody deciduous (**b**) and woody evergreen (**c**) plants derived from Generalised Additive Models (GAMs). Six major traits are leaf area (LA, mm^2), leaf mass per unit area (LMA, kg/m^2), leaf nitrogen content per unit area (N_{area}, g/m^2), stem specific density (SSD, g/cm^3), plant height (H, m) and diaspore mass (DM, mg). All six traits were natural-log transformed before the analysis. Moisture availability, log-transformed moisture index (ln MI, unitless); winter temperature, mean temperature of the coldest month (MTCO, °C); growing-season temperature, mean growing-season temperature (MGST, °C) (see Methods for definition). All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[151].

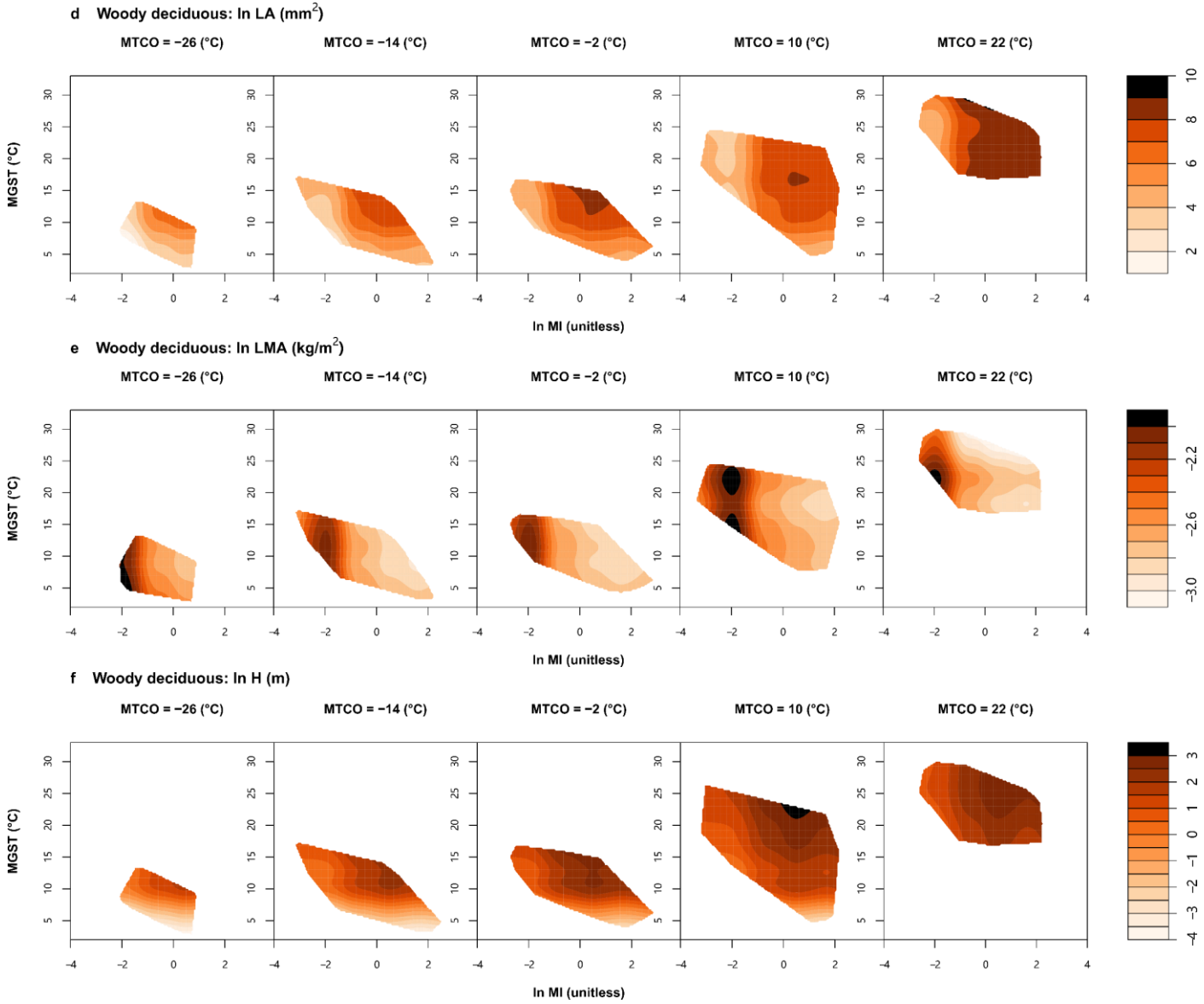
We visualised the trait distributions in the global climate space defined by three bioclimatic variables for each of the six major FTs (Figs. 2.3 and A1.8). This allowed for a more intuitive comparison of trait-climate relationships among the FTs and three plant groups. FTs related to the same principal axis of trait variation presented similar distribution patterns in the global climate

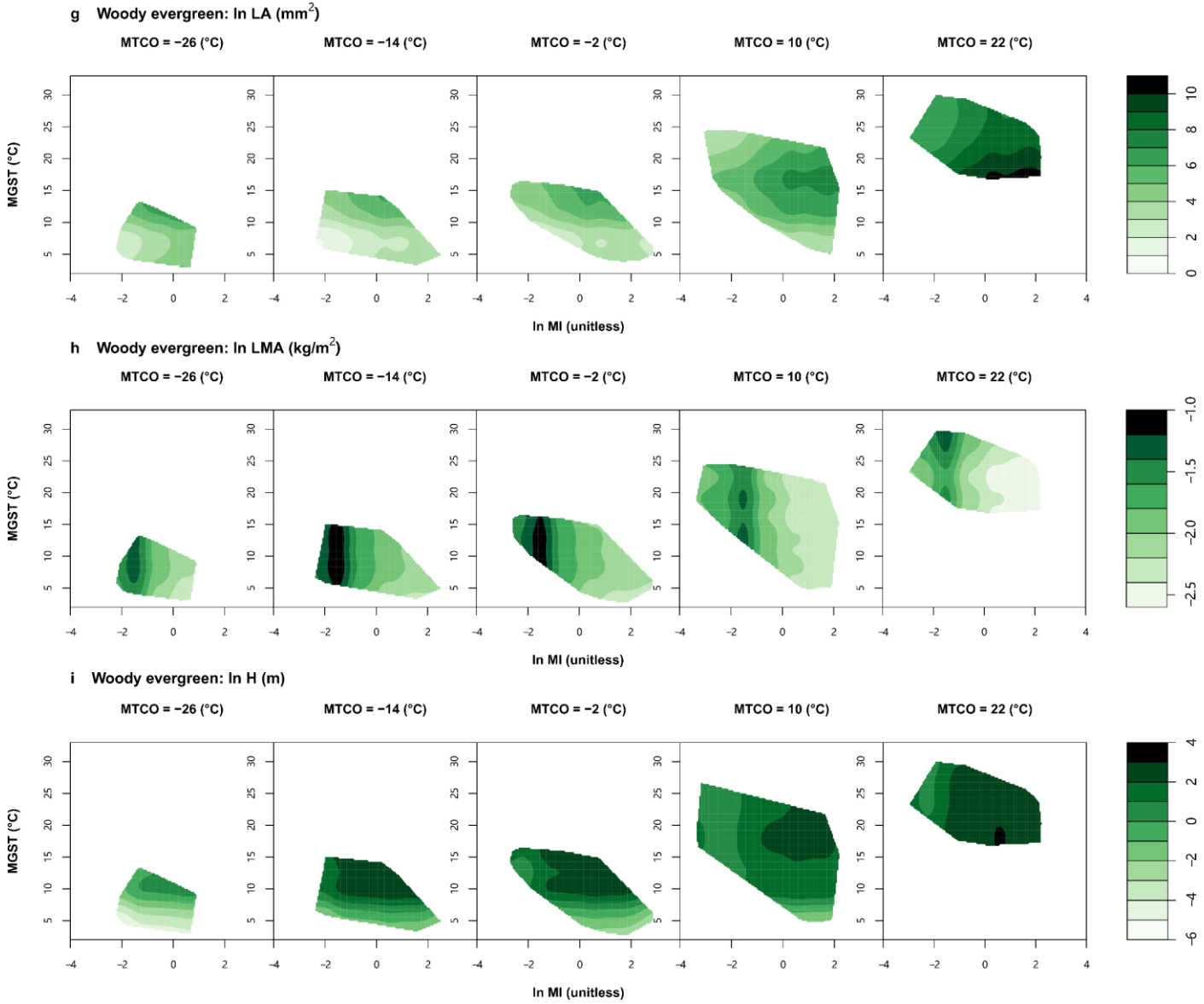
space. However, the GAMs reveal many more subtle distinctions. For example, among FTs related to the LES, Fig. 2.3b, e and h shows that LMA in all plant groups increases towards drier climates – but LMA in non-woody plants tends to decrease with growing-season temperature in climates with cold winters (MTCO ≤ –2 °C in Fig. 2.3b). LMA in woody plants is less sensitive to temperature (Fig. 2.3e and h); however woody deciduous plants show highest LMA in climates with warmer winters (MTCO ≥ 10 °C in Fig. 2.3e), while woody evergreen plants show highest LMA in climates with colder winters (MTCO ≤ –2 °C in Fig. 2.3h). Among FTs representing plant size (Fig. 2.3c, f and i), H is greatest in climates that are both wet (ln MI > 1) and warm (MGST > 15 ˚C, MTCO ≥ 10 °C) within all groups.

LA showed a general pattern of increase with moisture within all three plant groups (Fig. 2.3a, d and g), but LA was less sensitive to moisture in woody plants (Fig. 2.3d and g) compared to non-woody plants (Fig. 2.3a; also confirmed by Fig. 2.1d–f). Among woody evergreens (Fig. 2.3g), the largest leaves are found in wet climates with warm winters (MTCO > 10 ˚C) as well as warm summers (MGST > 15 ˚C). Consistent with their habit, woody deciduous (Fig. 2.3d) and non-woody plants (Fig. 2.3a) do not show the same strong sensitivity of LA to cold winters.

Fig. 2.3: Examples of climate space diagrams.


a Non−woody: ln LA (mm^2)
MTCO = −26 (°C)
MTCO = −14 (°C)
MTCO = −2 (°C)
MTCO = 10 (°C)
MTCO = 22 (°C)
MGST (°C)
ln MI (unitless)
b Non−woody: ln LMA (kg/m^2)
MTCO = −26 (°C)
MTCO = −14 (°C)
MTCO = −2 (°C)
MTCO = 10 (°C)
MTCO = 22 (°C)
MGST (°C)
ln MI (unitless)
c Non−woody: ln H (m)
MTCO = −26 (°C)
MTCO = −14 (°C)
MTCO = −2 (°C)
MTCO = 10 (°C)
MTCO = 22 (°C)
MGST (°C)
ln MI (unitless)


d Woody deciduous: ln LA (mm^2)
MTCO = −26 (°C)
MTCO = −14 (°C)
MTCO = −2 (°C)
MTCO = 10 (°C)
MTCO = 22 (°C)
MGST (°C)
ln MI (unitless)
e Woody deciduous: ln LMA (kg/m^2)
f Woody deciduous: ln H (m)


g Woody evergreen: ln LA (mm^2)
MTCO = −26 (°C)
MTCO = −14 (°C)
MTCO = −2 (°C)
MTCO = 10 (°C)
MTCO = 22 (°C)
MGST (°C)
ln MI (unitless)
h Woody evergreen: ln LMA (kg/m^2)
i Woody evergreen: ln H (m)

Climate space diagrams of non-woody (**a**–**c**), woody deciduous (**d**–**f**) and woody evergreen (**g**–**i**) plants: showing distributions of natural-log transformed leaf area (LA, mm^2) (**a**, **d** and **g**), leaf mass per unit area (LMA, kg/m^2) (**b**, **e** and **h**) and plant height (H, m) (**c**, **f** and **i**) in the global climate space defined by three bioclimatic variables. Fitted trait values are presented as contours, with darker colours in the right colour bar representing higher trait values. Values of ln MI and MGST vary continuously along the horizontal and vertical axes, respectively. Each slice is created at an exact MTCO value. Abbreviations and units of traits are shown in Table 2.1. ln MI, log-transformed moisture index; MTCO, mean temperature of the coldest month; MGST, mean growing-season temperature (see Methods for definition). Climate space diagrams for all six major FTs are shown in Fig. A1.8. All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[151].

2.3.3 Global maps for the six major traits

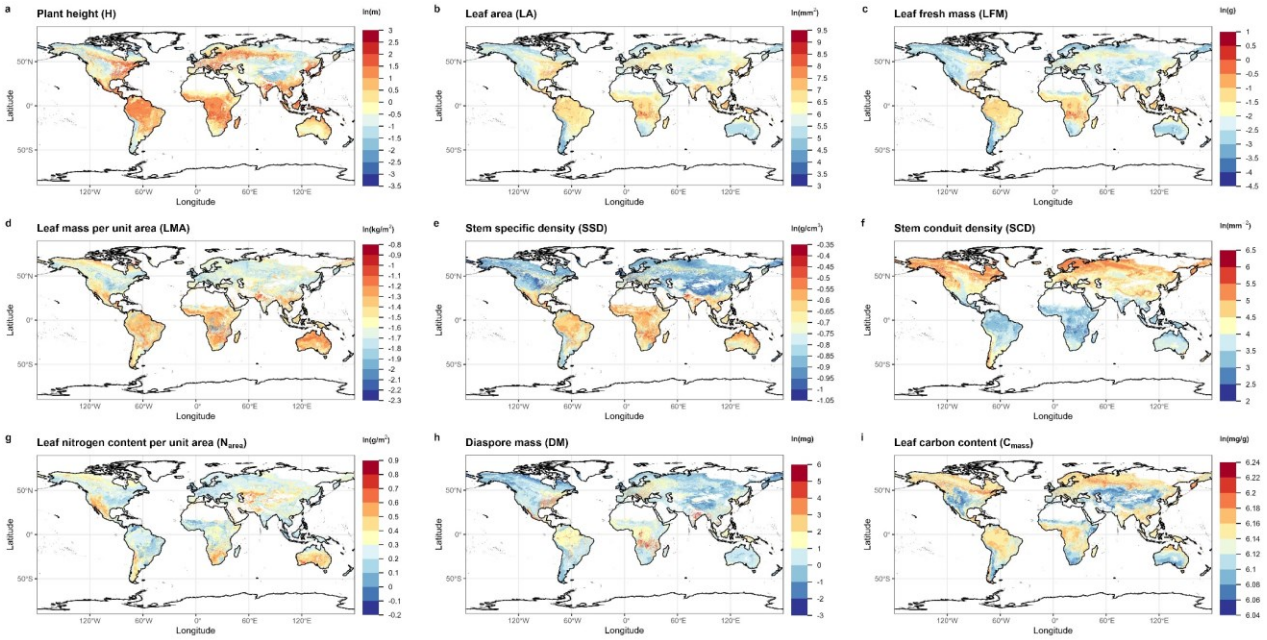
For each of the three plant groups, we upscaled the GAMs based on three bioclimatic variables to generate global distribution maps for the six major FTs at 0.1° spatial resolution, separately. The separate trait maps for each plant group are shown in Fig. A1.9.

2.3.4 Global maps for all traits

In order to predict global patterns for all 16 FTs having plot-level mean values (CWMs) in the sPlotOpen dataset[83], we fitted new GAMs based on the same three bioclimatic variables, combined with the (remotely-sensed) fractional cover of the three plant groups as additional predictors. All FTs were natural-log transformed prior to the analyses. These GAMs explained up to 77% of the global variation of plot-level mean trait values, with a median adjusted R^2 of 52% (average median calculated from 54% and 49%) over all 16 FTs. Best fits ($R_{adj}^2 \geq 49\%$) were obtained for the following nine FTs: H (77%), SSD (68%), DM (64%), SCD (61%), LMA (57%), LA (56%), LFM (55%), C_{mass} (54%) and N_{area} (49%) (see Table A1.11 for a summary). We created global trait maps for all 16 FTs based on these GAMs at 0.1° spatial resolution (Fig. 2.4; see Fig. A1.12 for all 16 FTs). H, LA and LFM (Fig. 2.4a–c) were predicted to be higher in tropical regions and warm and humid environments, with lower values found in cold high-latitude regions and dry temperate environments. The global distribution of LMA (Fig. 2.4d) might be related to the joint effects of

global patterns of LA (Fig. 2.4b) and LFM (Fig. 2.4c). Plants with higher LMA were predicted to inhabit tropical regions and drier environments, while plants with lower LMA mainly appeared in humid northern temperate regions. Tropical zones subject to high moisture stress harbour species with higher SSD but lower SCD (Fig. 2.4e–f). However, plants in humid northern temperate regions and boreal forests had relatively higher values of SCD and lower SSD (Fig. 2.4e–f). Although FTs within the same trait dimension did not show identical global patterns, their maximum values tended to occur under similar environmental conditions. Like high-LMA plants (Fig. 2.4d), plants with higher N_{area} (Fig. 2.4g) were generally favoured in hot and dry environments, supporting earlier findings[153–155]. DM (Fig. 2.4h) was predicted to be notably high in equatorial regions and warm and humid temperate regions with hot summers, coinciding with H (Fig. 2.4a). C_{mass} showed little variation across the globe (Fig. 2.4i). It was found to be slightly lower in deserts and arid temperate regions.

Fig. 2.4: Examples of global trait maps.



Global patterns of plant height (**a**), leaf area (**b**), leaf fresh mass (**c**), leaf mass per unit area (**d**), stem specific density (**e**), stem conduit density (**f**), leaf nitrogen content per unit area (**g**), diaspore mass (**h**) and leaf carbon content (**i**), produced by the best-fit Generalised Additive Models (GAMs, $R^2 \geq 49\%$) based on three bioclimatic variables combined with remotely-sensed fractional cover

data for the three plant groups (see Methods for details). The best-fit model here refers to the model with adjusted R^2 above the median level of the adjusted R^2 over all 16 FTs. Colour gradients from blue to red indicate the increase in trait values. All traits are natural-log transformed. All maps are at 0.1° resolution. Global trait maps of all 16 functional traits are shown in Fig. A1.12. For maps in GeoTiff format, refer to the Data availability statement. All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[151].

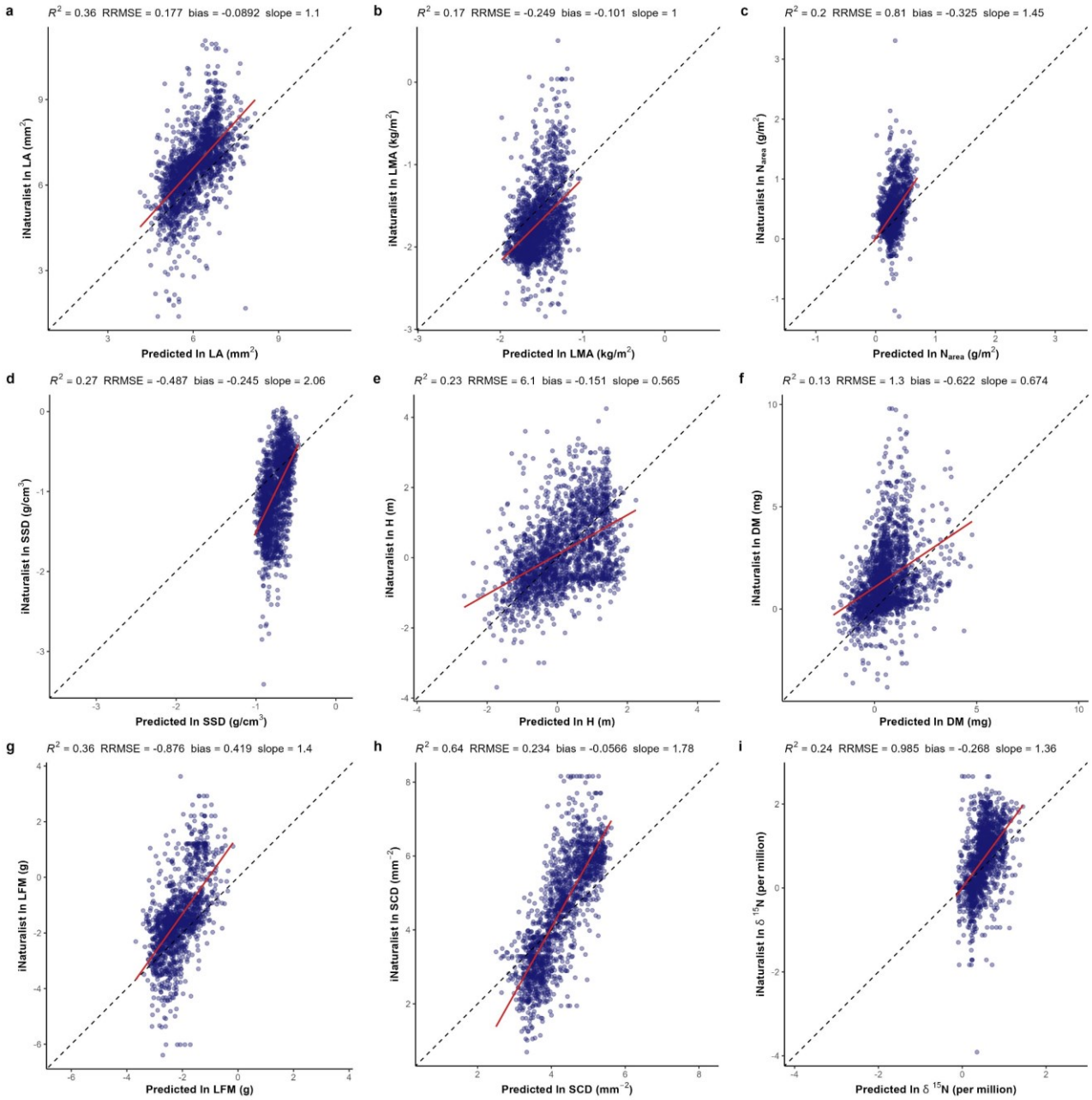
2.3.5 Comparison to iNaturalist data

Independent global trait maps[86] generated by linking plant observations from the iNaturalist citizen-science project to trait measurements from the TRY database were used as a benchmark for our GAM-based predictions, as the iNaturalist data included all 16 FTs considered in our study. We evaluated the pixel-by-pixel agreement between our GAM predicted maps with the iNaturalist maps at 2° resolution, chosen because some FTs presented the highest correlations between sPlotOpen and iNaturalist at this spatial scale[86].

The highest R^2 and median R^2 of linear regressions between our GAM-derived maps and the iNaturalist maps, over all 16 FTs, were 64% and 15%. Fig. 2.5 shows nine FTs that present better agreement than the remaining FTs ($R^2 \geq 13\%$ and slope > 0.5 and < 2): SCD, LA, LFM, SSD, $\delta^{15}N$, H, N_{area}, LMA and DM (see Table A1.13 for more information).

The slope of the linear regression also indicates the direction in which the GAM predicted values deviate from the iNaturalist gridded values. According to Fig. 2.5, GAMs tended to estimate higher LMA and lower LA and N_{area} in comparison to the iNaturalist data. GAM tended to overestimate SSD at lower values and underestimate SSD at higher values compared with the iNaturalist data. For larger values, GAMs predicted higher H and DM but predicted lower LFM, SCD and $\delta^{15}N$ than the iNaturalist data.

Fig. 2.5: Comparison between predicted trait values and the iNaturalist estimated trait values.



a, leaf area (LA); **b**, leaf mass per unit area (LMA); **c**, leaf nitrogen content per unit area (N_{area}); **d**, stem specific density (SSD); **e**, plant height (H); **f**, diaspore mass (DM); **g**, leaf fresh mass (LFM); **h**, stem conduit density (SCD); **i**, leaf nitrogen isotope ratio ($\delta^{15}N$). Red lines visualise the linear regression between model-predicted map pixel values and the iNaturalist map pixel values estimated at a 2° spatial resolution. The dotted line is the 1:1 line. R^2 is the coefficient of determination; RRMSE is the root-mean-square error expressed as a proportion of the observed

mean trait value (here observations are represented by iNaturalist trait values); bias is the difference between observed and model-predicted mean values, as a proportion of the observed mean trait value (here observed values are iNaturalist trait values); slope is the slope of the linear regression of iNaturalist estimations against model predictions. The nine traits displayed here showed higher agreement between model-predicted maps and the iNaturalist maps compared to the remaining traits (R^2 ≥ 13% and slope > 0.5 and < 2). All traits are natural-log transformed. Scatter plots of model-predicted trait values versus iNaturalist estimated trait values for all 16 traits are shown in Fig. A1.13 and parameters for evaluating linear regressions for all 16 traits are shown in Table A1.13. Abbreviations and units of traits are shown in Table 2.1. All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[151].

2.4 Discussion

2.4.1 Ecological strategies reflected by trait-trait and trait-climate relationships

Analysing six major FTs in non-woody, woody deciduous and woody evergreen plants separately, we found that the two orthogonal dimensions of covariation in plant size traits (H–DM) and LES traits (LMA–N_{area}), first identified by Díaz and colleagues[63], also apply within each of the three plant groups (Fig. 2.1a–c). Although Díaz *et al*.[63] showed that non-woody and woody plants form two almost disjunct hotspots along the H–DM axis[88], our study shows nonetheless that the covariation of H and DM follows almost the same pattern within the non-woody and woody taxon groups (Fig. 2.1a–c). FTs from the same dimensions of trait covariation tended to be influenced by climate in similar ways (Figs. 2.1–2.3 and A1.8). H relates to light competition and DM to dispersal[146,147]. In resource-rich environments (wetter, warmer and more productive sites), plants make more carbon profit that can be invested in stems and also used for seed production. Consequently, they can grow taller to have greater access to light, and produce larger seeds, promoting both a strong competitive status and a high survival probability[65,146,147,156,157]. LMA and N_{area} represents resource acquisition strategies[62,63]. LMA reflects a trade-off between the amount of carbon invested in leaves and their lifespan[158]. N_{area} has been considered as the sum of a metabolic component and a structural component, the former proportional to photosynthetic capacity, the

latter to LMA[149,150]. Leaves with high LMA and N_{area} are favoured in hot and dry climates because this combination can conserve water to allow the plants to maintain a relatively high photosynthetic rate in the face of increased evapotranspiration[153–155].

Our analysis also revealed a differentiation in trait combinations among the three plant groups, specifically in the covariation of SSD with other FTs. SSD in non-woody plants was strongly correlated with LES traits (Fig. 2.1a). SSD in woody deciduous plants was relatively independent (Fig. 2.1b), while SSD in woody evergreen plants was more correlated with plant size traits (Fig. 2.1c). This novel finding is explicable in adaptive terms. In dry climates, some herbaceous and succulent plants (typically with low SSD) can fix CO_2 though their green stems, reducing water loss from the leaves and providing an extra source of carbon[159–161]. The variation of SSD did not show stronger correlation with a single of the two major trait dimensions in the global spectrum of plant form and function[63]. Previous studies were either conducted on global vascular plants[63] or analysed trait data at the community level by using community-weighted means[67,138]. These studies did not identify the functional differences among the non-woody, woody deciduous and woody evergreen plants, consequently failing to detect the distinctive trait combinations associated with SSD. We also extracted pixel values from the GAM-based global maps of the six major traits in each plant group and performed PCA and RDA on them to test these observed trait combinations were still present in the predicted trait values (Fig. A1.16). The distinct patterns of SSD among the three plant groups were not evident in the predicted trait values – particularly in non-woody plants, where the previously strong correlation between SSD and LES traits was now absent (Fig. A1.16a). Trait data used in original PCA and RDA were calculated from the database of Díaz *et al.*[85] (see Methods) in which observed records for SSD are available only for a few non-woody species; missing values of SSD were imputed via leaf dry matter content (LDMC)[85], which was found to be closely related to both LMA and N_{area}[138]: perhaps accounting for the strong correlation between (LDMC-derived) SSD and the LES traits.

Correlations between the six major FTs and the three bioclimatic variables varied among non-woody, woody deciduous and woody evergreen plants. The relative importances of predictors were different among the groups, as were patterns of trait distribution in climate space (Figs. 2.1–2.3 and A1.8). Bioclimatic variables explained more trait variance for woody than non-woody

plants (Fig. 2.1d–f, Table A1.7). One explanation is that woody plants have permanent organs that must endure climatic conditions year-round over the years. In addition, short-stature non-woody plants are closer to the ground and experience different physical conditions compared with tall-statured woody plants[88]; therefore, the microclimate of non-woody plants in the understorey cannot be directly represented by macroclimate variables[88].

Different trait-climate correlations among the three plant groups are associated with their life history strategies. For non-woody plants, LMA was less constrained by plant-available moisture (represented by ln MI) than woody plants (Fig. 2.2), and their DM showed little correlation with all three bioclimatic variables (Figs. 2.2–2.3 and A1.8). Such discrepancies may arise because non-woody species tend to occupy smaller, more specialized and diverse niches than woody species[88,162]. Once climatic conditions are outside their tolerance, they undergo dieback[163]. Therefore, they must grow quickly and reproduce efficiently within the favourable portion of a short lifespan. Such a strategy precludes tall stature[164], supports lower-LMA leaves (thus lower leaf N_{area}) even in drier climates in order to maintain high photosynthetic and respiration rates[163], and supports the production of copious small seeds to maximise fitness under high resource availability[157,165,166]. The short lifespan of many non-woody plants also represents an adaptive strategy that facilitates their survival in demanding environments. They can live their short lives between fires, physical disturbances, droughts, or annual freezes[164].

Tall woody plants, on the other hand, are successful competitors for light, persisting across years in changing environments[164,165]. However, deciduous and evergreen woody plants also adopt different ecological strategies. We found that LES traits (LMA–N_{area}) of woody deciduous plants were insensitive to growing-season temperature (MGST) but were strongly affected by plant-available moisture (ln MI) and winter temperature (MTCO) (Figs. 2.3 and A1.8). This distinction corresponds to the fact that deciduous plants shed their leaves in response to dry or cold conditions[167,168]. As a result, the leaf life cycle of woody deciduous species is tied to the growing season[91]; their lower LMA (and thus lower N_{area}) represents an adaptation to a short lifespan[88,91]. In contrast, woody evergreen plants, which need to maintain leaves and vascular networks over successive seasons[164], typically adopt more conservative nutrient-use strategies (higher LMA and N_{area}) and may require greater protection against drought and/or frost damage[62,67,139].

LA in woody plants was positively correlated with winter temperature (MTCO) (Fig. 2.3), but the relative importance of winter cold for LA in evergreen plants was higher than that in deciduous plants (Fig. 2.2). Evergreen woody plants with small leaves (low LA) are better adapted to harsh winters. Larger leaves are more prone to frost damage because of their thicker insulating boundary layers of still air that slow sensible heat exchanges with their surroundings[169]. This means that their ability of heat uptake from their surroundings to compensate for the energy lost to the night-time sky is diminished, leaving them vulnerable to frost[169]. Although all three plant groups showed the highest SSD at the same moisture level, the SSD in evergreen plants was higher (Fig. A1.8) and was less controlled by moisture (Fig. 2.2) than that of the other two groups. SSD reflects wood density and is related to drought tolerance of plants[145]. Evergreen trees with higher wood density are less vulnerable to drought-induced embolism than co-existing drought-deciduous trees with lower wood density in arid environments[170]. The hydraulic architecture of high-SSD evergreen woody plants, including fibres and thick-walled vessels, allows them to maintain water transport to the canopy under greater xylem pressure caused by water shortage and to continue growing in seasonally dry climates[145,170].

2.4.2 Evaluations of global trait maps

Mindful of potential differences in trait-climate correlations among non-woody, woody deciduous and woody evergreen plants, as shown for the six major FTs, we fitted a set of GAMs that included fractional abundances of these groups to simulate global patterns of all 16 FTs. These GAMs predicted up to 77% of variation in our global data sample correctly (Table A1.11). We evaluated the agreement of our GAM predicted maps with the iNaturalist maps[86]. The iNaturalist[86] data represents a large and independent trait dataset based on citizen-science plant observations and trait measurements from the TRY database[84]. The data included all 16 FTs considered in our study. Previously published global trait maps generated by statistical modelling or machine learning methods[65,72,74,76–80] have only simulated global patterns for a subset of FTs we studied here; moreover, these maps show large differences in data sources, methodology and results[81,87]. We also compared the general patterns of global trait maps generated herein with those from other published products for six major traits and LDMC. The predictability of our GAMs for the common six major

traits was significantly stronger than models of other products (Table A1.12). Almost all six major traits presented similar global spatial patterns to that of previous studies. Nevertheless, regional differences emerged between our results and other products. For example, N_{area} values in Central Africa were at intermediate level in our GAM-predicted map (Fig. 2.4g), contrasting with the relatively lowest values observed in the theory-based map by Dong *et al.*[81]; plant height (H) was predicted to be higher in Europe in this study compared to Schiller *et al.*'s predictions[78]. The predictability of GAM of this study for LDMC was relatively weaker (R_{adj}^2 = 0.39, Table A1.12) and the global pattern of LDMC simulated in this study suggested significant deviation from the previous remote-sensing-based projections[74].

Overall, our GAM-predicted maps showed variably strong correlations with the iNaturalist maps, with the highest R^2 of 64% but a median R^2 of only 15% (Table A1.13). They showed varying degrees of overestimation or underestimation compared to the iNaturalist maps (Fig. A1.13). We first checked which regions across the globe showed overestimation and underestimation of our predicted FTs compared to iNaturalist data (Fig. A1.14). Our GAMs tended to predict higher LMA, DM and SSD and lower SCD than iNaturalist data in the northern temperate regions, but lower N_{area}, SSD and LFM than iNaturalist data in southern tropical and temperate regions. H was overestimated globally by the GAM compared to iNaturalist, especially in the northern high latitudes.

Plot-level mean values of 16 FTs in our study were obtained from the sPlotOpen dataset. In principle, the above disagreements may arise from the respective sampling biases of iNaturalist and sPlotOpen. The iNaturalist contributors are by no means evenly sampled across growth forms (tree/shrub/herb)[86], while the sPlotOpen data cover a relatively small portion of the global climate space. The coverage of growth form in each grid cell of the iNaturalist maps indicated a bias towards herbaceous plants in the northern regions, and towards woody plants further south. This is because plant observations in iNaturalist are largely governed by individual decisions[78,86] and are *a priori* likely to underrepresent grasses[86], for example. The data are nonetheless of value because of their independence from other data sources.

The sPlotOpen dataset has both commonalities and disagreements with the iNaturalist data[86]. sPlotOpen is an aggregation of vegetation plots from various studies with specific and distinct

research aims across global ecosystems that do not share the same criteria in the sampling approach, which may also lead to biases in this dataset[82,83]. Large areas of the globe, including Canada, Central South America, Central Africa, Central Asia, East Asia, and Siberia, are undersampled (Fig. A1.4). This sampling bias affects the robustness of predicting trait distribution for large areas. Undersampling of specific climatic conditions (e.g., high-latitude and tropical climates) can result in an overrepresentation of trait values from well-sampled temperate ecosystems, which can introduce bias into global trait-climate relationships. Predicted trait distributions in regions with limited coverage of certain vegetation types are subject to greater uncertainty, further reducing the robustness of global predictions. Furthermore, while sPlotOpen provides plot-level means for each trait, not all species within every vegetation plot have been recorded. 16.5% of plots have all vascular plant records, plant records of 72.1% plots are not specified, and the remaining 11.4% of plots have only certain species records (e.g. only woody plants or dominant plants) in the final plot-level mean trait dataset used in this study (Table A1.2). As a result, plot-level trait means sampled for specific target plant in a vegetation plot cannot accurately reflect the CWMs of traits within the corresponding regions.

The selection of predictor variables in GAMs might contribute to the disagreement between the two set of global trait maps. Our study applied a consistent set of three bioclimatic variables to each FT, thereby preserving model simplicity and providing a consistent baseline for comparison. The three selected variables reflect three most important aspects of climate that govern plant distributions, i.e., winter cold, summer warmth and moisture availability[11,14,152]. These three variables effectively explained global patterns of different FTs and sufficiently distinguished the main differences in trait-climate relationships among three plant groups. However, some FTs can be influenced by other factors in addition to the bioclimatic variables used in this study. Soil fertility factors and soil age have been indicated to correlate with leaf nutrient traits[67,79,171–173]. Land use and disturbance regimes can also affect trait-environment relationships[72,77,174]. Disturbance by fire or grazing and browsing by megaherbivores have strong effects on traits of seed dispersal and plant regeneration[72]. Atmospheric CO_2 concentration is also an important driver of some FTs, directly affecting the growth rates and productivity of plants[72,175]. Evolutionary history effects (represented

by phylogeny) were reported to explain on average more than two-thirds of the variability of the foliar concentrations of N, P and K[79].

The scale at which trait variation occurs may be another reason for the lower correlations between GAM maps and the iNaturalist data for some FTs. Some FTs lack macroecological patterns, such as C_{mass}, which showed little variation on a global scale (Fig. 2.4i). In addition, our community-level prediction ignores intraspecific trait variation. A global meta-analysis based on 36 plant traits indicated that traits can vary substantially among individuals within species, and that the relative amount of intraspecific trait variation does not vary with plant growth forms or climate[176].

This study has characterised universal relationships among FTs and between FTs and climates within non-woody, woody deciduous and woody evergreen plants, and simulated global patterns of FTs based on these relationships. It has highlighted certain key differences between the three plant groups in their ecological strategies as represented by trait-climate correlations, indicating that it is useful to differentiate between non-woody, woody deciduous and woody evergreen plants in large-scale, trait-based studies. Based on three bioclimatic variables and global vegetation cover and their interaction effects, we can explain up to three-quarters (on average about a half) of global variation of community-weighted means for all 16 FTs. By generating global maps for all 16 FTs at 0.1° resolution, we have provided the most comprehensive set of trait maps based on statistical upscaling approach to date.

2.5 Methods

2.5.1 Plant functional traits and three plant groups

The FT data were collected from the sPlotOpen dataset[83] and the enhanced species-level trait dataset of Díaz *et al*[85] (see Table 1).

The sPlotOpen dataset provides plot-level trait means (as community-weighted means, CWMs) of 18 FTs, as well as plant taxa within each vegetation plot and their relative abundances[83]. We calculated the leaf mass per unit area (LMA) and area-based leaf phosphorus content (P_{area}) from specific leaf area (SLA) and mass-based leaf phosphorus content (P_{mass}) in the sPlotOpen dataset according to equations (1) and (2). After removing some redundant traits, such as SLA, P_{mass},

mass-based leaf nitrogen content (N_{mass}) and leaf N:P ratio (which can be directly calculated from N_{area} and P_{area}), 16 FTs were selected for analysis (Table 2.1). All 16 FTs were natural-log transformed before analysis. Here we focus on N_{area} and P_{area}, because Osnas *et al.*[177] showed that leaf nitrogen and phosphorus contents are approximately distributed proportional to leaf area rather than mass. N_{area} and P_{area} were found to be proportional to LMA and photosynthetic capacity of plants[149,150,178]. Both LMA and photosynthetic capacity are quantitatively predictable from climate[91,178], which may lead to potentially high predictability of N_{area} and P_{area} from climates. In contrast, values of N_{mass} and P_{mass} may be more conservative under climate change. We also conducted the following analyses on N_{mass} and P_{mass} (see A1 Supplementary information for Chapter 2 A1.2 for details) and found that climate is not good predictor of their distributions.

$$LMA\ (kg/m^2) = \frac{1}{SLA\ (m^2/kg)} \quad (1)$$

$$P_{area}\ (g/m^2) = \frac{P_{mass}\ (mg/g)}{SLA\ (m^2/kg)} \quad (2)$$

All plant taxa within each vegetation plot in the sPlotOpen data were extracted. We collected plant information including woodiness, plant growth form and leaf phenology for each taxon from public databases, regional floras, and published papers (see Data availability for relevant references of plant information). Each plant taxon was then assigned to a growth form according to the collected plant information to create a plant growth form dataset. There are 42,676 unique taxa in the dataset. Table A1.3 presents 16 growth forms in total and the proportion of each growth form in the plant growth form dataset.

We classified all taxa as non-woody, woody deciduous or woody evergreen according to their growth forms. Some minority growth forms, including ferns, palms, cycads, bamboos, cacti and succulents, were omitted from subsequent analyses. We then used the relative coverage of each taxon within a vegetation plot provided by the sPlotOpen data and the assigned probability of each taxon being in each plant group (Table A1.4) to calculate the relative coverage of each of the three plant groups within a plot. The calculation followed equations (3) and (4).

$$\mathrm{RA(group)} = \sum \big(\mathrm{RC(taxon)} \times \mathrm{P(group)}\big) \quad (3)$$

$$\mathrm{RC(group)} = \frac{\mathrm{RA(group)}}{\sum(\mathrm{RA(group)})} \quad (4)$$

where $\mathrm{RA(group)}$ is the relative abundance of each plant group within a vegetation plot; $\mathrm{RC(taxon)}$ is the relative coverage of each taxon within a vegetation plot provided by the sPlotOpen dataset; $\mathrm{P(group)}$ is the probability of each taxon being in each plant group, which is derived from the probability of each growth form being in each plant group; $\mathrm{RC(group)}$ is the relative coverage of each plant group within a vegetation plot.

The species-level trait dataset from Díaz *et al*.[85] provides species-mean values of six major FTs that constitute the global spectrum of plant form and function[63]. We converted N_{mass} to area-based leaf nitrogen content (N_{area}) based on the raw data, according to the equation (5). After merging this species-level trait dataset with the sPlotOpen dataset and the plant growth form dataset, a total of 35,773 common species were explicitly classified into non-woody, woody deciduous and woody evergreen groups. The scientific names of all taxa from different datasets were aligned before integration by using Taxonomic Name Resolution Service V5.2 (TNRS, https://tnrs.biendata.org/)[179]. We then calculated CWMs for the six major FTs of each of the three plant groups within each vegetation plot based on their species-level means and the relative coverage of each plant group. These CWMs were then natural-log transformed prior to further analyses. All data manipulations and following analyses were conducted in R v4.2.2 (https://www.r-project.org/)[151].

$$N_{area}\ (g/m^2) = \frac{N_{mass}\ (mg/g)}{SLA\ (m^2/kg)} \quad (5)$$

2.5.2 Bioclimatic variables

Three bioclimatic variables were selected[11,14,152] (Fig. A1.2): i) moisture index (MI, unitless) representing plant-available moisture – the natural-log transformed MI (ln MI) was used in the analysis to emphasise differences at the dry end of the moisture range[180]; ii) mean temperature of the coldest month (MTCO, °C) representing winter cold; iii) mean growing-season temperature (MGST, °C) representing summer warmth. The three selected variables correspond to the three

recognised climatic dimensions that regulate global geographic distribution of vegetation[11,14,152]. They control the plant distribution limits by affecting the plant attributes that determine physiological processes and adaptive strategies, such as life form, leaf phenology, leaf size and stomatal conductance[11,14,152]. Hence, the distribution of some key plant traits is likely to be a function of these bioclimatic variables, and thus may be quantitatively predicted by these variables[152].

We collected gridded climatological data for further calculations, with a spatial resolution of 30 arc-seconds (*ca* 0.01°). Growing degree days (heat sum) above 0 °C (GDD_0, °C), number of growing degree days above a base level of 0 °C (NGD_0, number of days), monthly averaged mean daily air temperature (tas, K/10) and mean monthly precipitation amount (pr, kg m^{-2} $month^{-1/100}$) were obtained from CHELSA V2.1 (https://chelsa-climate.org/)[181,182], covering the period from 1981 to 2010. Downward total shortwave solar radiation data accumulated over the month (srad, kJ m^{-2}) were downloaded from the CHELSA V1.2 (https://chelsa-climate.org/)[181,182], covering the period from 1979 to 2013. Global elevation derived from the SRTM elevation data (elev, m) was acquired from WorldClim v.2.1 (https://www.worldclim.org/)[183].

We used the Simple Process-Led Algorithms for Simulating Habitats (SPLASH) model[184,185] with climatic variables as inputs to calculate the mean monthly potential evapotranspiration (PET, mm). We extracted the climatic values for each vegetation plot according to its longitude and latitude. The 'splash.point()' function in R *rsplash* package[184,185] was applied to calculate the PET of each plot. We then decreased the resolution of raw climatic raster data from 0.01° to 0.1° and fitted them into the SPLASH model by 'splash.grid()' function in R *rsplash* package[184,185] to calculate the global PET. Three bioclimatic variables (Fig. A1.2) were then calculated as follows:

$$MI\ (\text{unitless}) = \frac{\sum_{k=1}^{12} pr_k}{\sum_{k=1}^{12} PET_k} \tag{6}$$

$$MTCO\ (°\text{C}) = \min_{1 \le k \le 12} tas_k \tag{7}$$

$$MGST\ (°\text{C}) = \frac{GDD_0}{NGD_0} \tag{8}$$

where pr_k (mm) is mean monthly precipitation; PET_k (mm) is mean monthly potential evapotranspiration calculated from the SPLASH model; tas_k (°C) is monthly average temperature;

GDD_0 (°C) are growing degree days heat sum above 0 °C; and NGD_0 (number of days) is number of growing degree days above a base level of 0 °C.

2.5.3 Vegetation coverage

We checked the 'naturalness' of each vegetation plot in our trait dataset against the global map of land cover provided by the ESA CCI-LC database (https://www.esa-landcover-cci.org/?q=node/164)[186] at 300 m spatial resolution in 2010. For plots belonging to all types of croplands, urban areas, bare areas, water bodies and permanent snow and ice in the ESA CCI-LC database, we defined them as 'unnatural vegetation' and removed those plots from the trait dataset used for data analysis. After matching the trait dataset with bioclimate data, 77,074 natural vegetation plots remained in the final dataset for our study.

The ESA CCI-LC global land cover data were also used to calculate the global fractional cover of three plant groups. The spatial resolution of the raw land cover raster was decreased from 300 m to 0.1° to match with the resolution of bioclimate raster for further analyses. This ESA CCI-LC dataset contains 23 classes including 38 sub-classes of global land cover (Table A1.5). We first selected the vegetation plots that contained records of 'all vascular plants' in our trait dataset (12,743 plots in total). We then identified the land cover class and sub-class of each of these vegetation plots according to the ESA CCI-LC maps at original 300 m resolution and accordingly calculated the median relative cover of the three plant groups for each sub-class. The coverages of three plant groups of artificial ecosystems and non-natural and non-terrestrial vegetations were set to be NA. We then replaced grid values representing sub-classes of global land cover in the ESA CCI-LC maps with the median relative cover of the three plant groups within each sub-class and calculated the average values at 0.1° resolution to produce the global fractional cover of the three plant groups.

2.5.4 Multivariate analysis

The natural-log transformed CWMs of six major FTs of each plant group were rescaled to a mean of 0 and a standard deviation of 1 and then subjected to a Principal Component Analysis (PCA) and a Redundancy Analysis (RDA), conducted by 'prcomp()' function and 'rda()' function in the R *vegan* package[187], respectively. PCA was used to evaluate the covariation of trait

combinations among six FTs in different plant groups. RDA was applied to describe the extent to which the variation of trait combinations (response variables) can be explained by bioclimatic variables (explanatory variables).

2.5.5 Generalised Additive Models (GAMs)

Correlations between FTs and three bioclimatic variables were characterised and visualized by Generalised Additive Models (GAMs). GAMs were fitted for natural-log transformed CWMs of six major FTs in each of the three plant groups with three bioclimatic variables as explanatory variables, without interaction terms, using the 'gam()' function in R *mgcv* package[188]. We used the RMEL (restricted maximum likelihood) as smoothing parameter to control the smoothness of the predictive functions. We then calculated the relative importance of each bioclimatic variable in the GAMs to determine which of the predictors is more significantly related to the distribution of each FT. The concurvity of three bioclimatic variables in GAMs was checked before fitting models (mean values: ln MI = 0.487 ± 0.06, MTCO = 0.651 ± 0.07, MGST = 0.867 ± 0.02). MGST presented relatively high concurvity, which make it challenging to interpret the individual importance of variables in trait prediction. Nonetheless, the overall predictive performance of the GAMs remains robust, with the combined effect of the predictors effectively capturing the variation in FTs.

We sampled the bioclimates at regular intervals from minimum to maximum over the globe, e.g., sampling at increments of 0.05 (unitless) for ln MI, 0.25 °C for MTCO, and 0.25 °C for MGST, to generate three sample datasets simulating the global climate space. The sample data and GAMs were then used to predict the distribution of trait values in the global climate space for each plant group. The response surfaces of trait values (natural-log transformed) resulting from GAMs were presented as contours in the three-variable climate space in the form of two-dimensional slices. Convex hulls generated by 'convhulln()' function in the R *geometry* package[189] were used to avoid representing parts of the fitted surface that are not constrained by data. For each plant group separately, we also predicted and mapped the global distribution of natural-log transformed values of six major FTs, as predicted by the GAMs, based on global raster data for the three bioclimatic variables at 0.1° spatial resolution.

2.5.6 Global trait maps and evaluation

In order to simulate global trait patterns for all plant taxa, we fitted new GAMs for all 16 FTs with natural-log transformed plot-level means (CWMs). In the new GAMs, trait values were predicted by not only three bioclimatic variables, but also global fractional cover of plant groups, and interactions between each bioclimatic variable with the cover fraction of each plant group. As the sum of the global fractional cover for the three plant groups is 1, we only considered the coverage of two of these groups (woody deciduous and woody evergreen) when fitting the new GAMs. The new GAMs also used the RMEL as smoothing parameter and were fitted using the 'gam()' function in the R *mgcv* package[188]. We then generated global trait maps for all 16 FTs at 0.1° spatial resolution.

Although using a weighted average of three maps from different plant groups should be a more straightforward way to generate global maps, we did not adopt this approach. This is because simulating separate global trait patterns for each of the three plant groups, and calculating weighted averages of them with fractional cover of plant groups as weights, would only be applicable where traits possess species-level values. In this study, only the six major traits are provided with species-level mean values. The remaining traits have plot-level mean values, which are weighted averages of trait values for all plant taxa within the plot. It is almost impossible to predict global trait patterns for each plant group from such community weighted mean trait values. However, using the GAMs including fractional covers of plant groups provides a more universal alternative approach, which is directly applicable to plot-level means.

We tested our global trait maps based on GAMs predictions against another set of global trait maps[86] based on the combinations of citizen-science plant observations from the iNaturalist project and trait measurements from the TRY database. The iNaturalist-based maps provided all FTs available for the calculation of the same 16 FTs as in our study. The plot-level means of all 16 FTs in our study were taken from the sPlotOpen, and the global trait patterns estimated by the iNaturalist tend to most strongly resemble that of the sPlotOpen at 2° resolution[86]. Therefore, we aggregated and reprojected our global trait maps according to the iNaturalist maps (2° spatial resolution) using bilinear interpolation by 'projectRaster()' function in the R *raster* package[190].

Then we assessed the pixel-by-pixel agreement between our trait maps with the iNaturalist trait maps by linear regression using the 'lm()' function in the R *stats* package.

2.5.7 Statistics and reproducibility

All of the above statistical analyses are reproducible by following the procedures in Methods. All data and R scripts for carrying out all the above analyses are openly available at: https://doi.org/10.5281/zenodo.13325275.

Chapter 3: Deep-time geographic dynamics of climate shape global vascular plant diversity

3.1 Abstract

Climate has been proposed to shape global biodiversity through multiple, competing mechanisms that emphasise either magnitude, spatial distribution, or temporal stability of climatic variables[105–107,131], without integrating these dimensions into a unified spatio-temporal framework. Here we quantify spatio-temporal climate change in both climatic and geographic spaces by tracing when and where climatic conditions have emerged, persisted, and shifted across the Earth's surface over the past 50 million years. We compile the most comprehensive global vascular plant diversity map to date (350,864 species) to test how deep-time geographic dynamics of climate have influenced plant diversity patterns. We find that species-rich and evolutionarily diverse plant communities are associated with climates that originated early, remained geographically extensive over long time, and experienced moderate geographic shifts. These climates are primarily warm and humid, now dominant in the tropics but previously extending into high latitudes. Models incorporating climate and its deep-time geographic dynamics explain up to 95% of the global variation in plant diversity, with geographic extent and mobility of climate contributing over half the explanatory power for species richness, turnover, and phylogenetic diversity. This work addresses a fundamental question that has long been debated regarding the spatio-temporal impacts of ongoing climate change on biodiversity.

3.2 Main

That warm, humid tropical regions harbour more species than cold, dry higher latitudes is one of the fundamental patterns of life on Earth[104–108]. It is accepted that climate plays a key role in shaping global biodiversity patterns by affecting the speciation, dispersal, and extinction[32,34,191–193]

of animals[117,121], plants[194,195], and microbes[196]. However, the underlying mechanisms remain the subject of ongoing debate[106,109,110].

A well-studied explanation for how climate shapes latitudinal biodiversity gradient is the 'species-energy hypothesis', which posits that environments with greater energy and water availability (e.g., temperature, precipitation, and productivity) support more species[111,113,197–199]. Yet, accumulating evidence suggests that the geography and history of climate may be of equal or greater importance than any intrinsic property of climate[105,131]. The importance of climate geography was first formalized in the 'area hypothesis', stating that larger area of tropics provides greater carrying capacity for higher species richness[118–120]. More recently, this idea was refined into the 'area-isolation hypothesis' by Coelho and colleagues[121]. They further incorporated the geographic isolation of a climate – measured as the average distance among disconnected fragments of the same climate – as a driver of speciation[121]. In parallel, other hypotheses focus on the climate history of locations. The 'evolutionary-time hypothesis' argues that the tropics, being older than higher latitudes, have hosted lineages that originated in tropical climates for longer, allowing more time for diversification and species accumulation[109,114–116]. The 'stability hypothesis' proposes that the tropics, having been less affected by past glaciations and large-scale climatic disruptions, offered lineages relatively continuous opportunities for diversification, whereas higher latitudes experienced repeated extinctions and recolonisation[27,35,110,117].

None of these hypotheses simultaneously considers the dual dynamics of climate in space and time. Hypotheses focused on climate geography are temporally static. They explain current biodiversity patterns using present-day geographic attributes of climate (e.g., area and isolation) but ignore how these attributes evolved over millions of years[121,131]. A climatic condition that is widespread today may have experienced contraction and fragmentation in the past[33]. Conversely, hypotheses focused on climate history are spatially static. They measure value changes of climatic variables at fixed locations (e.g., within latitudinal bands), ignoring the key fact that species can track migrating climates suitable for their survival[26]. A climatic condition, defined by a stable set of climatic variables, can persist through time, though its geographic position may shift across the Earth's surface[33]. The limitations of hypotheses based on the geography or history of climate are complementary and together reveal a unifying perspective: the spatio-temporal dynamics of climate

may be a primary driver of global biodiversity patterns. From this perspective, climates that remained geographically widespread over long timescales can provide larger and more stable evolutionary stages for persistence, diversification, and accumulation of species[105,131,200]. Simultaneously, geographic expansion, contraction, and migration of climate over evolutionary time are expected to influence species dispersal, extinction, and speciation[26,33]. However, how the deep-time geographic dynamics of climate have shaped global biodiversity patterns remain unknown.

Quantifying the spatio-temporal influence of climate on global biodiversity requires a shift in analytical framework – from geographic to climatic space. This can be achieved by constructing a multi-dimensional space using climatic variables, which is then discretised into a grid of 'climatic bins', each representing a unique climatic condition[121,131]. One climatic bin may span disconnected geographic regions across both present and past, enabling the identification of historical geographic distributions and the quantification of spatio-temporal dynamics for each climatic condition. Within this framework, vascular plants are central. As primary producers constituting over 80% of Earth's biomass, vascular plants form the structural and trophic foundation of terrestrial ecosystems[50]. Variation in plant diversity can influence the composition and diversity of other terrestrial trophic levels[201–203]. Fossil pollen and macrofossil records also demonstrate that plants can track climate change via continental-scale migration[26,29].

A further challenge lies in defining an appropriate timescale. Phylogenomics of angiosperms reveals an early burst of diversification during the Late Jurassic and Early Cretaceous, which gave rise to more than 80% of extant angiosperm orders[204]. This was followed by a long period of steady, slower diversification for at least 80 million years through the remaining Mesozoic, during which around three-quarters of all modern angiosperm families originated[204]. A second surge in diversification rates occurred in the Cenozoic, concurrent with a period of global cooling that may have acted as a driver for this resurgence[204]. From the Early Eocene to the Late Pliocene (*ca*. 55–2.6 Ma), long-term cooling eliminated high-latitude tropical climates and expanded temperate climates poleward, contributing to the modern latitudinal diversity gradient[33]. The subsequent Quaternary was marked by glacial-interglacial cycles with large fluctuations in temperature, precipitation, and CO_2 concentration, imposing strong environmental filtering on species[26]. The

Holocene and the pre- and post-industrial periods reflect post-glacial warming and the increasing impact of human activities[205].

Here, we investigated how climate and its deep-time geographic dynamics have shaped present-day global patterns of vascular plant diversity. We hypothesised that (1) climates that have remained widespread over evolutionary time harbour more species, and (2) the historical stability of climate' geographic extent influences species composition. By compiling global occurrence records for 350,864 vascular plant species, we generated the most comprehensive data-based map of global plant diversity to date. We defined a climatic space based on recent (1901–1989) temperature and plant-available moisture data. Within this space, analogues of each climatic condition were identified across key geological periods for which multiple palaeoclimate model simulations are available: the Eocene (*ca*. 50 Ma), Mid-Pliocene (3.3–3.0 Ma), Last Glacial Maximum (LGM, 22–18 ka), Mid-Holocene (*ca*. 6 ka), and pre-industrial (pre-1850) periods. We quantified the geographic extent of climatic conditions, as well as their persistence and shift, across the selected periods. Finally, we assessed the relative contributions of contemporary climate and its deep-time geographic dynamics in explaining global vascular plant diversity today.

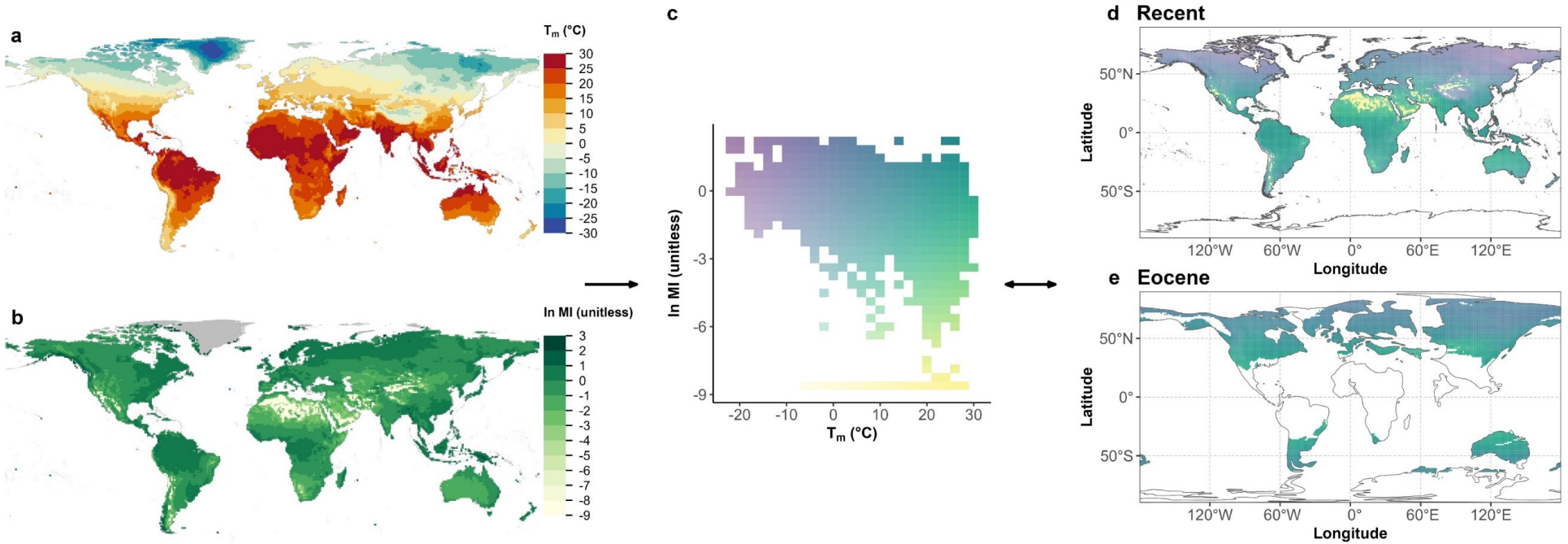
3.3 Results

3.3.1 Historical geography of contemporary climates

A temperature index (T_m, °C) and a log-transformed moisture index (ln MI, unitless) were used to define a two-dimensional climatic space (Fig. 3.1a–c). These two bioclimatic variables represent major independent climatic controls on the global geographic distribution of plants[11,14,152,206], with T_m capturing the latitudinal temperature gradient[207] and ln MI representing plant-available moisture[11,14,152,206] (see Methods). The climatic space was constructed using climate data from 1901 to 1989, a recent period selected to reflect post-industrial conditions prior to the onset of accelerated anthropogenic influence around 1990[208]. Multiple grid resolutions of the climatic space were generated by dividing each axis into equal intervals (see Methods); all main-text results are based on the 30-interval resolution, with results from alternative resolutions (40, 50, and 60 intervals) reported in the A2 Supplementary information for Chapter 3. Each grid cell in the

climatic space – hereafter referred to as a ‘climatic bin’ – represents a distinct recent climatic condition[121,131,134] and is visualised by a unique colour (Fig. 3.1c, and see Figs. A2.1 and A2.2 for alternative colour gradient schemes). Each bin may correspond to multiple geographic locations across different time periods[121,131,134]. This property allows for the tracing of the historical geography of recent climatic conditions. We then identified the geographic locations of these climatic conditions in the recent period (1901–1989, Fig. 3.1d) and in five geohistorical periods (Extended Data Fig. 3.1).

Fig. 3.1: Mapping climates in geographic and climatic spaces.



a and **b**, Geographic distributions of the two bioclimatic variables used to define the climatic space: the temperature index (**a**, T_m, °C) and the natural-log-transformed moisture index (**b**, ln MI, unitless), for the recent reference period (1901–1989). **c**, The two-dimensional climatic space defined by the recent T_m and ln MI, partitioned into 30 equal intervals along each axis. Each coloured grid cell ('climatic bin') represents a unique climatic condition (405 cells in total). Alternative colour gradient schemes are provided in Figs. A2.1 and A2.2. **d** and **e**, The application of the framework to trace historical geography of recent climates. The map in **d** shows the geographic locations of all climatic bins during the recent period (1901–1989). The map in **e** shows the geographic locations where those same climatic conditions (defined in **c**) existed during the Eocene (*ca.* 50 Ma); uncoloured

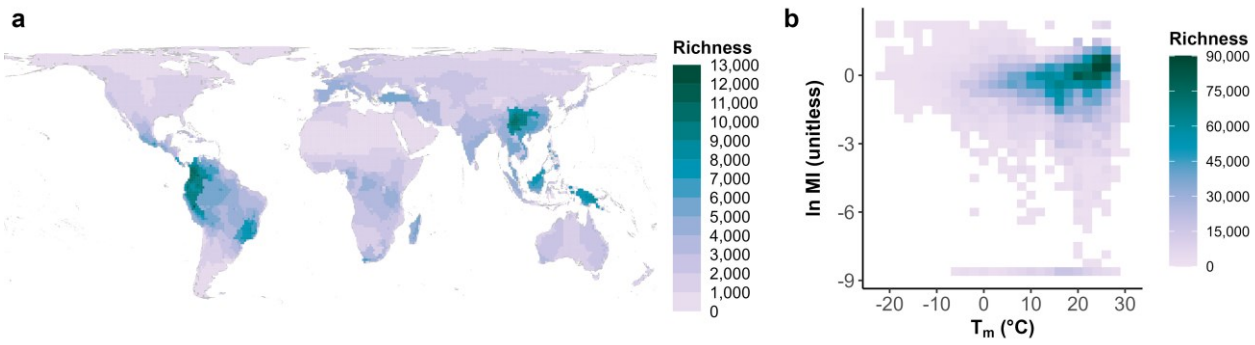
areas denote climatic conditions absent during this period. Colours correspond to those in **c**, with identical colours indicating the same climatic conditions. The same colour gradient scheme is consistently applied across subsequent figures. The geographic distribution of recent climatic conditions in other geological periods is shown in Extended Data Fig. 3.1. All maps are shown at a 1° spatial resolution.

3.3.2 Global patterns of vascular plant diversity

By compiling native occurrence records for 350,864 vascular plant species (see Methods), representing over 98% of accepted species in the World Checklist of Vascular Plants (WCVP)[209], we generated a data-based global map of plant species richness at 1° spatial resolution (Fig. 3.2a, see Methods for details). To map plant diversity patterns in climatic space, we counted the number of species occurring within each climatic bin defined by T_m and ln MI (Fig. 3.2b). While global patterns of plant diversity have previously been explored through model-based or empirical approaches[194,209–215], our map represents the most taxonomically complete global dataset compiled to date at fine spatial resolution (Table A2.1).

Hotspots of plant diversity (Figs. 3.2a and A2.3) are concentrated in Central and South America, including the Andes-Amazonia and Atlantic Forest in Brazil; in the mountains of Southwest China; across Southeast Asia, extending from the Indochina and Malay Archipelago to the island of New Guinea; in parts of Africa, encompassing Cameroon, Tanzania, Madagascar and the Cape region of South Africa; and in parts of Europe and West Asia including the Alps, the Mediterranean basin and the Caucasus mixed forests. In climatic space (Fig. 3.2b), species richness peaks under warm conditions ($T_m \approx 19–28$ °C) and humid conditions (MI > 1). These patterns are consistent with previous estimates[194,209–216].

Fig. 3.2: Global patterns of vascular plant diversity across geographic and climatic spaces.



a, Global map of vascular plant species richness in the Behrmann equal-area projection at ~110 km resolution (~1° resolution), based on native distributions derived from occurrence records of 350,864 species. **b**, Global patterns of vascular plant species richness in the climatic space defined

in **Fig. 3.1**. T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index.

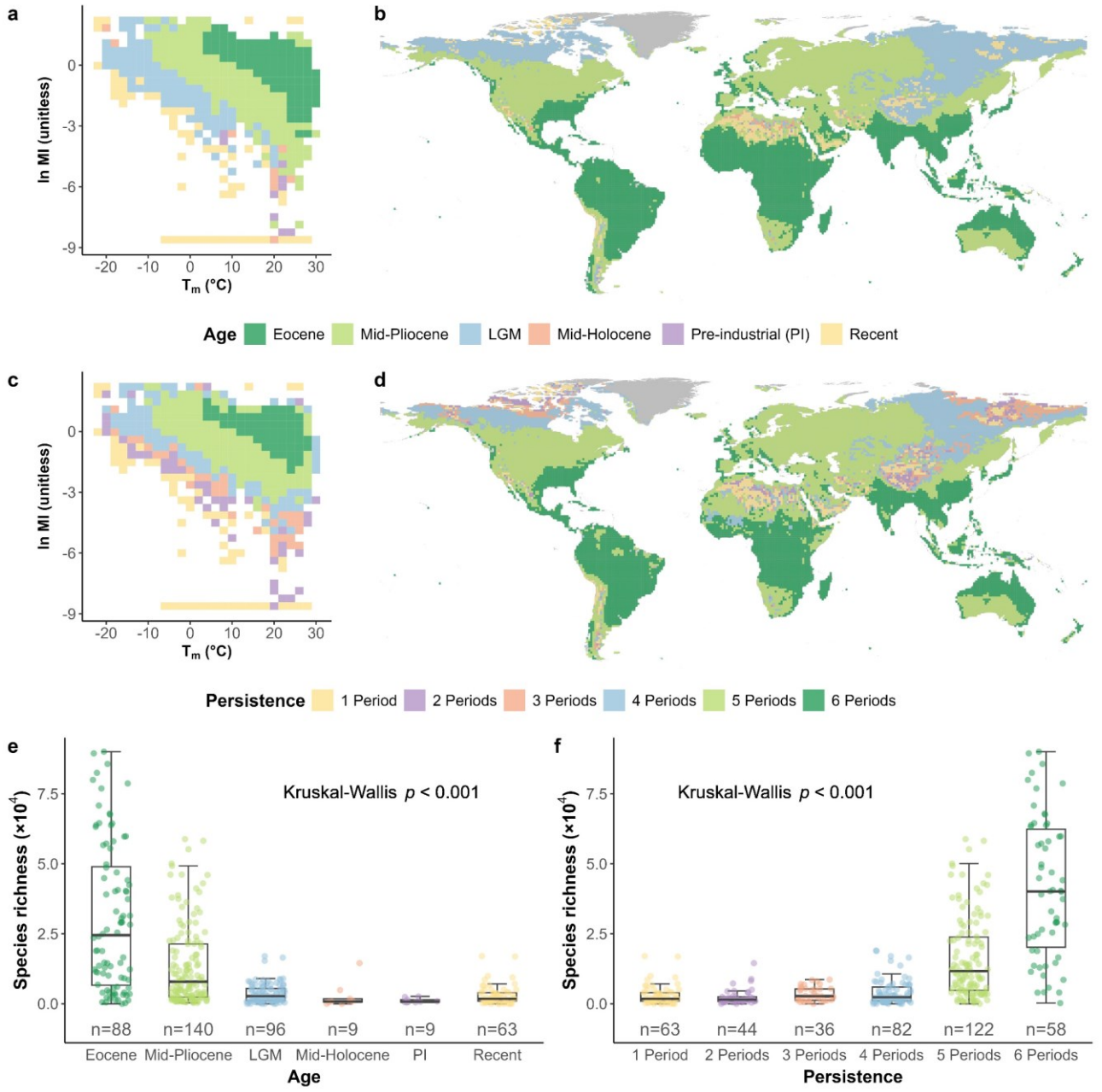
3.3.3 Old and persistent climates harbour more plant species

We evaluated how the age and persistence of recent climatic conditions relate to contemporary plant diversity. 'Age' was defined as the earliest of six selected geological periods during which a recent climatic condition first appeared. 'Persistence' was calculated as the number of these periods in which a climatic condition occurred.

Age and Persistence of recent climates show similar distribution patterns across both climatic and geographic spaces (Fig. 3.3a–d). Warm and humid climates tend to be both older and more persistent, originating as early as (or before) the Eocene and persisting to the present (Fig. 3.3a, c). These climatic conditions are predominantly found in present-day tropical regions (Fig. 3.3b, d). As the Earth cooled from the Eocene to the Pliocene[33], climatic conditions typical of today's temperate zones began to emerge, with many first appearing in the Mid-Pliocene and continuing through to the present (Fig. 3.3b, d). In contrast, cold and dry climates (Fig. 3.3a, c) near the poles (Fig. 3.3b, d) originated more recently, commonly first appearing during Quaternary glacial periods as illustrated by the LGM. Hot and arid climates (Fig. 3.3a, c) in equatorial regions (Fig. 3.3b, d) – such as the Sahara, Arabian Peninsula, and Mongolian Plateau – emerged recently.

To assess how the age and persistence of recent climates have influenced recent species richness, we compared richness across climatic bins grouped by Age and Persistence. Species richness varies significantly with both factors ($p < 0.001$ in Fig. 3.3e–f). Climates with older origins, especially those dating back to the Eocene (median $\approx 2.5 \times 10^4$, $p < 0.001$ in Fig. 3.3e), and climates that persisted through all six periods (median $\approx 4 \times 10^4$, $p < 0.001$ in Fig. 3.3f), harbour substantially larger numbers of vascular plant species.

Fig. 3.3: Age and persistence of climate and their associations with recent species richness.



a and **b**, Age of recent climatic conditions mapped in climatic (**a**) and geographic (**b**) spaces, indicating when each climatic bin first emerged across the six selected geological periods (Eocene to recent). **c** and **d**, Persistence of climates, defined as the number of periods in which a recent climatic condition was present, shown in climatic (**c**) and geographic (**d**) spaces. **e** and **f**, Relationships between plant species richness and Age (**e**) or Persistence (**f**) of climatic conditions were assessed using Kruskal-Wallis tests. Horizontal lines and boxes indicate medians and

interquartile ranges. Sample sizes per group are shown below the boxes. T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index.

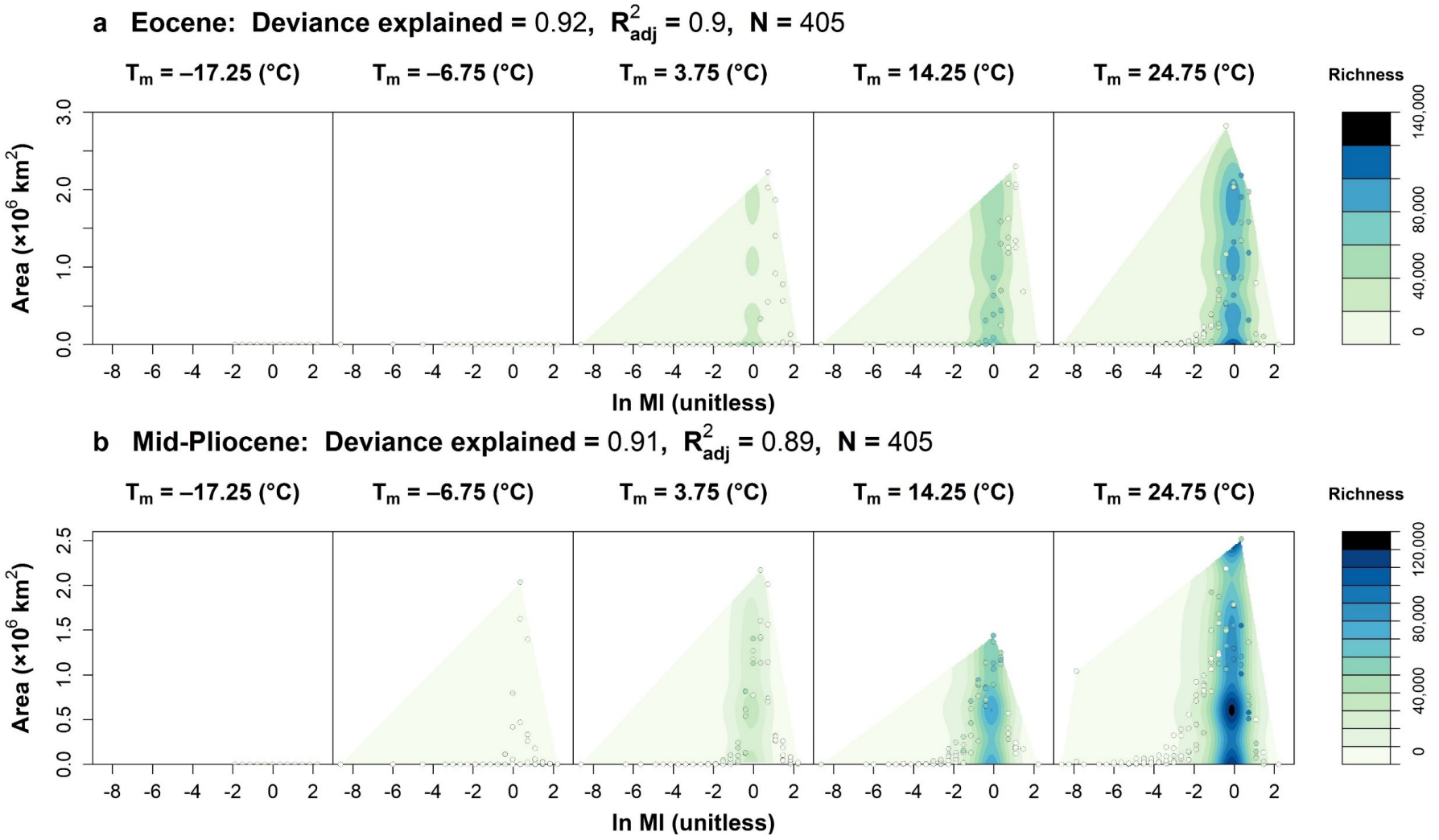
3.3.4 Persistently widespread climates support higher plant species richness

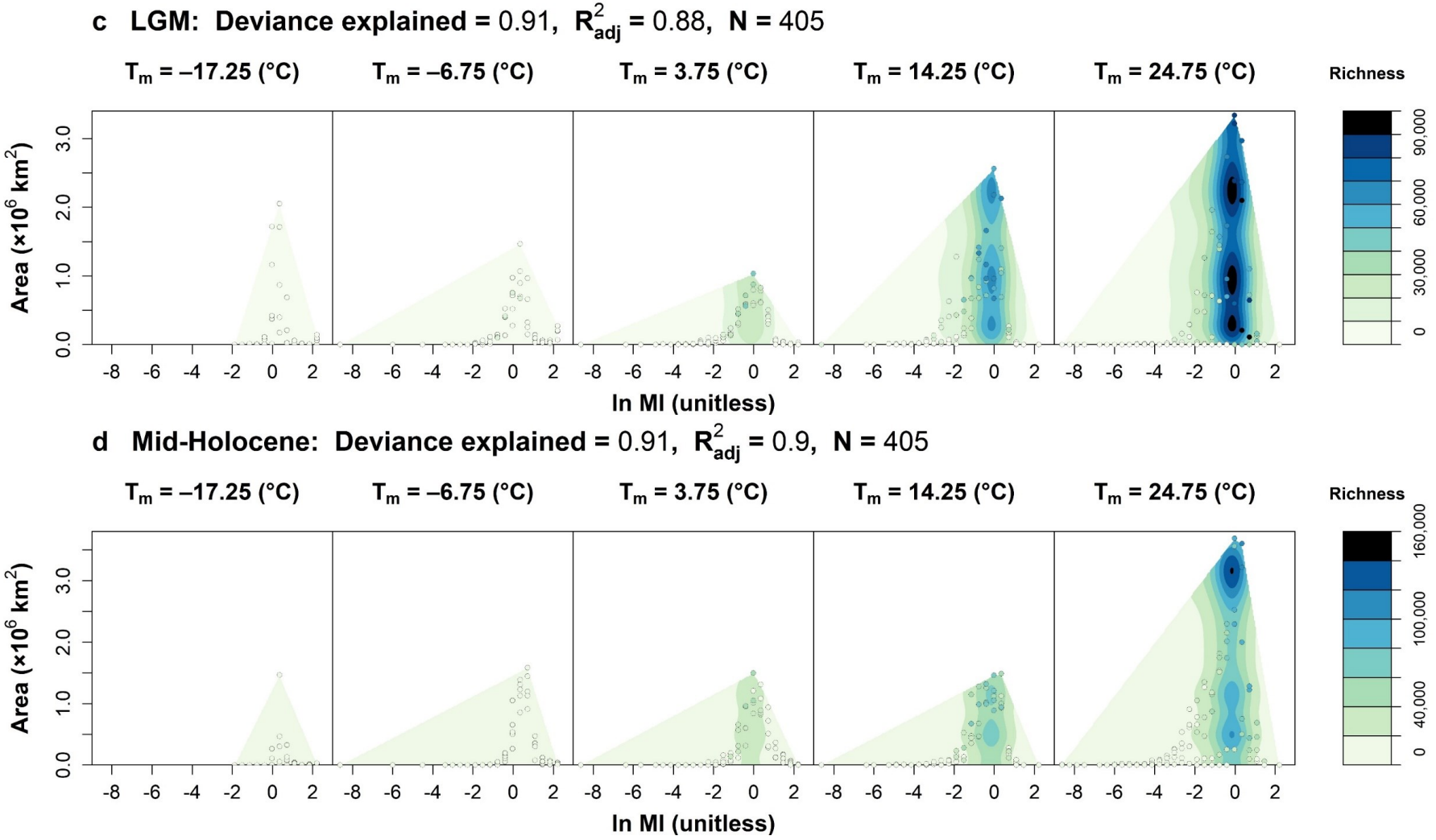
To assess how the geographic extent of climatic conditions influences species richness, we quantified the total land area occupied by each recent climatic condition during five geohistorical periods and the recent period (1901–1989), referred to as 'Area' (km^2). We also calculated the mean 'Area' since the first appearance of each climatic condition, representing its long-term spatial persistence.

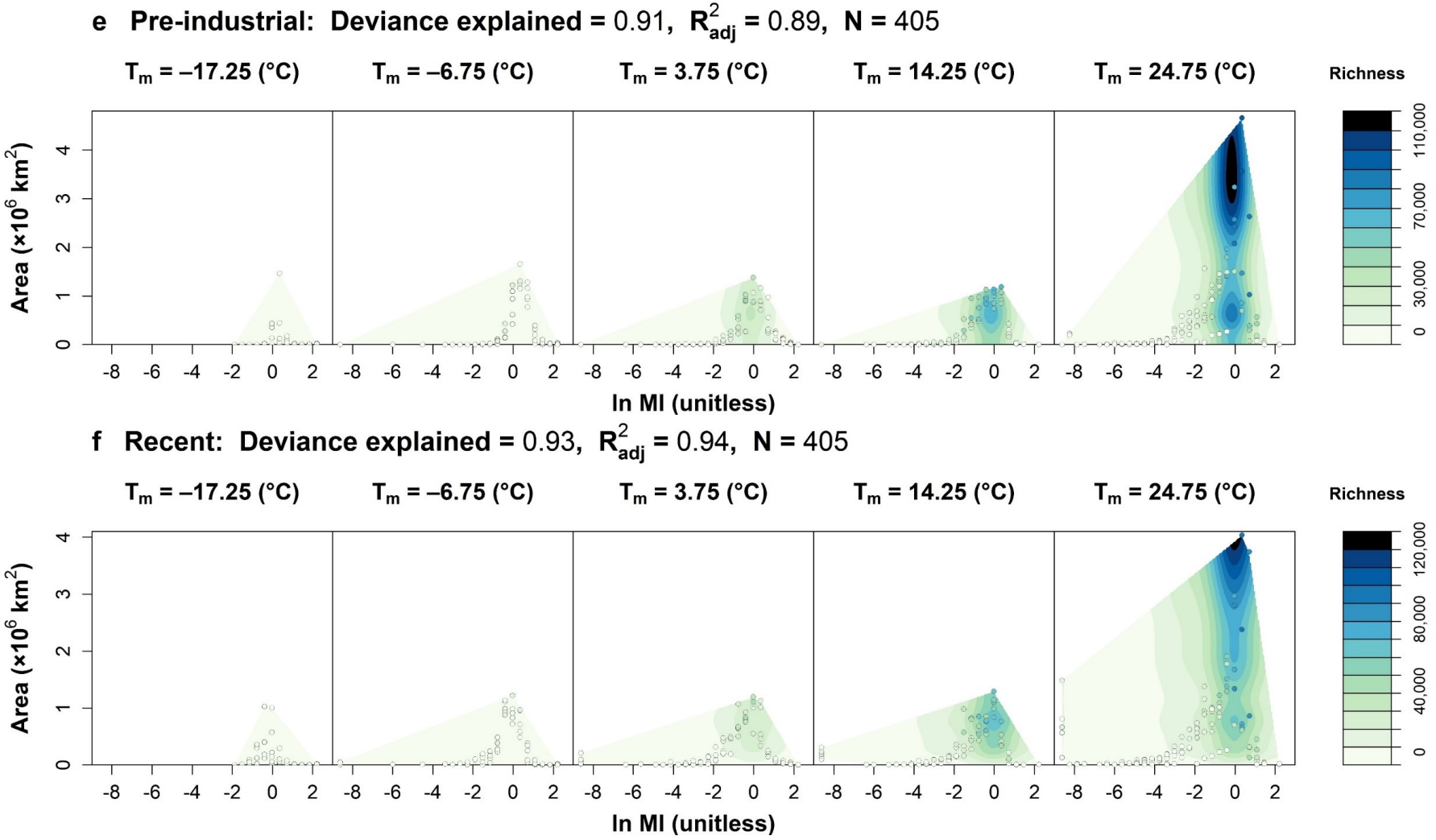
We fitted Generalised Additive Models (GAMs) to model species richness in climatic bins as a function of three explanatory variables: two bioclimatic variables (T_m and ln MI) and Area (either from individual periods or averaged across periods). Separate GAMs were fitted for each of the six periods and for mean Area. Across all models, the explanatory variables explain over 90% of the deviance in species richness (Fig. 3.4 and Tables A2.3–A2.9).

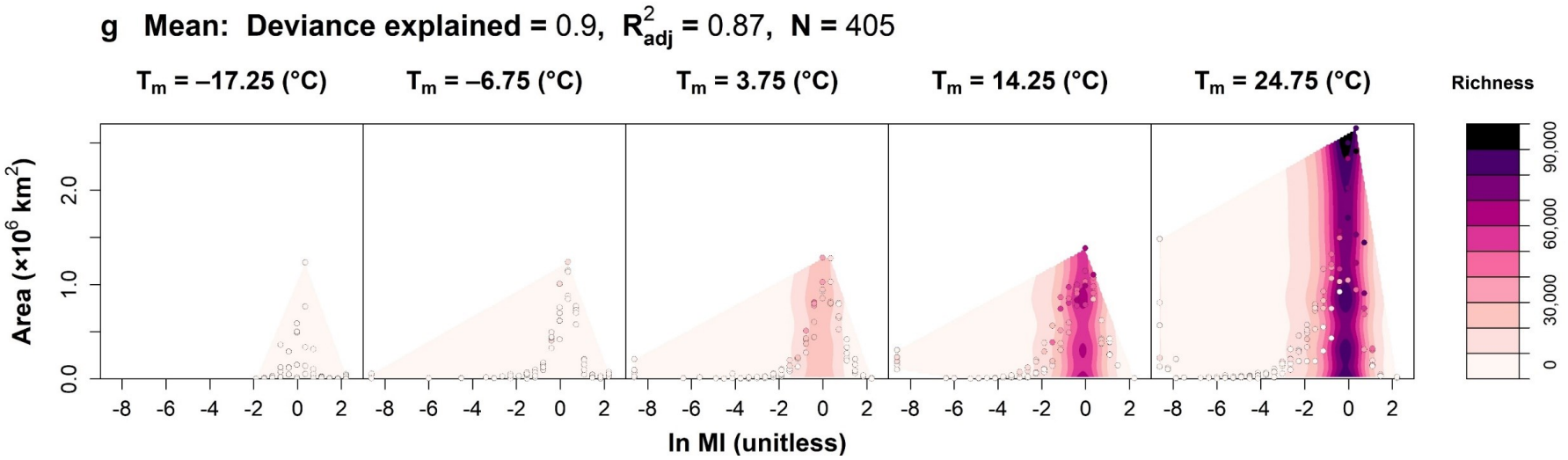
We visualised the relationships between species richness and the three explanatory variables using contour plots (Fig. 3.4). Across all periods, species richness consistently peaks under warm and humid climatic conditions (T_m = 24.75 °C; MI = 1). However, species richness shows different relationships with Area across periods (Fig. 3.4). During the Eocene (Fig. 3.4a), Mid-Pliocene (Fig. 3.4b), and LGM (Fig. 3.4c), species richness was relatively insensitive to the geographic extent of climate. From the Mid-Holocene onwards (Fig. 3.4d–f), species richness was higher in climatic bins occupying larger geographic areas (Area $\geq 3.0 \times 10^6$ km^2 in Fig. 3.4d–f). This pattern is also evident for mean Area across periods (Area $\geq 2.0 \times 10^6$ km^2 in Fig. 3.4g). These results suggest that the long-term persistence of large geographic area, rather than initial spatial extent at the time of emergence, is more important in shaping contemporary plant diversity. Combined with the findings on climate age and persistence (Fig. 3.3), these results indicate that warm, moist climates that have remained geographically extensive for longer periods tend to support more diverse plant communities.

Fig. 3.4: Relationships between recent plant species richness and three explanatory variables across selected geological periods.




c LGM: Deviance explained = 0.91, R^2_{adj} = 0.88, N = 405
T_m = –17.25 (°C)
T_m = –6.75 (°C)
T_m = 3.75 (°C)
T_m = 14.25 (°C)
T_m = 24.75 (°C)
Richness
0
30,000
60,000
90,000
Area (×10^6 km^2)
0.0
1.0
2.0
3.0
-8 -6 -4 -2 0 2
ln MI (unitless)
d Mid-Holocene: Deviance explained = 0.91, R^2_{adj} = 0.9, N = 405
T_m = –17.25 (°C)
T_m = –6.75 (°C)
T_m = 3.75 (°C)
T_m = 14.25 (°C)
T_m = 24.75 (°C)
Richness
0
40,000
100,000
160,000
Area (×10^6 km^2)
0.0
1.0
2.0
3.0
-8 -6 -4 -2 0 2
ln MI (unitless)


e Pre-industrial: Deviance explained = 0.91, R^2_{adj} = 0.89, N = 405
T_m = –17.25 (°C)
T_m = –6.75 (°C)
T_m = 3.75 (°C)
T_m = 14.25 (°C)
T_m = 24.75 (°C)
Richness
0
30,000
70,000
110,000
Area (×10^6 km^2)
ln MI (unitless)
f Recent: Deviance explained = 0.93, R^2_{adj} = 0.94, N = 405
T_m = –17.25 (°C)
T_m = –6.75 (°C)
T_m = 3.75 (°C)
T_m = 14.25 (°C)
T_m = 24.75 (°C)
Richness
0
40,000
80,000
120,000
Area (×10^6 km^2)
ln MI (unitless)



Contour plots show fitted species richness from Generalised Additive Models (GAMs), with species richness modelled as a function of T_m, ln MI, and Area. Panels correspond to: **a**, Eocene (*ca.* 50 Ma); **b**, Mid-Pliocene (3.3–3.0 Ma); **c**, Last Glacial Maximum (LGM, 22–18 ka); **d**, Mid-Holocene (*ca.* 6 ka); **e**, pre-industrial period (pre-1850); **f**, recent period (1901–1989); and **g**, mean Area since the first appearance of each climatic condition across all selected geological periods. Fitted species richness values are shown as contour levels; observed species richness values are overlaid as dots. Darker colours on the right corresponding scale bar indicating higher species richness. Values of ln MI and Area vary continuously along the horizontal and vertical axes, respectively, with each slice created at a specific T_m value. T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Area (km^2), total land surface area occupied by each climatic condition during the corresponding geological period.

3.3.5 Moderate climatic shifts promote plant diversity

A climatic condition that has occupied different geographic locations through time, and its expansion, contraction, or migration, may influence plant community composition. To quantify these dynamics, we developed two metrics. First, we calculated the change in total land area occupied by each climatic condition between its earliest occurrence (among six selected periods) and the recent period, denoted as 'ΔArea' (km^2). Positive values indicate spatial expansion, while negative values indicate contraction. Second, we measured the geographic shift of each climatic condition using the average great-circle distance between its initial and recent geographic distributions, termed 'Movement' (km). In addition to species richness, we computed three more diversity metrics within climatic bins. We measured species turnover among geographic grid cells sharing the same climatic condition, indicating how species composition varies geographically under constant climate. We also calculated Faith's phylogenetic diversity (PD, Ma)[100], representing the total evolutionary history accumulated within each climatic bin. Given the strong dependence of PD on species richness, we also derived relative phylogenetic diversity (RPD, Ma) by regressing PD on richness and extracting the residuals. Large positive RPD values indicate species in a climatic bin are more distantly related than expected, while negative values suggest phylogenetic clustering.

We fitted GAMs to examine how each diversity metric responds to five explanatory variables: two bioclimatic variables (T_m and ln MI), Area, ΔArea, and Movement. Area was drawn from the recent period, as it served as the reference point for calculating ΔArea. Pairwise correlations among these explanatory variables are low, reducing concerns of concurvity (Fig. A2.12). The GAMs explain 95% of the deviance in species richness ($R_{pred}^2 = 0.95$, $R_{adj}^2 = 0.95$), 74% of the deviance in species turnover ($R_{pred}^2 = 0.77$, $R_{adj}^2 = 0.75$), 85% of the deviance in PD ($R_{pred}^2 = 0.85$, $R_{adj}^2 = 0.84$) and 53% of the deviance in RPD ($R_{pred}^2 = 0.53$, $R_{adj}^2 = 0.49$).

We also assessed the effects of climate age and persistence on species turnover, PD, and RPD (Extended Data Fig. 3.3). Older and more persistent climates tend to support greater species turnover and higher PD ($p < 0.001$ in Extended Data Fig. 3.3 a–d). Climates that were present as early as the Eocene and persisted to the present harboured more distantly related and/or

evolutionarily ancient species (RPD > 0, Extended Data Fig. 3.3 e–f), whereas those first appearing in the Mid-Pliocene hosted more phylogenetically clustered and/or evolutionarily younger assemblages (RPD < 0, Extended Data Fig. 3.3 e–f). In contrast, more recently emerged climates showed little phylogenetic structuring (RPD ≈ 0, Extended Data Fig. 3.3 e–f). Based on these findings, we also fitted separate GAMs for climatic bins grouped by their Age, excluding groups postdating the Mid-Holocene due to small sample sizes.

Partial residual plots were generated to visualise the independent relationships between each explanatory variable and species richness (Fig. 3.5a–e), species turnover (Fig. 3.5f–j), PD (Fig. 3.5k–o) and RPD (Fig. 3.5p–t). Nearly all diversity metrics increase significantly with both temperature and moisture, except turnover, which shows no relationship with ln MI (Fig. 3.5g). Recent Area is positively correlated with species richness (Fig. 3.5c), turnover (Fig. 3.5h), and PD (Fig. 3.5m), but negatively with RPD (Fig. 3.5r). These patterns are robust across Age groups (Extended Data Figs. 3.4, 3.6 and 3.8).

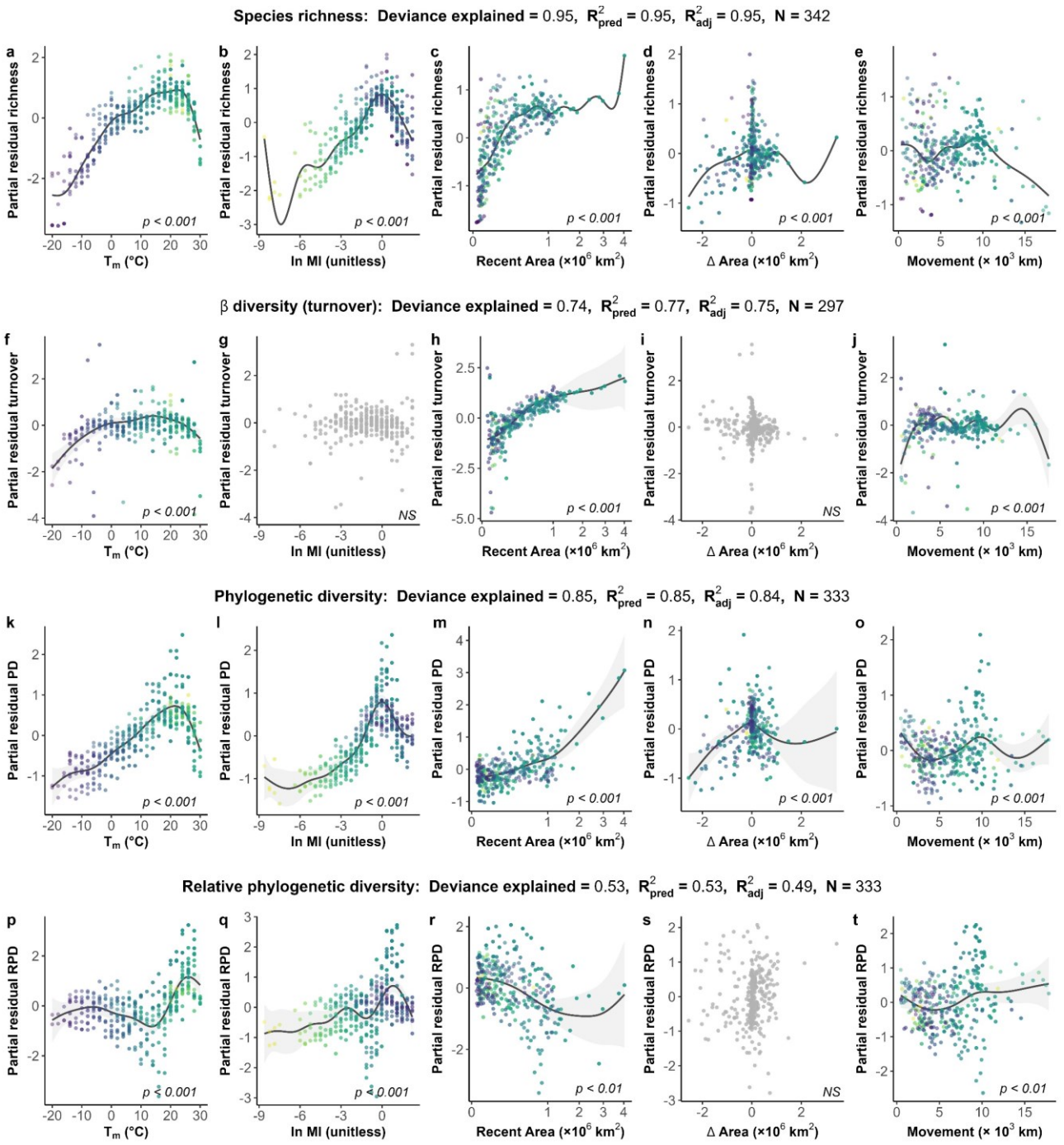
Climatic conditions experiencing geographic contraction (i.e. ΔArea < 0) are associated with lower species richness and PD, while geographic expansion (i.e. ΔArea > 0) does not consistently increase diversity (Fig. 3.5d and n). Instead, peak species richness and PD occur in climates with relatively stable area through time, i.e. with ΔArea close to 0 km^2 (Fig. 3.5d and n). This pattern suggests that long-term geographic stability, rather than expansion *per se*, has been key to fostering plant diversity.

The influence of Movement on plant diversity is more complicated. Moderate geographic shifts (5,000–10,000 km) are associated with higher species richness and PD, whereas both metrics decline when Movement exceeds this range (Fig. 3.5e and o). These negative effects are largely attributable to the warm, humid climates that occurred at high latitudes during the Eocene but now correspond to tropical regions in the present day (Extended Data Fig. 3.4e). In contrast, climates originating in the Mid-Pliocene or LGM showed positive associations between Movement and richness when Movement remained below 10,000 km (Extended Data Figs. 3.6e and 3.8e). Species turnover increases with Movement up to ~5,000 km, then plateaus (Fig. 3.5j), particularly evident for climatic conditions that originated during the Mid-Pliocene and LGM (Extended Data Figs. 3.6j and 3.8j). RPD shows a weak but statistically significant positive correlation with Movement (Fig.

3.5t), although this pattern does not persist within Age groups (Extended Data Figs. 3.4t, 3.6t and 3.8t).

We further quantified the relative importance of each explanatory variable in determining plant diversity (Fig. 3.6). Among the five variables, recent Area and Movement consistently contribute the most explanatory power for species richness, species turnover, and PD, exceeding the combined contribution of the other three predictors. In contrast, RPD is primarily influenced by climatic conditions themselves (T_m and ln MI), while recent Area, ΔArea, and Movement have little effect. This suggests that environmental filtering by present-day climate, rather than historical geographic dynamics of climate, plays a stronger role in shaping phylogenetic structure of plant communities[217].

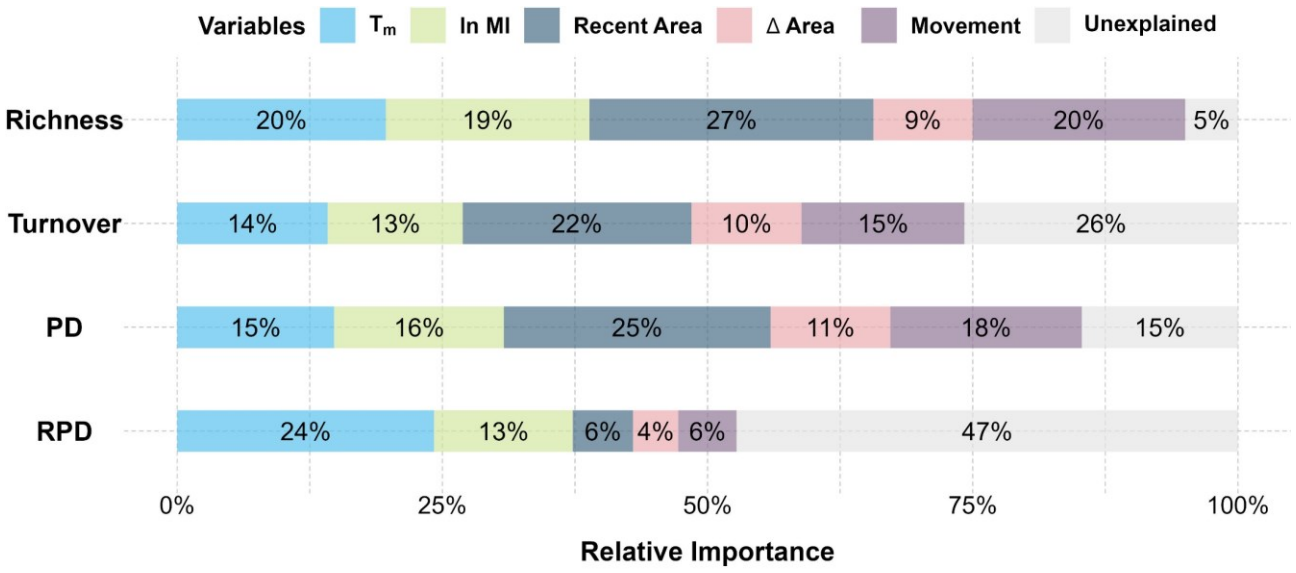
Fig. 3.5: Influence of climate and its deep-time geographic dynamics on plant diversity.



Partial residual plots from Generalised Additive Models (GAMs) showing the independent relationships between four metrics of vascular plant diversity and five explanatory variables. Rows correspond to the four diversity metrics: species richness (**a**–**e**), species turnover (**f**–**j**), phylogenetic diversity (PD, **k**–**o**), and relative phylogenetic diversity (RPD, **p**–**t**). Columns correspond to the five explanatory variables: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent

natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence to the recent period. Black lines represent smooth curves of the GAM regression. Shaded areas indicate 95% confidence intervals, which are not visible in some cases due to narrow uncertainty. Each coloured dot represents a climatic bin, coloured by the gradient used in **Fig. 3.1c–e**. Statistically non-significant terms ($p > 0.05$) are shown as grey dots with no fitted curve. Model performance is reported as deviance explained, predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2).

Fig. 3.6: Relative importance of climate and its deep-time geographic dynamics in explaining plant diversity.



Relative importance of five explanatory variables estimated by partitioning deviance explained by Generalised Additive Models (GAMs). T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence to the recent period. Each stacked bar represents the total deviance in a given diversity metric, with coloured segments

indicating the proportion attributed to each explanatory variable and the grey segment showing unexplained deviance.

3.4 Discussion

Contemporary climates that support the highest plant diversity – primarily warm and humid conditions – tend to share four key characteristics: they are ancient, persistent, geographically extensive, and have undergone moderate geographic shifts over time. Many of these climates had already emerged by the Eocene (*ca*. 50 Ma), and although their geographic locations have changed, they have remained present until today and are now mainly found in tropical regions. Climatic conditions that have consistently occupied large geographic areas over long timescales, rather than those that expanded more recently without long-term persistence, support higher species richness, species turnover, and phylogenetic diversity. Moderate geographic shifts have further promoted diversity, whereas extensive movements are associated with diversity loss. Collectively, our results highlight that the early emergence, persistently large spatial extents, and moderate mobility of climatic conditions jointly contribute to the formation and maintenance of rich and evolutionarily diverse plant communities.

Climates that have remained steadily widespread over deep time likely preserve ancient lineages and reduce extinction risks[32,218]. Because many species retain ancestral ecological requirements (niche conservatism), spatially extensive climates that emerged early and persisted longer provide more time for lineages to diversify within their ancestral environments[193]. Such climatic conditions may act as evolutionary 'museums'[32], conserving both species and phylogenetic diversity. In addition, a large geographic extent facilitates species survival under environmental fluctuations by increasing opportunities for migration rather than extinction. Species-area relationships in climatic space may also reflect capacity rules[121]: climates occupying larger areas tend to support larger populations, which are generally associated with lower extinction rates and potentially higher speciation rates[120]. Palaeoclimate events such as glacial-interglacial cycles likely introduced temporal heterogeneity in non-climatic environmental factors such as soils, topography, and hydrology. These changes may have reshaped species compositions within the same climatic

conditions across time through repeated colonisation, niche differentiation, and community reassembly.

Moderate geographic movement of climate can further promote diversification by fragmenting populations and facilitating species dispersal[218]. As climatic conditions shift spatially, populations may become fragmented across regions, creating opportunities for allopatric speciation[219,220]. At the same time, lineages with sufficient dispersal capacity can track their climatic niches and colonise new regions, while those with intermediate levels of adaptive niche evolution may persist or diversify *in situ* under the novel climatic conditions[32]. By contrast, extensive climatic shifts may have negative effects on biodiversity. Movements exceeding species' dispersal capacities or their adaptive niche evolution rates can lead to local extinction or metapopulation collapse[32,221]. For example, warm and humid climates now widespread in the tropics were distributed at high latitudes during the Eocene. The equatorward displacement of these climates, which occurred ~30–40 million years ago, may have contributed to the extinction of some 'tropical' species and clades previously found at higher latitudes[33,105].

While our study reveals how deep-time geographic dynamics of climate have shaped present-day global patterns of plant diversity, it does not incorporate changes in taxonomic richness and floristic composition through time. Fossil records offer valuable information about past floristic diversity and vegetation structure[205,222]. Understanding how the taxonomic composition of past communities under similar climatic conditions differs from that of modern assemblages helps reveal spatio-temporal dynamics of plant diversity. Such comparisons, capturing turnover in community composition between ancient and modern climate analogues, can help identify which climatic conditions have historically functioned as 'cradles' (promoting high speciation), 'museums' (preserving ancient lineages), or 'graves' (marking high extinction) of plant diversity[32,218].

Current global-scale palaeoclimate reconstructions remain limited by spatial resolution. At the spatial resolution of available deep-time climatic reconstructions, tropical montane climates are often merged with those of adjacent lowland regions. Tropical mountains support exceptionally high plant diversity, while their climatic conditions are often more similar to temperate climates than to lowland tropics. This may inflate species richness in lowland tropical regions at coarse-grained scales, because montane species are effectively counted within surrounding lowland

climatic conditions. These effects highlight the importance of spatial grain in studies linking climate to global biodiversity patterns. Future analyses could address this issue by using finer-grained palaeoclimate reconstructions or by accounting for elevational structure in species distributions. A further consequence of working at coarse spatial resolutions is that major Earth system processes, such as continental drift and mountain uplift, are difficult to assess in detail. Yet long-term climate change and Earth-system dynamics interact over geological time, and both are recognised as important drivers of biodiversity patterns[223,224]. As high-resolution palaeoclimate reconstructions become increasingly available, future work could investigate the geographic dynamics of climate along elevational gradients. Such exploration may clarify how interactions between deep-time climate change and landscape development have shaped montane biodiversity. Nevertheless, our study presents a framework for integrating the spatial and temporal dimensions of climate dynamics: moving beyond conventional assessments of climatic variable variability in geographic space, to a climate-space approach that quantifies the geographic distribution of climatic conditions and their emergence, persistence, and movement over millions of years.

Today, biodiversity loss and climate change are increasingly recognised as 'inseparable challenges' facing our planet[225]. The 16th United Nations Biodiversity Conference of the Parties (COP16) in 2024 reaffirmed the urgency of aligning biodiversity conservation with climate action[226]. Our findings contribute to this goal by identifying the deep-time geographic attributes of climate that are critical for plant diversity conservation, thereby informing implementation of the Kunming-Montreal Global Biodiversity Framework (Targets 1–3, 8 and 21) and the new Global Strategy for Plant Conservation[226,227]. Climates that have remained widespread over deep time represent long-term reservoirs of biodiversity, offering valuable opportunities for restoring and conserving native vegetation. Conservation efforts should also account for both the diversification potential and extinction risks associated with ongoing geographic shifts of climate. Effective conservation must therefore move beyond a static, location-based framework by integrating the spatio-temporal dynamics of climate – tracking the emergence, persistence, extent, and movement of climatic conditions over time – as essential indicators for biodiversity monitoring and planning in a changing world[228].

3.5 Methods

3.5.1 Bioclimatic variables

Two bioclimatic variables were selected for this research: a temperature index (T_m, °C), calculated as the average of mean temperature of the coldest month (MTCO, °C) and the warmest month (MTWA, °C); and a moisture index (MI, unitless), derived from the ratio of mean annual precipitation (MAP, mm) to the mean annual potential evapotranspiration (PET, mm). The natural-log transformed MI (ln MI) was used in the analysis to emphasise differences at the dry end of the moisture range[206]. The variables correspond to the two recognised climatic dimensions that regulate the global geographic distribution of plants[11,14,152,206], with T_m reflecting the latitudinal gradient of temperature[207] and MI (ln MI) representing plant-available moisture[11,14,152,206].

We selected six representative geological periods across the Cenozoic Era: the Eocene (*ca.* 50 Ma), Mid-Pliocene (3.3–3.0 Ma), Last Glacial Maximum (LGM, 22–18 ka), Mid-Holocene (*ca.* 6 ka), pre-industrial period (pre-1850), and a recent period from 1901 to 1989. For each of these periods, the two bioclimatic variables (T_m and ln MI) were calculated based on climatological data of four raw climatic variables: monthly averaged mean daily air temperature (*tas*, °C), mean monthly precipitation (*pr*, mm), monthly accumulated surface downwelling shortwave radiation (*rsds*, W m^{-2}), and elevation (*elev*, m).

We collected the required gridded climate data from various databases for the selected periods (Table A2.2). Climatological simulations of the four raw climatic variables in the Eocene were collected from the Deep-Time Model Intercomparison Project (DeepMIP version 1.0; http://www.deepmip.org)[229]. For the Mid-Pliocene, LGM and Mid-Holocene, simulations of the four variables were derived from ensemble model under the Paleoclimate Modelling Intercomparison Project Phase 4 (PMIP4; https://pmip.lsce.ipsl.fr/)[230]. Pre-industrial simulations of the four raw climatic variables were obtained from the Coupled Model Intercomparison Project Phase 6 (CMIP6)[231]. For the recent period (1901–1989): *tas* and *pr* data were downloaded from the Climatic Research Unit (CRU) Time-Series (TS) database (version 4.09: https://catalogue.ceda.ac.uk/uuid/9cf07e92afaa405da4f40b6733f362d3)[232]; *elev* data derived from

the SRTM elevation data was acquired from WorldClim v.2.1 (https://www.worldclim.org/)[183]; *rsds* data were obtained from CMIP6[231].

We retained the climate model simulations that have shown strong agreement with available observations or proxy-based climate reconstructions for each selected geological period (Table A2.2). The multi-model ensemble mean of these selected simulations was then calculated for the four climatic variables to represent the climatology of each period. Detailed descriptions of the paleoclimate simulations and climate databases used are provided in Table A2.2. We utilised vector maps of land boundaries for the Eocene and Mid-Pliocene epochs from Knowing Earth (http://www.knowing.earth)[233] and raster map of land boundary for other periods from Natural Earth (https://www.naturalearthdata.com) to derive global land-surface climate data. All gridded climate data were aggregated and reprojected to a spatial resolution of 1° using bilinear interpolation, implemented via the 'projectRaster()' function in the *raster* R package[234] for further calculations.

We applied the Simple Process-Led Algorithms for Simulating Habitats (SPLASH) model with the above-mentioned four climatic variables as inputs to calculate the mean monthly potential evapotranspiration (PET, mm) using the 'splash.grid()' function in the R package *rsplash*[184,185]. The two bioclimatic variables – T_m and ln MI – for all geological periods were then calculated as follows:

$$T_m\ (℃) = \frac{\underset{1\le k\le 12}{min}\ tas_k + \underset{1\le k\le 12}{\max}\ tas_k}{2} \quad (1)$$

$$MI\ (\text{unitless}) = \frac{\sum_{k=1}^{12} pr_k}{\sum_{k=1}^{12} PET_k} \quad (2)$$

where tas_k (°C) is monthly average temperature; pr_k (mm) is mean monthly precipitation; PET_k (mm) is mean monthly potential evapotranspiration calculated from the SPLASH model using *tas*, *pr*, *rsds* and *elev* as inputs.

Reconstructed MI values for each geological period were subsequently corrected for the physiological effects of past atmospheric CO_2 concentration, following the methodology proposed by Prentice *et al.* (see ref. [235] for details). This correction is essential because conventional climatic moisture indices do not correctly represent the environment as experienced by plants. These indices do not account for how atmospheric CO_2 concentration modulates plant water-use efficiency via the stomatal regulation of carbon uptake and water loss[235–238]. For instance, during low-CO_2 periods

such as glacial intervals, plants must increase stomatal conductance to acquire sufficient carbon, which increases transpirational water loss[235–238]. The core principle is expressed by the equation[235]:

$$e(T_1, MI_1, c_{a1}) = e(T_0, MI_0, c_{a0}) \quad (3)$$

where e is the ratio of stomatal water loss to carbon uptake; T_1 (°C) is the past growing-season mean temperature; MI_1 (unitless) is the 'true' past value of MI (to be estimated); c_{a1} (ppm) is the past ambient CO_2 concentration; T_0 (°C) is the recent growing-season mean temperature; MI_0 (unitless) is the reconstructed past value of MI; and c_{a0} (ppm) is the recent ambient CO_2 concentration.

We used climatological data for the period 1981–2010 as a reference. For the reference period and the six representative geological periods, we calculated the mean growing-season temperature, defined as the average temperature during days of a year with climatological temperatures above 0 °C. We also reconstructed the MI for all periods using the SPLASH model and Eq. (2). The *tas*, *pr* and *rsds* of the reference period were obtained from CHELSA V2.1 (https://chelsa-climate.org/)[182], at a spatial resolution of 30 arc-seconds (*ca*. 0.01°). Global elevation data for the reference period (*elev*, m) were acquired from WorldClim v.2.1 (https://www.worldclim.org/)[183], also at 30 arc-seconds. Global atmospheric CO_2 concentration data for the reference period (1981–2010) and for the pre-industrial (pre-1850) and recent (1901–1989) periods, were downloaded from NOAA (https://doi.org/10.15138/9N0H-ZH07)[239]. For the Eocene, average atmospheric CO_2 concentration was derived from refs. [240] and [241]. The Mid-Pliocene CO_2 concentration was obtained from refs. [242] and [243], the LGM CO_2 concentration was acquired from ref. [244], and the Mid-Holocene CO_2 concentration was downloaded from NOAA Paleoclimatology Data[245]. All raster data were rescaled to a uniform spatial resolution of 1° using the 'projectRaster()' function in the R package *raster*[234]. We then applied 'corrected_mi()' function in the R package *codos*[246], which implements the equation of Prentice *et al.*[235], to correct MI values for the selected six geological periods. All data manipulations and the following analyses were conducted in R v4.3.2 (https://www.r-project.org/)[247].

3.5.2 Quantification of spatio-temporal climate dynamics

A two-dimensional climatic space was defined using recent T_m and ln MI data from 1901 to 1989. This period was chosen to reflect baseline climatic conditions after industrialisation but before the

rapid anthropogenic influence – such as accelerated greenhouse gas emissions and a step-up in warming rate – that began around 1990[208]. The climatic space was constructed at multiple resolutions by dividing each axis into 30, 40, 50 or 60 equal intervals. Resolutions below 30 intervals yielded insufficient data for robust analysis, while those exceeding 60 intervals provided no substantial increase in sample size. Analyses were performed across all resolutions. Results based on the 30-interval climatic space are presented in the main text, with results from other resolutions detailed in the A2 Supplementary information for Chapter 3.

The climatic space was discretised into a grid of 'climatic bins', each representing a unique recent climatic condition. The geographic distribution of each of these climatic conditions was then identified across all six geological periods. Specifically, for each period, we determined the set of geographic grid cells in the geographic space that correspond to each climatic bin in the climatic space – that is, all geographic locations sharing the same climatic conditions.

To characterise the deep-time geographic dynamics of climate, we defined and calculated five metrics. First, we identified the earliest occurrence of each climatic condition across the six geological periods, referred to as its 'Age'. We then counted the number of periods in which each climatic condition was present across all six periods. This metric was termed 'Persistence'. The geographic extent of each climatic condition within a given period was measured as the total land surface area it occupied, referred to as 'Area' (Fig. A2.9). The remaining two metrics quantified changes in the geographic distribution of each climatic condition between the period of its first occurrence (as defined by the metric 'Age') and the recent period. The first, 'ΔArea' (Fig. A2.10), represents the change in land area occupied by a given climatic condition. Positive values of ΔArea indicate geographic expansion, while negative values indicate contraction. The second metric, 'Movement' (Fig. A2.11), denotes the average geographic shift of a climatic condition between its earliest appearance (Age) and the recent period, calculated using great-circle distances.

3.5.3 Species distribution data

We compiled occurrence records for each vascular plant species based on the accepted species listed in the World Checklist of Vascular Plants (WCVP)[209] from six openly accessible databases: (1) Global Biodiversity Information Facility (GBIF, https://www.gbif.org; accessed in Feb–June 2024)

using the R package *rgbif*[248]; (2) Botanical information and Ecology Network (BIEN version 4.2, https://bien.nceas.ucsb.edu/bien/; accessed in Feb–March 2024) using the R package *BIEN*[249,250]; (3) Atlas of Living Australia (ALA, https://www.ala.org.au/; accessed in Feb–March 2024); (4) RAINBIO, a mega-database of tropical African vascular plants distributions (https://gdauby.github.io/rainbio/index.html; accessed in Feb–March 2024)[251]; (5) Flora of Russia (FoR, https://zenodo.org/records/4061848; accessed in Feb–March 2024)[252]; and also (6) WCVP (accessed in Feb–June 2024)[209]. The scientific names of species across all databases were standardised and aligned to the accepted species names in the WCVP before integration.

We subsequently cleaned the species occurrence data using the R packages *dplyr*[253] and *CoordinateCleaner*[254], removing unreliable and non-native data points according to the following process:

(1) Records with erroneous coordinates were excluded, including: no latitude or longitude record, coordinates equal to 0, identical latitude and longitude values, or coordinates falling outside the valid geographic range.

(2) Records with unsuitable data sources, particularly fossil records, were removed, retaining only data points where the 'basis of record' was an observation or preserved specimen.

(3) Records flagged as 'managed', 'introduced', 'invasive' or 'naturalised' in the 'establishment means' field were discarded. Additionally, non-native data points were removed according to the native distribution ranges provided by the WCVP.

(4) Records dated during or before 1945 were excluded, as these older records are generally considered less reliable[255].

(5) Records with 'coordinate uncertainty' greater than 100 km or 'coordinate precision' less than 0.01 were removed.

(6) Records with suspicious 'individual counts' were discarded, retaining records with $0 <$ 'individual counts' ≤ 100.

(7) The automatic cleaning algorithm by the 'clean_coordinates()' function in the *CoordinateCleaner* R package[254] was run to flag and delete error data points, including: sea coordinates, zero coordinates, mismatched countries, coordinates assigned to country and province

centroids, coordinates within city areas, outlier coordinates, and coordinates assigned to biodiversity institutions.

Occurrence records for the same species across different databases were merged. After data cleaning, we retained native occurrence records for 350,864 unique species, representing over 98% of the 357,347 accepted species listed in the WCVP (accessed on 19 February 2024).

3.5.4 Vascular plant diversity

Following the IUCN Red List of Categories and Criteria[256–258], we calculated the extent of occurrence (EOO) for species with ≥3 occurrence records by constructing the minimum convex polygon encompassing all recorded locations. Portions of the polygons falling outside the native distribution range, as provided by the WCVP, were cropped and excluded. For species with < 3 occurrence records (2,527 species in total), for which EEO could not be calculated, we measured the area of occupancy (AOO), defined as all geographic grid cells at a specific spatial resolution that contain the species' occurrence records[256–258]. In this study, occurrence data were gridded at 1° spatial resolution to estimate AOO. We then created presence-absence raster data at a 1° resolution from the EOO or AOO of each species, with a value of 1 denoting presence (i.e., EOO or AOO) and 0 denoting absence. To visualise global richness patterns in geographic space (Fig. 3.2), native species distributions were mapped using the Behrmann equal-area projection at ~110 km resolution (~1° resolution), thereby facilitating comparison across latitudes. By stacking the resulting presence-absence rasters of all species, we produced a global map of land vascular plant species richness. Regions where species richness exceeded the global 90th percentile were defined as plant diversity hotspots[210] (Fig. A2.3).

For the core analyses, species distributions were then mapped into climatic rather than geographic space. Based on the EOO or AOO of each species, we identified the climatic conditions in which each species occurs (representing climatic coordinates) and generated corresponding presence-absence rasters in climatic space. At this step, native species distributions were retained at the original 1° resolution to match the resolution of the climatic rasters and to ensure consistent spatial alignment between species distributions and climatic variables. By stacking the presence-absence rasters in climatic space, we counted the number of species present in each

climatic bin, thereby mapping plant species richness in climatic space (Fig. A2.4). Consequently, the latitudinal variation in the geographic area of 1° cells does not influence the estimated relationships between species richness and climate.

The species-level presence-absence data in climatic space were also used to compute multi-site β-diversity within each climatic bin by 'beta.multi()' function in the *betapart* R package[259]. This function calculates the turnover component of β-diversity (Fig. A2.5) based on the Simpson dissimilarity index, which reflects the degree of species composition turnover among geographic grid cells sharing the same climatic conditions.

To evaluate phylogenetic diversity in climatic space, we constructed a time-calibrated phylogenetic tree for all 350,864 vascular plant species. The initial tree was generated using the 'phylo.maker' function in the *V.PhyloMaker2* R package[260]. The backbone for this tree was the GBOTB.extended megatree, a time-calibrated phylogeny of vascular plants that integrates the Smith and Brown's phylogeny of seed plants[261] and Zanne *et al.*'s plant phylogeny including pteridophyte species[262]. Species absent from the megatree were grafted onto this backbone using a hierarchical approach that places species at the basal node of their genus or, if the genus is missing, at a calculated position within their family[260,263]. This resulting phylogeny, comprising 350,705 successfully placed tips, was then completed to cover all 350,864 species using the 'phylo.maker' function from the *U.PhyloMaker* R package[263].

Based on the comprehensive phylogeny of vascular plants, two phylogenetic diversity metrics were calculated. First, we calculated Faith's phylogenetic diversity (PD, Ma)[100] for each climatic bin, defined as the total branch length of the minimum phylogenetic subtree connecting all species present in that bin (Fig. A2.6). Higher PD values thus represent greater accumulation of evolutionary history within the climatic bin. To account for the strong, intrinsic correlation between PD and species richness, we also calculated relative phylogenetic diversity (RPD, Ma)[264,265]. RPD was defined as the residuals from an ordinary least squares (OLS) regression of PD against species richness across all climatic bins (Fig. A2.7). Higher positive RPD values therefore indicate that species in a climatic bin are more phylogenetically overdispersed than expected given their richness (i.e. species are more distantly related and/or more evolutionarily ancient than expected), whereas

negative values indicate phylogenetic clustering (i.e. more closely related and/or evolutionarily younger than expected).

3.5.5 Statistical analysis

We first assessed whether recent plant diversity in climatic bins was associated with the earliest occurrence (Age) and temporal persistence (Persistence) of climatic conditions. These relationships were tested using the Kruskal-Wallis test, implemented with the 'kruskal.test()' function in the R *stats* package[247].

The effects of climate and its deep-time geographic dynamics on plant diversity were characterised and visualised using Generalised Additive Models (GAMs) fitted in climatic space. All GAMs excluded interaction terms and employed restricted maximum likelihood (RMEL) to estimate smoothing parameters. Model fitting was performed using the 'gam()' function in the R package *mgcv*[266,267]. Concurvity among explanatory variables was examined prior to model fitting and found to be relatively low (Fig. A2.12 and Tables A2.3–A2.9, A2.13–A2.16 and A2.20–A2.31).

To estimate the relative contribution of each explanatory variable, we partitioned the total explained deviance into additive components following an approach conceptually analogous to the Lindeman-Merenda-Gold (LMG) method[268]. The relative importance of each variable was calculated as the average of its individual contributions across all possible orderings of variable inclusion.

To test whether climates that were more geographically extensive in the past support greater present-day species richness, we fitted Poisson-distributed GAMs separately for each of the six selected geological periods. In these models, species richness was the response variable, and three explanatory variables were included: two bioclimatic variables (T_m and ln MI), and the geographic extent of each climatic condition (Area) during the corresponding period. In addition, we calculated the mean Area across all periods from the emergence of each climatic condition onward and included it in a separate model. We created contour plots to visualise the relationship between species richness and three explanatory variables. The response surfaces of species richness resulting from the Poisson GAMs were presented as contours in the three-variable space in the form of two-dimensional slices. Convex hulls generated by 'convhulln()' function in the R package

geometry[189] were applied to avoid representing parts of the fitted surface that are not constrained by data.

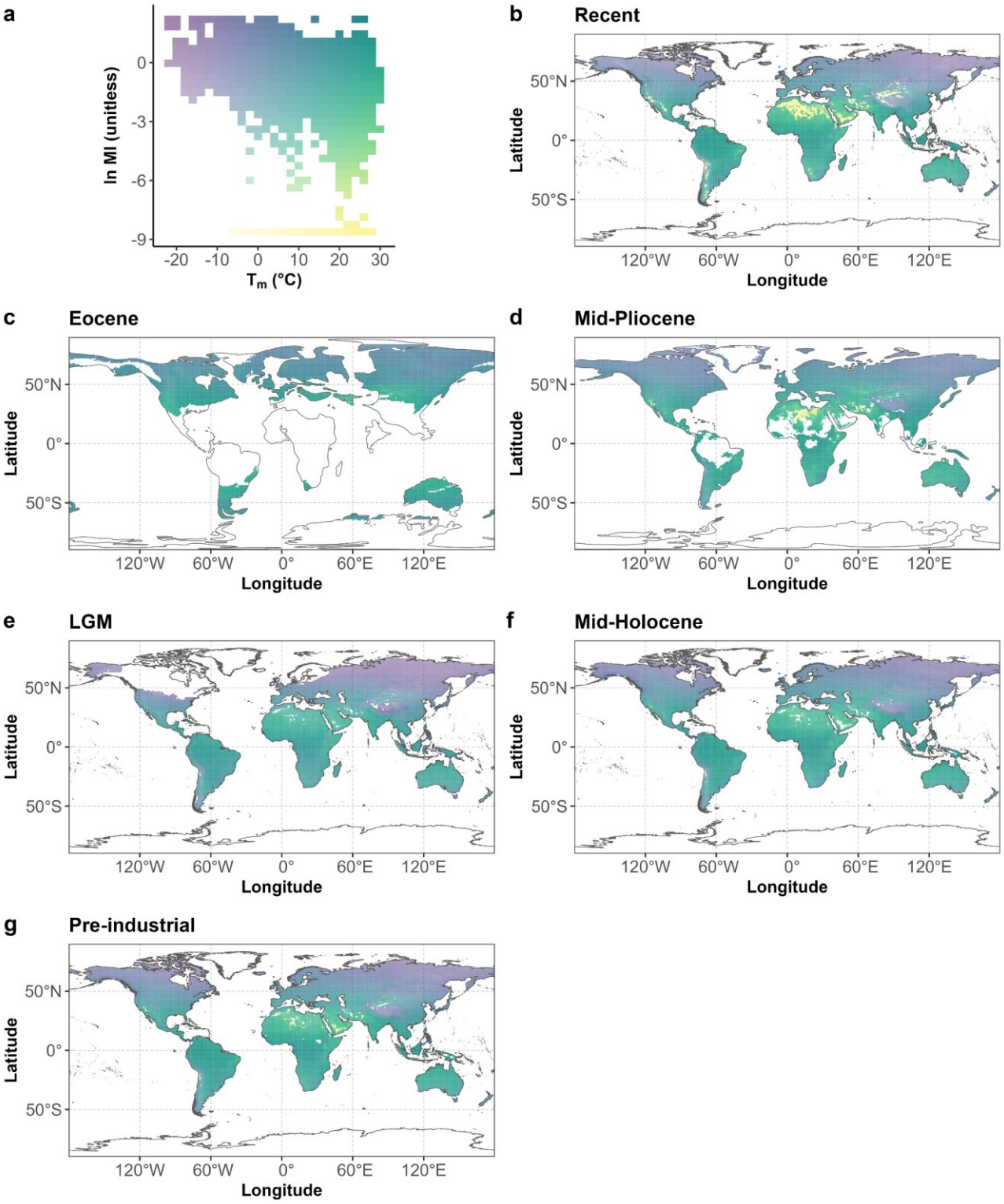
To characterise the relationship between plant diversity and the persistence and shifts of climate's geographic extent, we fitted a second set of GAMs. Response variables included species richness, species turnover, and two phylogenetic diversity metrics. Explanatory variables included T_m, ln MI, and Area during the recent period, as well as the ΔArea and Movement of each climatic condition from its first appearance to the recent period. Poisson GAMs were used for species richness, while beta-distributed GAMs with a logit link were used for species turnover, which is ranged between 0 and 1. To test the consistency of these relationships across climate Age groups, we grouped climatic bins by Age and fitted the same GAM structure within each group. Groups postdating the Mid-Holocene were excluded due to insufficient sample size. Partial residual plots were used to visualise the independent relationships between each explanatory variable and each diversity metric. Relative variable importance was calculated as described above.

To test for robustness across modelling approaches, we also fitted Generalised Linear Models (GLMs) with the same response and explanatory variables as the GAMs (Tables A2.11, A2.12, A2.18 and A2.19). Additionally, we fitted quadratic GLMs incorporating second-order polynomial terms to account for potential non-linear effects. All GLMs were implemented using the 'glm()' function in the *stats* R package[247].

We assessed potential spatial autocorrelation in our analyses, which typically leads to an overestimation of the significance and coefficient estimates of explanatory variables, thereby introducing bias. We performed Moran's *I* test on the residuals of all models using the 'moran.test()' function in the *spdep* R package[269]. All *Moran's I* values were low and/or not statistically significant (Tables A2.32–A2.36), suggesting that model results are robust to spatial structure in the data.

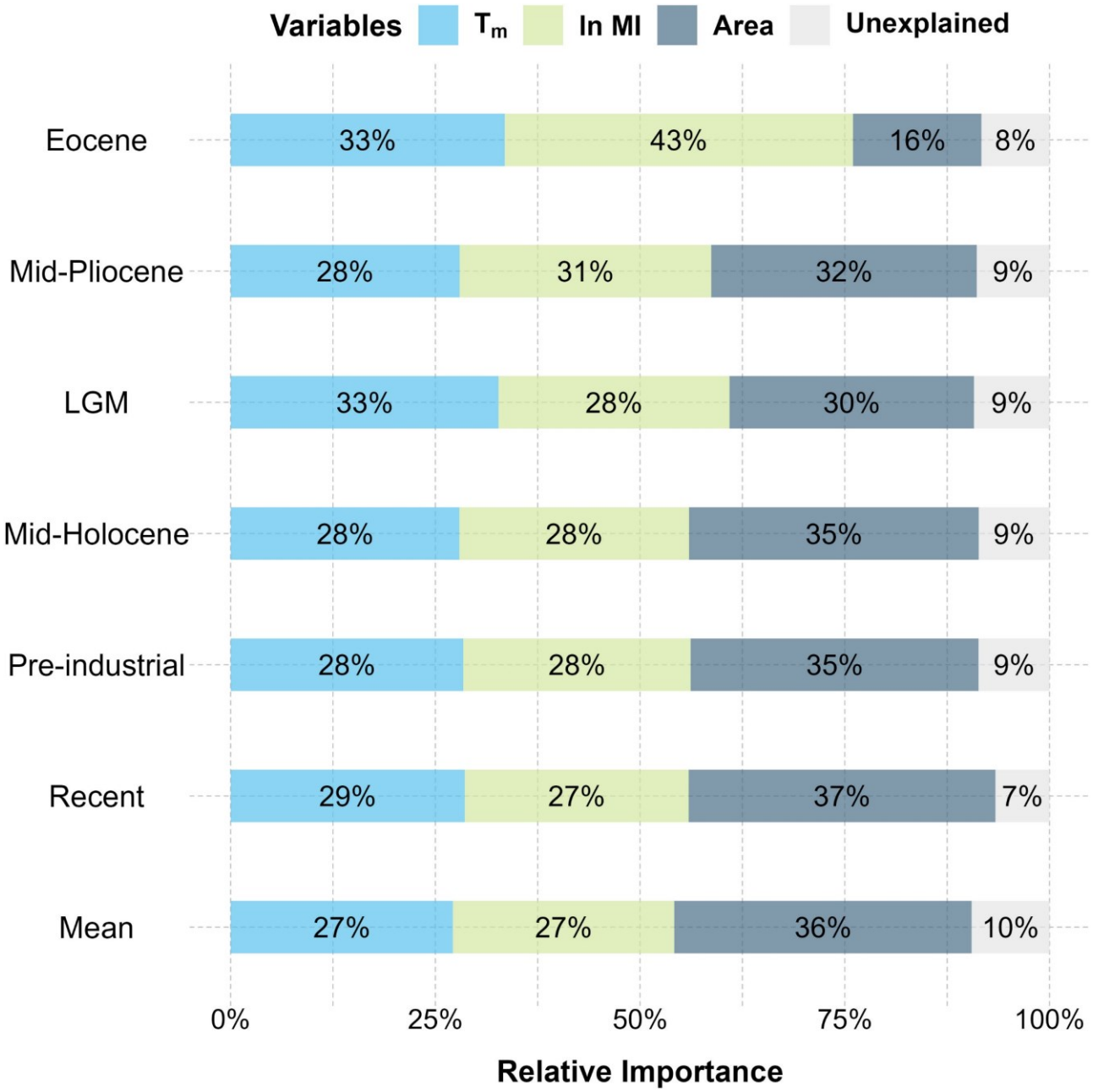
3.6 Extended data figures and tables

Extended Data Fig. 3.1 Historical geography of recent climates.



a, The two-dimensional climatic space defined by the recent (1901–1989) climatic conditions (T_m and ln MI), partitioned into 30 equal intervals along each axis. Each coloured grid cell ('climatic bin') represents a unique climatic condition (405 cells in total). Alternative colour gradient schemes are provided in Figs. A2.1 and A2.2. **b**–**g**, Geographic locations where those recent climatic conditions existed during the recent period (1901–1989, **b**), Eocene (*ca.* 50 Ma, **c**), Mid-Pliocene (3.3–3.0 Ma, **d**), Last Glacial Maximum (LGM, 22–18 ka, **e**), Mid-Holocene (*ca.* 6 ka, **f**) and pre-industrial period (pre-1850, **g**). Uncoloured areas denote climatic conditions absent during this period. Colours correspond to those in **a**, with identical colours indicating the same climatic conditions. All maps are shown at a 1° spatial resolution.

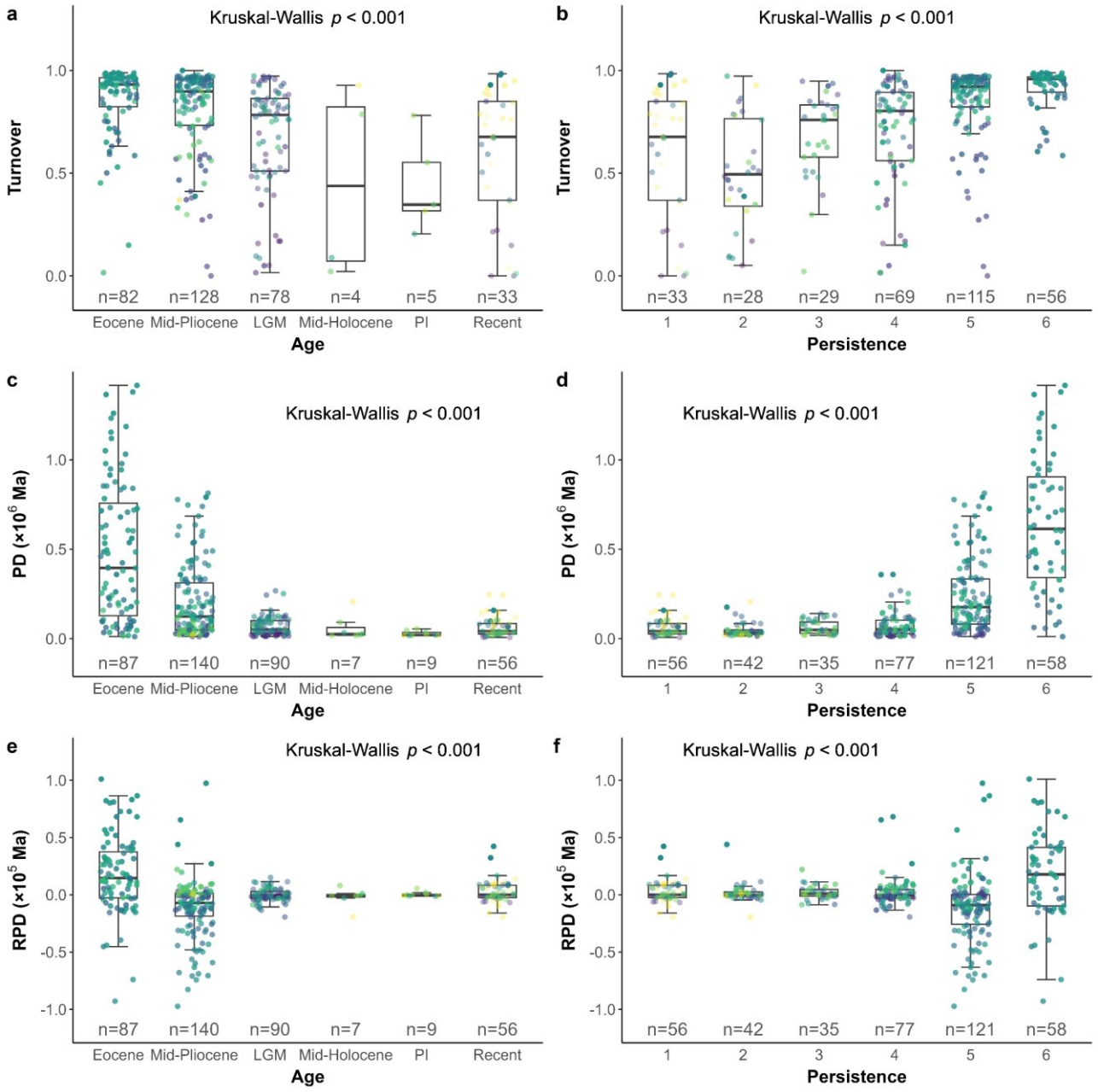
Extended Data Fig. 3.2 Relative importance of climate and its historical geographic extent in explaining plant species richness.



Relative importance of three explanatory variables estimated by partitioning deviance explained by Generalised Additive Models (GAMs). T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Area (km^2), total land surface area occupied by each climatic condition during the corresponding geological period. Mean Area since the first appearance of each climatic condition across all selected geological periods was fitted in a separate model. Each stacked bar represents the total deviance in species richness, with coloured

segments indicating the proportion attributed to each explanatory variable and the grey segment showing unexplained deviance.

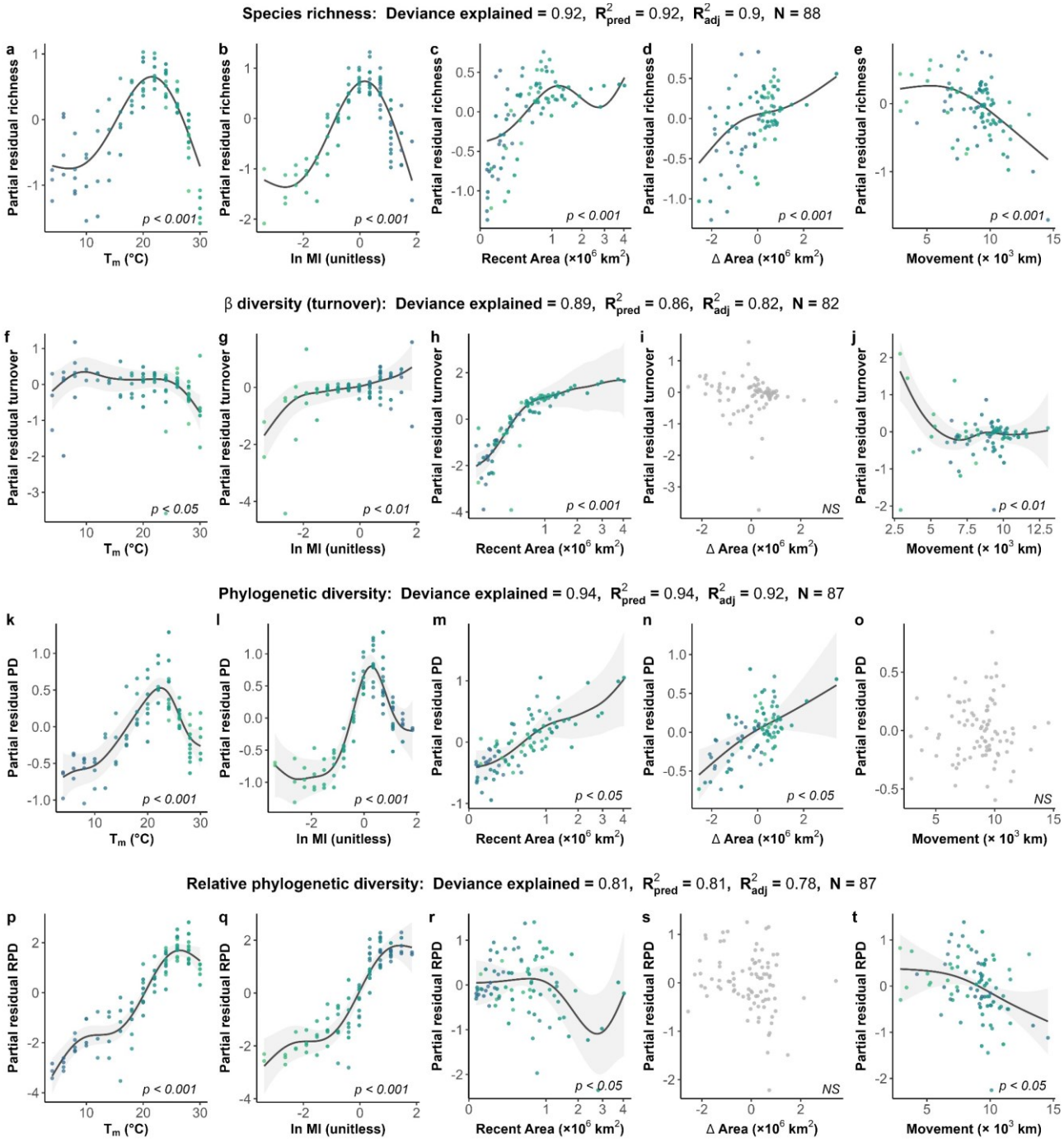
Extended Data Fig. 3.3 Relationships between recent plant diversity and age or persistence of climate.



a–**b**, Relationships between species turnover and Age (**a**) or Persistence (**b**) of climatic conditions. **c**–**d**, Relationships between Faith's phylogenetic diversity (PD) and Age (**c**) or Persistence (**d**) of climatic conditions. **e**–**f**, Relationships between relative phylogenetic diversity (RPD) and Age (**e**) or Persistence (**f**) of climatic conditions. *p*-values were assessed using Kruskal-Wallis tests. Each coloured dot represents a climatic bin, with colours following the gradient used in **Fig. 3.1c–e**; horizontal lines and boxes indicate medians and interquartile ranges. Sample sizes per group are

shown below the boxes. T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index.

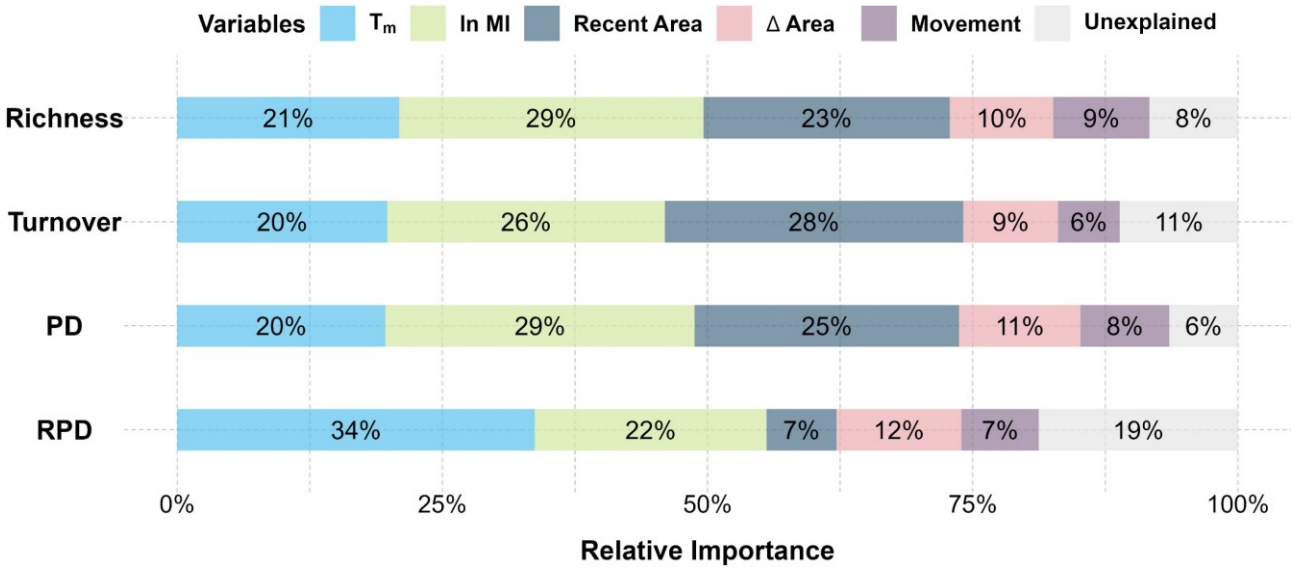
Extended Data Fig. 3.4 Influence of climatic conditions first occurring in the Eocene (*ca*. 50 Ma) and their deep-time geographic dynamics on plant diversity.



Partial residual plots from Generalised Additive Models (GAMs) showing the independent relationships between four metrics of vascular plant diversity and five explanatory variables. Rows correspond to the four diversity metrics: species richness (**a**–**e**), species turnover (**f**–**j**), phylogenetic diversity (PD, **k**–**o**), and relative phylogenetic diversity (RPD, **p**–**t**). Columns correspond to the five

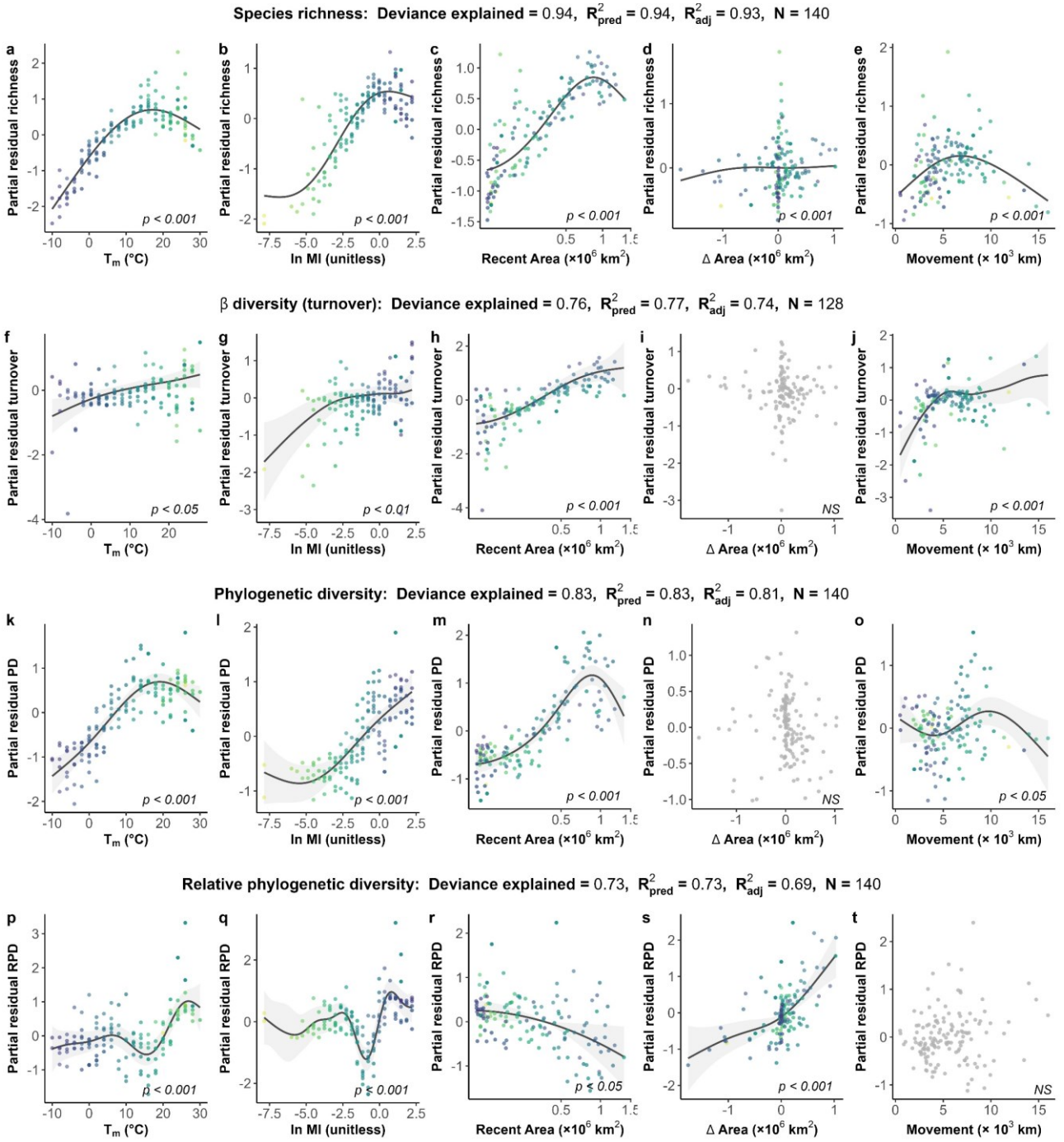
explanatory variables: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence during the Eocene to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence during the Eocene to the recent period. Black lines represent smooth curves of the GAM regression. Shaded areas indicate 95% confidence intervals, which are not visible in some cases due to narrow uncertainty. Each coloured dot represents a climatic bin, coloured by the gradient used in **Fig. 3.1c–e**. Statistically non-significant terms ($p > 0.05$) are shown as grey dots with no fitted curve. Model performance is reported as deviance explained, predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2).

Extended Data Fig. 3.5 Relative importance of climate climatic conditions first occurring in the Eocene (*ca.* 50 Ma) and their deep-time geographic dynamics in explaining plant diversity.



Relative importance of five explanatory variables estimated by partitioning deviance explained by Generalised Additive Models (GAMs). T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence during the Eocene to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence during the Eocene to the recent period. Each stacked bar represents the total deviance in a given diversity metric, with coloured segments indicating the proportion attributed to each explanatory variable and the grey segment showing unexplained deviance.

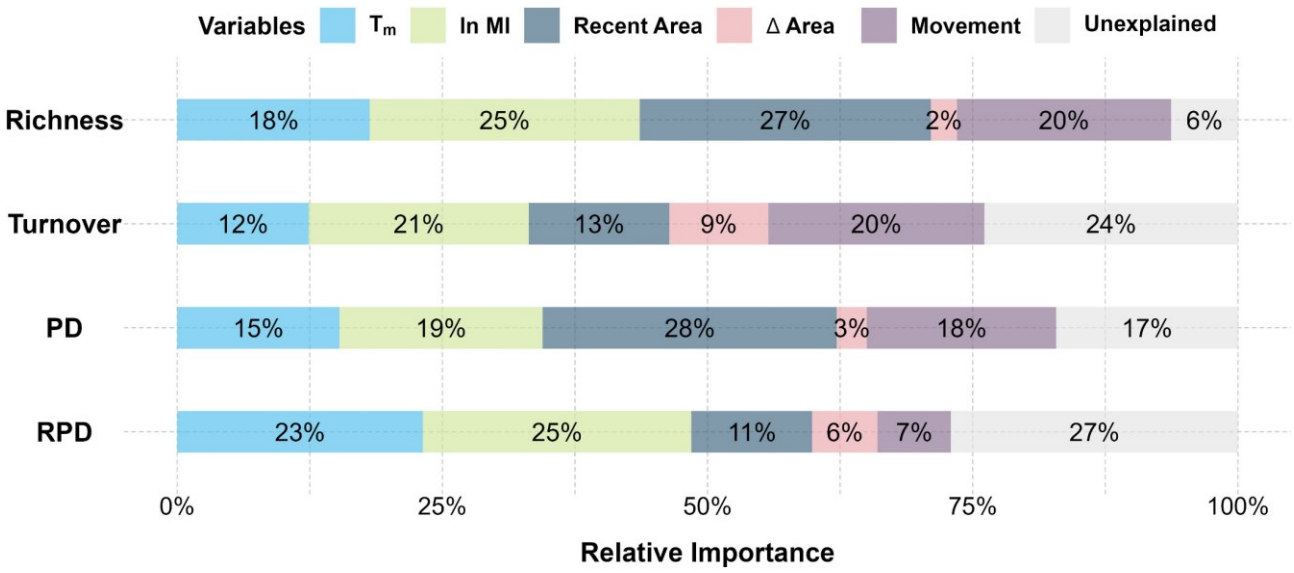
Extended Data Fig. 3.6 Influence of climatic conditions first occurring in the Mid-Pliocene (3.3–3.0 Ma) and their deep-time geographic dynamics on plant diversity.



Partial residual plots from Generalised Additive Models (GAMs) showing the independent relationships between four metrics of vascular plant diversity and five explanatory variables. Rows correspond to the four diversity metrics: species richness (**a**–**e**), species turnover (**f**–**j**), phylogenetic diversity (PD, **k**–**o**), and relative phylogenetic diversity (RPD, **p**–**t**). Columns correspond to the five

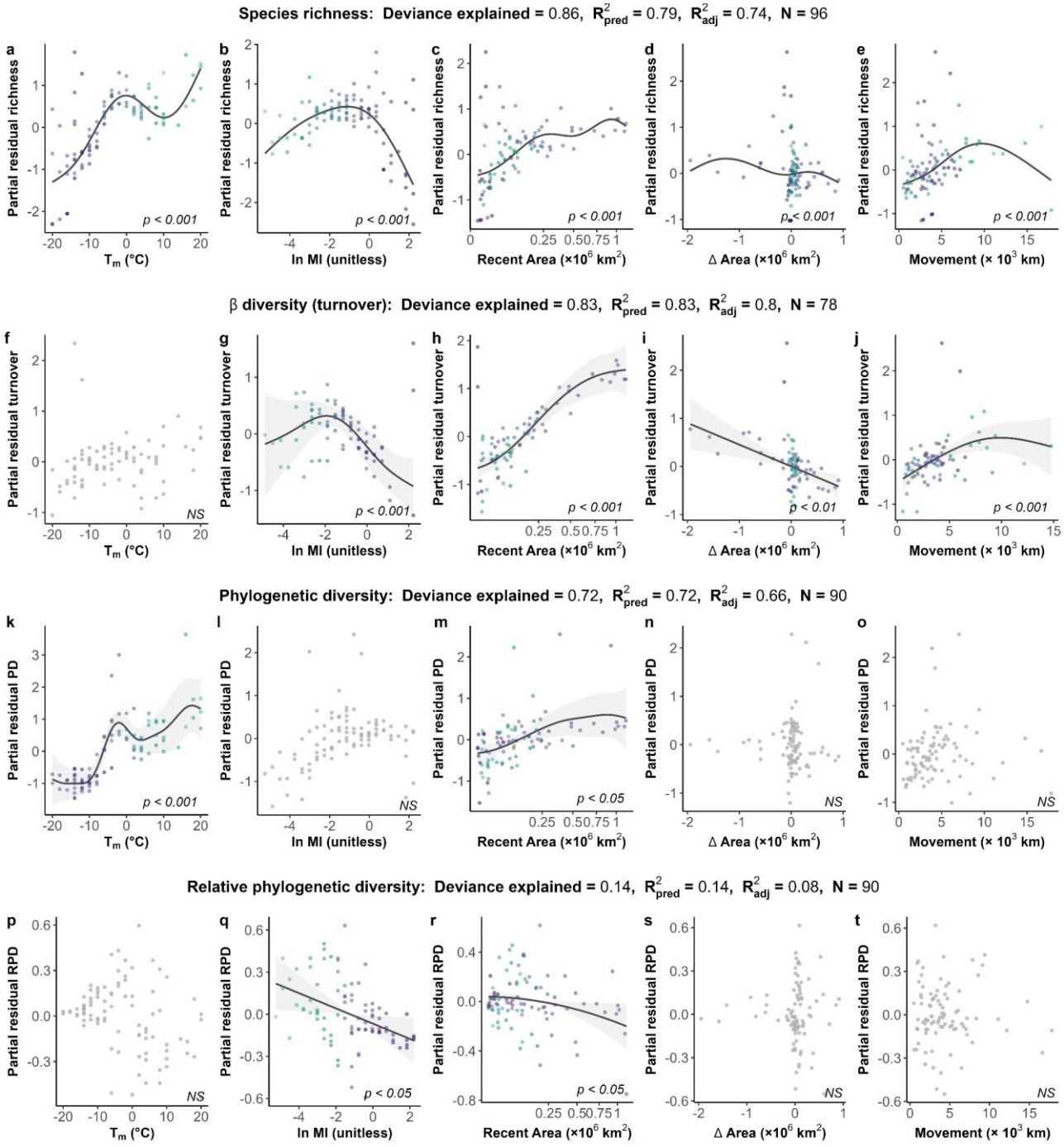
explanatory variables: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence during the Mid-Pliocene to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence during the Mid-Pliocene to the recent period. Black lines represent smooth curves of the GAM regression. Shaded areas indicate 95% confidence intervals, which are not visible in some cases due to narrow uncertainty. Each coloured dot represents a climatic bin, coloured by the gradient used in **Fig. 3.1c–e**. Statistically non-significant terms ($p > 0.05$) are shown as grey dots with no fitted curve. Model performance is reported as deviance explained, predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2).

Extended Data Fig. 3.7 Relative importance of climate climatic conditions first occurring in the Mid-Pliocene (3.3–3.0 Ma) and their deep-time geographic dynamics in explaining plant diversity.



Relative importance of five explanatory variables estimated by partitioning deviance explained by Generalised Additive Models (GAMs). T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence during the Mid-Pliocene to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence during the Mid-Pliocene to the recent period. Each stacked bar represents the total deviance in a given diversity metric, with coloured segments indicating the proportion attributed to each explanatory variable and the grey segment showing unexplained deviance.

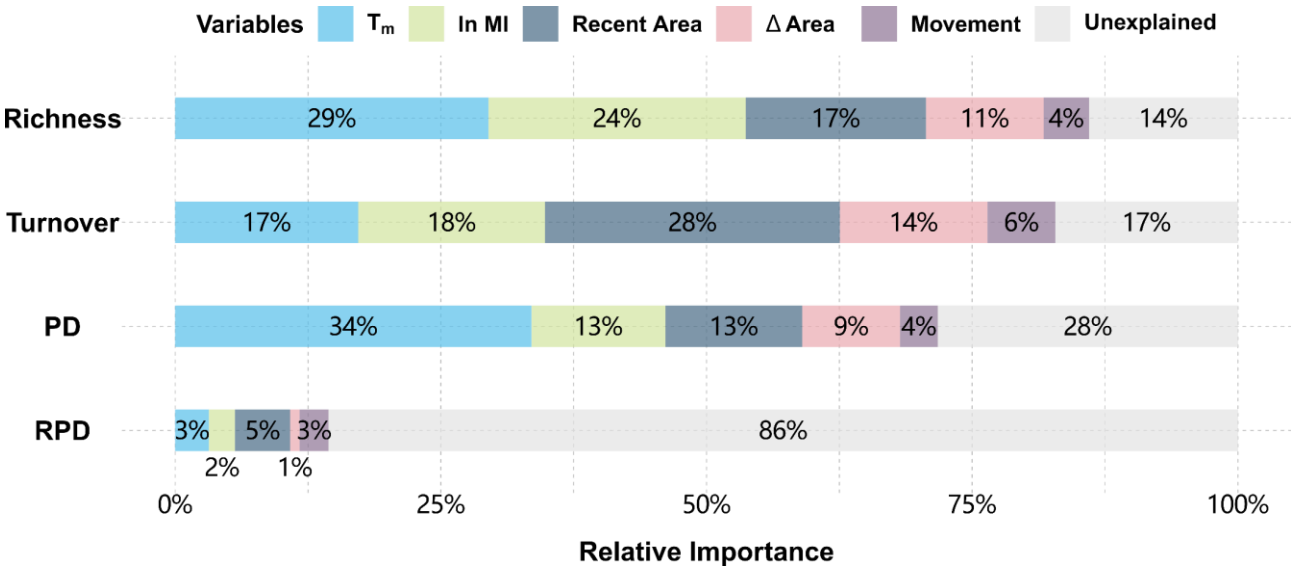
Extended Data Fig. 3.8 Influence of climatic conditions first occurring in the Last Glacial Maximum (LGM, 22–18 ka) and their deep-time geographic dynamics on plant diversity.



Partial residual plots from Generalised Additive Models (GAMs) showing the independent relationships between four metrics of vascular plant diversity and five explanatory variables. Rows correspond to the four diversity metrics: species richness (**a**–**e**), species turnover (**f**–**j**), phylogenetic diversity (PD, **k**–**o**), and relative phylogenetic diversity (RPD, **p**–**t**). Columns correspond to the five

explanatory variables: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence during the LGM to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence during the LGM to the recent period. Black lines represent smooth curves of the GAM regression. Shaded areas indicate 95% confidence intervals, which are not visible in some cases due to narrow uncertainty. Each coloured dot represents a climatic bin, coloured by the gradient used in **Fig. 3.1c–e**. Statistically non-significant terms ($p > 0.05$) are shown as grey dots with no fitted curve. Model performance is reported as deviance explained, predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2).

Extended Data Fig. 3.9 Relative importance of climate climatic conditions first occurring in the Last Glacial Maximum (LGM, 22–18 ka) and their deep-time geographic dynamics in explaining plant diversity.



Relative importance of five explanatory variables estimated by partitioning deviance explained by Generalised Additive Models (GAMs). T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence during the LGM to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence during the LGM to the recent period. Each stacked bar represents the total deviance in a given diversity metric, with coloured segments indicating the proportion attributed to each explanatory variable and the grey segment showing unexplained deviance.

Chapter 4: Global biogeography of C4 plants through evolutionary time

4.1 Abstract

A comprehensive global biogeography of the C4 photosynthetic pathway across its evolutionary history does not exist due to incomplete C4 species records and geographically sparse and temporally fragmented palaeoecological evidence. Here, we address this gap by integrating the most comprehensive global C4 plant diversity data to date with eco-evolutionary optimality (EEO) model simulations of their biogeographic dynamics over the past 50 million years. Our analyses, conducted in both geographic and climatic space, reveal that C4 diversity is organised into two distinct biogeographic patterns: (1) hot-arid 'refugia' characterised by persistently high C4 abundance since the Mid-Pliocene, low species richness but high C4 diversity fraction, and high, positive relative phylogenetic diversity (RPD); and (2) 'radiation centres' with high species richness but neutral or negative RPD. We show these patterns are a legacy of historical dynamics. Declining atmospheric CO_2 was the primary driver for the widespread expansion of C4 plants, while rising CO_2 since the Last Glacial Maximum has driven their contraction in tropical regions and a retreat to their core hot-arid niche. These findings provide a deep-time context for C4 evolution, important for understanding ecosystem function and biodiversity responses to an ever-changing world.

4.2 Main

Terrestrial plants fix atmospheric CO_2 via diverse photosynthetic pathways. The C4 pathway employs a CO_2-concentrating mechanism (CCM) to elevate CO_2 levels around the carboxylating enzyme Rubisco[38,122,123]. The CCM reduces carbon loss and energy waste caused by photorespiration, thereby enhancing photosynthetic efficiency, particularly under low ambient CO_2 conditions[38,41,122,123]. This mechanism also confers high water-use and nitrogen-use efficiency, enabling C4 species to thrive in warm, arid, and nutrient-poor environments[40,41]. Consequently,

despite comprising only ~3% of land plant species[41,128], C4 plants dominate tropical and subtropical grasslands and savannas[38,42,127] and contribute about 20% of global gross primary production[43–47]. The importance of C4 plants also extends to human societies, as they include major food and energy crops (e.g. maize, sorghum, sugarcane, and *Miscanthus*) as well as many widespread alien invasive species[38,41].

The rise of C4 photosynthesis is generally interpreted as an evolutionary response to major environmental shifts over geological timescales. For most of Earth's history, elevated atmospheric CO_2 levels limited photorespiration, so the C3 pathway was sufficient for plant carbon fixation[270]. However, a long-term decline in atmospheric CO_2 since the Eocene (~50 million years ago) created a selective pressure favouring the C4 pathway[38,125,126]. Fossil evidence, isotopic records and molecular clocks suggest that the earliest C4 lineages may have originated during the Oligocene (*ca*. 24–35 million years ago), with a marked diversification occurring throughout the Miocene and Pliocene[38,40]. These periods coincided with declining CO_2 levels and increasing aridity, conditions that likely facilitated the emergence and ecological expansion of early C4 lineages[38,40]. A more recent C4 diversification phrase likely occurred during the Pleistocene, characterised by strong glacial-interglacial climate oscillations and periodic exceptionally low atmospheric CO_2 levels (~180 ppm), increasing the ecological competitiveness of C4 species[38,40,271]. Entering the Holocene, atmospheric CO_2 concentrations began to gradually increase, while anthropogenic influences, including agriculture and industrialisation, may have further reshaped the competitive balance between C3 and C4 plants[46,272].

The biogeography of C4 plants is governed by their physiological responses to key environmental factors. While atmospheric CO_2 concentration is a primary determinant of C4 evolution and competitiveness[38,273,274], other factors are also in play. Temperature affects Rubisco performance[44,274]. Higher temperatures increase the affinity of Rubisco for O_2 and reduce CO_2 availability within C3 leaf cells, further amplifying the advantages of the C4 pathway[44,274]. Water availability influences resource-use strategies[41,275]. Due to their higher water-use efficiency, C4 plants are often better adapted to arid or seasonally dry environments[41,275]. Light intensity modulates the energetic trade-offs[41,274]. C4 photosynthesis requires greater energy investment than the C3 pathway, making C4 species more responsive to high irradiance but more constrained under

low-light or shaded conditions due to energy limitation[41,274]. The interplay of these factors has likely driven the adaptive success and dynamic distribution of C4 plants across space and time.

The relative advantage of C4 photosynthesis has continuously shifted across space and time in response to ongoing environmental change. In the context of rapid global change during the Anthropocene, understanding the responses of this key functional group is critical for predicting future ecosystem structure and function[41]. Research to date has advanced along several lines. Biodiversity studies have focused on continuously updating C4 species lists and mapping their biogeographical hotspots[128,276–278]. Models predicting modern C4 distribution have evolved from early 'crossover temperature' approaches defining climates favourable to C4 plants[44,279], to recent integrations of optimality theory and remote sensing that reveal dynamic changes in C4 abundance[46]. On longer timescales, stable carbon isotope analyses of fossil herbivore tooth enamel have been used to reconstruct past C4 distribution, revealing a global expansion of C4-dominated biomes during the late Miocene[36,42].

Although previous studies have advanced our understanding of C4 plant biogeography, key gaps remain. First, a global database of C4 species encompassing all major evolutionary lineages and their distributions has been lacking. Second, our understanding of C4 historical dynamics is spatially and temporally incomplete. Large-scale vegetation models primarily focus on contemporary patterns, while palaeoecological records, such as fossil and isotopic evidence, are geographically sparse and temporally fragmented. As a result, the long-term environmental drivers of C4 evolution and expansion remain poorly characterised. These knowledge gaps limit our ability to understand the origins of contemporary C4 plant biogeography from a deep-time perspective. To address these limitations, vegetation models based on eco-evolutionary optimality (EEO) theory offer a promising approach[47,280–284]. These models assume that plants optimise the trade-off between carbon gain and water loss to maximise photosynthetic production[280,285]. With minimal input requirements and broad applicability, EEO models can simulate C3-C4 competition across varying environmental scenarios[47]. While EEO models have performed well in reconstructing global vegetation dynamics during the late Quaternary[47,286], their potential to simulate C4 plant biogeography over evolutionary history remains unexplored.

Here, we integrate an EEO-based model with deep-time paleoclimate simulations and a newly compiled database of global C4 plant diversity to explore global biogeography of C4 plants across geological timescales. We conduct our analyses in parallel across both geographic space and a climatic space defined by temperature and plant-available moisture to provide complementary perspectives. By compiling distributional records for over 7,000 recognised C4 species, we identify contemporary global hotspots of species richness and phylogenetic diversity. We then focus on six representative time periods corresponding to key stages of C4 evolution and expansion: the Eocene (*ca*. 50 Ma), Mid-Pliocene (3.3–3.0 Ma), Last Glacial Maximum (LGM, 22–18 ka), Mid-Holocene (*ca*. 6 ka), pre-industrial (pre-1850), and recent (1901–1989) periods. For the geohistorical periods, environmental conditions are represented by outputs from multiple climate model simulations, whereas for the recent period, we use instrumental climate data. The model simulations provide the spatially contiguous global datasets necessary for our analysis, complementing insights from the more spatially sparse proxy-based fossil record[287]. Using environmental variables as inputs, we employ the EEO models to simulate spatial patterns and temporal shifts in C4 abundance fraction to investigate the historical origins of modern diversity centres. Finally, we quantify the relative influence of four environmental drivers: CO_2, temperature, moisture, and light on C4 fraction across time.

4.3 Results

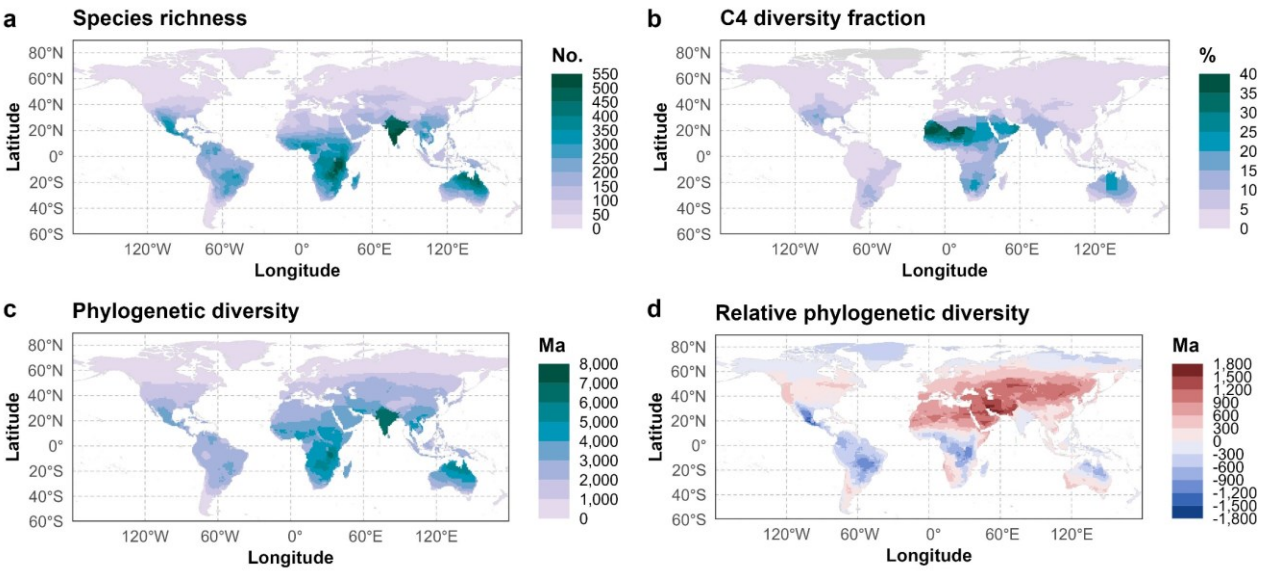
4.3.1 Contemporary global biogeography of C4 plants

We identified 7,156 land plant species that apply the C4 photosynthetic pathway. Based on the native geographic ranges of these species, we mapped global patterns of C4 species richness at 1° spatial resolution (Fig. 4.1a). To assess the contribution of C4 plants to local floras, we calculated the C4 diversity fraction by dividing C4 species richness by total vascular plant richness (from ref. [288], see Chapter 3; Fig. 4.1b). To evaluate phylogenetic diversity, we constructed a time-calibrated phylogenetic tree for all 7,156 species and estimated regional values using Faith's phylogenetic diversity (PD, Ma) metricc (Fig. 4.1c). As PD is tightly correlated with species richness, we further

computed relative phylogenetic diversity (RPD) – the residuals from a PD-richness regression – which captures deviations from the expected PD given the number of species (Fig. 4.1d).

Hotspots of C4 species richness are concentrated in tropical and subtropical regions, notably Mesoamerica, central South America, central and eastern Africa (excluding the Horn of Africa), South Asia, and northern Australia (Fig. 4.1a). The C4 diversity fraction peaks in a different set of regions, including the Sahel, southern Africa, the Arabian Peninsula, and the Northern Territory of Australia, where overall C4 richness is relatively low (Fig. 4.1a) but C4 species comprise more than 25% of the flora (Fig. 4.1b). While spatial patterns of PD (Fig. 4.1c) are nearly identical to those of species richness (Fig. 4.1a), the RPD map reveals important differences in evolutionary structure (Fig. 4.1d). Northern Africa and western Asia with high C4 diversity fractions (Fig. 4.1b) show higher positive RPD values (Fig. 4.1d), suggesting that their C4 floras comprise phylogenetically distant and/or evolutionarily ancient lineages. By contrast, species-rich tropical regions (Fig. 4.1a) such as Mesoamerica, central South America, and eastern Africa show negative RPD values (Fig. 4.1d), implying more recent radiations from relatively young clades. South Asia, despite harbouring the highest C4 species richness (Fig. 4.1a), shows near-zero RPD (Fig. 4.1d), suggesting that its diversity reflects neither a strong bias towards younger lineages nor towards older ones, but rather a phylogenetic composition broadly proportional to its species richness.

Fig. 4.1: Contemporary global patterns of C4 plant diversity in geographic space.



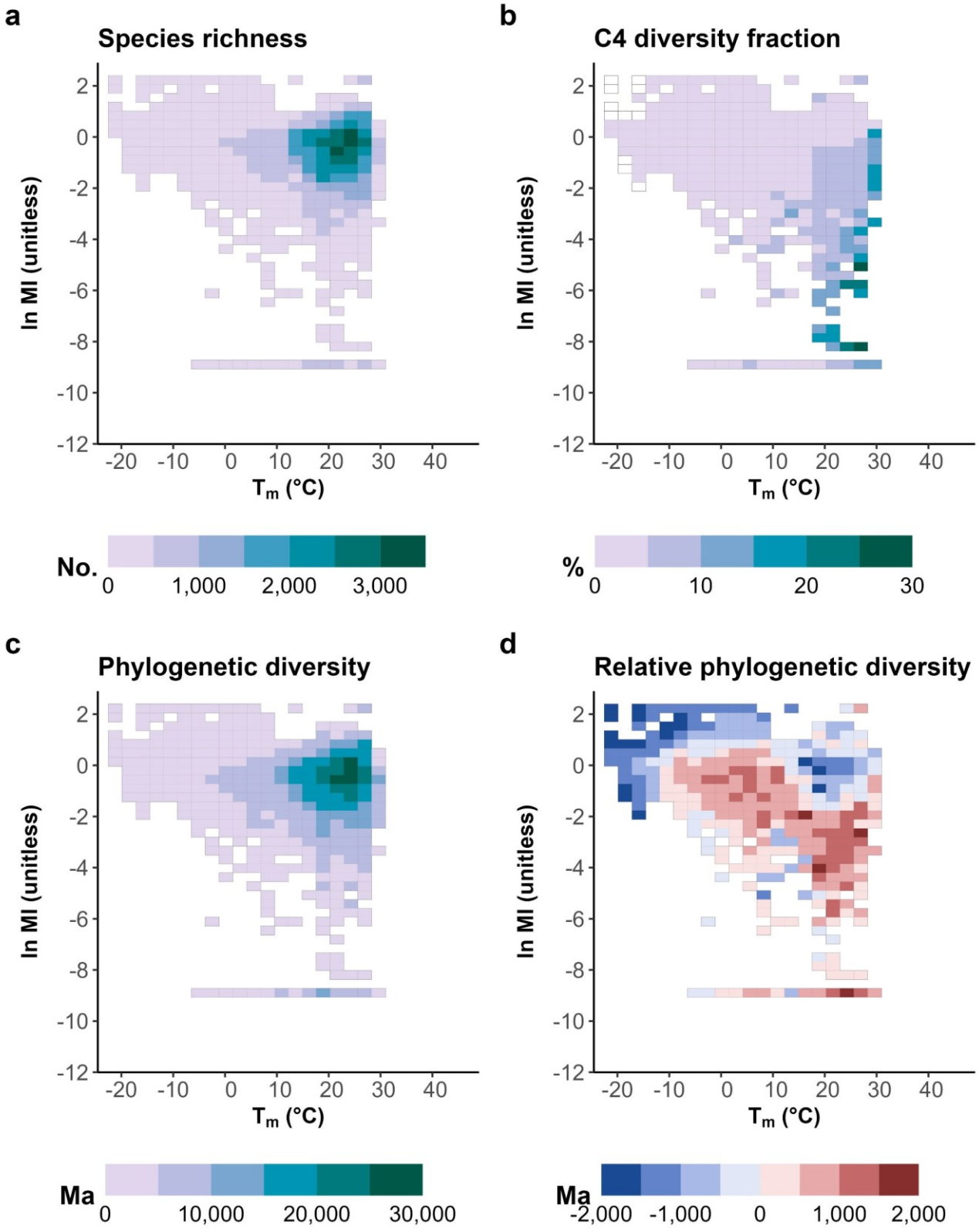
Global C4 diversity at 1° spatial resolution, based on native distributions of 7,156 C4 species and a time-calibrated phylogeny comprising all these C4 species. **a**, C4 species richness. **b**, C4 diversity fraction (%), calculated as the proportion of vascular plant species using the C4 pathway. **c**, Phylogenetic diversity (Faith's PD, Ma)[100]. **d**, Relative phylogenetic diversity (RPD, Ma), calculated as the residual from a PD-richness regression.

Because Poales (e.g., grasses and sedges) constitute the majority (~80%) of C4 species in our dataset, we also visualised global diversity patterns for Poales and non-Poales separately. These maps revealed distinct biogeographic patterns between the two groups (Extended Data Figs. 4.1 and 4.3). For Poales, spatial patterns of species richness, PD, and RPD (Extended Data Fig. 4.1a, c, and d) are broadly consistent with those observed for all C4 species combined (Fig. 4.1 a, c, and d). Outside of the high-latitude regions of Eurasia, Poales comprise the dominant proportion of the C4 flora across most tropical and subtropical regions (Extended Data Fig. 4.3b). Similar to Poales, non-Poales species richness peaks in Mesoamerica and northern Australia (Extended Data Fig. 4.3a). However, these richness hotspots present negative RPD values for non-Poales (Extended Data Fig. 4.3d). In contrast to Poales, western Asia represents an additional hotspot of non-Poales species richness (Extended Data Fig. 4.3a), associated with near-zero RPD (Extended Data Fig. 4.3d). Although non-Poales species richness in Africa is comparatively low, PD values in Africa are elevated relative to richness (Extended Data Fig. 4.3c), resulting in high positive RPD values (Extended Data Fig. 4.3d). Non-Poales also shows much greater variation in RPD values than Poales, with both positive and negative values reaching 3–4 times the amplitude of Poales. By contrast, Poales shows comparatively small RPD amplitudes, reflecting a closer proportionality between richness and PD.

We also visualised global C4 diversity patterns into a two-dimensional climatic space defined by mean annual temperature (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless) (Fig. 4.2; see Methods). These two variables represent independent axes of temperature and plant-available water availability, the dominant climatic controls on plant distribution globally[11,14,152,206,288] (see Methods). The climatic space was partitioned into 40 equal intervals along each axis, and diversity metrics were computed for each grid cell (i.e. 'climatic bin') based on the climatic preference of every C4 species inferred from its native distribution. Results from alternative resolutions (50 and 60 intervals) are reported in the A3 Supplementary information for Chapter 4.

Species richness (Fig. 4.2a) and PD (Fig. 4.2c) of C4 plants both peak under hot ($T_m \approx 25$ °C) and dry (MI ≈ 0.35) conditions, reflecting their broad ecological success in warm, water-limited ecosystems. The proportion of C4 species (C4 diversity fraction) is highest under even hotter and drier climates (Fig. 4.2b). In climatic bins with $T_m > 25$ °C and MI approaching 0, C4 species comprise more than 20% of all vascular plant species (Fig. 4.2b). Arid environments (MI < 0.15) are linked to higher positive RPD values (Fig. 4.2d), indicating that these climatic bins are occupied by evolutionarily distinct and/or ancient C4 lineages. The climatic bins with higher C4 species richness ($15 < T_m < 25$ °C and $0.35 < MI < 1$) show negative RPD values, suggesting dominance by recently diversified and closely related clades (Fig. 4.2d).

Fig. 4.2: Contemporary global patterns of C4 plant diversity in climatic space.



Global patterns of C4 plant diversity in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). The climatic space is divided into 40 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic

combination. All metrics are calculated based on native distributions of 7,156 C4 species and a time-calibrated phylogeny comprising all C4 species. **a**, C4 species richness, i.e. number of species per climatic bin. **b**, C4 diversity fraction (%), calculated as the proportion of vascular plant species using the C4 pathway within each bin. **c**, Phylogenetic diversity (Faith's PD, Ma)[100]. **d**, Relative phylogenetic diversity (RPD, Ma), calculated as the residual from a PD-richness regression. Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

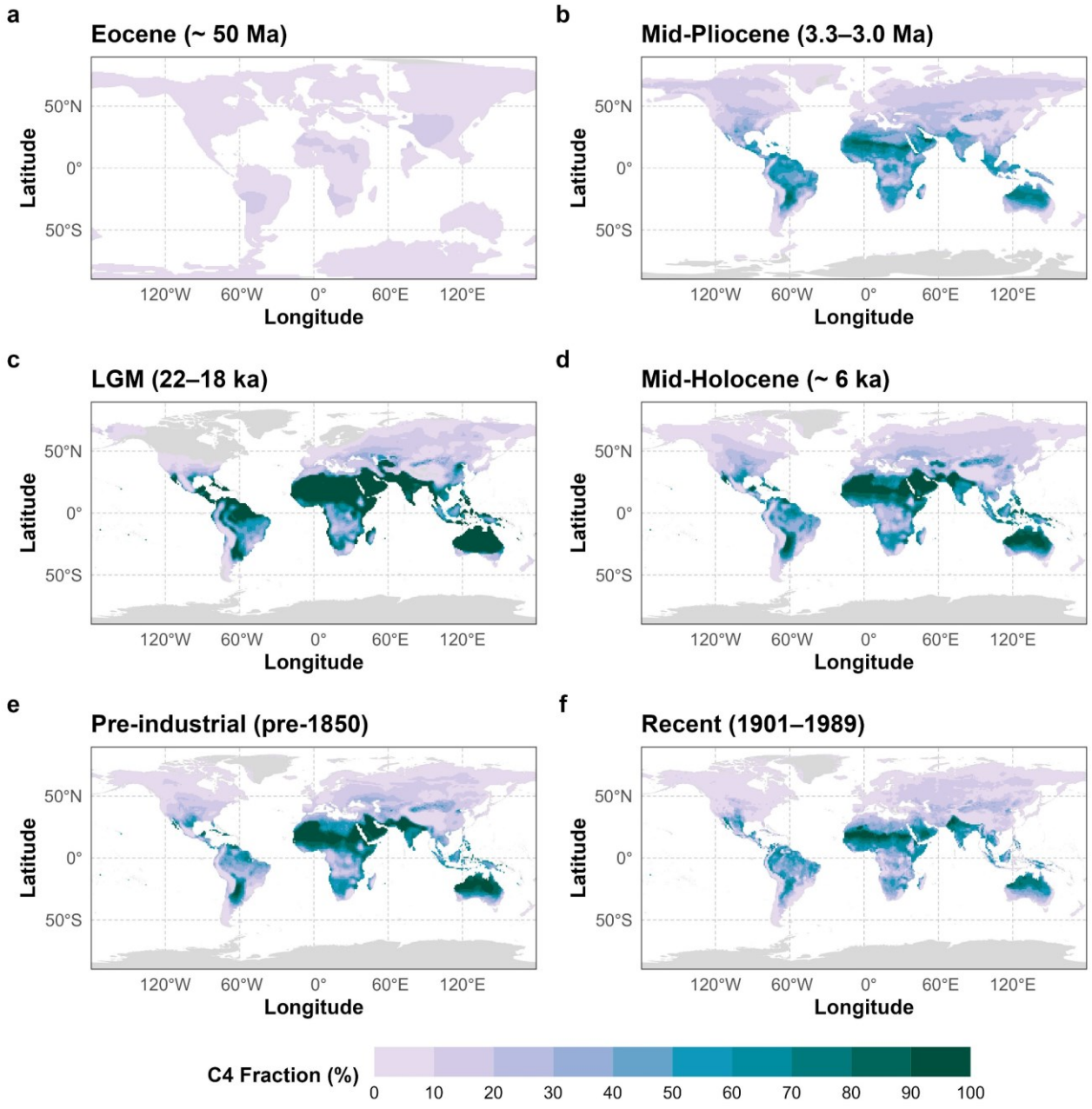
In climatic space, Poales and non-Poales also present distinct diversity patterns (Extended Data Figs. 4.2 and 4.4). For Poales, both species richness and PD reach their highest values under hot (T_m ≈ 25 °C) and dry (MI ≈ 0.35) conditions (Extended Data Fig. 4.2a and c), consistent with the peaks observed when aggregating all C4 species (Fig. 4.2a and c). Across most climatic bins, Poales constitute the dominant proportion of C4 species, with their contribution increasing further towards the hot-wet corner of climatic space (Extended Data Fig. 4.2b). In contrast to the aggregated C4 patterns, climatic bins with the highest richness and PD for Poales show RPD values close to zero or slightly positive (Extended Data Fig. 4.2d). For non-Poales, their richness and PD peak under cooler (T_m ≈ 17.5 °C) and drier (MI ≈ 0.15) conditions than for Poales (Extended Data Fig. 4.4a and c). Climatic bins corresponding to these peak values show negative RPD (Extended Data Fig. 4.4d) for non-Poales, consistent with the aggregated C4 patterns (Fig. 4.2d).

4.3.2 Historical global biogeography of C4 plants

To investigate the deep-time origins of contemporary C4 plant diversity patterns, we applied a set of eco-evolutionary optimality (EEO) models to simulate the historical C4 plant distribution across six key time slices: the Eocene (*ca.* 50 Ma), Mid-Pliocene (3.3–3.0 Ma), Last Glacial Maximum (LGM, 22–18 ka), Mid-Holocene (*ca.* 6 ka), pre-industrial period (pre-1850), and the recent period (1901–1989) (Fig. 4.3). For each period, we compiled environmental variables from climate model simulations or instrumental data, and used these as inputs to simulate the fractional abundance of C4 plants (hereafter C4 fraction) within the total C3 and C4 vegetation across the global land surface.

EEO-model simulations indicate that suitable habitats for C4 plants are virtually absent in the Eocene (Fig. 4.3a). By the Mid-Pliocene, however, the emergence of C4 vegetation is observed in several regions, particularly Central America, central South America, the Sahel, southern Africa, the Arabian Peninsula, South Asia, and northern Australia (Fig. 4.3b). The LGM marks a major expansion of C4 vegetation, with the model simulating widespread areas across the tropics, subtropics, and warm temperate regions, where C4 plants account for over 90% of predicted vegetation cover (Fig. 4.3c). Following this peak, however, the spatial extent and fractional dominance of C4 vegetation continuously decrease (Fig. 4.3d–f). From the Mid-Holocene through to the pre-industrial and recent periods (Fig. 4.3d–f), although tropical and subtropical C4 hotspots persist, both the geographic extent and the fractional abundance of C4 vegetation decline.

Fig. 4.3: Simulated historical biogeography of C4 plants in geographic space.



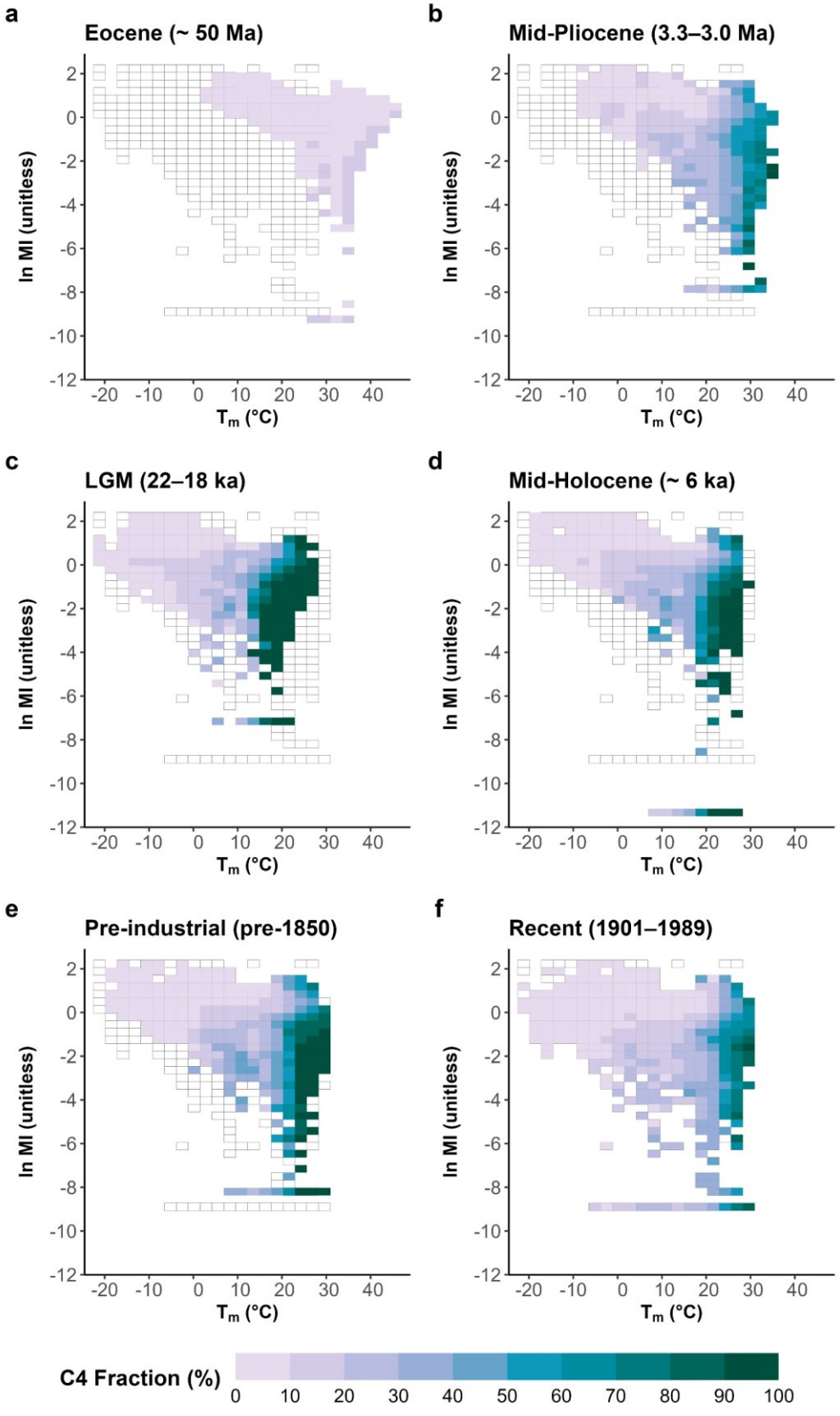
Modelled C4 plant fraction (% of combined C3 and C4 plant abundance) at 1° spatial resolution across six geological slices, reconstructed using models based on eco-evolutionary optimality (EEO) theory. Panels correspond to: **a**, Eocene (*ca.* 50 Ma); **b**, Mid-Pliocene (3.3–3.0 Ma); **c**, Last Glacial Maximum (LGM, 22–18 ka); **d**, Mid-Holocene (*ca.* 6 ka); **e**, pre-industrial period (pre-1850); **f**, recent period (1901–1989).

We then projected the modelled C4 fractions into climatic space for each of the six time slices (Fig. 4.4). Because a single climatic bin may correspond to multiple geographic grid cells[121,131,134]

(see Chapter 1 and Chapter 3), we calculated the average C4 fraction for all geographic grid cells falling within each climatic bin, thus capturing the dynamics of the realised climatic niche occupied by C4 plants during each historical period.

Consistent with results in geographic space, C4 fraction across the climatic space is near zero in the Eocene (Fig. 4.4a), reflecting the unfavourable high-CO_2, warm, and moist global conditions of the time for C4 plants. By the Mid-Pliocene, C4 plants begin to occupy hot ($T_m > 28$ °C), arid to semi-arid ($MI < 0.15$) climatic conditions (Fig. 4.4b). During the LGM, the C4 realised climatic niche reaches its maximum extent (Fig. 4.4c). High C4 fractions ($\geq 90\%$) extend well beyond their Mid-Pliocene climatic limits, encompassing cooler (T_m as low as ~15 °C) and more humid (MI up to 2.72) conditions (Fig. 4.4c). Following the LGM, the climatic niche of C4 plants contracts (Fig. 4.4d–f). From the Mid-Holocene to the pre-industrial and recent periods (Fig. 4.4d–f), C4 fractions decline in cooler and wetter bins, gradually retreating toward the warmer and drier parts of climatic space which they initially occupied in the Mid-Pliocene (Fig. 4.4b).

Fig. 4.4: Simulated historical biogeography of C4 plants in climatic space.



Modelled C4 plant fraction (% of combined C3 and C4 plant abundance) across six geological slices, visualised in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). Historical C4 plant fractions are simulated using models based on eco-evolutionary optimality (EEO) theory. The climatic space is divided into 40 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. Panels correspond to: **a**, Eocene (*ca.* 50 Ma); **b**, Mid-Pliocene (3.3–3.0 Ma); **c**, Last Glacial Maximum (LGM, 22–18 ka); **d**, Mid-Holocene (*ca.* 6 ka); **e**, pre-industrial period (pre-1850); **f**, recent period (1901–1989). Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

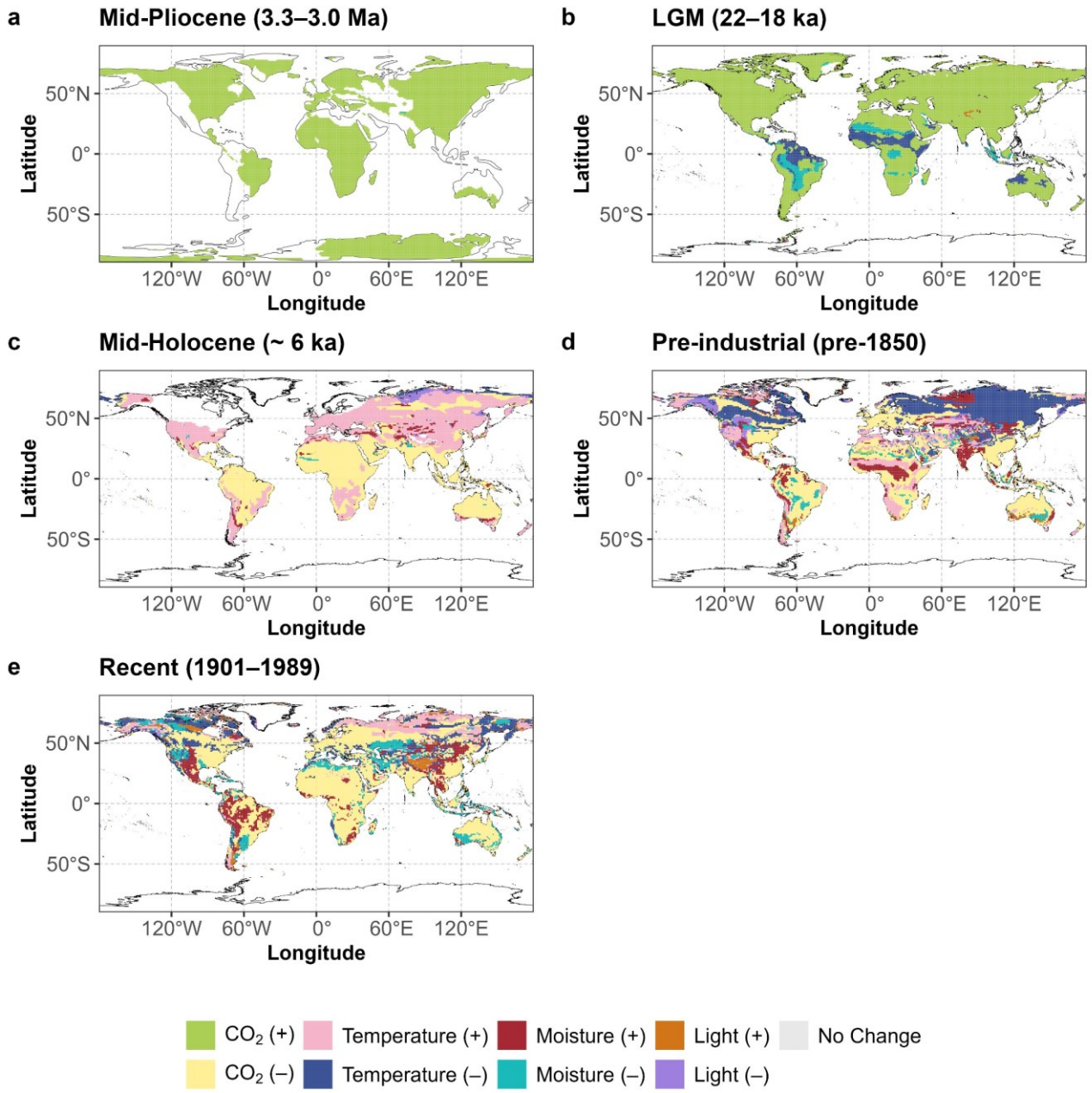
4.3.3 Environmental drivers of C4 plant biogeography

To identify the dominant environmental drivers behind the major transitions in C4 fraction, we first grouped the input variables of EEO models into four categories: CO_2, temperature, moisture, and light (see Methods). We then used the Stein-Alpert decomposition method[289] to quantify the independent effect of each driver category on the change in C4 fraction between consecutive time periods: Mid-Pliocene relative to Eocene, LGM relative to Mid-Pliocene, Mid-Holocene relative to LGM, pre-industrial period (pre-1850) relative to Mid-Holocene, and the recent period (1901–1989) relative to pre-industrial (pre-1850) conditions. The results of this analysis identify the dominant environmental drivers responsible for each major phase of C4 plant expansion and contraction (Figs. 4.5–4.6).

From the Eocene to the Mid-Pliocene, the emergence of C4 vegetation during the Mid-Pliocene is driven almost exclusively by the positive effect of declining atmospheric CO_2 across all regions (Fig. 4.5a), consistent with the known sensitivity of the C4 pathway to low-CO_2 conditions[38,273,274]. From the Mid-Pliocene to the LGM, further global CO_2 decline continues to be the dominant positive driver of increased C4 fraction (Fig. 4.5b). However, global cooling imposes localised negative effects in central part of the Amazonia, the Sahel, and Western Australia, while increased moisture reduces C4 fraction across Amazonia, the Pantanal, the southern Sahara desert, equatorial Africa, and Southeast Asia (Fig. 4.5b). From the LGM to the Mid-Holocene, postglacial warming emerges as the primary driver of increased C4 fraction in high-latitude regions, while in most

tropical and subtropical regions, rising CO_2 exerts a dominant negative effect, reducing the C4 fraction (Fig. 4.5c). Between the Mid-Holocene and the pre-industrial period, the spatial patterns of driver dominance become highly fragmented (Fig. 4.5d). Cooling is the primary negative driver in the northern high latitudes (Fig. 4.5d). In many tropical and subtropical regions, the negative effect of rising CO_2 competes with positive effects from warming or declining moisture availability (Fig. 4.5d). The most recent transition from pre-industrial to the recent period (1901–1989) also shows a heterogeneous pattern (Fig. 4.5e). CO_2 rise remains the principal negative influence across most tropical and subtropical regions, whereas reduced moisture emerges as the dominant positive driver in parts of western United States, the Andes, South Africa, much of temperate East Asia, and mainland Southeast Asia (Fig. 4.5e). Increased moisture suppresses C4 fraction in several regions including Central Asia, southwestern Australia, and the Mediterranean coast of North Africa (Fig. 4.5e).

Fig. 4.5: Dominant environmental drivers of simulated C4 plant biogeography in geographic space.

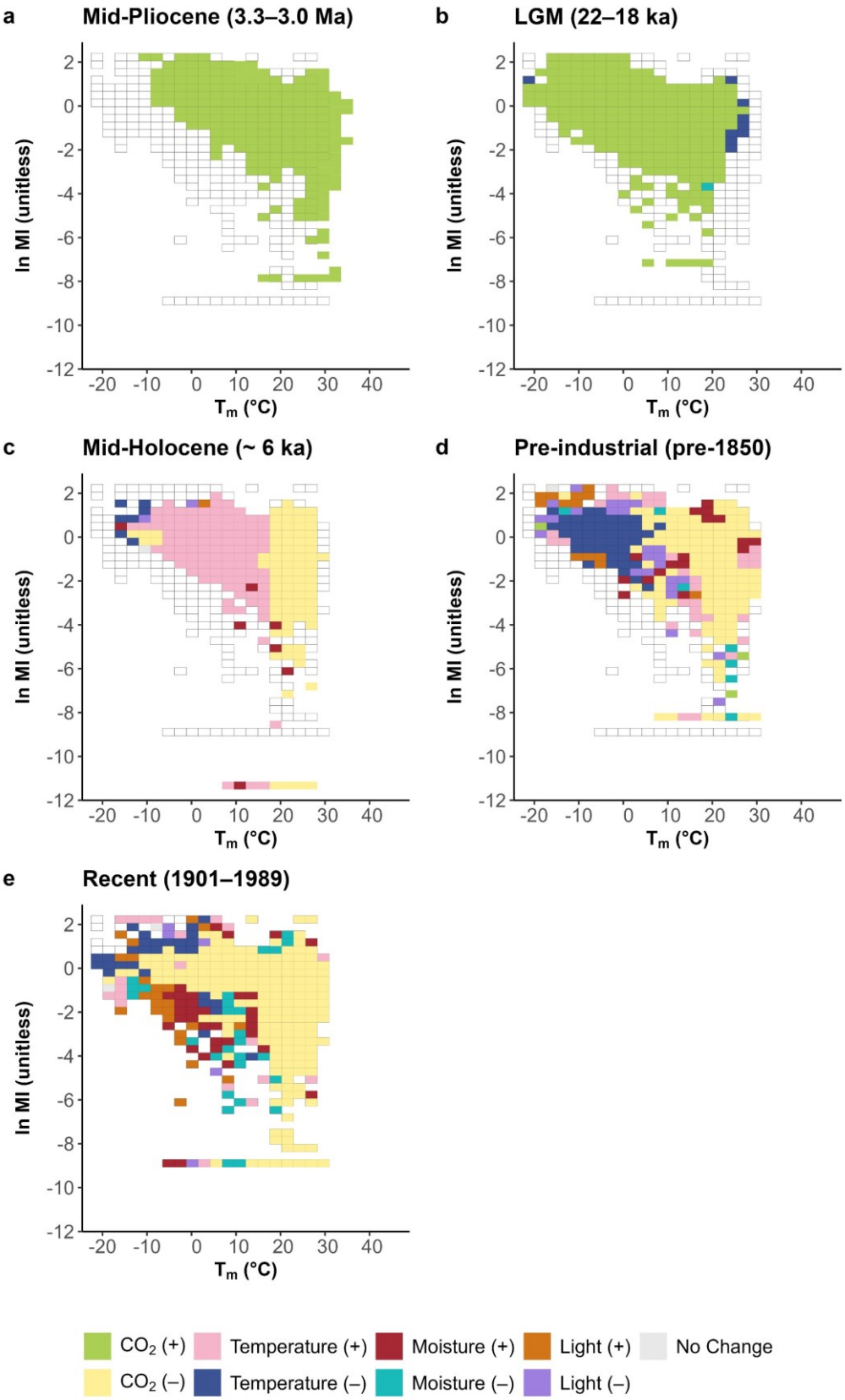


Spatial patterns of the dominant environmental driver influencing changes in modelled C4 plant fraction, mapped at 1° spatial resolution. Independent driver contribution is calculated by Stein-Alpert decomposition of model simulations. Panels correspond to: **a**, Mid-Pliocene (3.3–3.0 Ma); **b**, Last Glacial Maximum (LGM, 22–18 ka); **c**, Mid-Holocene (*ca.* 6 ka); **d**, pre-industrial period (pre-1850); **e**, recent period (1901–1989). Each grid cell is assigned to the environmental variable

(i.e. CO_2, temperature, moisture and light) that exerted the strongest net effect on simulated C4 fraction changes. Colour hues denote the identity of the dominant driver; (±) indicates whether the change in that driver increased (+) or decreased (–) the local C4 fraction. Grey indicates no detectable change.

We extended the dominant driver analysis to climatic space to identify the dominant environmental factors driving historical shifts in the realised climatic niche of C4 plants (Fig. 4.6). The establishment of the C4 climatic niche in the Mid-Pliocene, relative to the Eocene, is entirely driven by the positive effect of declining atmospheric CO_2 (Fig. 4.6a). Similarly, the subsequent expansion of the C4 climatic niche from the Mid-Pliocene to the LGM is primarily promoted by continued CO_2 decline (Fig. 4.6b). Following the LGM, the contraction of the C4 climatic niche towards hotter and drier conditions is primarily driven by the negative effect of rising CO_2 (Fig. 4.6c–e). While rising CO_2 is the persistent driver of this overall contraction, the spatial patterns of other dominant drivers become increasingly fragmented from the Mid-Holocene onwards (Fig. 4.6d–e).

Fig. 4.6: Dominant environmental drivers of simulated C4 plant biogeography in climatic space.



Spatial patterns of the dominant environmental driver influencing changes in modelled C4 plant fraction, mapped in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). The climatic space is divided into 40 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. Independent driver contribution is calculated by Stein-Alpert decomposition of model simulations. Panels correspond to: **a**, Mid-Pliocene (3.3–3.0 Ma); **b**, Last Glacial Maximum (LGM, 22–18 ka); **c**, Mid-Holocene (*ca.* 6 ka); **d**, pre-industrial period (pre-1850); **e**, recent period (1901–1989). Each climatic bin is assigned to the environmental variable (i.e. CO_2, temperature, moisture and light) that exerted the strongest net effect on simulated C4 fraction changes. Colour hues denote the identity of the dominant driver; (±) indicates whether the change in that driver increased (+) or decreased (–) the local C4 fraction. Grey indicates no detectable change. Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

4.4 Discussion

Our study investigated contemporary C4 plant diversity patterns and the historical biogeographic dynamics that shaped them. We reconstructed the major expansions and contractions of C4 vegetation and identified the dominant environmental drivers governing these transitions. This research provides a deep-time context for understanding how past environmental changes have shaped the assembly, diversity, and distribution of C4 vegetation worldwide.

We compiled the most comprehensive C4 species list (7,156 species) to date and constructed global native ranges for each species from multiple global and regional databases, producing the first global map of contemporary C4 diversity. This C4 diversity data encompasses the full set of recognised C4 lineages currently documented in the literature. Because Poales account for roughly 80% of C4 species in our dataset (5,724 species), we examined Poales and non-Poales separately. This comparison shows that the aggregated C4 diversity signal is largely shaped by Poales, whereas non-Poales contribute additional and distinct diversity patterns. For Poales, phylogenetic diversity (PD) increases in proportion to species richness, which produces small absolute values of relative phylogenetic diversity (RPD) across both geographic and climatic space (Extended Data Figs. 4.1

and 4.2). This pattern reflects the phylogenetic structure and evolutionary history of C4 Poales. All C4 Poales species belong to the Poaceae (i.e., grasses) and Cyperaceae (i.e., sedges) families and share a long evolutionary backbone and relatively short terminal branches. Although C4 Poales originated earlier (in the Oligocene)[290], the major radiation and global expansion of C4 grasslands occurred in the Late Miocene[38,40,42], producing species-rich but phylogenetically shallow assemblages dominated by recently diverged taxa. Under such conditions, increases in richness contribute little to deep evolutionary history. Consequently, observed PD remains very close to the expected PD, and RPD values stay close to zero. In contrast, non-Poales C4 species are polyphyletic and spread across more than 30 angiosperm families, with uneven diversification histories ranging from recent species-rich radiations to isolated, evolutionarily deep lineages[38]. This combination of phylogenetically disparate origins and heterogeneous speciation rates produces much larger differences between richness and PD, resulting in RPD values that are three to four times higher than those of Poales (Extended Data Figs. 4.3 and 4.4). Non-Poales also occupy a distinct climatic space: their richness and PD peak under cooler and drier conditions than those of Poales, and many occur in deserts, saline habitats, and other high-stress environments[38], where distantly related taxa persist under similar conditions, contributing to elevated RPD.

The contemporary hotspots of C4 species richness and phylogenetic diversity are concentrated in tropical and subtropical regions, notably in Mesoamerica, central South America, central and eastern Africa, South Asia, and northern Australia (Fig. 4.1a and c). In many of these areas, such as South Asia and northern Australia, C4 vegetation remains dominant today (Fig. 4.3f). Our model simulations indicate that the origins of these hotspots can be traced back to the Mid-Pliocene (Fig. 4.3b), a finding consistent with geological records. Stable carbon isotopic data ($\delta^{13}C$) from fossil tooth enamel document a worldwide expansion of C4 grasslands through the displacement of C3 vegetation during the Late Miocene and Pliocene[36]. In Mesoamerica, isotopic analysis of Pliocene herbivore teeth from Yepómera, Mexico, reveals a diet incorporating a substantial C4 vegetation component within a dry, open savanna or grassland[291]. Similarly, in central South America, $\delta^{13}C$ records from soil carbonates in northwest Argentina indicate the presence of C4 plants as early as 7.3–6.7 Ma[292], with fossil teeth confirming that C4 plants became a major dietary component for herbivores during the Pliocene[293]. In eastern Africa, isotopic data show that various large herbivores,

including gomphotheriids and elephantids, became increasingly reliant on C4 resources between the late Miocene and Pliocene, indicating that C4 grasses had become a stable and abundant feature of the ecosystem[294]. The C4 expansion in South Asia was driven by the development of seasonal climates with warm-season precipitation linked to the uplift of the Himalaya-Tibetan plateau, which intensified fire cycles that favoured a forest-to-grassland transition[39,42,295,296]. In northern Australia, carbon isotope ratios of plant waxes from ocean sediments pinpoint the onset of C4 expansion at ~3.5 Ma, coinciding with the development of a summer monsoon climate[297].

Our analysis also identifies modern hotspots of high C4 diversity fraction, including the Sahel, southern Africa, the Arabian Peninsula, and northern Australia (Fig. 4.1b), despite their relatively low overall C4 richness (Fig. 4.1a). EEO model simulations show that these regions have maintained a high fractional abundance of C4 plants throughout the geological periods studied (Fig. 4.3). The persistence of high C4 abundance in these regions, particularly in the Sahel and the Arabian Peninsula, is coupled with high positive RPD values (Fig. 4.1d), suggesting they harbour evolutionarily distinct and/or ancient C4 lineages. Palaeoecological evidence supports the long-term persistence of C4-dominated vegetations in these regions. In the Sahel, carbon isotopic data from Pliocene hominins in the Chad basin indicate a diet almost entirely dependent on C4 resources by ~3.5 Ma, pointing to a vast, C4-dominated open ecosystem[298]. This pattern continued through the Pleistocene, with palynological data showing a peak in C4 grass pollen during the LGM[299], and geochemical records linking large-scale C3 to C4 vegetation changes to continental aridity driven by the African monsoon[300]. Similarly, in southern Africa, the inclusion of C4 tropical grasses and sedges in the diet of Pliocene australopithecines from ~3 Ma onwards marks the establishment of open savanna or grassland ecosystems[301]. Marine sediments from the southeast Atlantic continental margin confirm a persistent latitudinal gradient of increasing C4 indicators southwards across multiple glacial-interglacial cycles in southern Africa[237,302]. Although the deep-time palaeoecological record for the Arabian Peninsula is sparse, pollen records show the continuous dominance of C4-associated desert herbaceous plant communities since the Mid-Holocene[303]. For northern Australia, the development of highly seasonal monsoon climates around 3.5 Ma created the seasonal dryness that has favoured C4 vegetation[297,304], a condition further intensified by aridity from weakened monsoons during subsequent glacial periods[304].

Our dominant driver analysis identifies atmospheric CO_2 as the primary global control over the major biogeographic transitions of C4 plants. The decline in CO_2 concentration ([CO_2]) below a critical physiological threshold is a necessary pre-condition for the widespread success of C4 plants[36–39]. This is evident in our EEO-model simulations, which show C4 plants were virtually absent during the high-CO_2 Eocene ([CO_2] ≈ 1850 ppm, Fig. 4.3a) but widespread in the tropics and subtropics during the Mid-Pliocene as [CO_2] fell to ~400 ppm (Fig. 4.3b) – a transition driven almost exclusively by declining CO_2 (Fig. 4.5a). The subsequent expansion of C4 vegetation peaked during the LGM (Fig. 4.3c), when historically low [CO_2] of ~190 ppm gave C4 plants a significantly competitive advantage over CO_2-starved C3 plants (Fig. 4.5b), leading to C4 fractions exceeding 90% across vast tropical, subtropical and warm temperate regions (Fig. 4.3c). This pattern is consistent with pollen and carbon isotope records in tropical regions[305]. Conversely, rising CO_2 since the LGM has been the dominant negative driver of C4 fraction, particularly in tropical and subtropical climates, despite temperature and moisture acting as secondary drivers shaping regional boundaries (Fig. 4.5c–e). Similar dynamics are shown in climate space, where the realised climatic niche of C4 plants expanded from a hot-arid core in the Mid-Pliocene into cooler and more humid conditions during the low-CO_2 LGM, before contracting back towards this core as post-glacial CO_2 levels rose (Figs. 4.4 and 4.6). These transitions are consistent with the 'crossover temperature' concept, whereby rising CO_2 increases the temperature at which C4 photosynthesis becomes advantageous over C3[279]. The persistent occupation of these hot-arid climatic conditions, which correspond to climatic bins with higher positive RPD (Fig. 4.2d), indicates they have acted as long-term refugia for evolutionarily distinct C4 lineages. These results suggest that contemporary C4 biogeographic patterns retain strong legacy signals from their CO_2-driven historical dynamics.

Our study reveals two distinct contemporary biogeographic patterns of C4 diversity: (1) hot-arid 'refugia' with low species richness, but high diversity fraction and positive RPD, and persistently high C4 abundance; and (2) 'radiation centres' with high richness but neutral or negative RPD. While this provides a new perspective on C4 macroecology, it also highlights the need to differentiate among the adaptive strategies of distinct evolutionary lineages. C4 photosynthesis has evolved independently over 60 times across different geological periods and geographic locations[38,128]. The two biogeographic patterns we observe likely reflect the divergent evolutionary

trajectories among C4 lineages. Future research could explore why some lineages originated in and remain low-diversity endemics within hot-arid 'refugia', whereas others, especially ancient grass lineages, apparently used these areas as springboards to colonise and radiate into more productive 'radiation centres'. Exploring the historical and ecological determinants of these divergent pathways will be key to understanding the macroevolutionary dynamics of C4 plants.

Our study also has several limitations that should be acknowledged. First, the phylogenetic relationships used in this analysis were derived from a megatree onto which species absent from the backbone were grafted at the genus or family level. While this approach captures the deep phylogenetic structure most relevant to large-scale patterns, it does not resolve species-level topology or divergence times. These uncertainties may influence the magnitude and spatial distribution of RPD. Second, although our analyses demonstrate that Poales drive the dominant contemporary C4 diversity signal and that non-Poales contribute additional phylogenetic and ecological complexity, the eco-evolutionary optimality (EEO) models that we employed to reconstruct deep-time C4 biogeography treat C4 plants as a single functional group. This approach emphasises environmental controls on C3-C4 competition, but does not differentiate among C4 lineages with distinct environmental requirements, such as Poales and non-Poales. As a result, potential differences in the environmental drivers underlying the expansion of different C4 clades cannot be evaluated. These limitations highlight the need for future work to incorporate more complete dated phylogenies and more advanced modelling frameworks that can explicitly represent phylogenetic and lineage-level environmental variation.

This study compiles the most comprehensive C4 plant diversity data to date and, for the first time, simulates the deep-time dynamics of global C4 biogeography since *ca*. 50 Ma using eco-evolutionary optimality theory. The results confirm the central role of atmospheric CO_2 change in shaping the biogeographic history of C4 plants, and reveal distinct refugia and radiation centres of C4 evolution in both geographic and climatic spaces. These findings are of direct relevance to applied and conservation science: C4 plants underpin global food security through staple crops, dominate the structure and functioning of grassland and savanna ecosystems, and include invasive species whose spread may be intensified by climate change. Moreover, the deep-time biogeographic dynamics of C4 plants provide crucial context for understanding the evolution and biogeography of

C4-dependent fauna. Integrating macroevolutionary perspectives with contemporary diversity patterns is therefore essential for predicting and managing C4 plant responses to environmental change and for conserving biodiversity in an ever-changing world.

4.5 Methods

4.5.1 C4 plant diversity

We assembled a list of 7,223 C4 plant species from public databases and published literature (see Data availability for the complete references). To ensure taxonomic consistency, all scientific names were standardised against the World Checklist of Vascular Plants (WCVP)[209] using the Taxonomic Name Resolution Service V5.3.1 (TNRS, https://tnrs.biendata.org/)[179]. We then cross-referenced this C4 species list with a comprehensive vascular plant distribution database from ref. [288] (Chapter 3) to obtain native distributional data for each C4 species. This process resulted in 7,156 species with matched presence-absence raster data at a 1° resolution, where a cell value of 1 denotes presence and 0 denotes absence. By stacking the individual species rasters, we generated a global map of C4 plant species richness. Furthermore, we calculated the C4 diversity fraction by dividing C4 species richness by total vascular plant richness, with the latter also sourced from the comprehensive vascular plant diversity data compiled in ref. [288] (Chapter 3).

To evaluate phylogenetic diversity, we constructed a time-calibrated phylogenetic tree for the 7,156 C4 plant species with distributional data. The phylogeny was generated using the 'phylo.maker' function in the *U.PhyloMaker* R package[263], with a comprehensive, time-calibrated vascular plant phylogeny developed in ref. [288] (Chapter 3) serving as the backbone. Species not present in the backbone megatree were grafted onto it using a hierarchical approach that places them at the basal node of their genus or, if the genus is absent, at a calculated position within their family[260,263].

Using this comprehensive C4 phylogeny, we calculated two metrics of phylogenetic diversity. First, we calculated Faith's phylogenetic diversity (PD, Ma)[100] for each 1° × 1° grid cell, defined as the total branch length of the minimum phylogenetic subtree connecting all species present in that grid. Higher PD values thus represent greater accumulation of evolutionary history within the grid.

To control for the strong, intrinsic correlation between PD and species richness, we also calculated relative phylogenetic diversity (RPD, Ma)[264,265]. RPD is defined as the residuals from an ordinary least squares (OLS) regression of PD against species richness across all global grid cells. Larger positive RPD values thus indicate that the species assemblage in a grid cell is phylogenetically overdispersed, i.e., species are more distantly related and/or more evolutionarily ancient than expected given their richness. Larger negative values indicate phylogenetic clustering, i.e. more closely related than expected, implying more recent radiations from younger clades.

Because Poales (e.g., grasses and sedges) account for 5,724 of 7,156 C4 species in our dataset (*ca.* 80%), we calculated the same set of diversity metrics separately for Poales and non-Poales C4 taxa to test whether patterns in species richness and phylogenetic diversity were mainly driven by the diversification of Poales.

4.5.2 Modelling framework

To reconstruct C4 plant biogeography, we simulated vegetation dynamics for six representative geological time slices cover the C4 evolutionary history: the Eocene (*ca.* 50 Ma), Mid-Pliocene (3.3–3.0 Ma), Last Glacial Maximum (LGM, 22–18 ka), Mid-Holocene (*ca.* 6 ka), pre-industrial period (pre-1850), and a recent period from 1901 to 1989. Our approach integrates a suite of models based on eco-evolutionary optimality (EEO) theory (Fig. A3.1)[47,280–284], implemented in Python v3.12 using the *pyrealm* package (https://pyrealm.readthedocs.io/en/latest/index.html).

The core of our framework is a C3/C4 competition model estimating the C4 fraction[47]. This C4 fraction is defined as the areal proportion of vegetation utilising the C4 photosynthesis pathway within each grid cell. The model calculates C4 fraction based on the relative advantage of C4 over C3 plants, which is derived from gross primary production (GPP) of the two pathways under given environmental conditions. This principle is similar to that used in ref. [46], although our algorithm, as shown in ref. [47], more accurately reproduces observed soil $\delta^{13}C$ patterns. The 'C4 advantage' is then converted to the share of C4 plants in the total GPP by fitting a logistic curve between the modelled C4 advantage and observed C4 abundance data. However, when water and nutrients are not limiting, C3 trees can outcompete C4 plants through shading, even where C4 photosynthesis would be favoured. To account for this shading effect, the C3/C4 competition model incorporates tree cover

as an additional required input. This tree cover is fitted in a power function to evaluate the proportional contribution from C3 trees in total GPP which is subsequently subtracted from the C4 GPP share. The resulting potential C4 share is first constrained by temperature by setting to zero in areas where the mean temperature of the coldest month falls below −24°C, a known physiological limit for C4 plant growth[308]. The temperature-corrected C4 share is then converted to a potential areal fraction of C4 plants by dividing it by an empirically derived factor of 1.13, following the emergent constraint identified by Luo and colleagues[46]. Finally, this potential C4 fraction is corrected for land use by subtracting the areas occupied by C3 and C4 crops and urban areas, yielding the final natural C4 fraction used in our analysis. The entire modelling process is therefore driven by three primary inputs: C3 GPP, C4 GPP and fractional tree cover.

We used an optimality-based light-use efficiency P model[280–282] to simulate the potential GPP for C3 and C4 plants separately. The potential GPP is defined as the theoretical productivity assuming complete light absorption by the canopy (i.e., fraction of absorbed photosynthetically active radiation, fAPAR = 1). The P-model also requires daily inputs of atmospheric CO_2 concentration ([CO_2], ppm), air temperature (*tas*, °C), atmospheric pressure (*ps*, Pa), vapour pressure deficit (*vpd*, Pa) derived from *tas*, *ps* and specific humidity (*hus*, unitless), and incident photosynthetic photon flux density (*ppfd*, $\mu mol \cdot m^{-2} \cdot s^{-1}$) derived from surface downwelling shortwave radiation (*rsds*, $W \cdot m^{-2}$). The resulting potential GPP is then adjusted for water limitation using a soil moisture stress factor calculated by the Simple Process-Led Algorithms for Simulating Habitats (SPLASH) model[184,185], which in turn requires daily inputs of air temperature (*tas*, °C), precipitation (*pr*, mm), elevation (*elev*, m; or using surface altitude, *orog*, m), and fractional hours of bright sunshine (*sf*, unitless) derived from total cloud cover percentage (*clt*, %).

To convert the potential GPP to actual GPP, the fixed fAPAR value of 1 was replaced with a dynamic seasonal cycle of fAPAR. This dynamic fAPAR is derived from the seasonal cycle of leaf area index (LAI). The process began by simulating the seasonal maximum fAPAR ($fAPAR_{max}$) using an $fAPAR_{max}$ model[283] that incorporates a water-carbon trade-off, defined as the lesser of an energy-limited (maximising GPP) and a water-limited (maximising available moisture derived from: precipitation, *pr*, mm; and potential evapotranspiration, *pet*, mm) estimate. This seasonal $fAPAR_{max}$ was then converted to a seasonal maximum LAI using Beer's law[309]. Subsequently, we applied a

prognostic LAI model[284] based on the EEO theory that the display of foliage should follow, with some time lag, the seasonal time course of potential GPP. The model simulates the daily time course of LAI based on the seasonal maximum LAI and the potential GPP derived from climate variables. It uses a moving average to represent the time lag between the modelled steady-state LAI and the actual allocation of carbon to leaves. The seasonal cycle of LAI was then used to calculate the seasonal cycle of fAPAR, which was integrated back into the P-model to yield the final actual GPP for C3 and C4 plants, respectively. These actual GPP were then used as inputs for the C3/C4 competition model.

Another critical input of the C3/C4 competition model, tree cover, is not available from reconstructions for deep-time periods. We therefore developed a semi-empirical model to simulate it based on the net carbon profit (P_n) of C3 trees (Extended Data Fig. 4.5) – a metric reflecting the carbon remaining for tree growth after accounting for the costs of constructing and maintaining leaves[310]. P_n has been shown to influence tree diameter and height, which in turn affect crown area and stand density, factors that collectively determine tree cover[311]. We calculated P_n based on the potential GPP of C3 trees, seasonal maximum fAPAR and LAI derived from the P-model and the $fAPAR_{max}$ model (Fig. A3.2). To parameterise this relationship, we first calculated a reference P_n for the period 2001–2010 (Extended Data Fig. 4.5a) using remote-sensing fAPAR and LAI data from the BESS v2.0 database[312], alongside the environmental inputs collected from the WFDE5 version 2.0 database[313]. For the reference tree cover (Extended Data Fig. 4.5b), we used a global potential tree cover map estimated by integrating satellite-based tree cover observations into a machine learning framework[314]. We identified a strong logistic relationship between the reference P_n and tree cover (Extended Data Fig. 4.5c). This relationship was formalised by fitting a robust non-linear regression model using the 'nlrob()' function from the *robustbase* package in R[315]. This function uses M-estimators and employs iterated reweighted least squares to minimise the effects of outliers, thereby generating the final predictive equation for tree cover (Extended Data Fig. 4.5d).

We then executed the full modelling chain for each of the six periods, using the corresponding environmental variables as inputs to estimate the global C4 fraction across the six periods.

4.5.3 Environmental variables

We collected gridded environmental data required by the EEO models from various databases for each selected period (Tables A3.1–A3.2) and converted them to a daily timestep. These variables are atmospheric CO_2 concentration ([CO_2], ppm), air temperature (*tas*, °C), precipitation (*pr*, mm), atmospheric pressure (*ps*, Pa), specific humidity (*hus*, unitless), surface downwelling shortwave radiation (*rsds*, $W \cdot m^{-2}$), total cloud cover percentage (*clt*, %), as well as elevation for the present day (*elev*, m), and palaeo-elevation for geohistorical periods, represented by modelled surface altitude (*orog*, m).

Global [CO_2] data for the pre-industrial (pre-1850) and recent (1901–1989) periods were downloaded from NOAA (https://doi.org/10.15138/9N0H-ZH07)[239]. For the Eocene, average [CO_2] was derived from refs. [240] and [241]. The Mid-Pliocene [CO_2] was obtained from refs. [242] and [243], the LGM [CO_2] was acquired from ref. [244], and the Mid-Holocene [CO_2] was downloaded from NOAA Paleoclimatology Data[245].

Simulations of the other seven required environmental variables for the Eocene were sourced from the Deep-Time Model Intercomparison Project (DeepMIP version 1.0; http://www.deepmip.org)[229]. For the Mid-Pliocene, LGM and Mid-Holocene, simulations of the environmental variables were derived from ensemble model under the Paleoclimate Modelling Intercomparison Project Phase 4 (PMIP4; https://pmip.lsce.ipsl.fr/)[230]. Pre-industrial simulations of the environmental variables were obtained from the Coupled Model Intercomparison Project Phase 6 (CMIP6)[231]. For the recent period (1901–1989): *tas*, *pr* and *clt* data were downloaded from the Climatic Research Unit (CRU) Time-Series (TS) database (version 4.09: https://catalogue.ceda.ac.uk/uuid/9cf07e92afaa405da4f40b6733f362d3)[232]; *elev* data derived from the SRTM elevation data was acquired from WorldClim v.2.1 (https://www.worldclim.org/)[183]; and *hus*, *ps*, and *rsds* data were obtained from CMIP6[231].

For each period, we selected model simulations that showed strong agreement with available observations or proxy-based reconstructions (Table A3.2) and calculated a multi-model ensemble mean. We utilised vector maps of land boundaries for the Eocene and Mid-Pliocene epochs from Knowing Earth (http://www.knowing.earth)[233] and raster map of land boundary for other periods from Natural Earth (https://www.naturalearthdata.com) to derive global land-surface environmental data. All gridded environmental data were aggregated and reprojected to a spatial resolution of 1°

using bilinear interpolation, implemented via the 'projectRaster()' function in the *raster* R package[234] for further calculations.

4.5.4 Climatic space

We evaluated contemporary C4 plant diversity and historical C4 fraction within a two-dimensional climatic space, defined using two bioclimatic variables following the approach in ref. [288] (Chapter 3). One variable is a temperature index (T_m, °C), calculated as the average of mean temperature of the coldest month and the warmest month (°C). The other one is a moisture index (MI, unitless), derived from the ratio of mean annual precipitation (mm) to the mean annual potential evapotranspiration (mm). The MI was subsequently corrected for the physiological effects of past atmospheric CO_2 concentration using 'corrected_mi()' function in the R package *codos*[246], following the methodology proposed by Prentice *et al.* (see ref. [235] for details). We used the natural-log transformed MI (ln MI) to emphasise differences at the dry end of the moisture range[206]. These two variables correspond to the two recognised climatic dimensions that regulate the global geographic distribution of plants[11,14,152,206,288], with T_m reflecting the latitudinal gradient of temperature[207] and MI (ln MI) representing plant-available moisture[11,14,152,206]. We calculated these two bioclimatic variables for all selected periods, and the final climatic space was constructed to encompass the full range of climatic conditions present across all periods.

The climatic space was constructed at multiple resolutions by dividing each axis into 40, 50 or 60 equal intervals. Although analyses were performed across all resolutions, with consistent findings (A3 Supplementary information for Chapter 4), results presented in the main text are based on the 40-interval resolution. The climatic space was discretised into a grid of 'climatic bins', each representing a unique recent climatic condition.

Using the presence-absence rasters in geographic space, we identified the climatic conditions occupied by each C4 species and generated corresponding presence-absence rasters in the climatic space. By stacking these rasters, we mapped C4 species richness within each climatic bin. This procedure was repeated for all vascular plants to compute the C4 diversity fraction (C4 richness / total vascular plant richness) in each bin. Subsequently, we calculated PD and RPD within each climatic bin.

A single climatic bin can correspond to multiple geographic grid cells[121,131,134] (see Chapter 1 and Chapter 3). Specifically, for each period, we determined the set of geographic grid cells in the geographic space that correspond to each climatic bin in the climatic space – that is, all geographic locations sharing the same climatic conditions. To evaluate the C4 fraction in each climatic bin, we calculated the average C4 fraction for all geographic grid cells falling within that bin for each period. Mapping the C4 fraction in climatic space captures the dynamics of the realised climatic niche occupied by C4 plants during each geological period.

4.5.5 Dominant driver analysis

We quantified the relative contributions of different environmental factors to the major historical transitions in C4 fraction. The environmental input variables for the EEO models were grouped into four driver categories: CO_2 (i.e. [CO_2] and *ps*); temperature (i.e. *tas*); moisture (i.e. *pr*, *pet*, *vpd* and soil moisture stress factor); and light (i.e. *rsds* and tree cover). We then employed the Stein-Alpert decomposition method[289] to partition the change in C4 fraction between consecutive time periods. This method quantifies the independent effect of each driver category.

For each transition from a reference period (*r*) to a target period (*t*) (e.g. Eocene to Mid-Pliocene), the calculation proceeded as follows:

(1) Reference simulation

We first calculated the baseline C4 fraction for the reference period,

$$F_r = F\left(CO_{2_r}, Temperature_r, Moisture_r, Light_r\right)$$

(2) Driver-specific substitutions

For each driver category *k*, we recalculated the C4 fraction after replacing all variables in that category with their target-period values, while keeping the other categories at reference-period values. For example, for moisture:

$$F_{Moisture|t} = F\left(CO_{2_r}, Temperature_r, \boldsymbol{Moisture_t}, Light_r\right)$$

(3) Category effect (grid-cell level)

The effect of category *k* was then calculated as the difference between the substituted and reference simulations:

$$\Delta_k = F_{k|t} - F_r$$

A positive Δ_k means that the change in category *k* increased the C4 fraction in that grid cell, while a negative Δ_k means it decreased the C4 fraction.

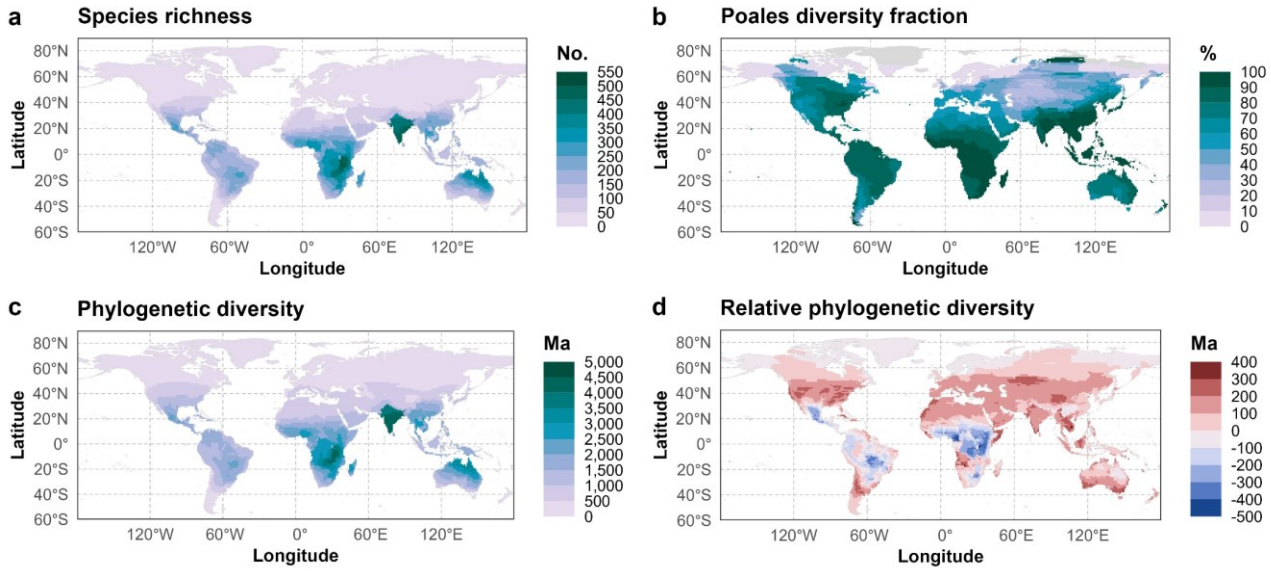
(4) Dominant driver identification

For each grid cell, the dominant driver of the transition was identified as the category with the largest absolute effect $|\Delta_k|$. The sign of $|\Delta_k|$ indicates whether the driver contributed positively or negatively to the change.

The analysis was applied sequentially to five key transitions: Mid-Pliocene relative to the Eocene, LGM relative to the Mid-Pliocene, Mid-Holocene relative to the LGM, pre-industrial period (pre-1850) relative to Mid-Holocene, and the recent period (1901–1989) relative to pre-industrial conditions. This sequential decomposition allowed us to identify the dominant environmental drivers responsible for each major phase of C4 plant expansion and contraction.

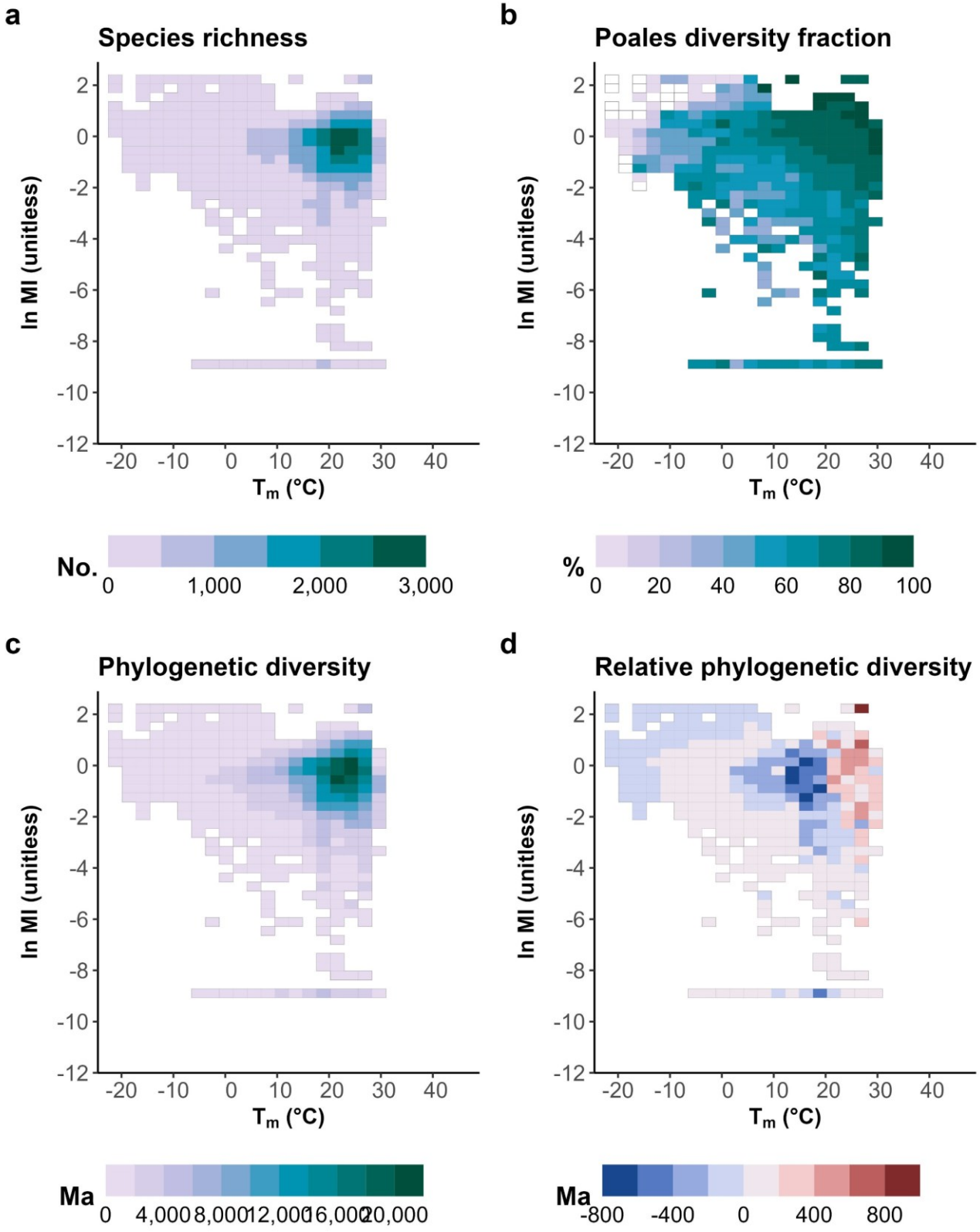
4.6 Extended data figures and tables

Extended Data Fig. 4.1 Contemporary global patterns of C4 Poales diversity in geographic space.



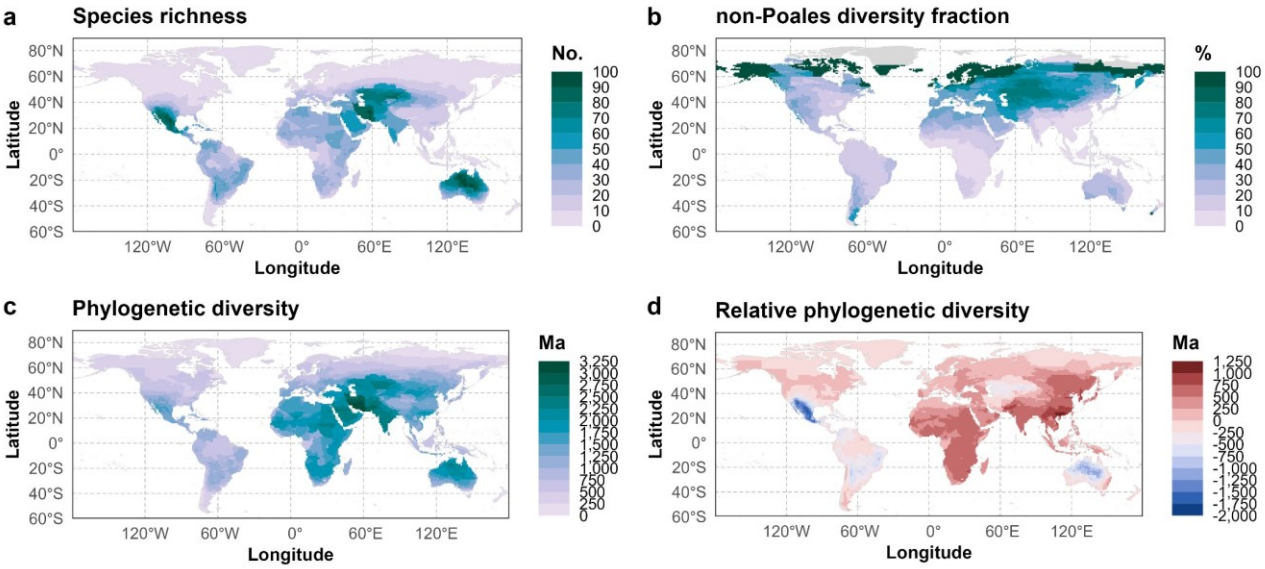
Global C4 Poales diversity at 1° spatial resolution, based on native distributions of 5,724 C4 Poales species and a time-calibrated phylogeny comprising all C4 Poales species. **a**, Species richness of C4 Poales. **b**, Poales diversity fraction (%), calculated as the proportion of local C4 species that belong to Poales. **c**, Phylogenetic diversity (Faith's PD, Ma)[100]. **d**, Relative phylogenetic diversity (RPD, Ma), calculated as the residual from a PD-richness regression.

Extended Data Fig. 4.2 Contemporary global patterns of C4 Poales diversity in climatic space.



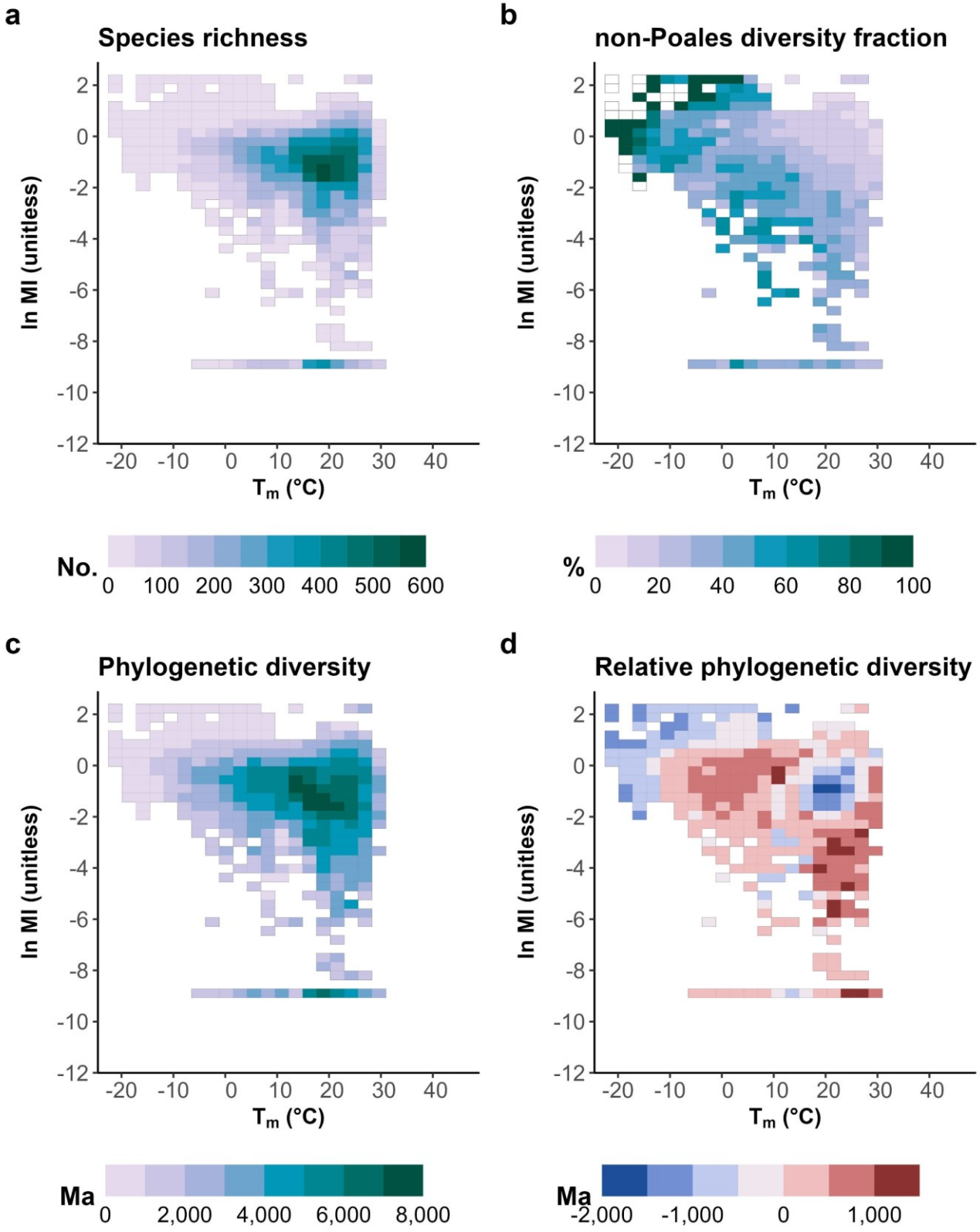
Global patterns of C4 Poales diversity in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). The climatic space is divided into 40 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. All metrics are calculated based on native distributions of 5,724 C4 Poales species and a time-calibrated phylogeny comprising all C4 Poales species. **a**, Species richness of C4 Poales, i.e. number of C4 Poales species per climatic bin. **b**, C4 diversity fraction (%), calculated as the proportion of local C4 species that belong to Poales within each bin. **c**, Phylogenetic diversity (Faith's PD, Ma)[100]. **d**, Relative phylogenetic diversity (RPD, Ma), calculated as the residual from a PD-richness regression. Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

Extended Data Fig. 4.3 Contemporary global patterns of non-Poales C4 diversity in geographic space.



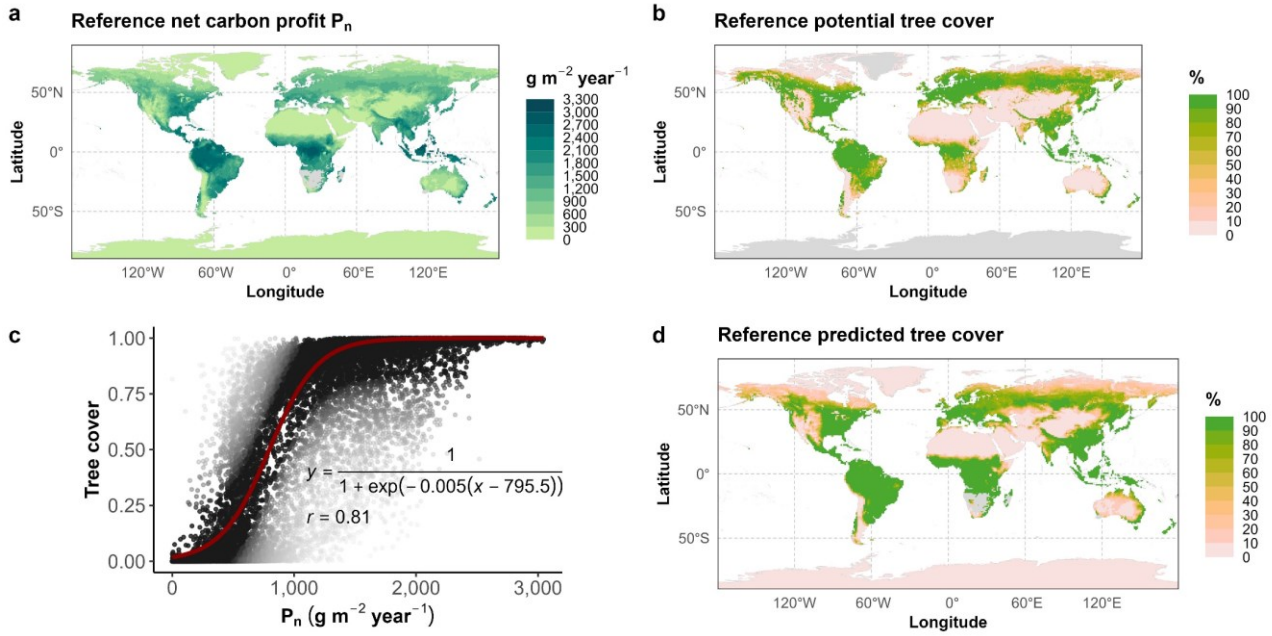
Global non-Poales C4 diversity at 1° spatial resolution, based on native distributions of 1,432 non-Poales C4 species and a time-calibrated phylogeny comprising all non-Poales C4 species. **a**, species richness of non-Poales C4 plants. **b**, non-Poales diversity fraction (%), calculated as the proportion of local C4 species that do not belong to Poales. **c**, Phylogenetic diversity (Faith's PD, Ma)[100]. **d**, Relative phylogenetic diversity (RPD, Ma), calculated as the residual from a PD-richness regression.

Extended Data Fig. 4.4 Contemporary global patterns of non-Poales C4 diversity in climatic space.



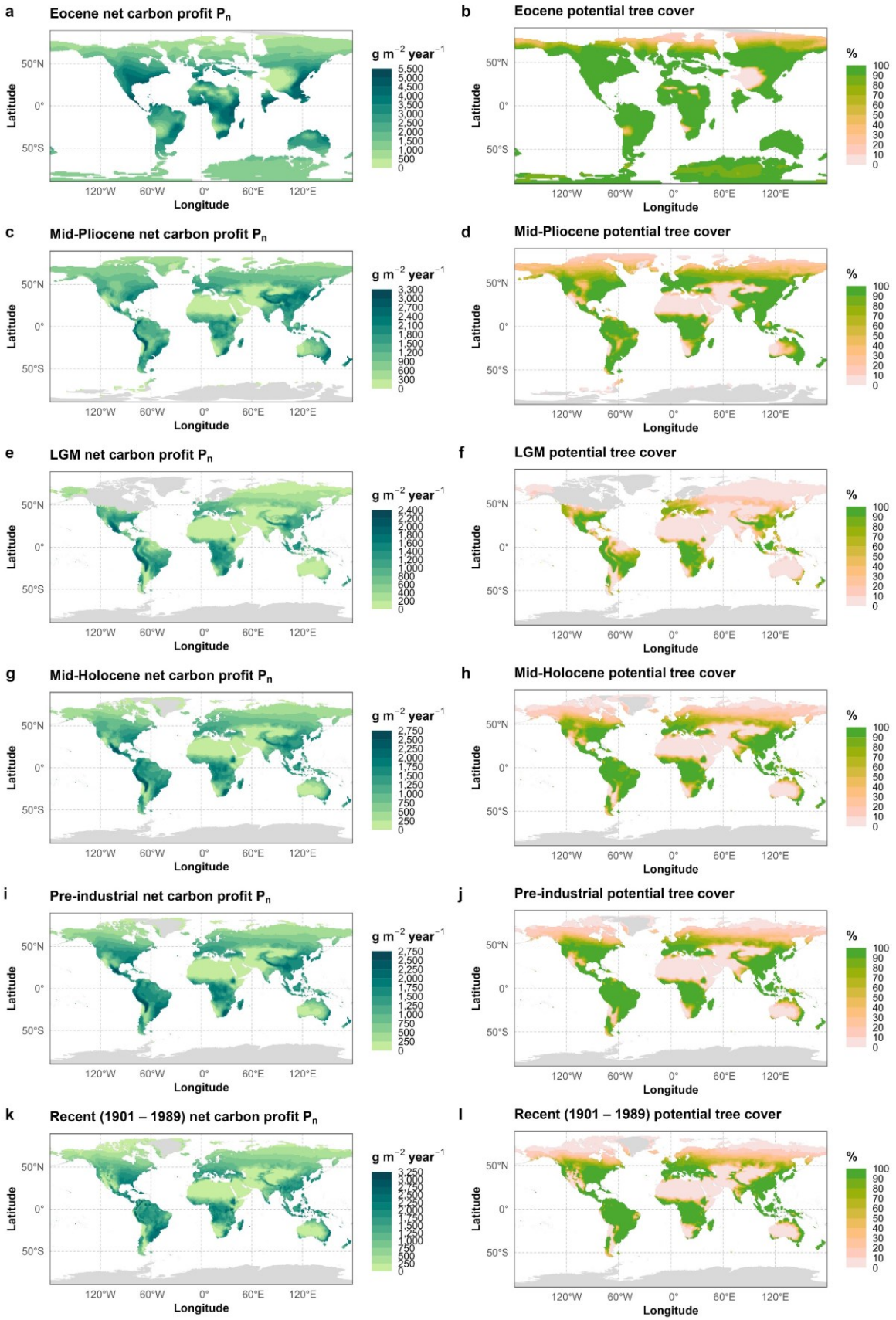
Global patterns of non-Poales C4 diversity in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). The climatic space is divided into 40 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. All metrics are calculated based on native distributions of 1,432 non-Poales C4 species and a time-calibrated phylogeny comprising all non-Poales C4 species. **a**, Species richness of non-Poales C4, i.e. number of non-Poales C4 species per climatic bin. **b**, C4 diversity fraction (%), calculated as the proportion of local C4 species that do not belong to Poales. **c**, Phylogenetic diversity (Faith's PD, Ma)[100]. **d**, Relative phylogenetic diversity (RPD, Ma), calculated as the residual from a PD-richness regression. Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

Extended Data Fig. 4.5 Simulating potential tree cover based on net carbon profit of trees.



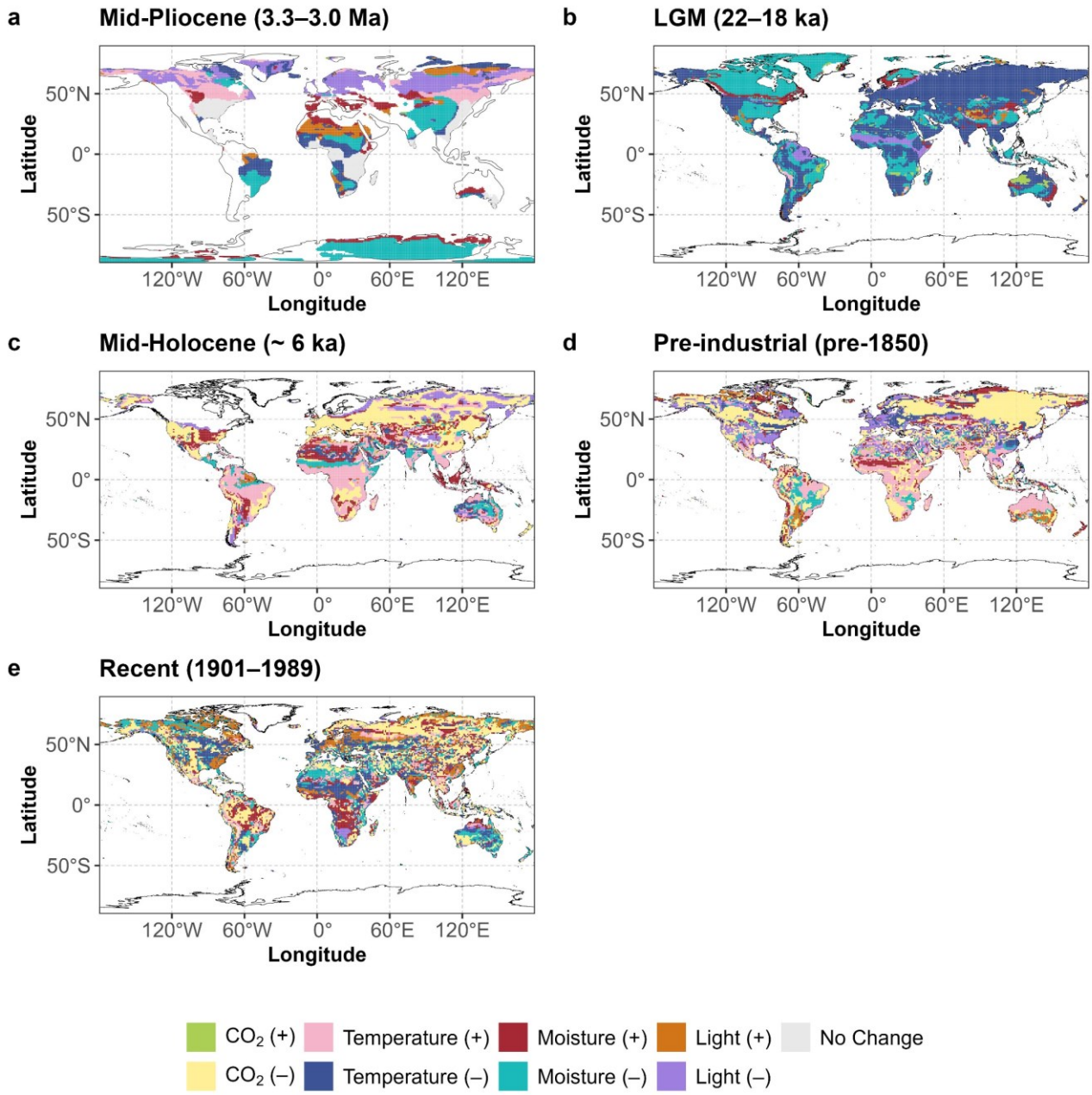
a, Reference net carbon profit (P_n) for trees, calculated using environmental data from 2001–2010 (see Methods). **b**, Reference potential tree cover from ref. [314]. **c**, Logistic function fitted to the empirical relationship between P_n and observed tree cover at the reference period, visualised as red curve (equation and correlation coefficient, *r*, are shown). **d**, Modelled potential tree cover estimated by applying the fitted logistic model to P_n across global geographic grid cells under reference environmental conditions.

Extended Data Fig. 4.6 Simulated net carbon profit for trees and tree cover across geological time slices.



Net carbon profit (P_n) for trees and the corresponding potential tree cover across six representative time periods. P_n (left column) is estimated for each period using the eco-evolutionary optimality (EEO) models and environmental inputs from either model simulations or instrumental datasets specific to that time slice (see Methods). Potential tree cover (right column) is derived by applying the logistic relationship between P_n and tree cover (see Extended Data Fig. 4.5c) to the simulated P_n. Panels correspond to: **a**–**b**, Eocene (*ca.* 50 Ma); **c**–**d**, Mid-Pliocene (3.3–3.0 Ma); **e**–**f**, Last Glacial Maximum (LGM, 22–18 ka); **g**–**h**, Mid-Holocene (*ca.* 6 ka); **i**–**j**, pre-industrial period (pre-1850); **k**–**l**, recent period (1901–1989).

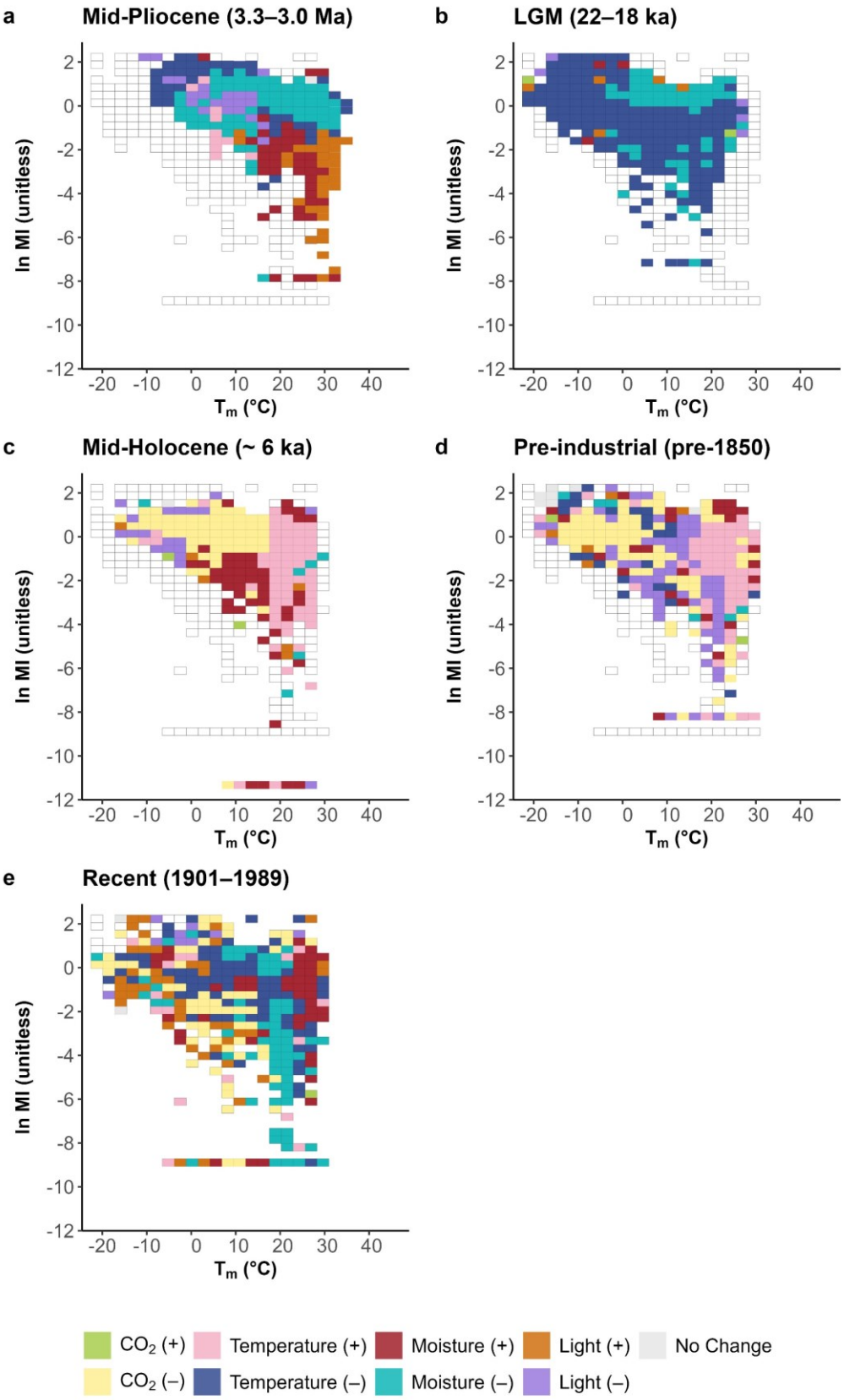
Extended Data Fig. 4.7 Second dominant environmental drivers of simulated C4 plant biogeography in geographic space.



Spatial patterns of the second dominant environmental driver influencing changes in modelled C4 plant fraction, mapped at 1° spatial resolution. Independent driver contribution is calculated by Stein-Alpert decomposition of model simulations. Panels correspond to: **a**, Mid-Pliocene (3.3–3.0 Ma); **b**, Last Glacial Maximum (LGM, 22–18 ka); **c**, Mid-Holocene (*ca.* 6 ka); **d**, pre-industrial period (pre-1850); **e**, recent period (1901–1989). Each grid cell is assigned to the environmental variable

(i.e. CO_2, temperature, moisture and light) that exerted the second strongest net effect on simulated C4 fraction changes. Colour hues denote the identity of the second dominant driver; (±) indicates whether the change in that driver increased (+) or decreased (–) the local C4 fraction. Grey indicates no detectable change.

Extended Data Fig. 4.8 Second dominant environmental drivers of simulated C4 plant biogeography in climatic space.



Spatial patterns of the second dominant environmental driver influencing changes in modelled C4 plant fraction, mapped in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). The climatic space is divided into 40 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. Independent driver contribution is calculated by Stein-Alpert decomposition of model simulations. Panels correspond to: **a**, Mid-Pliocene (3.3–3.0 Ma); **b**, Last Glacial Maximum (LGM, 22–18 ka); **c**, Mid-Holocene (*ca.* 6 ka); **d**, pre-industrial period (pre-1850); **e**, recent period (1901–1989). Each climatic bin is assigned to the environmental variable (i.e. CO_2, temperature, moisture and light) that exerted the second strongest net effect on simulated C4 fraction changes. Colour hues denote the identity of the second dominant driver; (±) indicates whether the change in that driver increased (+) or decreased (–) the local C4 fraction. Grey indicates no detectable change. Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

Chapter 5: Conclusion

5.1 Key findings

Climate governs the assembly of plant communities by filtering for specific trait combinations that enhance survival and reproduction. This thesis show that this process operates differently across plant groups (**Chapter 2**). Global plant trait variation is structured along two universal and orthogonal dimensions[63]: one defined by plant size and the other by the leaf economics spectrum[62]. However, the way in which different functional groups (non-woody, woody deciduous and evergreen) respond to climatic gradients reflects their contrasting life-history strategies. For example, leaf area in non-woody and woody deciduous plants is mainly associated with higher moisture availability, while in woody evergreens it is closely related to warmer winter temperature. These differences arise because the life-history strategy of a functional group dictates the types of climatic stress it experiences and thus the traits required to cope with them. Woody evergreen plants, by retaining leaves year-round, must withstand stresses outside the growing season, potentially favouring traits that confer tolerance to frost and drought[139,164,169]. In contrast, both woody deciduous and non-woody plants restrict their activity to growing seasons. Deciduous trees achieve this by shedding their leaves[91,168,316], while non-woody plants complete their life cycle and survive harsh conditions as seeds (annuals) or die back to protected underground organs (perennials)[157,165,166]. Both strategies favour traits that maximise rapid growth and resource acquisition. Consequently, divergent trait-trait and trait-climate relationships emerge among groups, representing alternative adaptive solutions to environmental challenges. This differential selection across functional groups provides the mechanistic foundation of community assembly: it determines which adaptive strategies persist under particular climatic conditions and thereby shapes the functional composition of ecosystems.

The operation of this adaptive mechanism through geological time leads to the central finding of this thesis (**Chapter 3**): global biodiversity patterns cannot be explained as a straightforward response to intrinsic properties of modern climate alone, but instead represent a historical legacy of

the spatio-temporal dynamics of Earth's climates. This thesis shows that, beyond the influence of contemporary climate, the deep-time geographic dynamics of climatic conditions – their origin, geographic extent, and mobility – better account for present-day plant diversity. The most species-rich and evolutionarily diverse communities are consistently associated with climates that originated early, maintained broad geographic ranges over long periods, while undergoing only moderate shifts in location. Such antiquity, persistence and extent provide the necessary time and space for speciation and the preservation of lineages[32,218]. At the same time, moderate shifts in distribution create opportunities for population isolation and allopatric speciation, further promoting diversification[218].

The biogeographic history of C4 plants provides an example of how these principles operate (**Chapter 4**). This chapter supports the widely accepted hypothesis that the Cenozoic decline in atmospheric CO_2 acted as the primary driver of C4 photosynthesis – a key functional innovation – while subsequent CO_2 fluctuations governed its global expansion and contraction[38,126]. Declining CO_2 facilitated the radiation of C4 plants beyond their ancestral hot-arid niche, whereas rising CO_2 after the Last Glacial Maximum caused regional contractions and a retreat towards hot-arid environments. These deep-time climatic dynamics produced two distinct contemporary biogeographic patterns of C4 plants. The first are hot-arid 'refugia', which are characterised by persistently high C4 abundance since the Mid-Pliocene, low species richness but a high fraction of C4 diversity, and high, positive relative phylogenetic diversity (RPD). The second are 'radiation centres' with the highest species richness but neutral or negative RPD. This case study illustrates how a long-term environmental driver selected for a specific functional adaptation, and how the legacy of CO_2-driven climatic shifts has constrained subsequent global distribution, resulting in the distinct biogeographic patterns observed today.

5.2 Limitations and future directions

5.2.1 Data limitations

Large-scale and deep-time biogeographical analyses are fundamentally constrained by data availability. Although major progress has been made in assembling biodiversity and environmental datasets at global scales, substantial gaps remain.

First, plant functional trait data remain limited in both taxonomic and geographic coverage. Only a few leaf traits are represented by species-level trait measurements, and these traits are present in a minority of species, often less than 10% of known vascular plants[85]. The geographic coverage of community-weighted mean trait data from sPlotOpen is uneven geographically, with undersampling in regions of high latitude and tropical climates. Such biases can introduce uncertainty into trait-climate relationships by overrepresenting temperate vegetation. Furthermore, these community-weighted means do not always include all species within a plot because many plots only record dominant or woody taxa. Consequently, community-weighted means may not fully characterise the local trait structure of undersampled vegetation types.

Second, palaeoenvironmental reconstructions remain limited by coarse spatial resolution. At coarse-grained scales, tropical montane climatic gradients cannot be resolved, and tropical high-elevation climates are frequently merged with adjacent lowland conditions. Yet tropical montane climates are often more similar to temperate climates than to lowland tropics. Such simplifications constrain the representation of environmental heterogeneity in complex landscapes.

Third, although this thesis compiled the most comprehensive dataset of native plant distributions currently available by integrating multiple global and regional sources, these data also reflect incomplete sampling and temporal aggregation. For computational tractability and to enable global comparisons, species ranges were rasterised to 1° resolution and derived from post-1945 occurrence records without explicitly incorporating temporal change. Future analyses could adopt finer spatial grains and consider the temporal dynamics of range shifts, especially for taxa with well-resolved historical biogeography.

Fourth, phylogenetic relationships among vascular plants remain incompletely resolved. This thesis relied on a megatree backbone, with species absent from the backbone grafted to the basal node of their genus, or, when necessary, to calculated positions within their family. Only ~20% of species were present in the backbone tree, and the remainder were added following hierarchical placement. Although deep-time phylogenetic structure at the family and genus levels is most

relevant for large-scale patterns of plant phylogenetic diversity, more complete species-level phylogenies would reduce topological uncertainty and improve the representation of evolutionary relationships among taxa.

Improvements in these data domains will allow future research to combine plant functional, phylogenetic, and biogeographic information with greater spatial and temporal precision, expanding the scope of global biodiversity analysis.

5.2.2 Mechanistic limitations and future research directions

This thesis demonstrates that contemporary trait covariations and trait-climate correlations differ depending on the life-history strategies of major functional groups. The functional trait composition of a region may also reflect its long-term evolutionary and biogeographical history. For instance, a region's historical role as a 'refugium' or a 'radiation centre' may have shaped the adaptive strategies of its resident flora. Future research could test the hypothesis that communities persisting in historically stable climates (refugia) have convergently evolved more conservative functional traits compared with communities from regions that experienced more dynamic climatic histories, even under similar recent climates. Such an approach would extend beyond static trait-climate correlations towards a predictive, evolutionary framework for functional biogeography.

This thesis characterises palaeoclimate primarily using annual means of a temperature index and a moisture index. The temperature index was calculated as the average of the mean temperatures of the coldest and warmest months, rather than treating these seasonal extremes separately as in the analysis of recent climate in Chapter 2. While this simplification captures broad climatic trends, it does not resolve changes in seasonality – a natural consequence of changes in the Earth's orbit and a determinant on the species composition of vegetation[29]. Future work should therefore aim to construct climatic spaces that incorporate seasonal variables, such as the mean temperature of the coldest and warmest months and metrics of seasonal moisture availability. Such an approach would provide a more mechanistic test of phylogenetic niche conservatism, by distinguishing between species' tolerances to seasonal extremes.

Due to data availability, this thesis analysed deep-time climate dynamics using a few selected geological periods rather than continuous palaeoclimate reconstructions. This time-slice approach

successfully captured the influence of long-term climatic trends on global plant diversity patterns. However, this approach does not resolve the potential impact of high-frequency, abrupt climatic events. Such events, like severe droughts or rapid warming episodes, could in principle exceed ecosystem resilience, inducing abrupt and potentially irreversible shifts to alternative stable states[5]. For example, a forest ecosystem may shift to a stable grassland state following an extreme drought that causes widespread tree mortality. This shift is often maintained by new positive feedback loops, such as increased fire frequency and grass competition, which prevent forest recovery. Because of hysteresis, ecosystems may not revert to their original composition even if climate returns to its prior long-term means, resulting in mismatches between climatic potential and realised vegetation states. The duration of such mismatches remains incompletely understood, as it depends on both the frequency of extreme events and the resilience of different ecosystems. A key future direction would therefore be to investigate how climatic variability and extreme events at different temporal scales influence plant function and diversity. As continuous palaeoclimate reconstructions become available, future work will be able to quantify the persistence of such climate-vegetation mismatches and incorporate the frequency and magnitude of abrupt events into the spatio-temporal framework. This approach would help to disentangle the relative roles of long-term mean climatic state shifts in driving macro-evolutionary processes (speciation and extinction) versus short-term shocks in driving community reassembly and ecological state transitions.

5.3 Implications

Understanding how climate shapes the structure and function of terrestrial ecosystems is a foundational scientific question in ecology and biogeography. Long-standing hypotheses link climate to global biodiversity patterns and plant functional trait distributions, yet their interplay and underlying mechanisms remain debated. Progress toward a synthetic understanding has been limited by the absence of comprehensive global datasets and analytical frameworks capable of testing these principles across geological timescales. Recent advances, including the expansion of large-scale ecological databases and increased computational capacity, now make such analyses possible. By compiling some of the most extensive global datasets to date and applying an integrated

spatio-temporal framework across both geographic and climatic spaces, this thesis establishes a new empirical and theoretical foundation for explaining the climatic drivers of plant function and diversity.

The theoretical advances in this thesis are grounded in the compilation of several large-scale ecological datasets. **Chapter 2** produced a global growth form database for 42,676 plant taxa and global maps for 16 key plant functional traits, providing critical empirical inputs for the parameterisation of Dynamic Global Vegetation Models (DGVMs) and Earth System Models (ESMs). **Chapter 3** assembled the most comprehensive distribution dataset of vascular plants, covering 350,864 species (>98% vascular plant species), creating an unprecedented resource for macro-ecological and biogeographical analyses. **Chapter 4** generated the most comprehensive global database of C4 plants, encompassing 7,156 species, thereby filling a major gap for studying the evolution, ecological dominance, and climate responses of this key functional group. Collectively, these datasets provide a lasting foundation for advancing ecological modelling, testing ecological and biogeographical theory, and informing conservation assessments.

These comprehensive datasets enabled the development of a novel analytical framework to investigate the spatio-temporal climate impacts on plant function and diversity across both geographic and climatic spaces. This framework provides a means to evaluate how climatic dynamics shape biodiversity, allowing different hypotheses to be tested within a single integrative approach. For instance, debates concerning the roles of intrinsic climatic properties, climatic history, and climatic geography in shaping global patterns such as the Latitudinal Diversity Gradient (LDG) can now be addressed in combination rather than in isolation. More broadly, the framework offers a means to reconstruct the historical development of functional groups' diversification, linking the biogeographic histories of functional groups with their contemporary diversity patterns. Its wider significance lies in the potential to extend across taxa, traits, and timescales, providing a versatile tool for macroecology, biogeography, and Earth system science.

This research also provides three advances for Earth system modelling. First, it delivers the most comprehensive global maps of plant functional traits generated through statistical upscaling to date, which should help in the development of next-generation DGVMs. Second, by using palaeoclimate inputs, it offers an 'out-of-sample' test for the ESMs used to project future scenarios: a model's

ability to accurately simulate these different past worlds is an important measure of its robustness. Third, the demonstration that modern biodiversity reflects deep-time climatic legacies highlights the need for ESMs to consider historical constraints, potentially enabling more realistic projections of biosphere resilience and adaptive capacity.

Global climate change is ongoing, and its ecological and evolutionary consequences are already evident across terrestrial ecosystems. As new datasets and analytical approaches become available, there is a growing need to apply these advances to questions concerning the future of plant diversity and ecosystem function, and to inform conservation decisions. The spatio-temporal framework developed in this thesis could be extended to assess how future warming and moisture trends may alter the distribution of climatic conditions across the globe. A key question is the degree to which future climates resemble historically favourable (analogue) or unfavourable (no-analogue) conditions. Climates that have remained widespread over deep time may act as reservoirs of evolutionary potential, harbouring ancient genetic diversity and demonstrating resilience to past environmental change, whereas novel no-analogue climates may pose challenges for both species persistence and conservation planning. Comparing future climatic projections with the climatic conditions that have historically supported diversification or long-term persistence would help identify regions where biodiversity may be most vulnerable to the loss or displacement of favourable climates. The use of future climate analogues for ecological restoration or managed relocation would benefit from incorporating information about the long-term size, stability, and connectivity of target climatic areas. Moreover, the eco-evolutionary optimality (EEO) models used in this thesis could also be applied to simulate the future biogeography of C4 plants under global change scenarios. Given the central role of C4 plants in food security, the functioning of grassland and savanna ecosystems, and the dynamics of invasive species, applying such models to future climates could provide valuable insight into how global change may alter the ecological roles and societal relevance of C4 vegetation.

Extending this thesis into forward-time projections would therefore help to link deep-time perspectives with conservation-focused forecasts, thereby bridging the gap between macroevolutionary research and real-world questions concerning ecosystem resilience, climate adaptation and the future of biodiversity. Such projections could also contribute to assessments of

climate-driven risk, conservation and restoration planning, and adaptation strategies under global policy frameworks aimed at safeguarding biodiversity.

Bibliography

1. Humboldt, A. von. Essai sur la géographie des plantes. (1807).
2. Whittaker, R. H. Vegetation of the great smoky mountains. *Ecol. Monogr.* **26**, 2–80 (1956).
3. Sankaran, M. *et al.* Determinants of woody cover in African savannas. *Nature* **438**, 846–849 (2005).
4. Grimm, N. B. *et al.* The impacts of climate change on ecosystem structure and function. *Front. Ecol. Environ.* **11**, 474–482 (2013).
5. Mucina, L. Biome: evolution of a crucial ecological and biogeographical concept. *New Phytol.* **222**, 97–114 (2019).
6. Stephenson, N. L. Climatic control of vegetation distribution: the role of the water balance. *Am. Nat.* **135**, 649–670 (1990).
7. Vitousek, P. M. *et al.* Human alteration of the global nitrogen cycle: sources and consequences. *Ecol. Appl.* **7**, 737–750 (1997).
8. Schimel, D. S. *et al.* Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. *Nature* **414**, 169–172 (2001).
9. Nemani, R. R. *et al.* Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science.* **300**, 1560–1563 (2003).
10. Whittaker, R. H. *Communities and Ecosystems. Macmillan* (1970).
11. Woodward, F. I. *Climate and Plant Distribution*. (Cambridge University Press, 1987).
12. HilleRisLambers, J., Adler, P. B., Harpole, W. S., Levine, J. M. & Mayfield, M. M. Rethinking community assembly through the lens of coexistence theory. *Annu. Rev. Ecol. Evol. Syst.* **43**, 227–248 (2012).
13. Lieth, H. Modeling the primary productivity of the world. in *Primary productivity of the biosphere* 237–263 (Springer, 1975).
14. Prentice, I. C. *et al.* Special paper: a global biome model based on plant physiology and dominance, soil properties and climate. *J. Biogeogr.* 117–134 (1992).

15. Raich, J. W. & Schlesinger, W. H. The global carbon dioxide flux in soil respiration and its relationship to vegetation and climate. *Tellus B* **44**, 81–99 (1992).
16. Richards, P. W. *The Tropical Rain Forest: An Ecological Study.* (Cambridge University Press, 1952).
17. Vitousek, P. M. & Sanford, R. L. Nutrient cycling in moist tropical forest. *Annu. Rev. Ecol. Syst.* 137–167 (1986).
18. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science.* **281**, 237–240 (1998).
19. Gaston, K. J. Global patterns in biodiversity. *Nature* **405**, 220–227 (2000).
20. Pan, Y. *et al.* A large and persistent carbon sink in the world's forests. *Science.* **333**, 988–993 (2011).
21. Malhi, Y., Gardner, T. A., Goldsmith, G. R., Silman, M. R. & Zelazowski, P. Tropical forests in the Anthropocene. *Annu. Rev. Environ. Resour.* **39**, 125–159 (2014).
22. Noy-Meir, I. Desert ecosystems: environment and producers. *Annu. Rev. Ecol. Syst.* 25–51 (1973).
23. Billings, W. D. & Mooney, H. A. The ecology of arctic and alpine plants. *Biol. Rev.* **43**, 481–529 (1968).
24. Hobbie, S. E., Schimel, J. P., Trumbore, S. E. & Randerson, J. R. Controls over carbon storage and turnover in high-latitude soils. *Glob. Chang. Biol.* **6**, 196–210 (2000).
25. Körner, C. & Kèorner, C. *Alpine Plant Life: Functional Plant Ecology of High Mountain Ecosystems*. vol. 4 (Springer, 1999).
26. Davis, M. B. & Shaw, R. G. Range shifts and adaptive responses to Quaternary climate change. *Science.* **292**, 673–679 (2001).
27. Sandel, B. *et al.* The influence of Late Quaternary climate-change velocity on species endemism. *Science.* **334**, 660–664 (2011).
28. Judd, E. J. *et al.* A 485-million-year history of Earth's surface temperature. *Science.* **385**, eadk3705 (2024).
29. Harrison, S. P. *et al.* Paleoclimate Perspectives on Contemporary Climate Change. *Annu. Rev. Environ. Resour.* **50**, (2025).

30. Cohen, K.M., Finney, S.C., Gibbard, P.L. & Fan, J.-X. The ICS International Chronostratigraphic Chart. vol. Episodes 3 199–204 (2024).
31. Dynesius, M. & Jansson, R. Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proc. Natl. Acad. Sci.* **97**, 9115–9120 (2000).
32. Rangel, T. F. *et al.* Modeling the ecology and evolution of biodiversity: Biogeographical cradles, museums, and graves. *Science.* **361**, eaar5452 (2018).
33. Meseguer, A. S. & Condamine, F. L. Ancient tropical extinctions at high latitudes contributed to the latitudinal diversity gradient. *Evolution (N. Y).* **74**, 1966–1987 (2020).
34. Urban, M. C. Climate change extinctions. *Science.* **386**, 1123–1128 (2024).
35. Hewitt, G. The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913 (2000).
36. Cerling, T. E. *et al.* Global vegetation change through the Miocene/Pliocene boundary. *Nature* **389**, 153–158 (1997).
37. Sage, R. F. Environmental and evolutionary preconditions for the origin and diversification of the C4 photosynthetic syndrome. *Plant Biol.* **3**, 202–213 (2001).
38. Sage, R. F. The evolution of C4 photosynthesis. *New Phytol.* **161**, 341–370 (2004).
39. Osborne, C. P. & Beerling, D. J. Nature's green revolution: the remarkable evolutionary rise of C4 plants. *Philos. Trans. R. Soc. B Biol. Sci.* **361**, 173–194 (2006).
40. Christin, P. & Osborne, C. P. The evolutionary ecology of C4 plants. *New Phytol.* **204**, 765–781 (2014).
41. Monson, R. K. *et al.* C4 photosynthesis, trait spectra, and the fast-efficient phenotype. *New Phytol.* **246**, 879–893 (2025).
42. Edwards, E. J. *et al.* The origins of C4 grasslands: integrating evolutionary and ecosystem science. *Science.* **328**, 587–591 (2010).
43. Lloyd, J. & Farquhar, G. D. 13C discrimination during CO2 assimilation by the terrestrial biosphere. *Oecologia* **99**, 201–215 (1994).
44. Still, C. J., Berry, J. A., Collatz, G. J. & DeFries, R. S. Global distribution of C3 and C4 vegetation: carbon cycle implications. *Global Biogeochem. Cycles* **17**, 1–6 (2003).

45. Yan, H. *et al.* Improved global simulations of gross primary product based on a new definition of water stress factor and a separate treatment of C3 and C4 plants. *Ecol. Modell.* **297**, 42–59 (2015).

46. Luo, X. *et al.* Mapping the global distribution of C4 vegetation using observations and optimality theory. *Nat. Commun.* **15**, 1219 (2024).

47. Lavergne, A., Harrison, S., Atsawawaranunt, K., Dong, N. & Prentice, I. Recent C4 vegetation decline is imprinted in atmospheric carbon isotopes. *Res. Sq.* (2024) doi:https://doi.org/10.21203/rs.3.rs-3966918/v1.

48. Morris, J. L. *et al.* The timescale of early land plant evolution. *Proc. Natl. Acad. Sci.* **115**, E2274–E2283 (2018).

49. Gensel, P. G. & Edwards, D. *Plants Invade the Land: Evolutionary and Environmental Perspectives*. (Columbia University Press, 2001).

50. Bar-On, Y. M., Phillips, R. & Milo, R. The biomass distribution on Earth. *Proc. Natl. Acad. Sci.* **115**, 6506–6511 (2018).

51. Beer, C. *et al.* Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. *Science.* **329**, 834–838 (2010).

52. Peng, Y. *et al.* Global terrestrial nitrogen uptake and nitrogen use efficiency. *J. Ecol.* **111**, 2676–2693 (2023).

53. Lenton, T. M. *et al.* Earliest land plants created modern levels of atmospheric oxygen. *Proc. Natl. Acad. Sci.* **113**, 9704–9709 (2016).

54. Beerling, D. J. & Berner, R. A. Feedbacks and the coevolution of plants and atmospheric CO2. *Proc. Natl. Acad. Sci.* **102**, 1302–1305 (2005).

55. Bonan, G. B. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science.* **320**, 1444–1449 (2008).

56. Staal, A. *et al.* Forest-rainfall cascades buffer against drought across the Amazon. *Nat. Clim. Chang.* **8**, 539–543 (2018).

57. Friedlingstein, P. *et al.* Global carbon budget 2024. *Earth Syst. Sci. Data Discuss.* **2024**, 1–133 (2024).

58. Violle, C. *et al.* Let the concept of trait be functional! *Oikos* **116**, 882–892 (2007).

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

60. Funk, J. L. *et al.* Revisiting the Holy Grail: using plant functional traits to understand ecological processes. *Biol. Rev.* **92**, 1156–1173 (2017).

61. Yang, Y. *et al.* Quantifying leaf-trait covariation and its controls across climates and biomes. *New Phytol.* **221**, 155–168 (2019).

62. Wright, I. J. *et al.* The worldwide leaf economics spectrum. *Nature* **428**, 821–827 (2004).

63. Díaz, S. *et al.* The global spectrum of plant form and function. *Nature* **529**, 167–171 (2016).

64. Moritz, C. & Agudo, R. The future of species under climate change: resilience or decline? *Science.* **341**, 504–508 (2013).

65. Madani, N. *et al.* Future global productivity will be affected by plant trait response to climate. *Sci. Rep.* **8**, 1–10 (2018).

66. Diaz, S., Cabido, M. & Casanoves, F. Plant functional traits and environmental filters at a regional scale. *J. Veg. Sci.* **9**, 113–122 (1998).

67. Joswig, J. S. *et al.* Climatic and soil factors explain the two-dimensional spectrum of global plant trait variation. *Nat. Ecol. Evol.* **6**, 36–50 (2022).

68. Dawson, T. P., Jackson, S. T., House, J. I., Prentice, I. C. & Mace, G. M. Beyond predictions: biodiversity conservation in a changing climate. *Science.* **332**, 53–58 (2011).

69. Prentice, I. C. *et al.* Dynamic global vegetation modeling: quantifying terrestrial ecosystem responses to large-scale environmental change. in *Terrestrial Ecosystems in a Changing World* 175–192 (Springer, 2007).

70. Butler, E. E. *et al.* Increasing functional diversity in a global land surface model illustrates uncertainties related to parameter simplification. *J. Geophys. Res. Biogeosciences* **127**, e2021JG006606 (2022).

71. Van Bodegom, P. M. *et al.* Going beyond limitations of plant functional types when predicting global ecosystem–atmosphere fluxes: exploring the merits of traits-based approaches. *Glob. Ecol. Biogeogr.* **21**, 625–636 (2012).

72. Van Bodegom, P. M., Douma, J. C. & Verheijen, L. M. A fully traits-based approach to modeling global vegetation distribution. *Proc. Natl. Acad. Sci.* **111**, 13733–13738 (2014).

73. Homolová, L., Malenovský, Z., Clevers, J. G. P. W., García-Santos, G. & Schaepman, M. E. Review of optical-based remote sensing for plant trait mapping. *Ecol. Complex.* **15**, 1–16 (2013).

74. Moreno-Martínez, Á. *et al.* A methodology to derive global maps of leaf traits using remote sensing and climate data. *Remote Sens. Environ.* **218**, 69–88 (2018).

75. Aguirre-Gutiérrez, J. *et al.* Pantropical modelling of canopy functional traits using Sentinel-2 remote sensing data. *Remote Sens. Environ.* **252**, 112122 (2021).

76. Butler, E. E. *et al.* Mapping local and global variability in plant trait distributions. *Proc. Natl. Acad. Sci.* **114**, E10937–E10946 (2017).

77. Boonman, C. C. F. *et al.* Assessing the reliability of predicted plant trait distributions at the global scale. *Glob. Ecol. Biogeogr.* **29**, 1034–1051 (2020).

78. Schiller, C., Schmidtlein, S., Boonman, C., Moreno-Martínez, A. & Kattenborn, T. Deep learning and citizen science enable automated plant trait predictions from photographs. *Sci. Rep.* **11**, 16395 (2021).

79. Vallicrosa, H. *et al.* Global maps and factors driving forest foliar elemental composition: the importance of evolutionary history. *New Phytol.* (2021).

80. Vallicrosa, H., Sardans, J., Maspons, J. & Peñuelas, J. Global distribution and drivers of forest biome foliar nitrogen to phosphorus ratios (N:P). *Glob. Ecol. Biogeogr.* **31**, 861–871 (2022).

81. Dong, N., Dechant, B., Wang, H., Wright, I. J. & Prentice, I. C. Global leaf-trait mapping based on optimality theory. *Glob. Ecol. Biogeogr.* **32**, 1152–1162 (2023).

82. Bruelheide, H. *et al.* sPlot–A new tool for global vegetation analyses. *J. Veg. Sci.* **30**, 161–186 (2019).

83. Sabatini, F. M. *et al.* sPlotOpen-An environmentally balanced, open-access, global dataset of vegetation plots. *Glob. Ecol. Biogeogr.* **30**, 1740–1764 (2021).

84. Kattge, J. *et al.* TRY plant trait database-enhanced coverage and open access. *Glob. Chang. Biol.* **26**, 119–188 (2020).

85. Díaz, S. *et al.* The global spectrum of plant form and function: enhanced species-level trait dataset. *Sci. Data* **9**, 755 (2022).

86. Wolf, S. *et al.* Citizen science plant observations encode global trait patterns. *Nat. Ecol. Evol.* 1–10 (2022).

87. Dechant, B. *et al.* Intercomparison of global foliar trait maps reveals fundamental differences and limitations of upscaling approaches. *Remote Sens. Environ.* **311**, 114276 (2024).

88. Šímová, I. *et al.* Spatial patterns and climate relationships of major plant traits in the New World differ between woody and herbaceous species. *J. Biogeogr.* **45**, 895–916 (2018).

89. Kikuzawa, K., Onoda, Y., Wright, I. J. & Reich, P. B. Mechanisms underlying global temperature-related patterns in leaf longevity. *Glob. Ecol. Biogeogr.* **22**, 982–993 (2013).

90. Reich, P. B., Rich, R. L., Lu, X., Wang, Y.-P. & Oleksyn, J. Biogeographic variation in evergreen conifer needle longevity and impacts on boreal forest carbon cycle projections. *Proc. Natl. Acad. Sci.* **111**, 13703–13708 (2014).

91. Wang, H. *et al.* Leaf economics fundamentals explained by optimality principles. *Sci. Adv.* **9**, eadd5667 (2023).

92. Laliberté, E., Zemunik, G. & Turner, B. L. Environmental filtering explains variation in plant diversity along resource gradients. *Science.* **345**, 1602–1605 (2014).

93. Kraft, N. J. B., Godoy, O. & Levine, J. M. Plant functional traits and the multidimensional nature of species coexistence. *Proc. Natl. Acad. Sci.* **112**, 797–802 (2015).

94. Holden, E. M. & Cahill Jr, J. F. Plant trait dissimilarity increases competitive interactions among co-occurring plants. *Funct. Ecol.* **38**, 1464–1474 (2024).

95. Suding, K. N. *et al.* Scaling environmental change through the community-level: A trait-based response-and-effect framework for plants. *Glob. Chang. Biol.* **14**, 1125–1140 (2008).

96. Pausas, J. G. & Keeley, J. E. Evolutionary ecology of resprouting and seeding in fire-prone ecosystems. *New Phytol.* **204**, 55–65 (2014).

97. Gotelli, N. J. & Colwell, R. K. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.* **4**, 379–391 (2001).

98. Tilman, D. *et al.* The influence of functional diversity and composition on ecosystem processes. *Science.* **277**, 1300–1302 (1997).

99. Díaz, S. & Cabido, M. Vive la différence: plant functional diversity matters to ecosystem processes. *Trends Ecol. Evol.* **16**, 646–655 (2001).

100. Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10 (1992).

101. Whittaker, R. H. Evolution and measurement of species diversity. *Taxon* **21**, 213–251 (1972).

102. Magurran, A. E. Measuring biological diversity. *Curr. Biol.* **31**, R1174–R1177 (2021).

103. Le Bagousse-Pinguet, Y. *et al.* Phylogenetic, functional, and taxonomic richness have both positive and negative effects on ecosystem multifunctionality. *Proc. Natl. Acad. Sci.* **116**, 8419–8424 (2019).

104. Hawkins, B. A. Ecology's oldest pattern? *Trends Ecol. Evol.* **16**, 470 (2001).

105. Wiens, J. J. & Donoghue, M. J. The origins of the latitudinal diversity gradient: revisiting the tropical conservatism hypothesis. *J. Biogeogr.* e15172 (2025).

106. Willig, M. R., Kaufman, D. M. & Stevens, R. D. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annu. Rev. Ecol. Evol. Syst.* **34**, 273–309 (2003).

107. Turner, J. R. G. Explaining the global biodiversity gradient: energy, area, history and natural selection. *Basic Appl. Ecol.* **5**, 435–448 (2004).

108. Forster, J. R. *Observations Made During a Voyage Round the World*. (London, G. Robinson, 1778). doi:https://doi.org/10.5962/bhl.title.50538.

109. Mittelbach, G. G. *et al.* Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecol. Lett.* **10**, 315–331 (2007).

110. Fine, P. V. A. Ecological and evolutionary drivers of geographic variation in species diversity. *Annu. Rev. Ecol. Evol. Syst.* **46**, 369–392 (2015).

111. Wright, D. H. Species-energy theory: an extension of species-area theory. *Oikos* 496–506 (1983).

112. Allen, A. P., Brown, J. H. & Gillooly, J. F. Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science.* **297**, 1545–1548 (2002).

113. Hawkins, B. A. *et al.* Energy, water, and broad-scale geographic patterns of species richness. *Ecology* **84**, 3105–3117 (2003).

114. Wallace, A. R. *Tropical Nature, and Other Essays*. (Macmillan and Company, 1878).

115. Fischer, A. G. Latitudinal variations in organic diversity. *Evolution (N. Y).* **14**, 64–81 (1960).

116. Wiens, J. J. & Donoghue, M. J. Historical biogeography, ecology and species richness. *Trends Ecol. Evol.* **19**, 639–644 (2004).

117. Brown, S. C., Wigley, T. M. L., Otto-Bliesner, B. L., Rahbek, C. & Fordham, D. A. Persistent Quaternary climate refugia are hospices for biodiversity in the Anthropocene. *Nat. Clim. Chang.* **10**, 244–248 (2020).

118. Terborgh, J. On the notion of favorableness in plant ecology. *Am. Nat.* **107**, 481–501 (1973).

119. Rosenzweig, M. L. *Species Diversity in Space and Time*. (Cambridge University Press, 1995). doi:https://doi.org/10.1017/CBO9780511623387.

120. Storch, D. & Okie, J. G. The carrying capacity for species richness. *Glob. Ecol. Biogeogr.* **28**, 1519–1532 (2019).

121. Coelho, M. T. P. *et al.* The geography of climate and the global patterns of species diversity. *Nature* **622**, 537–544 (2023).

122. Hatch, M. D. C4 photosynthesis: a unique elend of modified biochemistry, anatomy and ultrastructure. *Biochim. Biophys. Acta (BBA)-Reviews Bioenerg.* **895**, 81–106 (1987).

123. Langdale, J. A. C4 cycles: past, present, and future research on C4 photosynthesis. *Plant Cell* **23**, 3879–3892 (2011).

124. Chollet, R. & Ogren, W. L. Regulation of photorespiration in C3 and C4 species. *Bot. Rev.* **41**, 137–179 (1975).

125. Pagani, M., Zachos, J. C., Freeman, K. H., Tipple, B. & Bohaty, S. Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. *Science.* **309**, 600–603 (2005).

126. Palazzesi, L., Hidalgo, O., Barreda, V. D., Forest, F. & Höhna, S. The rise of grasslands is linked to atmospheric CO2 decline in the late Palaeogene. *Nat. Commun.* **13**, 293 (2022).

127. Sage, R. F. & Monson, R. K. *C4 Plant Biology*. (Academic Press, 1998).

128. Sage, R. F. A portrait of the C4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. vol. 67. *J Exp Bot* 4039–4056 (2016).

129. Soberón, J. & Peterson, A. T. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodivers. Informatics* **2**, 1–10 (2005).

130. Peterson, A. T. *et al. Ecological Niches and Geographic Distributions (MPB-49)*. (Princeton University Press, 2011). doi:https://www.jstor.org/stable/j.ctt7stnh.

131. Graham, C. H. *et al.* Biodiversity patterns redefined in environmental space. *Ecol. Lett.* **28**, e70008 (2025).

132. Grinnell, J. The niche-relationships of the California Thrasher. *Auk* **34**, 427–433 (1917).

133. Hutchinson, G. E. Concluding Remarks. *Cold Spring Harb. Symp. Quant. Biol.* **22**, 415–427 (1957).

134. Colwell, R. K. & Rangel, T. F. Hutchinson's duality: the once and future niche. *Proc. Natl. Acad. Sci.* **106**, 19651–19658 (2009).

135. Reich, P. B. Key canopy traits drive forest productivity. *Proc. R. Soc. B Biol. Sci.* **279**, 2128–2134 (2012).

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

137. Kraft, N. J. B. *et al.* Community assembly, coexistence and the environmental filtering metaphor. *Funct. Ecol.* **29**, 592–599 (2015).

138. Bruelheide, H. *et al.* Global trait–environment relationships of plant communities. *Nat. Ecol. Evol.* **2**, 1906–1917 (2018).

139. Thomas, H. J. D. *et al.* Global plant trait relationships extend to the climatic extremes of the tundra biome. *Nat. Commun.* **11**, 1351 (2020).

140. Maynard, D. S. *et al.* Global relationships in tree functional traits. *Nat. Commun.* **13**, 3185 (2022).

141. Wright, I. J. *et al.* Relationships among ecologically important dimensions of plant trait variation in seven Neotropical forests. *Ann. Bot.* **99**, 1003–1015 (2007).

142. Swenson, N. G. The assembly of tropical tree communities–the advances and shortcomings of phylogenetic and functional trait analyses. *Ecography.* **36**, 264–276 (2013).

143. Williams, J. W., Jackson, S. T. & Kutzbach, J. E. Projected distributions of novel and disappearing climates by 2100 AD. *Proc. Natl. Acad. Sci.* **104**, 5738–5742 (2007).

144. Yang, Y. *et al.* A novel approach for modelling vegetation distributions and analysing vegetation sensitivity through trait-climate relationships in China. *Sci. Rep.* **6**, 24110 (2016).

145. Chave, J. *et al.* Towards a worldwide wood economics spectrum. *Ecol. Lett.* **12**, 351–366 (2009).

146. Moles, A. T. *et al.* Global patterns in plant height. *J. Ecol.* **97**, 923–932 (2009).

147. Moles, A. T. *et al.* Global patterns in seed size. *Glob. Ecol. Biogeogr.* **16**, 109–116 (2007).

148. Craine, J. M. *et al.* Ecological interpretations of nitrogen isotope ratios of terrestrial plants and soils. *Plant Soil* **396**, 1–26 (2015).

149. Dong, N. *et al.* Leaf nitrogen from the perspective of optimal plant function. *J. Ecol.* **110**, 2585–2602 (2022).

150. Dong, N. *et al.* Leaf nitrogen from first principles: field evidence for adaptive variation with climate. *Biogeosciences* **14**, 481–495 (2017).

151. R Core Team, R. F. for S. C. R: A Language and Environment for Statistical Computing. (2022).

152. Harrison, S. P. *et al.* Ecophysiological and bioclimatic foundations for a global plant functional classification. *J. Veg. Sci.* **21**, 300–317 (2010).

153. Wright, I. J. *et al.* Modulation of leaf economic traits and trait relationships by climate. *Glob. Ecol. Biogeogr.* **14**, 411–421 (2005).

154. Wieczynski, D. J. *et al.* Climate shapes and shifts functional biodiversity in forests worldwide. *Proc. Natl. Acad. Sci.* **116**, 587–592 (2019).

155. Dong, N. *et al.* Components of leaf-trait variation along environmental gradients. *New Phytol.* **228**, 82–94 (2020).

156. Thomson, F. J., Moles, A. T., Auld, T. D. & Kingsford, R. T. Seed dispersal distance is more strongly correlated with plant height than with seed mass. *J. Ecol.* **99**, 1299–1307 (2011).

157. Adler, P. B. *et al.* Functional traits explain variation in plant life history strategies. *Proc. Natl. Acad. Sci.* **111**, 740–745 (2014).

158. Poorter, H., Niinemets, Ü., Poorter, L., Wright, I. J. & Villar, R. Causes and consequences of variation in leaf mass per area (LMA): a meta-analysis. *New Phytol.* **182**, 565–588 (2009).

159. Osmond, C. B., Smith, S. D., Gui-Ying, B. & Sharkey, T. D. Stem Photosynthesis in a Desert Ephemeral, Eriogonum inflatum. Characterization of Leaf and Stem CO_2 Fixation and H_2O Vapor Exchange under Controlled Conditions. *Oecologia* 542–549 (1987).

160. Ávila, E., Herrera, A. & Tezara, W. Contribution of stem CO 2 fixation to whole-plant carbon balance in nonsucculent species. *Photosynthetica* **52**, 3–15 (2014).

161. Ávila-Lovera, E., Zerpa, A. J. & Santiago, L. S. Stem photosynthesis and hydraulics are coordinated in desert plant species. *New Phytol.* **216**, 1119–1129 (2017).

162. Ricklefs, R. E. & Latham, R. E. Intercontinental correlation of geographical ranges suggests stasis in ecological traits of relict genera of temperate perennial herbs. *Am. Nat.* **139**, 1305–1321 (1992).

163. Reich, P. B. The world-wide 'fast–slow'plant economics spectrum: a traits manifesto. *J. Ecol.* **102**, 275–301 (2014).

164. Zanne, A. E. *et al.* Functional biogeography of angiosperms: life at the extremes. *New Phytol.* **218**, 1697–1709 (2018).

165. Westoby, M., Falster, D. S., Moles, A. T., Vesk, P. A. & Wright, I. J. Plant ecological strategies: some leading dimensions of variation between species. *Annu. Rev. Ecol. Syst.* **33**, 125–159 (2002).

166. Coomes, D. A. & Grubb, P. J. Colonization, tolerance, competition and seed-size variation within functional groups. *Trends Ecol. Evol.* **18**, 283–291 (2003).

167. Cavender-Bares, J. & Holbrook, N. M. Hydraulic properties and freezing-induced cavitation in sympatric evergreen and deciduous oaks with contrasting habitats. *Plant. Cell Environ.* **24**, 1243–1256 (2001).

168. Mota-Gutiérrez, D. *et al.* Seasonal variation in native hydraulic conductivity between two deciduous oak species. *J. Plant Ecol.* **13**, 78–86 (2020).

169. Wright, I. J. *et al.* Global climatic drivers of leaf size. *Science.* **357**, 917–921 (2017).

170. Choat, B., Ball, M. C., Luly, J. G. & Holtum, J. A. M. Hydraulic architecture of deciduous and evergreen dry rainforest tree species from north-eastern Australia. *Trees* **19**, 305–311 (2005).

171. Ordoñez, J. C. *et al.* A global study of relationships between leaf traits, climate and soil measures of nutrient fertility. *Glob. Ecol. Biogeogr.* **18**, 137–149 (2009).

172. Maire, V. *et al.* Global effects of soil and climate on leaf photosynthetic traits and rates. *Glob. Ecol. Biogeogr.* **24**, 706–717 (2015).

173. Simpson, A. H., Richardson, S. J. & Laughlin, D. C. Soil–climate interactions explain variation in foliar, stem, root and reproductive traits across temperate forests. *Glob. Ecol. Biogeogr.* **25**, 964–978 (2016).

174. Chen, S. *et al.* Plant diversity enhances productivity and soil carbon storage. *Proc. Natl. Acad. Sci.* **115**, 4027–4032 (2018).

175. Scheiter, S. *et al.* Climate change promotes transitions to tall evergreen vegetation in tropical Asia. *Glob. Chang. Biol.* **26**, 5106–5124 (2020).

176. Siefert, A. *et al.* A global meta-analysis of the relative extent of intraspecific trait variation in plant communities. *Ecol. Lett.* **18**, 1406–1419 (2015).

177. Osnas, J. L. D., Lichstein, J. W., Reich, P. B. & Pacala, S. W. Global leaf trait relationships: mass, area, and the leaf economics spectrum. *Science.* **340**, 741–744 (2013).

178. Peng, Y., Bloomfield, K. J., Cernusak, L. A., Domingues, T. F. & Colin Prentice, I. Global climate and nutrient controls of photosynthetic capacity. *Commun. Biol.* **4**, 462 (2021).

179. Boyle, B. L. *et al.* Taxonomic Name Resolution Service, version 5.3, Botanical Information and Ecology Network. (2025) doi:https://tnrs.biendata.org/.

180. Prentice, I. C., Cleator, S. F., Huang, Y. H., Harrison, S. P. & Roulstone, I. Reconstructing ice-age palaeoclimates: Quantifying low-CO2 effects on plants. *Glob. Planet. Change* **149**, 166–176 (2017).

181. Karger, D. N. *et al.* Climatologies at high resolution for the earth's land surface areas. (2021) doi:10.16904/envidat.228.

182. Karger, D. N. *et al.* Climatologies at high resolution for the earth's land surface areas. *Sci. data* **4**, 1–20 (2017).

183. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).

184. Davis, T. W. *et al.* Simple process-led algorithms for simulating habitats (SPLASH v. 1.0): robust indices of radiation, evapotranspiration and plant-available moisture. *Geosci. Model Dev.* **10**, 689–708 (2017).

185. Sandoval, D., Prentice, I. C. & Nóbrega, R. L. B. Simple process-led algorithms for simulating habitats (SPLASH v. 2.0): robust calculations of water and energy fluxes. *Geosci. Model Dev.* **17**, 4229–4309 (2024).

186. Defourny, P., Lamarche, C., Brockmann, C., Boettcher, M., Bontemps, S., De Maet, T., Duveiller, G. L. Harper, K., Hartley A., Kirches, G., Moreau, I., Peylin, P., Ottlé, C., Radoux J., Van Bogaert, E., Ramoino, F., Albergel, C., and Arino, O. Observed annual global land-use change from 1992 to 2020 three times more dynamic than reported by inventory-based statistics. *Prep.* (2023).

187. Oksanen, J. *et al.* Package 'vegan'. *R package* vol. 2 1–295 (2013).

188. Wood, S. & Wood, M. S. Package 'mgcv'. *R package* vol. 1 729 (2015).

189. Roussel, J.-R. *et al.* Package 'geometry'. *R package* (2023).

190. Hijmans, R. J. *et al.* Package 'raster'. *R package* vol. 734 473 (2015).

191. Norberg, J., Urban, M. C., Vellend, M., Klausmeier, C. A. & Loeuille, N. Eco-evolutionary responses of biodiversity to climate change. *Nat. Clim. Chang.* **2**, 747–751 (2012).

192. Saupe, E. E. *et al.* Spatio-temporal climate change contributes to latitudinal diversity gradients. *Nat. Ecol. Evol.* **3**, 1419–1429 (2019).

193. Wiens, J. J. *et al.* Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters* vol. 13 1310–1324 (2010).

194. Kreft, H. & Jetz, W. Global patterns and determinants of vascular plant diversity. *Proc. Natl. Acad. Sci.* **104**, 5925–5930 (2007).

195. Tordoni, E., Carmona, C. P., Toussaint, A., Tamme, R. & Pärtel, M. Global patterns and determinants of multiple facets of plant diversity. *Glob. Ecol. Biogeogr.* **33**, e13823 (2024).

196. Thompson, L. R. *et al.* A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457–463 (2017).

197. Hutchinson, G. E. Homage to Santa Rosalia or why are there so many kinds of animals? *Am. Nat.* **93**, 145–159 (1959).

198. Currie, D. J. Energy and large-scale patterns of animal-and plant-species richness. *Am. Nat.* **137**, 27–49 (1991).

199. Evans, K. L., Warren, P. H. & Gaston, K. J. Species–energy relationships at the macroecological scale: a review of the mechanisms. *Biol. Rev.* **80**, 1–25 (2005).

200. Fine, P. V. A. & Ree, R. H. Evidence for a time-integrated species-area effect on the latitudinal gradient in tree diversity. *Am. Nat.* **168**, 796–804 (2006).

201. Schneider, H. *et al.* Ferns diversified in the shadow of angiosperms. *Nature* **428**, 553–557 (2004).

202. Hughes, C. E., Nyffeler, R. & Linder, H. P. Evolutionary plant radiations: where, when, why and how? *New Phytol.* **207**, 249–253 (2015).

203. Tietje, M. *et al.* Global hotspots of plant phylogenetic diversity. *New Phytol.* **240**, 1636–1646 (2023).

204. Zuntini, A. R. *et al.* Phylogenomics and the rise of the angiosperms. *Nature* **629**, 843–850 (2024).

205. Gordon, J. D., Fagan, B., Milner, N. & Thomas, C. D. Floristic diversity and its relationships with human land use varied regionally during the Holocene. *Nat. Ecol. Evol.* **8**, 1459–1471 (2024).

206. Li, J. & Prentice, I. C. Global patterns of plant functional traits and their relationships to climate. *Commun. Biol.* **7**, 1136 (2024).

207. Ghalambor, C. K., Huey, R. B., Martin, P. R., Tewksbury, J. J. & Wang, G. Are mountain passes higher in the tropics? Janzen's hypothesis revisited. *Integr. Comp. Biol.* **46**, 5–17 (2006).

208. Samset, B. H. *et al.* Steady global surface warming from 1973 to 2022 but increased warming rate after 1990. *Commun. Earth Environ.* **4**, 400 (2023).

209. Govaerts, R., Nic Lughadha, E., Black, N., Turner, R. & Paton, A. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Sci. data* **8**, 215 (2021).

210. Cai, L. *et al.* Global models and predictions of plant diversity based on advanced machine learning techniques. *New Phytol.* **237**, 1432–1445 (2023).

211. Rafiqpoor, D., Kier, G. & Kreft, H. Global centers of vascular plant diversity. *Nov. Acta Leopoldina NF* **92**, 61–83 (2005).

212. Kier, G. *et al.* Global patterns of plant diversity and floristic knowledge. *J. Biogeogr.* **32**, 1107–1116 (2005).

213. Weigelt, P., König, C. & Kreft, H. GIFT–A global inventory of floras and traits for macroecology and biogeography. *J. Biogeogr.* **47**, 16–43 (2020).

214. Brummitt, N., Araújo, A. C. & Harris, T. Areas of plant diversity—What do we know? *Plants, People, Planet* **3**, 33–44 (2021).

215. Sabatini, F. M. *et al.* Global patterns of vascular plant alpha diversity. *Nat. Commun.* **13**, 4683 (2022).

216. Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).

217. Cai, L. *et al.* Environmental filtering, not dispersal history, explains global patterns of phylogenetic turnover in seed plants at deep evolutionary timescales. *Nat. Ecol. Evol.* **9**, 314–324 (2025).

218. Stebbins, G. L. *Flowering plants: evolution above the species level*. (Harvard University Press, 1974).

219. Haffer, J. Speciation in Amazonian Forest Birds: Most species probably originated in forest refuges during dry climatic periods. *Science.* **165**, 131–137 (1969).

220. Svenning, J.-C., Eiserhardt, W. L., Normand, S., Ordonez, A. & Sandel, B. The influence of paleoclimate on present-day patterns in biodiversity and ecosystems. *Annu. Rev. Ecol. Evol. Syst.* **46**, 551–572 (2015).

221. Corlett, R. T. & Westcott, D. A. Will plant movements keep up with climate change? *Trends Ecol. Evol.* **28**, 482–488 (2013).

222. Smith, J. *et al.* BioDeepTime: A database of biodiversity time series for modern and fossil assemblages. *Glob. Ecol. Biogeogr.* **32**, 1680–1689 (2023).

223. Cai, L. *et al.* Environmental filtering, not dispersal history, explains global patterns of phylogenetic turnover in seed plants at deep evolutionary timescales. *Nat. Ecol. Evol.* 1–11 (2024).

224. Marder, E., Smiley, T. M., Yanites, B. J. & Kravitz, K. Direct effects of mountain uplift and topography on biodiversity. *Science.* **387**, 1287–1291 (2025).

225. Baste, I. A. & Watson, R. T. *Making peace with nature: a scientific blueprint to tackle the climate, biodiversity and pollution emergencies*. *United Nations Environment Programme (UNEP)* https://www.unep.org/resources/making-peace-nature (2021).

226. United Nations Convention on Biological Diversity. *Biodiversity COP 16: Important Agreement Reached Towards Goal of 'Making Peace with Nature'*. https://www.cbd.int/article/agreement-reached-cop-16 (2024).

227. UN Environment Programme. Kunming-Montreal Global Biodiversity Framework (KM-GBF). in *Fifteenth meeting of the Conference of the Parties to the Convention on Biological Diversity (Part Two) Decision 15/4.* (2022).

228. Lawler, J. J., Gross, J. E., Foden, W. B. & Midgley, G. F. Whole-Earth: A Conservation-Planning Paradigm for a Changing Climate. *Glob. Chang. Biol.* **31**, e70328 (2025).

229. Steinig, S.; Abe-Ouchi, A.; de Boer, A.M.; Chan, W.-L.; Donnadieu, Y.; Hutchinson, D.; Knorr, G.; Ladant, J.-B.; Morozova, P.; Niezgodzki, I.; Poulsen, C.J.; Volodin, E.M.; Zhang, Z.; Zhu, J.; Lunt, D. J. Deep-Time Model Intercomparison Project (DeepMIP) Eocene model data version 1.0. *NERC EDS Cent. Environ. Data Anal.* (2024) doi:10.5285/95aa41439d564756950f89921b6ef215.

230. Kageyama, M. *et al.* PMIP4-CMIP6: the contribution of the Paleoclimate Modelling Intercomparison Project to CMIP6. *Geosci. Model Dev. Discuss.* **11**, 1033–1057 (2018).

231. Eyring, V. *et al.* Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).

232. University of East Anglia Climatic Research Unit; Harris, I.C.; Jones, P.D.; Osborn, T. CRU TS4.09: Climatic Research Unit (CRU) Time-Series (TS) version 4.09 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901- Dec. 2024). *NERC EDS Cent. Environ. Data Anal.* (2025).

233. Markwick, P. J. The palaeogeographic and palaeoclimatic significance of climate proxies for data-model comparisons. (2007).

234. Hijmans, R. J. Geographic Data Analysis and Modeling [R package raster version 3.6-30]. (2024).

235. Prentice, I. C., Villegas-Diaz, R. & Harrison, S. P. Accounting for atmospheric carbon dioxide variations in pollen-based reconstruction of past hydroclimates. *Glob. Planet. Change* **211**, 103790 (2022).

236. Harrison, S. P. & Prentice, C. I. Climate and CO2 controls on global vegetation distribution at the last glacial maximum: analysis based on palaeovegetation data, biome modelling and palaeoclimate simulations. *Glob. Chang. Biol.* **9**, 983–1004 (2003).

237. Bragg, F. J. *et al.* Stable isotope and modelling evidence for CO 2 as a driver of glacial–interglacial vegetation shifts in southern Africa. *Biogeosciences* **10**, 2001–2010 (2013).

238. Keenan, T. F. *et al.* Increase in forest water-use efficiency as atmospheric carbon dioxide concentrations rise. *Nature* **499**, 324–327 (2013).

239. Lan, X., Tans, P. and K. W. T. Trends in globally-averaged CO2 determined from NOAA Global Monitoring Laboratory measurements. doi:https://doi.org/10.15138/9N0H-ZH07.

240. Lunt, D. J. *et al.* DeepMIP: Model intercomparison of early Eocene climatic optimum (EECO) large-scale climate features and comparison with proxy data. *Clim. Past* **17**, 203–227 (2021).

241. Anagnostou, E. *et al.* Changing atmospheric CO2 concentration was the primary driver of early Cenozoic climate. *Nature* **533**, 380–384 (2016).

242. Burke, K. D. *et al.* Pliocene and Eocene provide best analogs for near-future climates. *Proc. Natl. Acad. Sci.* **115**, 13288–13293 (2018).

243. Burton, L. E. *et al.* On the climatic influence of CO 2 forcing in the Pliocene. *Clim. Past* **19**, 747–764 (2023).

244. Bouttes, N., Paillard, D., Roche, D. M., Brovkin, V. & Bopp, L. Last Glacial Maximum CO2 and δ13C successfully reconciled. *Geophys. Res. Lett.* **38**, (2011).

245. Bereiter, B. *et al.* Revision of the EPICA Dome C CO2 record from 800 to 600 kyr before present. *Geophys. Res. Lett.* **42**, 542–549 (2015).

246. Villegas-Diaz, S. P. H. and I. C. P. and R. codos: CO2 Correction Tools. (2022).

247. R Core Team. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria.* (2024) doi:https://www.R-project.org/.

248. Chamberlain, S. *et al.* rgbif: Interface to the Global Biodiversity Information Facility API. (2024).

249. Maitner, B. S. *et al.* The bien r package: A tool to access the Botanical Information and Ecology Network (BIEN) database. *Methods Ecol. Evol.* **9**, 373–379 (2018).

250. Maitner, B. BIEN: Tools for Accessing the Botanical Information and Ecology Network Database. *R Package* (2023).

251. Dauby, G. *et al.* RAINBIO: a mega-database of tropical African vascular plants distributions. *PhytoKeys* **74**, 1–18 (2016).

252. Seregin, A. P. *et al.* ' Flora of Russia' on iNaturalist: a dataset. *Biodivers. Data J.* **8**, e59249 (2020).

253. Wickham, H. *et al.* dplyr: A Grammar of Data Manipulation. (2023).

254. Zizka, A. *et al.* CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods Ecol. Evol.* **10**, 744–751 (2019).

255. Poppenwimer, T., Mayrose, I. & DeMalach, N. Revising the global biogeography of annual and perennial plants. *Nature* **624**, 109–114 (2023).

256. Gaston, K. J. & Fuller, R. A. The sizes of species' geographic ranges. *J. Appl. Ecol.* **46**, 1–9 (2009).

257. Keith, D. A. *et al.* Scientific foundations for an IUCN Red List of Ecosystems. *PLoS One* **8**, e62111 (2013).

258. Bland, L. M., Keith, D. A., Miller, R. M., Murray, N. J. & Rodríguez, J. P. *Guidelines for the application of IUCN Red List of Ecosystems Categories and Criteria, version 1.1*. *International Union for the Conservation of Nature, Gland, Switzerland* (2017).

259. Baselga, A. & Orme, C. D. L. betapart: an R package for the study of beta diversity. *Methods Ecol. Evol.* **3**, 808–812 (2012).

260. Jin, Y. & Qian, H. V. PhyloMaker2: An updated and enlarged R package that can generate very large phylogenies for vascular plants. *Plant Divers.* **44**, 335–339 (2022).

261. Smith, S. A. & Brown, J. W. Constructing a broadly inclusive seed plant phylogeny. *Am. J. Bot.* **105**, 302–314 (2018).

262. Zanne, A. E. *et al.* Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**, 89–92 (2014).

263. Jin, Y. & Qian, H. U. PhyloMaker: An R package that can generate large phylogenetic trees for plants and animals. *Plant Divers.* **45**, 347–352 (2023).

264. Forest, F. *et al.* Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* **445**, 757–760 (2007).

265. Davies, T. J. & Buckley, L. B. Phylogenetic diversity as a window into the evolutionary and biogeographic histories of present-day richness gradients for mammals. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 2414–2425 (2011).

266. Wood, S. N. *Generalized additive models: an introduction with R*. (Chapman and Hall/CRC, 2017). doi:https://doi.org/10.1201/9781315370279.

267. Wood, S. mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. (2023) doi:10.32614/CRAN.package.mgcv.

268. Lindeman, R. H., Merenda, P. F. & Gold, R. Z. *Introduction to bivariate and multivariate analysis*. (Glenview, Ill. [etc.] : Scott, Foresman and Comp, 1980).

269. Bivand, R. *et al.* spdep: Spatial dependence: weighting schemes, statistics and models. *R package* (2024).

270. Sage, R. F. & Sharkey, T. D. The effect of temperature on the occurrence of O2 and CO2 insensitive photosynthesis in field grown plants. *Plant Physiol.* **84**, 658–664 (1987).

271. Da, J., Zhang, Y. G., Li, G., Meng, X. & Ji, J. Low CO2 levels of the entire Pleistocene epoch. *Nat. Commun.* **10**, 4342 (2019).

272. Beerling, D. J. & Royer, D. L. Convergent cenozoic CO2 history. *Nat. Geosci.* **4**, 418–420 (2011).

273. Ehleringer, J. & Pearcy, R. W. Variation in quantum yield for CO2 uptake among C3 and C4 plants. *Plant Physiol.* **73**, 555–559 (1983).

274. Ehleringer, J. R., Cerling, T. E. & Helliker, B. R. C4 photosynthesis, atmospheric CO2, and climate. *Oecologia* **112**, 285–299 (1997).

275. Edwards, E. J. & Still, C. J. Climate, phylogeny and the ecological distribution of C4 grasses. *Ecol. Lett.* **11**, 266–276 (2008).

276. Osborne, C. P. *et al.* A global database of C4 photosynthesis in grasses. *New Phytol.* **204**, 441–446 (2014).

277. Zhang, A., Yang, Z., Zuo, Y., Ma, L. & Zhang, H. Geographic distribution of C4 species and its phylogenetic structure across China. *Front. Plant Sci.* **14**, 1214980 (2023).

278. Berasategui, J. A., Žerdoner Čalasan, A., Zizka, A. & Kadereit, G. Global distribution, climatic preferences and photosynthesis-related traits of C4 eudicots and how they differ from those of C4 grasses. *Ecol. Evol.* **13**, e10720 (2023).

279. Collatz, G. J., Berry, J. A. & Clark, J. S. Effects of climate and atmospheric CO2 partial pressure on the global distribution of C4 grasses: present, past, and future. *Oecologia* **114**, 441–454 (1998).

280. Prentice, I. C., Dong, N., Gleason, S. M., Maire, V. & Wright, I. J. Balancing the costs of carbon gain and water transport: testing a new theoretical framework for plant functional ecology. *Ecol. Lett.* **17**, 82–91 (2014).

281. Wang, H. *et al.* Towards a universal model for carbon dioxide uptake by plants. *Nat. Plants* **3**, 734–741 (2017).

282. Stocker, B. D. *et al.* P-model v1. 0: An optimality-based light use efficiency model for simulating ecosystem gross primary production. *Geosci. Model Dev.* **13**, 1545–1581 (2020).

283. Cai, W. *et al.* A unifying principle for global greenness patterns and trends. *Commun. Earth Environ.* **6**, 19 (2025).

284. Zhou, B. *et al.* A general model for the seasonal to decadal dynamics of leaf area. *Glob. Chang. Biol.* **31**, e70125 (2025).

285. Harrison, S. P. *et al.* Eco-evolutionary optimality as a means to improve vegetation and land-surface models. *New Phytol.* **231**, 2125–2141 (2021).

286. Zhao, J., Zhou, B., Harrison, S. P. & Prentice, I. C. Eco-evolutionary modelling of global vegetation dynamics and the impact of CO2 during the late Quaternary: insights from contrasting periods. *Earth Syst. Dyn.* **16**, 1655–1669 (2025).

287. Utescher, T. & Mosbrugger, V. Eocene vegetation patterns reconstructed from plant diversity—a global perspective. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **247**, 243–271 (2007).

288. Li, J. & Prentice, I. C. Deep-time geographic dynamics of climate shape global vascular plant diversity. *Prep.* (2025).

289. Stein, U. & Alpert, P. Factor separation in numerical simulations. *J. Atmos. Sci.* **50**, 2107–2115 (1993).

290. Christin, P.-A. *et al.* Oligocene CO2 decline promoted C4 photosynthesis in grasses. *Curr. Biol.* **18**, 37–43 (2008).

291. Hannold, C., Wang, Y., Wang, X. & Carranza-Castaneda, O. Isotopic evidence for mammalian diets and environment in Early Pliocene Yepómera, Mexico. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **578**, 110569 (2021).

292. Latorre, C., Quade, J. & McIntosh, W. C. The expansion of C4 grasses and global change in the late Miocene: stable isotope evidence from the Americas. *Earth Planet. Sci. Lett.* **146**, 83–96 (1997).

293. MacFadden, B. J., Cerling, T. E. & Prado, J. Cenozoic terrestrial ecosystem evolution in Argentina: evidence from carbon isotopes of fossil mammal teeth. *Palaios* 319–327 (1996).

294. Uno, K. T. *et al.* Late Miocene to Pliocene carbon isotope record of differential diet change among East African herbivores. *Proc. Natl. Acad. Sci.* **108**, 6509–6514 (2011).

295. Quade, J., Cerling, T. E. & Bowman, J. R. Development of Asian monsoon revealed by marked ecological shift during the latest Miocene in northern Pakistan. *Nature* **342**, 163–166 (1989).

296. An, Z., Kutzbach, J. E., Prell, W. L. & Porter, S. C. Evolution of Asian monsoons and phased uplift of the Himalaya–Tibetan plateau since Late Miocene times. *Nature* **411**, 62–66 (2001).

297. Andrae, J. W. *et al.* Initial expansion of C4 vegetation in Australia during the late Pliocene. *Geophys. Res. Lett.* **45**, 4831–4840 (2018).

298. Lee-Thorp, J. *et al.* Isotopic evidence for an early shift to C4 resources by Pliocene hominins in Chad. *Proc. Natl. Acad. Sci.* **109**, 20369–20372 (2012).

299. Hooghiemstra, H., Lézine, A.-M., Leroy, S. A. G., Dupont, L. & Marret, F. Late Quaternary palynology in marine sediments: a synthesis of the understanding of pollen distribution patterns in the NW African setting. *Quat. Int.* **148**, 29–44 (2006).

300. Schefuß, E., Schouten, S., Jansen, J. H. F. & Sinninghe Damsté, J. S. African vegetation controlled by tropical sea surface temperatures in the mid-Pleistocene period. *Nature* **422**, 418–421 (2003).

301. Lee-Thorp, J. A., Sponheimer, M., Passey, B. H., de Ruiter, D. J. & Cerling, T. E. Stable isotopes in fossil hominin tooth enamel suggest a fundamental dietary shift in the Pliocene. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 3389–3396 (2010).

302. Rommerskirchen, F., Eglinton, G., Dupont, L. & Rullkötter, J. Glacial/interglacial changes in southern Africa: Compound-specific δ13C land plant biomarker and pollen records from southeast Atlantic continental margin sediments. *Geochemistry, Geophys. Geosystems* **7**, (2006).

303. Lézine, A.-M., Ivory, S. J., Braconnot, P. & Marti, O. Timing of the southward retreat of the ITCZ at the end of the Holocene Humid Period in Southern Arabia: Data-model comparison. *Quat. Sci. Rev.* **164**, 68–76 (2017).

304. Gallagher, S. J. *et al.* Cenozoic history of the Australian Monsoon. *Prog. Earth Planet. Sci.* **11**, 1–32 (2024).

305. Prentice, I. C. & Harrison, S. P. Ecosystem effects of CO 2 concentration: evidence from past climates. *Clim. Past* **5**, 297–307 (2009).

306. Edwards, E. J. & Smith, S. A. Phylogenetic analyses reveal the shady history of C4 grasses. *Proc. Natl. Acad. Sci.* **107**, 2532–2537 (2010).

307. Pau, S., Edwards, E. J. & Still, C. J. Improving our understanding of environmental controls on the distribution of C3 and C4 grasses. *Glob. Chang. Biol.* **19**, 184–196 (2013).

308. Sage, R. F., de Melo Peixoto, M., Friesen, P. & Deen, B. C4 bioenergy crops for cool climates, with special emphasis on perennial C4 grasses. *J. Exp. Bot.* **66**, 4195–4212 (2015).

309. Swinehart, D. F. The beer-lambert law. *J. Chem. Educ.* **39**, 333 (1962).

310. Zhu, Z. *et al.* Optimality principles explaining divergent responses of alpine vegetation to environmental change. *Glob. Chang. Biol.* **29**, 126–142 (2023).

311. Ding, R., Nóbrega, R. L. B. & Prentice, I. C. A minimalist model for forest biomass dynamics based on net carbon profit. *Prep.*

312. Li, B. *et al.* BESSv2. 0: A satellite-based and coupled-process model for quantifying long-term global land–atmosphere fluxes. *Remote Sens. Environ.* **295**, 113696 (2023).

313. Cucchi, M. *et al.* Near surface meteorological variables from 1979 to 2019 derived from bias-corrected reanalysis, Version 2.0, Copernicus Climate Change Service (C3S) Climate Data Store (CDS)[data set]. (2021).

314. Roebroek, C. T. J. *et al.* Potential tree cover under current and future climate scenarios. *Sci. Data* **12**, 564 (2025).

315. Maechler, M. *et al.* Package 'robustbase'. *Basic Robust Statistics* vol. 2021 (2021).

316. Cavender-Bares, J. & Holbrook, N. M. Hydraulic properties and freezing-induced cavitation in sympatric evergreen and deciduous oaks with contrasting habitats. *Plant, Cell Environ.* **24**, 1243–1256 (2001).

317. Hurrell, J. W. *et al.* The community earth system model: a framework for collaborative research. *Bull. Am. Meteorol. Soc.* **94**, 1339–1360 (2013).

318. Delworth, T. L. *et al.* GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Clim.* **19**, 643–674 (2006).

319. Bentsen, M. *et al.* The Norwegian Earth System Model, NorESM1-M–Part 1: description and basic evaluation of the physical climate. *Geosci. Model Dev.* **6**, 687–720 (2013).

320. Danabasoglu, G. *et al.* The community earth system model version 2 (CESM2). *J. Adv. Model. Earth Syst.* **12**, e2019MS001916 (2020).

321. Kelley, M. *et al.* GISS-E2. 1: Configurations and climatology. *J. Adv. Model. Earth Syst.* **12**, e2019MS002025 (2020).

322. Miller, R. L. *et al.* CMIP6 historical simulations (1850–2014) with GISS-E2. 1. *J. Adv. Model. Earth Syst.* **13**, e2019MS002034 (2021).

323. Andrews, M. B. *et al.* Historical simulations with HadGEM3-GC3. 1 for CMIP6. *J. Adv. Model. Earth Syst.* **12**, e2019MS001995 (2020).

324. Semmler, T. *et al.* Simulations for CMIP6 with the AWI climate model AWI-CM-1-1. *J. Adv. Model. Earth Syst.* **12**, e2019MS002009 (2020).

325. Mauritsen, T. *et al.* Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1. 2) and its response to increasing CO2. *J. Adv. Model. Earth Syst.* **11**, 998–1038 (2019).

326. Yukimoto, S. *et al.* The Meteorological Research Institute Earth System Model version 2.0, MRI-ESM2. 0: Description and basic evaluation of the physical component. *J. Meteorol. Soc. Japan. Ser. II* **97**, 931–965 (2019).

327. Cao, J. *et al.* The NUIST Earth System Model (NESM) version 3: description and preliminary evaluation. *Geosci. Model Dev.* **11**, 2975–2993 (2018).

328. Seland, Ø. *et al.* Overview of the Norwegian Earth System Model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations. *Geosci. Model Dev.* **13**, 6165–6200 (2020).

329. Bi, D. *et al.* Configuration and spin-up of ACCESS-CM2, the new generation Australian community climate and earth system simulator coupled model. *J. South. Hemisph. Earth Syst. Sci.* **70**, 225–251 (2020).

330. Tatebe, H. *et al.* Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geosci. Model Dev.* **12**, 2727–2765 (2019).

Appendices

A1 Supplementary information for Chapter 2

A1.1 Analysis for 16 plant functional traits

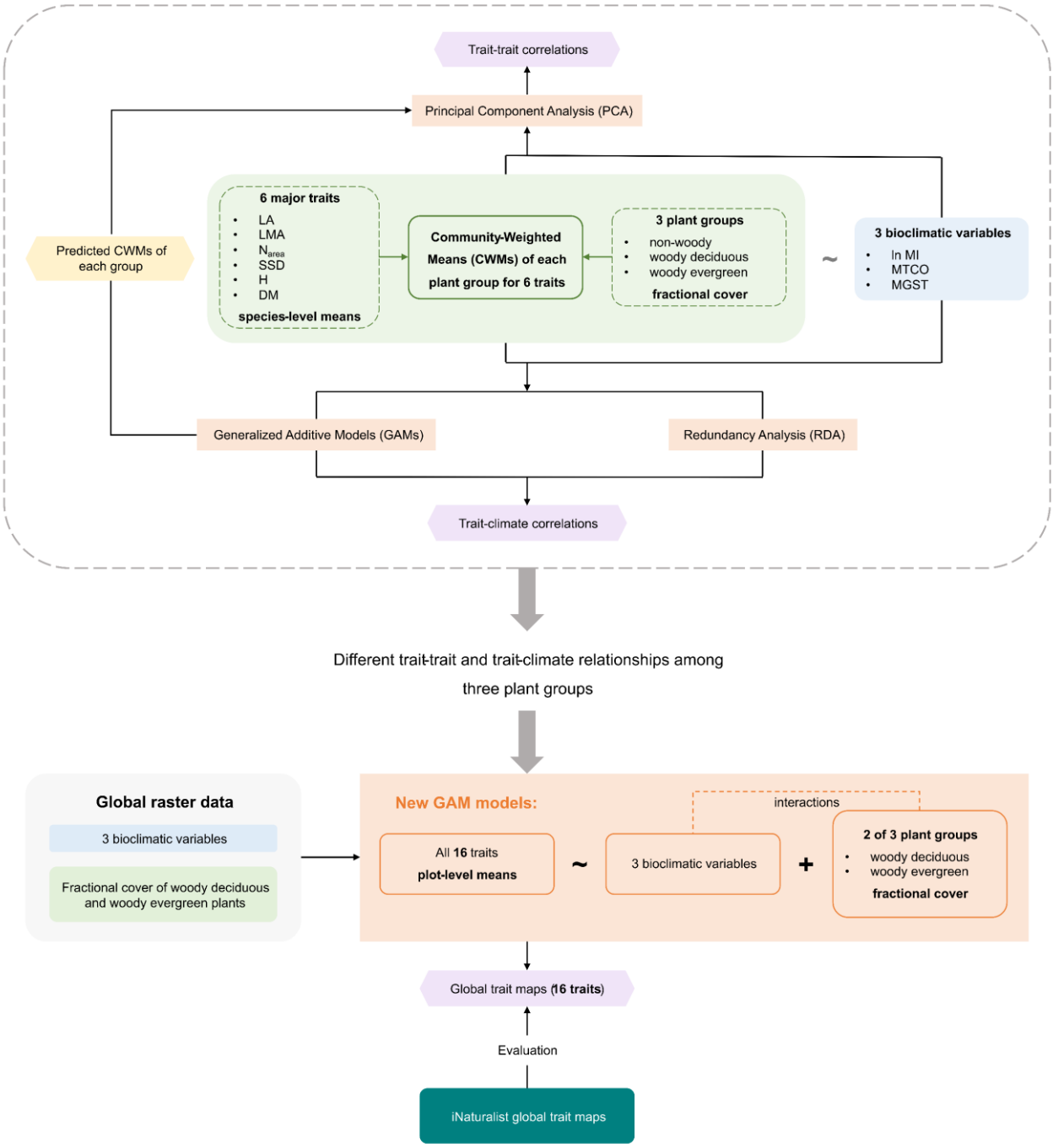


Fig. A1.1 | Overview of the study. All elements of the figure were created by the authors.

Table A1.1 | Sample size of all data used in this study. [a] Plant functional traits used in this study. [b] Abbreviation and [c] unit of each of the 16 traits. [d] Number of plots having plot-level mean trait values. [e] Number of species having species-level mean trait values. [f] Number of pixels of iNaturalist trait maps at 2° spatial resolution. [g–h] Number of pixels of global trait maps upscaled from Generalised Additive Models (GAMs) at [g] 0.1° spatial resolution and [h] 2° spatial resolution. [i] Number of pixels having both iNaturalist trait values and GAM trait values when making comparisons at 2° spatial resolution. The first six traits are major traits having both species-level[1] and plot-level trait[2] means. The remaining traits only have plot-level trait means (community-weighted means, CWMs)[2].

Trait[a]	Abbreviation[b]	Unit[c]	N_{plot}[d]	$N_{species}$[e]	$N_{iNaturalist}$[f]	N_{GAM01}[g]	N_{GAM2}[h]	$N_{Comparison}$[i]
Leaf area	LA	mm^2			3,381			1,851
Leaf mass per unit area	LMA	kg/m^2			3,573			1,916
Leaf nitrogen content per unit area	N_{area}	g/m^2		35,773	3,433			1,871
Stem specific density	SSD	g/cm^3			3,198			1,802
Plant height	H	m			3,732			1,980
Diaspore mass	DM	mg			3,749			1,987
Leaf fresh mass	LFM	g			2,476			1,417
Leaf phosphorus content per unit area	P_{area}	g/m^2			3,421			1,856
Leaf carbon content	C_{mass}	mg/g	77,074		3,326	830,086	2,397	1,818
Leaf dry matter content	LDMC	g/g			3,286			1,800
Stem conduit (vessel and tracheid) element length	WVL	µm			2,314			1,320
Stem conduit density	SCD	mm^{-2}		0	2,869			1,622
Seed number per reproductive unit	SN	NA			2,707			1,502
Seed length	SL	mm			2,745			1,528
Dispersal unit length	DUL	mm			2,818			1,534
Leaf nitrogen isotope ratio	$\delta^{15}N$	per million			2,774			1,602

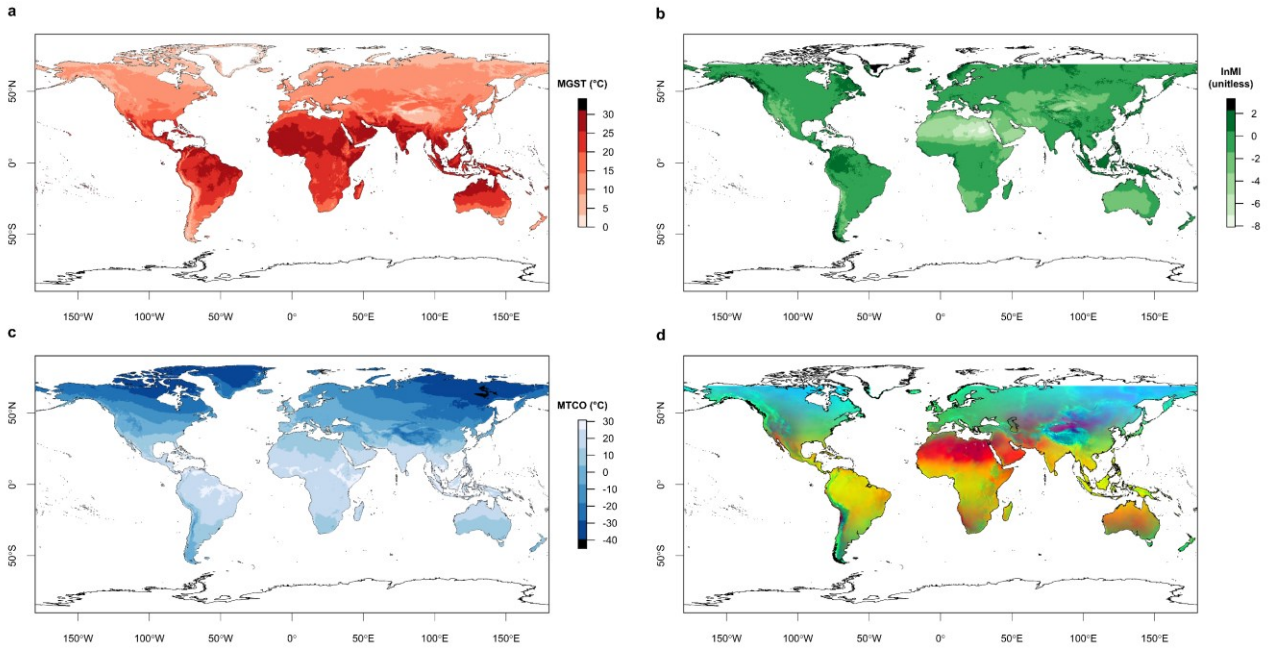


Fig. A1.2 | Global climates and global climate classes. a – c, Gridded global distribution of three bioclimatic variables at 0.1° spatial resolutions: **a**, mean growing-season temperature (MGST, ℃); **b**, natural-log transformed moisture index (ln MI, unitless); **c**, mean temperature of the coldest month (MTCO, ℃). **d**, map of global climate classes. The global climate classes were defined by three bioclimatic variables as follows: the grid values of each bioclimatic raster were rescaled into 0 – 255 and were then combined as RGB colour codes with red representing MGST, green representing ln MI and blue representing MTCO by the 'plotRGB()' function in R *raster* package[3].

Table A1.2 | Recorded plants and their relative proportion in the plot-level trait dataset. [a] Categories of recorded plants from the sPlotOpen database. [b] Number of plots recorded in each category in the plot-level trait dataset . [c] Proportion of each category to all records in the plot-level trait dataset.

Recorded plants[a]	N_{plot}[b]	Proportion[c]
All trees & dominant understory	110	0.14%
All vascular plants	12,743	16.53%
All woody plants	3,037	3.94%
Dominant trees	1	0.00%
Dominant woody plants >= 2.5 cm dbh	293	0.38%
Not specified	55,587	72.12%
Only dominant species	1,208	1.57%
Woody plants >= 1 cm dbh	14	0.02%
Woody plants >= 1 m height	24	0.03%
Woody plants >= 10 cm dbh	678	0.88%
Woody plants >= 2.5 cm dbh	2,837	3.68%
Woody plants >= 20 cm dbh	38	0.05%
Woody plants >= 5 cm dbh	504	0.65%

Table A1.3 | Summary of plant growth form dataset. [a] 16 plant growth forms defined based on the woodiness, plant growth form and leaf phenology of plants. [b] Brief description of each growth form. [c] Number of plant taxa within each growth form. [d] Proportion of plant taxa within each growth form to all plant taxa in the plant growth form dataset.

Plant growth form[a]	Description[b]	N_{taxa}[c]	Proportion[d]
non-woody	All non-woody taxa, e.g., herb, herbaceous vine…	21,483	50.34%
woody deciduous	All woody taxa with deciduous leaves; also including semi-deciduous or rarely/briefly evergreen taxa	3,659	8.57%
woody evergreen	All woody species with evergreen leaves; also including semi-evergreen or rarely/briefly deciduous taxa	9,385	21.99%
woody mix	All woody taxa with both deciduous and evergreen members	512	1.20%
woody NA	Woody plants with unavailable leaf phenology	5,083	11.91%
non-woody/woody deciduous	Taxa having both non-woody members and woody members with deciduous leaves	121	0.28%
non-woody/woody evergreen	Taxa having both non-woody members and woody members with evergreen leaves	151	0.35%
non-woody/woody mix	Taxa having both non-woody members and woody members with deciduous or evergreen leaves	58	0.14%
non-woody/woody NA	Taxa having both non-woody members and woody members with unavailable leaf phenology	198	0.46%
fern	All ferns	1,222	2.86%
bamboo	All bamboos	42	0.10%
palm	All palms	270	0.63%
cycad	All cycads	18	0.04%
succulent	All succulents	374	0.88%
cactus	All cacti	89	0.21%
NA	Lichen, liverworts, moss, clubmoss, fungi, algae…	11	0.03%

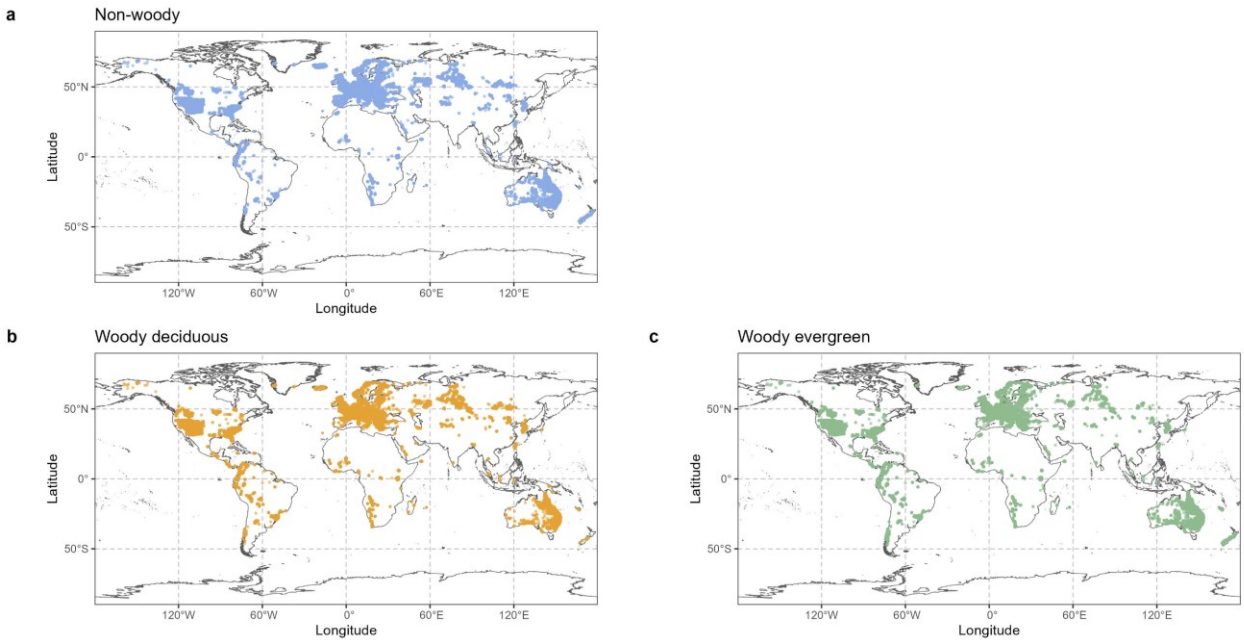


Fig. A1.3 | Global distributions of vegetation plots having community-weighted means (CWMs) of three plant groups for six major plant functional traits. a, global distribution of plots with trait CWMs of non-woody plants (number of plots: 66,066); **b**, global distribution of plots with trait CWMs of woody deciduous plants (number of plots: 46,659); **c**, global distribution of plots with trait CWMs of woody evergreen plants (number of plots: 52,802). All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

Table A1.4 | Probability of each plant growth form being three plant groups. [a] 16 plant growth forms used in this study, see Table A1.3 for more information. [b] The probability that a growth form is assumed to be non-woody. [c] The probability that a growth form is assumed to be woody deciduous. [d] The probability that a growth form is assumed to be woody evergreen.

Plant growth form[a]	Probability of being non-woody[b]	Probability of being woody deciduous[c]	Probability of being woody evergreen[d]
non-woody	1	0	0
woody deciduous	0	1	0
woody evergreen	0	0	1
woody mix	0	0.5	0.5
woody NA	0	0.5	0.5
non-woody/woody deciduous	0.5	0.5	0
non-woody/woody evergreen	0.5	0	0.5
non-woody/woody mix	0.33	0.33	0.33
non-woody/woody NA	0.33	0.33	0.33
fern	0	0	0
bamboo	0	0	0
palm	0	0	0
cycad	0	0	0
succulent	0	0	0
cactus	0	0	0
NA	0	0	0

Table A1.5 | Fractional covers of three plant groups for all land cover classes and sub-classes. [a] 23 classes of global land cover in the ESA CCI-LC global land cover dataset. [b] 38 sub-classes of global land cover presented as grid values in the ESA CCI-LC maps. [c] Brief description of each land cover sub-class. [d] Fractional cover of non-woody plants in each land cover sub-class. [e] Fractional cover of woody deciduous plants in each land cover sub-class. [f] Fractional cover of woody evergreen plants in each land cover sub-class. The detailed calculation of global fractional covers of three plant groups are shown in Methods.

Class[a]	Sub-class[b]	Description[c]	Non-woody[d]	Woody deciduous[e]	Woody evergreen[f]
No data	0	No data	NA	NA	NA
Rainfed cropland	10	Cropland, rainfed	NA	NA	NA
	11	10 - Herbaceous cover	NA	NA	NA
	12	10 - Tree or shrub cover	NA	NA	NA
Irrigated or post-flooding cropland	20	Cropland, irrigated or post-flooding	NA	NA	NA
Mosaic cropland	30	Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%)	0.7133	0.0163	0.2056
Mosaic natural vegetation	40	Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%)	0.6092	0.0166	0.2521
Broadleaved evergreen tree cover	50	Tree cover, broadleaved, evergreen, closed to open (>15%)	0.3608	0.0215	0.5658

Broadleaved deciduous tree cover	60	Tree cover, broadleaved, deciduous, closed to open (>15%)	0.0655	0.7072	0.0983
	61	60 - Tree cover, broadleaved, deciduous, closed (>40%)	0.9730	0.0051	0.0180
	62	60 - Tree cover, broadleaved, deciduous, open (15-40%)	0.3535	0.0212	0.5315
Needleleaved evergreen tree cover	70	Tree cover, needleleaved, evergreen, closed to open (>15%)	0.3182	0.2431	0.2528
	71	70 - Tree cover, needleleaved, evergreen, closed (>40%)	NA	NA	NA
	72	70 - Tree cover, needleleaved, evergreen, open (15-40%)	NA	NA	NA
Needleleaved deciduous tree cover	80	Tree cover, needleleaved, deciduous, closed to open (>15%)	0.3091	0.2690	0.0783
	81	80 - Tree cover, needleleaved, deciduous, closed (>40%)	NA	NA	NA
	82	80 - Tree cover, needleleaved, deciduous, open (15-40%)	NA	NA	NA

Mixed-leaf-type tree cover	90	Tree cover, mixed leaf type (broadleaved and needleleaved)	0.0197	0.6726	0.2346
Mosaic woody cover	100	Mosaic tree and shrub (>50%) / herbaceous cover (<50%)	0.6212	0.0081	0.3014
Mosaic herbaceous cover	110	Mosaic herbaceous cover (>50%) / tree and shrub (<50%)	0.6519	0.0254	0.2462
Shrubland	120	Shrubland	0.5631	0.0241	0.2857
	121	120 - Evergreen shrubland	NA	NA	NA
	122	120 - Deciduous shrubland	0.5667	0.0119	0.3867
Grassland	130	Grassland	0.9356	0.0056	0.0242
Lichens and mosse	140	Lichens and mosses	NA	NA	NA
Sparse vegetation	150	Sparse vegetation (tree, shrub, herbaceous cover) (<15%)	0.6986	0.0085	0.2281
	151	150 - Sparse tree (<15%)	NA	NA	NA
	152	150 - Sparse shrub (<15%)	NA	NA	NA
	153	150 - Sparse herbaceous cover (<15%)	NA	NA	NA
Fresh/brackish-water-flooded tree cover	160	Tree cover, flooded, fresh or brackish water	0.1282	0.6121	0.1372

Saline-water-flooded tree cover	170	Tree cover, flooded, saline water	0.4492	0.0004	0.4974
Flooded shrub or herbaceous cover	180	Shrub or herbaceous cover, flooded, fresh/saline/brackish water	0.5843	0.0149	0.2801
Urban areas	190	Urban areas	NA	NA	NA
Bare areas	200	Bare areas	NA	NA	NA
	201	200 - Consolidated bare areas	NA	NA	NA
	202	200 - Unconsolidated bare areas	NA	NA	NA
Water bodies	210	Water bodies	NA	NA	NA
Permanent snow and ice	220	Permanent snow and ice	NA	NA	NA

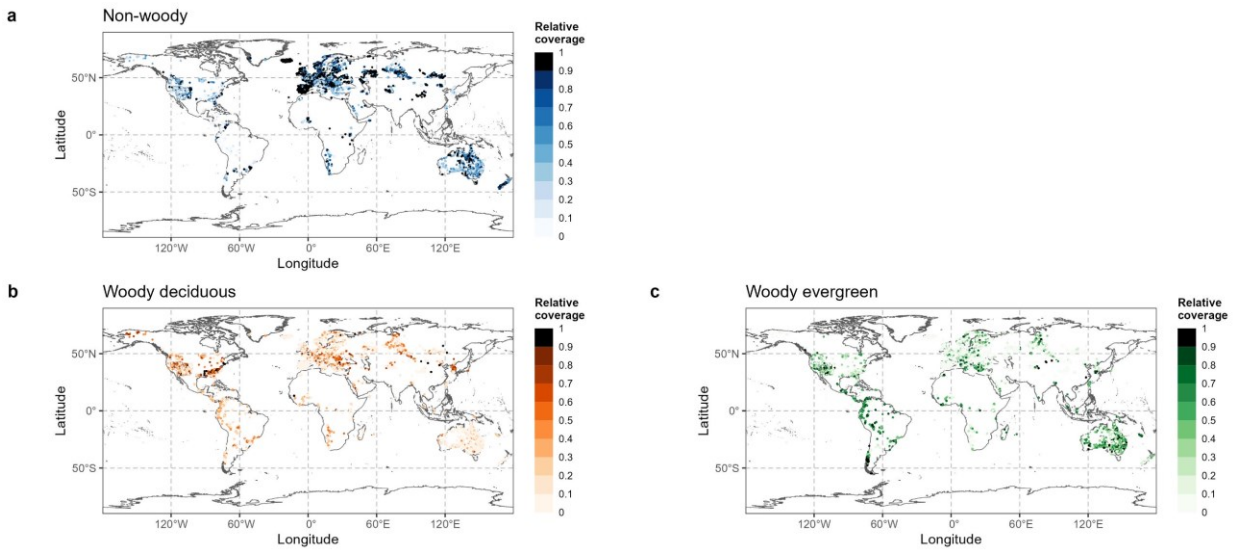


Fig. A1.4 | Relative coverages of three plant groups in all vegetation plots. Global distribution of all 77,074 natural vegetation plots used in this study and the relative coverage of non-woody plants (**a**), woody deciduous plants (**b**) and woody evergreen plants (**c**) within each plot. All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

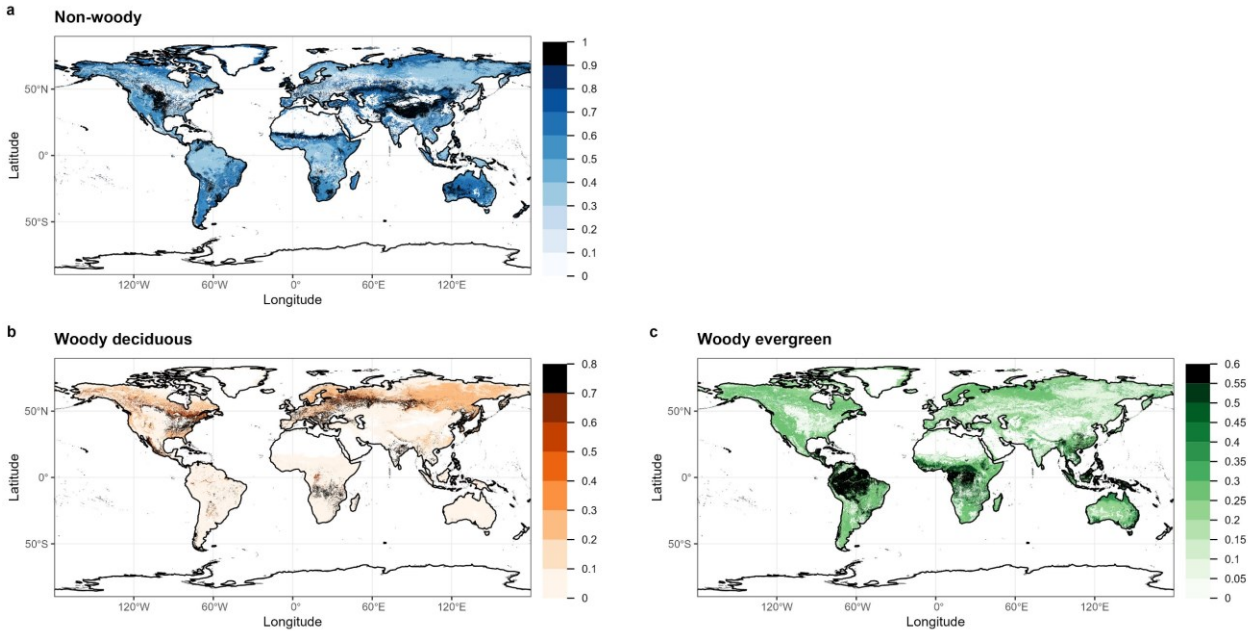


Fig. A1.5 | Global fractional covers of three plant groups. Global maps of estimated coverage for non-woody plants (**a**), woody deciduous plants (**b**) and woody evergreen plants (**c**). All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

Table A1.6 | Summary of principal component analysis (PCA) for six major plant functional traits in three plant groups. Including trait loadings, eigenvalues, and the proportion of trait variation explained by the first three successive principal components (PC1 – 3). All six traits were natural-log transformed before the analysis. Abbreviations and units of traits are shown in Table 2.1.

Variables	Non-woody			Woody deciduous			Woody evergreen		
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
LA	0.4698	-0.2594	0.3953	-0.4208	0.4291	-0.2152	0.3358	-0.5588	0.0321
LMA	-0.5366	-0.3003	0.2688	-0.2061	-0.6237	-0.1612	0.3440	0.5765	0.0350
N_{area}	-0.4923	-0.2349	0.4974	-0.2510	-0.5759	-0.3588	0.4110	0.5029	0.0511
SSD	-0.4145	-0.1838	-0.6624	-0.3689	-0.2112	0.8494	0.3151	-0.1420	0.8831
H	0.1551	-0.6698	-0.0265	-0.5352	0.1751	-0.2668	0.5046	-0.0684	-0.3877
DM	0.2304	-0.5522	-0.2910	-0.5430	0.1413	0.0796	0.4961	-0.2785	-0.2548
Eigenvalue	1.5292	1.1704	0.9324	1.5803	1.4284	0.8655	1.5436	1.3350	0.9421
Proportion Explained	0.3898	0.2283	0.1449	0.4162	0.3400	0.1249	0.3971	0.2970	0.1479
Cumulative Proportion	0.3898	0.6181	0.7630	0.4162	0.7563	0.8812	0.3971	0.6942	0.8421

Table A1.7 | Summary of redundancy analysis (RDA) for six major plant functional traits in three plant groups. Including trait loadings, eigenvalues, and the proportion of trait variation explained by three successive redundancy analysis axes (RDA1 – 3) (constrained by climate) and the first three residual principal components (PC1 – 3) (unrelated to climate). All six traits were natural-log transformed before the analysis. Abbreviations and units of traits are shown in Table 2.1.

Variables	Non-woody						Woody deciduous						Woody evergreen					
	RDA1	RDA2	RDA3	PC1	PC2	PC3	RDA1	RDA2	RDA3	PC1	PC2	PC3	RDA1	RDA2	RDA3	PC1	PC2	PC3
LA	-1.7527	-1.9862	0.6077	6.4243	-4.2646	2.1138	0.6330	3.1101	0.2625	-6.4770	1.8414	1.5128	6.0419	-1.6884	0.3859	-2.7686	4.0152	-0.0701
LMA	1.7818	0.3995	-0.2258	-7.7911	-3.7350	1.2750	4.3021	-1.2954	0.1559	1.5417	-6.1753	1.3505	-1.3114	4.8577	0.3665	-4.4959	-4.8433	0.5587
N_{area}	0.9525	1.4543	0.2692	-7.0363	-3.9918	3.3756	4.5567	-1.2580	0.0525	0.7945	-5.6931	2.7729	-0.0050	5.0672	0.0288	-4.8712	-4.1793	1.6609
SSD	1.9513	1.8881	0.4025	-5.4205	-0.8533	-6.6063	2.7458	-0.1354	-0.1255	-2.9954	-4.0098	-5.8023	5.9326	2.2868	-1.0091	-0.3092	-1.7593	-5.5515
H	3.9885	-1.9976	-0.1889	2.5715	-5.5155	-0.6782	1.9768	1.5890	0.9888	-6.7423	-1.1033	2.0997	1.9235	0.5278	1.6352	-7.2074	1.0865	-1.6330
DM	2.5686	-0.5040	0.4590	3.9751	-5.0898	-3.5118	2.8736	2.2854	-0.9349	-6.1808	-1.0550	-0.3705	3.4255	0.5883	0.2891	-6.1413	3.8639	0.5012
Eigenvalue	0.3631	0.1522	0.0098	2.2054	1.1405	0.8020	0.8746	0.3049	0.0289	2.0267	1.3602	0.7367	1.1243	0.7336	0.0514	1.7887	0.9648	0.4658
Proportion Explained	0.0605	0.0254	0.0016	0.3676	0.1901	0.1337	0.1458	0.0508	0.0048	0.3378	0.2267	0.1228	0.1874	0.1223	0.0086	0.2981	0.1608	0.0776
Cumulative Proportion	0.0605	0.0859	0.0875	0.4551	0.6452	0.7788	0.1458	0.1966	0.2014	0.5392	0.7659	0.8887	0.1874	0.3096	0.3182	0.6163	0.7771	0.8548

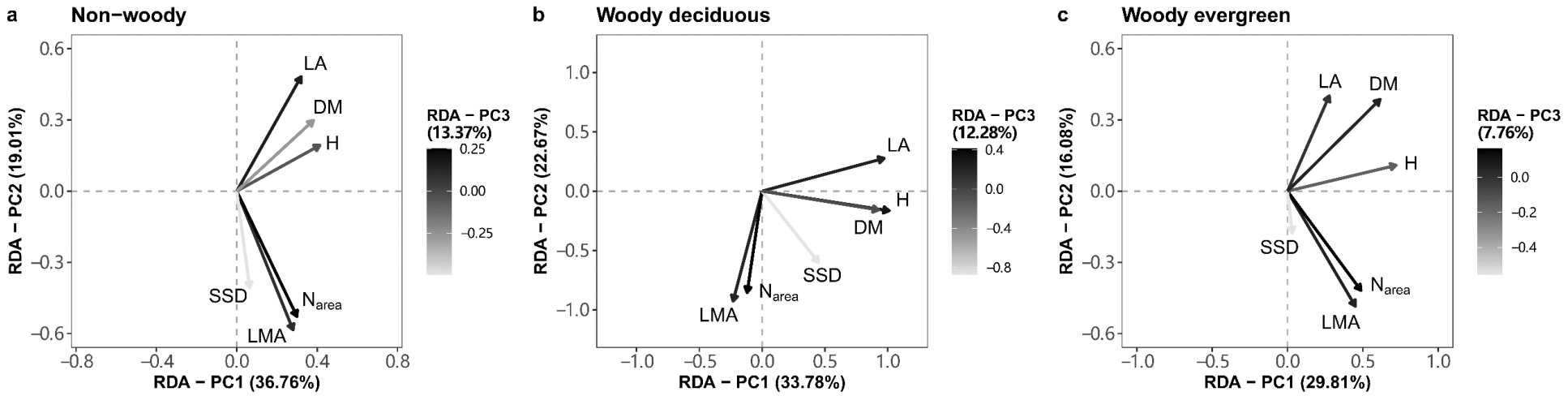


Fig. A1.6 | Residual (climate-independent) dimensions of trait variation from redundancy analysis for six major plant functional traits of non-woody (a), woody deciduous (b) and woody evergreen (c) plants. The orientation of axes has been rotated according to the Figure 2 in ref. [5]. Colour bars indicate the value and direction of the axis 3 for RDA–PC; gray scales indicate the loadings on the third axis. Arrows are upscaled or downscaled to fit the size of the graph. All six traits are natural-log transformed. Natural-log transformed traits were rescaled to a mean of 0 and a standard deviation of 1 before analyses. Abbreviations and units of traits are shown in Table 2.1. ln MI, log-transformed moisture index; MTCO, mean temperature of the coldest month; MGST, mean growing-season temperature (see Methods for definition). All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

Table A1.8 | Summary of Generalised Additive Models (GAMs) based on three bioclimatic variables for six major plant functional traits in three plant groups. Including number of vegetation plots containing community-weighted means (CWMs) of three plant groups for each trait (Sample size), and explained deviance, adjusted R^2 and P values of each GAM model. All six traits were natural-log transformed before the analysis. Abbreviations and units of traits are shown in Table 2.1.

Trait	Non-woody				Woody deciduous				Woody evergreen			
	Sample size	Deviance explained	Adjusted R^2	*P* value	Sample size	Deviance explained	Adjusted R^2	*P* value	Sample size	Deviance explained	Adjusted R^2	*P* value
LA	60,730	24.10%	0.2406	$< 2 \times 10^{-16}$ ***	39,541	53.12%	0.5309	$< 2 \times 10^{-16}$ ***	46,509	60.64%	0.6062	$< 2 \times 10^{-16}$ ***
LMA	60,872	22.60%	0.2257	$< 2 \times 10^{-16}$ ***	39,197	28.26%	0.2821	$< 2 \times 10^{-16}$ ***	46,432	31.51%	0.3147	$< 2 \times 10^{-16}$ ***
N_{area}	57,520	14.22%	0.1418	$< 2 \times 10^{-16}$ ***	36,509	30.63%	0.3058	$< 2 \times 10^{-16}$ ***	43,537	41.17%	0.4114	$< 2 \times 10^{-16}$ ***
SSD	52,256	23.10%	0.2307	$< 2 \times 10^{-16}$ ***	33,384	17.57%	0.1751	$< 2 \times 10^{-16}$ ***	44,035	57.54%	0.5752	$< 2 \times 10^{-16}$ ***
H	64,114	28.83%	0.2880	$< 2 \times 10^{-16}$ ***	43,528	46.26%	0.4622	$< 2 \times 10^{-16}$ ***	50,953	34.50%	0.3447	$< 2 \times 10^{-16}$ ***
DM	64,211	8.12%	0.0808	$< 2 \times 10^{-16}$ ***	43,146	45.04%	0.4501	$< 2 \times 10^{-16}$ ***	50,182	31.40%	0.3136	$< 2 \times 10^{-16}$ ***

Table A1.9 | Relative importance of each bioclimatic variable in predicting six major plant functional traits of three plant groups. All six traits were natural-log transformed before the analysis. Abbreviations and units of traits are shown in Table 2.1. ln MI, log-transformed moisture index; MTCO, mean temperature of the coldest month; MGST, mean growing-season temperature (see Methods for definition).

Trait	Non-woody			Woody deciduous			Woody evergreen		
	ln MI	MTCO	MGST	ln MI	MTCO	MGST	ln MI	MTCO	MGST

LA	64.84%	8.97%	26.19%	62.28%	14.81%	22.91%	3.89%	44.65%	51.46%
LMA	34.65%	19.43%	45.92%	72.60%	3.36%	24.04%	64.13%	24.45%	11.42%
N_{area}	57.92%	13.19%	28.89%	76.19%	3.91%	19.90%	64.10%	15.92%	19.98%
SSD	54.39%	4.82%	40.79%	42.94%	21.70%	35.36%	16.87%	40.33%	42.80%
H	11.79%	25.25%	62.96%	15.34%	21.41%	63.25%	19.17%	19.26%	61.56%
DM	49.47%	15.41%	35.12%	14.05%	44.61%	41.33%	7.68%	29.58%	62.74%

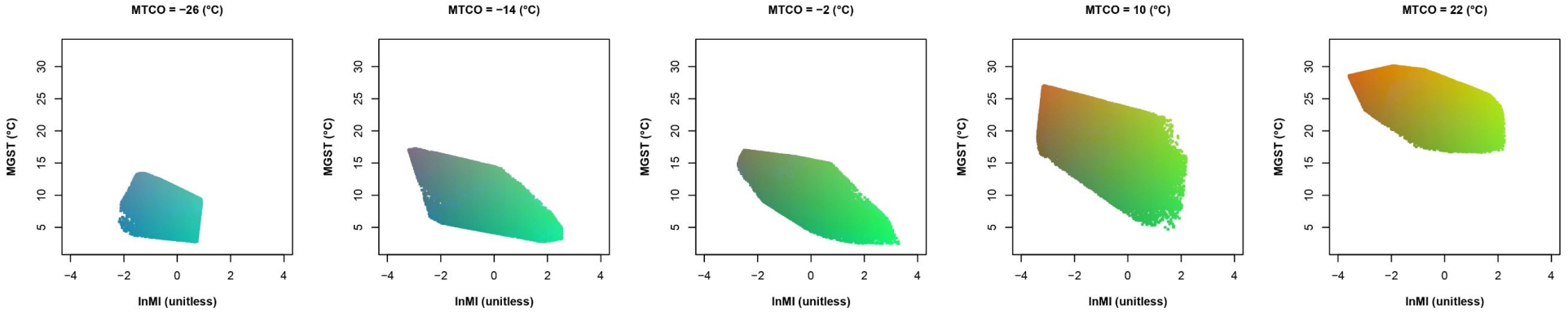


Fig. A1.7 | Global climate classes in the global climate space. The values of three bioclimatic variables over the globe were rescaled into 0 – 255, respectively, and were then combined as RGB colour codes with red representing mean growing-season temperature (MGST, ℃), green representing natural-log transformed moisture index (ln MI, unitless) and blue representing mean temperature of the coldest month (MTCO, ℃). All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

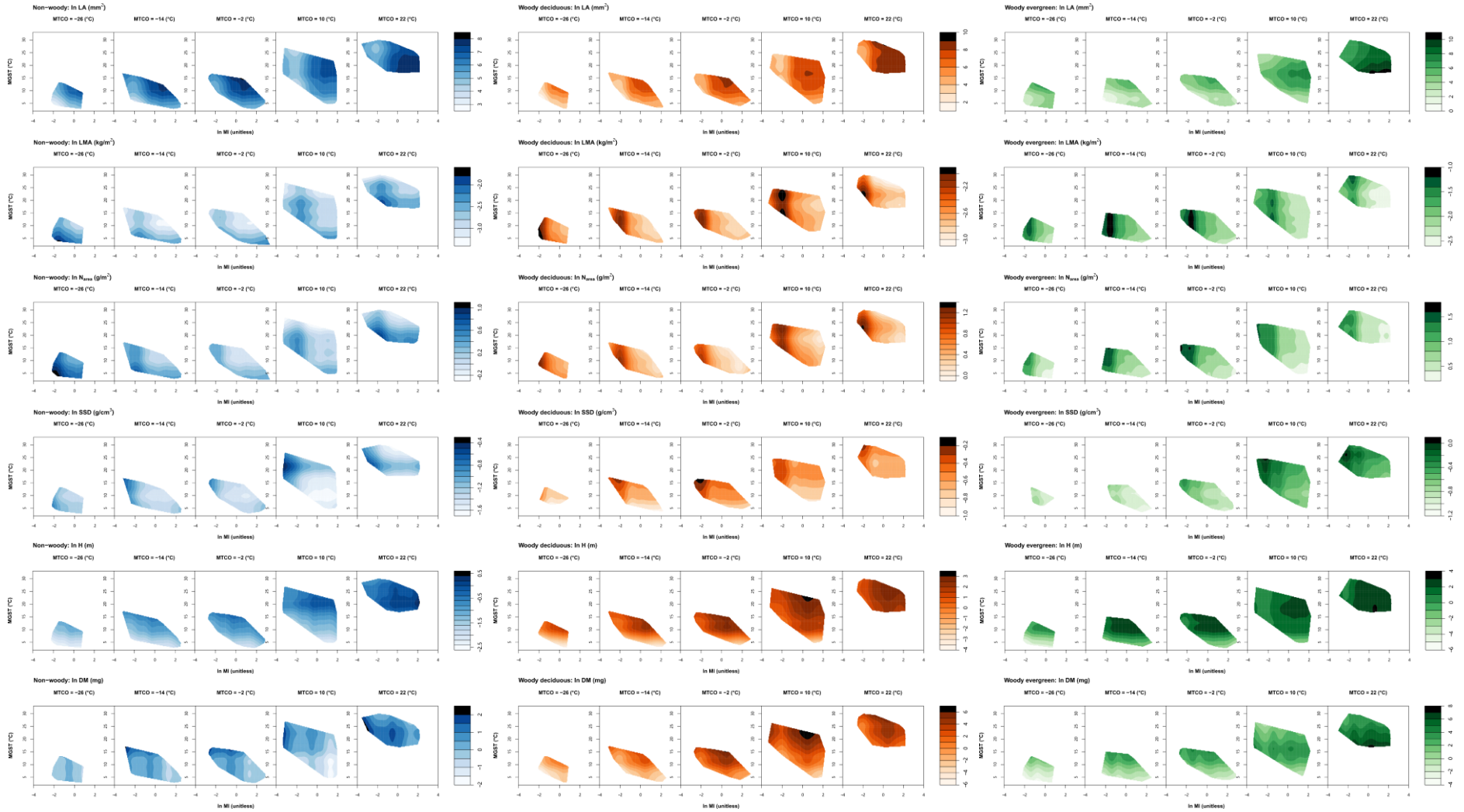


Fig. A1.8 | Climate space diagrams for six major plant functional traits of non-woody, woody deciduous and woody evergreen plants. Showing distributions of natural-log transformed trait values of three plant groups in the global climate space defined by three bioclimatic variables. Fitted trait values are presented as contours, with darker colours in the right colour bar representing higher trait values. Abbreviations and units of

traits are shown in Table 2.1. ln MI, log-transformed moisture index; MTCO, mean temperature of the coldest month; MGST, mean growing-season temperature (see Methods for definition). All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

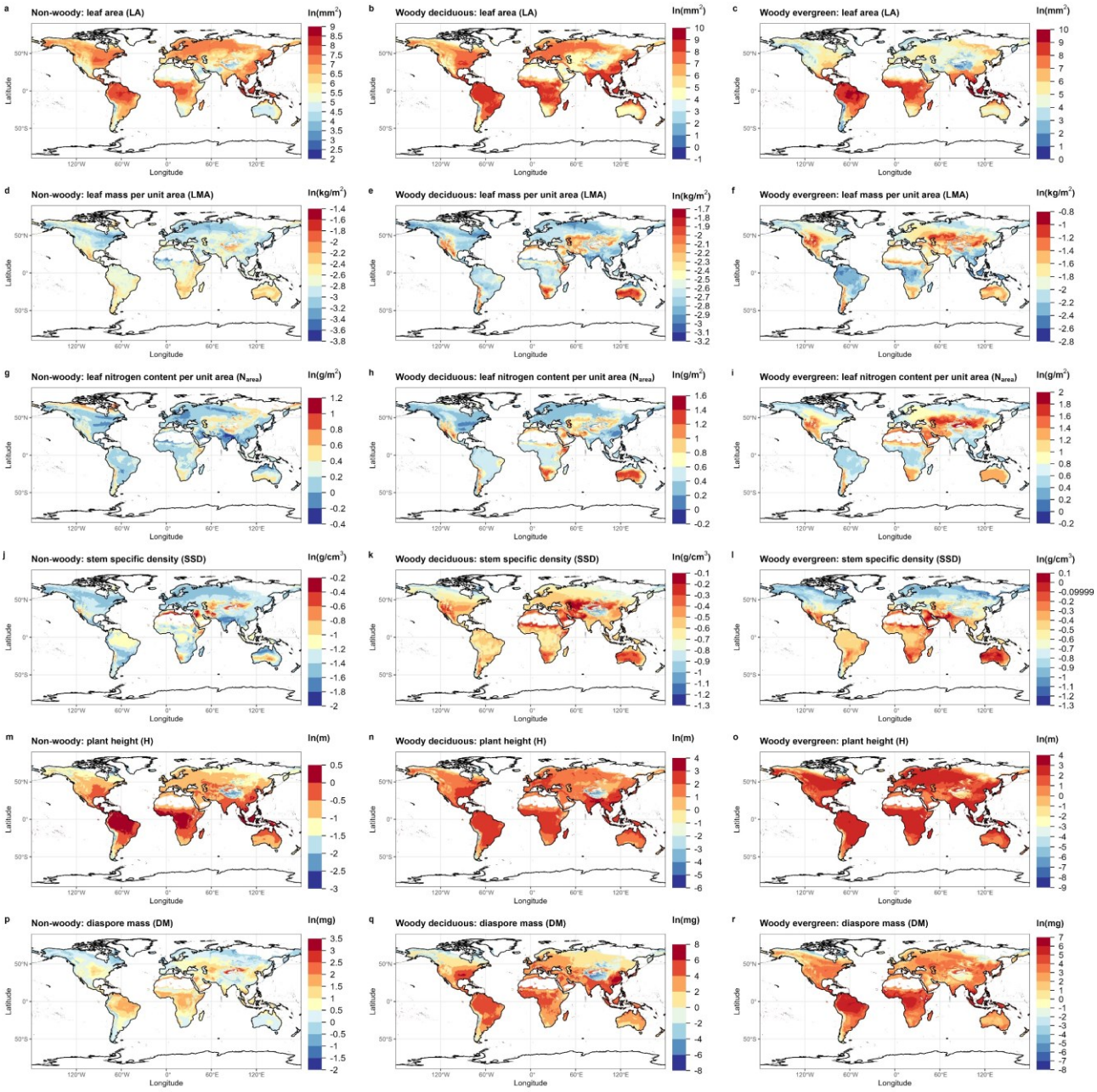


Fig. A1.9 | Separate global trait maps for six major plant functional traits of three plant groups. Global patterns of natural-log transformed trait values predicted by Generalised Additive Models (GAMs) based on three bioclimatic variables. The maps in the left column are global patterns of the six traits of non-woody plants. The middle-column maps are global patterns of the six traits of woody deciduous plants. The maps in the right column are global patterns of the six traits of woody evergreen plants. All maps are at a 0.1° spatial resolution. For maps in GeoTiff format, refer to the Data availability statement. Abbreviations of traits are shown in Table 2.1. All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

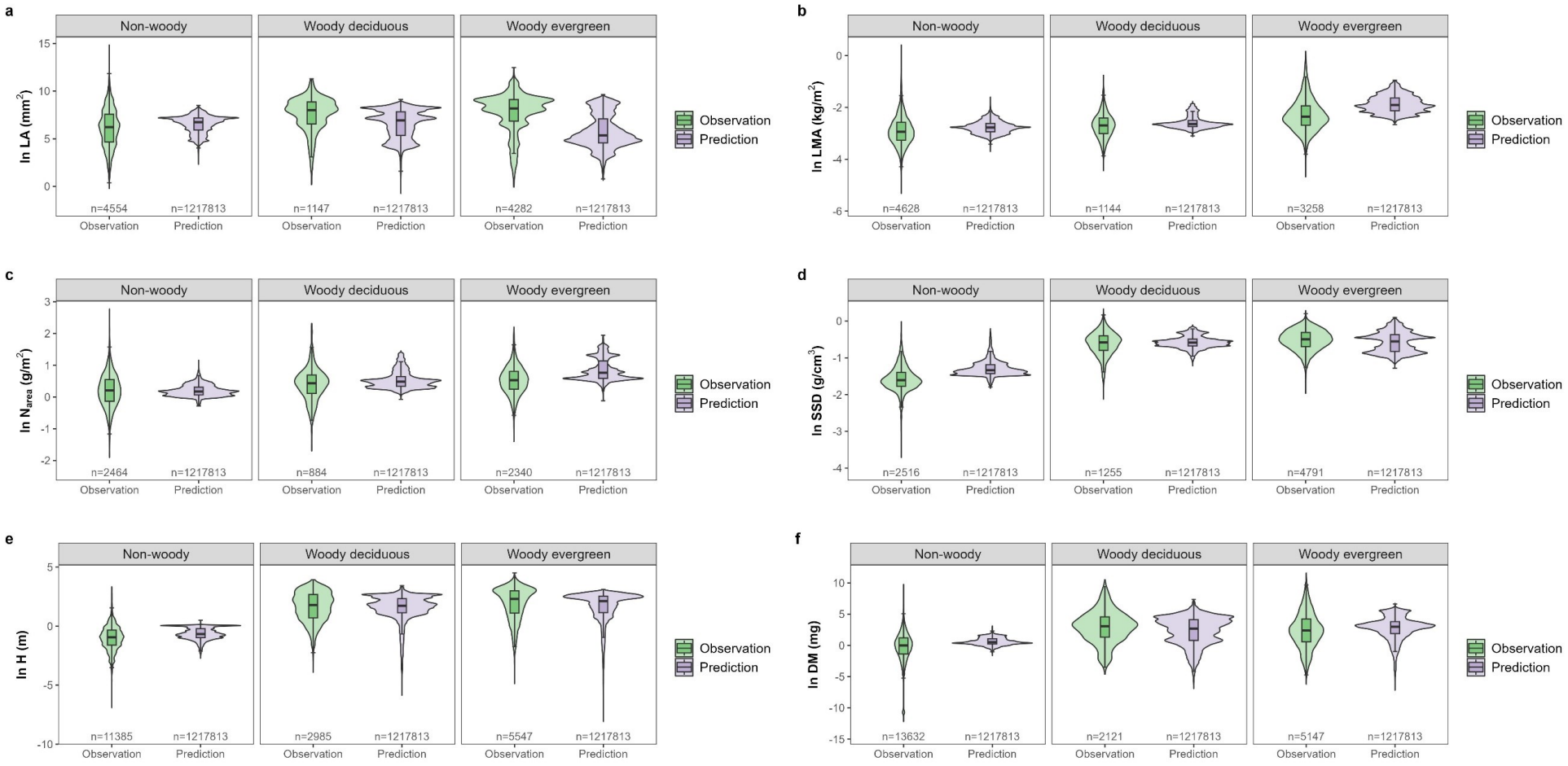


Fig. A1.10 | Violin plots showing comparisons between observed and predicted values of six major plant functional traits in three plant groups. All six traits are natural-log transformed. Green violin plots show community-weighted means (CWMs) of three plant groups for six traits used to fit Generalised Additive Models (GAMs). Purple violin plots show pixel values of separate global trait maps for each of the three plant groups.

Abbreviations of traits are shown in Table 2.1. All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

Table A1.10 | Summary of redundancy analysis (RDA) for all 16 plant functional traits. Including trait loadings, eigenvalues, and the proportion of trait variation explained by six successive redundancy analysis axes (RDA1 – 5) (constrained by climate and fractional covers of two of three plant groups). All traits were natural-log transformed before the analysis. Abbreviations and units of traits are shown in Table 2.1.

Trait	RDA1	RDA2	RDA3	RDA4	RDA5
LA	2.2875	4.1351	0.3359	0.5152	0.3765
LMA	3.7922	-3.2668	-0.5596	-1.6822	0.2021
N_{area}	3.6466	-3.3197	-0.6980	-0.2319	0.2207
SSD	6.3465	-0.5341	-1.0836	-0.1120	-0.5942
H	6.6650	0.3411	0.8414	0.2652	-0.1898
DM	5.7106	0.6421	0.2900	0.5213	-0.2975
LFM	3.3591	3.4485	-0.0480	0.0290	0.2440
P_{area}	0.9985	-4.4006	0.0584	0.9611	0.0768
C_{mass}	3.3403	-0.4903	3.5505	-1.1895	-0.2349
LDMC	3.7144	-1.3566	-0.4824	-0.4554	1.3840
WVL	0.4457	2.5943	1.2475	-1.4273	-0.3608
SCD	-2.1652	-3.6858	3.1341	0.6424	-0.1564
SN	-1.1684	-2.4867	0.2448	0.2550	-0.1338
SL	4.5236	0.4195	0.4516	0.8147	0.0015
DUL	4.2096	0.1319	0.7707	1.3317	0.0602
$\delta^{15}N$	1.9252	-1.3106	-2.7136	-0.1797	-0.8025
Eigenvalue	3.3945	1.4603	0.5090	0.1590	0.0514
Proportion Explained	0.2122	0.0913	0.0318	0.0099	0.0032
Cumulative Proportion	0.2122	0.3034	0.3352	0.3452	0.3484

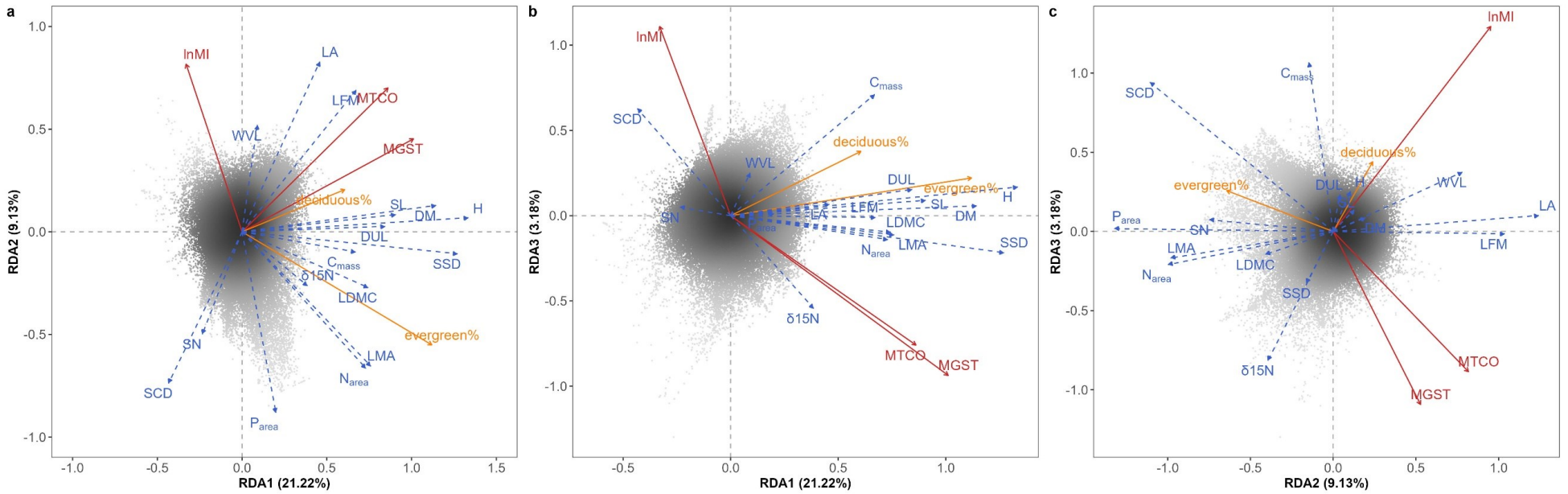


Fig. A1.11 | Redundancy analysis (RDA) for all 16 plant functional traits. a, axis 1 vs axis 2. **b**, axis 1 vs axis 3. **c**, axis 2 vs axis 3. Each point represents a vegetation plot, and the darker colour indicates plot clusters. Correlations among traits and between traits and climates and fractional covers of plant groups are represented by the direction of vectors, where trait vectors are shown in blue arrows, and vectors of bioclimatic variables and fractional covers are shown in red and orange arrows, respectively. Arrows are upscaled or downscaled to fit the size of the graph. As the sum of the global fractional cover for the three plant groups is 1, here only considered the coverage of two of three groups (woody deciduous and woody evergreen). All 16 traits are natural-log transformed. Both bioclimatic variables, fractional covers of plant groups and natural-log transformed traits were rescaled to a mean of 0 and a standard deviation of 1 before analyses. Abbreviations and units of traits are shown in Table 2.1. ln MI, log-transformed moisture index; MTCO, mean temperature of the coldest month; MGST, mean growing-season temperature (see Methods for definition). deciduous% and evergreen% are fractional covers of woody deciduous and evergreen plants within a vegetation plot, respectively. All

elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

Table A1.11 | Summary of the Generalised Additive Models (GAMs) based on three bioclimatic variables and global fractional covers of two of three plant groups for all 16 plant functional traits in 77,074 natural vegetation plots. Including explained deviance, adjusted R^2 and P values of each GAM model. All traits were natural-log transformed before the analysis. Abbreviations and units of traits are shown in Table 2.1.

Trait	Deviance explained	Adjusted R^2	*P* value
LA	56.50%	0.5640	$< 2 \times 10^{-16}$ ***
LMA	57.53%	0.5743	$< 2 \times 10^{-16}$ ***
N_{area}	49.16%	0.4905	$< 2 \times 10^{-16}$ ***
SSD	67.67%	0.6760	$< 2 \times 10^{-16}$ ***
H	76.75%	0.7669	$< 2 \times 10^{-16}$ ***
DM	63.72%	0.6364	$< 2 \times 10^{-16}$ ***
LFM	54.68%	0.5457	$< 2 \times 10^{-16}$ ***
P_{area}	47.10%	0.4697	$< 2 \times 10^{-16}$ ***
C_{mass}	54.09%	0.5398	$< 2 \times 10^{-16}$ ***
LDMC	39.44%	0.3930	$< 2 \times 10^{-16}$ ***
WVL	32.16%	0.3202	$< 2 \times 10^{-16}$ ***
SCD	61.01%	0.6092	$< 2 \times 10^{-16}$ ***
SN	26.59%	0.2641	$< 2 \times 10^{-16}$ ***
SL	46.23%	0.4611	$< 2 \times 10^{-16}$ ***
DUL	44.93%	0.4481	$< 2 \times 10^{-16}$ ***
$\delta^{15}N$	30.30%	0.3014	$< 2 \times 10^{-16}$ ***

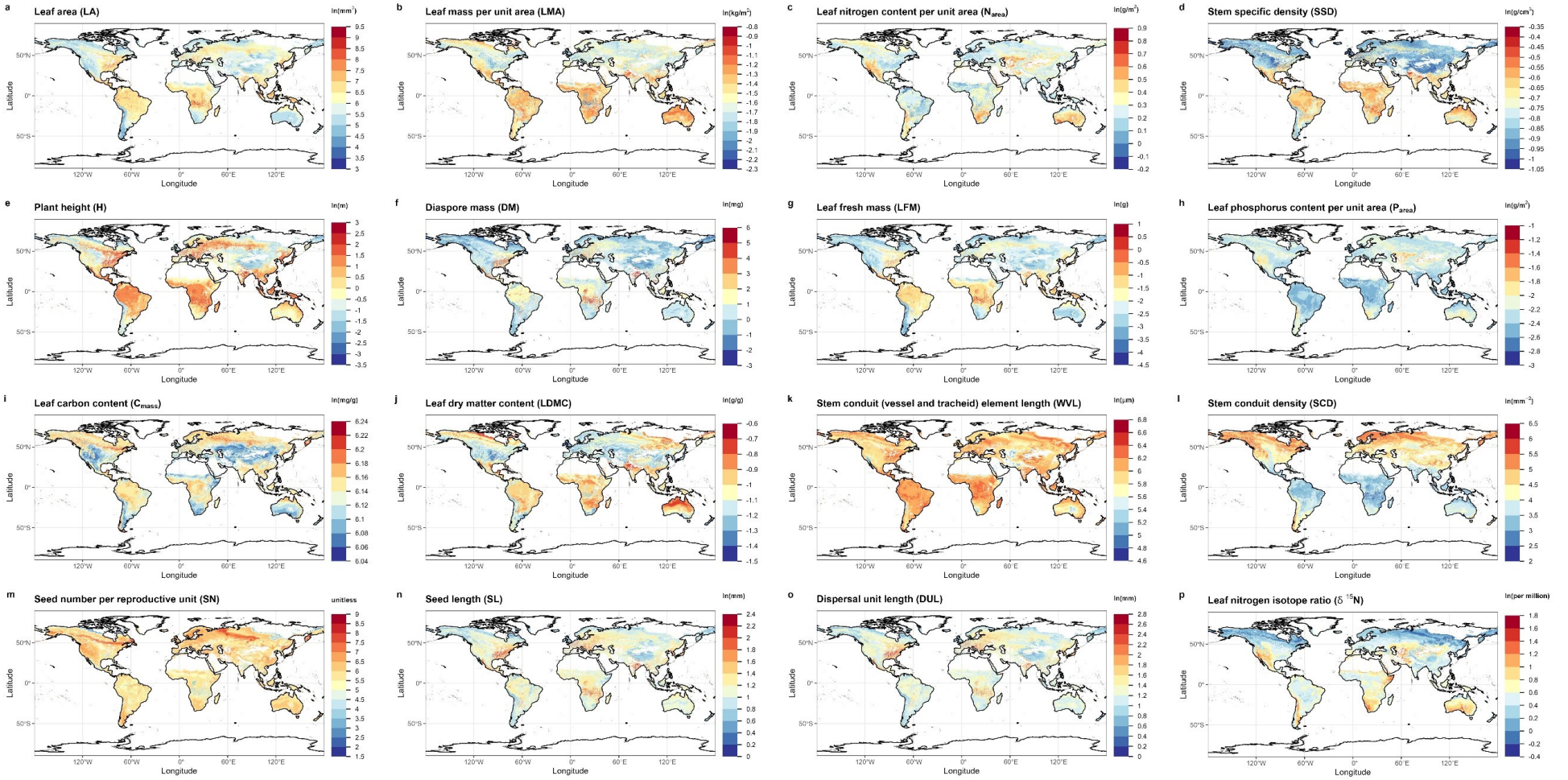


Fig. A1.12 | Global trait maps for all 16 plant functional traits. Global patterns of natural-log transformed trait values predicted by the Generalised Additive Models (GAMs) based on three bioclimatic variables and global fractional covers of two of three plant groups (see Methods). All maps are at a 0.1° spatial resolution. For maps in GeoTiff format, refer to the Data availability statement. Abbreviations of traits are shown in Table 2.1. All

elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

Table A1.12 | Model predictabilities of trait maps in this study and previously published trait maps. R^2_{GAM} represents adjusted R^2 of Generalised Additive Models (GAMs) in this study. R^2_{Pre} or R_{Pre} indicate adjusted R^2 or correlation coefficient R in previously published studies. Abbreviations and units of traits are shown in Table 2.1.

Paper	LA		LMA		N_{area}		SSD		H		DM		LDMC	
	R^2_{GAM}	R^2_{Pre}	R^2_{GAM}	R^2_{Pre}	R^2_{GAM}	R^2_{Pre}	R^2_{GAM}	R^2_{Pre}	R^2_{GAM}	R^2_{Pre}	R^2_{GAM}	R^2_{Pre}	R^2_{GAM}	R_{Pre}
Van Bodegom *et al*. (2014)[6]		--		0.3500		--		0.5200		--		0.4100		--
Madani *et al*. (2018)[7]		--		--		--		--		0.6620		0.4550		--
Moreno-Martínez *et al*. (2018)[8]	0.5640	--	0.5743	--	0.4905	--	0.6760	--	0.7669	--	0.6364	--	0.3930	0.7180
Boonman *et al*. (2020)[9]		--		--		--		--		0.6800		--		--
Schiller *et al*. (2021)[10]		0.4503		--		--		0.2002		0.5777		0.2495		--
Dong *et al*. (2023)[11]		--		--		0.2800		--		--		--		--

Table A1.13 | Summary of comparison between model-predicted values and iNaturalist estimated values for all 16 plant functional traits. Including four parameters for assessing agreements between model-predicted map pixel values and the iNaturalist map pixel values. The agreements were estimated by linear regression at a 2° spatial resolution. R^2 is the coefficient of determination; RRMSE is the root-mean-square error, as a proportion of the observed mean trait value (here observations are represented by iNaturalist trait values); bias is the difference between observed and model-predicted mean values, as a proportion of the observed mean trait value (here observed values are iNaturalist trait values); slope is the slope of the linear regression of iNaturalist estimation against model prediction. All traits are natural-log transformed. Abbreviations and units of traits are shown in Table 2.1.

Trait	R^2	RRMSE	bias	slope
LA	0.36	0.177	-0.0892	1.1
LMA	0.17	-0.249	-0.101	1
N_{area}	0.20	0.81	-0.325	1.45
SSD	0.27	-0.487	-0.245	2.06
H	0.23	6.1	-0.151	0.565
DM	0.13	1.3	-0.622	0.674
LFM	0.36	-0.876	0.419	1.4
P_{area}	0.037	-0.164	0.0171	0.48
C_{mass}	0.045	0.0108	0.00562	0.446
LDMC	0.12	-0.27	-0.209	0.701
WVL	0.062	0.057	0.023	0.464
SCD	0.64	0.234	-0.0566	1.78
SN	0.0091	0.369	-0.25	0.332
SL	0.024	0.729	0.229	0.418
DUL	0.00061	0.77	0.251	0.0656
$\delta^{15}N$	0.24	0.985	-0.268	1.36

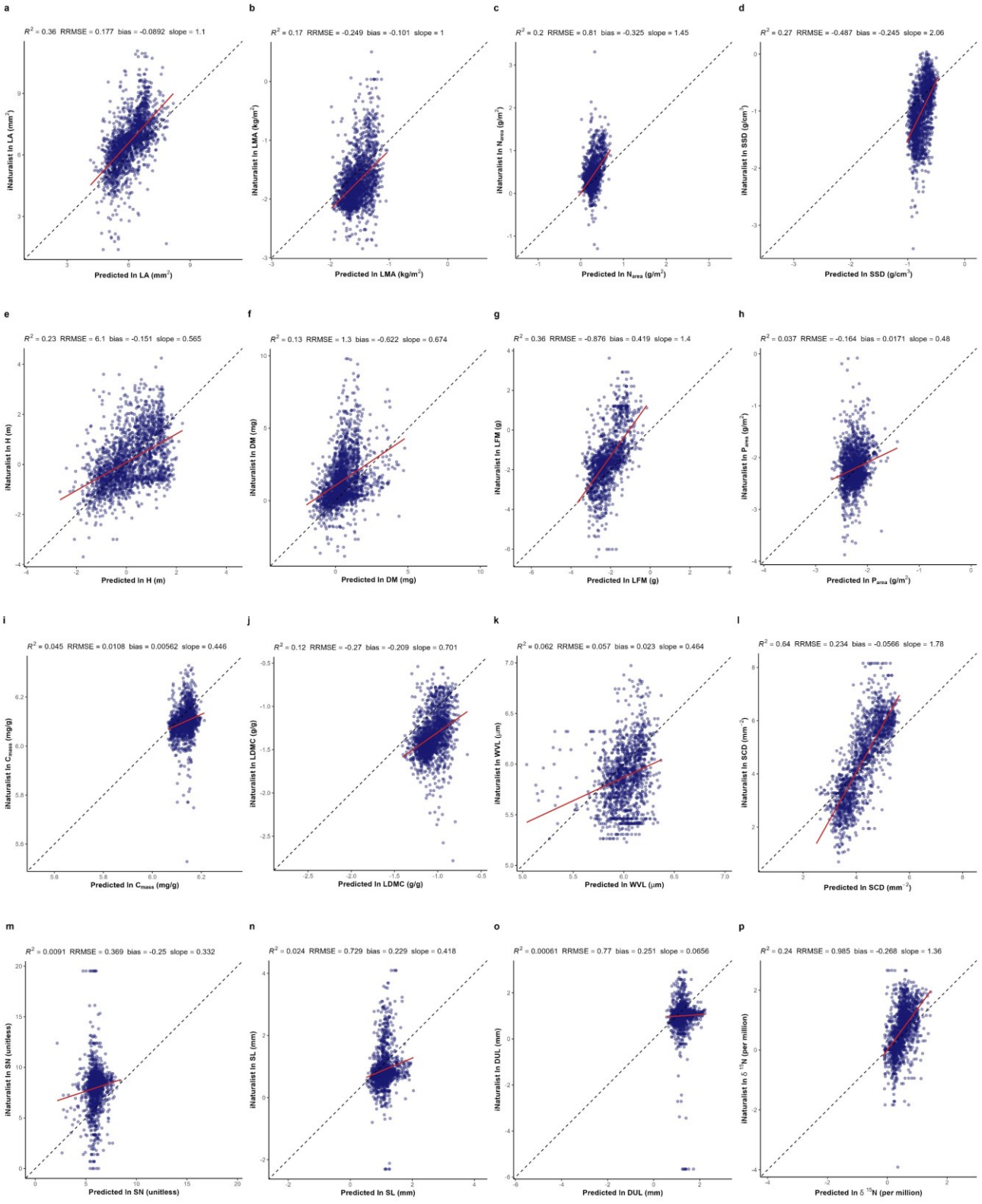


Fig. A1.13 | Scatter plots of comparison between predicted trait values in this study and estimated trait values from the iNaturalist for all 16 plant functional traits. Red lines visualise the linear regression between model-predicted map pixel values and the iNaturalist map pixel values estimated at a 2° spatial resolution. The dotted line is the 1:1 line. Parameters on each of the

above plots for assessing agreements between global trait maps in this study and the iNaturalist trait maps are listed in Table A1.13. All traits are natural-log transformed. Abbreviations and units of traits are shown in Table 2.1. All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

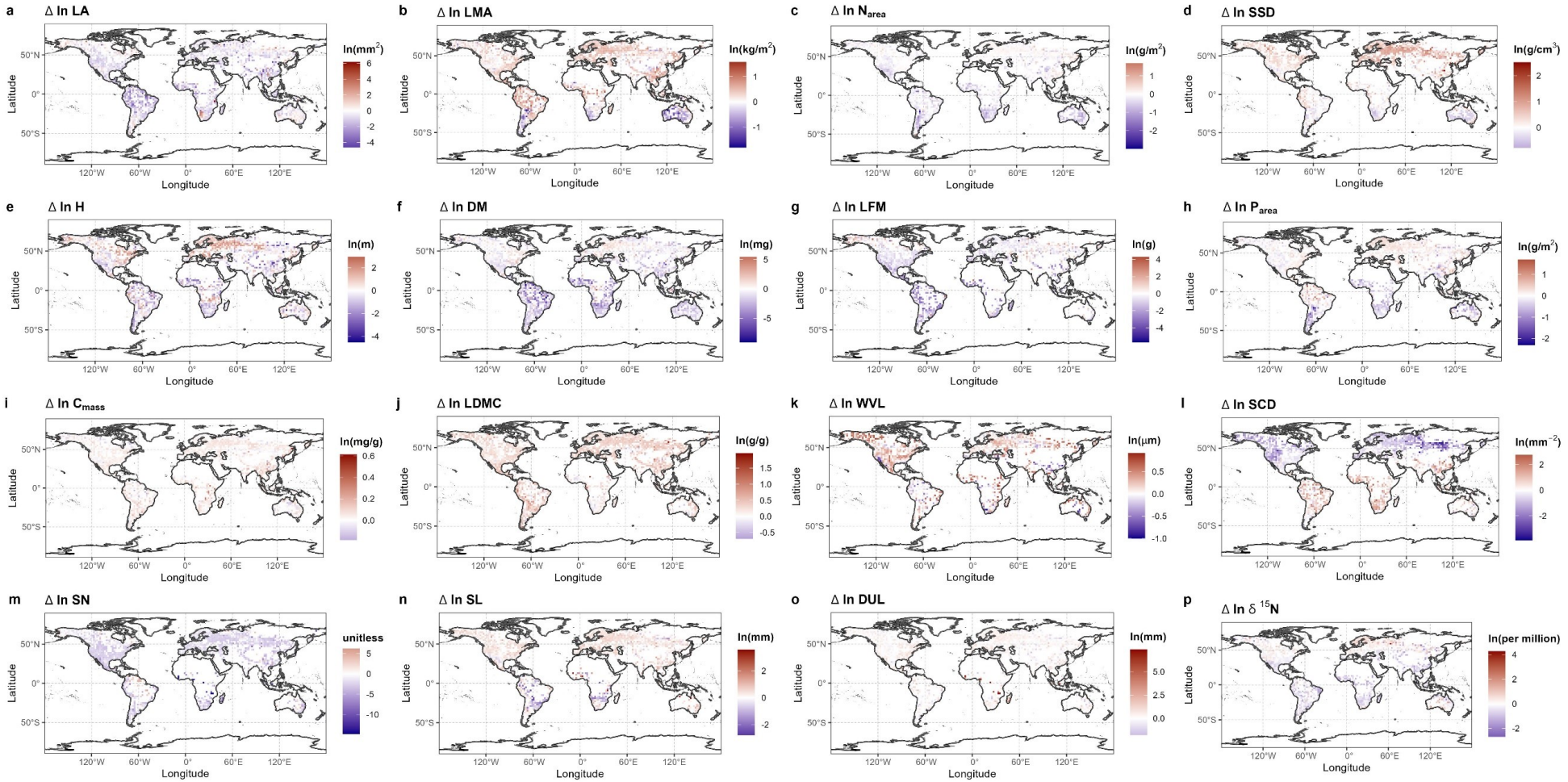


Fig. A1.14 | Global maps of differentiation between predicted trait values in this study and estimated trait values from the iNaturalist for all 16 plant functional traits at a 2° spatial resolution. Darker colour suggests greater difference between model prediction and iNaturalist estimation. All traits are natural-log transformed. Abbreviations and units of traits are shown in Table 2.1. All elements of the figure were created by the authors

using R v4.2.2 (https://www.r-project.org/)[4].

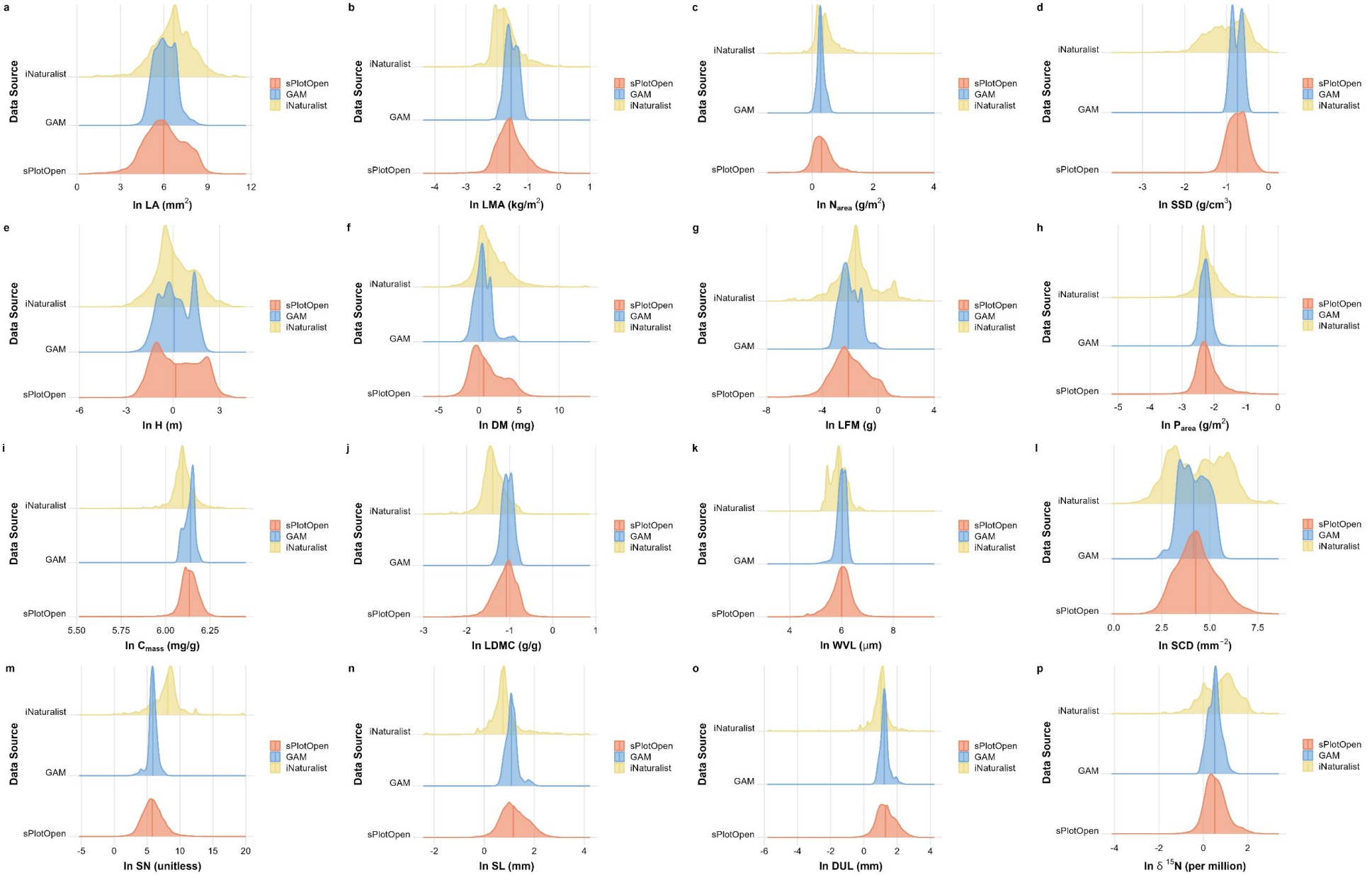

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
Data Source
iNaturalist
GAM
sPlotOpen
ln LA (mm^2)
ln LMA (kg/m^2)
ln N_{area} (g/m^2)
ln SSD (g/cm^3)
ln H (m)
ln DM (mg)
ln LFM (g)
ln P_{area} (g/m^2)
ln C_{mass} (mg/g)
ln LDMC (g/g)
ln WVL (μm)
ln SCD (mm^{-2})
ln SN (unitless)
ln SL (mm)
ln DUL (mm)
ln δ ^{15}N (per million)

Fig. A1.15 | Comparisons of trait value distributions among sample trait data based on the sPlotOpen database, model prediction in this study and iNaturalist estimation for all 16 plant functional traits. All traits are natural-log transformed. Abbreviations and units of traits are shown in Table 2.1. All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

Table A1.14 | Summary of principal component analysis (PCA) for six major plant functional traits in three plant groups based on the global predicted values from the generalised additive modes (GAMs). Including trait loadings, eigenvalues, and the proportion of trait variation explained by the first three successive principal components (PC1 – 3). All six traits are natural-log transformed generated from the GAMs based on bioclimatic variables and global fractional cover of plant groups. Abbreviations and units of traits are shown in Table 2.1.

Variables	Non-woody			Woody deciduous			Woody evergreen		
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
LA	0.5307	0.0385	-0.0086	0.5465	-0.1058	0.3372	0.5496	-0.1457	0.2114
LMA	-0.4379	-0.2658	0.6750	-0.4538	-0.3630	0.2334	-0.2548	0.5839	-0.1700
N_{area}	-0.5260	-0.1662	-0.0762	-0.4386	-0.3486	0.5610	-0.1265	0.6563	0.0151
SSD	-0.2027	-0.5526	-0.5827	-0.1314	-0.5744	-0.6722	0.2975	0.3562	0.7601
H	0.3736	-0.4690	0.4383	0.4116	-0.4124	0.2470	0.4777	0.1890	-0.5461
DM	0.2631	-0.6123	-0.0828	0.3412	-0.4853	-0.0653	0.5480	0.2109	-0.2242
Eigenvalue	1.5791	1.3959	0.8900	1.7461	1.5445	0.5658	1.6475	1.4528	0.8879
Proportion Explained	0.4156	0.3248	0.1320	0.5082	0.3976	0.0534	0.4524	0.3518	0.1314
Cumulative Proportion	0.4156	0.7404	0.8724	0.5082	0.9057	0.9591	0.4524	0.8042	0.9356

Table A1.15 | Summary of redundancy analysis (RDA) for six major plant functional traits in three plant groups based on the global predicted values from the generalised additive modes (GAMs). Including trait loadings, eigenvalues, and the proportion of trait variation explained by three successive redundancy analysis axes (RDA1 – 3) (constrained by climate) and the first three residual principal components (PC1 – 3) (unrelated to climate). All six traits are natural-log transformed generated from the GAMs based on bioclimatic variables and global fractional cover of plant groups. Abbreviations and units of traits are shown in Table 2.1.

Variables	Non-woody						Woody deciduous						Woody evergreen					
	RDA1	RDA2	RDA3	PC1	PC2	PC3	RDA1	RDA2	RDA3	PC1	PC2	PC3	RDA1	RDA2	RDA3	PC1	PC2	PC3
LA	-0.9235	15.9216	1.1417	10.5836	-6.8149	4.3864	8.5197	-14.3082	0.6441	12.1757	-0.1670	3.7394	18.8901	-0.2156	2.7081	-2.6101	5.8584	-5.5943
LMA	6.1636	-4.8034	-4.8625	-17.2671	1.4817	7.1244	-14.9008	2.8655	1.1425	-10.2029	10.0346	1.6198	-11.7650	-11.6984	0.9828	-6.3945	-10.3815	-4.1197
N_{area}	-3.9410	-6.4101	1.2422	-17.4243	-7.7709	1.1072	-15.4736	-0.1830	-0.4310	-12.1676	6.0929	4.4843	-8.5521	-16.3188	1.4709	-6.6564	-6.1646	-0.1144
SSD	7.7137	-9.5300	3.0702	-4.4597	-14.2480	-7.6612	-16.0419	-3.9914	-0.9034	3.5148	10.5275	-6.8297	11.0343	-16.2638	-2.1281	2.2477	-2.0315	3.3990
H	17.1876	6.8216	-0.3177	5.0706	-3.7691	5.5367	-0.8075	-14.0303	-2.4601	12.9179	7.0835	4.7074	10.5037	-1.6239	-0.0084	-17.4171	1.4373	4.7707
DM	14.2727	-1.7295	1.2400	6.7665	-12.1366	4.9724	-5.5766	-15.4951	2.0819	11.3135	4.4236	-0.9941	13.4332	-5.7025	-0.2563	-14.0743	4.4426	-2.3977
Eigenvalue	1.3606	1.0166	0.0834	1.7872	1.0511	0.4113	1.8271	1.4779	0.0291	1.5707	0.7068	0.2364	2.1773	1.5601	0.0334	1.3285	0.4573	0.1961
Proportion Explained	0.2268	0.1694	0.0139	0.2979	0.1752	0.0685	0.3045	0.2463	0.0048	0.2618	0.1178	0.0394	0.3629	0.2600	0.0056	0.2214	0.0762	0.0327
Cumulative Proportion	0.2268	0.3962	0.4101	0.7080	0.8831	0.9517	0.3045	0.5508	0.5557	0.8175	0.9353	0.9747	0.3629	0.6229	0.6285	0.8499	0.9261	0.9588

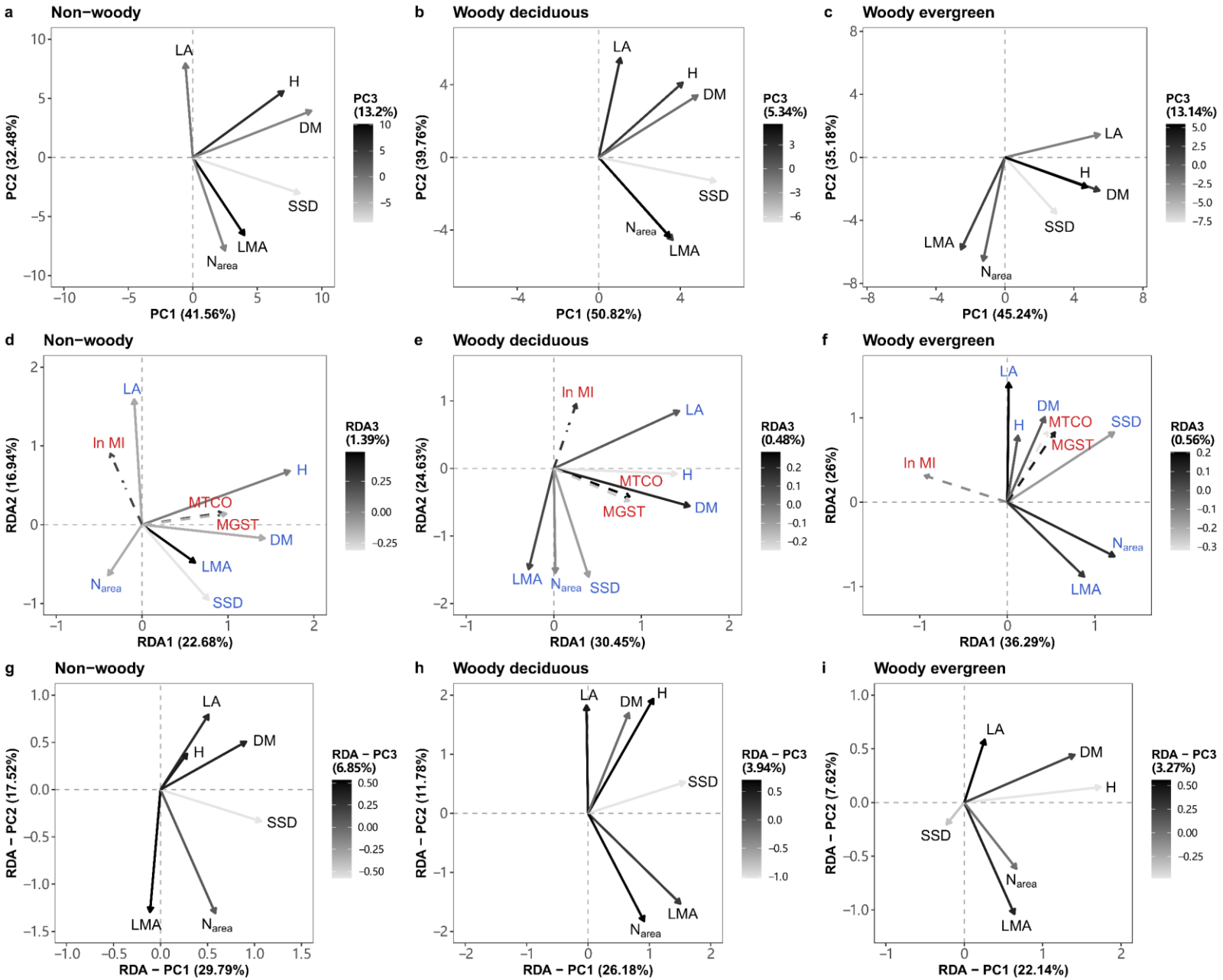

a
Non−woody
PC2 (32.48%)
PC1 (41.56%)
PC3 (13.2%)
LA
H
DM
SSD
LMA
N_area
b
Woody deciduous
PC2 (39.76%)
PC1 (50.82%)
PC3 (5.34%)
c
Woody evergreen
PC2 (35.18%)
PC1 (45.24%)
PC3 (13.14%)
d
Non−woody
RDA2 (16.94%)
RDA1 (22.68%)
RDA3 (1.39%)
In MI
MTCO
MGST
e
Woody deciduous
RDA2 (24.63%)
RDA1 (30.45%)
RDA3 (0.48%)
f
Woody evergreen
RDA2 (26%)
RDA1 (36.29%)
RDA3 (0.56%)
g
Non−woody
RDA − PC2 (17.52%)
RDA − PC1 (29.79%)
RDA − PC3 (6.85%)
h
Woody deciduous
RDA − PC2 (11.78%)
RDA − PC1 (26.18%)
RDA − PC3 (3.94%)
i
Woody evergreen
RDA − PC2 (7.62%)
RDA − PC1 (22.14%)
RDA − PC3 (3.27%)

Fig. A1.16 | Principal component analysis (PCA) (a – c), redundancy analysis (RDA) (d – f) and residual (climate-independent) dimensions of trait variation from redundancy analysis (g – i) for six major plant functional traits of non-woody (a, d and g), woody deciduous (b, e and h) and woody evergreen (c, f and i) plants based on the global predicted values from the generalised additive modes (GAMs). The orientation of axes has been rotated according to the Figure 2 in ref. [5]. Gray scales indicate the loadings on the third axis. Solid arrows (with blue labels in RDA) represent traits and dashed arrows (with red labels in RDA) represent bioclimatic variables. All six traits are natural-log transformed generated from the GAMs based on bioclimatic variables and global fractional cover of plant groups. Both log-transformed traits and bioclimatic variables were rescaled to a mean of 0 and a standard deviation of 1 before analysis. Abbreviations and units of traits are shown in Table 2.1. ln MI, log-transformed moisture index; MTCO, mean temperature of the coldest month; MGST, mean growing-season temperature (see Methods for definition). All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

A1.2 Analysis for mass-based leaf nitrogen content (N_{mass}) and leaf phosphorus content (P_{mass})

Table A1.16 | Sample size of data used in this study for mass-based leaf nitrogen content (N_{mass}) and leaf phosphorus content (P_{mass}). [a] Plant functional traits used in this study. [b] Abbreviation and [c] unit of each of the 16 traits. [d] Number of plots having plot-level mean trait values. [e] Number of species having species-level mean trait values. [f] Number of pixels of iNaturalist trait maps at 2° spatial resolution. [g–h] Number of pixels of global trait maps upscaled from Generalised Additive Models (GAMs) at [g] 0.1° spatial resolution and [h] 2° spatial resolution. [i] Number of pixels having both iNaturalist trait values and GAM trait values when making comparisons at 2° spatial resolution. N_{mass} has both species-level[1] and plot-level trait[2] means. P_{mass} only have plot-level trait means (community-weighted means, CWMs)[2].

Trait[a]	Abbreviation[b]	Unit[c]	N_{plot}[d]	$N_{species}$[e]	$N_{iNaturalist}$[f]	N_{GAM01}[g]	N_{GAM2}[h]	$N_{Comparison}$[i]
Leaf nitrogen content per unit mass	N_{mass}	mg/g	77,074	35,773	3,568	830,086	2,397	1,924
Leaf phosphorus content per unit mass	P_{mass}	mg/g		0	3,490			1,881

Table A1.17 | Summary of principal component analysis (PCA) for six major plant functional traits in three plant groups. Including trait loadings, eigenvalues, and the proportion of trait variation explained by the first three successive principal components (PC1 – 3). All six traits were natural-log transformed before the analysis. Abbreviations and units of traits are shown in Table 2.1.

Variables	Non-woody			Woody deciduous			Woody evergreen		
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
LA	0.4998	-0.2762	0.2842	-0.5059	0.3695	-0.1543	0.5494	-0.1859	0.0642
LMA	-0.5141	-0.2084	0.1387	-0.0397	-0.7282	0.2084	0.0064	0.7093	0.0361
N_{mass}	0.4485	0.3251	-0.3971	-0.1885	0.0794	0.9328	0.2025	-0.5937	-0.0853
SSD	-0.4702	-0.1538	-0.5005	-0.2968	-0.5653	-0.2220	0.3370	0.0788	0.8877
H	0.0954	-0.6949	0.2567	-0.5589	0.0555	0.0365	0.4726	0.3127	-0.3668
DM	0.2336	-0.5178	-0.6526	-0.5536	-0.0655	-0.1095	0.5659	0.0769	-0.2545
Eigenvalue	1.4888	1.1982	0.9019	1.5796	1.1849	0.9916	1.4941	1.2736	0.9433
Proportion Explained	0.3694	0.2393	0.1356	0.4158	0.2340	0.1639	0.3720	0.2703	0.1483
Cumulative Proportion	0.3694	0.6087	0.7443	0.4158	0.6498	0.8137	0.3720	0.6424	0.7907

Table A1.18 | Summary of redundancy analysis (RDA) for six major plant functional traits in three plant groups. Including trait loadings, eigenvalues, and the proportion of trait variation explained by three successive redundancy analysis axes (RDA1 – 3) (constrained by climate) and the first three residual principal components (PC1 – 3) (unrelated to climate). All six traits were natural-log transformed before the analysis. Abbreviations and units of traits are shown in Table 2.1.

Variables	Non-woody						Woody deciduous						Woody evergreen					
	RDA1	RDA2	RDA3	PC1	PC2	PC3	RDA1	RDA2	RDA3	PC1	PC2	PC3	RDA1	RDA2	RDA3	PC1	PC2	PC3
LA	-1.4986	-2.1922	0.5450	6.7914	-3.7532	2.0324	1.2672	2.9152	0.2329	-6.4083	1.4896	-2.2159	6.1039	-1.4540	0.5490	-3.5106	2.1884	-2.7798
LMA	1.7223	0.6232	-0.1858	-7.0882	-2.7177	0.7665	3.9240	-2.1743	0.2668	1.3469	-3.9399	3.5221	-1.5264	4.7889	0.2231	-3.6463	-4.9494	1.4602
N_{mass}	-1.4116	0.8649	0.1843	6.4414	4.2198	-2.2445	1.9550	-0.6374	-0.2276	-2.1050	4.9285	5.7803	2.2747	-0.9434	-0.8206	1.2113	7.6262	3.0084
SSD	1.6655	2.1140	0.4748	-6.0969	-2.0206	-4.0442	2.6515	-0.6947	-0.0648	-2.8142	-5.8111	2.0031	5.8841	2.5167	-0.9290	-0.1593	-1.6434	3.5972
H	4.2170	-1.4731	-0.1569	2.0779	-6.3099	2.8154	2.2474	1.1576	1.0082	-6.8594	-0.0239	0.3785	1.8677	0.6256	1.6664	-7.4256	0.2394	2.4245
DM	2.5977	-0.1717	0.4880	4.1366	-4.8746	-5.9032	3.3074	1.6100	-0.9043	-6.1223	-1.4226	-0.2379	3.3681	0.6989	0.3593	-6.6437	2.7223	-1.5801
Eigenvalue	0.3739	0.1366	0.0092	2.1283	1.1585	0.7463	0.6453	0.2656	0.0297	2.0571	1.1455	0.8086	1.1862	0.4174	0.0604	1.5913	1.229	0.5064
Proportion Explained	0.0623	0.0228	0.0015	0.3547	0.1931	0.1244	0.1076	0.0443	0.0049	0.3428	0.1909	0.1348	0.1977	0.0696	0.0101	0.2652	0.2048	0.0844
Cumulative Proportion	0.0623	0.0851	0.0866	0.4413	0.6344	0.7588	0.1076	0.1518	0.1568	0.4996	0.6905	0.8253	0.1977	0.2673	0.2773	0.5425	0.7474	0.8318

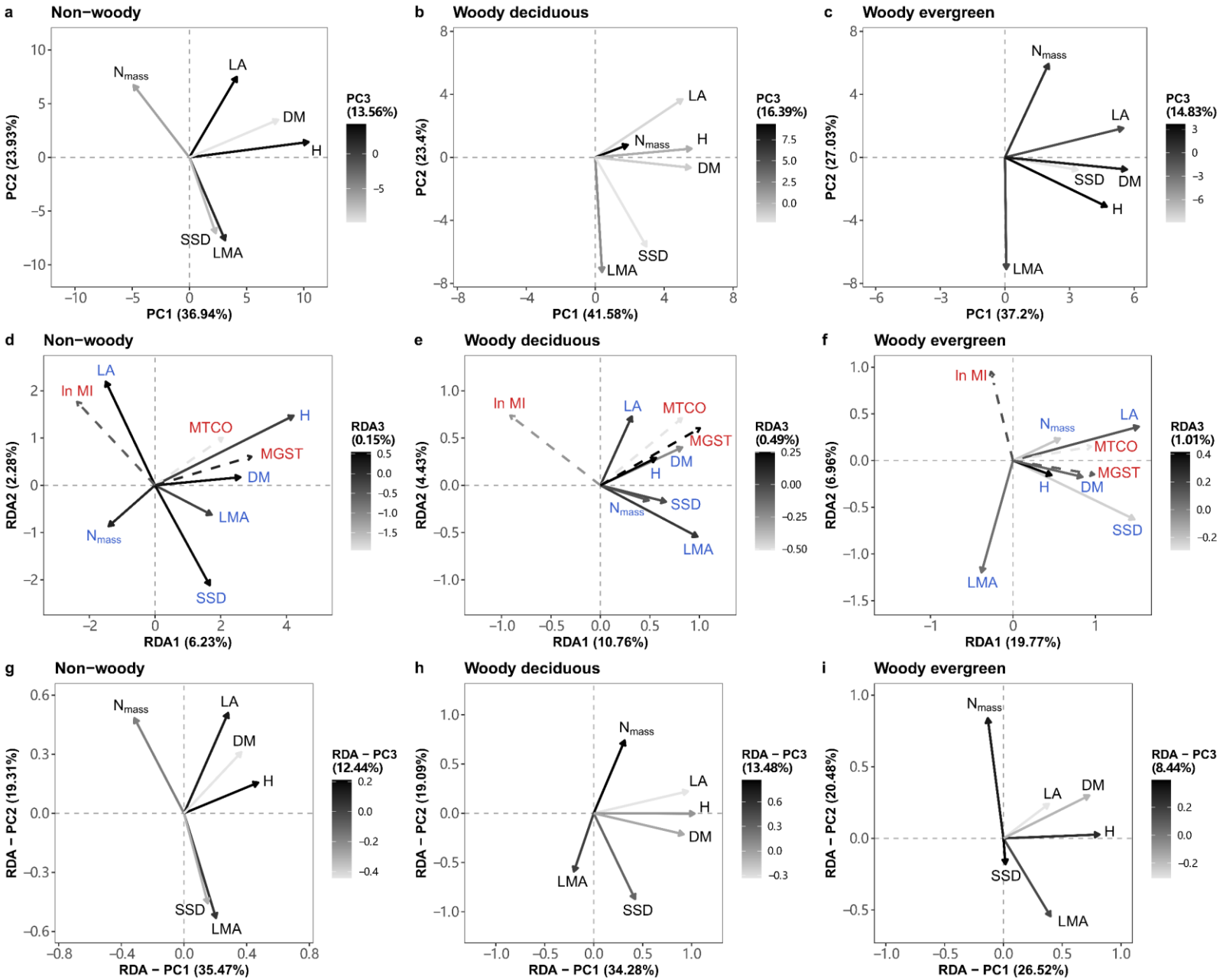

a
Non−woody
PC2 (23.93%)
PC1 (36.94%)
PC3 (13.56%)
b
Woody deciduous
PC2 (23.4%)
PC1 (41.58%)
PC3 (16.39%)
c
Woody evergreen
PC2 (27.03%)
PC1 (37.2%)
PC3 (14.83%)
d
Non−woody
RDA2 (2.28%)
RDA1 (6.23%)
RDA3 (0.15%)
e
Woody deciduous
RDA2 (4.43%)
RDA1 (10.76%)
RDA3 (0.49%)
f
Woody evergreen
RDA2 (6.96%)
RDA1 (19.77%)
RDA3 (1.01%)
g
Non−woody
RDA − PC2 (19.31%)
RDA − PC1 (35.47%)
RDA − PC3 (12.44%)
h
Woody deciduous
RDA − PC2 (19.09%)
RDA − PC1 (34.28%)
RDA − PC3 (13.48%)
i
Woody evergreen
RDA − PC2 (20.48%)
RDA − PC1 (26.52%)
RDA − PC3 (8.44%)
Nmass
LA
DM
H
SSD
LMA
In MI
MTCO
MGST

Fig. A1.17 | Principal component analysis (PCA) (a – c), redundancy analysis (RDA) (d – f) and residual (climate-independent) dimensions of trait variation from redundancy analysis (g – i) for six major plant functional traits of non-woody (a, d and g), woody deciduous (b, e and h) and woody evergreen (c, f and i) plants. The orientation of axes has been rotated according to the Figure 2 in ref. [5]. Here we replaced area-based leaf nitrogen content (N_{area}) with mass-based leaf nitrogen content (N_{mass}). Gray scales indicate the loadings on the third axis. Solid arrows (with blue labels in RDA) represent traits and dashed arrows (with red labels in RDA) represent bioclimatic variables. All six traits were natural-log transformed before analysis. Both log-transformed traits and bioclimatic variables were rescaled to a mean of 0 and a standard deviation of 1 before analysis. Abbreviations and units of traits are shown in Table 2.1. ln MI, log-transformed moisture index; MTCO, mean temperature of the coldest month; MGST, mean growing-season temperature (see Methods for definition). All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

Table A1.19 | Summary of Generalised Additive Models (GAMs) based on three bioclimatic variables for mass-based leaf nitrogen content (N_{mass}) in three plant groups. N_{mass} is natural-log transformed. Including number of vegetation plots containing community-weighted means (CWMs) of three plant groups for each trait (Sample size), and explained deviance, adjusted R^2 and P values of each GAM model.

Plant groups	N_{mass}			
	Sample size	Deviance explained	Adjusted R^2	*P* value
Non-woody	59,964	17.42%	0.1739	$< 2 \times 10^{-16}$ ***
Woody deciduous	38,037	8.98%	0.0892	< 2 × 10-16 ***
Woody evergreen	46,629	16.60%	0.1656	< 2 × 10-16 ***

Table A1.20 | Relative importance of each bioclimatic variable in predicting mass-based leaf nitrogen content (N_{mass}) of three plant groups. N_{mass} is natural-log transformed. ln MI, log-transformed moisture index; MTCO, mean temperature of the coldest month; MGST, mean growing-season temperature (see Methods for definition).

Plant groups	N_{mass}		
	ln MI	MTCO	MGST
Non-woody	11.32%	38.80%	49.88%
Woody deciduous	11.54%	54.70%	33.75%
Woody evergreen	29.47%	38.00%	32.53%

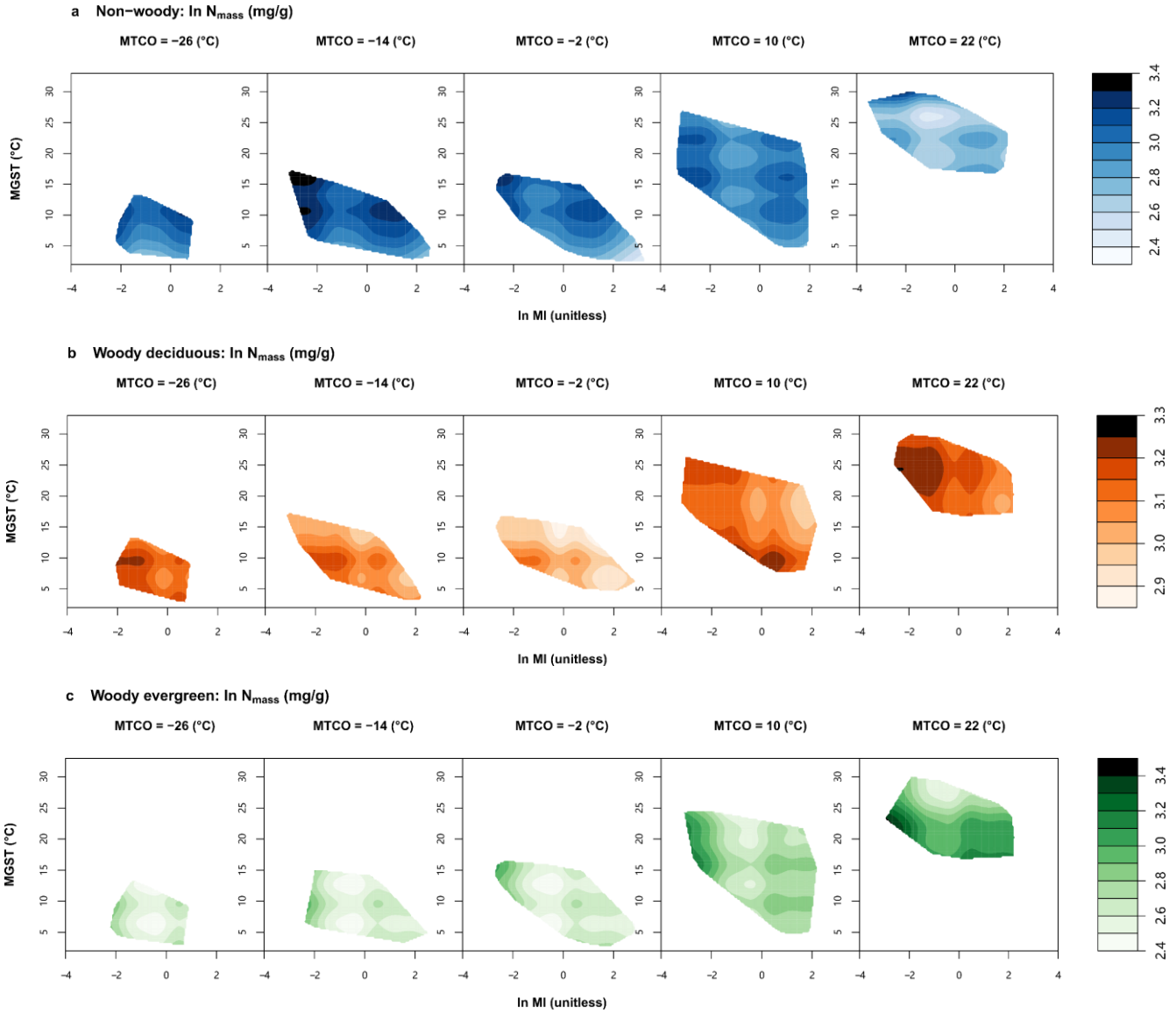

a Non−woody: ln N$_{mass}$ (mg/g)
MTCO = −26 (°C)
MTCO = −14 (°C)
MTCO = −2 (°C)
MTCO = 10 (°C)
MTCO = 22 (°C)
MGST (°C)
ln MI (unitless)
b Woody deciduous: ln N$_{mass}$ (mg/g)
c Woody evergreen: ln N$_{mass}$ (mg/g)

Fig. A1.18 | Climate space diagrams of mass-based leaf nitrogen content (N_{mass}) for (a) non-woody, (b) woody deciduous and (c) woody evergreen plants. Showing distributions of natural-log transformed N_{mass} values of three plant groups in the global climate space defined by three bioclimatic variables. The predicted trait values are presented as contours with darker colour representing higher values. ln MI, natural-log transformed; MTCO, mean temperature of the coldest month; MGST, mean growing-season temperature (see Methods for definition). All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

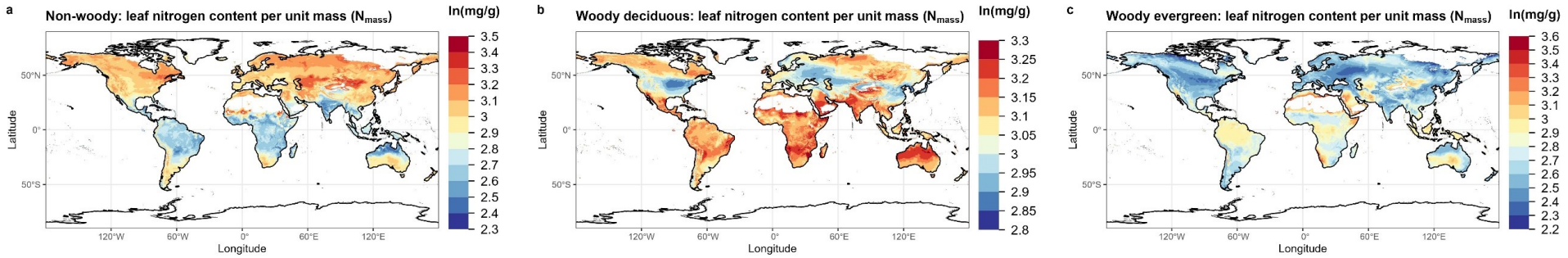


Fig. A1.19 | Separate global trait maps of mass-based leaf nitrogen content (N_{mass}) for non-woody (a), woody deciduous (b) and woody evergreen (c) plants. Global patterns of natural-log transformed N_{mass} values predicted by Generalised Additive Models (GAMs) based on three bioclimatic variables. All maps are at a 0.1° spatial resolution. For maps in GeoTiff format, refer to the Data availability statement. All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

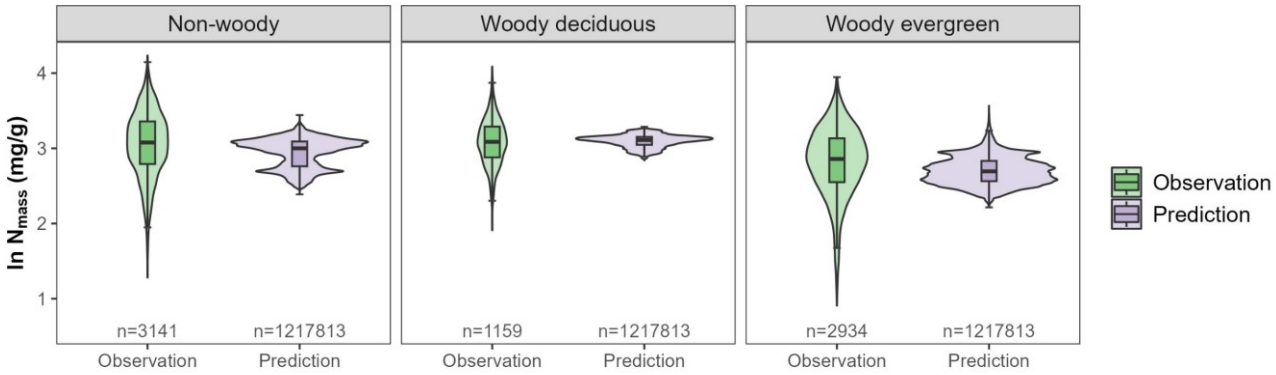


Fig. A1.20 | Violin plots showing comparisons between observed and predicted values of mass-based leaf nitrogen content (N_{mass}) in three plant groups. N_{mass} is natural-log transformed. Green violin plots show community-weighted means (CWMs) of N_{mass} in three plant groups to fit Generalised Additive Models (GAMs). Purple violin plots show pixel values of separate global trait maps for each of the three plant groups. All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

Table A1.21 | Summary of the Generalised Additive Models (GAMs) based on three bioclimatic variables and global fractional covers of two of three plant groups for mass-based leaf nitrogen content (N_{mass}) and leaf phosphorus content (P_{mass}) in 77,074 natural vegetation plots.

N_{mass} and P_{mass} were natural-log transformed before analysis. Including explained deviance, adjusted R^2 and P values of each GAM model.

Trait	Deviance explained	Adjusted R^2	*P* value
N_{mass}	44.78%	0.4464	$< 2 \times 10^{-16}$ ***
P_{mass}	48.60%	0.4847	$< 2 \times 10^{-16}$ ***

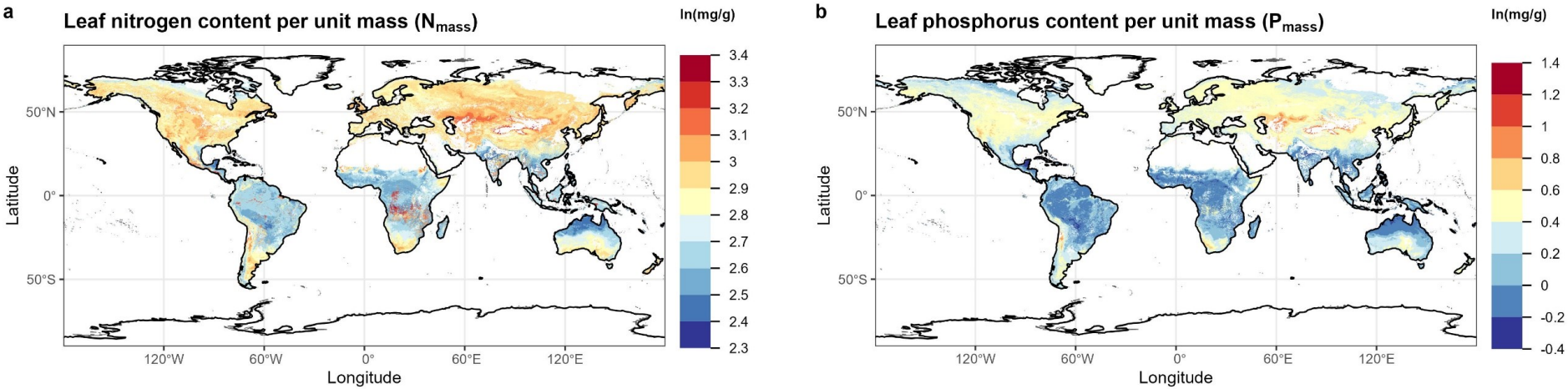


Fig. A1.21 | Global trait maps of mass-based leaf nitrogen content (N_{mass}, a) and mass-based leaf phosphorus content (P_{mass}, b). Global patterns of natural-log transformed trait values predicted by the Generalised Additive Models (GAMs) based on three bioclimatic variables and global fractional covers of two of three plant groups (see Methods). Both maps are at a 0.1° spatial resolution. For maps in GeoTiff format, refer to the Data availability statement. All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

Table A1.22 | Summary of comparison between model-predicted values and iNaturalist estimated values of mass-based leaf nitrogen content (N_{mass}) and leaf phosphorus content (P_{mass}). Including parameters for assessing agreements between model-predicted map pixel values and the iNaturalist map pixel values. The agreements were estimated by linear regression at a 2° spatial resolution. R^2 is the coefficient of determination; RRMSE is the root-mean-square error, as a proportion of the observed mean trait value (here observations are represented by iNaturalist trait values); bias is the difference between observed and model-predicted mean values, as a proportion of the observed mean trait value (here observed values are iNaturalist trait values); slope is the slope of the linear regression of iNaturalist estimation against model prediction. Both traits are natural-log transformed.

Trait	R^2	RRMSE	bias	slope
N_{mass}	0.024	0.105	-0.0744	0.21
P_{mass}	0.071	0.769	-0.43	0.373

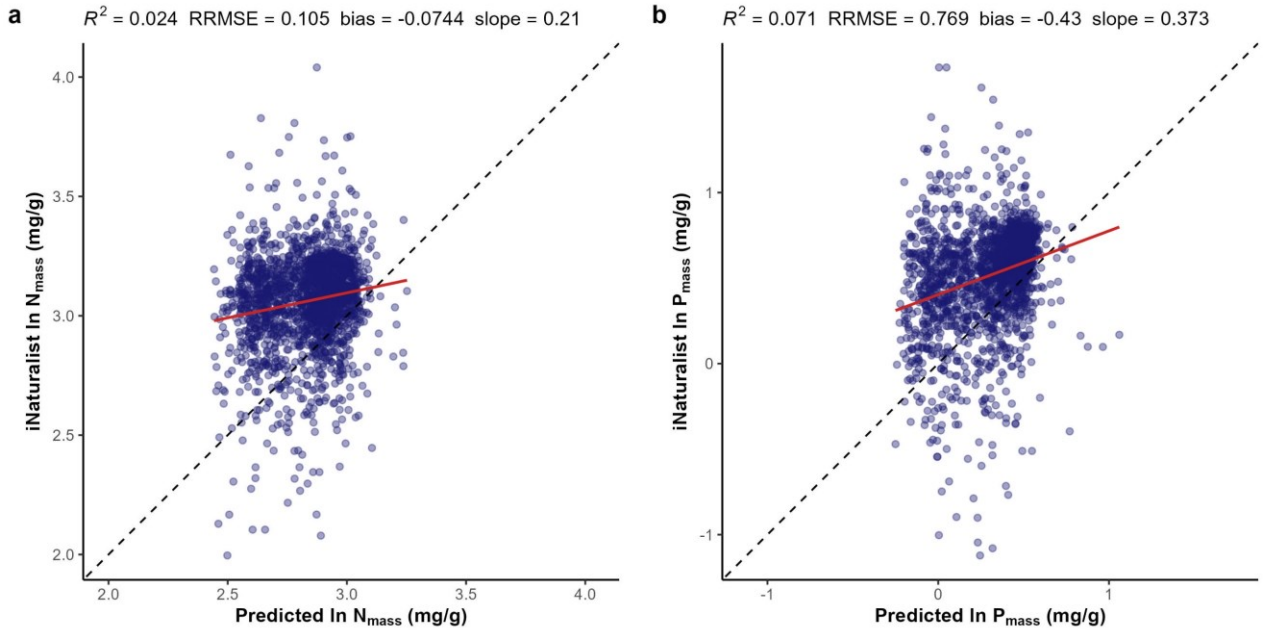


Fig. A1.22 | Scatter plots of comparison between predicted trait values in this study and estimated trait values from the iNaturalist for mass-based leaf nitrogen content (N_{mass}, a) and mass-based leaf phosphorus content (P_{mass}, b). Red lines represent correlations between model-predicted map pixel values and the iNaturalist map pixel values estimated by linear regression at a 2° spatial resolution. The dotted line is the 1:1 line. Parameters on each of the above plots for assessing agreements between global trait maps in this study and the iNaturalist trait maps are listed in the caption of the Table A1.13. Both traits are natural-log transformed. All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

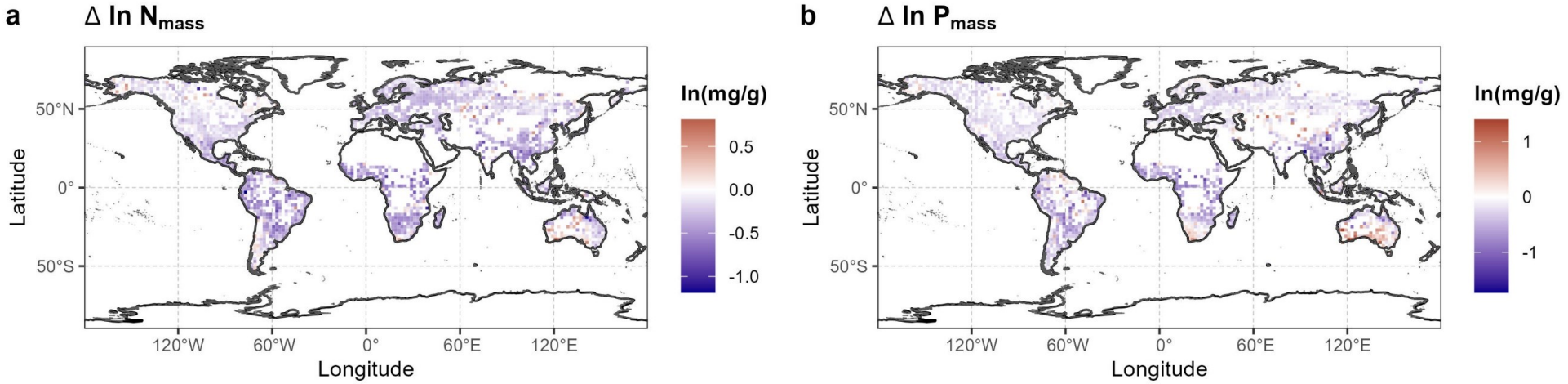


Fig. A1.23 | Global maps of differentiation between predicted trait values in this study and estimated trait values from the iNaturalist at a 2° spatial resolution. a, mass-based leaf nitrogen content (N_{mass}) and **b**, mass-based leaf phosphorus content (P_{mass}). Darker colour suggests greater difference between model prediction and iNaturalist estimation. Both traits are natural-log transformed. All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

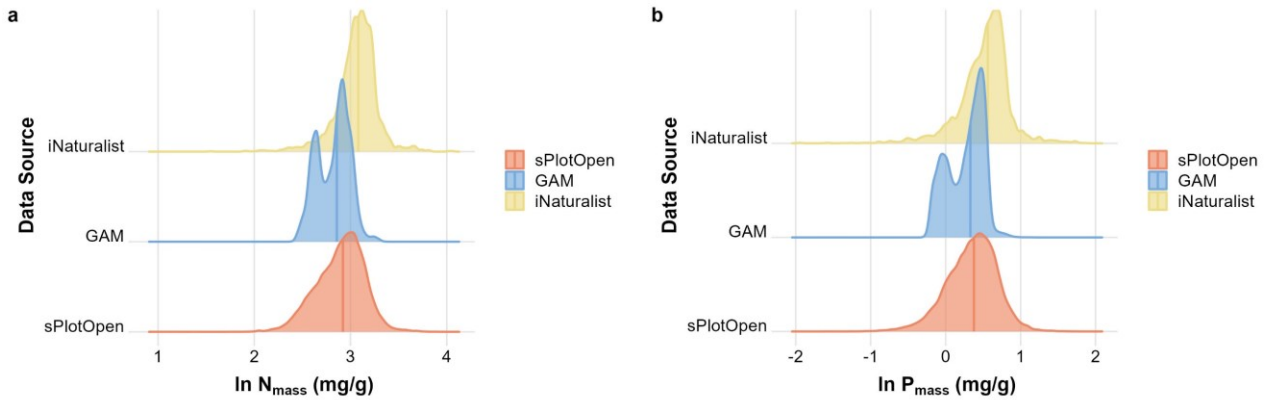
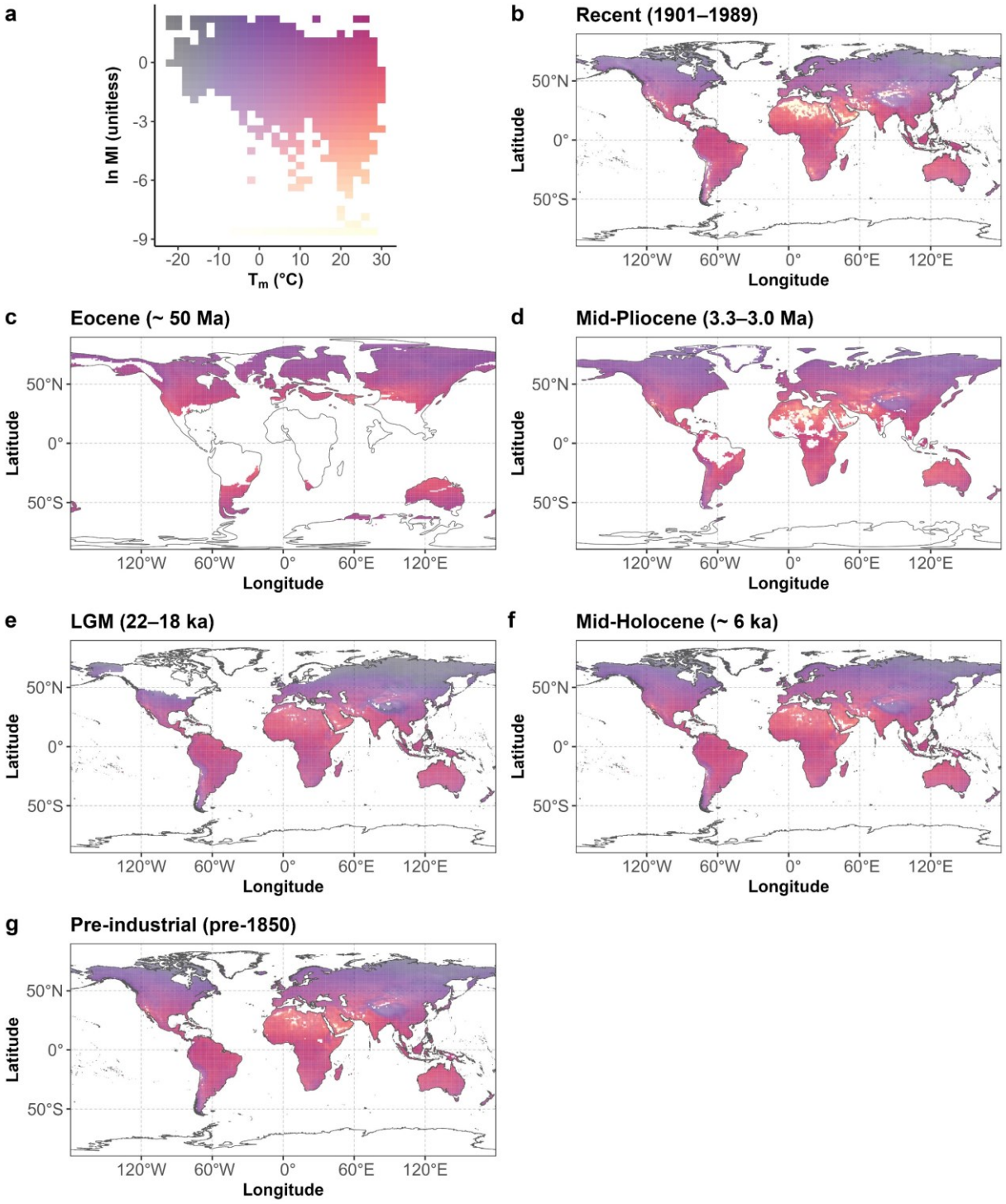


Fig. A1.24 | Comparisons of trait value distributions among sample trait data based on the sPlotOpen database, model prediction in this study and iNaturalist estimation for mass-based leaf nitrogen content (N_{mass}, a) and mass-based leaf phosphorus content (P_{mass}, b). Both traits are natural-log transformed. All elements of the figure were created by the authors using R v4.2.2 (https://www.r-project.org/)[4].

A2 Supplementary information for Chapter 3

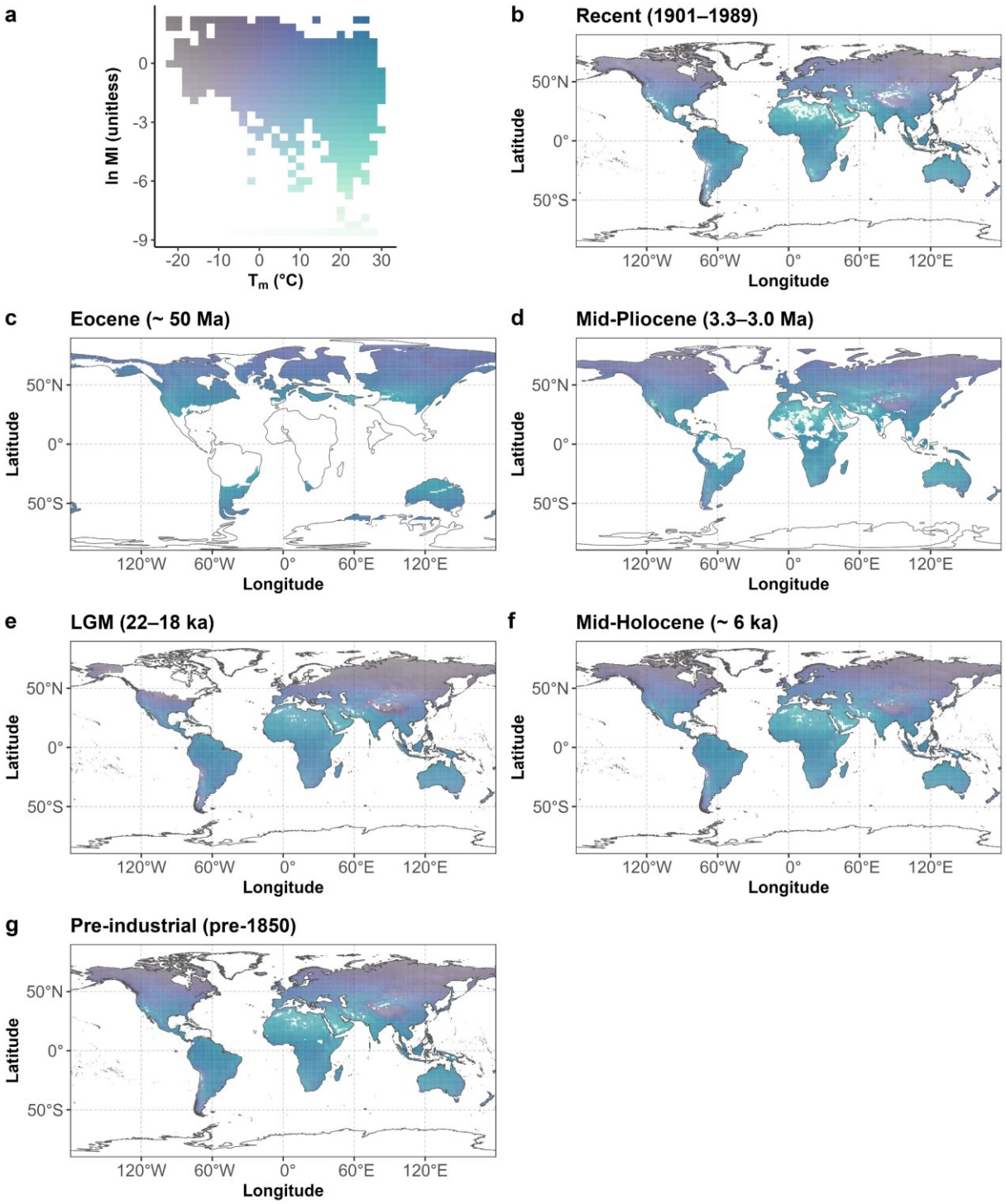
A2.1 Supplementary figures

Fig. A2.1: Alternative colour gradient scheme illustrating the historical geography of recent climates.



a, The two-dimensional climatic space defined by the recent (1901–1989) climatic conditions (T_m and ln MI), partitioned into 30 equal intervals along each axis. Each coloured grid cell ('climatic bin') represents a unique climatic condition (405 cells in total). **b**–**g**, Geographic locations where those recent climatic conditions existed during the recent period (1901–1989, **b**), Eocene (*ca.* 50 Ma, **c**), Mid-Pliocene (3.3–3.0 Ma, **d**), Last Glacial Maximum (LGM, 22–18 ka, **e**), Mid-Holocene (*ca.* 6 ka, **f**) and pre-industrial period (pre-1850, **g**). Uncoloured areas denote climatic conditions absent during this period. Colours correspond to those in **a**, with identical colours indicating the same climatic conditions. All maps are shown at a 1° spatial resolution.

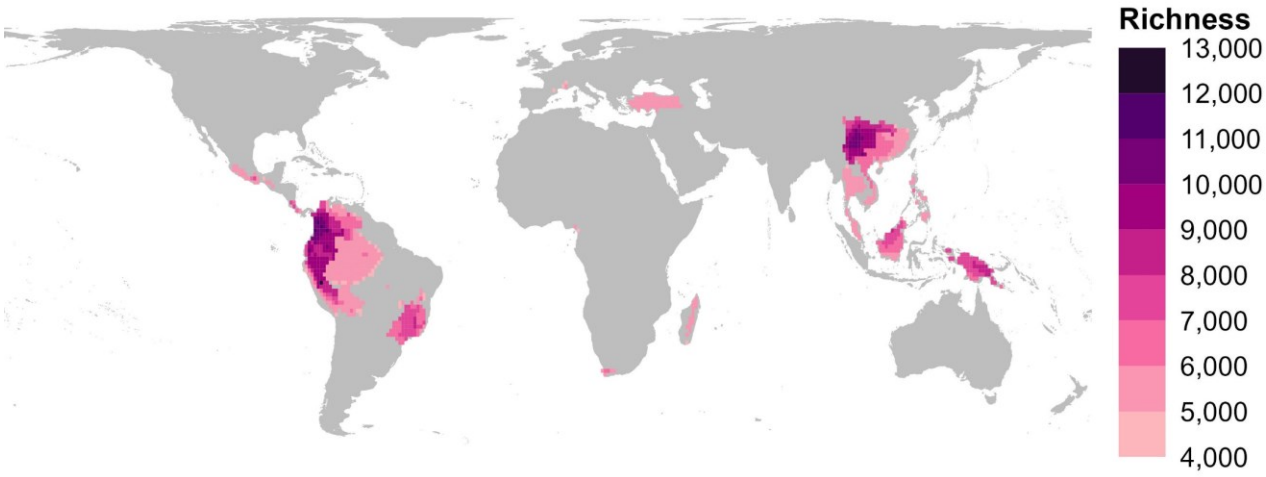
Fig. A2.2: Another alternative colour gradient scheme illustrating the historical geography of recent climates.



a, The two-dimensional climatic space defined by the recent (1901–1989) climatic conditions (T_m and ln MI), partitioned into 30 equal intervals along each axis. Each coloured grid cell ('climatic bin') represents a unique climatic condition (405 cells in total). **b**–**g**, Geographic locations where those

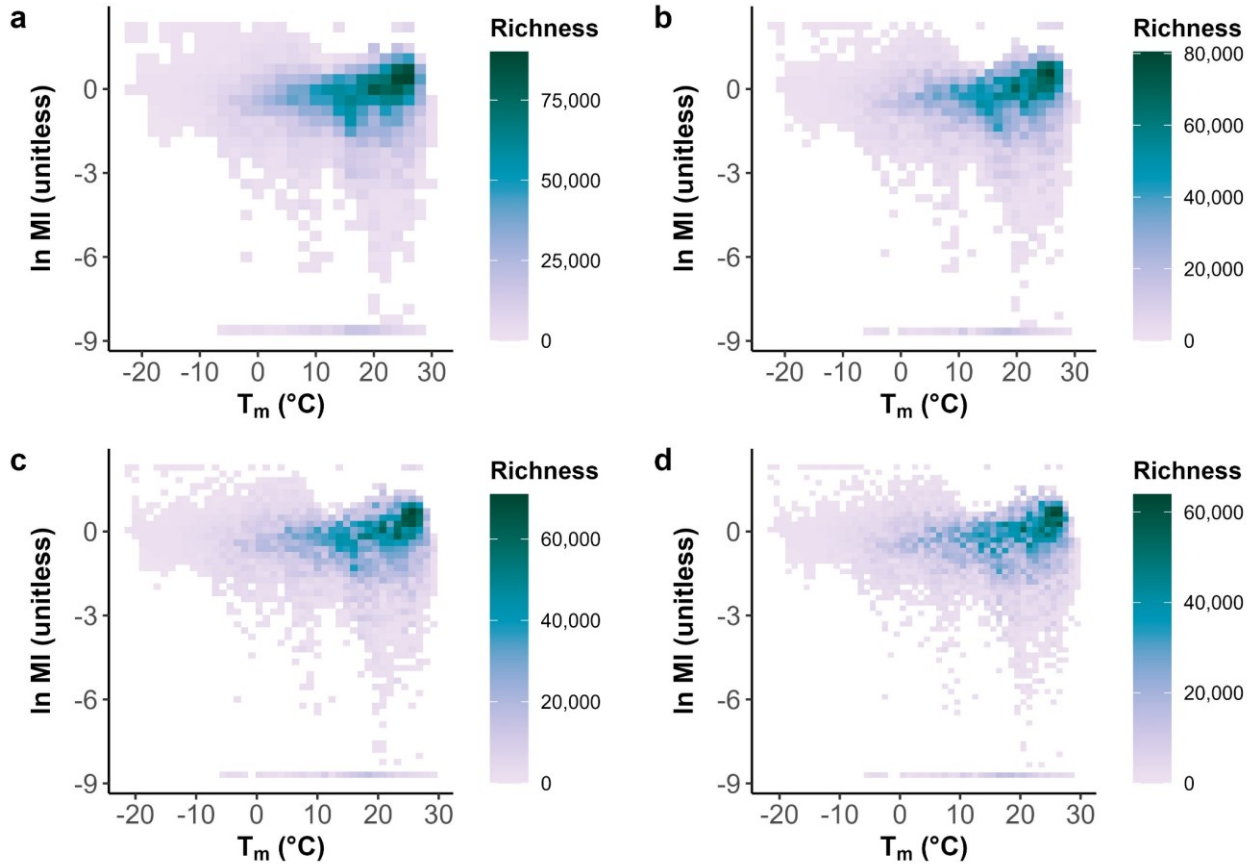
recent climatic conditions existed during the recent period (1901–1989, **b**), Eocene (*ca.* 50 Ma, **c**), Mid-Pliocene (3.3–3.0 Ma, **d**), Last Glacial Maximum (LGM, 22–18 ka, **e**), Mid-Holocene (*ca.* 6 ka, **f**) and pre-industrial period (pre-1850, **g**). Uncoloured areas denote climatic conditions absent during this period. Colours correspond to those in **a**, with identical colours indicating the same climatic conditions. All maps are shown at a 1° spatial resolution.

Fig. A2.3: Global hotspots of vascular plant species richness.



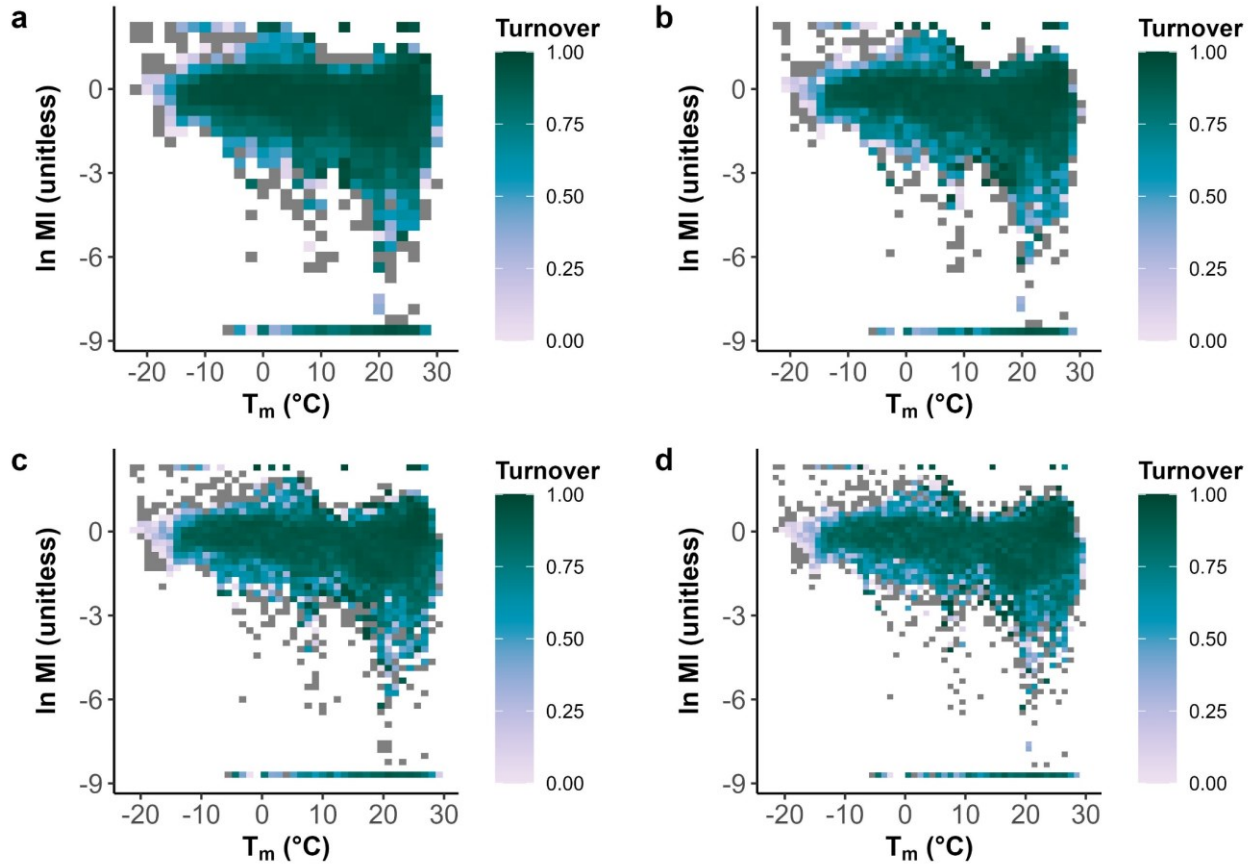
Global hotspots of vascular plant diversity using the Behrmann equal-area projection at ~110 km resolution (~1° resolution), defined as regions where species richness exceeds the 90th global quantile[12].

Fig. A2.4: Vascular plant species richness in climatic space at different resolutions.



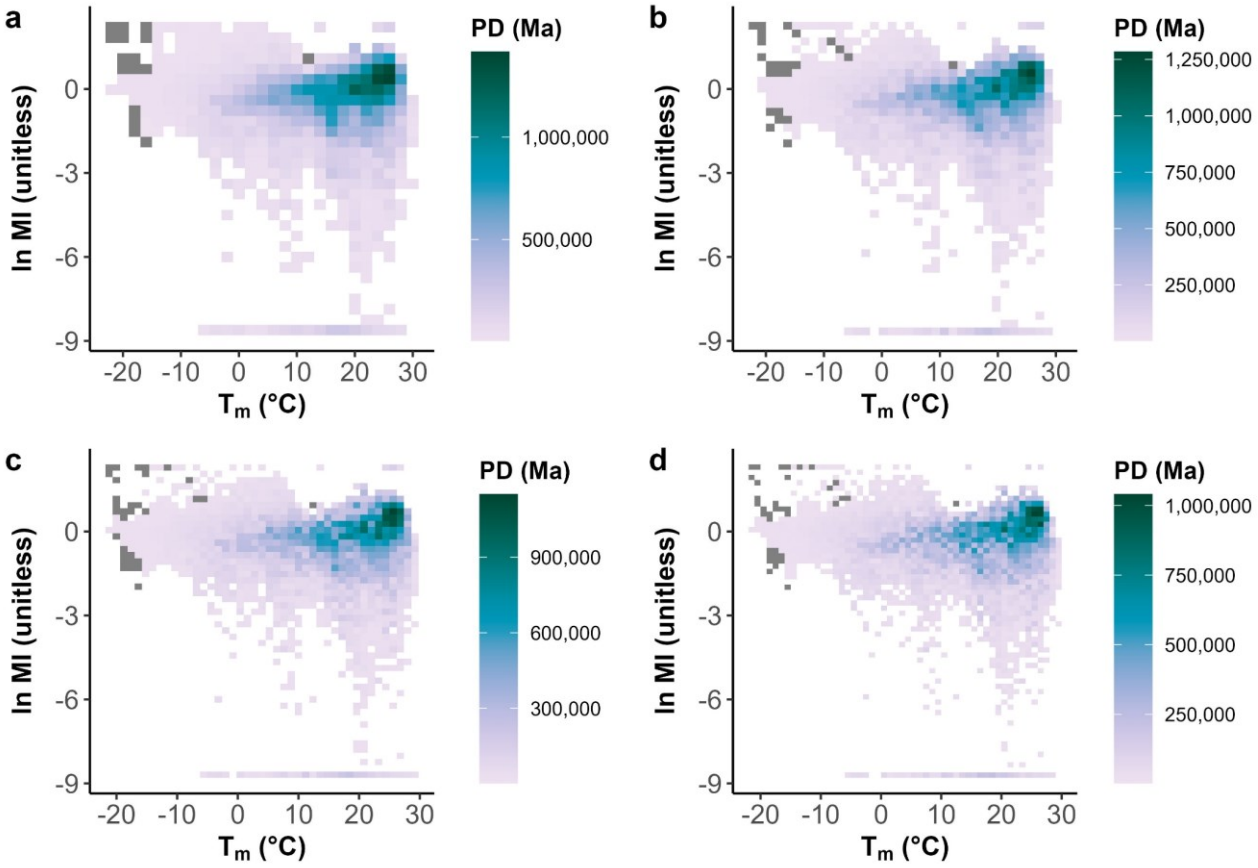
Species richness is shown per grid cell in climatic space defined by recent climatic conditions (T_m and ln MI), at four resolutions: 30 (**a**, 405 cells), 40 (**b**, 628 cells), 50 (**c**, 896 cells) and 60 (**d**, 1182 cells) equal intervals along each axis. Each grid cell ('climatic bin') represents a distinct climatic condition; colours indicate species richness. T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index.

Fig. A2.5: Vascular plant species turnover in climatic space at different resolutions.



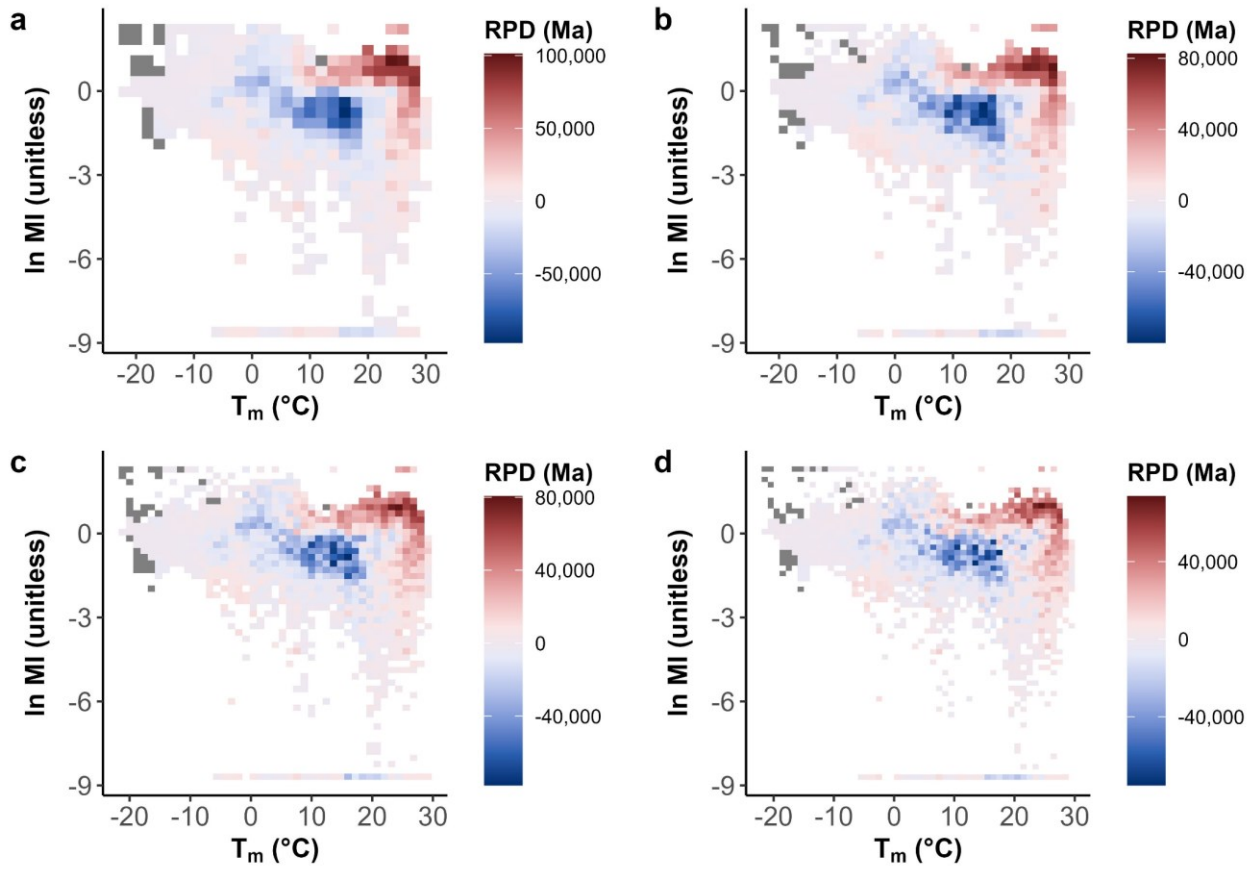
Species turnover (the turnover component of beta-diversity) is shown per grid cell in climatic space defined by recent climatic conditions (T_m and ln MI), at four resolutions: 30 (**a**, 405 cells), 40 (**b**, 628 cells), 50 (**c**, 896 cells) and 60 (**d**, 1182 cells) equal intervals along each axis. Each grid cell ('climatic bin') represents a distinct climatic condition; colours indicate species richness. T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index.

Fig. A2.6: Phylogenetic diversity of vascular plants in climatic space at different resolutions.



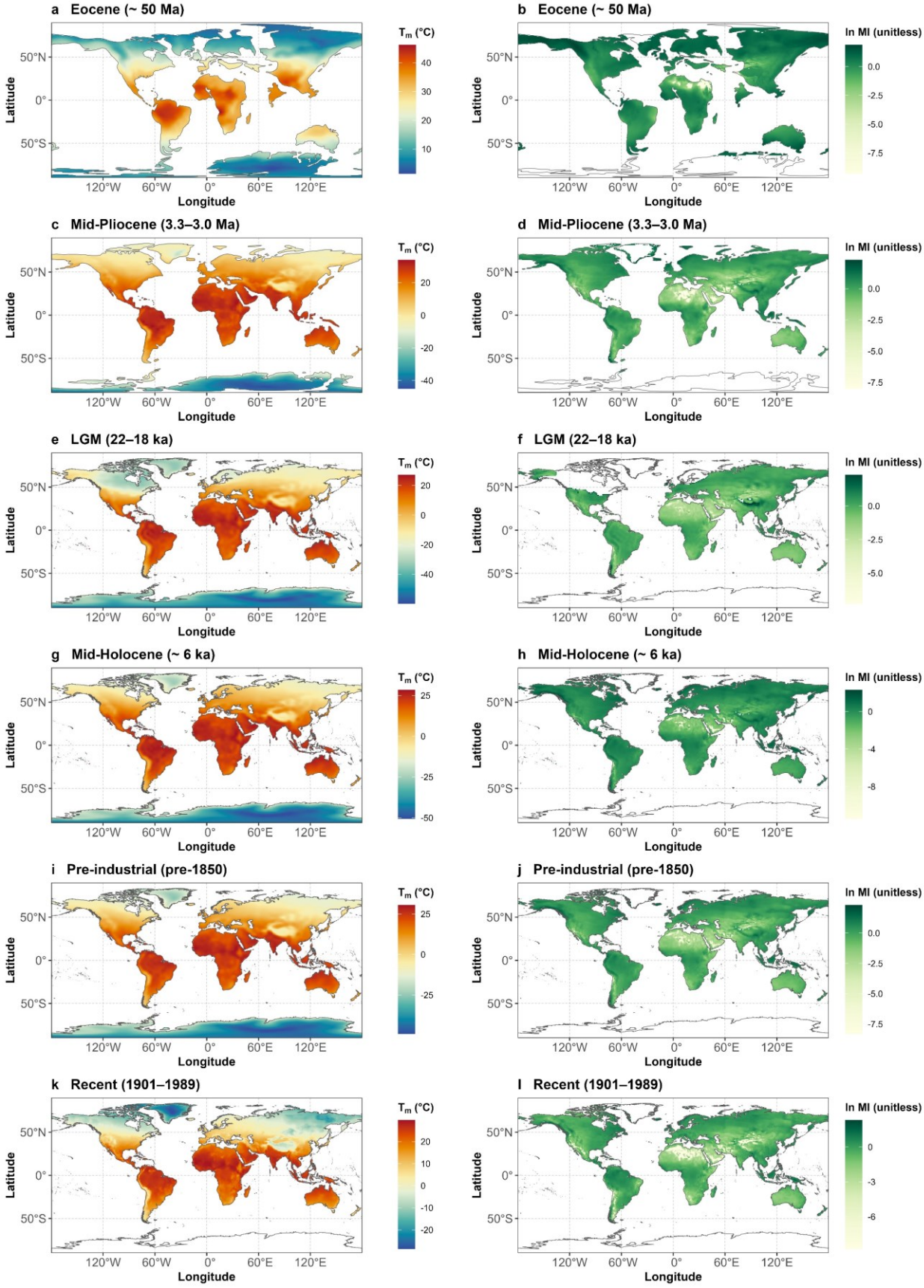
Phylogenetic diversity (Faith's phylogenetic diversity[13], PD, Ma) is shown per grid cell in climatic space defined by recent climatic conditions (T_m and ln MI), at four resolutions: 30 (**a**, 405 cells), 40 (**b**, 628 cells), 50 (**c**, 896 cells) and 60 (**d**, 1182 cells) equal intervals along each axis. Each grid cell ('climatic bin') represents a distinct climatic condition; colours indicate species richness. T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index.

Fig. A2.7: Relative phylogenetic diversity of vascular plants in climatic space at different resolutions.



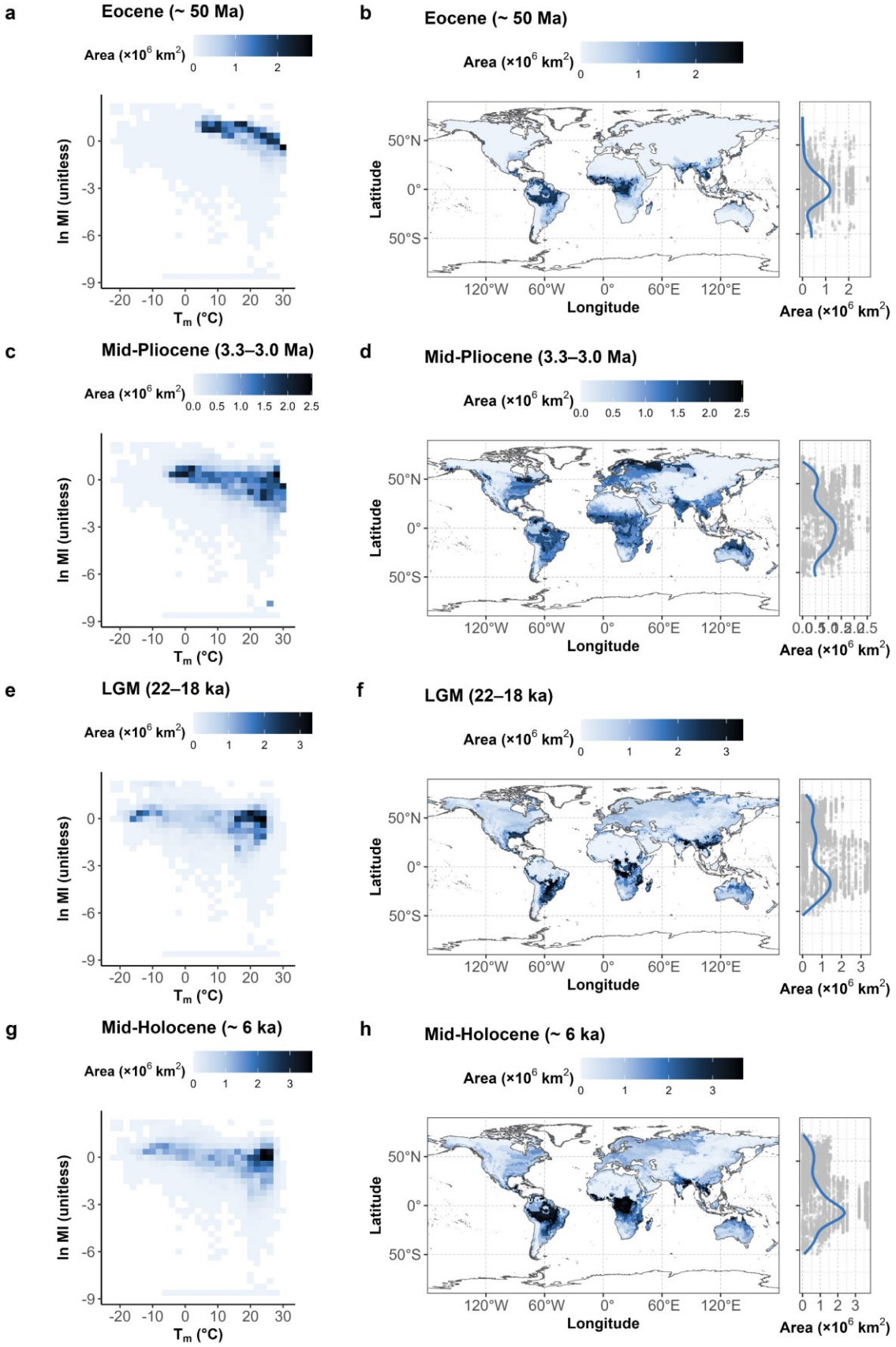
Relative phylogenetic diversity (calculated as the residuals from regressing phylogenetic diversity against species richness, Ma) is shown per grid cell in climatic space defined by recent climatic conditions (T_m and ln MI), at four resolutions: 30 (**a**, 405 cells), 40 (**b**, 628 cells), 50 (**c**, 896 cells) and 60 (**d**, 1182 cells) equal intervals along each axis. Each grid cell ('climatic bin') represents a distinct climatic condition; colours indicate species richness. T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index.

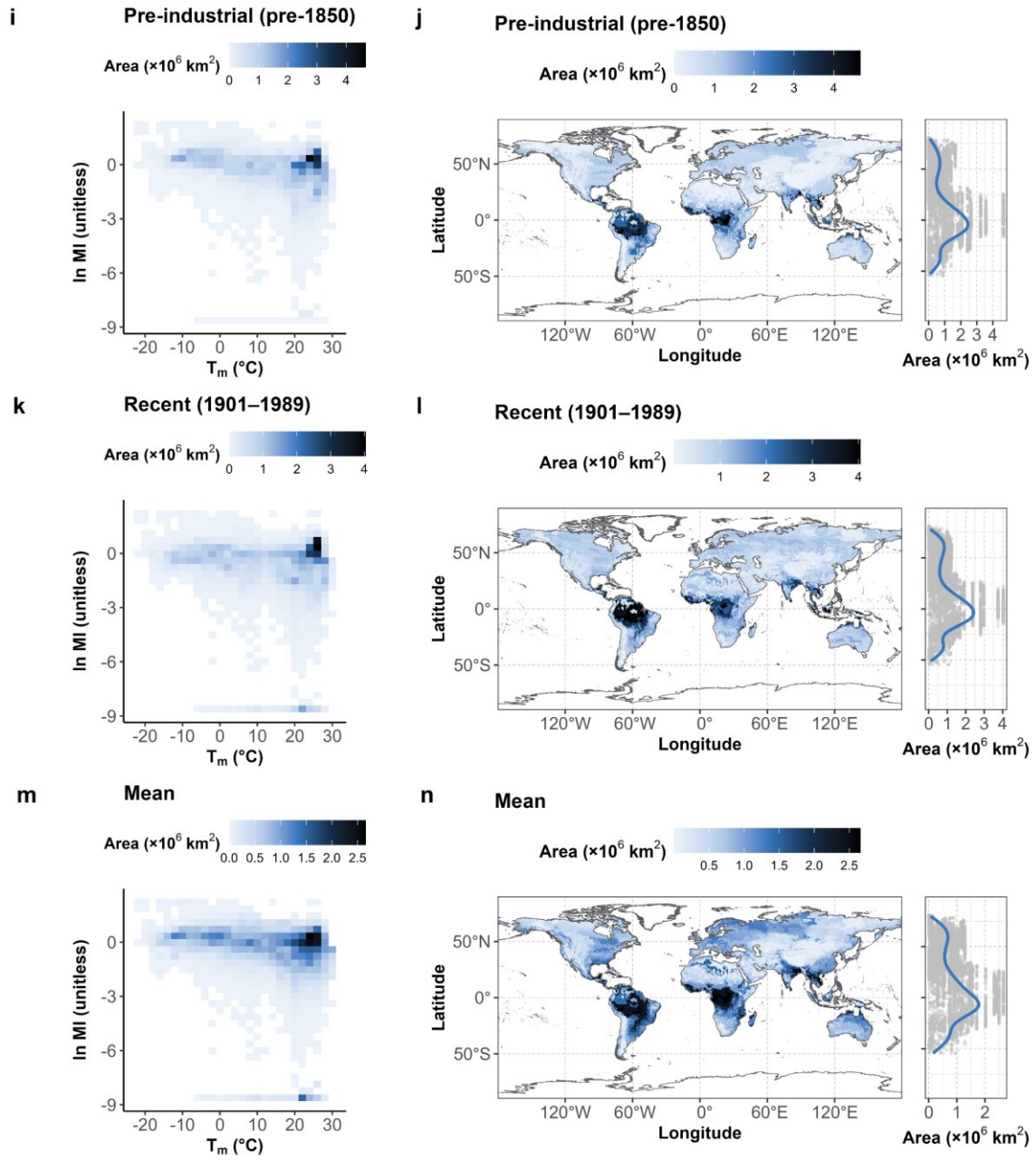
Fig. A2.8: Global distribution of bioclimatic variables in the selected geological periods at a 1° spatial resolution.



Gridded global distribution of: **a**, temperature index (T_m, °C) and **b**, natural-log-transformed moisture index (ln MI, unitless) of the Eocene (*ca.* 50 Ma); **c**, T_m and **d**, ln MI of the Mid-Pliocene (3.3–3.0 Ma); **e**, T_m and **f**, ln MI of the Last Glacial Maximum (LGM, 22–18 ka); **g**, T_m and **h**, ln MI of the Mid-Holocene (*ca.* 6 ka); **i**, T_m and **j**, ln MI of the pre-industrial period (pre-1850); **k**, T_m and **l**, ln MI of the recent period (1901–1989).

Fig. A2.9: Geographic extent of climate across geological periods.

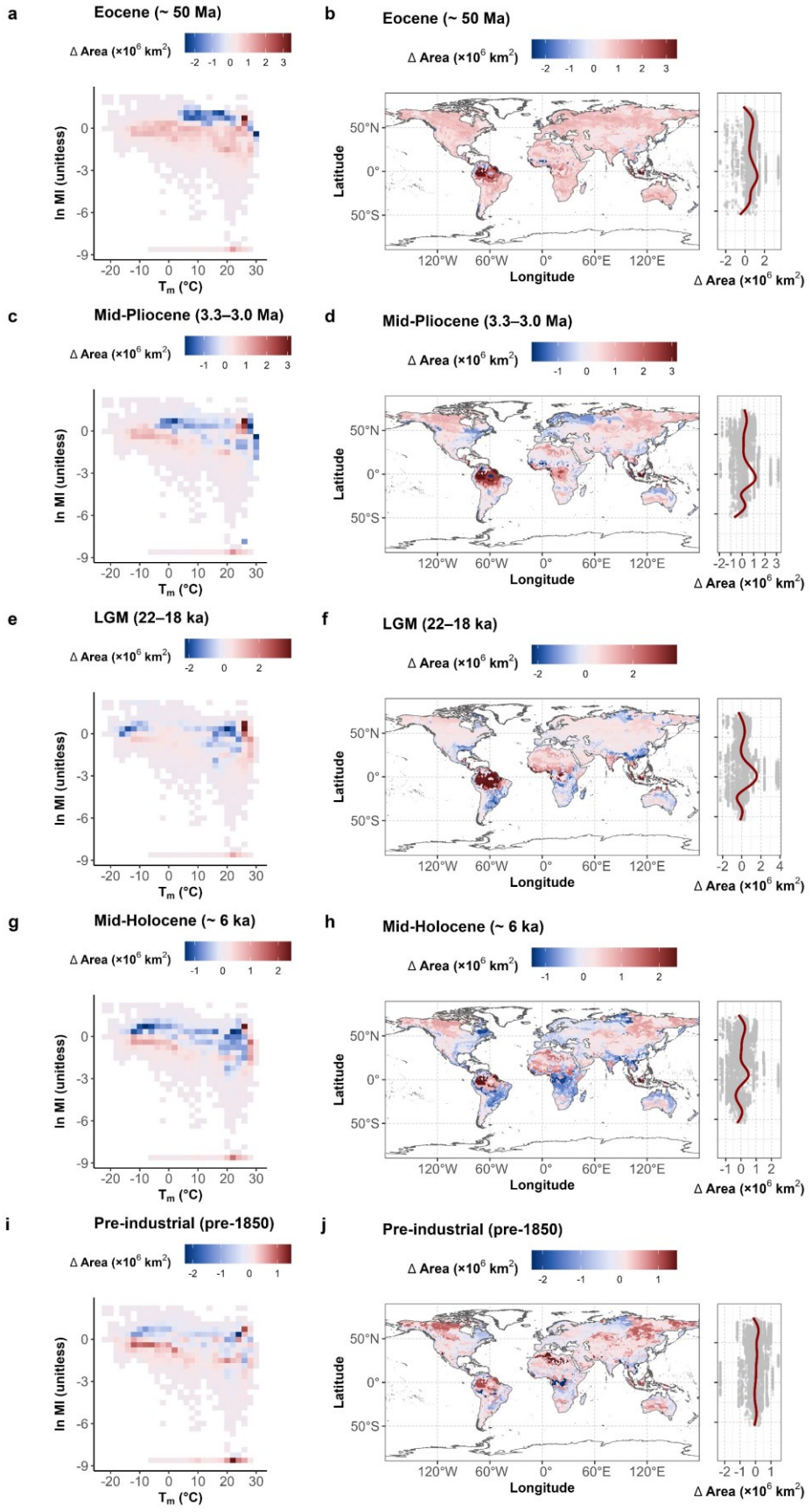




a, **c**, **e**, **g**, **i**, **k** and **m**, Patterns of the geographic extent of climate in climatic space across different geological periods. The geographic extent of climate is calculated as the total land surface area occupied by each climatic condition during the corresponding climatic period, referred to as 'Area' (km^2). Climatic space is defined by the recent (1901–1989) climatic conditions (T_m and ln MI), partitioned into 30 equal intervals along each axis. **b**, **d**, **f**, **h**, **j**, **l** and **n**, Global patterns and latitudinal distributions of Area in recent geographic space at a 1° spatial resolution. **a** and **b**, Eocene (*ca.* 50 Ma); **c** and **d**, Mid-Pliocene (3.3–3.0 Ma); **e** and **f**, Last Glacial Maximum (LGM, 22–18 ka); **g** and **h**, Mid-Holocene (*ca.* 6 ka); **i** and **j**, pre-industrial period (pre-1850); **k** and **l**, recent

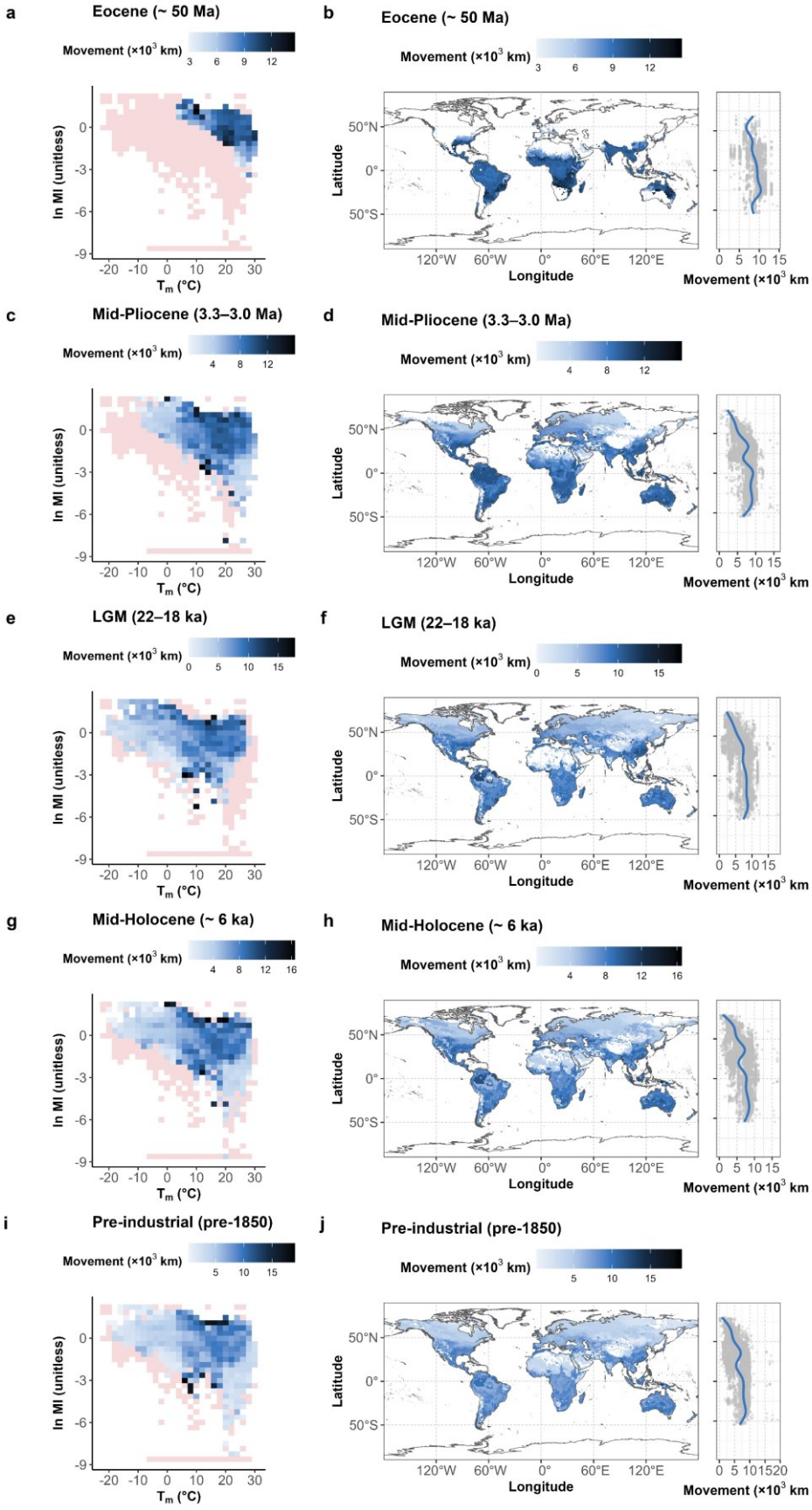
period (1901–1989); **m** and **n**, mean Area since the first appearance of each climatic condition across all selected climatic periods. T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index.

Fig. A2.10: Changes in the geographic extent of recent climatic conditions from geohistorical periods to the recent period.



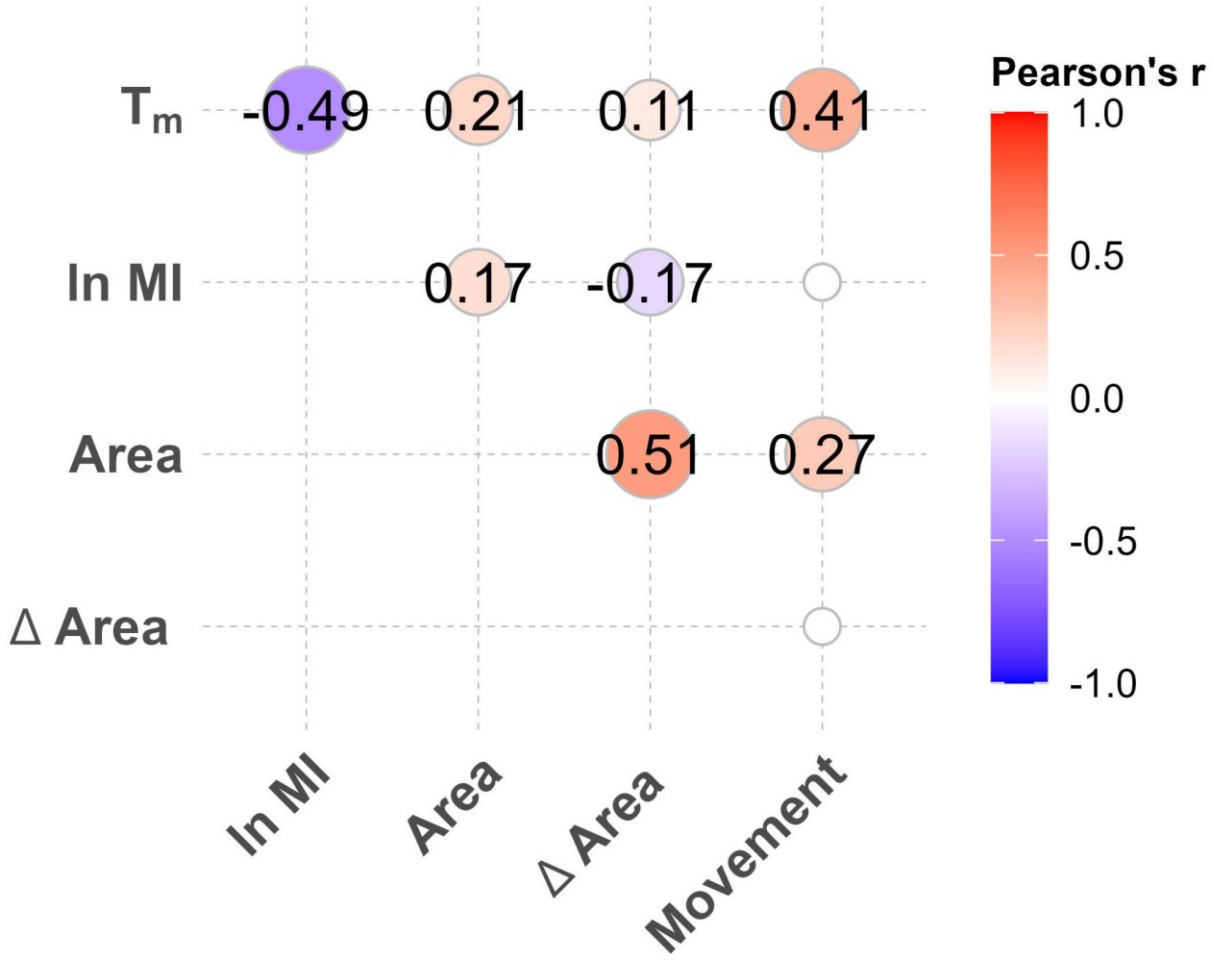
a, **c**, **e**, **g** and **i**, Change in total land area occupied by each climatic condition from its earliest occurrence to the recent period (referred to as ΔArea, km^2) in climatic space. Climatic space is defined by the recent (1901–1989) climatic conditions (T_m and ln MI), partitioned into 30 equal intervals along each axis. **b**, **d**, **f**, **h** and **j**, Global patterns and latitudinal distributions of ΔArea in recent geographic space at a 1° spatial resolution. **a** and **b**, Eocene (*ca.* 50 Ma); **c** and **d**, Mid-Pliocene (3.3–3.0 Ma); **e** and **f**, Last Glacial Maximum (LGM, 22–18 ka); **g** and **h**, Mid-Holocene (*ca.* 6 ka); **i** and **j**, pre-industrial period (pre-1850). T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index.

Fig. A2.11: Geographic shift of recent climatic conditions from geohistorical periods to the recent period.



a, **c**, **e**, **g** and **i**, Geographic shift of a climatic condition since its earliest occurrence to the recent period (Movement, km) in climatic space. Climatic space is defined by the recent (1901–1989) climatic conditions (T_m and ln MI), partitioned into 30 equal intervals along each axis. The blue colour gradient indicates climatic conditions shared by a geohistorical period and the recent period, while the pink background highlights climatic conditions present in the recent period but absent in the geohistorical period. **b**, **d**, **f**, **h** and **j**, Global patterns and latitudinal distribution of Movement in recent geographic space at a 1° spatial resolution. **a** and **b**, Eocene (*ca.* 50 Ma); **c** and **d**, Mid-Pliocene (3.3–3.0 Ma); **e** and **f**, Last Glacial Maximum (LGM, 22–18 ka); **g** and **h**, Mid-Holocene (*ca.* 6 ka); **i** and **j**, pre-industrial period (pre-1850). T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index.

Fig. A2.12: Pairwise correlation among explanatory variables.



Pairwise Pearson's correlation coefficients among all explanatory variables. The absolute values of all coefficients are below or around 0.5, indicating a low risk of concurvity or collinearity. T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence to the recent period.

A2.2 Supplementary tables

Table A2.1 | Comparison with previous global plant diversity products. Summary of data sources [a], methods for generating global plant diversity maps [b], the number of species included [c], and the spatial resolution of each map [d].

Data source [a]	Type [b]	Number of species [c]	Spatial Resolution [d]
Our research	Data-based	350,864	Any resolution (using 1° as an example)
Barthlott *et al.* (2005)[14]	Data-based	> 3,300 from different regions	10,000 km^2
Kier *et al.* (2005)[15]	Data-based	Not available – estimated for each selected geographical unit	867 Terrestrial ecoregions of the world
Kreft & Jetz (2007)[16]	Model-based	> 3,300 from 1,032 geographic units	110 × 110 km (*ca.* 1°)
Brummitt *et al.* (2021)[17]	Data-based and Model-based	Not available – estimated for each selected geographical unit (data-based); 1,250 (model-based)	Level 3 geo-political regions of the world and 867 ecoregions of WWF (data-based), 10 km^2 (model-based)
Sabatini *et al.* (2022)[18]	Model-based	53,271	400 m^2, 1,000 m^2 and 1 ha for forests, and 10 m^2, 100 m^2 and 1,000 m^2 for non-forests
Cai *et al.* (2023)[12]	Model-based	298,087	7,774 km^2, 23,322 km^2, 69,967 km^2, and 209,903 km^2

World Checklist of Vascular Plants (WCVP)[19]	Database	357,347 accepted species (Feb–June 2024) Notice: > 340,000 species with distribution information available	Level 3 of the Biodiversity Information Standards (TDWG) geographical codes
Global Inventory of Floras and Traits (GIFT)[20]	Database	303,713 vascular plants (v.3.1) Notice: distribution data for some species are not available	3,627 geographic regions

Table A2.2 | Climate data used for this study. Databases and model simulations used to extract each climatic variable in different geological periods.

Project	Version	Experiment	Geological period	Model	Simulation	Ensemble	Variable	Resolution	Reference
Deep-Time Model Intercomparison Project (DeepMIP)[21]	version 1.0	Eocene	Eocene (ca. 50 Ma)	CESM1.2-CAM5	× 6 CO_2	r1i1p1f1	elevation/surface altitude, precipitation, surface downwelling shortwave radiation, air temperature	2.5° × 1.9°	Ref [22]
				COSMOS-landveg-r2413	× 4 CO_2		elevation/surface altitude, precipitation, surface downwelling shortwave radiation, air temperature	3.75° × 3.71°	Ref [23]

				GFDL-CM2.1	× 4 CO_2, × 6 CO_2		elevation/surface altitude, precipitation, surface downwelling shortwave radiation, air temperature	3.75° × 3.02°	Ref [24]
				INM-CM4-8	× 6 CO_2		elevation/surface altitude, precipitation, surface downwelling shortwave radiation, air temperature	2.0° × 1.5°	Ref [23]
				NorESM1-F	× 4 CO_2		elevation/surface altitude, precipitation, surface downwelling shortwave radiation, air temperature	2.5° × 1.9°	Ref [25]
Paleoclimat Modeling Intercomparison Project (PMIP)[26]	Phase 4 (PMIP4)	midPliocene-eoi400	Mid-Pliocene (3.3–3.0 Ma)	CESM2	--	r1i1p1f1	elevation/surface altitude, precipitation, surface downwelling shortwave radiation, air temperature	100 km × 100 km	Ref [27]
				GISS-E2-1-G			precipitation, surface downwelling shortwave radiation, air temperature	250 km × 250 km	Refs [28,29]
				HadGEM3-GC3 1-LL			precipitation, air temperature	250 km × 250 km	Ref [30]
				NorESM1-F			elevation/surface altitude, precipitation, surface downwelling	250 km × 250 km	Ref [25]

			shortwave radiation, air temperature		
lgm	Last Glacial Maximum (LGM, 22–18 ka)	AWI-ESM-1-1-LR	elevation/surface altitude, precipitation, surface downwelling shortwave radiation, air temperature	250 km × 250 km	Ref [31]
		MPI-ESM1-2-LR	elevation/surface altitude, precipitation, surface downwelling shortwave radiation, air temperature	250 km × 250 km	Ref [32]
midHolocene	Mid-Holocene (ca. 6 ka)	AWI-ESM-1-1-LR	elevation/surface altitude, precipitation, surface downwelling shortwave radiation, air temperature	250 km × 250 km	Ref [31]
		CESM2	elevation/surface altitude, precipitation, surface downwelling shortwave radiation, air temperature	100 km × 100 km	Ref [27]
		MPI-ESM1-2-LR	elevation/surface altitude, precipitation, surface downwelling shortwave radiation, air temperature	250 km × 250 km	Ref [32]
		MRI-ESM2-0	elevation/surface altitude, precipitation, surface downwelling	100 km × 100 km	Ref [33]

						shortwave radiation, air temperature		
				NESM3		precipitation, surface downwelling shortwave radiation, air temperature	250 km × 250 km	Ref [34]
				NorESM1-F		surface downwelling shortwave radiation	250 km × 250 km	Ref [25]
				NorESM2-LM		elevation/surface altitude, precipitation, air temperature	250 km × 250 km	Ref [35]
Coupled Model Intercomparison Project (CMIP)[36]	Phase 6 (CMIP6)	piControl	Pre-industrial era (before 1850)	ACCESS-CM2	--	elevation/surface altitude, precipitation, surface downwelling shortwave radiation, air temperature	250 km × 250 km	Ref [37]
				CESM2		elevation/surface altitude, precipitation, surface downwelling shortwave radiation, air temperature	100 km × 100 km	Ref [27]
				MIROC6		elevation/surface altitude, precipitation, surface downwelling shortwave radiation, air temperature	250 km × 250 km	Ref [38]
				MPI-ESM1-2-LR		elevation/surface altitude, precipitation, surface downwelling	250 km × 250 km	Ref [32]

			shortwave radiation, air temperature		
		MRI-ESM2-0	elevation/surface altitude, precipitation, surface downwelling shortwave radiation, air temperature	100 km × 100 km	Ref [33]
		NESM3	precipitation, surface downwelling shortwave radiation, air temperature	250 km × 250 km	Ref [34]
		NorESM2-LM	elevation/surface altitude, precipitation, surface downwelling shortwave radiation, air temperature	250 km × 250 km	Ref [35]
historical	Recent period (1901–1989)	ACCESS-CM2	surface downwelling shortwave radiation	250 km × 250 km	Ref [37]
		CESM2	surface downwelling shortwave radiation	100 km × 100 km	Ref [27]
		MIROC6	surface downwelling shortwave radiation	250 km × 250 km	Ref [38]
		MPI-ESM1-2-LR	surface downwelling shortwave radiation	250 km × 250 km	Ref [32]
		MRI-ESM2-0	surface downwelling shortwave	100 km ×	Ref [33]

						radiation	100 km	
			NESM3			surface downwelling shortwave radiation	250 km × 250 km	Ref [34]
			NorESM2-LM			surface downwelling shortwave radiation	250 km × 250 km	Ref [35]
Climatic Research Unit (CRU) Time-Series (TS)[39]	version 4.09	--	Recent period (1901–1989)	Observation-based gridded dataset	--	--	precipitation, air temperature	0.5° × 0.5°	Ref [39]
WorldClim[40]	version 2.1						elevation/surface altitude	30 arc-second grid	Ref [40]

Table A2.3 | Summary of Generalised Additive Models (GAMs) incorporating bioclimatic variables and the geographic extent of climate during the Eocene (*ca.* 50 Ma) across multiple climatic space resolutions. Summary statistics of GAMs fitted with species richness as the response variable. Explanatory variables include recent (1901–1989) temperature index (T_m, °C) and natural-log-transformed moisture index (ln MI, unitless), and total land area occupied by each climatic condition (Area, km^2) during the Eocene. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.43	1826164.57	< 0.001				
ln MI	0.29	2621414.09	< 0.001	0.92	0.91	0.90	405
Area	0.28	110054.76	< 0.001				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.42	2528553.22	< 0.001				
ln MI	0.26	3152404.27	< 0.001	0.89	0.88	0.87	628
Area	0.26	153392.87	< 0.001				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.40	2905884.78	< 0.001				
ln MI	0.24	3663735.85	< 0.001	0.87	0.84	0.84	896
Area	0.23	194660.03	< 0.001				
60 equal interval resolution							

Predictor	Concurvity	χ^2	*p*-value	Deviance	R_{pred}^2	R_{adj}^2	N
T_m	0.42	3391694.10	< 0.001				
ln MI	0.23	3897622.26	< 0.001	0.84	0.80	0.79	1182
Area	0.21	159522.60	< 0.001				

Table A2.4 | Summary of Generalised Additive Models (GAMs) incorporating bioclimatic variables and the geographic extent of climate during the Mid-Pliocene (3.3–3.0 Ma) across multiple climatic space resolutions. Summary statistics of GAMs fitted with species richness as the response variable. Explanatory variables include recent (1901–1989) temperature index (T_m, °C) and natural-log-transformed moisture index (ln MI, unitless), and total land area occupied by each climatic condition (Area, km^2) during the Mid-Pliocene. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.61	1326565.81	< 0.001				
ln MI	0.41	1110697.66	< 0.001	0.91	0.89	0.89	405
Area	0.42	70793.87	< 0.001				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.60	1742019.22	< 0.001				
ln MI	0.37	1494216.90	< 0.001	0.89	0.85	0.85	628
Area	0.39	73973.80	< 0.001				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**

T_m	0.56	2276038.16	< 0.001				
ln MI	0.35	2033348.82	< 0.001	0.86	0.82	0.81	896
Area	0.31	89872.99	< 0.001				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.59	2722005.60	< 0.001				
ln MI	0.33	2466028.68	< 0.001	0.83	0.78	0.78	1182
Area	0.32	92430.90	< 0.001				

Table A2.5 | Summary of Generalised Additive Models (GAMs) incorporating bioclimatic variables and the geographic extent of climate during the Last Glacial Maximum (LGM, 22–18 ka) across multiple climatic space resolutions. Summary statistics of GAMs fitted with species richness as the response variable. Explanatory variables include recent (1901–1989) temperature index (T_m, °C) and natural-log-transformed moisture index (ln MI, unitless), and total land area occupied by each climatic condition (Area, km^2) during the LGM. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.25	1464573.86	< 0.001				
ln MI	0.32	701089.22	< 0.001	0.91	0.89	0.88	405
Area	0.29	43087.69	< 0.001				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.25	2092171.65	< 0.001	0.88	0.85	0.84	628

ln MI	0.28	1260857.35	< 0.001				
Area	0.29	26796.12	< 0.001				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.26	2525533.82	< 0.001				
ln MI	0.25	1610612.17	< 0.001	0.85	0.81	0.80	896
Area	0.26	20900.22	< 0.001				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.25	3001882.59	< 0.001				
ln MI	0.24	1830820.53	< 0.001	0.83	0.78	0.77	1182
Area	0.23	23292.10	< 0.001				

Table A2.6 | Summary of Generalised Additive Models (GAMs) incorporating bioclimatic variables and the geographic extent of climate during the Mid-Holocene (*ca*. 6 ka) across multiple climatic space resolutions. Summary statistics of GAMs fitted with species richness as the response variable. Explanatory variables include recent (1901–1989) temperature index (T_m, °C) and natural-log-transformed moisture index (ln MI, unitless), and total land area occupied by each climatic condition (Area, km^2) during the Mid-Holocene. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.40	1149298.76	< 0.001	0.91	0.91	0.90	405
ln MI	0.38	720379.97	< 0.001				

Area	0.28	91266.85	< 0.001				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.39	1542196.05	< 0.001				
ln MI	0.35	996440.98	< 0.001	0.88	0.85	0.84	628
Area	0.30	34971.45	< 0.001				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.39	1855885.52	< 0.001				
ln MI	0.33	1321145.55	< 0.001	0.86	0.82	0.81	896
Area	0.25	75771.48	< 0.001				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.41	2161235.79	< 0.001				
ln MI	0.30	1553946.16	< 0.001	0.83	0.78	0.78	1182
Area	0.23	46616.82	< 0.001				

Table A2.7 | Summary of Generalised Additive Models (GAMs) incorporating bioclimatic variables and the geographic extent of climate during the pre-industrial period (pre-1850) across multiple climatic space resolutions. Summary statistics of GAMs fitted with species richness as the response variable. Explanatory variables include recent (1901–1989) temperature index (T_m, °C) and natural-log-transformed moisture index (ln MI, unitless), and total land area occupied by each climatic condition (Area, km^2) during the pre-industrial period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.48	1240918.46	< 0.001				
ln MI	0.34	671161.66	< 0.001	0.91	0.90	0.89	405
Area	0.27	88409.00	< 0.001				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.48	1652612.87	< 0.001				
ln MI	0.33	971836.13	< 0.001	0.89	0.87	0.87	628
Area	0.23	153045.85	< 0.001				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.46	2047651.57	< 0.001				
ln MI	0.30	1270973.81	< 0.001	0.87	0.83	0.83	896
Area	0.21	154572.78	< 0.001				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.49	2470587.64	< 0.001				
ln MI	0.29	1595683.32	< 0.001	0.84	0.79	0.78	1182
Area	0.22	124551.12	< 0.001				

Table A2.8 | Summary of Generalised Additive Models (GAMs) incorporating bioclimatic variables and the geographic extent of climate during the recent period (1901–1989) across multiple climatic space resolutions. Summary statistics of GAMs fitted with species richness as the response variable. Explanatory variables include recent (1901–1989) temperature index (T_m, °C) and natural-log-transformed moisture index (ln MI, unitless), and total land area occupied by each climatic condition (Area, km^2) during the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model

include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.48	1252944.99	< 0.001				
ln MI	0.21	625065.58	< 0.001	0.93	0.94	0.94	405
Area	0.24	250319.58	< 0.001				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.43	1745430.59	< 0.001				
ln MI	0.18	877454.57	< 0.001	0.92	0.91	0.91	628
Area	0.27	352504.99	< 0.001				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.44	2073100.64	< 0.001				
ln MI	0.16	1046454.26	< 0.001	0.90	0.90	0.90	896
Area	0.24	562136.56	< 0.001				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.44	2360696.24	< 0.001				
ln MI	0.15	1236657.04	< 0.001	0.89	0.88	0.88	1182
Area	0.22	752415.89	< 0.001				

Table A2.9 | Summary of Generalised Additive Models (GAMs) incorporating bioclimatic variables and the average geographic extent of climate across all selected geological periods at multiple climatic space resolutions. Summary statistics of GAMs fitted with species richness as the response variable. Explanatory variables include recent (1901–1989) temperature

index (T_m, °C) and natural-log-transformed moisture index (ln MI, unitless), and the mean total land area occupied by each climatic condition since its first appearance across all selected climatic periods (Area, km^2). Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.56	1095137.37	< 0.001				
ln MI	0.26	540113.96	< 0.001	0.90	0.88	0.87	405
Area	0.38	21689.69	< 0.001				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.54	1441724.54	< 0.001				
ln MI	0.23	665310.71	< 0.001	0.88	0.85	0.84	628
Area	0.38	29428.30	< 0.001				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.51	1593777.78	< 0.001				
ln MI	0.21	821338.15	< 0.001	0.86	0.82	0.81	896
Area	0.38	86385.99	< 0.001				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.54	1830603.38	< 0.001				
ln MI	0.20	903978.79	< 0.001	0.84	0.79	0.78	1182
Area	0.35	115607.59	< 0.001				

Table A2.10 | Relative importance of explanatory variables in Generalised Additive Models (GAMs) incorporating bioclimatic variables and the geographic extent of climate. Summary of the relative importance of each explanatory variable in GAMs fitted across different geological periods and multiple climatic space resolutions. Climatic space is defined by the recent (1901–1989) climatic conditions (T_m and ln MI), partitioned into 30, 40, 50 or 60 equal intervals along each axis. Species richness is the response variable. Explanatory variables include recent (1901–1989) temperature index (T_m, °C) and natural-log-transformed moisture index (ln MI, unitless), and total land area occupied by each climatic condition (Area, km^2) during different geological periods. The method for calculating the relative importance of each explanatory variable is described in the Methods section of the main text.

30 equal interval resolution				
Geological period	**T_m**	**ln MI**	**Area**	**Unexplained**
Eocene (*ca.* 50 Ma)	33.48%	42.52%	15.69%	8.31%
Mid-Pliocene (3.3–3.0 Ma)	27.97%	30.72%	32.43%	8.88%
Last Glacial Maximum (LGM, 22–18 ka)	32.73%	28.20%	29.85%	9.22%
Mid-Holocene (*ca.* 6 ka)	27.94%	28.03%	35.39%	8.64%
Pre-industrial (pre-1850)	28.41%	27.77%	35.16%	8.67%
Recent period (1901–1989)	28.60%	27.34%	37.46%	6.60%
Mean	27.15%	27.03%	36.31%	9.51%
40 equal interval resolution				
Geological period	**T_m**	**ln MI**	**Area**	**Unexplained**
Eocene	34.76%	39.48%	15.18%	10.58%

(*ca.* 50 Ma)				
Mid-Pliocene (3.3–3.0 Ma)	28.15%	28.90%	31.59%	11.37%
Last Glacial Maximum (LGM, 22–18 ka)	34.46%	28.02%	25.66%	11.87%
Mid-Holocene (*ca.* 6 ka)	28.77%	26.54%	32.91%	11.79%
Pre-industrial (pre-1850)	29.04%	26.40%	33.84%	10.72%
Recent period (1901–1989)	30.01%	25.84%	35.67%	8.48%
Mean	27.96%	25.14%	35.04%	11.85%
50 equal interval resolution				
Geological period	**T_m**	**In MI**	**Area**	**Unexplained**
Eocene (*ca.* 50 Ma)	33.23%	38.20%	15.55%	13.02%
Mid-Pliocene (3.3–3.0 Ma)	28.23%	28.95%	28.93%	13.89%
Last Glacial Maximum (LGM, 22–18 ka)	34.00%	27.65%	23.84%	14.51%
Mid-Holocene (*ca.* 6 ka)	28.32%	26.29%	31.33%	14.06%
Pre-industrial (pre-1850)	28.95%	26.10%	31.50%	13.45%
Recent period (1901–1989)	29.42%	25.11%	35.60%	9.88%
Mean	26.88%	24.45%	34.71%	13.96%
60 equal interval resolution				

Geological period	T_m	ln MI	Area	Unexplained
Eocene (*ca.* 50 Ma)	33.67%	35.90%	14.40%	16.02%
Mid-Pliocene (3.3–3.0 Ma)	28.77%	28.33%	26.38%	16.53%
Last Glacial Maximum (LGM, 22–18 ka)	34.59%	26.37%	21.98%	17.06%
Mid-Holocene (*ca.* 6 ka)	28.45%	25.14%	29.52%	16.89%
Pre-industrial (pre-1850)	29.64%	25.23%	28.79%	16.34%
Recent period (1901–1989)	29.51%	23.96%	35.19%	11.34%
Mean	27.12%	22.95%	33.56%	16.37%

Table A2.11 | Summary of Generalised Linear Models (GLMs) incorporating bioclimatic variables and the geographic extent of climate. Summary of both linear GLMs and quadratic GLMs incorporating second-order polynomial terms to account for non-linear relationships, developed for each geological period. All models were constructed based on a climatic space defined by the recent (1901–1989) climatic conditions (T_m and ln MI), partitioned into 30 equal intervals along each axis. Species richness is the response variable. Explanatory variables include recent (1901–1989) temperature index (T_m, °C) and natural-log-transformed moisture index (ln MI, unitless), and total land area occupied by each climatic condition (Area, km^2) during different geological periods. The table reports key model statistics, including deviance, McFadden's R^2, Akaike information criterion (AIC) and sample size (N) for each model.

Geological period	Model	Deviance	McFadden's R^2	AIC	N
Eocene (ca. 50 Ma)	Linear	4557951.76	0.42	4562008.83	405
	Quadratic	3350951.63	0.57	3355014.69	

Mid-Pliocene (3.3–3.0 Ma)	Linear	3794049.22	0.52	3798106.28	405
	Quadratic	2330398.35	0.70	2334461.41	
Last Glacial Maximum (LGM, 22–18 ka)	Linear	3044519.46	0.61	3048576.53	405
	Quadratic	1948587.22	0.75	1952650.29	
Mid-Holocene (ca. 6 ka)	Linear	3165583.22	0.60	3169640.28	405
	Quadratic	1781927.91	0.77	1785990.98	
Pre-industrial (pre-1850)	Linear	3455835.36	0.56	3459892.42	405
	Quadratic	1751546.27	0.78	1755609.34	
Recent period (1901–1989)	Linear	3466057.64	0.56	3470114.70	405
	Quadratic	1414545.75	0.82	1418608.81	
Mean	Linear	2836269.57	0.64	2840326.64	405
	Quadratic	1427989.61	0.82	1432052.68	

Table A2.12 | Coefficients of explanatory variables in Generalised Linear Models (GLMs) incorporating bioclimatic variables and the geographic extent of climate. Summary of the coefficients (Estimate and *p*-value) of explanatory variables in both linear GLMs and quadratic GLMs incorporating second-order polynomial terms to account for non-linear relationships, developed for each geological period. All models were constructed based on a climatic space defined by the recent (1901–1989) climatic conditions (T_m and ln MI), partitioned into 30 equal intervals along each axis. Species richness is the response variable. Explanatory variables include recent (1901–1989) temperature index (T_m, °C) and natural-log-transformed moisture index (ln MI, unitless), and total land area occupied by each climatic condition (Area, km^2) during different geological periods.

Geological period	Model	Coefficient	T_m	T_m^2	ln MI	ln MI^2	Area	Area^2
Eocene (ca. 50 Ma)	Linear	Estimate	5.54×10^{-2}	– –	2.28×10^{-1}	– –	1.55×10^{-7}	– –
		p-value	< 0.001	– –	< 0.001	– –	< 0.001	– –
	Quadratic	Estimate	1.26×10^{-1}	-3.26×10^{-3}	7.20×10^{-2}	-4.57×10^{-2}	8.98×10^{-7}	-4.04×10^{-13}
		p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Mid-Pliocene (3.3–3.0 Ma)	Linear	Estimate	4.07×10^{-2}	– –	1.72×10^{-1}	– –	6.82×10^{-7}	– –
		p-value	< 0.001	– –	< 0.001	– –	< 0.001	– –
	Quadratic	Estimate	1.10×10^{-1}	-3.27×10^{-3}	9.18×10^{-2}	-1.12×10^{-2}	2.11×10^{-6}	-7.46×10^{-13}
		p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Last Glacial Maximum	Linear	Estimate	4.92×10^{-2}	– –	1.83×10^{-1}	– –	6.22×10^{-7}	– –

(LGM, 22–18 ka)		*p*-value	< 0.001	– –	< 0.001	– –	< 0.001	– –
	Quadratic	Estimate	9.80×10^{-2}	-2.17×10^{-3}	1.61×10^{-1}	-4.08×10^{-3}	1.63×10^{-6}	-3.91×10^{-13}
		p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Mid-Holocene (ca. 6 ka)	Linear	Estimate	4.28×10^{-2}	– –	1.68×10^{-1}	– –	5.73×10^{-7}	– –
		p-value	< 0.001	– –	< 0.001	– –	< 0.001	– –
	Quadratic	Estimate	1.11×10^{-1}	-3.22×10^{-3}	1.26×10^{-1}	-5.27×10^{-3}	1.46×10^{-6}	-2.66×10^{-13}
		p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Pre-industrial (pre-1850)	Linear	Estimate	4.33×10^{-2}	– –	1.75×10^{-1}	– –	5.07×10^{-7}	– –
		p-value	< 0.001	– –	< 0.001	– –	< 0.001	– –
	Quadratic	Estimate	1.26×10^{-1}	-4.04×10^{-3}	1.10×10^{-1}	-6.66×10^{-3}	1.50×10^{-6}	-2.43×10^{-13}
		p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Recent period (1901–1989)	Linear	Estimate	4.34×10^{-2}	– –	1.88×10^{-1}	– –	5.37×10^{-7}	– –
		p-value	< 0.001	– –	< 0.001	– –	< 0.001	– –
	Quadratic	Estimate	1.32×10^{-1}	-4.21×10^{-3}	1.88×10^{-1}	4.04×10^{-4}	1.88×10^{-6}	-3.59×10^{-13}
		p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Mean	Linear	Estimate	3.72×10^{-2}	– –	1.39×10^{-1}	– –	9.43×10^{-7}	– –
		p-value	< 0.001	– –	< 0.001	– –	< 0.001	– –
	Quadratic	Estimate	1.06×10^{-1}	-3.28×10^{-3}	3.49×10^{-2}	-1.43×10^{-2}	2.26×10^{-6}	-5.26×10^{-13}
		p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Table A2.13 | Summary of the Generalised Additive Models (GAMs) with species richness as the response variable and climate and its deep-time geographic dynamics as explanatory variables for all recent climatic conditions across multiple climatic space resolutions. Summary statistics of GAMs fitted with species richness as the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence to the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.68	485359.95	< 0.001				
ln MI	0.57	243133.72	< 0.001				
Recent Area	0.84	106864.19	< 0.001	0.95	0.95	0.95	342
ΔArea	0.62	35550.61	< 0.001				
Movement	0.42	41212.21	< 0.001				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.63	656361.87	< 0.001				
ln MI	0.53	341180.83	< 0.001				
Recent Area	0.76	213932.33	< 0.001	0.93	0.93	0.92	529
ΔArea	0.55	37366.92	< 0.001				

Movement	0.43	28029.86	< 0.001				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.68	824598.86	< 0.001				
ln MI	0.50	465665.46	< 0.001				
Recent Area	0.76	325886.77	< 0.001	0.91	0.91	0.91	751
ΔArea	0.55	20190.12	< 0.001				
Movement	0.40	55998.89	< 0.001				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.61	885554.69	< 0.001				
ln MI	0.48	591684.02	< 0.001				
Recent Area	0.77	381718.80	< 0.001	0.90	0.90	0.89	986
ΔArea	0.55	40336.41	< 0.001				
Movement	0.39	90781.19	< 0.001				

Table A2.14 | Summary of the Generalised Additive Models (GAMs) with species richness as the response variable and climate and its deep-time geographic dynamics as explanatory variables for recent climatic conditions that first appeared in the Eocene (*ca.* 50 Ma), across multiple climatic space resolutions. Summary statistics of GAMs fitted with species richness as the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence (i.e. Eocene) to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence (i.e. Eocene) to the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each

explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.69	208593.97	< 0.001				
ln MI	0.82	161160.60	< 0.001				
Recent Area	0.77	29935.28	< 0.001	0.92	0.92	0.90	88
ΔArea	0.79	15949.11	< 0.001				
Movement	0.47	37054.35	< 0.001				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.75	156233.82	< 0.001				
ln MI	0.82	146852.11	< 0.001				
Recent Area	0.87	85445.79	< 0.001	0.95	0.96	0.94	137
ΔArea	0.72	13256.22	< 0.001				
Movement	0.66	16157.92	< 0.001				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.70	228305.67	< 0.001				
ln MI	0.80	144820.57	< 0.001				
Recent Area	0.83	111722.55	< 0.001	0.94	0.95	0.93	195
ΔArea	0.70	19006.82	< 0.001				
Movement	0.59	16238.15	< 0.001				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.66	245215.03	< 0.001	0.93	0.93	0.91	258
ln MI	0.77	219098.10	< 0.001				

Recent Area	0.85	132211.70	< 0.001
ΔArea	0.72	14352.42	< 0.001
Movement	0.45	30592.89	< 0.001

Table A2.15 | Summary of the Generalised Additive Models (GAMs) with species richness as the response variable and climate and its deep-time geographic dynamics as explanatory variables for recent climatic conditions that first appeared in the Mid-Pliocene (3.3–3.0 Ma), across multiple climatic space resolutions. Summary statistics of GAMs fitted with species richness as the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence (i.e. Mid-Pliocene) to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence (i.e. Mid-Pliocene) to the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.74	124085.91	< 0.001				
ln MI	0.69	80856.61	< 0.001				
Recent Area	0.56	109540.94	< 0.001	0.94	0.94	0.93	140
ΔArea	0.22	497.03	< 0.001				
Movement	0.48	9512.98	< 0.001				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.78	171294.45	< 0.001	0.94	0.95	0.93	212

ln MI	0.66	127666.29	< 0.001				
Recent Area	0.81	32466.22	< 0.001				
ΔArea	0.60	10527.97	< 0.001				
Movement	0.45	15544.55	< 0.001				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.74	248967.86	< 0.001				
ln MI	0.64	180867.76	< 0.001				
Recent Area	0.71	57224.37	< 0.001	0.93	0.93	0.92	307
ΔArea	0.52	20857.56	< 0.001				
Movement	0.42	26931.11	< 0.001				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.73	229332.72	< 0.001				
ln MI	0.64	227880.25	< 0.001				
Recent Area	0.84	69952.47	< 0.001	0.90	0.89	0.88	411
ΔArea	0.63	22303.51	< 0.001				
Movement	0.40	32335.03	< 0.001				

Table A2.16 | Summary of the Generalised Additive Models (GAMs) with species richness as the response variable and climate and its deep-time geographic dynamics as explanatory variables for recent climatic conditions that first appeared in the Last Glacial Maximum (LGM, 22–18 ka), across multiple climatic space resolutions. Summary statistics of GAMs fitted with species richness as the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence (i.e. LGM) to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence (i.e. LGM) to the recent period. Different GAMs were constructed based on climatic

spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.87	29297.00	< 0.001				
ln MI	0.86	10462.70	< 0.001				
Recent Area	0.85	4936.89	< 0.001	0.86	0.79	0.74	96
ΔArea	0.53	435.82	< 0.001				
Movement	0.54	6219.60	< 0.001				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.90	33161.67	< 0.001				
ln MI	0.86	7298.11	< 0.001				
Recent Area	0.88	9135.40	< 0.001	0.93	0.94	0.91	146
ΔArea	0.61	2011.42	< 0.001				
Movement	0.61	3888.90	< 0.001				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.86	43972.48	< 0.001				
ln MI	0.80	6645.51	< 0.001				
Recent Area	0.86	8718.15	< 0.001	0.89	0.88	0.84	197
ΔArea	0.62	2742.06	< 0.001				
Movement	0.39	6989.94	< 0.001				
60 equal interval resolution							

Predictor	Concurvity	χ^2	*p*-value	Deviance	R_{pred}^2	R_{adj}^2	N
T_m	0.87	55468.33	< 0.001				
ln MI	0.77	9558.44	< 0.001				
Recent Area	0.86	14550.80	< 0.001	0.89	0.88	0.85	255
ΔArea	0.57	2113.34	< 0.001				
Movement	0.31	13362.34	< 0.001				

Table A2.17 | Relative importance of explanatory variables in Generalised Additive Models (GAMs) with species richness as the response variable and climate and its deep-time geographic dynamics as explanatory variables. Summary of the relative importance of each explanatory variable in GAMs fitted for all recent climatic conditions, as well as separately according to their first appearance. Species richness is the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence to the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. The method for calculating the relative importance of each explanatory variable is described in the Methods section of the main text.

30 equal interval resolution						
Geological period	**T_m**	**ln MI**	**Recent Area**	**ΔArea**	**Movement**	**Unexplained**
All periods	19.68%	19.19%	26.77%	9.34%	20.05%	4.97%
Eocene (*ca.* 50 Ma)	20.93%	28.68%	23.22%	9.76%	9.07%	8.33%
Mid-Pliocene (3.3–3.0 Ma)	18.15%	25.46%	27.45%	2.45%	20.21%	6.29%
Last Glacial Maximum	29.47%	24.21%	16.96%	11.09%	4.26%	14.01%

(LGM, 22–18 ka)

40 equal interval resolution						
Geological period	**T_m**	**In MI**	**Area**	**ΔArea**	**Movement**	**Unexplained**
All periods	20.24%	18.24%	25.90%	9.24%	19.25%	7.13%
Eocene (*ca.* 50 Ma)	23.53%	25.84%	25.23%	10.18%	10.57%	4.65%
Mid-Pliocene (3.3–3.0 Ma)	17.59%	25.68%	26.08%	6.47%	18.62%	5.56%
Last Glacial Maximum (LGM, 22–18 ka)	28.99%	21.79%	22.54%	14.51%	5.17%	7.00%
50 equal interval resolution						
Geological period	**T_m**	**In MI**	**Area**	**ΔArea**	**Movement**	**Unexplained**
All periods	19.80%	17.71%	25.36%	9.86%	18.75%	8.52%
Eocene (*ca.* 50 Ma)	21.89%	24.49%	26.06%	11.10%	10.51%	5.94%
Mid-Pliocene (3.3–3.0 Ma)	18.11%	24.24%	22.87%	9.50%	18.66%	6.62%
Last Glacial Maximum (LGM, 22–18 ka)	30.20%	21.19%	19.61%	13.88%	4.25%	10.86%
60 equal interval resolution						
Geological period	**T_m**	**In MI**	**Area**	**ΔArea**	**Movement**	**Unexplained**
All periods	19.85%	16.96%	24.59%	9.85%	19.12%	9.63%
Eocene (*ca.* 50 Ma)	22.60%	24.90%	24.78%	10.14%	10.14%	7.44%
Mid-Pliocene (3.3–3.0 Ma)	15.84%	22.54%	23.56%	9.80%	17.84%	10.44%
Last Glacial Maximum (LGM, 22–18 ka)	28.87%	20.58%	19.49%	13.07%	6.84%	11.15%

Table A2.18 | Summary of Generalised Linear Models (GLMs) with species richness as the response variable and climate and its deep-time geographic dynamics as explanatory variables. Summary of both linear GLMs and quadratic GLMs incorporating second-order polynomial terms to account for non-linear relationships, developed for all recent climatic conditions, as well as separately according to their first appearance. Species richness is the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence to the recent period. Different GLMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. The table reports key model statistics, including deviance, McFadden's R^2, Akaike information criterion (AIC) and sample size (N) for each model.

Geological period	Model	Deviance	McFadden's R^2	AIC	N
All periods	Linear	2865670.43	0.58	2869198.44	342
	Quadratic	1051793.47	0.85	1055331.47	
Eocene (ca. 50 Ma)	Linear	1196024.18	0.43	1197042.33	88
	Quadratic	355709.28	0.83	356737.43	
Mid-Pliocene (3.3–3.0 Ma)	Linear	669166.38	0.65	670670.58	140
	Quadratic	126914.60	0.93	128428.79	
Last Glacial Maximum (LGM, 22–18 ka)	Linear	147733.41	0.48	148620.13	96
	Quadratic	64219.10	0.77	65115.82	

Table A2.19 | Coefficients of explanatory variables in Generalised Linear Models (GLMs) with species richness as the response variable and climate and its deep-time geographic dynamics as explanatory variables. Summary of the coefficients (Estimate and *p*-value) of explanatory variables in both linear GLMs and quadratic GLMs incorporating second-order polynomial terms to account for non-linear relationships, developed for all recent climatic conditions, as well as separately according to their first appearance. Species richness is the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence to the recent period. Different GLMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals.

Geological period	Model	Coefficient	T_m	T_m^2	ln MI	ln MI^2	Recent Area	Recent Area^2	ΔArea	ΔArea^2	Movement	Movement^2
All periods	Linear	Estimate	3.24×10^{-2}	– –	1.78×10^{-1}	– –	5.21×10^{-7}	– –	-4.11×10^{-8}	– –	8.70×10^{-5}	– –
		p-value	< 0.001	– –	< 0.001	– –	< 0.001	– –	< 0.001	– –	< 0.001	– –
	Quadratic	Estimate	1.27×10^{-1}	-3.94×10^{-3}	1.22×10^{-1}	-4.05×10^{-2}	1.51×10^{-6}	-2.81×10^{-13}	4.03×10^{-8}	-3.64×10^{-15}	2.05×10^{-4}	-1.35×10^{-8}
		p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Eocene (ca. 50 Ma)	Linear	Estimate	3.55×10^{-3}	– –	9.83×10^{-2}	– –	3.88×10^{-7}	– –	1.05×10^{-7}	– –	8.71×10^{-5}	– –
		p-value	< 0.001	– –	< 0.001	– –	< 0.001	– –	< 0.001	– –	< 0.001	– –

	Quadratic	Estimate	3.94×10^{-1}	-9.90×10^{-3}	1.77×10^{-1}	-3.11×10^{-1}	7.32×10^{-7}	-1.33×10^{-13}	1.50×10^{-7}	-1.83×10^{-14}	-2.11×10^{-4}	5.30×10^{-9}
		p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Mid-Pliocene (3.3–3.0 Ma)	Linear	Estimate	3.66×10^{-2}	– –	1.87×10^{-1}	– –	1.54×10^{-6}	– –	-1.62×10^{-7}	– –	1.20×10^{-4}	– –
		p-value	< 0.001	– –	< 0.001	– –	< 0.001	– –	< 0.001	– –	< 0.001	– –
	Quadratic	Estimate	1.43×10^{-1}	-4.23×10^{-3}	1.78×10^{-1}	-3.00×10^{-2}	3.87×10^{-6}	-2.16×10^{-12}	2.53×10^{-8}	-3.59×10^{-15}	2.20×10^{-4}	-1.42×10^{-8}
		p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.14	< 0.001	< 0.001
Last Glacial Maximum (LGM, 22–18 ka)	Linear	Estimate	7.04×10^{-2}	– –	1.07×10^{-1}	– –	1.19×10^{-6}	– –	2.25×10^{-7}	– –	-5.49×10^{-5}	– –
		p-value	< 0.001	– –	< 0.001	– –	< 0.001	– –	< 0.001	– –	< 0.001	– –
	Quadratic	Estimate	6.88×10^{-2}	-1.16×10^{-3}	-3.17×10^{-1}	-1.08×10^{-1}	3.86×10^{-6}	-2.54×10^{-12}	-9.89×10^{-9}	-5.62×10^{-14}	5.59×10^{-5}	-5.16×10^{-9}
		p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.38	< 0.001	< 0.001	< 0.001

Table A2.20 | Summary of the Generalised Additive Models (GAMs) with species turnover as the response variable and climate and its deep-time geographic dynamics as explanatory variables for all recent climatic conditions across multiple climatic space resolutions. Summary statistics of GAMs fitted with turnover component of β-diversity as the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence to the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.65	63.00	< 0.001				
ln MI	0.54	7.31	0.11				
Recent Area	0.84	122.41	< 0.001	0.74	0.77	0.75	297
ΔArea	0.63	2.28	0.13				
Movement	0.46	61.57	< 0.001				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.66	114.00	< 0.001				
ln MI	0.51	17.04	< 0.05	0.72	0.75	0.73	464
Recent Area	0.76	159.96	< 0.001				
ΔArea	0.55	7.96	< 0.001				

Movement	0.49	93.95	< 0.001				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.67	107.91	< 0.001				
ln MI	0.55	18.51	< 0.01				
Recent Area	0.76	119.50	< 0.001	0.64	0.65	0.63	638
ΔArea	0.58	14.60	< 0.001				
Movement	0.49	58.31	< 0.001				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.66	127.13	< 0.001				
ln MI	0.53	40.57	< 0.001				
Recent Area	0.77	94.97	< 0.001	0.57	0.63	0.62	830
ΔArea	0.57	2.57	0.14				
Movement	0.46	70.44	< 0.001				

Table A2.21 | Summary of the Generalised Additive Models (GAMs) with species turnover as the response variable and climate and its deep-time geographic dynamics as explanatory variables for recent climatic conditions that first appeared in the Eocene (*ca.* 50 Ma), across multiple climatic space resolutions. Summary statistics of GAMs fitted with turnover component of β-diversity as the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence (i.e. Eocene) to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence (i.e. Eocene) to the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2

and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.78	14.00	< 0.05				
ln MI	0.85	15.94	< 0.001				
Recent Area	0.91	61.21	< 0.001	0.89	0.86	0.82	82
ΔArea	0.80	0.63	0.43				
Movement	0.73	19.84	< 0.001				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.77	81.62	< 0.001				
ln MI	0.86	9.20	< 0.05				
Recent Area	0.87	49.77	< 0.001	0.77	0.76	0.71	127
ΔArea	0.75	1.83	0.31				
Movement	0.74	18.76	< 0.001				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.72	22.73	< 0.001				
ln MI	0.82	1.43	0.23				
Recent Area	0.82	45.56	< 0.001	0.52	0.60	0.54	184
ΔArea	0.71	3.20	0.07				
Movement	0.65	9.09	< 0.001				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.72	63.45	< 0.001				
ln MI	0.79	6.74	0.13	0.60	0.66	0.61	236
Recent Area	0.85	21.21	< 0.001				

ΔArea	0.73	0.05	0.82				
Movement	0.54	43.59	< 0.001				

Table A2.22 | Summary of the Generalised Additive Models (GAMs) with species turnover as the response variable and climate and its deep-time geographic dynamics as explanatory variables for recent climatic conditions that first appeared in the Mid-Pliocene (3.3–3.0 Ma), across multiple climatic space resolutions. Summary statistics of GAMs fitted with turnover component of β-diversity as the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence (i.e. Mid-Pliocene) to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence (i.e. Mid-Pliocene) to the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.80	10.01	< 0.05				
ln MI	0.75	12.77	< 0.01				
Recent Area	0.82	46.59	< 0.001	0.76	0.77	0.74	128
ΔArea	0.63	0.89	0.66				
Movement	0.52	33.66	< 0.001				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.80	47.19	< 0.001	0.85	0.84	0.82	198

ln MI	0.68	71.69	< 0.001				
Recent Area	0.82	15.10	< 0.001				
ΔArea	0.65	4.91	0.10				
Movement	0.49	68.29	< 0.001				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.76	2.36	0.12				
ln MI	0.67	27.29	< 0.001				
Recent Area	0.74	40.08	< 0.001	0.61	0.63	0.60	267
ΔArea	0.60	1.12	0.29				
Movement	0.45	68.67	< 0.001				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.75	0.56	0.54				
ln MI	0.67	54.01	< 0.001				
Recent Area	0.85	32.17	< 0.001	0.59	0.60	0.58	368
ΔArea	0.67	0.92	0.34				
Movement	0.42	115.65	< 0.001				

Table A2.23 | Summary of the Generalised Additive Models (GAMs) with species turnover as the response variable and climate and its deep-time geographic dynamics as explanatory variables for recent climatic conditions that first appeared in the Last Glacial Maximum (LGM, 22–18 ka), across multiple climatic space resolutions. Summary statistics of GAMs fitted with turnover component of β-diversity as the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence (i.e. LGM) to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence (i.e. LGM) to the recent period. Different GAMs were constructed based on

climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.88	10.47	0.07				
ln MI	0.85	22.07	< 0.001				
Recent Area	0.85	67.36	< 0.001	0.83	0.83	0.80	78
ΔArea	0.58	10.75	< 0.001				
Movement	0.60	17.49	< 0.001				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.91	14.54	< 0.05				
ln MI	0.84	10.54	< 0.05				
Recent Area	0.90	60.04	< 0.001	0.78	0.80	0.77	117
ΔArea	0.68	0.78	0.38				
Movement	0.56	6.53	< 0.05				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.89	32.01	< 0.001				
ln MI	0.79	19.57	< 0.001				
Recent Area	0.87	106.63	< 0.001	0.79	0.80	0.77	160
ΔArea	0.65	6.32	< 0.05				
Movement	0.38	25.18	< 0.001				
60 equal interval resolution							

Predictor	Concurvity	χ^2	*p*-value	Deviance	R_{pred}^2	R_{adj}^2	N
T_m	0.88	37.93	< 0.001				
ln MI	0.77	42.87	< 0.001				
Recent Area	0.87	58.25	< 0.001	0.81	0.78	0.75	194
ΔArea	0.63	5.23	< 0.05				
Movement	0.34	31.93	< 0.001				

Table A2.24 | Summary of the Generalised Additive Models (GAMs) with phylogenetic diversity as the response variable and climate and its deep-time geographic dynamics as explanatory variables for all recent climatic conditions across multiple climatic space resolutions. Summary statistics of GAMs fitted with Faith's phylogenetic diversity[13] as the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence to the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.64	197.55	< 0.001				
ln MI	0.56	139.60	< 0.001				
Recent Area	0.84	48.21	< 0.001	0.85	0.85	0.84	333
ΔArea	0.62	41.54	< 0.001				
Movement	0.41	37.37	< 0.001				

40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.61	325.36	< 0.001				
ln MI	0.53	237.09	< 0.001				
Recent Area	0.76	55.06	< 0.001	0.83	0.83	0.82	516
ΔArea	0.55	41.80	< 0.001				
Movement	0.42	40.35	< 0.001				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.67	382.79	< 0.001				
ln MI	0.49	247.36	< 0.001				
Recent Area	0.76	86.85	< 0.001	0.82	0.82	0.81	734
ΔArea	0.55	40.24	< 0.001				
Movement	0.40	57.51	< 0.001				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.63	472.51	< 0.001				
ln MI	0.47	340.21	< 0.001				
Recent Area	0.77	133.54	< 0.001	0.81	0.81	0.81	968
ΔArea	0.56	54.97	< 0.001				
Movement	0.39	102.72	< 0.001				

Table A2.25 | Summary of the Generalised Additive Models (GAMs) with phylogenetic diversity as the response variable and climate and its deep-time geographic dynamics as explanatory variables for recent climatic conditions that first appeared in the Eocene (*ca.* 50 Ma), across multiple climatic space resolutions. Summary statistics of GAMs fitted with Faith's phylogenetic diversity[13] as the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2),

change in total land area occupied by a climatic condition from its earliest occurrence (i.e. Eocene) to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence (i.e. Eocene) to the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.76	92.73	< 0.001				
ln MI	0.86	130.50	< 0.001				
Recent Area	0.91	10.64	< 0.05	0.94	0.94	0.92	87
ΔArea	0.79	10.77	< 0.05				
Movement	0.66	0.10	0.88				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.75	89.69	< 0.001				
ln MI	0.82	114.57	< 0.001				
Recent Area	0.87	60.67	< 0.001	0.90	0.90	0.88	136
ΔArea	0.72	0.20	0.65				
Movement	0.66	1.01	0.40				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.70	99.30	< 0.001				
ln MI	0.80	130.64	< 0.001	0.89	0.89	0.87	194
Recent Area	0.83	83.61	< 0.001				
ΔArea	0.70	1.58	0.21				

Movement	0.59	3.99	0.09				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.66	127.88	< 0.001				
ln MI	0.77	166.98	< 0.001				
Recent Area	0.85	95.08	< 0.001	0.87	0.87	0.86	257
ΔArea	0.72	3.04	0.08				
Movement	0.46	5.36	0.22				

Table A2.26 | Summary of the Generalised Additive Models (GAMs) with phylogenetic diversity as the response variable and climate and its deep-time geographic dynamics as explanatory variables for recent climatic conditions that first appeared in the Mid-Pliocene (3.3–3.0 Ma), across multiple climatic space resolutions. Summary statistics of GAMs fitted with Faith's phylogenetic diversity[13] as the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence (i.e. Mid-Pliocene) to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence (i.e. Mid-Pliocene) to the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.74	73.23	< 0.001	0.83	0.83	0.81	140
ln MI	0.69	53.54	< 0.001				

Recent Area	0.56	119.42	< 0.001				
ΔArea	0.22	4.61	0.13				
Movement	0.48	8.07	< 0.05				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.78	253.68	< 0.001				
ln MI	0.65	223.38	< 0.001				
Recent Area	0.81	48.73	< 0.001	0.88	0.88	0.86	210
ΔArea	0.61	2.53	0.40				
Movement	0.45	0.82	0.37				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.74	392.41	< 0.001				
ln MI	0.64	320.08	< 0.001				
Recent Area	0.71	31.27	< 0.001	0.85	0.85	0.84	305
ΔArea	0.52	28.10	< 0.001				
Movement	0.42	3.24	0.24				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.73	266.27	< 0.001				
ln MI	0.63	236.90	< 0.001				
Recent Area	0.84	69.56	< 0.001	0.81	0.81	0.80	407
ΔArea	0.64	13.99	< 0.01				
Movement	0.39	8.88	< 0.05				

Table A2.27 | Summary of the Generalised Additive Models (GAMs) with phylogenetic diversity as the response variable and climate and its deep-time geographic dynamics as explanatory variables for recent climatic conditions that first appeared in the Last Glacial

Maximum (LGM, 22–18 ka), across multiple climatic space resolutions. Summary statistics of GAMs fitted with Faith's phylogenetic diversity[13] as the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence (i.e. LGM) to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence (i.e. LGM) to the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.91	60.30	< 0.001				
ln MI	0.88	8.68	0.05				
Recent Area	0.95	9.26	< 0.05	0.72	0.72	0.66	90
ΔArea	0.72	0.00	0.99				
Movement	0.62	3.91	0.30				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.90	92.99	< 0.001				
ln MI	0.87	7.79	0.12				
Recent Area	0.89	31.31	< 0.001	0.77	0.77	0.73	137
ΔArea	0.65	0.10	0.75				
Movement	0.61	11.60	< 0.05				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**

T_m	0.87	119.01	< 0.001				
ln MI	0.79	8.16	0.09				
Recent Area	0.86	31.75	< 0.001	0.70	0.71	0.67	187
ΔArea	0.64	0.57	0.45				
Movement	0.39	10.83	< 0.05				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.87	181.31	< 0.001				
ln MI	0.76	10.83	< 0.05				
Recent Area	0.86	68.59	< 0.001	0.74	0.74	0.72	245
ΔArea	0.59	0.77	0.38				
Movement	0.31	38.22	< 0.001				

Table A2.28 | Summary of the Generalised Additive Models (GAMs) with relative phylogenetic diversity as the response variable and climate and its deep-time geographic dynamics as explanatory variables for all recent climatic conditions across multiple climatic space resolutions. Summary statistics of GAMs fitted with relative phylogenetic diversity (calculated as the residuals from regressing phylogenetic diversity against species richness) as the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence to the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.64	157.07	< 0.001				
ln MI	0.56	56.06	< 0.001				
Recent Area	0.84	15.17	< 0.001	0.53	0.53	0.49	333
ΔArea	0.62	2.94	0.45				
Movement	0.41	17.34	< 0.001				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.61	250.31	< 0.001				
ln MI	0.53	114.22	< 0.001				
Recent Area	0.76	19.89	< 0.001	0.56	0.56	0.53	516
ΔArea	0.55	12.51	< 0.05				
Movement	0.42	36.63	< 0.001				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.67	383.95	< 0.001				
ln MI	0.49	179.03	< 0.001				
Recent Area	0.76	64.01	< 0.001	0.55	0.55	0.53	734
ΔArea	0.55	12.45	< 0.05				
Movement	0.40	28.34	< 0.001				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.63	478.01	< 0.001				
ln MI	0.47	218.37	< 0.001				
Recent Area	0.77	53.55	< 0.001	0.52	0.52	0.50	968
ΔArea	0.56	19.31	< 0.01				
Movement	0.39	54.78	< 0.001				

Table A2.29 | Summary of the Generalised Additive Models (GAMs) with relative phylogenetic diversity as the response variable and climate and its deep-time geographic dynamics as explanatory variables for recent climatic conditions that first appeared in the Eocene (*ca.* 50 Ma), across multiple climatic space resolutions. Summary statistics of GAMs fitted with relative phylogenetic diversity (calculated as the residuals from regressing phylogenetic diversity against species richness) as the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence (i.e. Eocene) to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence (i.e. Eocene) to the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.69	192.95	< 0.001				
ln MI	0.82	118.19	< 0.001				
Recent Area	0.80	11.44	< 0.05	0.81	0.81	0.78	87
ΔArea	0.73	1.16	0.29				
Movement	0.53	7.65	< 0.05				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.75	221.22	< 0.001				
ln MI	0.82	221.46	< 0.001	0.83	0.83	0.80	136
Recent Area	0.87	8.93	0.06				

ΔArea	0.72	5.32	0.20				
Movement	0.66	3.72	0.06				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.70	318.68	< 0.001				
ln MI	0.80	232.70	< 0.001				
Recent Area	0.83	15.48	< 0.001	0.77	0.77	0.74	194
ΔArea	0.70	0.58	0.45				
Movement	0.59	2.71	0.24				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.66	252.31	< 0.001				
ln MI	0.77	157.38	< 0.001				
Recent Area	0.85	5.59	< 0.05	0.69	0.69	0.66	257
ΔArea	0.72	9.25	0.12				
Movement	0.46	1.98	0.35				

Table A2.30 | Summary of the Generalised Additive Models (GAMs) with relative phylogenetic diversity as the response variable and climate and its deep-time geographic dynamics as explanatory variables for recent climatic conditions that first appeared in the Mid-Pliocene (3.3–3.0 Ma), across multiple climatic space resolutions. Summary statistics of GAMs fitted with relative phylogenetic diversity (calculated as the residuals from regressing phylogenetic diversity against species richness) as the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence (i.e. Mid-Pliocene) to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence (i.e. Mid-Pliocene) to the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic

conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.79	54.18	< 0.001				
ln MI	0.75	69.90	< 0.001				
Recent Area	0.81	5.19	< 0.05	0.73	0.73	0.69	140
ΔArea	0.59	36.49	< 0.001				
Movement	0.51	1.17	0.28				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.78	120.94	< 0.001				
ln MI	0.65	105.55	< 0.001				
Recent Area	0.81	17.03	< 0.001	0.74	0.74	0.71	210
ΔArea	0.61	42.19	< 0.001				
Movement	0.45	15.69	< 0.001				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.74	160.61	< 0.001				
ln MI	0.64	103.56	< 0.001				
Recent Area	0.71	47.79	< 0.001	0.70	0.70	0.67	305
ΔArea	0.52	35.96	< 0.001				
Movement	0.42	16.54	< 0.001				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**

T_m	0.73	194.06	< 0.001				
ln MI	0.63	114.24	< 0.001				
Recent Area	0.84	34.78	< 0.001	0.65	0.65	0.62	407
ΔArea	0.64	52.50	< 0.001				
Movement	0.39	17.71	< 0.001				

Table A2.31 | Summary of the Generalised Additive Models (GAMs) with relative phylogenetic diversity as the response variable and climate and its deep-time geographic dynamics as explanatory variables for recent climatic conditions that first appeared in the Last Glacial Maximum (LGM, 22–18 ka), across multiple climatic space resolutions. Summary statistics of GAMs fitted with relative phylogenetic diversity (calculated as the residuals from regressing phylogenetic diversity against species richness) as the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence (i.e. LGM) to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence (i.e. LGM) to the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Reported metrics for each model include concurvity, χ^2 and *p*-value for each explanatory variable, and explained deviance (Deviance), predicted R^2 (R_{pred}^2) and adjusted R^2 (R_{adj}^2) of each GAM model, and the sample size (N) fitted into each model.

30 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.91	3.38	0.23				
ln MI	0.88	5.27	< 0.05	0.14	0.14	0.08	90
Recent Area	0.95	4.64	< 0.05				
ΔArea	0.72	0.01	0.94				

Movement	0.62	0.08	0.78				
40 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.90	40.20	< 0.001				
ln MI	0.87	35.50	< 0.001				
Recent Area	0.89	5.96	0.08	0.45	0.46	0.37	137
ΔArea	0.65	0.05	0.83				
Movement	0.61	6.64	0.14				
50 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.87	41.92	< 0.001				
ln MI	0.79	42.50	< 0.001				
Recent Area	0.86	2.06	0.15	0.33	0.34	0.28	187
ΔArea	0.64	0.01	0.92				
Movement	0.39	3.08	0.08				
60 equal interval resolution							
Predictor	**Concurvity**	**χ^2**	***p*-value**	**Deviance**	**R_{pred}^2**	**R_{adj}^2**	**N**
T_m	0.87	74.10	< 0.001				
ln MI	0.76	55.39	< 0.001				
Recent Area	0.86	17.00	< 0.001	0.39	0.39	0.34	245
ΔArea	0.59	4.51	0.35				
Movement	0.31	4.73	< 0.05				

Table A2.32 | Spatial auto-correlation test for Generalised Additive Models (GAMs) incorporating bioclimatic variables and the geographic extent of climate at multiple climatic space resolutions. Summary of spatial autocorrelation tests for the residuals of GAMs fitted across different geological periods. Species richness is the response variable. Explanatory variables include recent (1901–1989) temperature index (T_m, °C) and natural-log-transformed moisture index (ln MI, unitless), and total land area occupied by each climatic condition (Area, km^2) in different

geological periods. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Potential spatial autocorrelation was assessed using Moran's *I* test on the residuals of each GAM. Moran's *I* index and corresponding *p*-value of each GAM are presented in the table.

30 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
Eocene (*ca.* 50 Ma)	0.26	< 0.001
Mid-Pliocene (3.3–3.0 Ma)	0.29	< 0.001
Last Glacial Maximum (LGM, 22–18 ka)	0.33	< 0.001
Mid-Holocene (*ca.* 6 ka)	0.30	< 0.001
Pre-industrial (pre-1850)	0.28	< 0.001
Recent period (1901–1989)	0.32	< 0.001
Mean	0.30	< 0.001
40 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
Eocene (*ca.* 50 Ma)	0.33	< 0.001
Mid-Pliocene (3.3–3.0 Ma)	0.34	< 0.001
Last Glacial Maximum (LGM, 22–18 ka)	0.40	< 0.001

Mid-Holocene (*ca.* 6 ka)	0.40	< 0.001
Pre-industrial (pre-1850)	0.34	< 0.001
Recent period (1901–1989)	0.38	< 0.001
Mean	0.39	< 0.001
50 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
Eocene (*ca.* 50 Ma)	0.38	< 0.001
Mid-Pliocene (3.3–3.0 Ma)	0.39	< 0.001
Last Glacial Maximum (LGM, 22–18 ka)	0.43	< 0.001
Mid-Holocene (*ca.* 6 ka)	0.41	< 0.001
Pre-industrial (pre-1850)	0.36	< 0.001
Recent period (1901–1989)	0.40	< 0.001
Mean	0.40	< 0.001
60 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
Eocene (*ca.* 50 Ma)	0.36	< 0.001
Mid-Pliocene (3.3–3.0 Ma)	0.35	< 0.001

Last Glacial Maximum (LGM, 22–18 ka)	0.40	< 0.001
Mid-Holocene (*ca.* 6 ka)	0.39	< 0.001
Pre-industrial (pre-1850)	0.34	< 0.001
Recent period (1901–1989)	0.37	< 0.001
Mean	0.38	< 0.001

Table A2.33 | Spatial auto-correlation test for Generalised Additive Models (GAMs) with species richness as the response variable and climate and its deep-time geographic dynamics as explanatory variables. Summary of spatial autocorrelation tests for the residuals of GAMs fitted for all recent climatic conditions, as well as separately according to their first appearance. Species richness is the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence to the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Potential spatial autocorrelation was assessed using Moran's *I* test on the residuals of each GAM. Moran's *I* index and corresponding *p*-value of each GAM are presented in the table.

30 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
All periods	0.17	< 0.001
Eocene	0.01	0.38

(*ca.* 50 Ma)		
Mid-Pliocene (3.3–3.0 Ma)	0.10	< 0.05
Last Glacial Maximum (LGM, 22–18 ka)	0.04	0.22

40 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
All periods	0.24	< 0.001
Eocene (*ca.* 50 Ma)	–0.14	0.99
Mid-Pliocene (3.3–3.0 Ma)	0.06	0.07
Last Glacial Maximum (LGM, 22–18 ka)	–0.05	0.80

50 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
All periods	0.27	< 0.001
Eocene (*ca.* 50 Ma)	–0.02	0.58
Mid-Pliocene (3.3–3.0 Ma)	0.03	0.17
Last Glacial Maximum (LGM, 22–18 ka)	0.01	0.39

60 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
All periods	0.23	< 0.001
Eocene (*ca.* 50 Ma)	–0.01	0.59

Mid-Pliocene (3.3–3.0 Ma)	0.02	0.22
Last Glacial Maximum (LGM, 22–18 ka)	–0.05	0.87

Table A2.34 | Spatial auto-correlation test for Generalised Additive Models (GAMs) with species turnover as the response variable and climate and its deep-time geographic dynamics as explanatory variables. Summary of spatial autocorrelation tests for the residuals of GAMs fitted for all recent climatic conditions, as well as separately according to their first appearance. Turnover component of β-diversity is the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence to the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Potential spatial autocorrelation was assessed using Moran's *I* test on the residuals of each GAM. Moran's *I* index and corresponding *p*-value of each GAM are presented in the table.

30 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
All periods	0.07	< 0.05
Eocene (*ca.* 50 Ma)	–0.17	0.98
Mid-Pliocene (3.3–3.0 Ma)	–0.05	0.79
Last Glacial Maximum (LGM, 22–18 ka)	0.02	0.34

40 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
All periods	0.06	< 0.05
Eocene (*ca.* 50 Ma)	–0.07	0.85
Mid-Pliocene (3.3–3.0 Ma)	0.02	0.28
Last Glacial Maximum (LGM, 22–18 ka)	–0.08	0.87
50 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
All periods	0.01	0.27
Eocene (*ca.* 50 Ma)	–0.05	0.82
Mid-Pliocene (3.3–3.0 Ma)	0.10	< 0.01
Last Glacial Maximum (LGM, 22–18 ka)	–0.03	0.66
60 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
All periods	–0.02	0.82
Eocene (*ca.* 50 Ma)	–0.10	0.99
Mid-Pliocene (3.3–3.0 Ma)	–0.01	0.62
Last Glacial Maximum (LGM, 22–18 ka)	–0.11	0.99

Table A2.35 | Spatial auto-correlation test for Generalised Additive Models (GAMs) with phylogenetic diversity as the response variable and climate and its deep-time geographic dynamics as explanatory variables. Summary of spatial autocorrelation tests for the residuals of GAMs fitted for all recent climatic conditions, as well as separately according to their first appearance. Faith's phylogenetic diversity[13] is the response variable. Explanatory variables include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence to the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Potential spatial autocorrelation was assessed using Moran's *I* test on the residuals of each GAM. Moran's *I* index and corresponding *p*-value of each GAM are presented in the table.

30 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
All periods	0.25	< 0.001
Eocene (*ca.* 50 Ma)	0.04	0.24
Mid-Pliocene (3.3–3.0 Ma)	0.11	0.02
Last Glacial Maximum (LGM, 22–18 ka)	–0.07	0.83
40 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
All periods	0.34	< 0.001
Eocene (*ca.* 50 Ma)	0.11	< 0.05

Mid-Pliocene (3.3–3.0 Ma)	0.01	0.39
Last Glacial Maximum (LGM, 22–18 ka)	–0.05	0.81
50 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
All periods	0.32	< 0.001
Eocene (*ca.* 50 Ma)	0.12	< 0.01
Mid-Pliocene (3.3–3.0 Ma)	0.14	< 0.001
Last Glacial Maximum (LGM, 22–18 ka)	–0.00047	0.46
60 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
All periods	0.31	< 0.001
Eocene (*ca.* 50 Ma)	0.14	< 0.001
Mid-Pliocene (3.3–3.0 Ma)	0.04	0.12
Last Glacial Maximum (LGM, 22–18 ka)	0.05	0.10

Table A2.36 | Spatial auto-correlation test for Generalised Additive Models (GAMs) with relative phylogenetic diversity as the response variable and climate and its deep-time geographic dynamics as explanatory variables. Summary of spatial autocorrelation tests for the residuals of GAMs fitted for all recent climatic conditions, as well as separately according to their first appearance. Relative phylogenetic diversity (calculated as the residuals from regressing phylogenetic diversity against species richness) is the response variable. Explanatory variables

include: T_m (°C), recent (1901–1989) temperature index; ln MI (unitless), recent natural-log-transformed moisture index; Recent Area (km^2), total land area occupied by a climatic condition in the recent period; ΔArea (km^2), change in total land area occupied by a climatic condition from its earliest occurrence to the recent period; Movement (km), geographic shift of a climatic condition since its earliest occurrence to the recent period. Different GAMs were constructed based on climatic spaces defined by the recent climatic conditions (T_m and ln MI) at four resolutions, with each climatic axis divided into 30, 40, 50 or 60 equal intervals. Potential spatial autocorrelation was assessed using Moran's *I* test on the residuals of each GAM. Moran's *I* index and corresponding *p*-value of each GAM are presented in the table.

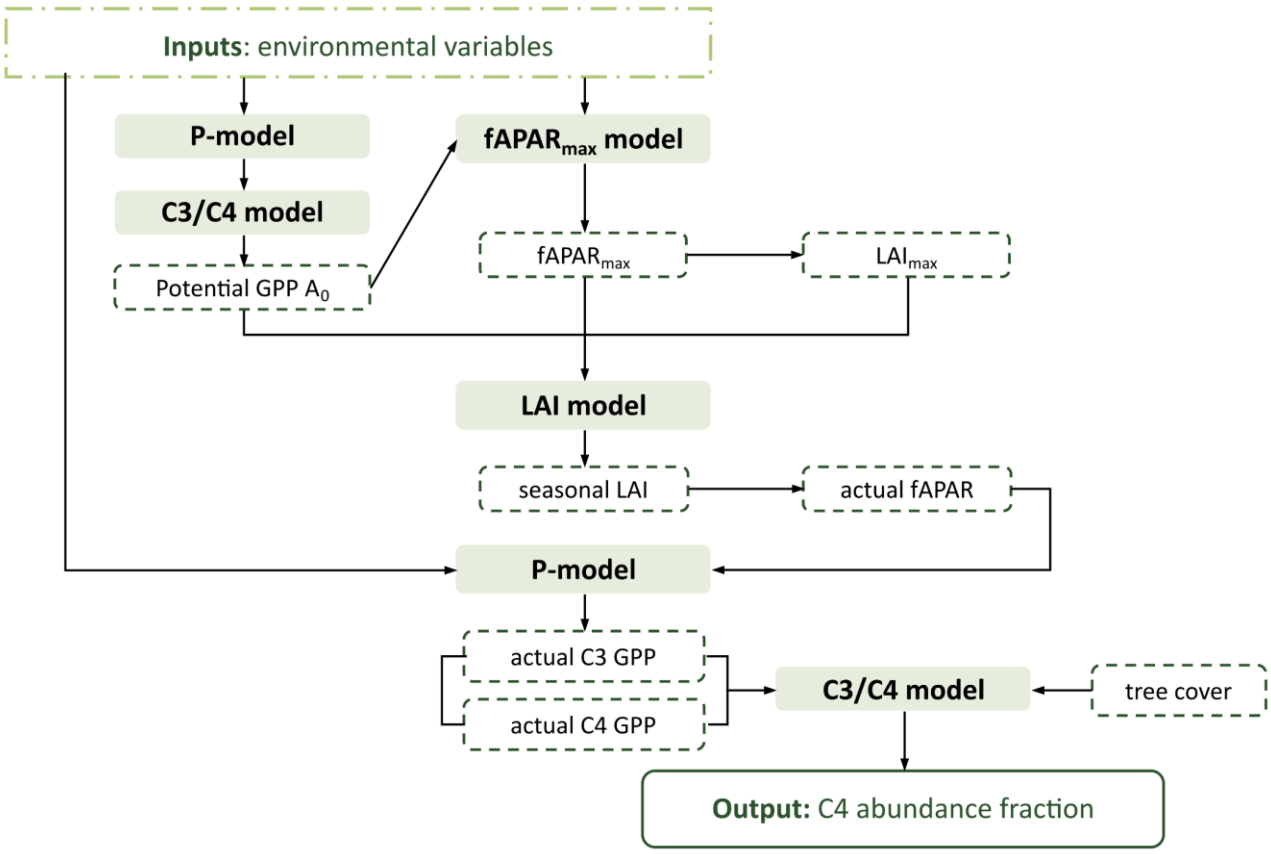
30 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
All periods	0.39	< 0.001
Eocene (*ca.* 50 Ma)	0.01	0.38
Mid-Pliocene (3.3–3.0 Ma)	0.18	< 0.001
Last Glacial Maximum (LGM, 22–18 ka)	–0.10	0.91
40 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
All periods	0.45	< 0.001
Eocene (*ca.* 50 Ma)	0.02	0.34
Mid-Pliocene (3.3–3.0 Ma)	0.27	< 0.001
Last Glacial Maximum (LGM, 22–18 ka)	–0.00021	0.45

50 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
All periods	0.48	< 0.001
Eocene (*ca.* 50 Ma)	0.06	0.08
Mid-Pliocene (3.3–3.0 Ma)	0.29	< 0.001
Last Glacial Maximum (LGM, 22–18 ka)	0.11	< 0.01
60 equal interval resolution		
Geological period	**Moran's *I***	***p*-value**
All periods	0.44	< 0.001
Eocene (*ca.* 50 Ma)	0.02	0.28
Mid-Pliocene (3.3–3.0 Ma)	0.19	< 0.001
Last Glacial Maximum (LGM, 22–18 ka)	0.11	< 0.01

A3 Supplementary information for Chapter 4

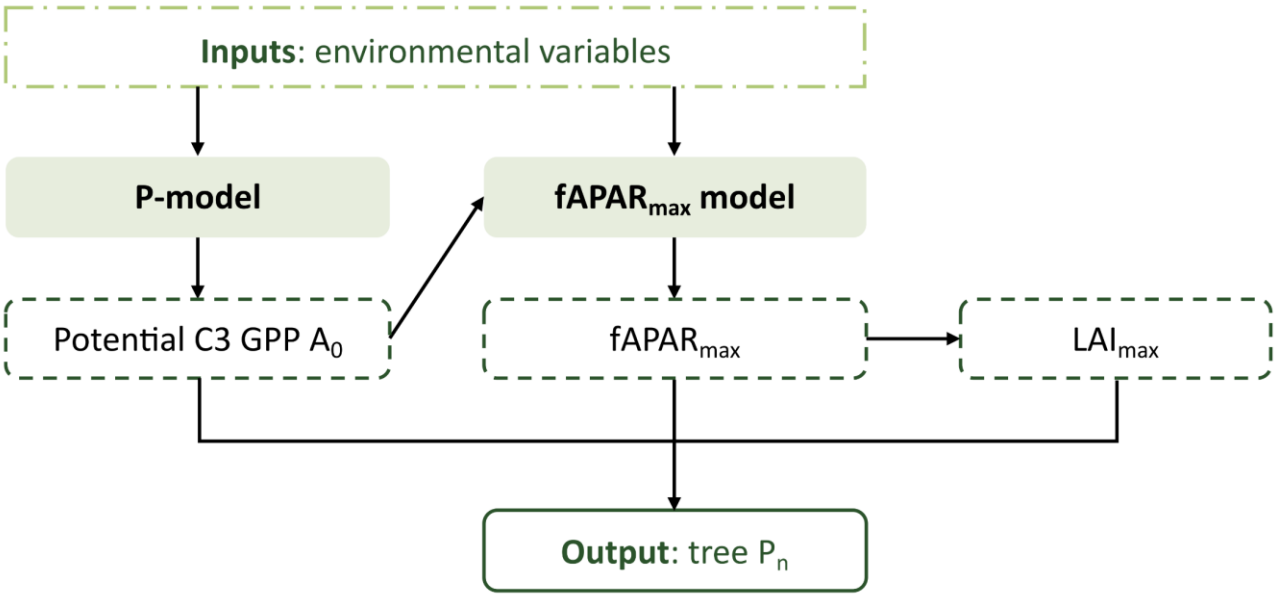
A3.1 Supplementary figures

Fig. A3.1: Modelling framework for simulating C4 abundance fraction.



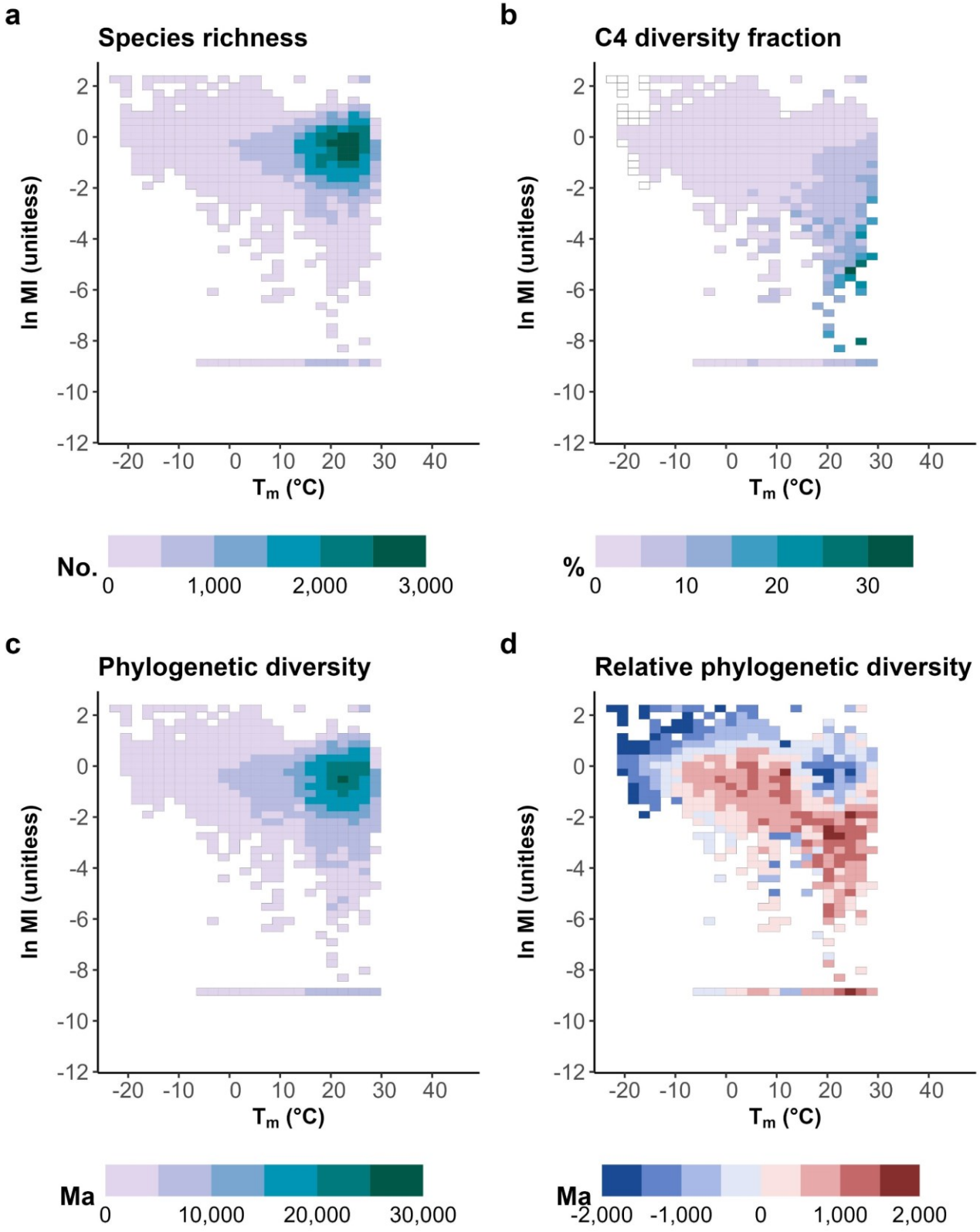
Flow chart of the modelling procedure. Details of input environmental variables and model implementation are provided in the Methods.

Fig. A3.2: Calculation of net carbon profit (P_n) of C3 trees.



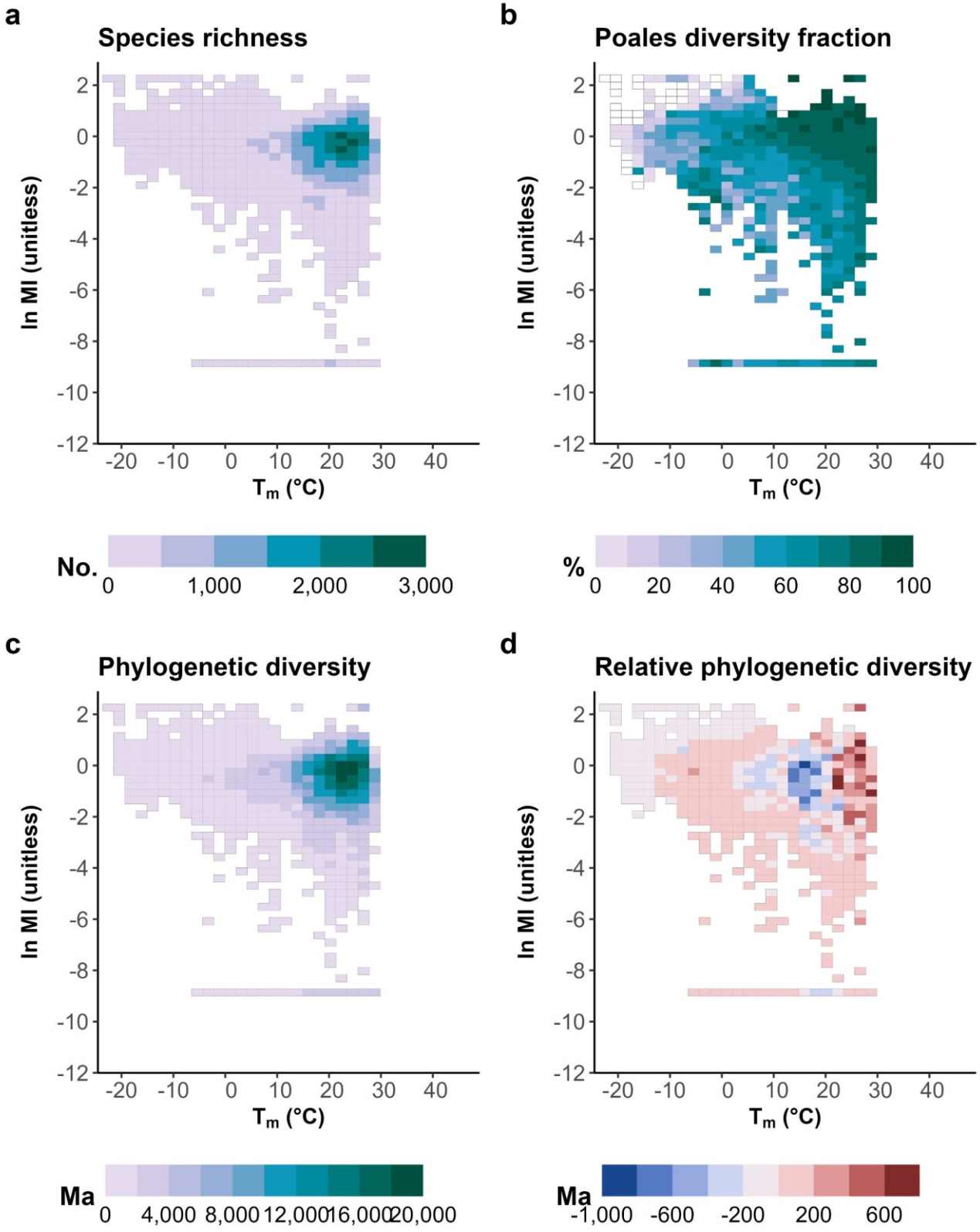
Flow chart illustrating the procedure for calculating the net carbon profit (P_n) of C3 trees. Details of input environmental variables and model implementation are provided in the Methods.

Fig. A3.3: Contemporary global patterns of C4 plant diversity in climatic space (50 intervals per axis).



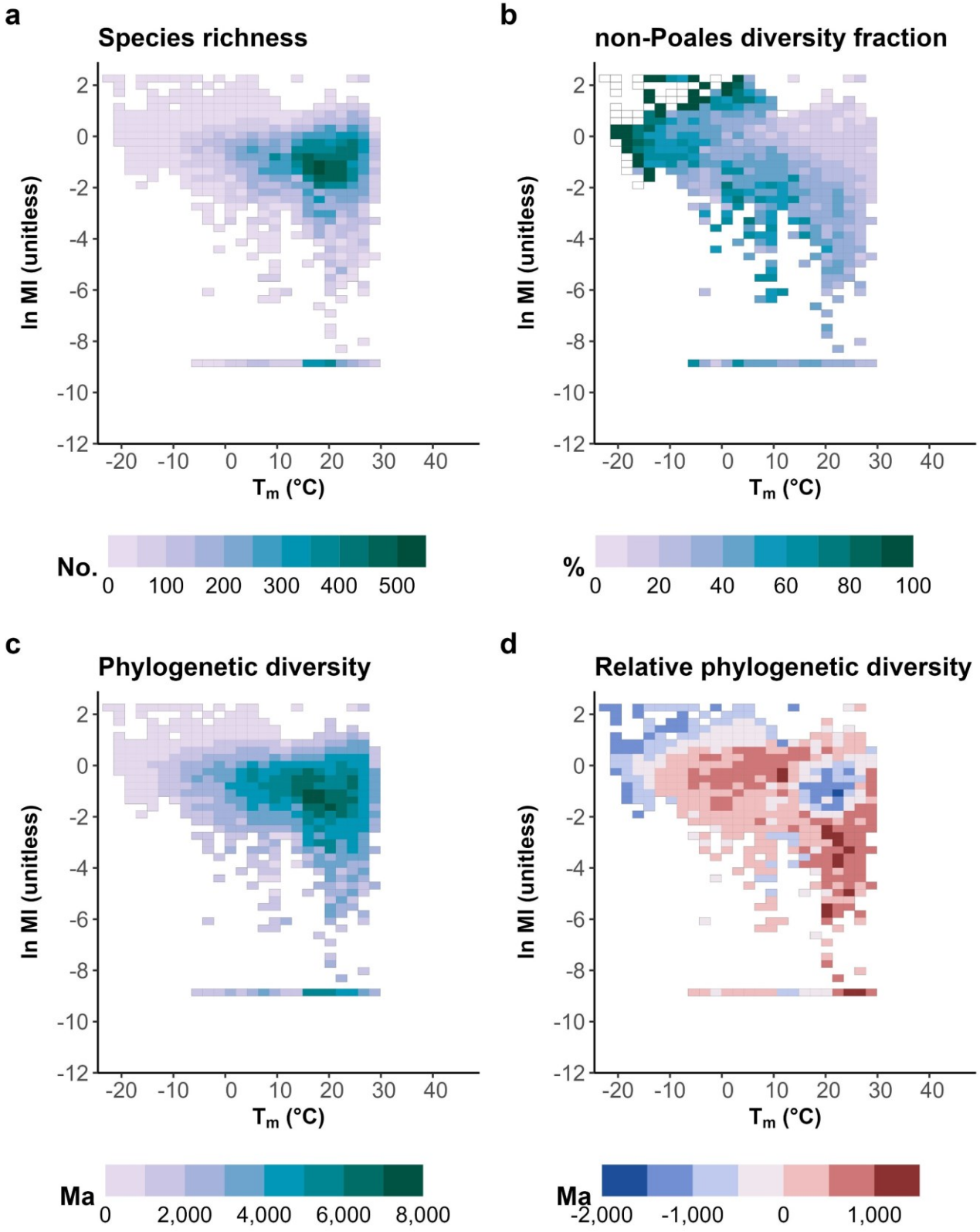
Global patterns of C4 plant diversity in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). The climatic space is divided into 50 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. All metrics are calculated based on native distributions of 7,156 C4 species and a time-calibrated phylogeny comprising all C4 species. **a**, C4 species richness, i.e. number of species per climatic bin. **b**, C4 diversity fraction (%), calculated as the proportion of vascular plant species using the C4 pathway within each bin. **c**, Phylogenetic diversity (Faith's PD, Ma)[13]. **d**, Relative phylogenetic diversity (RPD, Ma), calculated as the residual from a PD-richness regression. Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

Fig. A3.4: Contemporary global patterns of C4 Poales diversity in climatic space (50 intervals per axis).



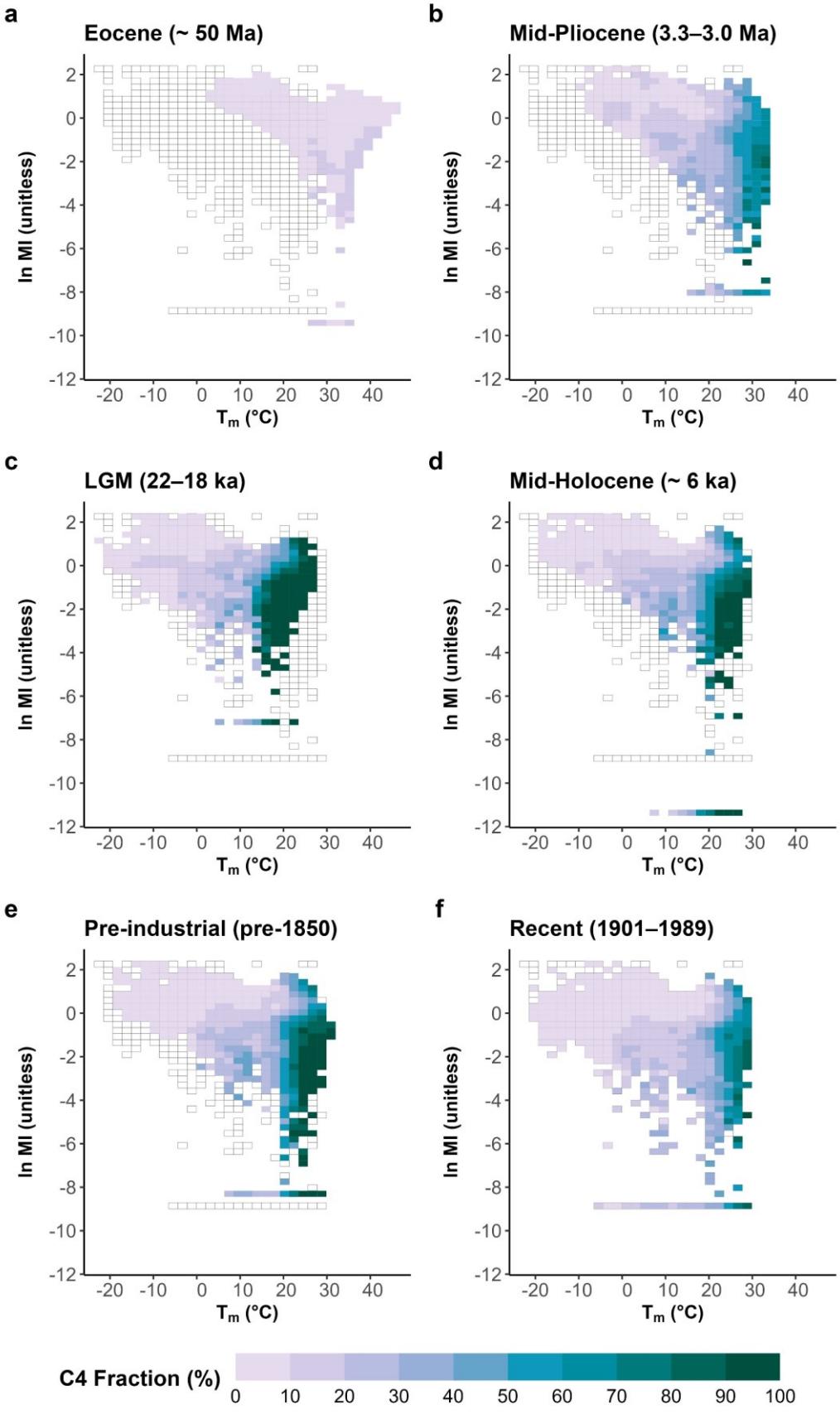
Global patterns of C4 Poales diversity in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). The climatic space is divided into 50 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. All metrics are calculated based on native distributions of 5,724 C4 Poales species and a time-calibrated phylogeny comprising all C4 Poales species. **a**, Species richness of C4 Poales, i.e. number of C4 Poales species per climatic bin. **b**, C4 diversity fraction (%), calculated as the proportion of local C4 species that belong to Poales within each bin. **c**, Phylogenetic diversity (Faith's PD, Ma)[13]. **d**, Relative phylogenetic diversity (RPD, Ma), calculated as the residual from a PD-richness regression. Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

Fig. A3.5: Contemporary global patterns of non-Poales C4 diversity in climatic space (50 intervals per axis).



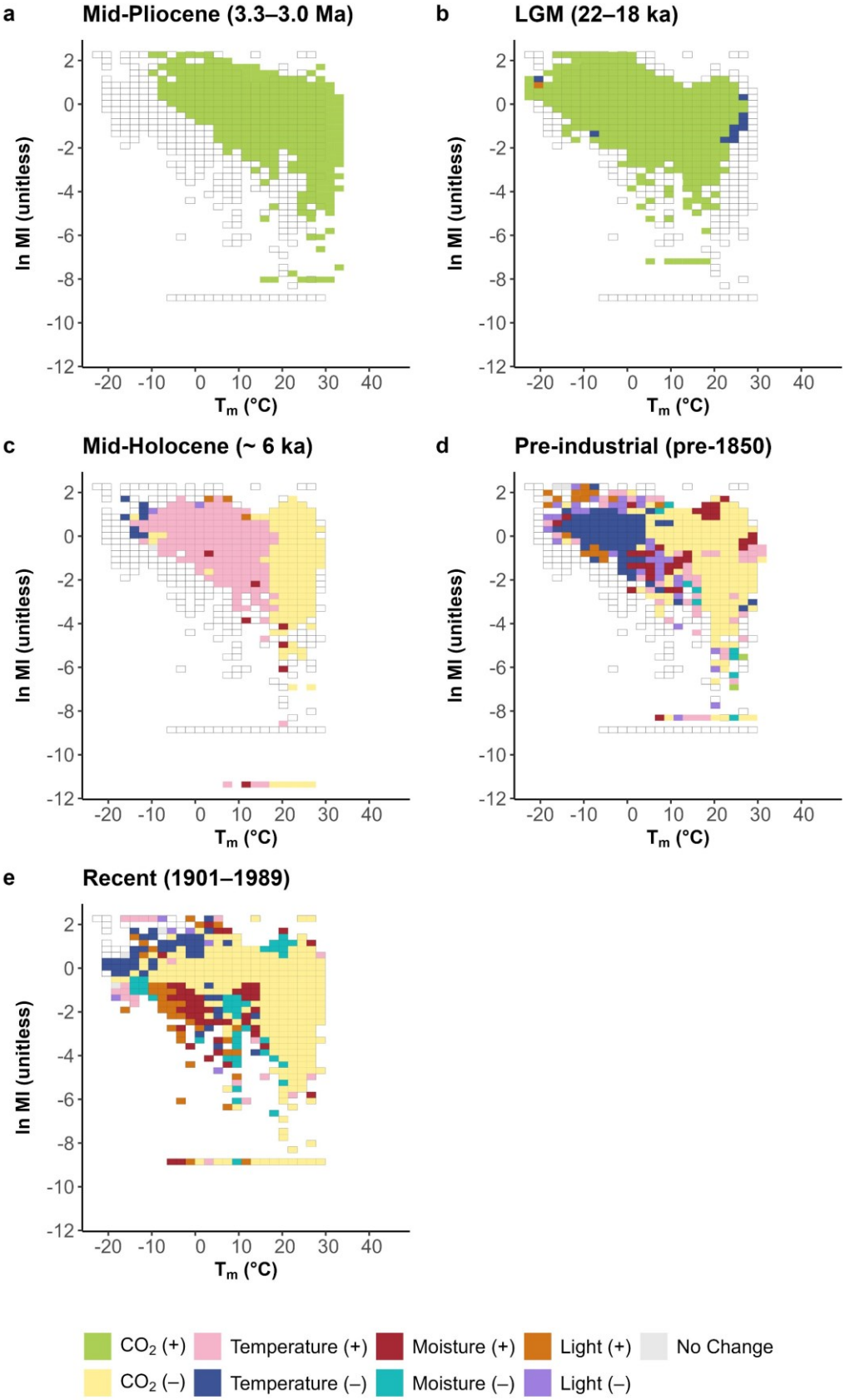
Global patterns of non-Poales C4 diversity in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). The climatic space is divided into 50 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. All metrics are calculated based on native distributions of 1,432 non-Poales C4 species and a time-calibrated phylogeny comprising all non-Poales C4 species. **a**, Species richness of non-Poales C4, i.e. number of non-Poales C4 species per climatic bin. **b**, C4 diversity fraction (%), calculated as the proportion of local C4 species that do not belong to Poales. **c**, Phylogenetic diversity (Faith's PD, Ma)[13]. **d**, Relative phylogenetic diversity (RPD, Ma), calculated as the residual from a PD-richness regression. Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

Fig. A3.6: Simulated historical biogeography of C4 plants in climatic space (50 intervals per axis).



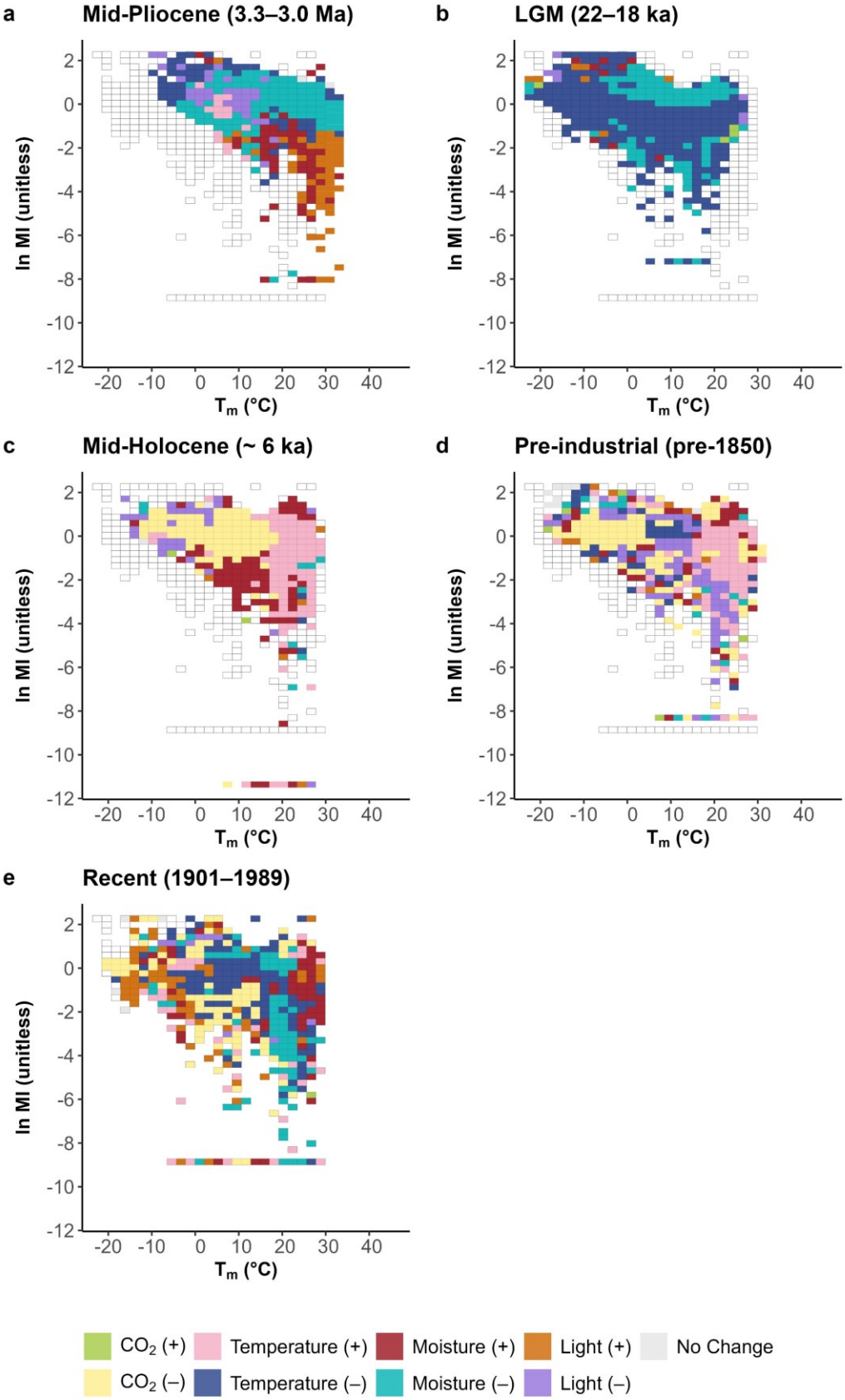
Modelled C4 plant fraction (% of combined C3 and C4 plant abundance) across six geological slices, visualised in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). Historical C4 plant fractions are simulated using models based on eco-evolutionary optimality (EEO) theory. The climatic space is divided into 50 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. Panels correspond to: **a**, Eocene (*ca.* 50 Ma); **b**, Mid-Pliocene (3.3–3.0 Ma); **c**, Last Glacial Maximum (LGM, 22–18 ka); **d**, Mid-Holocene (*ca.* 6 ka); **e**, pre-industrial (pre-1850); **f**, recent period (1901–1989). Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

Fig. A3.7: Dominant environmental drivers of simulated C4 plant biogeography in climatic space (50 intervals per axis).



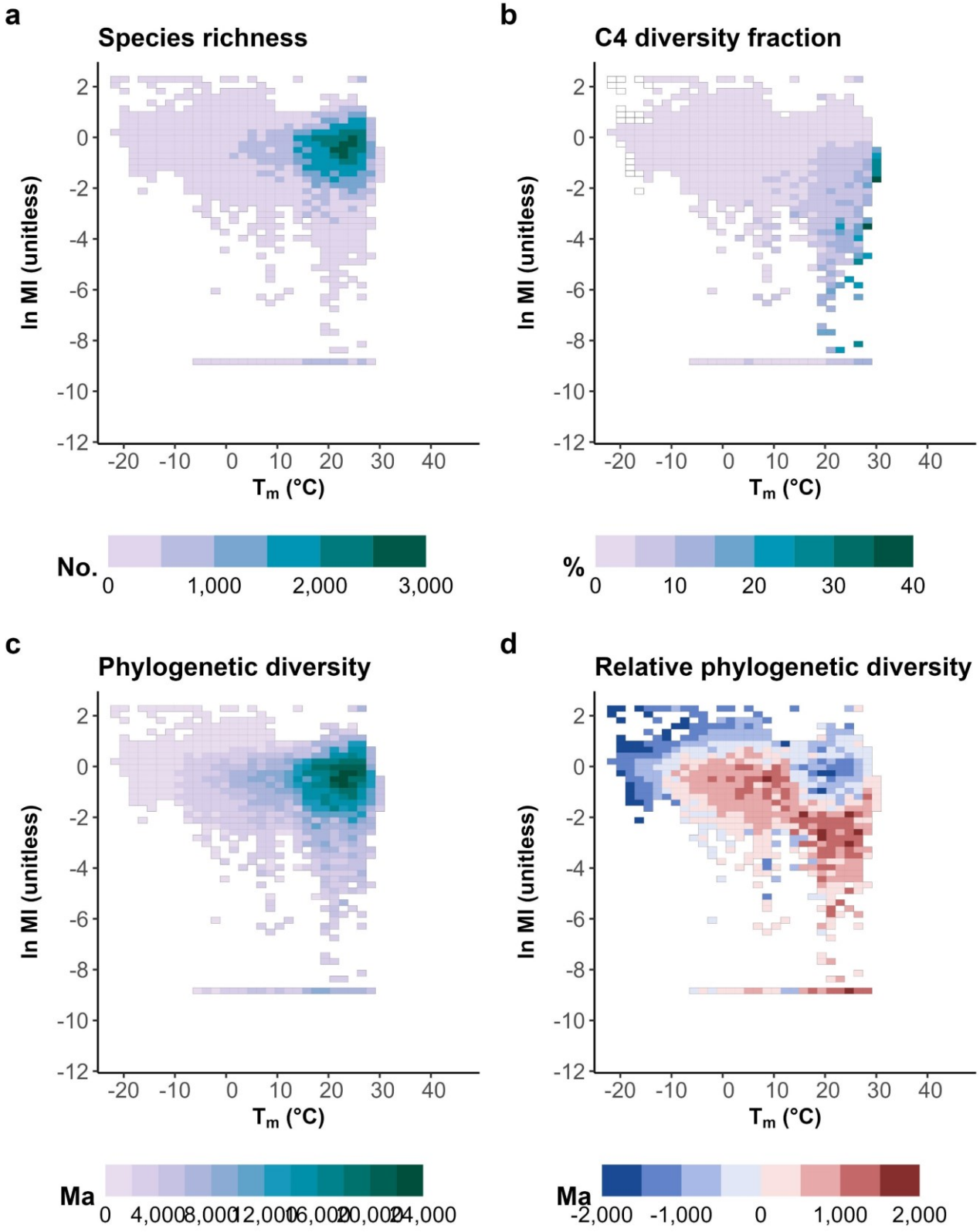
Spatial patterns of the dominant environmental driver influencing changes in modelled C4 plant fraction, mapped in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). The climatic space is divided into 50 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. Independent driver contribution is calculated by Stein-Alpert decomposition of model simulations. Panels correspond to: **a**, Mid-Pliocene (3.3–3.0 Ma); **b**, Last Glacial Maximum (LGM, 22–18 ka); **c**, Mid-Holocene (*ca.* 6 ka); **d**, pre-industrial (pre-1850); **e**, recent period (1901–1989). Each climatic bin is assigned to the environmental variable (i.e. CO_2, temperature, moisture and light) that exerted the strongest net effect on simulated C4 fraction changes. Colour hues denote the identity of the dominant driver; (±) indicates whether the change in that driver increased (+) or decreased (–) the local C4 fraction. Grey indicates no detectable change. Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

Fig. A3.8: Second dominant environmental drivers of simulated C4 plant biogeography in climatic space (50 intervals per axis).



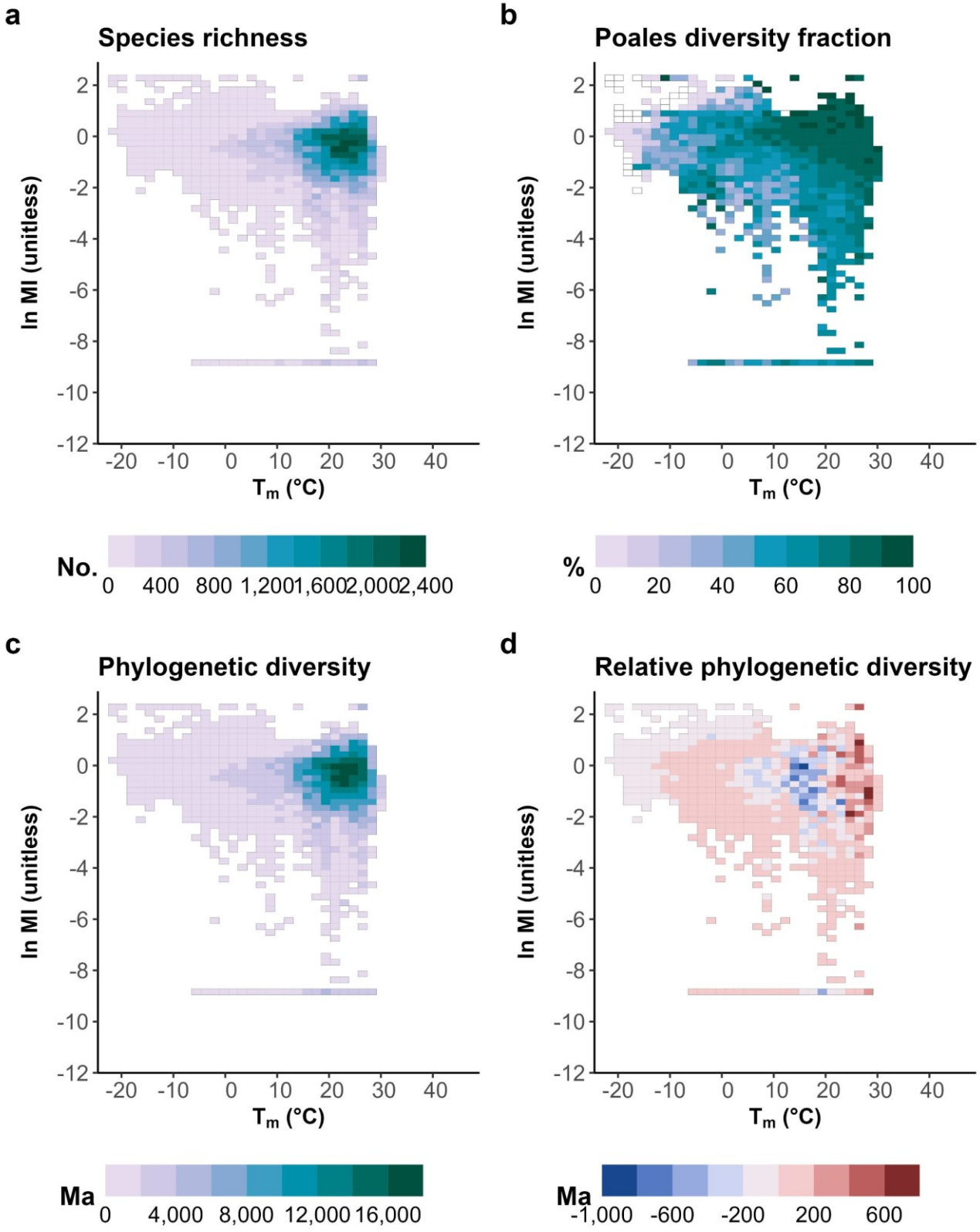
Spatial patterns of the second dominant environmental driver influencing changes in modelled C4 plant fraction, mapped in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). The climatic space is divided into 50 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. Independent driver contribution is calculated by Stein-Alpert decomposition of model simulations. Panels correspond to: **a**, Mid-Pliocene (3.3–3.0 Ma); **b**, Last Glacial Maximum (LGM, 22–18 ka); **c**, Mid-Holocene (*ca.* 6 ka); **d**, pre-industrial (pre-1850); **e**, recent period (1901–1989). Each climatic bin is assigned to the environmental variable (i.e. CO_2, temperature, moisture and light) that exerted the second strongest net effect on simulated C4 fraction changes. Colour hues denote the identity of the second dominant driver; (±) indicates whether the change in that driver increased (+) or decreased (–) the local C4 fraction. Grey indicates no detectable change. Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

Fig. A3.9: Contemporary global patterns of C4 plant diversity in climatic space (60 intervals per axis).



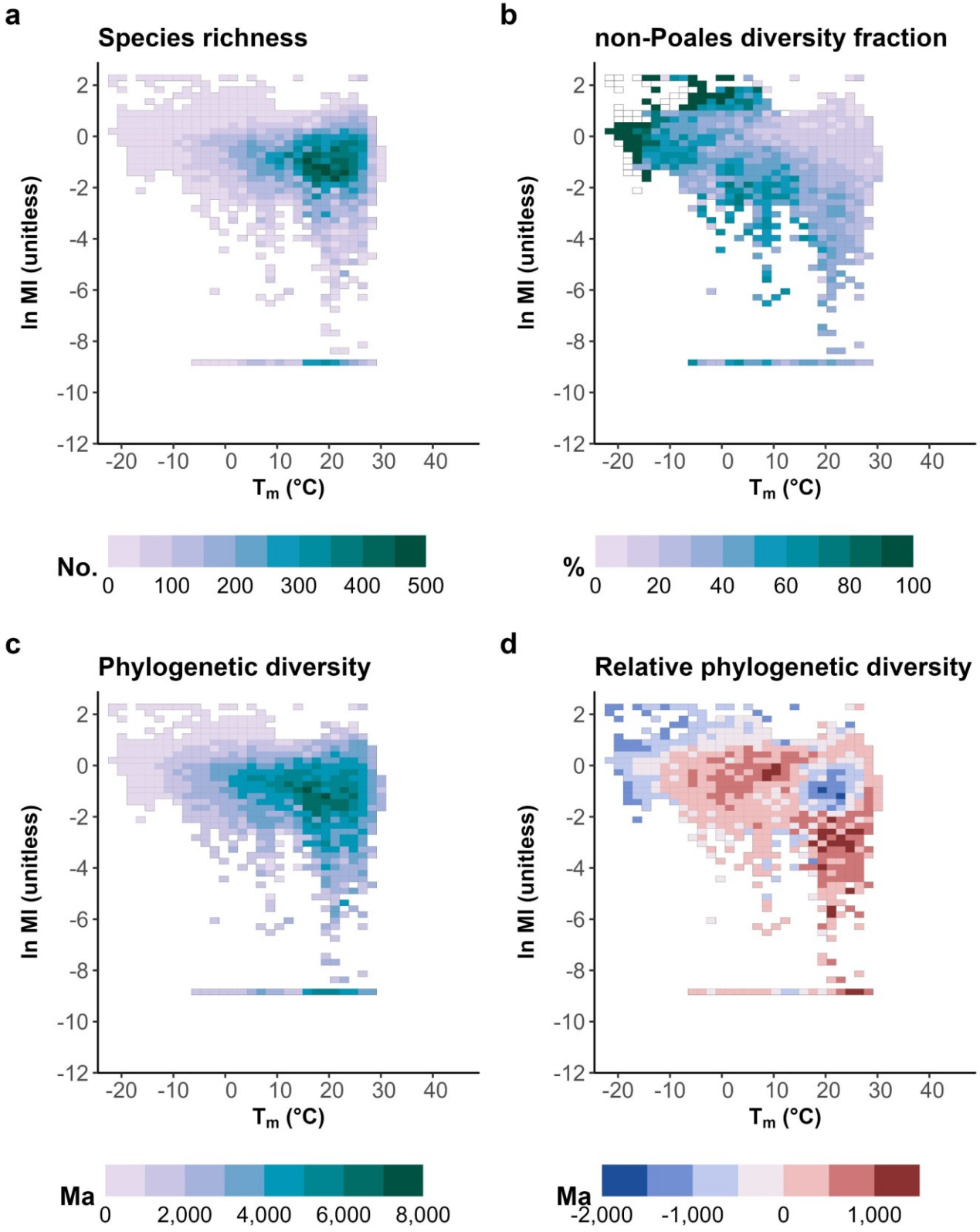
Global patterns of C4 plant diversity in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). The climatic space is divided into 60 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. All metrics are calculated based on native distributions of 7,156 C4 species and a time-calibrated phylogeny comprising all C4 species. **a**, C4 species richness, i.e. number of species per climatic bin. **b**, C4 diversity fraction (%), calculated as the proportion of vascular plant species using the C4 pathway within each bin. **c**, Phylogenetic diversity (Faith's PD, Ma)[13]. **d**, Relative phylogenetic diversity (RPD, Ma), calculated as the residual from a PD-richness regression. Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

Fig. A3.10: Contemporary global patterns of C4 Poales diversity in climatic space (60 intervals per axis).



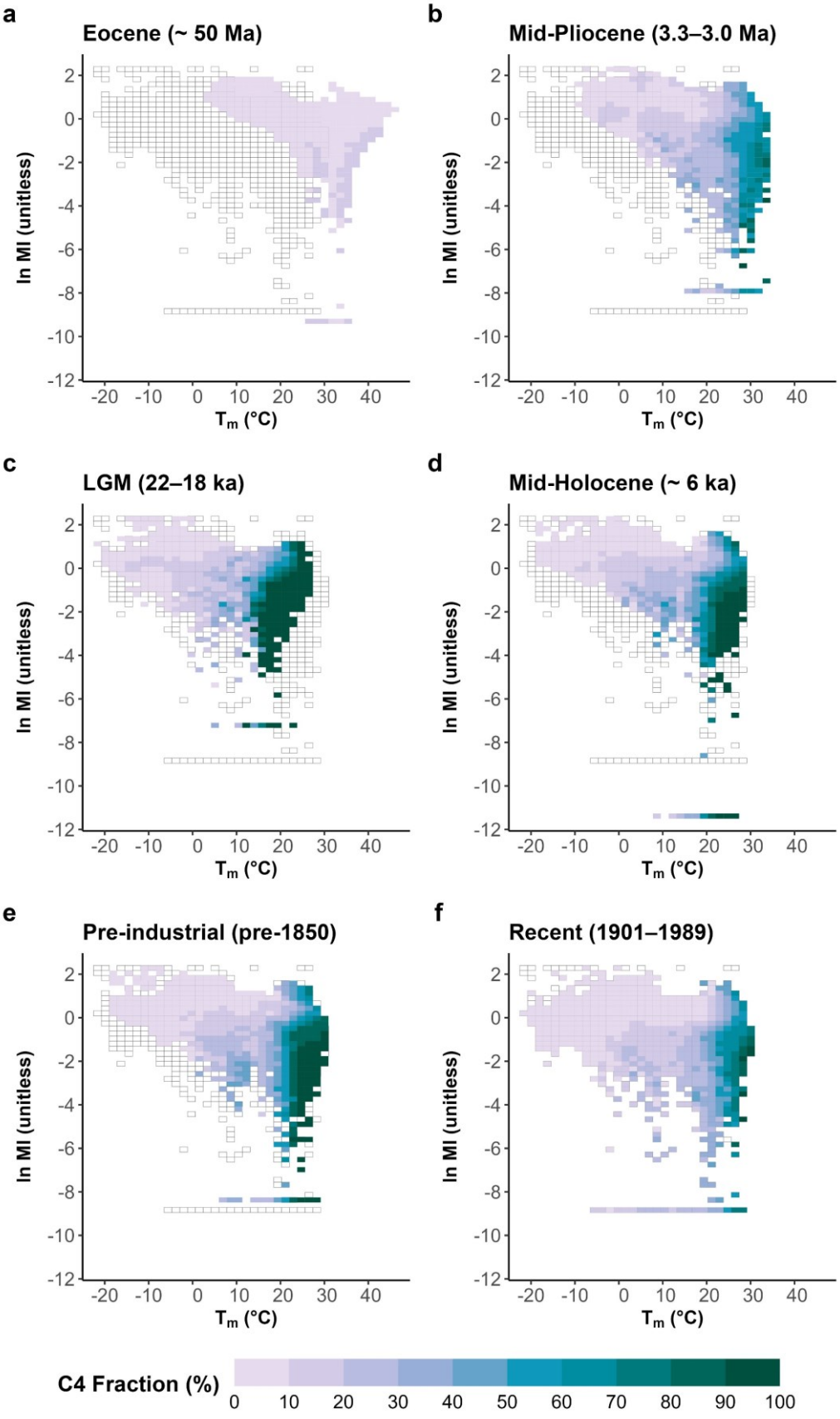
Global patterns of C4 Poales diversity in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). The climatic space is divided into 60 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. All metrics are calculated based on native distributions of 5,724 C4 Poales species and a time-calibrated phylogeny comprising all C4 Poales species. **a**, Species richness of C4 Poales, i.e. number of C4 Poales species per climatic bin. **b**, C4 diversity fraction (%), calculated as the proportion of local C4 species that belong to Poales within each bin. **c**, Phylogenetic diversity (Faith's PD, Ma)[13]. **d**, Relative phylogenetic diversity (RPD, Ma), calculated as the residual from a PD-richness regression. Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

Fig. A3.11: Contemporary global patterns of non-Poales C4 diversity in climatic space (60 intervals per axis).



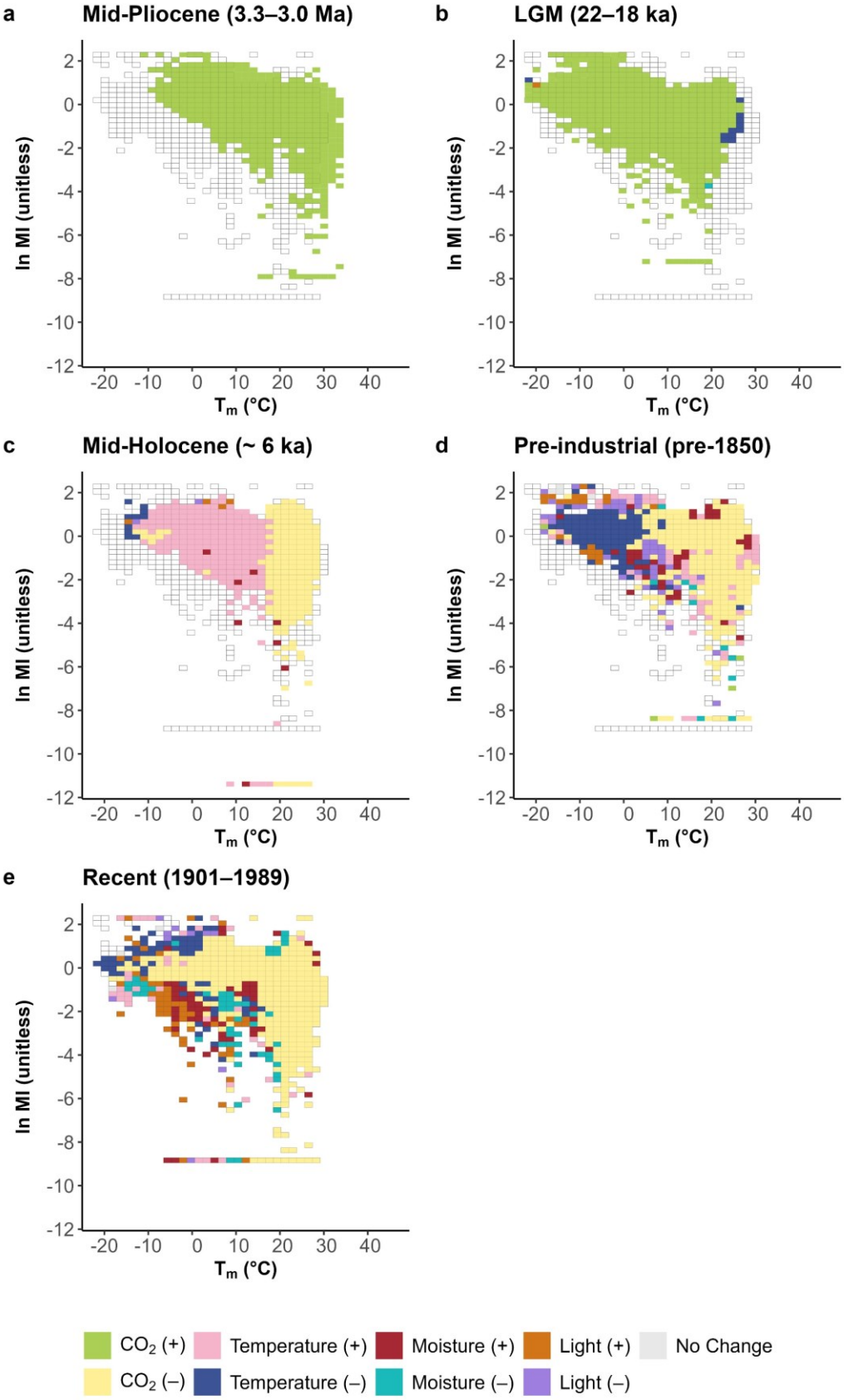
Global patterns of non-Poales C4 diversity in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). The climatic space is divided into 60 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. All metrics are calculated based on native distributions of 1,432 non-Poales C4 species and a time-calibrated phylogeny comprising all non-Poales C4 species. **a**, Species richness of non-Poales C4, i.e. number of non-Poales C4 species per climatic bin. **b**, C4 diversity fraction (%), calculated as the proportion of local C4 species that do not belong to Poales. **c**, Phylogenetic diversity (Faith's PD, Ma)[13]. **d**, Relative phylogenetic diversity (RPD, Ma), calculated as the residual from a PD-richness regression. Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

Fig. A3.12: Simulated historical biogeography of C4 plants in climatic space (60 intervals per axis).



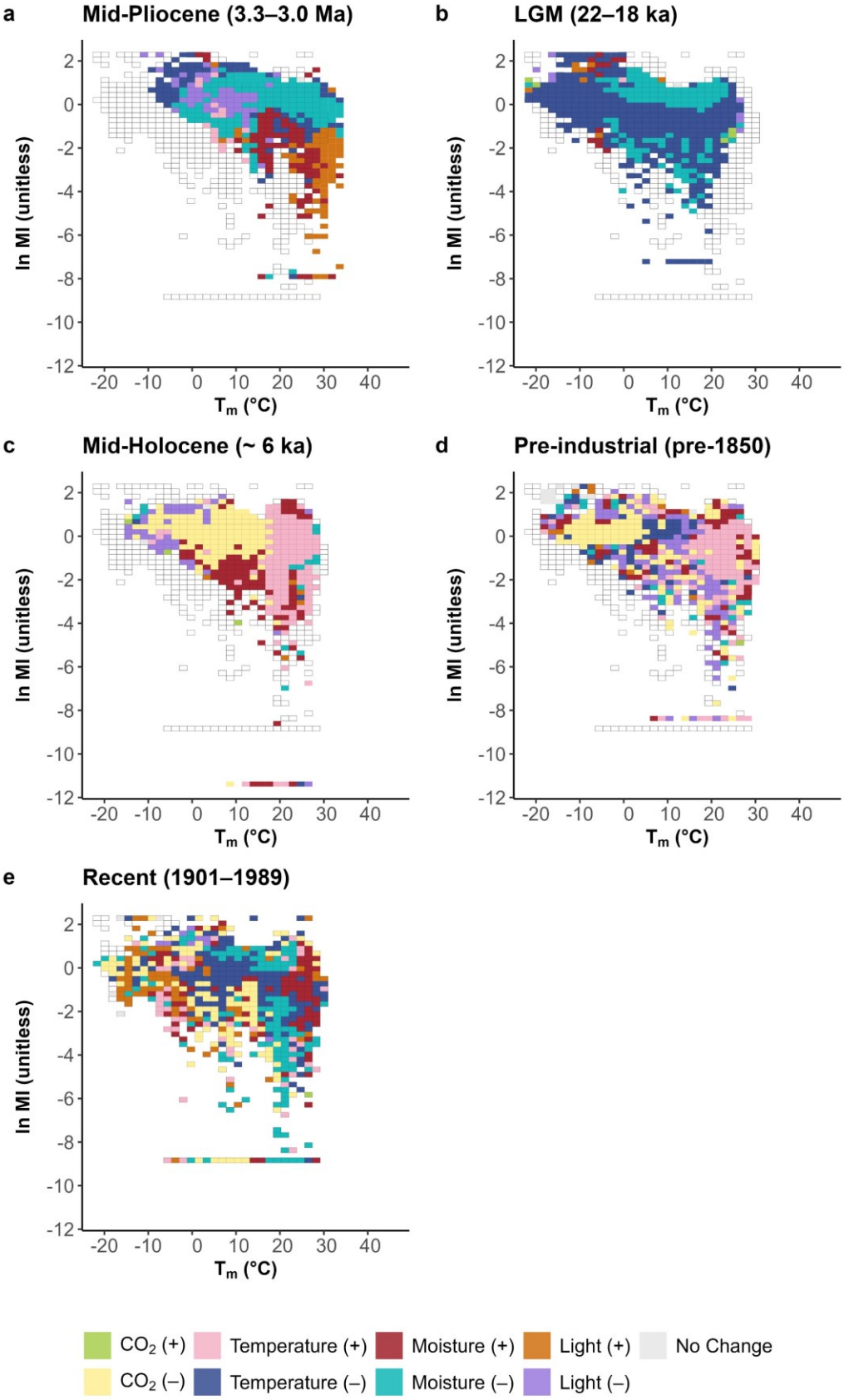
Modelled C4 plant fraction (% of combined C3 and C4 plant abundance) across six geological slices, visualised in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). Historical C4 plant fractions are simulated using models based on eco-evolutionary optimality (EEO) theory. The climatic space is divided into 60 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. Panels correspond to: **a**, Eocene (*ca.* 50 Ma); **b**, Mid-Pliocene (3.3–3.0 Ma); **c**, Last Glacial Maximum (LGM, 22–18 ka); **d**, Mid-Holocene (*ca.* 6 ka); **e**, pre-industrial (pre-1850); **f**, recent period (1901–1989). Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

Fig. A3.13: Dominant environmental drivers of simulated C4 plant biogeography in climatic space (60 intervals per axis).



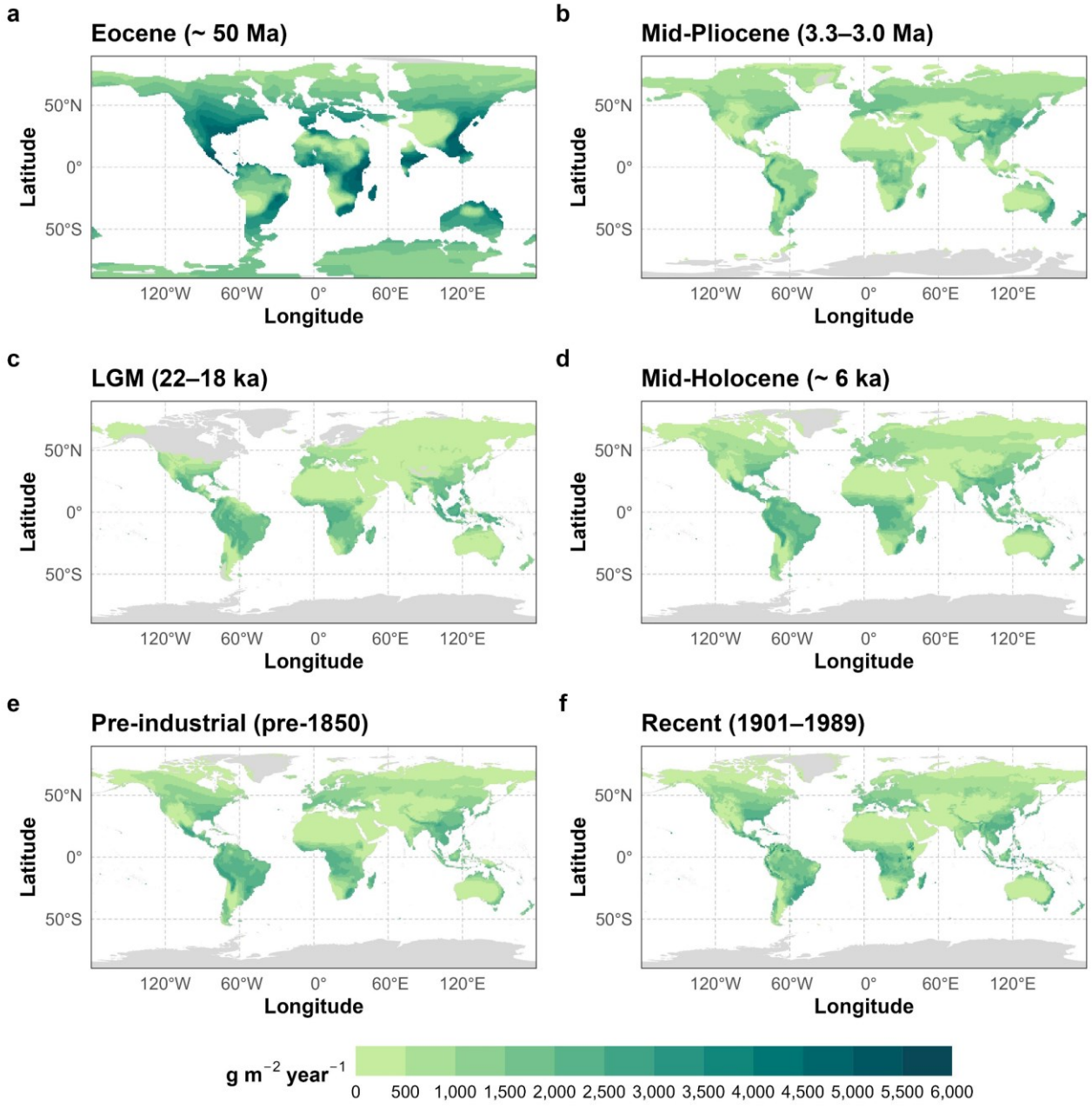
Spatial patterns of the dominant environmental driver influencing changes in modelled C4 plant fraction, mapped in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). The climatic space is divided into 60 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. Independent driver contribution is calculated by Stein-Alpert decomposition of model simulations. Panels correspond to: **a**, Mid-Pliocene (3.3–3.0 Ma); **b**, Last Glacial Maximum (LGM, 22–18 ka); **c**, Mid-Holocene (*ca.* 6 ka); **d**, pre-industrial (pre-1850); **e**, recent period (1901–1989). Each climatic bin is assigned to the environmental variable (i.e. CO_2, temperature, moisture and light) that exerted the strongest net effect on simulated C4 fraction changes. Colour hues denote the identity of the dominant driver; (±) indicates whether the change in that driver increased (+) or decreased (–) the local C4 fraction. Grey indicates no detectable change. Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

Fig. A3.14: Second dominant environmental drivers of simulated C4 plant biogeography in climatic space (60 intervals per axis).



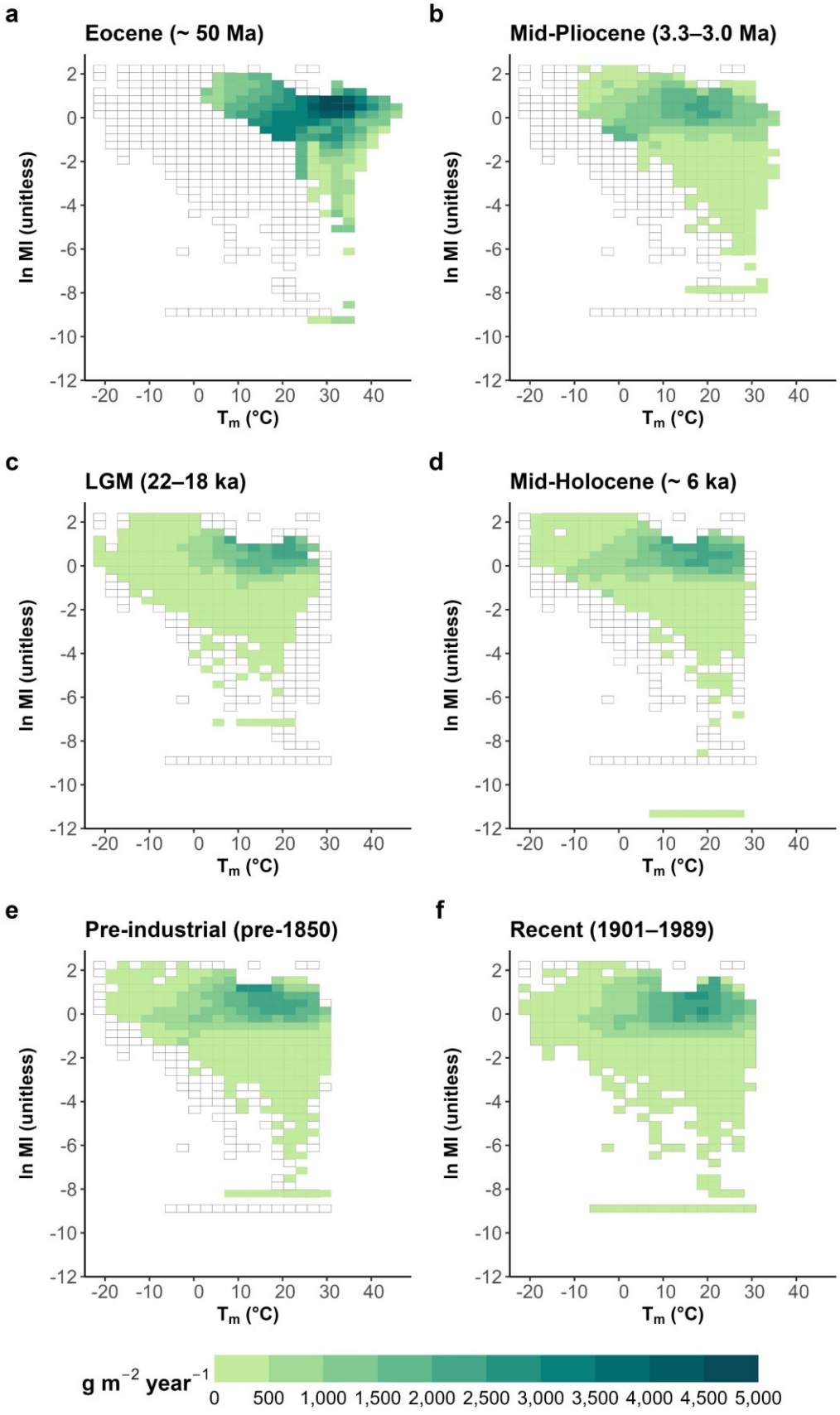
Spatial patterns of the second dominant environmental driver influencing changes in modelled C4 plant fraction, mapped in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). The climatic space is divided into 60 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. Independent driver contribution is calculated by Stein-Alpert decomposition of model simulations. Panels correspond to: **a**, Mid-Pliocene (3.3–3.0 Ma); **b**, Last Glacial Maximum (LGM, 22–18 ka); **c**, Mid-Holocene (*ca.* 6 ka); **d**, pre-industrial (pre-1850); **e**, recent period (1901–1989). Each climatic bin is assigned to the environmental variable (i.e. CO_2, temperature, moisture and light) that exerted the second strongest net effect on simulated C4 fraction changes. Colour hues denote the identity of the second dominant driver; (±) indicates whether the change in that driver increased (+) or decreased (–) the local C4 fraction. Grey indicates no detectable change. Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

Fig. A3.15: Simulated global mean annual gross primary production (GPP).



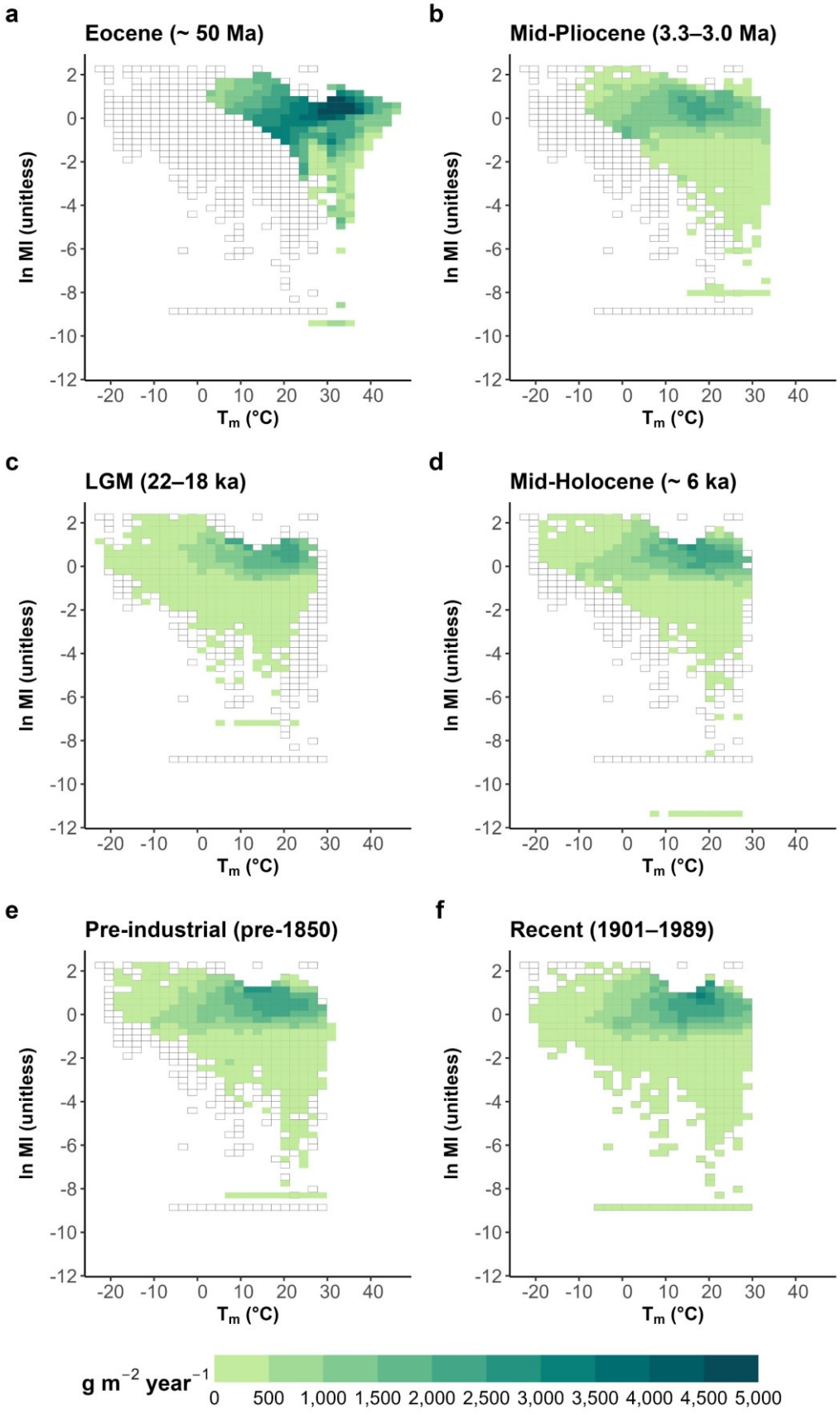
Modelled global mean annual gross primary production (GPP) at 1° spatial resolution across six geological slices, reconstructed using models based on eco-evolutionary optimality (EEO) theory. Panels correspond to: **a**, Eocene (*ca.* 50 Ma); **b**, Mid-Pliocene (3.3–3.0 Ma); **c**, Last Glacial Maximum (LGM, 22–18 ka); **d**, Mid-Holocene (*ca.* 6 ka); **e**, pre-industrial (pre-1850); **f**, recent period (1901–1989).

Fig. A3.16: Simulated global mean annual gross primary production (GPP) in climatic space (40 intervals per axis).



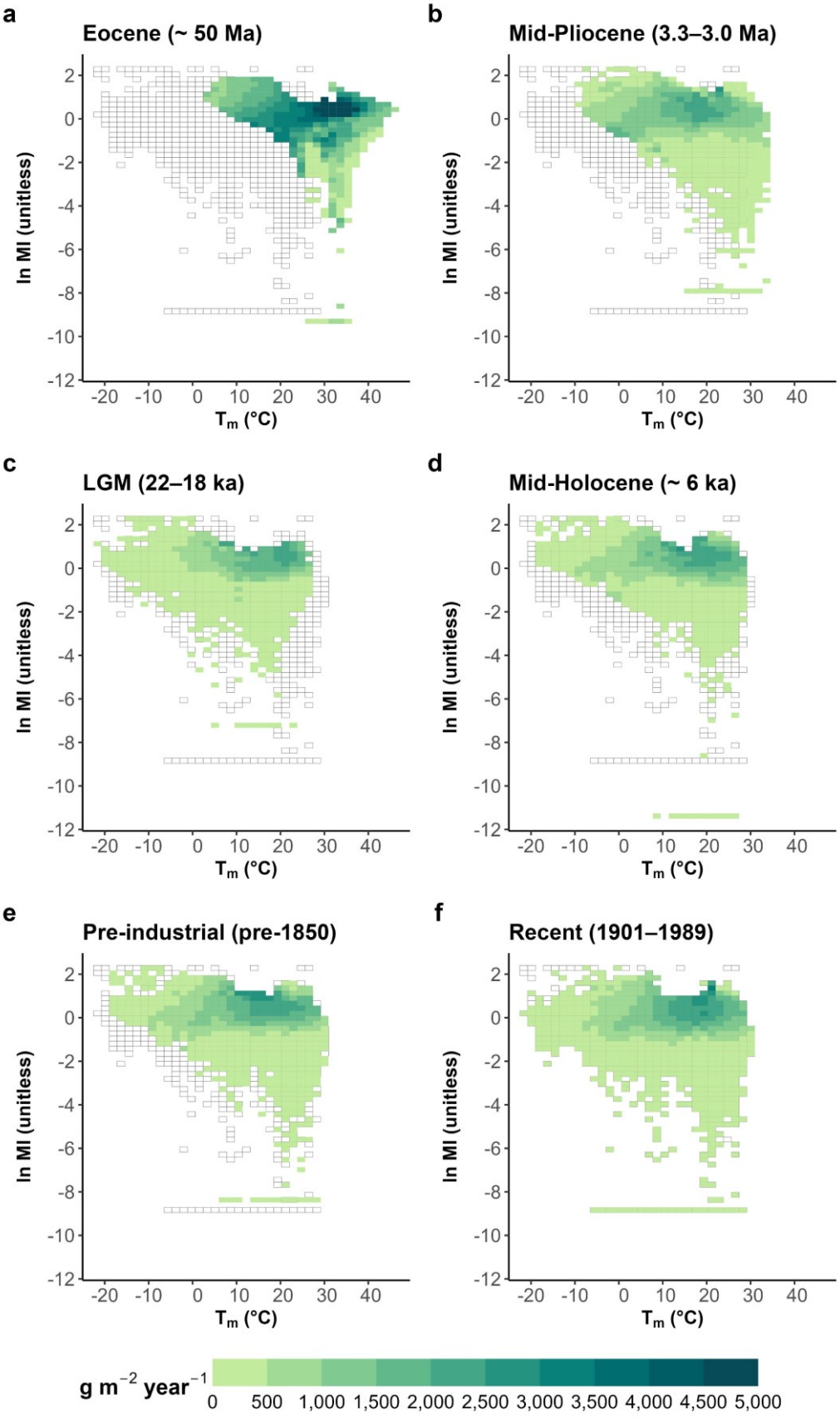
Global mean annual gross primary production (GPP) simulated using eco-evolutionary optimality (EEO) models across six geological slices, visualised in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). The climatic space is divided into 40 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. Panels correspond to: **a**, Eocene (*ca.* 50 Ma); **b**, Mid-Pliocene (3.3–3.0 Ma); **c**, Last Glacial Maximum (LGM, 22–18 ka); **d**, Mid-Holocene (*ca.* 6 ka); **e**, pre-industrial (pre-1850); **f**, recent period (1901–1989). Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

Fig. A3.17: Simulated global mean annual gross primary production (GPP) in climatic space (50 intervals per axis).



Global mean annual gross primary production (GPP) simulated using eco-evolutionary optimality (EEO) models across six geological slices, visualised in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). The climatic space is divided into 50 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. Panels correspond to: **a**, Eocene (*ca.* 50 Ma); **b**, Mid-Pliocene (3.3–3.0 Ma); **c**, Last Glacial Maximum (LGM, 22–18 ka); **d**, Mid-Holocene (*ca.* 6 ka); **e**, pre-industrial (pre-1850); **f**, recent period (1901–1989). Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

Fig. A3.18: Simulated global mean annual gross primary production (GPP) in climatic space (60 intervals per axis).



Global mean annual gross primary production (GPP) simulated using eco-evolutionary optimality (EEO) models across six geological slices, visualised in a climatic space defined using the temperature index (T_m, °C) and the natural-log-transformed moisture index (ln MI, unitless). The climatic space is divided into 60 equal intervals along each axis; each grid cell ('climatic bin') represents a unique climatic combination. Panels correspond to: **a**, Eocene (*ca.* 50 Ma); **b**, Mid-Pliocene (3.3–3.0 Ma); **c**, Last Glacial Maximum (LGM, 22–18 ka); **d**, Mid-Holocene (*ca.* 6 ka); **e**, pre-industrial (pre-1850); **f**, recent period (1901–1989). Black outlined bins indicate climatic conditions present during the recent period (1901–1989), shown as background for reference.

A3.2 Supplementary tables

Table A3.1 | Environmental variables used in this study. [a] Abbreviations of environmental variables. [b] Brief description of each variable. [c] Original units of these variables as provided in the source databases. [d] Standardised units applied for model fitting or analyses. [e] Variable type, distinguishing between primary variables obtained directly from databases and derived variables calculated from them for use in models or analyses.

Variable [a]	Description [b]	Original Unit [c]	Standard Unit [d]	Type [e]
clt	Total Cloud Cover Percentage	%	%	Primary Variables
[CO_2]	Atmospheric CO_2 Concentration	ppm	ppm	
hus	Specific Humidity	unitless	unitless	
elev	Elevation	m	m	
orog	Surface Altitude	m	m	
pr	Precipitation	$kg \cdot m^{-2} \cdot s^{-1}$	mm (aggregated totals)	
ps	Surface Air Pressure	Pa	Pa	
rsds	Surface Downwelling Shortwave Radiation	$W \cdot m^{-2}$	$W \cdot m^{-2}$	
tas	Near-Surface Air Temperature	K	°C	
treeFrac	Tree Cover Percentage	%	%	
aet	Actual evapotranspiration	mm (aggregated totals)	mm (aggregated totals)	Derived Variables
pet	Potential Evapotranspiration	mm (aggregated totals)	mm (aggregated totals)	

ppfd	Photosynthetic Photon Flux Density	$\mu mol \cdot m^{-2} \cdot s^{-1}$	$\mu mol \cdot m^{-2} \cdot s^{-1}$
sf	Sunshine Fraction	%	%
sm	Soil Moisture	mm	mm
vpd	Vapor Pressure Deficit	Pa	Pa

Table A3.2 | Environmental data information. Databases and model simulations used to extract each environmental variable in different geological periods. Abbreviations of variables are shown in Table A3.1.

Project	Version	Experiment	Geological period	Model	Simulation	Ensemble	Variable	Resolution	Reference
Deep-Time Model Intercomparison Project (DeepMIP)[21]	version 1.0	Eocene	Eocene (ca. 50 Ma)	CESM1.2-CAM5	× 6 CO_2	r1i1p1f1	clt, orog, pr, ps, rsds, tas	2.5° × 1.9°	Ref [22]
				COSMOS-landveg-r2413	× 4 CO_2		clt, orog, pr, ps, rsds, tas	3.75° × 3.71°	Ref [23]
				GFDL-CM2.1	× 4 CO_2, × 6 CO_2		clt, orog, pr, ps, rsds, tas	3.75° × 3.02°	Ref [24]
				INM-CM4-8	× 6 CO_2		hus, orog, pr, ps, rsds, tas	2.0° × 1.5°	Ref [23]
				NorESM1-F	× 4 CO_2		clt, orog, pr, ps, rsds, tas	2.5° × 1.9°	Ref [25]
Paleoclimat Modeling Intercomparison Project (PMIP)[26]	Phase 4 (PMIP4)	midPliocene-eoi400	Mid-Pliocene (3.3–3.0 Ma)	CESM2	--	r1i1p1f1	clt, orog, pr, ps, rsds, tas	100 km × 100 km	Ref [27]
				GISS-E2-1-G			clt, pr, ps, rsds, tas	250 km × 250 km	Refs [28,29]
				HadGEM3-GC31-LL			pr, tas	250 km × 250 km	Ref [30]

		NorESM1-F	clt, hus, orog, pr, ps, rsds, tas	250 km × 250 km	Ref [25]
lgm	Last Glacial Maximum (LGM, 22–18 ka)	AWI-ESM-1-1-LR	clt, hus, orog, pr, ps, rsds, tas	250 km × 250 km	Ref [31]
		MPI-ESM1-2-LR	clt, hus, orog, pr, ps, rsds, tas, treeFrac	250 km × 250 km	Ref [32]
midHolocene	Mid-Holocene (ca. 6 ka)	AWI-ESM-1-1-LR	clt, hus, orog, pr, ps, rsds, tas	250 km × 250 km	Ref [31]
		CESM2	clt, hus, orog, pr, ps, rsds, tas	100 km × 100 km	Ref [27]
		MPI-ESM1-2-LR	clt, hus, orog, pr, ps, rsds, tas, treeFrac	250 km × 250 km	Ref [32]
		MRI-ESM2-0	clt, hus, orog, pr, ps, rsds, tas	100 km × 100 km	Ref [33]
		NESM3	clt, hus, pr, ps, rsds, tas	250 km × 250 km	Ref [34]
		NorESM1-F	hus, ps, rsds	250 km × 250 km	Ref [25]

				NorESM2-LM			clt, orog, pr, tas	250 km × 250 km	Ref [35]
Coupled Model Intercomparison Project (CMIP)[36]	Phase 6 (CMIP6)	piControl	Pre-industrial era (before 1850)	ACCESS-CM2	--	r1i1p1f1	clt, hus, orog, pr, ps, rsds, tas, treeFrac	250 km × 250 km	Ref [37]
				CESM2			clt, hus, orog, pr, ps, rsds, tas	100 km × 100 km	Ref [27]
				MIROC6			clt, hus, orog, pr, ps, rsds, tas	250 km × 250 km	Ref [38]
				MPI-ESM1-2-LR			clt, hus, orog, pr, ps, rsds, tas, treeFrac	250 km × 250 km	Ref [32]
				MRI-ESM2-0			clt, hus, orog, pr, ps, rsds, tas	100 km × 100 km	Ref [33]
				NESM3			clt, hus, pr, ps, rsds, tas	250 km × 250 km	Ref [34]
				NorESM2-LM			orog, pr, rsds, tas	250 km × 250 km	Ref [35]
		historical	Recent period (1901–1989)	ACCESS-CM2			hus, ps, rsds	250 km × 250 km	Ref [37]

				CESM2			hus, ps, rsds	100 km × 100 km	Ref [27]
				MIROC6			hus, ps, rsds	250 km × 250 km	Ref [38]
				MPI-ESM1-2-LR			hus, ps, rsds	250 km × 250 km	Ref [32]
				MRI-ESM2-0			hus, ps, rsds	100 km × 100 km	Ref [33]
				NESM3			hus, ps, rsds	250 km × 250 km	Ref [34]
				NorESM2-LM			hus, ps, rsds	250 km × 250 km	Ref [35]
Climatic Research Unit (CRU) Time-Series (TS)[39]	version 4.09	--	Recent period (1901–1989)	Observation-based gridded dataset	--	--	clt, pr, tas	0.5° × 0.5°	Ref [39]
WorldClim[40]	version						orog	30	Ref [40]

	2.1			arc-second grid	
Potential tree cover under current and future climate scenarios[41]	version 1.0	Machine learning model based on satellite observations	treeFrac	1 km × 1 km	Ref [41]

Appendix bibliography

1. Díaz, S. *et al.* The global spectrum of plant form and function: enhanced species-level trait dataset. *Sci. Data* **9**, 755 (2022).
2. Sabatini, F. M. *et al.* sPlotOpen-An environmentally balanced, open-access, global dataset of vegetation plots. *Glob. Ecol. Biogeogr.* **30**, 1740–1764 (2021).
3. Hijmans, R. J. *et al.* Package 'raster'. *R package* vol. 734 473 (2015).
4. R Core Team, R. F. for S. C. R: A Language and Environment for Statistical Computing. (2022).
5. Díaz, S. *et al.* The global spectrum of plant form and function. *Nature* **529**, 167–171 (2016).
6. Van Bodegom, P. M., Douma, J. C. & Verheijen, L. M. A fully traits-based approach to modeling global vegetation distribution. *Proc. Natl. Acad. Sci.* **111**, 13733–13738 (2014).
7. Madani, N. *et al.* Future global productivity will be affected by plant trait response to climate. *Sci. Rep.* **8**, 1–10 (2018).
8. Moreno-Martínez, Á. *et al.* A methodology to derive global maps of leaf traits using remote sensing and climate data. *Remote Sens. Environ.* **218**, 69–88 (2018).
9. Boonman, C. C. F. *et al.* Assessing the reliability of predicted plant trait distributions at the global scale. *Glob. Ecol. Biogeogr.* **29**, 1034–1051 (2020).
10. Schiller, C., Schmidtlein, S., Boonman, C., Moreno-Martínez, A. & Kattenborn, T. Deep learning and citizen science enable automated plant trait predictions from photographs. *Sci. Rep.* **11**, 16395 (2021).
11. Dong, N., Dechant, B., Wang, H., Wright, I. J. & Prentice, I. C. Global leaf-trait mapping based on optimality theory. *Glob. Ecol. Biogeogr.* **32**, 1152–1162 (2023).
12. Cai, L. *et al.* Global models and predictions of plant diversity based on advanced machine learning techniques. *New Phytol.* **237**, 1432–1445 (2023).
13. Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10 (1992).

14. Rafiqpoor, D., Kier, G. & Kreft, H. Global centers of vascular plant diversity. *Nov. Acta Leopoldina NF* **92**, 61–83 (2005).

15. Kier, G. *et al.* Global patterns of plant diversity and floristic knowledge. *J. Biogeogr.* **32**, 1107–1116 (2005).

16. Kreft, H. & Jetz, W. Global patterns and determinants of vascular plant diversity. *Proc. Natl. Acad. Sci.* **104**, 5925–5930 (2007).

17. Brummitt, N., Araújo, A. C. & Harris, T. Areas of plant diversity—What do we know? *Plants, People, Planet* **3**, 33–44 (2021).

18. Sabatini, F. M. *et al.* Global patterns of vascular plant alpha diversity. *Nat. Commun.* **13**, 4683 (2022).

19. Govaerts, R., Nic Lughadha, E., Black, N., Turner, R. & Paton, A. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Sci. data* **8**, 215 (2021).

20. Weigelt, P., König, C. & Kreft, H. GIFT–A global inventory of floras and traits for macroecology and biogeography. *J. Biogeogr.* **47**, 16–43 (2020).

21. Steinig, S.; Abe-Ouchi, A.; de Boer, A.M.; Chan, W.-L.; Donnadieu, Y.; Hutchinson, D.; Knorr, G.; Ladant, J.-B.; Morozova, P.; Niezgodzki, I.; Poulsen, C.J.; Volodin, E.M.; Zhang, Z.; Zhu, J.; Lunt, D. J. Deep-Time Model Intercomparison Project (DeepMIP) Eocene model data version 1.0. *NERC EDS Cent. Environ. Data Anal.* (2024) doi:10.5285/95aa41439d564756950f89921b6ef215.

22. Hurrell, J. W. *et al.* The community earth system model: a framework for collaborative research. *Bull. Am. Meteorol. Soc.* **94**, 1339–1360 (2013).

23. Lunt, D. J. *et al.* DeepMIP: Model intercomparison of early Eocene climatic optimum (EECO) large-scale climate features and comparison with proxy data. *Clim. Past* **17**, 203–227 (2021).

24. Delworth, T. L. *et al.* GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Clim.* **19**, 643–674 (2006).

25. Bentsen, M. *et al.* The Norwegian Earth System Model, NorESM1-M–Part 1: description and basic evaluation of the physical climate. *Geosci. Model Dev.* **6**, 687–720 (2013).

26. Kageyama, M. *et al.* PMIP4-CMIP6: the contribution of the Paleoclimate Modelling Intercomparison Project to CMIP6. *Geosci. Model Dev. Discuss.* **11**, 1033–1057 (2018).

27. Danabasoglu, G. *et al.* The community earth system model version 2 (CESM2). *J. Adv. Model. Earth Syst.* **12**, e2019MS001916 (2020).

28. Kelley, M. *et al.* GISS-E2. 1: Configurations and climatology. *J. Adv. Model. Earth Syst.* **12**, e2019MS002025 (2020).

29. Miller, R. L. *et al.* CMIP6 historical simulations (1850–2014) with GISS-E2. 1. *J. Adv. Model. Earth Syst.* **13**, e2019MS002034 (2021).

30. Andrews, M. B. *et al.* Historical simulations with HadGEM3-GC3. 1 for CMIP6. *J. Adv. Model. Earth Syst.* **12**, e2019MS001995 (2020).

31. Semmler, T. *et al.* Simulations for CMIP6 with the AWI climate model AWI-CM-1-1. *J. Adv. Model. Earth Syst.* **12**, e2019MS002009 (2020).

32. Mauritsen, T. *et al.* Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1. 2) and its response to increasing CO2. *J. Adv. Model. Earth Syst.* **11**, 998–1038 (2019).

33. Yukimoto, S. *et al.* The Meteorological Research Institute Earth System Model version 2.0, MRI-ESM2. 0: Description and basic evaluation of the physical component. *J. Meteorol. Soc. Japan. Ser. II* **97**, 931–965 (2019).

34. Cao, J. *et al.* The NUIST Earth System Model (NESM) version 3: description and preliminary evaluation. *Geosci. Model Dev.* **11**, 2975–2993 (2018).

35. Seland, Ø. *et al.* Overview of the Norwegian Earth System Model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations. *Geosci. Model Dev.* **13**, 6165–6200 (2020).

36. Eyring, V. *et al.* Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).

37. Bi, D. *et al.* Configuration and spin-up of ACCESS-CM2, the new generation Australian community climate and earth system simulator coupled model. *J. South. Hemisph. Earth Syst. Sci.* **70**, 225–251 (2020).

38. Tatebe, H. *et al.* Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geosci. Model Dev.* **12**, 2727–2765 (2019).

39. University of East Anglia Climatic Research Unit; Harris, I.C.; Jones, P.D.; Osborn, T. CRU TS4.09: Climatic Research Unit (CRU) Time-Series (TS) version 4.09 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901- Dec. 2024). *NERC EDS Cent. Environ. Data Anal.* (2025).

40. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).

41. Roebroek, C. T. J. *et al.* Potential tree cover under current and future climate scenarios. *Sci. Data* **12**, 564 (2025).