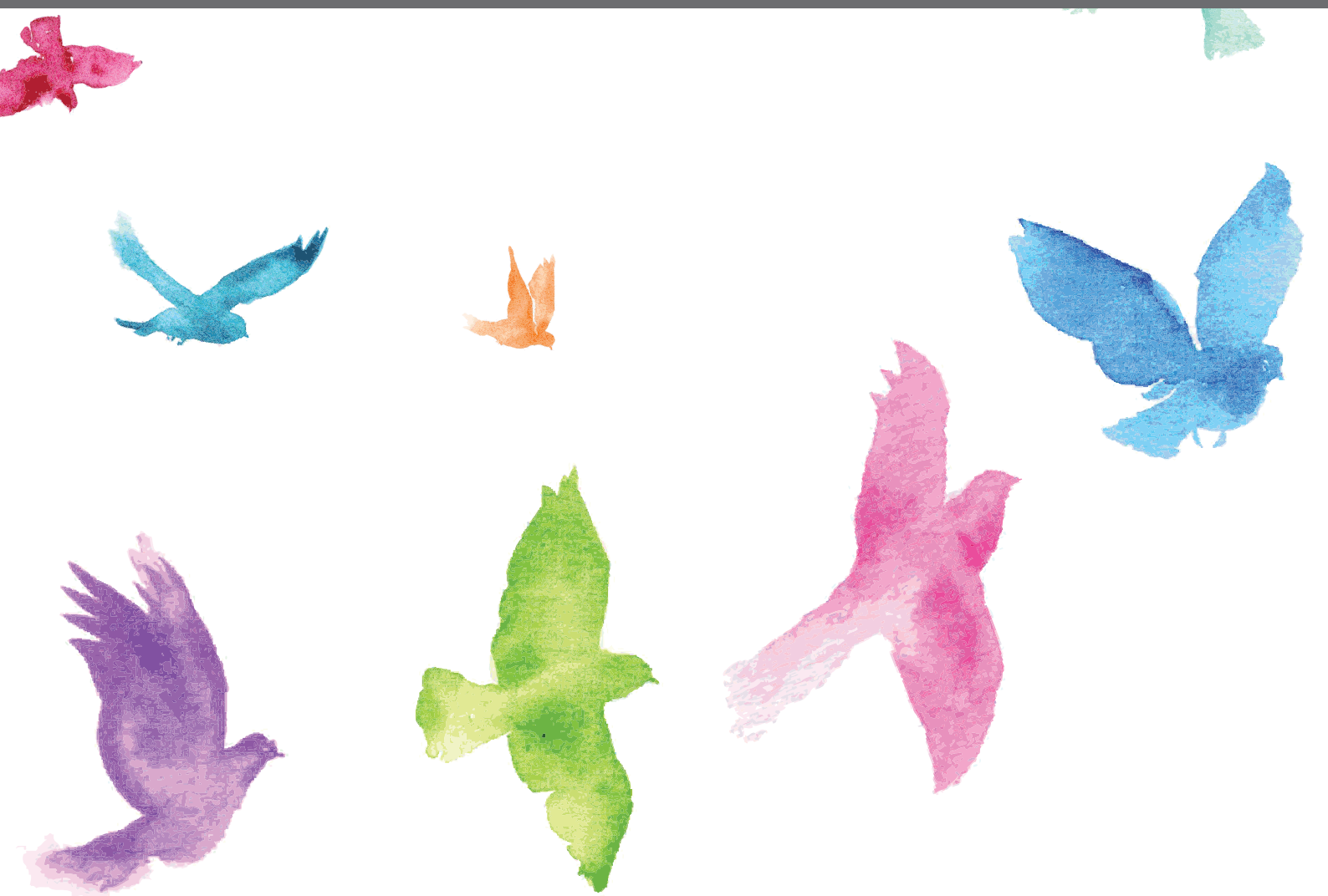
TEMPORAL AND LARGE-SCALE SPATIAL PATTERNS OF PLANT DIVERSITY AND DIVERSIFICATION

EDITED BY: Dimitar Dimitrov, Danilo M. Neves and Xiaoting Xu
PUBLISHED IN: Frontiers in Ecology and Evolution



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88976-336-8
DOI 10.3389/978-2-88976-336-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

TEMPORAL AND LARGE-SCALE SPATIAL PATTERNS OF PLANT DIVERSITY AND DIVERSIFICATION

Topic Editors:
Dimitar Dimitrov, University Museum of Bergen, Norway
Danilo M. Neves, Federal University of Minas Gerais, Brazil
Xiaoting Xu, Sichuan University, China

Citation: Dimitrov, D., Neves, D. M., Xu, X., eds. (2022). Temporal and Large-Scale Spatial Patterns of Plant Diversity and Diversification. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-336-8

Table of Contents

05 ***Editorial: Temporal and Large-Scale Spatial Patterns of Plant Diversity and Diversification***
Dimitar Dimitrov, Danilo M. Neves and Xiaoting Xu

08 ***Chemical Similarity of Co-occurring Trees Decreases With Precipitation and Temperature in North American Forests***
Brian E. Sedio, Marko J. Spasojevic, Jonathan A. Myers, S. Joseph Wright, Maria D. Person, Hamssika Chandrasekaran, Jack H. Dwenger, María Laura Prechi, Christian A. López, David N. Allen, Kristina J. Anderson-Teixeira, Jennifer L. Baltzer, Norman A. Bourg, Buck T. Castillo, Nicola J. Day, Emily Dewald-Wang, Christopher W. Dick, Timothy Y. James, Jordan G. Kueneman, Joseph LaManna, James A. Lutz, Ian R. McGregor, Sean M. McMahon, Geoffrey G. Parker, John D. Parker and John H. Vandermeer

26 ***Genome Skimming Reveals Widespread Hybridization in a Neotropical Flowering Plant Radiation***
Oriane Loiseau, Talita Mota Machado, Margot Paris, Darina Koubínová, Kyle G. Dexter, Leonardo M. Versieux, Christian Lexer and Nicolas Salamin

39 ***Geographic Pattern of Bryophyte Species Richness in China: The Influence of Environment and Evolutionary History***
Xiaotong Song, Wenzhuo Fang, Xiulian Chi, Xiaoming Shao and Qinggang Wang

48 ***Generality and Shifts in Leaf Trait Relationships Between Alpine Aquatic and Terrestrial Herbaceous Plants on the Tibetan Plateau***
Lei Yang, Haocun Zhao, Zhenjun Zuo, Xiangyan Li, Dan Yu and Zhong Wang

62 ***Evolutionary Diversity Peaks at Mid-Elevations Along an Amazon-to-Andes Elevation Gradient***
Andy R. Griffiths, Miles R. Silman, William Farfan-Rios, Kenneth J. Feeley, Karina García Cabrera, Patrick Meir, Norma Salinas, Ricardo A. Segovia and Kyle G. Dexter

72 ***Evolution of Dispersal, Habit, and Pollination in Africa Pushed Apocynaceae Diversification After the Eocene-Oligocene Climate Transition***
Cássia Bitencourt, Nicolai M. Nürk, Alessandro Rapini, Mark Fishbein, André O. Simões, David J. Middleton, Ulrich Meve, Mary E. Endress and Sigrid Liede-Schumann

91 ***Prediction of Potentially Suitable Distributions of* Codonopsis pilosula *in China Based on an Optimized MaxEnt Model***
Huyong Yan, Jiao He, Xiaochuan Xu, Xinyu Yao, Guoyin Wang, Lianggui Tang, Lei Feng, Limin Zou, Xiaolong Gu, Yingfei Qu and Linfa Qu

108 ***Defining Biologically Meaningful Biomes Through Floristic, Functional, and Phylogenetic Data***
Domingos Cardoso, Peter W. Moonlight, Gustavo Ramos, Graeme Oatley, Christopher Dudley, Edeline Gagnon, Luciano Paganucci de Queiroz, R. Toby Pennington and Tiina E. Särkinen

124 ***The Origins and Historical Assembly of the Brazilian Caatinga Seasonally Dry Tropical Forests***
Moabe F. Fernandes, Domingos Cardoso, R. Toby Pennington and Luciano P. de Queiroz

frontiers | Frontiers in Ecology and Evolution

EDITORIAL
published: 18 May 2022
doi: 10.3389/fevo.2022.894234

Editorial: Temporal and Large-Scale Spatial Patterns of Plant Diversity and Diversification

Dimitar Dimitrov[1]*, *Danilo M. Neves*[2]* and *Xiaoting Xu*[3]*

[1] Department of Natural History, University Museum of Bergen, University of Bergen, Bergen, Norway, [2] Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Brazil, [3] Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, China

Keywords: species richness, environmental gradients, traits, macroecology, macroevolution

Editorial on the Research Topic

Temporal and Large-Scale Spatial Patterns of Plant Diversity and Diversification

PLANT DIVERSITY AND DIVERSIFICATION

Plants have successfully colonized almost all of the Earth ecosystems and are among the most diverse eukaryote phyla (Mora et al., 2011; Pimm and Joppa, 2015). They also play a key role in the Earth system as a major actor in atmospheric, water and nutrient cycles that make possible life as we know it (Lucas, 2001; Payne et al., 2020). Numerous studies focused on plant diversity and evolution have been published and a large body of knowledge on plant diversity, evolutionary history, traits, physiology, and distribution has been generated. Such accumulation of knowledge combined with methodological and computational advances, and the ever-expanding biological databases (e.g., GenBank, TRY, GBIF) have allowed the study of plant diversity at large geographical and temporal scales as exemplified on **Figure 1** (Zanne et al., 2014; Smith and Brown, 2018; Igea and Tanentzap, 2020; Neves et al., 2021). However, despite numerous advances, plant diversity and diversification patterns remain contentions. For example, there is still no consensus about the crown age of angiosperms (flowering plants, the most diverse plant lineage) (e.g., Barba-Montoya et al., 2018; Sauquet et al., 2021). Many species remain undescribed and a large fraction of the described species have never been included in phylogenetic datasets (Stevens, 2017). Likewise, large gaps remain in distributional and trait databases (Cornwell et al., 2019).

The aim of this Research Topic is to bring together research that furthers our understanding of large-scale plant diversity and diversification patterns. It includes nine papers that tackle these issues from different evolutionary and ecological perspectives using diverse methods and study systems.

EVOLUTIONARY AND ECOLOGICAL DRIVERS OF PLANT DIVERSITY PATTERNS

Most of the articles in this Research Topic focus on the drivers of plant diversity and distribution patterns. Loiseau et al., using genome skimming data of the hyperdiverse genus *Vriesea* (Bromeliaceae), show that hybridization have likely played a major role in the diversification of this genus. Their results provide further evidence for the potential importance of hybridization for the diversification of Bromeliaceae.

Dispersal abilities and phenotypic adaptations are essential for the success of taxa in changing environments. Bitencourt et al. study the evolution of dispersal, habit, and pollination in Apocynaceae. They find that evolution of wind-dispersed seeds, climbing growth form, and pollinia have jointly shaped the diversification and diversity patterns of this family across Africa.

OPEN ACCESS

Edited and Reviewed by:
Jairo Patiño, Spanish National Research Council (CSIC), Spain

***Correspondence:**
Dimitar Dimitrov
dimitard.gwu@gmail.com
Danilo M. Neves
dneves@icb.ufmg.br
Xiaoting Xu
xiaotingxu@scu.edu.cn

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *11 March 2022*
Accepted: *04 May 2022*
Published: *18 May 2022*

Citation:
Dimitrov D, Neves DM and Xu X (2022) Editorial: Temporal and Large-Scale Spatial Patterns of Plant Diversity and Diversification. Front. Ecol. Evol. 10:894234. doi: 10.3389/fevo.2022.894234

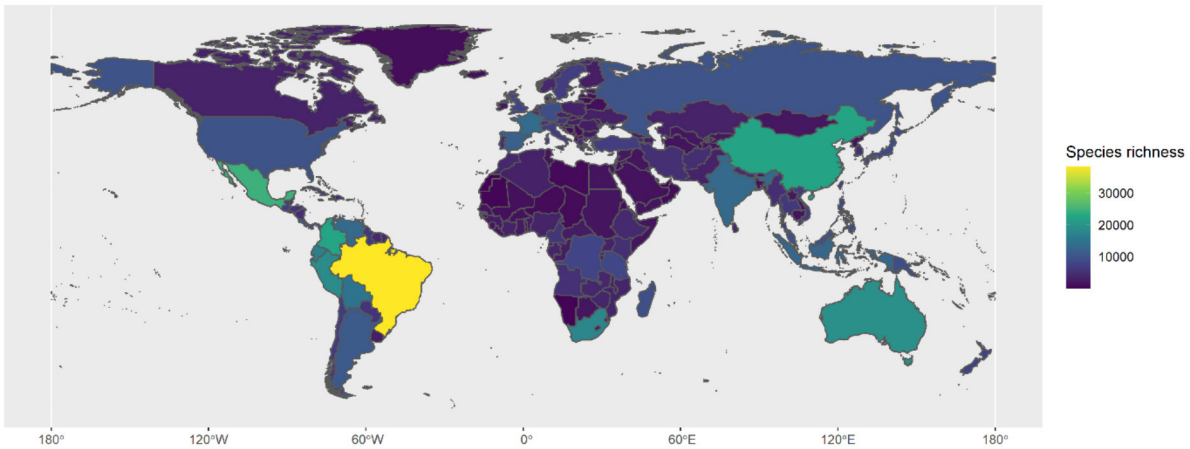


FIGURE 1 | Global patterns of terrestrial plant species richness per country (adapted from data available in BIEN—Botanical Information and Ecology Network; Maitner et al., 2018).

Fernandes et al. take a different approach and study the processes that have shaped the flora in a geographical region. They show that most lineages restricted to the Brazilian Caatinga seasonally dry forest emerged only recently during the Miocene when the region became drier. They also discuss the importance of local diversification processes driven by regional-scale geological history and local adaptations.

Diversity patterns along elevational gradients have attracted biologists since the time of Humboldt and Bonpland (1805). Griffiths et al. use plot inventory data and published phylogenies to investigate how evolutionary diversity varies along an elevation gradient. Similarly to taxonomic diversity (Rahbek, 1995), they find a mid-elevation pick in evolutionary diversity and they interpret this as a support for the Environmental Crossroads Hypothesis (Neves et al., 2020).

Song et al. study the patterns of species richness of bryophytes in China and find that the tropical and subtropical montane regions in China are both centers of diversification and refugia for bryophyte species. Different functional traits allow plant species to establish in diverse environments and co-exists in complex communities.

Yang et al. contribution investigates the patterns of leaf traits relationships in the Tibetan plateau, and the importance of particular traits for the adaptation of plants to aquatic environments.

Sedio et al. study chemical similarity among co-occurring trees and based on the comparison of more than 13,000 leaf metabolites conclude that plant metabolomes play an important role in community assembly. However, their importance is not homogenous along environmental gradients and is higher in wetter and warmer climates.

DEFINING BIOMES AND PREDICTING FUTURE DISTRIBUTIONS

Defining biomes and predicting suitable areas for species are central for conservation planning in the context of ongoing global warming.

Cardoso et al. present a new approach to biome delimitation that incorporates floristic, functional, and phylogenetic data and environmentally trained species distribution models. Their comprehensive analyses support a new, biologically meaningful delimitation of terrestrial biomes in eastern Brazil.

Global warming may cause major changes in the areas suitable for the growth and cultivation of plant species. Yan et al. estimate the areas that would be potentially suitable for *Codonopsis pilosula* (a medicinal plant from China) under global warming and suggest which regions may be optimal to cultivate this species.

CONCLUDING REMARKS

The articles in this Research Topic shed light on different aspects of plant diversity and diversification patterns in different regions of the world, across environmental gradients or trough evolutionary time. This Research Topic has highlighted the potential of integrative approaches that combine evolutionary, ecological, environmental, and geological data. It has also opened new questions and has provided new tools for future studies, and we hope that readers will find inspiration in the research presented here.

AUTHOR CONTRIBUTIONS

DD, DMN and XX wrote the article together and all authors approved the submitted version.

FUNDING

DMN acknowledges support from the US NSF (DEB-1556651) and Instituto Serrapilheira/Brazil (Serra-1912-32082). XX acknowledges support from the National Science Foundation of China (31770566).

REFERENCES

Barba-Montoya, J., Reis, M., dos Schneider, H., Donoghue, P. C. J., and Yang, Z. (2018). Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution. *New Phytol.* 218, 819–834. doi: 10.1111/nph.15011

Cornwell, W. K., Pearse, W. D., Dalrymple, R. L., and Zanne, A. E. (2019). What we (don't) know about global plant diversity. *Ecography* 42, 1819–1831. doi: 10.1111/ecog.04481

Humboldt, A., and Bonpland, A. (1805). *Essai sur la géographie des plantes*. Paris: Schoell et Cie.

Igea, J., and Tanentzap, A. J. (2020). Angiosperm speciation cools down in the tropics. *Ecol. Lett.* 23, 692–700. doi: 10.1111/ele.13476

Lucas, Y. (2001). The role of plants in controlling rates and products of weathering: importance of biological pumping. *Annu. Rev. Earth Planet. Sci.* 29, 135–163. doi: 10.1146/annurev.earth.29.1.135

Maitner, B. S., Boyle, B., Casler, N., Condit, R., Donoghue, I. I., J., et al. (2018). The bien r package: a tool to access the Botanical Information and Ecology Network (BIEN) database. *Methods Ecol. Evol.* 9, 373–379. doi: 10.1111/2041-210X.12861

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., and Worm, B. (2011). How many species are there on earth and in the ocean? *PLOS Biol.* 9, e1001127. doi: 10.1371/journal.pbio.1001127

Neves, D. M., Dexter, K. G., Baker, T. R., Coelho de Souza, F., Oliveira-Filho, A. T., Queiroz, L. P., et al. (2020). Evolutionary diversity in tropical tree communities peaks at intermediate precipitation. *Sci. Rep.* 10, 1188. doi: 10.1038/s41598-019-55621-w

Neves, D. M., Kerkhoff, A. J., Echeverría-Londoño, S., Merow, C., Morueta-Holme, N., Peet, R. K., et al. (2021). The adaptive challenge of extreme conditions shapes evolutionary diversity of plant assemblages at continental scales. *Proc. Natl. Acad. Sci.* 118, e2021132118. doi: 10.1073/pnas.2021132118

Payne, J. L., Bachan, A., Heim, N. A., Hull, P. M., and Knope, M. L. (2020). The evolution of complex life and the stabilization of the Earth system. *Interface Focus* 10, 20190106. doi: 10.1098/rsfs.2019.0106

Pimm, S. L., and Joppa, L. N. (2015). How many plant species are there, where are they, and at what rate are they going extinct? *Ann. Mol. Bot. Gard.* 100, 170–176. doi: 10.3417/2012018

Rahbek, C. (1995). The elevational gradient of species richness: a uniform pattern? *Ecography* 18, 200–205

Sauquet, H., Ramírez-Barahona, S., and Magallón, S. (2021). The age of flowering plants is unknown. *EcoEvoRxiv*. doi: 10.32942/osf.io/n4v6b

Smith, S. A., and Brown, J. W. (2018). Constructing a broadly inclusive seed plant phylogeny. *Am. J. Bot.* 105, 302–314. doi: 10.1002/ajb2.1019

Stevens, P. F. (2017). *Angiosperm Phylogeny Website. Version 14.* Available online at: http://www.mobot.org/MOBOT/research/APweb/ (accessed November 13, 2015).

Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. A., FitzJohn, R. G., et al. (2014). Three keys to the radiation of angiosperms into freezing environments. *Nature* 506, 89–92. doi: 10.1038/nature12872

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 26 May 2021
doi: 10.3389/fevo.2021.679638

Chemical Similarity of Co-occurring Trees Decreases With Precipitation and Temperature in North American Forests

Brian E. Sedio[1,2], Marko J. Spasojevic[3], Jonathan A. Myers[4], S. Joseph Wright[2], Maria D. Person[5], Hamssika Chandrasekaran[5], Jack H. Dwenger[1], María Laura Prechi[2], Christian A. López[2], David N. Allen[6], Kristina J. Anderson-Teixeira[7], Jennifer L. Baltzer[8], Norman A. Bourg[7], Buck T. Castillo[9], Nicola J. Day[10], Emily Dewald-Wang[11], Christopher W. Dick[2,9], Timothy Y. James[9], Jordan G. Kueneman[2], Joseph LaManna[12], James A. Lutz[13], Ian R. McGregor[7], Sean M. McMahon[14], Geoffrey G. Parker[14], John D. Parker[14] and John H. Vandermeer[9]*

[1] Department of Integrative Biology, University of Texas at Austin, Austin, TX, United States, [2] Smithsonian Tropical Research Institute, Panama City, Panama, [3] Department of Evolution, Ecology, and Organismal Biology, University of California, Riverside, Riverside, CA, United States, [4] Department of Biology, Washington University in St. Louis, St. Louis, MO, United States, [5] Proteomics Facility, University of Texas at Austin, Austin, TX, United States, [6] Department of Biology, Middlebury College, Middlebury, VT, United States, [7] Conservation Ecology Center, Smithsonian Conservation Biology Institute, Front Royal, VA, United States, [8] Department of Biology, Wilfrid Laurier University, Waterloo, ON, Canada, [9] Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, United States, [10] School of Biological Sciences, Victoria University of Wellington, Wellington, New Zealand, [11] Department of Integrative Biology, University of California, Berkeley, Berkeley, CA, United States, [12] Department of Biological Sciences, Marquette University, Milwaukee, WI, United States, [13] Department of Wildland Resources, Utah State University, Logan, UT, United States, [14] Forest Ecology Group, Smithsonian Environmental Research Center, Edgewater, MD, United States

OPEN ACCESS

Edited by:
Danilo M. Neves, Federal University of Minas Gerais, Brazil

Reviewed by:
Emmanuel Defossez, Université de Neuchâtel, Switzerland
Jose Eduardo Meireles, University of Maine, United States

***Correspondence:**
Brian E. Sedio
sediob@utexas.edu

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *12 March 2021*
Accepted: *29 April 2021*
Published: *26 May 2021*

Citation:
Sedio BE, Spasojevic MJ, Myers JA, Wright SJ, Person MD, Chandrasekaran H, Dwenger JH, Prechi ML, López CA, Allen DN, Anderson-Teixeira KJ, Baltzer JL, Bourg NA, Castillo BT, Day NJ, Dewald-Wang E, Dick CW, James TY, Kueneman JG, LaManna J, Lutz JA, McGregor IR, McMahon SM, Parker GG, Parker JD and Vandermeer JH (2021) Chemical Similarity of Co-occurring Trees Decreases With Precipitation and Temperature in North American Forests. Front. Ecol. Evol. 9:679638. doi: 10.3389/fevo.2021.679638

Plant diversity varies immensely over large-scale gradients in temperature, precipitation, and seasonality at global and regional scales. This relationship may be driven in part by climatic variation in the relative importance of abiotic and biotic interactions to the diversity and composition of plant communities. In particular, biotic interactions may become stronger and more host specific with increasing precipitation and temperature, resulting in greater plant species richness in wetter and warmer environments. This hypothesis predicts that the many defensive compounds found in plants' metabolomes should increase in richness and decrease in interspecific similarity with precipitation, temperature, and plant diversity. To test this prediction, we compared patterns of chemical and morphological trait diversity of 140 woody plant species among seven temperate forests in North America representing 16.2°C variation in mean annual temperature (MAT), 2,115 mm variation in mean annual precipitation (MAP), and from 10 to 68 co-occurring species. We used untargeted metabolomics methods based on data generated with liquid chromatography-tandem mass spectrometry to identify, classify, and compare 13,480 unique foliar metabolites and to quantify the metabolomic similarity of species in each community with respect to the whole metabolome and each of five broad classes of metabolites. In addition, we compiled morphological trait data from existing databases and field surveys for three commonly measured traits (specific leaf

area [SLA], wood density, and seed mass) for comparison with foliar metabolomes. We found that chemical defense strategies and growth and allocation strategies reflected by these traits largely represented orthogonal axes of variation. In addition, functional dispersion of SLA increased with MAP, whereas functional richness of wood density and seed mass increased with MAT. In contrast, chemical similarity of co-occurring species decreased with both MAT and MAP, and metabolite richness increased with MAT. Variation in metabolite richness among communities was positively correlated with species richness, but variation in mean chemical similarity was not. Our results are consistent with the hypothesis that plant metabolomes play a more important role in community assembly in wetter and warmer climates, even at temperate latitudes, and suggest that metabolomic traits can provide unique insight to studies of trait-based community assembly.

Keywords: metabolomics, chemical ecology, ForestGEO, species diversity gradient, climate, biotic interactions, functional traits, temperate forest

INTRODUCTION

Plant diversity varies immensely over large-scale climatic gradients at regional and global scales. Climates that are wetter, warmer, and less seasonal consistently exhibit greater species richness than climates that are drier, colder, and more seasonal. At global scales, tree diversity at the 0.25 km^2-scale increases by two orders of magnitude over 60° latitude from the Canadian Taiga Plains to the Amazon Basin (Anderson-Teixeira et al., 2015; Chu et al., 2019; Davies et al., 2021). At regional scales, woody plant diversity increases with precipitation and temperature (Gentry, 1988; Hawkins et al., 2003; Kreft and Jetz, 2007; Esquivel-Muelbert et al., 2017). Prominent hypotheses for these large-scale diversity gradients propose that environmental filtering reduces community membership, and hence diversity, in abiotically stressful climates and that increased specialization of biotic interactions increases species richness in less stressful climates (Schemske et al., 2009; Lamanna et al., 2014; Chu et al., 2019). These hypotheses make testable predictions regarding relationships between climatic gradients and the traits that mediate plant interactions with the abiotic and biotic environment.

Morphological and physiological traits mediate plant interactions with the abiotic and biotic environment and define fundamental tradeoffs in resource-use and life history strategy (Wright et al., 2004; Rüger et al., 2018). Variation in community weighted mean trait values can reveal climatically driven variation in suitable morphological traits. Dispersion and richness of morphological traits may reveal diversity in resource-use and life-history strategies among species that co-occur at a site, or the effects of environmental filters that limit the breadth of trait values with which plants can tolerate a climate (Lamanna et al., 2014; Harrison et al., 2020). For decades, researchers have sought to link community variation in plant functional traits to the processes of community assembly (Weiher and Keddy, 1995), yet a focus on coarse morphological traits (e.g., specific leaf area) that are only loosely correlated with ecological functions (Shipley et al., 2016) has resulted in some cases where there are strong trait-environmental linkages, and other cases where the same traits are weakly linked to the environment, making generalizations challenging.

In addition to morphological variation, much of the functional trait variation found in plant communities is due to variation in small organic molecules that make up the metabolome of the plant (Kursar et al., 2009; Salazar et al., 2018; Sedio et al., 2020, 2017). The plant metabolome includes primary metabolites involved in core metabolic processes and the basic building blocks of large organic polymers such as nucleotides, amino acids, and mono- and disaccharides. However, much of the astonishing diversity of plant metabolomes is made up of secondary metabolites that are typically not involved in core metabolism, but rather in specialized functions. Secondary metabolites can mediate plant response to abiotic stresses, such as those exerted by ultraviolet radiation and freezing temperatures (Close and McArthur, 2002; Rasmann et al., 2014; Schneider et al., 2019), as well as biotic stresses imposed by herbivores and microbial pathogens (Coley, 1983; Wink, 2003; Fine et al., 2006; Salazar et al., 2018; Sedio et al., 2020). Unlike abiotic stressors, biotic agents of selection on plant defenses are capable of reciprocal coevolution, that is, counter-adaptations on the part of herbivores and pathogens to the chemical defenses evolved by plant hosts (Schemske et al., 2009) (Schemske et al., 2009). Ehrlich and Raven (1964) proposed that such coevolution of plants and their natural enemies is a major driver of diversification in both groups. Hence, variation over climatic gradients in the relative importance of natural enemies as agents of selection on plants may contribute to the parallel diversity gradients in woody plants and their natural enemies (Dyer et al., 2007; Schemske et al., 2009), with consequences for variation in plant metabolomes (Defossez et al., 2018; Moreira et al., 2018; Sedio et al., 2018b; Volf et al., 2020).

In light of large-scale climate-associated diversity gradients, biologists have long assumed that plant investment in chemical defenses would increase with pest and pathogen pressure, and hence increase with precipitation and temperature and decrease with seasonality (Coley and Aide, 1991; Coley and Barone, 1996; Dirzo and Boege, 2008; Pellissier et al., 2014; Sam et al., 2020).

However, at regional scales, plants tend to invest more in defense in environments in which the cost of tissue loss to natural enemies is greater (Givnish, 1999; Fine et al., 2006; Defossez et al., 2018; Volf et al., 2020). Hence, recent studies suggest that the cost of tissue loss to herbivores or pathogens runs counter to that of pest and pathogen abundance, and hence decreases with precipitation, temperature and edaphic resources (Fine et al., 2006; Sam et al., 2020; Volf et al., 2020). The fitness landscape of plant secondary metabolites likely varies with multiple biotic and abiotic gradients over broad climatic gradients at global and regional scales, and plant responses may favor greater investment in stressful abiotic environments where tissue loss is costly, yet greater diversity or interspecific divergence in chemical defense composition in less stressful abiotic environments with high herbivore and pathogen pressure (Volf et al., 2020). Few studies have examined the relationship between plant metabolomic composition and major climatic or environmental gradients for specific woody plant lineages (Fine et al., 2006; Sam et al., 2020; Volf et al., 2020) and even fewer have done so at the community scale (Sedio et al., 2018b).

Here, we compare metabolomic and morphological traits among 140 species from seven forests in the US and Canada, representing 218 unique species-site combinations. We use untargeted metabolomics methods based on data generated with liquid chromatography-tandem mass spectrometry to identify, classify, and compare foliar metabolites and to quantify the metabolomic similarity of species in each forest community. We compare metabolomic variation with that of morphological traits to ask: (i) do metabolomic and morphological traits co-vary among species or communities?; (ii) do metabolomic and morphological traits show similar patterns of phylogenetic signal across species?; and (iii) do metabolomic and morphological traits show similar associations with large-scale gradients in precipitation and temperature? Finally, we test the hypotheses that abiotic stress reduces morphological and metabolomic trait diversity but increases investment in foliar metabolites in cold and dry climates and that biotic interactions increase metabolomic diversity and decrease similarity in warm and wet climates in North American tree communities.

MATERIALS AND METHODS

Study Sites and Species

We studied the chemical and functional ecology of seven large-scale (median 20 ha) forest dynamics plots coordinated by the Smithsonian Forest Global Earth Observatory (hereafter ForestGEO) in the U.S. and Canada. In each plot, all free-standing woody stems ≥1 cm diameter at breast height (dbh, defined as 1.3 m height) are identified, measured, and mapped at regular census intervals (Anderson-Teixeira et al., 2015; Davies et al., 2021). The seven plots span 23.6° latitude, 16.2°C mean annual temperature, and 2,115 mm mean annual precipitation, and support from 10 to 68 free-standing woody species ≥1 cm dbh (**Table 1**; Davies et al., 2021).

Three western forests are characterized by predominantly coniferous overstories. Scotty Creek, Northwest Territories (NT), Canada is a subarctic boreal forest dominated by *Picea mariana* (black spruce) on permafrost plateaus and *Larix laricina* (larch) on permafrost-free wetlands, with dense patches of *Betula glandulosa* (dwarf birch) (Dearborn et al., 2020). Cedar Breaks, UT, is a subalpine spruce-fir forest dominated by *Abies bifolia* (subalpine fir), *Populus tremuloides* (quaking aspen), *Picea engelmannii* (Engelmann spruce), *Pinus flexilis* (limber pine), *Pinus longaeva* (bristlecone pine), and *Picea pungens* (blue spruce) (Furniss et al., 2017). Wind River, WA, is a moist montane forest characterized by an overstory of *Tsuga heterophylla* (western hemlock) and *Pseudotsuga menziesii* (Douglas-fir) and an understory of diverse Ericaceae (*Gaultheria*, *Menziesia*, *Rhododendron*, and several *Vaccinium*) (Lutz et al., 2013).

Four eastern forests are characterized by predominantly broad-leaved overstory trees. Tyson Research Center, MO, and Michigan Big Woods, MI, are both characterized by several species of *Quercus* (oaks) and *Carya* (hickories) in the overstory (Spasojevic et al., 2016; Allen et al., 2020, 2018), though the Michigan plot may be undergoing a shift in composition, as recruitment is dominated by *Acer rubrum* (red maple) and *Prunus serotina* (black cherry) (Allen et al., 2018). The Smithsonian Conservation Biology Institute (SCBI) forest plot represents the eastern deciduous forest of the Blue Ridge, with abundant *Liriodendron tulipifera* (tulip poplar), *Quercus* and *Carya* (Bourg et al., 2013). The Smithsonian Environmental Research Center (SERC) forest plot is composed of upland secondary forest characterized by *Liriodendron tulipifera*, *Quercus*, *Carya*, *Fagus grandifolia* (beech), and *Liquidambar styraciflua* (sweetgum) in the overstory, but is bisected by a floodplain forest distinguished by a canopy of *Fraxinus* (ashes), *Platanus occidentalis* (sycamore), and *Ulmus americana* (elm).

Sampling and Extraction

We collected expanding, unlignified leaves that were between 50 and 90% fully expanded from saplings, shrubs, and small trees in the forest understory where possible at each site between June 2018 and July 2019. Three to five leaves from an individual were placed in a manila coin envelope and immediately flash frozen in liquid nitrogen in the field using a portable dewar (PrincetonCryo, Pipersville, PA, United States). Where possible, we sampled 5 individuals of each species recorded in each forest plot. In total, we sampled 140 species from the seven forest plots, representing 218 unique species-site combinations (**Supplementary Table 1**). In each of the seven plots, the species we sampled represented ≥64.0% of the species (median 91.1%) and ≥98.8% of the individuals (median 99.9%) recorded.

Leaf samples were stored at −80°C and shipped on dry ice to the Sedio Lab at the University of Texas at Austin, where they were freeze-dried and a 50–100-mg sample was pulverized using a Qiagen TissueLyser ball mill (Hilden, Germany). A 10-mg subsample was weighed and extracted with 1.8 ml 90:10 v/v methanol:water pH 5 overnight at 4°C and 300 rpm, centrifuged at 14,000 rpm for 30 min, and the supernatant removed and filtered for analysis using LC-MS.

Liquid Chromatography-Mass Spectrometry

We optimized UHPLC-MS parameters to detect and fragment metabolites representing a wide range in polarity and mass (Sedio et al., 2018a). Metabolomic extracts were separated using a Thermo Fisher Scientific (Waltham, MA, United States) Vanquish Horizon Duo ultra-high performance liquid chromatography (UHPLC) system with an Accucore C18 column with 150 mm length, 2.1 mm internal diameter, and 2.6-μm particle size. UHPLC buffer A (0.1% v/v formic acid in water) and buffer B (0.1% v/v formic acid in methanol) were employed in a solvent gradient from 5 to 100% buffer B over 18 min.

Separation of metabolites by UHPLC was followed by heated electrospray ionization (HESI) in positive mode using full scan MS1 and data-dependent acquisition of MS2 (dd-MS2) on a Thermo Fisher Scientific Q Exactive hybrid quadrupole-orbitrap mass spectrometer. We analyzed samples of three types: individual trees; species pools, consisting of pooled aliquots of all conspecific individuals for each species; and quality controls (QC), consisting of pools of aliquots of all species pools analyzed concurrently. For individuals, we collected an MS1 scan (115–1,725 *m/z*) at a resolution of 140,000. For species pools and QC samples, the MS1 full scan was at 70,000 resolution, followed by dd-MS2 at 17,500 resolution on the five most abundant precursors found in the MS1 spectrum. Automatic gain control target values were 1e6 for full scan MS1 and 1e5 for dd-MS2. Maximum ion injection times were 200 ms for full scan MS1, 100 ms for QC MS1, and 50 ms for MS2. For dd-MS2, we set the isolation window to 1.5 m/z and stepped collision energy at 20, 40, and 60. QC pooled samples were used to account for fluctuations in total ion intensity due to changes in temperature and atmospheric pressure over time.

Metabolomics Data Analysis

Raw data from the UHPLC-MS analyses were centroided and processed for peak detection, peak alignment, and peak filtering using MZmine2 (Pluskal et al., 2010), which groups chromatographic features into putative compounds based on molecular mass and LC retention time. We used the following parameter settings: mass detector method "centroid" with MS1 noise threshold 50,000 ion counts and MS2 noise threshold 0 ion counts; chromatogram builder module minimum highest intensity 35,000, absolute *m/z* tolerance 0.002 and ppm tolerance 10.0; deconvolution module method "local minimum search" with chromatographic threshold 0.9, minimum relative height 0.05, and peak duration range 0.0–0.3 min; isotope grouper module using absolute *m/z* tolerance 0.0015, ppm tolerance 7.0, retention-time tolerance 0.5; join aligner module using absolute *m/z* tolerance 0.0015, ppm tolerance 7.0, retention-time tolerance 0.5, weight for *m/z* vs. retention time 3:1, isotope absolute *m/z* tolerance 0.001, ppm tolerance 5.0, minimum absolute intensity 3,000, minimum score 0.6; Sirius export module merge mode "do not merge," absolute *m/z* tolerance 0.001, ppm tolerance 20.0; GNPS export module filter rows "only with MS2."

MZmine2 processing was performed for 11 batches of species. Each batch was uploaded to the Global Natural Products Social (GNPS) Molecular Networking platform (Wang et al., 2016) and used to generate a molecular network using the "feature-based molecular networking" method (Nothias et al., 2020) using the following parameter settings: precursor ion mass tolerance 0.02 Da, fragment ion mass tolerance 0.2 Da, minimum pairs cosine 0.7, minimum matched fragment ions 6, maximum shift between precursors 500 Da, network topK 10, maximum connected component size 100.

Each batch was then re-analyzed in GNPS using the "qemistree" tool to simultaneously infer molecular structures using Sirius (Dührkop et al., 2019) and CSI:FingerID (Dührkop et al., 2015), classify metabolites using ClassyFire (Djoumbou Feunang et al., 2016) and build a hierarchical dendrogram that reflects the structural similarity of unique metabolites using Qemistree (Tripathi et al., 2021). The output of the 11 batch analyses was integrated using the Qemistree plugin for Qiime (Bolyen et al., 2019; Tripathi et al., 2021). Compounds that occurred in blanks, such as contaminants and industrial surfactants, were removed from further analyses. The complete, combined foliar metabolome and Qemistree dendrogram can be found in **Supplementary Tables 2**, **3**, respectively. The complete dataset is available as a MassIVE dataset on GNPS[1]. Links to the Feature-Based Molecular Network and Qemistree results on GNPS for each of the 11 batches can be found in **Supplementary Table 4**.

ClassyFire's chemotaxonomic scheme places heterocyclic alkaloids in the superclass "organoheterocyclic compounds" rather than "alkaloids and derivatives." Hence, to reorganize the alkaloids and reduce the number of chemical classes for downstream analyses, we modified the chemotaxonomic classes provided by ClassyFire to derive our own classes as follows:

[1]ftp://massive.ucsd.edu/MSV000087217/;doi:10.25345/C5HV49

TABLE 1 | Variation in latitude, mean annual temperature, mean annual precipitation, census area, and species richness of seven forest plots.

Plot, location	Lat (°N)	MAT (°C)	MAP (mm)	Elevation (m)	Area (ha)	S
Scotty Creek, NT, Canada	61.3	−2.7	380	266	9.6	10
Wind River, WA, United States	45.8	9.2	2,495	369	25	25
Michigan Big Woods, MI, United States	42.5	8.4	792	288	23	41
SCBI, VA, United States	38.9	12.9	1,001	306	25.6	65
SERC, MD, United States	38.9	13.2	1,068	8	16	68
Tyson Research Center, MO, United States	38.5	13.5	957	205	20	42
Cedar Breaks, UT, United States	37.7	4.0	849	3,084	15.32	17

First, we defined "lipids" as the ClassyFire superclass "lipids and lipid-like molecules," including terpenes and steroids. Second, we defined "organic acids" as the ClassyFire superclass "organic acids and derivatives." Third, we defined "benzenoids" as all compounds in the ClassyFire superclasses "benzenoids" and "phenylpropanoids and polyketides," including phenols, polyphenols, and flavonoids. Fourth, we defined "alkaloids" as all compounds in the ClassyFire superclass "alkaloids and derivatives," as well as compounds in the superclass "organoheterocyclic compounds" that contain nitrogen. Finally, we defined "heterocyclic compounds" as all remaining compounds in the ClassyFire superclass "organoheterocyclic compounds." The classifications we used can be found in the column "customclass" in **Supplementary Table 2**. Chemical superclasses and relative ion intensity of compounds that occurred in each forest plot were visualized on the Qemistree hierarchical dendrogram (Tripathi et al., 2021) using the Interactive Tree of Life web tool (**Figure 1**; Letunic and Bork, 2019). We did not define "lignans, neolignans, and related compounds" as a distinct superclass in downstream calculations because they were absent from many species in the young, expanding leaves we analyzed. Other chemical superclasses we included only in the whole-metabolome analyses included "allenes," "carbides," "hydrocarbon derivatives," "hydrocarbons," "nucleosides, nucleotide, and analogs," "organic 1,3-dipolar compounds," "organic oxygen compounds," "organophosphorus compounds," "organosulfur compounds," and unclassified compounds. All five of the superclasses we examined include numerous defensive compounds, as common defenses such as polyphenols and flavonoids are included in the "benzenoids" superclass; terpenes, terpenoids, and steroids are included in the "lipids" superclass, and other anti-herbivore and antimicrobial compounds are included in the "heterocyclic compounds"and "organic acids" superclasses.

Sedio et al. (2017) developed the chemical structural-compositional similarity (CSCS) metric to quantify the similarity of samples or species in a manner that accounts for the presence of identical compounds as well as the presence of structurally similar unique compounds in two samples. Here, we calculated CSCS for each species pair as the mean structural similarity of every pair of compounds found in either species, weighted by the ion intensity of each compound pair in the species compared. We calculated CSCS chemical similarity for every species pair in each forest plot for the whole metabolome and for each of the five major chemical classes.

Morphological Traits

To compare patterns in metabolomic trait diversity with patterns of morphological trait diversity we focused on three putatively important morphological traits: specific leaf area (SLA), wood density, and seed mass. Specific leaf area is associated with resource uptake strategy and tissue N, where high SLA represents a strategy to maximize carbon gain and relative growth rate (Reich et al., 1997). Wood density is associated with mortality rate, hydraulic lift and the relative mechanical strength of the plant (Enquist et al., 1999). Seed mass is related to dispersal ability and a reproductive strategy where species that produce few large seeds are thought to better tolerate poor site conditions and those that produce many small seeds are thought to be better dispersers (Muller-Landau, 2010).

Between 2012 and 2014 we measured morphological traits at three of our seven sites: Wind River, Tyson and SERC. At these three sites we collected sun-exposed leaves with minimal damage or senescence from each of 5 to 10 representative individuals of each species and fully hydrated the leaves in water tubes. For all species except conifers, we collected 3 leaves per individual. Due to the small size of conifer leaves (needles) we collected approximately 0.5 g of leaves (between 40 and 70 needles) per individual per species. We calculated leaf area (cm^2) from scanned leaves and petioles using Image-J (Rasband, 2018). For compound species, we calculated leaf area as the mean leaflet area per leaf including petiolules (Pérez-Harguindeguy et al., 2016). For needle leaves we calculated the total area of all leaves and divided that area by the number of leaves collected. We calculated SLA (cm^2/g) as leaf area per unit dry mass after leaves were dried in an oven at 60°C for 4 days. Using the branches from which leaves were collected we cut out a section that was 2.5 cm long and at least 1 cm in diameter. We calculated wood density (g/cm^3) as the volume of the branch section per unit of dry mass after branch sections were dried in an oven at 60°C for 4 days. Seed mass data were compiled from the Kew Royal Botanical Gardens Seed Information Database[2]. For the other four sites morphological trait data was either downloaded from existing databases (Harrison et al., 2020; Kattge et al., 2020) or we used species means from other sites (i.e., mean trait values for *Fraxinus americana* at Tyson were applied to Michigan Big Woods as this species is found at both sites). Trait values used in the present study can be found in **Supplementary Table 5**.

Statistical Analysis

To examine patterns of covariance among metabolomic and morphological traits and describe variation in species traits among sites we conducted a principal components analysis (PCA). Total ion abundance of all compounds in each superclass was summed for each species-by-site and all traits were scaled prior to analysis. Principal components analysis was conducted using the "principal" function in R version 3.6.2 (R Core Team, 2020). Species scores and trait loadings can be found in **Supplementary Tables 5**, **6**, respectively. For comparison, we conducted a phylogenetic PCA using the R package phytools (Revell, 2012). However, this necessitated excluding species with missing values.

To generate a phylogeny representing all 140 species in the seven forest plots, we queried the Zanne et al. (2014) megaphylogeny using "phylomatic" (Webb and Donoghue, 2005). We then used phylogenetically independent contrasts (PICs) to examine the relationship between chemical similarity and phylogeny, and to test for phylogenetic signal. We calculated the PIC for each node in the phylogeny as the mean CSCS for all pairs of species for which the node represents the most recent common ancestor. Sedio et al. (2018b) refer to this metric as CSCSmrca. Here, we calculated 1–CSCSmrca

[2]http://data.kew.org/sid/

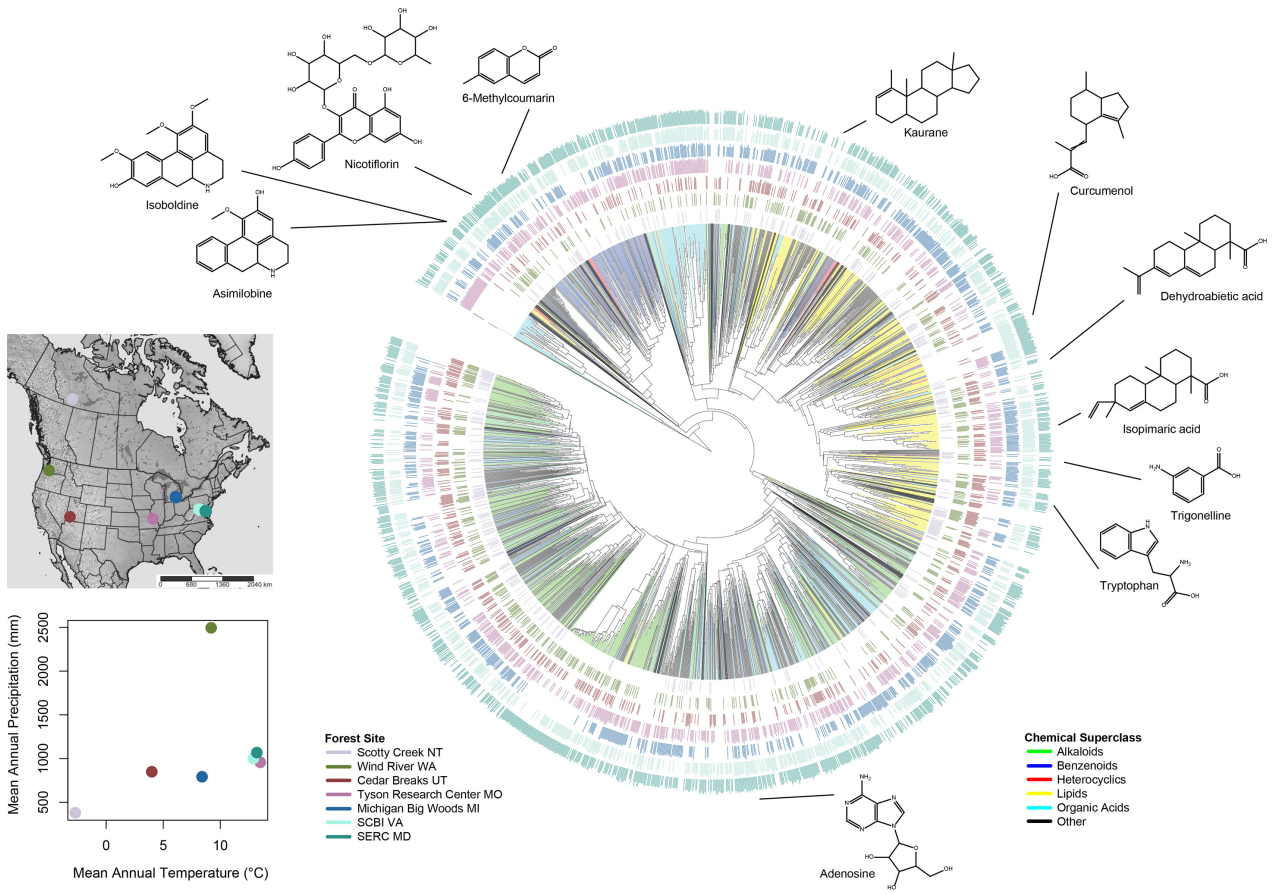


FIGURE 1 | Hierarchical dendrogram of 13,480 foliar metabolites. Qemistree (Tripathi et al., 2021) represents the structural similarity of unique metabolites in the form of a "phylogeny," a hierarchical dendrogram in which structurally related metabolites form sister pairs and clades. Branch color indicates the chemotaxonomic classification of the metabolite in one of five major chemical superclasses; black branches represent all other chemical classes. Colored interior branches represent branches for which all "descendent" branches were classified as the same superclass. Seven concentric bar plots represent the relative log ion intensity of metabolites in each of seven forest plots. The geographic location of each forest plot, mean annual precipitation (mm), and mean annual temperature (°C) are indicated in the inset.

to represent metabolomic disparity for ease of comparison with morphological disparity, which we calculated as the difference between the trait values of the descendent nodes or terminal taxa for each node in the phylogeny (i.e., the phylogenetic contrast in trait values). To evaluate phylogenetic signal, we regressed metabolomic phylogenetic contrasts (1-CSCSmrca) and morphological phylogenetic contrasts against log-transformed phylogenetic distance. Tests of phylogenetic signal have little statistical power when applied to samples with few taxa (Revell, 2010), hence we assessed phylogenetic signal for metabolic and morphological traits for 140 species, rather than for each community. We corroborated our CSCSmrca method by measuring Blomberg's K (Blomberg et al., 2003) for the three morphological traits.

To examine the differences in interspecific variation and the chemical space occupied by species in each forest plot and in taxonomic clades, we used non-metric multidimensional scaling (NMDS) to reduce the chemical space into two dimensions using the MASS package in R (Ripley et al., 2013).

For each plot, we calculated complementary metrics of functional trait diversity that are appropriate for metabolomic and morphological traits, respectively. For the metabolomic traits we calculated the abundance-weighted mean CSCS among co-occurring species ($CSCS_{mean}$). First, we calculated the mean CSCS for each species to all other species, weighted by abundance. We then calculated the mean of these species-level similarity scores, weighted by abundance. We also considered metabolomic richness, which we defined as the number of compounds detected in each forest community.

For morphological traits we calculated three complementary metrics of functional diversity: Community weighted mean (CWM) trait values, functional richness (FRic), and functional dispersion (FDis). Community-weighted mean (CWM) trait values describe the functional composition of a community and are calculated as the abundance-weighted trait averages (Garnier et al., 2004) of each trait. Functional richness measures the ranges of trait values in a community, and is calculated as the standardized difference between the maximum and minimum

functional values present in the community (Mason et al., 2005). Functional dispersion, which indicates the degree of trait dissimilarity among species, is calculated as the abundance-weighted mean distance of each species in trait space to the centroid of all species and is statistically independent of species richness (Laliberté and Legendre, 2010). We repeated metabolomic and morphological community calculations using species basal area to weight the CSCS, CWM, and FDis scores and found similar patterns as the abundance weighted patterns, so we focus only on abundance weighted results for simplicity.

To assess relationships among chemical and morphological traits of woody plant assemblages and climatic variables, we regressed plot-level measures of the diversity and similarity of traits with log-transformed mean annual precipitation ($\log_{10}$MAP), mean annual temperature (MAT), and species richness. We reported significant relationships for $p < 0.05$ and marginally significant relationships for $p < 0.07$.

RESULTS

We detected 13,480 compounds, ranging from 116.0705 to 1181.8350 Daltons (Da), in foliar extracts from all 140 species. We generated a predicted molecular structure and a chemotaxonomic classification for 11,415 and 10,376 of these compounds, respectively. A total of 8,323 classified compounds were represented by the superclasses "alkaloids," "benzenoids," "lipids," "organic acids," and "heterocyclic compounds" as we defined them (62%; **Figure 1** and **Supplementary Table 2**). Among compounds classified into one of these five superclasses, we generated a predicted molecular structure for 8,323 compounds (99.6%), but retrieved an MS2 library match from GNPS spectral libraries for only 1,032 compounds (12.3%).

We found that the five metabolomic superclasses and the three morphological traits we considered represented largely orthogonal axes of variation in the PCA (**Figure 2** and **Supplementary Table 6**), but there was some covariation among metabolomic and morphological traits (particularly on PC3). Alkaloids, lipids, and organic acids loaded positively on the first PCA axis, which explained 26% of the variation among species. SLA and wood density loaded on PCA axis 2, which explained 16% of the variation among species. Lastly, heterocyclic compounds, seed mass and benzenoids loaded on PCA axis 3, which explained another 16% of the variation among species (**Supplementary Table 6**). Phylogenetic PCA carried out on a subset of species for which all traits were measured or available was broadly concordant with non-phylogenetic PCA, as metabolomic and morphological traits represented orthogonal axes of variation (**Supplementary Figure 1**).

Phylogenetic signal varied among metabolomic and morphological traits. Metabolomic disparity for all foliar metabolites increased with log phylogenetic distance (**Figure 3A**). Likewise, disparity increased with log phylogenetic distance when measured in terms of alkaloids, benzenoids, lipids, and SLA (**Figures 3B,C,E,G**). In contrast, metabolomic or morphological disparity was unrelated to phylogeny for heterocyclic compounds, organic acids, seed mass, and wood density (**Figures 3D,F,H,I**). Blomberg's K and test for phylogenetic signal for morphological traits corroborated our approach, as SLA exhibited weak, but significant phylogenetic signal ($K = 0.086$, $p = 0.002$), whereas seed mass ($K = 0.034$, $p = 0.573$) and wood density ($K = 0.024$, $p = 0.213$) did not.

The NMDS ordination illustrates the chemical space represented by the 218 unique species-by-site combinations (**Figures 4**, **5**), as well as highlighted clades (**Figure 4**) and each forest community (**Figure 5**). Clades varied considerably in the breadth of the chemical space they occupied. Gymnosperms occupied a consistent neighborhood in NMDS chemical space, with the exception of the chemically very distinct *Taxus brevifolia* (**Figure 4A**). In contrast, Betulaceae and Ericaceae occupied broad areas of chemical space, reflecting the low metabolomic similarity among species in those families (**Figures 4C,D**).

Abundance-weighted metabolomic similarity of co-occurring species decreased with MAP among the seven forest plots (**Figure 6A**). Results for chemical superclasses largely reflected those for the whole metabolome (**Supplementary Figure 2**). Species were less similar with respect to alkaloids in forest plots with greater MAP (**Supplementary Figure 2A**). However, co-occurring species were only marginally less similar with respect to heterocyclic compounds and organic acids in forests with greater MAP (**Supplementary Figure 2C,E**), and similarity with respect to benzenoids and lipids was not related to precipitation among the seven sites (**Figures 6B,D**).

Abundance-weighted metabolomic similarity of co-occurring species also decreased with MAT among the seven forest plots (**Figure 6B**). Likewise, similarity with respect to all five chemical superclasses decreased with MAT among the forest sites (**Supplementary Figures 2F–J**). In contrast, abundance-weighted metabolomic similarity was not related to species richness (**Figure 6C**), though similarity with respect to benzenoids and lipids decreased with species richness among the sites (**Supplementary Figures 2L,N**).

Metabolomic richness was unrelated to MAP among the seven sites (**Figure 6D**). Likewise, richness of compounds in all five chemical superclasses was unrelated to MAP (**Supplementary Figures 3A–E**). In contrast with MAP, metabolomic richness increased significantly with MAT among the seven forest plots (**Figure 6E**). Alkaloid richness and that of benzenoids also increased with MAT (**Supplementary Figures 3F,G**), while the increase in richness of heterocyclic compounds and organic acids was marginally significant with MAT among sites (**Supplementary Figures 3H,J**). Only lipid richness was unrelated to variation in MAT (**Supplementary Figure 3I**). Metabolic richness as a whole (**Figure 6F**), and with respect to all five chemical superclasses (**Supplementary Figures 3K–O**), increased significantly with species richness among forest plots.

Morphological trait dispersion with respect to SLA increased significantly with MAP among the seven forest plots (**Figure 7A**). However, dispersion with respect to seed mass and wood density was unrelated to MAP (**Figures 7B,C**). Morphological trait dispersion was unrelated to MAT for all three morphological traits (**Figures 7D–F**).

Morphological trait richness was unrelated to MAP for all three morphological traits (**Figures 8A–C**). Trait richness defined

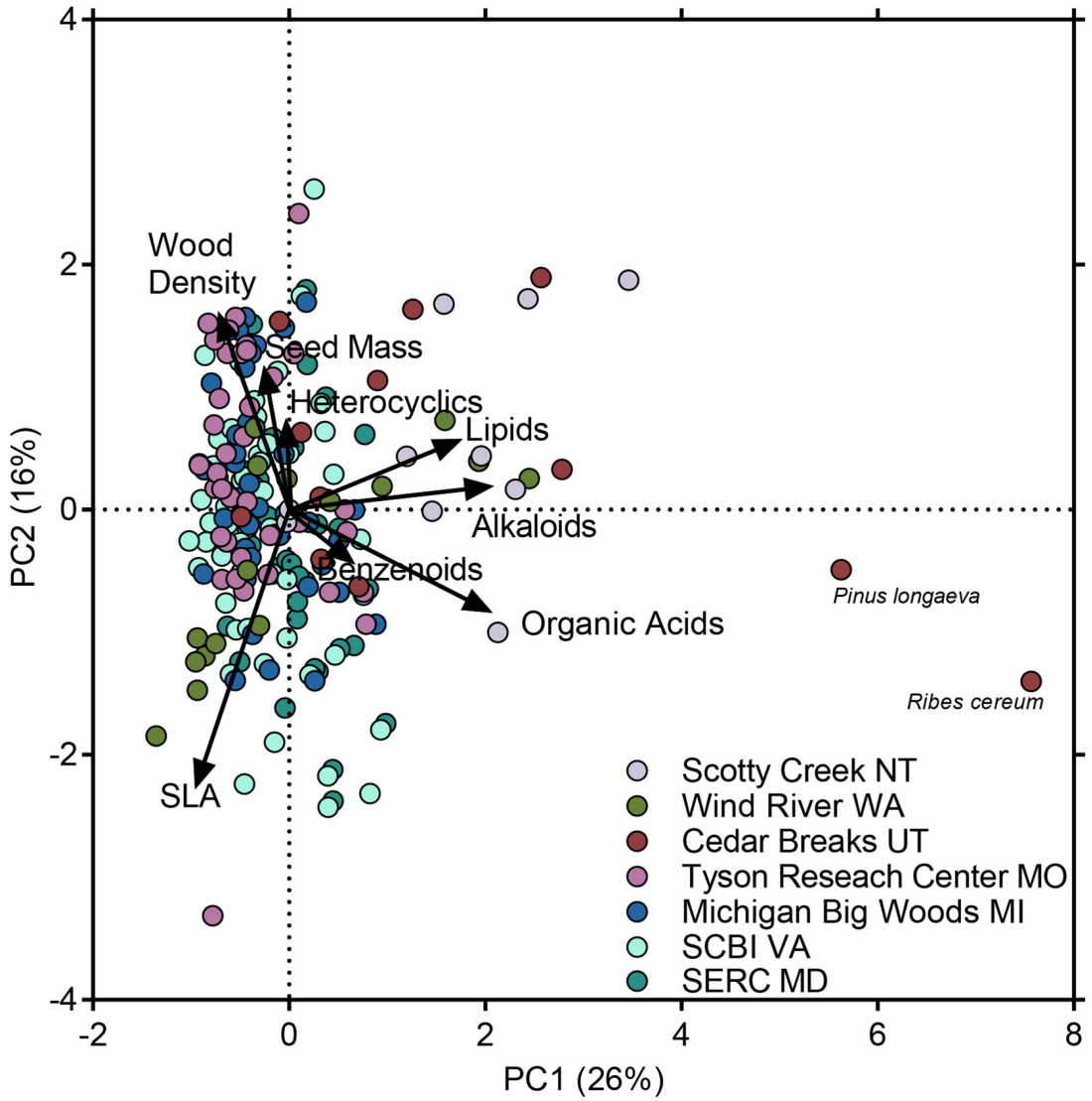


FIGURE 2 | Principal components analysis of variation among species in three morphological traits and in total ion intensity in five chemical superclasses. Loadings for chemical superclasses reflect species means in total ion intensity of all compounds classified in each superclass. Loadings for morphological traits reflect species means. Points represent 140 unique species; colors represent species occurrences in seven forest plots. Species that occurred in >1 forest plot contributed only a single observation (the species mean) to the principal components analysis, but are visually represented by jittered points, the colors of which reflect the species incidence in each forest plot. See **Supplementary Table 5** for species scores and **Supplementary Table 6** for trait loadings.

in terms of SLA was unrelated to MAT (**Figure 8D**). However, seed mass richness increased marginally significantly with MAT (**Figure 8E**), and wood density richness increased significantly with MAT (**Figure 8F**).

DISCUSSION

A Comparison of Metabolomic and Morphological Trait Variation

The seven forest plots represented a wide range of variation in precipitation, temperature and species richness within the temperate and boreal zones (**Table 1**; Anderson-Teixeira et al., 2015). Accordingly, the functional strategies of species present at each of these seven sites varied along these major gradients. This was reflected in variation in both the morphological traits, and in quantitative investment in broad chemical superclasses, which represent largely orthogonal axes of variation (**Figure 2**). Alkaloids, lipids, and organic acids loaded on the first PCA axis. All three superclasses include numerous anti-herbivore and antimicrobial defense compounds. These include terpenes, terpenoids, and steroids classified among the lipids. The lipid superclass also contains fatty acids that make up oils and waxes that are important mediators of tolerance to abiotic stresses such as drought and freezing temperatures (Uemura and Steponkus, 1994). Interestingly,

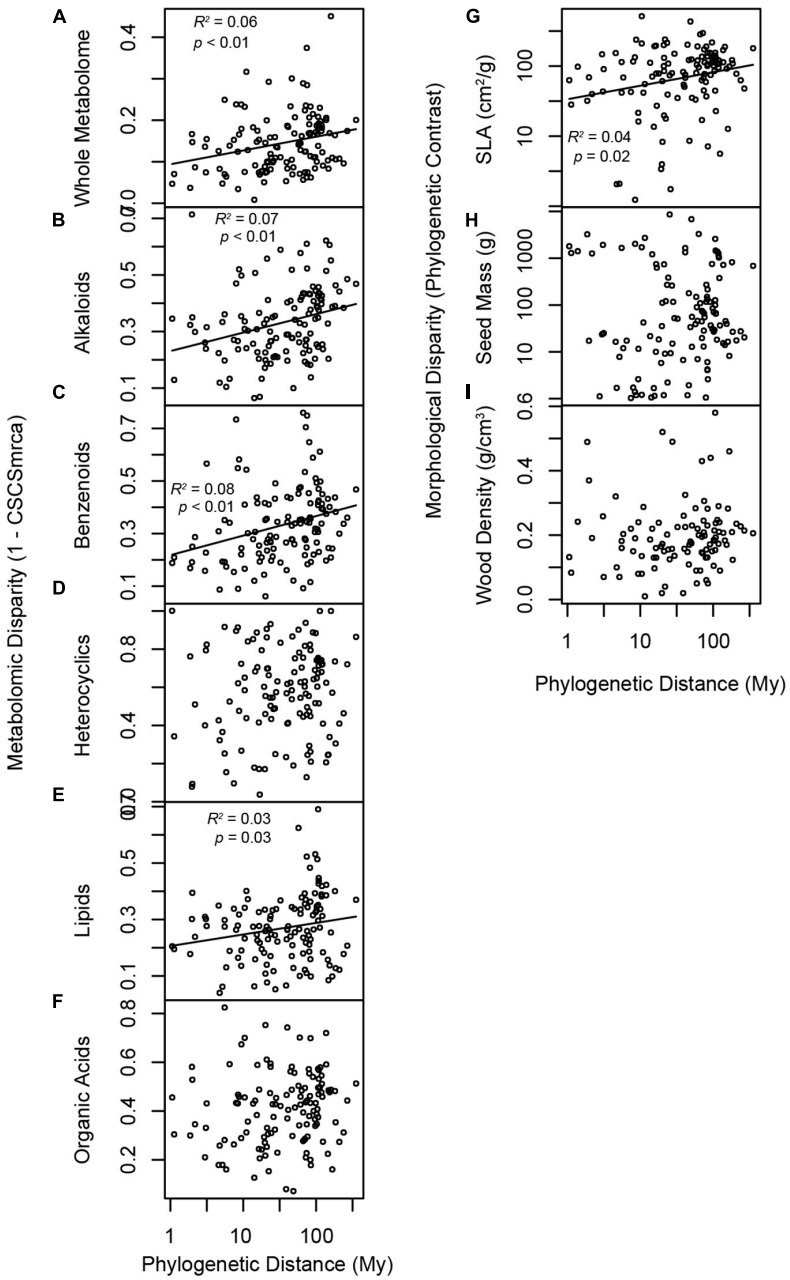


FIGURE 3 | Phylogenetic signal in the species pool and morphological traits. Panels illustrate relationships between phylogenetic distance and species disparity with respect to **(A)** the whole metabolome, **(B)** alkaloids, **(C)** benzenoids, **(D)** heterocyclic compounds, **(E)** lipids, **(F)** organic acids, **(G)** SLA, **(H)** seed mass, and **(I)** wood density. Metabolomic disparity is 1-CSCSmrca, where CSCSmrca is the mean chemical similarity of species for which a phylogenetic node represents the most recent common ancestor. Morphological disparity is the phylogenetic contrast in trait values at a phylogenetic node. Traits that evolve according to Brownian motion, or phenotypic drift after cladogenesis permits departure from a shared ancestral state, are expected to exhibit increasing disparity with log-transformed phylogenetic distance.

two key morphological traits, SLA and wood density, both loaded on PCA axis 2 (**Figure 2**) representing an uncorrelated orthogonal axis to the metabolomic traits on PC1. Specific leaf area is typically associated with resource uptake strategy, where high SLA represents a strategy to maximize carbon gain and relative growth rate (Reich et al., 1997), while wood density is associated with mortality rate, hydraulic lift and the relative mechanical strength of the plant (Enquist et al., 1999). These patterns of orthogonal chemical and morphological trait axes suggest that there is little covariation between chemical defense strategies represented by investment in alkaloids, lipids, and organic acids and growth and resource

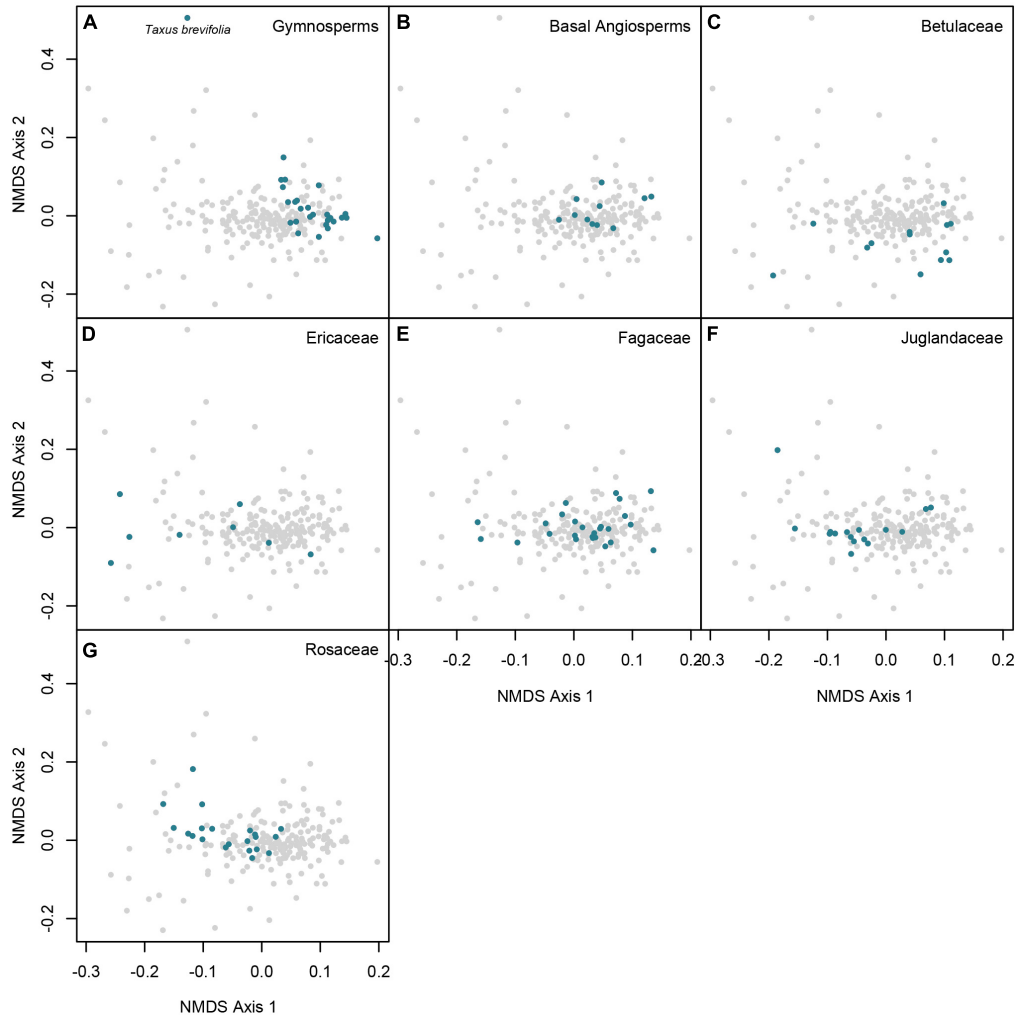


FIGURE 4 | Non-metric multidimensional scaling of pairwise CSCS chemical similarity for 218 populations of 140 woody plant species. Panels highlight species that are **(A)** gymnosperms, **(B)** basal Angiosperms, **(C)** Betulaceae, **(D)** Ericaceae, **(E)** Fagaceae, **(F)** Juglandaceae, and **(G)** Rosaceae. Each point represents a species-site population, and the distances between points reflect the pairwise CSCS chemical similarity between all pairs of species-site populations, represented in two dimensions. Colors represent the taxonomic family or clade of the species.

allocation strategies represented by SLA and wood density, and hence that metabolomic traits represent novel functional space that is not captured by commonly measured morphological traits (Díaz et al., 2016). However, other morphological traits we did not consider here may yet exhibit correlations with metabolomic features.

Benzenoids, heterocyclic compounds, and seed mass all loaded on the third PCA axes (**Figure 2**), with benzenoids and heterocyclic compounds positively covarying with each other and both negatively covarying with seed mass. The chemical superclass benzenoids includes defensive compounds such as polyphenols, as well as secondary metabolites that serve multiple functions, such as flavonoids that function in both defense from herbivory and protection against ultraviolet light (Close and McArthur, 2002; Schneider et al., 2019). Heterocyclic compounds, as we have defined them, include many compounds with potential defensive functions that lack the characteristic functional groups or metabolic origins used to define the other superclasses. Seed mass is generally associated with recruitment life history strategy, where large-seeded species tolerate suboptimal conditions for recruitment, but small-seeded species achieve greater fecundity (Muller-Landau, 2010). The association between seed mass, and quantitative investment in benzenoids and heterocyclics is likely driven by the abundance of broad-leafed, deciduous, large-seeded angiosperm trees, such as *Carya* and *Quercus*, in the four eastern forests, and the greater investment in benzenoid metabolites and heterocyclic compounds in more environmentally and climatically stressful boreal and alpine forests.

In general, we found that phylogenetic signal varied among metabolomic and morphological traits, supporting research demonstrating that phylogenetic signal is dependent on both

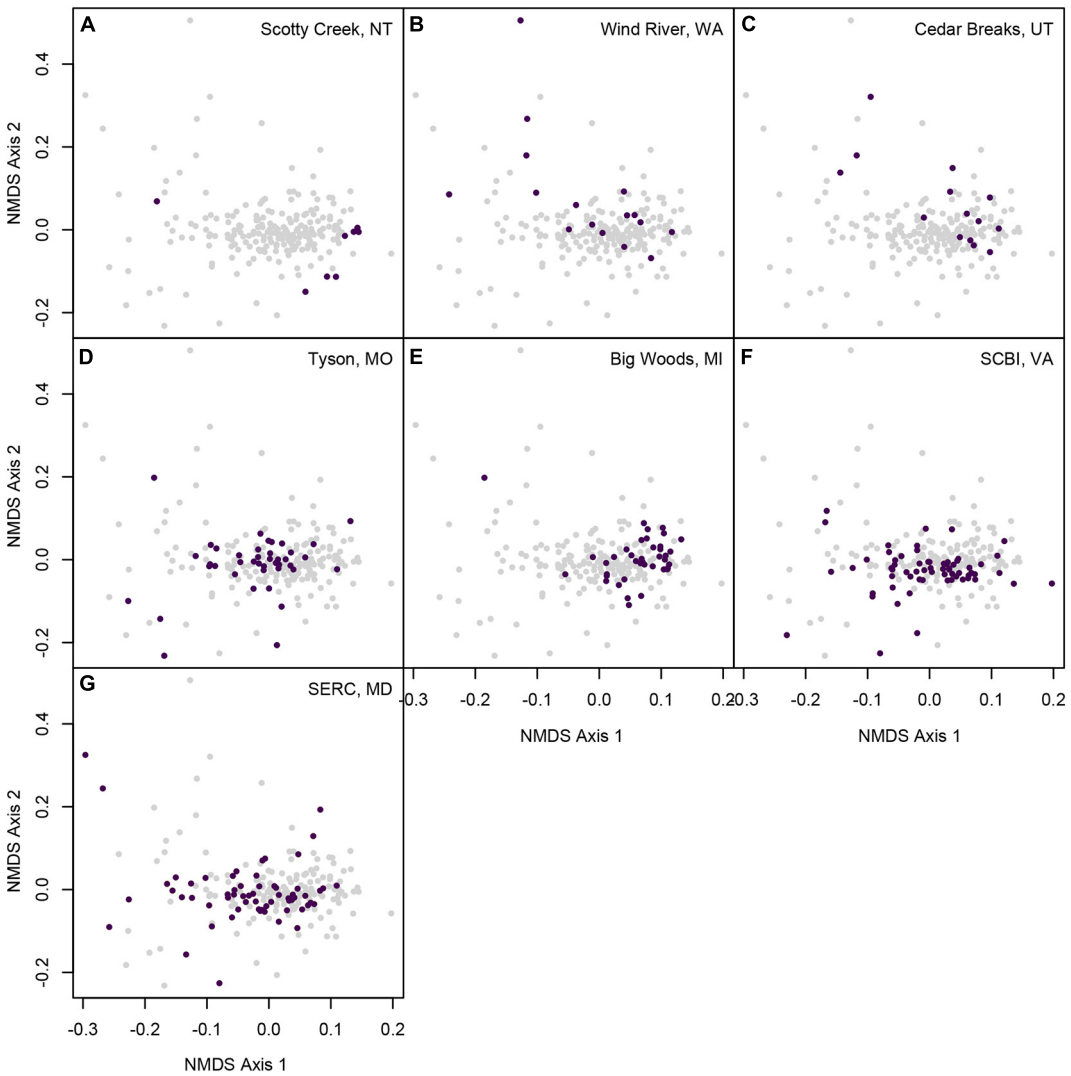


FIGURE 5 | Non-metric multidimensional scaling of pairwise CSCS chemical similarity for 218 populations of 140 woody plant species. Panels illustrate the chemical similarity of species that occur at **(A)** Scotty Creek, NT, **(B)** Wind River, WA, **(C)** Cedar Breaks, UT, **(D)** Tyson Research Center, MO, **(E)** Michigan Big Woods, Michigan, **(F)** SCBI, VA, and **(G)** SERC, MD. Each point represents a species-site population, and the distances between points reflect the pairwise CSCS chemical similarity between all pairs of species-site populations, represented in two dimensions. Colors represent the forest plot in which the species population occurred.

the taxonomic and spatial scales studied (Cavender-Bares et al., 2006). While wood density (Swenson and Enquist, 2007), SLA (Ackerly and Reich, 1999), and seed mass (Moles et al., 2007) are all generally phylogenetically conserved when examining broad scale phylogenetic patterns involving thousands of species, wood density and seed mass did not exhibit phylogenetic signal among the 140 species in our study. Interestingly, these results suggest that phylogenetic signal for wood density and seed mass may be weaker than for the foliar secondary metabolome at the community scale in North America. However, the relative phylogenetic signal displayed by morphological and metabolomic traits in temperate forests may be reversed in tropical forests, which exhibit much less metabolomic phylogenetic signal than their temperate counterparts (Sedio et al., 2018b).

Variation in the species composition of the seven forest plots resulted in striking differences in the chemical space occupied by each site. For example, the Scotty Creek community was represented by chemically similar gymnosperms and birches (Betulaceae) (**Figure 5A**) despite the much broader chemical space represented by species in those lineages at other sites (**Figures 4A,C**). In contrast, western coniferous forests at Wind River and Cedar Breaks, while relatively species poor (**Table 1**), were highly chemically diverse and composed of chemically distinct species. This was a result of two factors. First, the understory flora of these forests is composed of chemically distinctive shrubs, particularly species in the Rosaceae such as species of *Amelanchier* and *Holodiscus*, and species in the Ericaceae such as species of *Gaultheria*, *Menziesia*, *Rhododendron*, and *Vaccinium* (**Figures 4C,G**, **5**). Second,

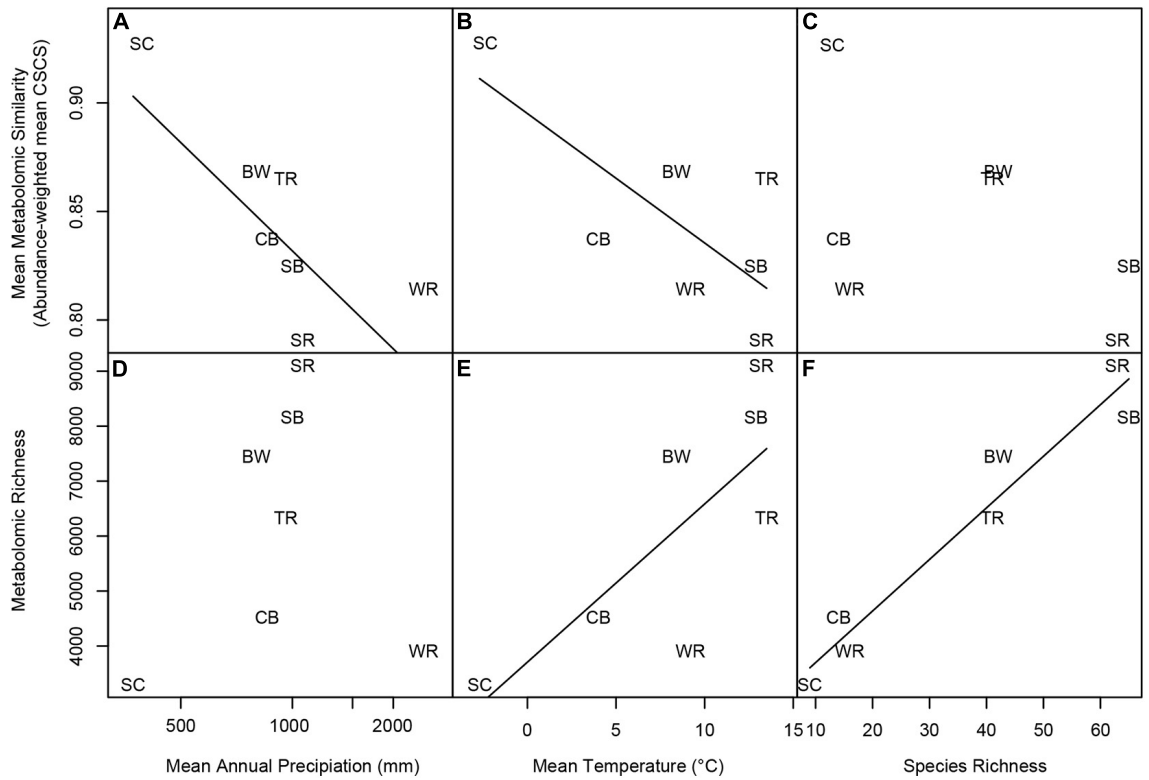


FIGURE 6 | Variation in metabolomic similarity and richness vs. $\log_{10}$ MAP, MAT, and species richness. Panels represent linear regressions between **(A)** mean metabolomic similarity and MAP, **(B)** mean metabolomic similarity and MAT, **(C)** mean metabolomic similarity and species richness, **(D)** metabolomic richness and MAP, **(E)** metabolomic richness and MAT, and **(F)** metabolomic richness and species richness. Panels a, b, and c represent regressions of abundance-weighted CSCS metabolomic similarity of co-occurring species at each site vs. $\log_{10}$ MAP, MAT, and species richness, respectively. Panels d, e, and f represent regressions of metabolite richness with $\log_{10}$ MAP, MAT, and species richness, respectively. Significant regressions are indicated by a solid line. Forest plot codes are: SC, Scotty Creek, NT; WR, Wind River, WA; CB, Cedar Breaks, UT; TR, Tyson Research Center, MO; BW, Michigan Big Woods, MI; SB, Smithsonian Conservation Biology Institute (SCBI), VA; and SR, Smithsonian Environmental Research Center (SERC), MD.

abundant overstory trees such as *Abies* and *Pinus* species were less chemically similar in the temperate rainforest at Wind River than at the drier and colder Cedar Breaks and Scotty Creek (**Figures 5B,C**, **6A,B**).

In contrast to the western North American plots, many species in the four eastern forest plots occupied the center of chemical space (**Figures 5D–G**). This pattern was driven by the species richness and relative chemical similarity of species in the families Fagaceae, Juglandaceae, and some Rosaceae (**Figures 4E–G**). However, the density of points in the center of chemical space that represent the four eastern forest plots in the NMDS may overrepresent the chemical similarity of species in those forests, as the NMDS ordination relies only on species pairwise CSCS chemical similarity and does not account for or visually represent variation in abundance.

Forest Metabolomes and Climatic Gradients

Abundance-weighted whole metabolomic similarity of co-occurring species strongly declined with log MAP and MAT among the seven forest plots (**Figures 6A,B**). This result is consistent with the prediction that chemically mediated biotic interactions are more specialized in abiotically mild, wetter and warmer climates (Mittelbach et al., 2007; Schemske et al., 2009). In contrast, the PCA (**Figure 2** and **Supplementary Table 5**) illustrates that quantitative investment in most classes of metabolites was greatest in species inhabiting abiotically stressful environments such as the boreal forest at Scotty Creek, NT, or the subalpine forest at Cedar Breaks, UT, which includes one of the most extreme stress tolerators, bristlecone pine (*Pinus longaeva*). Together, these results are consistent with the prediction that quantitative investment in chemical defense is greatest is stressful environments in which tissue loss to herbivores and pathogens is costly, but qualitative divergence in the composition of chemical defense is greatest in relatively benign environments where plant-enemy coevolution may occur at an accelerated rate (Rasmann et al., 2014; Volf et al., 2020).

Sedio et al. (2018b) found that tree species were more chemically similar in the temperate forest at SERC than in a tropical moist forest at Barro Colorado Island (BCI), Panama. Furthermore, species of *Quercus* (oaks), *Carya* (hickories), and *Viburnum* were chemically similar at SERC, whereas congeneric species in the seven most-diverse tree genera at BCI were chemically dissimilar (Sedio et al., 2018b). The metabolomic

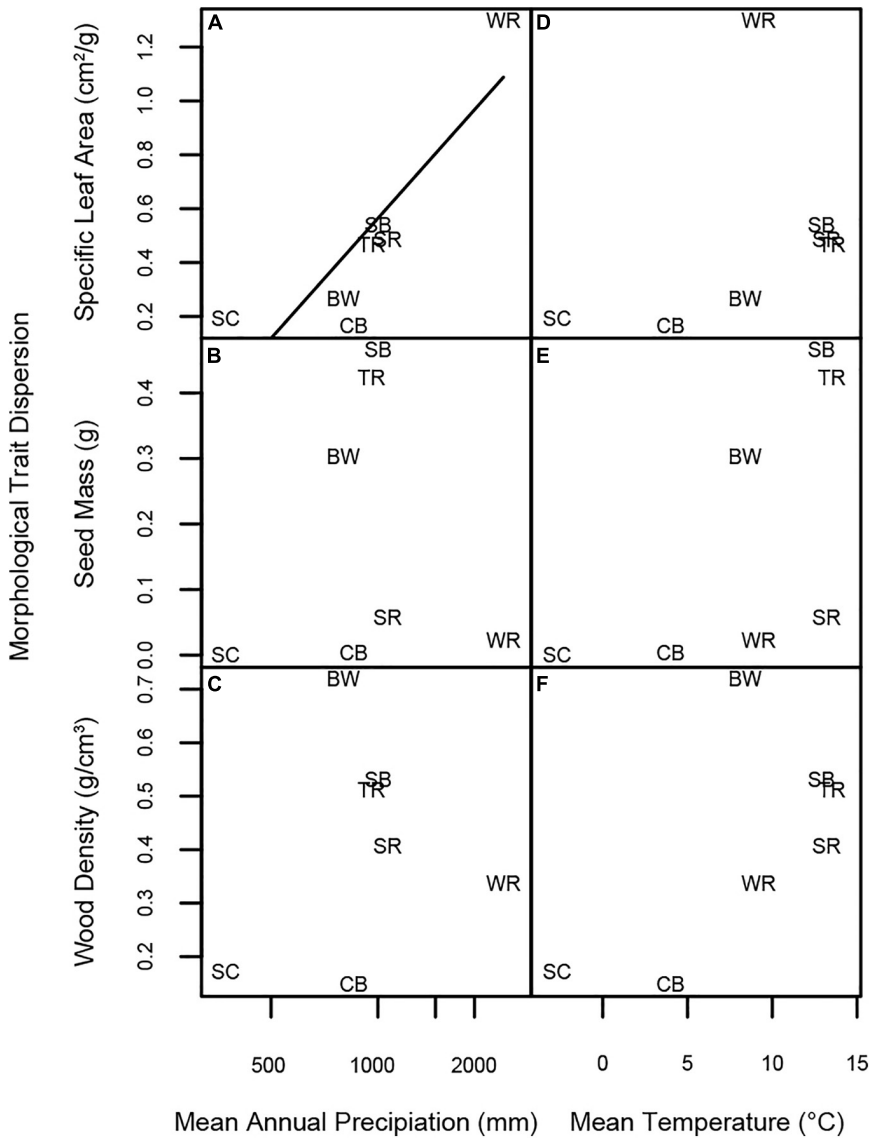


FIGURE 7 | Linear regressions of morphological trait dispersion at each forest site vs. log MAP, MAT, and species richness. Panels represent linear regressions between trait dispersion of **(A)** SLA and MAP, **(B)** seed mass and MAP, **(C)** wood density and MAP, **(D)** SLA and MAT, **(E)** seed mass and MAT, and **(F)** wood density and MAT, respectively. Significant regressions are indicated by a solid line; non-significant regressions with $p < 0.08$ are represented by a dashed line. Forest plot codes are: SC, Scotty Creek, NT; WR, Wind River, WA; CB, Cedar Breaks, UT; TR, Tyson Research Center, MO; BW, Michigan Big Woods, MI; SB, Smithsonian Conservation Biology Institute (SCBI), VA; and SR, Smithsonian Environmental Research Center (SERC), MD.

differences observed among closely related species of *Eugenia*, *Inga*, *Ocotea*, *Piper*, *Protium*, and *Psychotria* at BCI suggest that interspecific variation in foliar secondary metabolites may directly promote greater species richness by reducing host ranges of insect herbivores and microbial pathogens (Kursar et al., 2009; Salazar et al., 2018; Volf et al., 2018), and thereby reducing enemy-mediated competitive interactions and imposing stronger conspecific negative density-dependent effects on recruitment relative to the effects of heterospecific neighbors (Sedio and Ostling, 2013; Forrister et al., 2019). Whereas the results of Sedio et al. (2018b) suggest that the metabolomic similarity of co-occurring woody plant species may differ profoundly between the temperate and tropical regions, our results suggest that metabolomic similarity among co-occurring plant species decreases with precipitation and temperature even within the boreal and temperate regions of North America and hence may contribute to latitudinal and climatic diversity gradients even outside the tropics.

Future Directions

Our results indicate that metabolomic diversity and disparity increased with precipitation and temperature (**Figures 6A,G**, **7A,G**). Although these gradients are associated with increased species richness at regional and

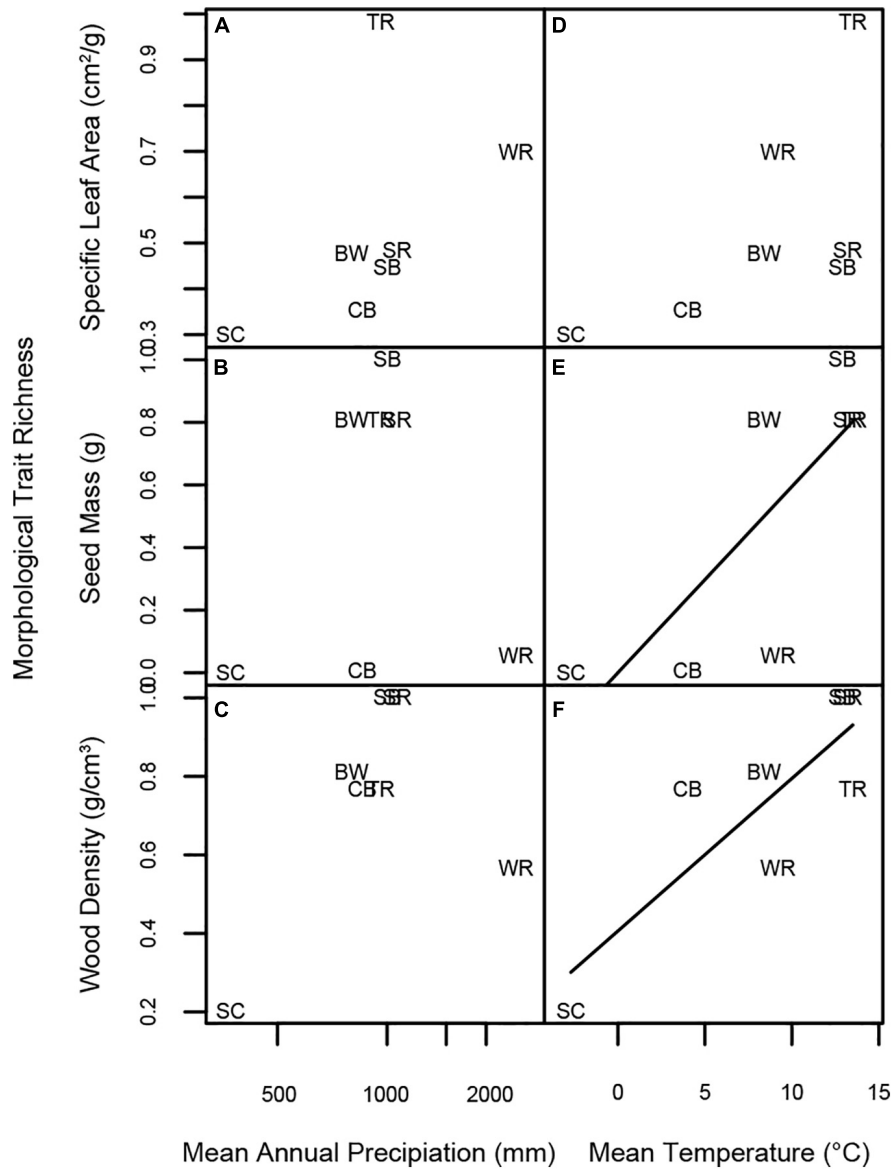


FIGURE 8 | Linear regressions of morphological trait richness at each forest site vs. log MAP, MAT, and species richness. Panels represent linear regressions between trait richness of **(A)** SLA and MAP, **(B)** seed mass and MAP, **(C)** wood density and MAP, **(D)** SLA and MAT, **(E)** seed mass and MAT, and **(F)** wood density and MAT, respectively. Significant regressions are indicated by a solid line; non-significant regressions with $p < 0.08$ are represented by a dashed line. Forest plot codes are: SC, Scotty Creek, NT; WR, Wind River, WA; CB, Cedar Breaks, UT; TR, Tyson Research Center, MO; BW, Michigan Big Woods, MI; SB, Smithsonian Conservation Biology Institute (SCBI), VA; and SR, Smithsonian Environmental Research Center (SERC), MD.

global scales, in our results metabolomic richness and disparity were not directly related to species richness among our seven forest sites (**Figures 6M**, **7M**). Linking community variation in the metabolome to mechanisms that may drive variation in species richness over ecological and evolutionary timescales will require the integration of metabolomics with the study of plant-enemy associations and plant performance at community scales over broad climatic gradients. Nevertheless, our results illustrate the scalability of metabolomics methods based on mass spectrometry (Dührkop et al., 2015, 2019; Wang et al., 2016; Nothias et al., 2020; Tripathi et al., 2021) and suggest their utility in the pursuit of chemical community ecology of global gradients in diversity and climate.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

BES, MJS, JAM, and SJW designed the research. BES, MDP, HC, JHD, CAL, and MLP developed chemical and instrumental methods. DNA, KJA-T, JLB, NAB, BTC, NJD, ED-W, CWD, TYJ, JGK, JAL, IRM, SMM, JAM, GGP, JDP, and JHV collected forest community data, identified species, and collected samples. BES, MJS, JAM, MDP, HC, and JHD collected chemical and trait data. BES and MJS analyzed the data and wrote the manuscript. All authors revised the manuscript.

FUNDING

This work was supported by a Corteva Agrisciences grant to the Smithsonian Tropical Research Institute and by the University of Texas at Austin. The Wind River Forest Dynamics Plot is a collaborative project of Utah State University and the USDA Forest Service Pacific Northwest Research Station. Funding was provided by the Smithsonian ForestGEO, Utah State University, the Utah Agricultural Experiment Station, and the National Science Foundation. We acknowledge the Gifford Pinchot National Forest and the Forest Service Wind River Field Station for providing logistical support, and the students, volunteers and staff individually listed at http://wfdp.org for data collection. The Utah Forest Dynamics Plot is a collaborative project of Utah State University and the Utah Agricultural Experiment Station. Funding was provided by Utah State University, the Ecology Center at Utah State University, and the Utah Agricultural Experiment Station. We thank Cedar Breaks National Monument for providing logistical support, and the students, volunteers and staff individually listed at http://ufdp.org for data collection. The plot census at the Michigan Big Woods Forest Dynamics Plot was supported by a USDA McIntire-Stennis Grant and the Edwin S. George Reserve Fund at the University of Michigan. Funding for the Tyson Research Center Plot was provided by the International Center for Advanced Renewable Energy and Sustainability (I-CARES) at Washington University in St. Louis, the National Science Foundation (DEB 1557094 to JAM and MJS), the Smithsonian ForestGEO, and Tyson Research Center. We thank the Tyson Research Center staff for providing logistical support, and the more than 100 high school students, undergraduate students, and researchers that have contributed to the project. Funding for trait-data collection at the Smithsonian Environmental Research Center, Tyson Research Center, and Wind River was provided by a ForestGEO Research Grants Program award to MJS and JAM. Funding for chemical-ecology research at Tyson Research Center was provided by a Washington University Environmental Studies Grant for Student Research awarded to ED-W. Funding for the Scotty Creek Forest Dynamics Plot was provided to JLB by the Natural Sciences and Engineering Research Council of Canada Discovery Grants program, Global Water Futures project Northern Water Futures, Canada Foundation for Innovation, Canada Foundation for Climate and Atmospheric Sciences, the Northern Student Training Program (support for field assistants), and the Smithsonian ForestGEO program. Funding for the SCBI ForestGEO plot was provided by ForestGEO, the Smithsonian Institution, and the HSBC Climate Partnership.

ACKNOWLEDGMENTS

We thank P. Dorrestein, E. A. Herre, J.-P. Salminen, V. Swamy, A. Tripathi, and M. Volf for helpful discussion and D. Espinoza, K. Greig, L. Hart, F. MacNeill, and K. Richardson for assistance in the laboratory; MiAmbiente Ministry of the Environment of the Republic of Panama for supporting the broader research endeavor of which this study is a part; the Texas Advanced Computing Center for assistance and support of high-performance computing for metabolomics; C. Catano, R. Swing, and D. Vela Díaz for assistance with trait sampling at Wind River and SERC, and Erika Gonzalez-Akre for assistance with leaf sampling at SCBI. We thank Scotty Creek Research Station, which provided accommodations during fieldwork. and the Government of the Northwest Territories—Wilfrid Laurier University partnership, which provided logistical support for fieldwork. We are grateful to the Dehcho First Nations for permission to conduct research on their lands (Aurora Research Institute license numbers 15413 and 16431). Finally, R. Patankar, K. Dearborn, and C. Wallace were central to successful establishment and remeasurement of the Scotty Creek plot.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.679638/full#supplementary-material

Supplementary Figure 1 | Phylogenetic principal components analysis of variation among species in three morphological traits and in total ion intensity in five chemical superclasses. Loadings for chemical superclasses reflect species means in total ion intensity of all compounds classified in each superclass. Loadings for morphological traits reflect species means. Points represent 91 unique species; colors represent species occurrences in seven forest plots. Species that occurred in >1 forest plot contributed only a single observation (the species mean) to the principal components analysis, but are visually represented by jittered points, the colors of which reflect the species incidence in each forest plot. Only species for which every trait was recorded are included.

Supplementary Figure 2 | Linear regressions of metabolomic similarity of five chemical superclasses vs. $\log_{10}$ MAP, MAT, and species richness. Significant regressions are indicated by a solid line; non-significant regressions with $p < 0.07$ are represented by a dashed line. Forest plot codes are: SC, Scotty Creek, NT; WR, Wind River, WA; CB, Cedar Breaks, UT; TR, Tyson Research Center, MO; BW, Michigan Big Woods, MI; SB, Smithsonian Conservation Biology Institute (SCBI), VA; and SR, Smithsonian Environmental Research Center (SERC), MD.

Supplementary Figure 3 | Linear regressions of metabolomic richness of five chemical superclasses vs. $\log_{10}$ MAP, MAT, and species richness. Significant regressions are indicated by a solid line; non-significant regressions with $p < 0.07$ are represented by a dashed line. Forest plot codes are: SC, Scotty Creek, NT; WR, Wind River, WA; CB, Cedar Breaks, UT; TR, Tyson Research Center, MO; BW, Michigan Big Woods, MI; SB, Smithsonian Conservation Biology Institute (SCBI), VA; and SR, Smithsonian Environmental Research Center (SERC), MD.

Supplementary Figure 4 | Linear regressions of community abundance-weighted mean morphological trait values at each forest site vs. log MAP, MAT, and species richness. Significant regressions are indicated by a solid

line. Forest plot codes are: SC, Scotty Creek, NT; WR, Wind River, WA; CB, Cedar Breaks, UT; TR, Tyson Research Center, MO; BW, Michigan Big Woods, MI; SB, Smithsonian Conservation Biology Institute (SCBI), VA; and SR, Smithsonian Environmental Research Center (SERC), MD.

Supplementary Table 1 | Sampling and species composition at seven forest sites. Columns represent: Latin, each species taxonomic name; Code, six-character species code; Plot Sampled, the site where the species/population was sampled. Seven columns labeled with forest plots contain a binary value (1/0) that indicates which sampled population was used to represent the species at the site (e.g., *Populus grandidentata* sampled at Michigan Big Woods was used to represent the species at Michigan Big Woods and SERC).

Supplementary Table 2 | Combined foliar metabolome of 140 species sampled in seven forest plots. Columns represent: id, unique alphanumeric compound identifier; ms2_library_match, named compound matched in a public mass spectral library queried using GNPS; parent_mass, the mass of the compound prior to MS2 fragmentation; retention_time, the time at which the compound was eluted from the UHPLC column during liquid chromatography; smiles, a SMILES text string representing the molecular structure inferred using either CSI:fingerID and Sirius or the MS2 library match; structure_source, whether the SMILES structure was inferred using CSI:fingerID and Sirius or represents that of the compound matched in the MS2 library on GNPS. Subsequent columns represent chemotaxonomic classifications at five levels of organization (kingdom, superclass, class, subclass, and direct_parent) made using ClassyFire. The column "customclass" reflects the five chemotaxonomic superclasses the authors defined for analyses presented for publication. The columns "ScottyCreek," "WindRiver," "CedarBreaks," "Tyson," "MichiganBigWoods," "SCBI," and "SERC" indicate whether the compound occurred in a plant species recorded in each of the seven forest plots. The remaining columns, with column names ending in "mzXML," indicate the ion abundance of each compound in each species-by-site pooled sample.

Supplementary Table 3 | Hierarchical dendrogram of 13,480 foliar metabolites. Qemistree (Tripathi et al., 2021) represents the structural similarity of unique metabolites in the form of a "phylogeny," a hierarchical dendrogram in which structurally related metabolites form sister pairs and clades. Tip labels correspond to compound identifiers in **Supplementary Table 2**.

Supplementary Table 4 | Links to GNPS Feature-Based Molecular Network and Qemistree results for 11 batches of species.

Supplementary Table 5 | Trait values for specific leaf Area (SLA), wood density and seed mass and PCA scores (PCA axes 1–3) for each unique site by species combination. Measurements that were not taken at each site have their source listed. All seed mass data is from the KEW seed information database.

Supplementary Table 6 | Trait loadings for each of three axes from a Principal Components Analysis.

REFERENCES

Ackerly, D. D., and Reich, P. B. (1999). Convergence and correlations among leaf size and function in seed plants: a comparative test using independent contrasts. *Am. J. Bot.* 86, 1272–1281. doi: 10.2307/2656775

Allen, D., Dick, C. W., Strayer, E., Perfecto, I., and Vandermeer, J. (2018). Scale and strength of oak–mesophyte interactions in a transitional oak–hickory forest. *Can. J. For. Res.* 48, 1366–1372. doi: 10.1139/cjfr-2018-0131

Allen, D., Dick, C., Burnham, R. J., Perfecto, I., and Vandermeer, J. (2020). *The Michigan Big Woods Research Plot at the Edwin s. George Reserve*. Pinckney, MI: University of Michigan.

Anderson-Teixeira, K. J., Davies, S. J., Bennett, A. C., Gonzalez-Akre, E. B., Muller-Landau, H. C., and Joseph Wright, S. (2015). CTFS-ForestGEO: a worldwide network monitoring forests in an era of global change. *Glob. Change Biol.* 21, 528–549.

Blomberg, S. P., Garland, T., and Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57, 717–745. doi: 10.1111/j.0014-3820.2003.tb00285.x

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857.

Bourg, N. A., McShea, W. J., Thompson, J. R., McGarvey, J. C., and Shen, X. (2013). Initial census, woody seedling, seed rain, and stand structure data for the SCBI SIGEO large forest dynamics plot. *Ecology* 94, 2111–2112. doi: 10.1890/13-0010.1

Cavender-Bares, J., Keen, A., and Miles, B. (2006). Phylogenetic structure of Floridian plant communities depends on taxonomic and spatial scale. *Ecology* 87, S109–S122.

Chu, C., Lutz, J. A., Král, K., Vrška, T., Yin, X., and Myers, J. A. (2019). Direct and indirect effects of climate on richness drive the latitudinal diversity gradient in forest trees. *Ecol. Lett.* 22, 245–255.

Close, D. C., and McArthur, C. (2002). Rethinking the role of many plant phenolics – protection from photodamage not herbivores? *Oikos* 99, 166–172. doi: 10.1034/j.1600-0706.2002.990117.x

Coley, P. D. (1983). Herbivory and defensive characteristics of tree species in a Lowland tropical forest. *Ecol. Monogr.* 53, 209–233. doi: 10.2307/1942495

Coley, P. D., and Aide, T. M. (1991). "Comparisons of herbivory and plant defenses in temperate and tropical broad-leaved forests," in *Plant-Animal Interactions: Evolutionary Ecology in Tropical and Temperate Regions*, eds P. W. Price, T. M. Lewinsohn, G. W. Fernandes, and W. W. Benson (New York: Wiley), 25–49.

Coley, P. D., and Barone, J. A. (1996). Herbivory and plant defenses in tropical forests. *Annu. Rev. Ecol. Syst.* 27, 305–335. doi: 10.1146/annurev.ecolsys.27.1.305

Davies, S. J., Abiem, I., Abu Salim, K., Aguilar, S., Allen, D., and Alonso, A. (2021). ForestGEO: understanding forest diversity and dynamics through a global observatory network. *Biol. Conserv.* 253:108907.

Dearborn, K. D., Wallace, C. A., Patankar, R., and Baltzer, J. L. (2020). Permafrost thaw in boreal peatlands is rapidly altering forest community composition. *J. Ecol.* 109, 1452–1467. doi: 10.1111/1365-2745.13569

Defossez, E., Pellissier, L., and Rasmann, S. (2018). The unfolding of plant growth form-defence syndromes along elevation gradients. *Ecol. Lett.* 21, 609–618. doi: 10.1111/ele.12926

Díaz, S., Kattge, J., Cornelissen, J. H., Wright, I. J., Lavorel, S., Dray, S., et al. (2016). The global spectrum of plant form and function. *Nature* 529, 167–171.

Dirzo, R., and Boege, K. (2008). "Patterns of herbivory and defense in tropical dry and rain forests," in *Tropical Forest Community Ecology*, eds W. P. Carson and S. A. Schnitzer (Oxford: Wiley-Blackwell).

Djoumbou Feunang, Y., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., et al. (2016). ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminformatics* 8:61.

Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A. A., Melnik, A. V., Meusel, M., et al. (2019). SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* 16, 299–302. doi: 10.1038/s41592-019-0344-8

Dührkop, K., Shen, H., Meusel, M., Rousu, J., and Böcker, S. (2015). Searching molecular structure databases with tandem mass spectra using CSI: FingerID. *Proc. Natl. Acad. Sci.* 112, 12580–12585. doi: 10.1073/pnas.1509788112

Dyer, L. A., Singer, M. S., Lill, J. T., Stireman, J. O., Gentry, G. L., Marquis, R. J., et al. (2007). Host specificity of Lepidoptera in tropical and temperate forests. *Nature* 448, 696–U9.

Ehrlich, P. R., and Raven, P. H. (1964). Butterflies and plants - a study in coevolution. *Evolution* 18, 586–608. doi: 10.2307/2406212

Enquist, B. J., West, G. B., Charnov, E. L., and Brown, J. H. (1999). Allometric scaling of production and life-history variation in vascular plants. *Nature* 401, 907–911. doi: 10.1038/44819

Esquivel-Muelbert, A., Baker, T. R., Dexter, K. G., Lewis, S. L., Steege, H., and Lopez-Gonzalez, G. (2017). Seasonal drought limits tree species across the Neotropics. *Ecography* 40, 618–629.

Fine, P. V. A., Miller, Z. J., Mesones, I., Irazuzta, S., Appel, H. M., Stevens, M. H. H., et al. (2006). The growth-defense trade-off and habitat specialization by plants in Amazonian forests. *Ecology* 87, S150–S162.

Forrister, D. L., Endara, M. J., Younkin, G. C., Coley, P. D., and Kursar, T. A. (2019). Herbivores as drivers of negative density dependence in tropical forest saplings. *Science* 363, 1213–1216. doi: 10.1126/science.aau9460

Furniss, T. J., Larson, A. J., and Lutz, J. A. (2017). Reconciling niches and neutrality in a subalpine temperate forest. *Ecosphere* 8:e1847.

Garnier, E., Cortez, J., Billès, G., Navas, M.-L., Roumet, C., Debussche, M., et al. (2004). Plant functional markers capture ecosystem properties during secondary succession. *Ecology* 85, 2630–2637. doi: 10.1890/03-0799

Gentry, A. H. (1988). Changes in plant community diversity and floristic composition on environmental and geographical gradients. *Ann. Mo. Bot. Gard.* 75, 1–34. doi: 10.2307/2399464

Givnish, T. J. (1999). On the causes of gradients in tropical tree diversity. *J. Ecol.* 87, 193–210. doi: 10.1046/j.1365-2745.1999.00333.x

Harrison, S., Spasojevic, M. J., and Li, D. (2020). Climate and plant community diversity in space and time. *Proc. Natl. Acad. Sci.* 117, 4464–4470. doi: 10.1073/pnas.1921724117

Hawkins, B. A., Field, R., Cornell, H. V., Currie, D. J., Guégan, J.-F., Kaufman, D. M., et al. (2003). Energy, water, and broad-scale geographic patterns of species richness. *Ecology* 84, 3105–3117. doi: 10.1890/03-8006

Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., et al. (2020). TRY plant trait database–enhanced coverage and open access. *Glob. Change Biol.* 26, 119–188.

Kreft, H., and Jetz, W. (2007). Global patterns and determinants of vascular plant diversity. *Proc. Natl. Acad. Sci.* 104, 5925–5930. doi: 10.1073/pnas.0608361104

Kursar, T. A., Dexter, K. G., Lokvam, J., Pennington, R. T., Richardson, J. E., Weber, M. G., et al. (2009). The evolution of antiherbivore defenses and their contribution to species coexistence in the tropical tree genus Inga. *Proc. Natl. Acad. Sci. U. S. A.* 106, 18073–18078. doi: 10.1073/pnas.0904786106

Laliberté, E., and Legendre, P. (2010). A distance-based framework for measuring functional diversity from multiple traits. *Ecology* 91, 299–305. doi: 10.1890/08-2244.1

Lamanna, C., Blonder, B., Violle, C., Kraft, N. J. B., Sandel, B., and Imova, I. (2014). Functional trait space and the latitudinal diversity gradient. *Proc. Natl. Acad. Sci.* 111, 13745–13750.

Letunic, I., and Bork, P. (2019). Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259.

Lutz, J. A., Larson, A. J., Freund, J. A., Swanson, M. E., and Bible, K. J. (2013). The importance of large-diameter trees to forest structural heterogeneity. *PLoS One* 8:e82784. doi: 10.1371/journal.pone.0082784

Mason, N. W. H., Mouillot, D., Lee, W. G., and Wilson, J. B. (2005). Functional richness, functional evenness and functional divergence: the primary components of functional diversity. *Oikos* 111, 112–118. doi: 10.1111/j.0030-1299.2005.13886.x

Mittelbach, G. G., Schemske, D. W., Cornell, H. V., Allen, A. P., Brown, J. M., and Bush, M. B. (2007). Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecol. Lett.* 10, 315–331.

Moles, A. T., Ackerly, D. D., Tweddle, J. C., Dickie, J. B., Smith, R., Leishman, M. R., et al. (2007). Global patterns in seed size. *Glob. Ecol. Biogeogr.* 16, 109–116.

Moreira, X., Nell, C. S., Katsanis, A., Rasmann, S., and Mooney, K. A. (2018). Herbivore specificity and the chemical basis of plant–plant communication in Baccharis salicifolia (Asteraceae). *New Phytol.* 220, 703–713. doi: 10.1111/nph.14164

Muller-Landau, H. C. (2010). The tolerance–fecundity trade-off and the maintenance of diversity in seed size. *Proc. Natl. Acad. Sci.* 107, 4242–4247. doi: 10.1073/pnas.0911637107

Nothias, L.-F., Petras, D., Schmid, R., Dührkop, K., Rainer, J., Sarvepalli, A., et al. (2020). Feature-based molecular networking in the GNPS analysis environment. *Nat. Methods* 17, 905–908.

Pellissier, L., Roger, A., Bilat, J., and Rasmann, S. (2014). High elevation *Plantago lanceolata* plants are less resistant to herbivory than their low elevation conspecifics: is it just temperature? *Ecography* 37, 950–959. doi: 10.1111/ecog.00833

Pérez-Harguindeguy, N., Díaz, S., Garnier, E., Lavorel, S., Poorter, H., and Jaureguiberry, P. (2016). Corrigendum to: New handbook for standardised measurement of plant functional traits worldwide. *Aust. J. Bot.* 64, 715–716. doi: 10.1071/bt12225_co

Pluskal, T., Castillo, S., Villar-Briones, A., and Orešič, M. (2010). MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinform.* 11:395.

R Core Team (2020). *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rasband, W. S. (2018). *ImageJ. U. S. National Institutes of Health*. MD, United States: Bethesda. Available online at: https://imagej.nih.gov/ij/

Rasmann, S., Buri, A., Gallot-Lavallée, M., Joaquim, J., Purcell, J., and Pellissier, L. (2014). Differential allocation and deployment of direct and indirect defences by *Vicia sepium* along elevation gradients. *J. Ecol.* 102, 930–938. doi: 10.1111/1365-2745.12253

Reich, P. B., Walters, M. B., and Ellsworth, D. S. (1997). From tropics to tundra: global convergence in plant functioning. *Proc. Natl. Acad. Sci.* 94, 13730–13734. doi: 10.1073/pnas.94.25.13730

Revell, L. J. (2010). Phylogenetic signal and linear regression on species data. *Methods Ecol. Evol.* 1, 319–329. doi: 10.1111/j.2041-210x.2010.00044.x

Revell, L. J. (2012). Phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3, 217–223. doi: 10.1111/j.2041-210x.2011.00169.x

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., et al. (2013). Package 'mass.'. *Cran R* 538, 113–120.

Rüger, N., Comita, L. S., Condit, R., Purves, D., Rosenbaum, B., Visser, M. D., et al. (2018). Beyond the fast–slow continuum: demographic dimensions structuring a tropical tree community. *Ecol. Lett.* 21, 1075–1084. doi: 10.1111/ele.12974

Salazar, D., Lokvam, J., Mesones, I., Pilco, M. V., Zuñiga, J. M. A., Valpine, P., et al. (2018). Origin and maintenance of chemical diversity in a species-rich tropical tree lineage. *Nat. Ecol. Evol.* 2, 983–990. doi: 10.1038/s41559-018-0552-0

Sam, K., Koane, B., Sam, L., Mrazova, A., Segar, S., Volf, M., et al. (2020). Insect herbivory and herbivores of Ficus species along a rain forest elevational gradient in Papua New Guinea. *Biotropica* 52, 263–276. doi: 10.1111/btp.12741

Schemske, D. W., Mittelbach, G. G., Cornell, H. V., Sobel, J. M., and Roy, K. (2009). Is there a latitudinal gradient in the importance of biotic interactions? *Annu. Rev. Ecol. Evol. Syst.* 40, 245–269. doi: 10.1146/annurev.ecolsys.39.110707.173430

Schneider, G. F., Coley, P. D., Younkin, G. C., Forrister, D. L., Mills, A. G., and Kursar, T. A. (2019). Phenolics lie at the centre of functional versatility in the responses of two phytochemically diverse tropical trees to canopy thinning. *J. Exp. Bot.* 70, 5853–5864. doi: 10.1093/jxb/erz308

Sedio, B. E., and Ostling, A. M. (2013). How specialised must natural enemies be to facilitate coexistence among plants? *Ecol. Lett.* 16, 995–1003. doi: 10.1111/ele.12130

Sedio, B. E., Boya, P. C. A., and Rojas Echeverri, J. C. (2018a). A protocol for high-throughput, untargeted forest community metabolomics using mass spectrometry molecular networks. *Appl. Plant Sci.* 6:e1033. doi: 10.1002/aps3.1033

Sedio, B. E., Devaney, J. L., Pullen, J., Parker, G. G., Wright, S. J., and Parker, J. D. (2020). Chemical novelty facilitates herbivore resistance and biological invasions in some introduced plant species. *Ecol. Evol.* 10, 8770–8792. doi: 10.1002/ece3.6575

Sedio, B. E., Echeverri, J. C. R., Boya, P. C. A., and Wright, S. J. (2017). Sources of variation in foliar secondary chemistry in a tropical forest tree community. *Ecology* 98, 616–623. doi: 10.1002/ecy.1689

Sedio, B. E., Parker, J. D., McMahon, S. M., and Wright, S. J. (2018b). Comparative foliar metabolomics of a tropical and a temperate forest community. *Ecology* 99, 2647–2653. doi: 10.1002/ecy.2533

Shipley, B., De Bello, F., Cornelissen, J. H. C., Laliberté, E., Laughlin, D. C., and Reich, P. B. (2016). Reinforcing loose foundation stones in trait-based plant ecology. *Oecologia* 180, 923–931. doi: 10.1007/s00442-016-3549-x

Spasojevic, M. J., Turner, B. L., and Myers, J. A. (2016). When does intraspecific trait variation contribute to functional beta-diversity? *J. Ecol.* 104, 487–496. doi: 10.1111/1365-2745.12518

Swenson, N. G., and Enquist, B. J. (2007). Ecological and evolutionary determinants of a key plant functional trait: wood density and its community-wide variation across latitude and elevation. *Am. J. Bot.* 94, 451–459. doi: 10.3732/ajb.94.3.451

Tripathi, A., Vázquez-Baeza, Y., Gauglitz, J. M., Wang, M., Dührkop, K., and Nothias-Esposito, M. (2021). Chemically informed analyses of metabolomics

mass spectrometry data with Qemistree. *Nat. Chem. Biol.* 17, 146–151. doi: 10.1038/s41589-020-00677-3

Uemura, M., and Steponkus, P. L. (1994). A contrast of the plasma membrane lipid composition of Oat and Rye leaves in relation to freezing tolerance. *Plant Physiol.* 104, 479–496. doi: 10.1104/pp.104.2.479

Volf, M., Laitila, J. E., Kim, J., Sam, L., Sam, K., and Isua, B. (2020). Compound specific trends of chemical defences in Ficus along an elevational gradient reflect a complex selective landscape. *J. Chem. Ecol.* 46, 442–454. doi: 10.1007/s10886-020-01173-7

Volf, M., Segar, S. T., Miller, S. E., Isua, B., Sisol, M., and Aubona, G. (2018). Community structure of insect herbivores is driven by conservatism, escalation and divergence of defensive traits in Ficus. *Ecol. Lett.* 21, 83–92. doi: 10.1111/ele.12875

Wang, M. X., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., and Peng, Y. (2016). Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nat. Biotechnol.* 34, 828–837.

Webb, C. O., and Donoghue, M. J. (2005). Phylomatic: tree assembly for applied phylogenetics. *Mol. Ecol. Notes* 5, 181–183. doi: 10.1111/j.1471-8286.2004.00829.x

Weiher, E., and Keddy, P. A. (1995). Assembly rules, null models, and trait dispersion: new questions from old patterns. *Oikos* 74, 159–164. doi: 10.2307/3545686

Wink, M. (2003). Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* 64, 3–19. doi: 10.1016/s0031-9422(03)00300-5

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., and Bongers, F. (2004). The worldwide leaf economics spectrum. *Nature* 428, 821–827.

Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. A., and FitzJohn, R. G. (2014). Three keys to the radiation of angiosperms into freezing environments. *Nature* 506, 89–92. doi: 10.1038/nature12872

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 28 May 2021
doi: 10.3389/fevo.2021.668281

Genome Skimming Reveals Widespread Hybridization in a Neotropical Flowering Plant Radiation

Oriane Loiseau[1,2], Talita Mota Machado[3], Margot Paris[4], Darina Koubínová[1], Kyle G. Dexter[2], Leonardo M. Versieux[5], Christian Lexer[6†] and Nicolas Salamin[1]*

[1] Department of Computational Biology, University of Lausanne, Lausanne, Switzerland, [2] School of Geosciences, University of Edinburgh, Edinburgh, United Kingdom, [3] Programa de Pós Graduação em Biologia Vegetal, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, [4] Department of Biology, Unit Ecology and Evolution, University of Fribourg, Fribourg, Switzerland, [5] Departamento de Botânica e Zoologia, Centro de Biociências, Universidade Federal do Rio Grande do Norte, Natal, Brazil, [6] Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria

The tropics hold at least an order of magnitude greater plant diversity than the temperate zone, yet the reasons for this difference are still subject to debate. Much of tropical plant diversity is in highly speciose genera and understanding the drivers of such high species richness will help solve the tropical diversity enigma. Hybridization has recently been shown to underlie many adaptive radiations, but its role in the evolution of speciose tropical plant genera has received little attention. Here, we address this topic in the hyperdiverse Bromeliaceae genus *Vriesea* using genome skimming data covering the three genomic compartments. We find evidence for hybridization in ca. 11% of the species in our dataset, both within the genus and between *Vriesea* and other genera, which is commensurate with hybridization underlying the hyperdiversity of *Vriesea*, and potentially other genera in Tillandsioideae. While additional genomic research will be needed to further clarify the contribution of hybridization to the rapid diversification of *Vriesea*, our study provides an important first data point suggesting its importance to the evolution of tropical plant diversity.

Keywords: Bromeliaceae, genome skimming, phylogenomics, hybridization, *Vriesea*

OPEN ACCESS

Edited by:
Danilo M. Neves, Federal University of Minas Gerais, Brazil

Reviewed by:
Clarisse Palma-Silva, State University of Campinas, Brazil
Tim Barraclough, University of Oxford, United Kingdom

****Correspondence:***
Oriane Loiseau
oriane.loiseau@unil.ch

† Deceased

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *15 February 2021*
Accepted: *30 April 2021*
Published: *28 May 2021*

Citation:
Loiseau O, Mota Machado T, Paris M, Koubínová D, Dexter KG, Versieux LM, Lexer C and Salamin N (2021) Genome Skimming Reveals Widespread Hybridization in a Neotropical Flowering Plant Radiation. Front. Ecol. Evol. 9:668281. doi: 10.3389/fevo.2021.668281

INTRODUCTION

Hybridization, the exchange of genetic material between lineages, is a widespread evolutionary phenomenon of crucial importance in the evolution of diversity (Marques et al., 2019). Gene flow between species may have deleterious effects on species cohesion by breaking species boundaries and causing genetic homogenization. There is however increasing evidence that hybridization can promote genetic diversity, adaptation and speciation, although the exchange of genetic material *per se* might not be sufficient to create new diversity without a key role of natural selection acting on the relevant parts of the genome to maintain them (Schumer et al., 2018). In particular, hybridization has been shown to be involved in many adaptive radiations, such as Darwin's finches (Grant and Grant, 2019), African rift lakes cichlid fishes (Meier et al., 2017; Irisarri et al., 2018; Svardal et al., 2020), *Heliconius* butterflies (Kozak et al., 2021), and Hawaiian silver swords

(Barrier et al., 1999). In comparison with the low rate of mutation, hybridization can immediately elevate the standing genetic variation upon which selection can act (Seehausen, 2013). This has the potential to accelerate the adaptive divergence of populations in face of environmental variability, novel habitats or previously underutilized niches. Elevated rates of speciation following hybridization can create a positive feedback as the evolution of many closely related species in one area may in turn provide new opportunities for hybridization (Seehausen, 2004).

Interspecific gene flow can promote diversification through the direct generation of a new species via hybrid speciation or through adaptive introgression, the latter being the transfer of only parts of the genome which may confer a selective advantage (Harrison and Larson, 2014). The identification of these mechanisms has been helped by two important advances in the field. First, the emergence of the genome scale datasets that are becoming standard in evolutionary studies provided the raw data to tackle these questions. Second, the development of statistical methods that can formally test the presence of introgression against a null model that involves only incomplete lineage sorting (ILS; e.g., Joly et al., 2009; Green et al., 2010; Blanco-Pastor et al., 2012; Blischak et al., 2018). A remaining open question is the prevalence of hybridization or adaptive introgression in the context of diversifying lineages (Pfennig, 2021).

Although hybridization has been detected across a wide range of organisms, it is particularly frequent in plants, and is acknowledged as one of the main drivers of the evolution of angiosperms (Soltis and Soltis, 2009). However, little is known about the contribution of hybridization in the evolution of diverse tropical clades which make up a large part of plant diversity. In this study, we focus on the Bromeliaceae, which is, with ca. 3,500 species, one of the most diverse and iconic plant families of the Neotropical flora, particularly abundant in Andean and Atlantic rainforests (Smith and Downs, 1974; Benzing, 2000; Gouda et al., 2018). The family harbors several exceptionally diverse young clades, such as the core tillandsioids and the tank-forming epiphytic bromelioids (Givnish et al., 2014). The increased rates of diversification which underlie these rapid radiations may have been spurred by several key innovations such as the tank habit, CAM photosynthesis or hummingbird pollination (Givnish et al., 2014; Silvestro et al., 2014). Hybridization is also thought to have promoted bromeliad diversity (Palma-Silva et al., 2016) but although both artificial (Vervaeke et al., 2004; Wagner et al., 2015; de Souza et al., 2017) and natural hybrids have been reported (e.g., Palma-Silva et al., 2011), the frequency and exact contribution of this phenomenon to bromeliad diversity are still to be determined. While some studies suggest that hybridization generates intraspecific genetic diversity in isolated populations (Lexer et al., 2016; Neri et al., 2018; Mota et al., 2019), it may also prevent genetic differentiation and blur species boundaries (Goetze et al., 2017). To date, most of the research on hybridization in bromeliads focuses on the population level and studies with a macroevolutionary perspective are lacking.

Here, we take a phylogenomic approach to the study of hybridization in the Tillandsioideae, the largest of the three subfamilies of bromeliads, with a focus on the diverse genus *Vriesea* (231 species, Gouda et al., 2018). Using genome skimming we provide an unprecedented molecular dataset for 108 species of *Vriesea* and 39 species from related genera. We use phylogenetic inference and coalescent-based approaches to test for the presence of hybridization during the evolutionary history of the group. Our results provide evidence for clear cyto-nuclear discordance in the phylogenetic history of Tillandisoideae. We further detect signatures of ancient hybridization at both the inter- and intrageneric level. Our study suggests that hybridization has occurred repeatedly throughout the evolutionary history of bromeliads and may have played an important role in triggering the rapid evolution of this group.

MATERIALS AND METHODS

Genomic Dataset

We used an extended version of the chloroplast genomic dataset generated with genome skimming by Machado et al. (2020) containing nuclear, chloroplast and mitochondrial data. Genome skimming, also referred as low-coverage whole genome sequencing, allows easy and cost-effective obtention of DNA sequences for phylogenomic analyses across a wide range of evolutionary divergences (Dodsworth, 2015). This technique is particularly effective for the high-depth sequencing of the high-copy fraction of the genome such as plastids and ribosomal nuclear DNA (Coissac et al., 2016). It also offers the possibility to obtain shallow coverage of single copy nuclear DNA, and thousands of low-depth nuclear markers can be obtained with a sufficient sequencing effort. Sampling included 147 species of 11 genera in the subfamily Tillandsioideae, plus *Ananas comosus* as an outgroup. In total, there were 108 species of *Vriesea s.l.*, representing c. 46% of the total diversity of the genus. For the other genera the number and proportion of species were as follows (species numbers according to Gouda et al., 2018): *Alcantarea* (11 species; 26% of the diversity), *Goudaea* (2; 100%), *Guzmania* (4; 1%), *Lutheria* (2; 50%), *Mezobromelia* (1; 20%), *Racinaea* (1; 78%), *Stigmatodon* (4; 22%), *Tillandsia* (9; 1%), *Werauhia* (3; 1%), *Zizkaea* (1; 100%). Protocols of DNA extraction, library preparation and sequencing, as well as the bioinformatic pipeline for quality checking of reads, and SNP calling are described in Machado et al. (2020). Briefly, library preparation was performed using KAPA LTP library preparation kits (Roche, Basel, Switzerland) and unique barcodes for each library were chosen from a set of 60 dual-index primers designed in Loiseau et al. (2019) to allow multiplexing. The use of a high number of Illumina dual-indexed primers, designed with a minimum edit distance of 4 between each of the 7-bp indexes, reduces the probability of conversion by sequencing, amplification errors and index hopping (Kircher et al., 2012; Costello et al., 2018). After sequencing in an Illumina HiSeq 3000 Genome Analyzer (Illumina, San Diego, California, United States), quality controlling and read trimming, pair-end reads (2 × 150 bp) were mapped onto the *Tillandsia adpressiflora* Mez pseudo-reference genome built in de la Harpe et al. (2020) using BOWTIE2 v.2.2.5 (Langmead and Salzberg, 2012). After identification of PCR duplicates, realignment of reads around indels and base calibration, SNPs were called using UnifiedGenotyper of GATK v.3.6 (McKenna et al.,

2010) and the EMIT_ALL_SITES option to recover both variant and invariant sites. We used Vcftools v.0.1.13 (Danecek et al., 2011) to filter only high-quality sites called with Q values higher than 20 and less than 50% missing data. For phylogenomic inferences invariant sites were filtered out and separate alignments of SNPs were produced for the nuclear, chloroplast and mitochondrial genomes.

Phylogenetic Inferences

We performed phylogenetic reconstruction using both maximum-likelihood and coalescent-based methods. Since using only variable sites for phylogenetic inference can lead to incorrect estimation of branch lengths and topology (Leaché et al., 2015), we employed methods designed for SNP datasets. First, we used RAxML v.8.2.10 (Stamatakis, 2014) to infer separate phylogenetic trees for the alignments of chloroplast, mitochondrial and nuclear SNPs. For each analysis we applied a GTR + G model of substitution and an ascertainment bias correction to account for the absence of constant sites in the alignments. Bootstrap replicates were performed using the -autoMRE option which executes a maximum of 1,000 bootstrap replicate searches but stops if support values reach convergence earlier. Secondly, to account for any potential effects of ILS in phylogeny estimation, we used SVDquartet (Chifman and Kubatko, 2014, 2015), a coalescent-based method designed for SNP data implemented in PAUP* (Swofford, 2001) to infer the nuclear phylogenetic tree. SVDquartet is a coalescent method for inferring phylogenetic relationships based on phylogenetic invariants which are mathematical functions representing the expected frequencies of site patterns in an alignment. It first infers the quartet topology of all subsets of four taxon in the dataset and then agglomerate them in order to obtain the species tree with all taxa. Although SVDquartet can also estimate phylogenetic relationships for non-recombining loci, we did not run it on our chloroplast or mitochondrial datasets because they contained $<$ 10,000 SNPs and the performance of this method is highly dependent on the number of SNPs.

Divergence Time Estimates

We temporally calibrated the nuclear phylogenetic tree using penalized-likelihood implemented in the program treePL v1.0 (Smith and O'meara, 2012). We chose this approach because joint estimation of the topology and divergence times from SNP data (e.g., SNAPP, Bryant et al., 2012) was computationally too intensive to be applied to our large dataset. Given the absence of fossil data for bromeliads, we used a secondary calibration to perform the divergence time estimation. We implemented the age estimate of the crown node of core Tillandsioideae from Bouchenak-Khelladi et al. (2014) as a calibration for the root node of our tree, which encompasses the core Tillandsioideae, using the boundaries of the 95% HPD as minimum and maximum constraints (7.2417–12.984 My). Although their study is a monocotyledon-wide analysis, their sampling of Tillandsioideae is as good as other studies focusing on Bromeliaceae, and their dating analysis is based on nine fossil calibration points. Their sampling did not include any member of the tribe Vrieseeae, and we were unable to calibrate any shallower node in our tree. We performed cross-validation to estimate the value of the best-fitting smoothing parameter whose value can vary between 0 (each branch has its own substitution rate) and 1 (strict clock model).

Hybridization Detection

In order to estimate levels of hybridization across our dataset, while accounting for ILS, we used the program HyDe v0.4.3 (Blischak et al., 2018), a coalescent method which tests for hybridization in the presence of ILS using phylogenetic invariants (Kubatko and Chifman, 2019), similarly to the D statistics (Patterson et al., 2012). HyDe can detect both recent and ancient hybridization without the need for an *a priori* hypothesis, considers ambiguous sites, and can be applied to large genomic datasets. For all possible quartets in a dataset comprising one outgroup population, two parental populations (P2 and P3) and one hybrid population (P1), HyDe estimates the amount of admixture using a gamma statistic. Gamma values are comprised between 0 and 1, where γ =0.5 means 50:50 hybrid and values close to 0 or 1 indicate a low level of asymmetrical admixture by introgression. We tested all four-taxon combinations comprising one outgroup and a triplet of ingroup taxa. For all tests Ananas comosus was set as the outgroup taxon using the python script run_hyde.py[1]. The program applies a conservative Bonferronni correction and outputs the combinations with a significant signal for hybridization at a level of $<$ 0.05. Unless it applied to population data, this approach does not determine whether the significantly introgressed individuals represent the results of a hybrid speciation event, but suggests at a minimum that these individuals have genetic material from at least two different lineages in the phylogeny. We summarized the result of HyDe by recording the genera of the two parental species for each significant triplet, and counted for each hybrid the frequency of the different generic combinations. This allowed us to evaluate whether the detected hybrids most likely reflected intra- or intergeneric hybridization. We then constructed a network diagram using the fonction qgraph from the R package Qgraph v.1.5 (Epskamp et al., 2012) in order to display the frequency at which each genus was involved as a parental species (P2 or P3) for hybrids in other genera across all significant combinations.

RESULTS

Sequencing

After base calling and filtering, 496,594 high-quality sites were obtained for the nuclear genome with a coverage depth of 22.2 reads per position and 19.3% missing data. For the chloroplast and mitochondrial genomes, 68,854 high-quality bases at 53.9X (5.8% missing data) and 19,403 high quality bases at 11.2X (18.2% of missing data) were recovered, respectively. From the SNP calling 116,478 SNPs were obtained for the nuclear, 5,171 SNPs for the chloroplast and 1,069 SNPs for the mitochondrial genomes. Sites presenting a pattern of a single nucleotide and an ambiguity code are considered by RaxML as invariant sites and we excluded them from the alignments, which led to a total

[1]https://github.com/pblischak/HyDe

dataset of 103,244 nuclear SNPs, 3,913 chloroplast SNPs, and 970 mitochondrial SNPs.

Phylogenetic Relationships

Intergeneric phylogenetic relationships were well-supported in all inferred trees. The chloroplast topology is similar to the one recovered in Machado et al. (2020), with subtribe Cipuropsidinae (composed of the genera *Goudaea, Zizkaea, Lutheria,* and *Werauhia)* sister to subtribe Vrieseinae (composed of *Vriesea s.str., Stigmatodon,* and *Alcantarea*, **Figure 1**). Our study adds new results for the mitochondrial phylogeny, which has a similar topology to the plastid tree. However, the nuclear topology resolves the subtribe Cipuropsidinae as sister to the tribe Tillandsieae (**Figure 1**). Furthermore, within the Cipuropsidinae, *Lutheria* is recovered as sister to *Werauhia* in the chloroplast and mitochondrial trees (with bootstrap support of 100 and 98, respectively, **Supplementary Figures 1, 2**), whereas in the nuclear tree, it forms a clade together with *Goudea*, *Zizkaea* and the *Cipuropsis-Mezobromelia* complex (with bootstrap support of 98; **Figure 1** and **Supplementary Figure 3**). The genus *Goudea,* represented by the two species *G. ospinae* and *G. chrysostachys,* is recovered as monophyletic in the nuclear phylogeny but not in the chloroplast and mitochondrial trees (**Figure 1** and **Supplementary Figures 1–3**). The coalescent-based phylogenetic tree obtained with SVDquartet from the nuclear data has a similar topology to the phylogenetic tree inferred with RAxML but an overall lower bootstrap support (**Supplementary Figure 4**). It differs, however, in the position of *Werauhia*, which is found in a sister position to the tribe Tillandsieae, and *Guzmania*, which is recovered nested within *Tillandsia* (with low support, **Supplementary Figure 4**). The monophyletic *Vriesea s. str* is characterized by small branch lengths and low internal support in all phylogenetic trees, yet most of the 12 clades described in Machado et al. (2020), which correspond to morphological or geographic groups, are recovered in the chloroplast and nuclear phylogenies, but not in the mitochondrial phylogeny.

Divergence Times

The best smoothing parameter found by treePL was 0.01, which allows for substantial substitution rate variation among lineages. The estimated crown age of the core Tillandsioideae was 12.98 My. The two main clades, the tribe Tillandsieae + the Cipuropsidinae and subtribe Vrieseinae have crown ages of 11.24 and 7.66 My, respectively (**Figure 2**). The genus *Vriesea s. str* has a crown age of 6.18 My and following the initial divergence a *V. drepanocarpa,* the rest of the group began to diversify 3.91 My ago and most of its diversification occurred during the Pleistocene.

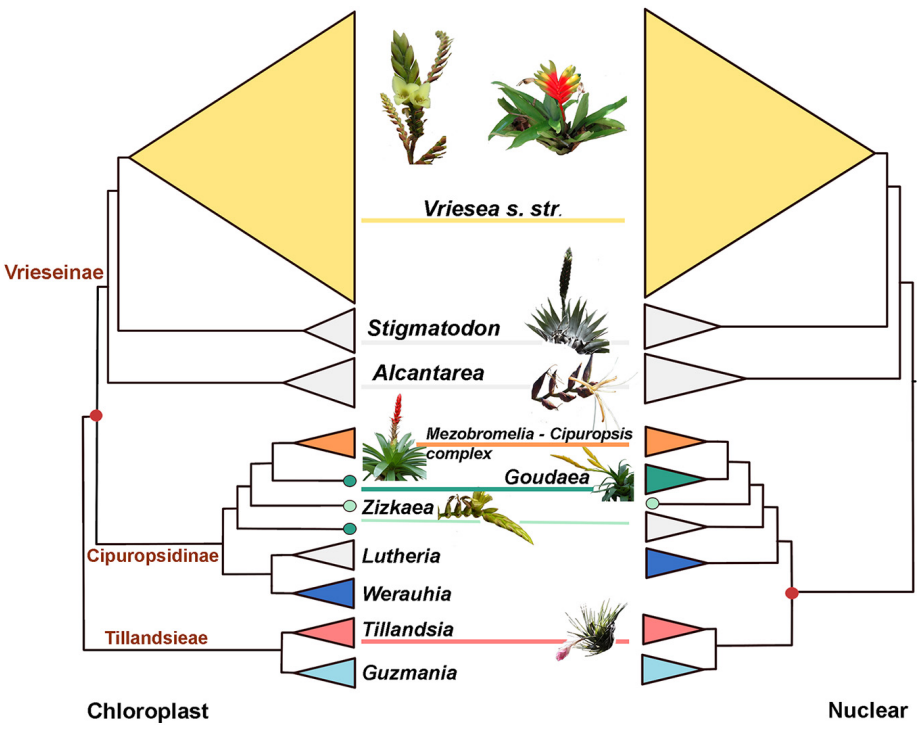


FIGURE 1 | Phylogenetic incongruence in core Tillandsioideae. Summary cladograms of the chloroplast **(left)** and nuclear **(right)** phylogenetic trees. The red dot indicates the incongruent node associated with the placement of subtribe Cipuropsidinae. Genera indicated as follow (from up to down): *V. s. str.*, *Vriesea sensu stricto; Sti*, *Stigmatodon; Alc*, *Alcantarea; Cip- Mez*, *Vriesea—Cipuropsis/Mezobromelia complex; Gou*, *Goudeae; Ziz*, *Zizkaea; Lut*, *Lutheria; Wer*, *Werauhia; Till*, *Tillandsia; Guz*, *Guzmania.* Colors indicate genera where hybrids were detected (see **Figure 2**). Photos by Eric Gouda (*Goudaea, Mezobromelia and Zizkaea*) and Talita Mota Machado (*Alcantarea, Stigmatodon, Tillandsia and Vriesea s. str.*).

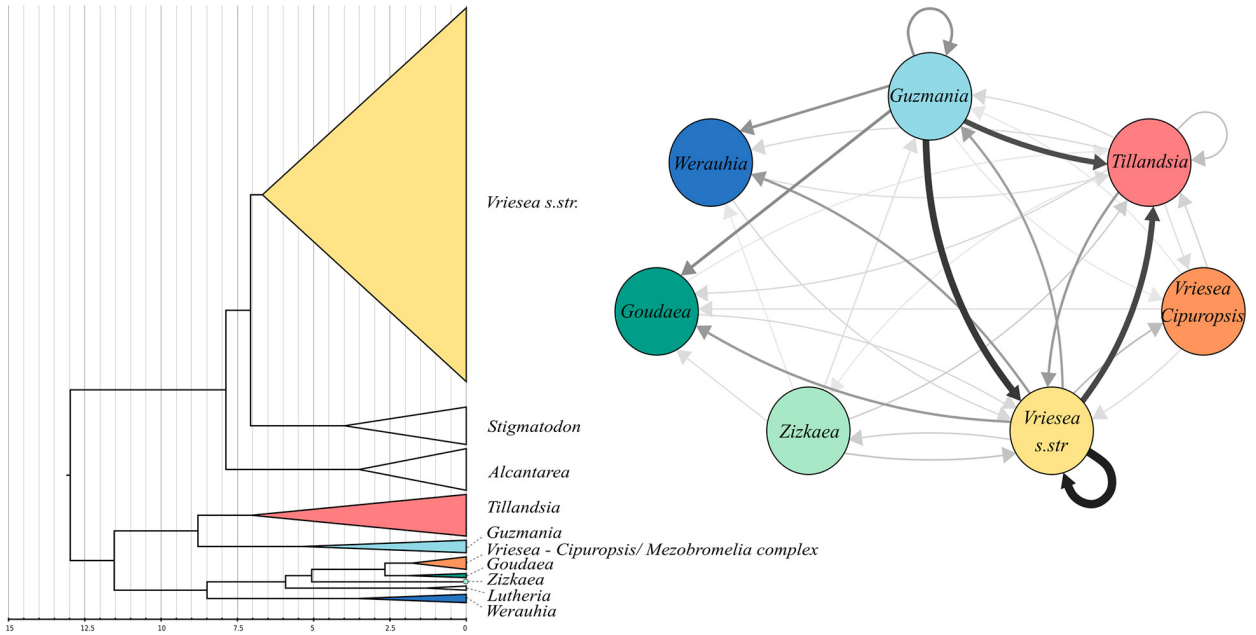


FIGURE 2 | Left: Dated nuclear phylogeny of coreTillandsioids (intrageneric branches not shown). Colors indicate clades were hybrids have been detected. **Right:** network diagram of the putative origin of the detected hybrids across all significant triplets identified by HyDe. The width and color of the arrows are proportional to the frequency at which one genus is detected as a parental lineageaccross all hybrids.

Introgression and Hybridization

The HyDe analyses detected significant introgression in 287 out of the 1,687,425 possible triplets, after the conservative Bonferroni correction (**Supplementary Table 1**). For simplicity, we will call these cases "hybrids," although we fully recognize that it is difficult to distinguish between hybrids and introgressive hybridization from the analyses performed. These 287 combinations contain 17 species identified as potentially of hybrid origin, which included nine species of *Vriesea s.str.*, the species *V. dubia* that belongs to the *Vriesea-Cipuropsis* clade, three species of *Tillandsia* as well as *Goudaea chrysostachys*, *Guzmania berteroniana, Werauhia gladioliflora*, and *Zizkaea tuerckheimii* (**Table 1**). Nine of them represented products of intergeneric hybridization. The boxplots of gamma values from all significant tests for each of the 17 hybrids are shown in **Figure 3**. Nine of these (representing *V. medusae* plus the intergeneric hybrids, except *Vriesea roberto-seidelli)* are centered around 0.5, indicating that these taxa are potentially 50:50 hybrids (**Figure 3**). The remaining hybrids were characterized by a smaller introgression proportion as indicated by their gamma values. Six hybrids had median gamma values between 0.5 and 0.7 and two (*Vriesea friburgensis* and *Zizkaea tuerckheimii)* had median values above 0.7 indicating a low level of admixture (**Figure 3**). *Vriesea billbergioides* and *Tillandsia juncea* were identified as the parental species of the hybrid *Vriesea roberto-seidelli.* All other hybrids had multiple significant triplets involving different P2 and P3, indicating that the parental species of most hybrid taxa could not be unambiguously identified (**Table 1** and **Supplementary Table 1**). When HyDe identifies multiple significant triplets involving several putative parental taxa for a given hybrid, it points to ancestral hybridization which remains detectable in several descendant lineages, even after considerable evolutionary divergence. Although the exact origin of the detected hybrids could not be uncovered, most of them had a signal of admixture biased toward one combination of parental lineages (**Table 1**). For example, the majority of the significant combinations involved a species from *Vriesea s.str* and a species from *Guzmania* in eight out of the nine intergeneric hybrids (**Figure 4**), while four hybrids of *Vriesea* had both parental species from *Vriesea. s. str* in most of their combinations. This prevalence of some lineages in the significant triplets is reflected in **Figure 2** (right), where the thickness and darkness of the arrows is proportional to the number of time a genera was identified as a parental lineage for the hybrids in other lineages. This network diagram clearly depicts *Guzmania* and *Vriesea s.str* as the main putative parents for *Vriesea s.str, Tillandsia, Goudaea,* and *Werauhia* hybrids.

DISCUSSION

Hybridization as a Widespread Phenomenon in Bromeliaceae

Our study points to widespread hybridization in subfamily Tillandsioideae, both within and among genera. By recovering mitochondrial, chloroplast and nuclear DNA in a single sequencing experiment, we were able to clearly demonstrate incongruent evolutionary histories between the three genomic compartments. Discrepancy between organellar and nuclear

TABLE 1 | Frequency of the possible combinations for the generic origin of the parental species for each hybrid across all significant triplets identified by HyDe.

Hybrid (P3)	Genera of parental species (P1,P2)	Number of significant triplets with this combination
Goudaea chrysostachys	*Vriesea s.str.—Guzmania*	21
	Guzmania—Tillandsia	3
	Guzmania—Vriesea Cipuropsis	2
	Guzmania—Zizkaea	2
Guzmania berteroniana	*Vriesea s.str.—Guzmania*	18
	Guzmania—Tillandsia	3
	Guzmania—Zizkaea	2
	Guzmania—Vriesea Cipuropsis	1
Tillandsia geminiflora	*Vriesea s.str.—Guzmania*	23
	Guzmania—Tillandsia	3
	Guzmania—Zizkaea	2
	Guzmania—Vriesea Cipuropsis	1
Tillandsia juncea	*Vriesea s.str.—Guzmania*	27
	Guzmania—Tillandsia	3
	Guzmania—Vriesea Cipuropsis	3
	Guzmania—Zizkaea	3
Tillandsia tenuifolia	*Vriesea s.str.—Vriesea s.str.*	9
	Vriesea s.str.—Tillandsia	2
	Vriesea s.str.—Werauhia	2
	Vriesea s.str.—Goudaea	1
	Vriesea s.str.—Guzmania	1
Vriesea botafoguensis	*Vriesea s.str.—Vriesea s.str.*	4
	Vriesea s.str.—Goudaea	1
	Vriesea s.str.—Tillandsia	1
	Vriesea s.str.—Werauhia	1
Vriesea dubia	*Vriesea s.str.—Vriesea s.str.*	3
	Vriesea s.str.—Tillandsia	2
	Vriesea s.str.—Guzmania	1
Vriesea friburgensis	*Vriesea s.str.—Vriesea s.str.*	2
	Vriesea s.str.—Tillandsia	1
Vriesea medusa	*Tillandsia—Tillandsia*	1
	Tillandsia—Vriesea Cipuropsis	1
	Vriesea s.str.—Tillandsia	1
Vriesea oxapampae	*Vriesea s.str.—Guzmania*	21
	Guzmania—Tillandsia	3
	Guzmania—Zizkaea	2
	Guzmania—Vriesea Cipuropsis	1
Vriesea roberto-seindelli	*Vriesea s.str.—Tillandsia*	1
Vriesea rodigasiana	*Vriesea s.str.—Guzmania*	25
	Guzmania—Tillandsia	3
	Guzmania—Zizkaea	3
Vriesea saundersi	*Vriesea s.str.—Vriesea s.str.*	2
	Vriesea s.str.—Tillandsia	1
Vriesea saxicola	*Vriesea s.str.—Vriesea s.str.*	6
	Vriesea s.str.—Tillandsia	4
	Vriesea s.str.—Goudaea	2
	Vriesea s.str.—Guzmania	2
	Vriesea s.str.—Werauhia	2
Vriesea simulans	*Vriesea s.str.—Guzmania*	21
	Guzmania—Tillandsia	3
	Guzmania—Zizkaea	2
Werauhia gladioliflora	*Vriesea s.str.—Guzmania*	20
	Guzmania—Tillandsia	3
	Guzmania—Zizkaea	1
Zizkaea tuerckheimii	*Vriesea s.str.—Vriesea s.str.*	2
	Vriesea s.str.—Tillandsia	1

phylogenies is common in plants and can be caused either by ILS or hybridization (Naciri and Linder, 2015). By applying a method which uses phylogenetic invariants to detect hybridization in the presence of ILS, we were able to detect a significant level of admixture in the nuclear genome of 17 of the 148 species included in our phylogeny. Four of these (*Goudaea chrysostachys, Vriesea dubia, Werauhia gladioliflora,* and *Zizkaea tuerckheimii)* belong to subtribe Cipuropsidinae and two of these originated from ancestral hybridization between the *Vriesea s.str.* and Tillandsieae lineages (**Figure 4**). This finding supports the hypothesis that hybridization is responsible for the incongruent placement of subtribe Cipuropsidinae which is recovered as either closely related to the Vrieseinae (in the plastid and mitochondrial phylogenies) or alternatively to the Tillandsieae (in the nuclear phylogeny). In bromeliads, a similar example of cyto-nuclear discordance was found in the genus *Puya* and was explained by a mixture of ancient and recent hybridization events (Jabaily and Sytsma, 2010; Schulte et al., 2010). Our results add to the growing body of evidence that hybridization may have been a widespread phenomenon and an important driving force in the evolution of bromeliads. Indeed, interspecific and intergeneric artificial crosses of bromeliads are a common horticultural practice (Vervaeke et al., 2004; de Souza et al., 2017) and several studies report the permeable nature of species boundaries among bromeliads (Palma-Silva et al., 2011; Mota et al., 2019). For example, a study on pre-pollination isolating mechanisms in an assemblage of 42 sympatric species of bromeliads in Brazil found that neither microhabitat preferences, flowering phenology nor pollinators contributed to prezygotic reproductive isolation (Wendt et al., 2008). Another study looked at post-pollination barriers among 13 species of bromeliads and found that despite pollen tube growth inhibition being an important mechanism, nearly 25% of interspecific and intergeneric crosses lead to normal pollen tube growth (Matallana et al., 2016). Additionally, three sympatric species of *Tillandsia* in Mexico showed overlap in flowering time, similarity in reproductive organs morphology and manual crosses among them resulted in high fruit production and viable seeds (Ramírez-Rosas et al., 2020). There is further morphological and genetic evidence for natural introgression or hybrid speciation in several bromeliad genera, including

Aechmea (Goetze et al., 2017), *Alcantarea* (Versieux et al., 2012; Lexer et al., 2016), *Fosterella* (Paule et al., 2017), *Pitcairnia* (Palma-Silva et al., 2011; Mota et al., 2020), *Puya* (Jabaily and Sytsma, 2010; Schulte et al., 2010), *Tillandsia* (Gardner, 1984; Gonçalves and de Azevêdo-Gonçalves, 2009), and *Vriesea* (Matos et al., 2016; Zanella et al., 2016; Neri et al., 2018).

Despite this wealth of evidence for a strong hybridization potential in bromeliads, population genetic studies generally have found low levels of interspecific gene flow in extant populations of bromeliads. For example, in *Vriesea s.str.,* gene flow has been reported between *V. scalaris* and *V. simplex* (Neri et al., 2018) and between *V. carinata* and *V. incurvata* (Zanella et al., 2016) but all species exhibited high genetic structure, a low proportion of admixed individuals in the populations (between 8 and 12%) and low level of introgression. In these cases, the interaction between multiple reproductive barriers such as different reproductive systems, variation in floral traits, temporal flowering differences and low hybrid seed viability may have limited interspecific crosses and helped to maintain species integrity (Zanella et al., 2016; Neri et al., 2017). Similar results have been found in other genera such as *Pitcarnia* (Palma-Silva et al., 2011; Mota et al., 2020). However, these studies were done at the population level and explicitly targeted recent or ongoing gene-flow within small sympatric populations, whereas we focused on the long-term evolutionary signature of introgression across the whole subfamily. Our results indicate that 11% of the species in our dataset exhibit a signature of ancient hybridization with high levels of admixture. Therefore, while exchange of genetic material seems to be restricted to low level of introgression in the extant populations of bromeliads that have been studied so far, our study provides evidence that gene flow occurred between the ancestors of extant lineages of core Tillandsioideae that have considerably diverged from each other. Whether these events represent cases of hybrid speciation or introgressive hybridization remains to be investigated. Nevertheless, our results suggest that hybridization may have been more frequent in bromeliads than what has been suggested so far. If this is the case, it could have played an essential role in promoting speciation and generating diversity in bromeliads over evolutionary time, similarly to what has been demonstrated in other evolutionary radiations. In fact, the emerging view of bromeliad evolution is that the existence of generally strong, yet occasionally permeable reproductive barriers maintains species cohesion, while allowing the spreading of advantageous alleles (Neri et al., 2017). This is in line with the "speciation with gene flow" concept or the idea of the "porous genome," in which some parts of the genome can be easily introgressed while genes essential to species cohesion are resistant to gene flow (Wu, 2001). Uniformity of chromosomal numbers in Tillandsioideae may enhance chromosome rearrangements and recombination, promoting speciation in the presence of gene flow (Rieseberg, 2001).

Although our results suggest that hybridization is a widespread phenomenon in Tillandisoideae, it does not rule out the existence of ILS as an additional source of phylogenetic incongruence. In fact, the lower branch support of the phylogenetic tree inferred using a method which models ILS (SVDquartet), compared to a method which does not (RAxML), suggests that ILS is indeed present in the group. Yet, given that HyDe is based on a theoretical frameworks that considers both ILS and hybridization, the presence of ILS is unlikely to have mislead our inference of hybridization. However, we cannot rule out the possibility that other factors related to data acquisition impacted our analyses. First, index hopping, an inherent source of errors linked to sequencing technologies, can potentially lead to mis-assignment of reads between multiplexed samples. In this study, we used a set of 60 dual-indexed Illumina adapters to limit the redundancy of the indexes and reduce this potential error, as recommended by Costello et al. (2018). Ros-Freixedes et al. (2018) also showed that index hopping, when present, has little impact on SNP call accuracy from low-coverage sequence data, and we are therefore confident that this potential bias has a limited impact on our analyses. Bias toward reference allele due to alignment also can have an effect on SNP calling, even if it was shown to be negligible with genome skimming on pigs (Ros-Freixedes et al., 2018). To limit this bias, we used the *Tillandsia adpressiflora* pseudo-reference genome constructed in de la Harpe et al. (2020) for mapping our samples. The method consisted in incorporating specific variation of a *T. adpressiflora* individual sequenced at high coverage into the high quality *Ananas comosus* genome (Ming et al., 2015). This strategy has the advantage of improving the global mapping efficiency of our samples for both reference and alternative alleles, but we cannot exclude that some of our SNP calls were impacted by biases linked to the quality and distance of the reference genome. In addition, because our genome skimming approach favors the detection of the high copy fraction of the genome, our sequence data may include duplicated genes or transposable elements. It is possible that the recovery of alternative paralogs in different lineages can affect the inference of phylogenetic trees and the estimation of the proportion of introgression. A *Vriesea de novo* genome assembly using long reads sequencing would be very helpful to detect and filter such multicopy genomic regions.

The method that we used to detect introgression has been shown through simulations to perform well in identifying introgressed individuals but may select the wrong parental species in the case of ancestral hybridization, by favoring parental species belonging to the sister clade of the true parent (Kubatko and Chifman, 2019). This is why here we focus on the identification of the hybrids rather than on the exact determination of their origin, which could also be affected the unbalanced taxonomic sampling of our study. Although it is unclear how the limited sampling outside of *Vriesea* may have impacted the detection of hybridization, a deeper sampling of all Tillandsioid genera would allow to further investigate intergeneric hybridization and confirm that the repeated signal for ancestral hybridization between *Vriesea* and tribe Tillandsieae is not an artifact due to the lack of intrageneric sampling. Finally, our results suggest that ancient hybrid speciation has occurred in core Tillandsioids, but the analysis we performed does not allow to discriminate between hybrid speciation and introgressive hybridization. A deeper investigation of the exact origin of the hybrids detected in this study would require additional analyses based on whole genome data in order to carry out a proper test of these two scenarios.

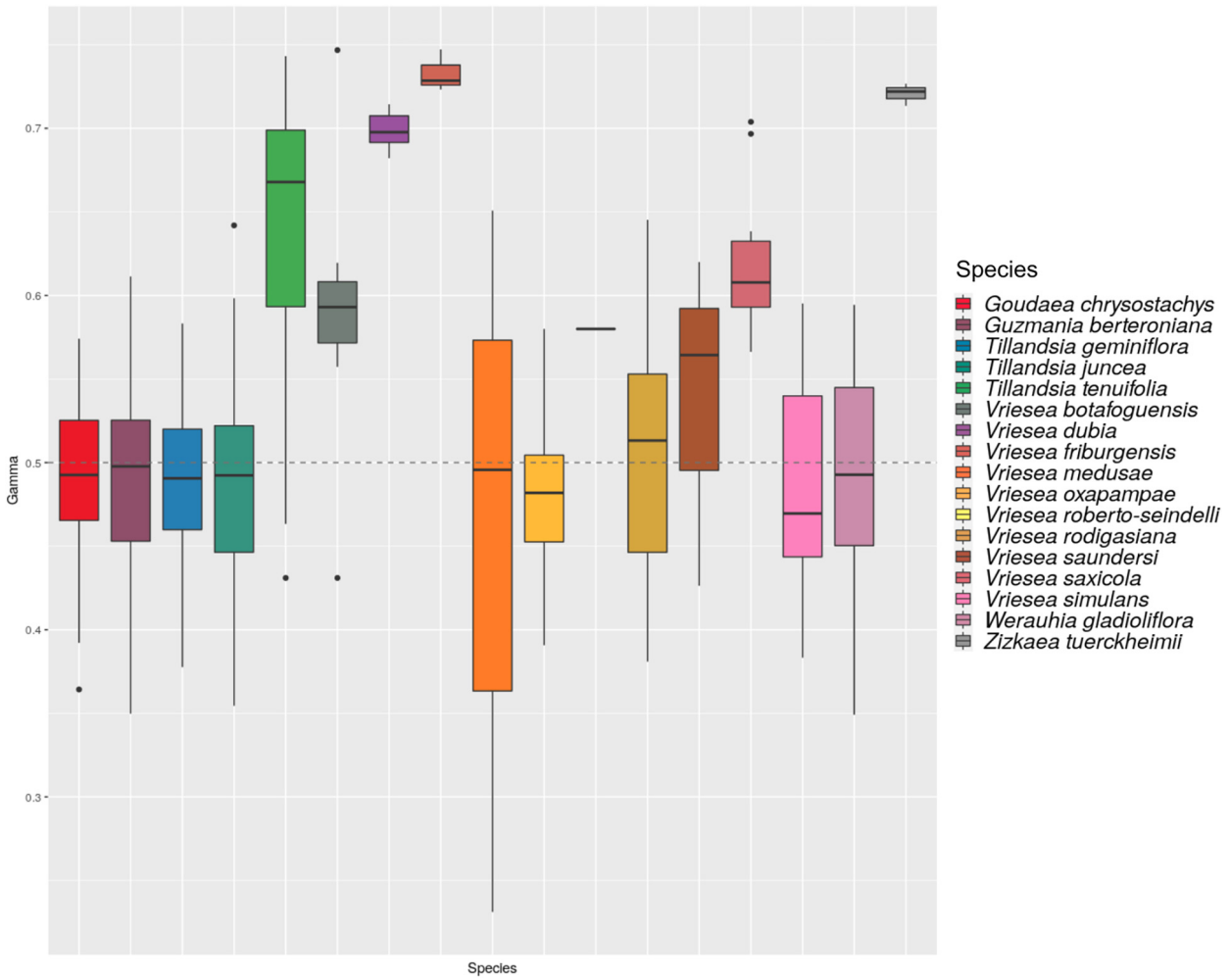


FIGURE 3 | Histogram of the distribution of the gamma statistics across all significant triplets for each detected species. The dashed line indicate the value for 50/50 hybridization.

Biogeographic History of Tillandsioideae and Opportunities for Hybridization

Kessous et al. (2020) inferred a broad ancestral area for the core Tillandsioideae, spreading across the Atlantic Forest (AF), the Andes and the Chacoan dominion (i.e., the South American "Dry Diagonal," Neves et al., 2015). The authors hypothesized that tectonic and climatic events during the Miocene (the formation of the Paranean Sea and the Dry Diagonal) likely caused the vicariance between the Andean lineages (Tillandsieae + Cipuropsidinae) and the Vrieseinae in the AF (Kessous et al., 2020). Although the Andean and Amazonian rainforests are at present separated from the AF by the drier biomes of the Dry Diagonal, there is considerable evidence in the literature that these wet forest biomes were in contact during the Miocene and Pleistocene, allowing for migration and biotic exchange (Batalha-Filho et al., 2013; Trujillo-Arias et al., 2017). Our analysis revealed a strong signal of hybridization between *Vriesea s.str.* and the tribe Tillandsieae in nine species, a finding coherent with the biogeographic history of the group. Indeed, hybridization between *Vriesea s.str.* and *Guzmania* is unlikely to be the result of recent hybridization given that these two lineages diverged from each other more than 12 My ago and have their respective centers of diversity in the Brazilian Atlantic Forest and the Andes. It is therefore plausible that hybridization occurred between ancestral populations of different core-tillandsioid genera with overlapping broad distributions or between populations that had geographically diverged between the Andes and the Amazon but were in contact during the Miocene. More recent hybridization events between *Vriesea s.str* and the Andean lineages are less likely but cannot be ruled out given that *Vriesea* has dispersed to the Amazonian and Andean regions during the Pleistocene (a few species have a disjunct distribution spanning the Dry Diagonal, with others being restricted to the Andes/Amazon and absent from the AF). In contrast, we did not detect hybridization between *Vriesea* and its two most closely genera, *Alcantarea* and *Stigmatodon,* which are also distributed in the

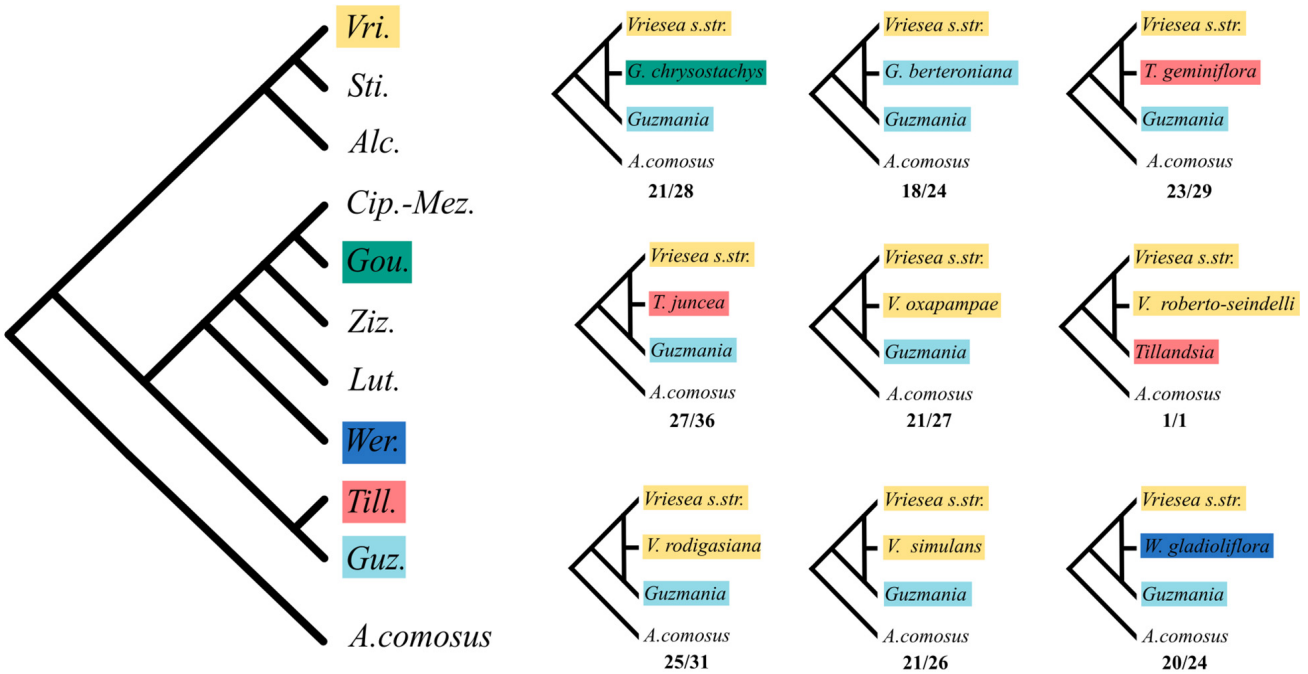


FIGURE 4 | Most likely ancestral intergeneric hybridization events detected by HyDe. **Left:** Cladogram of the intergeneric phylogenetic relationships within core Tillandsioideae. Colored squares show genera where hybrids have been detected. **Right:** Most frequent combinations for the generic origin of the nine inter-generic hybrids inferred by HyDe. Numbers below each quartet indicate the number of time the depicted combination was found over the total number of significant triplets for that hybrid.

AF and have species occuring in sympatry, nor did we find signal for intrageneric hybridization within these two genera. This result could by explained either by our limited taxonomic sampling, insufficient genomic data, limitation of the method used to detect hybridization, or alternatively by the existence of strong reproductive barriers preventing interspecific crosses among *Vriesea*, *Alcantarea* and *Stigmatodon*. Little is known about *Stigmatodon,* a recently segregated genus, but studies of *Alcantarea* have demonstrated the existence of introgression among several *Alcantarea* species, in spite of high population genetic divergence (Barbará et al., 2009; Lexer et al., 2016). However, because our phylogenetic approach does not include intraspecific sampling and uses only a limited part of the genome, it cannot detect such geographically localized and low level of interspecific gene flow. Despite a difference in style length which could act as a prezygotic reproduction barrier, natural hybridization between *Vriesea* and *Alcantarea* cannot be ruled out given that it is at least experimentally feasible under certain conditions (e.g., with *Alcantarea* as the male donor; de Souza et al., 2017). The main habitat of *Alcantarea* and *Stigmatodon* are inselbergs, considered terrestrial islands due to their ecological and spatial isolation which reduces the connectivity between populations (Porembski, 2007). It has been proposed that low level of introgressive hybridization could contribute to the maintenance of the genetic diversity of these isolated populations, thereby balancing the lack of intraspecific gene flow observed in inselberg bromeliad species (Palma-Silva et al., 2011; Mota et al., 2019). Thus, considering that *Alcantarea, Stigmatodon* and *Vriesea* are often found in sympatry in the Atlantic Forest inselbergs, hybridization between them seems at least plausible. Further investigation is required to elucidate whether or not gene flow between these lineages occurred at some point during their evolutionary history.

Bromeliads Phylogenomics

Bromeliaceae are known for their low rate of molecular evolution compared to other Poales (Gaut et al., 1992; Smith and Donoghue, 2008), resulting in a lack of resolution in species-level phylogenies reconstructed from a small number of plastid and sometimes nuclear markers (e.g., Sass and Specht, 2010; Versieux et al., 2012; Goetze et al., 2016). In this study, we aimed at circumventing this limitation by using genome skimming, a method which allows the obtention of a large amount of DNA from different genomic compartments without the need for upstream marker development. We obtained thousands of SNPs from the chloroplast, mitochondrial and nuclear genomes for nearly 150 species of Tillandsioideae, representing the largest genomic dataset for bromeliads to date. Using these data to infer a phylogeny, we were able to confidently resolve intergeneric relationships, indicating that the method is well-adapted for intermediate levels of divergence in bromeliads. However, within genera, branch support values ranged from high (*Alcantarea*) to intermediate (the *Stigmatodon* clade), to very low (*Vriesea* s.str.). This lack of power is particularly critical in our chloroplast dataset where only 8.3% of sites are variable despite a high sequencing depth, 53.9x on average, obtained with our genome

skimming method. Chloroplast genes have been the most common markers used for phylogenetics analyses in bromeliads so far, but our results suggest that even the full plastome would be unlikely to contain enough information to resolve phylogenetic relationships within *Vriesea*. Simpson et al. (2017) faced a similar problem using genome skimming approach to obtain plastid genomes for the phylogenomic analysis of a large clade of Boraginaceae. Coissac et al. (2016) suggest that entire plastid genomes obtained with genome skimming would not be enough for "difficult" groups, and that the obtention of hundreds of nuclear loci is likely to be required. They suggested that shallow-pass nuclear loci obtained with genome skimming could provide enough phylogenetic informative data in such cases. Despite considerable genomic data in our study, we obtained low support within the focal clade *Vriesea s.str*. This is probably due to (1) the substantial amount of missing data (c. 22%) in the alignment of nuclear SNPs, and (2) the fact that only 40% of the nuclear SNPs are parsimony-informative. Indeed, even though 48% of our detected high-quality nuclear sites are polymorphic, SNPs were found mainly at low frequencies (median frequency of 5.9%) and were rarely informative. Hence, while being advantageous in term of cost and time, genome skimming is probably not the most efficient approach for shallow-scale phylogenetic studies in Bromeliaceae, particularly in young and speciose clades such as *Vriesea*, which diversified into c. 300 species in less than 8 My. Obtaining more informative nuclear SNPs, would require increasing the sequencing depth of each sample and therefore dramatically increasing the cost for large scale phylogenomic studies. For these reasons, we argue that more specific approaches such as target-capture sequencing of low-copy nuclear genes could be the way forward to obtain better-resolved phylogenies of large bromeliad genera such as *Guzmania*, *Tillandsia* or *Vriesea*. Even if the development of such target-capture kits is complex and challenging (de La Harpe et al., 2019; Andermann et al., 2020), it would benefit a wide research community working on many fascinating groups of bromeliads.

CONCLUSION

By generating an unprecedented genomic dataset for the largest sub-family of Bromeliaceae, our study sheds lights on the evolutionary history of an important floristic component of Neotropical rainforests. We pinpointed the incongruence of the nuclear phylogeny with the chloroplast and mitochondrial phylogenetic trees regarding the position of the subtribe Cipuropsidinae.

Our finding of 17 substantially introgressed taxa, potentially the products of hybrid speciation, in our dataset of 116,478 nuclear SNPs for 148 species of Tillandsioideae suggest that hybridization is a plausible explanation for this incongruence. Furthermore, the signal for hybridization between the ancestral lineages of *Vriesea* and tribe Tillandsieae suggests that hybridization may be widespread phenomenon in core tillandsioids, and to a greater extent, in bromeliads. Thus, our study adds to the growing body of evidence that hybridization is ubiquitous across this diverse Neotropical plant family and may have fueled the diversification of the most diverse clades of Bromeliaceae such as the core Tillandsioideae and core Bromelioideae. Future genomic studies with deeper sequencing across a wide taxonomic range in bromeliads would likely reveal more hybridization and help to tell apart the respective contribution of hybrid speciation and adaptive introgression to the evolution of bromeliad diversity. Further joint investigation of hybridization at the micro- and macro- evolutionary level will be necessary to clarify the exact mechanisms through which it may have promoted genetic diversity, adaptation and speciation, and ultimately contributed to the adaptive radiation and ecological success of bromeliads.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: NCBI (accession: PRJNA554504).

AUTHOR CONTRIBUTIONS

NS and CL designed the study. TM performed fieldwork and sample collection. MP led the sequencing experiments and post sequencing bioinformatics. TM and MP did the labwork. DK and OL did phylogenetic analyses. OL did hybridization test and led the writing with significant contributions from all co-authors. All authors commented and agreed on the last version of the manuscript.

FUNDING

This work was funded by the Sinergia grant (CRSII3-147630) from the Swiss National Science Foundation to NS and CL as well as funding from the University of Lausanne. TM was supported by Ph.D. fellowship grants from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and CNPq [Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPq-SWE (205660/2014-2)/CNPq (142354/2016-3)]. LV received funding from the CNPq (455510/2014-8 and 303794/2019-4). Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures, and by the computing infrastructure (DCSR) of the University of Lausanne.

ACKNOWLEDGMENTS

The authors thank the Botanical Gardens of the University of Vienna (WU-HBV), Botanical Garden of Rio de Janeiro (RBvb), Marie Selby Botanical Garden (SEL), and Jardin des Serres d'Auteuil (P) for providing plant material, ICMBIO, and IEF-MG for collection permits. TM thanks all the amazing people who collaborated in the fieldwork and other stages of this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.668281/full#supplementary-material

Supplementary Figure 1 | Phylogenetic tree of core Tillandsioideae inferred with RaxML using the concatenated mitochondrial SNPs.

Supplementary Figure 2 | Phylogenetic tree of core Tillandsioideae inferred with RaxML using the concatenated chloroplast SNPs.

Supplementary Figure 3 | Phylogenetic tree of core Tillandsioideae inferred with RaxML using the concatenated nuclear SNPs.

Supplementary Figure 4 | Phylogenetic tree of core Tillandsioideae inferred with SVDquartet using the concatenated nuclear SNPs.

Supplementary Table 1 | List of all significant triplets identified by HyDe.

REFERENCES

Andermann, T., Torres Jiménez, M. F., Matos-Maraví, P., Batista, R., Blanco-Pastor, J. L., Gustafsson, A. L. S., et al. (2020). A guide to carrying out a phylogenomic target sequence capture project. *Front. Genet.* 10:1407.

Barbará, T., Martinelli, G., Palma-Silva, C., Fay, M. F., Mayo, S., and Lexer, C. (2009). Genetic relationships and variation in reproductive strategies in four closely related bromeliads adapted to neotropical 'inselbergs': *Alcantarea glaziouana, A. regina, A. geniculata* and *A. imperialis* (Bromeliaceae). *Ann. Bot.* 103, 65–77. doi: 10.1093/aob/mcn226

Barrier, M., Baldwin, B. G., Robichaux, R. H., and Purugganan, M. D. (1999). Interspecific hybrid ancestry of a plant adaptive radiation: allopolyploidy of the *Hawaiian silversword* alliance (Asteraceae) inferred from floral homeotic gene duplications. *Mol. Biol. Evol.* 16, 1105–1113. doi: 10.1093/oxfordjournals.molbev.a026200

Batalha-Filho, H., Fjeldså, J., Fabre, P.-H., and Miyaki, C. Y. (2013). Connections between the Atlantic and the Amazonian forest avifaunas represent distinct historical events. *J. Ornithol.* 154, 41–50. doi: 10.1007/s10336-012-0866-7

Benzing, D. H. (2000). *Bromeliaceae: Profile of an Adaptive Radiation*. Cambridge: Cambridge University Press.

Blanco-Pastor, J. L., Vargas, P., and Pfeil, B. E. (2012). Coalescent simulations reveal hybridization and incomplete lineage sorting in Mediterranean Linaria. *PLoS One* 7:e39089. doi: 10.1371/journal.pone.0039089

Blischak, P. D., Chifman, J., Wolfe, A. D., and Kubatko, L. S. (2018). HyDe: a Python package for genome-scale hybridization detection. *Syst. Biol.* 9, 821–829. doi: 10.1093/sysbio/syy023

Bouchenak-Khelladi, Y., Muasya, A. M., and Linder, H. P. (2014). A revised evolutionary history of poales: origins and diversification: evolutionary history of poales. *Bot. J. Linn. Soc.* 175, 4–16. doi: 10.1111/boj.12160

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., and RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol. Biol. Evol.* 29, 1917–1932. doi: 10.1093/molbev/mss086

Chifman, J., and Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30, 3317–3324. doi: 10.1093/bioinformatics/btu530

Chifman, J., and Kubatko, L. (2015). Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *J. Theor. Biol.* 374, 35–47. doi: 10.1016/j.jtbi.2015.03.006

Coissac, E., Hollingsworth, P. M., Lavergne, S., and Taberlet, P. (2016). From barcodes to genomes: extending the concept of DNA barcoding. *Mol. Ecol.* 25, 1423–1428. doi: 10.1111/mec.13549

Costello, M., Fleharty, M., Abreu, J., Farjoun, Y., Ferriera, S., Holmes, L., et al. (2018). Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. *BMC Genomics* 19:332.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330

de La Harpe, M., Hess, J., Loiseau, O., Salamin, N., Lexer, C., and Paris, M. (2019). A dedicated target capture approach reveals variable genetic markers across micro- and macro-evolutionary time scales in palms. *Mol. Ecol. Resour.* 19, 221–234. doi: 10.1111/1755-0998.12945

de la Harpe, M., Paris, M., Hess, J., Barfuss, M. H. J., Serrano-Serrano, M. L., Ghatak, A., et al. (2020). Genomic footprints of repeated evolution of CAM photosynthesis in a Neotropical species radiation. *Plant Cell Environ.* 43, 2987–3001. doi: 10.1111/pce.13847

de Souza, E. H., Versieux, L. M., Souza, F. V. D., Rossi, M. L., Costa, M. A. P., and Martinelli, A. P. (2017). Interspecific and intergeneric hybridization in Bromeliaceae and their relationships to breeding systems. *Sci. Hortic.* 223, 53–61. doi: 10.1016/j.scienta.2017.04.027

Dodsworth, S. (2015). Genome skimming for next-generation biodiversity analysis. *Trends Plant Sci.* 20, 525–527. doi: 10.1016/j.tplants.2015.06.012

Epskamp, S., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D., and Borsboom, D. (2012). qgraph: network visualizations of relationships in psychometric data. *J. Stat. Softw.* 48, 1–18.

Gardner, C. S. (1984). Natural hybridization in "Tillandsia" subgenus "Tillandsia". *Selbyana* 7, 380–393.

Gaut, B. S., Muse, S. V., Clark, W. D., and Clegg, M. T. (1992). Relative rates of nucleotide substitution at the rbcL locus of monocotyledonous plants. *J. Mol. Evol.* 35, 292–303. doi: 10.1007/bf00161167

Givnish, T. J., Barfuss, M. H. J., Ee, B. V., Riina, R., Schulte, K., Horres, R., et al. (2014). Molecular Phylogenetics and evolution adaptive radiation, correlated and contingent evolution, and net species diversification in Bromeliaceae. *Mol. Phylogenet. Evol.* 71, 55–78. doi: 10.1016/j.ympev.2013.10.010

Goetze, M., Schulte, K., Palma-Silva, C., Zanella, C. M., Büttow, M. V., Capra, F., et al. (2016). Diversification of Bromelioideae (Bromeliaceae) in the Brazilian Atlantic rainforest: a case study in *Aechmea subgenus* Ortgiesia. *Mol. Phylogenet. Evol.* 98, 346–357. doi: 10.1016/j.ympev.2016.03.001

Goetze, M., Zanella, C. M., Palma-Silva, C., Büttow, M. V., and Bered, F. (2017). Incomplete lineage sorting and hybridization in the evolutionary history of closely related, endemic yellow-flowered Aechmea species of subgenus Ortgiesia (Bromeliaceae). *Am. J. Botany* 104, 1073–1087. doi: 10.3732/ajb.1700103

Gonçalves, C. N., and de Azevêdo-Gonçalves, C. F. (2009). A new hybrid bromeliad from southernmost Brazil, Tillandsia ×baptistana. *Novon J. Bot. Nomenclature* 19, 353–356. doi: 10.3417/2001154

Gouda, E. J., Butcher, D., and Gouda, C. S. (2018). *Encyclopaedia of Bromeliads, Version 4*. Utrecht: University Botanic Gardens.

Grant, P. R., and Grant, B. R. (2019). Hybridization increases population variation during adaptive radiation. *Proc. Natl. Acad. Sci. USA* 116:23216.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., et al. (2010). A draft sequence of the neandertal genome. *Science* 328, 710–722.

Harrison, R. G., and Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *J. Heredity* 105, 795–809. doi: 10.1093/jhered/esu033

Irisarri, I., Singh, P., Koblmüller, S., Torres-Dowdall, J., Henning, F., Franchini, P., et al. (2018). Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake *Tanganyika cichlid* fishes. *Nat. Commun.* 9:3159.

Jabaily, R. S., and Sytsma, K. J. (2010). Phylogenetics of Puya (Bromeliaceae): placement, major lineages, and evolution of Chilean species. *Am. J. Bot.* 97, 337–356. doi: 10.3732/ajb.0900107

Joly, S., McLenachan, P. A., and Lockhart, P. J. (2009). A statistical approach for distinguishing hybridization and incomplete lineage sorting. *Am. Nat.* 174, E54–E70.

Kessous, I. M., Neves, B., Couto, D. R., Paixão-Souza, B., Pederneiras, L. C., Moura, R. L., et al. (2020). Historical biogeography of a Brazilian lineage of Tillandsioideae (subtribe Vrieseinae, Bromeliaceae): the Paranaean Sea hypothesized as the main vicariant event. *Bot. J. Linn. Soc.* 192, 625–641.

Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* 40:e3. doi: 10.1093/nar/gkr771

Kozak, K. M., Joron, M., McMillan, W. O., and Jiggins, C. D. (2021). Rampant genome-wide admixture across the *Heliconius* radiation. *Genome Biol. Evol.* evab099. doi: 10.1093/gbe/evab099

Kubatko, L. S., and Chifman, J. (2019). An invariants-based method for efficient identification of hybrid species from large-scale genomic data. *BMC Evol. Biol.* 19:112.

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357. doi: 10.1038/nmeth.1923

Leaché, A. D., Banbury, B. L., Felsenstein, J., de Oca, A. N., and Stamatakis, A. (2015). Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Syst. Biol.* 64, 1032–1047. doi: 10.1093/sysbio/syv053

Lexer, C., Marthaler, F., Humbert, S., Barbará, T., de la Harpe, M., Bossolini, E., et al. (2016). Gene flow and diversification in a species complex of Alcantarea inselberg bromeliads. *Bot. J. Linn. Soc.* 181, 505–520. doi: 10.1111/boj.12372

Loiseau, O., Olivares, I., Paris, M., de La Harpe, M., Weigand, A., Koubínová, D., et al. (2019). Targeted capture of hundreds of nuclear genes unravels phylogenetic relationships of the diverse neotropical Palm Tribe Geonomateae. *Front. Plant Sci.* 10:864.

Machado, T. M., Loiseau, O., Paris, M., Weigand, A., Versieux, L. M., Stehmann, J. R., et al. (2020). Systematics of Vriesea (Bromeliaceae): phylogenetic relationships based on nuclear gene and partial plastome sequences. *Bot. J. Linn. Soc.* 192, 656–674. doi: 10.1093/botlinnean/boz102

Marques, D. A., Meier, J. I., and Seehausen, O. (2019). A combinatorial view on speciation and adaptive radiation. *Trends Ecol. Evol.* 34, 531–544. doi: 10.1016/j.tree.2019.02.008

Matallana, G., Oliveira, P. E., Silva, P. R. R., and Wendt, T. (2016). Post - pollination barriers in an assemblage of Bromeliaceae in south - eastern Brazil. *Bot. J. Linn. Soc.* 181, 521–531. doi: 10.1111/boj.12406

Matos, J. Z., Juan, A., Agulló, J. C., and Crespo, M. B. (2016). Morphological features, nuclear microsatellites and plastid haplotypes reveal hybridisation processes between two sympatric Vriesea species in Brazil (Bromeliaceae). *Phytotaxa* 261, 58–74. doi: 10.11646/phytotaxa.261.1.2

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., and Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* 8:14363.

Ming, R., VanBuren, R., Wai, C. M., Tang, H., Schatz, M. C., Bowers, J. E., et al. (2015). The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.* 47, 1435–1442. doi: 10.1038/ng.3435

Mota, M. R., Pinheiro, F., Leal, B. S. S., Wendt, T., and Palma-Silva, C. (2019). The role of hybridization and introgression in maintaining species integrity and cohesion in naturally isolated inselberg bromeliad populations. *Plant Biol.* 21, 122–132. doi: 10.1111/plb.12909

Mota, M. R., Pinheiro, F., Leal, B. S., Sardelli, C. H., Wendt, T., and Palma-Silva, C. (2020). From micro- to macroevolution: insights from a Neotropical bromeliad with high population genetic structure adapted to rock outcrops. *Heredity* 125, 353–370. doi: 10.1038/s41437-020-0342-8

Naciri, Y., and Linder, H. P. (2015). Species delimitation and relationships: the dance of the seven veils. *Taxon* 64, 3–16. doi: 10.12705/641.24

Neri, J., Wendt, T., Leles, B., dos Santos, M. F., and Palma-Silva, C. (2017). Variation in reproductive systems facilitates species boundaries of sympatric *Vriesea* (Bromeliaceae). *Bot. J. Linn. Soc.* 184, 272–279. doi: 10.1093/botlinnean/box026

Neri, J., Wendt, T., and Palma-Silva, C. (2018). Natural hybridization and genetic and morphological variation between two epiphytic bromeliads. *AoB Plants* 10:lx061.

Neves, D. M., Dexter, K. G., Pennington, R. T., Bueno, M. L., and Filho, A. T. O. (2015). Environmental and historical controls of floristic composition across the South American dry diagonal. *J. Biogeogr.* 42, 1566–1576. doi: 10.1111/jbi.12529

Palma-Silva, C., Leal, B. S. S., Chaves, C. J. N., and Fay, M. F. (2016). Advances in and perspectives on evolution in Bromeliaceae. *Bot. J. Linn. Soc.* 181, 305–322. doi: 10.1111/boj.12431

Palma-Silva, C., Wendt, T., Pinheiro, F., Barbará, T., Fay, M. F., Cozzolino, S., et al. (2011). Sympatric bromeliad species (Pitcairnia spp.) facilitate tests of mechanisms involved in species cohesion and reproductive isolation in Neotropical inselbergs. *Mol. Ecol.* 20, 3185–3201. doi: 10.1111/j.1365-294x.2011.05143.x

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., et al. (2012). Ancient admixture in human history. *Genetics* 192, 1065–1093. doi: 10.1534/genetics.112.145037

Paule, J., Wagner, N. D., Weising, K., and Zizka, G. (2017). Ecological range shift in the polyploid members of the South American genus Fosterella (Bromeliaceae). *Ann. Bot.* 120, 233–243.

Pfennig, K. S. (2021). Biased hybridization and its impact on adaptive introgression. *Trends Ecol. Evol.* 36, 488–497.

Porembski, S. (2007). Tropical inselbergs: habitat types, adaptive strategies and diversity patterns. *Braz. J. Bot.* 30, 579–586. doi: 10.1590/s0100-84042007000400004

Ramírez-Rosas, K., Aguirre-Jaimes, A., Ramírez-Morillo, I. M., and García-Franco, J. G. (2020). Floral biology and potential hybridization of three sympatric epiphytic bromeliads in Veracruz, Mexico. *Plant Species Biol.* 35, 197–209. doi: 10.1111/1442-1984.12279

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* 16, 351–358. doi: 10.1016/s0169-5347(01)02187-5

Ros-Freixedes, R., Battagin, M., Johnsson, M., Gorjanc, G., Mileham, A. J., Rounsley, S. D., et al. (2018). Impact of index hopping and bias towards the reference allele on accuracy of genotype calls from low-coverage sequencing. *Genet. Sel. Evol.* 50:64.

Sass, C., and Specht, C. D. (2010). Phylogenetic estimation of the core Bromelioids with an emphasis on the genus Aechmea (Bromeliaceae). *Mol. Phylogenet. Evol.* 55, 559–571. doi: 10.1016/j.ympev.2010.01.005

Schulte, K., Silvestro, D., Kiehlmann, E., Vesely, S., Novoa, P., and Zizka, G. (2010). Detection of recent hybridization between sympatric Chilean Puya species (Bromeliaceae) using AFLP markers and reconstruction of complex relationships. *Mol. Phylogenet. Evol.* 57, 1105–1119. doi: 10.1016/j.ympev.2010.09.001

Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., et al. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* 360, 656–660. doi: 10.1126/science.aar3684

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends Ecol. Evol.* 19, 198–207. doi: 10.1016/j.tree.2004.01.003

Seehausen, O. (2013). Conditions when hybridization might predispose populations for adaptive radiation. *J. Evol. Biol.* 26, 279–281. doi: 10.1111/jeb.12026

Silvestro, D., Zizka, G., and Schulte, K. (2014). Disentangling the effects of key innovations on the diversification of bromelioideae (Bromeliaceae). *Evolution* 68, 163–175. doi: 10.1111/evo.12236

Simpson, M. G., Guilliams, C. M., Hasenstab-Lehman, K. E., Mabry, M. E., and Ripma, L. (2017). Phylogeny of the popcorn flowers: use of genome skimming to evaluate monophyly and interrelationships in subtribe Amsinckiinae (Boraginaceae). *Taxon* 66, 1406–1420. doi: 10.12705/666.8

Smith, L., and Downs, R. (1974). *Flora Neotropica Monograph no. 14.(Pitcairnioideae)(Bromeliaceae)*. New York, NY: Hafner Publishing Company.

Smith, S. A., and Donoghue, M. J. (2008). Rates of molecular evolution are linked to life history in flowering plants. *Science* 322, 86–89. doi: 10.1126/science.1163197

Smith, S. A., and O'meara, B. C. (2012). treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28, 2689–2690. doi: 10.1093/bioinformatics/bts492

Soltis, P. S., and Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* 60, 561–588. doi: 10.1146/annurev.arplant.043008.092039

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Svardal, H., Quah, F. X., Malinsky, M., Ngatunga, B. P., Miska, E. A., Salzburger, W., et al. (2020). Ancestral hybridization facilitated species diversification in the lake Malawi Cichlid fish adaptive radiation. *Mol. Biol. Evol.* 37, 1100–1113. doi: 10.1093/molbev/msz294

Swofford, D. L. (2001). *Paup*: Phylogenetic Analysis Using Parsimony (and other Methods) 4.0. B5*. Princeton: CiteSeerX.

Trujillo-Arias, N., Dantas, G. P. M., Arbeláez-Cortés, E., Naoki, K., Gómez, M. I., Santos, F. R., et al. (2017). The niche and phylogeography of a passerine reveal the history of biological diversification between the Andean and the Atlantic forests. *Mol. Phylogenet. Evol.* 112, 107–121. doi: 10.1016/j.ympev.2017.03.025

Versieux, L. M., Barbará, T., Wanderley, M., das, G. L., Calvente, A., Fay, M. F., et al. (2012). Molecular phylogenetics of the Brazilian giant bromeliads (Alcantarea, Bromeliaceae): implications for morphological evolution and biogeography. *Mol. Phylogenet. Evol.* 64, 177–189. doi: 10.1016/j.ympev.2012.03.015

Vervaeke, I., Wouters, J., Londers, E., Deroose, R., and Proft, M. P. D. (2004). Morphology of artificial hybrids of *Vriesea splendens* x *Tillandsia cyanea* and V . splendens x *Guzmania lingulata* (Bromeliaceae). *Ann. Bot. Fennici* 1840, 201–208.

Wagner, N. D., Wöhrmann, T., Öder, V., Burmeister, A., and Weising, K. (2015). Reproduction biology and chloroplast inheritance in Bromeliaceae: a case study in Fosterella (Pitcairnioideae). *Plant Syst. Evol.* 301, 2231–2246. doi: 10.1007/s00606-015-1226-x

Wendt, T., Coser, T. S., Matallana, G., and Guilherme, F. A. G. (2008). An apparent lack of prezygotic reproductive isolation among 42 sympatric species of Bromeliaceae in southeastern Brazil. *Plant Syst. Evol.* 275, 31–41. doi: 10.1007/s00606-008-0054-7

Wu, C. (2001). The genic view of the process of speciation. *J. Evol. Biol.* 14, 851–865. doi: 10.1046/j.1420-9101.2001.00335.x

Zanella, C. M., Palma-Silva, C., Goetze, M., and Bered, F. (2016). Hybridization between two sister species of Bromeliaceae: *Vriesea carinata* and V. incurvata. *Bot. J. Linn. Soc.* 181, 491–504. doi: 10.1111/boj.12424

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 15 June 2021
doi: 10.3389/fevo.2021.680318

Geographic Pattern of Bryophyte Species Richness in China: The Influence of Environment and Evolutionary History

*Xiaotong Song[1], Wenzhuo Fang[1,2], Xiulian Chi[3], Xiaoming Shao[1] and Qinggang Wang[1]**

[1] Department of Ecology and Ecological Engineering, College of Resources and Environmental Sciences, and Key Laboratory of Biodiversity and Organic Farming of Beijing City, China Agricultural University, Beijing, China, [2] Ministry of Education, Key Laboratory for Biodiversity Science and Ecological Engineering, Department of Ecology and Evolutionary Biology, School of Life Sciences, Fudan University, Shanghai, China, [3] State Key Laboratory Breeding Base of Dao-di Herbs, National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China

How contemporary environment interacts with macroevolutionary processes to generate the geographic pattern of bryophyte species is still unresolved. China is very rich in bryophytes, with more than 3,000 bryophytes covering 70% of the families in the world. In this study, we assessed the effects of the contemporary environment (average temperature of the coldest season TCQ, precipitation of the warmest season PWQ, and elevational range) and the recent diversification rates (estimated as mean species number per genus, MSG) on the geographical pattern of species richness for bryophytes and two groups (i.e., liverworts and mosses) in China. We compiled the provincial level distribution of bryophyte species and estimated the geographic pattern of the recent diversification rate by MSG for species in China. Univariate, multivariate regressions and path model analyses were used to assess the relationships between species richness, MSG, and their potential environmental drivers. Species richness of all bryophytes and liverworts significantly increased with the increase of MSG, either in regressions or path analyses, indicating that provinces with high bryophyte richness were mainly inhabited by species (especially liverworts) from lineages with particularly high MSG. In contrast, the species richness of mosses was insignificantly decreased with MSG in univariate regression or insignificantly increased with MSG in path analysis. Both species richness and MSG of all bryophytes and liverworts increased with the increase in energy and water availability. In contrast, for mosses, the species richness significantly increased with the increase of energy and water availability, while MSG decreased with the increase of energy and water availability. The MSG of liverworts increase with the increase of elevational range but the MSG of mosses decrease with the increase of elevational range. Our study suggests that the humid tropical and subtropical mountains in China are not only diversity hotspots for bryophytes, but also cradles for high recent diversification of liverworts, and refuges for mosses to hold many monotypic and oligotypic genera.

Keywords: liverworts, mosses, net diversification rate hypothesis, mountains, climate

OPEN ACCESS

Edited by:
Xiaoting Xu,
Sichuan University, China

Reviewed by:
Gang Feng,
Inner Mongolia University, China
Jihong Huang,
Research Institute of Forest Ecology, Environment and Protection, Chinese Academy of Forestry, China

****Correspondence:***
Qinggang Wang
wangqg@cau.edu.cn

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *14 March 2021*
Accepted: *19 May 2021*
Published: *15 June 2021*

Citation:
Song X, Fang W, Chi X, Shao X and Wang Q (2021) Geographic Pattern of Bryophyte Species Richness in China: The Influence of Environment and Evolutionary History. Front. Ecol. Evol. 9:680318. doi: 10.3389/fevo.2021.680318

INTRODUCTION

The huge spatialvariation in species richness and mechanisms underlying it have always been a hot topic of concern to ecologists and biogeographers (Rosenzweig, 1995). The contemporary environmental factors such as energy, water availability and habitat heterogeneity have been regarded as potential drivers of species richness pattern (Allen et al., 2002; Currie et al., 2004; Stein et al., 2014; Rana et al., 2019). However, the species richness of a region is ultimately determined by the evolutionary processes of speciation, dispersal and extinction (Mittelbach et al., 2007). The positive species richness-energy/water/habitat heterogeneity relationships must reflect evolutionary history (Rana et al., 2019). For example, regions with high energy and water availability have high primary productivity to expand the population size of animals and plants, and thus contains more genetic diversity to increase the rate of speciation (Hubbell, 2001; Allen et al., 2006; Clarke and Gaston, 2006). Habitat heterogeneity also can increase species richness by promoting the speciation rate and reducing species extinction (Turner, 2004; Fjeldsa et al., 2012; Hughes and Atchison, 2015).

Recently, the diversification rate hypothesis explaining species diversity patterns in terms of evolutionary processes has received renewed attention from researchers (Svenning et al., 2008; Condamine et al., 2012; Kozak and Wiens, 2016; Cai et al., 2020). According to the diversification rate hypothesis, species-rich areas have a high rate of net diversification owing to their higher speciation rates and/or lower extinction rates (Fischer, 1960). The high richness in low latitudes would result from the high diversification rate in warm and wet climates and/or in tropical and subtropical mountains due to fast molecular evolution or strengthened reproductive isolation (Steinbauer et al., 2016; Xing and Ree, 2017; Dagallier et al., 2020). Therefore, the diversification rate hypothesis would predict a positive species richness-diversification rate relationship, and positive diversification rate-climate or habitat heterogeneity relationships.

In contrast of the diversification rate hypothesis, the time-for-speciation hypothesis believes that species richness reflects the available time for diversification, and tropical regions have more species because they are occupied for a longer time, which provide more time for speciation (Fischer, 1960; Rohde, 1992; Dynesius and Jansson, 2000). According to the time for speciation hypothesis, old relic clades with low net diversification (high extinction and/or low speciation) rate mainly occupy in long-inhabited species-rich regions (Mittelbach et al., 2007), which causes that mean net diversification rates are not high in species-rich region. Therefore, the time for speciation would expect a positive or no relationship between species richness and net diversification rate, and negative relationships between diversification rate and climate or habitat heterogeneity. Previous studies have tested the effects of diversification rate hypothesis and/or time-for-speciation hypothesis on animals (e.g., amphibians, Stephens and Wiens, 2003; birds, Hawkins et al., 2005; marine bivalves, Jablonski et al., 2006; salamanders, Kozak and Wiens, 2016) and vascular plants (e.g., Arecaceae, Svenning et al., 2008; Rhododendron, Shrestha et al., 2018; Zygophyllaceae, Wang et al., 2018). However, few studies have evaluated how contemporary environments interacted with evolutionary processes to generate the geographic pattern of bryophyte species (Laenen et al., 2018).

Bryophytes, including about 20,000 species of hornworts, liverworts and mosses, are the second largest group of higher plants after angiosperms (Patiño and Vanderpoorten, 2018). They are widely distributed in terrestrial ecosystems (St Martin and Mallik, 2017). As the earliest land plants, the survival, growth and reproduction of bryophytes have strong dependence on water availibity, but they have considerable cold tolerance (Proctor et al., 2007; Vitt et al., 2014; Perera-Castro et al., 2020). Compared with hornworts and liverworts, mosses have better drought and cold tolerance, and many of them can occur in dry and/or cold climates (e.g., *Bryum pseudotriquetrum* and *Sanionia uncinate*, Cannone et al., 2017). It has been suggested that the low species richness of extant bryophyte might have resulted from mass extinction event, but their recent diversification rate since mid-Mesozoic were high (Laenen et al., 2014).

China is one of the countries with the richest species of bryophytes, with more than 3000 species of bryophytes covering 70% of the families in the world (Chen et al., 2015; Qian and Chen, 2016). The vast land area of mainland China has created enormous regional differences in climate and habitat. In this study, we complied the province level distribution of extant bryophyte species, and attempted to assess the effects of the contemporary environment (temperature, precipitation and habitat heterogeneity) and recent diversification rates (estimated as the mean species number per genus, MSG) on the geographical pattern of species richness for bryophytes and two groups (i.e., liverworts and mosses) in China. We also examined the associations between environmental factors and MSG.

DATA AND METHODS

Data

To estimate the province-level species richness pattern of bryophytes, we combined each municipality except Chongqing which has a very large area with one of its neighboring provinces. Specifically, we combined Beijing and Tianjin with Hebei Province, Shanghai with Zhejiang Province, and Hong Kong and Macau with Guangdong Province. On the final map, China was divided into 29 province-level geographic units ("province" hereafter) (see **Supplementary Figure 1**).

We compiled the distribution data of bryophytes in China from the online database of Catalog of Life China (CoLChina, http://www.sp2000.org.cn/CoLChina) and recent literatures (see **Supplementary Data**). We standardized nomenclature of species according to *CoLChina*. Finally, this study included a total of 3184 bryophytes in 601 genera and 157 families. Those bryophytes included 26 Hornworts in 8 genera and 3 families, 1102 liverworts in 155 genera and 63 families, and 2056 mosses in 438 genera and 91 families. Because of the small number of hornworts species, we incorporated them into liverworts for subsequent analyses. To eliminate the effect of area on species richness assessment, we calculated the area-correlated species richness as species number of a province being divided by the log-transformed area of that

province ("species richness" hereafter). Up to now, the dated phylogeny of bryophytes only covered limited species and genera (Laenen et al., 2014, 2018; Liu et al., 2019). To represent the net diversification rate within genera, we used the species number per genus following previous study (Svenning et al., 2008). The geographic pattern of diversification rate was estimated by MSG for the species occurring in each province.

Environmental variables included the average temperature of the coldest season (TCQ), the precipitation of the warmest season (PWQ), and the elevational range in each province. TCQ represented environmental energy and emphasized cold tolerance effects (Wang et al., 2011), PWQ represented water availability in growing season which is important for bryophytes as suggested by previous studies (Qian and Chen, 2016), and ER was frequently used to represent habitat heterogeneity. ER weakly correlates with TCQ and PWQ ($r = -0.40$ and -0.42), and TCQ moderate correlates with PWQ ($r = 0.76$, **Supplementary Table 2**). The data for TCQ and PWQ with a spatial resolution of ca 1×1 km^2 were downloaded from the *Worldclim* dataset[1] (Hijmans et al., 2005). We calculated each climate variable in each province as the average of all 1×1 km^2 grids in that province. We calculated elevational range as the difference between the minimum and maximum elevations of a province using the GTOPO30 dataset of the US Geographical Survey. To improve the normality, we log-transformed the elevational range data.

Statistical Analyses

The square root-transformed species richness and MSG for all bryophytes, liverworts or mosses within provinces were approximately normally distributed with skewness and kurtosis $<|1.0|$. Therefore, we conducted general linear models (GLM) to test the influence of MSG on species richness pattern of all bryophytes, liverworts or mosses. To handle the effects of spatial autocorrelation in significance testing, we conducted a modified *t*-test to test the significance of the GLMs (Clifford et al., 1989). The test corrects the correlation coefficient between response variable and predictor, and estimates the effective sample size that allows for the spatial structure by introducing estimated covariance matrices for distance classes (Dutilleul et al., 1993). The modified *t*-test was done using "modified.ttest" function in R package of *SpatialPack* (Osorio et al., 2020).

To evaluate the effects of three individual environmental variables (i.e., TCQ, PWQ, and elevational range) on geographical patterns in species richness and MSG for all bryophytes, liverworts or mosses, we ran GLMs using modified *t*-tests for significance testing. In addition, we built regression models with all possible combinations of three environmental predictors done model ranking based on the Bayesian Information Criterion (BIC), and conducted model averaging. We extracted predictor's importance values which were computed as the sum of the weights of models containing the corresponding predictors. Those analyses were done using "dredge" and "model.avg" functions in R package of *MuMIn* (Bartoń, 2020).

To further compare the effects of TCQ, PWQ, elevational range, and MSG on geographical pattern of species richness, we constructed a path model by assuming that MSG influence species richness directly, while TCQ, PWQ, and elevational range can influence species richness directly and also indirectly via its effects on MSG. To handle the effects of spatial autocorrelation, the significances of the path coefficients were tested using a bootstrap resampling method. We constructed the path models and done 199 times bootstrapping using the "sem" function in *lavaan* package of R (Rosseel et al., 2021).

RESULTS

Geographic Pattern of Species Richness

The average species richness of bryophytes varied greatly among the 29 provinces in China ranged from 11 to 137, with an average of 55 species (**Supplementary Table 1**). The species richness was markedly higher in southern China than in northern China (**Figure 1A**). The average species richness of all bryophytes per province in southern China (17 provinces with latitude lower than 35°) and northern China are 68 (±28) and 37 (±14), respectively. The greatest richness was found in the southwest and southeast provinces (**Figure 1A**), such as Yunnan, Xizang, Taiwan, Fujian. By contrast, the species richness was very low in the Qinghai, Shanxi, Ningxia and Henan. The patterns of species richness of liverworts and mosses were generally similar to those of all bryophytes (**Figures 1B,C**). The average species richness of liverworts and mosses among provinces are 15 (±12) and 39 (±16). The Pearson correlation coefficients between bryophyte richness and liverwort or moss richness are 0.95 and 0.98, respectively. The Pearson correlation coefficients between liverwort richness and moss richness is 0.87.

Effects of MSG on Species Richness Pattern

The geographic pattern of MSG for all bryophytes was roughly coincided with its species richness pattern (**Figure 1D**). MSG for liverworts and mosses showed very different geographic patterns (**Figures 1E,F**). Species richness of all bryophytes significantly increased with the increase of MSG, either in univariate GLM analysis ($R^2 = 0.40$, $p < 0.0001$, **Figure 2A**) or path analysis (path coefficient = 0.43, $p < 0.05$, **Figure 3A**), indicating that provinces with high species richness of all bryophytes were dominated by species from lineages with particularly high MSG. Consistent with bryophytes, the species richness of liverworts was significantly increased with MSG (univariate GLM analysis $R^2 = 0.47$, $p < 0.0001$, **Figure 2B**; path coefficient = 0.37, $p < 0.001$, **Figure 3B**). However, the species richness of mosses was insignificantly decreased MSG in univariate GLM analysis ($R^2 = 0.11$, $p = 0.14$; **Figure 2C**) or insignificantly increased with MSG in path analysis (path coefficient = 0.13, $p > 0.05$, **Figure 3C**).

[1] www.worldclim.org

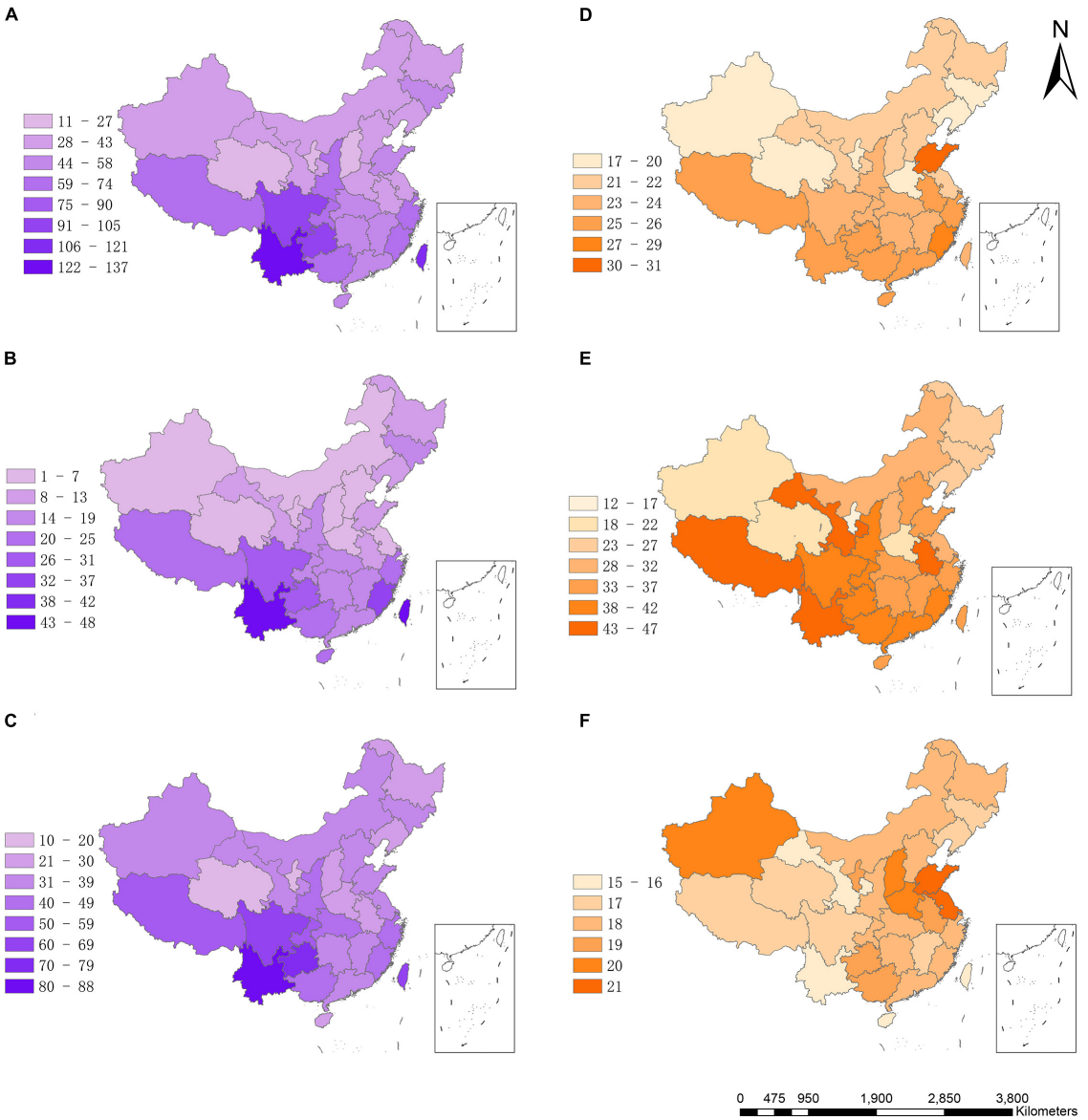


FIGURE 1 | Geographic pattern of species richness and species differentiation rate of bryophytes in China. **(A)** Species richness of all bryophytes, **(B)** species richness of liverworts, **(C)** species richness of mosses, **(D)** diversification rates of all bryophytes, **(E)** diversification rates of liverworts, **(F)** diversification rates. Species richness = number of species/ln area.

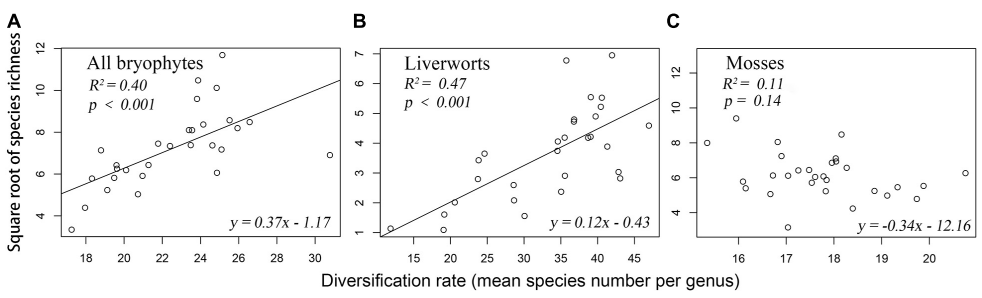


FIGURE 2 | Relationship between species richness and diversification rates of bryophyte. **(A)** All bryophytes **(B)** Liverworts **(C)** Mosses.

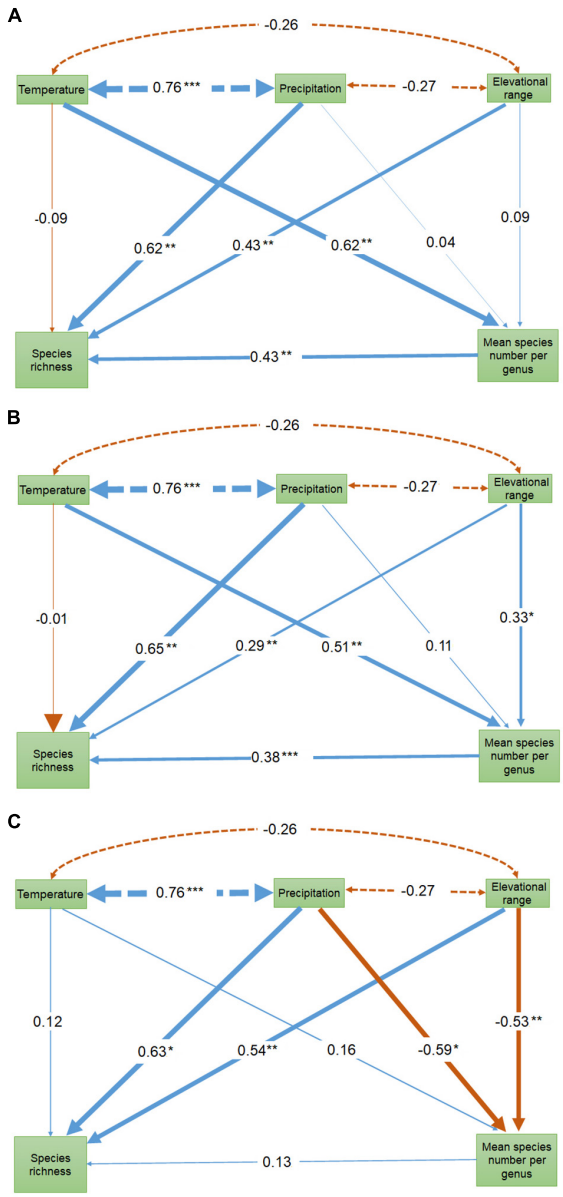


FIGURE 3 | Path diagram showing the interaction among the geographical patterns of temperature, precipitation, elevational range, mean number of species per genus and species richness. The values on the arrows are standardized path coefficients. Blue arrows indicate positive relationships and red values negative relationships. The line thickness of arrows represents the strength of the relationships. The dashed arrows show the Pearson correlation between temperature, precipitation and elevational range. Temperature and precipitation were estimated as the average temperature of the coldest season and the precipitation of the warmest season, respectively. *indicates $p < 0.1$, **$p < 0.05$, ***$p < 0.001$. **(A)** All bryophytes **(B)** Liverworts **(C)** Mosses.

Effects of Climatic Factor on Species Richness Pattern and MSG

For all bryophytes, both species richness and MSG significantly increased with the increase in energy and water availability ($p < 0.05$; **Supplementary Figure 2**), indicating that both species richness and MSG were higher in warm and wet environment. These relationships were generally consistent with the predictions from the diversification rate hypothesis. The results for liverworts and mosses were quite different. Generally consistent with results for all bryophytes, both species richness and MSG of liverworts increased with the increase in energy and water availability except (**Supplementary Figure 2**). In contrast, for mosses, the species richness significantly increased with the increase of energy and water availability, while MSG insignificantly decreased with the increase of energy and water availability (**Supplementary Figure 2**). There was a high number of moss genus containing one single species in southern China (**Supplementary Figure 3**). Genus containing single species of mosses significantly increased with the increase of the PWQ ($R^2 = 0.17$, $p = 0.04$) and elevational range ($R^2 = 0.29$, $p = 0.003$), and insignificantly increased with the increase of the average temperature of TCQ ($R^2 = 0.03$, $p = 0.40$). These results for mosses were inconsistent with the prediction of diversification rate hypothesis.

Multiple regressions and model average analyses showed that PWQ and elevational range had the highest effects on species richness of bryophytes, liverworts or mosses (**Table 1**). TCQ was the primary determinant of the MSG of all bryophytes and liverworts. The PWQ and elevational range were important predictors for MSG of mosses (**Table 1**). Both univariate and multiple regressions showed that explanatory powers of environmental variables on the species richness of mosses were lower than those of liverworts.

Effects of Elevational Range on Species Richness Pattern and MSG

Univariate GLM analyses showed insignificant effects and low explanatory powers of elevational range both on species richness pattern and MSG for all bryophytes, liverworts or mosses ($R^2 < 0.1$, $p > 0.05$), except MSG for mosses ($R^2 = 0.18$, $p = 0.02$) (**Supplementary Figure 2**). When considering the other explanatory variables in multiple regressions, elevational range showed high relative importance on species richness pattern of all bryophytes, liverworts or mosses (importance value 0.95–0.98, **Table 1**). Path analyses showed that species richness significantly increased with the increase of elevational range on (path coefficients = 0.29–0.54, $p < 0.1$, **Figure 3**). Path analyses also showed that significant positive relationship between MSG and elevational range for liverworts (path coefficients = 0.33, $p < 0.1$), and significantly negative relationship between MSG and elevational range for mosses (path coefficients = −0.53, $p < 0.05$, **Figure 3**).

DISCUSSION

The Support for the Diversification Rate Hypothesis Differed Between Liverworts and Mosses

The species richness of all bryophytes in China showed a clear latitudinal gradient (i.e., higher species richness in low latitudes than in high latitudes). Similar trend also was also reported for

TABLE 1 | Multiple regression between species richness and diversification rate of all bryophytes, liverworts and mosses at province level (response variables) and the average temperature of the coldest season (TCQ), the precipitation of the warmest season (PWQ), and the elevational range (ER) showing the estimated coefficients (ß), standard error, relative importance (RI) and R^2.

	Group		Intercept	TCQ	PWQ	ER	R^2
Species richness	All bryophytes	β	−6.82 (3.90)	8.05×10^{-4} (2.4×10^{-3})	6.12×10^{-3} (1.5×10^{-3})	1.44 (0.47)	0.62
		RI		0.23	**0.99**	**0.98**	
	Liverworts	β	−7.22 (3.02)	7.63×10^{-4} (1.9×10^{-3})	5.69×10^{-3} (1.1×10^{-3})	1.06 (0.37)	0.69
		RI		0.26	**1**	**0.98**	
	Mosses	β	−3.37 (3.40)	2.73×10^{-3} (3.5×10^{-3})	3.78×10^{-3} (1.0×10^{-3})	1.06 (0.35)	0.49
		RI		0.22	**0.96**	**0.95**	
Mean species number per genus	All bryophytes	β	21.71 (3.56)	0.02 (6.43×10^{-3})	2.64×10^{-4} (1.88×10^{-3})	0.09 (0.43)	0.40
		RI		**0.97**	0.18	0.18	
	Liverworts	β	11.78 (24.73)	0.04 (0.02)	2.51×10^{-3} (7.35×10^{-3})	2.60 (3.11)	0.35
		RI		**0.86**	0.26	**0.54**	
	Mosses	β	27.64 (3.71)	2.17×10^{-4} (1.73×10^{-3})	$^{-}2.52 \times 10^{-3}$ (1.39×10^{-3})	−1.11 (0.44)	0.40
		RI		0.22	**0.89**	**0.96**	

Species richness of bryophyte, liverworts, and mosses were all square root-transformed. Elevational range in each province was log-transformed. The models were constructed using multiple model selection and model averaging. Bold RI values indicated predictors with high relative importance.

woody plants in China and eastern Asia by previous studies (Wang et al., 2011; Su et al., 2020). In addition, we found that the geographical patterns of liverworts and mosses are quite similar. However, the evolutionary mechanisms underlying those patterns seem to be different. We found a positive relationship between the species richness pattern and MSG, a measure of recent net diversification, per province for all bryophytes and liverworts. This finding indicates that the regions with high species richness are mainly inhabited by bryophytes (especially liverworts) with high recent net diversification rate, and therefore is generally consistent with the diversification rate hypothesis. Similar positive relationship between species richness pattern and net diversification rate have also been reported for other plant groups (e.g., Arecaceae in the New World, Svenning et al., 2008; global Rhododendron, Shrestha et al., 2018; global Zygophyllaceae, Wang et al., 2018). We also found that MSG of all bryophytes tended to be high in warm and wet regions at low latitude. The high MSG in warm and wet climate may reflect the greater ecological success of bryophytes, especially liverworts in tropical climate, or increase of individual number in tropical climate which favor low extinction rate and high speciation rate (Hubbell, 2001; Svenning et al., 2008). In addition, this trend may result from the high recent speciation rates due to faster molecular evolution and increased biotic interaction in tropical climates (Mittelbach et al., 2007; Goldie et al., 2010). This is consistent with previous findings of Laenen et al. (2018), that tropical genera of liverworts have higher diversification rates and younger ages than temperate liverworts.

In contrast of the results of liverworts and all bryophytes, we found an insignificantly relationship between species richness pattern and MSG of mosses (**Figures 2C**, **3C**), suggesting that the high moss species richness in southern China is not likely the outcome of the recent high diversification. Furthermore, we found negative relationships between MSG of mosses and water availability and elevational range. First, the slightly high MSG in northern China may reflect that many lineages of mosses such as *Grimmia*, *Syntrichia*, and *Tortula* with fairly strong desiccation tolerance are well adapted to dry climate (Vitt et al., 2014). Some taxa have enhanced their drought tolerance with a suite of adaptive structures, such as leaf hair points and papillae which facilitate water retention and dew formation (Tao and Zhang, 2012; Vitt et al., 2014; Wu et al., 2014). In addition, a series of physiological mechanisms such as high content of non-reducing sugars and effective antioxidant allowing them to survive after rapid desiccation (Proctor et al., 2007). Those adaptabilities to dry climate of mosses may help them decrease the extinction rates and/or increase speciation rate in northern China. Second, those results for mosses are consistent with the prediction of time-for-speciation hypothesis. Humid mountainous areas of southern China may retain many ancient relict mosses with low diversification rates, therefore decreased the diversification rate there. We found that many mountainous provinces with a humid climate such as Yunnan, Sichuan, Xizang, Guizhou, Fujian, Taiwan have a large number of monotypic moss genera (**Supplementary Figure 3**). Future studies are required to reveal whether those monotypic genera are old as expected by time-for-speciation hypothesis.

A More Important Role of Water Availability Than Temperature

The univariate GLM, multiple regression and path analyses consistently showed that water availability related variable PWQ was the more important than energy-related variable TCQ in affecting the species richness of all bryophytes, liverworts or mosses. These findings consistent with previous studies on bryophytes conducted in Macaronesian Island (Aranda et al., 2014) and in China (Qian and Chen, 2016). Bryophytes are early land plants, and water availability is the primary limiting factor for the survival, growth and reproduction (Vitt et al., 2014; Sakakibara, 2016). First, as the poikilohydric plant, bryophyte have no vascular tissues to transport waters. Therefore, the water

content of bryophyte is directly regulated by ambient humidity, and most species rely on atmospheric precipitation to absorb water (Patiño and Vanderpoorten, 2018). Although bryophytes have evolved a certain degree of dehydration tolerance, long-term lack of water will cause them to dry and dormant and unable to perform normal physiological activities (Proctor et al., 2007). Second, the dispersal of bryophytes is primarily driven by spores, and the process of sexual reproduction and spore germination must depend on water (Proctor et al., 2007; Aranda et al., 2014).

In this study, we found a relative low importance of TCQ in regulating species richness pattern of bryophytes, which is in contrast with the previous finding that TCQ is the strongest predictors on species richness pattern of woody species in China (Wang et al., 2011). Those results probably reflected that bryophytes (especially mosses) have considerable cold tolerance compared with woody plants. Most mosses can grow at low temperature, and they still have net photosynthetic capacity even at relatively low temperatures (Patiño and Vanderpoorten, 2018). At the same time, the increase of the concentration of intracellular solute (such as soluble sugar) in cells of mosses decreased the intracellular freezing point (Nagao et al., 2006), which helps to prevent the formation of internal crystals and avoid the destruction of cell structure by freezing. These all increase the ability of mosses to tolerate low temperatures even in glacier. Liverworts showed stronger richness-water availability relationship and richness-energy relationship than mosses, which were consistent with the results of the previous studies (Aranda et al., 2014; Qian and Chen, 2016). This probably reflected that mosses generally have stronger cold and dry tolerance than liverworts.

The Importance of Habitat Heterogeneity

Multiple regressions and path analyses showed that elevational range is also an important contributor to the geographic pattern of bryophyte species richness, suggesting that provinces with high habitat heterogeneity hold more bryophytes. Firstly, mountains have strengthened the geographical isolation within species, therefore promoted the speciation rate (Tang et al., 2006; Xing and Ree, 2017). That is, mountains are "cradles" of biodiversity. In this study, we found that elevational range have a positive influence on the geographic pattern of MSG of liverworts, suggesting that habitat heterogeneity in mountains facilitated the recent speciation of liverworts. Secondly, mountainous areas serve as biological refuges, reducing the extinction of species. Bryophytes are very sensitive to environmental changes (Gignac, 2001). Mountain areas with diverse habitats may preserve many ancient and relict species of mosses, resulting in a lower MSG. This also explains why MSG of mosses and elevational range show a significant negative correlation (**Supplementary Figure 2**). Finally, path analysis also showed a significant direct effect of elevational ranges on species richness. This probably reflected that regions with high elevational range provided the diversified local climates, vegetation types and soil conditions (Stein et al., 2014), therefore creating more ecological niches for coexistence of more bryophytes. In addition, the effect may also involve the other evolutionary processes induced by habitat heterogeneity which are not considered in this analysis.

CONCLUSION

In this study, we found that the geographical pattern of liverwort richness in China is significantly associated with the high recent diversification rate (estimated by MSG) in warm, wet and humid mountain areas. In contrast, for mosses, the geographical pattern of species richness in China is not well-explained by the MSG patterns, and the MSG are low in humid mountains where monotypic and oligotypic genera are abundant. Our study suggests the humid tropical and subtropical mountains in China as not only diversity hotspots for bryophytes, but also cradles for high recent diversification of liverworts, and refuges for mosses to hold many monotypic and oligotypic genera. Our study provided important information on ecological and evolutionary causes of geographical pattern of bryophyte richness. However, the conclusions drawn in our study call for further tests especially using evolutionary information (e.g., evolutionary rates and evolutionary age) obtained from dated phylogeny of bryophytes in the future.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

QW conceived and designed the study. WF, QW, XSo, and XC compiled the data. QW and XSo analyzed the data and led the writing. All authors contributed to the writing and agreed to be accountable for all aspects of the work.

FUNDING

This work was supported by the National Key Research and Development Program of China (2019YFC0507501), Research on Theories and Methods of Evaluating the Outflow Risk of Important Wild Plant Genetic Resources (19-02), the National Natural Science Foundation of China (81603236 and 41771054), and the Fundamental Research Funds for the Central Public Welfare Research Institutes (ZZ13-YQ-087).

ACKNOWLEDGMENTS

We thank all bryologists in China who shared the data for this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.680318/full#supplementary-material

REFERENCES

Allen, A. P., Brown, J. H., and Gillooly, J. F. (2002). Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science* 297, 1545–1548. doi: 10.1126/science.1072380

Allen, A. P., Gillooly, J. F., Savage, V. M., and Brown, J. H. (2006). Kinetic effects of temperature on rates of genetic divergence and speciation. *Proc. Natl. Acad. Sci. U.S.A.* 103, 9130–9135. doi: 10.1073/pnas.0603587103

Aranda, S. C., Gabriel, R., Borges, P. A., Santos, A. M., de Azevedo, E. B., Patiño, J., et al. (2014). Geographical, temporal and environmental determinants of bryophyte species richness in the Macaronesian Islands. *PLoS One* 9:e101786. doi: 10.1371/journal.pone.0101786

Bartoń, K. (2020). *MuMIn: Multi-Model Inference*. Available online at: https://CRAN.R-project.org/package=MuMIn. (accessed January 15, 2021)

Cai, T., Shao, S., Kennedy, J. D., Alström, P., Moyle, R. G., Qu, Y., et al. (2020). The role of evolutionary time, diversification rates and dispersal in determining the global diversity of a large radiation of passerine birds. *J. Biogeogr.* 47, 1612–1625. doi: 10.1111/jbi.13823

Cannone, N., Corinti, T., Malfasi, F., Gerola, P., Vianelli, A., Vanetti, I., et al. (2017). Moss survival through in situ cryptobiosis after six centuries of glacier burial. *Sci. Rep.* 7, 1–7. doi: 10.1038/s41598-017-04848-6

Chen, S. B., Slik, F., Gao, J., Mao, L. F., Bi, M. J., Shen, M. W., et al. (2015). Latitudinal diversity gradients in bryophytes and woody plants: Roles of temperature and water availability. *J. Syst. Evol.* 53, 535–545. doi: 10.1111/jse.12158

Clarke, A., and Gaston, K. J. (2006). Climate, energy and diversity. *Proc. R. Soc. B* 273, 2257–2266. doi: 10.1098/rspb.2006.3545

Clifford, P., Richardson, S., and Hemon, D. (1989). Assessing the significance of the correlation between two spatial processes. *Biometrics* 45, 123–134. doi: 10.2307/2532039

Condamine, F. L., Sperling, F. A. H., Wahlberg, N., Rasplus, J. Y., and Kergoat, G. J. (2012). What causes latitudinal gradients in species diversity? Evolutionary processes and ecological constraints on swallowtail biodiversity. *Ecol. Lett.* 15, 267–277. doi: 10.1111/j.1461-0248.2011.01737.x

Currie, D. J., Mittelbach, G. G., Cornell, H. V., Field, R., Guegan, J. F., Hawkins, B. A., et al. (2004). Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecol. Lett.* 7, 1121–1134. doi: 10.1111/j.1461-0248.2004.00671.x

Dagallier, L. P. M., Janssens, S. B., Dauby, G., Blach—Overgaard, A., Mackinder, B. A., Droissart, V., et al. (2020). Cradles and museums of generic plant diversity across tropical Africa. *New Phytol.* 225, 2196–2213. doi: 10.1111/nph.16293

Dutilleul, P., Clifford, P., Hemon, D., and Richardson, S. (1993). Modifying the t-test for assessing the correlation between two spatial processes. *Biometreics* 49, 305–314. doi: 10.2307/2532625

Dynesius, M., and Jansson, R. (2000). Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proc. Natl. Acad. Sci. U.S.A.* 97, 9115–9120. doi: 10.1073/pnas.97.16.9115

Fischer, A. G. (1960). Latitudinal Variations In Organic Diversity. *Evolution* 14, 64–81. doi: 10.2307/2405923

Fjeldsa, J., Bowie, R. C. K., and Rahbek, C. (2012). The role of mountain ranges in the diversification of birds. *Annu. Rev. Ecol. Evol. Syst.* 43, 249–265. doi: 10.1146/annurev-ecolsys-102710-145113

Gignac, L. D. (2001). Bryophytes as indicators of climate change. *Bryologist* 104, 410–420. doi: 10.1639/0007-27452001104[0410:BAIOCC]2.0.CO;2

Goldie, X., Gillman, L., Crisp, M., and Wright, S. (2010). Evolutionary speed limited by water in arid Australia. *Proc. R. Soc. B* 277, 2645–2653. doi: 10.1098/rspb.2010.0439

Hawkins, B. A., Diniz-Filho, J. A. F., and Soeller, S. A. (2005). Water links the historical and contemporary components of the Australian bird diversity gradient. *J. Biogeogr.* 32, 1035–1042. doi: 10.1111/j.1365-2699.2004.01238.x

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978. doi: 10.1002/joc.1276

Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton: Princeton University Press.

Hughes, C. E., and Atchison, G. W. (2015). The ubiquity of alpine plant radiations: from the Andes to the Hengduan Mountains. *New Phytol.* 207, 275–282. doi: 10.1111/nph.13230

Jablonski, D., Roy, K., and Valentine, J. W. (2006). Out of the tropics: evolutionary dynamics of the latitudinal diversity gradient. *Science* 314, 102–106. doi: 10.1126/science.1130880

Kozak, K., and Wiens, J. (2016). Testing the relationships between diversification, species richness, and trait evolution. *Syst. Bot.* 65, 975–988.

Laenen, B., Patino, J., Hagborg, A., Desamore, A., Wang, J., Shaw, A. J., et al. (2018). Evolutionary origin of the latitudinal diversity gradient in liverworts. *Mol. Phylogenet. Evol.* 127, 606–612. doi: 10.1016/j.ympev.2018.06.007

Laenen, B., Shaw, B., Schneider, H., Goffinet, B., Paradis, E., Desamore, A., et al. (2014). Extant diversity of bryophytes emerged from successive post-Mesozoic diversification bursts. *Nat. Commun.* 5:5134. doi: 10.1038/ncomms6134

Liu, Y., Johnson, M. G., Cox, C. J., Medina, R., Devos, N., Vanderpoorten, A., et al. (2019). Resolution of the ordinal phylogeny of mosses using targeted exons from organellar and nuclear genomes. *Nat. Commun.* 10, 1–11. doi: 10.1038/s41467-019-09454-w

Mittelbach, G. G., Schemske, D. W., Cornell, H. V., Allen, A. P., Brown, J. M., Bush, M. B., et al. (2007). Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecol. Lett.* 10, 315–331. doi: 10.1111/j.1461-0248.2007.01020.x

Nagao, M., Oku, K., Minami, A., Mizuno, K., Sakurai, M., Arakawa, K., et al. (2006). Accumulation of theanderose in association with development of freezing tolerance in the moss Physcomitrella patens. *Phytochemistry* 67, 702–709. doi: 10.1016/j.phytochem.2006.01.031

Osorio, F., Vallejos, R., Cuevas, F., and Mancilla, D. (2020). *SpatialPack: Tools for assessment the association between two spatial processes*. Available online at: https://CRAN.R-project.org/package=SpatialPack. (accessed January 15, 2021)

Patiño, J., and Vanderpoorten, A. (2018). Bryophyte biogeography. *Crit. Rev. Plant Sci.* 37, 175–209. doi: 10.1080/07352689.2018.1482444

Perera-Castro, A. V., Flexas, J., González-Rodríguez, ÁM., and Fernández-Marín, B. (2020). Photosynthesis on the edge: photoinhibition, desiccation and freezing tolerance of Antarctic bryophytes. *Photosynth. Res.* 2020, 1–19. doi: 10.1007/s11120-020-00785-0

Proctor, M., Oliver, M., Wood, A., Alpert, P., Stark, L., Cleavitt, N. L., et al. (2007). Desiccation-tolerance in bryophytes: a review. *Bryologist* 110, 595–621. doi: 10.1639/0007-27452007110[595:DIBAR]2.0.CO;2

Qian, H., and Chen, S. B. (2016). Reinvestigation on species richness and environmental correlates of bryophytes at a regional scale in China. *J. Plant Ecol.* 9, 734–741. doi: 10.1093/jpe/rtw001

Rana, S. K., Price, T. D., and Qian, H. (2019). Plant species richness across the Himalaya driven by evolutionary history and current climate. *Ecosphere* 10:e02945. doi: 10.1002/ecs2.2945

Rohde, K. (1992). Latitudinal gradients in species diversity: the search for the primary cause. *Oikos* 1992, 514–527. doi: 10.2307/3545569

Rosenzweig, M. L. (1995). *Species diversity in space and time*. Cambridge, MA: Cambridge University Press.

Rosseel, Y., Jorgensen, T., Rockwood, N., Oberski, D., Byrnes, J., Vanbrabant, L., et al. (2021). *lavaan: Latent variable analysis*. Available online at: https://CRAN.R-project.org/package=lavaan. (accessed January 15, 2021)

Sakakibara, K. (2016). Technological innovations give rise to a new era of plant evolutionary developmental biology. *Adv. Bot. Res.* 78, 3–35. doi: 10.1016/bs.abr.2016.01.001

Shrestha, N., Wang, Z. H., Su, X. Y., Xu, X. T., Lyu, L. S., Liu, Y. P., et al. (2018). Global patterns of Rhododendron diversity: The role of evolutionary time and diversification rates. *Glob. Ecol. Biogeogr.* 27, 913–924. doi: 10.1111/geb.12750

St Martin, P., and Mallik, A. U. (2017). The status of non-vascular plants in trait-based ecosystem function studies. *Perspect. Plant Ecol. Evol. Syst.* 27, 1–8. doi: 10.1016/j.ppees.2017.04.002

Stein, A., Gerstner, K., and Kreft, H. (2014). Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecol. Lett.* 17, 866–880. doi: 10.1111/ele.12277

Steinbauer, M. J., Field, R., Grytnes, J. A., Trigas, P., Ah—Peng, C., Attorre, F., et al. (2016). Topography—driven isolation, speciation and a global increase of endemism with elevation. *Glob. Ecol. Biogeogr.* 25, 1097–1107. doi: 10.1111/geb.12469

Stephens, P. R., and Wiens, J. J. (2003). Explaining species richness from continents to communities: the time-for-speciation effect in emydid turtles. *Am. Nat.* 161, 112–128. doi: 10.1086/345091

Su, X., Shrestha, N., Xu, X., Sandanov, D., Wang, Q., Wang, S., et al. (2020). Phylogenetic conservatism and biogeographic affinity influence woody plant species richness–climate relationships in eastern Eurasia. *Ecography* 43, 1027–1040. doi: 10.1111/ecog.04839

Svenning, J. C., Borchsenius, F., Bjorholm, S., and Balslev, H. (2008). High tropical net diversification drives the New World latitudinal gradient in palm (Arecaceae) species richness. *J. Biogeogr.* 35, 394–406. doi: 10.1111/j.1365-2699.2007.01841.x

Tang, Z. Y., Wang, Z. H., Zheng, C. Y., and Fang, J. Y. (2006). Biodiversity in China's mountains. *Front. Ecol. Environ.* 4:347–352. doi: 10.1890/1540-92952006004[0347:BICM]2.0.CO;2

Tao, Y., and Zhang, Y. M. (2012). Effects of leaf hair points of a desert moss on water retention and dew formation: implications for desiccation tolerance. *J. Plant Res.* 125, 351–360. doi: 10.1007/s10265-011-0449-3

Turner, J. R. G. (2004). Explaining the global biodiversity gradient: energy, area, history and natural selection. *Basic Appl. Ecol.* 5, 435–448. doi: 10.1016/j.baae.2004.08.004

Vitt, D., Crandall-Stotler, B., and Wood, A. (2014). "Bryophytes: survival in a dry world through tolerance and avoidance," in *Plant Ecology and Evolution in Harsh Environments*, eds N. Rajakaruna, R. Boyd, and T. Harris (New York, NY: Nova Publishers), 267–295.

Wang, Z., Fang, J., Tang, Z., and Lin, X. (2011). Patterns, determinants and models of woody plant diversity in China. *Proc. R. Soc. B* 278, 2122–2132. doi: 10.1098/rspb.2010.1897

Wang, Q., Wu, S., Su, X., Zhang, L., Xu, X., Lyu, L., et al. (2018). Niche conservatism and elevated diversification shape species diversity in drylands: evidence from Zygophyllaceae. *Proc. R. Soc. B* 285:1742. doi: 10.1098/rspb.2018.1742

Wu, N., Zhang, Y. M., Downing, A., Aanderud, Z. T., Tao, Y., and Williams, S. (2014). Rapid adjustment of leaf angle explains how the desert moss, *Syntrichia caninervis*, copes with multiple resource limitations during rehydration. *Funct. Plant Biol.* 1, 168–177. doi: 10.1071/FP13054

Xing, Y., and Ree, R. H. (2017). Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. *Proc. Natl. Acad. Sci. U.S.A.* 114, E3444–E3451. doi: 10.1073/pnas.1616063114

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 25 June 2021
doi: 10.3389/fevo.2021.706237

Generality and Shifts in Leaf Trait Relationships Between Alpine Aquatic and Terrestrial Herbaceous Plants on the Tibetan Plateau

Lei Yang, Haocun Zhao, Zhenjun Zuo, Xiangyan Li, Dan Yu and Zhong Wang**

Department of Ecology, College of Life Sciences, Wuhan University, Wuhan, China

Plant traits mirror both evolutionary and environmental filtering process with universal trait-trait relationships across plant groups. However, plants also develop unique traits precisely to different habitats, inducing deviations of the trait coupling relations. In this study, we aimed to compare the differences in leaf traits and examine the generality and shifts of trait-trait relationships between alpine aquatic and terrestrial herbaceous plants on the Tibetan Plateau, to explore the precise adaptive strategies of aquatic and terrestrial plants for its habitats. We measured mass-based and area-based leaf N and P concentrations, N:P ratios and specific leaf area (SLA) of aquatic and terrestrial herbaceous plants. Standardized major axis analysis were applied to build the correlations for every trait pairs of each plant group, and then to compare the differences in the trait-trait correlations among different plant groups. Leaf N_{mass} and P_{mass} of two groups of aquatic plants (emergent and submerged plants) were higher, but N:P ratios were lower than those of two groups of terrestrial plants (sedges and grasses). Submerged plants had extremely high SLA, while grasses had the lowest SLA. N_{mass} positively correlated with P_{mass} in three out of four plant groups. The two terrestrial plant groups had positive N_{mass}-SLA relationships but these two traits coupled weakly in aquatic plants. P_{mass} showed positive relationships to SLA in three out of four plant groups. Significant shifts of trait-trait relationships between aquatic and terrestrial plants were observed. In general, aquatic plants, especially submerged plants, are characterized by higher SLA, greater leaf $nutrient_{mass}$ than terrestrial plants, tend to pursue fast-return investment strategies, and represent the acquisitive end of leaf economics spectrum. The deviations of trait-trait relationships between different plant groups reveal the precise adaptions of submerged plants to the unique aquatic habitats.

Keywords: alpine wetlands, functional traits, leaf economics spectrum, leaf N and P concentrations, specific leaf area, stoichiometry

OPEN ACCESS

Edited by:
Xiaoting Xu,
Sichuan University, China

Reviewed by:
Lin Zhang,
Institute of Tibetan Plateau Research, Chinese Academy of Sciences (CAS), China
Yan Geng,
Beijing Forestry University, China

****Correspondence:***
Zhong Wang
wangzhong@whu.edu.cn
Dan Yu
lakeyd@163.com

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *07 May 2021*
Accepted: *02 June 2021*
Published: *25 June 2021*

Citation:
Yang L, Zhao H, Zuo Z, Li X, Yu D and Wang Z (2021) Generality and Shifts in Leaf Trait Relationships Between Alpine Aquatic and Terrestrial Herbaceous Plants on the Tibetan Plateau. Front. Ecol. Evol. 9:706237. doi: 10.3389/fevo.2021.706237

INTRODUCTION

Functional traits are known as quantitative indicators of plant fitness that developed from both evolution and ecological filtering (Chapin et al., 1993), of which leaf nitrogen (N) and phosphorus (P) concentrations, N:P ratios and specific leaf area (SLA) attract the most attention (Díaz et al., 2016; Moor et al., 2017). Leaf N and P concentrations (mass-based or area-based) represent the

total amounts of N and P per unit dry mass (mass-based, mg g^{-1}) or per unit leaf area (area-based, g m^{-2}), while SLA is the one-sided leaf area per unit dry mass (cm^2 g^{-1}) (Pérez-Harguindeguy et al., 2013), reflecting the expected return (light-intercepting area) on per unit of resource (carbon and nutrients) investments (Poorter et al., 2009). Both SLA and leaf nutrient concentrations are species-specific traits, but meanwhile, vary along environmental gradients (Güsewell and Koerselman, 2002; Poorter et al., 2009). Species (or populations) native to cold or arid habitats often have greater leaf N_{mass} and P_{mass} (Reich and Oleksyn, 2004; He et al., 2008; Wang et al., 2015), and lower SLA (Poorter et al., 2009). Plants evolve a suite of trait combinations acclimatizing itself to native habitats.

Despite the significant differences in leaf traits among species and habitats, the bivariate traits relationships are often similar and species-independent, indicating the convergent evolution and generality in plant adaptation (Reich et al., 1997). Numerous reviews and study cases described the generally accepted correlations of leaf trait pairs (Chapin et al., 1993; Reich et al., 1999, 2010; Wright et al., 2004; He et al., 2006; Reich, 2014; Díaz et al., 2016; Pan et al., 2020a; Thomas et al., 2020). High SLA always couples with greater leaf N_{mass} and P_{mass}, while the two photosynthesis-related leaf nutrients (N and P) show a positive relationship with each other. Such acquisitive traits have faster return on investment of resources and make plants growing with high rate. On the contrary, lower SLA and lower leaf nutrients are conservative traits with longer leaf life span, result in lower photosynthetic rate and growth rate. Variations in leaf traits from conservative to acquisitive ones form a continuous gradient of leaf economics spectrum (LES). Each plant species has its position on the universal trade-off surface (Reich et al., 1997).

Most studies of LES focused on terrestrial ecosystems. Although aquatic habitats are special, plants surviving in water show common physiological and structural adaptive strategies that reflect fundamental trade-offs in economics, and can be compared with terrestrial species directly in perspective of LES (Pierce et al., 2012; Onoda et al., 2017). In aquatic habitats, plants, especially submerged species, have developed strategies to maximize the ability for carbon gain and light interception, maintaining high return on investment (Moor et al., 2017; Maberly and Gontero, 2018). Aquatic plants suffer from very different environmental stressors from those in terrestrial habitats, such as low light intensity, slow diffusion of gases (CO_2 in water and O_2 in water-saturated substrate), drastic water-level fluctuation caused by flooding, competition with phytoplankton, concentrated N, P and other substances in substrate and surrounding water that leached from soil and anthropic pollutants. In previous studies, aquatic plants showed greater leaf N_{mass} and P_{mass} (Xia et al., 2014; Wang et al., 2015), extremely high SLA (Poorter et al., 2009; Liu et al., 2021), and extended the acquisitive end of LES (Pierce et al., 2012; Pan et al., 2020a). However, variations in different leaf traits along a same environmental gradient or among plant species should be allometric. For example, SLA increased several hundred-fold from xerophyte to submerged plants (Poorter et al., 2009), while leaf N varied 32-fold and P by 200-fold in mean values among different plant functional groups globally (Reich et al., 2010). Therefore, although all plants followed the fundamental rules of bivariate traits relationships, differences should also exist among plant functional groups, especially between terrestrial and aquatic plants. Such deviations can mirror the precise adaptive strategies of plants to the unique environmental stressors of local habitats. Pan et al. (2020a) compared the general relationships of leaf traits among two plant groups in wetland and non-wetland species on a global scale, and found significant shifts along the common slopes between different plant groups. However, limited by data availability, aquatic plants in alpine wetlands on the Tibetan Plateau were absent in the global analysis.

Alpine *Kobresia* meadow and *Stipa* steppe constitute the major vegetation types of alpine grassland regions along large environmental gradients on the Tibetan Plateau (Tibetan Plateau Scientific Expedition and Research Team (TPSERT).,, 1992; Zhao, 2009). Alpine wetlands that embedded in grassland regions provide an ideal platform for comparative studies on the different adaptive strategies between aquatic and terrestrial plants in the same climatic but different habitat conditions (Wang et al., 2015). We hypothesized that aquatic plants had similar trends but different scales of trait-trait correlations from terrestrial herbaceous plants. Variations in SLA were involved, as the most reactive trait, in plant adaptation to the different conditions between aquatic and terrestrial habitats, resulting in different scaling trait-trait correlations. In this study, we aimed to (1) compare the differences in leaf traits, and (2) examine the generality of trait-trait relationships and the shifts of common slopes among two aquatic plant groups (emergent plants and submerged plants) and two terrestrial plant groups (sedges, mainly *Kobresia* species and grasses, mainly *Stipa* species), to explore the adaptive strategies of aquatic and terrestrial plants for its habitats.

MATERIALS AND METHODS

Study Area

This study was carried out on the Tibetan Plateau, which is the highest plateau with an average elevation of 4000 m a.s.l. and an area of 2.5 million km^2. On Tibetan Plateau, low temperature, low precipitation, strong wind and strong solar radiation together characterize the alpine climate. Detailed topographical feature and climatic characteristics were described in Wang et al. (2015). Briefly, obstruction of the Himalayas along the southern edge of the plateau to oceanic warm and humid air current, together with the northwestward lifted elevation, form the decreasing trends of both annual precipitation (800 mm to 20 mm) and annual mean air temperature (11°C to −5°C), resulting in a series of vegetation type of montane forest, shrub, alpine meadow, alpine grassland, and alpine desert from southeast to northwest (Zheng and Zhao, 2017).

The sample sites in this study covered 23° (79.7 E–102.7 E) in longitude, 10.7° (27.5 N–38.2 N) in latitude and 2991 m (2194–5185 m) in altitude (**Figure 1**). We mainly focused on the regions dominated by herbaceous plants. Therefore, we modified the original vegetation regionalization scheme of the whole plateau to four vegetation regions: forest region, meadow

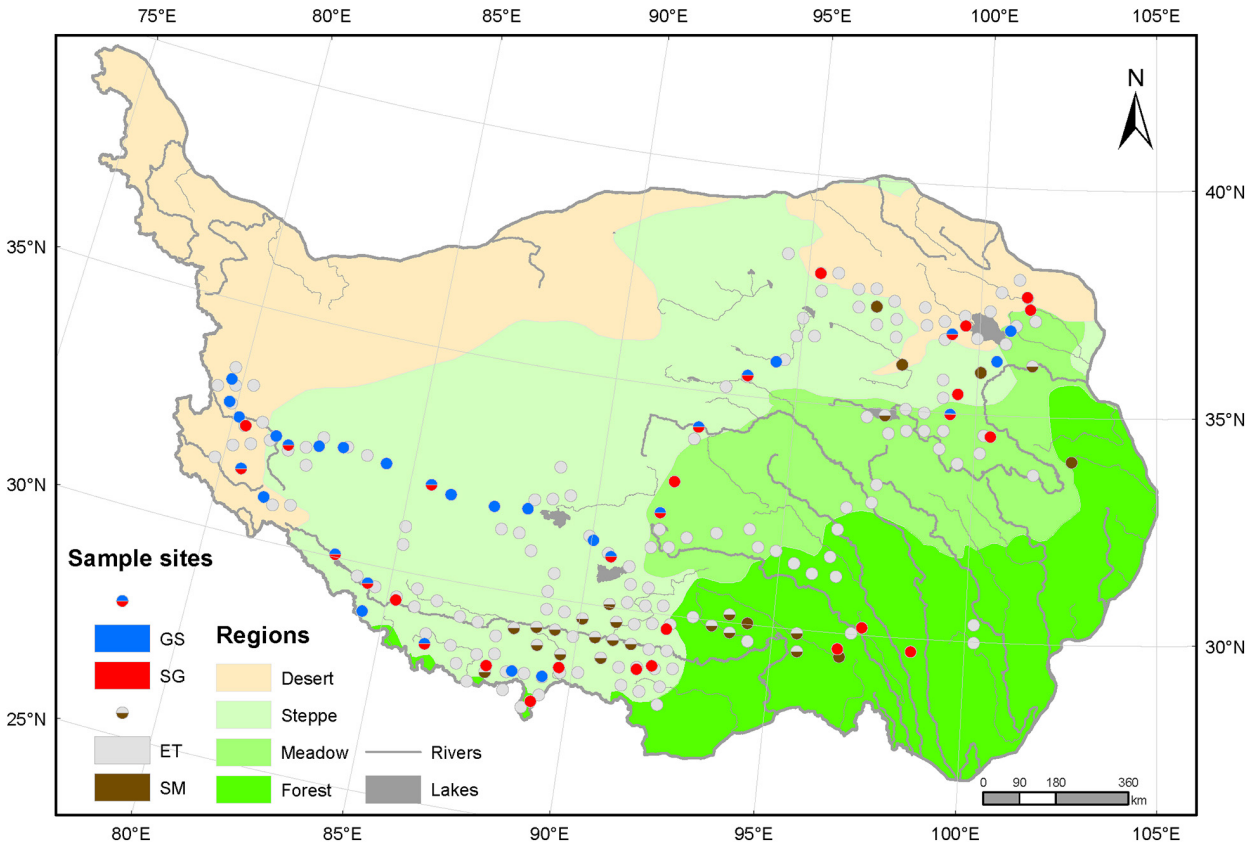


FIGURE 1 | Sample sites on the Tibetan Plateau. Based on the vegetation regionalization scheme, the Tibetan Plateau was modified to four vegetation regions: forest, meadow, steppe grassland, and desert. The raw data for the origin scheme were download from the Resource and Environment Science and Data Center (http://www.resdc.cn). All 218 sample sites were divided into two groups: aquatic plots (170) and terrestrial plots (48). Aquatic plots were tagged as three types: emergent plants (ET) only, submerged plants (SM) only and ET and SM mixed plots, depending on whether a certain group of plants are present or not. The different portions in each terrestrial plot indicated the important values of sedges (SG) versus grasses (GS) and reflect the relative importance of the plant type in this plot. The SG and GS plots in forest region located in alpine meadow belt at high altitude above treeline, and the aquatic plots there were non-forest wetland habitats.

region, steppe region, and desert region, which roughly reflected the water availability and the species composition. The raw data for vegetation regionalization were applied from Resource and Environment Science and Data Center[1] in October 2020. The meadow region is comprised of alpine meadow and alpine shrub mixed zones and typical alpine meadow zones, where the dominant herbaceous species are sedges of genera *Kobresia* or *Carex*. The steppe region includes alpine grassland zones and temperate steppe zones, and is dominated by grasses of genera *Stipa* or *Poa*. The desert region consists of alpine desert zones and temperate desert zones, which is covered by sparse vegetation of shrubs and grasses.

Field Sampling

In total, we investigated 218 1 m × 1 m plots in July and August of 2018. According to soil water level and species composition, the 218 plots were classified into two groups: aquatic plots (170) and terrestrial plots (48). All aquatic plots were in water-saturated habitats with overlying water while terrestrial ones were not.

[1]http://www.resdc.cn

For terrestrial plots, each 1 m × 1 m plot was divided into 100 10 cm × 10 cm grids. The plant percentage coverage of each species in each grid was estimated by eye and plant height was measured by a ruler. Alpine grassland communities are floristically rich. However, such communities are dominated by a few species while the other species play relatively weaker roles in community structure and function than dominant species. In this study, two indices were introduced to identify the dominant species. The first index was the importance value (IV) of each species and its proportions in the total IV. The importance values were calculated as follows (Fang et al., 2009).

$$IV = (relativecoverage + relativefrequency + relativeheight)/3$$

The second one was plant volume of each species and its proportion in the total volume. As a surrogate of plant aboveground biomass, the plant volumes were calculated by multiplying plant coverage by height (Wang et al., 2013). According to the important values and plant volumes, two groups of plants were identified as dominant species: sedges of genera

Kobresia, *Carex*, or *Scirpus* and grasses of genera *Stipa* or *Poa* (**Supplementary Appendix Table 1**). Aboveground materials of the dominant species (three of grasses and five of sedges) were collected for further analysis.

The aquatic habitats we investigated included the shallow areas (generally no deeper than 1.5 m) of lakes, rivers, marshes as well as streams, ponds, and channels. Temporary water bodies were excluded during field sampling, ensuring that aquatic plants could complete their life cycles. For aquatic plots, it cannot be divided into 10 cm × 10 cm grids. We estimated the percentage coverage and measured the height of each species at the whole 1 m × 1 m plot level, and collected all the aboveground materials. There were two plant forms in aquatic plots: submerged plants and emergent plants. Submerged plants are defined as the rooted aquatic macrophytes with root systems in sediments. The whole body of submerged plants, include leaves, are submersed in water while only lift the reproductive organs out of the water. Emergent plants are those growing in waterlogged soil with roots submerging in water and leaves in air. Finally, aboveground materials were collected from four common submerged species and 15 common emergent species.

Soil samples were collected in each plot at depth of 0–20 cm by a soil auger with the diameter of 38 mm. Because all aquatic plots were investigated in shallow areas, we waded into the water to dig the sediments using the same soil auger. Three soil samples were cored randomly in the plot, mixed and air dried for soil nutrients analysis.

Laboratory Analysis

The aboveground materials were separated by species. Firstly, 30–40 mature and intact fresh leaves of each species were picked out and expanded on a scanner (Canon Inc., Japan), to scan the flat area (cm^2) of each species. The leaves were then oven dried at 70°C for 72 h to constant weight. The specific leaf area (SLA) was determined as leaf area per unit dry weight ($cm^2\ g^{-1}$). Then, all the other biomass materials were also oven dried at 70°C for 72 h, ground and sieved through a 0.15 mm mesh sieve to measure the mass-based leaf total nitrogen concentration (N_{mass}, $mg\ g^{-1}$) and phosphorus concentration (P_{mass}, $mg\ g^{-1}$), respectively. N_{mass} were determined using Vario MACRO cube elemental analyzer (Elementar Analysensysteme GmbH, Germany), while P_{mass} were measured via the molybdate/stannous chloride method (Kuo, 1997) following H_2SO_4-H_2O_2-HF digestion (Bowman, 1988). The area-based leaf nitrogen contents (N_{area}, $g\ m^{-2}$) and leaf phosphorus contents (P_{area}, $g\ m^{-2}$) were calculated via mass-based element concentrations divided by SLA.

The air-dried soil samples for nutrients analysis were ground, sieved by 0.15 mm mesh sieve to measure soil total nitrogen content (STN) and soil total phosphorus content (STP). Methods of STN and STP analysis were the same for plant samples, respectively.

Climate Variables

Growing season mean air temperature (GST, °C), growing season precipitation (GSP, mm), and growing season mean solar radiation (GSR, $kJ\ m^{-2}\ day^{-1}$) were introduced to explain the effects of climate variables on plant traits and their bivariate traits correlations. Growing season on the Tibetan Plateau was defined as the 5 months from May to September (Wang et al., 2013). The climate variables were extracted from the climatic raster database that was download from the world climate website[2].

Data Analysis

Plant Grouping

Terrestrial plots dominated by sedges mainly locate in east part of the Tibetan Plateau, while the plots dominated by grasses in west part, reflecting the decreased trends of water availability from east to west (Zheng and Zhao, 2017). Aquatic plots are habitat-dependent and embed in all vegetation regions. According to plant life forms, we further divided all plants into four groups: emergent plants and submerged plants from aquatic plots, sedges, and grasses from terrestrial plots. Then, we compared the mean values of the three climatic variables (GST, GSP, and GSR) and three soil properties (N, P contents and N:P ratios) among the plots dominated by different plants. The results showed that GST of aquatic plants were higher than that of terrestrial plots, while the lowest GSP but highest GSR were observed in plots of grasses (**Figures 2A–C**). Aquatic substrates concentrated more N and P than terrestrial soil but the N:P ratios did not show significant differences between the four types of plots (**Figures 2D–F**). Thus, we considered that these four plant groups face different combinations of environmental factors, and may evolve varied adaptive strategies. Because aquatic substrates are always water-saturated, GSP was not suitable as a climatic factor for both emergent plants and submerged plants. Meanwhile, considering that solar radiation is influenced by a variety of factors (water depth, turbidity, etc.) and loses rapidly in water, GSR was excluded when analyzing the effects of climatic factors on traits of submerged plants.

Comparison of Mean Values

In terms of plant traits, leaf N_{mass}, P_{mass}, and SLA were measured by plant sample analysis directly, while N:P ratios, N_{area}, and P_{area} were calculated by mass-based traits. We defined the first three traits as measured plant traits, and the last three ones as transformed traits. We firstly tested if the plant traits data within each plant group followed normal distribution by Shapiro–Wilk normality test. The results showed that all traits were non-normal distribution ($p < 0.05$). Then, we compared the mean values of leaf N_{mass}, P_{mass}, SLA, N:P ratios, N_{area}, and P_{area} among the four plant groups by Kruskal–Wallis test and Dunn's *Post Hoc* Multiple Comparisons with R (version 3.6.3) packages *stats* and *PMCMR* (Pohlert, 2014).

Standardized Major Axis (SMA) Analysis

To compare the differences in the trait-trait correlations among the four plant groups, and the differences in the effects of environmental factors on plant traits, we applied SMA analysis using software R (version 3.6.3) with its package *smatr* (Warton et al., 2012). The analyses were performed by following the steps below.

[2]https://www.worldclim.org

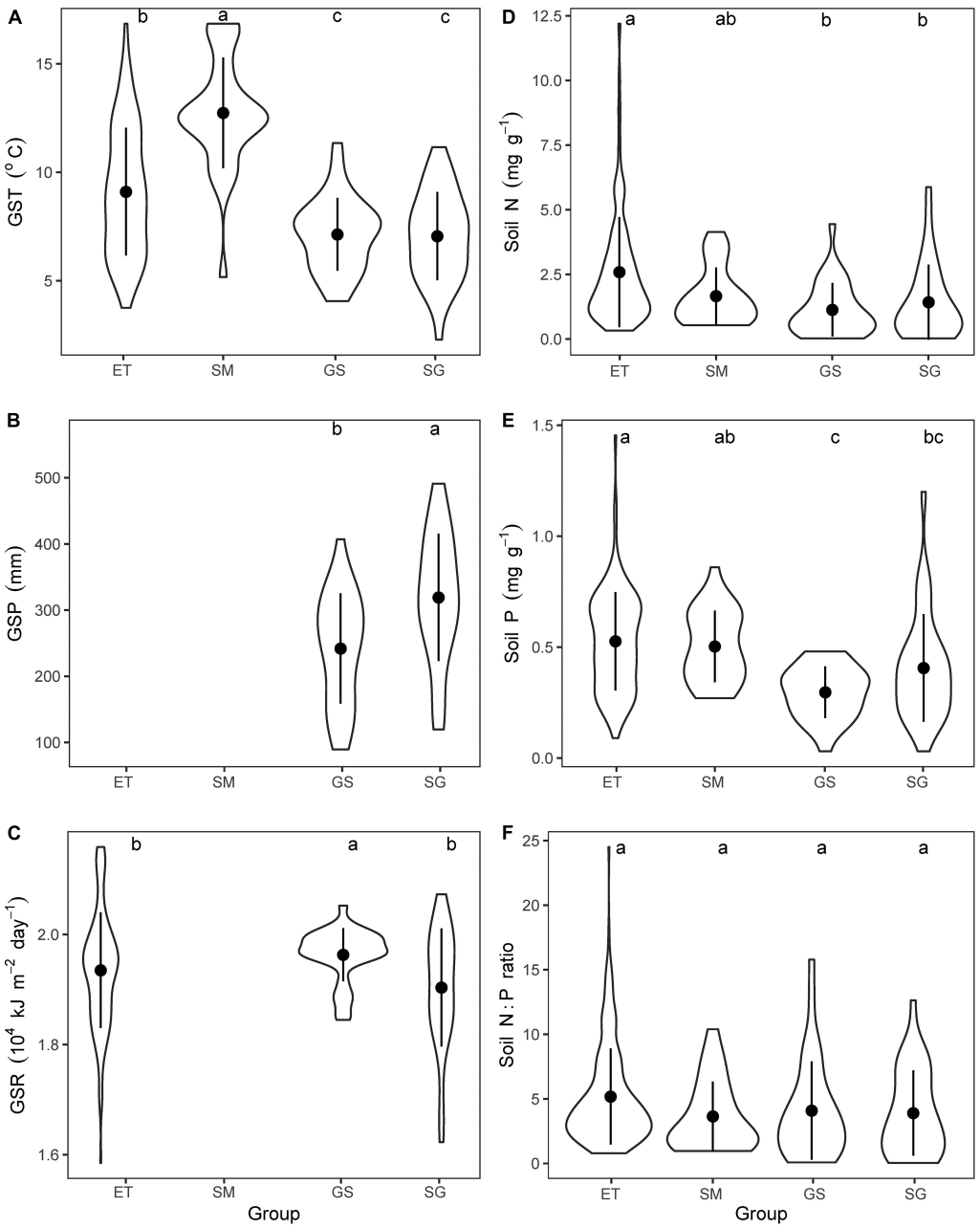


FIGURE 2 | Comparing the mean values of climatic variables and soil properties of nutrients. **(A)** Growing season mean air temperature (GST), **(B)** growing season precipitation, **(C)** growing season mean solar radiation, **(D)** soil total N contents, **(E)** soil total P contents, **(F)** soil N:P ratios for plots dominated by the four plant groups. ET, emergent plants, GS, grasses, SG, sedges, and SM, submerged plants. GSP for both emergent plants and submerged plants, and GSR for submerged plants were excluded. The lowercase letters on the top of each panel showed the results of Kruskal–Wallis test and Dunn's *Post Hoc* Multiple Comparisons ($p < 0.05$).

(i) All the plant traits data were log-transformed before analysis. Then, we built bivariate correlations within the three measured traits for each plant group, respectively, and tested if the correlations were significant ($p < 0.05$). If a certain correlation was non-significant, this correlation should be excluded from further analysis of slopes comparisons among plant groups.

(ii) We further fitted all six plant traits against climatic variables to explore the effects of environmental factors on plant traits. Similarly, slopes were compared only if the regressions were significant ($p < 0.05$).

(iii) Three steps were performed in SMA analysis (Warton et al., 2006; Pan et al., 2020a). Test 1 for comparing the differences in slopes by running code sma(y∼x × groups,

multcomp = TRUE). Only if the correlation lines among groups were different significantly ($p < 0.05$), the multiple pair-wise comparisons (the argument multcomp = TRUE) in slopes were applied. If the slopes between plant groups were equal ($p > 0.05$), we then tested the shifts along the common slope and elevation differences among the parallel slopes.

Test 2 for testing the shifts along the common slope by sma(y∼x + groups, type = "shift," multcomp = TRUE).

Test 3 for comparing the differences in elevations among the parallel slopes between groups, by sma(y∼x + groups, type = "elevation," multcomp = TRUE).

Only if all three tests were non-significant, should we define that one fit curve was not differ from the other one significantly.

RESULTS

Traits Values of Different Plant Groups

The mean values of leaf traits (N_{mass}, P_{mass}, and SLA) showed significant differences ($p < 0.001$) among different groups (**Figure 3** and **Table 1**).

(i) Emergent plants had significantly higher leaf N_{mass} than those of other groups, but there were no significant differences between submerged plants and their terrestrial counterparts (**Figure 3A**). Aquatic plants had higher leaf P_{mass} (**Figure 3B**) and lower leaf N:P ratios (**Figure 3D**) than terrestrial grassland species.

(ii) Grasses that well adapted to arid and semiarid climate had the lowest SLA, while the submerged plants showed extremely high SLA (**Figure 3C**). Therefore, when mass-based leaf N and P were transformed to area-based traits via dividing by SLA, submerged plants showed significantly lower leaf N_{area} and P_{area} (**Figures 3E,F**). SLA and area-based traits of emergent plants were closer to those of terrestrial plants than submerged plants.

Trait-Trait Correlations Between Leaf N_{mass}, P_{mass} and SLA

By applying the standardized major axis (SMA) analysis, we built the correlations for every trait pairs (**Figure 4**) and compared the pair-wise differences in slopes of the four groups (**Table 2**).

(i) Three out of four groups showed significantly positive leaf N_{mass} – P_{mass} correlations. In general, SMA slopes of the three groups were unequal ($p < 0.01$). The multiple comparisons revealed that the slope of sedges was significantly flatter ($p < 0.01$), while those of the other two plant groups did not differ from each other ($p > 0.05$).

(ii) Leaf N_{mass} of grasses and sedges was positively correlated to SLA ($p < 0.05$). SMA slopes were significantly different between the two groups ($p = 0.007$).

(iii) In terms of the bivariate traits of leaf P_{mass}-SLA, three out of four groups were positive ($p < 0.01$). All SMA slopes of the three groups were equal ($p = 0.247$), which means there were no significant differences among them. Tests of the slope shifts and elevation differences for all undifferentiated slopes showed that the fit curves were different significantly ($p < 0.05$ for shift and/or elevation in **Table 1**), indicating that the trait-trait correlations of different plant groups were separate from each other.

Effects of Environmental Factors on Plant Traits

We further tested the effects of GST, GSP, and GSR on measured plant traits (**Figure 5**) and transformed plant traits (**Figure 6**).

(i) For GST (**Figures 5A,D,G**), leaf N_{mass} of both terrestrial and aquatic plants, leaf P_{mass} of aquatic plants and sedges decreased with increasing GST. Only leaf P_{mass} of grasses increased with increasing GST. SLA of terrestrial plants and aquatic plants showed reversed trends along the temperature gradient.

(ii) For GSP (**Figures 5B,E,H**), leaf N_{mass}, P_{mass}, and SLA of grasses showed negative relationships with GSP, while all these three traits of sedges showed reversed trends.

TABLE 1 | Leaf N, P concentrations on mass-based (mg g^{-1}) and area-based (g m^{-2}), Specific leaf area (SLA, cm^2 g^{-1}) and N:P ratios for two groups of aquatic plants (emergent, ET and submerged, SM) and two groups of terrestrial plants (grasses, GS and sedges, SG).

Leaf traits	Groups	Mean	Std.	Median	CV
N_{mass}					
	GS	22.78^b	3.53	22.20	0.15
	SG	24.99^b	3.76	25.32	0.15
	ET	32.00^a	8.19	32.35	0.26
	SM	24.55^b	7.77	25.86	0.32
P_{mass}					
	GS	1.60^b	0.63	1.54	0.39
	SG	1.74^b	0.67	1.60	0.38
	ET	3.10^a	1.39	2.88	0.45
	SM	2.82^a	1.13	2.76	0.40
SLA					
	GS	125.58^d	19.98	130.98	0.16
	SG	170.08^c	53.61	187.39	0.32
	ET	205.28^b	75.90	197.41	0.37
	SM	661.38^a	173.27	650.38	0.26
N_{area}					
	GS	1.85^a	0.33	1.85	0.18
	SG	1.62^a	0.56	1.51	0.35
	ET	1.77^a	0.79	1.59	0.44
	SM	0.39^b	0.14	0.38	0.35
P_{area}					
	GS	0.13^b	0.07	0.11	0.51
	SG	0.11^b	0.04	0.10	0.38
	ET	0.16^a	0.07	0.15	0.42
	SM	0.04^c	0.02	0.04	0.37
Mass-based N:P ratio					
	GS	15.90^a	5.22	14.74	0.33
	SG	15.98^a	4.90	15.56	0.31
	ET	12.04^b	5.23	11.22	0.43
	SM	9.37^c	2.82	8.99	0.30

Number of observations (Num.), arithmetic mean values (Mean) and standard deviations (Std.), median values, and CV were listed. The coefficients of variation (CV) were calculated by dividing Std. by Mean, and represented the variability in leaf nutrient elements within plant group. Different letters that marked on Mean indicated significant differences in mean leaf N_{mass}, P_{mass}, SLA, N_{area}, P_{area}, and N:P ratio between different plant groups, respectively ($p < 0.05$). Number of observations (n): ET, 269; SM, 30; GS, 31, SG, 34.

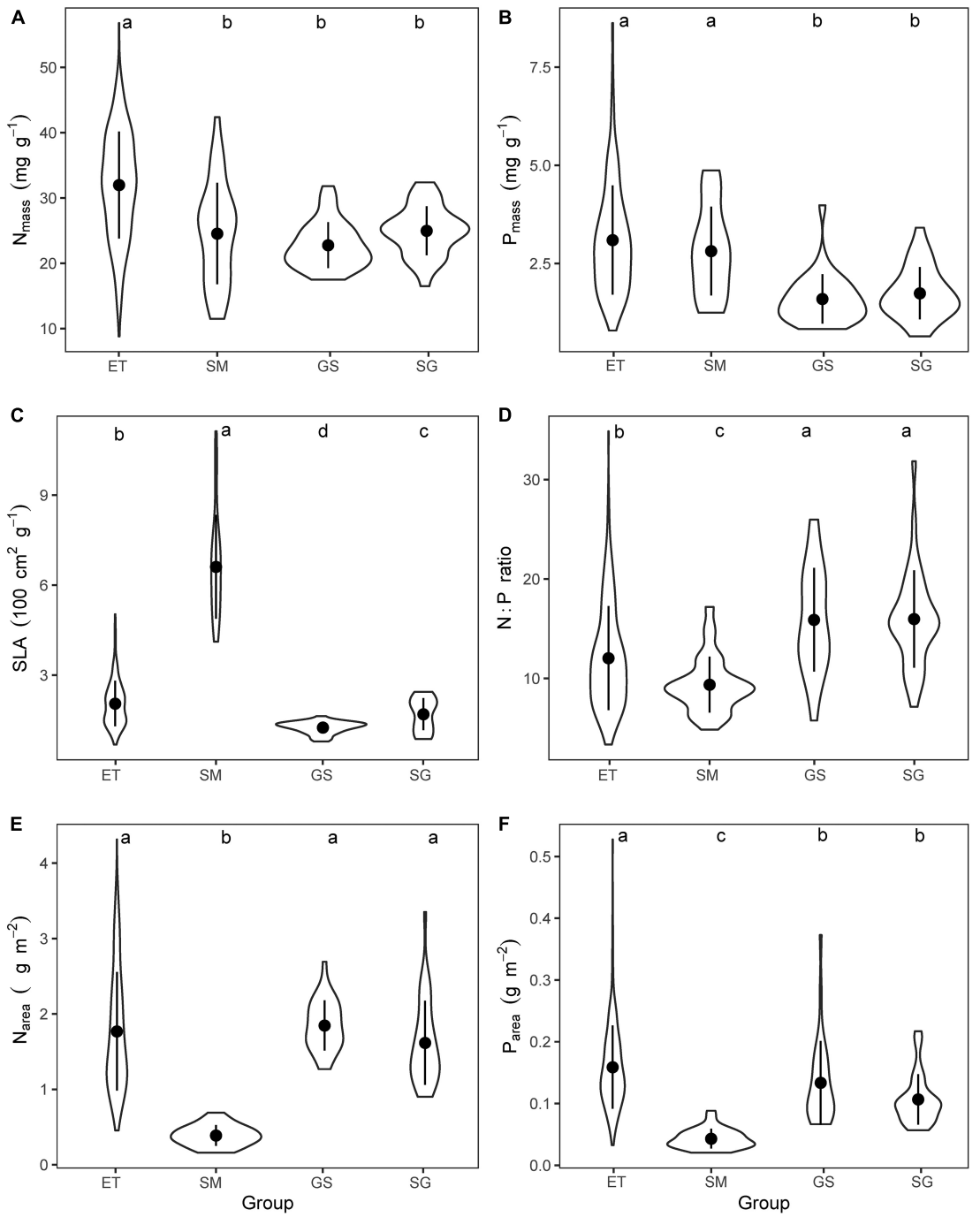


FIGURE 3 | Comparing the mean values of **(A)** mass-based leaf nitrogen concentration (N_{mass}), **(B)** mass-based leaf phosphorus concentration (P_{mass}), **(C)** specific leaf area (SLA), **(D)** mass-based N:P ratios, **(E)** area-based leaf nitrogen content (N_{area}), and **(F)** area-based leaf phosphorus content (P_{area}) for the four plant groups. ET, emergent plants, GS, grasses, SG, sedges, and SM, submerged plants. The lowercase letters on the top of each panel showed the results of Kruskal–Wallis test and Dunn's *Post Hoc* Multiple Comparisons ($p < 0.05$).

(iii) For GSR (**Figures 5C,F,I**), the effects of radiation on SLA were more consistent and significant. SLA of emergent plants ($p = 0.066$) and both terrestrial plants ($p < 0.05$) decreased toward intense radiation.

(iv) In terms of transformed plant traits, each of them was related to two measured traits (mass-based traits and SLA). Affected by the different variation rates (slopes of the regression lines) of single measured traits, transformed traits showed

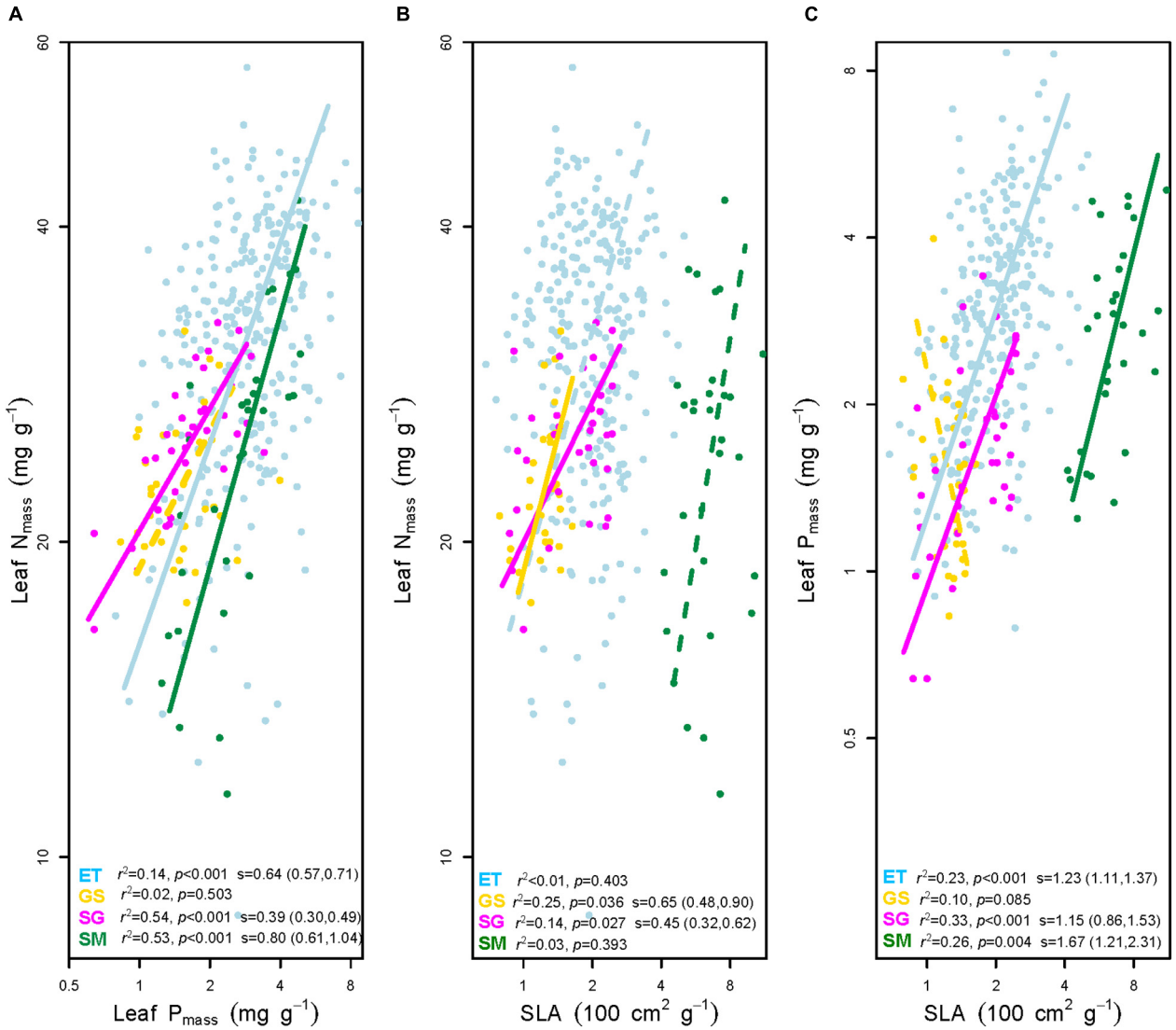


FIGURE 4 | Bivariate correlations between mass-based leaf nitrogen concentration (N_{mass}), mass-based leaf phosphorus concentration (P_{mass}), and specific leaf area (SLA). **(A)** Leaf N_{mass} vs. leaf P_{mass}. **(B)** Leaf N_{mass} vs. SLA. **(C)** Leaf P_{mass} vs. SLA. ET, emergent plants, GS, grasses, SG, sedges, and SM, submerged plants. Solid lines denoted that the two traits correlated with each other at significance level $p < 0.05$, while the dashed lines showed non-significant correlations. The coefficients of determination (r^2) and p values for trait-trait correlations were given on the bottom of each panel. The slopes (s) with 95% confidence interval of each regression line were added only if $p < 0.05$.

different patterns from measured ones along environmental gradients (**Figure 6**).

DISCUSSION

Plant functional traits and trait syndromes reflect both evolutionary trade-offs and ecological fitness to ambient environment (Pérez-Harguindeguy et al., 2013; Díaz et al., 2016). The major differences between aquatic and terrestrial habitats are soil water conditions and the low light intensity, slow diffusion of gases (CO_2 and O_2) underwater, resulting in different trait values and scales of trait-trait correlations between aquatic and terrestrial plants. In this study, we mainly focused on the differences and similarities in plant functional traits and the bivariate traits correlations between aquatic and terrestrial herbaceous plants in harsh environmental conditions on the Tibetan Plateau.

Differences in Leaf Traits Between Aquatic and Terrestrial Plants

First, we found that differences existed in N_{mass}, P_{mass}, SLA, N_{area}, and P_{area} between plant groups. Terrestrial plant groups have lower N_{mass}, P_{mass}, and SLA than aquatic plants (except N_{mass} of submerged plants), of which grasses have the lowest traits values (**Figure 3**). Water availability is the primary climatic factor

TABLE 2 | The multiple comparisons of the trait-trait correlations between every two plant groups.

	Log P_{mass}			Log SLA			Group pair
	Slope	**Shift**	**Elevation**	**Slope**	**Shift**	**Elevation**	
log N_{mass}	–	–	–	–	–	–	ET-GS
	<0.001	/	/	–	–	–	ET-SG
	0.117	0.003	<0.001	–	–	–	ET-SM
	–	–	–	0.007	/	/	GS-SG
	–	–	–	–	–	–	GS-SM
	<0.001	/	/	–	–	–	SG-SM
log P_{mass}				–	–	–	ET-GS
				0.630	<0.001	<0.001	ET-SG
				0.085	<0.001	<0.001	ET-SM
				–	–	–	GS-SG
				–	–	–	GS-SM
				0.089	<0.001	<0.001	SG-SM

The plant groups which had no significant trait-trait correlations were excluded from slope comparisons between groups. Tests for the shifts among the common slope and the elevation differences among the parallel slopes were applied only if the slopes were undifferentiated. "–" denotes that there is at least one plant group in each group pair had no significant trait-trait correlation, and thus cannot fit a meaningful correlation line for the trait pair. "/" denotes that the difference in slopes of the two plant groups is significant and the tests of slope shifts and elevation differences are not applicable.

in determining the distribution of herbaceous species on the Tibetan Plateau (Zheng and Zhao, 2017). Grass of *Stipa purpurea* is the most dominant herbaceous species of alpine steppe in western part of the Tibetan Plateau, where the environmental conditions are characterized by low precipitation, intense solar radiation and lack of soil nutrient availability (Zheng and Zhao, 2017). All of these environmental factors force *Stipa purpurea* constructing leaves with lower SLA (thicker or denser leaves) and lower N_{mass} and P_{mass} to survive there (Poorter et al., 2009; Reich, 2014). Water content of the leaf determines SLA (Shipley, 1995). In arid and semi-arid regions, increasing temperature results in higher evapotranspiration and aggravate the existing water deficiency. This may explain the lowest SLA of grasses and the downward trend of SLA within this plant group with increasing GST (**Figure 5G**), rather than limited increase of GSP in areas where evapotranspiration is far greater than precipitation. Sedges of genera *Kobresia or Carex* dominate the community of alpine meadow, where water stress is relatively alleviated. SLA of sedges are higher than that of grasses, and increase with increasing GSP. Light regime is another environmental factor in affecting plant leaf construction (Long et al., 2011; Dalke et al., 2018). An *in situ* field experiment showed that enhanced radiation induced thicker leaves of two common but not dominant species, *Saussurea superba* and *Gentiana straminea* on the Tibetan Plateau (Shi et al., 2004), resulting in lower SLA. In this study, leaves of two terrestrial plant groups (grasses and sedges) and emergent plants (weak correlation, $p = 0.066$) lower their SLA toward strong radiation (**Figure 5I**). Hu et al. (2012) found that both leaf thickness and leaf density were negatively correlated with SLA of *Stipa purpurea* on the Tibetan Plateau. Whether the variation in SLA is determined by leaf density or by leaf thickness in this study needs further study. Submerged plants have extremely high SLA, partly because thinner leaves are helpful to light capture and gas exchange in low light and low CO_2 pressure environments (Pierce et al., 2012; Liu et al., 2021), and partly because of the reduced investments in support tissues under water (Poorter et al., 2009). Emergent plants, which only the roots are inundated while the leaves stay in air facing atmospheric surrounding conditions directly, have similar SLA to their terrestrial counterparts.

N_{mass}, P_{mass}, and N:P ratios also showed significant differences among plant groups. N_{mass}, P_{mass}, and N:P ratios in this study were nearly identical to our previous data of both emergent plants and submerged plants (Wang et al., 2015), and the data of genera of *Stipa* and *Kobresia* on the Tibetan Plateau (He et al., 2008). N_{mass} of emergent plants are higher than those of the other three plant groups, while P_{mass} of two groups of aquatic plants are significantly higher than two terrestrial plant groups in this study (**Figures 3A,B**). Correspondingly, Substrates of emergent plants contain the highest soil N contents among the four groups, and soil P contents in two groups of aquatic plants substrates are significantly enriched than terrestrial substrate (**Figures 2D,E**). Thus, the differences in leaf nutrient$_{mass}$ between aquatic and terrestrial plants may be caused by the differences in soil nutrients conditions. With respect to plant N:P ratio, it is sometimes regarded as an indicator assessing whether the availability of N or P is more limiting for plant growth (Güsewell and Koerselman, 2002; Pérez-Harguindeguy et al., 2013). The plant N:P ratios 14∼16 are critical values, of which lower than 14, higher than 16, and between 14 and 16, indicates N limitation, P limitation, and co-limiting by N and P, respectively (Koerselman and Meuleman, 1996). However, this empirical threshold should only be used when factors other than N and P are unlikely to limit plant growth (Güsewell and Koerselman, 2002). The grassland region of the Tibetan Plateau is a youthful geologic unit with soil types of Xerosols and Cambisols. Low vegetation coverage (especially in alpine steppe regions) limits the input of organic litters, together with the low rate of nitrogen mineralization induced by cold weather, resulting in low soil available N and extensive N limitation of plants (Zheng and Zhao, 2017; Kou et al., 2020). In contrast, P comes from the weathering of soil parent material and

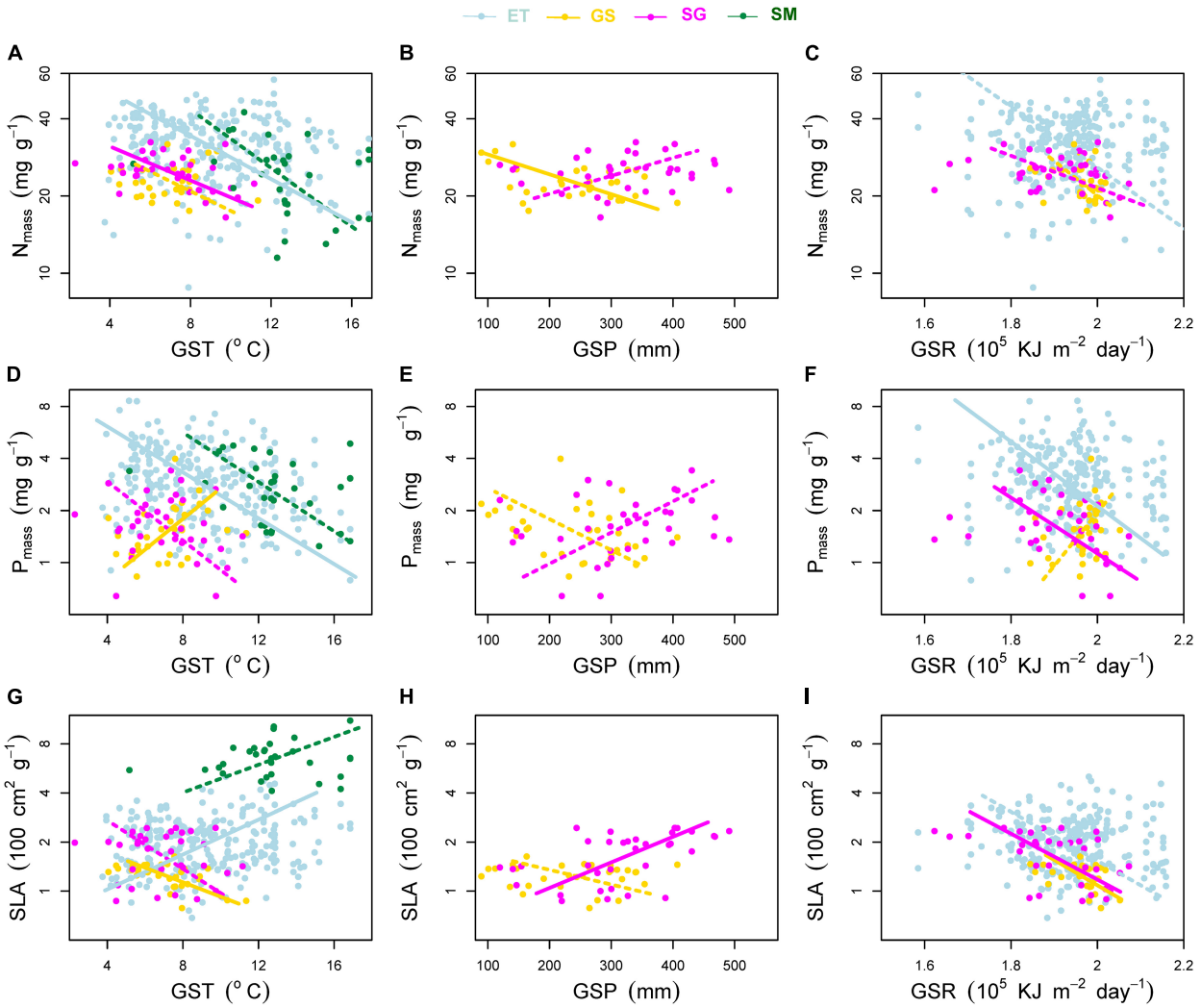


FIGURE 5 | Effects of environmental factors on **(A–C)** mass-based leaf nitrogen concentration (N_{mass}), **(D–F)** mass-based leaf phosphorus concentration (P_{mass}) and **(G–I)** specific leaf area (SLA). Environmental factors include growing season mean air temperature (GST), growing season mean precipitation (GSP), and growing season mean solar radiation (GSR). ET, emergent plants, GS, grasses, SG, sedges, and SM, submerged plants. Solid lines denoted that the environmental factors affected the plant traits significantly at significance level $p < 0.05$, while the dashed lines showed non-significant relationships.

thus young soil on the Tibetan Plateau maintains high soil P but low N contents (Han et al., 2005), resulting in very low soil N:P ratios. Although the mean N:P ratio of grasses (15.90) and sedges (15.98) are very close to 16, lack of N supply, rather than P, should be the primary factor limiting plant growth in grassland regions on the Tibetan Plateau. Low N:P ratios of aquatic plants are generated by the high P_{mass}, which is caused by concentrated P contents in aquatic substrates. Substrate in aquatic habitat always receives leached nutrient elements as a sink from terrestrial soils by surface runoff, resulting in nutrients enrichment, especially P, in aquatic substrates (Reynolds and Davies, 2001). Another potential explanation is that lower N:P ratio (together with high SLA), as adaption strategies, helps submerged plants maintain rapid growth rate in stressful conditions (e.g., low light) (Reich and Oleksyn, 2004; Pierce et al., 2012).

Area-based leaf nutrient contents, which generally are converted from dividing $nutrient_{mass}$ by SLA, are positively correlated with potential photosynthetic capacity per unit leaf area (Reich, 2014; Onoda et al., 2017). In this study, both N_{mass} and P_{mass} of grasses were relatively lower in mean values than those of sedges. However, significantly lower SLA of grasses generated higher N_{area} and P_{area} (**Figure 3**). The allometric variations in $nutrient_{mass}$ and SLA help species that survives in arid regions maintain high photosynthetic rates when stomatal conductance is depressed (Farquhar et al., 2002; Onoda et al., 2017). Aquatic plants always exhibit high acquisitive strategies with high $nutrient_{mass}$ and high SLA, and grow faster (Pierce et al., 2012). Extending leaf area or thinning leaf thickness to increase SLA are the most cost-effective strategies that maximize light capture and CO_2 absorption in maintaining relative growth

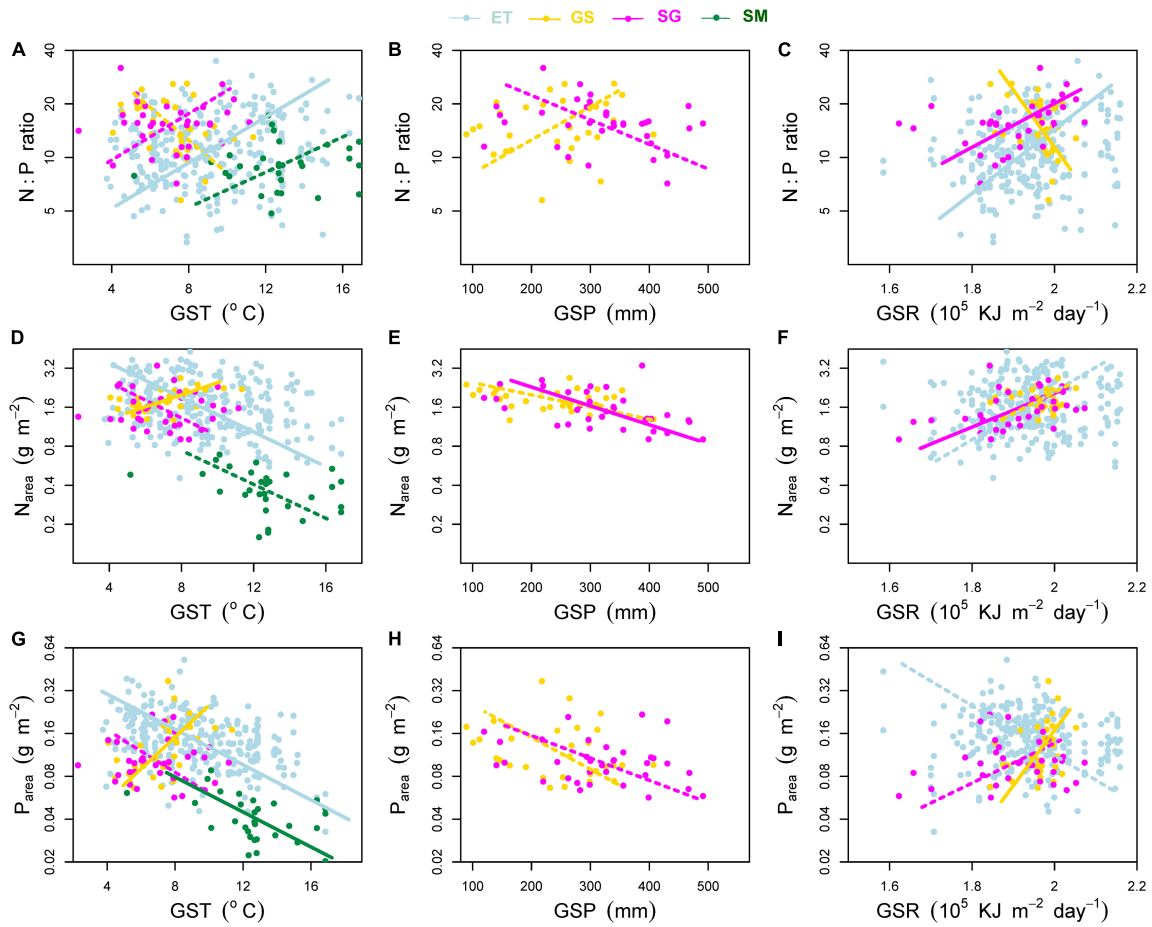


FIGURE 6 | Effects of environmental factors on **(A–C)** mass-based N:P ratio, **(D–F)** area-based leaf nitrogen content (N_{area}) and **(G–I)** area-based leaf phosphorus content (P_{area}). Environmental factors include growing season mean air temperature (GST), growing season mean precipitation (GSP), and growing season mean solar radiation (GSR). ET, emergent plants, GS, grasses, SG, sedges, and SM, submerged plants. Solid lines denoted that the environmental factors affected the plant traits significantly at significance level $p < 0.05$, while the dashed lines showed non-significant relationships.

rate in low irradiance conditions (e.g., underwater) (Pierce et al., 2012; Maberly and Gontero, 2018). Compared to terrestrial plants, the increases in SLA of submerged plants were much greater than that of nutrient$_{mass}$, resulting in low nutrient$_{area}$ (**Figure 3**), i.e., low photosynthetic rate per unit leaf area (Onoda et al., 2017). However, extremely high SLA, together with high nutrient$_{mass}$ would counterbalance the negative influences of low nutrient$_{area}$ on photosynthetic capacity of submerged plants, and keep relatively high plant growth rate.

Although He et al. (2008) found that precipitation and temperature had weak influence on N_{mass} and N:P ratios, our results showed that both mass- and area-based N and P concentrations of all plant groups (except grasses) had weakly increased trends, and N:P ratios decreased with decreasing temperature (**Figures 5A,D, 6A,D,G**), consisting with our previous study of aquatic plants (Wang et al., 2015). The temperature-plant physiological hypothesis (TPPH), which suggests that plants require more N and P (and low N:P ratio) to counterbalance the depressed biochemical efficiency caused by low temperature (Reich and Oleksyn, 2004), provides a reasonable explanation for our results. Soil N and P contents had no significant effects on leaf nutrient$_{mass}$ of other three plant groups and N_{mass} of grasses (data not shown). With respect to leaf P_{mass} of grasses, significant positive effect of soil P contents on it ($r^2 = 0.33$, $p < 0.01$) induced its different pattern along climatic gradients from other plant groups.

Generality of Trait-Trait Correlations and Shifts in Common Slopes Between Aquatic and Terrestrial Plants

Our results verified the general bivariate relationships between N_{mass}, P_{mass}, and SLA, which have been well documented both on global (Güsewell and Koerselman, 2002; Reich et al., 2010; Pan et al., 2020a) and regional (He et al., 2006; Xia et al., 2014; Dalke et al., 2018) scales. Thomas et al. (2020) extended the plant trait relationships to tundra biome, and found that tundra plants demonstrated similar resource economic traits in extremely harsh environments, just like that in the alpine climate in this study. He et al. (2006) reported the generality of interspecific leaf

trait relationships in herbaceous plants on the Tibetan Plateau. In this study, leaf nutrient$_{mass}$ positively correlated with SLA (except P$_{mass}$ – SLA of grasses) in both terrestrial and aquatic plant groups (**Figure 4**), supporting the convergent evolution in plant functioning (Reich et al., 1997). However, due to the unique habitat conditions, significant differences in trait-trait correlations also can be observed between terrestrial and aquatic plants (Pan et al., 2020a). For example, submerged plants face very different surrounding environments from terrestrial plants, of which pH is one of the most important factors. Because submerged plants are submersed in water, the water pH controls the balance of different forms of dissolved inorganic carbon (DIC) directly and thus affects the photosynthesis of submerged plants (Su and Li, 2005). In addition, the accumulation of leaf N and P of aquatic plants on the Tibetan Plateau were significantly influenced by water pH (Wang et al., 2015). The differences of conditions between aquatic and terrestrial habitats force the plants adjust the trait coupling relationships precisely to survive in certain habitats (Pan et al., 2020b).

N and P are the two most interested photosynthesis-related macroelements but often limit plant growth as limiting nutrients. Based on more than 9300 observations, Reich et al. (2010) generalized that the positive relation of leaf N$_{mass}$ and P$_{mass}$ had a uniform 2/3 scaling across taxonomical groups and biomes. However, in this study, the SMA slops of N$_{mass}$-P$_{mass}$ relationships showed significant deviation from the value of 2/3, and differed from each other between terrestrial and aquatic plants (**Figure 4** and **Table 1**). The disproportionate accumulation of leaf N and P most likely caused the deviation of slopes (Wang et al., 2015; Pan et al., 2020a). For example, the mean leaf N$_{mass}$ of submerged plants was nearly equal to that of sedges (24.55 vs. 24.99 mg g^{-1}), while the mean leaf P$_{mass}$ was 62% higher (2.82 vs. 1.74 mg g^{-1}, **Supplementary Appendix Table 1**). As mentioned in Section "Differences in Leaf Traits Between Aquatic and Terrestrial Plants," soils on the Tibetan Plateau are characterized by fertile P but lack of N. Furthermore, both soil N and P contents were relatively lower in terrestrial soils than aquatic substrates (**Figure 2**). The much severer N limitation in terrestrial plants resulted in smaller varied range of N at a given range of P variation (flatter slope) than aquatic plants (**Figure 4**).

Another generally accepted bivariate trait correlation is the positive SLA-nutrient$_{mass}$ relationship across plant groups (Onoda et al., 2017; Pan et al., 2020a; Thomas et al., 2020). With respect to SLA, it is very variable between species and even between replicates, and the affecting factors of SLA are more complex than those of leaf N$_{mass}$ (Wilson et al., 1999), especially in aquatic habitats. When compared the variations in leaf traits between terrestrial and submerged plants by mean values, we found that N$_{mass}$, P$_{mass}$, and SLA varied no differences, less than 2-fold and almost 5-fold, respectively. In the case of limited variation ranges of leaf nutrient$_{mass}$, increasing SLA is considered to be one of the most favorable strategies to adapt to the submerged environment (Pierce et al., 2012), inducing the allometric variation rates in leaf nutrient$_{mass}$ and SLA between submerged plants and other plant groups. At given leaf nutrient$_{mass}$, submerged plants can invest C to construct thinner leaves with high SLA.

In general, aquatic plants, especially submerged plants, are characterized by higher SLA, greater leaf nutrient$_{mass}$ than terrestrial plants, tend to pursue fast-return investment strategies, and represent the acquisitive end of leaf economics spectrum. The deviations of trait-trait relationships between different plant groups reveal the precise adaptions of submerged plants to the unique aquatic habitats.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

LY, ZW, and DY designed the study. LY, HZ, ZZ, and ZW performed the field investigation. LY and ZW analyzed the data and wrote the manuscript. All authors worked together to produce the final submitted version of the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant 31300296) and Plateau Ecology Youth Innovative Fund of Wuhan University (413100105).

ACKNOWLEDGMENTS

We thank Xinwei Xu in Wuhan University, Xiujuan Qiao in Wuhan Botanical Garden, Chinese Academy of Sciences, for the valuable comments on the manuscript, Yingji Pan in Leiden University, Netherlands, for the helpful discussion on data analysis. All the leaf and soil samples were measured in the National Field Station of Freshwater Ecosystem of Liangzi Lake, Wuhan University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.706237/full#supplementary-material

REFERENCES

Bowman, R. (1988). A rapid method to determine total phosphorus in soils. *Soil Sci. Soc. Am. J.* 52, 1301–1304. doi: 10.2136/sssaj1988.03615995005200050016x

Chapin, F. S. III., Autumn, K., and Pugnaire, F. (1993). Evolution of suites of traits in response to environmental stress. *Am. Nat.* 142, S78–S92. doi: 10.1086/285524

Dalke, I. V., Novakovskiy, A. B., Maslova, S. P., and Dubrovskiy, Y. A. (2018). Morphological and functional traits of herbaceous plants with different functional types in the European Northeast. *Plant Ecol.* 219, 1295–1305. doi: 10.1007/s11258-018-0879-2

Díaz, S., Kattge, J., Cornelissen, J. H. C., Wright, I. J., Lavorel, S., Dray, S., et al. (2016). The global spectrum of plant form and function. *Nature* 529, 167–171. doi: 10.1038/nature16489

Fang, J., Wang, X., Shen, Z., Tang, Z., He, J., Yu, D., et al. (2009). Methods and protocols for plant community inventory. *Biodivers. Sci.* 17, 533–548. doi: 10.3724/SP.J.1003.2009.09253

Farquhar, G. D., Buckley, T. N., and Miller, J. M. (2002). Optimal stomatal control in relation to leaf area and nitrogen content. *Silva Fenn.* 36, 625–637. doi: 10.14214/sf.530

Güsewell, S., and Koerselman, W. (2002). Variation in nitrogen and phosphorus concentrations of wetland plants. *Perspect. Plant Ecol. Evol. Syst.* 5, 37–61. doi: 10.1078/1433-8319-0000022

Han, W. X., Fang, J. Y., Guo, D. L., and Zhang, Y. (2005). Leaf nitrogen and phosphorus stoichiometry across 753 terrestrial plant species in China. *New Phytol.* 168, 377–385. doi: 10.1111/j.1469-8137.2005.01530.x

He, J., Wang, L., Flynn, D. F. B., Wang, X., Ma, W., and Fang, J. (2008). Leaf nitrogen:phosphorus stoichiometry across Chinese grassland biomes. *Oecologia* 155, 301–310. doi: 10.1007/s00442-007-0912-y

He, J., Wang, Z., Wang, X., Schmid, B., Zuo, W., Zhou, M., et al. (2006). A test of the generality of leaf trait relationships on the Tibetan plateau. *New Phytol.* 170, 835–848. doi: 10.1111/j.1469-8137.2006.01704.x

Hu, M., Zhang, L., Luo, T., and Shen, W. (2012). Variations in leaf functional traits of stipa purpurea along a rainfall gradient in Xizang, china. *Chin. J. Plant Ecol.* 36, 136–143. doi: 10.3724/SP.J.1258.2012.00136

Koerselman, W., and Meuleman, A. (1996). The vegetation N:P ratio: a new tool to detect the nature of nutrient limitation. *J. Appl. Ecol.* 33, 1441–1450. doi: 10.2307/2404783

Kou, D., Yang, G., Li, F., Feng, X., Zhang, D., Mao, C., et al. (2020). Progressive nitrogen limitation across the Tibetan alpine permafrost region. *Nat. Commun.* 11:3331. doi: 10.1038/s41467-020-17169-6

Kuo, S. (1997). "Phosphorus," in *Methods of Soil Analysis Part 3. Chemical Methods*, ed. J. Bigham (Madison: Soil Science Society of America), 869–919.

Liu, H., Liu, G., and Xing, W. (2021). Functional traits of submerged macrophytes in eutrophic shallow lakes affect their ecological functions. *Sci. Total Environ.* 760:143332. doi: 10.1016/j.scitotenv.2020.143332

Long, W., Zang, R. G., Schamp, B. S., and Ding, Y. (2011). Within- and among-species variation in specific leaf area drive community assembly in a tropical cloud forest. *Oecologia* 167, 1103–1113. doi: 10.1007/s00442-011-2050-9

Maberly, S. C., and Gontero, B. (2018). "Trade-offs and synergies in the structural and functional characteristics of leaves photosynthesizing in aquatic environments," in *The Leaf: A Platform for Performing Photosynthesis*, eds W. W. Adams III and I. Terashima (Berlin: Springer), 307–343. doi: 10.1007/978-3-319-93594-2_11

Moor, H., Rydin, H., Hylander, K., Nilsson, M. B., Lindborg, R., and Norberg, J. (2017). Towards a trait-based ecology of wetland vegetation. *J. Ecol.* 105, 1623–1635. doi: 10.1111/1365-2745.12734

Onoda, Y., Wright, I. J., Evans, J. R., Hikosaka, K., Kitajima, K., Niinemets, U., et al. (2017). Physiological and structural tradeoffs underlying the leaf economics spectrum. *New Phytol.* 214, 1447–1463. doi: 10.1111/nph.14496

Pan, Y., Cieraad, E., Armstrong, J., Armstrong, W., Clarkson, B., Colmer, T., et al. (2020a). Global patterns of the leaf economics spectrum in wetlands. *Nat. Commun.* 11:4519. doi: 10.1038/s41467-020-18354-3

Pan, Y., Cieraad, E., Clarkson, B. R., Colmer, T. D., Pedersen, O., Visser, E. J. W., et al. (2020b). Drivers of plant traits that allow survival in wetlands. *Funct. Ecol.* 34, 956–967. doi: 10.1111/1365-2435.13541

Pérez-Harguindeguy, N., Díaz, S., Garnier, E., Lavorel, S., Poorter, H., Jaureguiberry, P., et al. (2013). New handbook for standardised measurement of plant functional traits worldwide. *Aust. J. Bot.* 61, 167–234. doi: 10.1071/BT12225

Pierce, S., Brusa, G., Sartori, M., and Cerabolini, B. E. L. (2012). Combined use of leaf size and economics traits allows direct comparison of hydrophyte and terrestrial herbaceous adaptive strategies. *Ann. Bot.* 109, 1047–1053. doi: 10.1093/aob/mcs021

Pohlert, T. (2014). *The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR)*. Available Online at: https://CRAN.R-project.org/package=PMCMR. R package.

Poorter, H., Niinemets, U., Poorter, L., Wright, I. J., and Villar, R. (2009). Causes and consequences of variation in leaf mass per area (lma): a meta-analysis. *New Phytol.* 182, 565–588. doi: 10.1111/j.1469-8137.2009.02830.x

Reich, P. B. (2014). The world-wide 'fast-slow' plant economics spectrum: a traits manifesto. *J. Ecol.* 102, 275–301. doi: 10.1111/1365-2745.12211

Reich, P. B., Ellsworth, D. S., Walters, M. B., Vose, J. M., Gresham, C., Volin, J. C., et al. (1999). Generality of leaf trait relationships: a test across six biomes. *Ecology* 80, 1955–1969. doi: 10.2307/176671

Reich, P. B., and Oleksyn, J. (2004). Global patterns of plant leaf N and P in relation to temperature and latitude. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11001–11006. doi: 10.1073/pnas.0403588101

Reich, P. B., Walters, M. B., and Ellsworth, D. S. (1997). From tropics to tundra: global convergence in plant functioning. *Proc. Natl. Acad. Sci. U. S. A.* 94, 13730–13734. doi: 10.1073/pnas.94.25.13730

Reich, P. B., Wright, I. J., Niklas, J. O., Hedin, K. J., and Elser, J. J. (2010). Evidence of a general 2/3-power law of scaling leaf nitrogen to phosphorus among major plant groups and biomes. *Proc. Biol. Sci.* 277, 877–883. doi: 10.1098/rspb.2009.1818

Reynolds, C. S., and Davies, P. S. (2001). Sources and bioavailability of phosphorus fractions in freshwaters: a British perspective. *Biol. Rev. Camb. Philos. Soc.* 76, 27–64. doi: 10.1111/j.1469-185X.2000.tb00058.x

Shi, S., Zhu, W. Y., Li, H. M., Zhou, D. W., Han, F., Zhao, X. Q., et al. (2004). Photosynthesis of *Saussurea superba* and *Gentiana straminea* is not reduced after long-term enhancement of UV-B radiation. *Environ. Exp. Bot.* 51, 75–83. doi: 10.1016/S0098-8472(03)00062-5

Shipley, B. (1995). Structured interspecific determinants of specific leaf area in 34 species of herbaceous angiosperms. *Funct. Ecol.* 9, 312–319. doi: 10.2307/2390579

Su, R. L., and Li, W. (2005). Advances in research on photosynthesis of submerged macrophytes. *Chin. Bull. Botany* 22, 128–138.

Thomas, H. J. D., Bjorkman, A. D., Myers-Smith, I. H., Elmendorf, S. C., Kattge, J., Diaz, S., et al. (2020). Global plant trait relationships extend to the climatic extremes of the tundra biome. *Nat. Commun.* 11:1351. doi: 10.1038/s41467-020-15014-4

Tibetan Plateau Scientific Expedition and Research Team (TPSERT). (1992). *Tibetan Grassland*. Beijing: Science Press.

Wang, Z., Luo, T., Li, R., Tang, Y., and Du, M. (2013). Causes for the unimodal pattern of biomass and productivity in alpine grasslands along a large altitudinal gradient in semi-arid regions. *J. Veg. Sci.* 24, 189–201. doi: 10.1111/j.1654-1103.2012.01442.x

Wang, Z., Xia, C., Yu, D., and Wu, Z. (2015). Low-temperature induced leaf elements accumulation in aquatic macrophytes across Tibetan plateau. *Ecol. Eng.* 75, 1–8. doi: 10.1016/j.ecoleng.2014.11.015

Warton, D. I., Duursma, R. A., Falster, D. S., and Taskinen, S. (2012). smatr 3 - an r package for estimation and inference about allometric lines. *Methods Ecol. Evol.* 3, 257–259. doi: 10.1111/j.2041-210X.2011.00153.x

Warton, D. I., Wright, I. J., Falster, D. S., and Westoby, M. (2006). Bivariate line-fitting methods for allometry. *Biol. Rev. Camb. Philos. Soc.* 81, 259–291. doi: 10.1017/S1464793106007007

Wilson, P. J., Thompson, K., and Hodgson, J. G. (1999). Specific leaf area and leaf dry matter content as alternative predictors of plant strategies. *New Phytol.* 143, 155–162. doi: 10.1046/j.1469-8137.1999.00427.x

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., et al. (2004). The worldwide leaf economics spectrum. *Nature* 428, 821–827. doi: 10.1038/nature02403

Xia, C., Yu, D., Wang, Z., and Xie, D. (2014). Stoichiometry patterns of leaf carbon, nitrogen and phosphorous in aquatic macrophytes in eastern china. *Ecol. Eng.* 70, 406–413. doi: 10.1016/j.ecoleng.2014.06.018

Zhao, X. Q. (2009). *Alpine Meadow Ecosystem and Global Change*. Beijing: Science Press.

Zheng, D., and Zhao, D. (2017). Characteristics of natural environment of the Tibetan plateau. *Sci. Technol. Rev.* 35, 13–22.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 24 August 2021
doi: 10.3389/fevo.2021.680041

Evolutionary Diversity Peaks at Mid-Elevations Along an Amazon-to-Andes Elevation Gradient

Andy R. Griffiths[1], Miles R. Silman[2], William Farfan-Rios[3,4,5], Kenneth J. Feeley[6,7], Karina García Cabrera[2], Patrick Meir[1,8], Norma Salinas[9], Ricardo A. Segovia[10,11,12] and Kyle G. Dexter[1,13]*

[1] School of Geosciences, University of Edinburgh, Edinburgh, United Kingdom, [2] Biology Department, Center for Energy, Environment and Sustainability, Wake Forest University, Winston Salem, NC, United States, [3] Living Earth Collaborative, Washington University in St. Louis, St. Louis, MO, United States, [4] Center for Conservation and Sustainable Development, Missouri Botanical Garden, St. Louis, MO, United States, [5] Herbario Vargas (CUZ), Escuela Profesional de Biología, Universidad Nacional de San Antonio Abad del Cusco, Cusco, Peru, [6] Department of Biology, University of Miami, Coral Gables, FL, United States, [7] Fairchild Tropical Botanic Garden, Coral Gables, FL, United States, [8] Research School of Biology, Australian National University, Canberra, ACT, Australia, [9] Instituto de Ciencias de la Naturaleza, Territorio y Energías Renovables, Pontificia Universidad Católica del Peru, Lima, Peru, [10] Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción, Concepción, Chile, [11] Instituto de Ecología y Biodiversidad, Santiago, Chile, [12] Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral, Valdivia, Chile, [13] Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom

OPEN ACCESS

Edited by:
Danilo M. Neves, Federal University of Minas Gerais, Brazil

Reviewed by:
Felix Forest, Royal Botanic Gardens, Kew, United Kingdom
Sebastian Gonzalez, National University of Colombia, Medellin, Colombia

***Correspondence:**
Andy R. Griffiths
agriff16@hotmail.com

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *13 March 2021*
Accepted: *13 July 2021*
Published: *24 August 2021*

Citation:
Griffiths AR, Silman MR, Farfan-Rios W, Feeley KJ, Cabrera KG, Meir P, Salinas N, Segovia RA and Dexter KG (2021) Evolutionary Diversity Peaks at Mid-Elevations Along an Amazon-to-Andes Elevation Gradient. Front. Ecol. Evol. 9:680041. doi: 10.3389/fevo.2021.680041

Elevation gradients present enigmatic diversity patterns, with trends often dependent on the dimension of diversity considered. However, focus is often on patterns of taxonomic diversity and interactions between diversity gradients and evolutionary factors, such as lineage age, are poorly understood. We combine forest census data with a genus level phylogeny representing tree ferns, gymnosperms, angiosperms, and an evolutionary depth of 382 million years, to investigate taxonomic and evolutionary diversity patterns across a long tropical montane forest elevation gradient on the Amazonian flank of the Peruvian Andes. We find that evolutionary diversity peaks at mid-elevations and contrasts with taxonomic richness, which is invariant from low to mid-elevation, but then decreases with elevation. We suggest that this trend interacts with variation in the evolutionary ages of lineages across elevation, with contrasting distribution trends between younger and older lineages. For example, while 53% of young lineages (originated by 10 million years ago) occur only below ∼1,750 m asl, just 13% of old lineages (originated by 110 million years ago) are restricted to below ∼1,750 m asl. Overall our results support an Environmental Crossroads hypothesis, whereby a mid-gradient mingling of distinct floras creates an evolutionary diversity in mid-elevation Andean forests that rivals that of the Amazonian lowlands.

Keywords: diversity gradient, lineage diversity, lineage age, tropical montane forest, TILD, environmental crossroads

INTRODUCTION

Environmental gradients are commonly associated with trends in diversity, and the elevational diversity gradient (EDG) is one of the most widely discussed and enigmatic patterns in ecology (Humboldt and Bonpland, 1805; Rahbek, 1995; Guo et al., 2013). In the simplest terms, diversity broadly declines with elevation, yet there is substantial variation amongst taxa and across systems, with many non-monotonic and non-linear diversity trends (Rahbek, 1995; Guo et al., 2013).

In systems with complex climatic trends superimposed on elevation, such as tropical montane forests (TMF), the EDG is nuanced and not fully understood, particularly at middle elevations (Lieberman et al., 1996; Kessler, 2000; Nottingham et al., 2018). While many investigations of diversity variation focus on taxonomic measures such as, species richness, interest has expanded to include quantification of evolutionary diversity (e.g., Faith, 1992; Dexter et al., 2019), providing potential for a new temporal perspective to the EDG. Possible contrasts between evolutionary diversity and measures of current taxonomic richness have yet to be fully explored in the context of the EDG. Further, the interaction between the EDG and variation in the evolutionary ages of lineages at different elevation is still unclear.

South American TMF provide an excellent context in which to investigate environmental and evolutionary aspects of the EDG. Within TMF, environmental conditions and ecological processes vary substantially with elevation (Sundqvist et al., 2013), many species occupy narrow thermal niches (Perez et al., 2016), and closely related tree lineages tend to occur at similar elevations (Griffiths et al., 2021). In addition, TMFs are globally outstanding centers of endemism and biodiversity (Myers et al., 2000) and provide vital ecosystem services such as regulation of hydrological processes (Bruijnzeel et al., 2011), soil stabilization, and carbon storage (Zimmermann et al., 2010). They also encompass distinctive tropical montane cloud forests, with the cloud-base ecotone representing an understudied ecological phenomenon (Griffiths et al., 2021). A deeper exploration of the evolutionary dimension of the EDG of trees in TMF can provide important context to our ecological understanding of TMFs, as well as the wider spatial organization of diversity.

TMF elevational diversity patterns for plants vary. However, linear decreases with elevation (Gentry, 1988; Vazquez and Givnish, 1998; Jankowski et al., 2013), plateaus (Tiede et al., 2015), and mid-elevation peaks (Kessler, 2000; Girardin et al., 2014), are consistent with wider EDG variation (Rahbek, 1995; McCain and Grytnes, 2010; Guo et al., 2013). The broad influence of climatic factors, such as temperature and precipitation, on diversity trends (Currie, 1991; Currie et al., 2004; Kreft and Jetz, 2007) are likely to be key factors influencing patterns within TMF. However, tolerance to freezing or other factors (Zanne et al., 2014; Segovia et al., 2020), and eco-geographic barriers, such as the subtropical arid belt (Smith et al., 2012; Bacon et al., 2013) are also thought to modulate broad scale spatial variation in diversity. Similarly, along TMF elevation gradients the environmental shifts occurring at the cloud-base ecotone (Bruijnzeel, 2001; Halladay et al., 2012; Fadrique et al., 2018), such as light availability (Fyllas et al., 2017) and precipitation regime (Rapp and Silman, 2012), can influence the distribution patterns of taxa (Griffiths et al., 2021), and as such may interact with the shape of diversity patterns. To our knowledge there has been no substantial consideration of a cloud-base ecotone influence on diversity patterns.

Integration of an evolutionary dimension into our understanding of spatial variation in diversity has found a fertile testing ground in the warm-to-cool thermal gradients of latitude and elevation. The Tropical Niche Conservatism (TNC) (Wiens and Donoghue, 2004) and Out of the Tropics (OTT) (Jablonski et al., 2006) hypotheses suppose a warm-tropical ancestral origin for the angiosperm clade (Crane et al., 1995). TNC supposes niche conservatism of warm-tropical ancestral characteristics with recent and infrequent dispersal of lineages into cool climates, resulting in higher species richness, evolutionary diversity, and older clade ages in present tropical climates (Wiens and Donoghue, 2004; Holt, 2009). OTT, or Out of the Tropical Lowlands (OTL) in the context of tropical elevation gradients (Qian and Ricklefs, 2016), proposes niche convergence and random, early dispersal into cooler climates followed by slow diversification, resulting in a predominance of older lineages in cool climates, with lower evolutionary relatedness among lineages present and relatively higher evolutionary diversity (Jablonski et al., 2006; Qian and Ricklefs, 2016). Observations to date provide mixed support for the TNC, OTT, and OTL hypotheses: latitudinal decreases in evolutionary diversity (Kerkhoff et al., 2014) and angiosperm family ages (Hawkins et al., 2011; Romdal et al., 2013) have been reported, but also an increase in basal−weighted phylogenetic beta diversity to high latitudes (McFadden et al., 2019). Evolutionary diversity has been found to decrease with elevation in tropical montane forest (Bergamin et al., 2020), but also to increase along with increasing family age (Qian et al., 2014; Tiede et al., 2015).

Perspectives beyond the TNC and OTT/OTL hypotheses have also been explored. For example, biome affiliation and climate may be the key determinants of evolutionary diversity trends in woody plants (Rezende et al., 2017), while the Environmental Crossroads (EC) hypothesis suggests that intermediate conditions along environmental gradients should support the highest diversity due to coexistence of distinct evolutionary lineages from gradient extremes (Neves et al., 2020). The EC hypothesis has yet to be applied to EDGs. By weighing the expectations of the EC hypothesis against predictions of decreases in evolutionary diversity and mean lineage age with elevation (TNC), or increase in evolutionary diversity and mean lineage age with elevation (OTL), we can gain greater insight into the EDG in TMF.

In this paper, we explore the taxonomic and evolutionary diversity of trees along a large Amazon-to-Andes tropical montane forest elevation gradient. While we do not aim to consider patterns of phylogenetic community structure, we do integrate a deeper evolutionary perspective through the analysis of lineage age trends. Framed by the context of broad scale diversity trends predicted by the TNC, Out of the Tropical Lowlands, and Environmental Crossroads hypotheses, our analyses aim to address three core questions: (1) What is the shape of the EDG within tropical montane forest? (2) Do elevational trends of taxonomic and phylogenetic diversity differ along this gradient? (3) Is there elevational variation in lineage age patterns along this gradient?

MATERIALS AND METHODS

Study Site and Plot Inventories

We utilize plot census data for all woody stems $\geq$ 10 cm diameter at breast height (130 cm above the ground; DBH) in a network

of 23 one-hectare forest plots located along an Amazon-to-Andes elevation gradient spanning 425–3625 m asl. This plot network was established and is maintained by the Andes Biodiversity and Ecosystem Research Group (ABERG)[1] and is centered on the Kosñipata valley on the edge of Manu National Park in south-eastern Peru. Given numerous indeterminate species identifications, all analyses presented here were conducted at the genus-level.

From lowland/sub-montane forests at 425 m asl, up to the tropical montane treeline at 3,625 m, the gradient encompasses large variation in habitat and environmental variables (Girardin et al., 2010). The cloud-base ecotone at around 1,500–2,000 m asl marks the lower limit of the tropical montane cloud forest (Girardin et al., 2010; Rapp and Silman, 2012), with a characteristic cloud immersion which reaches peak frequency between 2,000 and 3,500 m asl (Halladay et al., 2012). Along the gradient, mean annual temperatures range from c.24°C at low elevations, to c.9°C at higher elevation (Malhi et al., 2017). Mean total annual precipitation follows a non-linear trend from c.3,000 mm yr^{-1} at low elevations to a peak c.5,000 mm yr^{-1} at around 1,000 m asl before declining monotonically to c.1,000 mm yr^{-1} at the very highest elevations (Rapp and Silman, 2012), though precipitation exceeds potential evapotranspiration at all elevations. The edaphic character of this gradient is complex and varies non-linearly (Zimmermann et al., 2010; Nottingham et al., 2015).

Taxonomic and Phylogenetic Diversity

We quantified taxonomic diversity in terms of genus richness, the number of genera in each plot. We quantified three metrics of phylogenetic diversity for each plot, using a temporally calibrated phylogeny (**Figure 1**), based on matK and rbcL sequences, representing 275 angiosperm, gymnosperm, and tree fern genera, and published in Griffiths et al. (2021). (1) Faith's phylogenetic diversity (PD), the sum of all the branch lengths of a phylogeny for taxa in a community, including the root branch (Faith, 1992). PD is weighted toward diversity in recent evolutionary time (Dexter et al., 2019). (2) The standard effect size of phylogenetic diversity (sesPD), which is obtained by comparing the observed phylogenetic diversity versus. expected phylogenetic diversity. Expected diversity is estimated by random shuffling of the phylogeny tips. Positive values of sesPD indicate greater phylogenetic diversity than expected given species richness, negative values indicate less diversity than expected. (3) time-integrated lineage diversity (TILD) which evenly weights phylogenetic diversity across recent and deep evolutionary time and is calculated by integrating the area under a lineage through time plot (*sensu* Yguel et al., 2016), after log-transforming the y-axis, or the number of lineages at each time point (Dexter et al., 2019). PD and sesPD were calculated using the "picante" package (Kembel et al., 2010), TILD was calculated using a new R function (Dexter et al., 2019; **Supplementary Material**).

To ensure comparability, all diversity metrics were calculated on data subset to the 275 genera in our phylogeny. There is large variation in the number of individual stems sampled across plots for our phylogenetic subset (min = 320, max = 1,509), and richness is influenced by sample size (Colwell et al., 2012). Therefore, before calculation of richness metrics, we rarefied plot data to 320 individual stems per plot, as this was the lowest number of stems recorded among plots in our phylogenetic subset. We conducted 100 rarefactions and from these calculated the mean of each metric for each plot. Some genera are present in plots but not in the phylogeny (n = 26). Therefore, to check genus richness calculated from the phylogenetic subset is representative, we correlated the rarefied genus richness of our phylogenetic subset with the rarefied genus richness of data for all genera (rarefied to 459 individuals per plot, the lowest number of individuals in a plot for the full dataset). Rarefactions were conducted using the "rrarefy" function in the "vegan" package (Oksanen et al., 2018). We then conducted pairwise correlations of the metrics, based on Spearman's rank correlation coefficient.

We applied linear, quadratic, and breakpoint regression models to identify which pattern best describes the elevational richness gradient for each metric. Linear and quadratic regression were conducted using the "lm" function in the R package stats (R Core Team, 2018). Breakpoint regressions were conducted using the "segmented" function in the R package segmented (Muggeo, 2008). The model with the best fit was established based on a comparison of adjusted R^2 values. Due to the potentially strong influence of the long branch lengths leading to tree ferns and gymnosperm, all analyses were repeated excluding these lineages and considering only angiosperms.

Evolutionary Age Patterns

We further investigated the evolutionary dimension of the observed diversity patterns by considering the elevational distribution trends of lineages at different evolutionary depths. Evolutionary depths here represent lineage age time slices on our temporally calibrated phylogeny for every 10 million years from the present day to 382 million years old—our age estimate for the split between seed plants and ferns (Silvestro et al., 2015). This does not reflect an ancestral range reconstruction, but rather the present distribution of lineages at different evolutionary depths. We calculated the number of lineages and elevational range of each lineage for each 10-million-year time slice. We additionally calculated which elevations are occupied by the highest number of lineages at each 10-million-year time slice. As an illustration, at an evolutionary depth of 200 million years, there are a total of three lineages in our phylogeny. One branch leading to all angiosperms, one branch leading to all gymnosperms, and one branch leading to all tree ferns. In terms of present elevational range, angiosperms occur from 425 to 3,625 m asl, gymnosperms occur from 1,250 to 3,000 m asl, and tree ferns occur from 800 to 3,450 m asl. Thus, the maximum lineage diversity for an evolutionary depth of 200 million years is found between 1,250 and 3,000 m asl, where all three lineages are present, while outside this range there are only one or two lineages at this evolutionary depth.

[1] www.andesconservation.org

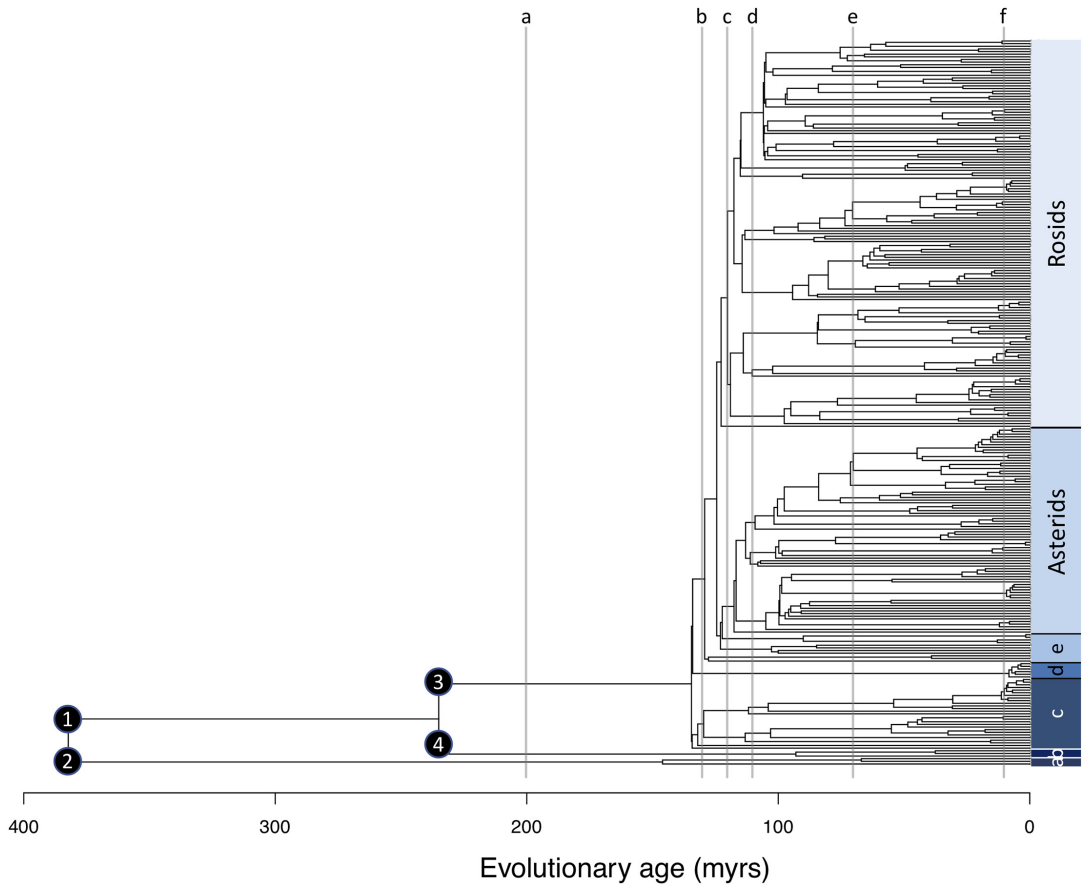


FIGURE 1 | Genus level phylogeny for 275 genera occurring across 23 plots along an elevation gradient from 425 to 3,625 m asl. Numbered nodes relate to lineages mentioned in the text: 1 = seed plants, 2 = tree ferns, 3 = angiosperms, 4 = gymnosperms. Gray vertical lines relate to evolutionary depths of 10, 70, 110, 120, 130, and 200 myrs with letters relating to the panels presented in **Figure 3**. Major lineages are indicated in a bar on the right where a = tree ferns, b = gymnosperms, c = Magnoliids and *Hedyosmum*, d = Monocots, and e = basal Eudicots.

RESULTS

Taxonomic and Phylogenetic Richness

Across all 23 plots, excluding individuals not identified to genus (n = 458), a total of 20,926 individuals were recorded, belonging to 301 genera and 100 families. The most abundant genera were *Cyathea* (2,410 individuals), *Weinmannia* (1,843 individuals), and *Miconia* (1,530 individuals). Based on the raw data, the highest generic richness occurs in plot SPD-02 at 1,500 m asl, with a total of 109 genera. After 100 rarefactions, accounting for the effect of variation in stem density, the highest generic richness remains in plot SPD-02, at 1,500 m asl (mean = 78 genera SD = 3.8; **Figure 1a**). After 100 rarefactions, the highest mean values for phylogenetic richness metrics occur in plot SAI-01 at 1,500 m asl for PD (mean = 6,227 myrs, SD = 288; **Figure 1b**), plot TRU-01 at 3,450 m asl for sesPD (mean = 1.93, SD = 0.23; **Figure 1c**), and plot SAI-01 at 1,500 m asl for TILD (mean = 663 [log(myrs)], SD = 33; **Figure 1d**).

Pairwise correlations show that all metrics, with the exception of sesPD with TILD, display significant positive or negative correlations with each other. Genus richness has a positive correlation with PD (rho = 0.91, $p < 0.001$), a positive correlation with TILD (rho = 0.54, $p < 0.01$), and a negative correlation with sesPD (rho = −0.66, $p < 0.001$). PD is negatively correlated with sesPD (rho = −0.5, $p < 0.05$), and positively correlated with TILD (rho = 0.76, $p < 0.001$). sesPD and TILD are not correlated (rho = 0.04, $p > 0.05$). Further, rarefied genus richness based on our phylogenetic subset in highly correlated with rarefied genus richness for all genera (rho = 1, $p < 0.001$), suggesting genus richness calculated based on our phylogenetic subset is representative of the complete genera data.

Despite these pairwise correlations, the elevational patterns for the different richness metrics differ in terms of the best fitting model (**Table 1**). The elevational pattern of genus richness is best described by a breakpoint regression (adjusted R^2 = 0.9), though the amount of variation explained by this model is not markedly better than a linear regression (adjusted R^2 = 0.87) or quadratic regression (adjusted R^2 = 0.87). The genus richness of plots is relatively constant up until an estimated breakpoint at 1,383 m asl (± 294 m), above which genus richness decreases with elevation (**Figure 2A**). A break point regression also provides the best fit for the elevational pattern of PD (adjusted R^2 = 0.94); PD increases

TABLE 1 | Comparison of regression model fits for elevational patterns of genus richness, phylogenetic diversity (PD), standardized effect size phylogenetic diversity (sesPD), and time-integrated lineage diversity (TILD).

	Linear regression		Quadratic regression		Breakpoint regression	
	Adjusted R^2	*p*-value	Adjusted R^2	*p*-value	Adjusted R^2	*p*-value
Genus richness	0.87	<0.001	0.87	<0.001	**0.89**	N/A
PD	0.81	<0.001	0.89	< 0.001	**0.94**	N/A
sesPD	0.46	<0.001	**0.75**	<0.001	0.73	N/A
TILD	0.19	0.022	**0.77**	<0.001	0.66	N/A

Numbers in bold indicate highest model R^2 value for each diversity measure.

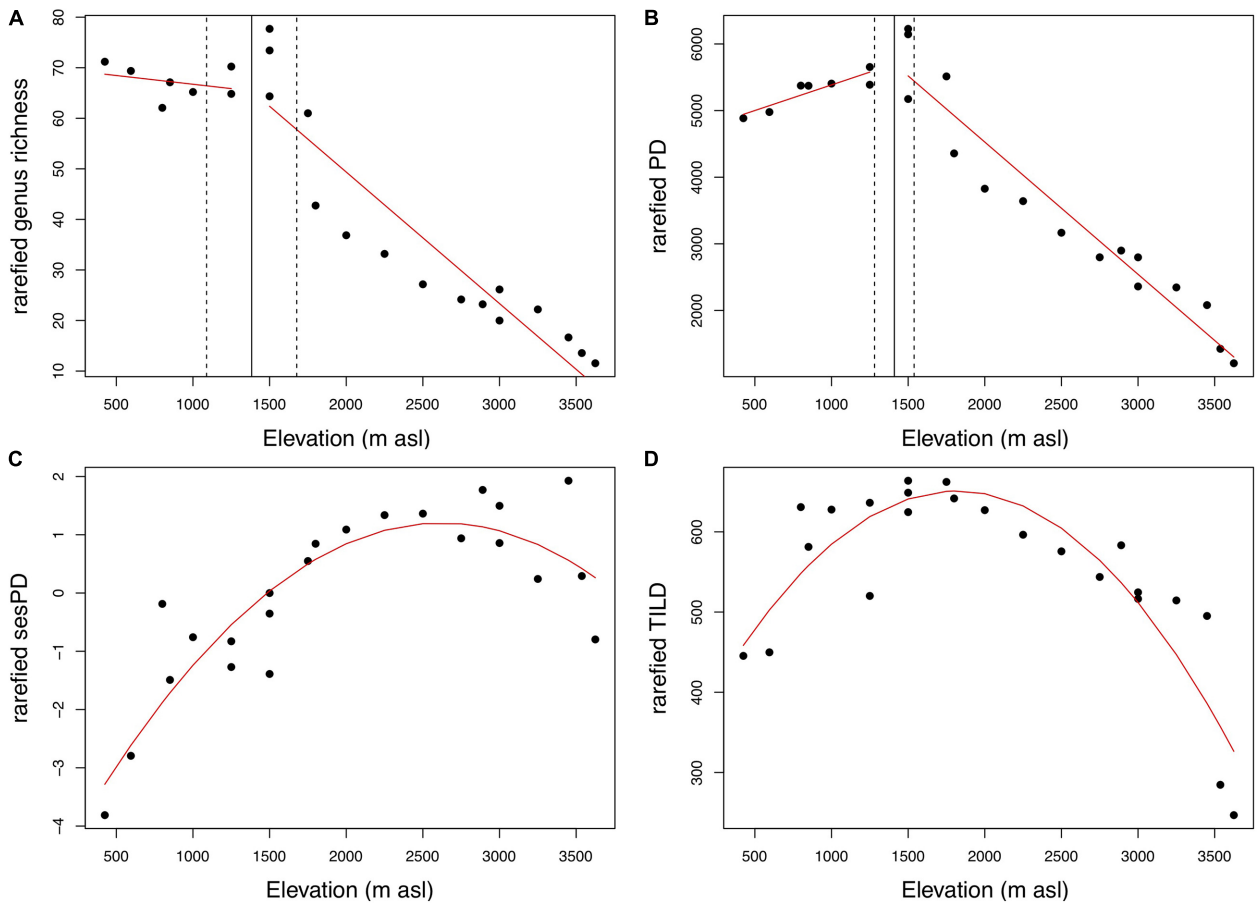


FIGURE 2 | The shape of the elevational richness gradient varies between metrics. **(A)** Genus richness and **(B)** phylogenetic diversity (PD) are best described by breakpoint regression models. **(C)** Standardized effect size phylogenetic diversity (sesPD) and **(D)** time integrated lineage diversity (TILD), are best described by quadratic regression models. Points are mean rarefied metric values for each plot. Red lines represent regression lines from the best fitting model for each metric. For figures a and b, the solid vertical line indicates the estimated breakpoint, with respective standard error represented by vertical dashed lines.

up to an estimated breakpoint at 1,410 m asl (± 129m), and then decreases with elevation (**Figure 2B**). The pattern of sesPD across elevation (**Figure 2C**) is best described by a quadratic regression model (adjusted $R^2 = 0.75$). TILD follows a hump-shaped pattern across elevation (**Figure 2D**), with a quadratic regression giving the best fit to the data (adjusted $R^2 = 0.77$). These results are not notably different when considering only angiosperm lineages, though the mid-elevation peak for evolutionary diversity metrics is less pronounced.

Evolutionary Age Patterns

Only extant lineages are included in our phylogeny. Therefore, the number of lineages at different evolutionary depths necessarily increases from two 382 million years old lineages (1 lineage leading to tree ferns, 1 lineage leading to seed plants), to 275 modern lineages, equal to the total number of genera in plots that are present in our phylogeny (**Figure 1**). The pattern of elevational range distribution of lineages and peak in number of lineages varies at different evolutionary

TABLE 2 | Summary of lineage elevational range trends and richness at six illustrative evolutionary depths.

Evolutionary depth (myrs)	Total no. lineages	No. lineages only <1,750 m asl	No. lineages only >1,750 m asl	Elevation of peak lineage richness (m asl)	No. lineages at richness peak
10	238	125	17	1,300–1,500	117
70	101	33	9	1,300–1,500	65
110	30	4	1	1,300–1,500	27
120	16	1	1	1,300–1,900	13
130	8	0	1	1,500–2,000	7
200	3	0	0	1,300–3,000	3

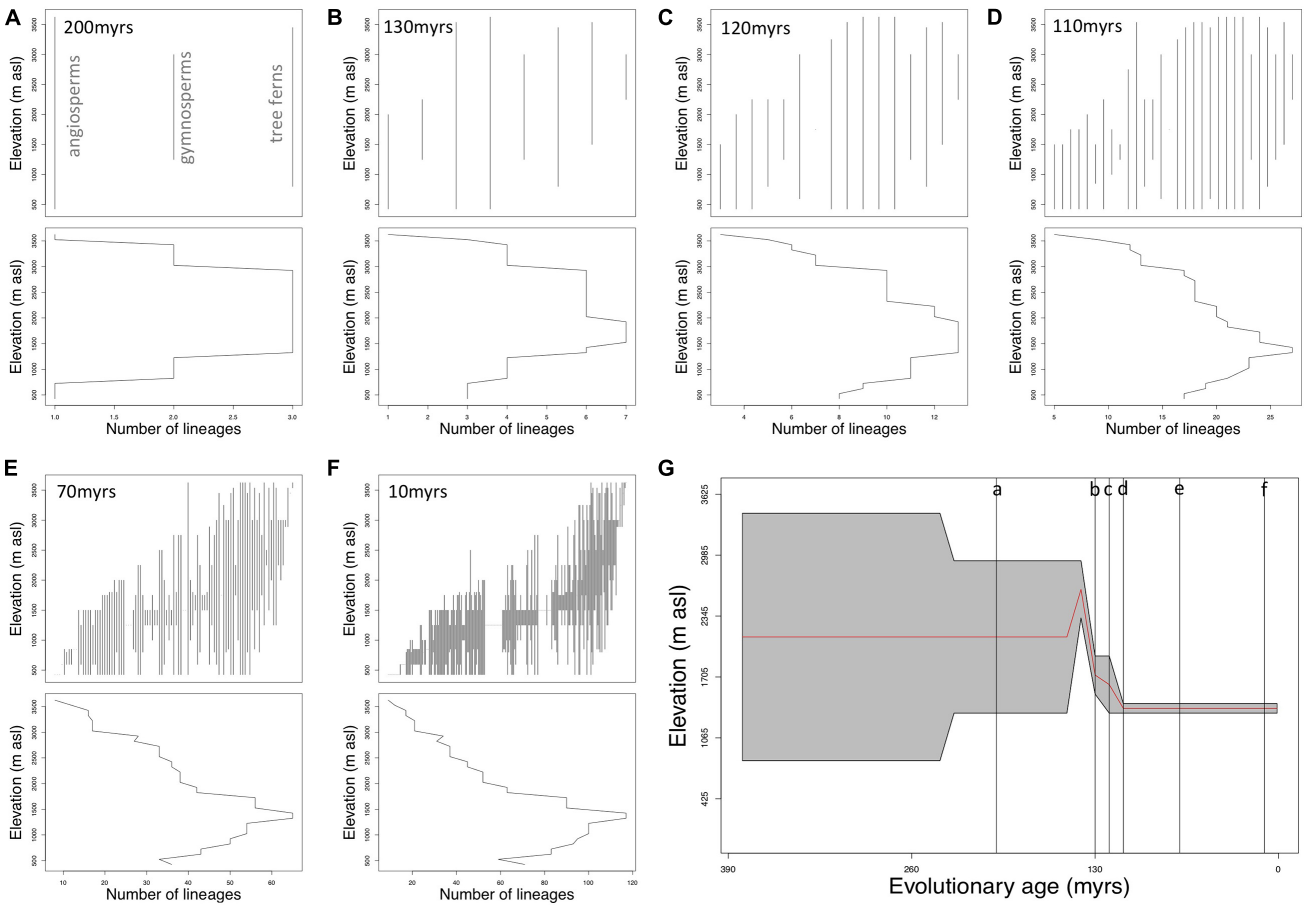


FIGURE 3 | (**A–F)** Illustrate how evolutionary age structure varies across elevation and at six notable evolutionary depths **(A)** 200 myrs, **(B)** 130 myrs, **(C)** 120 myrs, **(D)** 110 myrs, **(E)** 70 myrs, and **(F)** 10 myrs. Within each panel, the upper plot illustrates the elevational ranges of lineages. Vertical lines represent each individual lineage present at the respective evolutionary depth. Lower plots represent the number of lineages (x-axis) across elevation (y-axis) at each of the size evolutionary depths. **(G)** Illustrates how the elevations at which lineage richness is most numerous changes across different evolutionary depths. The gray shaded area represents the elevations most numerous in lineages across evolutionary depths. The red line illustrates mean elevation of the peak number of lineages. Vertical gray lines in **(G)** with letters relate to **(A–F)**.

depths (**Table 2** and **Figure 3**). The highest number of younger lineages is found at lower elevations, with many younger lineages restricted to lower elevations. Older lineages are more prevalent at mid-elevations. Fewer older lineages are restricted to lower elevations and a greater percentage extend to higher elevations.

Six illustrative evolutionary time slices are presented in **Table 2** and **Figure 3**. These time slices were chosen to capture the main points of variation across the evolutionary depth of our phylogeny. For example, at an evolutionary depth of 10 myrs (**Figure 3F**), there are 238 lineages that presently occur along the elevation transect. One-hundred and twenty-five of those lineages do not presently occur above 1,750 m asl, the mid-point of our cloud-base ecotone estimate. Seventeen lineages at 10 myrs evolutionary depth only occur above 1,750 m asl. The peak in the number of 10 myr old lineages

(n = 117) is between 1,300 and 1,500 m asl (**Figures 3F,G**). Data for each 10-million-year time slices is presented in **Supplementary Material**.

DISCUSSION

Our results demonstrate a discernible contrast between patterns of taxonomic and phylogenetic diversity of trees along an Amazon-to-Andes elevation gradient. Average genus richness is essentially invariant with elevation below 1,383 m asl (± 294 m) but decreases with elevation above this point. Evolutionary richness metrics peak at mid-elevations, a pattern driven by a greater number of older lineages at mid to high elevations, and many more-recently diverged lineages failing to occur above 1,500–1,750 m asl. These findings do not match the linear expectations of the TNC or Out of the Tropical Lowlands (OTL) hypotheses. The observed prevalence of older lineages at higher elevations is anticipated by the OTL hypothesis, but our findings are most closely aligned with the Environmental Crossroads hypothesis (EC; Neves et al., 2020), which predicts the highest levels of evolutionary diversity at intermediate environmental conditions.

The observed trend in genus richness—a relatively invariant pattern up to mid-elevations and then steep diversity decline above that (**Figure 2A**)—matches previous species-level trends reported for this gradient (Jankowski et al., 2013), and is consistent with patterns in neo TMF more broadly (Gentry, 1988; Vazquez and Givnish, 1998), though these reports have sometimes been incorrectly interpreted as linear decreases. Many factors, such as richness-area effects (MacArthur and Wilson, 1963; Rosenzweig, 1995), the association between environmental heterogeneity and richness (Kreft and Jetz, 2007; Antonelli et al., 2018), and a non-linear precipitation gradient (Rapp and Silman, 2012), are likely to modulate the observed equivalence of low elevation and mid-elevation taxonomic richness; additional research will be needed to tease apart the contributions of these different factors.

Contrasting the trend in taxonomic richness, the evolutionary diversity metrics PD and TILD both show increasing diversity up to mid-elevations, followed by a subsequent decrease toward higher elevations. A similar peak in evolutionary diversity at mid-elevations has recently been reported in Tropical Andean forests in Colombia (González-Caro et al., 2020). In our data, PD displays a distinct mid elevation peak between c.1,300–1,500 m asl (**Figure 2B**); TILD follows a more gradual hump-shaped pattern across elevation, with the highest values observed across a broad range of intermediate elevations (**Figure 2D**). TILD values only drop markedly at the extremes of the elevational gradient. The elevational pattern for sesPD is a non-linear increase with elevation with a decrease at the highest elevations sampled (**Figure 2C**). However, it has been suggested that variation in standardized phylogenetic diversity metrics may be an artifact of variation in taxonomic richness (Sandel, 2018) and that sesPD may be a poor estimator of the richness dimension of diversity (Tucker et al., 2017; Dexter et al., 2019). These results must be put in the context of an analysis focused on trees. Given some lineages encompass multiple forms, studies incorporating life forms, such as lianas or shrubs, may find different patterns.

The nuanced variation observed between PD and TILD is likely to be associated with the way each metric weights the different evolutionary depths. TILD gives equal weight to diversity across all the evolutionary depths of a phylogeny, while PD gives greater weight to more-recently diverged lineages (Dexter et al., 2019). Basically, the contrast in results between TILD and PD show that older evolutionary lineages (picked up by TILD, but not represented well by PD) are present across a broad range of elevations and only really drop out at the extremes of the gradient.

Our exploration of patterns in evolutionary age of tree lineages across elevation reveals further trends. The tendency for older lineages to be found at mid and high elevations, and younger lineages to be found at lower elevations, is demonstrated by the changing number, and distribution pattern, of lineages at different evolutionary depths (**Figures 3A–F**). The elevational ranges of lineages at younger evolutionary depths (**Figures 3E,F**) indicate that for example, 53% of the lineages (n = 125) that are 10 million years old are not found above 1,750 m asl, and 33% of lineages (n = 33) that are 70 million years old, are not found above 1,750 m asl. This pattern in many younger lineages could reflect a lowland tropical origin and failure to adapt to environmental conditions at higher elevations. At older evolutionary depths, though older lineages do occur at lower elevations, fewer are restricted there and a greater percentage extend to higher elevations. For example, just 13% (n = 4) of lineages that are 110 million years old are restricted below ∼1,750 m asl, and no lineages older than 130 million years are restricted to a distribution below 1,750 m asl. Between evolutionary depths of 110 and 130 million years, there is a discernible shift in the elevation with the peak number of lineages. For lineages that are 130 million years old, the highest number of lineages occurs between 1,500 and 2,000 m asl, but for lineages that are 110 million years old the elevation where the highest number of lineages occur is lower, between 1,300 and 1,500 m asl (**Figure 3G**).

The prevalence of older lineages at mid and high elevations, combined with the restriction of many young lineages to lower elevations, is consistent both with expectations of the OTL hypothesis (Qian and Ricklefs, 2016) *and* previous findings suggesting a positive association between elevation and lineage age, as quantified by mean family ages (Segovia and Armesto, 2015; Tiede et al., 2015; Qian and Ricklefs, 2016). However, we suggest that our analysis based on phylogenetic depth may be more evolutionarily meaningful than previously utilized family ages, which are based on relatively subjective taxonomic circumscription of families.

Integrating lineage age into our understanding of the EDG clearly demonstrates the biodiversity significance of mid-elevation TMF. However, variation in evolutionary age trends across elevation is not, in and of itself the driver of the diversity trend. The observation of highest evolutionary diversity in mid-elevation TMF may be explained by the meeting of distinct evolutionary floras at the intermediate region of an

environmental gradient as posited by the EC hypothesis (Neves et al., 2020). This explanation is supported by the occurrence of a tropical montane flora that is evolutionarily distinct to that found in the lowlands (Segovia and Armesto, 2015; Griffiths et al., 2021) as has also been suggested for Andean forests in Colombia (González-Caro et al., 2020). Further, TMF appear to comprise a unique tree flora that is also evolutionarily distinct from the more temperate affiliated tree flora found at the highest Andean elevations (Segovia et al., 2020).

Along this Amazon-to-Andes gradient, an environmental crossroads type process may also intersect with the presence of the cloud-base ecotone at mid-elevations around 1,500–2,000 m asl (Halladay et al., 2012; Rapp and Silman, 2012; Fadrique et al., 2018). Ecotonal transitions may represent areas of high diversity (Odum, 1971; Risser, 1995; Kernaghan and Harper, 2001; Silva-Pereira et al., 2020) with the possible influence of mass effects, where individuals are able to occur in marginal habitat on the edge of their species' core range (Shmida and Wilson, 1985). Moreover, ecotones may actually act as generators of diversity (Moritz et al., 2000). Although this has yet to be explicitly tested, the cloud-base ecotone, with its environmental transitions, could represent a "leaky barrier" influencing the mixing of representatives from both the tropical montane flora above, and the lowland flora below, and providing a boost to evolutionary diversity.

While the cloud-base ecotone may play some part in the shape of the TMF EDG, other processes may be of stronger and broader influence. Historical processes, such as Andean uplift are recognized as of broad influence on the biogeographic patterns of the South American flora (Antonelli et al., 2009). The evolutionary ages of many lineages at higher elevations substantially exceeds the relatively recent uplift of the Andes (Graham, 2009) and fits suggestions that colonization by older lineages, preadapted to extratropical environments, likely played an important role in the assembly of TMF communities (Gentry, 1982; Segovia and Armesto, 2015). Mid-elevation TMF may represent the region where these colonizing lineages mix with those taxa from lowland Amazonian communities that adapted to a cooler climate during the Pleistocene, and reach their elevational limit around 1,700 m asl (Silman, 2007).

The shape of the EDG for trees clearly depends upon the dimension of diversity that is quantified. Our results demonstrate clear contrasts between taxonomic and evolutionary patterns of alpha diversity, as well as variation across different evolutionary depths. The high values of evolutionary diversity at mid-elevations support the Environmental Crossroads hypothesis (Neves et al., 2020) more strongly than competing hypotheses, and we argue that our observations support the hypothesized mixing of evolutionarily distinct tropical montane and tropical lowland floras at mid-elevations. Given the potential impact of environmental changes, such as a shift in cloud-base dynamics under climate change (Helmer et al., 2019), broader acknowledgment of this globally unique diversity is urgent. Lowland Amazonia has historically been considered as containing the most diverse forests on the planet (Henderson et al., 1991), yet as revealed here, lowland richness is rivaled by that of mid-elevation montane forests, especially when factoring in the full evolutionary depth of lineages.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data is provided in **Supplementary Material** for Griffiths et al. (2021), Biotropica. Including GenBank accession numbers.

AUTHOR CONTRIBUTIONS

AG and KD conceived the manuscript. MS, WF-R, and PM designed and established the plot network. KF, WF-R, KC, and NS contributed to data collection and taxonomic identification. AG, KD, and RS conducted the analyses. Writing of the manuscript was led by AG with input from all co-authors. All authors contributed to the article and approved the submitted version.

FUNDING

AG was supported by the NERC-E3DTP studentship (grant no. NERC NE/L002558/1) and a grant from the Moss family. PM was supported by NERC (grant no. NE/G018278/1) and ARC (grant no. DP170104091). KF was supported by the US National Science Foundation (grant no. NSF DEB LTREB 1754664). MS was supported by NSF DEB LTREB (grant no. 1754647). Andes Biodiversity and Ecosystem Research Group (ABERG) tree plots were funded by the Gordon and Betty Moore Foundation Andes-Amazon Program, ForestPlots, and US National Science Foundation LTREB grant no. 1754647.

ACKNOWLEDGMENTS

This manuscript is a product of the Andes Biodiversity and Ecosystem Research Group (ABERG). We thank the administration and staff of Manu National Park, the Peruvian Protected Areas Service (SERNANP) for logistical support and permission to carry out the work. Special thanks to the hundreds of young Peruvian scientists who have given body and soul to carry out the field work in this forbidding but essential environment. In addition we thank the two reviewers for their contributions toward improving this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.680041/full#supplementary-material

REFERENCES

Antonelli, A., Kissling, W. D., Flantua, S. G. A., Bermúdez, M. A., Mulch, A., Muellner-Riehl, A. N., et al. (2018). Geological and climatic influences on mountain biodiversity. *Nat. Geosci.* 11, 718–725. doi: 10.1038/s41561-018-0236-z

Antonelli, A., Nylander, J. A. A., Persson, C., and Sanmartin, I. (2009). Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proc. Natl. Acad Sci.* 106, 9749–9754. doi: 10.1073/pnas.0811421106

Bacon, C. D., Mora, A., Wagner, W. L., and Jaramillo, C. A. (2013). Testing geological models of evolution of the Isthmus of Panama in a phylogenetic framework. *Bot. J. Linn. Soc.* 171, 287–300. doi: 10.1111/j.1095-8339.2012.01281.x

Bergamin, R. S., Seger, G. D. S., Carlucci, M. B., Molz, M., Mello, R. S. P., Martins, R., et al. (2020). Elevational shifts in phylogenetic diversity of angiosperm trees across the subtropical Brazilian Atlantic Forest. *Austral Ecol.* 46:12996. doi: 10.1111/aec.12996

Bruijnzeel, L. A. (2001). Hydrology of tropical montane cloud forests: A Reassessment. *Land Use Water Resour. Res.* 1:1.

Bruijnzeel, L. A., Mulligan, M., and Scatena, F. N. (2011). Hydrometeorology of tropical montane cloud forests: Emerging patterns. *Hydrol. Process.* 25, 465–498. doi: 10.1002/hyp.7974

Colwell, R. K., Chao, A., Gotelli, N. J., Lin, S.-Y., Mao, C. X., Chazdon, R. L., et al. (2012). Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *J. Plant Ecol.* 5, 3–21. doi: 10.1093/jpe/rtr044

Crane, P. R., Friis, E. M., and Pedersen, K. R. (1995). The origin and early diversification of angiosperms. *Nature* 374, 27–33. doi: 10.1038/374027a0

Currie, D. J. (1991). Energy and large-scale patterns of animal- and plant-species richness. *Am. Nat.* 137, 27–49. doi: 10.1086/285144

Currie, D. J., Mittelbach, G. G., Cornell, H. V., Field, R., Guégan, J.-F., Hawkins, B. A., et al. (2004). Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecol. Lett.* 7, 1121–1134. doi: 10.1111/j.1461-0248.2004.00671.x

Dexter, K. G., Segovia, R. A., and Griffiths, A. R. (2019). Exploring the concept of lineage diversity across north american forests. *Forests* 10:520. doi: 10.3390/f10060520

Fadrique, B., Báez, S., Duque, A., Malizia, A., Blundo, C., Carilla, J., et al. (2018). Widespread but heterogeneous responses of Andean forests to climate change. *Nature* 564, 207–212. doi: 10.1038/s41586-018-0715-9

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* 61, 1–10. doi: 10.1016/0006-3207(92)91201-3

Fyllas, N. M., Bentley, L. P., Shenkin, A., Asner, G. P., Atkin, O. K., Díaz, S., et al. (2017). Solar radiation and functional traits explain the decline of forest primary productivity along a tropical elevation gradient. *Ecol. Lett.* 20, 730–740. doi: 10.1111/ele.12771

Gentry, A. (1982). Neotropical floristic diversity: phytogeographical connections between central and south america. pleistocene climatic fluctuations, or an accident of the andean orogeny? *Ann. Mo. Bot. Gard.* 69, 557–593. doi: 10.2307/2399084

Gentry, A. H. (1988). Changes in plant community diversity and floristic composition on environmental and geographical gradients. *Ann. Mo. Bot. Gard* 75, 1–34. doi: 10.2307/2399464

Girardin, C. A. J., Farfan-Rios, W., Garcia, K., Feeley, K. J., Jørgensen, P. M., Murakami, A. A., et al. (2014). Spatial patterns of above-ground structure, biomass and composition in a network of six Andean elevation transects. *Plant Ecol. Divers.* 7, 161–171. doi: 10.1080/17550874.2013.820806

Girardin, C. A. J., Malhi, Y., Aragão, L. E. O. C., Mamani, M., Huaraca Huasco, W., Durand, L., et al. (2010). Net primary productivity allocation and cycling of carbon along a tropical forest elevational transect in the Peruvian Andes. *Glob Change Biol.* 16, 3176–3192. doi: 10.1111/j.1365-2486.2010.02235.x

González-Caro, S., Duque, A., Feeley, K. J., Cabrera, E., Phillips, J., Ramirez, S., et al. (2020). The legacy of biogeographic history on the composition and structure of Andean forests. *Ecology* 101:e03131. doi: 10.1002/ecy.3131

Graham, A. (2009). The Andes: A Geological Overview from a Biological Perspective. *Ann. Mo. Bot. Gard.* 96, 371–385. doi: 10.3417/2007146

Griffiths, A. R., Silman, M. R., Farfán Rios, W., Feeley, K. J., García Cabrera, K., Meir, P., et al. (2021). Evolutionary heritage shapes tree distributions along an Amazon—to—Andes elevation gradient. *Biotropica* 53, 38–50. doi: 10.1111/btp.12843

Guo, Q., Kelt, D. A., Sun, Z., Liu, H., Hu, L., Ren, H., et al. (2013). Global variation in elevational diversity patterns. *Sci. Rep.* 3:3007. doi: 10.1038/srep03007

Halladay, K., Malhi, Y., and New, M. (2012). Cloud frequency climatology at the Andes/Amazon transition: 2. Trends and variability. *J. Geophys. Res. Atmospheres* 117:D23103. doi: 10.1029/2012JD017789

Hawkins, B. A., Rodriguez, M. A., and Weller, S. G. (2011). Global angiosperm family richness revisited: Linking ecology and evolution to climate. *J. Biogeogr.* 38, 1253–1266. doi: 10.1111/j.1365-2699.2011.02490.x

Helmer, E. H., Gerson, E. A., Baggett, L. S., Bird, B. J., Ruzycki, T. S., and Voggesser, S. M. (2019). Neotropical cloud forests and páramo to contract and dry from declines in cloud immersion and frost. *PLoS One* 14:e0213155. doi: 10.1371/journal.pone.0213155

Henderson, A., Churchill, S. P., and Luteyn, J. L. (1991). Neotropical plant diversity. *Nature* 351:21.

Holt, R. D. (2009). Bringing the hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. *Proc. Nat. Acad. Sci.* 106, 19659–19665. doi: 10.1073/pnas.0905137106

Humboldt, A. V., and Bonpland, A. (1805). *Essai sur la géographie des plantes*. Whitefish, MT: Kessinger Publishing

Jablonski, D., Roy, K., and Valentine, J. W. (2006). Out of the tropics: evolutionary dynamics of the latitudinal diversity gradient. *Science* 314, 102–106. doi: 10.1126/science.1130880

Jankowski, J. E., Merkord, C. L., Rios, W. F., Cabrera, K. G., Revilla, N. S., and Silman, M. R. (2013). The relationship of tropical bird communities to tree species composition and vegetation structure along an Andean elevational gradient. *J. Biogeogr.* 40, 950–962. doi: 10.1111/jbi.12041

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., et al. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26, 1463–1464. doi: 10.1093/bioinformatics/btq166

Kerkhoff, A. J., Moriarty, P. E., and Weiser, M. D. (2014). The latitudinal species richness gradient in New World woody angiosperms is consistent with the tropical conservatism hypothesis. *Proc. Nat. Acad. Sci.* 111, 8125–8130. doi: 10.1073/pnas.1308932111

Kernaghan, G., and Harper, K. A. (2001). Community structure of ectomycorrhizal fungi across an alpine/subalpine ecotone. *Ecography* 24, 181–188. doi: 10.1034/j.1600-0587.2001.240208.x

Kessler, M. (2000). Elevational gradients in species richness and endemism of selected plant groups in the central Bolivian Andes. *Plant Ecol.* 149, 181–193. doi: 10.1023/A:1026500710274

Kreft, H., and Jetz, W. (2007). Global patterns and determinants of vascular plant diversity. *Proc. Nat. Acad. Sci.* 104, 5925–5930. doi: 10.1073/pnas.0608361104

Lieberman, D., Lieberman, M., Peralta, R., and Hartshorn, G. S. (1996). Tropical forest structure and composition on a large-scale altitudinal gradient in costa rica. *J. Ecol.* 84, 137–152. doi: 10.2307/2261350

MacArthur, R. H., and Wilson, E. O. (1963). An equilibrium theory of insular zoogeography. *Evolution* 17:373. doi: 10.2307/2407089

Malhi, Y., Girardin, C. A. J., Goldsmith, G. R., Doughty, C. E., Salinas, N., Metcalfe, D. B., et al. (2017). The variation of productivity and its allocation along a tropical elevation gradient: A whole carbon budget perspective. *New Phytol.* 214, 1019–1032. doi: 10.1111/nph.14189

McCain, C. M., and Grytnes, J.-A. (2010). *Elevational Gradients in Species Richness*. (Hoboken.NJ: John Wiley & Sons, Ltd (Ed.)), doi: 10.1002/9780470015902.a0022548

McFadden, I. R., Sandel, B., Tsirogiannis, C., Morueta Holme, N., Svenning, J., Enquist, B. J., et al. (2019). Temperature shapes opposing latitudinal gradients of plant taxonomic and phylogenetic β diversity. *Ecol. Lett.* 22, 1126–1135. doi: 10.1111/ele.13269

Moritz, C., Patton, J. L., Schneider, C. J., and Smith, T. B. (2000). Diversification of rainforest faunas: an integrated molecular approach. *Ann. Rev. Ecol. Syst.* 31, 533–563. doi: 10.1146/annurev.ecolsys.31.1.533

Muggeo, V. M. (2008). Segmented: An R package to fit regression models with broken-line relationships. *R News* 8, 20–25.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. B., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature* 403:853. doi: 10.1038/35002501

Neves, D. M., Dexter, K. G., Baker, T. R., Coelho de Souza, F., Oliveira-Filho, A. T., Queiroz, L. P., et al. (2020). Evolutionary diversity in tropical tree communities peaks at intermediate precipitation. *Sci. Rep.* 10:1188. doi: 10.1038/s41598-019-55621-w

Nottingham, A. T., Fierer, N., Turner, B. L., Whitaker, J., Ostle, N. J., McNamara, N. P., et al. (2018). Microbes follow Humboldt: Temperature drives plant and soil microbial diversity patterns from the Amazon to the Andes. *Ecology* 99, 2455–2466. doi: 10.1002/ecy.2482

Nottingham, A. T., Whitaker, J., Turner, B. L., Salinas, N., Zimmermann, M., Malhi, Y., et al. (2015). Climate warming and soil carbon in tropical forests: insights from an elevation gradient in the peruvian andes. *BioScience* 65, 906–921. doi: 10.1093/biosci/biv109

Odum, E. P. (1971). *Fundamentals of ecology* (3d ed). Belmont, CA: Thomson Brooks/Cole.

Oksanen, J., Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2018). *Vegan: Community Ecology Package. R package version 2.5-2*, 2018.

Perez, T. M., Stroud, J. T., and Feeley, K. J. (2016). Thermal trouble in the tropics. *Science* 351, 1392–1393. doi: 10.1126/science.aaf3343

Qian, H., Hao, Z., and Zhang, J. (2014). Phylogenetic structure and phylogenetic diversity of angiosperm assemblages in forests along an elevational gradient in Changbaishan. *China. J. Plant Ecol* 7, 154–165. doi: 10.1093/jpe/rtt072

Qian, H., and Ricklefs, R. E. (2016). Out of the tropical lowlands: latitude versus elevation. *Trends Ecol Evol* 31, 738–741. doi: 10.1016/j.tree.2016.07.012

R Core Team (2018). *R: A language and environment for statistical computing*. Vienna: R Core Team.

Rahbek, C. (1995). The elevational gradient of species richness: A uniform pattern? *Ecography* 18, 200–205. doi: 10.1111/j.1600-0587.1995.tb00341.x

Rapp, J., and Silman, M. (2012). Diurnal, seasonal, and altitudinal trends in microclimate across a tropical montane cloud forest. *Clim. Res.* 55, 17–32. doi: 10.3354/cr01127

Rezende, V. L., Dexter, K. G., Pennington, R. T., and Oliveira-Filho, A. T. (2017). Geographical variation in the evolutionary diversity of tree communities across southern South America. *J. Biogeogr.* 44, 2365–2375. doi: 10.1111/jbi.13013

Risser, P. G. (1995). The Status of the Science Examining Ecotones. *BioScience* 45, 318–325. doi: 10.2307/1312492

Romdal, T. S., Araújo, M. B., and Rahbek, C. (2013). Life on a tropical planet: Niche conservatism and the global diversity gradient. *Glob. Ecol. Biogeogr.* 22, 344–350. doi: 10.1111/j.1466-8238.2012.00786.x

Rosenzweig, M. L. (1995). *Species Diversity in Space and Time*. Cambridge, MA: Cambridge University Press.

Sandel, B. (2018). Richness-dependence of phylogenetic diversity indices. *Ecography* 41, 837–844. doi: 10.1111/ecog.02967

Segovia, R. A., and Armesto, J. J. (2015). The gondwanan legacy in south american biogeography. *J. Biogeogr.* 42, 209–217. doi: 10.1111/jbi.12459

Segovia, R. A., Pennington, R. T., Baker, T. R., Coelho de Souza, F., Neves, D. M., Davis, C. C., et al. (2020). Freezing and water availability structure the evolutionary diversity of trees across the Americas. *Sci. Adv.* 6:eaaz5373. doi: 10.1126/sciadv.aaz5373

Shmida, A., and Wilson, M. V. (1985). Biological determinants of species diversity. *J. Biogeogr.* 12, 1–20. doi: 10.2307/2845026

Silman, M. R. (2007). "Plant species diversity in Amazonian forests," in *Tropical Rainforest Responses to Climatic Change*, eds M. B. Bush and J. R. Flenley (Berlin: Springer), 269–294. doi: 10.1007/978-3-540-48842-2_10

Silva-Pereira, I., Meira-Neto, J. A. A., Rezende, V. L., and Eisenlohr, P. V. (2020). Biogeographic transitions as a source of high biological diversity: Phylogenetic lessons from a comprehensive ecotone of South America. *Perspect. Plant Ecol. Evol. Syst.* 44:125528. doi: 10.1016/j.ppees.2020.125528

Silvestro, D., Cascales-Miñana, B., Bacon, C. D., and Antonelli, A. (2015). Revisiting the origin and diversification of vascular plants through a comprehensive Bayesian analysis of the fossil record. *New Phytol.* 207, 425–436. doi: 10.1111/nph.13247

Smith, B. T., Amei, A., and Klicka, J. (2012). Evaluating the role of contracting and expanding rainforest in initiating cycles of speciation across the Isthmus of Panama. *Proc. Biol Sci.* 279, 3520–3526. doi: 10.1098/rspb.2012.0706

Sundqvist, M. K., Sanders, N. J., and Wardle, D. A. (2013). Community and ecosystem responses to elevational gradients: processes. *mechanisms, and insights for global change. Ann. Rev. Ecol. Evol. Syst.* 44, 261–280. doi: 10.1146/annurev-ecolsys-110512-135750

Tiede, Y., Homeier, J., Cumbicus, N., Peña, J., Albrecht, J., Ziegenhagen, B., et al. (2015). Phylogenetic niche conservatism does not explain elevational patterns of species richness, phylodiversity and family age of tree assemblages in Andean rainforest. *Erdkunde* 70, 83–106. doi: 10.3112/erdkunde.2016.01.06

Tucker, C. M., Cadotte, M. W., Carvalho, S. B., Davies, T. J., Ferrier, S., Fritz, S. A., et al. (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology: A guide to phylogenetic metrics for ecology. *Biol. Rev.* 92, 698–715. doi: 10.1111/brv.12252

Vazquez, J. A., and Givnish, T. J. (1998). Altitudinal gradients in tropical forest composition. *structure, and diversity in the sierra de manantlan. J. Ecol.* 86, 999–1020. doi: 10.1046/j.1365-2745.1998.00325.x

Wiens, J. J., and Donoghue, M. J. (2004). Historical biogeography, ecology and species richness. *Trends Ecol. Evol.* 19, 639–644. doi: 10.1016/j.tree.2004.09.011

Yguel, B., Jactel, H., Pearse, I. S., Moen, D., Winter, M., Hortal, J., et al. (2016). The evolutionary legacy of diversification predicts ecosystem function. *Am. Nat.* 188, 398–410. doi: 10.1086/687964

Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. A., FitzJohn, R. G., et al. (2014). Three keys to the radiation of angiosperms into freezing environments. *Nature* 506, 89–92. doi: 10.1038/nature12872

Zimmermann, M., Meir, P., Silman, M. R., Fedders, A., Gibbon, A., Malhi, Y., et al. (2010). No differences in soil carbon stocks across the tree line in the peruvian andes. *Ecosystems* 13, 62–74. doi: 10.1007/s10021-009-9300-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 04 October 2021
doi: 10.3389/fevo.2021.719741

Evolution of Dispersal, Habit, and Pollination in Africa Pushed Apocynaceae Diversification After the Eocene-Oligocene Climate Transition

Cássia Bitencourt[1†‡], Nicolai M. Nürk[2†], Alessandro Rapini[1†], Mark Fishbein[3†], André O. Simões[4†], David J. Middleton[5†], Ulrich Meve[2†], Mary E. Endress[6] and Sigrid Liede-Schumann[2*†]*

[1] Programa de Pós-Graduação em Botânica, Departamento de Ciências Biológicas, Universidade Estadual de Feira de Santana, Feira de Santana, Brazil, [2] Department of Plant Systematics, Bayreuth Center of Ecology and Environmental Research, BayCEER, University of Bayreuth, Bayreuth, Germany, [3] Department of Plant Biology, Ecology, and Evolution, Oklahoma State University, Stillwater, OK, United States, [4] Programa de Pós-Graduação em Biologia Vegetal, Departamento de Biologia Vegetal, Universidade Estadual de Campinas, São Paulo, Brazil, [5] Singapore Botanic Gardens, National Parks Board, Singapore, Singapore, [6] Department of Systematic and Evolutionary Botany, University of Zurich, Zurich, Switzerland

Apocynaceae (the dogbane and milkweed family) is one of the ten largest flowering plant families, with approximately 5,350 species and diverse morphology and ecology, ranging from large trees and lianas that are emblematic of tropical rainforests, to herbs in temperate grasslands, to succulents in dry, open landscapes, and to vines in a wide variety of habitats. Despite a specialized and conservative basic floral architecture, Apocynaceae are hyperdiverse in flower size, corolla shape, and especially derived floral morphological features. These are mainly associated with the development of corolline and/or staminal coronas and a spectrum of integration of floral structures culminating with the formation of a gynostegium and pollinaria—specialized pollen dispersal units. To date, no detailed analysis has been conducted to estimate the origin and diversification of this lineage in space and time. Here, we use the most comprehensive time-calibrated phylogeny of Apocynaceae, which includes approximately 20% of the species covering all major lineages, and information on species number and distributions obtained from the most up-to-date monograph of the family to investigate the biogeographical history of the lineage and its diversification dynamics. South America, Africa, and Southeast Asia (potentially including Oceania), were recovered as the most likely ancestral area of extant Apocynaceae diversity; this tropical climatic belt in the equatorial region retained the oldest extant lineages and these three tropical regions likely represent museums of the family. Africa was confirmed as the cradle of pollinia-bearing lineages and the main source of Apocynaceae intercontinental dispersals. We detected 12 shifts toward accelerated species diversification, of which 11 were in the APSA clade (apocynoids, Periplocoideae, Secamonoideae, and Asclepiadoideae), eight of these in

OPEN ACCESS

Edited by:
Danilo M. Neves, Federal University of Minas Gerais, Brazil

Reviewed by:
Edeline Gagnon, Royal Botanic Garden Edinburgh, United Kingdom
Veronica Aydos Thode, University of São Paulo, Brazil

***Correspondence:**
Alessandro Rapini
rapinibot@yahoo.com.br
Sigrid Liede-Schumann
sigrid.liede@uni-bayreuth.de

†ORCID:
Cássia Bitencourt
orcid.org/0000-0001-9141-2323
Nicolai M. Nürk
orcid.org/0000-0002-0471-644X
Alessandro Rapini
orcid.org/0000-0002-8758-9326
Mark Fishbein
orcid.org/0000-0003-3099-4387
André O. Simões
orcid.org/0000-0001-6555-8759
David J. Middleton
orcid.org/0000-0003-3754-1452
Ulrich Meve
orcid.org/0000-0001-7395-5199
Sigrid Liede-Schumann
orcid.org/0000-0003-2707-0335

‡Present address:
Cássia Bitencourt, Royal Botanic Gardens, Kew, Richmond, United Kingdom

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *03 June 2021*
Accepted: *27 August 2021*
Published: *04 October 2021*

the pollinia-bearing lineages and six within Asclepiadoideae. Wind-dispersed comose seeds, climbing growth form, and pollinia appeared sequentially within the APSA clade and probably work synergistically in the occupation of drier and cooler habitats. Overall, we hypothesize that temporal patterns in diversification of Apocynaceae was mainly shaped by a sequence of morphological innovations that conferred higher capacity to disperse and establish in seasonal, unstable, and open habitats, which have expanded since the Eocene-Oligocene climate transition.

Keywords: apocynoids, APSA clade, Asclepiadoideae, biogeography, Gondwana, Laurasia, long-distance dispersal, rauvolfioids

INTRODUCTION

The early diversification of flowering plants (angiosperms), a clade comprising approximately 90% of extant land plant species (Hernández-Hernández and Wiens, 2020), was marked by a rapid increase in lineage diversity, abundance, and distribution during the Late Cretaceous, 100–66 million years ago (Ma; Friis et al., 2011). More than half of angiosperm families diverged between 100 and 90 Ma (stem node ages) and diversified in the Cenozoic (66–2.6 Ma; crown node ages; Ramírez-Barahona et al., 2020). The abrupt collapse of ecosystems caused by the Cretaceous-Paleogene (K-Pg) mass extinction at ca. 65 Ma catalyzed drastic floristic changes and extirpated up to 57% of plant species, mostly affecting lineages dispersed and pollinated by animals (McElwain and Punyasena, 2007). The recovery of plant diversity in the early Paleocene coincided with the establishment and diversification of modern tropical rainforests (e.g., Morley, 2000; Wang et al., 2012; Meseguer et al., 2018) and, in the early Eocene climatic optimum (52–50 Ma), rainforests reached higher latitudes and formed a relatively continuous boreotropical flora (Wolfe, 1978). The global cooling and increased aridification that started mainly from the Eocene-Oligocene climatic transition (34 Ma) and intensified after the mid-Miocene climatic optimum (15 Ma), however, fostered the establishment and expansion of modern cold and dry biomes and narrowed rainforests to lower latitudes, which resulted in a collapse of intercontinental routes suitable for tropical plants (i.e., megathermal plants; Folk et al., 2020). While, in general, the recent temperate habitats show higher rates of speciation and turnover, older tropical rainforests have been less affected by extinction, show higher rates of net diversification (Igea and Tanentzap, 2020), and harbor more ancient families of angiosperms (Ramírez-Barahona et al., 2020). Although the accumulation of extant species increased mainly since the Pliocene (5 Ma) and, more intensively, during the Quaternary (2.6 Ma), the impact of Pleistocene climatic oscillations on the diversification of angiosperms, in particular within highly diverse tropical regions, such as the lowland Amazonian rainforests and campos rupestres highlands, remains debatable (Rapini et al., 2021).

Exceptionally diverse lineages are scattered across the angiosperm tree of life, but are also often concentrated in particular clades, such as the core asterids, which comprises six of the ten extremely species-rich orders of angiosperms (Magallón and Sanderson, 2001). Although a great variety of possible drivers of species diversification have been discussed, including, among others, biotic interactions, morphological key innovations, and whole-genome duplications (e.g., Soltis et al., 2019), shifts in diversification rates often depend on a complex and intricate net of interactions between intrinsic biological traits and extrinsic (biotic and abiotic) conditions (Davies et al., 2004; Vamosi and Vamosi, 2011; Bouchenak-Khelladi et al., 2015; Donoghue and Sanderson, 2015; Magallón et al., 2019; Nürk et al., 2020). Geographic range size is often positively associated with species diversity and was recently indicated as the trait that most explains overall diversification rates of orders and families within angiosperms (Hernández-Hernández and Wiens, 2020). Theory predicts that widely distributed lineages are less prone to extinction and more susceptible to fragmentation and ecological specialization, favoring higher diversification rates. In turn, higher diversification rates may generate key innovations that will boost range expansion under favorable conditions (Vamosi and Vamosi, 2011). In this relationship, the precedence between range size and key innovations is not clear-cut, encouraging spatiotemporal investigations with highly diverse groups of angiosperms (Hernández-Hernández and Wiens, 2020). Lineage-specific studies are essential to understand the inescapable idiosyncrasy associated with the evolution of biodiversity (Donoghue and Sanderson, 2015) and, here, we explore the historical biogeography and diversification of Apocynaceae (Gentianales), one of the ten largest families of angiosperms (Christenhusz and Byng, 2016).

Apocynaceae, the dogbane and milkweed family, consists of approximately 5,350 species and 378 genera (Endress et al., 2018-2019). The lineage is essentially pantropical, with centers of diversity in Southeast Asia, tropical America, and southern Africa (Ollerton et al., 2019), but otherwise distributed almost worldwide. It comprises large trees, treelets, shrubs, lianas (twining woody climbers), and more rarely epiphytes in tropical rainforests, succulents in dry, open landscapes, perennial herbs in temperate grasslands, to vines (herbaceous or succulent twiners) in a wide variety of habitats. The flowers are usually pentamerous (tetramerous in *Leuconotis*, and calyx tetramerous in *Parahancornia* and a few species of *Aspidosperma*), isostemonous, bicarpellate (tri- to pentacarpellate in *Lepinia*, *Lepiniopsis*, and some species of *Pleiocarpa*), and hyperdiverse in size and corolla shape (**Figure 1**), ranging from reflexed (e.g., *Asclepias*) or rotate, star-like (e.g., *Stapelia*) to campanulate (e.g., *Beaumontia* and *Mandevilla*), urceolate

FIGURE 1 | Diversity of flowers in Apocynaceae. **(A)** *Aspidosperma pyrifolium*, Aspidospermateae; **(B)** *Hancornia speciosa*, Willughbeieae; **(C)** *Couma rigida*, Willughbeieae; **(D)** *Tabernaemontana solanifolia*, Tabernaemontaneae; **(E)** *Condylocarpon isthmicum*, Alyxieae; **(F)** *Thevetia peruviana*, Plumerieae; **(G)** *Allamanda puberula*, Plumerieae, **(H)** *Wrightia coccinea*, Wrightieae; **(I)** *Adenium obesum*, Nerieae; **(J)** *Mandevilla atroviolacea*, Mesechiteae; **(K)** *Pentopetia grevei*, Periplocoideae; **(L)** *Secamone parviflora*, Secamonoideae; **(M)** *Ceropegia tihamana*, Ceropegieae; **(N)** *Orbea lutea*, Ceropegieae; **(O)** *Calotropis procera*, Asclepiadinae; **(P)** *Oxypetalum rusticum*, Oxypetalinae [Photos of panels **(A–G,I,J,O,P)** by Rapini; panels **(H,K–N)** by Meve].

(e.g., *Urceola* and *Dischidia*), salverform (e.g., *Carissa* and *Tabernaemontana*), cylindrical pitfalls (*Ceropegia*), or even fig-like (*Heterostemma ficoides* A. Kidyoo; Kidyoo, 2019). They often produce corolline and/or staminal coronas and show a spectrum of integration between anthers and style-head that culminated with the formation of a gynostegium and five pollinaria, formed by a translator derived from gynoecium exudates and pollinia from the androecium (Endress, 2016; Endress et al., 2018-2019). Several species are planted as ornamentals, mainly because of their showy flowers, such as allamanda (*Allamanda cathartica*), Madagascar periwinkle (*Catharanthus roseus*), frangipani (*Plumeria rubra*), Easter lily vine (*Beaumontia*

grandiflora), oleander (*Nerium oleander*), milkweeds (*Asclepias* spp.), and desert rose (*Adenium obesum*), but also due to a cactus-like habit, particularly the stapeliads.

Most Apocynaceae species are pollinated by insects, such as butterflies, moths, bees, wasps, beetles, flies, even cockroaches (Xiong et al., 2020), and very rarely by birds (Pauw, 1998). Almost 20% of the species studied so far are insect generalists, i.e., pollinated by more than three of the seven big classes of insects listed above, and floral complexity is not paralleled by a specialization in pollinators (Ollerton et al., 2019). Anti-herbivory and medicinal compounds, particularly alkaloids and cardenolides, are widespread in the family (e.g., Agrawal et al., 2012) and are key in several fascinating plant-animal-animal interactions. The sequestration of glycosides from milkweeds (*Asclepias* spp.), for instance, causes the emblematic monarch butterfly (*Danaus plexipus*) to be unpalatable to birds (Agrawal, 2017), whereas other groups of Lepidoptera sequester pyrrolizidine alkaloids from various genera of apocynoids to produce pheromones for courtship displays (Edgar et al., 2007). The cardiac glycoside ouabain, extracted from the bark of *Acokanthera*, is used by the African crested rat (*Lophiomys imhausi*) against larger mammals, but also by traditional East African hunters to kill elephants (Kingdon et al., 2011). However, not all interactions with animals are made through phytochemicals. For example, leaves in some epiphytic species of *Dischidia* and *Hoya* are modified into "pitchers" that are inhabited by ants. These plants offer room for the ants to raise their young and in return, the ants bring nutrients through their debris and confer protection to the plant (Janzen, 1974; Peeters and Wiwatwitaya, 2014).

The species in Apocynaceae are classified in five major groups: three monophyletic subfamilies (Periplocoideae, Secamonoideae, and Asclepiadoideae) and two non-monophyletic groups informally treated as rauvolfioids and apocynoids (Endress et al., 2018-2019). These five divisions are defined by the direction of corolla lobe aestivation, type of pollen apertures, level of reproductive synorganization, presence of translators, and number and type of pollinia. Rauvolfioids and apocynoids do not produce a morphologically differentiated translator and were placed in the traditional Apocynaceae s.str. These two groups, however, form a grade toward the pollinia-bearing subfamilies: Periplocoideae, Secamonoideae, and Asclepiadoideae (e.g., Sennblad and Bremer, 1996; Fishbein et al., 2018; Antonelli et al., 2021). The pollinia-bearing subfamilies produce differentiated translators (acellular bodies derived from the hardening of style-head exudates), which apparently evolved more than once (Livshultz, 2010; Straub et al., 2014). The clade consisting of the apocynoids along with the two pollinia-bearing lineages Periplocoideae and Secamonoideae-Asclepiadoideae is known as the APSA clade (Livshultz et al., 2007; see also Endress et al., 2018-2019). The style-head is free from the anthers in rauvolfioids, but almost always adnate to the anthers, forming a gynostegium in the APSA clade. Besides a derived floral synorganization, the APSA clade is also mostly characterized by wind-dispersed seeds with a coma at the micropylar end, produced in follicles. Fruits of rauvolfioids, in contrast, are highly variable and include dehiscent fruits with ejected or wind-blown seeds, dehiscent fruits with arillate seeds dispersed by animals, indehiscent fleshy fruits with seeds also dispersed by animals, and indehiscent fruits adapted to disperse seeds by water, even sea water (Simões et al., 2016; Endress et al., 2018-2019; **Figure 2**).

The oldest fossil assigned to the family seems to be the *Alyxia*-like pollen of *Psilodiporites wolfendenii* from the Paleocene of West Africa (Movshovich, 1975) and the Eocene of tropical Asia (Indonesia and Borneo; Muller, 1968; Morley, 1998). Comose seeds (seeds bearing a tuft of hairs at the end) remain the most convincing fossil evidence of Apocynaceae so far, indicating the presence of the APSA clade in Africa since the Eocene (47 Ma; *Apocynospermum* spp.; Collinson et al., 2010). However, most fossils from the Eocene to the Oligocene are restricted to the Northern Hemisphere, mainly to Europe and Asia (Martínez-Millán, 2010; Del Rio et al., 2020), such as the recently described fossil flower of the rauvolfioid *Maryendressantha succinifera* (Singh et al., 2021). American fossils assigned to Apocynaceae, such as the North American wood of *Paraapocynaceoxylon barghoornii*, dating from the late Cretaceous (Wheeler et al., 1987), and the well-preserved flower of the Neotropical *Discoflorus neotropicus*, dated to the mid-Tertiary (Poinar, 2017), are scant and taxonomically questionable. Pollen tetrads of *Polyporotetradites laevigatus* (Salard-Cheboldaeff, 1978) from the Oligocene-Miocene of Africa remain the oldest fossil evidence of Periplocoideae and comose seeds from the early Eocene were recently discovered in China and assigned to Asclepiadoideae (*Asclepiadospermum* spp.; Del Rio et al., 2020). Estimates for the Apocynaceae crown diversification based on molecular data are divergent, ranging from the early Eocene (Rapini et al., 2007) to the late Cretaceous (Fishbein et al., 2018; see also Ramírez-Barahona et al., 2020), with estimates exactly in between (Ribeiro et al., 2014; Pugliesi and Rapini, 2015). Although discrepant for ancient nodes, dating for intermediate nodes tends to be comparable among studies, as noted by Fishbein et al. (2018).

The biogeography of the family has been largely neglected from a phylogenetic perspective. Comprehensive inferences about the ancestral distributions of Apocynaceae and their major clades remain sparse and are mainly based on informal observations. Simões et al. (2007) suggested a tropical origin for the Apocynaceae, but did not postulate a place for their emergence, and Meve and Liede (2002b); Goyder (2006), Rapini et al. (2007), and Livshultz et al. (2007, 2011) suggested an African origin for the pollinia-bearing lineages. Pollinia provide additional protection against pollen desiccation, confer higher efficiency to pollination under low population densities, and may have reduced the extent of extinction in drier climate niches, contributing to the high diversity of milkweeds (Livshultz et al., 2011). However, formal biogeographic analyses in the family are few and restricted to particular asclepiad groups (e.g., Liede-Schumann et al., 2014, 2016), providing limited insights into a global understanding of the family's diversification. Spatiotemporal patterns have not been formally investigated for the entire Apocynaceae, resulting in uncertainties regarding their place of origin and the emergence of the modern pantropical distribution.

Tropical plant families with a crown age younger than the continuing breakup of Gondwana, which started in the

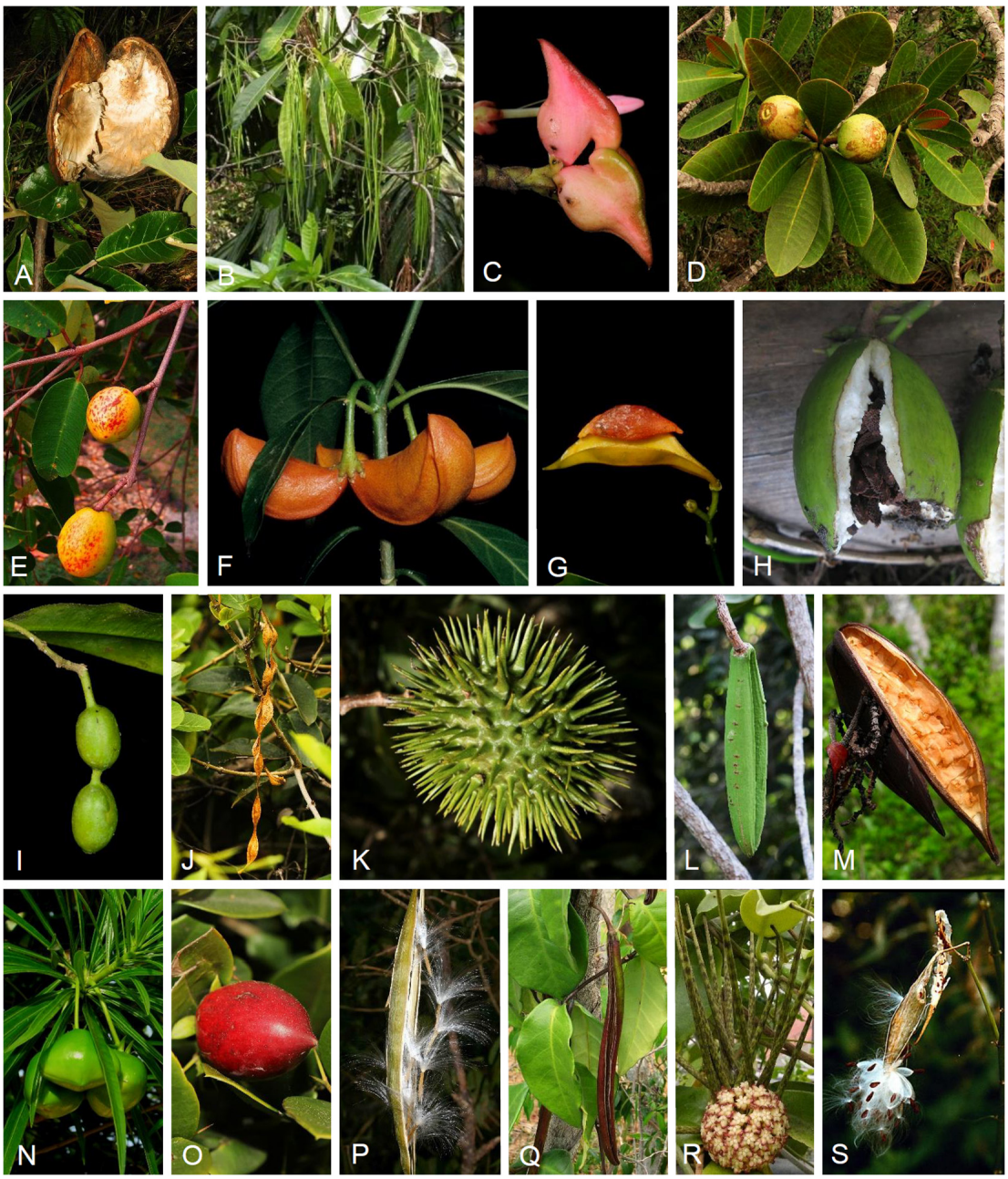


FIGURE 2 | Diversity of fruits and seeds in Apocynaceae. **(A)** Open follicle with winged seeds of *Aspidosperma macrocarpon*, Aspidospermateae; **(B)** Follicles of *Alstonia macrophylla*, Alstonieae; **(C)** One-seeded drupes of *Kopsia pauciflora*, Vinceae; **(D)** Berries of *Couma rigida*, Willughbeieae; **(E)** Berries of *Hancornia speciosa*, Willughbeieae; **(F)** Follicles of *Tabaernaemontana pauciflora*, Tabernaemontaneae; **(G)** Dehisced follicle with arillate seed of *T. bovina*, Tabernaemontaneae; **(H)** Berry with seeds of *Spongiosperma grandiflorum*, Tabernaemontaneae; **(I)** Torulose drupe with two articles of *Alyxia siamensis*, Alyxieae; **(J)** Torulose drupes with several articles of *Condylocarpon isthmicum*, Alyxieae; **(K)** Spiny capsule of *Allamanda cathartica*, Plumerieae; **(L)** Pair of fused immature follicles of *Himatanthus drasticus*, Plumerieae; **(M)** Open follicle with winged seeds of *H. bracteatus*, Plumerieae; **(N)** Drupaceous fruits of *Thevetia peruviana*, Plumerieae; **(O)** Berry of *Carissa macrocarpa*, Carisseae; **(P)** Open follicle with comose seeds of *Wrightia pubescens* subsp. *lanitii,* Wrightieae; **(Q)** Fused follicles of *Temnadenia violacea,* Echiteae; **(R)** Inflorescence with flowers and follicles of *Hoya incrassata*, Marsdenieae; **(S)** Comose seeds from an old fruit of *Asclepias curassavica,* Asclepiadinae [Photos of panels **(A,D,E,J–O,Q,S)** by Rapini; panels **(B,G)** by Middleton; panels **(C,I)** by Preecha Karaket, with permission from The Forest Herbarium Bangkok; panels **(F,R)** by Meve; panel **(H)** by Simões; panel **(P)** by Leonid Averyanov].

Jurassic, with greater peaks of continental fragmentation in early and late Cretaceous to Paleocene (McLoughlin, 2001), reached their ancestral distribution *via* two main historical routes: a southern Gondwanan (e.g., Axelrod and Raven, 1978; Linder and Crisp, 1995; van den Ende et al., 2017) or a northern Laurasian route (e.g., Davis et al., 2002; Ruhfel et al., 2016).

Globally warmer climates, especially from the Paleocene–Eocene thermal maximum and early Eocene climatic optimum (ca. 56–48 Ma) to the Eocene-Oligocene climate transition (ca. 34 Ma; the terminal Eocene event, sensu Wolfe, 1978), allowed plants to disperse between the Old and the New World using more poleward routes and possibly land bridges to cross between continents in both Hemispheres (Wolfe, 1975, 1978; Axelrod and Raven, 1978; Tiffney, 1985b,a; McLoughlin, 2001; Morley, 2003). The global cooling and aridification after the Eocene-Oligocene climate transition closed these routes for megathermal plants and resulted in the expansion of temperate regions and the establishment of modern grasslands and deserts (Simon et al., 2009; Arakaki et al., 2011; Gagnon et al., 2019; Folk et al., 2020).

In this study, we provide the first attempt to integrate the most comprehensive time-calibrated phylogeny and morphological character evolution estimates in Apocynaceae (Fishbein et al., 2018) with an updated monographic work for the family (Endress et al., 2018-2019) to investigate (1) the historical biogeography of major lineages and (2) the dynamics of species diversification in the family. More specifically, we ask (1) whether a Gondwanan (Southern Hemisphere) or Laurasian (Northern Hemisphere) route was prevalent in shaping the distribution of Apocynaceae and (2) how variation in diversification rates underpinned the evolution of the family. First, we estimate ancestral areas under several models to produce new evidence on the origin of the family and the main routes of historical dispersals. Then, we estimate rates of species diversification across the Apocynaceae and test for shifts that may hint toward potential drivers of diversification. Finally, we associate estimates of diversification rates, historical biogeography, and morphological evolution (Endress et al., 2018-2019; Fishbein et al., 2018) to discuss the evolutionary history of Apocynaceae and to provide an overview of potential events that boosted the diversification of the family, from the uncertain origin to the uneven distribution of extant diversity.

MATERIALS AND METHODS

Phylogenetic Tree, Species Numbers, and Distributions

We make use of the time-calibrated phylogenetic tree of Apocynaceae published in Fishbein et al. (2018) that covers all major lineages in the family and includes 1,041 species (ca. 20% of species richness) and 50% of the genera (ca. 180 out of 378 currently accepted sensu Endress et al., 2018-2019). The tree is a chronogram based on maximum likelihood analysis of 21 concatenated plastid loci (for more details, see Fishbein et al., 2018). Due to the incomplete and biased sampling, an inevitable limitation under such a broad coverage (e.g., Magallón et al., 2019; Mendel et al., 2019; Sun et al., 2020), we restricted our interpretations to Apocynaceae major clades, mainly above the genus level. Numbers of known species per clade and their distributions are recorded from Endress et al. (2018-2019) and updated according to new data and specialists' knowledge (**Supplementary Table 1**; all data and scripts used in this study are available at our DRYAD repository).

Biogeographic Hypothesis Testing

Currently, three common biogeographic models are available for phylogenetic comparative inference of ancestral areas among species: dispersal and vicariance analysis (DIVA; Cronquist, 1997), the dispersal-extinction-cladogenesis (DEC; Ree et al., 2005; Ree and Smith, 2008), and BayArea models (Landis et al., 2013). The last model was the first to include a distance-dependent parameter, which can be set according to, for example, ecological distances, making possible direct comparison of biologically informed hypotheses (e.g., Pirie et al., 2019). These three biogeographic models are implemented in a likelihood framework in BioGeoBEARS (Matzke, 2013), which can also add the founder event (+J) parameter to all models, allowing speciation and dispersal to occur simultaneously. It has been argued that likelihoods under the classical models and their +J variants cannot be compared because +J models are biased toward having higher likelihoods (see discussion in Matzke, 2021). Since we found only minor differences in ancestral area estimates under the + J models compared to estimates without this parameter (**Supplementary Tables 2A,B** and **Supplementary Figures 2–5**, besides that model fit was generally much better for the + J models), we here present and discuss the estimates of the classical models (full results are available at DRYAD). For model comparison, we relied on log likelihoods and the sample size corrected Akaike information criterion (AICc), also calculating the ΔAICc (ΔAICc > 10 was taken as good evidence to reject a model in favor of the best one; Burnham and Anderson, 2002) and Akaike weights, which indicate the relative fit of a model in relation to all models in the set (i.e., cumulative Akaike weights of all models = 1).

We designed models to estimate ancestral areas during the diversification of Apocynaceae and to test specific hypotheses about (1) the geographical origin of the lineage and (2) macroevolutionary patterns of dispersal in BioGeoBEARS. Based on centers of species richness, relevant geological time periods, and theories of the major biogeographic realms, we divided the globe into six major terrestrial areas: Africa, including the Arabian Peninsula and Madagascar (A), temperate Eurasia (E), North America (N), Oceania, including Australasia and eastern Malesia (O), South America (S), and tropical Asia (T; justification, references, and maps are provided in Methods S1). We created a default model (M1), which allowed unconstrained and time homogeneous dispersal probabilities, and set the maximum number of areas to four (the maximum inhabited by extant species). To model shifts in dispersal probabilities over time due to plate movements, changing sea levels, and other major geological processes, we then defined a time-stratified model (M2) with four time periods (time slices: 0–30, 30–60, 60–80, 80–100 Ma; for details, see **Supplementary Methods 1** and **Supplementary Figure 1**. N.B., the M2 model also restricts most dispersals between Hemispheres and is intended to illuminate the importance of such interhemispheric dispersal pathways in the biogeographic history of Apocynaceae). Additionally, in a "restricted range"

modification of M1 and M2, we excluded unrealistic ancestral population distributions (e.g., Africa–North America) from the state space of geographic ranges (M1s, M2s models; for details, see R scripts at DRYAD). Additionally, to test hypotheses about the origin of Apocynaceae (i.e., the ancestral geographic range at the crown node of the family), we enforced the following constrained "test" models: South America (M3), Africa (M4), West Gondwana (M5), Gondwana (M6), tropical Asia (M7), East Gondwana (M8), Palaearctic (M9), Laurasia (M10), and two pantropical models, including West Gondwana plus tropical Asia (Pantropical 1, M11) and Gondwana plus tropical Asia (Pantropical 2, M12), respectively (for details, see **Supplementary Table 3**). At the time of crown group origin (85.6 Ma), these were the relevant landmasses according to plate tectonic reconstructions. We evaluated the test models with default (M3–M12) and restricted range (M3s–M12s) state spaces of geographic ranges. Each of the 24 models was run in BioGeoBEARS, using the DEC, BayArea-like, and DIVA-like models.

We estimated the frequency of dispersal events between areas with biogeographical stochastic mapping (BSM; Dupin et al., 2016; Matzke, 2016) using the model that best fits the data without constraining the area of the Apocynaceae crown node (M1s.dec model; see section "Results"; for BSM results including the + J parameter, see **Supplementary Table 4**). We conducted 1,000 MCMC stochastic mappings of occupied ranges and summarized dispersal events from the posterior distribution (R scripts and full analysis output are available at DRYAD).

Diversification Analyses

To detect variation in species diversification rates among lineages in the Apocynaceae phylogeny, we used maximum likelihood (ML) and Bayesian approaches. First, we used ML in a character-free version of the HiSSE model (Beaulieu and O'Meara, 2016; Caetano et al., 2018), the so-called missing-state SSE (MiSSE) model, which implements averaging of rate estimates weighted according to goodness-of-fit of the models to the data (model averaging for short). We ran 52 MiSSE models varying in the number of "hidden" states (i.e., the number of different diversification regimes) with 1–26 regimes considered, and by setting the ratio of extinction to speciation (extinction fraction) to either 0.0 or 0.9, representing extreme diversification scenarios (Magallón and Sanderson, 2001). Sampling fraction was set to 20% for the entire tree, based on the empirical level of sampling. We used the functions generateMiSSEGreedyCombinations, MiSSEGreedy, and MarginReconMiSSE to obtain log likelihoods for the 52 models, and averaged rate estimates using Akaike weights based on the AICc (Burnham and Anderson, 2002).

Second, we used Bayesian analysis of macroevolutionary mixture (BAMM 2.5; Rabosky, 2014) to reconstruct the positions of significant shifts in diversification regimes during the evolution of Apocynaceae. Sampling fractions per clade were provided in order to account for uneven representation of species that are included in the tree (**Supplementary Table 1**). Appropriate priors were estimated in BAMMtools v2.1.7 (Rabosky et al., 2014) and a series of different priors on the expected number of shifts were tested in preliminary analyses (0.1–12; because results were stable among tested priors, we show and discuss results using a shift prior of 1; for comparison of the prior and posterior distributions of number of shifts, see **Supplementary Figure 6**). We ran four simultaneous Markov chain Monte Carlo (MCMC) chains for 20 million generations, saving estimates every 10 thousand generations and discarding the first 10% as burn-in before assessing convergence of chains; mixing of chains was assessed visually and effective sample sizes calculated using coda v0.19-4 (Plummer et al., 2006) were >200.

RESULTS

Ancestral Area Estimates and Biogeographic Tests

Of the 72 biogeographical models analyzed, the DEC model with restricted ranges and no constraint on the Apocynaceae ancestral area (M1s.dec) was the best fit by far, with an Akaike weight of 0.92 (**Figure 3**). Only one other model, the unrestricted DEC (M1.dec), had a ΔAICc < 10 (4.9), with an Akaike weight 0.08. All other models performed substantially worse (ΔAICc > 200; **Supplementary Table 2A**), including the time-stratified model M2 and M2s, which indicates the importance of interhemispheric dispersals in the history of the Apocynaceae. According to the M1s.dec model, the most recent common ancestor of Apocynaceae was most likely distributed in a geographic region that encompassed South America, Africa, and tropical Asia (S + A + T: state probability P = 0.62; **Figure 3** and **Supplementary Figure 2**), whereas according to the M1.dec model, it would additionally include North America (S + A + T + N: P = 0.26; **Supplementary Figure 3**; state probabilities for all nodes are available at DRYAD). Uncertainties in ancestral ranges are apparent for other backbone nodes near the root, with South America, Africa, and/or tropical Asia as the most likely ranges (**Figure 3** and **Supplementary Figure 2**). However, the crown node of early diverging clades was often associated with particular geographic regions generally with high degrees of certainty: e.g., the diversification of the Aspidospermateae (in South and North America, P = 0.94), Alstonieae and Vinceae (in tropical Asia, P = 0.91 and P = 0.81, respectively), Willughbeieae–Tabernaemontaneae (in South America, P = 0.24), and Melodineae–Asclepiadoideae (in Africa, P = 0.96).

Models constraining the ancestral range of Apocynaceae (M2–M12 and M2s–M12s) had much poorer fits than the less constrained models (M1 and M1s; **Supplementary Table 2A**). A comparison of the model fit among the 10 ancestral constrained "text" models (M3–M12) yields similar equivocal support for the range at the node of crown group (**Table 1**). Irrespective of the biogeographic model used (DEC, BayArea-like, or DIVA-like; **Supplementary Table 5**), only two constraints under the DEC model, the Pantropical model 2 (S + A + O + T; Akaike weight 0.50) and the Pantropical model 1 (S + A + T; Akaike weight 0.42), accounted for a cumulative AICc weight of 0.92 (**Table 1** and **Figure 3**). Other constrained models with ΔAICc < 10 included the Laurasia model (N + E + T; Akaike weight 0.07,

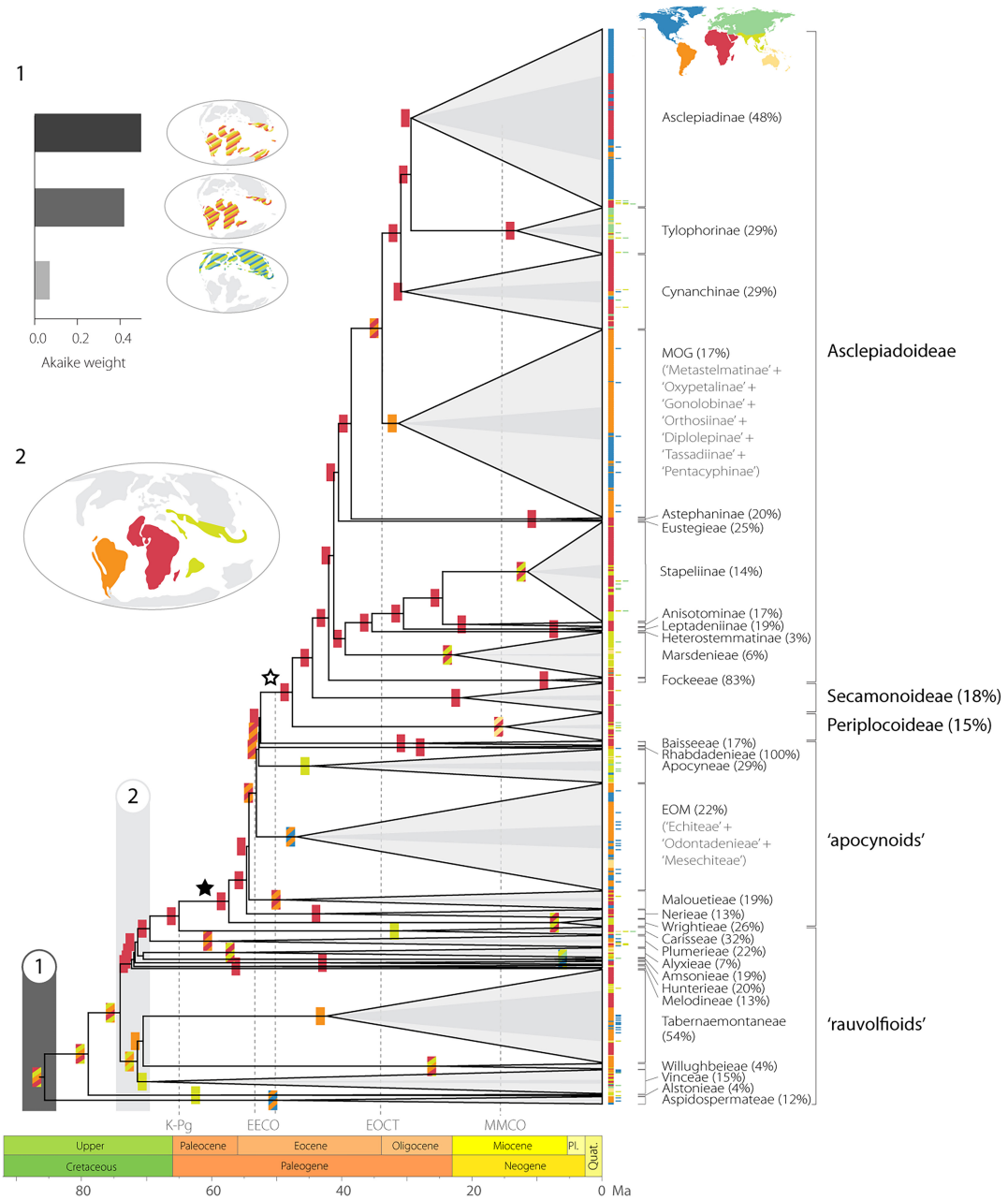


FIGURE 3 | Historical biogeography of Apocynaceae. Colored boxes at the tree nodes indicate the most probable ancestral geographic range inferred by the best biogeographic model (out of 72: M1s-DEC, Akaike weight 0.92; **Supplementary Table 2A**). Current distribution of species is given to the right of the tips following the color code in the map inset top right. Named clades according to Fishbein et al. (2018) are collapsed and sampling fraction is given after the clade name. The superimposed triangles in darker gray are scaled to the proportion of the sampling fraction per clade. The filled star indicates the APSA clade and the open star the pollinia-bearing lineages. (1) Results of biogeographic hypothesis tests of the ancestral area of Apocynaceae showing the model fit (Akaike weight): Pantropical 2, Pantropical 1, and Laurasia (**Table 2**). (2) The inferred ancestral distribution of Apocynaceae at around 72 Ma is indicated in the map. Pl., Pliocene; Quat., Quaternary; K-Pg, Cretaceous-Paleogene boundary; EECO, early Eocene climatic optimum; EOCT, Eocene-Oligocene climate transition; MMCO, middle Miocene climatic optimum.

ΔAICc 4; **Figure 3**), and the Gondwana model (S + A + O; Akaike weight 0.01, ΔAICc 8; **Table 1**).

Throughout the evolution of Apocynaceae, Africa is estimated by both ancestral area model fits (**Figure 3** and **Supplementary Figure 2**) and BSM (**Table 2**) as the main source of dispersals to other areas with 64.2 events in total (ca. 0.8 events per million years; E/Myr), mainly to tropical Asia (ca. 0.3 E/Myr) and temperate Eurasia (ca. 0.2 E/Myr). The highest rates of lineage dispersals are estimated to have occurred between the two New World areas, with ca. 0.5 E/Myr from South to North America,

TABLE 1 | Results of biogeographic hypothesis tests on the origin of Apocynaceae detailing fit to the data of the ten constrained models (M3–M12) using the dispersal-extinction-cladogenesis model (DEC).

Model	Hypothesis	Range	lnL	DF	AICc	ΔAICc	Akaike weights
M12	Pantropical 2	S + A + O + T	−1,464.25	2	2,933	0	0.498
M11	Pantropical 1	S + A + T	−1,464.43	2	2,933	0	0.418
M10	Laurasia	N + E + T	−1,466.22	2	2,936	4	0.070
M6	Gondwana	S + A + O	−1,468.33	2	2,941	8	0.009
M5	West Gondwana	S + A	−1,469.16	2	2,942	10	$3.68e^{-03}$
M8	East Gondwana	T + O	−1,471.89	2	2,948	15	$2.40e^{-04}$
M3	South America	S	−1,473.31	2	2,951	18	$5.80e^{-05}$
M7.	Tropical Asia	T	−1,473.42	2	2,951	18	$5.20e^{-05}$
M9	Palearctic	N + E	−1,473.22	2	2,950	18	$6.36e^{-05}$
M4	Africa	A	−1,473.99	2	2,952	19	$2.94e^{-05}$

Given is the log likelihood (lnL) and the degree of freedom (DF) of models, the sample size corrected Akaike Information Criterion (AICc), delta AICc (ΔAICc) values, and Akaike weights. The table is sorted according to model fit. A, Africa; E, temperate Eurasia; N, North America; O, Oceania incl. Australasia; S, South America; T, tropical Asia.

and ca. 0.2 E/Myr for the reverse. North America is estimated as the main sink receiving ca. 0.5 E/Myr, closely followed by tropical Asia (**Table 2**).

Rates of Species Diversification

A model of constant diversification rate for the Apocynaceae was strongly rejected in both MiSSE and BAMM analyses. A single diversification regime (without a shift) was not included in the posterior of BAMM, and the MiSSE models with a single regime (one state in parameter "turnover") were rejected by a ΔAICc of 776 for an extinction fraction (eps) of 0.9 and 1,236 for eps = 0.0 compared to the best model (**Supplementary Table 6**). Instead, the best-fitting MiSSE model assumed 18 diversification regimes and an extinction fraction of 0.9, with an Akaike weight of 0.13. Of the 52 MiSSE models tested, 16 models assuming 11–26 distinct diversification regimes received a ΔAICc < 10, all of which with eps = 0.9, with a cumulative AICc weight of 0.9998. The nine top models with ΔAICc < 2 assumed 14–22 distinct diversification rates with a cumulative Akaike weight of 0.8674 (**Supplementary Table 6**). The model-averaged MiSSE rate estimates of net diversification in Apocynaceae are given in **Figure 4**.

The BAMM analysis produced a single best shift configuration that includes 12 shifts toward accelerated species diversification (i.e., 13 different regimes; **Figure 4** and **Table 3**). Shift configurations with 13–16 shifts in diversification rates are sampled with even higher frequencies in the posterior of the BAMM analysis (**Supplementary Figure 4**), which agree with our MiSSE results. Of the 12 shifts inferred in the best shift configuration (**Figure 4** and **Table 3**), 11 occurred in the APSA clade: one in Wrightieae, two within the EOM clade and eight in pollinia-bearing lineages, including one leading to Periplocoideae, another to Secamonoideae, and the other six within Asclepiadoideae. Shifts toward higher net diversification rates occurred from the mid-Eocene onward, the oldest within the EOM clade (ca. 45 Ma) and the most recent within *Mandevilla* (ca. 3.3 Ma), with seven shifts during the Miocene. Although not detected as rate shifts by BAMM, the MiSSE results additionally indicate two older accelerations in species diversification rates, in the ancient radiations of rauvolfioids ca. 72–69 Ma (R1) and apocynoids ca. 55–52 Ma (R2; **Figure 4**).

DISCUSSION

The Origin of Apocynaceae

The complex breakup sequence of the southern supercontinent Gondwana started 180–150 Ma, involved approximately 90 fragments (including regions in Southeast Asia, Europe, and Florida), and caused disjunctions in lineages of angiosperms between land masses that currently correspond mainly to South America, Africa and Madagascar, Australia and New Zealand, and Antarctica (McLoughlin, 2001). Disjunct distributions in the Southern Hemisphere, therefore, are often interpreted as evidence of a Gondwanan origin followed by vicariance (e.g., Linder and Crisp, 1995), although this may not always be the most plausible explanation (e.g., Davis et al., 2002). Studies questioning the primacy of tectonic vicariance over oceanic dispersals in historical biogeography (e.g., Richardson et al., 2004; de Queiroz, 2005) and reinforcing the importance of dispersals on the distribution of tropical lineages since the

TABLE 2 | Numbers of dispersal events between the six biogeographic areas estimated by biogeographical stochastic mapping [BSM using M1s.dec; sum all dispersals (d) and range switching (a) events].

Area	N	S	A	T	O	E	Source
N		14.6	0.5	0.1	0.0	1.7	17.0
S	41.0		2.9	2.6	2.5	0.0	49.0
A	0.9	7.6		27.9	10.0	17.8	64.2
T	0.6	2.7	5.5		16.3	11.4	36.5
O	0.2	0.4	1.1	7.3		1.0	9.9
E	3.1	0.0	2.6	6.4	1.7		13.7
Sink	45.7	25.4	12.6	44.2	30.6	31.9	

Dispersals from an area (source) in rows to an area (sink) in columns. A, Africa; E, temperate Eurasia; N, North America; O, Oceania incl. Australasia; S, South America; T, tropical Asia. Note that the approximate rate of dispersals per million year is the sum of events divided by the crown age of the Apocynaceae.

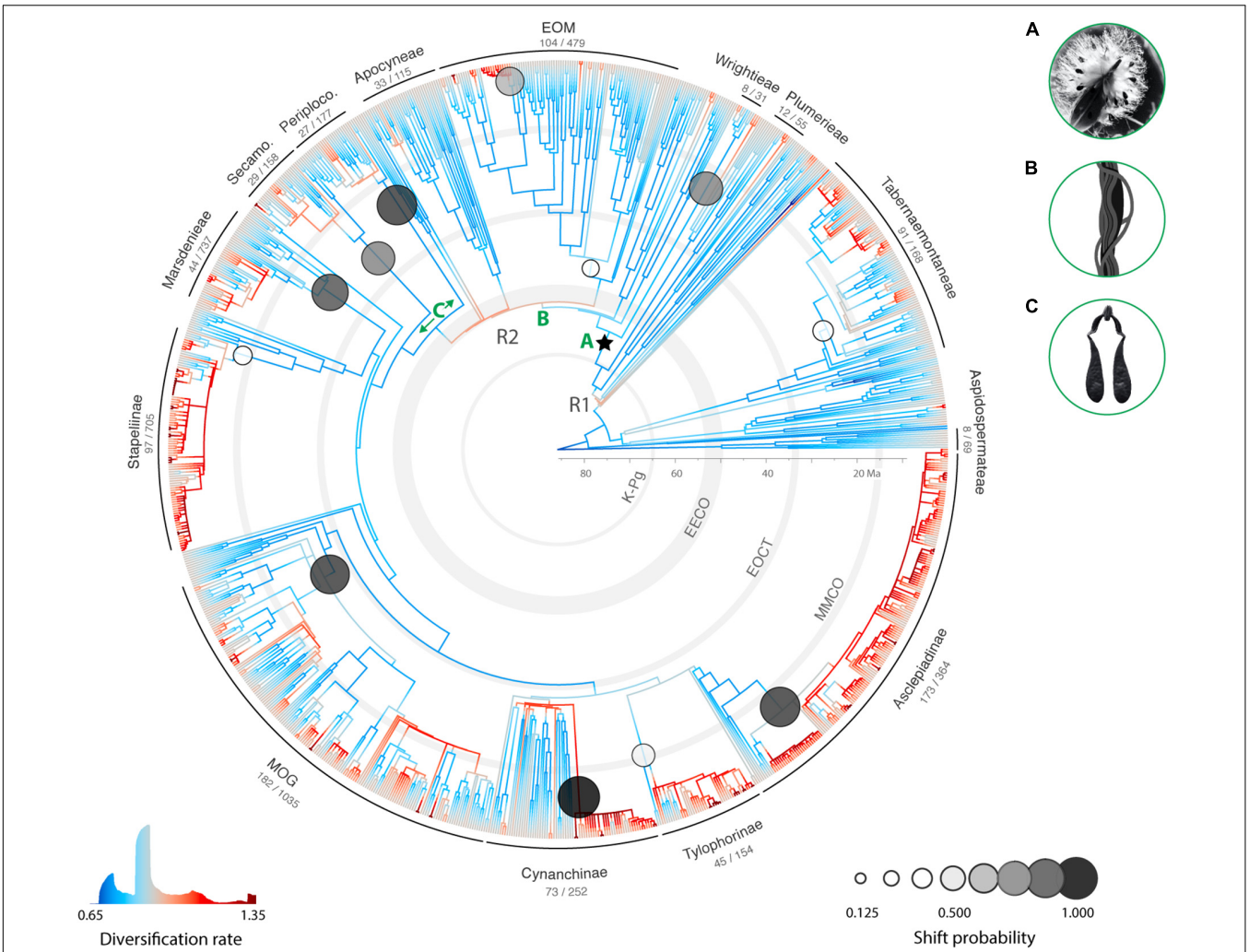


FIGURE 4 | Species diversification in Apocynaceae. Colors on tree branches are averaged net diversification rate estimates inferred by the 52 models varying in the number of rate regimes ("hidden states"; 1–26) with extinction fraction (ratio of extinction to speciation) kept at 0.0 (26 models) or 0.9 (26 models) using MiSSE (rates > 1.35 species per million years are not shown in the color scale). Circles in grayscale superimposed on the tree indicate the nodes for which a shift in diversification rates has been inferred by BAMM and are scaled in gray and size to marginal likelihoods (shift probability; **Table 1**). Selected clade names and sampling fractions (spp. sampled/spp. known; for details see **Figure 3** and **Supplementary Table 1**) are indicated. The ancient radiations (internodes with warmer colors) of rauvolfioids (close to R1) and apocynoids (close to R2) are indicated. The star indicates the APSA clade. Secamo., Secamonoideae; Periploco., Periplocoideae. EOM, clade comprising Echiteae, Odontadenieae, and Mesechiteae; MOG, clade comprising Metastelmatinae, Oxypetalinae, and Gonolobinae (plus other five smaller Neotropical subtribes). K-Pg, Cretaceous-Paleogene boundary; EECO, early Eocene climatic optimum; EOCT, Eocene-Oligocene climate transition; MMCO, middle Miocene climatic optimum. Insets top right are potential drivers of diversification: A, dry fruits with comose seeds; B, climbing habit (twining growth form) and; C. pollinarium with a pair of pollinia, indicating the two pollinia-bearing lineages as discussed in the text.

late Cretaceous (e.g., Renner, 2004; Ruhfel et al., 2016) are increasingly being published. The predominantly austrotropical distribution of Apocynaceae fits the pattern classically associated with an initial diversification in tropical West Gondwana, i.e., South America-Africa in the late Cretaceous, ca. 85 Ma (Axelrod and Raven, 1978). Our results, however, refute the Gondwana or West Gondwana origin of Apocynaceae and did not support the prevalence of any route. Instead, they retrieved a tropical climatic belt in the equatorial region (the pantropical models), encompassing part of Gondwana (western landmasses) and part of Laurasia (Southeast Asia), as the most likely ancestral area of Apocynaceae, potentially also including Oceania (**Figure 3**). Uneven extinction rates and high rates of dispersal in plants obscure vicariance patterns and have often distorted biogeographic analyses of extant lineages, leading to unclear and equivocal reconstructions of ancestral areas especially at deeper nodes (Sanmartín and Ronquist, 2004; Meseguer et al., 2015, 2018; Rose et al., 2018; Larridon et al., 2020). An ancient pantropical distribution is shared by several other families (Procheş and Ramdhani, 2020) and the ancestral area estimated for Apocynaceae here reinforces the great importance of intercontinental dispersals in plant distribution since the late Cretaceous (e.g., Renner, 2004; de Queiroz, 2005).

TABLE 3 | Shifts in net diversification rates inferred in the Bayesian analysis of macroevolutionary mixture (BAMM).

Clade	Node	Crown age	Shift age	Shift probability	Net diversification rate (95% HPD)
Apocynaceae	Crown	85.6	—	—	0.08 (0.07, 0.09)
EOM	*Parsonsia–Mandevilla*	42.9	45.0	0.32	0.14 (0.10, 0.16)
MOG	*Monsanima–Blepharodon*	27.0	28.9	**0.93**	0.23 (0.19, 0.26)
Secamonoideae	Crown	21.6	28.8	**0.77**	0.19 (0.13, 0.26)
Marsdenieae	Crown	22.8	25.4	**0.86**	0.28 (0.21, 0.34)
Periplocoideae	Crown	14.9	22.6	**0.95**	0.23 (0.17, 0.28)
Tabernaemontaneae	*Tabernaemontana–Voacanga*	20.8	21.8	0.42	0.17 (0.14, 0.21)
Wrigthieae	Crown	6.4	19.7	**0.78**	0.12 (0.03, 0.24)
Tylophorinae	Crown	13.2	16.3	0.48	0.31 (0.21, 0.39)
Stapeliinae	Crown	11.5	13.5	0.37	0.47 (0.35, 0.59)
Asclepiadinae	*Asclepias coulteri–A. elata*	9.9	10.8	**0.95**	0.43 (0.38, 0.48)
Cynanchinae	*Cynanchum viminale– C. implicatum*	5.4	8.5	**1.00**	0.42 (0.32, 0.52)
EOM: *Mandevilla*	*M. moricandiana–M. pohliana*	2.4	3.3	**0.63**	0.84 (0.14, 1.27)

Shift probabilities > 0.5 are in bold font. Shifts are sorted according to ages from the maximum a posteriori probability configuration; ages are in million years; HPD, highest posterior density; EOM, clade comprising Echiteae, Odontadenieae, and Mesechiteae; MOG, clade comprising Metastelmatinae, Oxypetalinae, and Gonolobinae (plus five other smaller Neotropical subtribes). Note that for Apocynaceae the background net diversification rate is given (i.e., mean rate estimates excluding clades for which a shift is detected).

Intercontinental dispersals occurred during the entire history of the family and, in a greenhouse world, we might expect that the ancestor of Apocynaceae was potentially widespread in austrotropical floras in South America, Africa, and Southeast Asia, with gene flow facilitated either through southern transoceanic land bridges (Morley, 2003) or through the boreotropical flora (Wolfe, 1975). The cool to warm temperate climate, with short periods of sunlight during winter months in Antarctica probably prevented most tropical lineages from using a southernmost route since the late Cretaceous (Sanmartín and Ronquist, 2004; van den Ende et al., 2017). North America, Europe, and Asia shared similar floristic compositions during the warmer early to mid-Eocene, suggesting a wide dispersal in tropical and subtropical floras of the Northern Hemisphere (Tiffney, 1985a,b; Su et al., 2020). The boreotropical flora reached paleolatitudes of approximately 60° N and persisted in a relatively uniform composition, from the late Cretaceous to mid-Eocene or later, with a shallow latitudinal temperature gradient (Wolfe, 1978). Eurasia was connected to eastern North America through the Northern Atlantic Land Bridge until the early Eocene and to western North America at higher latitudes by the Bering Strait, although with minor importance for megathermal plants after the early Eocene (Morley, 2003). These northern connections provided a pathway for disjunct austrotropical floras in South America, Africa, and Southeast Asia. Hence, the boreotropical flora represents an alternative hypothesis in place of oceanic long-distance dispersals to explain austrotropical floristic disjunctions that are too young to have been caused by West Gondwana fragmentation (Davis et al., 2002).

Tropical lineages of angiosperms may have used the boreotropical pathway between the New and Old World until the Eocene-Oligocene climate transition (Wolfe, 1978). The decrease of temperatures and increase of seasonality after the early Eocene climatic optimum (55 Ma), but mainly since the Eocene-Oligocene climate transition (ca. 34 Ma), caused the deterioration of the boreotropical flora, which contracted to refugia in lower latitudes, including Southeast Asia (Tiffney, 1985a; Morley, 2003). Lineages not adapted to cooler conditions were extirpated from the Northern Hemisphere during the Neogene, but part of the microthermal flora adapted to cooler climates derived from the preceding Paleogene local flora (e.g., Wolfe, 1975; Tiffney, 1985a; Nürk et al., 2015, 2018; Meseguer et al., 2018). The dramatic floristic change in the Northern Hemisphere since the Eocene-Oligocene climate transition may have blurred and distorted biogeographic results (e.g., Meseguer and Condamine, 2020), but fossil records from the Eocene of Europe, North America, and Asia (Martínez-Millán, 2010; Endress et al., 2018-2019; Del Rio et al., 2020; Singh et al., 2021) support the presence of Apocynaceae in the Northern Hemisphere at least since approximately 52 Ma. Therefore, data available so far cannot confidently indicate a center of origin for the Apocynaceae. The regions encompassing tropical Asia, South America, and Africa recovered in the ancestral area reconstruction could be interpreted as "museums" (sensu Stebbins, 1974) of Apocynaceae as they have apparently retained the oldest extant lineages of the family since the earliest diversification.

Apocynaceae Spatial Diversification: Tempo and Mode

The highest diversification rates in Apocynaceae are concentrated in relatively recent lineages, although ancient rauvolfioids and apocynoids also showed episodes of higher diversification rates in the MiSSE analysis, giving birth to lineages that are often recognized at tribe and subfamily levels (**Figure 4**). The oldest tribe of Apocynaceae (Aspidospermateae, 69 spp.) diversified exclusively in the New World, while the second oldest tribe (Alstonieae, 45 spp.) diversified exclusively in the Old World (**Figure 3**; for complete ancestral area estimates on the models with the best fit, see **Supplementary Figures 2**, **3**).

In contrast, another early diverging clade, Vinceae (156 spp.; **Figure 3**), apparently showed a higher capacity of dispersal and is distributed worldwide. Vinceae shows multiple intercontinental dispersal events and reached the Hawaiian Islands more than once, possibly from the southern and western Pacific and from the Caribbean islands (Simões et al., 2016). Although follicular fruits are found in some genera, drupaceous fruits are prevalent and the tribe is mainly dispersed by animals and water. Vinceae may have occupied boreo- and austrotropical forests during the Eocene and the largest genus, *Rauvolfia*, reached South America from the Paleotropics more than once according to our reconstruction (**Supplementary Figure 2**), starting in the Miocene, probably through transoceanic dispersals, and thence to North America, although Simões et al. (2016) supported only a single dispersal to the Neotropics. The sister clades Willughbeieae (144 spp.) and Tabernaemontaneae (168 spp.) share South American ancestral and current pantropical distributions (**Figure 3**). Both dispersed to the Paleotropics during the Oligocene-Miocene boundary, but a shift in diversification rates was detected only within Tabernaemontaneae (**Figure 4**). This is the only shift in a lineage of Apocynaceae consisting predominantly of trees and shrubs producing ecomose seeds. The shift may be associated with a change from tall trees to treelets and shrubs growing at the lower canopy and below but also with a change in the dispersal mode as it roughly coincided with a shift from berries to dehiscent follicles with arillate seeds, characteristic of most derived Tabernaemontaneae.

The remaining Apocynaceae began to diverge in Africa and a few small tribes appeared in quick succession during the ancient rauvolfioid radiation in the late Cretaceous (**Figures 3**, **4**). Melodineae (I and II; 30 spp.) and Hunterieae (20 spp.) remained restricted to the Paleotropics. Conversely, Amsonieae (16 spp.) shows a Laurasian distribution (North America, Europe, and Japan), possibly representing a relict component of the Eocene boreotropical flora, adapted to the cooler Neogene in higher latitudes and able to disperse through the Bering Strait, still available for subtropical plants during the late Miocene (Graham, 2011, 2018). Alyxieae (138 spp.) dispersed to Southeast Asia and Oceania, reaching South America from the Paleotropics in the Eocene, Plumerieae (55 spp.) dispersed to South America in the Paleocene, showing independent diversifications in the New and Old Worlds since the late Eocene, and Carisseae (25 spp.) remained restricted to the Old World, with *Carissa* probably dispersing to Asia and Australia before the reduction of forests and expansion of grasslands, caused by an increase of aridity in northern and eastern India after the mid-Miocene climatic optimum (Chen et al., 2019).

The oldest apocynoid tribes also diverged in Africa (**Figure 3**). Wrightieae (31 spp.), sister to the rest of the APSA clade, diverged in the Paleocene, dispersed eastward to tropical Asia (and Australia), and showed a shift in diversification rates in the Miocene (**Figure 4**). Nerieae (71 spp.) remained mainly restricted to Africa, independently reaching Europe and Asia, whereas in Malouetieae (95 spp.), early diversification occurred in Africa and South America, during the Eocene, with later dispersals to tropical Asia and North America (see also Livshultz et al., 2007). Apart from the Paleocene appearance of Wrightieae, Nerieae, and Malouetieae, the ancient apocynoid radiation coincided with the early Eocene climatic optimum. The increase of diversification rate estimated by MiSSE during this warm period may be associated with a sequential expansion from Africa to South America and Southeast Asia followed by retractions (**Figures 3**, **4**), giving rise to the EOM clade and Rhabdadenieae in the New World, as well as Apocyneae, Baisseeae, and the pollinia-bearing lineages in the Old World. The EOM clade (479 spp.) consists mainly of Neotropical lianas and comprises the tribes Echiteae, Odontadenieae, and Mesechiteae, although Odontadenieae was not recovered as monophyletic in the large Apocynaceae tree of Fishbein et al. (2018) used here. Outside Asclepiadoideae, this is the most diverse lineage of Apocynaceae in the New World. A lineage of the EOM clade dispersed to the Paleotropics between the Oligocene and Miocene and became especially diverse in Oceania (*Artia-Parsonsia*, ca. 85 spp.). The oldest and the most recent shifts in diversification rates are located in the EOM clade (**Figure 4**). The more recent and likely rate shift in the late Pliocene is located in a lineage of *Mandevilla* with flowers bearing only a pair of nectaries alternate with the carpels (Clade III p.p. in Simões et al., 2006). Apocyneae (115 spp.) originated and diversified mainly in tropical Asia, but became widespread in the Old World, reaching North America (*Apocynum*), whereas Rhabdadenieae (3 spp.) diversified in South America, dispersing northward, and Baisseeae (29 spp.) remained restricted to Africa and Madagascar.

Two pollinia-bearing lineages emerged in Africa, independently derived from the apocynoid ancient radiation (Straub et al., 2014), although they formed a clade in the Apocynaceae tree used here (Fishbein et al., 2018; **Figures 3**, **4**). Livshultz et al. (2011) hypothesized that the more efficient pollination mechanism conferred by pollinia contributed to reduce the Allee effects caused by low population densities and fewer pollinators available in cooler and seasonally dry forests in Africa after the early Eocene climatic optimum (Jacobs, 2004). Periplocoideae (177 spp.) showed a shift in diversification rates during the early Miocene, and remained distributed mainly in Africa and Madagascar, reaching Asia, Australia, and Europe. The subfamily consists mainly of lianas, and pollinia evolved more than once in the group (Ionta and Judd, 2017). Secamonoideae (158 spp.) is sister to Asclepiadoideae and showed a shift in diversification rates during the Oligocene. The lineage remains restricted to the Old World, and an origin in Madagascar, which still harbors three relict genera (*Calyptranthera*, *Pervillaea*, and *Secamonopsis*), is most likely (Lahaye, 2005). From there, it probably dispersed to Africa and, subsequently, to Oceania as well as to Asia. *Secamone* (90 spp.), the largest genus of Secamonoideae, is also most diverse in Madagascar, possibly due to pulses of vicariance and ecological specializations (Lahaye et al., 2007).

Six of the 12 shifts in Apocynaceae diversification rates are reconstructed in Asclepiadoideae, the largest subfamily, with about 60% of the species (**Figure 4**); none of them in the oldest tribe Fockeeae (nine spp.). Two independent shifts were detected in the sister tribes Marsdenieae and Ceropegieae and four in Asclepiadeae, with one in Tylophorinae, one in the MOG clade,

one in Cynanchinae, and one in Asclepiadinae. Marsdenieae (737 spp.) crown diversification is marked by a shift in diversification rates in the late Oligocene, associated with its expansion across the Paleotropics. The tribe shows more than one dispersal to Asia, with one of the lineages becoming highly diverse in tropical Asia and Oceania (*Hoya-Dischidia*, > 400 spp.). It also reached South America in the mid to late Miocene, giving rise to an exclusively American clade of Marsdenieae (ca. 120 spp., segregated as *Ruehssia*; Espírito Santo et al., 2019), which has the highest diversification rate among the four Neotropical asclepiad lineages (Rapini et al., 2007).

Within Ceropegieae (Stapeliinae + Anisotominae + Leptadeniinae + Heterostemmatinae; 786 spp.), the unresolved relationships among more than 30 genera of Stapeliinae (stapeliads; Bruyns, 2000; Meve and Liede, 2002a) suggested a possible rapid diversification, resulting in a controversial proposal to include all species of the subtribe in a single genus (Bruyns et al., 2017). Prior to the availability of dated phylogenies of Apocynaceae, a Gondwanan distribution for some stapeliads was proposed due to their occurrence in Africa and India (Bruyns, 2000). However, the diversification of Stapeliinae has been estimated to be in the Miocene or later (Rapini et al., 2007; Bruyns et al., 2015; Fishbein et al., 2018), too recent to support that hypothesis. Meve et al. (2016) proposed an origin and early diversification of Ceropegieae in wet tropical forests of Southeast Asia, with at least two dispersals westward, to northeastern and eastern Africa. However, we reconstruct the origin and early diversification of Ceropegieae in Africa starting in the late Eocene, with two dispersals to Southeast Asia. The first gave rise to *Heterostemma* (30 spp.) and the second occurred within the Stapeliinae (705 spp.). The origin of Stapeliinae was associated with a shift in diversification rates soon after the mid-Miocene, with nested radiations in the Pliocene (**Figure 4**), associated with the establishment of the succulent biome (Arakaki et al., 2011; Gagnon et al., 2019). *Ceropegia* radiated in tropical Asia, while the remaining Stapeliinae, with succulent stems, diversified in a drier Africa (Bruyns et al., 2014, 2015; Meve et al., 2016), with several dispersals to Asia (**Supplementary Figure 2**). The appearance in Stapeliinae of fleshy and soft roots from ancestral hard and wiry roots provided an adaptation to drier habitats, which was followed by the characteristic stem succulence of the clade (Bruyns et al., 2017).

The remaining Apocynaceae fall into two clades, the depauperate, African Eustegieae (4 spp.) and its sister group, Asclepiadeae (1,820 spp.), the largest tribe of the family, which retains an ancestral African distribution, as does the early-diverging Astephaninae (15 spp.) (**Figure 3**). A dispersal from Africa to South America at the end of the Eocene gave rise to the MOG clade (1,053 spp.). As the North Atlantic Land Bridge broke up in the early Eocene and the Bering Strait at higher latitude had limited importance for evergreen plants (Tiffney, 1985b; Morley, 2003), a transatlantic dispersal may be the most likely mode of arrival in the New World (Rapini et al., 2007). The MOG lineage, the most species-rich of the Neotropical clades (about 20% of the family), consists of eight subtribes, the three largest giving the clade its designation: Metastelmatinae, Oxypetalinae, and Gonolobinae. A shift in diversification rates is inferred early in the radiation of the MOG clade during the Oligocene, and several rate increases are apparent within it (**Figure 4**). At least four dispersals to Mesoamerica were inferred starting in the Miocene: once each in Gonolobinae and Oxypetalinae (ca. 20 Ma), Metastelmatinae (ca 10 Ma), and Orthosiinae (after 3 Ma). Of these, only Gonolobinae shows significant diversity in North America with dispersal back to South America (**Supplementary Figure 2**).

Cynanchinae (252 spp.) consists of *Schizostephanus* (2 spp.) and *Cynanchum* (250 spp.). The subtribe is inferred to have originated and begun to diversify in Africa during the Oligocene (**Figure 3**). *Cynanchum* consists of two major lineages with strongly contrasting evolutionary trajectories. For the first, a brief period of rapid diversification in Africa was followed by low diversification rates, with one of its lineages dispersing northward to Eurasia and from there to the New Word, possibly by the Bering Strait, reaching South America at the end of the Oligocene (**Supplementary Figure 2**). The other major lineage of *Cynanchum* dispersed eastward, reaching Madagascar, Southeast Asia, and Oceania. It consists of succulent species previously assigned to different genera and radiated after a shift in diversification rates during the late Miocene (**Figure 4**). This diversification was relatively synchronous with the diversification of Stapeliinae in Africa and the diversification of other succulent plant lineages with the establishment of modern deserts (Arakaki et al., 2011). The succulent group comprises one-half of the 100 Madagascan species of *Cynanchum* (Liede and Täuber, 2002) and dispersed from there back to Africa and other Paleotropical regions (Meve and Liede, 2002b; Khanum et al., 2016).

Tylophorinae (154 spp.) consists of two genera, *Pentatropis* (4 spp.) and *Vincetoxicum* s.l. (150 spp.), including *Tylophora* and all other genera traditionally recognized in Tylophorinae (Liede-Schumann and Meve, 2018). The subtribe emerged and began to diversify in the Oligocene in Africa, although most of its diversification followed a rate shift near the mid-Miocene (**Figures 3**, **4**). Although Liede-Schumann et al. (2012) estimated a younger origin for this group, the hypothesized major diversification of *Vincetoxicum* in Asia starting at the mid-Miocene climatic optimum agrees with the estimate obtained here. Both *Pentatropis* and *Vincetoxicum* are inferred to have dispersed to Asia, where *Vincetoxicum* radiated and subsequently dispersed to Europe (Liede-Schumann et al., 2012, 2016). The sequence of short branches with low support due to pulses of rapid diversification makes ancestral area reconstructions uncertain for recent events, but is consistent with this scenario.

Like Cynanchinae and Tylophorinae, Asclepiadinae (364 spp.) is inferred to have originated and begun to diversify in Africa, in this case more recently, in the late Miocene (**Figure 3**). Most of the species belong to the *Asclepias* s.l. radiation, which may consist of up to 400 species in Africa and the Americas. The origin of *Asclepias* s.l. coincides with a shift in diversification rates (**Figure 4**), which resulted in the largest temperate radiation in the family. The African and American species probably form sister clades restricted to each area (Fishbein et al., 2011; Chuba et al., 2017), but this dichotomy is not resolved in the large phylogeny of Fishbein et al. (2018) used here. The dispersal of Asclepiadinae to North America in the late Miocene and

subsequent diversification were likely favored by the expansion of grasslands in Africa and North America (Fishbein et al., 2011). The lineage would have reached North America through the Bering Strait (Rapini et al., 2007; Fishbein et al., 2011), which would have been available as a route for subtropical plants to the late Miocene (Graham, 2011, 2018), suggesting that the lineage was present in Asia and became extinct during the cooler Quaternary, without leaving a trace of this dispersal route. After diversifying in North America, *Asclepias* dispersed to South America either by a stepping-stone route across the Central America Seaway (Rapini et al., 2007) or after the complete emergence of the Panama Isthmus and subsequent expansion of savannas (Bacon et al., 2016).

Potential Key Drivers of Apocynaceae Diversification

In Apocynaceae, only a diversification rate shift within Tabernaemontaneae (**Figure 4**) might be associated with a shift between biotic dispersal modes, from seeds embedded in the pulp of indehiscent fruits to exposed arillate seeds in dehiscent fruits (shift from fleshy to dry fruits in Fishbein et al., 2018). This shift is associated with a change in the dispersal propagule, which is the entire fruit in berries and each seed with the associated aril in follicles. Fruits and dispersal modes are heterogeneous in rauvolfioid lineages, but follicles producing wind-dispersed comose seeds are highly conservative in the APSA clade. Simões et al. (2007) recovered a dry dehiscent fruit as ancestral in Apocynaceae (see also Zhang et al., 2020), from which berries and drupes derived several times in rauvolfioid lineages. The apparent phylogenetic conservatism of dry fruits coupled with the great flexibility of fruits in rauvolfioids, however, led Fishbein et al. (2018) to estimate at least five origins of dry fruits from rauvolfioid ancestors with fleshy fruits. Among the several groups that produce dry fruits in the family, significant accelerations of diversification are concentrated in only one, the APSA clade. Eleven of the 12 rate shifts in the family occurred in this essentially anemochorous lineage (**Figure 4**). The 20-million-years lag between the appearance of comose seeds in the APSA clade, near the Cretaceous-Paleogene boundary, and diversification rate shifts after the mid-Eocene does not suggest a direct effect of dispersal modes in the Apocynaceae diversification. However, the impact of dispersal syndromes in diversification rates is circumstantial (Herrera, 1989; Tiffney and Mazer, 1995) and the higher capacity of dispersal of comose seeds outside wet forests might have pushed the expansion of the APSA clade in open habitats and indirectly favored its diversification in the drier Miocene.

The ancient apocynoid radiation in the Early Eocene was accompanied by a shift from self-supporting (trees) to climbing habit (lianas and vines) in the APSA clade (Fishbein et al., 2018). Anatomical evolutionary trends in wood, such as the reduction of vessel element length and formation of larger vessel clusters, marked the transition from erect to climbing habit and might have favored the occupation of drier habitats (Lens et al., 2008, 2009; Endress et al., 2018-2019). The wind-dispersed comose seeds, then, might have favored twining plants to disperse and occupy open and semi-arid habitats. Twining plants seem to have greater capacities for the uptake and transportation of water and are more competitive than trees in high-light, nutrient-rich environments, such as the seasonally dry forests (Medina-Vega et al., 2021). Xylem adaptations to drought and cold show similar trends and possibly contributed to rapid diversification events in several lineages of angiosperms in association with the global cooling and aridification, especially after the mid-Miocene climatic optimum (Folk et al., 2020).

The gynostegium is a synapomorphy of the APSA clade (Endress et al., 2018-2019) and a precursor (sensu Donoghue and Sanderson, 2015) of pollinaria (specialized pollen dispersal units; Endress, 2016). Each pollinarium carries the content of two or four pollen sacs in tetrads or, more often, packaged in pollinia. Otherwise known only in Orchidaceae, pollinia evolved more than once in the APSA clade (Straub et al., 2014) and there were no reversals (Fishbein et al., 2018). They confer higher pollen transfer efficiency to pollinia-bearing lineages, as a single pollinium transfer in this group can produce many more seeds than a pollination by means of single pollen grains (free monads). Pollinia-bearing lineages are thus thought to be less impacted by lower densities of mates and pollinators (mate-finding Allee effect), and to have been positively selected for during the African aridification (Livshultz et al., 2011).

Rather than a single key innovation, the high diversity of Apocynaceae can be mostly explained by a sequence of interacting innovations (i.e., a synnovation, sensu Donoghue and Sanderson, 2015) that conferred higher capacity to disperse and establish in dry, open, and unstable habitats, in agreement with Stebbins' classic paper on aridity as a stimulus to evolution (Stebbins, 1952). Wind-dispersed comose seeds, a twining growth form, and pollinia appeared sequentially and probably work synergistically in the occupation of drier habitats, shaping most of the extant diversity of Apocynaceae. The long-term global decline of atmospheric CO_2, temperature, and humidity after the Eocene-Oligocene climate transition promoted the expansion of temperate regions and the establishment of grass and succulent biomes during the Miocene (e.g., Schrire et al., 2005, 2009; Simon et al., 2009; Arakaki et al., 2011; Gagnon et al., 2019; Folk et al., 2020), and fostered the geographic expansion of the APSA clade. Wider distributions may then have fueled accelerations in diversification, for example, due to more opportunities for the occupation of new habitats and specialization or environmental changes and fragmentation (Vamosi and Vamosi, 2011; Nürk et al., 2020). Following the terminology of Bouchenak-Khelladi et al. (2015) and Donoghue and Sanderson (2015), we hypothesize that the diversification of the APSA clade was triggered by the key confluence of a background synnovation and the long-term environmental changes since the Eocene-Oligocene climatic transition, which might also have served as modulators of diversification in the family.

Apocynaceae Biogeography and Diversification: The Big Picture

This is the first investigation of the biogeography and the tempo of diversification in Apocynaceae to consider all major lineages

within the family. The cumulative noise caused by dispersals and extinctions may add uncertainty to ancestral area estimates at deep nodes, and uncertainties remain on the area of origin for Apocynaceae. However, tropical South America, Southeast Asia, and Africa have retained the oldest lineages since the early diversification (**Figure 3**) and can be considered museums of Apocynaceae diversity (Stebbins, 1974). Africa was confirmed as the cradle of pollinia-bearing lineages and the main source of Apocynaceae intercontinental dispersals (**Figure 3** and **Table 3**). From Africa, Apocynaceae may have reached Asia by birds across the Indian Ocean (e.g., *Petchia*) or used land pathways (e.g., oceanic coastlines, as discussed for stem-succulent *Cynanchum* spp.; Meve and Liede, 2002b), either the boreotropical flora (before the Miocene) or the Arabian Peninsula (in the early and mid-Miocene for lineages inhabiting forests or in the late Miocene, after the expansion of savannas and semi-arid environments, for more xeric lineages; Chen et al., 2019). From tropical Asia, Apocynaceae dispersed to Oceania, temperate Eurasia (e.g., *Vincetoxicum*; Liede-Schumann et al., 2016), and North America, possibly through the Bering Strait. As discussed above, this northern subtropical to temperate route was probably used by *Amsonia* (Amsonieae), *Apocynum* (Apocyneae), *Cynanchum* (Cynanchinae), and *Asclepias* (Asclepiadinae). Most of the Apocynaceae diversity in the New World, however, dispersed from tropical Africa to South America, possibly *via* transatlantic dispersal or, before the mid-Eocene, also through the boreotropical flora. Two of the dispersals resulted in the largest American clades, the apocynoid EOM and the asclepiad MOG, but the rauvolfioid tribes Alyxieae and Plumerieae as well as the apocynoid tribes Malouetieae and Rhabdadenieae also reached South America from Africa (**Supplementary Figure 2**). Migrations from the New to Old World were rare, mostly restricted to the rauvolfioid tribes Willughbeieae and Tabernaemontaneae and the apocynoid EOM clade. Exchanges between South and North American Apocynaceae spanned most of the family history, at least since the Eocene, suggesting that the Greater and Lesser Antilles acted as plant sources for continental America (Liede-Schumann et al., 2014; Nieto-Blázquez et al., 2017) and that narrow seaways are not an effective barrier for plants (Joyce et al., 2021).

This Apocynaceae saga can be roughly divided in two major momenta ecologically driven by the increasingly dry Cenozoic climate and a sequence of key morphological innovations. Diversification rate shifts are concentrated after the Eocene-Oligocene climate transition, almost exclusively in lineages of the APSA clade, and mainly in groups with pollinia. From ancestors in closed tropical rainforests, the Apocynaceae phylogenetic backbone followed a general escalation toward the highly diverse wind-dispersed, twining, pollinia-bearing lineages in more stressful, unstable habitats. The increase of diversification rates in Apocynaceae, therefore, was probably stimulated by an evolutionary sequence of morphological innovations: (1) fixation of seeds with a micropylar coma (anemochory) and (2) transition from a self-supported to a climbing habit during the apocynoid early diversification, followed by (3) the advent of pollinia, first in Asclepiadoideae-Secamonoideae and, more recently, in Periplocoideae. From the Cretaceous through most of the Eocene, the climate was warmer, with a lower latitudinal gradient, allowing forests to spread over much of the Earth during the more humid Eocene. In this greenhouse world, Apocynaceae was mainly distributed in austro- and boreotropical rainforests. Following an abrupt decline of atmospheric CO_2 after the Eocene-Oligocene climate transition, global temperature and humidity dropped, temperate and seasonally dry habitats expanded, while rainforests retracted to lower latitudes, mainly in the Southern Hemisphere. In this cooling world, the diversity of Apocynaceae was mainly dominated by the wind-dispersed, twining, pollinia-bearing lineages in more unstable seasonal and semi-arid habitats, whereas most rauvolfioid and ancient apocynoid lineages found refuge mainly in stable tropical forests.

The approaches assumed here are not immune to, for example, biases introduced by species sampling and model choice, which can affect results. In addition, traits apart from the evolutionary sequence toward pollinia-bearing lineages may also have effects on diversification of Apocynaceae, especially at more recent phylogenetic scales. Biotic interactions, in particular associated with the complex evolution of chemical defenses against herbivory (e.g., Agrawal et al., 2009; Livshultz et al., 2018), may be additionally involved in the diversification history of the family. Hence, although the big picture shown here may represent an oversimplification of Apocynaceae evolution, we believe that it nevertheless serves as a useful roadmap to highlight major biogeographic patterns and outline potential key drivers to higher diversification rates, which are now open to be tested.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**/DRYAD repository (https://doi.org/10.5061/dryad.vq83bk3sj), further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The research reported in this paper was exempt from ethical approval procedures.

AUTHOR CONTRIBUTIONS

CB and AR conceived the study and prepared the dataset, which was revised by all authors. CB and NN designed the analyses with contributions from AR. NN ran the analyses. CB, NN, and AR interpreted the results and prepared an initial draft. All authors revised the manuscript critically for intellectual content.

FUNDING

CB acknowledges the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code #1514632 (DS), an International Association of Plant Taxonomy (IAPT, 2016) grant, and the Neotropical Grasslands Conservancy Memorial (NGC, 2016) and Shirley A. Graham (Missouri

Botanical Garden, 2018) Fellowships. AR is supported by CNPQ (Productivity Fellowship no. 307396/2019-3). NN acknowledges the support of the Deutsche Forschungsgemeinschaft (DFG grant number NU292/4-1). This publication was funded by the German Research Foundation (DFG) and the University of Bayreuth in the funding program Open Access Publishing.

ACKNOWLEDGMENTS

This work is part of the Ph.D. thesis of CB, developed at Programa de Pós-Graduação em Botânica da Universidade Estadual de Feira de Santana. We would like to thank Moabe Fernandes for assistance with the dataset, Luciano Paganucci de Queiroz for suggesting our study as a potential contribution to this research topic, Danilo Neves for inviting us to join the article collection "Temporal and Large-Scale Spatial Patterns of Diversity and Diversification" at Frontiers in Ecology and Evolution, and the two reviewers for insightful suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.719741/full#supplementary-material

Supplementary Table 1 | Species data: taxon names, clade assignment, species richness, sampling fractions, and species distribution.

Supplementary Table 2 | Biogeographic analyses: model comparison results (a: considered; b: incl. + J).

Supplementary Table 3 | Biogeographic test: model definition.

Supplementary Table 4 | BSM: dispersals estimated under the M1s.DEC + J model.

Supplementary Table 5 | Biogeographic tests: model comparison results (all test models).

Supplementary Table 6 | Diversification rates analysis (MiSSE): model comparison results.

Supplementary Methods 1 | Biogeographic analysis: Delimitation of areas; Paleogeographic model; References.

Supplementary Figure 1 | Paleogeographic model: used in the time-stratified analysis (M2, M2s models).

Supplementary Figure 2 | Biogeographic analyses: ancestral area estimates produced by the M1s.DEC model showing **(A)** the best estimates, and **(B)** the state probabilities.

Supplementary Figure 3 | Biogeographic analyses: ancestral area estimates produced by the M1.DEC model, showing **(A)** the best estimates, and **(B)** the state probabilities.

Supplementary Figure 4 | Biogeographic analyses: ancestral area estimates produced by the M1.DEC + j model, showing **(A)** the best estimates, and **(B)** the state probabilities.

Supplementary Figure 5 | Biogeographic analyses: ancestral area estimates produced by the M1s.DEC + j dec model, showing **(A)** the best estimates, and **(B)** the state probabilities.

Supplementary Figure 6 | Diversification rate shifts analysis (BAMM): Prior and posterior distributions of number of shifts.

REFERENCES

Agrawal, A. A. (2017). *Monarchs and Milkweed: A Migrating Butterfly, A Poisonous Plant, and Their Remarkable Story of Coevolution*. Princeton, NJ: Princeton University Press.

Agrawal, A. A., Fishbein, M., Halitschke, R., Hastings, A. P., Rabosky, D. L., and Rasmann, S. (2009). Evidence for adaptive radiation from a phylogenetic study of plant defenses. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18067–18072. doi: 10.1073/pnas.0904862106

Agrawal, A. A., Petschenka, G., Bingham, R. A., Weber, M. G., and Rasmann, S. (2012). Toxic cardenolides: chemical ecology and coevolution of specialized plant–herbivore interactions. *New Phytol.* 194, 28–45. doi: 10.1111/j.1469-8137.2011.04049.x

Antonelli, A., Clarkson, J. J., Kainulainen, K., Maurin, O., Brewer, G. E., Davis, A. P., et al. (2021). Settling a family feud: a high-level phylogenomic framework for the Gentianales based on 353 nuclear genes and partial plastomes. *Am. J. Bot.* 108, 1143–1165. doi: 10.1002/ajb2.1697

Arakaki, M., Christin, P.-A., Nyffeler, R., Lendel, A., Eggli, U., Ogburn, R. M., et al. (2011). Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8379–8384. doi: 10.1073/pnas.1100628108

Axelrod, D. I., and Raven, P. H. (1978). "Late cretaceous and tertiary vegetation history of Africa," in *Biogeography and Ecology of Southern Africa*, eds M. J. A. Werger and A. C. van Bruggen (The Hague: Dr W. Junk bv Publishers), 77–130. doi: 10.1007/978-94-009-9951-0_5

Bacon, C. D., Molnar, P., Antonelli, A., Crawford, A. J., Montes, C., and Vallejo-Pareja, M. C. (2016). Quaternary glaciation and the Great American Biotic Interchange. *Geology* 44, 375–378. doi: 10.1130/G37624.1

Beaulieu, J. M., and O'Meara, B. C. (2016). Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Syst. Biol.* 65, 583–601. doi: 10.1093/sysbio/syw022

Bouchenak-Khelladi, Y., Onstein, R. E., Xing, Y., Schwery, O., and Linder, H. P. (2015). On the complexity of triggering evolutionary radiations. *New Phytol.* 207, 313–326. doi: 10.1111/nph.13331

Bruyns, P. V. (2000). Phylogeny and biogeography of stapeliads. *Plant Syst. Evol.* 221, 199–226. doi: 10.1007/bf01089294

Bruyns, P. V., Klak, C., and Hanáèek, P. (2014). Evolution of the stapeliads (Apocynaceae – Asclepiadoideae) – repeated major radiation across Africa in an Old World group. *Mol. Phylogen. Evol.* 77, 251–263. doi: 10.1016/j.ympev.2014.03.022

Bruyns, P. V., Klak, C., and Hanáèek, P. (2015). Recent radiation of *Brachystelma* and *Ceropegia* (Apocynaceae) across the Old World against a background of climatic change. *Mol. Phylogen. Evol.* 90, 49–66. doi: 10.1016/j.ympev.2015.04.015

Bruyns, P. V., Klak, C., and Hanáèek, P. (2017). A revised, phylogenetically-based concept of *Ceropegia* (Apocynaceae). *S. Afr. J. Bot.* 112, 399–436. doi: 10.1016/j.sajb.2017.06.021

Burnham, K. P., and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York, NY: Springer-Verlag.

Caetano, D. S., O'Meara, B. C., and Beaulieu, J. M. (2018). Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution* 72, 2308–2324. doi: 10.1111/evo.13602

Chen, J., Thomas, D. C., and Saunders, R. M. K. (2019). Geographic range and habitat reconstructions shed light on palaeotropical intercontinental disjunction and regional diversification patterns in *Artabotrys* (Annonaceae). *J. Biogeogr.* 46, 2690–2705. doi: 10.1111/jbi.13703

Christenhusz, M. J. M., and Byng, J. W. (2016). The number of known plants species in the world and its annual increase. *Phytotaxa* 261, 201–217. doi: 10.11646/phytotaxa.261.3.1

Chuba, D., Goyder, D. J., Chase, J. M., and Fishbein, M. (2017). Phylogenetics of the African *Asclepias* complex (Apocynaceae) based on three plastid DNA regions. *Syst. Bot.* 42, 148–159. doi: 10.1600/036364417X694539

Collinson, M. E., Manchester, S. R., Wilde, V., and Hayes, P. (2010). Fruit and seed floras from exceptionally preserved biotas in the European Paleogene. *Bull. Geosci.* 85, 155–162. doi: 10.3140/bull.geosci.1155

Cronquist, F. (1997). Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Syst. Biol.* 46, 195–203. doi: 10.1093/sysbio/46.1.195

Davies, T. J., Barraclough, T. G., Chase, M. W., Soltis, P. S., Soltis, D. E., and Savolainen, V. (2004). Darwin's abominable mystery: insights from a supertree of the angiosperms. *Proc. Natl. Acad. Sci. U.S.A.* 101, 1904–1909. doi: 10.1073/pnas.0308127100

Davis, C. C., Bell, C. D., Mathews, S., and Donoghue, M. J. (2002). Laurasian migration explains Gondwanan disjunctions: evidence from Malpighiaceae. *Proc. Natl. Acad. Sci. U.S.A.* 99, 6833–6837. doi: 10.1073/pnas.102175899

de Queiroz, A. (2005). The resurrection of oceanic dispersal in historical biogeography. *Trends Ecol. Evol.* 20, 68–73. doi: 10.1016/j.tree.2004.11.006

Del Rio, C., Wang, T.-X., Liu, J., Liang, S.-Q., Spicer, R. A., Wu, F.-X., et al. (2020). *Asclepiadospermum* gen. nov., the earliest fossil record of Asclepiadoideae (Apocynaceae) from the early Eocene of central Qinghai-Tibetan Plateau, and its biogeographic implications. *Am. J. Bot.* 107, 126–138. doi: 10.1002/ajb2.1418

Donoghue, M. J., and Sanderson, M. J. (2015). Confluence, synnovation, and depauperons in plant diversification. *New Phytol.* 207, 260–274. doi: 10.1111/nph.13367

Dupin, J., Matzke, N. J., Särkinen, T., Knapp, S., Olmstead, R., Bohs, L., et al. (2016). Bayesian estimation of the global biogeographic history of the Solanaceae. *J. Biogeogr.* 44, 887–899. doi: 10.1111/jbi.12898

Edgar, J. A., Boppré, M., and Kaufmann, E. (2007). Insect-synthesised retronecine ester alkaloids: precursors of the common arctiine (Lepidoptera) pheromone hydroxydanaidal. *J. Chem. Ecol.* 33, 2266–2280. doi: 10.1007/s10886-007-9378-y

Endress, M. E., Meve, U., Middleton, D. J., and Liede-Schumann, S. (2018-2019). "Apocynaceae," in *Flowering Plants. Eudicots. Apiales and Gentianales (Except Rubiaceae) Families and Genera of Vascular Plants*, Vol. 15, eds J. W. Kadereit and V. Bittrich (Berlin: Springer), 208–411. doi: 10.1007/978-3-319-93605-5_3

Endress, P. K. (2016). Development and evolution of extreme synorganization in angiosperm flowers and diversity: a comparison of Apocynaceae and Orchidaceae. *Ann. Bot.* 117, 749–767. doi: 10.1093/aob/mcv119

Espírito Santo, F. S., Rapini, A., Ribeiro, P. L., Liede-Schumann, S., Goyder, D. J., and Fontella-Pereira, J. (2019). Phylogeny of the tribe Marsdenieae (Apocynaceae), reinstatement of *Ruehssia* and the taxonomic treatment of the genus in Brazil. *Kew Bull.* 74:30. doi: 10.1007/s12225-019-9807-4

Fishbein, M., Chuba, D., Ellison, C., Mason-Gamer, R. J., and Lynch, S. P. (2011). Phylogenetic relationships of *Asclepias* (Apocynaceae) inferred from non-coding chloroplast DNA sequences. *Syst. Bot.* 36, 1008–1023. doi: 10.1600/036364411x605010

Fishbein, M., Livshultz, T., Straub, S. C. K., Simões, A. O., Boutte, J., McDonnell, A., et al. (2018). Evolution on the backbone: Apocynaceae phylogenomics and new perspectives on growth forms, flowers, and fruits. *Am. J. Bot.* 105, 495–513. doi: 10.1002/ajb2.1067

Folk, R. A., Siniscalchi, C. M., and Soltis, D. E. (2020). Angiosperms at the edge: extremity, diversity, and phylogeny. *Plant Cell Environ.* 43, 2871–2893. doi: 10.1111/pce.13887

Friis, E. M., Crane, P., and Pedersen, K. (2011). *Early Flowers and Angiosperm Evolution*. Cambridge: Cambridge University Press.

Gagnon, E., Ringelberg, J. J., Bruneau, A., Lewis, G. P., and Hughes, C. E. (2019). Global succulent biome phylogenetic conservatism across the pantropical *Caesalpinia* group (Leguminosae). *New Phytol.* 222, 1994–2008. doi: 10.1111/nph.15633

Goyder, D. J. (2006). "An overview of asclepiad biogeography," in *Taxonomy and Ecology of African Plants, Their Conservation and Sustainable Use*, eds S. A. Ghazanfar and H. J. Beentje (Kew: Royal Botanic Gardens), 205–214.

Graham, A. (2011). The age and diversification of terrestrial New World ecosystems through Cretaceous and Cenozoic time. *Am. J. Bot.* 98, 336–351. doi: 10.3732/ajb.1000353

Graham, A. (2018). *Land Bridges: Ancient Environments, Plant Migration and New World Connections*. Chicago, IL: Chicago Press.

Hernández-Hernández, T., and Wiens, J. J. (2020). Why are there so many flowering plants? A multiscale analysis of plant diversification. *Am. Nat.* 195, 948–963. doi: 10.1086/708273

Herrera, C. M. (1989). Seed dispersal by animals: a role in angiosperm diversification? *Am. Nat.* 133, 309–322. doi: 10.1086/284921

Igea, J., and Tanentzap, A. J. (2020). Angiosperm speciation cools down in the tropics. *Ecol. Lett.* 23, 692–700. doi: 10.1111/ele.13476

Ionta, G. M., and Judd, W. S. (2017). Phylogenetic relationships in Periplocoideae (Apocynaceae s.l.) and insights into the origin of pollinia. *Ann. Missouri Bot. Gard.* 94, 360–375. doi: 10.3417/0026-6493(2007)94[360:pripas]2.0.co;2

Jacobs, B. F. (2004). Palaeobotanical studies from tropical Africa: relevance to the evolution of forest, woodland and savannah biomes. *Philos. Trans. R. Soc. Lond. B* 359, 1573–1583. doi: 10.1098/rstb.2004.1533

Janzen, D. H. (1974). Epiphytic myrmecophytes in Sarawak: mutualism through the feeding of plants by ants. *Biotropica* 6, 237–259. doi: 10.2307/2989668

Joyce, E. M., Thiele, K. R., Slik, J. W. F., and Crayn, D. M. (2021). Plants will cross the lines: climate and available land mass are the major determinants of phytogeographical patterns in the Sunda–Sahul convergence zone. *Biol. J. Linn. Soc.* 132, 374–387. doi: 10.1093/biolinnean/blaa194

Khanum, R., Surveswaran, S., Meve, U., and Liede-Schumann, S. (2016). *Cynanchum* (Apocynaceae: Asclepiadoideae): a pantropical asclepiadoid genus revisited. *Taxon* 65, 467–486. doi: 10.12705/653.3

Kidyoo, A. (2019). *Heterostemma* ficoides (Apocynaceae: Asclepiadoideae), a new species with fig-like flowers from Northern Thailand. *Kew Bull.* 74:26. doi: 10.1007/S12225-019-9815-4

Kingdon, J., Agwanda, B., Kinnaird, M., O'Brien, T., Holland, C., Gheysens, T., et al. (2011). A poisonous surprise under the coat of the African crested rat. *Proc. R. Soc. B* 279, 675–680. doi: 10.1098/rspb.2011.1169

Lahaye, R. (2005). *Phylogenie Moleculaire des Secamonoideae (Apocynaceae s.l.): Histoire Biogéographique et Évolution des Formes de Croissance.* [Thèse de Doctorat en Systématique Végétale. Ecologie et Evolution]. Toulouse: Toulouse III.

Lahaye, R., Klackenberg, J., Källersjö, M., van Campo, E., and Civeyrel, L. (2007). Phylogenetic relationships between derived Apocynaceae s.l. and within Secamonoideae based on chloroplast sequences. *Ann. Missouri Bot. Gard.* 94, 376–391. doi: 10.3417/0026-6493(2007)94[376:prbdas]2.0.co;2

Landis, M. J., Matzke, N. J., Moore, B. R., and Huelsenbeck, J. P. (2013). Bayesian analysis of biogeography when the number of areas is large. *Syst. Biol.* 62, 789–804. doi: 10.1093/sysbio/syt040

Larridon, I., Díaz, J. G., Bauters, K., and Escudero, M. (2020). What drives diversification in a pantropical plant lineage with extraordinary capacity for long-distance dispersal and colonization? *J. Biogeogr.* 48, 64–77. doi: 10.1111/jbi.13982

Lens, F., Endress, M. E., Baas, P., Jansen, S., and Smets, E. (2008). Wood anatomy of Rauvolfioideae (Apocynaceae): a search for meaningful non-DNA characters at the tribal level. *Am. J. Bot.* 95, 1199–1215. doi: 10.3732/ajb.0800159

Lens, F., Endress, M. E., Baas, P., Jansen, S., and Smets, E. (2009). Vessel grouping patterns in subfamilies Apocynoideae and Periplocoideae confirm phylogenetic value of wood structure within Apocynaceae. *Am. J. Bot.* 96, 2168–2183. doi: 10.3732/ajb.0900116

Liede, S., and Täuber, A. (2002). Circumscription of the genus *Cynanchum* (Apocynaceae-Asclepiadoideae). *Syst. Bot.* 27, 789–800. doi: 10.1043/0363-6445-27.4.789

Liede-Schumann, S., and Meve, U. (2018). *Vincetoxicum* (Apocynaceae—Asclepiadoideae) expanded to include *Tylophora* and allies. *Phytotaxa* 369, 129–184. doi: 10.11646/phytotaxa.369.3.1

Liede-Schumann, S., Khanum, R., Mumtaz, A. S., Gherghel, I., and Pahlevani, A. (2016). Going west – a subtropical lineage (*Vincetoxicum*, Apocynaceae: Asclepiadoideae) expanding into Europe. *Mol. Phylogen. Evol.* 94, 436–446. doi: 10.1016/j.ympev.2015.09.021

Liede-Schumann, S., Kong, H.-H., Meve, U., and Thiv, M. (2012). *Vincetoxicum* and *Tylophora* (Apocynaceae: Asclepiadoideae: Asclepiadeae) – two sides of the same medal: independent shifts from tropical to temperate habitats. *Taxon* 61, 803–825. doi: 10.1002/tax.614007

Liede-Schumann, S., Nikolaus, M., Soares e Silva, U. C., Rapini, A., Mangelsdorff, R. D., and Meve, U. (2014). Phylogenetics and biogeography of the genus *Metastelma* (Apocynaceae-Asclepiadoideae-Asclepiadeae: Metastelmatinae). *Syst. Bot.* 39, 594–612. doi: 10.1600/036364414x680708

Linder, H. P., and Crisp, M. D. (1995). *Nothofagus* and Pacific biogeography. *Cladistics* 11, 5–32. doi: 10.1111/j.1096-0031.1995.tb00002.x

Livshultz, T. (2010). The phylogenetic position of milkweeds (Apocynaceae subfamilies Secamonoideae and Asclepiadoideae): evidence from the nucleus and chloroplast. *Taxon* 59, 1016–1030. doi: 10.1002/tax.594003

Livshultz, T., Kaltenegger, E., Straub, S. C. K., Weitemier, K., Hirsch, E., Koval, K., et al. (2018). Evolution of pyrrolizidine alkaloid biosynthesis in Apocynaceae: revisiting the defence de-escalation hypothesis. *New Phytol.* 218, 762–773. doi: 10.1111/nph.15061

Livshultz, T., Mead, J. V., Goyder, D. J., and Brannin, M. (2011). Climate niches of milkweeds with plesiomorphic traits (Secamonoideae; Apocynaceae) and the milkweed sister group link ancient African climates and floral evolution. *Am. J. Bot.* 98, 1966–1977. doi: 10.3732/ajb.1100202

Livshultz, T., Middleton, D. J., Endress, M. E., and Williams, J. (2007). Phylogeny of Apocynoideae and the APSA clade. *Ann. Missouri Bot. Gard.* 94, 323–361.

Magallón, S., and Sanderson, M. J. (2001). Absolute diversification rates in angiosperm clades. *Evolution* 55, 1762–1780. doi: 10.1554/0014-3820(2001)055[1762:adriac]2.0.co;2

Magallón, S., Sánchez-Reyes, L. L., and Gómez-Acevedo, S. L. (2019). Thirty clues to the exceptional diversification of flowering plants. *Ann. Bot.* 123, 491–503. doi: 10.1093/aob/mcy182

Martínez-Millán, M. (2010). Fossil record and age of the Asteridae. *Bot. Rev.* 76, 83–135. doi: 10.1007/s12229-010-9040-1

Matzke, N. J. (2013). *BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts*. R Package, Version 0.2.1. Available online at: http://CRAN.R-project.org/package=BioGeoBEARS (accessed July 27, 2013)

Matzke, N. J. (2016). *Stochastic Mapping Under Biogeographical Models*. Available online at: http://phylo.wikidot.com/bio-geobears#stochastic_mapping (accessed January 2, 2021)

Matzke, N. J. (2021). Statistical comparison of DEC and DEC+J is identical to comparison of two ClaSSE submodels, and is therefore valid. *OSF [Preprints]* doi: 10.31219/osf.io/vqm7r

McElwain, J. C., and Punyasena, S. W. (2007). Mass extinction events and the plant fossil record. *Trends Ecol. Evol.* 22, 548–557. doi: 10.1016/j.tree.2007.09.003

McLoughlin, S. (2001). The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Aust. J. Bot.* 49, 271–300. doi: 10.1071/bt00023

Medina-Vega, J. A., Bongers, F., Poorter, L., Schitzer, S. A., and Sterck, F. J. (2021). Lianas have more acquisitive traits than trees in a dry but not in a wet forest. *J. Ecol.* 109, 2367–2384. doi: 10.1111/1365-2745.13644

Mendel, J. R., Dikow, R. B., Siniscalchi, C. M., Thapa, R., Watson, L. E., and Funk, V. A. (2019). A fully resolved backbone phylogeny reveals numerous dispersals and explosive diversifications throughout the history of Asteraceae. *Proc. Natl. Acad. Sci. U.S.A.* 116, 14083–14088. doi: 10.1073/pnas.1903871116

Meseguer, A. S., and Condamine, F. L. (2020). Ancient tropical extinctions at high latitudes contributed to the latitudinal diversity gradient. *Evolution* 74, 1966–1987. doi: 10.1111/evo.13967

Meseguer, A. S., Lobo, J. M., Cornuault, J., Beerling, D., Ruhfel, B. R., Davis, C. C., et al. (2018). Reconstructing deep-time palaeoclimate legacies in the clusioid Malpighiales unveils their role in the evolution and extinction of the boreotropical flora. *Global Ecol. Biogeogr.* 27, 616–628. doi: 10.1111/geb.12724

Meseguer, A. S., Lobo, J. M., Ree, R. H., Beerling, D. J., and Sanmartin, I. (2015). Integrating fossils, phylogenies, and niche models into biogeography to reveal ancient evolutionary history: the case of *Hypericum* (Hypericaceae). *Syst. Biol.* 64, 215–232. doi: 10.1093/sysbio/syu088

Meve, U., and Liede, S. (2002a). A molecular phylogeny and generic rearrangement of the stapelioid Ceropegieae (Apocynaceae-Asclepiadoideae). *Plant Syst. Evol.* 234, 171–209. doi: 10.1007/s00606-002-0220-2

Meve, U., and Liede, S. (2002b). Floristic exchange between mainland Africa and Madagascar: case studies in Apocynaceae-Asclepiadoideae. *J. Biogeogr.* 29, 865–873. doi: 10.1046/j.1365-2699.2002.00729.x

Meve, U., Heiduk, A., and Liede-Schumann, S. (2016). Origin and early evolution of Ceropegieae (Apocynaceae-Asclepiadoideae). *Syst. Biodivers.* 15, 143–155. doi: 10.1080/14772000.2016.1238019

Morley, R. J. (1998). "Palynological evidence for tertiary plant dispersal in the SE Asian region in relation to plate tectonics and climate," in *Biogeography and Geological Evolution of SE Asia*, eds R. Hall and J. D. Halloway (Leiden: Backhuys), 211–234.

Morley, R. J. (2000). *Origin and Evolution of Tropical Rain Forests*. Chichester: Wiley.

Morley, R. J. (2003). Interplate dispersal paths for megathermal angiosperms. *Perspect. Plant Ecol. Evol. Syst.* 6, 5–20. doi: 10.1078/1433-8319-00039

Movshovich, E. B. (1975). Fundamentals of elaboration of a general classification for gas and oil traps. *Izv. Akad. Nauk SSSR Ser. Geol.* 5, 89–98.

Muller, J. (1968). Palynology of the Pedawan and Plateau sandstone formations (Cretaceous-Eocene) in Sarawak, Malaysia. *Micropaleontology* 14, 1–37. doi: 10.2307/1484763

Nieto-Blázquez, M. E., Antonelli, A., and Roncal, J. (2017). Historical biogeography of endemic seed plant genera in the Caribbean: did GAARlandia play a role? *Ecol. Evol.* 23, 10158–10174. doi: 10.1002/ece3.3521

Nürk, N. M., Linder, H., Onstein, R., Larcombe, M., Hughes, C., Piñeiro Fernández, L., et al. (2020). Diversification in evolutionary arenas: assessment and synthesis. *Ecol. Evol.* 10, 6163–6182. doi: 10.1002/ece3.6313

Nürk, N. M., Michling, F., and Linder, P. (2018). Are the radiations of temperate lineages in tropic-alpine ecosystems pre-adapted? *Glob. Ecol. Biogeogr.* 27, 334–345. doi: 10.1111/geb.12699

Nürk, N. M., Uribe-Convers, S., Gehrke, B., Tank, D. C., and Blattner, F. (2015). Oligocene niche shift, Miocene diversification – cold tolerance and accelerated speciation rates in the St. John's Worts (*Hypericum*, Hypericaceae). *BMC Evol. Biol.* 15:80. doi: 10.1186/s12862-015-0359-4

Ollerton, J., Liede-Schumann, S., Endress, M. E., Meve, U., Rech, A. R., Shuttleworth, A., et al. (2019). The diversity and evolution of pollination systems in large plant clades: Apocynaceae as a case study. *Ann. Bot.* 123, 311–325. doi: 10.1093/aob/mcy127

Pauw, A. (1998). Pollen transfer on bird's tongues. *Nature* 394, 731–732. doi: 10.1038/29421

Peeters, C., and Wiwatwitaya, D. (2014). *Philidris* ants living inside *Dischidia* epiphytes from Thailand. *Asian Myrmecol.* 6, 49–61.

Pirie, M. D., Kandziora, M., Nürk, N. M., Le Maitre, N. C., de Kuppler, A. M., Gehrke, B., et al. (2019). Leaps and bounds: geographical and ecological distance constrained the colonisation of the Afrotemperate by *Erica*. *BMC Evol. Biol.* 19:222. doi: 10.1186/s12862-019-1545-6

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R News* 6, 7–11.

Poinar, G. O. (2017). Ancient termite pollinator of milkweed flowers in Dominican amber. *Am. Entomol.* 63, 52–56. doi: 10.1093/ae/tmx011

Procheş, Ş, and Ramdhani, S. (2020). A global regionalisation based on the present-day distribution of broad plant lineages. *Phytotaxa* 442, 20–26. doi: 10.11646/phytotaxa.442.1.3

Pugliesi, L., and Rapini, A. (2015). Tropical refuges with exceptionally high phylogenetic diversity reveal contrasting phylogenetic structures. *Int. J. Biodivers.* 2015:e758019. doi: 10.1155/2015/758019

Rabosky, D. L. (2014). Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS One* 9:e89543. doi: 10.1371/journal.pone.0089543

Rabosky, D. L., Grundler, M., Anderson, C., Title, P., Shi, J. J., Brown, J. W., et al. (2014). BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods Ecol. Evol.* 5, 701–707. doi: 10.1111/2041-210X.12199

Ramírez-Barahona, S., Sauquet, H., and Magallón, S. (2020). The delayed and geographically heterogeneous diversification of flowering plant families. *Nat. Ecol. Evol.* 4, 1232–1238. doi: 10.1038/s41559-020-1241-3

Rapini, A., Bitencourt, C., Luebert, F., and Cardoso, D. (2021). An escape-to-radiate model for explaining the high plant diversity and endemism in campos rupestres. *Biol. J. Linn. Soc.* 133, 481–498. doi: 10.1093/biolinnean/blaa179

Rapini, A., van den Berg, C., and Liede-Schumann, S. (2007). Diversification of Asclepiadoideae (Apocynaceae) in the New World. *Ann. Missouri Bot. Gard.* 94, 407–422. doi: 10.3417/0026-6493(2007)94[407:doaait]2.0.co;2

Ree, R. H., and Smith, S. A. (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst. Biol.* 57, 4–14. doi: 10.1080/10635150701883881

Ree, R. H., Moore, B. R., Webb, C. O., and Donoghue, M. J. (2005). A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution* 59, 2299–2311. doi: 10.1111/j.0014-3820.2005.tb00940.x

Renner, S. (2004). Plant dispersal across the tropical Atlantic by wind and sea currents. *Int. J. Plant Sci.* 165(4 Suppl.), S23–S33. doi: 10.1086/383334

Ribeiro, P. L., Rapini, A., Damascena, L. S., and van den Berg, C. (2014). Plant diversification in the Espinhaço range: insights from the biogeography of *Minaria* (Apocynaceae). *Taxon* 63, 1253–1264. doi: 10.12705/636.16

Richardson, J. E., Chatrou, L. W., Mols, J. B., Erkens, R. H. J., and Pirie, M. D. (2004). Historical biogeography of two cosmopolitan families of flowering plants: Annonaceae and Rhamnaceae. *Philos. Trans. R. Soc. Lond. B* 359, 1495–1508. doi: 10.1098/rstb.2004.1537

Rose, J. P., Kleist, T. J., Löfstrand, S. D., Drewd, B. T., Schönenberger, J., and Sytsma, K. J. (2018). Phylogeny, historical biogeography, and diversification of angiosperm order Ericales suggest ancient Neotropical and East Asian connections. *Mol. Phylogen. Evol.* 122, 59–79. doi: 10.1016/j.ympev.2018.01.014

Ruhfel, B. R., Bove, C. P., Philbrick, C. T., and Davis, C. C. (2016). Dispersal largely explains the Gondwanan distribution of the ancient tropical clusioid plant clade. *Am. J. Bot.* 103, 1117–1128. doi: 10.3732/ajb.1500537

Salard-Cheboldaeff, M. (1978). Sur la palynoflore maestrichtiene et tertiaires du Bassin sédimentaire littoral du Cameroun. *Pollen Spores* 20, 215–260.

Sanmartín, I., and Ronquist, F. (2004). Southern Hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Syst. Biol.* 53, 216–243. doi: 10.1080/10635150490423430

Schrire, B. D., Lavin, M., Nigel, P. B., and Forest, F. (2009). Phylogeny of the tribe Indigofereae (Leguminosae – Papilionoideae): geographically structured more in succulent-rich and temperate settings than in grass-rich environments. *Am. J. Bot.* 96, 816–852. doi: 10.3732/ajb.0800185

Schrire, B. D., Lewis, G. P., and Lavin, M. (2005). "Biogeography of the Leguminosae," in *Legumes of the World*, eds G. Lewis, B. Schrire, B. Mackinder, and M. Lock (Richmond, VA: Kew Publishing), 21–54.

Sennblad, B., and Bremer, B. (1996). The familial and subfamilial relationships of Apocynaceae and Asclepiadaceae evaluated data with *rbc*L data. *Plant Syst. Evol.* 202, 153–175. doi: 10.1007/bf00983380

Simões, A. O., Endress, M. E., van der Niet, T., Kinoshita, L. S., and Conti, E. (2006). Is *Mandevilla* (Apocynaceae, Mesechiteae) monophyletic? Evidence from five plastid DNA loci and morphology. *Ann. Missouri Bot. Gard.* 93, 565–591. doi: 10.3417/0026-6493(2006)93[565:imamme]2.0.co;2

Simões, A. O., Kinoshita, L. S., Koch, I., Silva, M. J., and Endress, M. E. (2016). Systematics and character evolution of Vinceae (Apocynaceae). *Taxon* 65, 99–122. doi: 10.12705/651.7

Simões, A. O., Livshultz, T., and Endress, M. E. (2007). Phylogeny and systematics of the Rauvolfioideae (Apocynaceae) based on molecular and morphological evidence. *Ann. Missouri Bot. Gard.* 94, 268–297. doi: 10.3417/0026-6493(2007)94[268:pasotr]2.0.co;2

Simon, M. F., Grether, R., Queiroz, L. P., Skema, C., Pennington, R. T., and Hughes, C. E. (2009). Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20359–20364. doi: 10.1073/pnas.0903410106

Singh, H., Judd, W. S., Samant, B., Agnihotri, P., Grimaldi, D. A., and Manchester, S. R. (2021). Flowers of Apocynaceae in amber from the early Eocene of India. *Am. J. Bot.* 108, 1–10. doi: 10.1002/ajb2.1651

Soltis, P. S., Folk, R. A., and Soltis, D. E. (2019). Darwin review: angiosperm phylogeny and evolutionary radiations. *Proc. R. Soc. B* 286, e2019.0099. doi: 10.1098/rspb.2019.0099

Stebbins, G. L. (1952). Aridity as a stimulus to plant evolution. *Am. Nat.* 86, 33–44. doi: 10.1086/281699

Stebbins, G. L. (1974). *Flowering Plants: Evolution above the Species Level*. Cambridge, MA: Harvard University Press.

Straub, S. C., Moore, M. J., Soltis, P. S., Soltis, D. E., Liston, A., and Livshultz, T. (2014). Phylogenetic signal detection from an ancient rapid radiation: effects of noise reduction, long-branch attraction, and model selection in crown clade Apocynaceae. *Mol. Phylogen. Evol.* 80, 169–185. doi: 10.1016/j.ympev.2014.07.020

Su, T., Spicer, R. A., Wu, F.-X., Farnsworth, A., Huang, J., Del Rio, C., et al. (2020). A middle Eocene lowland humid subtropical "Shangri-La" ecosystem in central Tibet. *Proc. Natl. Acad. Sci. U.S.A.* 117, 32989–32995. doi: 10.1073/pnas.2012647117

Sun, M., Folk, R. A., Gitzendanner, M. A., Soltis, P. S., Chen, Z., Soltis, D. E., et al. (2020). Recent accelerated diversification in rosids occurred outside the tropics. *Nat. Commun.* 11:3333. doi: 10.1038/s41467-020-17116-5

Tiffney, B. H. (1985a). Perspectives on the origin of the floristic similarity between eastern Asia and eastern North America. *J. Arnold Arbor.* 66, 73–94. doi: 10.5962/bhl.part.13179

Tiffney, B. H. (1985b). The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography of the Northern Hemisphere. *J. Arnold Arbor.* 66, 243–273. doi: 10.5962/bhl.part.13183

Tiffney, B. H., and Mazer, S. J. (1995). Angiosperm growth habit, dispersal and diversification reconsidered. *Evol. Ecol.* 9, 93–117. doi: 10.1007/bf01237700

Vamosi, J. C., and Vamosi, S. M. (2011). Factors influencing diversification in angiosperms: at the crossroads of intrinsic and extrinsic traits. *Am. J. Bot.* 98, 460–471. doi: 10.3732/ajb.1000311

van den Ende, C., White, L. T., and van Welzen, P. C. (2017). The existence and break-up of the Antarctic land bridge as indicated by both amphi-Pacific distributions and tectonics. *Gondwana Res.* 44, 219–227. doi: 10.1016/j.gr.2016.12.006

Wang, W., Ortiz, R. C., Jacques, F. M. B., Xiang, X.-G., Li, H.-L., Lin, L., et al. (2012). Menispermaceae and the diversification of tropical rainforests near the Cretaceous-Paleogene boundary. *New Phytol.* 195, 470–478. doi: 10.1111/j.1469-8137.2012.04158.x

Wheeler, E., Lee, M., and Matten, L. C. (1987). Dicotyledonous woods from the Upper Cretaceous of southern Illinois. *Bot. J. Linn. Soc.* 95, 77–100. doi: 10.1111/j.1095-8339.1987.tb01990.x

Wolfe, J. A. (1975). Some aspects of plant geography of the Northern Hemisphere during the late Cretaceous and Tertiary. *Ann. Missouri Bot. Gard.* 62, 264–279. doi: 10.2307/2395198

Wolfe, J. A. (1978). A paleobotanical interpretation of Tertiary climates in the Northern Hemisphere. *Am. Sci.* 66, 694–703.

Xiong, W., Ollerton, J., Liede-Schumann, S., Zhao, W., Jiang, Q., Sun, H., et al. (2020). Specialized cockroach pollination in the rare and endangered plant *Vincetoxicum hainanense* in China. *Am. J. Bot.* 107, 1–11. doi: 10.1002/ajb2.1545

Zhang, C., Zhang, T., Luebert, F., Xiang, Y., Huang, C.-H., Hu, Y., et al. (2020). Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-genome duplications. *Mol. Biol. Evol.* 37, 3188–3210. doi: 10.1093/molbev/msaa160

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Citation: Bitencourt C, Nürk NM, Rapini A, Fishbein M, Simões AO, Middleton DJ, Meve U, Endress ME and Liede-Schumann S (2021) Evolution of Dispersal, Habit, and Pollination in Africa Pushed Apocynaceae Diversification After the Eocene-Oligocene Climate Transition. Front. Ecol. Evol. 9:719741. doi: 10.3389/fevo.2021.719741

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 22 November 2021
doi: 10.3389/fevo.2021.773396

Prediction of Potentially Suitable Distributions of *Codonopsis pilosula* in China Based on an Optimized MaxEnt Model

Huyong Yan[1,2], Jiao He[3], Xiaochuan Xu[4], Xinyu Yao[4], Guoyin Wang[5], Lianggui Tang[1,2], Lei Feng[6,7], Limin Zou[8], Xiaolong Gu[9,10], Yingfei Qu[1,2] and Linfa Qu[8]*

[1] Chongqing Engineering Laboratory for Detection, Control and Integrated System, Chongqing Technology and Business University, Chongqing, China, [2] School of Computer Science and Information Engineering, Chongqing Technology and Business University, Chongqing, China, [3] School of International Business and Management, Sichuan International Studies University, Chongqing, China, [4] State Grid Chongqing Electric Power Company, Chongqing, China, [5] Chongqing Key Laboratory of Computational Intelligence, College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China, [6] Online Monitoring Center of Ecological and Environmental of the Three Gorges Project, Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing, China, [7] College of Environment and Ecology, Chongqing University, Chongqing, China, [8] School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing, China, [9] College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China, [10] National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing, China

OPEN ACCESS

Edited by:
Xiaoting Xu,
Sichuan University, China

Reviewed by:
Mohamed Abdelaal,
Mansoura University, Egypt
Helena Romo,
Autonomous University of Madrid, Spain

***Correspondence:**
Jiao He
hj2897@163.com

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *09 September 2021*
Accepted: *26 October 2021*
Published: *22 November 2021*

Citation:
Yan H, He J, Xu X, Yao X, Wang G, Tang L, Feng L, Zou L, Gu X, Qu Y and Qu L (2021) Prediction of Potentially Suitable Distributions of Codonopsis pilosula in China Based on an Optimized MaxEnt Model. Front. Ecol. Evol. 9:773396. doi: 10.3389/fevo.2021.773396

Species distribution models are widely used in conservation biology and invasive biology. MaxEnt models are the most widely used models among the existing modeling tools. In the MaxEnt modeling process, the default parameters are used most often to build the model. However, these models tend to be overfit. Aiming at this problem, this study uses an optimized MaxEnt model to analyze the impact of past, present and future climate on the distributions of *Codonopsis pilosula*, an economic species, to provide a theoretical basis for its introduction and cultivation. Based on 264 distribution records and eight environmental variables, the potential distribution areas of *C. pilosula* in the last interglacial, middle Holocene and current periods and 2050 and 2070 were simulated. Combined with the percentage contribution, permutation importance, and jackknife test, the environmental factors affecting the suitable distribution area of this species were discussed. The results show that the parameters of the optimal model are: the regularization multiplier is 1.5, and the feature combination is LQHP (linear, quadratic, hinge, product). The main temperature factors affecting the distribution of *C. pilosula* are the annual mean temperature, mean diurnal range, and isothermality. The main precipitation factors are the precipitation seasonality, precipitation in the wettest quarter, and precipitation in the driest quarter, among which the annual average temperature contributes the most to the distribution area of this species. With climate warming, the suitable area of *C. pilosula* exhibits a northward expansion trend. It is estimated that

in 2070, the suitable area of this species will expand to its maximum, reaching 2.5108 million square kilometers. The highly suitable areas of *C. pilosula* are mainly in Sichuan, Gansu, Shaanxi, Shanxi, and Henan Provinces. Our findings can be used to provide theoretical support related to avoiding the blind introduction of *C. pilosula*.

Keywords: optimized MaxEnt, *Codonopsis pilosula*, regularization multiplier, feature combination, potential distribution, multivariate environmental similarity surface analysis

INTRODUCTION

Based on niche theory, a species distribution model analyses the main factors affecting a species distribution and predicts its potential distribution according to the known distribution sample data and corresponding environmental factor data (Zhu et al., 2013). With the sharing of global species distribution data and the rapid development of GIS technology, the species distribution model of correlation schemes has developed rapidly (Zhu et al., 2017). The MaxEnt model, as a commonly used species distribution model, has been widely used in invasive biology (Yan et al., 2020b) and conservation biology (Yan et al., 2020a) and in evaluating the impact of global climate change on species distributions due to its short computing time, stable computing results and simple operation (Li et al., 2018). The MaxEnt model can better predict the distribution of species than other models in the case of a small sample size and can test the prediction results (Estes et al., 2013), but in the prediction of a potential distribution, it is prone to overfitting and can lead to a decline in species transfer-ability (Porfirio et al., 2014).

At present, although MaxEnt is widely used in species distribution models, most of the research methods directly use the default parameters of the MaxEnt model and do not consider the optimization of the model (Warren and Seifert, 2011). The prediction results of the unoptimized model may have serious fitting deviations, such deviations will not only lead to the wrong assessment of the species niche, but also mislead the formulation of conservation management policies. Although MaxEnt software provides some options for model optimization, there is no clear standard for the setting of these parameters (Syfert et al., 2013).

MaxEnt model optimization has four main aspects: sampling deviation correction, model complexity adjustment, selection of species distribution judgment threshold, and model performance evaluation. This article is to optimize the model complexity. The optimizations of model complexity mainly include the optimization of feature combination and regularization multiplier. These optimizations have been a hot research area in Maxent model in recent years. Maxent provides five basic feature modes, such as Linear (L), quadratic (Q), product (P), hinge (H), and threshold (T) (Elith et al., 2011). In the case of small sample sizes, the default mode of MaxEnt will limit the use of complex feature modes, the L feature is always running, the Q feature starts to run at least 10 samples, the H function requires at least 15 samples, and the P and T features require more than 80 samples. In the case of using different control multipliers, the H model and the LQH model have similar trends, and there is little difference between the model's omission error and the AUC value (Shcheglovitova and Anderson, 2013). To balance the fit of the model, MaxEnt sets a regularization multiplier to constrain the weights of variables, so that the model does not have to accurately fit the modeling data, but sets an error bound for the modeling data, and the model is fitted according to the observation value containing the error bound. The adjustment of the regularization multiplier is essentially an adjustment of the error bound.

Although the MaxEnt is based on the default regularization multipliers set for the data of 226 species in 6 regions of the world, the impact of overfitting is greatly reduced (Phillips and Dudík, 2008), many studies have shown that it is necessary to select an individual regularization multiplier according to different species and their data structures (Elith et al., 2010, 2011; Anderson and Gonzalez, 2011; Muscarella et al., 2014; Moreno-Amat et al., 2015). A higher or lower regularization multiplier will lead to a decrease in the AUC of the modeling data and test data (Anderson and Gonzalez, 2011). The appropriate regularization multiplier is also affected by the feature model. The more feature modes, the more complex the data Gibson distribution, and the regularization multiplier should be increased accordingly (Phillips and Dudík, 2008). The regularization multiplier also implies the choice of the feature mode, for example, the regularization multiplier of some feature modes is set to zero (Elith et al., 2011). The sample size also affects the choice of the regularization multiplier. The smaller the amount of data, the narrower the estimation interval. Therefore, it is necessary to increase the regularization multiplier to increase the error bound (Phillips and Dudík, 2008). The error bound finally determined by the regularization multiplier is proportional to the standard error of the sample (Anderson and Gonzalez, 2011). This is also the reason why the regularization multiplier needs to be increased for small sample data. The most important thing is that different species have different patterns of model performance changes due to the adjustment of predictive variables, feature models, and regularization multiplier. In addition, the optimization of the model obtained by some studies may not be suitable for the target species (Shcheglovitova and Anderson, 2013), especially for the traditional Chinese medicine *C. pilosula*, there is almost no optimization model in this species. Therefore, it is a hot and cutting-edge issue to avoid overfitting of the MaxEnt model and to explore the possible influence of model parameters on the prediction of the Chinese medicine species.

Codonopsis pilosula is the dry root of *Codonopsis pilosula* (Franch.) Nannf., *C. pilosula* Nannf. var. *modesta* (Nannf.) L. T. Shen or *C. tangshen* Oliv. It is sweet and flat. It has the functions of strengthening the spleen, tonifying the lungs, nourishing the blood and promoting fluid production

(Bi et al., 2008). Modern pharmacological research has confirmed that *C. pilosula* has certain effects on the digestive system and vascular systems improves blood and hematopoietic function and immune function, has anti-aging benefits and so on (Feng et al., 2012). As a less expensive substitute for ginseng, *C. pilosula* is also widely used in diet and health care (Gong, 2011). With the increasing market demand from Japan and South Korea, the export volume of *C. pilosula* continues to increase in China, with promising market prospects (Guo et al., 2017). At present, *C. pilosula* is mainly produced in Shanxi, Shaanxi, Gansu, and Sichuan Provinces and in other places in China, with a large amount of cultivation all over the country. Due to the scattered wild resources of this species, the quantity of commodities it can provide is extremely limited. From the perspective of production and marketing over the years, the output of wild products has been declining and cultivated products have been increasing. Therefore, cultivation of *C. pilosula* will still be the main source of commercial *C. pilosula* in the future (Zhang, 2006). At present, there are few reports on *C. pilosula*, and most of them focus on chemical composition (Huang et al., 2018), pharmacological activity (Liu M. X. et al., 2018), and quality evaluation (Duan et al., 2012). Some studies use the MaxEnt model to predict the potential distribution area of *C. pilosula*, but they do not discuss the environmental factors that restrict the distribution of this species, and they do not consider the historic distribution area of *C. pilosula* in the past and the potential distribution area in the future (Wu et al., 2017). To avoid the loss caused by blind introduction, explore the main environmental factors limiting geographical distribution, and grasp the change trend in the distribution pattern of *C. pilosula* under past and future climate change, it is particularly important to guide its introduction and cultivation scientifically.

This study aims at the current MaxEnt model's failure to deeply discuss the impact of parameter settings on the geographical distribution of the simulated species. At the same time, combined with the problem of the scientific introduction of the current economic species *C. pilosula*, by setting MaxEnt different regularization multipliers and feature combinations, the simulation potential distribution area of this species under different climatic conditions in five periods (last interglacial period (LIG), middle Holocene (MHC), the current era and the projection for the periods 2041–2060 (2050), and 2061–2080 (2070), at the same time, according to the permutation feature importance, jackknife test and percentage contribution rate, this study discusses the environmental factors that affect *C. pilosula*, analyses the reasons for restricting the geographical spread of this species, explores the changes in the distribution pattern of *C. pilosula*, and provides a certain scientific basis for the resource protection and management and development of this species.

In view of the fact that the current MaxEnt model's failure to deeply discuss the impact of parameter settings on the geographical distribution of the simulated species, the Maxent model is used to simulate the suitable distribution of *C. pilosula* and how this might change under different periods after parameter optimization. The specific objectives of this study were: (1) to display the influence of different parameter settings on the performance of the Maxent model; (2) to identify the main environmental variables affecting the potential suitable distribution of *C. pilosula*; and (3) to predict distribution of this species in different time periods.

MATERIALS AND METHODS

The flowchart of this study's contribution is shown in **Figure 1**. Briefly, first, the distribution record data were summarized, and environmental data were obtained. Then, these data were processed. Finally, the forecast analysis was performed.

Data Source and Processing

Collection and Processing of the Geographic Data

By searching the Chinese Virtual Herbarium (CVH),[1] Global Biodiversity Information Facility (GBIF),[2] and National Specimen Information Infrastructure,[3] 1,432 data points were obtained. Duplicate samples and samples with unclear geographic colocations were removed and cultivation records were introduced manually. The latitude and longitude information of the sampling location was obtained through the Baidu coordinate pickup system, and Google satellite maps was used to check the accuracy of the information. After removing some samples with inaccurate position coordinates, 467 data points are finally obtained. The original spatial resolution is 30 arcsec (about 1 km^2), which was matched with climate data layers. To reduce the spatial autocorrelation (Fortin, 1999), only one point closest to the center in the 30 arcsec grid plots (approximately 1 km^2) is selected, and 264 effective data records are obtained and converted into csv format for storage.

Screening and Processing of the Environmental Variables

Through the World Climate Database,[4] the bioclimatic variables of the last interglacial (LIG), middle Holocene (MHC), current and future (2050, 2070) periods are downloaded. These variables (**Supplementary Table 1**), including 19 bioclimatic variables, were derived from a CMIP5 global circulation model, CCSM4 (Community Climate System Model version 4), generated by the National Center for Atmospheric Research (Colorado, United States). It is one of the most efficient global climate tools for the simulation of future conditions, which has been evaluated in China and successfully applied to forecast the influence of future changes on the distribution of plant species in similar circumstances (Abdelaal et al., 2019). In the 5th Assessment Report of the IPCC, four representative concentration pathways (RCPs), viz. RCP 2.6, RCP 4.5, RCP 6.0, and RCP 8.5, were defined using the total radiative forcing value in 2,100 according to the pre-industrial values of 2.6, 4.5, 6.0, and 8.5 W/m^2, respectively (Yan et al., 2020c). RCP 2.6 and RCP8.5 were selected to model the prospective potential distribution of *C. pilosula* under minimum and maximum emissions scenarios

[1] https://www.cvh.ac.cn
[2] http://www.gbif.org
[3] www.nsii.org.cn
[4] www.worldclimate.org

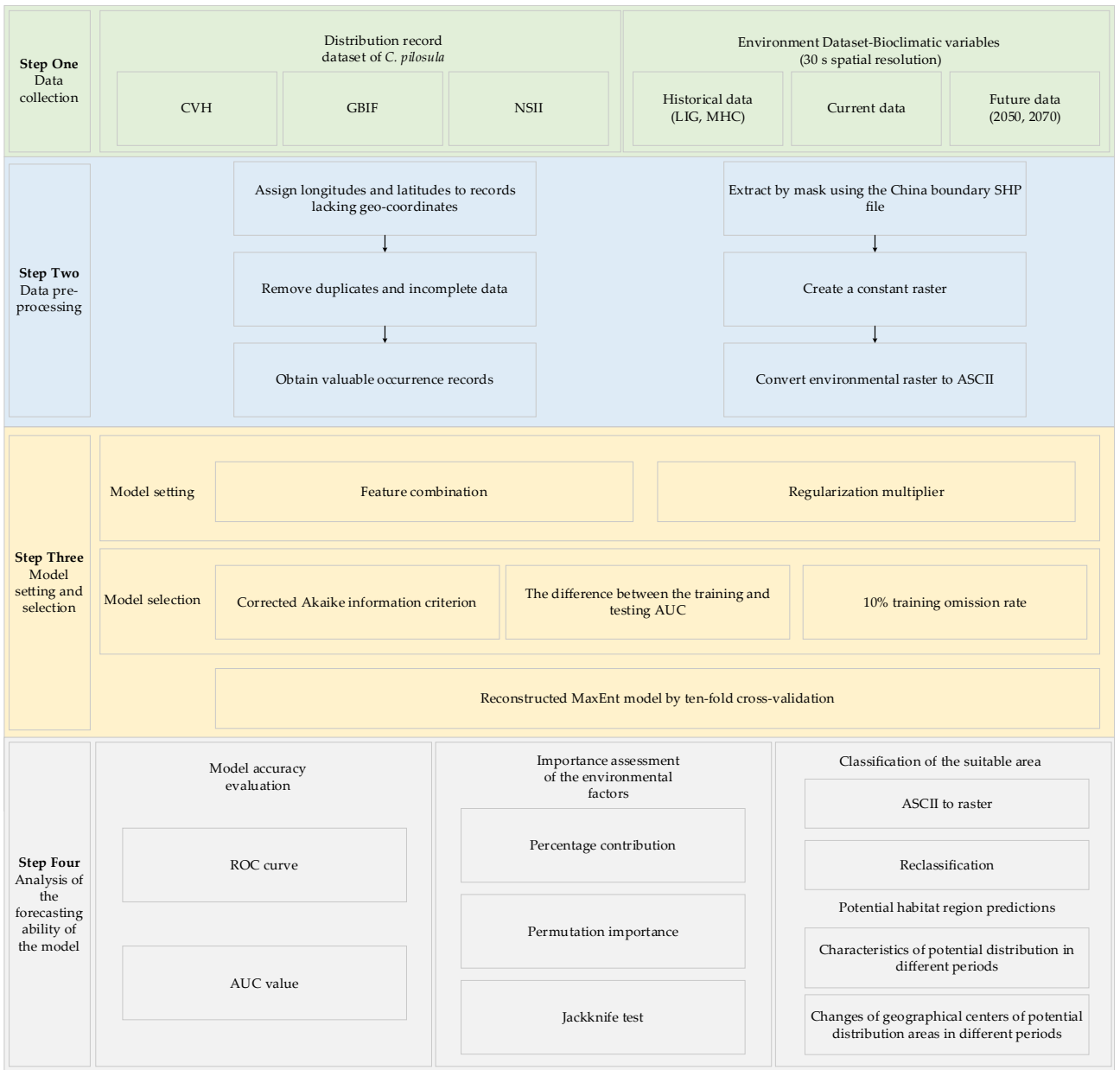


FIGURE 1 | The framework of this study's contribution.

hypothesis. The resolution of these bioclimatic variables is 30 arcsec. We use Pearson correlation coefficient analysis to test the multicollinearity among climate variables was applied (Yang et al., 2013). Among a group of climate variables with high correlation ($|r|$ 0.8), only one variable closely related to species distribution or easy to explain was selected for model prediction (Kumar and Stohlgren, 2009). R program methods was used to analyze the Pearson correlation coefficients. Eight variables were selected (**Supplementary Table 2**), which is convenient for model interpretation: annual mean temperature (bio01), mean diurnal range (bio02), isothermality (bio03), temperature seasonality (bio04), max temperature of the warmest month (bio05), precipitation seasonality (bio15), precipitation in the wettest quarter (bio16), and precipitation in the driest quarter (bio17).

Model Setting and Selection

This study used MaxEnt v3.4.1[5] to predict the potential geographic distribution of *C. pilosula* in different periods. The MaxEnt model is based on a certain algorithm to simulate a known real model with potential complexity. Its accuracy and authenticity cannot be estimated accurately. Therefore, it is necessary to take appropriate restrictive measures to reduce

[5]http://www.cs.princeton.edu/~schapire/maxent

overfitting. Specifically, the optimal model is selected by setting the regularization multiplier (RM) and feature combination (FC). RM can affect how focused the output distribution is. A smaller RM value will result in a more localized output distribution that fits the given occurrence records better, but is more prone to overfitting (Zhu et al., 2018). Overfitted models are poorly transferable to novel environments and, thus, not appropriate to project distribution changes under environmental change. A larger RM value will produce a prediction that can be applied more widely. FC constrain the computed probability distribution. In the default parameter setting, the selection of FC is related to the number of species distribution points, and the RM value is 1 (Zhu et al., 2018). In R v3.6.3,[6] the ENMeval data package[7] is used for optimization. The checkerboard2 method is used to divide the distribution data, set the aggregation factor to 5, and set the RM to change from 0.5 to 4.0 with an increase by 0.5 each time. Therefore, there are eight values of RM; FC selects 7 types for testing, namely, L, LQ, H, LQH, LPQ, LQHP, LQHPT. ENMeval tests the above 56 parameter combinations, and finally, we use the corrected Akaike information criterion (AICc) (Akaike, 1998; Muscarella et al., 2014), the difference between the training and testing AUCs (AUC_{DIFF}) (Warren and Seifert, 2011), and the 10% training omission rate (OR_{10}) (Pearson et al., 2007), to evaluate model performance. The AICc value comprehensively reflects the goodness of fit and the complexity of the model. It is an excellent standard for measuring the performance of the model. The model with the smallest AICc value, that is, the model with delta AICc = 0, is considered the best model, and all models with delta AICc < 2 are considered to have substantial support (Burnham and Anderson, 2004; Warren and Seifert, 2011). AUC_{DIFF} is expected to be positively associated with the degree of model overfitting (Warren and Seifert, 2011). OR_{10} represents the proportion of test localities with suitability values lower than that excluding the 10% of the training localities with the lowest predicted suitability. Omission rates greater than the expectation of 10% indicate model overfitting (Fielding and Bell, 1997; Peterson et al., 2011).

According to the above three evaluation indexes, the best model parameter combination is determined, and the best MaxEnt model is rebuilt with the eight selected variables. The number of best model repetitions is set to 10. The method of sampling the test samples is 10-fold cross-validation. This method divides all the current distribution records of *C. pilosula* into 10 subsets, each of which takes turns as the test set and the other nine subsets are as the training set, which improves the utilization ratio of the data and the persuasiveness of the results. The training data and test data of this species distribution records have three characteristics: serial number, longitude information, and latitude information.

Model Forecast Analysis

Optimized MaxEnt software is used to predict the potential geographical distribution of *C. pilosula* in the past, current, and future. According to the habitat suitability value of the distribution point and the normal distribution parameters μ and δ, the suitability of *C. pilosula* was divided into three levels: $[0,\mu-\delta]$ are unsuitable habitats, $(\mu-\delta,\mu)$ are potential habitats and $[\mu,1]$ are high-potential habitats (Wei et al., 2019). The high-potential habitats are areas in which the ecological factors are optimal, and there are few restrictions on the growth and development of this species; the potential habitats are areas in which the ecological factors are suitable for the growth of *C. pilosula*, but to a certain extent, they are slightly inferior to the high-potential habitats; the unsuitable habitats are areas in which the ecological factors are severely restricted and are unfavorable for the growth of this species, and they are not suitable for cultivation and planting.

The MaxEnt model software has three functions: permutation importance, percentage contribution rate and the jackknife test, which are used to evaluate the relative importance of the environmental factors to the distribution of *C. pilosula*. The permutation importance value is independent of the specific algorithm path of the software and only depends on the final results of the model. The contribution degree of each environmental variable depends on the random enumerated variable values on the training point (existence point and background value) and the resulting decrease in the training AUC value. The larger the AUC decrease, the more dependent the model is on the environment variable. The percentage contribution rate depends on the specific algorithm of the maximum entropy program to obtain the optimal solution. The MaxEnt model records the contribution of the environmental variables to the adaptive model during training (Phillips, 2006). By modifying the coefficient corresponding to each eigenvalue, the model gain value will increase, and the gain value will be allocated to the environmental variables on which each eigenvalue depends; finally, the environmental contribution will be converted to the percentage output. After the jackknife test runs a single environmental factor one by one and uses all environmental factors to obtain the results, the differences between the training gain value, test gain value and AUC value are compared to understand the more important environmental factors in the distribution of *C. pilosula*.

The accuracy of this model was evaluated by the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC). Its value is directly proportional to the accuracy of the model prediction. When the software runs 10 times, the average training AUC is calculated automatically. The AUC is generally between 0 and 1. An AUC between 0 and 0.5 indicates that the prediction effect of the model is very poor; an AUC between 0.6 and 0.9 indicates that the prediction effect is moderate; an AUC greater than 0.9 indicates that the prediction effect of the model is very good. The closer that this value is to 1, the better the performance of the model is (Phillips et al., 2006).

RESULTS

Model Optimization and Accuracy Evaluation

Based on 264 records of the current distribution of *C. pilosula* and eight layers of environmental variables, the current potential

[6] http://cran.r-project.org/

[7] http://cran.r-project.org/web/packages/ENMeval/index.html

distribution habitats of this species were simulated by the MaxEnt model. The default setting of MaxEnt is FC = LQHPT, RM = 1. At this time, the delta AIC_C was 39.240 (**Table 1**). After ENMeval optimization, the MaxEnt model is optimized as follows: the FC is LQHP, namely, linear, quadratic, fragmented, and product, and the RM is 1.5. At this time, the delta AICc is 0, and its value is less than 2 (**Figure 2A**), which indicates that the complexity of the model is in an acceptable range and has high reliability. Further comparison shows that the mean AUC_{DIFF} and OR_{10} of the optimized model are 0.011 (**Figure 2B**) and 0.118 (**Figure 2C**), respectively, which are 54.17 and 30.59% lower than the default setting, indicating that overfitting with the optimized parameter setting is low. Therefore, RM = 1.5 and FC = LQHP are selected as the final parameter settings. Under the condition of this parameter setting scheme, when the MaxEnt model simulates the current geographical distribution 10 times, the maximum AUC value is 0.926, the minimum AUC value is 0.846, and the average AUC value is 0.910 $\pm$ 0.009 (**Figure 3B**), which indicates that the prediction of the model is accurate. If the default parameters are used, the maximum AUC value is 0.878, the minimum AUC value is 0.826, and the average AUC value is 0. 865 $\pm$ 0.029 (**Figure 3A**). After using the optimized parameters to rebuild the model, the prediction accuracy is better than that of the default parameters.

The Main Environmental Variables Affecting This Species

The contribution rate and permutation importance of each environmental variable listed in **Table 2** indicated that bio01 plays an important role in controlling the distribution of *C. pilosula*. Its contribution rate reaches 42.1%, and the permutation importance reaches 36.8%. In addition, the cumulative contribution rate of bio01, bio02, and bio03 reaches 83.8%, and the cumulative permutation importance reaches 70.8%.

By analyzing the gain of each variable, it is found that the gain of using any environmental variable alone does not exceed the gain of using all the variables. Each environmental variable contributes to improving the prediction accuracy of the model, which is mutually verified with the calculation results of the contribution rate in **Table 2**. Among them, bio16, bio01, and bio17 are the top-three variables in terms of the normalized training gain, and bio16, bio17, and bio01 are the top-three variables in terms of the test gain and AUC value, so these variables have better matching and more effective information for the training data; when other variables are used, the three variables with the most decline in normalized training gain, test gain and AUC value are bio16, bio17, and bio15, indicating that the above variables contain valid information that other variables do not have. In conclusion, the main temperature factors affecting the distribution of *C. pilosula* are bio01, bio02, and bio03, and the main rainfall factors are bio15, bio16, and bio17.

The curves in **Figure 4** show how each environmental variable affects the MaxEnt model prediction. The curves show how the predicted probability of presence changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. The curves show the mean response of the 10 replicates. MaxEnt runs the mean (red) and +/− one standard deviation (blue). In a certain range, the distribution probability of *C. pilosula* increased with increasing bio01, bio02, bio03, bio15, bio16, and bio17. After reaching a certain peak, the distribution probability decreased with an increase in the values of the environmental factors. According to the response curve of the climate factors, we can judge the relationship between the existence of *C. pilosula* and the values of the ecological factors. When the distribution probability of this species is greater than 0.3324, its corresponding ecological factor value is suitable for the growth of *C. pilosula*. *C. pilosula* has certain adaptability to six major environmental factors, and the thresholds of each environmental factor are bio01 (4.02–23.17°C) (**Figure 4A**), bio02 (3.24–13.49°C) (**Figure 4B**), bio03 (23.69–40.46) (**Figure 4C**), bio15 (52.10–109.56) (**Figure 4D**), bio16 (242.83–861.02 mm) (**Figure 4E**), and bio17 (8.88–146.23 mm) (**Figure 4F**).

The contribution rate of the annual mean temperature to the distribution area of *C. pilosula* is the largest. When the annual mean temperature is less than 0°C, the distribution probability of this species is almost 0. When the annual mean temperature is more than 0°C, the distribution probability of *C. pilosula* rises sharply, reaches its peak value and optimum growth condition at 17.91°C. Then, the distribution probability of *C. pilosula* decreases rapidly.

Forecast of the Distribution of Suitable Geographical Areas for *Codonopsis pilosula*

Characteristics of the Current Potential Geographical Distribution

Among the 264 records of the *C. pilosula* distribution, 70.08% are potential habitats, and 12.12% are high-potential habitats, which shows that the current potential habitats simulated by MaxEnt can basically cover the current distribution points. Current potential habitats are mainly distributed in most provinces and cities of China, and Xinjiang and Hainan are unsuitable habitats (**Figure 5**). Due to the sparse data of individual sampling points, although there are distribution records in Heilongjiang Province, they are not predicted as potential habitats. The high-potential habitats are located in the north of Henan Province, the junction of Sichuan and Gansu Provinces, the junction of Sichuan and Shaanxi Provinces, the junction of Sichuan and Chongqing Provinces, the junction of Shaanxi and Shanxi Provinces, and the central region of Shaanxi Province. The average suitability degree of the 264 distribution records is 0.536, and the three locations with the highest suitability degrees are Guangyuan City, Lizhou District, Sichuan Province (0.856); Xi'an city, Hu County, Shaanxi Province (0.839); and Daofu County, Ganzi Prefecture, Sichuan Province (0.823). The three locations with the lowest suitability degrees are Manzhouli city, Hulunbuir city, Inner Mongolia (0.014); Shaoyang City, Xinning County, Hunan Province (0.038); and Menyuan Hui Autonomous County, Haibei Tibetan Autonomous Prefecture, Qinghai Province (0.053).

TABLE 1 | The performance of the MaxEnt model using the default parameter settings and the optimized parameter settings.

Settings	Regularization multiplier	Feature combination	Mean AUC_{DIFF}	OR_{10}	Delta AICc
Default	1	LQHPT	0.024	0.170	39.240
Optimized	1.5	LQHP	0.011	0.118	0

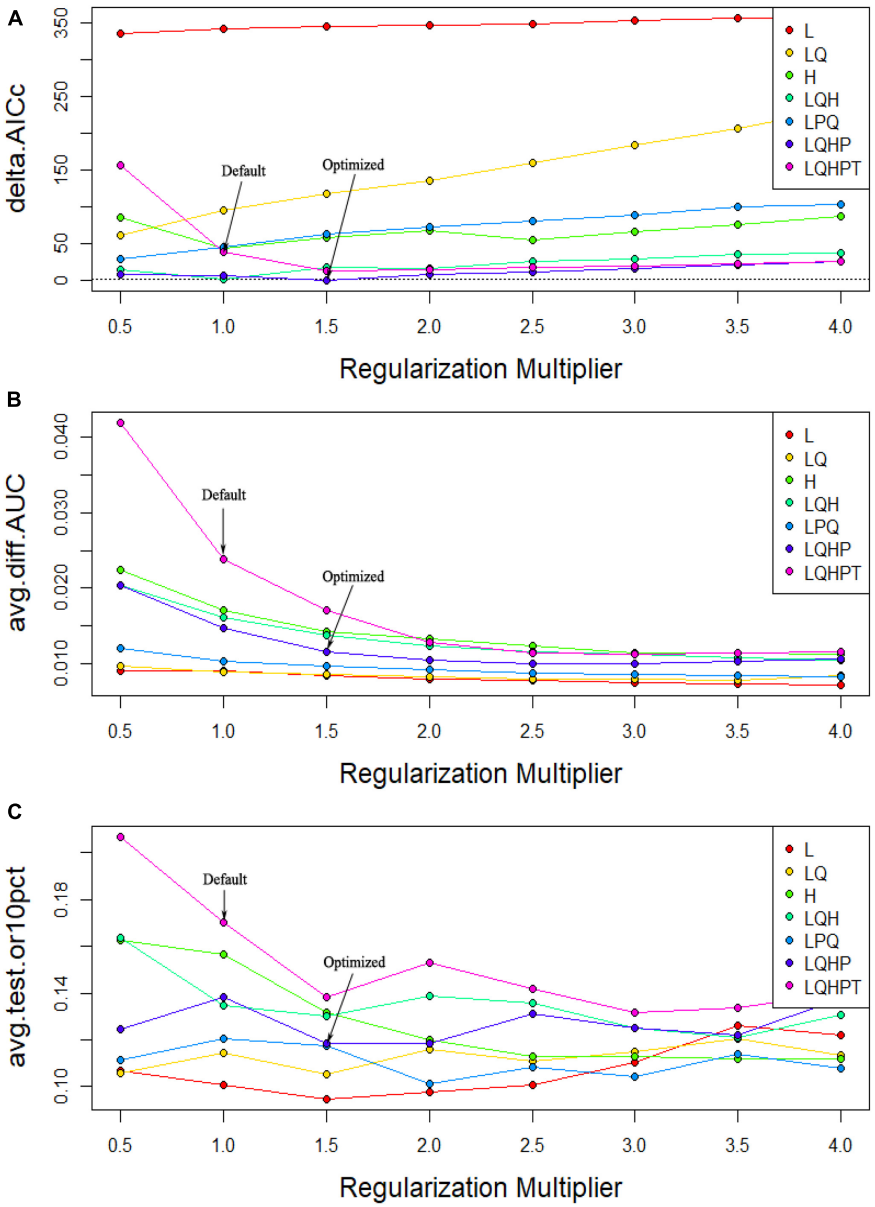


FIGURE 2 | Evaluation results of the MaxEnt model under different parameters settings. **(A)** Delta AICc; **(B)** AUC_{DIFF}; **(C)** OR_{10}.

Potential Distribution Characteristics of the Past and the Future

Based on the MaxEnt model, the last two periods are predicted, which are the past interglacial period (**Figure 6A**) and the middle Holocene period (**Figure 6B**). Compared with the current distribution, the LIG period changed considerably. The area of potential suitable distribution habitats and high-potential suitable distribution habitats both decreased, the potential habitats decreased by 5.20% and the high-potential suitable habitats decreased the most, reaching 13.00% (**Table 3**). Only in the north of Sichuan Province, the middle and south of Shaanxi Province, and the sporadic areas in Shanxi and Henan Provinces are highly suitable. In the middle Holocene, the area of potential and high-potential areas increased compared with that of the

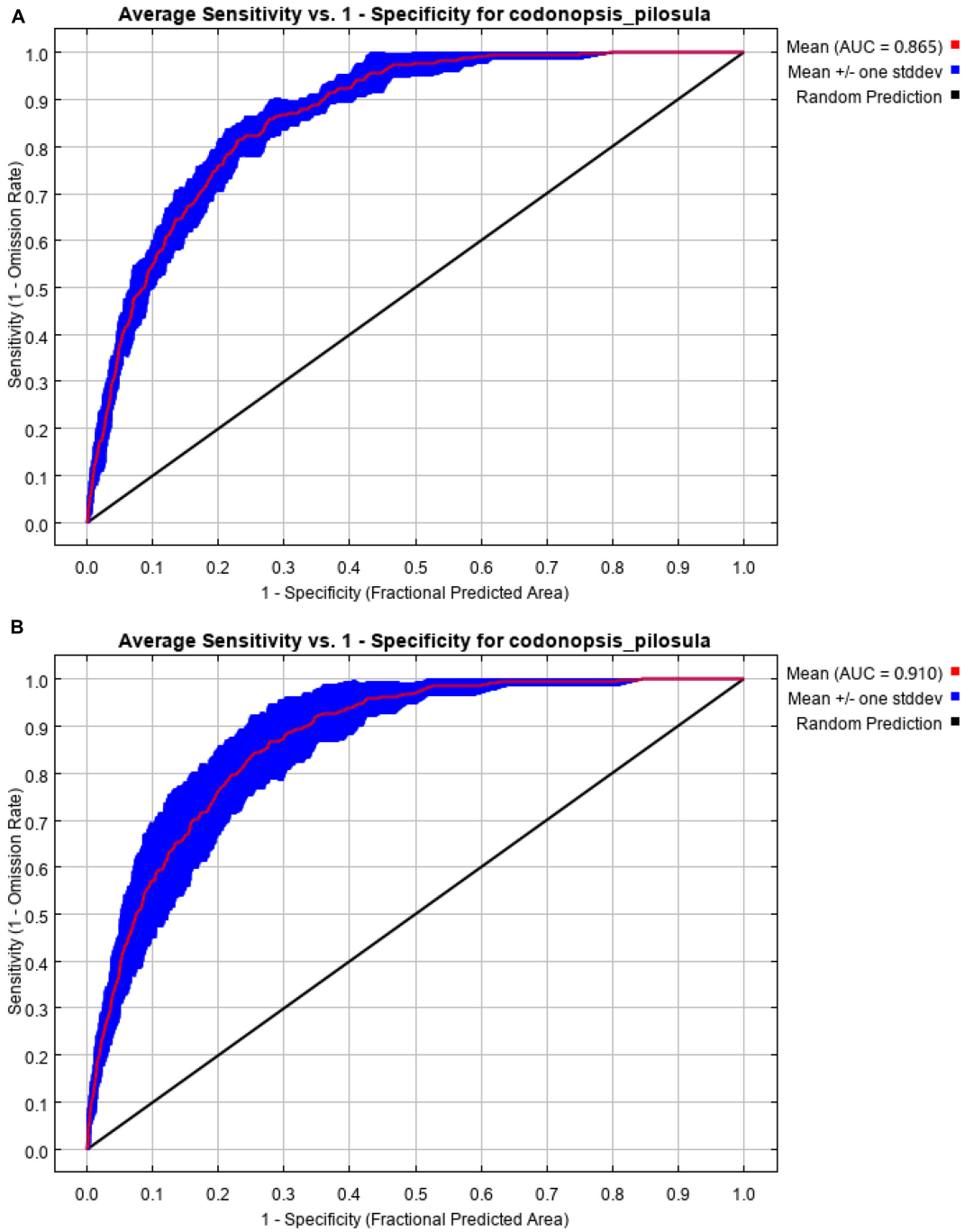


FIGURE 3 | ROC curve of the reconstructed MaxEnt model. **(A)** Defaulted model; **(B)** optimized model.

last interglacial period, and the outline was similar to that of the current outline, but both areas were smaller than that of the current distribution.

Based on the model of current distribution points, the future potential distribution areas of *C. pilosula* were predicted for 2050 and 2070 under two future climate scenarios (RCP 2.6 and RCP 8.5). Under the scenario with high levels of greenhouse gas emissions (RCP 8.5), it is estimated that by 2050, the range of potential habitats of this species will decrease by 2.03%, and the range of high-potential habitats will increase by 7.08%. The potential habitats will expand eastward, almost covering the whole province of Shandong, and the current high-potential habitats at the junction of Gansu and Sichuan Provinces will expand eastward to the junction of Gansu, Sichuan and Shaanxi

TABLE 2 | Important environmental variables and corresponding metrics.

Variables	Percentage contribution (%)	Permutation importance (%)	TNGw	TNGo	TGw	TGo	AUCw	AUCo
bio01	42.1	36.8	1.0384	0.342	1.0569	0.3418	0.8686	0.6939
bio02	29.1	11.6	1.0372	0.1884	1.0592	0.2038	0.87	0.6719
bio03	12.6	22.4	1.0402	0.0871	1.0643	0.1106	0.8703	0.6426
bio04	7.6	5.2	1.0103	0.115	1.0298	0.1181	0.8667	0.612
bio05	3.5	8.2	1.0356	0.1761	1.0526	0.1873	0.8687	0.6322
bio15	2.7	0.9	1.0016	0.1119	1.0184	0.1338	0.8653	0.6491
bio16	1.8	10	0.9392	0.449	0.9707	0.4613	0.8582	0.7255
bio17	0.5	5	0.9414	0.3259	0.975	0.342	0.8593	0.7143

TNG_w represents training gain without this variable, TNG_o represents training gain with only this variable, TG_w represents test gain without this variable, TG_o represents test gain with only this variable, AUC_w represents the AUC without this variable, and AUC_o represents the AUC with only this variable.

Provinces (**Figure 6C**). Compared with the current distribution, in 2070, the potential habitats of this species will increase by 0.60%, and the range of high-potential habitats will increase by 11.41%. The potential habitats gradually appeared in the central and northern parts of Yunnan, central Qinghai and central Gansu Provinces, and the changes in other potential habitats were small. Most of the areas at the junction of Gansu, Sichuan and Shaanxi Provinces have high-potential habitat conditions (**Figure 6D**). Under RCP 8.5, it is estimated that by 2050, the range of potential habitats and high-potential habitats of this species will increase by 22.91 and 202.99%, respectively. Potential habitats tend to spread to the southeast provinces, and the size of high-potential habitats has also expanded correspondingly, reaching 250,500 square kilometers (**Figure 6E**). Compared with the current distribution, in 2070, the potential habitats of this species will increase by 28.38%, and the range of high-potential habitats will increase by 191.29%. At that time, the area of potential habitats will be as high as 2.5108 million square kilometers (**Figure 6F**).

According to the statistics of the geographical centers of the suitable distribution areas in different periods, it can be seen that the geographical centers of the two historical periods are south of 33.5°N, and the change in the current geographical centers is relatively small (only 28.50 km to the northeast), and then the geographical centers will continue to move to the northwest in the next two periods. By 2070, the geographical centers will move to the north to 33.55°N (**Figure 7**).

From the perspective of the change in the suitable area in different periods, from the last interglacial period to 2070, the distribution of *C. pilosula* experienced an overall gradual increase. The potential suitable area in the five periods was the smallest in the last interglacial period and the largest in 2070. The high-potential habitats changed greatly in five periods. In the last interglacial period, the high-potential habitat area is the smallest (only 72,000 square kilometers). In 2050, the area is the largest, reaching 250,500 square kilometers.

Under the trend of global warming, most of the multivariate environmental similarity surface (MESS) of climate scenarios in the future multi-period range is between 0 and 10 in the current suitable region, and some areas are 10–20. The areas with a MESS greater than 20 will only occupy a small area part. Under climate scenario RCP 2.6-2050 (**Figure 8A**), the mean value of the MESS under this climate scenario is 14.13, the first quartile is 1.96, and the third quartile is 24.68; Under RCP 8.5-2050 (**Figure 8C**), the mean value of the MESS is 9.94, the first quantile is 0.17, and the third quantile is 14.95; Under RCP 2.6-2070 (**Figure 8B**), the mean value of the MESS under this climate scenario is 13.13, the first quantile is 1.55, and the third quantile is 24.08; Under RCP 8. 5-2070 (**Figure 8D**), the mean value of the MESS under this climate scenario is 6.34, the first quantile is −2.15, and the third quantile is 10.20. Combining the values of MESS and change maps in different periods under future climate change, it can be seen that in different climate change scenarios in the same period, in 2050, the MESS values from large to small are RCP2.6 > RCP8.5; in 2070, the MESS values from large to small is RCP2.6 > RCP8.5. When the climate change scenarios in different periods are the same, under the climate scenario RCP2.6, the mean value of the MESS in the 2050 period is greater than 2070; under the climate scenario RCP8.5, the mean value of the MESS in the 2050 period is greater than 2070. That is, when the period is the same, and the climate change scenarios are different, the MESS value is positively correlated with the greenhouse gas emission concentration; when the period is different, the climate change scenario is the same, the MESS value decreases with the passage of time, and there is a negative correlation between the two relations.

The mean range of the MESS in different periods under future climate change scenarios ranges from 6.34 to 14.13. The degree of climate anomaly of each climate change scenario is similar, but there are some differences. In general, under future climate change scenarios, areas with high climate anomalies are mainly the two provinces, Shanxi and Shaanxi, located in the northwest of the current potential habitats; most of these regions have MESS values less than −10, while the areas with low climate anomalies are mainly located in Hebei, Shandong, and Henan in the east of the current potential habitats. The range of the MESS in these provinces is 0–20. The RCP 2.6-2050 climate scenario has the highest MESS value and the lowest climate anomaly degree; while the RCP 8.5-2070 climate scenario has the lowest MESS value and the highest climate anomaly degree.

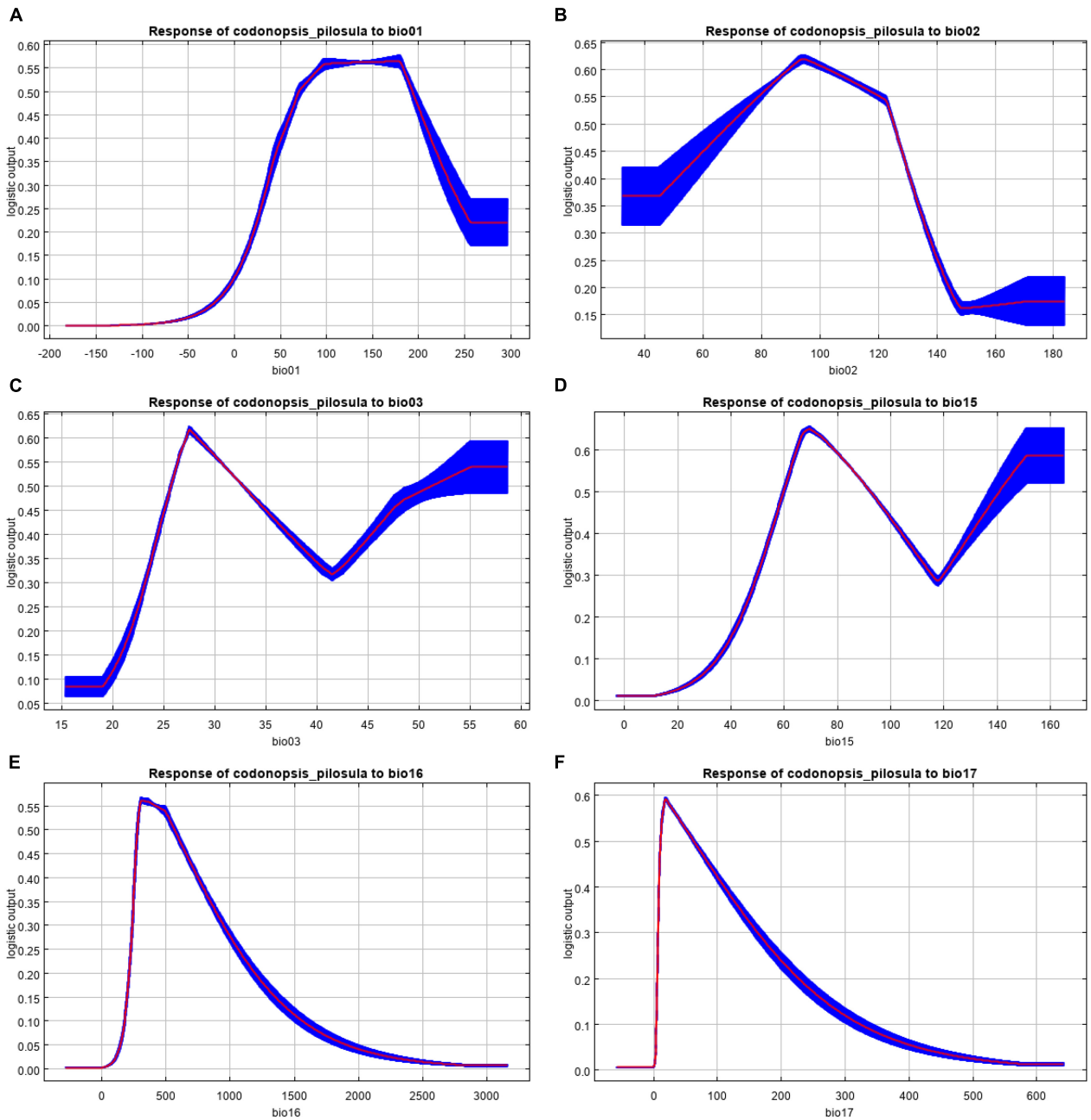


FIGURE 4 | Response of *Codonopsis pilosula* to six major environmental factors. **(A)** bio01; **(B)** bio02; **(C)** bio03; **(D)** bio15; **(E)** bio16; **(F)** bio17 (**A,B** represent temperature in degrees Celsius, and its value is magnified by 10 times due to data synthesis).

DISCUSSION

MaxEnt usually retains a random subset for data modeling and then uses the AUC to evaluate the model prediction ability, but the software has defects (Warren et al., 2014). First, when the training data and test data are affected by the sampling deviation, the AUC may overestimate the ability of the model (Veloz, 2009). Second, the MaxEnt model is a highly complex machine learning model. When simulating the potential distribution of species, it will lead to overfitting of the model, which will directly affect the transfer-ability of species (Zhu and Qiao, 2016). Although some domestic and foreign experts and scholars have carried out research on the prediction of the distribution area of the *C. pilosula* based on the MaxEnt model and achieved good prediction results, they did not consider the over-fitting phenomenon of maxent modeling, which caused the generalization ability of the model to be limited (Jie et al., 2017; Niu et al., 2021). The complexity of the model can be

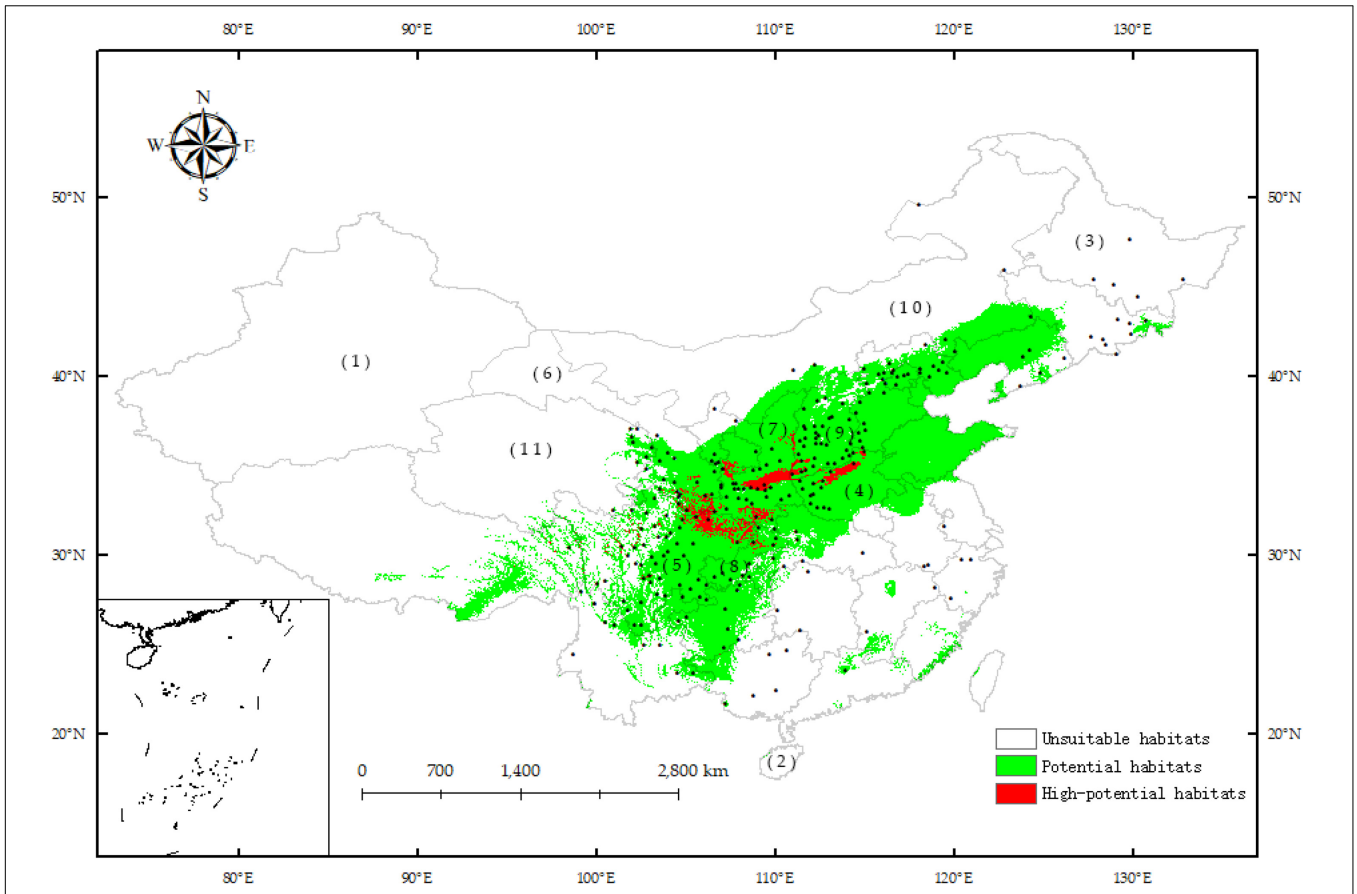


FIGURE 5 | Potential distribution for *C. pilosula* under the current conditions predicted by the MaxEnt model based on the optimized parameters. (1) Xinjiang; (2) Hainan; (3) Heilongjiang; (4) Henan; (5) Sichuan; (6) Gansu; (7) Shaanxi; (8) Chongqing; (9) Shanxi; (10) Inner Mongolia; (11) Qinghai.

constrained by adjusting the regularization multiplier in the MaxEnt model and using AICc to select the appropriate feature combination (Warren and Seifert, 2011). In this study, AICc, AUC and OR_{10} are used to select the best MaxEnt model with the smallest overfitting effect. The results show that RM = 1.5 and FC = LQHP have the lowest degree of overfitting. After using the optimized parameters to rebuild the model, the prediction of the *C. pilosula* habitats of the five periods is obtained, and the prediction accuracy is better than that of the default parameters.

The prediction results of the MaxEnt model show that the main ecological factor is temperature, followed by precipitation. The important factors affecting the geographical distribution of *C. pilosula* include the annual mean temperature (bio01), mean diurnal range (bio02), isothermality (bio03), precipitation seasonality (bio15), precipitation in the wettest quarter (bio16), and precipitation in the driest quarter (bio17). The suitable temperature range for growth in this species is 12.0–25.0°C, and its seed germination temperature is 18.0–20.0°C. When the temperature reaches 12.0–18.0°C, the underground stem sprouts. The temperature during which the fastest growing occurs is 15.0–25.0°C, and the most suitable temperature for flowering and seed setting is 15.0–22.0°C (Chang, 2013). The highest germination rates of *C. pilosula* seeds were 96.00 and 97.33%, respectively, at 20°C, and the germination duration was 9 days. At 30°C, the lowest germination rates were 17.00 and 28.00%, respectively, which indicated that the seeds of this species prefer cool but not hot temperatures (Su et al., 2012). *C. pilosula* is more cold resistant than other plants, and its growth above 30°C is limited (Zhao et al., 2006). The results show that the annual mean temperature range of *C. pilosula* is 4.02–23.17°C, and the best temperature is 17.91°C, which is in line with the suitable temperature for growth of this species and is very close to the temperature needed for seed germination. When *C. pilosula* is exposed to sustained hot temperatures in the main growth period, its aboveground parts will easily wither and suffer from diseases (Chang, 2013). Therefore, we should pay attention to the control of high temperatures in the cultivation of *C. pilosula*. The mean diurnal range plays an important role in the growth of *C. pilosula*. The suitable range of diurnal range is 3.24–13.49°C, the peak value is 9.26°C, and the higher diurnal range is conducive to the formation of tubers, promoting the growth of this species (Chang, 2013). In contrast, if the diurnal range is smaller than that above, the respiration function of *C. pilosula* will be strengthened. Under these conditions, the respiration loss is large, which is not conducive to the growth of tubers, and low drying rate of the tubers can easily

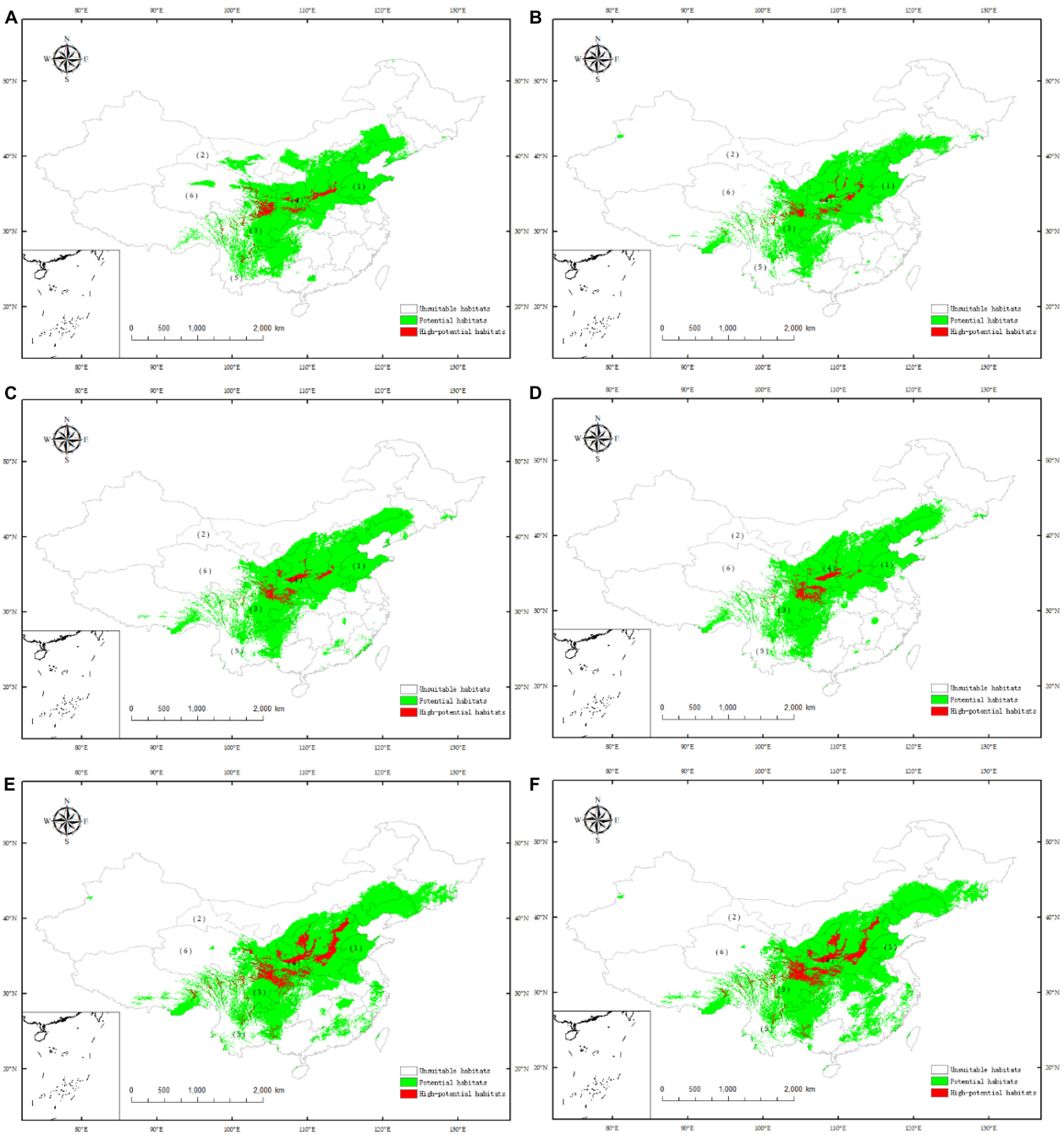


FIGURE 6 | Potential distribution for *C. pilosula* during different periods predicted by the MaxEnt model based on the optimized parameters. **(A)** LIG; **(B)** MHC; **(C)** RCP 8.5-2050; **(D)** RCP 8.5-2070; **(E)** RCP 2.6-2050; **(F)** RCP 2.6-2070 (1) Shandong; (2) Gansu; (3) Sichuan; (4) Shannxi; (5) Yunnan; (6) Qinghai.

occur. After processing, low-quality, inferior products are easily formed. The main precipitation factors affecting the growth of *C. pilosula* are the precipitation seasonality, the precipitation in the wettest quarter, and the precipitation in the driest quarter. The suitable range of the precipitation seasonality is 52.10–109.56, that of the precipitation in the driest season is 8.88–146.23 mm and that of the precipitation in the wettest season is 242.83–861.02 mm. The water demand of *C. pilosula* in different growth periods is different. In the sowing and transplanting periods, the water demand is higher, and the soil humidity needs to be approximately 70%. After planting, the water demand of *C. pilosula* is reduced, and if the soil is too wet, especially under high-temperature and high-humidity conditions, it is easy to have diseases and insect pests and rotten roots. In the mature period, the precipitation is too high, which is not conducive to the formation of the roots of this species (Wang et al., 2006).

TABLE 3 | Characteristics of the potential distribution in different periods.

Periods				Area (10^4 square kilometers)		Percentage change in the area compared to the current distribution (%)	
	①	②	③	Potential habitats	High-potential habitats	Potential habitats	High-potential habitats
LIG	10.55	498.33	328.93	185.51	7.20	−5.20	−13.00
MHC	10.63	804.49	456.11	185.95	7.79	−4.97	−5.80
Present	11.35	770.99	423.04	195.68	8.27	0.00	0.00
RCP 8.5-2050	14.10	814.84	452.70	191.70	8.86	−2.03	7.08
RCP 8.5-2070	15.21	847.84	478.66	196.85	9.22	0.60	11.41
RCP 2.6-2050	11.37	780.18	430.75	240.37	25.05	22.91	202.99
RCP 2.6-2070	12.33	784.12	432.78	251.08	24.08	28.38	191.29

The annual mean temperature and precipitation are the average values across the 264 C. pilosula data points. ①Annual mean temperature (°C); ②Annual precipitation (mm); ③Precipitation in the wettest quarter (mm).

The seasonality of precipitation reflects the different demands of *C. pilosula* on precipitation in different growth periods. Precipitation has the greatest influence on *C. pilosula* in the sowing and transplanting periods than the other variables. During these periods, we should pay attention to providing sufficient water for *C. pilosula*.

The climate of the last interglacial period, which began 120,000 years ago, is the most similar to the current climate, but it is still very different. The global temperature is approximately 5°C lower than it is now. At that time, China's climate was abnormal, cold and dry. China's glacial area was 8.4 times larger than it is now. The cold and cold climate had a general impact on animals and plants (Chen et al., 2011), and most of the plants significantly reduced their distribution area after the temperature dropped in the ice age (Xu et al., 2017). The suitable area for warm and humid plants moved to the south as a whole, and most of the northern areas were occupied by plants that can tolerate a cold and dry climate (Shepherd et al., 2017). The results showed that out of the five periods, the geographical center of the suitable habitats of *C. pilosula* in the last interglacial period was the southernmost, and the distribution area was the smallest. The middle Holocene period, is the most recent warm period in history. In this period, the whole suitable area and the highly suitable area were increased compared with the current area. The prediction results of the MaxEnt model also show that the geographical distribution center has a trend of moving northward, and the outline is basically similar to that of now. The study of paleoclimate reconstruction shows that the average increases in the area, precipitation and precipitation intensity in China's monsoon area during this period were 10.7, 18.7, and 7.3%, respectively, under different climatic patterns (Tian and Jiang, 2015). As a species that prefers a wet climate, the increase in rainfall in summer is beneficial to the growth of *C. pilosula*, and its suitable area and highly suitable area increased. Climate fluctuations caused the migration of the vegetation belt. Although most studies found that the vegetation belt moved northward in the peak of the warm period (Fang et al., 2011), the range of movement was not large (Xun, 2000). Therefore, compared with the current distribution, the potential suitable area of *C. pilosula* did not show a significant trend in movement, and the contour is close to the current contour. Compared with current times,

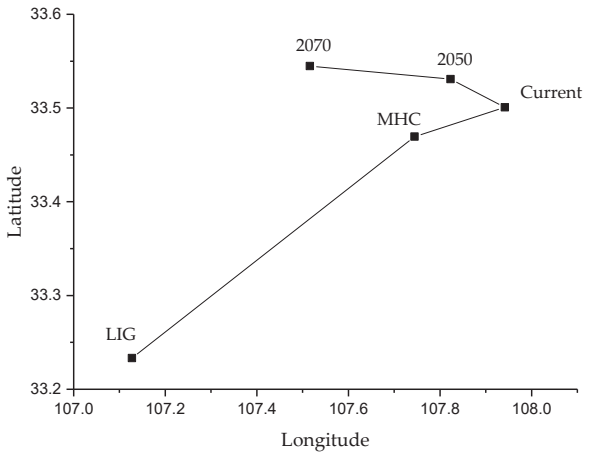


FIGURE 7 | Changes in geographical centers of the potential distribution areas of *C. pilosula* in the different periods.

when the greenhouse gas emission scenario is RCP 8.5, *C. pilosula* moved northward in 2050 and 2070, and its area expanded because, as a species that prefers a warm and humid climate, the increase in temperature and precipitation in the future may provide a more suitable growth environment for it. This finding is consistent with the conclusion that, on the premise that the temperature will continue to rise in the future, the suitable habitats of plants that prefer warm and humid climates will move to the north as a whole, and the zone of vegetation zone that adapts to cold and dry climates will gradually retreat (Zhang et al., 2017; Liu C. et al., 2018).

This study shows that since climate warming, the highly suitable habitats of *C. pilosula* are mainly in Sichuan, Gansu, Shaanxi, Shanxi, and Henan Provinces. The results are basically in line with the previous research results (Jie et al., 2017; Niu et al., 2021), which can provide a reference for the production layout of *C. pilosula* and the development of countermeasures against climate change. It should be pointed out that the potential suitable distribution of *C. pilosula* depends not only on the climatic conditions but also on the socio-economic structure,

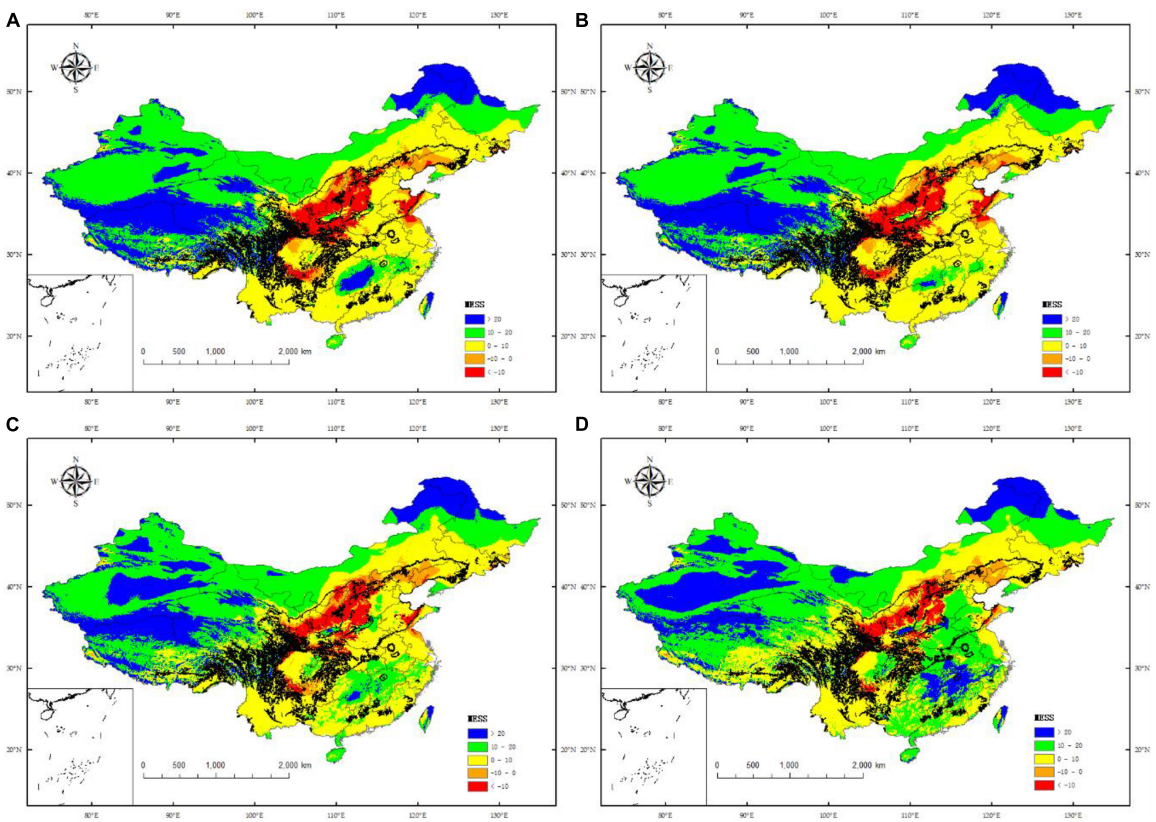


FIGURE 8 | Multivariate environmental similarity surface (MESS) analysis for *C. pilosula* in China under different combinations of climate change scenarios. **(A)** RCP 2.6-2050; **(B)** RCP 2.6-2070; **(C)** RCP 8.5-2050; **(D)** RCP 8.5-2070.

production technology level and other factors. At the same time, it is also affected by historical phenomena, human activities, geographical characteristics and genetic types. Therefore, there will be some deviation between the predicted results and the actual suitable distribution area of this species. Therefore, in actual production activities, the introduction, trial planting and planting of *C. pilosula* need to consider the role of various factors, especially the impact of economic factors and production factors.

A new method based on the MaxEnt model and ultra-high performance chromatographic fingerprint technique was developed to evaluate the effects of environmental factors on the quality of this species (Wan et al., 2021). Unlike this study, their research focuses on the distribution and quality of *C. pilosula* on a regional scale. In contrast, our research focuses on the distribution and influencing factors of this species on a national scale under climate change from the past to the future and considers modeling optimization. In addition, data sparsity is a problem in the MaxEnt modeling process (Haffner, 2006). Therefore, another future research direction is to combine this approach with other approaches, such as UPLC fingerprint technique, rough sets (Yan et al., 2016b,c, 2017a), Petri nets (Yan et al., 2016a), and a multidimensional cloud modeling (Yan et al., 2017b), which can effectively improve the space-time cost and processing speed in the age of big data (Wu et al., 2018a,b), and the results from these approaches can be compared with those provided by our model.

CONCLUSION

In this study, by setting two parameters of the MaxEnt model, namely, the regularization multiplier and the feature combination, we use the mean AUC_{DIFF}, OR_{10}, and delta AICc indicators to determine that the optimal regularization multiplier is 1.5, and the optimal parameter feature combination is LQHP. The main temperature factors affecting the distribution of *C. pilosula* are the annual mean temperature, mean diurnal range, and isothermality, and the main rainfall factors are the precipitation seasonality, precipitation in the wettest quarter, and precipitation in the driest quarter. Among these factors, the annual average temperature contributes the most to the distribution area of *C. pilosula*. When the annual average temperature is greater than 0°C, the distribution probability of *C. pilosula* rises sharply, reaches the peak value at 17.91°C, and reaches the optimum growth condition. Then, the distribution probability of *C. pilosula* decreases rapidly. The MaxEnt model predicts that the potential distribution habitats of *C. pilosula* early in the last interglacial period moved southward and that the

suitable area was the smallest. With the warming of the climate, the high-potential habitats gradual expanded to the north to Sichuan, Gansu, Shaanxi, Shanxi, and Henan Provinces, and the area is the largest in 2070. The introduction and cultivation in these areas cannot only avoid economic loss and resource waste caused by blind introduction but also improve the yield of this species.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

HY: conceptualization. HY and JH: funding acquisition. JH: project administration. GW: supervision. LT and LF: resources. LZ and LQ: methodology. XG: validation. XX and XY: data curation. HY: editing. YQ: visualization. All authors have read and agreed to the published version of the manuscript.

FUNDING

This study was supported by the Youth Fund of Chongqing Technology and Business University (grant no. 1952033) and the Opening Research Platform of Chongqing Technology and Business University (grant no. KFJJ2018058).

ACKNOWLEDGMENTS

We thank to the reviewers for the constructive and valuable comments and the editors for their assistance in refining this article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.773396/full#supplementary-material

REFERENCES

Abdelaal, M., Fois, M., Fenu, G., and Bacchetta, G. (2019). Using MaxEnt modeling to predict the potential distribution of the endemic plant *Rosa arabica* Crép. in Egypt. *Ecol. Inform.* 50, 68–75. doi: 10.1016/j.ecoinf.2019.01.003

Akaike, H. (1998). "Information theory and an extension of the maximum likelihood principle," in *Selected Papers of Hirotugu Akaike*, eds E. Parzen, K. Tanabe, and G. Kitagawa (New York, NY: Springer New York), 199–213. doi: 10.1007/978-1-4612-1694-0_15

Anderson, R. P., and Gonzalez, I. (2011). Species-specific tuning increases robustness to sampling bias in models of species distributions: an implementation with Maxent. *Ecol. Modell.* 222, 2796–2811. doi: 10.1016/j.ecolmodel.2011.04.011

Bi, H. Y., Zhang, L. P., Zhen, C., and Wu, B. (2008). An overview of research on germplasm resources of Radix Codonopsis and their utilization. *China J. Chin. Mater. Med.* 33, 590–594.

Burnham, K. P., and Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociol. Methods Res.* 33, 261–304. doi: 10.1177/0049124104268644

Chang, L. (2013). The climatic condition suitable for *Codonopsis pilosula* in changzi county Shanxi Province. *Chin. J. Agrometeorol.* 9, 152–153.

Chen, D. M., Kang, H. Z., and Liu, C. J. (2011). An overview on the potential quaternary glacial refugia of plants in China mainland. *Bull. Bot. Res.* 31, 623–632. doi: 10.7525/j.issn.1673-5102.2011.05.019

Duan, Q. M., Liang, Z. S., Yang, D. F., Liu, Y., and Liu, F. H. (2012). Establisthment of quality evaluation system and difference analysis on roots of *Codonopsis pilosula* from different habitats. *Chin. Tradit. Herb. Drugs* 43, 995–999.

Elith, J., Kearney, M., and Phillips, S. (2010). The art of modelling range-shifting species. *Methods Ecol. Evol.* 1, 330–342. doi: 10.1111/j.2041-210X.2010.00036.x

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., and Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* 17, 43–57. doi: 10.1111/j.1472-4642.2010.00725.x

Estes, L. D., Bradley, B. A., Beukes, H., Hole, D. G., Lau, M., Oppenheimer, M. G., et al. (2013). Comparing mechanistic and empirical model projections of crop suitability and productivity: implications for ecological forecasting. *Glob. Ecol. Biogeogr.* 22, 1007–1018. doi: 10.1111/geb.12034

Fang, Y. M., Wang, T., and Zhang, J. C. (2011). Analysis of quaternary vegetation changes along the Beijing-Hangzhou Grand Canal. *J. Nanjing For. Univ. Nat. Sci. Ed.* 35, 113–119.

Feng, P. P., Li, Z. X., and Yuan, Z. (2012). Review on the phytochemical and pharmacological study of *Codonopsis* Genus. *J. Shenyang Pharm. Univ.* 29, 61–65.

Fielding, A. H., and Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24, 38–49. doi: 10.1017/S0376892997000088

Fortin, M. J. (1999). Effects of sampling unit resolution on the estimation of spatial autocorrelation. *Ecoscience* 6, 636–641. doi: 10.1080/11956860.1999.11682547

Gong, J. W. (2011). Feasibility analysis of *Codonpsis pliosula* series developing to health care product. *J. Shanxi Med. Coll. Contin. Educ.* 21, 56–57.

Guo, J., Liu, X. P., Zhang, Q., Zhang, D. F., Xie, C. X., and Liu, X. (2017). Prediction for the potential distribution area of *Codonopsis pilosula* at global scale based on Maxent model. *Chin. J. Appl. Ecol.* 28, 992–1000. doi: 10.13287/j.1001-9332.201703.026

Haffner, P. (2006). "Fast transpose methods for kernel learning on sparse data," in *Proceedings of the 23rd International Conference on Machine Learning*, Pittsburgh, PA, 385–392.

Huang, Y. Y., Zhang, Y., Kang, L. P., Yu, Y., and Guo, L. P. (2018). Research progress on chemical constituents and their pharmacological activities of plant from *Codonopsis*. *Chin. Tradit. Herb. Drugs* 49, 239–250.

Jie, G., Xiao-ping, L., Qin, Z., Dong-fang, Z., Caixiang, X., and Xia, L. (2017). Prediction for the potential distribution area of *Codonopsis pilosula* at global scale based on Maxent model. *Chin. J. Appl. Ecol.* 28, 992–1000. doi: 10.13287/j.1001-9332.201703.026

Kumar, S., and Stohlgren, T. J. (2009). Maxent modeling for predicting suitable habitat for threatened and endangered tree *Canacomyrica monticola* in New Caledonia. *J. Ecol. Nat. Environ.* 1, 094–098.

Li, X., Li, Y., and Fang, Y. M. (2018). Prediction of potential suitable distribution areas of *Quercus fabri* in China based on an optimized Maxent model. *Sci. Silvae Sin.* 54, 153–164.

Liu, M. X., Qi, J., and Yu, B. Y. (2018). Research progress on the pharmacological activity of *Codonopsis pilosula*. *Strait Pharm. J.* 30, 36–39.

Liu, C., Huo, H. L., Tian, L. M., Dong, X. G., Qi, D., Zhang, Y., et al. (2018). Potential geographical distribution of *Pyrus calleryana* under different climate change scenarios based on the MaxEnt model. *Chin. J. Appl. Ecol.* 29, 3696–3704. doi: 10.13287/j.1001-9332.201811.016

Moreno-Amat, E., Mateo, R. G., Nieto-Lugilde, D., Morueta-Holme, N., Svenning, J.-C., and García-Amorena, I. (2015). Impact of model complexity on cross-temporal transferability in Maxent species distribution models: an assessment using paleobotanical data. *Ecol. Modell.* 312, 308–317. doi: 10.1016/j.ecolmodel.2015.05.035

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., et al. (2014). ENMeval: an R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods Ecol. Evol.* 5, 1198–1205. doi: 10.1111/2041-210X.12261

Niu, Y. T., Zhang, J., Wang, H. Z., Jin, L., Du, T., Song, P. S., et al. (2021). Study on planting suitability of *Codonopsis* radix. *Chin. J. Inform. TCM* 28, 13–16.

Pearson, R. G., Raxworthy, C. J., Nakamura, M., and Townsend Peterson, A. (2007). Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *J. Biogeogr.* 34, 102–117. doi: 10.1111/j.1365-2699.2006.01594.x

Peterson, A., Soberón, J., and Pearson, R. G. (2011). *Ecological Niches and Geographic Distributions*, Vol. 49. Princeton, NJ: Princeton University Press.

Phillips, S. J. (2006). A brief tutorial on Maxent. *AT T Res.* 190, 231–259.

Phillips, S. J., Anderson, R. P., and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* 190, 231–259. doi: 10.1016/j.ecolmodel.2005.03.026

Phillips, S. J., and Dudík, M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31, 161–175. doi: 10.1111/j.0906-7590.2008.5203.x

Porfirio, L. L., Harris, R. M. B., Lefroy, E. C., Hugh, S., Gould, S. F., Lee, G., et al. (2014). Improving the use of species distribution models in conservation planning and management under climate change. *PLoS One* 9:e113749. doi: 10.1371/journal.pone.0113749

Shcheglovitova, M., and Anderson, R. P. (2013). Estimating optimal complexity for ecological niche models: a jackknife approach for species with small sample sizes. *Ecol. Modell.* 269, 9–17. doi: 10.1016/j.ecolmodel.2013.08.011

Shepherd, L. D., de Lange, P. J., Perrie, L. R., and Heenan, P. B. (2017). Chloroplast phylogeography of New Zealand Sophora trees (Fabaceae): extensive hybridization and widespread last glacial maximum survival. *J. Biogeogr.* 44, 1640–1651. doi: 10.1111/jbi.12963

Su, N. N., Zhang, L. P., Wang, Y. F., and Shi, J. Q. (2012). Study on the seed germination percentage of *Codonopsis pilosula*. *Chin. Agric. Sci. Bull.* 28, 294–298.

Syfert, M. M., Smith, M. J., and Coomes, D. A. (2013). The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PLoS One* 8:e55158. doi: 10.1371/journal.pone.0055158

Tian, Z. P., and Jiang, D. B. (2015). Mid-Holocene and last glacial maximum changes in monsoon area and precipitation over China. *Chin. Sci. Bull.* 60, 400–410. doi: 10.1360/n972014-00718

Veloz, S. D. (2009). Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *J. Biogeogr.* 36, 2290–2299. doi: 10.1111/j.1365-2699.2009.02174.x

Wan, G.-Z., Wang, L., Jin, L., and Chen, J. (2021). Evaluation of environmental factors affecting the quality of *Codonopsis pilosula* based on chromatographic fingerprint and MaxEnt model. *Ind. Crops Prod.* 170:113783. doi: 10.1016/j.indcrop.2021.113783

Wang, Z. L., Wu, Z. H., Wu, Q. J., and Sun, C. (2006). Regional planning of ecological climate conditions suitable for planting Dangshen in Longnan of Gansu Province. *Chin. J. Agric. Resour. Reg. Plan.* 27, 32–35.

Warren, D. L., and Seifert, S. N. (2011). Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecol. Appl.* 21, 335–342. doi: 10.1890/10-1171.1

Warren, D. L., Wright, A. N., Seifert, S. N., and Shaffer, H. B. (2014). Incorporating model complexity and spatial sampling bias into ecological niche models of climate change risks faced by 90 California vertebrate species of concern. *Divers. Distrib.* 20, 334–343. doi: 10.1111/ddi.12160

Wei, B., Ma, S. M., Song, J., He, L. Y., and Li, X. C. (2019). Prediction of the potential distribution and ecological suitability of *Fritillaria walujewii*. *Acta Ecol. Sin.* 39, 228–234.

Wu, D., Luo, X., Wang, G., Shang, M., Yuan, Y., and Yan, H. Y. (2018a). A highly accurate framework for self-labeled semisupervised classification in industrial applications. *IEEE Trans. Ind. Inform.* 14, 909–920. doi: 10.1109/TII.2017.2737827

Wu, D., Shang, M. S., Luo, X., Xu, J., Yan, H. Y., Deng, W. H., et al. (2018b). Self-training semi-supervised classification based on density peaks of data. *Neurocomputing* 275, 180–191. doi: 10.1016/j.neucom.2017.05.072

Wu, X. J., Zhang, X. B., Guo, L. P., Yu, Y., and Hunag, L. Q. (2017). Study on distribution division of *Codonopsis* Radix. *China J. Chin. Mater. Med.* 42, 126–130. doi: 10.19540/j.cnki.cjcmm.2017.0186

Xu, Z. P., Zhang, J. Q., Wan, T., Cai, P., and Yi, W. D. (2017). Study on the history distribution pattern of *Gymnocarpos przewalskii* and refuge area. *Acta Bot. Boreali Occidentalia Sin.* 37, 2074–2081.

Xun, X. J. (2000). Pollen-based biome reconstruction at middle holocene (6ka BP) and last glacial maximum (18ka BP) in China. *Acta Bot. Sin.* 42, 1201–1209.

Yan, H. Y., Feng, L., Zhao, Y. F., Feng, L., Zhu, C. P., Qu, Y. F., et al. (2020b). Predicting the potential distribution of an invasive species, *Erigeron canadensis* L., in China with a maximum entropy model. *Glob. Ecol. Conserv.* 21:e00822. doi: 10.1016/j.gecco.2019.e00822

Yan, H. Y., Feng, L., Zhao, Y. F., Feng, L., Wu, D., and Zhu, C. P. (2020a). Prediction of the spatial distribution of *Alternanthera philoxeroides* in China based on ArcGIS and MaxEnt. *Glob. Ecol. Conserv.* 21:e00856. doi: 10.1016/j.gecco.2019.e00856

Yan, H. Y., He, J., Zhao, Y. F., Zhang, L., Zhu, C. P., and Wu, D. (2020c). Gentiana macrophylla response to climate change and vulnerability evaluation in China. *Glob. Ecol. Conserv.* 22:e00948. doi: 10.1016/j.gecco.2020.e00948

Yan, H. Y., Wang, G. Y., Zhang, X. R., Dong, J. H., Shan, K., Wu, D., et al. (2016b). A fast method to evaluate water eutrophication. *J. Cent. South Univ.* 23, 3204–3216. doi: 10.1007/s11771-016-3386-4

Yan, H. Y., Zhang, X. R., Dong, J. H., Shang, M. S., Shan, K., Wu, D., et al. (2016c). Spatial and temporal relation rule acquisition of eutrophication in Da'ning River based on rough set theory. *Ecol. Indic.* 66, 180–189. doi: 10.1016/j.ecolind.2016.01.032

Yan, H. Y., Huang, Y., Wang, G. Y., Zhang, X. R., Shang, M. S., Feng, L., et al. (2016a). Water eutrophication evaluation based on rough set and petri nets: a case study in Xiangxi-River, three gorges reservoir. *Ecol. Indic.* 69, 463–472. doi: 10.1016/j.ecolind.2016.05.010

Yan, H. Y., Wang, G. Y., Wu, D., Huang, Y., Shang, M. S., Xu, J. J., et al. (2017a). Water bloom precursor analysis based on two direction S-rough set. *Water Resour. Manag.* 31, 1435–1456. doi: 10.1007/s11269-017-1579-8

Yan, H. Y., Wu, D., Huang, Y., Wang, G. Y., Shang, M. S., Xu, J. J., et al. (2017b). Water eutrophication assessment based on rough set and multidimensional cloud model. *Chemometr. Intell. Lab. Syst.* 164, 103–112. doi: 10.1016/j.chemolab.2017.02.005

Yang, X.-Q., Kushwaha, S. P. S., Saran, S., Xu, J., and Roy, P. S. (2013). Maxent modeling for predicting the potential distribution of medicinal plant, *Justicia adhatoda* L. in lesser Himalayan foothills. *Ecol. Eng.* 51, 83–87. doi: 10.1016/j.ecoleng.2012.12.004

Zhang, D. F., Zhang, Q., Guo, J., Sun, C. Z., Wu, J., Nie, X., et al. (2017). Research on the global ecological suitability and characteristics of regions with *Angelica*

sinensis based on the MaxEnt model. *Acta Ecol. Sin.* 37, 5111–5120. doi: 10.5846/stxb201605030837

Zhang, Z. Q. (2006). *A Study on Standardization Cultivating Technology for Codonopsis pilosula Nannf var*. Xianyang: Northwest A&F University.

Zhao, G. F., Zhang, L. P., and Wu, B. (2006). Study on GAP of Radix *Codonopsis* and Draw up of Its SOP. *Res. Pract. Chin. Med.* 20, 13–16.

Zhu, G. P., Fan, J. Y., Wang, M. L., Chen, M., and Qiao, H. J. (2017). The importance of the shape of receiver operating characteristic(ROC) curve in ecological niche model evaluation - case study of *Hlyphantria cunea*. *J. Biosaf.* 26, 184–190.

Zhu, G. P., Liu, G. Q., Bu, W. J., and Gao, Y. B. (2013). Ecological niche modeling and its applications in biodiversity conservation. *Biodivers. Sci.* 21, 90–98. doi: 10.3724/sp.j.1003.2013.09106

Zhu, G. P., and Qiao, H. J. (2016). Maxent model's complexity on the prediction of species potential distributions. *Biodivers. Sci.* 24, 1189–1196. doi: 10.17520/biods.2016265

Zhu, G. P., Yuan, X. J., Fan, J. Y., and Wang, M. L. (2018). Effects of model parameters in MaxEnt modeling of ecological niche and geographic distribution: case study of the brown marmorated stink bug, *Halyomorpha haly*. *J. Biosaf.* 27, 118–123.

Conflict of Interest: XX and XY were employed by State Grid Chongqing Electric Power Company.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 24 December 2021
doi: 10.3389/fevo.2021.723558

Defining Biologically Meaningful Biomes Through Floristic, Functional, and Phylogenetic Data

Domingos Cardoso[1†‡], Peter W. Moonlight[2†‡], Gustavo Ramos[2,3†], Graeme Oatley[4†], Christopher Dudley[2], Edeline Gagnon[2†], Luciano Paganucci de Queiroz[5†], R. Toby Pennington[2,4†] and Tiina E. Särkinen[2†]*

[1] National Institute of Science and Technology in Interdisciplinary and Transdisciplinary Studies in Ecology and Evolution (INCT IN-TREE), Instituto de Biologia, Universidade Federal da Bahia, Salvador, Brazil, [2] Tropical Diversity Section, Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom, [3] Department of Biological Sciences, The University of Edinburgh, Edinburgh, United Kingdom, [4] Department of Geography, University of Exeter, Exeter, United Kingdom, [5] Departamento de Ciências Biológicas, Universidade Estadual de Feira de Santana, Feira de Santana, Brazil

While we have largely improved our understanding on what biomes are and their utility in global change ecology, conservation planning, and evolutionary biology is clear, there is no consensus on how biomes should be delimited or mapped. Existing methods emphasize different aspects of biomes, with different strengths and limitations. We introduce a novel approach to biome delimitation and mapping, based upon combining individual regionalizations derived from floristic, functional, and phylogenetic data linked to environmentally trained species distribution models. We define "core Biomes" as areas where independent regionalizations agree and "transition zones" as those whose biome identity is not corroborated by all analyses. We apply this approach to delimiting the neglected Caatinga seasonally dry tropical forest biome in northeast Brazil. We delimit the "core Caatinga" as a smaller and more climatically limited area than previous definitions, and argue it represents a floristically, functionally, and phylogenetically coherent unit within the driest parts of northeast Brazil. "Caatinga transition zones" represent a large and biologically important area, highlighting that ecological and evolutionary processes work across environmental gradients and that biomes are not categorical variables. We discuss the differences among individual regionalizations in an ecological and evolutionary context and the potential limitations and utility of individual and combined biome delimitations. Our integrated ecological and evolutionary definition of the Caatinga and associated transition zones are argued to best describe and map biologically meaningful biomes.

Keywords: biome delimitation, functional diversity, macroecology, phylogenetic diversity, SDTF, species distribution modeling, transition zones, bioregionalization

OPEN ACCESS

Edited by:
Danilo M. Neves, Federal University of Minas Gerais, Brazil

Reviewed by:
Thaís Guedes, State University of Maranhão, Brazil
Adeline Fayolle, University of Liège, Belgium

***Correspondence:**
Domingos Cardoso
cardosobot@gmail.com

***†ORCID**:*
Domingos Cardoso
orcid.org/0000-0001-7072-2656
Peter W. Moonlight
orcid.org/0000-0003-4342-2089
Gustavo Ramos
orcid.org/0000-0002-4539-394X
Graeme Oatley
orcid.org/0000-0003-0896-9994
Edeline Gagnon
orcid.org/0000-0003-3212-9688
Luciano Paganucci de Queiroz
orcid.org/0000-0001-7436-0939
R. Toby Pennington
orcid.org/0000-0002-8196-288X
Tiina E. Särkinen
orcid.org/0000-0002-6956-3093

‡These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *10 June 2021*
Accepted: *22 November 2021*
Published: *24 December 2021*

Citation:
Cardoso D, Moonlight PW, Ramos G, Oatley G, Dudley C, Gagnon E, Queiroz LP, Pennington RT and Särkinen TE (2021) Defining Biologically Meaningful Biomes Through Floristic, Functional, and Phylogenetic Data. Front. Ecol. Evol. 9:723558. doi: 10.3389/fevo.2021.723558

INTRODUCTION

Biomes are a key concept in ecology and biogeography (Higgins et al., 2016; Mucina, 2019) and have been largely used in global change ecology (Prentice et al., 2007; Williams et al., 2007; Lehmann et al., 2014; Moncrieff et al., 2016), conservation planning (Hoekstra et al., 2005), and evolutionary biology (Donoghue and Edwards, 2014; Landis et al., 2021a). Although biome definitions have differed (Mucina, 2019), the scientific community has generally settled on an

agreed biome definition: "a biotic community finding its expression at large geographic scales, shaped by climatic factors, and perhaps better characterized by physiognomy and functional aspects, rather than by species or life-form composition" (Mucina, 2019). Despite this accord over the definition of a biome, there remains no universally recognized method of delimiting and mapping biomes. Different approaches focus upon different elements of biomes – their physiognomic, floristic, environmental, or functional characteristics – which in turn produce different biome maps. Although such single-criterion-based biome schemes are helpful for understanding plant communities from an operational point of view, and at the local to the global scale (Conradi et al., 2020), they cannot define the nature of biomes through time and fail to capture the distribution, structure, and functioning of biomes in an evolutionary continuum.

Most recently, a global-scale conceptual view of biomes has been proposed which considers biomes as the confluence of ecology, evolution, and biogeography (Pennington et al., 2009, 2018; Oliveira-Filho et al., 2013a,b; Moncrieff et al., 2016; Pennington and Lavin, 2016; Mucina, 2019, 2020; Nürk et al., 2020; Ringelberg et al., 2020). While there is a consensus in ecology and biogeography that biomes should be defined based on physiognomy and functional aspects (Mucina, 2019; Pennington et al., 2018), an evolutionary dimension emphasizes the processes that have led to current biome distributions. This concept defines biomes as "evolutionary theaters" within which lineages interact and evolve through time, and as meta-communities regulated by community assembly at large spatial scale (Pennington et al., 2009; Oliveira-Filho et al., 2013b; Pennington and Lavin, 2016). The concept has emerged partly in response to increasing evidence for the prevalence of phylogenetic niche conservatism (Crisp et al., 2009; Pennington et al., 2009; Oliveira-Filho et al., 2013a,b; Kerkhoff et al., 2014; Gagnon et al., 2019; Ringelberg et al., 2020; Segovia et al., 2020). This tendency of plant lineages to inherit their overall ancestral environmental niches is based upon evidence that many plant lineages have dispersed across large distances over evolutionary timescales yet occupy similar ecological conditions. The general lack of dispersal limitation and difficulty of accruing novel environmental adaptations had led to the popularity of the phrase that for plants, it is "easier to move than evolve" (Donoghue, 2008).

That plant phylogenies are often more structured ecologically than geographically suggests that ecological gradients are evolutionarily important (Crisp et al., 2009; Oliveira-Filho et al., 2013a). The concept of phylogenetic niche conservatism is strongly linked to environmental filtering, i.e., the process whereby environmental gradients act as strong filters for species distributions (Cavender-Bares et al., 2009; Hardy et al., 2012; Blonder et al., 2015). For example, a previous study across plant lineages has shown that environmental filtering has played an important role in shaping the flora of the Galapagos Islands (Carvajal-Endara et al., 2017). In the presence of environmental filtering, lineages cannot successfully establish unless they have traits that leave them pre-adapted to pass environmental filters, leading to distinct biome assemblies (see also Donoghue and Edwards, 2014).

Biome conservatism and environmental filtering are of course not universal. For example, depending on the "evolutionary accessibility" of a new ecological setting or a lineage's biome affinity and location relative to the spatial distribution of other biomes at any given time, there is a varying spectrum of possible biogeographical scenarios, including those at which lineages can transcend ecological barriers more easily than geographical barriers (Edwards and Donoghue, 2013; Landis et al., 2021a). Evidence from global tropical grasslands shows that many lineages were able to colonize the biome over the past 10 million years from other biomes (Simon et al., 2009; Maurin et al., 2014), indicating that perhaps some biomes are more open to outsiders (i.e., non-native or pre-adapted lineages; Edwards and Donoghue, 2013; Donoghue and Edwards, 2014). Such evolutionary biome switching reflects that some environmental gradients (i.e., biome borders) are more easily crossed than others, perhaps due to the ease at which adaptations to these gradients can be acquired (Pennington et al., 2006; Simon and Pennington, 2012). The intermediate disturbance hypothesis (Connell, 1978) posits that if a disturbance is not too extreme, many plant lineages may already have or can evolve traits required to survive it, but the more extreme the disturbance (e.g., extreme drought and extreme cold or extreme heat), an increasingly small number of species will have these traits because they are hard to evolve. Quantitative evidence across floras is now needed to understand the relative roles of niche conservatism and the species-environment interactions (environmental filtering) across ecological gradients, particularly in the Neotropics where much biome complexity is found (Pennington et al., 2009; Hughes et al., 2013; Dexter et al., 2018; Silva-de-Miranda et al., 2018).

If phylogenetic niche conservatism and environmental filtering have shaped the macroevolutionary patterns of floristic and functional diversity that make up the evolutionary theaters or biomes, then exploring variation in taxonomic, functional trait, and community phylogenetic data may help delimiting more biologically meaningful biomes that are globally comparable. While biomes defined exclusively by individual, distinct operational criteria will result in different biome maps fit for different purposes such as comparative ecology and global change research (Conradi et al., 2020), comparisons of biomes delimited under different or combinations of criteria remain rare, particularly across geologically and ecologically complex biogeographic regions like the Caatinga domain of northeast Brazil (NE Brazil, **Figure 1**) (Queiroz et al., 2017; Fernandes et al., 2020).

The Caatinga region is often treated in biodiversity, evolutionary, conservation, and biogeographical studies of plants and animals as a single, homogeneous unit, generally termed a "biome" (Instituto Brasileiro de Geografia e Estatística [IBGE], 2012; The Brazil Flora Group [BFG], 2015; Garda et al., 2017; Mesquita et al., 2017; Araujo and Silva, 2017; Antonelli et al., 2018; Manhães et al., 2018; Nascimento et al., 2018; Silva and Souza, 2018; Medeiros et al., 2019; Prieto-Torres et al., 2019; Souza-e-Silva et al., 2019; Correia et al., 2020;

FIGURE 1 | The main biogeographical regions of northeast Brazil (including the state of Minas Gerais) where three biomes predominate: the geographically disjunct Amazonian and Atlantic rainforests [**(A)**, green; **(D)**, blue], the succulent-rich Caatinga seasonally dry tropical forests [**(B)**, red], and the grass-rich, fire-prone savannas of the Cerrado [**(C)**, yellow]. For geographic reference, the whole north and northeastern borders of the inset are the Atlantic Ocean. Photos by PWM **(A)** and DC **(B–D)**.

Dória and Dobrovolski, 2021). However, this definition of the Caatinga as a "biome" is in conflict with the more generally accepted definition of a biome at a global scale; it is in fact a biogeographic region. Whilst the Caatinga region may have a broadly similar, seasonally dry climate, it includes interdigitating, distinct biomes (as defined and recognized at global scale) such as rainforest and fire-prone savannas within a predominant matrix of Caatinga seasonally dry tropical forests (SDTF biome). The use of maps with a geographic delimitation of the Caatinga as a "biome" may impact upon downstream analyses aimed at disentangling the ecological and historical drivers that have shaped the evolutionary trajectories of all Caatinga species diversity (Queiroz, 2006; Cardoso and Queiroz, 2010; Queiroz et al., 2017; Guedes et al., 2014). It may also impact assessing priority areas for conservation in the severely impacted SDTF biome across Brazil, which it is not just confined to the Caatinga, but is found across the Cerrado and Pantanal regions (DRYFLOR, 2016; the Cerrado and Pantanal are also termed "biomes" by IBGE).

Here we explore how taxonomic, functional, and community phylogenetic data can be used to delimit biomes in NE Brazil and explore how biomes defined in these different ways are shaped by climatic variables. We aim to demonstrate that a Caatinga SDTF biome, lying at the extremely dry end of the tropical seasonality and rainfall gradient of NE Brazil, greatly differs from nearby and interdigitating ecologically and evolutionarily distinct rainforest and savanna biomes and occupies a unique environmental space. We use the three distinct data sources to measure the degree to which the Caatinga SDTF biome differs floristically, phylogenetically, and in functional trait composition, as first step to understand the ease at which these differences can and have been traversed adaptively by plant lineages through time (Wiens et al., 2010; Crisp and Cook, 2012; Donoghue and Edwards, 2014).

We use standard clustering algorithms, which in a geographic context have been termed regionalization approaches (Kreft and Jetz, 2010; Linder et al., 2012; Vilhena and Antonelli, 2015; Daru et al., 2017), to delimit clusters. Several authors have used biogeographic regionalization methods on a relatively similar (Linder et al., 2012; Fayolle et al., 2018; Aleman et al., 2020) or even larger scales (e.g., Kreft and Jetz, 2010; Holt et al., 2013; Vilhena and Antonelli, 2015; Ficetola et al., 2017) and argued that the results are comparable to biomes (Vilhena and Antonelli, 2015; Aleman et al., 2020). However, while extremely useful, previous regionalizations have been based on only a single type of data (e.g., floristic data; Linder et al., 2012) so cannot encompass all aspects of biomes. Our approach differs because we use three independent regionalizations, each based upon a different type of data (floristic, functional, and phylogenetic). By both combining and comparing the results of these regionalizations,

we argue that our results approach for the first time a data-driven and repeatable biome map that considers all facets of agreed biome definitions.

MATERIALS AND METHODS

Our approach differs from classical and modern approaches that fall under the umbrella of "biogeographical regionalization" or "bioregionalization" (e.g., Kreft and Jetz, 2010; Linder et al., 2012; Holt et al., 2013; Vilhena and Antonelli, 2015; Daru et al., 2016, 2017; Edler et al., 2017; Ficetola et al., 2017). Most such analyses are essentially focused on understanding the signature of historical (i.e., geographic barriers) processes in explaining the spatial distribution of specific "groups" of organisms across geographically confined areas, rather than delimiting biomes that are applicable across the tree of life. Those that do include broader ranges of taxa still rely only on floristic (e.g., Linder et al., 2012) or phylogenetic (e.g., Daru et al., 2017) data to classify the "bioregions," rather than attempting to integrate floristic, functional, and phylogenetic data. The resulting areas are more akin to "biogeographic regions" (sensu Wallace, 1876; Holt et al., 2013) than "biomes." In our study, biomes were delimited by comparing and combining independent regionalization analyses based upon floristic, functional, and phylogenetic data. We therefore consider the impact of geography and the evolutionary and functional distinctiveness of areas, as well as how the environment [here incorporated indirectly by the use of Species Distribution Models (SDMs)] defines where lineages are confined. We consider the results of our analyses biologically meaningful biomes.

Floristic data were based on thresholded, statistically significant SDMs for 3,457 flowering plant species, which were also included in a molecular phylogeny. SDMs were used to estimate species lists for all grid cells across the study area. Functional and phylogenetic data for each grid cell were based on seven functional traits and a community phylogeny generated for all the species present for the study area, respectively. Regionalization analyses were carried out using hierarchical clustering based upon floristic, functional, and phylogenetic distance among all cells. All three distance matrices (floristic, functional, and phylogenetic) were summed and used to delimit "total evidence" biomes.

Study Area

We applied our biome delimitation approaches to NE Brazil because of our long-term experience working on the taxonomy, distribution, ecology, and evolution of flowering plants in the region (e.g., Rocha et al., 2004; Queiroz, 2006; Queiroz et al., 2010, 2017; Särkinen et al., 2011; Santos et al., 2012; Cardoso et al., 2014; Fernandes et al., 2020; Moonlight et al., 2020). We defined our study area as NE Brazil (including the state of Minas Gerais) in order to include all areas defined as the Caatinga by Instituto Brasileiro de Geografia e Estatística [IBGE] (2012) and alternative biome classifications (e.g., Queiroz et al., 2017; Silva-de-Miranda et al., 2018; Moonlight et al., 2020). We have a particular interest in identifying "core" areas that are relatively homogenous in floristic, functional, and phylogenetic space, so it was important to include areas of all biomes that surround the Caatinga (i.e., areas known to differ in these respects). According to the Instituto Brasileiro de Geografia e Estatística [IBGE] (2012) classification, this includes the Cerrado (Brazilian savanna), Mata Atlantica (Brazilian Atlantic rainforest), and Amazonia (Amazon rainforest). Our study area has an area of 2.144 million km^2 and includes a buffer of at least 200 km around the IBGE definition of the Caatinga in all directions (**Figure 1**).

Species Distribution Data

The ongoing Flora do Brasil 2020 project (2021)[1] provides a robust taxonomic framework to work in the region with currently 16,351 species of flowering plants recorded for NE Brazil (Flora do Brasil 2020, 2021). A relatively large amount of species occurrence data is available for the region (**Supplementary Figure 1**) thanks to two dynamic network of local herbaria across Brazil available through CRIA speciesLink[2] and the Reflora specimen database[3].

In this study, we attempted to produce an SDM for every angiosperm species with recorded distributions in NE Brazil (Reflora, 2021) using data from CRIA speciesLink and the Reflora specimen database. While many NE Brazilian species have distributions outside Brazil, we only included Brazilian specimen records due to difficulties matching taxonomic backbones across countries. We used the latest version of Reflora (2021) to harmonize the taxonomy of the two specimen databases and update synonymy. Data were cleaned in six stages as described in Appendix S2 of Moonlight et al. (2020), which were designed to remove misidentified specimens and those with coordinate errors. We addressed environmental bias in occurrence data using spatial filtering by retaining a single specimen record within each 10 km radius for every species following Kramer-Schadt et al. (2013). We retained and attempted to model all species with $\geq$5 records collected in different grid cells at a 0.05° resolution, resulting in 296,439 unique occurrence records for 9,134 species.

Species Distribution Modeling

Climatic predictors were derived from remotely sensed temperature (MODIS; Wan and Dozier, 1996; Wan, 2014), rainfall (CHIRPS; Funk et al., 2014), and cloud cover data (MODCF; Wilson and Jetz, 2016). These data are calibrated with data from ground weather stations and outperform those derived from ground data alone (e.g., WorldClim) for SDMs (Deblauwe et al., 2016). All data were downloaded at a 0.05° resolution (c. 5.5 km^2 at the equator). Edaphic data were derived from the SoilGrids 250 m database (accessed February 2017)[4] interpolated to a 0.05° resolution. Edaphic factors are believed to be important in determining species distributions in NE Brazil (Queiroz et al., 2017) and have been shown to increase SDM performance in the tropical Americas (Figueiredo et al., 2017; Moulatlet et al., 2017;

[1]http://floradobrasil.jbrj.gov.br/
[2]http://www.splink.org.br/
[3]http://floradobrasil.jbrj.gov.br/reflora/herbarioVirtual/
[4]https://soilgrids.org/

Rapini et al., 2021). Climatic and edaphic predictors (35 and 55, respectively) were converted into two independent principal component analysis (PCA) axes. Four and five axes were selected, which each explained >80% of variation (see Appendix S4 of Moonlight et al., 2020). This process reduces the number of explanatory variables, thus minimizing collinearity (Dormann et al., 2013) and model overfitting (Peterson et al., 2007) while maximizing the explanatory data available for modeling.

SDMs for 9,134 were run using MaxEnt v.3.3.3 in the R package "dismo" (Hijmans et al., 2017). MaxEnt was chosen because it has been shown to outperform other SDM methodologies, particularly when species have few distribution points. MaxEnt was used with the default settings, with 5-fold cross-validation, and all feature classes allowed. Background data (also known as pseudo-absence data) were sampled from a unique area for each species, consisting of NE Brazil plus circles of 250 km around each species' known occurrence points. This was a compromise between predicting species into areas not covered by background points, providing a large number of climatically unsuitable points (Anderson and Raza, 2010), and including a biologically realistic extent for each species' model. We controlled for bias in sampling effort (Stolar and Nielsen, 2015) by selecting 10,000 background points for each species using an Epanechnikov kernel (Wiegand and Moloney, 2013) calculated from all angiosperm presence data for Brazil.

We evaluated model performance using the Continuous Boyce Index (CBI; Hirzel et al., 2006). This evaluation index that relies upon presence and pseudo-absence data is based upon the Boyce Index (Boyce et al., 2002), calculated with code available at https://rdrr.io/github/adamlilith/enmSdm/man/contBoyce.html. CBI has been shown to be less stochastic to variation at low numbers of presence points than alternative indices (Hirzel et al., 2006). A CBI of $\geq$0 indicates that a model is better than chance and from the SDMs of 9,134 species we retained all models with a mean CBI of $\geq$0.25 over the five replicates (6,823 species, 75%, see **Supplementary Table 3**). Retained replicates were summed and converted into binary presence-absence maps using the 10th percentile logistic threshold. This was chosen because it was the strictest of the commonly used threshold values, so it limits the well-documented over-estimated of summed SDMs. To maximize the compatibility of analyses, we retained models only for species included in the community phylogeny (see section "Phylogenetic Delimitation" below) for downstream analyses (3,457 species). Predicted species lists were estimated for every 0.05° grid cell in NE Brazil, which were aggregated to a 0.25° resolution due to computation constraints. Presence in a species list was assigned based upon predicted presence in any constituent cell at the original 0.05° resolution.

Cluster Delimitation

Floristic Delimitation

To delimit clusters based upon floristic data (floristic regionalization), a distance matrix was computed based on β-diversity (Simpson's dissimilarity: β-sim) in the R package "betapart" (Baselga et al., 2018). This approach was chosen because it measures floristic turnover (i.e., dissimilarity) between grid cells and not nestedness (Baselga, 2010). The floristic distance matrix was used in unbiased cluster analysis, where the row order of the distance matrix was randomized 100 times using the "recluster" package in R (Dapporto et al., 2013, 2015). RogueNaRok (Aberer et al., 2012; Available at: https://rnr.h-its.org/) was used to identify rogue grid cells responsible for reducing resolution in the resulting 50% majority rule consensus dendrogram. A total of 138 rogue grid cells were identified and removed. Clusters were mapped based upon a process of reciprocal illumination following Moonlight et al. (2020) and analogous to the approaches taken by similar analyses (Silva-de-Miranda et al., 2018; Moonlight et al., 2020). Clusters were labeled based on comparing the mapped distribution of sub-clusters to our biological knowledge of the vegetation patterns in NE Brazil. Our priorities were to delimit clusters that could be matched with confidence to four biomes recognized by the Instituto Brasileiro de Geografia e Estatística [IBGE] (2012) classification, and to maximize similarity with the mapped clusters from the phylogenetic and functional biome classifications (see below). We acknowledge that this approach is not fully repeatable, but argue that it is the best currently available considering that alternative approaches (e.g., k-means clustering, Amaral et al., 2017) also rely upon prior knowledge to define an expected number of biomes. We term this approach "floristic regionalization."

Phylogenetic Delimitation

To delimit clusters based upon phylogenetic distances between plant communities (phylogenetic regionalization), we produced a novel community phylogeny for the flora of NE Brazil based on DNA sequences mostly downloaded from GenBank (Benson et al., 2017). We attempted to download sequence data for all 16,351 species recorded in NE Brazil by Flora do Brasil 2020 (2021) for the following regions: *matK*, *atpB*, *ndhF*, *rbcL*, and *trnL*. Regions were chosen based upon: (i) wide use across angiosperms; (ii) ease of alignment across angiosperms; (iii) adequate level of sequence variation across orders, families, and genera.

To augment our species sampling in the community phylogeny, we have newly generated 445 *matK* and 444 *rbcL* sequences from herbarium and field-collected leaf tissues preserved through silica gel desiccation of 546 species. DNA extraction, PCR amplification, and robotic sequencing largely followed standard protocols of DNA barcoding for community phylogenetics (e.g., Kress et al., 2009). The newly generated DNA sequences are publicly available in GenBank (see **Supplementary Table 1** for voucher and accession number details)[5].

The R package "rentrex" (Winters, 2017) was used to query GenBank, specifying a sequence length of 500–5,000 base pairs. For species in NE Brazil for which no sequence data were available in GenBank, we repeated the steps above to locate sequences for congeners from outside of Brazil (for genera with one or two species within NE Brazil) because any one or two species for these genera would provide the same phylogenetic distance in our analyses as those present in NE Brazil. Alignments

[5]https://www.ncbi.nlm.nih.gov/genbank/

were done in Mafft v.7.450 for each DNA region with default settings with a maximum of six iterations performed per region. Data cleaning was done by identifying poorly aligned sequences based on visual assessment, using Vsearch v2.14.2 to identify highly variable sequences with <40% sequence similarity, and by identifying misplaced species based on neighbor-joining trees run for each DNA region with FastTree v2.1.10 (Price et al., 2010). Replacements for any species removed during the cleaning were searched in GenBank if alternative sequences were available. The final cleaned community phylogeny contained 10,279 sequences from GenBank and 662 newly generated sequences for a total of 6,296 species (**Supplementary Table 2**). All DNA regions were combined and analyzed using RaxML-HPC2 on XSEDE on the CIPRES Science Gateway v.3.3 on-line portal (Miller et al., 2010)[6]. The phylogeny was rooted with Nymphaeales. The community phylogeny included 6,296 species from 209 families and 1,775 genera (**Supplementary Figure 2**). The relationships among them were consistent with the phylogeny of flowering plants (The Angiosperm Phylogeny Group et al., 2016). A total of 3,457 species from 184 families and 1,325 genera in the phylogeny were also included in the SDMs so were retained for downstream analyses.

Phylogenetic distances between all grid cells in NE Brazil were calculated by estimating the phylogenetic β-diversity among estimated species lists. The phylogenetic Simpson's index was used following Chave et al. (2007) because it is comparable with β-sim (see above). Phylogenetic regionalization was carried out following the hierarchical clustering method described above under section "Floristic Delimitation." No rogue grid cells were identified and removed.

Functional Delimitation

To delimit clusters based upon functional distances between plant communities (functional regionalization), we scored seven independent plant traits for all 3,457 species for which SDMs were generated and which were included in the community phylogeny (100% coverage for 3,457 species for all seven traits). These included 931 species scored by Moonlight et al. (2020). Traits were chosen on the basis of: (i) that they were simple and unambiguous to score from herbarium specimens or literature; and (ii) had hypothesized links with environmental and functional differences among biomes in NE Brazil. The seven traits chosen were: (i) latex; (ii) corky bark; (iii) spines; (iv) compound leaves; (v) nitrogen nodulation; (vi) Crassulacean acid metabolism (CAM) photosynthesis; and (vii) C4 photosynthesis (**Supplementary Table 2**). The literature used to score traits is detailed in Appendix S7 of Moonlight et al. (2020).

Functional distance matrices between all grid cells in NE Brazil were created based upon estimated species lists. Euclidean distances of grid cells in 6-dimensional trait space based on the proportion of species with each trait in each grid cell was used for measuring functional distance. Published functional diversity metrics (e.g., Ricotta et al., 2016) were found inappropriate for measuring functional distance between plant communities because they are based on the presence rather than the proportion of a trait, which leads to overestimation of functional similarity at broad spatial scales where almost all grid cells had at least one predicted species with every trait. Functional regionalization was carried out following the hierarchical clustering method described above under section "Floristic Delimitation." No rogue grid cells were identified and removed.

Combined Cluster Delimitation

Combined analyses of floristic, phylogenetic, and functional distance matrices were run to delimit "total evidence" clusters based on all three approaches. All possible combinations of the three approaches were also run to see subsets of results (i.e., floristic + phylogenetic, floristic + functional, and phylogenetic + functional). Distance matrices were scaled from 0 to 1 to give equal weight to each matrix before being summed, so the distance matrix values could range from 0 to 2 in analyses with two approaches, or 0 to 3 in analyses with three approaches. Clusters were estimated, mapped, and named following the hierarchical clustering method described under section "Floristic Delimitation." The results of the "total evidence" clusters are named herein as biomes because they are the result of three, independent lines of evidence. No rogue grid cells were removed.

Comparison of Regionalizations: Clusters as Biomes

Comparisons of regionalizations based on individual approaches (functional, phylogenetic, and floristic) were carried out in both geographic and environmental space to highlight areas of "core" biomes (areas where all regionalizations analyses agreed on biome identity) and transition zones (areas where regionalizations differed between analyses). "Core" biome areas were visualized by highlighting areas of agreement between the three individual regionalizations (e.g., "core Caatinga"). Transitional biome areas were visualized by highlighting areas of disagreement where one or two approaches showed disagreement regarding biome distribution (e.g., "transitional Caatinga"). A raster file depicting the core Caatinga and associated transition zones is available as **Supplementary Figure 5**.

To investigate whether core biomes overlapped in environmental space, environmental comparison of biome delimitations was achieved by plotting clusters from different analyses in environmental space based on mean annual temperature (bio1) and rainfall data (bio12) extracted for each grid cell. The bio1 and bio12 values for each biome grid cell from each analysis were summed. The mean annual temperature (bio1) from MODIS temperature data (Wan and Dozier, 1996; Wan, 2014) and mean annual rainfall (bio12) from CHIRPS rainfall data (Funk et al., 2014) were used.

RESULTS

Major Biomes Across NE Brazil

All three individual regionalizations identified two clusters of Caatinga SDTF, two clusters as Mata Atlantica, and three clusters as Cerrado (**Figures 2A–C**). The functional analysis resulted in the identification of an additional cluster within both the Mata Atlantica and Cerrado, and an extra cluster that we were unable

[6]www.phylo.org

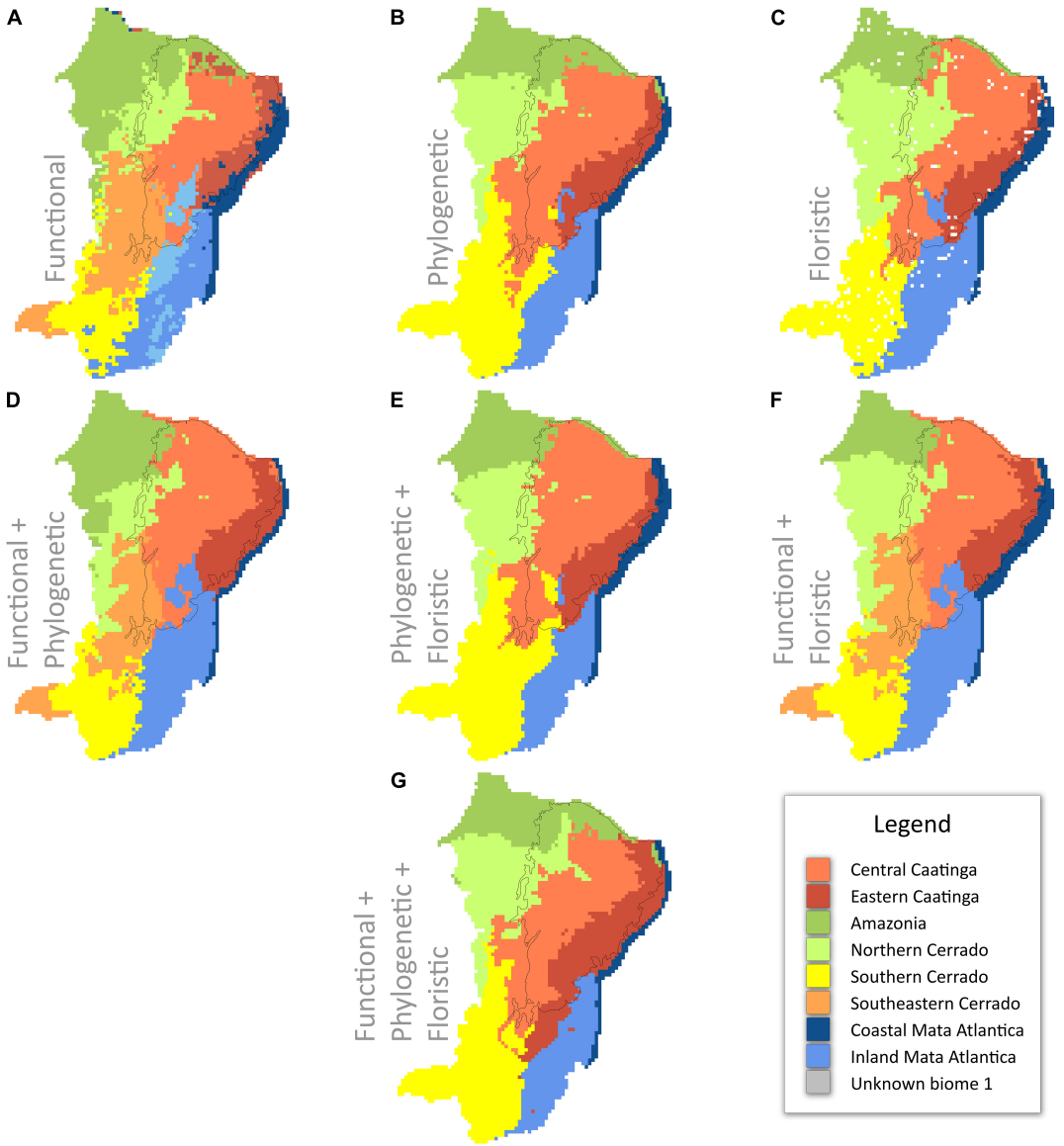


FIGURE 2 | Cluster classifications across northeast Brazil based upon hierarchical clustering analyses of functional, phylogenetic, and floristic data alone **(A–C)** and in combination **(D–F)**: **(A)** functional data; **(B)** phylogenetic data; **(C)** floristic data; **(D)** functional and phylogenetic data; **(E)** phylogenetic and floristic data; **(F)** functional and floristic data; and **(G)** functional, phylogenetic, and floristic data. The black line indicates the limits of the Caatinga domain sensu Instituto Brasileiro de Geografia e Estatística (IBGE) (2012).

to assign to any recognized biome (**Figure 2A**). The combined "total evidence" analyses of all three approaches suggested seven major clusters based on functional, phylogenetic, and floristic data (**Figure 2G**).

Differences Between Approaches

The areas where differences between the three regionalizations are seen can be considered as transitional areas (**Figure 3** and **Supplementary Figure 4**). Most disagreement between the three individual regionalizations is seen in the southern part of the Caatinga biogeographic domain in the Chapada Diamantina area (**Figures 2**, **3**). Functional and floristic regionalizations identify most of this area as Mata Atlantica but the phylogenetic regionalization identifies most of these areas as Cerrado (**Figures 2**, **3**).

Comparison of clusters delimited by single approaches shows that clusters delimited by floristic and phylogenetic approaches are highly similar both in spatial extent and number of major clusters where both analyses resolve seven major clusters in largely similar areas across NE Brazil (**Figures 2**, **3**). Functional clusters show largest differences compared to phylogenetic and floristic clusters, and resolve ten major clusters indicating higher resolution of functional data for biome delimitation within NE Brazil, enabling further splitting of vegetation types as

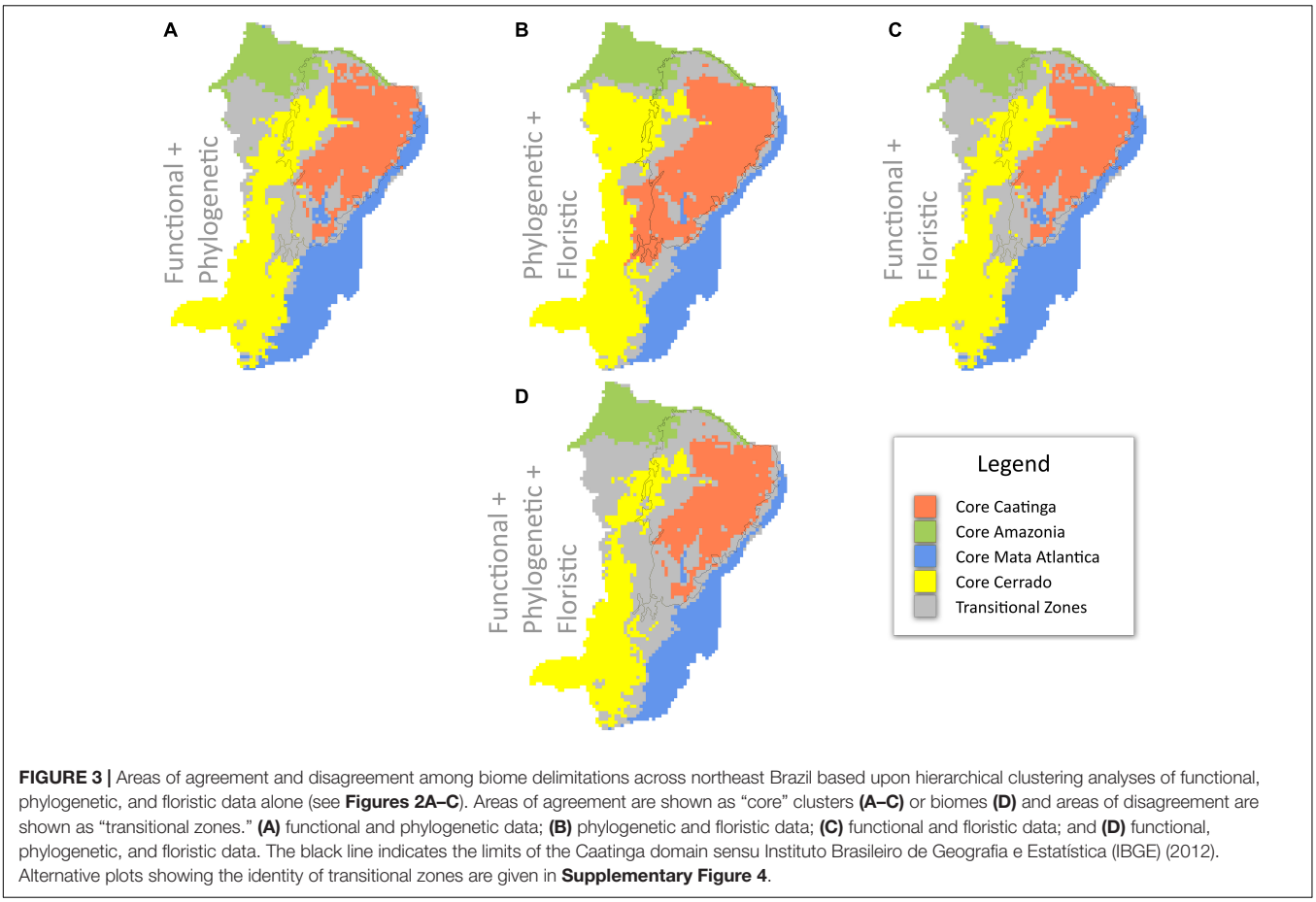


FIGURE 3 | Areas of agreement and disagreement among biome delimitations across northeast Brazil based upon hierarchical clustering analyses of functional, phylogenetic, and floristic data alone (see **Figures 2A–C**). Areas of agreement are shown as "core" clusters **(A–C)** or biomes **(D)** and areas of disagreement are shown as "transitional zones." **(A)** functional and phylogenetic data; **(B)** phylogenetic and floristic data; **(C)** functional and floristic data; and **(D)** functional, phylogenetic, and floristic data. The black line indicates the limits of the Caatinga domain sensu Instituto Brasileiro de Geografia e Estatística (IBGE) (2012). Alternative plots showing the identity of transitional zones are given in **Supplementary Figure 4**.

functionally distinct clusters (**Figures 2**, **3** and **Supplementary Figure 3**). This is despite our functional classification being based on only seven categorical traits.

The "Core" Caatinga

In terms of the SDTF biome, all three analyses identified two groups as Caatinga SDTF (**Figures 2A–C**). A large area within NE Brazil is identified here as the "core Caatinga" biome supported by all three regionalizations (functional, phylogenetic, and floristic) (**Figure 4**). Areas surrounding the "core" Caatinga are identified as "transitional Caatinga" (**Figure 4**): these include areas supported as Caatinga by one or two regionalizations but not all three. Transitional Caatinga is found across the western and southern borders of the "core" Caatinga, much less so along the eastern side along the boundary with the coastal Mata Atlantica domain (**Figures 3**, **4**). Floristic and phylogenetic data support transitional Caatinga in the South and West of the "core" Caatinga, but functional data identify these areas as savanna-like Cerrado (e.g., Chapada do Araripe; **Figures 2**, **3**). In common with the individual analyses, all mapped Caatinga groups were largely congruent with the Caatinga domain along its eastern border but included several differences along the eastern border, which are discussed in detail below.

Our integrated biome delimitation using a combination of floristic, functional, and phylogenetic data identifies the "core" Caatinga within the driest and hottest areas of NE Brazil (**Figure 4A**). These areas are dominated by a floristically, functionally, and phylogenetically distinct flora adapted to dry conditions with seasonal rainfall. This "core" Caatinga area is found in areas with 353–1,271 mm annual rainfall and with a mean annual temperature from 19.6 to 28.0°, and does not overlap with other core biomes in environmental space (**Figure 4A**). The flora of the "core" Caatinga is characterized by a high proportion of species that are succulent, nodulating (**Figure 4B**), and with spines but without corky bark (**Figure 4C**). The low rainfall in the "core" Caatinga is notable. Dry forests in the Americas are found in areas with up to 1800 mm rainfall (DRYFLOR, 2016; Dexter et al., 2018) but none of the "core" Caatinga dry forests approach this threshold, despite the prevalence of areas with higher rainfall in the wider study area of NE Brazil.

DISCUSSION

Core Biomes and Transition Zones

Biomes have been delimited in various ways, including based on the spatial distribution of physiognomic, floristic or functional discontinuities amongst plant communities (Mucina, 2019) or even dynamic global vegetation biome modeling involving the

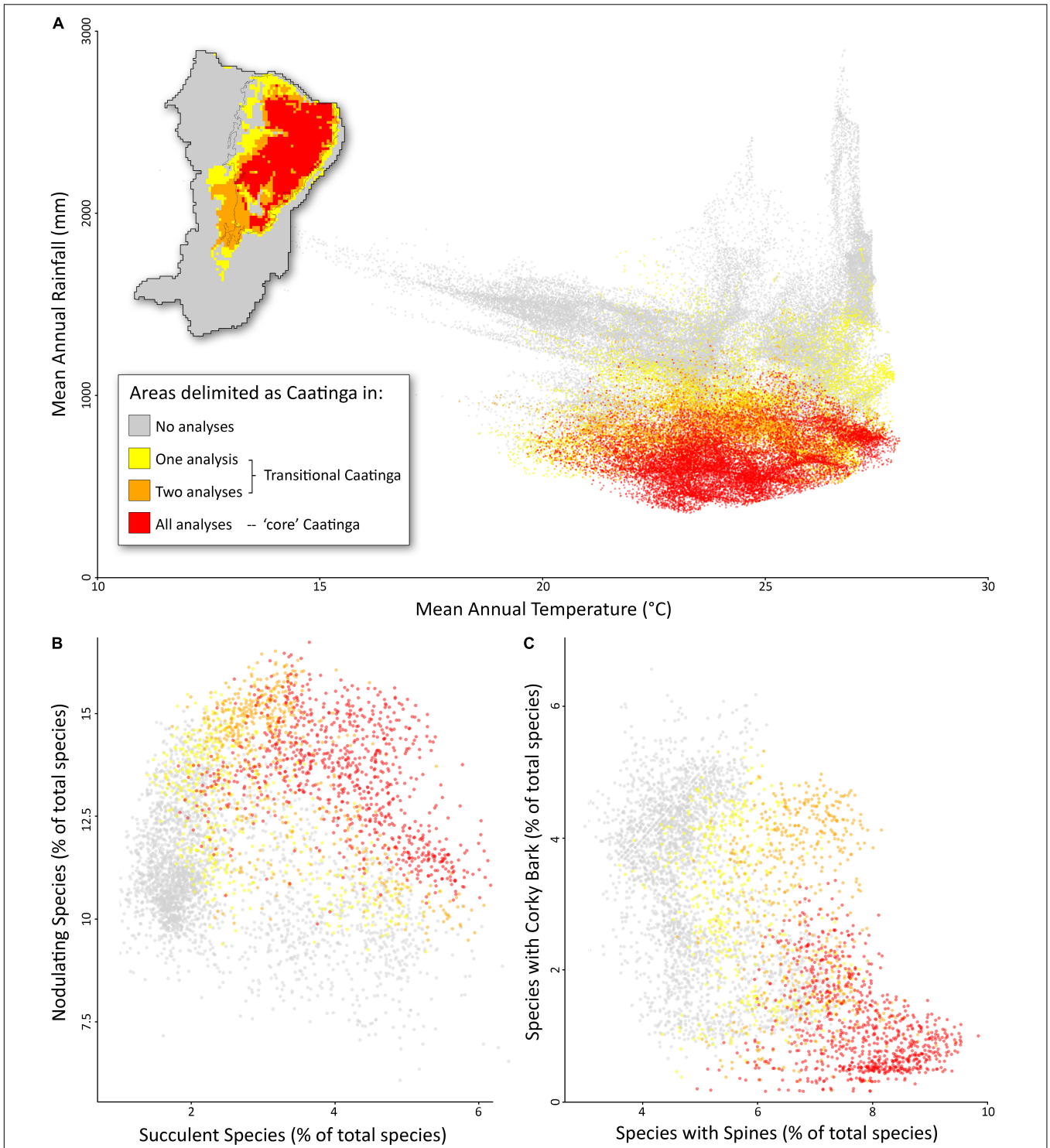


FIGURE 4 | Distribution of the Caatinga seasonally dry forest biome within NE Brazil in geographic, environmental **(A)** and trait **(B,C)** space. Colors indicate areas delimited as Caatinga in no analyses (gray), one or two analyses ("transitional Caatinga," yellow and orange), and three analyses ("core" Caatinga, red). The black line in the map indicates the limits of the Caatinga domain sensu Instituto Brasileiro de Geografia e Estatística (IBGE) (2012).

combination of physical environment, plant functional types, physiology, and biochemical fluxes (Prentice et al., 1992; Kaplan et al., 2003). Physiognomic delimitations have followed plant growth form and canopy structure (Woodward et al., 2004); floristic maps have focused on dominant plant families, genera, and species and the associated ecological correlates (e.g., in

Africa, White, 1983; Power et al., 2017; Aleman et al., 2020; South America, Silva-de-Miranda et al., 2018); and functional systems on the presence of functional groups, such as evergreen trees found in the tropical rainforest biome, succulents in the seasonally dry tropical forest or succulent biome, and a continuous grass layer in savannas (Whittaker, 1970; Scholes and Archer, 1997; Schrire et al., 2005). Additionally, strictly environmental-based delimitations of biomes based on climate, edaphic composition or degree of fire disturbance have provided important insights into our understanding of the ecological limits driving the assembly of plant communities (Archibald et al., 2013; Langan et al., 2017). These different approaches based upon different types of data produce different biome maps, but no map produced using a single type of data can possibly satisfy all aspects of the multifaceted definition of a biome expected to have shaped their biodiversity over evolutionary time, i.e., delimited biomes may not be floristically, functionally, or ecologically distinct.

Our approach differs in two ways from previous methods depicting global or regional distribution of biomes, including the Caatinga (e.g., Whittaker, 1970; White, 1983; Schrire et al., 2005; Instituto Brasileiro de Geografia e Estatística [IBGE], 2012; Conradi et al., 2020; Ringelberg et al., 2020): firstly, we identify "core" biome areas supported by three distinct lines of evidence (floristic, functional, and phylogenetic), and secondly, we identify transition zones that are supported by distinct data sources that point to biologically important areas of transition previously neglected by most biome maps. Our integrated biome analyses are able to highlight "core" areas where all three data sources (floristic, functional, and phylogenetic) agree on the distribution of the same biome (**Figure 4**). We suggest that such areas of congruence based on multiple lines of evidence may be of particular interest to biology and earth system science.

Ultimately, the identification of biologically relevant areas should depend upon its intended purpose. If you are interested in the response of the Caatinga to climate change, for example, you should focus on the areas that independent data sources agree are "core" Caatinga (**Figure 4**), not areas where data sources disagree (i.e., transition zones), where the species, ecological function, or phylogenetic diversity may overlap that of other biomes. Likewise, if you wish to measure the functional traits of a species, you measure the traits of an individual at the "core" of the species concept, not a hybrid of dubious identification. Also, because the transition zones may be particularly dynamic or vulnerable to climate change (i.e., "zones of tension," Clements, 1905), if your interest is in conserving the maximum species diversity and ecosystem function of Caatinga dry forests in the face of climate change, it may be best to model a set of species/geographic area/set of traits that all data agree are Caatinga.

The "Core" Caatinga

The Caatinga and other seasonally dry tropical forests have often been considered equivalent to the trans-continentally distributed "succulent biome," typically characterized by nutrient-rich substrates with little water holding capacity, and by a highly seasonal, drought-prone climate in which succulent plant lineages (e.g., columnar and arborescent members of Agavaceae, Bromeliaceae, Cactaceae, and Euphorbiaceae) evolved and diversified under a regime of regular and prolonged drought and without the influences of fire and other physical disturbances (Schrire et al., 2005; Pennington et al., 2009; Ringelberg et al., 2020). The climate space of the global succulent biome as recently modeled by the distribution of stem succulents involves a mean annual rainfall that closely matches that of South American SDTFs (Ringelberg et al., 2020), which also have a high proportion of stem succulents (DRYFLOR, 2016; Queiroz et al., 2017). Also at intercontinental scales, Segovia et al. (2020) underlines the structuring of evolutionary diversity of trees in the Neotropics along precipitation gradients. Precipitation-related bioclimatic variables were singled out as the most important precipitation measures predicting the succulent biome with a MaxEnt approach to large-scale biome distribution modeling of South American SDTFs (Särkinen et al., 2011). These same precipitation variables were found by Oliveira-Filho et al. (2013a) to be most important in distinguishing the succulent (including the Caatinga) and savanna (including the Cerrado) biomes. The analysis of community phylogenetic distances and the biome assignments of 466 floristic sites across eastern Brazil were best explained, of all the bioclimatic variables, by only annual precipitation at a threshold of <1,200 mm (Oliveira-Filho et al., 2013a). This ecological structure reflects the link between annual precipitation and phylogenetic niche conservatism (Oliveira-Filho et al., 2013a; Segovia et al., 2020). Substrate conditions were shown to be ecologically significant (with or without water-holding capacity), but this ecological variable may not mask the more important influence of annual precipitation as explanatory of the community phylogenetic structure (Oliveira-Filho et al., 2013a).

While comparative analyses have employed multiple taxa to interrogate biomes, other approaches seek to identify key taxa or functional groups that can be used to delimit biomes. One such technique was employed by Ringelberg et al. (2020), who used modeled distributions of a single functional group (i.e., stem succulents) to map the "succulent biome" (Schrire et al., 2005; Ringelberg et al., 2020). The results do not adequately capture the biome complexity in Brazil. For example, Ringelberg et al. (2020) delimit not just the Caatinga as succulent biome, but also parts of the Chaco and the campos rupestres of the Chapada Diamantina, which are ecologically, historically, and functionally distinct biomes (Pennington et al., 2000; DRYFLOR, 2016; Silva-de-Miranda et al., 2018; Rapini et al., 2021). Indeed, we demonstrate that not just the "core" Caatinga but also transitional and non-Caatinga areas have high proportions of stem succulents (**Figure 4B**). Such single-criterion approaches to biome delimitation do not differ much from an operational view of biomes (Conradi et al., 2020). If biomes are to be defined as evolutionary theaters, the distributions of as general group of taxa as possible should be examined in conjunction with their phylogenetic relationships and functional characteristics. Our integrated ecological and evolutionary approach involving multiple taxa across all growth forms and associated measurements of functional and phylogenetic diversity seem to better describe "core" evolutionary arenas.

Our results show that the "core" Caatinga area is in fact narrower than the widely used definition of Caatinga by Instituto

Brasileiro de Geografia e Estatística [IBGE], 2012. It excludes the wetter end of dry forests and transitional areas between savanna and dry forests (**Figure 4**). Not only that, but other major biomes exist within the Caatinga region (Queiroz et al., 2017). Disagreement between the three regionalizations in the Chapada Diamantina region, for example, highlights the biological reality of complex variation across environmentally highly heterogeneous areas, and should not be ignored. Thus, areas delimited by an essentially geographic approach such as that used by Instituto Brasileiro de Geografia e Estatística [IBGE] (2012) should not be termed "biomes." These approaches not only oversimplify the complexity of the interdigitating nature of biomes, but also disregard the global nature dimension of biomes, which is fundamental to a more biologically realistic use of biomes in ecology, conservation, and evolutionary biogeography.

Transitional Biome Areas

One of the perhaps most important messages of our study relates to areas where the three regionalizations disagree in their cluster delimitation (**Figure 3**). These areas of disagreement between floristic, functional, and phylogenetic data indicate the presence of transition zones between two or more biomes, presumably along environmental gradients. An alternative method of visualizing transition zones is presented in **Supplementary Figure 3**. Most previous biome delimitation analyses have not included transition zones, treating biomes as categorical variables with no intermediates (e.g., Instituto Brasileiro de Geografia e Estatística [IBGE] (2012); see **Figure 1**). Here we highlight transition zones as biologically interesting areas in their own right that cover a large percentage of our study area. For example, the area we delimit as "core" Caatinga covers 420k km^2 and the areas we delimit as transition zones cover 480k km^2 (254k km^2 delimited as Caatinga by two analyses; 227k km^2 delimited as Caatinga by one analysis), or 14% larger than the "core" Caatinga itself.

The existence and extent of transition zones raises an important question: what may underlie the differences among the results from our individual regionalizations and what may these differences tell us about biomes in NE Brazil? Transition zones may reveal areas along environmental gradients between "core" biomes where sets of traits from more than one biome may permit survival, as suggested by the "environmental crossroads hypothesis" (Neves et al., 2020). Additionally, we predict that differences in the adaptations required to successfully persist under different rainfall regimes lead to distinct floristic and functional compositions across biomes. The level of phylogenetic distinctiveness, on the other hand, will depend on the relative ease at which plants have acquired adaptations required for crossing environmental gradients, i.e., the level of phylogenetic niche and trait conservatism (Donoghue, 2008; Crisp et al., 2009).

Our results show that similar clusters are delimited using phylogenetic and floristic data across NE Brazil, some of which we suggest correspond to biomes that can be recognized at global scale, including SDTF, rainforest, and savanna. This indicates that phylogenetic niche conservatism is operating and preventing plant lineages from crossing environmental gradients across evolutionary time (e.g., extreme drought constrained species adaptations to successfully thrive in dry forests identified here as Caatinga). Functional clusters, however, show clear differences to floristic and phylogenetic clusters, both in the number of clusters and the geographic distribution of those clusters (**Figure 3**). The differences between the functional (**Figure 3A**) and phylogenetic (**Figure 3B**) regionalizations indicate that the functional traits included in our analyses are not conserved across the phylogeny of the angiosperms, or that phylogenetic trait conservatism acts differently for different traits. This may relate to the specific traits used here (e.g., we do not expect spines or corky bark to be conserved) but our analyses do include several well conserved traits (e.g., nodulation, latex, and C4 photosynthesis). The differences between the floristic (**Figure 3A**) and functional (**Figure 3C**) regionalizations may indicate that the possession of certain trait combinations may allow limited sets of species to span borders created by ecological and evolutionary processes.

The identity of clusters identified in the functional analysis that could not be linked to any previously indicated biomes is unclear. They might reflect that the functional data is capturing the complexity with respect to the high within-biomes habitat heterogeneity across NE Brazil. For example, each of the most predominant savanna, rainforest, and the "core" Caatinga seasonally dry forest biomes are not physiognomically, floristically, and edaphically homogeneous, rather they also exhibit highly specialized habitats. Just in the "core" Caatinga, plant communities seem to have been evolutionarily structured by major soil types like karst, sand, and crystalline (Santos et al., 2012; Moro et al., 2015; Queiroz et al., 2017); the savanna here encompasses all the campos rupestres vegetation on mountaintops of the Chapada Diamantina, which mostly involve fire-sensitive plant lineages, including succulents (Rapini et al., 2021); and the Mata Atlantica involves habitats as distinct as the more open coastal restingas (Scarano, 2002; Oliveira et al., 2014; Fernandes and Queiroz, 2015). These habitats may not be unique in their phylogenetic or floristic composition as potentially distinct biomes but are in their functional composition.

Agreement in clusters delimited by all three regionalizations in areas such as along the eastern border of the "core" Caatinga suggests ecological filtering is acting in conjunction with phylogenetic niche conservatism along the same environmental gradient (Caatinga-Mata Atlantica border; **Figure 4**). Transitional Caatinga areas in the southern parts of the Caatinga biogeographic domain indicate that ecological filtering is acting along different environmental gradients to phylogenetic niche conservatism, at least for some of the traits included in our analyses.

Biologically Meaningful Biomes in Evolutionary Biogeography

Comparative phylogenetic approaches of multiple taxa to understand the evolutionary history of the biodiversity in species-rich, yet geologically and climatically complex regions like the Neotropics have revealed important insights into how lineages and species interact with ecology and geography over evolutionary time. Such analyses often employ the reconstruction of ancestral areas across phylogenies to estimate

the rates of biotic interchange between biogeographic regions. For example, Antonelli et al. (2018) showed an impressively large number of dispersal events out of Amazonia to other major neotropical regions, where at least 85 species among angiosperms, birds, ferns, frogs, mammals, and squamates inhabiting the Caatinga region were inferred to have derived from an Amazonian ancestor. Biogeographic regions are often treated as homogeneous units (i.e., synonymous with biomes) in estimates of ancestral distributions. This approach cannot take into account the complexity of climatic, evolutionary, and functional spaces that confine species ecologically as we demonstrate for NE Brazil here. As such, it risks conflating the primary roles of geography and ecology, such that biome switching is likely to be overestimated in ecologically confined groups with broad distributions but underestimated in biogeographically confined groups with broad ecological distributions. Thus, such comparative approaches cannot allow us to deepen our understanding of whether the "core" Caatinga or the other putatively distinct biomes of NE Brazil (**Figure 3**) are more evolutionarily accessible to lineages from Amazonian tropical rainforests. In other words, some Amazon to Caatinga "switches" may be to "rainforests" within the Caatinga biogeographic region and thus represent geographic movements, not evolutionary switches. Likewise, most studies neglect that the grass-rich, fire-prone savanna biome that predominates in the Cerrado domain is criss-crossed by a network of gallery forests (i.e., rainforests; Silva-de-Miranda et al., 2018), as well as fire-sensitive seasonally dry, evergreen or semideciduous forests, depending on water availability, in patches of high fertility soils (Oliveira-Filho et al., 2013b; Bueno et al., 2018).

That lineages can transcend geographic barriers is also a product of the evolutionary accessibility of distinct biomes that create geographical opportunity (Edwards and Donoghue, 2013). So, by inferring biotic interchange across biogeographic regions (e.g., Antonelli et al., 2018), without considering the biotic complexity within them, it is impossible to uncover the true balance of phylogenetic biome conservatism versus evolutionary biome shifts. Fortunately, recent progress has been made toward developing more biogeographically realistic approaches that model how lineages shift between biomes depending on the temporal availability and geographical connectivity of biomes (Landis et al., 2021a,b). A definition that captures the most biologically meaningful nature of a biome in space and time is clearly critical for comparative analyses involving Landis et al.'s (2021a) phylogenetic biome shift model. Thus, we believe that our approach of combining the multiple dimensions describing the spatial distribution of biomes will help to more reliably map and understand the evolution and functioning of biodiversity.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

DC, PWM, and TS designed the study and led the writing with significant contributions from all co-authors. DC, GR, and PWM entered functional trait data. GO generated sequence data for community phylogeny. CD designed the pipeline for producing community phylogenies. PWM led all analyses. All authors commented and agreed on the last version of the manuscript.

FUNDING

This research was funded by the Natural Environment Research Council-Newton grant NE/N012526/1 and Fundação de Amparo à Pesquisa do Estado de São Paulo grant 2015/50488-5 "Nordeste: New Science for a Neglected Biome," and Royal Society Advanced Fellowship grant NAF/R1/180331. DC's research in plant biodiversity is also funded by Fundação de Amparo à Pesquisa do Estado da Bahia (Universal no. APP0037/2016) and Conselho Nacional de Desenvolvimento Científico e Tecnológico Research Productivity PQ-2 grant 308244/2018-4.

ACKNOWLEDGMENTS

We thank Centro de Referência em Informação Ambiental (CRIA) and Reflora for sharing their complete distribution datasets; Alexandre Brunello, Luiza Cosme, Moabe Fernandes, Raquel Miatto, Marcelo Mizushima, Tony Oliveira, Oliver Phillips, Desirée Ramos, Raphael Rocha, Rubens Santos, Valdemir Silva, and Elmar Veenendaal for helping collecting silica-gel dried leaf samples during fieldwork; Glaucia Drummond (Fundação Biodiversitas, Canudos) and Eudes Velozo (Fazenda Esperança, Boa Vista do Tupim) for their support during fieldwork; Matt Lavin and two reviewers for their valuable suggestions to improve the manuscript; and Associate Editor Danilo Neves for inviting us to contribute with the special issue "Temporal and large-scale spatial patterns of plant diversity and diversification." DC acknowledges Jon Lloyd for inviting him to be speaker of the 44th New Phytologist Symposium "Determinants of tropical vegetation structure and function" in Accra, Ghana, when most ideas of this article were originally presented.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.723558/full#supplementary-material

Supplementary Figure 1 | Species richness across northeast Brazil showing **(A)** "raw" species richness from cleaned occurrence data for 9,134 species with ≥5 records; **(B)** "raw" species richness from cleaned occurrence data for 3,547 species included in all analyses; **(C)** modeled species richness for 3,547 species included in all analyses.

Supplementary Figure 2 | Community phylogeny for northeast Brazil.

Supplementary Figure 3 | Dendrograms showing the major clusters from the cluster analyses based upon: **(A)** phylogenetic data; **(B)** floristic data; **(C)**

functional data; **(D)** taxonomic and functional data combined; **(E)** taxonomic and phylogenetic data combined; **(F)** functional and phylogenetic data combined; **(G)** taxonomic, functional, and phylogenetic data combined. Clusters are colored by their biome classification and are mapped in geographic space in **Figures 2A–G**.

Supplementary Figure 4 | Areas of agreement and disagreement among biome classifications across northeast Brazil based upon hierarchical clustering analyses of functional, phylogenetic, and floristic data alone (see **Figures 2A–C**). Areas of agreement ("core" biomes) are highlighted as per the legend and areas of disagreement ("transitional zones") are shown with intermediate colors between core biomes. **(A)** functional and phylogenetic data; **(B)** phylogenetic and floristic data; **(C)** functional and floristic data; and **(D)** functional, phylogenetic, and floristic data. The black line indicates the limits of the Caatinga domain sensu Instituto Brasileiro de Geografia e Estatística [IBGE] (2012).

Supplementary Figure 5 | Raster file used to produce **Figure 4A** that highlights the "core" Caatinga and associated transition zones.

Supplementary Table 1 | Voucher details for DNA sequences newly generated in this study.

Supplementary Table 2 | Species included in the analyses, including taxonomic family, functional traits, and model performance statistics.

Supplementary Table 3 | The relationship between Continuous Boyce Index (CBI) and the number of independent records by species.

REFERENCES

Aberer, A. J., Krompass, D., and Stamatakis, A. (2012). Pruning rogue taxa improves phylogenetic accuracy. *Syst. Biol.* 62, 162–166. doi: 10.1093/sysbio/sys078

Aleman, J. C., Fayolle, A., Favier, C., Staver, A. C., Dexter, K. G., Ryan, C. M., et al. (2020). Floristic evidence for alternative biome states in tropical Africa. *Proc. Natl. Acad. Sci. U.S.A.* 117, 28183–28190. doi: 10.1073/pnas.2011515117

Amaral, A. G., Munhoz, C. B. R., Walter, B. M. T., Aguirre-Gutiérrez, J., and Raes, N. (2017). Richness pattern and phytogeography of the Cerrado herb–shrub flora and implications for conservation. *J. Veg. Sci.* 28, 848–858. doi: 10.1111/jvs.12541

Anderson, R. P., and Raza, A. (2010). The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus Nephelomys) in Venezuela. *J. Biogeogr.* 37, 1378–1393. doi: 10.1111/j.1365-2699.2010.02290.x

Antonelli, A., Zizka, A., Carvalho, F. A., Scharn, R., Bacon, C. D., Silvestro, D., et al. (2018). Amazonia is the primary source of Neotropical biodiversity. *Proc. Natl. Acad. Sci. U.S.A.* 115, 6034–6039. doi: 10.1073/pnas.1713819115

Araujo, H. F. P., and Silva, J. M. C. (2017). "The avifauna of the Caatinga: biogeography, ecology, and conservation," in *Caatinga: The Largest Tropical Dry Forest Region in South America*, eds J. M. C. Silva I, R. Leal, and M. Tabarelli (Cham: Springer), 181–210.

Archibald, S., Lehmann, C. E. R., Gómez-Dans, J. L., and Bradstock, R. A. (2013). Defining pyromes and global syndromes of fire regimes. *Proc. Natl. Acad. Sci. U.S.A.* 110, 6442–6447. doi: 10.1073/pnas.1211466110

Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Glob. Ecol. Biogeogr.* 19, 134–143. doi: 10.1111/j.1466-8238.2009.00490.x

Baselga, A., Orme, D., Villéger, S., de Bortolli, J., and Leprieur, F. (2018). *Package 'betapart': Partitioning Beta Diversity into Turnover and Nestedness. R Package Version 1.5.0*. Available online at: https://CRAN.R-project.org/package=betapart (accessed April 11, 2021).

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., et al. (2017). GenBank. *Nucleic Acid Res.* 45, D37–D42. doi: 10.1093/nar/gkw1070

Blonder, B., Nogués-Bravo, D., Borregaard, M. K., Donoghue, J. C., Jørgensen, P. M., Kraft, N. J. B., et al. (2015). Linking environmental filtering and disequilibrium to biogeography with a community climate framework. *Ecology* 96, 972–985. doi: 10.1890/14-0589.1

Boyce, M. S., Vernier, P. R., Nielsen, S. E., and Schmiegelow, F. K. A. (2002). Evaluating resource selection functions. *Ecol. Modell.* 157, 281–300. doi: 10.1016/s0304-3800(02)00200-4

Bueno, M. L., Dexter, K. G., Pennington, R. T., Pontara, V., Neves, D. M., Ratter, J. A., et al. (2018). The environmental triangle of the Cerrado domain: ecological factors driving shifts in tree species composition between forests and savannas. *J. Ecol.* 106, 2109–2120. doi: 10.1111/1365-2745.12969

Cardoso, D. B. O. S., and Queiroz, L. P. (2010). "Caatinga no contexto de uma metacomunidade: evidências da biogeografia, padrões filogenéticos e abundância de espécies em leguminosas," in *Biogeografia da América do Sul: Padrões e Processos*, eds C. J. B. Carvalho and E. A. B. Almeida (São Paulo: Roca), 241–260.

Cardoso, D. B. O. S., Queiroz, L. P., and Lima, H. C. (2014). A taxonomic revision of the South American papilionoid genus *Luetzelburgia* (Fabaceae). *Bot. J. Linn. Soc.* 175, 328–375. doi: 10.1111/boj.12153

Carvajal-Endara, S., Hendry, A. P., Emery, N. C., and Davies, J. (2017). Habitat filtering not dispersal limitation shapes oceanic island floras: species assembly of the *Galápagos archipelago*. *Ecol. Lett.* 20, 495–504. doi: 10.1111/ele.12753

Cavender-Bares, J., Kozak, K., Fine, P., and Kembel, S. (2009). The merging of community ecology and phylogenetic biology. *Ecol. Lett.* 12, 693–715. doi: 10.1111/j.1461-0248.2009.01314.x

Chave, J., Chust, G., and Thébaud, C. (2007). "The importance of phylogenetic structure in biodiversity studies," in *Scaling Biodiversity*, eds D. Storch, P. Marquet, and J. H. Brown (Santa Fe: Institute Editions), 151–167.

Clements, F. E. (1905). *Research Methods in Ecology*. Lincoln, NB: University of Nebraska Publishing Company.

Connell, J. H. (1978). Diversity in tropical rain forests and coral reefs. *Science* 199, 1302–1310. doi: 10.1126/science.199.4335.1302

Conradi, T., Slingsby, J. A., Midgley, G. F., Nottebrock, H., Schweiger, A. H., and Higgins, S. I. (2020). An operational definition of the biome for global change research. *New Phytol.* 227, 1294–1306. doi: 10.1111/nph.16580

Correia, I., Nascimento, E. R., and Gouveia, S. F. (2020). Effects of climate and land-use gradients on avian phylogenetic and functional diversity in a tropical dry forest. *J. Arid Environ.* 173:104024. doi: 10.1016/j.jaridenv.2019.104024

Crisp, M. D., and Cook, L. G. (2012). Phylogenetic niche conservatism: what are the underlying evolutionary and ecological causes? *New Phytol.* 196, 681–694. doi: 10.1111/j.1469-8137.2012.04298.x

Crisp, M. D., Arroyo, M. T. K., Cook, L. G., Gandolfo, M. A., Jordan, G. J., McGlone, M. S., et al. (2009). Phylogenetic biome conservatism on a global scale. *Nature* 458, 754–756. doi: 10.1038/nature07764

Dapporto, L., Ramazzotti, M., Fattorini, S., Talavera, G., Vila, R., and Dennis, R. L. H. (2013). Package recluster: an unbiased clustering procedure for beta-diversity turnover. *Ecography* 36, 1070–1075. doi: 10.1111/j.1600-0587.2013.00444.x

Dapporto, L., Ramazzotti, M., Fattorini, S., Talavera, G., Vila, R., and Dennis, R. L. H. (2015). *Package Recluster: Ordination Methods for The analysis of Beta-Diversity Indices. R. Package Version 2.8*. Available online at: https://CRAN.R-project.org/package=recluster (accessed July 26, 2020).

Daru, B. H., Elliott, T. L., Park, D. S., and Davies, T. J. (2017). Understanding the processes underpinning patterns of phylogenetic regionalization. *Trends Ecol. Evol.* 32, 845–860. doi: 10.1016/j.tree.2017.08.013

Daru, B. H., Van der Bank, M., Maurin, O., Yessoufou, K., Schaefer, H., Slingsby, J. A., et al. (2016). A novel phylogenetic regionalization of the phytogeographic zones of southern Africa reveals their hidden evolutionary affinities. *J. Biogeogr.* 43, 155–166. doi: 10.1111/jbi.12619

Deblauwe, V., Droissart, V., Bose, R., Sonké, B., Blach-Overgaard, A., Svenning, J.-C., et al. (2016). Remotely sensed temperature and precipitation data improve species distribution modelling in the tropics. *Global Ecol. Biogeogr.* 25, 443–454. doi: 10.1111/geb.12426

Dexter, K. G., Pennington, R. T., Oliveira-Filho, A. T., Bueno, M. L., Silva de Miranda, P. L., and Neves, D. M. (2018). Inserting tropical dry forests into the discussion on biome transitions in the tropics. *Front. Ecol. Evol.* 6:104. doi: 10.3389/fevo.2018.00104

Donoghue, M. J. (2008). A phylogenetic perspective on the distribution of plant diversity. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11549–11555. doi: 10.1073/pnas.0801962105

Donoghue, M. J., and Edwards, E. J. (2014). Biome shifts and niche evolution in plants. *Annu. Rev. Ecol. Evol. Syst.* 45, 547–572. doi: 10.1146/annurev-ecolsys-120213-091905

Dória, T. A. F., and Dobrovolski, R. (2021). Improving post-2020 conservation of terrestrial vertebrates in Caatinga. *Biol. Conserv.* 253:108894. doi: 10.1016/j.biocon.2020.108894

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 27–46. doi: 10.1111/j.1600-0587.2012.07348.x

DRYFLOR. (2016). Plant diversity patterns in neotropical dry forests and their conservation implications. *Science* 353, 1383–1387. doi: 10.1126/science.aaf5080

Edler, D., Guedes, T., Zizka, A., Rosvall, M., and Antonelli, A. (2017). Infomap bioregions: interactive mapping of biogeographical regions from species distributions. *Syst. Biol.* 66, 197–204. doi: 10.1093/sysbio/syw087

Edwards, E. J., and Donoghue, M. J. (2013). Is it easy to move and easy to evolve? Evolutionary accessibility and adaptation. *J. Exp. Bot.* 64, 4047–4052. doi: 10.1093/jxb/ert220

Fayolle, A., Swaine, M. D., Aleman, J., Azihou, A. F., Bauman, D., te Beest, M., et al. (2018). A sharp floristic discontinuity revealed by the biogeographic regionalization of African savannas. *J. Biogeogr.* 46, 454–465. doi: 10.1111/jbi.13475

Fernandes, M. F., and Queiroz, L. P. (2015). Floristic surveys of Restinga forests in southern Bahia, Brazil, reveal the effects of geography on community composition. *Rodriguésia* 66, 51–73. doi: 10.1590/2175-7860201566104

Fernandes, M. F., Cardoso, D., and Queiroz, L. P. (2020). An updated plant checklist of the Brazilian Caatinga seasonally dry forests and woodlands reveals high species richness and endemism. *J. Arid Environ.* 174:104079. doi: 10.1016/j.jaridenv.2019.104079

Ficetola, G. F., Mazel, F., and Thuiller, W. (2017). Global determinants of zoogeographical boundaries. *Nat. Ecol. Evol.* 1:0089. doi: 10.1038/s41559-017-0089

Figueiredo, O. G., Zuquim, G., Tuomisto, H., Moulatlet, G. M., Balslev, H., and Costa, F. R. C. (2017). Beyond climate control on species range: the importance of soil data to predict distribution of Amazonian plant species. *J. Biogeogr.* 45, 190–200. doi: 10.1111/jbi.13104

Flora do Brasil 2020 (2021). *Jardim Botânico do Rio de Janeiro*. Available online at: http://floradobrasil.jbrj.gov.br/ (accessed Jun 04, 2021).

Funk, C. C., Peterson, P. J., Landsfeld, M. F., Pedreros, D. H., Verdin, J. P., Rowland, J. D., et al. (2014). A quasi-global precipitation time series for drought monitoring. *U.S. Geological Survey Data Series* 832, 1–4. doi: 10.3133/ds832

Gagnon, E., Ringelberg, J. J., Bruneau, A., Lewis, G. P., and Hughes, C. E. (2019). Global succulent biome phylogenetic conservatism across the pantropical *Caesalpinia* group (*Leguminosae*). *New Phytol.* 222, 1994–2008. doi: 10.1111/nph.15633

Garda, A. A., Stein, M. G., Machado, R. B., Lion, M. B., Juncá, F. A., and Napoli, M. F. (2017). "Ecology, biogeography, and conservation of amphibians of the Caatinga," in *Caatinga: The Largest Tropical Dry Forest Region in South America*, eds J. M. C. Silva I, R. Leal, and M. Tabarelli (Cham: Springer), 133–150. doi: 10.1007/978-3-319-68339-3_5

Guedes, T. B., Sawaya, R. J., and Nogueira, C. C. (2014). Biogeography, vicariance and conservation of snakes of the neglected and endangered Caatinga region, north-eastern Brazil. *J. Biogeogr.* 41, 919–931. doi: 10.1111/jbi.12272

Hardy, O. J., Couteron, P., Munoz, F., Ramesh, B. R., and Pélissier, R. (2012). Phylogenetic turnover in tropical tree communities: impact of environmental filtering, biogeography and mesoclimatic niche conservatism. *Global Ecol. Biogeogr.* 21, 1007–1016. doi: 10.1111/j.1466-8238.2011.00742.x

Higgins, S. I., Buitenwerf, R., and Moncrieff, G. R. (2016). Defining functional biomes and monitoring their change globally. *Glob. Change Biol.* 22, 3583–3593. doi: 10.1111/gcb.13367

Hijmans, R. J., Phillips, S., Leathwick, J., and Elith, J. (2017). *Package Dismo: Species Distribution Modelling. R Package Version 1.1-4*. Available online at: https://CRAN.R-project.org/package=dismo (accessed October 11, 2021).

Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., and Guizan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecol. Modell.* 199, 142–152. doi: 10.1016/j.ecolmodel.2006.05.017

Hoekstra, J. M., Boucher, T. M., Ricketts, T. H., and Roberts, C. (2005). Confronting a biome crisis: global disparities of habitat loss and protection. *Ecol. Lett.* 8, 23–29. doi: 10.1111/j.1461-0248.2004.00686.x

Holt, B. G., Lessard, J. P., Borregaard, M. K., Fritz, S. A., Araújo, M. B., Dimitrov, D., et al. (2013). An update of Wallace's zoogeographic regions of the world. *Science* 339, 74–78. doi: 10.1126/science.1228282

Hughes, C. E., Pennington, R. T., and Antonelli, A. (2013). Neotropical plant evolution: assembling the big picture. *Bot. J. Linn. Soc.* 171, 1–18. doi: 10.1111/boj.12006

Instituto Brasileiro de Geografia e Estatística [IBGE] (2012). *Manual Técnico da Vegetação Brasileira (2a Edição Revista e Ampliada)*. Rio de Janeiro: IBGE.

Kaplan, J. O., Bigelow, N. H., Prentice, I. C., Harrison, S. P., Bartlein, P. J., Christensen, T. R., et al. (2003). Climate change and arctic ecosystems: 2. Modeling, paleodata-model comparisons, and future projections. *J. Geophys. Res.* 108:8171. doi: 10.1029/2002JD002559

Kerkhoff, A. J., Moriarty, P. E., and Weiser, M. D. (2014). The latitudinal species richness gradient in New World woody angiosperms is consistent with the tropical conservatism hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8125–8130. doi: 10.1073/pnas.1308932111

Kramer-Schadt, S., Niedballa, J., Pilgrim, J. D., Schröder, B., Lindenborn, J., Reinfelder, V., et al. (2013). The importance of correcting for sampling bias in MaxEnt species distribution models. *Divers. Distrib.* 19, 1366–1379. doi: 10.1111/ddi.12096

Kreft, H., and Jetz, W. (2010). A framework for delineating biogeographical regions based on species distributions. *J. Biogeogr.* 37, 2029–2053. doi: 10.1016/j.ympev.2017.01.018

Kress, W. J., Erickson, D. L., Jones, F. A., Swenson, N. G., Perez, R., Sanjur, O., et al. (2009). Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18621–18626. doi: 10.1073/pnas.0909820106

Landis, M., Edwards, E. J., and Donoghue, M. J. (2021a). Modeling phylogenetic biome shifts on a planet with a past. *Syst. Biol.* 70, 86–107. doi: 10.1093/sysbio/syaa045

Landis, M., Eaton, D. A. R., Clement, W. L., Park, B., Spriggs, E. L., Sweeney, P., et al. (2021b). Joint phylogenetic estimation of geographic movements and biome shifts during the global diversification of *Viburnum*. *Syst. Biol.* 70, 67–85. doi: 10.1093/sysbio/syaa027

Langan, L., Higgins, S. I., and Scheiter, S. (2017). Climate-biomes, pedo-biomes or pyro-biomes: which world view explains the tropical forest–savanna boundary in South America? *J. Biogeogr.* 44, 2319–2330. doi: 10.1111/jbi.13018

Lehmann, C. E. R., Anderson, T. M., Sankaran, M., Higgins, S. I., Archibald, S., and Hoffmann, W. A. (2014). Savanna vegetation-fire-climate relationships differ among continents. *Science* 343, 548–552. doi: 10.1126/science.1247355

Linder, H. P., De Klerk, H. M., Born, J., Burgess, N. D., Fjeldsa, J., and Rahbek, C. (2012). The partitioning of Africa: statistically defined biogeographical regions in sub-Saharan Africa. *J. Biogeogr.* 39, 1189–1205. doi: 10.1111/j.1365-2699.2012.02728.x

Manhães, A. P., Loyola, R., Mazzochini, G. M., Ganade, G., Oliveira-Filho, A. T., and Carvalho, A. R. (2018). Low-cost strategies for protecting ecosystem services and biodiversity. *Biol. Conserv.* 217, 187–194. doi: 10.1016/j.biocon.2017.11.009

Maurin, O., Davies, T. J., Burrows, J. E., Daru, B. H., Muasya, K. Y. A. M., van der Bank, M., et al. (2014). Savanna fire and the origins of the 'underground forests' of Africa. *New Phytol.* 204, 201–214. doi: 10.1111/nph.12936

Medeiros, E. S. S., Machado, C. C. C., Galvíncio, J. D., Moura, M. S. B., and Araujo, H. F. P. (2019). Predicting plant species richness with satellite images in the largest dry forest nucleus in South America. *J. Arid Environ.* 166, 43–50. doi: 10.1016/j.dib.2019.104335

Mesquita, D. O., Costa, G. C., Garda, A. A., and Delfim, F. R. (2017). "Species composition, biogeography, and conservation of the Caatinga lizards," in *Caatinga: The Largest Tropical Dry Forest Region in South America*, eds J. M. C. Silva, I. R. Leal, and M. Tabarelli (Cham: Springer), 151–180. doi: 10.1007/978-3-319-68339-3_6

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). "Creating the CIPRES science gateway for inference of large phylogenetic trees," in *Proceedings of the Gateway Computing Environments Workshop (GCE)*, (New Orleans).

Moncrieff, G. R., Bond, W. J., and Higgins, S. I. (2016). Revising the biome concept for understanding and predicting global change impacts. *J. Biogeogr.* 43, 863–873. doi: 10.1111/jbi.12701

Moonlight, P. W., Silva-de-Miranda, P. L., Cardoso, D., Dexter, K. G., Oliveira-Filho, A. T., Pennington, R. T., et al. (2020). The strengths and weaknesses of

species distribution models in biome delimitation. *Global Ecol. Biogeogr.* 29, 1770–1784.

Moro, M. F., Silva, I. A., Araújo, F. S., Lughadha, E. N., Meagher, T. R., and Martins, F. R. (2015). The role of edaphic environment and climate in structuring phylogenetic pattern in seasonally dry tropical plant communities. *PLoS One* 10:e0119166. doi: 10.1371/journal.pone.0119166

Moulatlet, G. M., Zuquim, G., Figueiredo, F. O. G., Lehtonen, S., Emilio, T., Ruokolainen, K., et al. (2017). Using digital soil maps to infer edaphic affinities of plant species in Amazonia: problems and prospects. *Ecol. Evol.* 7, 8463–8477. doi: 10.1002/ece3.3242

Mucina, L. (2019). Biome: evolution of a crucial ecological and biogeographical concept. *New Phytol.* 222, 97–114. doi: 10.1111/nph.15609

Mucina, L. (2020). Biomes are everybody's kingdom: a platform where ecology and biogeography meet. *New Phytol.* 228, 1463–1466. doi: 10.1111/nph.16933

Nascimento, E. R., Correia, I., Ruiz-Esparza, J. M., and Gouveia, S. F. (2018). Disentangling phylogenetic from non-phylogenetic functional structure of bird assemblages in a tropical dry forest. *Oikos* 127, 1177–1185. doi: 10.1111/oik.04910

Neves, D. M., Dexter, K. G., Baker, T. R., Souza, F. C., Oliveira-Filho, A. T., Queiroz, L. P., et al. (2020). Evolutionary diversity in tropical tree communities peaks at intermediate precipitation. *Sci. Rep.* 10:1188. doi: 10.1038/s41598-019-55621-w

Nürk, N. M., Linder, H. P., Onstein, R. E., Larcombe, M. J., Hughes, C. E., Fernández, L. P., et al. (2020). Diversification in evolutionary arenas—Assessment and synthesis. *Ecol. Evol.* 10, 6163–6182. doi: 10.1002/ece3.6313

Oliveira, A. A., Vicentini, A., Chave, J., Castanho, C. T., Davies, S. J., Martini, A. M. Z., et al. (2014). Habitat specialization and phylogenetic structure of tree species in a coastal Brazilian white-sand forest. *J. Plant Ecol.* 7, 134–144. doi: 10.1093/jpe/rtt073

Oliveira-Filho, A. T., Cardoso, D., Schrire, B. D., Lewis, G. P., Pennington, R. T., Brummer, T. J., et al. (2013a). Stability structures tropical woody plant diversity more than seasonality: insights into the ecology of high legume-succulent-plant biodiversity. *S. Afr. J. Bot.* 89, 42–57. doi: 10.1016/j.sajb.2013.06.010

Oliveira-Filho, A. T., Pennington, R. T., Rotella, J., and Lavin, M. (2013b). "Exploring evolutionarily meaningful vegetation definitions in the tropics: a community phylogenetic approach," in *Forests and Global Change*, eds D. A. Coomes, D. F. R. P. Burslem, and W. D. Simonsen (Cambridge: Cambridge University Press), 239–260. doi: 10.1017/cbo9781107323506.012

Pennington, R. T., and Lavin, M. (2016). The contrasting nature of woody plant species in different neotropical forest biomes reflects differences in ecological stability. *New Phytol.* 210, 25–37. doi: 10.1111/nph.13724

Pennington, R. T., Lavin, M., and Oliveira-Filho, A. (2009). Woody plant diversity, evolution, and ecology in the tropics: perspectives from seasonally dry tropical forests. *Annu. Rev. Ecol. Evol. Syst.* 40, 437–457. doi: 10.1146/annurev.ecolsys.110308.120327

Pennington, R. T., Lehmann, C. E. R., and Rowland, L. M. (2018). Tropical savannas and dry forests. *Curr. Biol.* 28, R541–R545.

Pennington, R. T., Prado, D. E., and Pendry, C. A. (2000). Neotropical seasonally dry forests and Quaternary vegetation changes. *J. Biogeogr.* 27, 261–273. doi: 10.1046/j.1365-2699.2000.00397.x

Pennington, R. T., Richardson, J. E., and Lavin, M. (2006). Insights into the historical construction of species-rich biomes from dated plant phylogenies, neutral ecological theory and phylogenetic community structure. *New Phytol.* 172, 605–616. doi: 10.1111/j.1469-8137.2006.01902.x

Peterson, A. T., Pape?, M., and Eaton, M. (2007). Transferability and model evaluation in ecological niche modeling: a comparison of GARP and MaxEnt. *Ecography* 30, 550–560. doi: 10.1111/j.0906-7590.2007.05102.x

Power, S. C., Verboom, G. A., Bond, W. J., and Cramer, M. D. (2017). Environmental correlates of biome-level floristic turnover in South Africa. *J. Biogeogr.* 44, 1745–1757. doi: 10.1111/jbi.12971

Prentice, I. C., Bondeau, A., Cramer, W., Harrison, S. P., Hickler, T., Lucht, W., et al. (2007). "Dynamic global vegetation modeling: quantifying terrestrial ecosystem responses to large-scale environmental change," in *Terrestrial Ecosystems in a Changing World*, eds J. G. Canadell, D. E. Pataki, and L. F. Pitelka (Berlin: Springer), 175–192. doi: 10.1002/eap.2313

Prentice, I. C., Cramer, W., Harrison, S. P., Leemans, R., Monserud, R. A., and Solomon, A. M. (1992). A global biome model based on plant physiology and dominance, soil properties and climate. *J. Biogeogr.* 19, 117–134.

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 – Approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490

Prieto-Torres, D. A., Rojas-Soto, O. R., Bonaccorso, E., Santiago-Alarcon, D., and Navarro-Sigüenza, A. G. (2019). Distributional patterns of Neotropical seasonally dry forest birds: a biogeographical regionalization. *Cladistics* 35, 446–460. doi: 10.1111/cla.12366

Queiroz, L. P. (2006). "The Brazilian Caatinga: phytogeographical patterns inferred from distribution data of the *Leguminosae*," in *Neotropical Savannas and Seasonally Dry Forests: Plant Diversity, Biogeography, and Conservation*, eds R. T. Pennington, G. P. Lewis, and J. A. Ratter (London: CRC Press), 121–158. doi: 10.1201/9781420004496-6

Queiroz, L. P., Cardoso, D., Fernandes, M. F., and Moro, M. F. (2017). "Diversity and evolution of flowering plants of the Caatinga domain," in *Caatinga: The Largest Tropical Dry Forest Region in South America*, eds J. M. C. Silva, I. R. Leal, and M. Tabarelli (Cham: Springer), 23–63. doi: 10.1099/ijsem.0.004193

Queiroz, L. P., Lewis, G. P., and Wojciechowski, M. F. (2010). Tabaroa, a new genus of Leguminosae tribe Brongniartieae from Brazil. *Kew Bull.* 65, 189–203. doi: 10.1007/s12225-010-9202-7

Rapini, A., Bitencourt, C., Luebert, F., and Cardoso, D. (2021). An escape-to-radiate model for explaining the high plant diversity and endemism in campos rupestres. *Biol. J. Linn. Soc.* 133, 481–498. doi: 10.1093/biolinnean/blaa179

Ricotta, C., Podani, J., and Pavoine, S. (2016). A family of functional dissimilarity measures for presence and absence data. *Ecol. Evol.* 6, 5383–5389. doi: 10.1002/ece3.2214

Ringelberg, J. J., Zimmermann, N. E., Weeks, A., Lavin, M., and Hughes, C. E. (2020). Biomes as evolutionary arenas: convergence and conservatism in the trans-continental succulent biome. *Global Ecol. Biogeogr.* 29, 1100–1113. doi: 10.1111/geb.13089

Rocha, P. L. B., Queiroz, L. P., and Pirani, J. R. (2004). Plant species and habitat structure in a sand dune field in the Brazilian Caatinga: a homogeneous habitat harbouring an endemic biota. *Rev. Bras. Bot.* 27, 739–755. doi: 10.1590/s0100-84042004000400013

Reflora. (2021). *Virtual Herbarium*. Available online at: http://floradobrasil.jbrj.gov.br/reflora/herbarioVirtual/ (accessed June 04, 2021).

Santos, R. M., Oliveira-Filho, A. T., Eisenlohr, P. V., Queiroz, L. P., Cardoso, D. B. O. S., and Rodal, M. J. N. (2012). Identity and relationships of the Arboreal Caatinga among other floristic units of seasonally dry tropical forests (SDTFs) of north-eastern and Central Brazil. *Ecol. Evol.* 2, 409–428. doi: 10.1002/ece3.91

Särkinen, T. E., Iganci, J. R. V., Linares-Palomino, R., Simon, M. F., and Prado, D. E. (2011). Forgotten forests – issues and prospects in biome mapping using seasonally dry tropical forests as a case study. *BMC Ecol.* 11:27. doi: 10.1186/1472-6785-11-27

Scarano, F. R. (2002). Structure, function and floristic relationships of plant communities in stressful habitats marginal to the Brazilian Atlantic Rainforest. *Ann. Bot.* 90, 517–524. doi: 10.1093/aob/mcf189

Scholes, R. J., and Archer, S. R. (1997). Tree-grass interactions in savannas. *Annu. Rev. Ecol. Evol. Syst.* 28, 517–544. doi: 10.1146/annurev.ecolsys.28.1.517

Schrire, B. D., Lavin, M., and Lewis, G. P. (2005). Global distribution patterns of the Leguminosae: insights from recent phylogenies. *Biol. Skr.* 55, 375–422.

Segovia, R. A., Pennington, T., Baker, T. R., Souza, F. C., Neves, D. M., Davis, C. C., et al. (2020). Freezing and water availability structure the evolutionary diversity of trees across the Americas. *Sci. Adv.* 6:aaz5373. doi: 10.1126/sciadv.aaz5373

Silva, A. C., and Souza, A. F. (2018). Aridity drives plant biogeographical sub regions in the Caatinga, the largest tropical dry forest and woodland block in South America. *PLoS One* 13:e0196130. doi: 10.1371/journal.pone.0196130

Silva-de-Miranda, P., Oliveira-Filho, A. T., Pennington, R. T., Neves, D. M., Baker, R. T., and Dexter, K. G. (2018). Using tree species inventories to map biomes and assess their climatic overlaps in lowland tropical South America. *Global Ecol. Biogeogr.* 27, 899–912. doi: 10.1111/geb.12749

Simon, M. F., and Pennington, R. T. (2012). Evidence for adaptation to fire regimes in the tropical savannas of the Brazilian Cerrado. *Int. J. Plant Sci.* 173, 711–723. doi: 10.1086/665973

Simon, M. F., Grether, R., Queiroz, L. P., Skema, C., Pennington, R. T., and Hughes, C. E. (2009). Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20359–20364. doi: 10.1073/pnas.0903410106

Souza-e-Silva, J. L., Cruz-Neto, O., Peres, C. A., Tabarelli, M., and Lope, A. V. (2019). Climate change will reduce suitable Caatinga dry forest habitat for endemic plants with disproportionate impacts on specialized reproductive strategies. *PLoS One* 14:e0217028. doi: 10.1371/journal.pone.0217028

Stolar, J., and Nielsen, S. E. (2015). Accounting for spatially biased sampling effort in presence-only species distribution modelling. *Divers. Distrib.* 21, 595–608. doi: 10.1111/brv.12359

The Angiosperm Phylogeny Group, Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., Judd, W. S., et al. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* 181, 1–20. doi: 10.1111/boj.12385

The Brazil Flora Group [BFG] (2015). Growing knowledge: an overview of seed plant diversity in Brazil. *Rodriguesia* 66, 1085–1113. doi: 10.1590/2175-7860201566411

Vilhena, D. A., and Antonelli, A. (2015). A network approach for identifying and delimiting biogeographical regions. *Nat. Commun.* 6:6848. doi: 10.1038/ncomms7848

Wallace, A. R. (1876). *The Geographical Distribution of Animals: with a Study of the Relations of Living and Extinct Faunas as Elucidating the Past Changes of the Earth's Surface*. Cambridge: Cambridge University Press.

Wan, Z. (2014). New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote Sens. Environ.* 140, 36–45. doi: 10.1016/j.rse.2013.08.027

Wan, Z., and Dozier, J. (1996). A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Trans. Geosci. Remote Sens.* 34, 892–905. doi: 10.1109/36.508406

White, F. (1983). *The Vegetation of Africa: A Descriptive Memoir to Accompany the UNESCO/ETFAT/UNSO Vegetation Map of Africa*. Paris: UNESCO.

Whittaker, R. H. (1970). *Communities and Ecosystems*. New York, NY: MacMillan.

Wiegand, T., and Moloney, K. A. (2013). *Handbook of Spatial Point-Analysis in Ecology*. Boca Raton, FL: CRC Press.

Wiens, J. J., Ackerly, D. D., Allen, A. P., Buckley, L. B., Cornell, H. V., Damschen, E. I., et al. (2010). Niche conservatism as an emerging principle in ecology and conservation biology. *Ecol. Lett.* 13, 1310–1324. doi: 10.1111/j.1461-0248.2010.01515.x

Williams, J. W., Jackson, S. T., and Kutzbach, J. E. (2007). Projected distributions of novel and disappearing climates by 2100 AD. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5738–5742. doi: 10.1073/pnas.0606292104

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

Winters, D. J. (2017). rentrez: an R package for the NCBI eUtils API. *R J.* 9, 520–526. doi: 10.32614/rj-2017-058

Woodward, F. I., Lomas, M. R., and Kelly, C. K. (2004). Global climate and the distribution of plant biomes. *Phil. Trans. R. Soc. Lond. B* 359, 1465–1476. doi: 10.1098/rstb.2004.1525

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 24 February 2022
doi: 10.3389/fevo.2022.723286

The Origins and Historical Assembly of the Brazilian Caatinga Seasonally Dry Tropical Forests

Moabe F. Fernandes[1,2], Domingos Cardoso[3], R. Toby Pennington[2,4] and Luciano P. de Queiroz[1]*

[1] Departamento de Ciências Biológicas, Universidade Estadual de Feira de Santana, Feira de Santana, Brazil, [2] Department of Geography, University of Exeter, Exeter, United Kingdom, [3] National Institute of Science and Technology in Interdisciplinary and Transdisciplinary Studies in Ecology and Evolution (INCT IN-TREE), Instituto de Biologia, Universidade Federal da Bahia, Salvador, Brazil, [4] Tropical Diversity Section, Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom

The Brazilian Caatinga is considered the richest nucleus of the Seasonally Dry Tropical Forests (SDTF) in the Neotropics, also exhibiting high levels of endemism, but the timing of origin and the evolutionary causes of its plant diversification are still poorly understood. In this study, we integrate comprehensive sampled dated molecular phylogenies of multiple flowering plant groups and estimations of ancestral areas to elucidate the forces driving diversification and historical assembly in the Caatinga flowering plants. Our results show a pervasive floristic exchange between Caatinga and other neotropical regions, particularly those adjacent. While some Caatinga lineages arose in the Eocene/Oligocene, most dry-adapted endemic plant lineages found in region emerged from the middle to late Miocene until the Pleistocene, indicating that only during this period the Caatinga started to coalesce into a SDTF like we see today. Our findings are temporally congruent with global and regional aridification events and extensive denudation of thick layers of sediments in Northeast (NE) Brazil. We hypothesize that global aridification processes have played important role in the ancient plant assembly and long-term Caatinga SDTF biome stability, whereas climate-induced vegetation shifts, as well as the newly opened habitats have largely contributed as drivers of *in situ* diversification in the region. Patterns of phylogenetic relatedness of Caatinga endemic clades revealed that much modern species diversity has originated *in situ* and likely evolved via recent (Pliocene/Pleistocene) ecological specialization triggered by increased environmental heterogeneity and the exhumation of edaphically disparate substrates. The continuous assembly of dry-adapted flora of the Caatinga has been complex, adding to growing evidence that the origins and historical assembly of the distinct SDTF patches are idiosyncratic across the Neotropics, driven not just by continental-scale processes but also by unique features of regional-scale geological history.

Keywords: biogeography, biome evolution, biome reconstruction, landscape evolution, Northeastern South America, semi-arid, SDTF, speciation

OPEN ACCESS

Edited by:
Danilo M. Neves, Federal University of Minas Gerais, Brazil

Reviewed by:
Carlos Jaramillo, Smithsonian Tropical Research Institute, Panama
Fernanda Antunes Carvalho, Minas Gerais State University, Brazil

****Correspondence:***
Moabe F. Fernandes
moabeffernandes@gmail.com

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *10 June 2021*
Accepted: *27 January 2022*
Published: *24 February 2022*

Citation:
Fernandes MF, Cardoso D, Pennington RT and de Queiroz LP (2022) The Origins and Historical Assembly of the Brazilian Caatinga Seasonally Dry Tropical Forests. Front. Ecol. Evol. 10:723286. doi: 10.3389/fevo.2022.723286

INTRODUCTION

The Brazilian Caatinga is recognized as the largest and most species rich nucleus of the Seasonally Dry Tropical Forest (SDTF) biome in the New World (Pennington et al., 2000; Banda et al., 2016; Queiroz et al., 2017; Fernandes et al., 2020), yet it is one of the least studied neotropical biogeographic regions (Fiaschi and Pirani, 2009). The highly heterogeneous gradients of topography, geology, soils, and the spatial and temporal distribution of rainfall support a wide variety of dry-adapted plant communities in the region (Velloso et al., 2002; Silva and Souza, 2018). In particular, the nature of geological parent material composing the soils is suggested to be the most important constraint to plant diversity and endemism, producing an impressive turnover in the distribution of species and, consequently defining the arrangement of the floristic sets (Gomes et al., 2006; Cardoso and Queiroz, 2007; Santos et al., 2012; Moro et al., 2016; Queiroz et al., 2017). Caatinga plant communities have been classically divided into: (i) Crystalline Caatinga, established over fertile-soils primarily derived from Precambrian crystalline basement of the *Depressão Sertaneja,* ranging from woodlands (Caatinga *s.s.*) on driest sites to tall forests (Arboreal Caatinga) on moister areas; (ii) Sedimentary Caatinga, open scrub occupying the patchily distributed nutrient-poor sandy soils of ancient sedimentary basins and continental sand dunes and; (iii) the vegetation established over the karst deposits of the Bambuí and Jandaíra groups (**Figures 1A–F**; Queiroz, 2006; Moro et al., 2016; Queiroz et al., 2017).

The development of the Caatinga's diverse geomorphological systems has been mostly the result of large-scale regional denudation from the Miocene onward (**Figure 2**; Ab'Sáber, 1974; Peulvast and Claudino-Sales, 2004; Valadão, 2009; Japsen et al., 2012), caused by neotectonic reactivation (Oliveira and Medeiros, 2012; Rodríguez-Tribaldos et al., 2017; Klöckinga et al., 2020) and a series of global and regional events that exacerbated aridity (Zachos et al., 2001). For example, the rise of the Borborema Plateau at the eastern limits of the Caatinga formed an orographic barrier to air masses entering from the Atlantic Ocean, creating a rain shadow effect that drastically altered temperature and precipitation regimes in the region (Gottsberger and Silberbauer-Gottsberger, 2006; Hoorn et al., 2014). Ultimately, increased aridity resulted in massive erosion of a thick sedimentary cover accumulated during wetter and warmer climatic conditions of the Early Cenozoic (**Figure 2A**). This process culminated in the exhumation of the Paleo-Mesozoic sediments and the formation of the large-scale, lower-level peneplain of the *Depressão Sertaneja* (**Figure 2B**). Yet, post-Miocene uplift within the Araripe Plateau led to widespread fluvial incision (Peulvast et al., 2008) and dramatic changes in the course of the São Francisco River (once flowing toward equatorial Atlantic Ocean), creating an endorheic system (**Figure 2C**; Mabesoone, 1994). Later, in the Pleistocene, when the river acquired its current course toward eastern Atlantic, large amounts of sand accumulated in this "paleo-lake" were reshaped by wind, forming the modern São Francisco River dune system (**Figure 2D**; Barreto and Suguio, 1993; Mabesoone, 1994; Barreto, 1996). Also in the Pleistocene, the large karst systems belonging to Bambuí and Jandaíra groups were exhumed (Silva et al., 2017), adding to the increased environmental heterogeneity in the Caatinga (**Figure 2D**).

While major geological transformations have been relatively well characterized in the Caatinga, little is known about the origins and evolution of the Caatinga biota, particularly how the chronology of major geological and climatic events impacted its diverse and endemic plant assemblages. The Caatinga flora was long supposed to be composed of a minor subset of drought-tolerant elements from surrounding biomes (see Rizzini, 1979; Andrade-Lima, 1981). Other authorities interpreted it as derived from lineages occupying other nuclei of the neotropical SDTF (Sarmiento, 1975; Prado and Gibbs, 1993; Pennington et al., 2000) or the Chaco of southern South America (Andrade-Lima, 1982). An alternative, more integrative hypothesis that took into account the complex geological evolution of the Brazilian Northeast (NE) region was proposed by Queiroz (2006), which suggested that fragmentation of the previously continuous sedimentary cover would have promoted vicariance of isolated populations during the Miocene. Such fragmentation exposed the fertile-soils of the *Depressão Sertaneja*, allowing massive migration of dry-adapted elements that originated in other neotropical SDTF. A major issue with Queiroz's (2006) propositions, however, is that he considered peneplanation (long-term erosion that generated Caatinga stepped relief) as a single, temporally defined (Miocene) event, greatly contrasting with subsequent geological evidence of multiple events of massive erosion taking place over discontinuous periods (e.g., Valadão, 2009; Japsen et al., 2012). This more complex geological history suggests that Queiroz's (2006) hypothesis about the timing of origin and historical processes that shaped modern Caatinga's plant diversity deserves new scrutiny using rigorous phylogenetic testing.

Improved understanding of the timing and forces underlying species assembly and diversification in the Caatinga is not just limited and incomplete but has also been challenging because only few Caatinga plant lineages had DNA sequence data available or were included in molecular phylogenetic studies. Over the last decade, new densely sampled phylogenies of key Caatinga lineages have been published (e.g., Queiroz et al., 2010; Simon et al., 2011; Cardoso et al., 2013; Carvalho-Sobrinho et al., 2016; Almeida et al., 2018; Maya-Lastra and Steinmann, 2019), all of which now provide an excellent opportunity to evaluate the impact of the complex geological and climatic history on the Caatinga biodiversity and to unveil the spatial and temporal dynamics underlying floristic exchanges between the Caatinga SDTF and other neotropical regions. By exploring and comparing the biogeographical history of multiple, unrelated taxonomic groups during the complex landscape evolution of the Caatinga (**Figure 2**), here we also address whether (i) the Caatinga plant lineages constitute a subset of the surrounding flora (Rizzini, 1979; Andrade-Lima, 1981, 1982) and (ii) endemic lineages have emerged by Miocene vicariance (Queiroz, 2006). By analysing plant lineage diversification in a geological-evolutionary perspective that takes into account the synergistic effect of landscape and biological evolution, we expect to provide

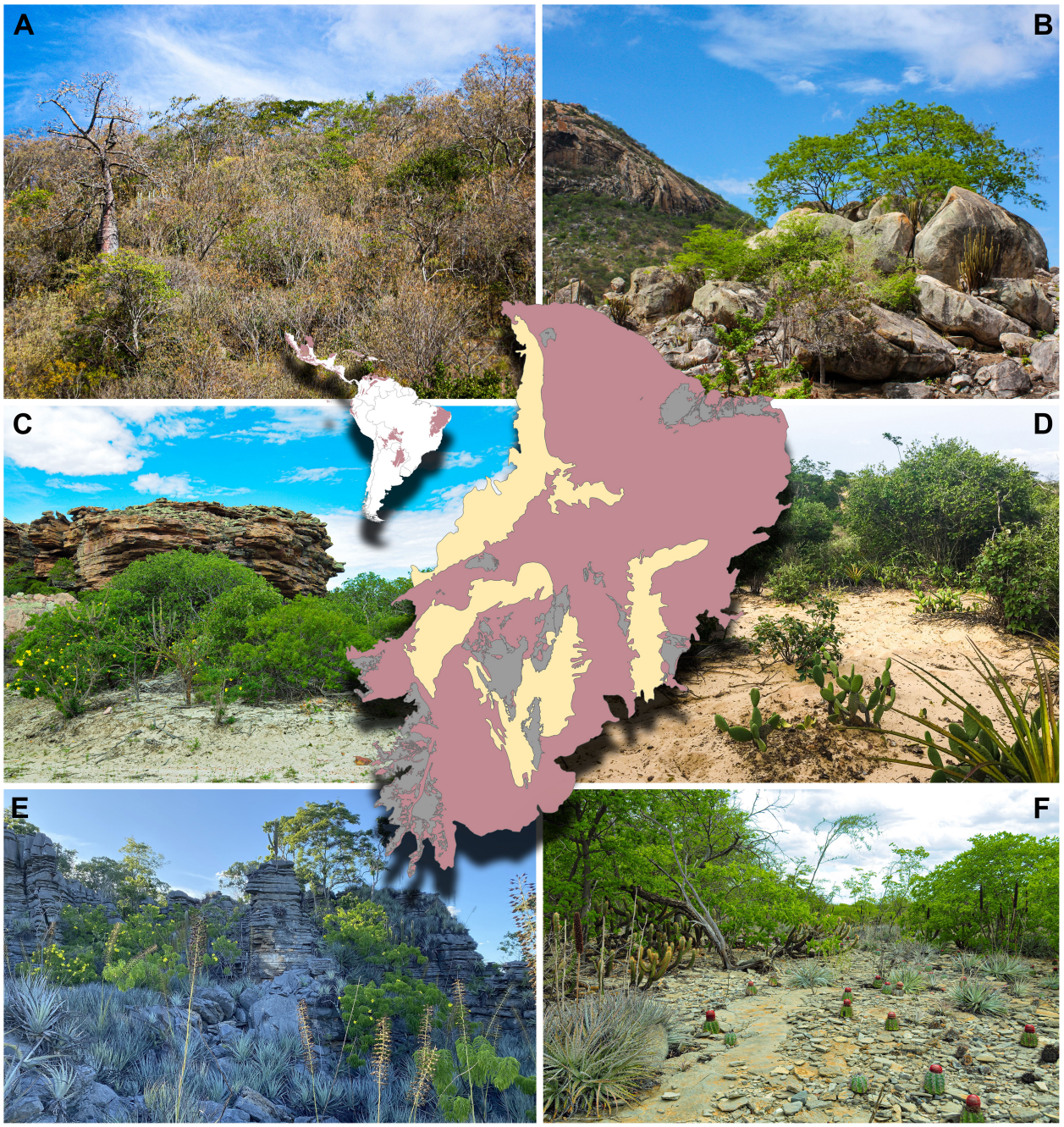


FIGURE 1 | Vegetation structure variation throughout the three main geological types of the Caatinga seasonally dry forests. The small side map shows the distribution of the Caatinga in Northeast Brazil in the context of the Seasonally Dry Tropical Forest biome (SDTF) in the Neotropics (Queiroz et al., 2017). The *Depressão Sertaneja* (red; **A,B**) is the most dominant and continuously distributed formation, while the Sedimentary Caatinga (pale yellow; **C,D**) and Karst Caatinga (gray; **E,F**) are patchily distributed. Representative images highlight the structurally heterogenous vegetation, with exposed crystalline rocks, white-sand soil, or limestones, from across these main geological formations at: **(A)** Parnamirim, Bahia, showing the Arboreal Caatinga with a leafless large tree of the Brazilian baobab *Cavanillesia umbellata* (Malvaceae); **(B)** Sobral, Ceará, showing a more open Crystalline Caatinga woodland with an individual of the dominant legume tree *Luetzelburgia auriculata* and the Caatinga endemic *Pilosocereus chrysostele* (Cactaceae); **(C)** Morro do Chapéu, Bahia, showing a Sedimentary Caatinga with the yellow-flowered shrub *Allamanda puberula*; **(D)** Barra, Bahia, showing a sand dune of the mid São Francisco River valley with the endemic opuntioid cactus *Tacinga inamoena*; **(E)** Parque Nacional Carvernas do Peruaçu, Minas Gerais, showing the exposed razor-sharp limestone outcrops, dominated by individuals of the Bromeliaceae *Encholirium luxor* and the karst endemic *Allamanda calcicola*; **(F)** Morro do Chapéu, Bahia, with individuals of the barrel Cactaceae *Melocactus pachyacanthus*, and the Caatinga endemic *Encholirium spectabile* and the columnar Cactaceae *Xiquexique gounellei*. All photos by Domingos Cardoso.

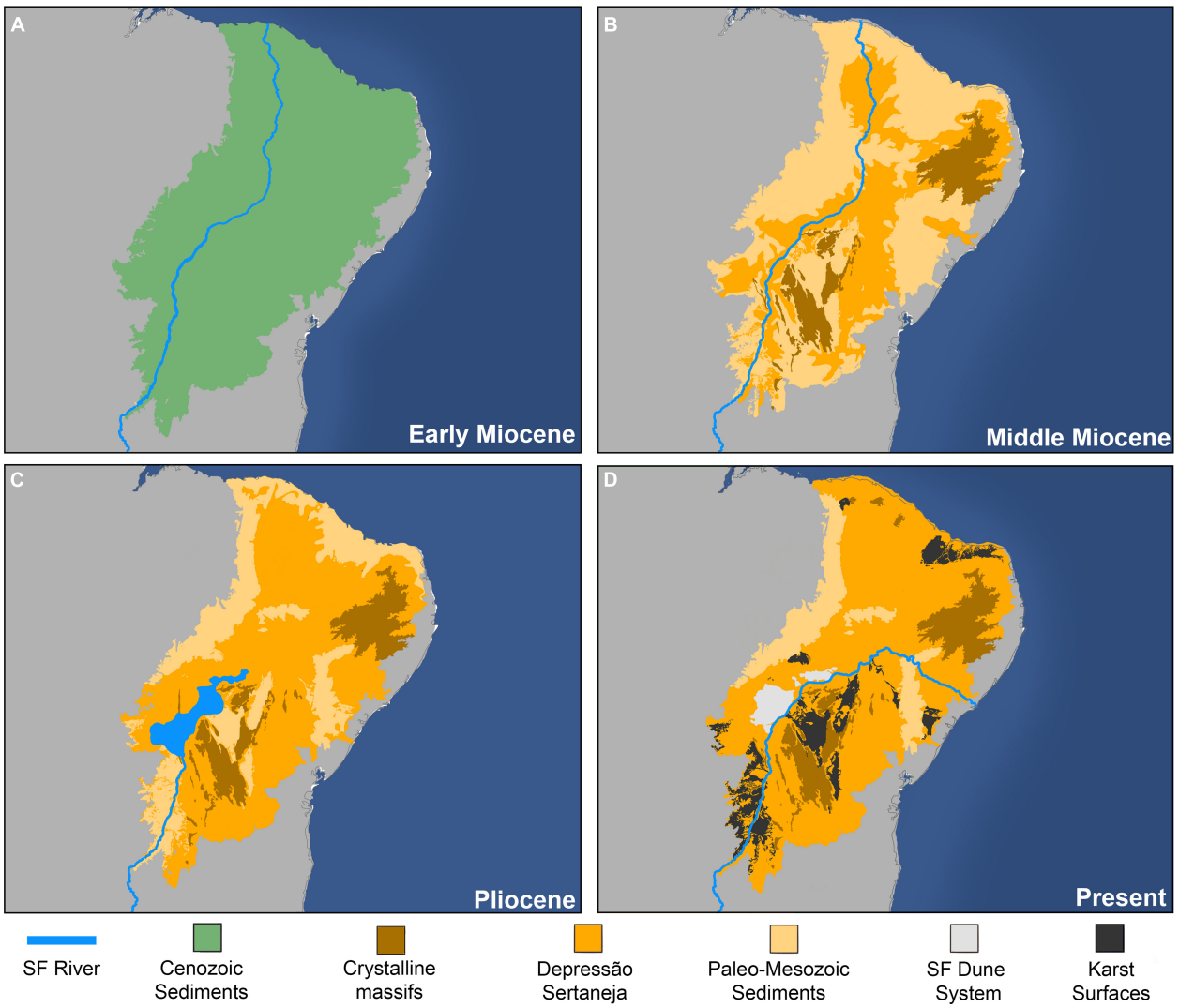


FIGURE 2 | Paleogeographic maps of the transition from the widespread paleo-sedimentary surface to modern heterogeneous Caatinga landscapes. **(A)** The sedimentary cover accumulated during previous times is maintained in the Caatinga due to the widespread dominance of tropical rainforests over the region at Early Cenozoic (Fine and Ree, 2006; Jaramillo and Cardenas, 2012). **(B)** Global climate changes and increased uplift and erosion at the Miocene (Rodríguez-Tribaldos et al., 2017; Klöckinga et al., 2020) exposed the Paleo-Mesozoic sediments accumulated due to post-rift sedimentation–after fragmentation of Gondwana supercontinent (Magnavita et al., 1994; Maia et al., 2010; Bezerra et al., 2011; Gurgel et al., 2013) and generated a new, low-level peneplain, known as Depressão Sertaneja. **(C)** Pliocene tectonic reactivation within the Araripe Plateau changed the course of São Francisco (SF) River, which had a paleo-course flowing toward the equatorial Atlantic; this event created an endorheic drainage at the central region of the Caatinga (Mabesoone, 1994). **(D)** In the Pleistocene, the lacustrine sediments, which remained after the São Francisco River acquired its current watercourse, were remodeled by wind and formed a system of continental dunes that is currently renowned by a highly endemic biota (Rocha et al., 2004; Queiroz et al., 2017). Also at this period, most of the karst systems were also exhumed (Silva et al., 2017).

a series of more focused pictures of the historical assembly of the Caatinga SDTF.

MATERIALS AND METHODS

Taxon Sampling and Molecular Data

We used the updated checklist of flowering plants of the Caatinga SDTF (Fernandes et al., 2020) to identify a set of lineages that are taxonomically and ecologically important in the dry formations of the region, i.e., representative plant clades in terms of species richness, endemism, or ecological dominance. We then compiled from the literature a set of densely sampled species-level phylogenies from the target lineages. Our final dataset encompasses 20 monophyletic groups, which include a total of 150 Caatinga-inhabiting species, from which 95 are endemics. By sampling widely across several families of flowering plants and a wide range of life forms (e.g., herbs and trees), we maximize the inclusion of lineages with contrasting biogeographical histories that may have colonized the Caatinga by different routes and

at different times, providing a multi-taxon picture of plant evolution in that region.

We retrieved complete alignments or individual DNA sequences for each selected lineage from TreeBASE[1] (Piel et al., 2009) and GenBank[2] (Benson et al., 2013), respectively (see **Supplementary Table 1** for details on study groups, genes, and sources and **Supplementary Table 2** for GenBank accessions). For downstream dating analyses, the originally retrieved TreeBASE alignments were not modified, whereas sequences of each individual region retrieved from GenBank were first aligned automatically using MAFFT (Katoh and Standley, 2013) with default settings and then, manually adjusted in Geneious (Drummond et al., 2012a).

Phylogenetic Reconstruction and Divergence Estimates

Phylogenetic relationships and divergence times were estimated using a Bayesian approach in BEAST v1.8.4 (Drummond et al., 2012b) as implemented in the CIPRES Science Gateway (Miller et al., 2010). We ran analyses using the Yule speciation model and uncorrelated relaxed molecular clock with a lognormal prior distribution. We estimated the best-fit nucleotide substitution models using AIC (Akaike Information Criterion) in jModelTest2 (Darriba et al., 2012). Depending on the availability of reliable fossil records for each group, we used primary (fossil) or secondary (in the absence of reliable fossils) calibration points, with lognormal and normal prior distributions, respectively (**Supplementary Table 3** details the calibration points). We performed a minimum of two independent runs of 10^7 generations of MCMC (Markov chain Monte Carlo), sampling parameters every 10^4 generations. Convergence of the runs was evaluated in Tracer v1.6 (Rambaut and Drummond, 2013) to ensure that effective sample sizes (ESS) reached values higher than 200 for all parameters. After the exclusion of 25% of the initial trees (burn-in) of each independent run, we combined tree files in LogCombiner v.1.8.2 (Drummond et al., 2012b) and used TreeAnnotator v1.8.4 (Drummond et al., 2012b) to summarize results into maximum clade credibility (MCC) trees. Prior to ancestral biome reconstruction analyses all outgroups were pruned from phylogenies.

Geographic Distribution

Defining biogeographical units for ancestral area reconstruction is a difficult task. Because they are highly dependent on the lineages under study, there is no universal way to define them. There are also multiple limitations associated with choosing a reasonable number of operational units to be used in the analyses, leading to highly polymorphic states (occurrence in multiple areas). Taking this into account, we aimed to define sets of regions that are biologically meaningful for all lineages under analysis, reflecting geography and the geological history.

We produced a classification for the New World based on previous regionalization proposed by Morrone (2014). Our classification is composed of nine regions (**Figure 3**), and includes: two regions dominated by tropical rainforests (Amazon, Atlantic Forest); one dominated by grass-rich and fire-prone savannas (Cerrado + Pampas, which here also includes the higher

[1]http://treebase.org

[2]https://www.ncbi.nlm.nih.gov/genbank/

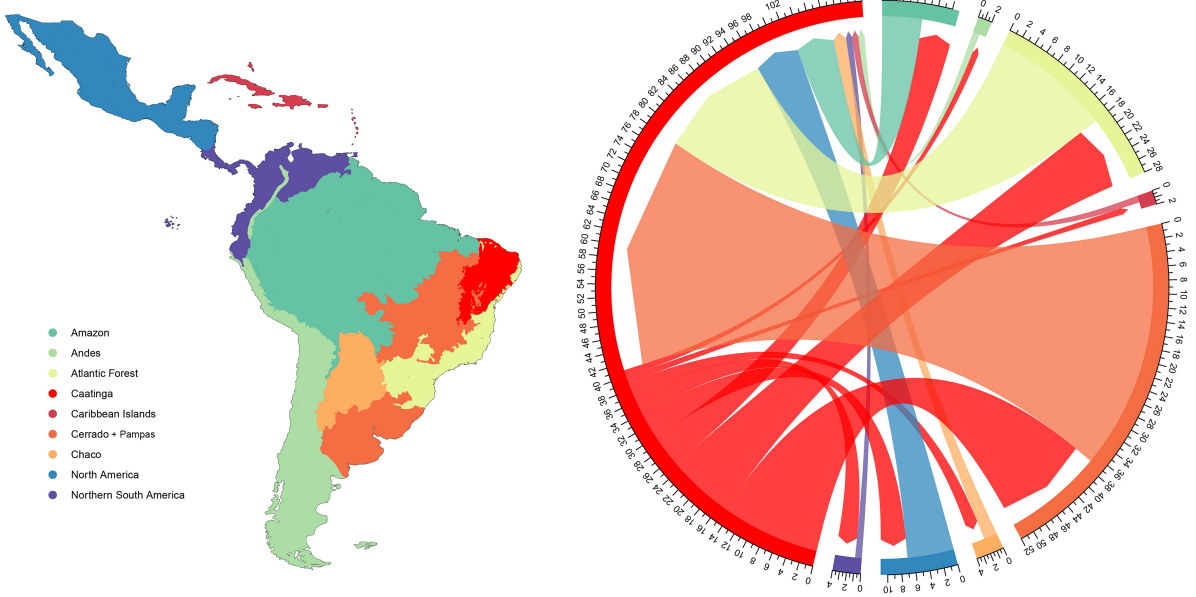


FIGURE 3 | Distribution of major areas used in our ancestral reconstructions based on reclassification of Morrone (2014) biogeographical regionalization of Neotropics and the floristic interchange among the Caatinga and other regions as estimated from DEC + j reconstructions for our sample composed of 20 lineages. Colors of the map and the chord-graph coincide and indicate the areas referred in the section "Materials and Methods." Shapefile from Löwenberg-Neto (2014).

elevations of the Chapada Diamantina mountain range, in the core of the Caatinga dry matrix and also the southern grasslands); the Caatinga, characterized by fire-sensitive, succulent-rich formations associated with the global SDTF biome; the Chaco, a region with great temperature seasonality that experiences regular frost, and potentially represents a distinct biome with more temperate affinities (Prado and Gibbs, 1993; Banda et al., 2016; Miranda et al., 2018); the other four regions (Andes, Caribbean Islands, North America, and Northern South America) are composed by a mix of distinct biome types. Since our aim is to assess the evolution of the dry-adapted flora, we considered Caatinga in its narrow sense (*sensu* Fernandes et al., 2020), including only vegetation types connected to the SDTF biome, which encompasses distinct formations that mostly occur on areas of low altitude (below 1,000 m) in the region. In this context, species occurring at the rainforest enclaves (locally known as *brejos de altitude*) were coded as Atlantic Domain, and those recorded for cerrado enclaves and *campos rupestres*, mostly occurring on the Chapada Diamantina were coded as Cerrado. Despite the disjunct distribution of the Cerrado and the Pampas (phytogeographic region located in Southern Brazil and Uruguay), both regions are characterized by fire-prone grasslands, and their lineages exhibit evident evolutionary affinities (see Simon et al. (2011)), which justifies their clustering into a single region.

The current distribution of each species within the areas and biomes were obtained primarily from specialized literature (taxonomic monographic works and floras), but we also used information available in *Flora do Brasil 2020*.[3] We coded all species in our phylogenies as present or absent within these regions.

Ancestral Area Estimation

We estimated ancestral areas through a likelihood approach as implemented in the R (R Core Team, 2020) package BioGeoBEARS (BioGeography with Bayesian, and Likelihood, Evolutionary Analysis in R Scripts - Matzke, 2013). We performed analyses using Dispersal-Extinction-Cladogenesis (DEC) model and the DEC + j model, which incorporates the jump parameter (j) that allows for a descendent lineage to occupy a new area that is different from its parental lineage by a founder event. We restricted the maximum number of areas of each analysis to the same as that the most widespread species in the phylogeny were distributed. As a rule, DEC + j reconstructions gave the highest likelihood values and, because of this, they will be discussed in this text. To quantify the number of transitions from all regions into Caatinga and from the Caatinga to all other regions, we counted the biogeographical events as inferred by our DEC + j reconstructions.

RESULTS

We identified a total of 175 events involving the Caatinga across all lineages. Approximately 65% (114 events) of this total represent biogeographical shifts, whereas the 35% (61 events) are related to *in situ* speciation (**Supplementary Table 4**). The majority of biogeographical shifts (73 events, or 64% of total shifts) represent transitions into the Caatinga and, only 41 events (36%) represent shifts from the Caatinga into other regions (**Figure 3**). Atlantic Forest and Cerrado + Pampas were the most important regions in terms of plant lineage exchange. Together, they account for 61% of all shifts into the Caatinga and 81% of total shifts from the Caatinga. The Amazon and North America exhibited moderate floristic exchange with the Caatinga, whereas the Andean region, Caribbean Islands, Chaco, and Northern South America were involved in few exchange events (**Figure 3**; see **Supplementary Figure 1** and **Supplementary Table 4** for further details on the types, areas, and ages involved in each of these biogeographical events as well as the ancestral area estimations). These numbers clearly demonstrate multiple independent origins for Caatinga plant lineages, and also show asymmetrical connectivity dynamics, in a way that Caatinga has received more lineages than being a source of plant diversity to other regions.

Temporal patterns show that colonization of the Caatinga by lineages from other regions precede events of *in situ* speciation (**Figure 4A**). They also show that the relevance of cladogenetic shifts (those generating endemic species through isolation of lineages in the Caatinga) from distinct regions changed markedly over evolutionary time (**Figures 4B,C**). Early divergence that formed the initial species pool of elements adapted to the seasonally dry conditions of the Caatinga started as early as the Eocene-Oligocene (estimated mean age of 41.9 Ma for the oldest transition in Bombacoids), from ancestors primarily associated with regions largely characterized by the dominance of rainforests (Amazon and Atlantic Forest). However, these events are rare and, only after the mid-Miocene, particularly in the last 10 Ma, we found a general increase in the relative number of events isolating lineages in the Caatinga (**Figure 4C**). This increased Miocene lineage immigration that continues up to the Pleistocene came from regions that harbor major areas of seasonally dry and/or open biomes (Cerrado + Pampas and North America). We also identified a significant number of transitions from the Atlantic Forest to the Caatinga during the Miocene-Pleistocene periods. However, many of these cases are lineages (e.g., *Amorimia*, *Luetzelburgia*, and *Pereskia*) that are largely confined to the drier formations within the Atlantic Forest region and so they perhaps are better characterized as phylogenetic biome conservatism than evolutionary biome-shift events.

Though significant, the number of independent colonization events of the Caatinga cannot fully account for its present-day plant diversity and our results (**Supplementary Table 4**) show that *in situ* speciation also played a key role in the evolution of the Caatinga flora, with most events having occurred within the last few million years (**Figure 4A**). *In situ* speciation events began in the late Miocene, with an increase at the Pliocene and peaking in the Pleistocene. From the Pliocene onward, *in situ* diversification became the dominant mode of species accumulation in the Caatinga (**Supplementary Table 4**), exceeding rates of transition from other regions. A common pattern that emerges from our analyses is that Caatinga endemic

[3]http://floradobrasil.jbrj.gov.br

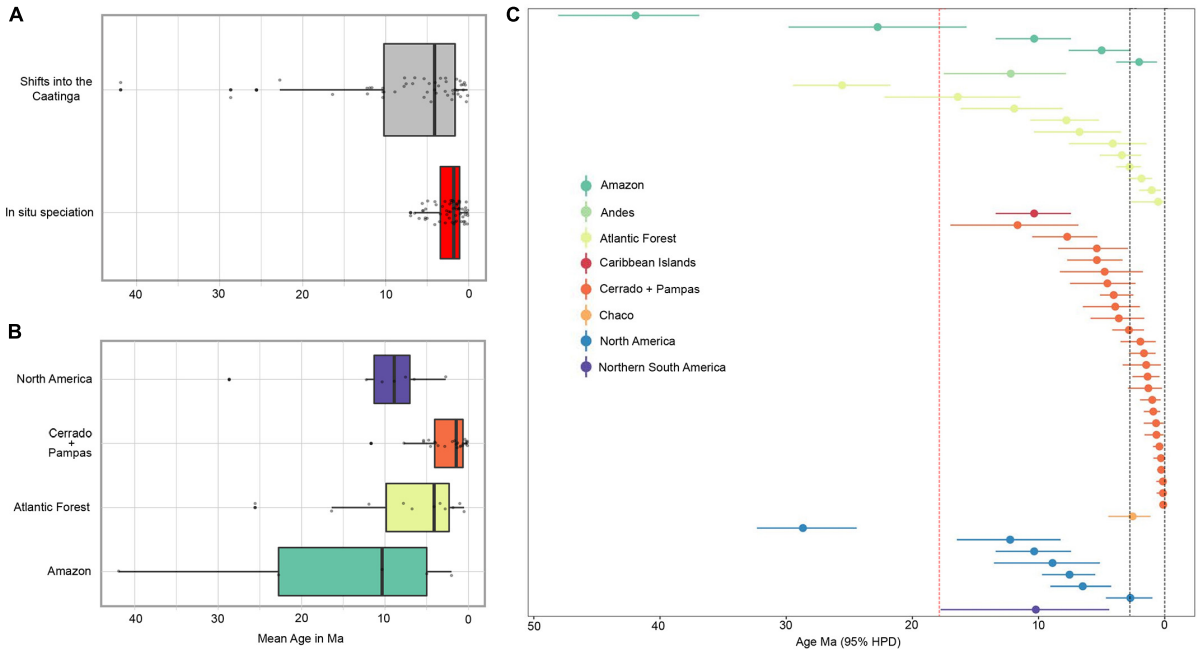


FIGURE 4 | Temporal patterns of biogeographical events generating Caatinga endemic lineages. Boxplots summarize ages for nodes with **(A)** highest probabilities of cladogenetic shifts into the Caatinga and *in situ* speciation and; **(B)** cladogenetic shifts from regions that are important providers of lineages to the Caatinga. **(C)** Node ages and confidence intervals of cladogenetic shifts (both dispersal and vicariance) leading to Caatinga crown clades as estimated by a DEC + j model. Ages are in decreasing order and grouped by inferred ancestral areas. Colors correspond to the same shown in **Figure 3**. Vertical dotted red line represents the onset of uplift and erosion at 18–15 Ma according to Japsen et al. (2012). Vertical dotted black lines indicate the Pleistocene period.

lineages generally represent small radiations of clades composed by a few endemic species, which are often restricted to distinct geological substrates (**Figure 5**). Therefore, after the early colonization and establishment of plant lineages in the Caatinga, many of them have diversified *in situ*, showing that relatively recent diversification has generated a great portion of endemic plant species that inhabit the region today.

DISCUSSION

Frequent Migration From Adjacent Regions Have Shaped the Assembly of the Caatinga Seasonally Dry Tropical Forests

The recruitment of dry-adapted Caatinga plant lineages was partly a by-product of complex dynamics of colonization from geographically adjacent areas. On one hand, our results reinforce a scenario of high biotic interchange among neotropical biogeographical regions and biomes (Queiroz et al., 2017; Antonelli et al., 2018). Thus, they partially support previous propositions (Rizzini, 1979; Andrade-Lima, 1981) that the Caatinga's plant diversity originated from a subset of drought-tolerant elements coming from the surroundings, particularly the Atlantic Forest and the Cerrado. On the other hand, the small proportion of events involving the Caribbean (in its broadest sense, including both the Caribbean Islands and the northern coast of South America) and the Chaco indicates that floristic exchanges with these regions have been less significant than previously thought (e.g., Sarmiento, 1975; Andrade-Lima, 1982). The large size of the Caatinga, bordered by the ecologically distinct, species-rich formations of Amazonia, Atlantic Forests, and the Cerrado have likely impacted the high levels of their historical biotic exchange. The highly heterogeneous environments found within the Caatinga region, including scattered rainforest and savanna enclaves more biogeographically related to the Atlantic Forest and the Cerrado (Ratter et al., 2006; Queiroz et al., 2017; Silveira et al., 2019), facilitates exchange with these neighboring biomes. This likely accounts for high biological exchange, allowing the successful establishment of ecologically labile species from other regions within the Caatinga.

The environmental differences and physiological tolerances of lineages inhabiting distinct biomes may account for differences recovered by our ancestral biome reconstructions (**Figure 3**). Despite the long-term presence of tropical rainforests in South America (Burnham and Johnson, 2004; Wing et al., 2009), and hence time for lineages to adapt to emerging dry climatic conditions of Northeastern Brazil, shifting events from the Amazon into the Caatinga drylands have been infrequent. Even arrivals from the Atlantic Forest often derived from ancestors that occupied SDTF enclaves there (e.g., *Amorimia* and the genus *Luetzelburgia* within the Vataireoids), not the rainforest biome that largely characterizes the region. This suggests that rainforest species are not able to easily overcome the barriers imposed by such a seasonally dry ecology. Indeed, rainforest species are

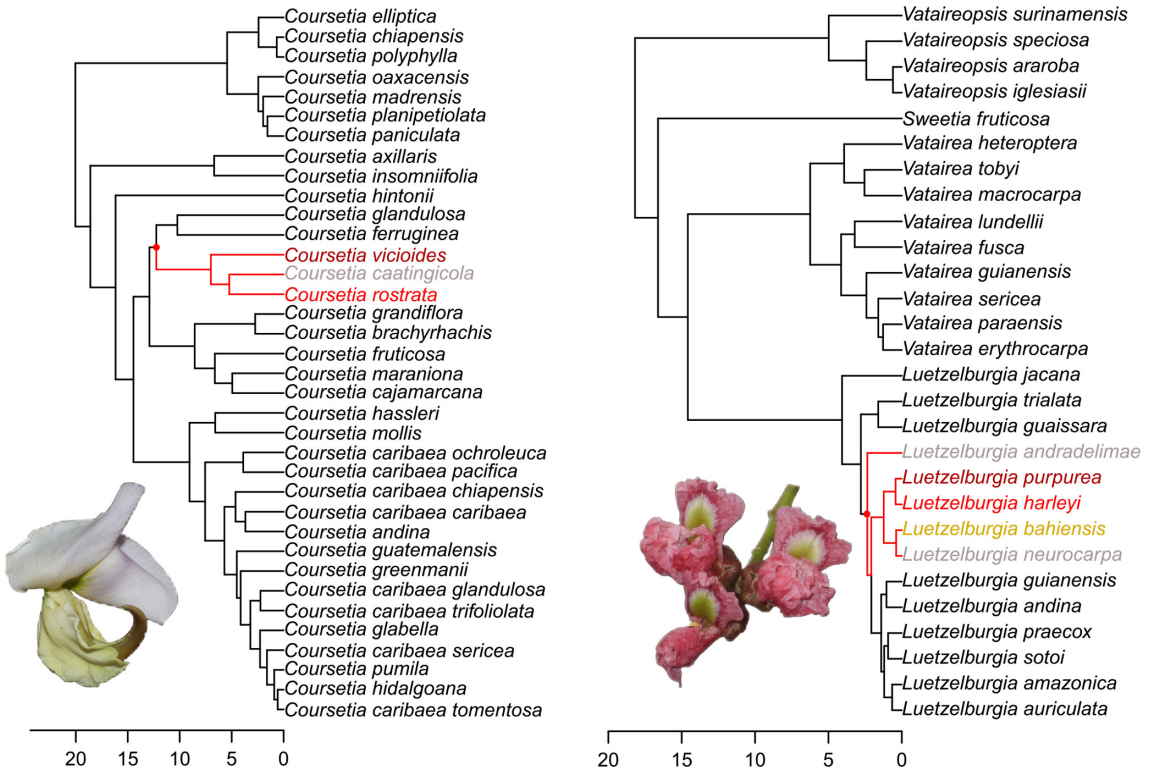


FIGURE 5 | BEAST-dated phylogenies showing examples of Caatinga endemic clades within the mostly SDTF endemic legume clades Coursetia (Queiroz and Lavin, 2011) and the Vataireoids (Cardoso et al., 2013; Cruz et al., 2018). Red dots indicate the nodes where isolation of lineages within the Caatinga occurs. Red lines indicate Caatinga endemic clades. Distinct name colors denote the type of geological substrate where the corresponding species is endemic: Arboreal Tall Caatinga (dark red); more open Crystalline Caatinga (light red); Sedimentary Caatinga (pale yellow); and Karst Caatinga (gray). Note that species do not group by geological substrate. Photos of Coursetia rostrata and Luetzelburgia neurocarpa by Domingos Cardoso.

highly constrained by low water-use efficiency, which means susceptibility to drought-induced mortality (McDowell et al., 2008). Many savanna lineages, on the other hand, may already possess the eco-physiological adaptations needed to cope with a seasonally dry climate, which may explain the greater amount of immigration into the Caatinga of lineages that previously inhabited the Cerrado + Pampas region. Such an active role for climatic similarity is also reflected in mechanisms regulating community-level diversity patterns and greater floristic similarity between SDTF and savannas (Dexter et al., 2015; Fernandes et al., 2020; Segovia et al., 2020).

The Origins of the Dry-Adapted Caatinga Flora Coincide With Regional Events of Massive Denudation and Climate-Induced Vegetation Shifts

Our results indicate that the dry-adapted species that inhabit present-day Caatinga began to emerge during the Eocene-Oligocene, but the low proportion of elements with such an old age (only three were recorded for this period; **Figure 4C**) suggest that a SDTF community resembling a modern one may not have existed at that time. It is possible that such early dry-adapted species remained restricted to suitable niches scattered across the rainforest matrix that dominated most of South America (Fine and Ree, 2006; Jaramillo and Cardenas, 2012). The early presence of rainforest immigrants (from the Amazon and the Atlantic Forest) that formed the initial species pool of modern dry-adapted Caatinga plant lineages suggests a gradual replacement of rainforests by SDTF in the region. These lineages could have gained advantage from an early presence that allowed adaptation to developing more seasonal, arid climates. The more frequent transitions into the Caatinga after the mid-Miocene (**Figure 4C**) are temporally congruent with the establishment of semi-arid climate conditions in Northeastern South America (~13 Ma; Harris and Mix, 2002) and the extensive process of geological denudation that culminated in the formation of the low-lying *Depressão Sertaneja* (Ab'Sáber, 1974; Japsen et al., 2012). The timing of diversification of SDTF plant lineages is dated to Miocene as evidenced from time-calibrated phylogenies (Arakaki et al., 2011) and from the fossil record (Burnham and Carranco, 2004). The high fraction of migrations into the Caatinga by lineages nested within clades largely restricted to SDTF (e.g., *Amorimia*, *Coursetia*, and *Leuenbegeria*) suggests niche conservatism as an important process in the historical assembly of the Caatinga endemic flora, as also observed at various spatial scales and ecological settings (Donoghue and Smith, 2004; Wiens and Donoghue, 2004; Ricklefs et al., 2006;

Donoghue, 2008; Crisp et al., 2009), particularly in the SDTF (Schrire et al., 2005; Lavin, 2006; Pennington et al., 2006, 2009; Govindarajulu et al., 2011; Oliveira-Filho et al., 2013; Gagnon et al., 2019; Ringelberg et al., 2020). As argued by Queiroz (2006), the exposure of the nutrient rich soils that characterizes the *Depressão Sertaneja*, and are preferred by many SDTF species (Pennington et al., 2000, 2004), opened the way for greater migration and successful establishment of dry-adapted lineages with a SDTF ecology that originated elsewhere. Toward the Pliocene, as drier seasonal conditions became more pronounced in South America (Fedorov et al., 2013; Jaramillo et al., 2020) and suitable edaphic conditions more widespread, Caatinga probably coalesced into a seasonally dry community physiognomically similar to the present-day vegetation. This pre-Pleistocene immigration of SDTF elements into the Caatinga adds to growing evidence that historical exchanges between different fragments of the SDTF biome have been primarily driven by long-distance dispersal (Mayle, 2004; Pennington et al., 2004), rather than by the fragmentation of a wider historical formation due to Pleistocene climatic fluctuations (Prado and Gibbs, 1993; Pennington et al., 2000; Prado, 2000).

The observed evolutionary transitions during the Pleistocene involving lineages that inhabit a wide array of different ecologies (e.g., Bombacoids, *Cereus*, *Ficus*, and *Manihot*) indicate that climate-induced vegetational shifts in the recent past may have provided opportunities for evolutionary biome shifts. Fossil evidence (e.g., Oliveira et al., 1999) and paleoecological reconstructions (e.g., Auler et al., 2004; Wang et al., 2004; Arruda et al., 2017) for the Caatinga indicate large shifts in vegetation during this period, which potentially created a complex mosaic of vegetation types representing distinct biomes. Biome-shift events are critically dependent on the spatial adjacency and dimensions of boundaries between biomes over time (Donoghue and Edwards, 2014). The historical climatic instability within the Caatinga might have promoted vegetation shifts that favored the exchange of lineages with other regions, resulting in the relatively high levels of transitions for particular plant clades, and thus facilitating the mixing of lineages with distinct geographical distributions and ecological preferences.

Recurrent Habitat Specialization Underlies the *in situ* Diversification of the Caatinga Endemic Plant Lineages

In the Caatinga, the influence of geology is not only reflected in the impressive species turnover even at local scale, but also on the disproportional contribution of species that are specialist to particular edaphic conditions to regional plant diversity and endemism (Queiroz, 2006; Moro et al., 2016; Queiroz et al., 2017; Fernandes et al., 2020). Such an influence of geological features on plant distribution patterns led Queiroz (2006) to postulate that a great proportion of Caatinga endemics had been generated by allopatric (vicariance) speciation due to Miocene fragmentation of the previously continuous sedimentary surface. This hypothesis clearly implies that pairs of sandy-soil species should occupy disjunct sedimentary basins, and their split should be dated within the Miocene.

General patterns that emerged from our dated phylogenies (**Figure 5**) largely rejected Queiroz's (2006) expectations. Rather, they suggest that parallel *in situ* diversification of plant lineages has probably evolved by fine-scale niche differentiation, reflecting adaptations of the regional flora to different conditions over evolutionary timescales. The time span of *in situ* speciation events coincides with changes in the geology of the Caatinga. Post-Miocene uplifting of the Araripe plateau, for example, induced dramatic changes in the course of the São Francisco River and the formation of an endorheic drainage (with no outflow to the ocean) in its middle course (Mabesoone, 1994; Maia et al., 2010; Barros-Corrêa et al., 2019). As a result of sedimentation of this ancient endorheic drainage and posterior action of wind, a system of continental sand dunes was formed later in the Pleistocene (Barreto and Suguio, 1993; Mabesoone, 1994; Barreto, 1996). It was also during the Pleistocene that most of the Caatinga karst systems, formed by previous marine incursions into the continent, were exhumed (Silva et al., 2017). The plant communities on both sand dunes and karstic rocks are physiognomically distinctive and rich in endemism (e.g., Rocha, 1995; Rocha et al., 2004; Queiroz et al., 2017; Fernandes et al., 2020).

Although greater topographic heterogeneity caused by continuation of regional denudation (**Figures 2A–D**) might have played a role in diversification of Caatinga endemics, the disproportional contribution of edaphic specialization to regional plant diversity suggests that the formation of soils with contrasting textures and fertility, derived from distinct parental material, have more likely acted as the trigger of Caatinga's plant diversification through ecological speciation. The emergence of completely new environments created ecological opportunities for divergence of lineages of the regional species-pool within these edaphic islands. Indeed, edaphic specialization has been previously designated as an important driver of diversification, endemism, and spatial genetic structure of plants (e.g., Fine et al., 2005, 2013; DRYFLOR et al., 2016; Cacho and Strauss, 2014), happening relatively rapidly in some lineages (Rajakaruna, 2017). Factors other than the chemical and textural nature of the substrate, however, may be responsible for the diversification, and the nature of the substrate may be only a surrogate for yet unmeasured factors. For example, the water retention capacity within the *Depressão Sertaneja* is critically impacted by the features of its rocky shallow soils, derived from non-porous crystalline basement rocks, a characteristic that is not present on more sandy soils or even on the deeper soils of mountains slopes (Queiroz et al., 2017).

CONCLUDING REMARKS

Relatively few studies, most of which are lineage-specific, have investigated plant evolution in the isolated patches of the SDTF biome. Despite their important biogeographical insights, we lacked basic understanding on the timing of origin and the evolutionary causes that shaped the high levels of diversity and endemism in the key SDTF nucleus of the Caatinga. Here, we used a multi-taxon phylogenetic approach to provide a

comprehensive assessment of the evolutionary dynamics shaping current patterns of plant diversity in the Caatinga. We showed that the historical assembly of plant lineages in this region has been more complex than previously proposed (e.g., Sarmiento, 1975; Rizzini, 1979; Andrade-Lima, 1981, 1982; Queiroz, 2006). The dry-adapted Caatinga plant communities resulted from a mosaic of apparently idiosyncratic evolutionary histories, involving pervasive, accumulating immigration from nearby biogeographic regions, and relatively recent *in situ* speciation events. Transitions into the Caatinga have been facilitated by both geographic accessibility (reflected in the large number of floristic exchanges with surrounding regions) and niche conservatism (reflected in the large number of Caatinga lineages nested within clades largely restricted to the SDTF). In contrast to what had been hypothesized by Queiroz (2006), our results show that vicariance processes did not drive allopatric speciation of Caatinga endemics but that the high levels of plant species endemism are more likely a by-product of ecological speciation. We hypothesize that global aridification processes may explain the antiquity of some Caatinga SDTF lineages, whereas the newly opened habitats in the region, as well as Pleistocene climate-induced biome reconfigurations and vegetation shifts, have jointly contributed to more recent *in situ* plant diversification.

The model of geological evolution for the Caatinga region summarized here (**Figure 2**) provides a starting point for research aimed at understanding the mechanisms underlying the assembly of many other organisms. For example, the Caatinga also harbors high diversity and endemism of aquatic plants (Queiroz et al., 2017) adapted to temporary lakes and flooded areas on the margins of rivers or lowered terrains (e.g., Moro et al., 2014; The Brazil Flora Group, 2015; Queiroz et al., 2017). Investigating whether the Pliocene São Francisco "paleo-lake" has influenced the evolution of these lineages, just like the Pebas system may have contributed to the evolution of aquatic plants in the Amazon basin (e.g., Hoorn et al., 2010), is an important future question. Investigations should also be extended to beyond plants, for example to the seasonal killifishes (Rivulideae) which are highly diverse in the Caatinga, with impressive rates of endemism in the middle São Francisco River basin (Costa et al., 2013).

While our comparative approach across multiple lineages of ecologically and geographically important Caatinga plant families has revealed a more complete picture of the evolutionary history in the Caatinga SDTF, many underlying evolutionary processes require further study. For example, most Sedimentary Caatinga endemic lineages in our sampling are restricted to the São Francisco Dune System and examples need expanding to the endemic flora of the Paleo-Mesozoic sedimentary basins. The inclusion of such species will be particularly important to determine the differential effect of geological history over the disjunctly distributed sedimentary plant communities.

Future studies should also focus on the remarkable cases of transitions into Caatinga without subsequent diversification, which generally encompass relatively old endemic genera (Pennington et al., 2006). For example, *Holoregmia* and *Tabaroa* were included in our sampling, but 29 endemic genera have already been recorded for the Caatinga (Queiroz et al., 2017; Fernandes et al., 2020). Species occurring in multiple Caatinga environments (e.g., crystalline, sedimentary, and arboreal) or disjunct nuclei of the same environment (e.g., sedimentary species occurring on *Tucano-Jatobá* and *Araripe* sedimentary basins) should also be focus of finer-scale studies on incipient and cryptic speciation, a pattern that seem to be relatively common in other neotropical SDTF (e.g., Särkinen et al., 2011). Detailed phylogenetic studies, including robust sampling with multiple accessions of individual species may reveal how frequent cryptic speciation is, shedding light on the processes driving high genetic divergence despite little morphological variation.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

MF and LQ designed the original project. MF performed the analyses, interpreted the results, and drafted the manuscript. LQ, RP, and DC supervised the study and revised the manuscript. All authors conceived the main underlying ideas, presentation of the study, and approved the final manuscript.

FUNDING

This work was part of MF Ph.D. research. MF acknowledges the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for a doctorate fellowship (#141560/2015-0) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for funding a "sandwich" scholarship (PDSE 88881.133545/2016-01) at the Royal Botanic Garden Edinburgh. DC and LQ acknowledge their research productivity fellowships (Grant Nos. 306736/2015-2 and 308244/2018-4, respectively) from CNPq. DC's research in plant biodiversity and evolution was also supported by the Royal Society (Newton Advanced Fellowship Grant No. NAF/R1/180331), CNPq (Universal Grant No. 422325/2018), and Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB, Universal Grant No. APP0037/2016). All authors acknowledge support of the NERC and FAPESP grant "NORDESTE" (NE/N012550/1, #2015/50488-5).

ACKNOWLEDGMENTS

We thank Lucas Cardoso (Universidade Federal do Maranhão) for helping with edition of figures, Cássia Bitencourt (Royal Botanic Gardens Kew) for assistance with edition of sequences, and to Alessandro Rapini (Universidade Estadual de Feira de Santana), Danilo Neves (Universidade Federal de Minas Gerais), Henrique Batalha-Filho (Universidade Federal da Bahia), Lígia Silveira Funch (Universidade Estadual de Feira de Santana),

Rubson Pinheiro Maia (Universidade Federal do Ceará), Carlos Jaramillo, and an anonymous reviewer for insightful suggestions. We also thank Danilo Neves for inviting us to join the special issue "Temporal and large-scale spatial patterns of plant diversity and diversification."

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2022.723286/full#supplementary-material

REFERENCES

Ab'Sáber, N. A. (1974). O domínio morfoclimático semiárido das Caatingas brasileiras. *Geomorfologia* 43, 1–39.

Almeida, R. F., Amorim, A. M. A., and van den Berg, C. (2018). Timing the origin and past connections between Andean and Atlantic Seasonally Dry Tropical Forests in South America: insights from the biogeographical history of Amorimia (Malpighiaceae). *Taxon* 67, 739–751. doi: 10.12705/674.4

Andrade-Lima, D. (1981). The caatingas dominium. *Rev. Bras. Bot.* 4, 149–153.

Andrade-Lima, D. (1982). "Present day forest refuges in Northeastern Brazil," in *Biological Diversification in the Tropics*, ed. G. T. Prance (New York, NY: Columbia University Press), 245–251.

Antonelli, A., Zizka, A., Carvalho, F. A., Scharn, R., Bacon, C. D., Silvestro, D., et al. (2018). Amazonia is the primary source of Neotropical biodiversity. *Proc. Natl. Acad. Sci. U.S.A.* 115, 6034–6039. doi: 10.1073/pnas.1713819115

Arakaki, M., Christin, P.-A., Nyffeler, R., Lendel, A., Eggli, U., Ogburn, R. M., et al. (2011). Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8379–8384. doi: 10.1073/pnas.1100628108

Arruda, D. M., Schaefer, C. E. G. R., Fonseca, R. S., Solar, R. R. C., and Fernandes-Filho, E. I. (2017). Vegetation cover of Brazil in the last 21 ka: new insights into the Amazonian refugia and Pleistocenic arc hypotheses. *Glob. Ecol. Biogeogr.* 27, 47–56.

Auler, A. S., Wang, X., and Edwards, R. L. (2004). Quaternary ecological and geomorphic changes associated with rainfall events in presently semi-arid north-eastern Brazil. *J. Quat. Sci.* 19, 693–701.

Banda, K., Delgado, A., Dexter, K., Linares, R., Oliveira, A., Prado, D., et al. (2016). Plant diversity patterns in neotropical dry forests and their conservation implications. *Science* 353, 1383–1387. doi: 10.1126/science.aaf5080

Barreto, A. M. F. (1996). *Interpretação Paleoambiental do Sistema de Dunas Fixadas do Médio Rio São Francisco, Bahia*. São Paulo :SP: Universidade de São Paulo. [PhD thesis].

Barreto, A. M. F., and Suguio, K. (1993). *Considerações Sobre a Idade e a Paleogeografia das Paleodunas do Médio Rio São Francisco, Bahia. Congresso da Associação Brasileira de Estudos do Quaternário*. São Paulo: ABEQUA, 11–11.

Barros-Corrêa, A. C., Tavares, B. A. C., Lira, D. R., Mutzenberg, D. S., and Cavalcanti, L. C. S. (2019). "The semi-arid domain of the northeast of brazil," in *The Physical Geography of Brazil. Geography of the Physical Environment,* eds A. Salgado, L. Santos, and J. Paisani J. (Cham: Springer International Publishing AG), 119–150. doi: 10.1007/978-3-030-04333-9_7

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., et al. (2013). GenBank. *Nucleic Acids Res.* 41, D36–D42.

Bezerra, F. H. R., Nascimento, A. F., Ferreira, J. M., Nogueira, F. C., Fuck, R. A., Neves, B. B. B., et al. (2011). Review of active faults in the Borborema Province, intraplate South America integration of seismological and paleoseismological data. *Tectonophysics* 510, 269–290. doi: 10.1016/j.tecto.2011.08.005

The Brazil Flora Group [BFG] (2015). Growing knowledge: an overview of seed plant diversity in Brazil. *Rodriguesia* 66, 1085–1113.

Burnham, R. J., and Carranco, N. L. (2004). Miocene winged fruits of Loxopterygium (Anacardiaceae) from the Ecuadorian Andes. *Am. J. Bot.* 91, 1767–1773. doi: 10.3732/ajb.91.11.1767

Burnham, R. J., and Johnson, K. R. (2004). South American paleobotany and the origins of Neotropical rainforests. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 1595–1610. doi: 10.1098/rstb.2004.1531

Cacho, N. I., and Strauss, S. Y. (2014). Occupation of bare habitats, an evolutionary precursor to soil specialization in plants. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15132–15137. doi: 10.1073/pnas.1409242111

Cardoso, D., Queiroz, L. P., Lima, H. C., Suganuma, E., van den Berg, C., and Lavin, M. (2013). A molecular phylogeny of the vataireoid legumes underscores floral evolvability that is general to many early-branching papilionoid lineages. *Am. J. Bot.* 100, 403–421. doi: 10.3732/ajb.1200276

Cardoso, D. B. O. S., and Queiroz, L. P. (2007). Diversidade de Leguminosae nas Caatingas de Tucano, Bahia: implicações para a fitogeografia do semiárido do Nordeste do Brasil. *Rodriguesia* 58, 379–391.

Carvalho-Sobrinho, J. G., Alverson, W. S., Alcantara, S., Queiroz, L. P., Mota, A. C., and Baum, D. A. (2016). Revisiting the phylogeny of Bombacoideae (Malvaceae): novel relationships, morphologically cohesive clades, and a new tribal classification based on multilocus phylogenetic analyses. *Mol. Phylogenet. Evol.* 101, 56–74. doi: 10.1016/j.ympev.2016.05.006

Costa, W. J. E. M., Amorim, P. F., and Bragança, P. H. N. (2013). Species limits and phylogenetic relationships of red-finned cryptic species of the seasonal killifish genus Hypsolebias from the Brazilian semi-arid Caatinga (Teleostei: Cyprinodontiformes: Rivulidae). *J. Zool. Syst. Evol. Res.* 52, 52–58.

Crisp, M., Arroyo, M. T., Cook, L. G., Gandolfo, M. A., Jordan, G. J., McGlone, M. S., et al. (2009). Phylogenetic biome conservatism on a global scale. *Nature* 458, 754–758. doi: 10.1038/nature07764

Cruz, D. T., Idárraga, A., Banda, K., Cogollo, A., van den Berg, C., Queiroz, L. P., et al. (2018). Ancient speciation of the papilionoid legume *Luetzelburgia jacana*, a newly discovered species in an inter-Andean seasonally dry valley of Colombia. *Taxon* 67, 931–943. doi: 10.12705/675.6

Darriba, D., Taboada, G., Doallo, R., and Posada, D. (2012). jModelTest2: more models, new heuristics and parallel computing. *Nat. Methods* 9:772. doi: 10.1038/nmeth.2109

Dexter, K. G., Smart, B., Baldauf, C., Baker, T. R., Balinga, M. P. B., Brienen, R. J. W., et al. (2015). Floristics and biogeography of vegetation in seasonally dry tropical regions. *Int. For. Rev.* 17, 10–32. doi: 10.1016/j.pld.2016.11.004

Donoghue, M. (2008). A phylogenetic perspective on the distribution of plant diversity. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11549–11555. doi: 10.1073/pnas.0801962105

Donoghue, M. J., and Edwards, E. J. (2014). Biome shifts and niche evolution in plants. *Annu. Rev. Ecol. Evol. Syst.* 45, 547–572. doi: 10.1146/annurev-ecolsys-120213-091905

Donoghue, M. J., and Smith, S. A. (2004). Patterns of the assembly of temperate forests around the northern hemisphere. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 1633–1644. doi: 10.1098/rstb.2004.1538

Drummond, A. J., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R., et al. (2012a). *Geneious Version 6.1.6*. http://www.geneious.com (accessed February 17, 2021).

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012b). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075

DRYFLOR, Karina Banda, R., Dexter, K. G., Linares-Palomino, R., Oliveira-Filho, A., Prado, D., et al. (2016). Plant diversity patterns in neotropical dry forests and their conservation implications. 353, 1383–1387. doi: 10.1371/journal.pone.0062639

Fedorov, A. V., Brierley, C. M., Lawrence, K. T., Liu, Z., Dekens, P. S., and Ravelo, A. C. (2013). Patterns and mechanisms of early Pliocene warmth. *Nature* 496, 43–49. doi: 10.1038/nature12003

Fernandes, M. F., Cardoso, D., and Queiroz, L. P. (2020). An updated plant checklist of the Brazilian Caatinga seasonally dry forests and woodlands reveals high species richness and endemism. *J. Arid Environ.* 174:104079. doi: 10.1016/j.jaridenv.2019.104079

Fiaschi, P., and Pirani, J. R. (2009). Review of plant biogeographic studies in Brazil. *J. Syst. Evol.* 47, 477–496.

Fine, P. V. A., Daly, D. C., Muñoz, G. V., Mesones, I., and Cameron, K. (2005). The contribution of edaphic heterogeneity to the evolution and diversity of Burseraceae trees in the Western Amazon. *Evolution* 59, 1464–1478. doi: 10.1554/04-745

Fine, P. V. A., and Ree, R. H. (2006). Evidence for a time-integrated species-area effect on the latitudinal gradient in tree diversity. *Am. Nat.* 168, 796–804. doi: 10.1086/508635

Fine, P. V. A., Zapata, F., Daly, D. C., Mesones, I., Misiewicz, T. M., Cooper, H. F., et al. (2013). The importance of environmental heterogeneity and spatial distance in generating phylogeographic structure in edaphic specialist and generalist tree species of Protium (Burseraceae) across the Amazon Basin. *J. Biogeogr.* 40, 646–661.

Gagnon, E., Ringelberg, J. J., Bruneau, A., Lewis, G. P., and Hughes, C. E. (2019). Global Succulent Biome phylogenetic conservatism across the pantropical Caesalpinia Group (Leguminosae). *New Phytol.* 222, 1994–2008. doi: 10.1111/nph.15633

Gomes, A. P. S., Rodal, M. J. N., and Melo, A. L. (2006). Florística e fitogeografia da vegetação arbustiva subcaducifólia da Chapada de São José. *Buíque PE Brasil Acta Bot. Bras.* 20, 37–48.

Gottsberger, G., and Silberbauer-Gottsberger, I. (2006). *Life in the Cerrado: a South American tropical seasonal ecosystem*. Germany: Reta Verlag.

Govindarajulu, R., Hughes, C. E., and Bailey, C. D. (2011). Phylogenetic and population genetic analyses of diploid Leucaena (Leguminosae) reveal cryptic species diversity and patterns of allopatric divergent speciation. *Am. J. Bot.* 98, 2049–2063. doi: 10.3732/ajb.1100259

Gurgel, S. P. P., Bezerra, F. H. R., Corrêa, A. C. B., Marques, F. O., and Maia, R. P. (2013). Cenozoic uplift and erosion of structural landforms in NE Brazil. *Geomorphology* 186, 68–84.

Harris, S. E., and Mix, A. C. (2002). Climate and tectonics drive continental erosion of tropical South America, 0-13 Ma. *Geology* 30, 447–450.

Hoorn, C., Bernardes-de-Oliveira, M. E. C., Dino, R., Garcia, M. J., Antonioli, L., Casado, F. C., et al. (2014). "Neogene climate evolution in Amazonia and the Brazilian Northeast," in *Paleontologia: Cenários de vida - Paleoclimas*, eds I. S. Carvalho, M. J. Garcia, O. Strohschoen, and C. C. Lana (Rio de Janeiro, RJ: Editora Interciência), 277–310.

Hoorn, C., Wesselingh, F. P., Ter Steege, H., Bermudez, M. A., Mora, A., Sevink, J., et al. (2010). Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* 330, 927–931. doi: 10.1126/science.1194585

Japsen, P., Bonow, J. M., Green, P. F., Cobbold, P. R., Chiossi, D., Lilletveit, R., et al. (2012). Episodic burial and exhumation in NE Brazil after opening of the South Atlantic. *Geol. Soc. Am. Bull.* 124, 800–816. doi: 10.1130/b30515.1

Jaramillo, C., and Cardenas, A. (2012). Global warming and Neotropical rainforests, a historical perspective. *Annu. Rev. Earth Planet. Sci.* 41, 741–766.

Jaramillo, C., Sepulchre, P., Cardenas, D., Correa-Metrio, A., Moreno, J. E., Trejos, R., et al. (2020). Drastic Vegetation Change in the Guajira Peninsula (Colombia) During the Neogene. Paleoceanogr. *Paleoclimatol* 35:e2020PA003933.

Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Klöckinga, M., Hoggard, M. J., Rodríguez-Tribaldos, V., Richards, F. D., Guimarães, A. R., Maclennan, J., et al. (2020). A tale of two domes: Neogene to recent volcanism and dynamic uplift of northeast Brazil and southwest Africa. *Earth Planet. Sci. Lett.* 547:116464. doi: 10.1016/j.epsl.2020.116464

Lavin, M. (2006). "Floristic and geographical stability of discontinuous seasonally dry tropical forests explains patterns of plant phylogeny and endemism," in *Neotropical savannas and seasonally dry forests: plant diversity, biogeography, and conservation*, eds G. P. Lewis and J. A. Ratter (London, UK: CRC Press), 433–447.

Löwenberg-Neto, P. (2014). Neotropical region: a shapefile of Morrone's (2014) biogeographical regionalisation. *Zootaxa* 3802:300. doi: 10.11646/zootaxa.3802.2.12

Mabesoone, J. M. (1994). *Sedimentary Basins of Northeast Brazil*. Recife: Editora Universitária UFPE.

Magnavita, L. P., Davison, I., and Kusznir, N. J. (1994). Rifting, erosion and uplift history of the Reconcavo–Tucano–Jatoba Rift, northeast Brazil. *Tectonics* 13, 367–388. doi: 10.1029/93tc02941

Maia, R. P., Bezerra, F. H., and Sales, V. C. (2010). Geomorfologia do Nordeste: concepções clássicas e atuais acerca das superfícies de aplainamento. *Rev. Bras. Geogr.* 27, 6–19.

Matzke, N. J. (2013). *BioGeoBEARS: Biogeography with Bayesian (and likelihood) Evolutionary Analysis in R Scripts*. Berkeley, CA: University of California.

Maya-Lastra, C. A., and Steinmann, V. W. (2019). Evolution of the untouchables: Phylogenetics and classification of Cnidoscolus (Euphorbiaceae). *Taxon* 68, 692–713.

Mayle, F. E. (2004). Assessment of the Neotropical dry forest refugia hypothesis in the light of palaeoecological data and vegetation model simulations. *J. Quat. Sci.* 19, 713–720.

McDowell, N., Pockman, W. T., Allen, C. D., Breshears, D. D., Cobb, N., Kolb, T., et al. (2008). Mechanisms of plant survival and mortality during drought: why do some plants survive while others succumb to drought? *New Phytol.* 178, 719–739. doi: 10.1111/j.1469-8137.2008.02436.x

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). *Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees*. Piscataway, NJ: IEEE, 1–8.

Miranda, P. L. S., Oliveira-Filho, A. T., Pennington, R. T., Neves, D. M., Baker, T. R., and Dexter, K. G. (2018). Using tree species inventories to map biomes and assess their climatic overlaps in lowland tropical South America. *Glob. Ecol. Biogeogr.* 27, 899–912. doi: 10.1111/geb.12749

Moro, M. F., Nic Lughadha, E., de Araújo, F. S., and Martins, F. R. (2016). A phytogeographical metaanalysis of the semiarid Caatinga Domain in Brazil. *Bot. Rev.* 82, 91–148. doi: 10.1007/s12229-016-9164-z

Moro, M. F., Sousa, D. J. L., and Matias, L. (2014). Rarefaction, richness estimation and extrapolation methods in the evaluation of unseen plant diversity in aquatic ecosystems. *Aquat. Bot.* 117, 48–55. doi: 10.1016/j.aquabot.2014.04.006

Morrone, J. J. (2014). Biogeographical regionalisation of the Neotropical region. *Zootaxa* 3782, 1–110. doi: 10.11646/zootaxa.3782.1.1

Oliveira, P. E., Barreto, A. M. F., and Suguio, K. (1999). Late Pleistocene-Holocene climatic and vegetational history of the Brazilian Caatinga: the fossil dunes of the middle São Francisco River. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 152, 319–337.

Oliveira, R. G., and Medeiros, W. E. (2012). Evidences of buried loads in the base of the crust of Borborema Plateau (NE Brazil) from Bouguer admittance estimates. *J. South Am. Earth Sci.* 37, 60–76.

Oliveira-Filho, A. T., Cardoso, D., Schrire, D., Lewis, G. P., Pennington, R. T., Brummer, T. J., et al. (2013). Stability structures tropical woody plant diversity more than seasonality: Insights into the ecology of high legume-succulent-plant biodiversity. *S. Afr. J. Bot.* 89, 42–57.

Pennington, R. T., Lavin, M., and Oliveira-Filho, A. T. (2009). Woody plant diversity, evolution and ecology in the tropics: perspectives from seasonally dry tropical forests. *Ann. Rev. Ecol. Evol. Syst.* 40, 437–457.

Pennington, R. T., Lavin, M., Prado, D. E., Pendry, C. A., Pell, S. K., and Butterworth, C. A. (2004). Historical climate change and speciation: neotropical seasonally dry forest plants show patterns of both tertiary and quaternary diversification. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 515–538. doi: 10.1098/rstb.2003.1435

Pennington, R. T., Lewis, G. P., and Ratter, J. A. (2006). "An overview of the plant diversity, biogeography and conservation of Neotropical savannas and seasonally dry forests," in *Neotropical Savannas and Seasonally Dry Forests: plant Diversity, Biogeography, and Conservation*, eds R. T. Pennington, G. P. Lewis, and J. A. Ratter (London, UK: CRC Press), 193–211.

Pennington, R. T., Prado, D. E., and Pendry, C. A. (2000). Neotropical seasonally dry forests and Pleistocene vegetation changes. *J. Biogeogr.* 27, 261–273.

Peulvast, J. P., and Claudino-Sales, V. (2004). Stepped surfaces and paleolandforms in the Northern Brazilian "Nordeste": constraints on models of morphotectonic evolution. *Geomorphology* 3, 89–122. doi: 10.1016/j.geomorph.2004.02.006

Peulvast, J. P., Claudino-Sales, V., Bétard, F., and Gunnell, Y. (2008). Low post-Cenomanian denudation depths across the Brazilian Northeast: implications for long-term landscape evolution at a transform continental margin. *Glob. Planet. Change* 62, 39–60. doi: 10.1016/j.gloplacha.2007.11.005

Piel, W. H., Chan, L., Dominus, M. J., Ruan, J., Vos, R. A., and Tannen, V. (2009). *v. 2: a database of phylogenetic knowledge. In: e-BioSphere 2009*. Available online at: https://treebase.org/treebase-web/reference.html (accessed February 17, 2021).

Prado, D. E. (2000). Seasonally dry forests of tropical South America: from forgotten ecosystems to a new phytogeographic unit. *Edinb. J. Bot.* 57, 437–461. doi: 10.1017/s096042860000041x

Prado, D. E., and Gibbs, P. E. (1993). Patterns of species distributions in the dry seasonal forests of South America. *Ann. Mo. Bot. Gard.* 80, 902–927.

Queiroz, L. P. (2006). "The Brazilian Caatinga: phytogeographical patterns inferred from distribution data of the Leguminosae," in *Neotropical Savannas and Seasonally Dry Forests: plant Diversity, Biogeography, and Conservation*, eds

R. T. Pennington, G. P. Lewis, and J. A. Ratter (London, UK: CRC Press), 121–158.

Queiroz, L. P., Cardoso, D., Fernandes, M. F., and Moro, M. F. (2017). "Diversity and evolution of flowering plants of the Caatinga Domain," in *Caatinga: the Largest Tropical Dry Forest Region in South America*, eds J. M. C. Silva, I. R. Leal, and M. Tabarelli (Cham, CH: Springer International Publishing AG), 23–63. doi: 10.1099/ijsem.0.004193

Queiroz, L. P., and Lavin, M. (2011). Coursetia (Leguminosae) from eastern Brazil: nuclear ribosomal and chloroplast DNA sequence analysis reveal the monophyly of three caatinga-inhabiting species. *Syst. Bot.* 36, 69–79.

Queiroz, L. P., Lewis, G. P., and Wojciechowski, M. F. (2010). Tabaroa, a new genus of Leguminosae tribe Brongniartieae from Brazil. *Kew Bull.* 65, 189–203. doi: 10.1007/s12225-010-9202-7

R Core Team (2020). *R: a Language and Environment for Statistical Computing*. Vienna, AUS: R Foundation for Statistical Computing.

Rajakaruna, N. (2017). Lessons on evolution from the study of edaphic specialization. *Bot. Rev.* 84, 1–40.

Rambaut, A., and Drummond, A. J. (2013). *Tracer version 1.6. Computer Program and Documentation Distributed by the Author*. http://beast.bio.ed.ac.uk/Tracer (accessed February 17, 2021).

Ratter, J. A., Bridgewater, S., and Ribeiro, J. F. (2006). "Biodiversity patterns of the woody vegetation of the Brazilian Cerrados," in *Neotropical Savannas and Seasonally Dry Forests: Plant Diversity, Biogeography and Conservation*, eds R. T. Pennington, J. A. Ratter, and G. P. Lewis (London, UK: CRC Press), 31–66. doi: 10.1201/9781420004496.ch2

Ricklefs, R. E., Schwarzbach, A. E., and Renner, S. S. (2006). Rate of lineage origin explains the diversity anomaly in the world's mangrove vegetation. *Am. Nat.* 168, 805–810. doi: 10.1086/508711

Ringelberg, J. J., Zimmermann, N. E., Weeks, A., Lavin, M., and Hughes, C. E. (2020). Biomes as evolutionary arenas: convergence and conservatism in the trans-continental succulent biome. *Glob. Ecol. Biogeogr.* 29, 1100–1113.

Rizzini, C. T. (1979). *Tratado de Fitogeografia do Brasil*. São Paulo: Editora da USP.

Rocha, P. L. (1995). Proechimys yonenagae, a new species of spiny rat (Rodentia: Echimyidae) from fossil sand dunes in Brazilian Caatinga. *Mammalia* 59, 537–550.

Rocha, P. L., Queiroz, L. P., and Pirani, J. R. (2004). Plant species and habitat structure in a sand dune field in the Brazilian Caatinga: a homogeneous habitat harbouring an endemic biota. *Rev. Bras. Bot.* 27, 739–755.

Rodríguez-Tribaldos, V., White, N. J., Roberts, G. G., and Hoggard, M. J. (2017). Spatial and temporal uplift history of South America from calibrated drainage analysis. *Geochem. Geophys. Geosyst.* 18, 2321–2353. doi: 10.1002/2017gc006909

Santos, R. M., Oliveira-Filho, A. T., Eisenlohr, P. V., Queiroz, L. P., Cardoso, D. B. O. S., and Rodal, M. J. N. (2012). Identity and relationships of the arboreal Caatinga among other floristic units of seasonally dry tropical forests (SDTFs) of north-eastern and Central Brazil. *Ecol. Evol.* 2, 409–428. doi: 10.1002/ece3.91

Särkinen, T., Marcelo-Peña, J. L., Yomona, A. D., Simon, M. F., and Pennington, R. T. (2011). Underestimated endemic species diversity in the dry inter-Andean valley of the Río Marañón, northern Peru: an example from Mimosa (Leguminosae. Mimosoideae). *Taxon* 60, 139–150.

Sarmiento, G. (1975). The dry plant formations of South America and their floristic connections. *J. Biogeogr.* 2, 233–251. doi: 10.1371/journal.pone.0233729

Schrire, B. D., Lavin, M., and Lewis, G. P. (2005). Global distribution patterns of the Leguminosae: insights from recent phylogenies. *Biol. Skr.* 55, 375–386.

Segovia, R. A., Pennington, R. T., Baker, T. R., Souza, F. C., Neves, D. M., Davis, C. C., et al. (2020). Freezing and water availability structure the evolutionary diversity of trees across the Americas. *Sci. Adv.* 6:eaaz5373. doi: 10.1126/sciadv.aaz5373

Silva, A. C., and Souza, A. F. (2018). Aridity drives plant biogeographical subregions in the Caatinga, the largest tropical dry forest and woodland block in South America. *PLoS One* 13:e0196130. doi: 10.1371/journal.pone.0196130

Silva, O. L., Bezerra, F. H. R., Maia, R. P., and Cazarin, C. L. (2017). Karst landforms revealed at various scales using LiDAR and UAV in semi-arid Brazil: Consideration on karstification processes and methodological constraints. *Geomorphology* 295, 611–630.

Silveira, M. H. B., Mascarenhas, R., Cardoso, D., and Batalha-Filho, H. (2019). Pleistocene climatic instability drove the historical distribution of forest islands in the northeastern Brazilian Atlantic Forest. *Palaeogeog. Palaeoclimatol. Palaeoecol.* 527, 67–76. doi: 10.1016/j.palaeo.2019.04.028

Simon, M. F., Grether, R., Queiroz, L. P., Särkinen, T. E., Dutra, V. F., and Hughes, C. E. (2011). The Evolutionary history of Mimosa (Leguminosae): toward a phylogeny of the sensitive plants. *Am. J. Bot.* 98, 1201–1221. doi: 10.3732/ajb.1000520

Valadão, R. C. (2009). Geodinâmica de superfícies de aplanamento, desnudação continental e tectônica ativa como condicionantes da megageomorfologia do Brasil oriental. *Braz. J. Geol.* 10, 77–90.

Velloso, A. L., Sampaio, E. V. S. B., Giulietti, A. M., Barbosa, M. R. V., Castro, A. A. J. F., Queiroz, L. P., et al. (2002). *Ecorregiões: propostas Para o Bioma Caatinga*. Recife: Associação Plantas do Nordeste and The Nature Conservancy do Brasil.

Wang, X., Auler, A. S., Edwards, R. L., Cheng, H., Cristalli, P. S., Smart, P. L., et al. (2004). Wet periods in Northeastern Brazil over the past 210 ka linked to distant climate anomalies. *Nature* 432, 740–743. doi: 10.1038/nature03067

Wiens, J. J., and Donoghue, M. J. (2004). Historical biogeography, ecology and species richness. *Trends Ecol. Evol.* 19, 639–644. doi: 10.1016/j.tree.2004.09.011

Wing, S. L., Herrera, F., Jaramillo, C. A., Gomez-Navarro, C., Wilf, P., and Labandeira, C. C. (2009). Late Paleocene fossils from the Cerrejón Formation, Colombia, are the earliest record of Neotropical rainforest. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18627–18632. doi: 10.1073/pnas.0905130106

Zachos, J. C., Pagani, M. O., Sloan, L. C., Thomas, E., and Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292, 686–693. doi: 10.1126/science.1059412

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership