Family Matters: Harnessing Big Data to Investigate How Phylogenetic Niche Conservatism Has Shaped the Distributions of Contemporary Angiosperms Across the Americas

by

Héctor Fox-Figueroa

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in the University of Michigan
2023

Doctoral Committee:
    Professor Stephen A. Smith, Chair
    Professor Chris W. Dick
    Professor Andrew Marshall
    Associate Professor Annette M. Ostling

Black Lives Matter

No One is Illegal on Stolen Land

Héctor Fox-Figueroa

hecfox@umich.edu

ORCID: 0000-0001-9785-8529

DEDICATION

"The universe is made of stories, not of atoms" –Muriel Rukeyser

This work is dedicated first to the Earth: that pale blue dot on which all our work and all our lives are spent, and from which all that we study and cherish was born.

Thank you to my ancestors, both living and dead, for they have guided me and helped me to find my path. Without them, I could not be.

This work is also dedicated to the generations yet to be. May the small bits of knowledge offered here help those to come to understand and appreciate this Earth—our home—just a little bit more.

ACKNOWLEDGEMENTS

Scientific research cannot (and in my opinion *should* not) be practiced alone by a single individual. Although I have led the research projects presented in this dissertation, I would not have been able to do any of this research without help—both professional and personal—from literally dozens, if not hundreds, of individuals. I attempt here to specifically thank some of the people and institutions that helped me to achieve my goal of obtaining a doctoral degree.

The dissertation projects presented here, specifically those detailed in Chapters 2-4, were funded, in whole or in part, by NSF grant DBI 1930030, "*Integrating data communities with BiotaPhy: a computational platform for data-intensive biodiversity research and training*". This grant helped introduce me to many of the coauthors and collaborators contributing to Chapters 2-4, and I am very grateful to each of them for their contributions on these projects. Specifically: CJ Grady, Dr. Jim Beach, and Aimee Stewart at Kansas State for aggregating literally millions of collection records into a single, cleaned dataset and then creating a smooth pipeline for turning those occurrence records into a massive niche model dataset; Drs. Pamela and Douglas Soltis at University of Florida and Dr. Ryan A. Folk at Mississippi State for their thoughtful critiques and suggestions on how best to present results and analyze data; and Maria Beatriz de Souza Cortez for excellent writing suggestions as well as discussions on how we, as the new generation of scientists, should advocate for our identities.

I would like to thank Dr. Hannah Marx, now at University of New Mexico, who helped me as my unofficial mentor throughout the latter phases of my dissertation, and who might love the mountains even more than I do.

I am deeply indebted to Dr. Maya Barak at University of Michigan Dearborn. She is not only my oldest friend but also a constant source of encouragement and support and our discussions on teaching philosophies and the culture of academia have helped shape me into the professional I am today.

I am deeply grateful to my two dissertation co-chairs, whose guidance has helped me to grow and develop into a successful ecologist and evolutionary biologist. Dr. Stephen A. Smith provided continual guidance and support throughout my graduate studies and encouraged me to pursue my own projects and ideas—even when such ideas might not have been the most productive. Dr. Annette M. Ostling helped me to grow as an ecologist and helped to sharpen my analytical and computational skills, without which I could not have achieved the technical feats of my dissertation. Both have helped me to solidify my philosophy of science as both an ecologist and an evolutionary biologist—not an easy task for someone originally trained as a biochemist.

Although I was trained in the Ecology and Evolutionary Biology Department, I am thankful to the professors in the Department of Anthropology, especially Drs. Laura MacLatchy, Andy Marshall, and Milford Wolpoff for their courses and guidance, direct or indirect, that helped to broaden my perspectives and ensure that I should always be thinking of how my own research fits into the larger picture.

I am also grateful to Drs. Chris Dick and Dan Rabosky for engaging and insightful discussions on the nature and meaning of science, as well as Chris' near encyclopedic knowledge of all things related to tropical forest biology.

I thank Minh Chau Ho for helping me to better express the messaging of my first chapter; Ursula Trigos-Raczkowski for helping me through the darker times of graduate school; and Sarah Orth for allowing me to be a mentor on her academic journey.

Finally, I would like to thank the University of Michigan, and all the members of the department of Ecology and Evolutionary Biology, for their support and the opportunity to earn my doctoral degree.

LAND ACKNOWLEDGEMENT

"The ground on which we stand is sacred ground. It is the dust and blood of our ancestors."

—Chief Plenty Coups

The research presented in this dissertation tells a story of American plant evolution. As such, this research has benefitted from, and is inherently complicit in, ongoing colonialism and imperialism across the Americas.

The research itself took place on the Ann Arbor campus of the University of Michigan, located within Washtenaw County. From this alone, it should be clear that appropriation of

Indigenous names has taken place at both the state and local levels. Further, however, I would like to acknowledge that the University of Michigan resides on the traditional Territories of the Three Fire Peoples—the Ojibwe, Odawa and Bodéwadmi—as well as others. Therefore, as I have worked, lived, played, and completed my research on these territories, it has been imperative to keep in mind the ongoing effects of colonization, community struggles for self-determination, colonial state violence, and the recognition of Indigenous sovereignty.

Although land acknowledgements are an important component of recognizing the impacts of colonialism in scientific research (as well as our daily lives), they are not enough. Action is also needed. Below I have included a short list of links to local organizations helping to improve quality of life for Indigenous communities in an attempt to encourage action in addition to acknowledgement.

Native Justice Coalition

"The Native Justice Coalition (https://www.nativejustice.org/who-we-are) was formed in 2016 with the intent of being a platform for healing, social, and racial justice for all Native American people. Our goal is to provide a safe and nurturing platform for Native people based in an anti-oppression framework. We seek to collaborate first and foremost with tribal governments, Native American non-profits, and other Native American led community organizations. Our goal is to bring resources, initiatives, and programming into our tribal communities that are creative, engaging, and transformative."

Inter-Tribal Council of Michigan (ITCMI)

"The Inter-Tribal Council of Michigan, Inc. (ITCM; https://www.itcmi.org) is a 501©(3) non-profit corporation duly organized under a State Charter filed April 16, 1968…It represents twelve federally recognized tribes in Michigan. [Our] mission statement [is]: To act as a forum for member tribes; To advocate for member tribes in the development of programs and policies which will improve the economy, education, and quality of life for Michigan's Native Americans; and to provide technical assistance to member tribes, assisting in the development of tribal regulations, ordinances, and policies applicable to health and human services."

American Indian Health and Family Services

"American Indian Health & Family Services [https://www.aihfs.org/] is a non-profit health center whose mission is to empower and enhance the physical, spiritual, emotional, and mental wellbeing of American Indian/Alaska Native individuals, families and other underserved populations in SE MI through culturally grounded health and family services."

PREFACE

"All things are our relatives. What we do to everything we do to ourselves. All is really one."

—Black Elk

The dissertation research offered here represents work that took place primarily over the course of four years, from 2017-2021 in the United States. In the wake of a divisive 2016 presidential election, these years saw a substantial rise in race-related tensions across the US, including numerous murders of Black and Brown individuals at the hands of state police officers. The video documentation of murders, such as that of George Floyd, garnered nation-wide support for movements such as Black Lives Matter and others. As protestors took to the streets in an attempt to raise awareness of the systemic racial disparities that form a foundation of US society, their confrontations with police officers were often marked by disproportionately violent responses from the police. At the same time, political divides intensified as Democrats and Republicans appealed to vastly different ideologies among their respective political bases.

On top of this growing sense of discontent, beginning around 2020, the world was gripped by the COVID-19 pandemic, which caused global shutdowns for months and heightened the already significant political divides. The US pandemic response was generally inadequate at the federal, state, and often local levels. As a consequence, universities were shut down for a substantial time and daily work activities were generally moved to remote, online-only formats. For numerous graduate students, this resulted in considerable alterations to their projected time-to-completion and forced many universities to consider alternative funding arrangements to handle both the increased need for graduate support and the hiring freezes that spread across the country.

In early 2021, the rollout of COVID-19 vaccines and the change in administration following a second divisive presidential election offered some individuals hope that the past years might not represent "the new normal." Nevertheless, the 2021 change of administration was marred by a violent insurrection in which an angry mob breached the United States Capitol

Building. This incident has been named the most severe assault on the capitol since the 1814 burning by the British army. The COVID-19 vaccine rollout was also burdened with infrastructure issues leading to vaccine shortages in some areas and massive surpluses in others. These logistical issues occurred amidst a backdrop of mixed feelings concerning the vaccine—in which some US citizens even refuse to be vaccinated. Further, these domestic vaccine-related issues were waged in the background of second and third COVID-19 waves across the globe, notably in India, where the country's healthcare system was overrun by virus-related deaths and oxygen tank shortages, leading to massive increases in daily deaths.

This is the social and political background in which the research offered in this dissertation took place. Thus, while I and other researchers battled police officers, white supremacists, and neo-Nazis in the streets, ate tear gas, and were often arrested or beaten, our research continued. I feel this context is important because the research contained within this dissertation fundamentally attempts to shed light on the historical context of the flora of the Americas. As such, this work is intimately tied to this land and its history. Simultaneously, the social movements and activism taking place during the formation of this work were writing a new chapter of American history that would impact the social and political landscape to come. The title of this dissertation proclaims that "family matters." For biology, this statement represents our scientific acknowledgement that all life—all life—is fundamentally related through common descent—in other words, through our shared and divergent histories. This statement is no less true for our social lives than for our scientific ones: each of us is related both through shared ancestry and through shared and divergent history. I sincerely hope that the recognition of our familial bond that forms such a fundamental backbone of the biological sciences also makes its way into the social and political consciousness in the years to come.

Héctor Fox
May 3, 2021
Ann Arbor, MI

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

APPENDIX A

APPENDIX B

APPENDIX C

APPENDIX D

LIST OF TABLES

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

APPENDIX D

LIST OF APPENDICES

ABSTRACT

Even as early as Darwin, biologists recognized that closely related species were likely to be more similar to one another than to more distantly related organisms. Modern phylogenetic community ecology has extended this framework to suggest that closely related species should be similar in functional traits, occupied niche space, and even geographic distributions. The extent to which this type of phylogenetic niche conservatism has shaped modern biodiversity patterns remains a contentious issue, with extreme views arguing that niche conservatism is either the sole driver of modern distributions or that abiotic filtering alone is responsible for explaining where contemporary species are found. In this dissertation, I investigate the degree to which phylogenetic niche conservatism is observed among American angiosperms. Notably, I adopt the perspective that both niche conservatism and niche shift are important, scale-dependent factors contributing to modern biodiversity patterns, and attempt to analyze the phylogenetic scale over which we can observe niche conservatism for different aspects of these organisms' niches.

In **CHAPTER ONE**, I investigate the phylogenetic scale over which variation in a multi-trait resource-acquisition strategy is best observed. I show that coordinated evolution of height, growth form, seed mass, and wood density is best observed among species from con-ordinal families. Outside this phylogenetic scale, quantitative, clade-specific, variation in trait values among different angiosperm orders, as well as convergent evolution to similar trait values obscure the ability to detect phylogenetic niche conservatism.

In **CHAPTER TWO**, I develop novel niche characterization methods in order to quantify the degree of climatic and elevational niche similarity across the ranges of American seed plants at different phylogenetic scales. I find strong statistical support for phylogenetic niche conservatism as an important contributor to plant diversity patterns, corroborating the hypothesis that plants may tend to track favorable climate. In particular, I observed greater conservatism in precipitation niches and niche breadth compared to temperature or elevation, and niche

conservatism was greater at smaller phylogenetic scales. Nevertheless, even among closely related species, differences in climatic niche were significantly greater than expected by chance and variance in niche similarity among close relatives was large, such that within-clade variation equaled or exceeded between-clade differences.

In the remaining chapters, I focus on American alpine communities across different mountain systems. I demonstrate that contrasting patterns of phylogenetic diversity and degree of alpine specialization differentiate the alpine communities occurring at opposite ends of the American mountain range system, despite broad climatic similarities in these two regions, thus suggesting an important role for historical factors in the assembly of their floras (**CHAPTER THREE**). I also use continuous character niche reconstruction to show how alpine lineages arise primarily from montane species, indicating that a perspective which views niche conservatism and niche shift as occurring on a continuous spectrum may be a more productive framework for understanding how species enter novel niche space (**CHAPTER FOUR**).

I **CONCLUDE** by discussing (1) the generality of an appropriate phylogenetic scale for observing niche conservatism among angiosperm lineages; (2) the degree to which process such as abiotic filtering, biogeographic history, and ecological dynamics have shaped different assemblages; and (3) how the age of “big data” can be used and abused to investigate big picture questions of community assembly and evolutionary history. This dissertation underscores the complex interplay of niche conservatism, abiotic filtering, and historical processes in shaping modern angiosperm biodiversity.

INTRODUCTION

"…[T]he great force of history comes from the fact that we carry it within us…"

-James Baldwin, *The White Man's Guilt* (1965)

"[W]here there has been close descent in common, there will certainly be close resemblance or affinity."

-Charles Darwin, *On the Origin of Species (6th ed.)*

The concept of relatedness through common descent is now an accepted paradigm of modern biology. Perhaps made most famous in Charles Darwin's *On the Origin of Species* (1859), the idea that all life on the planet is related revolutionized the biological sciences and suggested many novel expectations for the traits (or characters) and distributions of organisms of varying degrees of relatedness. One such expectation is "the tendency for related species to resemble each other more than they resemble species" chosen at random (Blomberg & Garland, 2002, p. 905). The manifestation of this expected pattern is termed 'phylogenetic signal' (Blomberg et al., 2003; Blomberg & Garland, 2002; Losos, 2008; Münkemüller et al., 2012), and its prevalence (or absence) is frequently used to make inferences about evolutionary processes shaping community assembly.

Often employed in investigations of traits such as body mass (Blomberg et al., 2003), plant height (Ackerly, 2009), or metabolic rate (Capellini et al., 2010), quantitative metrics of phylogenetic signal, such as Blomberg's K (Blomberg et al., 2003) or Pagel's λ (Pagel, 1999), attempt to describe the amount of similarity between organisms that would be expected due to their known (and presumably accurate) phylogenetic relationships (Losos, 2008; Münkemüller et al., 2012). Such similarity is expected to arise because even genetic drift alone will result in a graded degree of similarity among organisms arrayed along a hierarchical tree-like structure, such as a phylogeny (Blomberg & Garland, 2002).

This perspective views phylogeny somewhat as a nuisance parameter. The effect of phylogeny must be 'subtracted out' or 'controlled for' in order to obtain insight on the 'real' processes of interest, such as community assembly dynamics, strength of selection, or rates of evolutionary change. Although imperfect, this framework has provided a productive conceptual context for addressing many biological questions (Losos, 2008). For instance, Felsenstein (1985) introduced the method of standardized independent contrasts to obtain an estimate of the rate of evolutionary change for a quantitative trait under the assumption of a Brownian Motion model of trait evolution. This method uses the expected amount of change based on branch lengths and tree topology to arrive at an estimate of the 'true' rate of character evolution. Similarly, the metrics of community phylogenetic structure introduced by Webb (2000) attempted to quantify how much ecological similarity exists within assemblages beyond what would be expected from their phylogenetic relationships alone.

Despite the clear value in being able to quantitatively account for the expected degree of similarity among related species in a community, it is not always the case that phylogeny must be viewed as a nuisance parameter. For instance, if biologists are curious about which lineages are present in particular communities, rather than factoring out phylogeny, they may need to explicitly incorporate the biogeographic history of specific lineages. For example, clades with centers of diversification in the tropics (e.g., Solanaceae originating in South America; Olmstead, 2013) are often restricted in whole or in part to such environments due to physiological constraints rather than dispersal limitation (Donoghue & Edwards, 2014; Hawkins et al., 2011; Judd et al., 1994; Segovia et al., 2020; Zanne et al., 2014). In such cases, the modern distributions of these clades therefore reflect the influence of both abiotic filtering (limiting their range extents due to physiological constraints) as well as historical processes (e.g., substantially greater diversity in their centers of diversification).

Apart from not always being a necessary component of comparative studies on community assembly, quantitative metrics of phylogenetic signal can also fail when biological realities do not meet their underlying assumptions (Freckleton & Harvey, 2006). For instance, a Brownian motion model of trait evolution (reviewed in Harmon, 2019, chapters 3-5) predicts an ever-increasing variance in trait values over evolutionary time. However, physiological limits can constrain the functional trait values species may possess—wood density, as an example, has well-defined upper and lower bounds (Chave et al., 2009). Even apart from physiological constraints, a Brownian motion model of trait evolution implicitly assumes that phylogenetic scale is unimportant (Harmon, 2019, chapter 6). For example, convergent evolution to similar trait values among distantly related lineages, which might only be observed at particular phylogenetic scales, can violate these model assumptions. Although some effort has been made to incorporate additional complexity into models of trait evolution, such as the OU model (Harmon et al., 2010; Pennell & Harmon, 2013), or multi-rate Brownian motion (O'Meara et al., 2006), these models still fail to identify the important biological processes that generate research questions in the first place and their additional parameters and model complexity can create identifiability issues in model selection (e.g., Harmon, 2019; O'Meara et al., 2006; Revell et al., 2008).

Most importantly, all current models of trait evolution can only incorporate single-valued, scalar traits (such as seed mass, height, or body size). However, some traits are best expressed as continuously valued distributions, such as the (proportioned) climatic and elevational gradients that occur along a species' range (**CHAPTER TWO**)[*]. In other cases, some traits, such as those relating to secondary metabolism and chemical defense, do not have pre-defined ordination[†], and therefore defining their relative similarity may require alternative distance metrics not compatible with current models of trait evolution (**APPENDIX A**). As a result, although standard metrics of phylogenetic signal have their value, their use may be inappropriate in some cases (**CHAPTERS ONE** and **TWO**), and a critical evaluation of model assumptions in relation to the phylogenetic scale of the study are important considerations.

The neutral expectation of phylogenetic signal, which states that even in the absence of any selective forces closely related organisms should be more similar, has led some researchers to use the alternative term 'phylogenetic niche conservatism' (or PNC) to refer to only that amount of similarity which is *beyond* what would be expected solely due to phylogenetic relationships (e.g., from Brownian motion; Losos, 2008; but see also Wiens et al., 2010 for a more general framework). The inference here being that PNC then reflects some 'active' process causing species to be more similar than expected from their degree of relatedness alone.

---

[*] For example, imagine that the range of *species A* is 20% montane and always below 500 mm mean annual precipitation (MAP), and *species B* has a range that is 15% montane and bound between 300- and 600-mm MAP. Assuming these are congeners, are these two species more or less similar in climatic niche than expected by chance, given their phylogenetic relationship?

[†] For instance, is benzene more similar to salicylic acid or benzyl acetate?

The resulting framework on PNC helped to greatly expand the field of community phylogenetics and drive the development of novel methodologies (Cavender-Bares et al., 2009; Webb, 2000; Webb et al., 2002). Quantitative metrics such as mean nearest taxon distance (MNTD) and mean pairwise distance (MPD) were developed by Webb (2000) to quantify the extent to which communities showed either phylogenetic clustering (i.e., species more closely related than expected by chance) or phylogenetic overdispersion (also called evenness; i.e., species were less closely related than expected by chance). By integrating these metrics of phylogenetic community structure with trait data thought to represent important aspects of species' niches (e.g., plant height or wood density), the field of community phylogenetics expanded to include questions of phylogenetic niche conservatism and the extent to which species distributions were shaped by niche conservatism or by niche shift (Dick & Pennington, 2019; Donoghue & Edwards, 2014).

Numerous studies have attempted to infer either competitive exclusion or abiotic filtering as primary drivers of community structure based on whether communities were phylogenetically overdispersed or clustered, respectively (Cavender-Bares et al., 2006; Webb et al., 2002). However, such metrics, and the subsequent inferences regarding underlying community assembly processes, were quickly shown to depend on both the phylogenetic and spatial scale of the observations (Cavender-Bares et al., 2009; Swenson et al., 2006). For example, competitive interactions driving character displacement might be best observed among more closely related organisms but would be harder to detect in a more taxonomically broad community sample, where the range of trait values might be broader and convergent evolution might lead to similar trait values shared between distantly related organisms (Cavender-Bares et al., 2009).

The debates surrounding how to integrate pattern-based data on phylogenetic signal, PNC, and community phylogenetic structure in order to infer community assembly processes remain contentious (e.g., Crisp et al., 2009; Donoghue & Edwards, 2014; Segovia et al., 2020). Nevertheless, this field of research has raised the important question of whether phylogenetic niche conservatism or niche shift shapes modern plant diversity patterns. Donoghue (2008) suggested that plants may track favorable climate more easily than they would evolve tolerances to new climatic conditions (i.e., 'it is easier to move than to evolve,' p. 11549). This and subsequent prominent research (reviewed in Wiens et al., 2010) demonstrated that niche conservatism is a strong determinant of plant diversity patterns.

Despite the clear prevalence of niche conservatism, however, it is also true that plants do radiate into novel niche space and biome-switching does occur (e.g., Donoghue & Edwards, 2014). Switches from tropical to temperate environments have taken place (Judd et al., 1994), and temperate clades are often nested within tropical ones (Hawkins et al., 2011; Kerkhoff et al., 2014). In principle, there is no reason why modern distributions could not have been shaped by both niche conservatism and niche shift. It seems plausible that both processes have contributed to differing extents in different lineages throughout the course of their evolution (**CHAPTER TWO**).

This dissertation will explore the degree to which PNC is observed for various aspects of seed plant niches, with a focus on American angiosperms. At each stage, multiple spatial and phylogenetic scales will be incorporated into the analyses, as scale-dependence is a ubiquitous feature of ecological studies (Swenson et al., 2006). In particular, I have adopted the framework that *both* niche conservatism and niche shift are important processes contributing to modern biodiversity patterns, and that the degree to which PNC is observed will depend not only on the niche axes examined but also the spatial and phylogenetic scale of the observations.

For this work, I have chosen to focus on the Americas as the primary study area. Apart from being a well-studied region with ease of access to trait data and field sites, the Americas are well-suited to studies on PNC because of their unique geography and biogeographic history. Unlike the mountain ranges of Europe and the Himalayas, which have a more East-West orientation, American mountain systems tend to run North-South. This orientation is less likely to impede dispersal of plants and animals during glacial cycles, which has resulted in decreased extinction of tree species in the Americas compared to Europe during the Quaternary period (Bennett et al., 1991). Additionally, the nearly unbroken line of high-elevation habitat running from one pole to another provides a potential cooler climate corridor for plant dispersal along the American mountain system (Antonelli et al., 2009). Thus, overall, we would expect American plants to have had greater opportunity to disperse to and track favorable climate during glacial cycles and their present-day climatic niches and distributions might therefore be less impacted by extinction due to glaciations and habitat loss. This, in turn, could allow plants to have retained climatic and elevational niches to a greater extent than in Europe or parts of Asia (i.e., there has been greater opportunity for PNC in the Americas).

In this dissertation, I have tended toward a macroecological approach, choosing to synthesize data across the entire Americas rather than focusing on specific communities or sites within this study region. In many ways such choices at the level of the individual scientist are arbitrary, as there is value in both macro- and micro-ecological approaches. And, indeed, I have contributed to work examining niche partitioning among functional and chemical defense traits in smaller scale Neotropical forest communities (see **APPENDIX A**). Nevertheless, my particular research questions are best approached from a larger spatial scale. For instance, in **CHAPTER ONE** I ask how a coordinated suite of functional traits varies along latitude for different angiosperm clades. In **CHAPTER TWO**, I investigate the degree of climatic and elevational niche similarity across the entire ranges of seed plant species at different phylogenetic scales. And in **CHAPTER THREE**, I ask how phylogenetic diversity and alpine specialization vary among different American mountain ranges. Such questions require a more macroecological framework and larger spatial scales. This approach allows me to ask questions regarding what the "regional species pool" is for seed plants in different American ecosystems (Cornell & Harrison, 2014), and I have previously shown that incorporating knowledge of the regional scale community dynamics (**APPENDIX B**) or phylogenetic structure (**APPENDIX C**) into studies of local scale ecological systems can significantly alter even our neutral expectations of community structure.

Overall, in this dissertation I will examine key aspects of how phylogenetic niche conservatism has helped to shape the modern distributions of American seed plants, especially angiosperms. In **CHAPTER ONE**, I investigate the phylogenetic scale over which observable variation in functional trait syndromes occurs. This provides quantitative insight into how evolutionary distance relates to divergence in resource-acquisition traits and, therefore, the strength of PNC for these traits across the Americas. In **CHAPTER TWO**, I provide a quantitative assessment of the degree of climatic and elevational niche similarity across the ranges of all American seed plants at multiple phylogenetic scales. In particular, I highlight statistically significant differences in niche space occupied by plants at different phylogenetic scales, providing a comprehensive and robust assessment of the degree of climatic PNC across the Americas. In **CHAPTERS THREE** and **FOUR**, I focus on alpine species (whose niches are thus restricted in elevation). For **CHAPTER THREE**, I illustrate that contrasting patterns of phylogenetic diversity and alpine specialization imply different community assembly processes in otherwise climatically similar near-polar regions, suggesting that abiotic filtering alone cannot explain these diversity patterns (i.e., a historical perspective is also needed). Finally, in **CHAPTER FOUR**, I use continuous character reconstructions to show how alpine lineages arise predominantly from montane species, suggesting an important role for both PNC and historical factors in shaping the modern American flora at high elevations, and indicating that a perspective which views niche conservatism and niche shift as occurring on a continuous spectrum may be a more productive framework for understanding how species enter novel niche space.

REFERENCES

Ackerly, D. (2009). Conservatism and diversification of plant functional traits: Evolutionary rates versus phylogenetic signal. *Proceedings of the National Academy of Sciences*, *106*(Supplement_2), 19699–19706. https://doi.org/10.1073/pnas.0901635106

Antonelli, A., Nylander, J. A. A., Persson, C., & Sanmartín, I. (2009). Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proceedings of the National Academy of Sciences*, *106*(24), 9749–9754. https://doi.org/10.1073/pnas.0811421106

Bennett, K. D., Tzedakis, P. C., & Willis, K. J. (1991). Quaternary Refugia of North European Trees. *Journal of Biogeography*, *18*(1), 103–115. https://doi.org/10.2307/2845248

Blomberg, S., & Garland, T. (2002). Tempo and mode in evolution: Phylogenetic inertia, adaptation and comparative methods. *Journal of Evolutionary Biology*, *15*(6), 899–910. https://doi.org/10.1046/j.1420-9101.2002.00472.x

Blomberg, S., Garland, T., & Ives, A. (2003). Testing for Phylogenetic Signal in Comparative Data: Behavioral Traits Are More Labile. *Evolution*, *57*(4), 717–745.

Capellini, I., Venditti, C., & Barton, R. (2010). Phylogeny and metabolic scaling in mammals. *Ecology*, *91*(9), 2783–2793.

Cavender-Bares, J., Keen, A., & Miles, B. (2006). Phylogenetic structure of Floridian plant communities depends on taxonomic and spatial scale. *Ecology*, *87*(sp7), S109–S122. https://doi.org/10.1890/0012-9658(2006)87[109:PSOFPC]2.0.CO;2

Cavender-Bares, J., Kozak, K. H., Fine, P. V. A., & Kembel, S. W. (2009). The merging of community ecology and phylogenetic biology. *Ecology Letters*, *12*(7), 693–715. https://doi.org/10.1111/j.1461-0248.2009.01314.x

Chave, J., Coomes, D., Jansen, S., Lewis, S., Swenson, N., & Zanne, A. (2009). Towards a worldwide wood economics spectrum. *Ecology Letters*, *12*(4), 351–366. https://doi.org/10.1111/j.1461-0248.2009.01285.x

Cornell, H. V., & Harrison, S. P. (2014). What Are Species Pools and When Are They Important? *Annual Review of Ecology, Evolution, and Systematics*, *45*(1), 45–67. https://doi.org/10.1146/annurev-ecolsys-120213-091759

Crisp, M. D., Arroyo, M. T. K., Cook, L. G., Gandolfo, M. A., Jordan, G. J., McGlone, M. S., Weston, P. H., Westoby, M., Wilf, P., & Linder, H. P. (2009). Phylogenetic biome conservatism on a global scale. *Nature*, *458*(7239), 754–756. https://doi.org/10.1038/nature07764

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. John Murray.

Dick, C. W., & Pennington, R. T. (2019). History and Geography of Neotropical Tree Diversity. *Annual Review of Ecology, Evolution, and Systematics*, *50*(1), 279–301. https://doi.org/10.1146/annurev-ecolsys-110617-062314

Donoghue, M. J. (2008). A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences*, *105*(Supplement 1), 11549–11555. https://doi.org/10.1073/pnas.0801962105

Donoghue, M. J., & Edwards, E. J. (2014). Biome Shifts and Niche Evolution in Plants. *Annual Review of Ecology, Evolution, and Systematics*, *45*(1), 547–572. https://doi.org/10.1146/annurev-ecolsys-120213-091905

Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist*, *125*(1), 1–15. https://doi.org/10.1086/284325

Freckleton, R. P., & Harvey, P. H. (2006). Detecting Non-Brownian Trait Evolution in Adaptive Radiations. *PLoS Biology*, *4*(11), e373. https://doi.org/10.1371/journal.pbio.0040373

Harmon, L. (2019). *Phylogenetic Comparative Methods—Learning from trees.*

Harmon, L. J., Losos, J. B., Jonathan Davies, T., Gillespie, R. G., Gittleman, J. L., Bryan Jennings, W., Kozak, K. H., McPeek, M. A., Moreno-Roark, F., Near, T. J., Purvis, A., Ricklefs, R. E., Schluter, D., Schulte II, J. A., Seehausen, O., Sidlauskas, B. L., Torres-Carvajal, O., Weir, J. T.,

& Mooers, A. Ø. (2010). Early bursts of body size and shape evolution are rare in comparative data. *Evolution*, *64*(8), 2385–2396. https://doi.org/10.1111/j.1558-5646.2010.01025.x

Hawkins, B. A., Rodríguez, M. Á., & Weller, S. G. (2011). Global angiosperm family richness revisited: Linking ecology and evolution to climate. *Journal of Biogeography*, *38*(7), 1253–1266. https://doi.org/10.1111/j.1365-2699.2011.02490.x

Judd, W. S., Sanders, R., & Donoghue, M. J. (1994). Angiosperm family pairs: Preliminary phylogenetic analyses. *Harvard Papers in Botany*, *1*(5), 1–51.

Kerkhoff, A. J., Moriarty, P. E., & Weiser, M. D. (2014). The latitudinal species richness gradient in New World woody angiosperms is consistent with the tropical conservatism hypothesis. *Proceedings of the National Academy of Sciences*, *111*(22), 8125–8130. https://doi.org/10.1073/pnas.1308932111

Losos, J. B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, *11*(10), 995–1003. https://doi.org/10.1111/j.1461-0248.2008.01229.x

Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K., & Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, *3*(4), 743–756. https://doi.org/10.1111/j.2041-210X.2012.00196.x

Olmstead, R. G. (2013). Phylogeny and biogeography in Solanaceae, Verbenaceae and Bignoniaceae: A comparison of continental and intercontinental diversification patterns. *Botanical Journal of the Linnean Society*, *171*(1), 80–102. https://doi.org/10.1111/j.1095-8339.2012.01306.x

O'Meara, B. C., Cecile, A., Sanderson, M. J., & Wainwright, P. (2006). Testing for different rates of continuous trait evolution using likelihood. *Evolution*, *60*(5), 922–933.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, *401*(6756), 877–884. https://doi.org/10.1038/44766

Pennell, M. W., & Harmon, L. J. (2013). An integrative view of phylogenetic comparative methods: Connections to population genetics, community ecology, and paleobiology. *Annals of the New York Academy of Sciences*, *1289*(1), 90–105. https://doi.org/10.1111/nyas.12157

Revell, L. J., Harmon, L. J., & Collar, D. C. (2008). Phylogenetic Signal, Evolutionary Process, and Rate. *Systematic Biology*, *57*(4), 591–601. https://doi.org/10.1080/10635150802302427

Segovia, R. A., Pennington, R. T., Baker, T. R., Souza, F. C. de, Neves, D. M., Davis, C. C., Armesto, J. J., Olivera-Filho, A. T., & Dexter, K. G. (2020). Freezing and water availability structure the evolutionary diversity of trees across the Americas. *Science Advances*, *6*(19), eaaz5373. https://doi.org/10.1126/sciadv.aaz5373

Swenson, N. G., Enquist, B. J., Pither, J., Thompson, J., & Zimmerman, J. K. (2006). The problem and promise of scale dependency in community phylogenetics. *Ecology*, *87*(10), 2418–2424. https://doi.org/10.1890/0012-9658(2006)87[2418:TPAPOS]2.0.CO;2

Webb, C. O. (2000). Exploring the Phylogenetic Structure of Ecological Communities: An Example for Rain Forest Trees. *The American Naturalist*, *156*(2), 145–155. https://doi.org/10.1086/303378

Webb, C. O., Ackerly, D. D., McPeek, M. A., & Donoghue, M. J. (2002). Phylogenies and Community Ecology. *Annual Review of Ecology and Systematics*, *33*(1), 475–505. https://doi.org/10.1146/annurev.ecolsys.33.010802.150448

Wiens, J. J., Ackerly, D. D., Allen, A. P., Anacker, B. L., Buckley, L. B., Cornell, H. V., Damschen, E. I., Jonathan Davies, T., Grytnes, J.-A., Harrison, S. P., Hawkins, B. A., Holt, R. D., McCain, C. M., & Stephens, P. R. (2010). Niche conservatism as an emerging principle in ecology and conservation biology: Niche conservatism, ecology, and conservation. *Ecology Letters*, *13*(10), 1310–1324. https://doi.org/10.1111/j.1461-0248.2010.01515.x

Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. A., FitzJohn, R. G., McGlinn, D. J., O'Meara, B. C., Moles, A. T., Reich, P. B., Royer, D. L., Soltis, D. E., Stevens, P. F., Westoby, M., Wright, I. J., Aarssen, L., Bertin, R. I., Calaminus, A., Govaerts, R., … Beaulieu, J. M. (2014). Three keys to the radiation of angiosperms into freezing environments. *Nature*, *506*(7486), 89–92. https://doi.org/10.1038/nature12872

Chapter I

A Targeted Phylogenetic Approach Helps Explain New World Functional Diversity Patterns Of Two Eudicot Lineages[‡]

CHAPTER IN CONTEXT

In this study, I integrate data from multiple life-history traits, digitized occurrence records, and phylogenetic information to demonstrate a phylogenetically conserved and latitudinally graded resource-acquisition trait syndrome among two major eudicot lineages, Ericales (Asterids) and Fabales (Rosids). The functional traits involved—height, seed mass, wood density, and growth form—showed evidence of coordinated evolution, strengthening the argument that they jointly represent an underlying ecological strategy. Importantly, although both clades showed *qualitatively* similar patterns in the manifestation of this trait syndrome, *quantitative* differences in the trait values obtained by each clade were observed. For example, both Ericales and Fabales showed decreasing seed mass with increasing (absolute) latitude. However, for the same latitude, the two clades did not have the same distribution of seed mass values.

This work enhances our understanding of angiosperm biogeography in two primary ways. First, this study is one of the few (as of its 2021 publication date) to examine multiple functional traits simultaneously. The traits chosen here were not random, but rather represent traits strongly believed to be tied to ecological strategies impacting species distributions and community assembly and dynamics. By examining these traits together, I was able to provide insight into the underlying resource-acquisition strategy and conclude that habitats with relatively greater soil

---

[‡] This chapter was originally published in the *Journal of Biogeography* (2021; 48:202-215).

nutrient content and a shorter growing season favor angiosperm species with shorter stature, lower seed mass and wood density, and a shrubby or herbaceous growth form.

Secondly, this study places the results in a broader phylogenetic context, which is often lacking from research on the distributions of functional traits. This aspect is crucial as it demonstrates the phylogenetic scale over which variation in the trait syndrome can best be observed and therefore sheds light on the relevance and extent of phylogenetic niche conservatism for this study system. This is accomplished by assessing trait similarity across the entire phylogeny of each clade, not just for species which co-occur at a particular site. For example, in this dataset, congeners tended to have similar functional trait values and achieve similarly extreme poleward distributions. This observation makes a strong case for phylogenetic niche conservatism as a driver of where species are to be found. At the same time, however, convergent evolution to similar trait values can occur between distantly related groups, such as has occurred for *Primula* (Primulaceae; Ericales) and *Astragalus* (Fabaceae; Fabales). Therefore, without the broader phylogenetic context of how functional traits are distributed within Ericales and Fabales, analysis of a community composed of both *Primula* and *Astragalus* might not suggest that phylogenetic niche conservatism was an important community assembly component.

The degree of niche conservatism observed in this dataset suggested that, for this particular trait syndrome, variation was best observed among species from con-ordinal families. Species within the same genus tended to be quite similar in their functional trait values[§], whereas species from different taxonomic orders could converge to similar trait values, thus obscuring the relevance of phylogenetic niche conservatism. This insight into the phylogenetic scale over which trait variation is best observed must be related, at least in part, to the “evolvability” or “lability” of the traits involved. Such questions of trait lability—at least for traits strongly related to ecological function—are at the heart of the question of whether it is “easier to move or to evolve,” and thus to the question of how the process of phylogenetic niche conservatism helps to shape contemporary angiosperm distributions.

[§] This result will contrast with the findings for climatic niche similarity detailed in **CHAPTER TWO**, in which congeners often demonstrated much higher variability in niche similarity.

ABSTRACT

*Aim***:** Large-scale functional diversity studies typically examine isolated traits, often without phylogenetic context. Here, we integrate data from five life-history traits with phylogeny and occurrence records to assess: (1) correlated latitudinal gradients of trait combinations; (2) which traits show phylogenetic conservatism; and (3) quantitative, clade-specific differences in trait syndromes, illustrating the phylogenetic scale of observable variation in ecological strategies.

*Location & Taxon***:** The Americas; Ericales (Asterids) and Fabales (Rosids).

*Methods***:** We used publicly available trait data sets on height, seed mass, wood density, leaf mass per area (LMA), and growth form, an open-source phylogeny, and georeferenced occurrence records to investigate functional diversity patterns. We employed phylogenetic generalized least squares and phylogenetic principal components analyses (pPCA) to assess correlated trait evolution and quantify the trait syndrome, respectively. We employed the InfoMap Ecoregions web app to cluster species by bioregions. We used standard statistical tests and randomization simulations to assess statistical significance of results.

*Results***:** Ericales and Fabales exhibited a biogeographically consistent, phylogenetically conserved trait syndrome. Moving toward both poles, species exhibited progressively smaller trait values and more herbaceous and shrubby growth forms (except for LMA, which showed no consistent pattern). We quantified latitudinal variation in this trait syndrome using pPCA and provide evidence for correlated trait evolution.

*Main conclusions***:** We demonstrate a functional trait syndrome involving height, seed mass, wood density, and growth form, but not LMA. Functional trait values showed consistent latitudinal patterns and evidence of correlated evolution, suggesting an underlying ecological strategy. Further, the two clades showed quantitative differences in the manifestation of this trait syndrome. Variation in the syndrome was best observed among species from con-ordinal families. We interpret this trait syndrome as a strategy of resource acquisition in which habitats

with relatively greater soil nutrient content and a shorter growing season favor shorter stature, lower seed mass and wood density, and shrubby or herbaceous growth form.

INTRODUCTION

Functional traits mediate numerous aspects of plant ecology (Grime, 2006), and there has been extensive interest in categorizing the distributional patterns of functional trait values. Such patterns represent the confluence of differing ecological strategies and interactions, selection pressures, physiological constraints, abiotic filtering, and historical contingency (Reich et al., 2003). Nevertheless, understanding the underlying biological reasons for the observed patterns remains elusive. Strong latitudinal gradients in many life-history traits, sometimes assessed across all angiosperms or seed plants, have given researchers a first glimpse into continental-scale patterns of functional diversity (Moles et al., 2005; Moles et al., 2007, 2009; Swenson et al., 2012), which might provide crucial insight into the integrated ecological strategies employed by particular taxa (Grime, 2006).

Despite these advances, important gaps in our syntheses of these trait patterns remain (Moles, 2018). Large-scale analyses of functional diversity often look at only a single trait (Moles et al., 2005, 2007, 2009), or include phylogeny in only a limited context (Swenson et al., 2012; Swenson & Weiser, 2010). However, ecological strategies might manifest in the combination of traits a plant exhibits (Grime, 2006; Pierce et al., 2017), and functional trait values will vary among different clades, and can be influenced by evolutionary history (e.g. phylogenetic constraint). For example, although trees may be taller on average in the tropics, what constitutes tall may vary among different lineages. Thus, one important and outstanding question is the extent to which traditional functional traits show evidence of phylogenetic signal/ conservatism (Blomberg et al., 2003; Wiens et al., 2010). Indeed, phylogenetic niche conservatism, in general, has been upheld as the primary driver of present-day distributional patterns among New World trees (Segovia et al., 2020). Understanding the extent to which the functional trait values which contribute to a plant's niche are conserved, therefore, has significant implications for understanding modern biodiversity.

Another important consideration is the extent to which certain functional traits correlate, or group together, potentially indicating an integrated strategy. In part, this is a practical concern, since if two traits are tightly correlated, but only one of the pair is easily measured, then it may suffice to measure only that trait for which data acquisition is simpler. However, biologically, it might also be the case that a plant's ecological strategy is better understood in the context of multiple traits and that phylogenetic niche conservatism might manifest as coordinated evolution among multiple characters. Underpinning both the question of phylogenetic signal and coordinated evolution of traits is the question of appropriate evolutionary scale: at what phylogenetic scale is variation in ecological strategies observed?

To begin addressing these questions, we examined five life-history traits that relate to ecological strategies and are linked to plant and ecosystem function: height, seed mass, wood density, leaf mass per area (LMA), and growth form. Plant height is important for its relationship to carbon sequestration (Conti & Díaz, 2013; Moles et al., 2009) and competition for light (Moles & Leishman, 2008), as well as influencing habitat type available for animal species (e.g. closed canopy forest vs. open woodland). Seed mass is one determinant of dispersal syndrome (Muller-Landau et al., 2008; Westoby et al., 2002) and contributes to fecundity-survival (Muller-Landau, 2010) and shade tolerance trade-offs (Wright et al., 2010). Wood density provides mechanical support and forms a 'wood economics spectrum' (Chave et al., 2009). Leaf traits relate directly to photosynthetic strategy and the leaf economics spectrum (Wright et al., 2004; Wright et al., 2010). Leaf mass per area (LMA) represents a fundamental part of a plant's carbon-investment strategy and has been correlated with leaf lifespan in woody plants.

While previous work on continental-scale diversity patterns (Moles et al., 2007, 2009; Swenson et al., 2012; Swenson & Weiser, 2010) has found the signal of latitudinal gradients for several traits, the inclusion of phylogeny has been limited. On the other hand, studies that have related functional trait patterns to phylogeny (Moles et al., 2005; Zanne et al., 2014) have utilized very broad phylogenetic scales (the entire seed plant or angiosperm phylogeny; but see Hawkins et al., 2014). These extremely broad studies may obscure the underlying biological explanations for observed patterns. Here, we adopt a targeted approach that allows us to account for individual histories of particular lineages and provides a clearer understanding of empirical patterns.

We examine the geographic distribution of this suite of life-history traits across the Americas within two angiosperm clades: Ericales (Asterids) and Fabales (Rosids). Ericales is a large clade within the Asterids, containing 22 taxonomic families and comprised of ~12,000 known species (Stevens, 2001; APG IV, 2016). Although resolution of deep divergences remains contentious, the clade likely began diversifying between 100-125 million years ago (MYA; cf. Magallón et al., 2019; Rose et al., 2018; Stevens, 2001), providing not only ample opportunity for phenotypic diversification, but also for species to be exposed to several changes in habitat types and climate regimes. Similarly, Fabales represents a ~80-100 MYA clade (Li et al., 2015; Stevens, 2001; Wikström et al., 2001) that has diversified into numerous ecosystems across the globe. Additionally, Fabales are of particular note for their involvement with nitrogen-fixing bacteria (Sprent, 2009), of which the Fabaceae contributes greatly to the global nitrogen cycle.

Both Ericales and Fabales are present throughout the Americas and contribute to the structure and diversity of a wide range of ecosystems, including tropical rainforests, forest understories, open habitats, and polar environments (Hedwall et al., 2013; Moquet et al., 2017; Stevens, 2001; ter Steege et al., 2006). Therefore, we expected that each of these two clades would show sufficient variation in functional trait values to address the question of which traits show phylogenetic conservatism and evidence of correlated evolution. Nevertheless, because these clades are relatively distantly related (separated by over 100 MY), we expected clade-specific quantitative differences in the trait values characterizing each clade, even if overall qualitative phenotypic patterns were similar.

METHODS

Phylogenies & species lists.

We obtained species-level Ericales and Fabales phylogenies from the Smith and Brown (2018) seed plant phylogeny, and left polytomies unresolved. Species lists from this phylogeny were used to query trait and geographic occurrence record databases (see below). To investigate the interaction between traits and biogeographic distribution, we pruned each larger phylogeny (Ericales or Fabales) based on trait-data availability to create subtrees for each trait, yielding the following species-by-trait coverage after data cleaning and processing (percent indicates fraction of species for which trait data was available) . Growth form: 2362 (Ericales, 14.9%), 4052 (Fabales, 16.5%). Plant height: 558 (Ericales, 3.5%), 593 (Fabales, 2.4%). Seed mass: 412 (Ericales, 2.6%), 1014 (Fabales, 4.1%). Wood density: 314 (Ericales, 1.9%), 367 (Fabales, 2.4%). LMA: 451 (Ericales, 2.8%), 625 (Fabales, 2.6%).

We illustrated the correspondence between species' trait values and their farthest latitudinal extent from the equator (see *Occurrence Records* below) with phylogenetic heatmaps for which we painted ancestral state reconstructions on branches and latitudinal categories surrounding each phylogeny. Phylogenies were visualized using the R statistical computing environment (R Development Core Team. 2011) with 'GGTREE' (Yuet al., 2017) and 'TREEIO' (Wang et al., 2020) packages; reconstructions were calculated with 'PHYTOOLS' (Revell, 2012).

To determine phylogenetic clustering (niche conservatism) of species within latitudinal categories, we calculated the average phylogenetic distance among species found in each category. To assess whether these average distances were more extreme than expected by chance, we performed 1000 random draws of *n* taxa from the phylogeny and calculated the average distance among them for each draw (where *n* equals the number of species in a category). Observed phylogenetic distances were deemed significant if they were more extreme than 95% of distances obtained from these simulations.

We used phylogenetic principal component analyses (pPCA) to quantify our trait syndrome, calculated with 'PHYTOOLS' and visualized with 'PCA3D' (Weiner, 2015). pPCA accounts for the covariance structure of the phylogeny during ordination (Polly et al., 2013), allowing one to measure the degree of correlation among variables beyond what is expected from Brownian motion. We performed pPCA on the subsets of taxa for which complete trait data was available (86 Ericales, 220 Fabales species), though we did not include growth form (categorical trait) in the ordination. For these subsets, correlated evolution among traits was assessed by phylogenetic generalized least squares (PGLS), using the variance-covariance matrix of the subtree as the expected correlation structure, in order to quantify the degree of correlation between traits beyond what would be expected solely from a Brownian motion model of trait evolution.

Occurrence records.

Ericales and Fabales georeferenced occurrence records across all American countries were obtained from the Global Biodiversity Information Facility (GBIF) and processed using the 'COORDINATECLEANER' package (Zizka et al., 2019) in R to remove suspect coordinates, fossil records, and match taxa to country. Records were further processed by removing those with fewer than two decimal precision (~1.1 km) and matching all names to those from the Smith and Brown (2018) phylogeny. From these records, each species was assigned a unique latitudinal category based on the farthest *absolute* distance it was found from the equator. We required at least three records within a category in order to assign a species to it. Categories were as follows: beyond 60° (polar); between 60° and 50° (hemi-boreal); between 50° and 40° (temperate); between 40° and 20° (subtropical); between 20° and 0° (tropical). Results were robust to changes in the boundaries of these latitudinal categories (data not shown).

Latitudinal gradients were visualized with 'GGPLOT2' (Wickham, 2016) in R. Trait disparity maps were constructed as follows. For each clade and trait, we calculated standardized effect size as $SES = {(X - \bar{X})}/{\sigma_x}$, where X denotes species' trait value, $\bar{X}$ the mean clade-wide trait value (**Figure 1.S1**), and σ_x is the standard deviation of the trait distribution. We calculated the average SES found within a 0.5° grid cell in order to visualize 'hot spots' where trait values differ significantly from the overall clade average. Positive (negative) values indicate grid cells with average trait values much larger (smaller) than the clade-wide average. Scores are in units of standard deviations and can be interpreted as a Z-score. American terrain maps were obtained from stamen maps (Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under CC BY SA) via the 'GGMAP' API (Kahle & Wickham, 2013).

Trait and climate data.

Trait data were obtained from the TRY Plant Trait Database (Kattge et al., 2020), by querying all traits related to plant height, seed mass, wood density, LMA, and growth form for our Ericales and Fabales species lists. Trait data were converted to standard units (seed mass—mg; height—m; wood density—g/cm^3; LMA—mg/mm^2; growth form—unitless, see below), entries flagged with risk factor >4 by TRY removed, and remaining trait values averaged for each species. We log-transformed all traits, except growth form, prior to analyses. Growth form, instead, was coded numerically as: 0—herbaceous, 0.5—shrub/ climber; 1—tree; *NA*—other; however, we did not use such values as ordinated points.

We downloaded 30-sec climate data from worldclim.org (Fick & Hijmans, 2017) describing mean annual temperature (MAT), precipitation (MAP), and elevation (ELV) of the Americas. We then extracted MAT, MAP, and ELV values for each species' location using the 'RASTER' package in R, and calculated the minimum MAT, MAP, and maximum ELV for each species across all of its sites.

Bioregion mapping and trait distributions.

For those species with complete functional trait information, we used their georeferenced occurrence records to cluster species by biogeographical regions (bioregions), using the Infomap Bioregions web application (https://bioregions.mapequation.org). Briefly, this algorithm takes species' point occurrences as input and uses an information-theoretic clustering algorithm, Infomap (Edler et al., 2016; Rosvall & Bergstrom, 2008), to form a bipartite network mapping species to different bioregions. Rather than binning species by latitudinal extrema, this approach allowed the entirety of species' ranges to define potentially biologically relevant geographical boundaries that might separate groups of species using different ecological strategies. Parameters used for clustering each clade by bioregion are given in **Table 1.S1**. To examine the relationship between functional traits and the bioregions identified by the Infomap algorithm, we calculated mean and standard error of functional trait values for all species assigned to each bioregion.

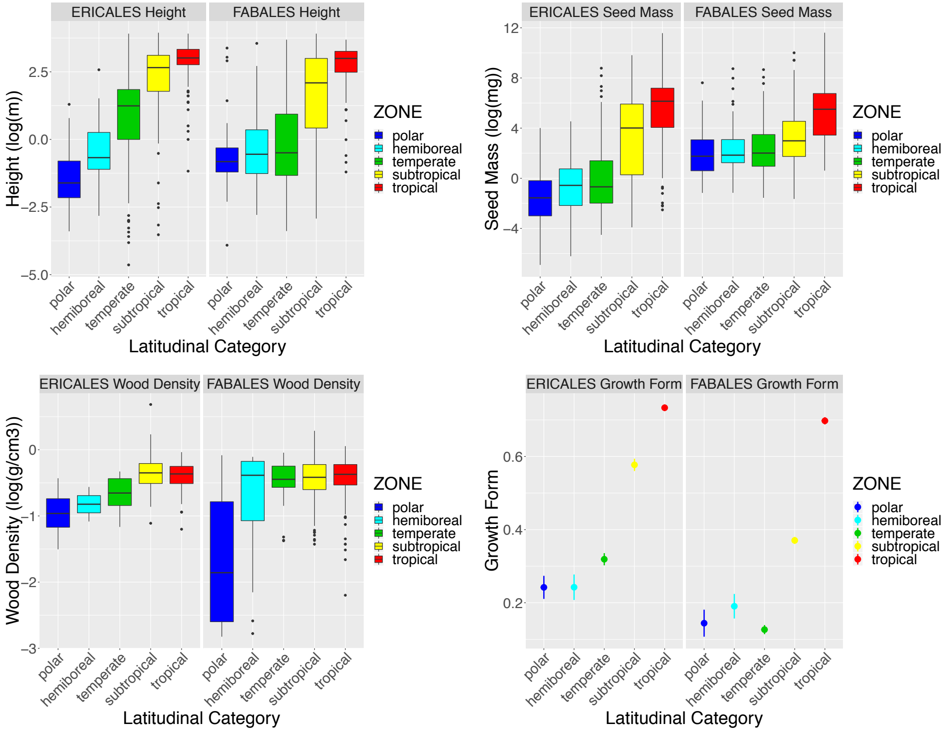


Figure 1.1. Latitudinal gradients for each clade (Ericales, *left panels*; Fabales, *right panels*) and functional trait (height, seed mass, wood density, growth form). For both clades, taxa confined to tropical latitudes show significantly higher trait values than those which reach polar latitudes ($p<0.05$ in all cases), and there is a tendency for a graded increase in trait values with decreasing distance from the equator. This trend is interpreted as a trait syndrome since it manifests across multiple traits. We omit LMA here because it did not show a strong latitudinal gradient (see **Figure 1.S2** for LMA results), though tropical taxa still possess higher median LMA values than polar ones.

Differences between Fabales and Ericales are discernible, indicating quantitative differences in the manifestation of this trait syndrome (see Discussion). *(Boxplots indicate median values. Growth form is plotted using average values (points) and standard deviations (lines) because this trait was treated differently than others in the analyses (see Methods).)*

RESULTS

Correspondence between geography & traits.

From an initial set of 1,854,067 GBIF records, data cleaning methods yielded a final set of 1,170,770 records (~63% retained). Among these, we observed a strong relationship between species' latitudinal category and most life-history trait values (**Figure 1.1**), LMA being an exception (**Figure 1.S2**). Both Ericales and Fables occurring at more extreme latitudes (particularly beyond 50° from the equator) exhibited a high-latitude trait syndrome characterized by the production of smaller seeds, shorter stature, lower wood density, being herbaceous or shrubby, and lower LMA. The converse was generally true for taxa confined to within 20° of the equator. Production of larger seeds, higher wood density and stature, and tree growth form typified the lower-latitude manifestation of this trait syndrome.

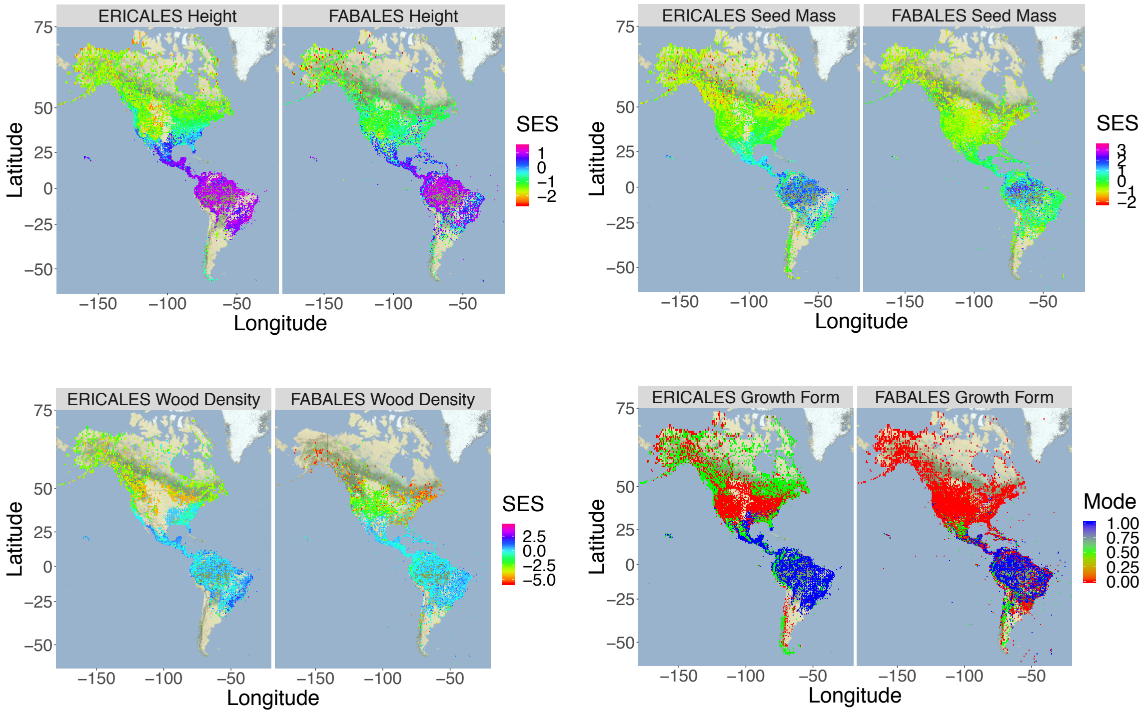


Figure 1.2. Geographic heatmaps showing the extent of trait disparity for each clade (Ericales, *left panels*; Fabales, *right panels*) and functional trait (height, seed mass, wood density, growth form). LMA is omitted because it did not show a strong latitudinal

gradient (see **Figure 1.S2**). Except for growth form, grid cell colors indicate the extent to which the average trait value in that grid cell differs from the overall average across the entire clade (standard deviation units, e.g., z-score; see Methods), with red (purple) grid cells indicating values much lower (higher) than the average. For growth form, grid cells are colored by the most common (modal) growth habit in that cell (blue: tree; green: shrub; red: herbaceous). In general, South American tropics have trait values higher than the average and are dominated by trees (though many more Fabales are herbs here), while polar latitudes have trait values lower than the average and a greater proportion of herbs (though many more Ericales are shrubs here). Maps for the same trait share a common legend, however, maps of different traits do not and so colors are not directly comparable between different traits.

Quantitatively, we found a latitudinal gradient in trait values for each trait and clade (**Figure 1.1**), with the exception of LMA (**Figure 1.S2**), which showed a comparatively modest trend. Although both clades showed a graded tendency toward lower trait values at progressively higher latitudes, the quantitative trait values within each latitudinal category differed between Fabales and Ericales. For example, for both groups, species at polar latitudes were significantly shorter than those at tropical latitudes ($p<0.05$); however, at polar latitudes Fabales species were significantly taller than Ericales species ($p<0.05$; see Discussion). We also note that Fabales tended to produce larger seeds, on average, than Ericales (**Figure 1.S1**). Further, there was a strong drop in median Fabales wood density at polar latitudes, which was not present for Ericales. To some extent, these clade-specific differences might be attributable to differences in how Fabales and Ericales are distributed with regards to temperature, precipitation, and elevation (**Figure 1.S3**).

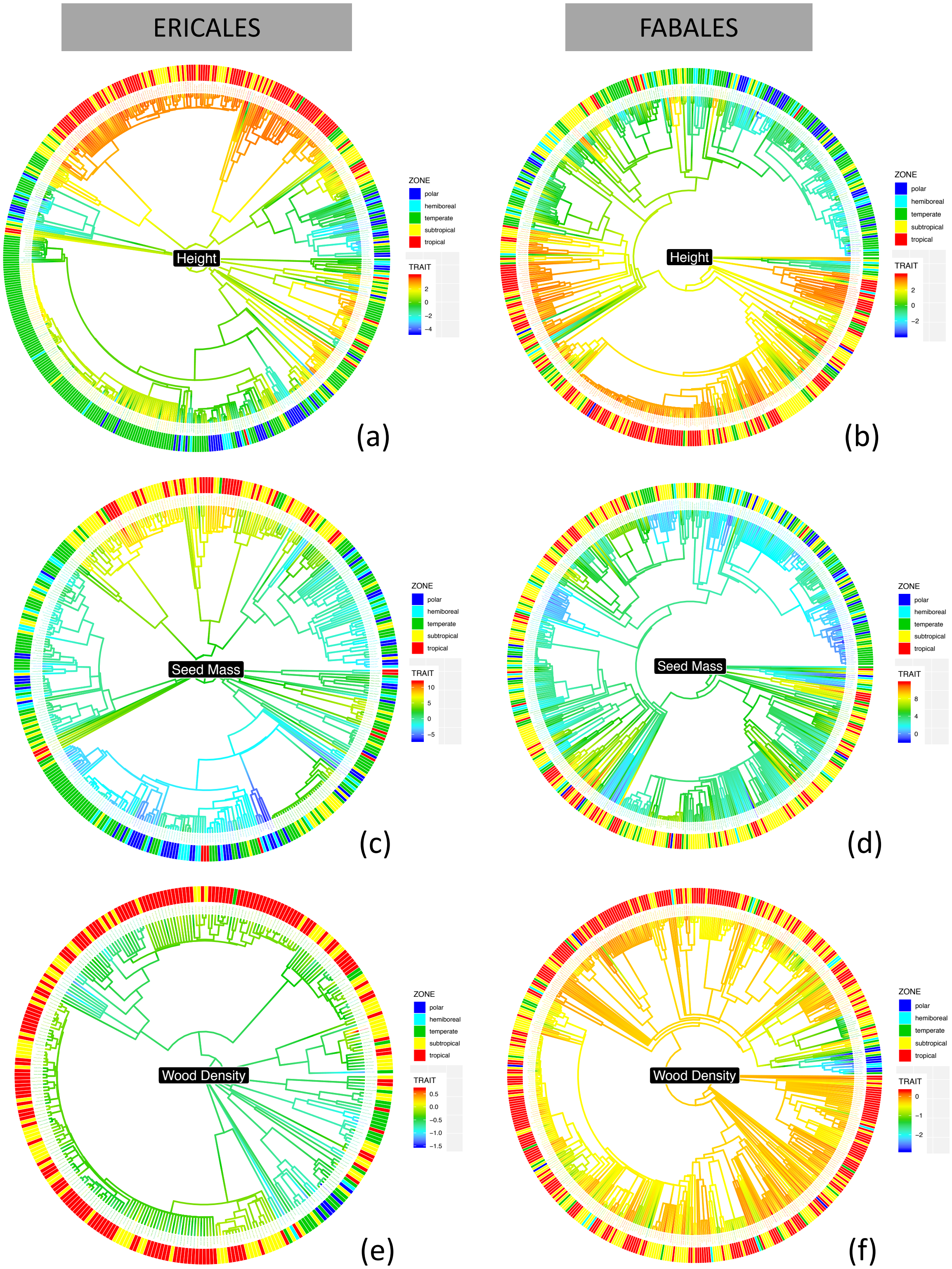
We found variation in the intensity of trait disparity for each clade and trait, using a standardized effect size metric (see Methods; **Figure 1.2**). In general, for both clades, grid cells at higher (lower) latitudes show average trait values that were significantly lower (higher) than the within-clade average. Clade-specific differences in the extent of relative trait disparity were also apparent. For example, while both groups had lower average wood density at polar latitudes, Fabales species at polar latitudes showed significantly lower wood density compared to Ericales taxa ($p<0.05$), and Ericales taxa had a higher proportion of shrub taxa in polar regions than Fabales.

Correspondence between traits, phylogeny, and geography.

Species demonstrated phylogenetically conserved life-history trait values (**Figure 1.3**). In addition to showing conserved *individual* trait values, clades within each lineage also showed conserved *suites* of trait values. For example, Sapotaceae (Ericales) tended to be tall trees with large seeds and higher wood density, while *Trifolium* (Fabales) tended to be shorter statured, herbaceous or with lower wood density, and produce smaller seeds. Overall, congeneric species were similar in their functional trait values, indicating that within-genus comparisons would generally not provide observable variation in the functional traits we examined here. However, comparisons of con-ordinal species from differing families (e.g., comparing Sapotaceae, Ericaceae, and Primulaceae) permitted observable variation in functional trait values, indicating that this phylogenetic scale is better suited for this trait syndrome. Comparisons made by grouping Ericales and Fabales together, however, could obscure the ability to see variation in this syndrome (see Discussion).

In general, there was also phylogenetic conservatism of latitudinal categories (colored bars surrounding phylogenies in **Figure 1.3**; **Figure 1.4**). Species tended to be significantly more closely related to species in the same latitudinal category (i.e., diagonal elements in **Figure 1. 4** tend to have lower values than other elements in the same row or column). Additionally, species inhabiting polar latitudes tended to be more closely related to each other than to species in other categories (i.e., the average polar-polar phylogenetic distance is usually among the smallest values in each table). Overall, therefore, taxa showed phylogenetically conserved life-history trait syndromes, which corresponded to the latitudinal-extrema at which the species are found. LMA was an exception to this trend, however; this trait showed little connection to phylogeny or latitudinal extrema.

Although data from both Ericales and Fabales was consistent with these trends, there were notable differences between the two clades in terms of how trait values and latitudinal extrema were distributed along their respective phylogenies. For example, although both clades had similar fractions of tropical species sampled for plant height (**Figure 1.S5** for proportion of species in each latitudinal category), tropical Fabales typically were clustered in the phylogeny, whereas tropical Ericales were distributed throughout multiple clades. The overall effect was that Ericales trait syndromes generally corresponded to Linnaean families, whereas the Fabales did not fall as easily into typical family or tribe categories (see Discussion).


ERICALES
FABALES
Height
ZONE
polar
hemiboreal
temperate
subtropical
tropical
TRAIT
2
0
−2
−4
(a)
Height
ZONE
polar
hemiboreal
temperate
subtropical
tropical
TRAIT
2
0
−2
(b)
Seed Mass
ZONE
polar
hemiboreal
temperate
subtropical
tropical
TRAIT
10
5
0
−5
(c)
Seed Mass
ZONE
polar
hemiboreal
temperate
subtropical
tropical
TRAIT
8
4
0
(d)
Wood Density
ZONE
polar
hemiboreal
temperate
subtropical
tropical
TRAIT
0.5
0.0
−0.5
−1.0
−1.5
(e)
Wood Density
ZONE
polar
hemiboreal
temperate
subtropical
tropical
TRAIT
0
−1
−2
(f)

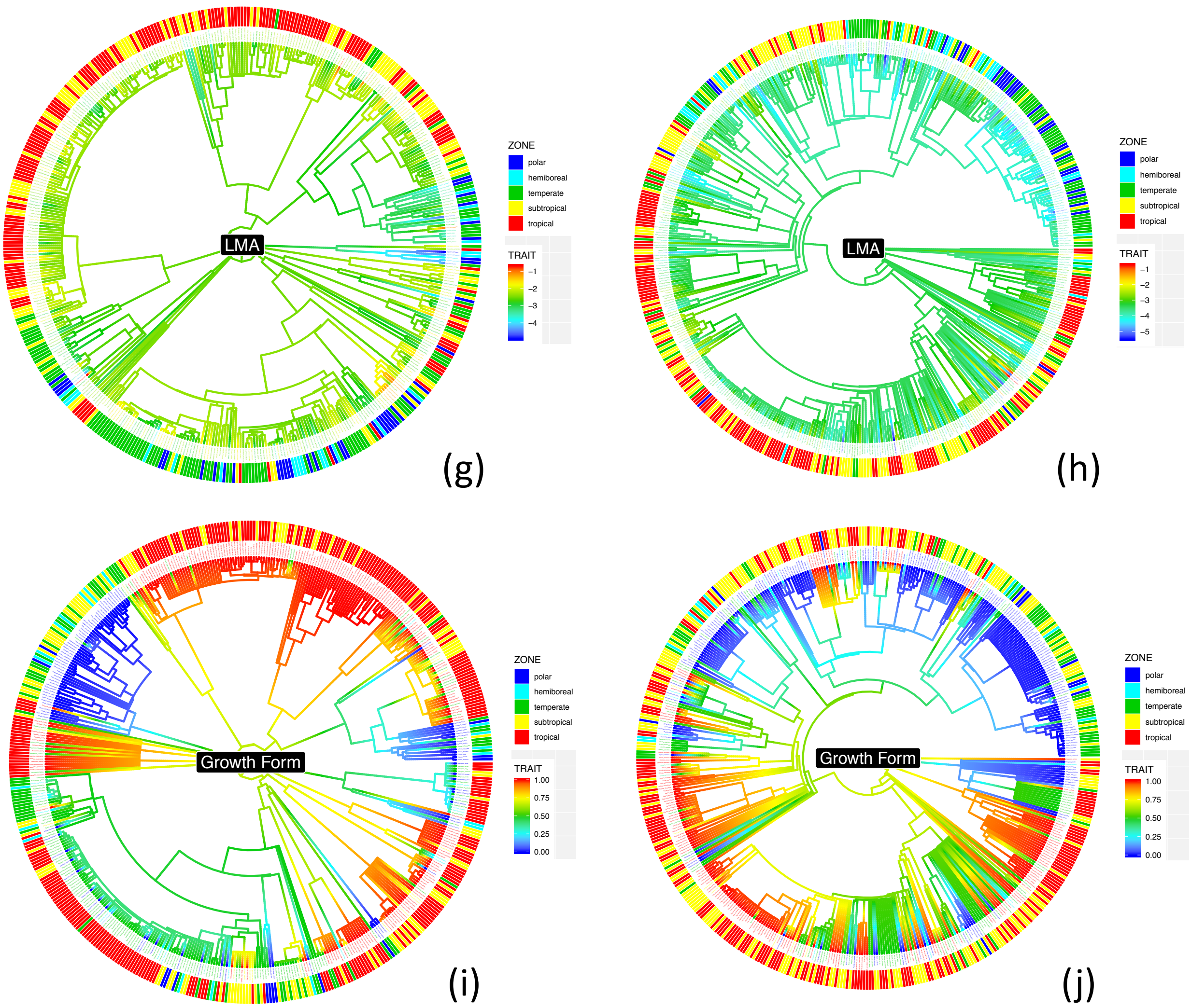
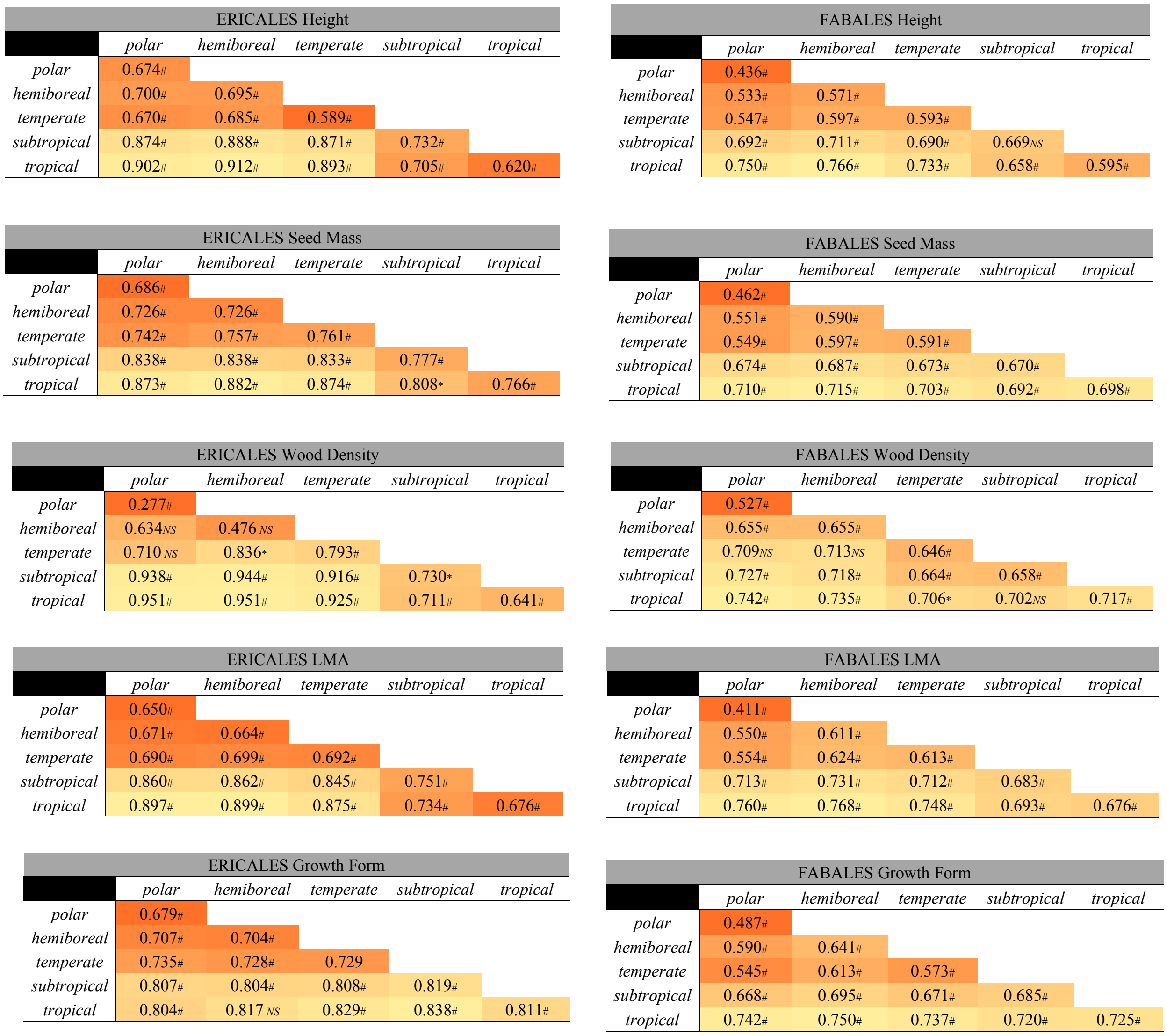


Figure 1.3. Phylogenetic conservatism of functional trait values and latitudinal extrema for Ericales (*left column*) and Fabales (*right column*). Species names are colored according to observed trait values. Colors along phylogenetic branches depict the ancestral state reconstruction for each functional trait, with warmer (cooler) colors indicating larger (smaller) trait values. Colored bars surrounding each phylogeny illustrate the assigned latitudinal category. Bar colors are as in **Figure 1.1**: polar (blue), hemiboreal (cyan), temperate (green), subtropical (yellow), tropical (red). In order to maintain readability, for traits with data available for more than 500 species, only a random sample of 500 tips are included in this figure (see online supplement to the published version of this chapter for phylogenies without any tips excluded). However, for all analyses, we included all species. For traits other than LMA, trait values are conserved, with closely related species having similar trait values. Species also tend to show conservation of latitudinal category, meaning that similarly colored bars around each phylogeny tend to be grouped together. Further, clades within each lineage tend to show conserved suites of traits when one looks across multiple traits within each column. The distribution of both functional trait values and of latitudinal category differs noticeably between Ericales and Fabales. Although similar fractions of each clade are represented in each latitudinal category (**Figure 1.S5**), extra-tropical Fabales taxa cluster

together in one portion of the phylogeny, whereas they are distributed throughout the Ericales phylogeny.



ERICALES Height

	polar	hemiboreal	temperate	subtropical	tropical
polar	0.674#				
hemiboreal	0.700#	0.695#			
temperate	0.670#	0.685#	0.589#		
subtropical	0.874#	0.888#	0.871#	0.732#	
tropical	0.902#	0.912#	0.893#	0.705#	0.620#

ERICALES Seed Mass

	polar	hemiboreal	temperate	subtropical	tropical
polar	0.686#				
hemiboreal	0.726#	0.726#			
temperate	0.742#	0.757#	0.761#		
subtropical	0.838#	0.838#	0.833#	0.777#	
tropical	0.873#	0.882#	0.874#	0.808*	0.766#

ERICALES Wood Density

	polar	hemiboreal	temperate	subtropical	tropical
polar	0.277#				
hemiboreal	0.634*NS*	0.476 *NS*			
temperate	0.710 *NS*	0.836*	0.793#		
subtropical	0.938#	0.944#	0.916#	0.730*	
tropical	0.951#	0.951#	0.925#	0.711#	0.641#

ERICALES LMA

	polar	hemiboreal	temperate	subtropical	tropical
polar	0.650#				
hemiboreal	0.671#	0.664#			
temperate	0.690#	0.699#	0.692#		
subtropical	0.860#	0.862#	0.845#	0.751#	
tropical	0.897#	0.899#	0.875#	0.734#	0.676#

ERICALES Growth Form

	polar	hemiboreal	temperate	subtropical	tropical
polar	0.679#				
hemiboreal	0.707#	0.704#			
temperate	0.735#	0.728#	0.729		
subtropical	0.807#	0.804#	0.808#	0.819#	
tropical	0.804#	0.817 *NS*	0.829#	0.838#	0.811#

FABALES Height

	polar	hemiboreal	temperate	subtropical	tropical
polar	0.436#				
hemiboreal	0.533#	0.571#			
temperate	0.547#	0.597#	0.593#		
subtropical	0.692#	0.711#	0.690#	0.669*NS*	
tropical	0.750#	0.766#	0.733#	0.658#	0.595#

FABALES Seed Mass

	polar	hemiboreal	temperate	subtropical	tropical
polar	0.462#				
hemiboreal	0.551#	0.590#			
temperate	0.549#	0.597#	0.591#		
subtropical	0.674#	0.687#	0.673#	0.670#	
tropical	0.710#	0.715#	0.703#	0.692#	0.698#

FABALES Wood Density

	polar	hemiboreal	temperate	subtropical	tropical
polar	0.527#				
hemiboreal	0.655#	0.655#			
temperate	0.709*NS*	0.713*NS*	0.646#		
subtropical	0.727#	0.718#	0.664#	0.658#	
tropical	0.742#	0.735#	0.706*	0.702*NS*	0.717#

FABALES LMA

	polar	hemiboreal	temperate	subtropical	tropical
polar	0.411#				
hemiboreal	0.550#	0.611#			
temperate	0.554#	0.624#	0.613#		
subtropical	0.713#	0.731#	0.712#	0.683#	
tropical	0.760#	0.768#	0.748#	0.693#	0.676#

FABALES Growth Form

	polar	hemiboreal	temperate	subtropical	tropical
polar	0.487#				
hemiboreal	0.590#	0.641#			
temperate	0.545#	0.613#	0.573#		
subtropical	0.668#	0.695#	0.671#	0.685#	
tropical	0.742#	0.750#	0.737#	0.720#	0.725#

Figure 1.4. Mean pairwise phylogenetic distances (MPD) among taxa within each latitudinal category, for Ericales (*top row*) and Fabales (*bottom row*). Matrices are symmetrical, and we show only the lower triangle of each. Darker (lighter) color indicates smaller (larger) phylogenetic distances, indicating species are more (less) closely related on average. In general, species in the same latitudinal category tend to be more closely related (i.e., the diagonal elements are usually darker color than other elements in the same row/column), and polar species tend to be more closely related to each other than any other combination (i.e. the polar-polar element is usually one of the smallest values in the table). More generally, species outside of the tropics tend to be more closely related to each other than to tropical or subtropical species (the converse is also true). Most values were statistically significant (see Methods). *($p<0.001$, #; $p< 0.05$, *; not significant, NS.)*

Categorizing trait syndromes.

Although not every species was represented for every trait (see Methods), by examining multiple traits across each plant clade, the correspondence between phylogeny and functional trait values suggested clades could be categorized by trait syndromes (**Figure 1.5a-b**). In general, the phylogenies showed that clusters of species exhibited shared suite of traits, indicating phylogenetic conservatism of this functional trait syndrome. LMA is an exception to this pattern, appearing instead as a 'barcode' pattern with very little relationship to phylogeny (**Figure 1.5a-b**). This trait syndrome was quantified using phylogenetic principal components analysis (pPCA; **Figure 1.6**). For both clades, the first two principal components (PCs) accounted for over 88% of the total variance, though we include the first three here (>95% variance explained; **Figure 1.6**). For Fabales, tropical and subtropical species clustered separately from temperate and boreal species, and this separation is visible in all 3 PCs, suggesting two broad trait patterns at the transition from tropical to temperate latitudes. For Ericales, there was greater separation between temperate taxa and those found in boreal latitudes (3 groups). Thus, the manifestation of the trait syndrome differs between these two clades (see Discussion).

We also examined the evidence for correlated evolution of functional trait values. We illustrated this graphically in **Figure 1.5c-d**, where phylogenies are mapped onto the trait space of each pairwise trait combination. For many trait combinations, there was a visually apparent correlation between the traits examined. Further, we utilized PGLS to ask whether or not these correlations were significant *beyond* what we would expect under a Brownian motion model of trait evolution (p-values in lower right corner of each plot). Most, but not all, PGLS correlations were found to be significant. LMA, in particular, did not show evidence of correlated evolution. Our results suggest that clades possess a shared suite of functional trait values that may have evolved in a correlated manner, though greater coverage of these functional traits would help strengthen this claim.

Bioregion mapping and trait distributions.

In addition to latitudinal categories, for those species with complete functional trait information, we also allowed species' ranges to define biogeographical clusters (see Methods), which in turn might be expected to use different functional-trait strategies. For both clades, we obtained a similar number of clusters (Ericales: 11; Fabales: 10), and clusters converged on similar geographic regions (**Figure 1.S6**). When measuring average trait values among all species assigned to each cluster, we found that clusters tended to exhibit consistent trait patterns (e.g., all low or all high trait values, excepting LMA, as in above analyses; insets of **Figure 1.S6**). Additionally, geographically adjacent regions tended to exhibit similar trait values. Seed mass, however, showed an abrupt transition dividing tropical clusters (large average seed mass) from extra-tropical (low average seed mass) ones. Both trait data limitations and sparse sampling from certain regions (e.g., southern South America) influenced the biogeographical clusters obtainable in this analysis and enhanced data collection should continuously update and refine the clusters thus identified.

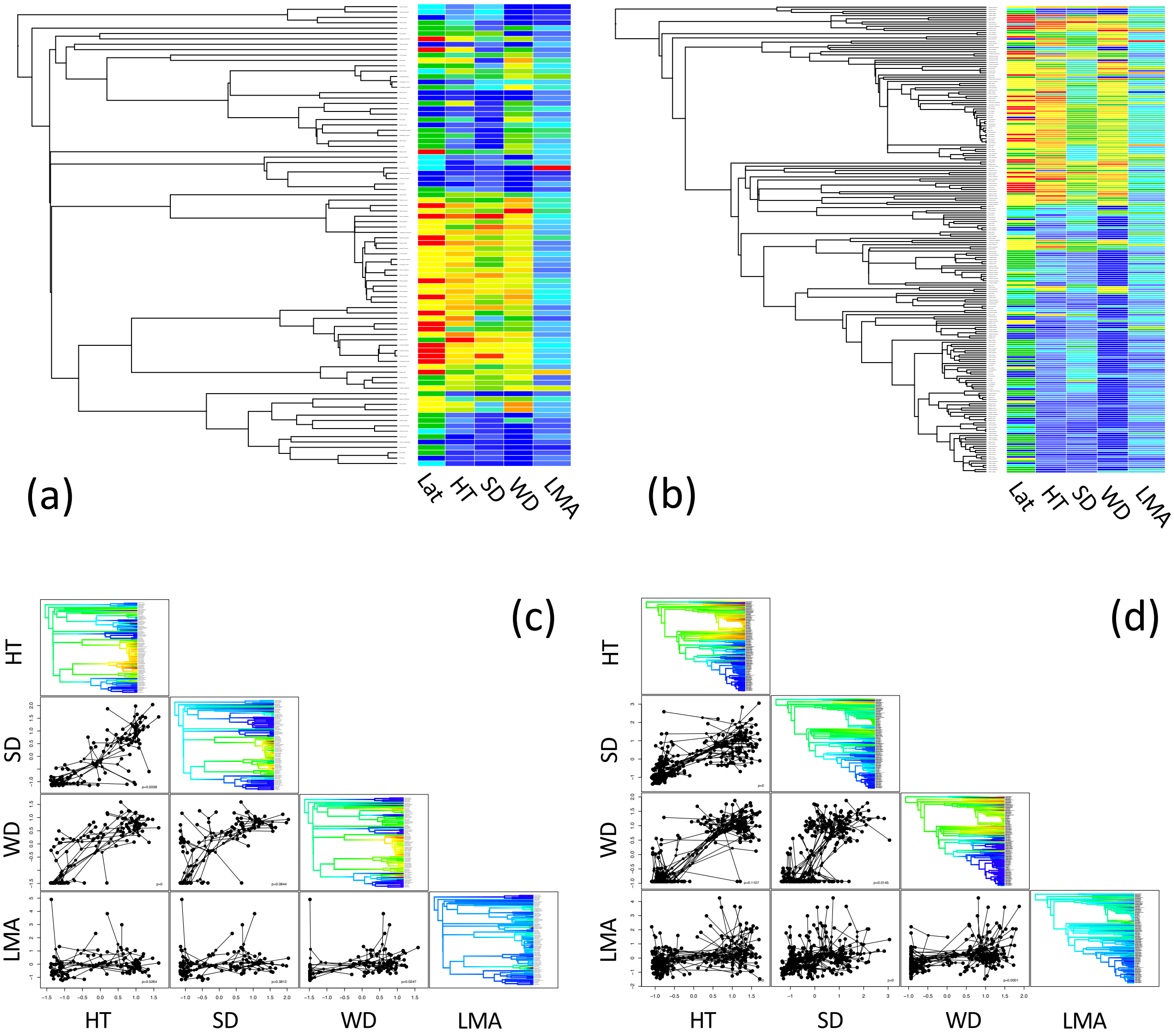


Figure 1.5. Evidence of correlated evolution among functional traits. *Top panels*: Phylogenetic heatmaps for (*a*) Ericales and (*b*) Fabales taxa for which complete trait information is available, indicating a conserved trait syndrome. The first column of trait values (*Lat.*) corresponds to the latitudinal extrema (colors are as in **Figures 1.1** and **1.3**). For functional traits, we include height (*HT*), seed mass (*SD*), wood density (*WD*), and *LMA* here because these traits were used in tests of correlated evolution (bottom panels) and phylogenetic principal components analysis (pPCA, **Figure 1.6**). Growth form was excluded in those analyses because it is not an ordinated trait. Colors for each trait are as in the ancestral state reconstructions of **Figure 1.3**, and bottom panels of this same figure. In general, but with the exception of LMA, species tend to show blocks of trait values consistent with a trait syndrome. This trait syndrome corresponds well to species assigned latitudinal category. *Bottom panels*: Tests for correlated evolution among those functional traits depicted in top panels for Ericales (*c*) and Fabales (*d*) taxa. The grids show correlations of each pairwise trait combination, with the phylogeny mapped onto the two-dimensional trait space. Along the diagonal of the grid, we plot ancestral state reconstructions for each trait on the phylogenies shown in (*a*) and (*b*) above. With the exception of LMA, correlations between traits are discernible. We utilized phylogenetic generalized least squares (PGLS) to determine if correlations were significant *beyond* what would be expected under a Brownian motion model of trait evolution (p-values indicated in lower right corners of each plot). Correlations with LMA were generally not significant, whereas other correlations were. Though relatively few species have sufficient trait data to allow for these comparative methods, this is consistent with correlated evolution of these traits—thus indicating that the relative trait values defining the syndrome we describe here might have evolved in concert.

DISCUSSION

Our results provide evidence among New World Ericales and Fabales of a latitudinally-graded (**Figures 1.1-1.2**) and phylogenetically conserved (**Figures 1.3-1.5**) functional-trait syndrome involving height, seed mass, wood density, and growth form, but not LMA. Moving poleward, species tend to exhibit progressively shorter stature, smaller seeds, lower wood density, and become predominantly herbaceous or shrubby. pPCA for those taxa with complete trait data suggest that this trait syndrome can be classified based on the clustering of species into separate groups in trait-space (**Figure 1.6**). Additionally, the comparative analyses provide evidence that the life-history traits involved in this syndrome may have evolved in a correlated manner (**Figure 1.5c-d**).

Our results also highlight quantitative, clade-specific, differences in the manifestation of this syndrome, illustrating the need for targeted phylogenetic approaches and examining potential patterns at different phylogenetic scales. While previous studies have made strides in examining broad intersections of ecological function and plant traits, they have typically incorporated phylogeny in a limited sense (e.g. Swenson et al., 2012), or used an extremely broad phylogenetic scale (e.g. Zanne et al., 2014). However, just as ecological studies have gained appreciation for the appropriate choice of spatial scale in study design, choice of phylogenetic scale can also be an important decision. To observe variation in the trait syndrome we describe here required looking beyond the level of congeners, which, for those traits we examined, tended to have very similar trait values, and our results suggest that comparisons of species sampled from con-ordinal families provided the best scale at which this trait syndrome could be quantified.

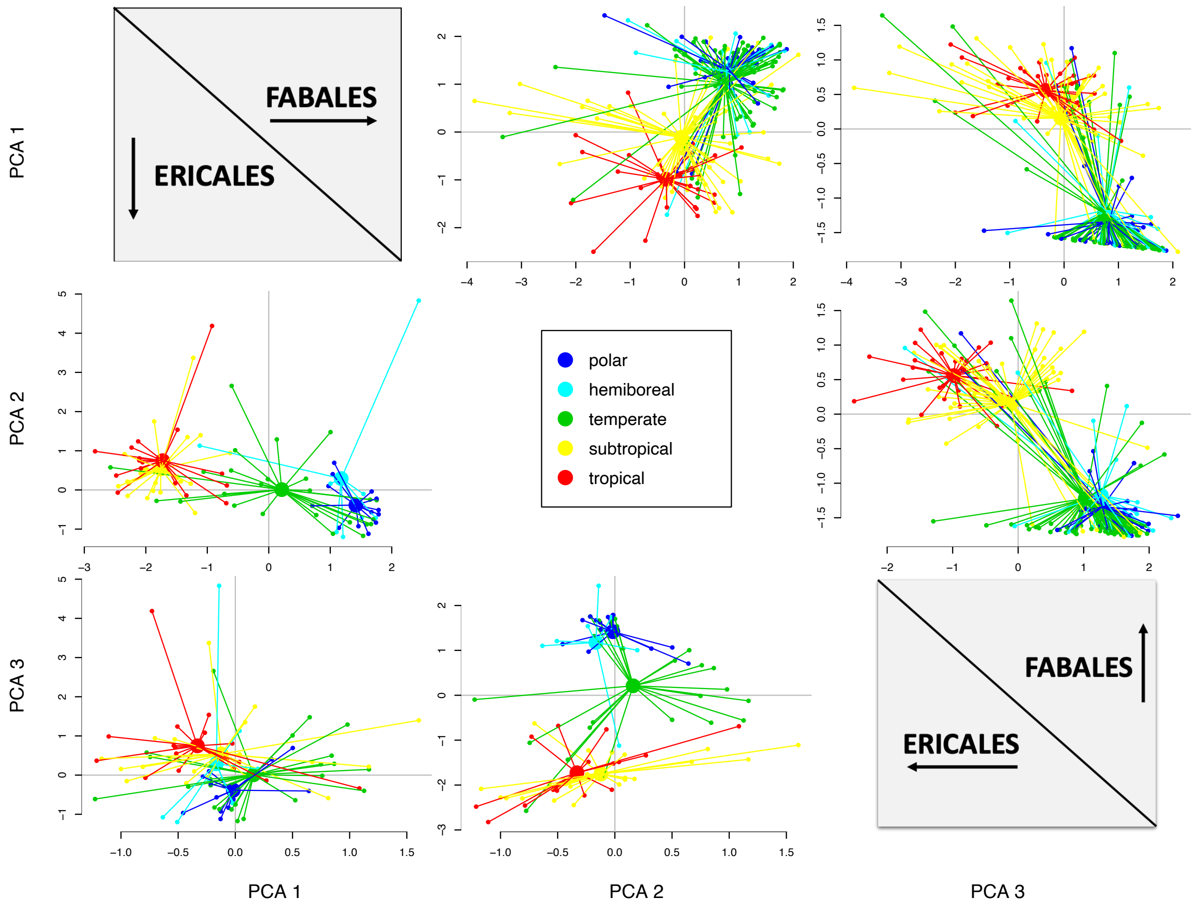


Figure 1.6. Phylogenetic principal components analysis (pPCA) among height, seed mass, wood density, and LMA, for Ericales and Fabales taxa (as indicated by arrows in the plot). Growth form was excluded in the ordination because it is not an ordinated trait. For both clades, the first two principal components capture >88% of the total variance, but we include the first three components here (>92% total variance). Small points indicate individual species, while larger circles indicate centroids of each group. The groups are colored based on latitudinal category (colors are as in **Figure 1.1**). For both Ericales and Fabales, species within different latitudinal categories cluster separately in

> the composite trait space. This is consistent with the latitudinally-graded multi-trait syndrome we describe here. However, there are differences between the two clades. Fabales shows two main clusters, with tropical and subtropical species clustering separately from temperate, hemiboreal, and polar ones. Ericales, however, shows greater separation between temperate and hemiboreal/polar taxa (three clusters). Thus, the manifestation of the trait syndrome differs between these two groups.

The importance of scale can be illustrated by examining seed mass measurements. If Fabales and Ericales seed mass measurements are lumped together, latitudinal differences among species whose ranges extend to temperate, boreal, and polar areas are largely obscured due to the statistically significantly differences in average seed mass between the two groups (**Figure 1.S1**). There are also clade-specific differences in plant height among taxa within the same latitudinal category. For example, among polar-latitude species, Fabales are significantly taller than Ericales ($p<0.05$, **Figure 1.1**, compare blue boxes), but at temperate latitudes, the situation is reversed and Ericales are taller ($p<0.05$, **Figure 1.1**, compare green boxes). In **Figure 1.S3**, we show that, at polar latitudes Ericales occur at sites with lower average temperature and higher average elevation than Fabales. In contrast, in temperate regions, Ericales tend to occur at sites with lower average elevation and higher average precipitation. These differences in how taxa are distributed according to smaller-scale climatic factors might help explain some of the clade-specific differences in the manifestation of the trait syndrome described here, particularly height and seed mass. We also note that other trait syndromes might be best observed at quite different phylogenetic scales. For instance, the disparity in traits relating to chemical defense against herbivory among co-occurring congeneric species of *Inga* is known to be quite high (Sedio et al., 2018). Thus, careful and study-specific choice of phylogenetic scale can be important in teasing out processes of interest.

Unlike Ericales (Asterids), Fabales (Rosids) are part of the nitrogen-fixing clade (Friis et al., 2011), and form root nodules in association with Rhizobia bacteria to fix atmospheric nitrogen. Further, in higher-latitude environments, Fabales tend to preferentially produce indeterminate nodules, whereas determinate nodules are utilized more in tropical forests (Sprent, 2009). Thus, Fabales taxa show a distinct high-latitude nitrogen-fixing phenotype. It is possible that such differences in the nitrogen-fixation mode also contribute to differences in how these two clades manifest this trait syndrome and how they are arrayed along elevational and precipitation gradients, particularly at higher latitudes, where nitrogen may be more limiting (Reich & Oleksyn, 2004).

Although we focused on latitudinal categories for species, we also clustered species into bioregions using the InfoMap algorithm (see Methods; **Figure 1.S6**; Edler et al., 2016). That both Ericales and Fabales species clustered into a similar number of roughly geographically convergent bioregions could suggest these regions represent areas with distinct ecological strategies. Indeed, the functional trait values for height, seed mass, and wood density within each bioregion tended to coincide with the presence of the functional trait syndrome we describe here, and LMA showed little relationship with any of the bioregions. Geographically proximate bioregions tended to have similar trait values, suggesting that clustering species by bioregions supplemented rather than supplanted the latitudinally-based analyses, and that both methods of organizing species illustrate the occurrence of a similar trait syndrome. Clade-specific differences were also apparent in the bioregion analysis. Most notably, tropical Ericales were divided into two bioregions, whereas tropical Fabales formed a single, large cluster.

The manifestation of this trait syndrome, occurring under a variety of geographically based delimiting methods, could have originated under a variety of scenarios. For instance, this trait syndrome may represent a resource acquisition strategy whose manifestation is primarily influenced by soil nutrient availability and growing season length (i.e., the time in which available nutrients can be utilized). In particular, we note that, at higher latitudes, the manifestation of this trait syndrome is qualitatively similar to the suite of traits associated with 'pioneer,' or early successional-stage, tropical species (e.g. smaller stature, lower seed mass, LMA, and wood density; Wright et al., 2010). 'Pioneer,' here, refers to those plants that colonize disturbed patches or gaps first and display enhanced growth rates under favorable conditions. In tropical soils, nutrient availability is generally lower than at higher absolute latitudes (Reich & Oleksyn, 2004). Treefall gaps can help to unlock some of these nutrients and make them available for early colonizers employing a resource-acquisitive strategy. As succession proceeds, plant growth within the gap again sequesters those nutrients and the manifestation of the resource-acquisition strategy changes, typically characterized by taller stature trees with larger seeds and higher wood density.

At higher latitudes, soil nutrients are relatively more abundant than in tropical soils (Reich & Oleksyn, 2004), and recent glaciation cycles have further helped to enhance deposition of these nutrients (Chadwick et al., 1999; Van Andel, 1994). However, also at progressively higher latitudes, the incident angle of light changes, resulting in a shorter effective growing season for plants. This, in turn, limits the time over which plants can utilize these available nutrients, and might favor a more resource-acquisitive strategy in these areas as well. The effects of light seasonality are latitudinally-graded and would predate the onset of global cooling in the Eocene (Graham, 2011). Additionally, temperature seasonality imposes limits on growing-season length at higher latitudes, favoring a more resource-acquisitive strategy as well.

The manifestation of the trait syndrome we describe, and its similarity to a tropical pioneer species strategy, might, therefore, be explained as a strategy to deal with both light and temperature seasonality in a relatively nutrient-rich environment with a short growing season. This trait syndrome could have arisen in several different ways. For example, it may have originated through the dispersal of tropical pioneer lineages preadapted to have a more resource-acquisitive strategy. Or it might have arisen independently at higher latitudes in response to light seasonality prior to the onset of global cooling, when the flora at such latitudes consisted of both boreal and tropical elements (e.g. boreotropical forests; Dick & Pennington, 2019; Graham, 2011). Under this hypothesis, lineages found in colder environments might represent longer-lasting endemics of past, non-analogous biomes, whose ecological strategies for addressing light seasonality became advantageous for temperature seasonality as well. Since light seasonality predates global cooling, and a similar trait syndrome occurs among tropical pioneers, where cold-tolerance is not a factor, it may be that the higher-latitude trait syndrome we observe is not a direct cold adaptation *per se*. Further work investigating timing of shifts in trait values with the biogeographic history of lineages undergoing those shifts would be invaluable to tease apart these scenarios or discover alternative explanations.

Our ability to describe the correlated evolution among this collection of (rather than individual) traits was limited significantly by data availability. In particular, wood density measurements were primarily for tropical tree species, and the extreme southern latitudes of South America and regions of the Amazon basin suffered from relatively poor species sampling. Given the destructive and labor-intensive methodology required for obtaining wood density, it is perhaps unsurprising this trait would be limiting. However, greater sampling of temperate and boreal wood densities for angiosperm species would have greatly enhanced our dataset. We suggest that as an avenue for further data collection, along with greater species sampling at extreme southern latitudes for Ericales and Fabales.

Conclusion

We suggest the tendency for a biogeographically consistent and phylogenetically conserved functional trait syndrome involving height, seed mass, wood density, and growth form (and excluding LMA) among Ericales (Asterids) and Fabales (Rosids) New World taxa. By focusing on specific lineages, we are able to illustrate that variation in this trait syndrome is best observed by sampling species from con-ordinal families. Quantitative differences in the manifestation of this trait syndrome among Ericales and Fabales taxa occurring in similar regions highlight the importance of matching phylogenetic scale with the research question(s) and pose potential new research avenues. We interpret the manifestation of this trait syndrome at higher latitudes as a resource-acquisitive strategy adapted to relatively high soil nutrient content coupled with a short growing season, a situation somewhat analogous to that faced by gap specialists in tropical forest communities which exhibit a similar suite of traits.

REFERENCES

Blomberg, S. P., Garland, T., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*, *57*(4), 717–745. https://doi.org/10.1111/j.0014-3820.2003.tb00285.x

Chadwick, O. A., Derry, L. A., Vitousek, P. M., Huebert, B. J., & Hedin, L. O. (1999). Changing sources of nutrients during four million years of ecosystem development. *Nature*, *397*(6719), 491–497. https://doi.org/10.1038/17276

Chave, J., Coomes, D., Jansen, S., Lewis, S. L., Swenson, N. G., & Zanne, A. E. (2009). Towards a worldwide wood economics spectrum. *Ecology Letters*, *12*(4), 351–366. https://doi.org/10.1111/j.1461-0248.2009.01285.x

Conti, G., & Díaz, S. (2013). Plant functional diversity and carbon storage—An empirical test in semi-arid forest ecosystems. *Journal of Ecology*, *101*(1), 18–28. https://doi.org/10.1111/1365-2745.12012

Dick, C. W., & Pennington, R. T. (2019). History and Geography of Neotropical Tree Diversity. *Annual Review of Ecology, Evolution, and Systematics*, *50*(1), 279–301. https://doi.org/10.1146/annurev-ecolsys-110617-062314

Edler, D., Guedes, T., Zizka, A., Rosvall, M., & Antonelli, A. (2016). Infomap Bioregions: Interactive Mapping of Biogeographical Regions from Species Distributions. *Systematic Biology*, syw087. https://doi.org/10.1093/sysbio/syw087

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Friis, E. M., Crane, P. R., & Pedersen, K. R. (2011). *Early flowers and angiosperm evolution*. Cambridge University Press.

Graham, A. (2011). *A Natural History of the New World.* University of Chicago Press.

Grime, J. P. (2006). *Plant Strategies, Vegetation Processes, and Ecosystem Properties.* (2nd ed.). John Wiley & Sons, Inc.

Hawkins, B. A., Rueda, M., Rangel, T. F., Field, R., & Diniz-Filho, J. A. F. (2014). Community phylogenetics at the biogeographical scale: Cold tolerance, niche conservatism and the structure of North American forests. *Journal of Biogeography*, *41*(1), 23–38. https://doi.org/10.1111/jbi.12171

Hedwall, P.-O., Brunet, J., Nordin, A., & Bergh, J. (2013). Changes in the abundance of keystone forest floor species in response to changes of forest structure. *Journal of Vegetation Science*, *24*(2), 296–306. https://doi.org/10.1111/j.1654-1103.2012.01457.x

Kahle, D., & Wickham, H. (2013). ggmap: Spatial Visualization with ggplot2. *The R Journal*, *5*(1), 144. https://doi.org/10.32614/RJ-2013-014

Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I., Leadley, P., & … Wirth, C. (2020). TRY plant trait database – enhanced coverage and open access. *Global Change Biology*, *26*, 119–188.

Li, H.-L., Wang, W., Mortimer, P. E., Li, R.-Q., Li, D.-Z., Hyde, K. D., Xu, J.-C., Soltis, D. E., & Chen, Z.-D. (2015). Large-scale phylogenetic analyses reveal multiple gains of actinorhizal nitrogen-fixing symbioses in angiosperms associated with climate change. *Scientific Reports*, *5*(1), 14023. https://doi.org/10.1038/srep14023

Magallón, S., Sánchez-Reyes, L. L., & Gómez-Acevedo, S. L. (2019). Thirty clues to the exceptional diversification of flowering plants. *Annals of Botany*, *123*, 491–503.

Moles, A. T., Ackerly, D. D., Webb, C. O., Tweddle, J. C., Dickie, J. B., & Westoby, M. (2005). A Brief History of Seed Size. *Science*, *307*(5709), 576–580. https://doi.org/10.1126/science.1104863

Moles, A. T. (2018). Being John Harper: Using evolutionary ideas to improve understanding of global patterns in plant traits. *Journal of Ecology*, *106*(1), 1–18. https://doi.org/10.1111/1365-2745.12887

Moles, A. T., Ackerly, D. D., Tweddle, J. C., Dickie, J. B., Smith, R., Leishman, M. R., Mayfield, M. M., Pitman, A., Wood, J. T., & Westoby, M. (2007). Global patterns in seed size. *Global Ecology and Biogeography*, *16*(1), 109–116. https://doi.org/10.1111/j.1466-8238.2006.00259.x

Moles, A. T., & Leishman, M. R. (2008). The seedling as part of a plant's life history strategy. In M. A. Leck, V. T. Parker, & R. L. Simpson (Eds.), *Seedling Ecology and Evolution* (pp. 217–238). Cambridge University Press. https://doi.org/10.1017/CBO9780511815133.012

Moles, A. T., Warton, D. I., Warman, L., Swenson, N. G., Laffan, S. W., Zanne, A. E., Pitman, A., Hemmings, F. A., & Leishman, M. R. (2009). Global patterns in plant height. *Journal of Ecology*, *97*(5), 923–932. https://doi.org/10.1111/j.1365-2745.2009.01526.x

Moquet, L., Vanderplanck, M., Moerman, R., Quinet, M., Roger, N., Michez, D., & Jacquemart, A.-L. (2017). Bumblebees depend on ericaceous species to survive in temperate heathlands. *Insect Conservation and Diversity*, *10*(1), 78–93. https://doi.org/10.1111/icad.12201

Muller-Landau, H. C. (2010). The tolerance-fecundity trade-off and the maintenance of diversity in seed size. *Proceedings of the National Academy of Sciences*, *107*(9), 4242–4247. https://doi.org/10.1073/pnas.0911637107

Muller-Landau, H. C., Wright, S. J., Calderón, O., Condit, R., & Hubbell, S. P. (2008). Interspecific variation in primary seed dispersal in a tropical forest. *Journal of Ecology*, *96*(4), 653–667. https://doi.org/10.1111/j.1365-2745.2008.01399.x

Pierce, S., Negreiros, D., Cerabolini, B. E. L., Kattge, J., Díaz, S., Kleyer, M., Shipley, B., Wright, S. J., Soudzilovskaia, N. A., Onipchenko, V. G., van Bodegom, P. M., Frenette-Dussault, C., Weiher, E., Pinho, B. X., Cornelissen, J. H. C., Grime, J. P., Thompson, K., Hunt, R., Wilson, P. J., … Tampucci, D. (2017). A global method for calculating plant CSR ecological strategies applied across biomes world-wide. *Functional Ecology*, *31*(2), 444–457. https://doi.org/10.1111/1365-2435.12722

Polly, P. D., Lawing, A. M., Fabre, A.-C., & Goswami, A. (2013). Phylogenetic Principal Components Analysis and Geometric Morphometrics. *Hystrix, the Italian Journal of Mammalogy*, *24*(1), 33–41. https://doi.org/10.4404/hystrix-24.1-6383

R Development Core Team. 2011. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. (n.d.).

Reich, P. B., & Oleksyn, J. (2004). Global patterns of plant leaf N and P in relation to temperature and latitude. *Proceedings of the National Academy of Sciences*, *101*(30), 11001–11006. https://doi.org/10.1073/pnas.0403588101

Reich, P. B., Wright, I. J., Cavender-Bares, J., Craine, J. M., Oleksyn, J., Westoby, M., & Walters, M. B. (2003). The Evolution of Plant Functional Variation: Traits, Spectra, and Strategies. *International Journal of Plant Sciences*, *164*(S3), S143–S164. https://doi.org/10.1086/374368

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things): phytools: R package. *Methods in Ecology and Evolution*, *3*(2), 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Rose, J. P., Kleist, T. J., Löfstrand, S. D., Drew, B. T., Schönenberger, J., & Sytsma, K. J. (2018). Phylogeny, historical biogeography, and diversification of angiosperm order Ericales suggest ancient Neotropical and East Asian connections. *Molecular Phylogenetics and Evolution*, *122*, 59–79. https://doi.org/10.1016/j.ympev.2018.01.014

Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, *105*(4), 1118–1123. https://doi.org/10.1073/pnas.0706851105

Sedio, B. E., Parker, J. D., McMahon, S. M., & Wright, S. J. (2018). *Comparative metabolomics of forest communities: Species differences in foliar chemistry are greater in the tropics.* https://doi.org/10.1101/271361

Segovia, R. A., Pennington, R. T., Baker, T. R., de Souza, F. C., Neves, D. M., Davis, C. C., Armesto, J. J., Olivera-Filho, A. T., & Dexter, K. G. (2020). Freezing and water availability structure the evolutionary diversity of trees across the Americas. *SCIENCE ADVANCES*, 9.

Smith, S. A., & Brown, J. W. (2018). Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany*, *105*(3), 302–314. https://doi.org/10.1002/ajb2.1019

Sprent, J. (2009). *Legume nodulation: A global perspective*. Wiley-Blackwell.

Stevens, P. F. (2001). *Angiosperm Phylogeny Website. Version 14, July 2017 [and more or less continuously updated since]*.

Swenson, N. G., Erickson, D. L., Mi, X., Bourg, N. A., Forero-Montaña, J., Ge, X., Howe, R., Lake, J. K., Liu, X., Ma, K., Pei, N., Thompson, J., Uriarte, M., Wolf, A., Wright, S. J., Ye, W., Zhang, J., Zimmerman, J. K., & Kress, W. J. (2012). Phylogenetic and functional alpha and beta diversity in temperate and tropical tree communities. *Ecology*, *93*(sp8), S112–S125. https://doi.org/10.1890/11-0402.1

Swenson, N. G., & Weiser, M. D. (2010). Plant geography upon the basis of functional traits: An example from eastern North American trees. *Ecology*, *91*(8), 2234–2241. https://doi.org/10.1890/09-1743.1

ter Steege, H., Pitman, N. C. A., Phillips, O. L., Chave, J., Sabatier, D., Duque, A., Molino, J.-F., Prévost, M.-F., Spichiger, R., Castellanos, H., von Hildebrand, P., & Vásquez, R. (2006). Continental-scale patterns of canopy tree composition and function across Amazonia. *Nature*, *443*(7110), 444–447. https://doi.org/10.1038/nature05134

The Angiosperm Phylogeny Group. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, *181*(1), 1–20. https://doi.org/10.1111/boj.12385

Van Andel, T. (1994). *New Views on an Old Planet: A History of Global Change.* (2nd ed.). Cambridge University Press.

Wang, L.-G., Lam, T. T.-Y., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C. W., Jones, B. R., Bradley, T., Zhu, H., Guan, Y., Jiang, Y., & Yu, G. (2020). treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. *Molecular Biology and Evolution*, *37*(2), 599–603. https://doi.org/10.1093/molbev/msz240

Weiner, J. (2015). *Pca3d: Three dimensional PCA plots. R package version 0.8, 484.*

Westoby, M., Falster, D. S., Moles, A. T., Vesk, P. A., & Wright, I. J. (2002). Plant Ecological Strategies: Some Leading Dimensions of Variation Between Species. *Annual Review of Ecology and Systematics*, *33*(1), 125–159. https://doi.org/10.1146/annurev.ecolsys.33.010802.150452

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag.

Wiens, J. J., Ackerly, D. D., Allen, A. P., Anacker, B. L., Buckley, L. B., Cornell, H. V., Damschen, E. I., Davies, T. J., Grytnes, J.-A., Harrison, S. P., Hawkins, B. A., Holt, R. D., McCain, C. M., & Stephens, P. R. (2010). Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters*, *13*(10), 1310–1324. https://doi.org/10.1111/j.1461-0248.2010.01515.x

Wikström, N., Savolainen, V., & Chase, M. W. (2001). Evolution of the angiosperms: Calibrating the family tree. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1482), 2211–2220. https://doi.org/10.1098/rspb.2001.1782

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J. H. C., Diemer, M., Flexas, J., Garnier, E., Groom, P. K., Gulias, J., Hikosaka, K., Lamont, B. B., Lee, T., Lee, W., Lusk, C., … Villar, R. (2004). The worldwide leaf economics spectrum. *Nature*, *428*(6985), 821–827. https://doi.org/10.1038/nature02403

Wright, S. J., Kitajima, K., Kraft, N. J. B., Reich, P. B., Wright, I. J., Bunker, D. E., Condit, R., Dalling, J. W., Davies, S. J., Díaz, S., Engelbrecht, B. M. J., Harms, K. E., Hubbell, S. P., Marks, C. O., Ruiz-Jaen, M. C., Salvador, C. M., & Zanne, A. E. (2010). Functional traits and the growth–mortality trade-off in tropical trees. *Ecology*, *91*(12), 3664–3674. https://doi.org/10.1890/09-2335.1

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2017). ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36. https://doi.org/10.1111/2041-210X.12628

Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. A., FitzJohn, R. G., McGlinn, D. J., O'Meara, B. C., Moles, A. T., Reich, P. B., Royer, D. L., Soltis, D. E., Stevens, P. F., Westoby, M., Wright, I. J., Aarssen, L., Bertin, R. I., Calaminus, A., Govaerts, R., … Beaulieu, J. M. (2014). Three keys to the radiation of angiosperms into freezing environments. *Nature*, *506*(7486), 89–92. https://doi.org/10.1038/nature12872

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V., & Antonelli, A. (2019). CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*, *10*(5), 744–751. https://doi.org/10.1111/2041-210X.13152

SUPPORTING INFORMATION

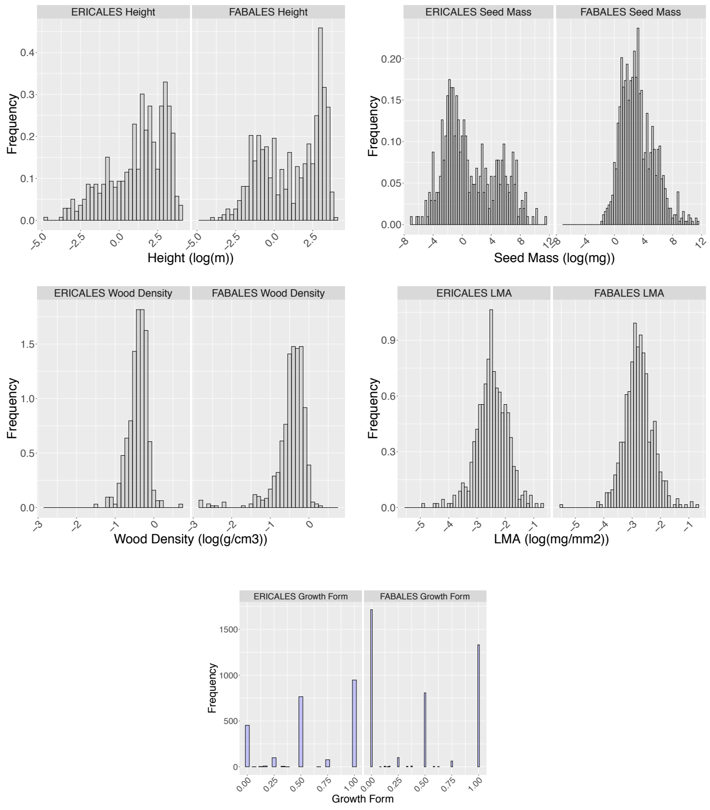


Figure 1.S1. Comparison of Ericales and Fabales trait distributions. In general, the overall distribution of trait values across each plant order is similar, indicating that differences between groups at similar latitudes (**Figure 1.1**) are not generally caused by average differences in trait values achieved between the two clades. Growth form is a categorical trait treated differently from the others. The three spikes for growth form indicate herbs (0), shrubs (0.5), and trees (1; see Methods), with bars at intermediate values indicating multiple growth forms listed for a given species. For two traits, significant clade-wide differences exist between Ericales and Fabales: (1) Fabales produce larger seeds on average than Ericales. This is due both to a lack of very small seeds among Fabales, and a greater proportion of small seeds produced by Ericales; (2) A greater overall proportion of Fabales are herbaceous, whereas a greater proportion of Ericales are trees.

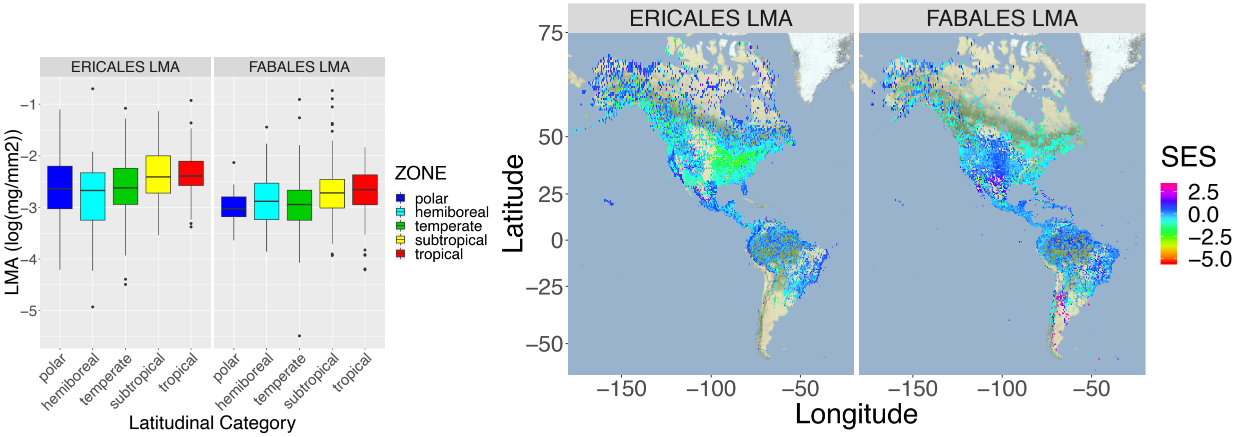


Figure 1.S2. Geographic results for LMA for both Ericales (*left panels*) and Fabales (*right panels*). Boxplots are as in **Figure 1.1**. Although there is some trend of tropical taxa having higher median LMA values than polar ones, the trend is quite slight and not significant. Maps are as in **Figure 1.2**. For both Ericales and Fabales, regardless of latitude, most grid cells have average values that are close to the overall LMA average. Fabales taxa do show some notable hotspots of particularly low LMA values—near the southwestern United States and northwestern/ central Mexico, which coincide with arid desert climates.

***** ***** *****

For Supplemental Figures 1.S3 and 1.S4, the reader is referred to the online supplement of the published version of this chapter (https://doi.org/10.1111/jbi.13993).

***** ***** *****

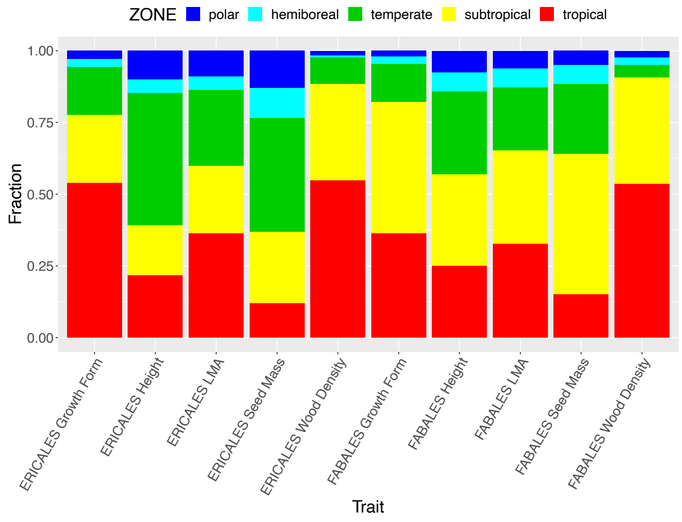
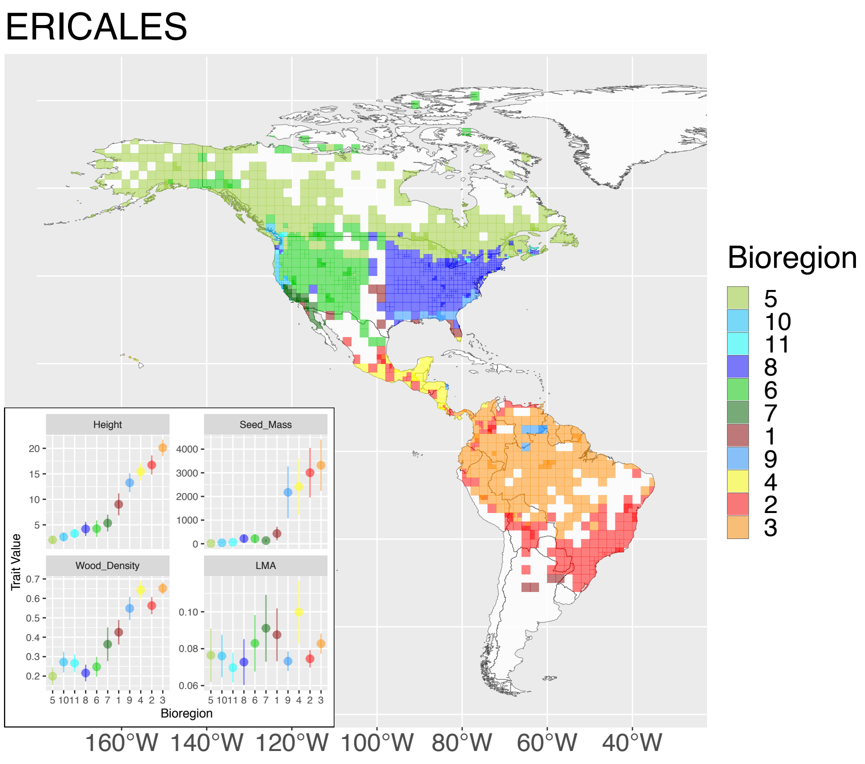


Figure 1.S5. Stacked bar plot showing the fraction of species in each latitudinal category for each trait among Ericales and Fabales taxa. When looking at the same trait, both groups showed similar fractions of species in each latitudinal category, despite there being more data for Fabales overall (see Methods).



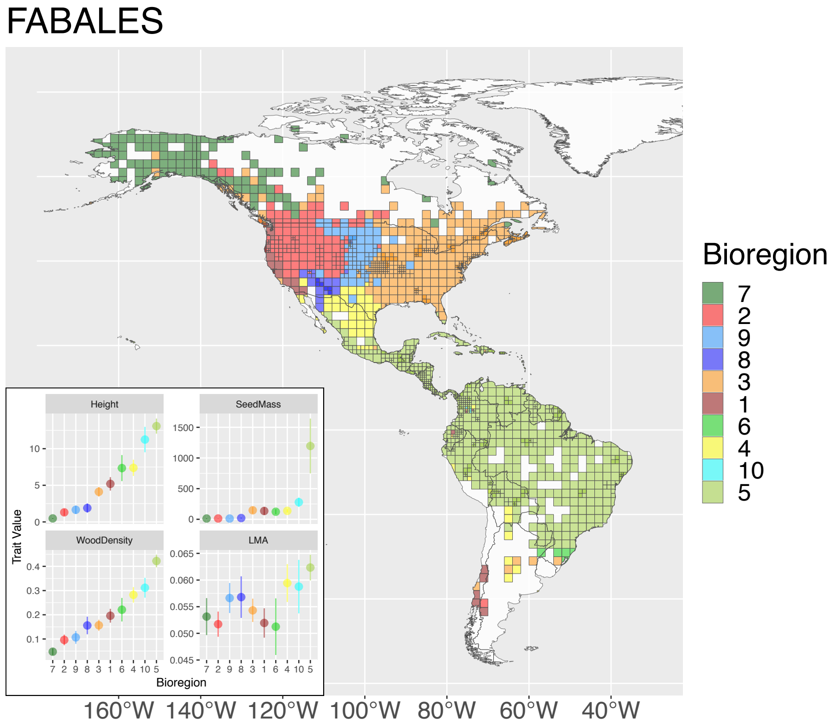


Figure 1.S6. Biogeographic clusters (bioregions) identified with the Infomap Bioregions algorithm (see Methods) for both Ericales (86 sp.) and Fabales (220 sp.) taxa with complete trait information. Bioregion colors and numbering were chosen arbitrarily by the algorithm for each plot; however, bioregions have been ordered in the legend and inset plots by increasing average height within that cluster. This generally corresponds also with increasing average seed mass and wood density, but not leaf mass per area (LMA). Inset plots show average (+/ SE) trait values within each cluster for each functional trait, as indicated. Both Ericales and Fabales show similar numbers of bioregions, and these also generally map to similar geographic areas. Bioregions tend to show consistent trait patterns, although there is a sharp discontinuity within both clades for seed mass, with tropical (extra-tropical) clusters showing much greater (lower) seed mass. Clade-specific differences in trait values associated with bioregions are also observed. Notably, Fabales taxa group primarily into a single tropical cluster, whereas Ericales do not.

Table 1.S1. InfoMap bioregions parameters.

	Cell size (deg.)	*Number of species per cell*
Max	2	100
Min	0.5	5

Clustering cost was left at the default value of 1. Number of clustering iterations: 10.

Chapter II

Within-Clade Niche Similarity is at Least as Great as Between-Clade Niche Differences for American Angiosperms**

CHAPTER IN CONTEXT

In this chapter, I use a large, macroecological niche model dataset to describe the temperature, precipitation, and elevation niches of 72,372 American seed plants. With this dataset, I reframe the perspective that phylogenetic niche conservatism and niche shifts are mutually exclusive processes shaping modern biodiversity patterns by instead assuming that both play major roles and I quantify the amount of niche similarity among New World seed plants' abiotic niches. I employ a conservative statistical methodology to identify statistically significant niche shifts between species at multiple phylogenetic scales. I found that within-clade niche differences were at least as great as between-clade niche differences, suggesting that even closely related species exhibit statistically significant niche shifts. In particular, precipitation niches were more conserved than elevation, and niche conservatism dominated at smaller phylogenetic scales whereas plant taxonomic orders may be overdispersed.

This work makes two important contributions to the literature. First, I analyze the observed degree of niche similarity within the broader phylogenetic context of American seed plants, which allows me to assess how similar closely related species are in their abiotic niches, regardless of whether those species co-occur at a particular site. This is a more appropriate test to evaluate phylogenetic niche conservatism, as PNC would predict that niche similarity would be greatest for closely related species, regardless of their geographic proximity. Consistent with

---

**A version of this chapter was originally submitted to *New Phytologist*.

these expectations, I found that the average niche similarity decreased with increasing phylogenetic distance.

Second, the broad sampling of this dataset permits an evaluation of the variance in niche similarity among closely related species. This is also crucial for understanding to what extent niche conservatism is present among American seed plants as the variance in niche similarity for a given phylogenetic scale indicates the extent to which niche shifts have taken place. In this study, I found that within-clade variance in niche similarity was at least as great as between-clade niche differences. Therefore, for example, the niche differences between all Ericales were as great as those among all seed plants. These results show that niche shifts are a widespread feature across the American flora, and that individual clades tend to occupy large and overlapping regions of climatic niche space.

Within the context of this dissertation, this chapter illustrates that neither niche conservatism nor niche shifts should be viewed as mutually exclusive processes acting to shape modern biodiversity patterns. Both processes contribute, and it is possible to quantify the degree to which each one contributes to particular lineages, at different phylogenetic scales, and for different niche variables. Further, the extent to which shifts in climatic niches occur can potentially be tied to questions of ‘lability’ and ‘evolvability’, similar to inferences made for the functional traits in **CHAPTER ONE**.

ABSTRACT

Phylogenetic niche conservatism is a primary force shaping modern biodiversity patterns. Nevertheless, niche shift must occur for plants to radiate into novel environmental conditions. Thus, although niche shift and niche conservatism are often framed as mutually exclusive, both processes could contribute to contemporary biodiversity patterns. Here, we employ novel methodologies in order to quantify the degree of climatic niche conservatism among American seed plants with regards to mean annual temperature (MAT) and precipitation (MAP) and analyze how climatic niche breadth (*B*) varies along elevational gradients. To incorporate the biological reality that species experience varied abiotic conditions across their range, we constructed niche models for each species and used these to characterize temperature, precipitation, and elevation niches as continuously valued distributions. We then quantified niche similarity (distributional overlap) and identified statistically significant differences compared to a randomized null. The degree of niche similarity differed among climate variables, plant lineages, and at different phylogenetic scales. For example, ~17% of all seed plants were significantly different in elevational niche from their closest relative(s), whereas for precipitation, this value was only ~4%. Average niche similarity decreased with increasing phylogenetic distance, consistent with niche conservatism; however, variance in niche similarity among close relatives was large, such that there always existed niche differences equaling those among distantly related species. Our results demonstrated that within-clade niche differences were at least as great as between-clade differences. This suggests that researchers should incorporate both niche shift and conservatism as important, scale-dependent factors shaping biodiversity patterns as these processes are not mutually exclusive, nor do they contribute equally to patterns among different plant lineages or niche variables.

INTRODUCTION

Abiotic conditions such as temperature, precipitation, and elevation change over evolutionary time (Nie et al., 2017; Zachos et al., 2001), providing potential selective pressures on plant lineages that might have influenced their modern distributions (Donoghue, 2008). Broadly, there are two strategies a plant lineage might employ when faced with climate changes, such as decreased temperature (whether from mountain uplift or global trends). One approach is to track suitable climates as they shift in geographic distribution, thereby retaining climatic niches. This is the scenario of phylogenetic niche conservatism (Wiens et al., 2010), which predicts that lineages tend to maintain climatic tolerances and niche breadth as diversification occurs. A second possibility is that the lineage will adapt to the changing conditions, employing new strategies to deal with changes in climatic variables or utilizing existing phenotypes in new ways that enhance survival. Under this scenario, niche shift occurs, and lineages evolve to tolerate new climatic conditions.

Undoubtedly, both strategies have been employed, to varying extents, by different plant lineages over the course of their evolution. Nevertheless, phylogenetic niche conservatism has emerged as an important driver of modern biodiversity patterns, particularly at larger biogeographic scales (Crisp et al., 2009; Gagnon et al., 2019; Segovia et al., 2020; Webb, 2000;

Wiens et al., 2010). For example, Gagnon et al. (2019) demonstrated that Succulent Biome members of the Caesalpinia Group (Leguminosae), which included trees, shrubs, and lianas, showed strong evidence of niche conservatism even across transcontinental disjunctions. Recently, Segovia et al. (2020) used a genus-level phylogeny of angiosperm trees to show that North American lineages could be partitioned into four, phylogenetically conserved, climatically defined types. They concluded that the phylogenetic conservatism of temperature and precipitation niche is the primary driver of North American tree diversity. Despite the clear prevalence of niche conservatism, it is also true that plants do radiate into novel climatic space and biome-switching occurs (e.g. Donoghue & Edwards, 2014). Switches from tropical to temperate environments have taken place (Judd et al., 1994), and temperate clades are often nested within tropical ones (Hawkins et al., 2011; Kerkhoff et al., 2014).

Rather than framing the issue of niche conservatism versus shift as a mutually exclusive dichotomy, it could be useful to quantify the extent to which such niche shifts have occurred across different seed plant lineages, and to what extent those shifts have helped to shape present-day patterns of biodiversity. For example, Crisp et al. (2009) investigated biome switching among 11,000 Southern Hemisphere vascular plant species and concluded that biome switching was relatively rare, occurring at approximately a 1:25 ratio against niche conservatism. If such tests are generalizable across all plant lineages, they might offer a window into the expected rate and prevalence of biome switching during the evolutionary history of seed plants.

Here, we quantify the degree of climatic niche similarity among American seed plant lineages and investigate how species found at different elevations compare in terms of temperature (B_{TEMP}) and precipitation (B_{PREC}) niche breadth. Although numerous climatic variables influence how plants are distributed, temperature and precipitation are generally considered among the most important (Clarke & Gaston, 2006; Macarthur, 1972; Whittaker, 1970). Temperature (and its correlate, latitude) influences processes such as plant growth and metabolic rates (Körner, 2003; Loveys et al., 2002). Mean annual temperature (MAT), in particular, has shown quantitative correlations with ecologically relevant plant traits (Moles et al., 2014), suggesting it may be an important determinant of ecological strategies. Mean annual precipitation (MAP) is important for its relation to drought tolerance (Craine et al., 2013; Jamieson et al., 2012; Xie et al., 2015), and changes in global precipitation and temperature have jointly helped to shape the development of modern American plant biomes, such as the emergence of arid grasslands (Graham, 2011). Elevation (ELV) also influences diversity patterns (Antonelli et al., 2009, 2018; Hughes & Atchison, 2015; Rahbek, Borregaard, Colwell, et al., 2019), and high-elevation habitats are among the most diverse in the world (e.g. Ding et al., 2020; Hughes & Eastwood, 2006; Muellner-Riehl et al., 2019; Rahbek, Borregaard, Antonelli, et al., 2019; Wen et al., 2014; Xing & Ree, 2017).

Mountain range uplift creates substantial topographic heterogeneity, providing a wide variety of microclimatic niche space in which plants can become established (Körner, 2003), as well as opportunities for isolation and allopatric speciation, which can promote diversity. Mountain ranges further offer high-elevation corridors for long-range dispersal (Antonelli et al., 2009; Rahbek, Borregaard, Antonelli, et al., 2019). Such routes have potentially exposed plants to freezing temperatures prior to the onset of global cooling (Hawkins et al., 2011, 2014; Qian, 2017). Because tolerance to freezing appears to be a major driver of the biogeographic distribution of plant lineages (Folk et al., 2020; Hawkins et al., 2014; Qian, 2017; Qian & Chen, 2016; Segovia et al., 2020; Sun et al., 2020; Zanne et al., 2014), understanding the relationship between adaptation to high elevation and adaptation to life in the cold could provide crucial insight into the factors shaping modern plant diversity. Further, the American mountain ranges run north-south in an almost unbroken line from one pole to another. This orientation could allow plants to track favorable climate during cycles of glaciation in the Americas, unlike the situation for European mountains, which are oriented east-west, and therefore significantly obstruct plant dispersal (Bennett et al., 1991; Rahbek, Borregaard, Antonelli, et al., 2019). Thus, during Quaternary glaciation cycles, American plants have potentially had greater opportunity to maintain their climatic niches. Therefore, we expect that the prevalence of niche shifts among American seed plants will be less influenced by habitat loss and extinctions during glacial cycles.

To quantify the degree of climatic niche similarity among montane and alpine plant lineages and compare the climatic niche breadth of taxa found at different elevations, we assembled a large-scale dataset of species distribution models (SDMs) describing the climatic and elevational niches of the ~70,000 seed plants known to occur across the Americas along with their phylogenetic relationships. Our taxonomically broad approach also allowed us to draw comparisons among gymnosperms and major clades of angiosperms (e.g., monocots and eudicots), groups whose divergent evolutionary histories might conceivably offer quite disparate pictures of the importance of niche conservatism, as well as dissecting variation at finer scales.

With this dataset, we examine five climatic variables (MAT, MAP, ELV, B_{TEMP}, and B_{PREC}) and ask the following questions: (1) Do we observe phylogenetic conservatism in climatic niche? (2) How prevalent are climatic niche shifts among American seed plants? (3) Are species found at high elevations climate generalists or specialists?

METHODS

Phylogeny, occurrence records, and climatic niche models.

We obtained a dated phylogeny for all seed plants from Smith & Brown (2018) (ALLMB phylogeny) and left polytomies unresolved. This phylogeny generated a species list with which to query American occurrence records from the Global Biodiversity Information Facility (GBIF; www.gbif.org) and Integrated Digitized Biocollections (iDigBio; www.idigbio.org). Records were then cleaned and filtered using the BiotaPhy Platform interface (https://biotaphy.github.io/, Soltis & Soltis, 2016), following their accepted best practices.

The full GBIF dataset ($N_{records}$=36,335,199) is described and accessible at https://doi.org/10.15468/dl.gtgtt5, as well as the **Online Supplement**. Briefly, GBIF records with the following flags were removed: TAXON-MATCH_FUZZY, TAXON_MATCH_HIGHER_RANK, TAXON_MATCH_NONE. Further processing was performed after aggregating GBIF and iDigBio records. For iDigBio, data cleaning and filtering produced a dataset of 13,667,523 records ($N_{initial}$=58,384,427; 23.4% retained). Briefly, initial records were filtered by removing those with any of the following flags: GEOPOINT_DATUM_MISSING, GEOPOINT_BOUNDS, GEOPOINT_DATUM_ERROR, GEOPOINT_SIMILAR_COORD, REV_GEOCODE_MISMATCH, REV_GEOCODE_FAILURE, GEOPOINT_0_COORD, TAXON_MATCH_FAILED, DWC_KINGDOM_SUSPECT, DWC_TAXONRANK_INVALID, DWC_TAXONRANK_REMOVED (see **Online Supplement** for full details).

Aggregated GBIF and iDigBio records were then further processed by excluding points with any of the following issues: (1) falling outside the study area (the Americas); (2) less than four decimal point precision (~11 m at the equator); (3) duplicate localities (rarefaction); (4) falling outside polygons describing accepted species' distributions (defined by Plants of the World Online, POWO, www.plantsoftheworldonline.org; Brummitt 2001; www.github.com/tdwg/wgsrpd); (5) species with fewer than twelve records (in order to build reliable niche models).

Cleaned records were then passed to MaxEnt (version 3.1.4; www.github.com/mrmaxent/Maxent; Phillips et al., 2006, 2004) along with 2.5’ resolution climate data from WorldClim (www.worldclim.org; Fick & Hijmans, 2017) in order to build species distribution models (SDMs; climate layers described in the **Online Supplement**). We chose to perform our analyses using SDMs rather than point occurrence records for two reasons. SDMs offer a probabilistic way of describing expected species’ ranges based on the climate from sites where the species has been observed. In this way, SDMs convert presence/ absence data into a continuously valued function, allowing us to ask how distributions are impacted by abiotic factors without having to arbitrarily bin species, as for example, alpine or montane. Second, using SDMs helps overcome some sampling limitations by providing insight into the climatic tolerances of where species might occur, even if they have not been sampled at that precise location. Although this could lead to erroneously predicting, for example, that a northern boreal species should occur at extreme southern latitudes, we overcame this obstacle by masking the SDMs with polygons provided by POWO that define geographically broad areas where each species occurs based on expert assessments. This approach thus constrained SDMs by both known areas of occurrence and climatic tolerances.

Calculating niche vectors.

We used the SDMs described above to characterize climatic niches of each species with respect to mean annual temperature (MAT), precipitation (MAP), and montane bioclimatic zone (ELV; GMBA, 2010). For MAT and MAP, we downloaded 2.5’ resolution WorldClim climate data. For ELV, we used the seven bioclimatic zones defined and made publicly available by the Global Mountain Biodiversity Assessment (GMBA, 2010; Körner et al., 2011, 2017; Körner & Paulsen, 2004). These zones integrate temperature, growing season length, and topographic information to provide a robust and biogeographically relevant assessment of the extent and type of montane and alpine habitat across the globe. They further permit an assignment of species distributions along elevational gradients despite latitudinal differences in what constitutes montane and alpine areas. Nevertheless, these bioclimatic categories are fundamentally defined by isothermal zones, and not therefore an independent axis from MAT (see Table 2 in Körner et al. (2011) for precise climatic zone definitions). However, our goal was not to define independent niche axes *per se*, but rather to characterize climatic niches in a biologically meaningful way. The GMBA-defined zones provide the current best estimate of a biogeographically relevant classification of mountain systems (Körner et al., 2011, 2017; Körner & Paulsen, 2004).

The seven GMBA bioclimatic zones used to define species' ELV niche were: 1—nival (perpetual snow); 2—upper alpine; 3—lower alpine; 4—upper montane; 5—lower montane; 6—mountain slope with frost; 7—mountain slope without frost or lowland. We chose to combine GMBA zone 7 with lowland areas for our study; thus, our assessment of lowland distributions was fundamentally linked with exposure to freezing temperatures.

For each climatic variable (MAT, MAP, or ELV), we calculated the fraction of SDM attributable to an increment of that variable. Increments were 1° C for MAT and 250 mm for MAP (ELV zones are already discretized, and SDMs were parsed into each zone). This method yielded three vectors that jointly quantified the distribution of niche space potentially occupied by each species (e.g., 10% probability a species will be present at a site with MAP of 500 mm). Figure 1 plots some representative examples and provides a conceptual illustration of our characterization of climatic niches. In this way, we avoided having to arbitrarily bin species into habitat categories, such as alpine or lowland, and instead could assess niches as continuously valued distributions of climatic space occupied by each species. Parsing of SDMs was performed using custom Python scripts (available in the **Online Supplement**).

Quantifying niche breadth.

Using the climatic niche vectors described above, we quantified the temperature and precipitation niche breadth of each species as follows, using custom scripts in the R statistical computing environment (R Core Team, 2020, v.4.0.2; scripts available in the **Online Supplement**). Having partitioned the SDMs among small increments of each climatic variable, we interpreted these climatic vectors as representing states a species could occupy within its range. We used Levins' (1968) niche breadth index to quantify SDM uniformity across the full suite of possible climatic states for either MAT or MAP using:

$$B_{i,c} = \frac{1}{\sum_{k}^{n} p_k^2}$$

Where $B_{i,c}$ is the niche breadth of species i with respect to climatic variable c, and p is the proportion of that species' SDM found in state k of the climatic variable (e.g., within each 1° C increment of temperature or 250 mm increment of precipitation). To facilitate comparison across species and different niche variables, we calculated a standardized niche breadth, $\hat{B}_{i,c}$, bound between zero and one, by dividing niche breadth by the number of possible states (n), corrected for the finite number of states:

$$\hat{B}_{i,c} = \frac{B_{i,c} - 1}{n - 1}$$

Phylogenetic niche conservatism and niche shift.

To quantify the similarity between climatic niche vectors of different species, we used their Pearson correlation (bounded between -1 and +1). In contrast, for niche breadth, which we measured as a scalar quantity, we used one minus the difference between pairwise niche breadths (bounded between 0 and +1). We calculated these indices among species at four different phylogenetic scales: closest relatives (nearest neighbors; NNs); congeneric species; species within four focal clades (see below); and among randomly sampled species from the entire seed plant phylogeny (null distribution, see below).

To form a null distribution of the expected degree of niche similarity, we randomly sampled 5,000 species from the seed plant phylogeny and calculated average similarity among them. We performed 500 such random samples, independently, for each climatic variable. The sample size of 5,000 species was chosen because further increases in sample size did not alter the resulting null distribution (data not shown). We used this null expectation as a benchmark of niche shift, interpreting similarity scores below the bottom 5% quantile of the null distribution as significantly more different (i.e., over-dispersed) in climatic niche than expected by chance (conditioned on the dataset).

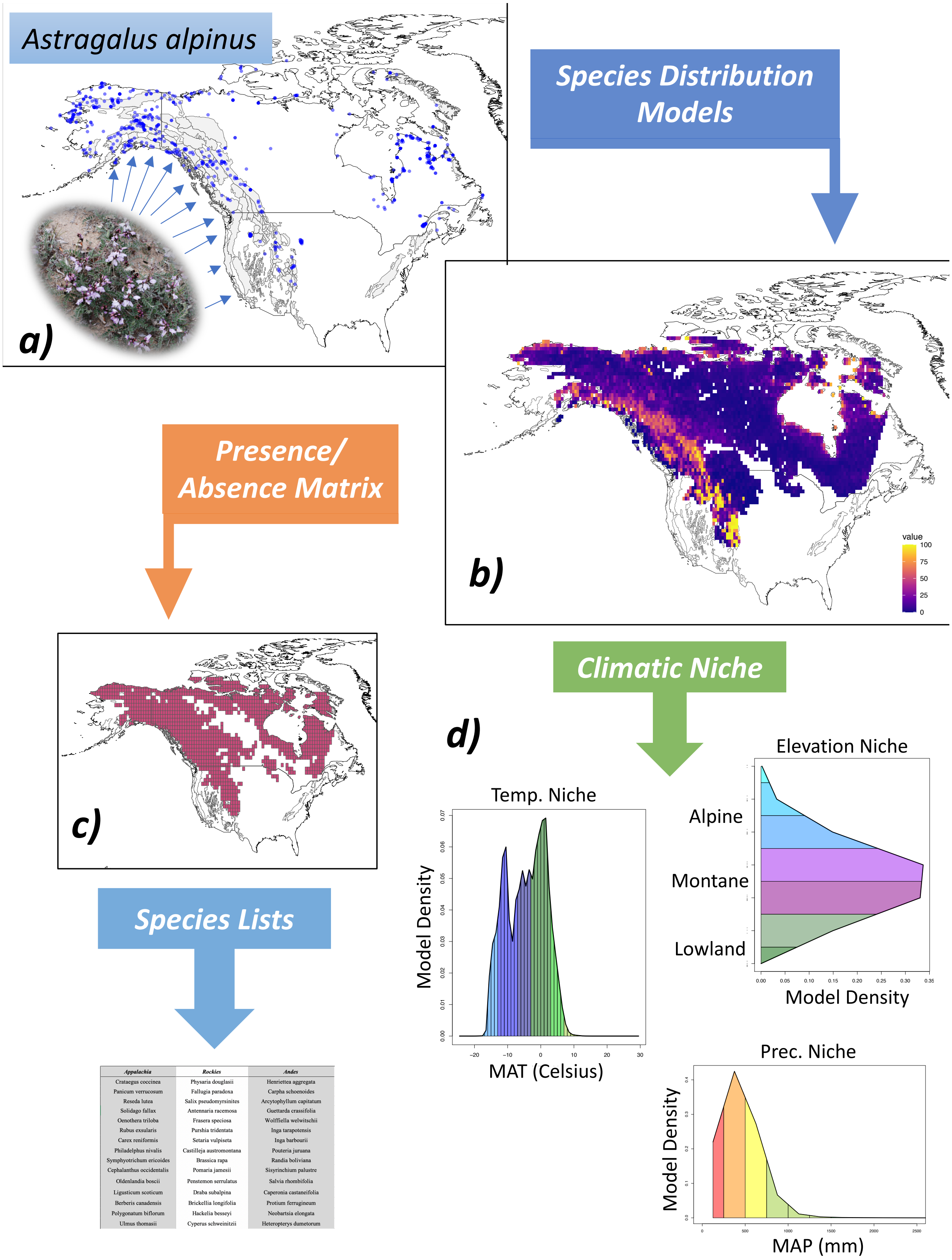
Our analyses emphasized comparisons between closely related taxa (NNs and congeners) for two reasons. First, we observed a steep decay in niche similarity with increasing phylogenetic distance (see Results), indicating that comparisons across extremely disparate groups might be of limited utility. Second, genus-level phylogenies have been used recently to demonstrate the importance of phylogenetic niche conservatism in shaping modern biodiversity (e.g., Segovia et al., 2020), carrying the implicit assumption that within-genus climatic niche variation is negligible. By focusing our analyses on closely related species, we emphasize the extent to which even members of the same genus can differ in climatic niche. Nevertheless, our analyses incorporate multiple scales in order to provide a nuanced description of how phylogenetic scale impacts the observed degree of niche similarity.

In addition to analyses involving the entire set of seed plants, we chose to investigate potential clade-specific differences by focusing on four major plant groups: gymnosperms, a monocot lineage, Poales, and two eudicot clades, Ericales (asterids), and Fabales (rosids). These groups were chosen because they represent several large and climatically widespread lineages found in numerous habitat types and represent a large portion of the total seed plant diversity. For each of these focal groups, we used the tree_subset function of the 'TREEIO' (Wang et al., 2020) package in R to quantify the relationship between climatic niche similarity for MAT, MAP, and ELV, and phylogenetic distance (measured as node depth). Briefly, for each focal species in a clade, we took the subset tree of all tips within *n* nodes of that focal species and calculated the average niche similarity score (similarity indices described above) among all pairwise species comparisons. We varied node depth from 1 to 65, though beyond ~20 nodes further increase in phylogenetic distance did not alter the average similarity scores obtained. For gymnosperms ($N_{species}$=301), we performed calculations for every species. However, for the angiosperm clades (range: 2790 to 6638 species), we performed calculations on a random sample of 1000 species for each clade. All calculations were performed using custom scripts in R, and results visualized with 'GGPLOT' (Wickham, 2016) and 'GGTREE' (Yu et al., 2017) packages. Disparity-Through-Time (DTT) plots were calculated with the disparity function in the 'GEIGER' package (Harmon et al. 2008), with some customization to handle the vectorized form of climatic niches in our dataset.

Assigning species lists to montane regions

GMBA (2010) provides shapefiles delimiting each mountain in the Americas, as well as larger shapefiles defining the greater American mountain ranges (e.g., Appalachians or the Andes). We compiled a first approximation of the potential seed plant species lists for all such mountains and mountain ranges with our dataset. To do this, we first parsed the Americas into half-degree grid cells and tabulated whether or not each species in our dataset was predicted to be present within each grid cell. Presence was defined by the SDM predicting at least a 5%

chance of finding the species within at least 25% of the area covered by the grid cell. These parameters, as implemented in the BiotaPhy Platform interface (https://biotaphy.github.io/, Soltis & Soltis, 2016), are completely flexible, and future users can modify them to address more specific questions of interest. Having thus obtained presence/ absence grids for each species, we then tabulated which species were recorded as being present in grid cells falling within each mountain and mountain range shape file. We also recorded which countries were associated with the species present in each mountain/ mountain range in order to permit questions of predicted co-occurrence to be asked (e.g., one could easily subset the list to ask which species are predicted to occur in the Colombian Andes). By summing the number of species predicted to occur in each grid cell, we also obtained predicted species richness values for all grid cells with sufficient data (**Figure 2.S1**). Expected species lists, cross-referenced by country, are available in the **Online Supplement**.



Appalachia	***Rockies***	***Andes***
Crataegus coccinea	Physaria douglasii	Henriettea aggregata
Panicum verrucosum	Fallugia paradoxa	Carpha schoenoides
Reseda lutea	Salix pseudomyrsinites	Arcytophyllum capitatum
Solidago fallax	Antennaria racemosa	Guettarda crassifolia
Oenothera triloba	Frasera speciosa	Wolffiella welwitschii
Rubus exsularis	Purshia tridentata	Inga tarapotensis
Carex reniformis	Setaria vulpiseta	Inga barbourii
Philadelphus nivalis	Castilleja austromontana	Pouteria juruana
Symphyotrichum ericoides	Brassica rapa	Randia boliviana
Cephalanthus occidentalis	Pomaria jamesii	Sisyrinchium palustre
Oldenlandia boscii	Penstemon serrulatus	Salvia rhombifolia
Ligusticum scoticum	Draba subalpina	Caperonia castaneifolia
Berberis canadensis	Brickellia longifolia	Protium ferrugineum
Polygonatum biflorum	Hackelia besseyi	Neobartsia elongata
Ulmus thomasii	Cyperus schweinitzii	Heteropterys dumetorum

Figure 2.1. Conceptual illustration of project workflow and niche characterization, using *Astragalus alpinus* (photo in *A*) as an example. (*A)* We began by aggregating georeferenced occurrence records from the Global Biodiversity Information Facility (GBIF) & Integrated Digitized Biocollections (iDigBio). Here, GBIF records within the study area for *A. alpinus* are shown in blue. North American mountain ranges are highlighted in grey. These records were then integrated with climate data from WorldClim, such as mean annual temperature, precipitation, elevation, and slope, to build species distribution models (SDMs) using MaxEnt (see Methods); (*B)* SDMs provided a probabilistic view of where we expected species to occur based on their known distributions. Each grid cell was assigned a score representing how likely one is to observe that species in that grid cell, given the climatic conditions where it is known to occur. In this example, warmer (cooler) colors represent higher (lower) probability of *A. alpinus* occurring at that site (this species primarily occurs in and around higher elevation habitat in the Northern Hemisphere). The SDMs were further parsed and analyzed in two different ways, as follows: (*C)* We assigned a presence/absence rating to each grid cell based on the SDM in order to predict where *A. alpinus* might occur. A grid cell was assigned a rating of "present" if the SDM score for that cell was at least 5% across at least 25% of the area covered by that grid cell. These presence/ absence matrices allowed us to compile expected species lists for each mountain and mountain range in the New World (only a small sample of the species' lists obtained are shown here for illustrative purposes; full species lists for each mountain and mountain range, cross-referenced by country are available in the **Online Supplement**), and to calculate expected species richness across different mountain ranges (**Figure 2.S1**). (*D)* Separately from (*C*), we also parsed each SDM by three climatic variables—temperature, precipitation, and elevation—in order to build a quantitative, continuously-valued description of each species realized climatic niche space (colors as in **Figure 2.2** and **Figure 2.S2**). In this case, the SDM of *A. alpinus* describes a range centered within montane habitat that is bounded between approx. -17° and +10° Celsius and generally below 1000 mm of precipitation. For temperature and precipitation, the horizontal axes in each plot indicate the value of the climatic variable and the vertical axis is the fraction of the SDM attributable to that value of the climatic variable. For elevation, axis orientation is switched, and the vertical axis represents the elevational category (i.e., higher elevations are also higher in the plot; alpine areas in blues; montane in purples; and slope or lowland in greens) while the horizontal axis is the SDM fraction falling into each category. In this way, the climatic niche of each species in our data set was described by three climate vectors. For most analyses, we used these vectors as "traits" for each species, rather than scalar values, such as the mean or median of each vector. Niche similarity, for example, was quantified as the degree of correlation between niche vectors.

RESULTS

Dataset construction, description of climatic niches, and geographic coverage.

Our data cleaning and filtering methods produced a dataset of 50,002,722 georeferenced occurrence records spanning 72,372 seed plant species, of which 68,241 (5397 genera; ~18% of total estimated seed plant lineages; Stevens, 2001, onwards) could be matched back to our phylogeny (Smith & Brown, 2018), with excellent geographic coverage of all American mountains and most of the American region in general (**Figure 2.S1**). There were notable geographic areas of poor sampling, however, including the Amazon basin and extreme southern and northern latitudes. We also used our dataset to compile potential species lists describing the expected composition within each mountain and mountain range, cross-referenced by country (available with the **Online Supplement**).

From these occurrence records, we built species distribution models (SDMs) to characterize niches with respect to mean annual temperature (MAT), precipitation (MAP), and elevation (ELV; see Methods for details on niche calculations). Importantly, these niches were characterized continuously and not as scalar values. **Figure 2.1** illustrates an example of this niche characterization for a single species, *Astragalus alpinus*, whose composite niche is centered within montane habitat (i.e., >50% of the modelled range falls within montane area) and is bounded between approximately -17° and +10° C, and generally below 1000 mm of precipitation. To assess processes such as niche conservatism and niche shift, we quantified the degree of niche similarity between species' niches by measuring the correlation in these niche distributions and compared the observed similarity to a null expectation.

Assessing phylogenetic conservatism of climatic niches.

Having thus characterized the distribution of climatic niche space for all species in our dataset, we then placed those results in their phylogenetic context to examine the correlation between niche similarity and relatedness (i.e., phylogenetic niche conservatism). **Figure 2.2** provides an illustrative example, wherein the phylogenetic relationships among 50 randomly selected taxa are shown along with each of their climatic niches. We do not provide an illustration for the climatic niches of all 68,241 species in the dataset; however, **Figure 2.S2** highlights the phylogenetic distribution of climatic niche space for four major plant groups: gymnosperms ($N_{species}$=301; 28%), two eudicot lineages (Ericales [$N_{species}$=2790; 23%], and Fabales [$N_{species}$=5579; 27%]), and one monocot lineage (Poales [$N_{species}$=6638; 35%]). (Percents here indicate the fraction of total estimated species in each lineage as reported in Stevens, 2001.)

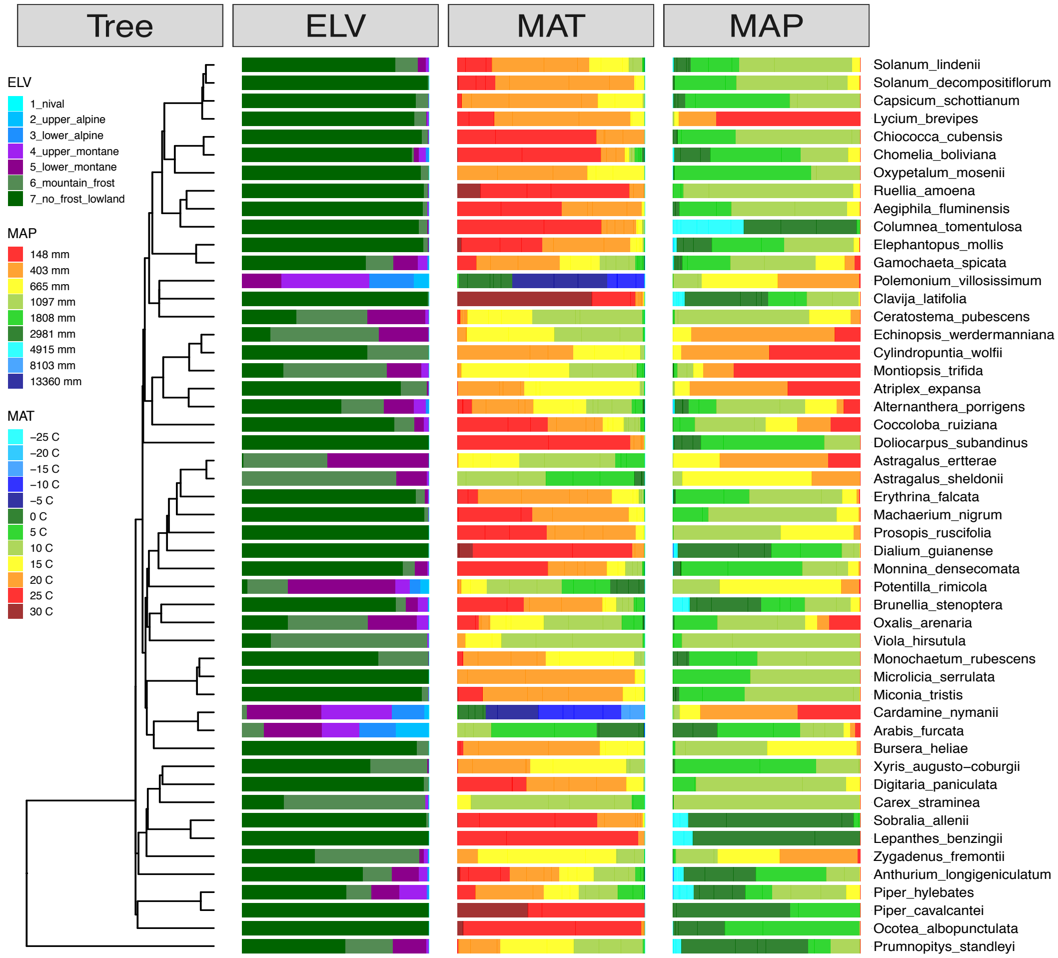


Figure 2.2. Climatic niches for 50 randomly selected seed plants in their phylogenetic context (*panel TREE*). For *ELV* niches, colors indicate the fraction of that species' range falling into each of the 7 bioclimatic zones defined by the Global Mountain Biodiversity Assessment (GMBA). For *MAT* niches, colors indicate the fraction of that species' range falling within a 1° Celsius increment of temperature, with warmer (cooler) colors indicating hotter (colder) temperatures. For *MAP* niches, colors indicate the fraction of that species' range falling within a 250 mm increment of precipitation, with warmer (cooler) colors indicating less (more) rainfall. The continuously valued characterization of climatic niches in this way allows us to incorporate the biological reality that most species experience varied climatic conditions along their range. It also permits categorization of species, for example, as alpine specialists or generalists based on what habitat their range is centered upon. Among this random subset, most species are primarily lowland (i.e., in the *ELV* panel the majority of each bar is usually dark green), with species' ranges bound between 15°-25° C mean annual temperature (MAT), and >1000 mm mean annual precipitation (MAP). There are ~5 species in this subset with significant fractions of their range found at higher elevations (purple and blue colors in *ELV* panel), and these are associated with decreased temperature and precipitation. **Figure 2.S2** illustrates the phylogenetic distribution of climatic niches within four larger clades (gymnosperms, Poales, Ericales, and Fabales).

Within each of these four focal clades, average climatic niche similarity decreased with decreasing relatedness (**Figure 2.3a**), measured by node depth, consistent with a model of phylogenetic niche conservatism. Disparity-Through-Time (DTT) plots (**Figure 2.3b**) provided more nuanced results. Average MAP niche similarity declined with time for all four groups (though with some increase at ~50 MYA). However, ELV niche similarity showed clade-specific variation, with a sharp discontinuity for Fabales at ~50 MYA, perhaps suggesting a single radiation into higher-elevation habitat, and undulating patterns for Ericales (periodicity ~45 MY) and Poales (periodicity ~5 MY), consistent with multiple radiations into higher elevations for these two lineages (see Discussion).

When looking across all seed plants, climatic niches were significantly more similar than expected by chance both among closest relatives (i.e., nearest neighbors, NNs; **Figure 2.4a**), and congeners (**Figure 2.4b**; p<0.001 in all cases, Wilcoxon signed-rank test). At the same time, however, for each climatic variable, a substantial fraction of NNs and congeners were actually *less* similar than expected by chance, potentially indicating niche shifts had occurred (see below).

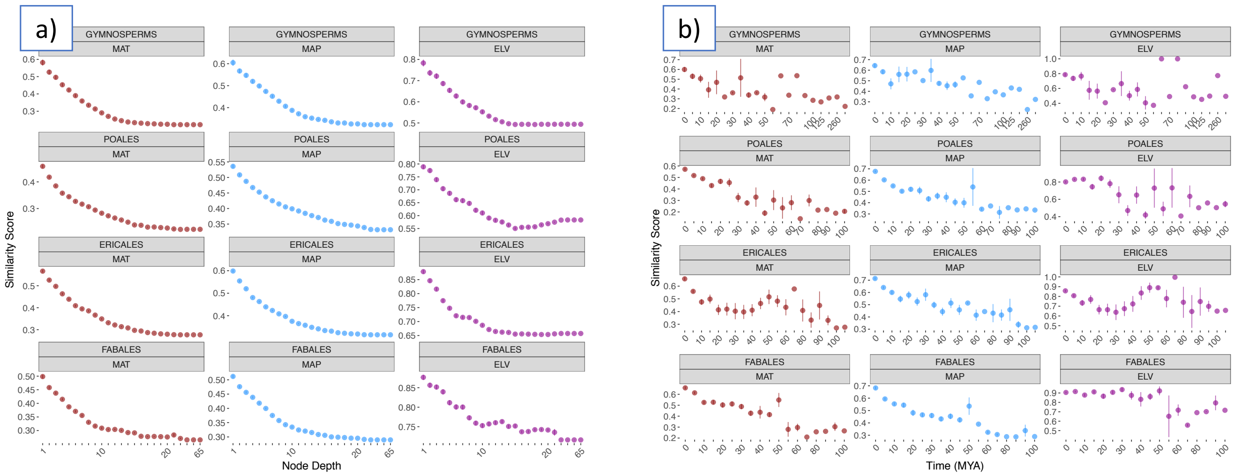


Figure 2.3. Disparity through time plots for abiotic niches. (*A*) Among gymnosperms and the angiosperm lineages, Poales (monocots), Ericales, and Fabales (both eudicots), average niche similarity scores for temperature (MAT), precipitation (MAP), and elevation (ELV) decreased with decreasing relatedness (measured here as node depth), consistent with a model of phylogenetic niche conservatism. Beyond a depth of ~20 nodes, further increases in phylogenetic distance did not change the observed similarity. Calculations were performed up to 65 nodes. The ELV niche of Poales showed a slight departure from this pattern, reaching a minimum similarity at ~15 nodes and then increasing somewhat before plateauing. (*B*) Disparity through time plots of climatic niche similarity for each focal group, as in (*A*). MAP shows a general decrease in niche

similarity with time, however other niche-clade combinations show more nuanced results. For example, ELV niche similarity showed clade-specific variation, with a sharp discontinuity for Fabales at ~50 MYA, perhaps suggesting a single radiation into higher elevation habitat, and undulating patterns for Ericales (periodicity ~45 MYA) and Poales (periodicity ~5MYA), consistent with multiple radiations into higher elevations for these two lineages. *Circles indicate the average similarity score among all sampled species; lines are standard error.*

Climatic niche breadth across elevational gradients.

In addition to characterizing the climatic niches (MAT, MAP, and ELV) of each species, we also assessed their temperature (B_{TEMP}) and precipitation (B_{PREC}) niche breadth. The distributions of average, standardized, niche breadths (bounded between 0 and 1) across all American seed plants were generally narrow ($B_{TEMP} = 0.130 \pm 0.08$; $B_{PREC} = 0.099 \pm 0.08$), but with long tails, indicating much greater niche breadth for some taxa (**Figure 2.S3**). Relatively few species had both large B_{TEMP} and B_{PREC}.

We observed strong spatial patterning to B_{TEMP} and B_{PREC} (**Figure 2.5**). Across the Americas, B_{TEMP} was consistently higher in mountains than surrounding lowlands, and higher at greater absolute latitudes. In contrast, B_{PREC} was largest near the equator, regardless of elevation. Most mountains showed comparatively lower B_{PREC}, exceptions being Central American mountains, the Northern Andes, and the southernmost Andes. As a result, the Northern Andes are notable here for having broad average niche breadth in both temperature and precipitation.

We also evaluated the degree of climatic niche specialization across the elevational gradients of each species' modelled range by comparing species' niche breadths to the proportion of their SDM falling within alpine, montane, or lowland habitat. B_{TEMP} was maximal for species with an intermediate fraction of their SDM in a given elevational category (**Figure 2.6**, *top row*). For instance, species with ~50% of their SDM in montane regions exhibited greater B_{TEMP} than species found exclusively in either lowland or upland areas. For alpine taxa, maximum B_{TEMP} occurred for species with ~10-30% of their SDM in alpine regions. B_{PREC}, in contrast, was generally narrow regardless of elevation. However, species in lowland areas did have a greater range of B_{PREC} than alpine or primarily montane species (**Figure 2.6**, *bottom row*). There was some increase in median B_{PREC} for species with ~20% of their SDM in montane regions, but it was not significant.

The degree of alpine specialization also showed phylogenetic conservatism (**Figure 2.S4**). Genera containing alpine specialists (where 'specialist' is defined as having at least 60%

SDM in a given habitat type; Körner 2003) also had a greater proportion of alpine generalist and montane specialist species, compared to the degree of specialization found across all seed plants. A similar, but less extreme, pattern for montane specialists was also observed.

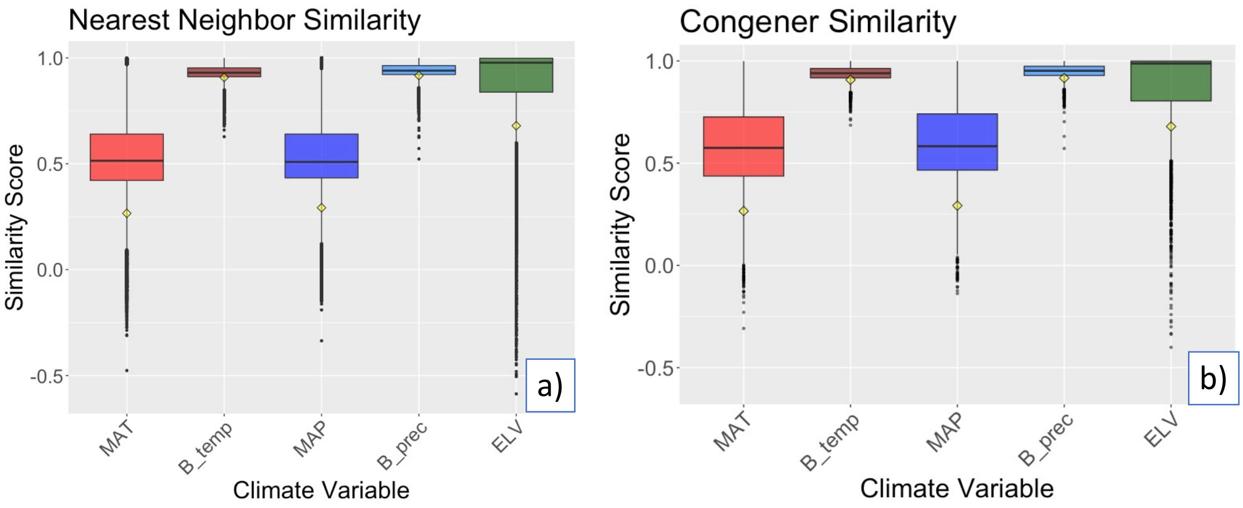


Figure 2.4. Abiotic niche similarity among close phylogenetic relatives. (*A*) Climatic niche similarity among closest relatives (nearest phylogenetic neighbors), for all five climatic variables (*B_temp* and *B_prec* indicate temperature and precipitation niche breadth, resp.). (*B*) Climatic niche similarity among congeneric species for all five climate variables. For both nearest neighbors and congeners, species were more similar in climatic niche than expected by chance. *Yellow diamonds* indicate the bottom 5% quantile of niche similarity scores among randomly sampled species pairs from across the entire seed plant phylogeny. Thus, points below yellow diamonds indicate close relatives or congeners that are *significantly more different* than expected by chance, indicating niche shifts have occurred (see **Table 2.1** for fraction of species in these categories). *(Similarity scores are either Pearson correlations (MAT, MAP, ELV) or one minus the scalar difference in niche breadth (B_temp* and *B_prec) among species.)*

Prevalence of climatic niche shift

We defined niche shift, for a given climatic variable, as two species being less similar in climatic niche than the bottom 5% quantile of the distribution formed by randomly sampled seed plants in our dataset. This method was deemed conservative because a classification of niche shift required that the species were significantly less similar than randomly chosen species pairs. This likely underestimated the extent of biologically relevant niche differences as this null was formed by comparing species from as disparate environments, for example, as alpine tundra vs. neotropical rainforests.

Applying this criterion among closest relatives, for each climatic variable (MAT, MAP, ELV, B_{TEMP}, B_{PREC}) we found that the majority of species fell within the same niche as their nearest phylogenetic neighbors (**Table 2.1**), indicating strong niche conservatism at these finer phylogenetic scales. Nevertheless, even among nearest neighbor species pairs, substantial fractions differed significantly for each climatic niche. For example, ~17% of nearest neighbors differed in ELV niche, and among the four focal clades, the extent of nearest neighbor ELV niche shift varied from 14.0% for Ericales to 27.2% for gymnosperms. MAP and MAT niche shifts, in contrast, were far less common (3.9% and 7.8%, respectively).

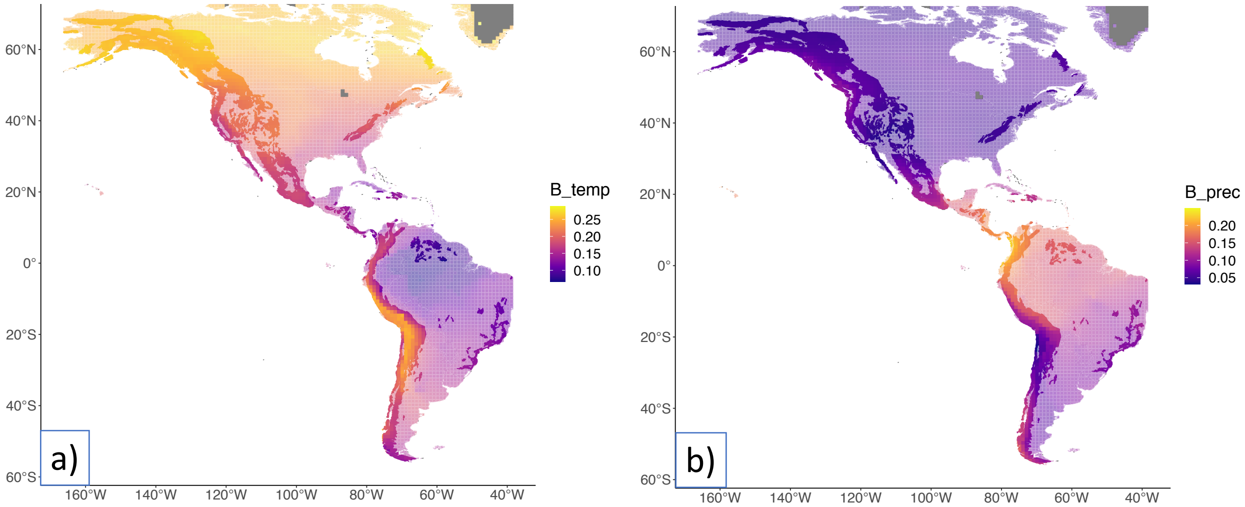


Figure 2.5. Heatmaps indicating the average niche breadth for temperature (*left; B_temp.*), or precipitation (*right; B_prec.*) in half degree grid cells across the study area. Cooler (warmer) colors indicate smaller (greater) niche breadth. *Foreground:* Mountain areas defined by the Global Mountain Biodiversity Assessment (GMBA). *Background:* non-mountainous areas in the sampling region. Grey cells indicate missing or insufficient data. Across the Americas, higher *absolute* latitudes tend to be occupied by species with greater temperature and lower precipitation niche breadth. Montane and alpine regions tend to have higher temperature niche breadth than surrounding lowland areas. This is especially apparent for the Andes mountains. The northern Andes, in particular, are also notable for being a hotspot wherein species tend to have *both* higher average temperature and precipitation niche breadths.

Among congeners, phylogenetic niche conservatism was also quite prevalent (**Table 2.1**). For MAT, 37.1% (N=1230) of multi-species genera (N=3312) demonstrated complete niche conservatism, meaning that all species within the genus were more similar in their temperature niche than expected by chance. At the other extreme, only 2.9% (N=95) of genera showed absolutely no niche conservatism, meaning all species differed from each other in temperature niche. However, of these genera, all had four or fewer species, and these therefore mostly represented one to three pairwise comparisons. The remaining genera (~60%, N=1987) showed at least some degree of niche shift among species. In other words, 60% of genera had at least some species which differed more in climatic niche than randomly chosen species pairs. For MAP, numerical results were nearly identical, again with those genera showing complete niche shift being dominated by clades with only 2 or 3 species. For ELV, by contrast, the vast majority of genera, 71.6% (N=2371), showed complete niche conservatism, with only 2.0% (N=65) showing complete niche shift (though, as with MAT and MAP, these genera typically had no more than 3 species).

Overall, therefore, the degree of observed niche shift varied with the phylogenetic scale used to make comparisons, with finer scales resulting in greater niche conservatism. In the **Online Supplement**, we provide a complete table of all genera along with their species richness, niche similarity scores, and the fraction of species showing significant niche shifts. There were some statistically significant correlations among the fractions of niche shift observed for each climatic variable (**Table 2.S1**); specifically, genera showing a high degree of MAT niche shift also showed higher prevalence of MAP or ELV shifts.

Clade-specific differences.

Finally, we asked whether or not the prevalence of climatic niche shifts differed among our four focal groups (gymnosperms, Poales, Ericales, and Fabales; **Figures 2.S5**, **2.S6**, and **Tables 2.1** and **2.S2**). Median similarity scores for all niches except ELV were higher for gymnosperms than the angiosperm groups. In contrast, median ELV niche similarity was lowest for gymnosperms (0.91) compared to the angiosperms (Poales: 0.93; Fabales: 0.98; Ericales: 0.98), though differences were slight. Statistically significant differences between groups are summarized in **Table 2.S2**.

Within each clade, species were, on average, significantly more similar in climatic niche to their closest relative(s) than species selected at random across the entire seed plant phylogeny (**Figure 2.S5**; $p<0.001$ in all cases, Wilcoxon signed-rank test). Nevertheless, substantial fractions of nearest-neighbor pairs within each group showed less similarity than expected by chance (**Table 2.1**), particularly for ELV niches and niche breadth metrics, indicating niche shifts.

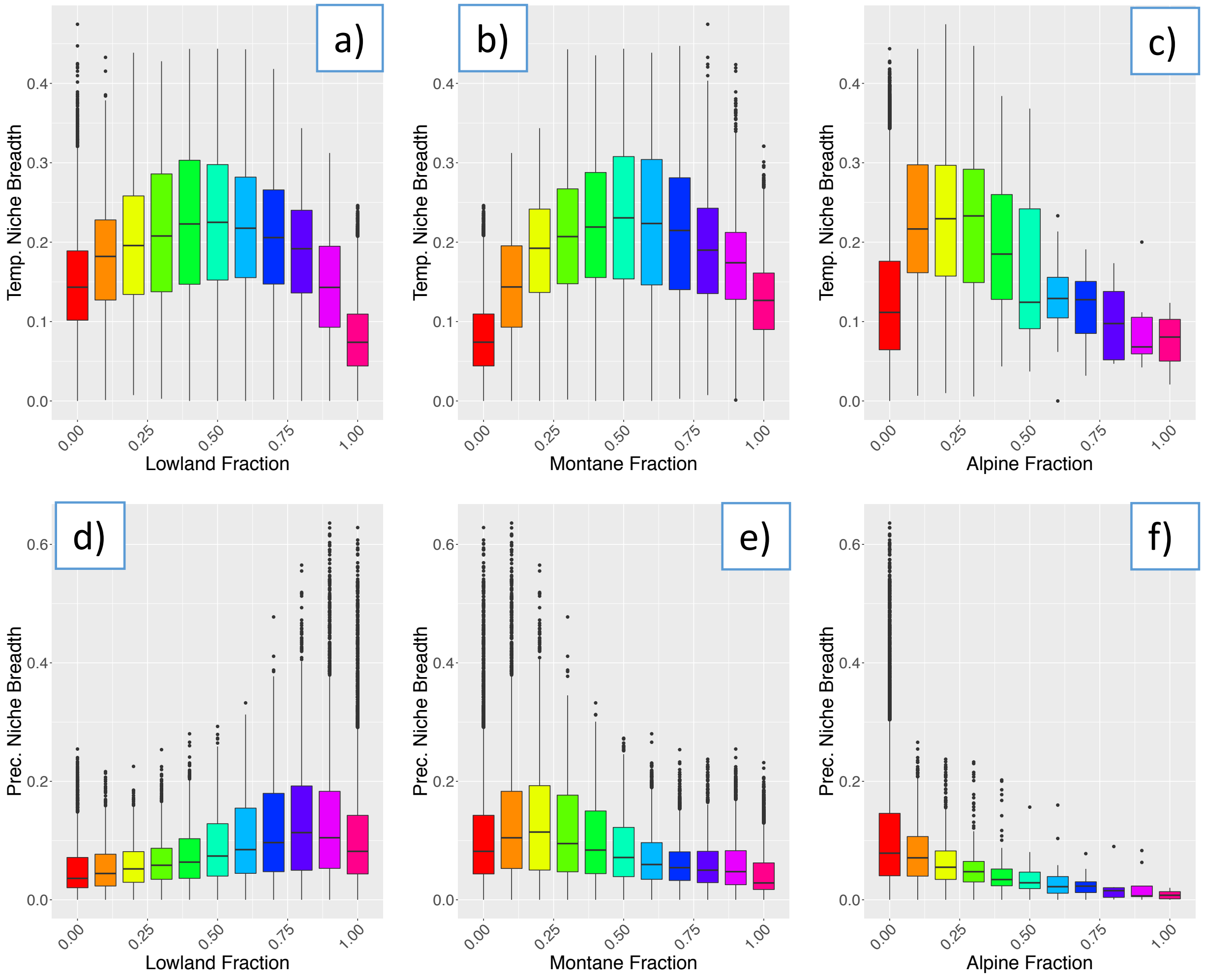


Figure 2.6. Relationships between niche breadth and elevational specialization. Temperature (*top row*) and precipitation (*bottom row*) niche breadth among species with varying proportion of their modelled range attributable to either lowland (*left column*), montane (*middle column*), or alpine (*right column*) habitat. Temperature niche breadth is maximal at intermediate range fractions in any given category, whereas precipitation niche breadth is consistently narrow regardless of elevational habitat, with lowland species having a wider range of precipitation niche breadth than alpine or primarily montane species.

We also compared nearest-neighbor pairs to the distribution of niche similarity scores within their particular focal clade (**Figure 2.S5**). With one exception (Poales B_{PREC} niche), nearest-neighbor pairs within each clade were, on average, significantly more similar in climatic niche than randomly chosen pairs within the same clade ($p<0.01$, Wilcoxon signed-rank test), though differences were not as extreme as when compared with the random pairs drawn from the entire seed plant phylogeny. However, the variation in nearest-neighbor similarity scores equaled the within-clade variation. Further, the mean within-clade niche similarity was generally at least as low as the average niche similarity across all seed plants. We discuss the significance of this below.

DISCUSSION

Phylogenetic niche conservatism.

Our results were consistent with phylogenetic niche conservatism playing a significant role in shaping the climatic niches of modern New World seed plants. Across disparate groups, such as gymnosperms, Poales (monocots), Ericales (asterids), and Fabales (rosids), average niche similarity decreased with decreasing relatedness (**Figure 2.3**), suggesting that niche conservatism is a general phenomenon contributing to the distribution of climatic niche space occupied by species within these groups. Additionally, at small phylogenetic scales (among closest relatives and among congeners), species were more similar to one another than expected by chance (**Figure 2.4**).

Table 2.1. Extent of climatic niche shifts among American seed plants.

	MAT	*Temp. Breadth*	*MAP*	*Prec. Breadth*	*ELV*
Congeners w/ complete conservatism[+]	37.1% (1230)	54.8% (1816)	37.1% (1228)	62.4% (2066)	71.6% (2371)
Congeners w/ complete niche shift[*]	2.9% (95)	1.6% (54)	2.4% (80)	0.9% (29)	2.0% (65)
Congeners w/ some niche shift	60.0% (1987)	43.6% (1442)	60.5% (2004)	36.7% (1217)	26.4% (876)
All sp. NNNS[#]	7.8%	21.6%	3.9%	19.0%	16.7%
Gymnosperms NNNS[#]	15.0%	22.0%	8.6%	13.3%	27.2%
Poales NNNS[#]	12.2%	19.6%	2.6%	8.4%	27.1%
Ericales NNNS[#]	6.2%	15.6%	3.3%	24.6%	14.0%
Fabales NNNS[#]	13.5%	10.6%	5.6%	9.1%	17.5%

[+] Complete conservatism (shift) means all species in the genus fall within the same (a different) niche. [*] Most genera in this category have 3 or fewer species and thus represent only one or two comparisons. [#] NNNS: Nearest-Neighbor Niche Shift indicates the fraction of species that differ in climatic niche from their closest phylogenetic relative(s). (*Total number of genera in data set with at least two species: 3312; total number of species: 68241. Numbers in parentheses indicate number of genera in that category.*)

These results represent a quantitative assessment of the degree of climatic niche similarity at four different phylogenetic scales: among closest relatives, among congeners, within broad clades (the angiosperm groups are defined as plant orders; however, gymnosperm is an unranked designation), and among all seed plants. Many biological processes are scale-dependent and drawing comparisons at multiple phylogenetic scales over such a broad dataset provided a more nuanced view of how niche conservatism manifests among different lineages and for different niche variables. For example, genus-level phylogenies that have been used previously to provide support for phylogenetic niche conservatism as a driver of biodiversity patterns (Segovia et al., 2020) carry with them the implicit assumption that within-genus variation in climatic niches is negligible.

Our analyses suggested, however, that even among congeners, statistically significant differences in climatic niches may exist (**Table 2.1**). For example, in our dataset, ~60% of genera (N=1987) contained species that differed significantly in temperature niche, and ~26% (N=876) of the same differed in their elevational niches. While the magnitudes of these differences varied (because we quantified niches as continuously valued functions), it demonstrated that even closely related species can show significant niche differences. This suggests that a strictly dichotomous paradigm, which views modern biodiversity as being shaped solely by niche conservatism or by niche shift, would fail to capture a substantial amount of observed biological variation. It is clear that phylogenetic niche conservatism is a major factor shaping modern plant biodiversity (Wiens et al., 2010), and recent work has shown that plants tend to retain their climatic niches over evolutionary time (Crisp et al., 2009; Gagnon et al., 2019; Segovia et al., 2020). However, it is also clear that species do exhibit niche shifts and biome switching (Donoghue & Edwards, 2014). Incorporating both of these views is thus crucial to our understanding of plant diversity, past and present.

Although in each of our analyses average niche similarity was greater for more closely related species, we also observed large variation in niche similarity scores. For example, within each focal clade, the variance in nearest-neighbor niche similarities equaled the total within-clade variance (**Figure 2.S5**). In other words, within each focal clade there existed some closest relatives that were as different in climatic niche as the most disparate members of that clade. Further, the average within-clade niche similarity was generally at least as low as the average niche similarity from randomly chosen species pairs across the seed plant phylogeny (**Figure**

2.S5). This suggests, for example, that the average temperature niche similarity among all Ericales is about as low as the average similarity from two randomly chosen seed plants, potentially indicating overdispersion in climatic niches at the scale of these focal clades. These results underscore how the prevalence of niche conservatism varies with phylogenetic scale.

We also found quantitative differences in the degree of niche similarity for different niche variables. For example, ~17% of species had elevational niches that differed significantly from their closest relatives, whereas for precipitation this number was only ~4%. In general, regardless of scale, precipitation niches were more conserved than elevational niches. Given the potential links between adaptive tolerance to cold and drought/ arid conditions (Folk et al., 2020), it could be that shifts in precipitation tolerances are less common and represent a greater abiotic filter that has a stronger influence on the distribution of seed plant lineages.

In addition to observing differences in niche similarity for different climatic variables, we observed substantial differences in DTT plots for each of our four focal clades (**Figure 2.3b**). In particular, Fabales niches showed a sharp discontinuity in average niche similarity at ~50-55 MYA. For temperature and precipitation niches, this time period is marked by an abrupt increase in niche similarity, whereas for elevational niches it marks a transition from high to low niche similarity. The elevational discontinuity might be explained by the relatively few, phylogenetically clustered transitions into higher-elevation habitat present in Fabales (**Figure 2.S2**). This result is consistent with contrasting patterns observed in the Ericales and Poales DTT plots of elevational niche, which showed undulating patterns of niche similarity. Both Ericales and Poales have multiple, independent transitions into higher elevations (**Figure 2.S2**), which might explain this observation. Since many of the angiosperm DTT plots show high to moderately high niche similarity at or around 50 MYA, it is tempting to explain this result in terms of decreasing global temperatures during this time, as well as the spread of more dry/ arid and open vegetation patterns (Graham, 2011). While this is certainly a likely contributor to the observed patterns, more explicit tests of this hypothesis, which incorporate knowledge from the fossil (pollen) record about the location of specific lineages over time, would be helpful to provide support for this potential explanation.

Habitat specialization and climatic niche breadth.

By quantifying the climatic niche space of species in a continuously valued way, we incorporated the biological reality that species experience variation in climate across their range. We could thus distinguish between species with ranges centered in alpine or montane habitat (i.e., higher elevation ‘specialists’) from those whose ranges extend into higher elevations but are centered in different habitat (i.e., higher elevation ‘generalists’). These distinctions are important for how biologists classify what is an ‘alpine species’ (Körner, 2003) and may reflect different adaptive responses and tolerances.

Here, we examined higher-elevation specialization by quantifying the relationship between climatic niche breadth and the fraction of species’ ranges falling within alpine, montane, or lowland habitat. We observed both specialist and generalist strategies for each of these three elevational categories, with species having anywhere from 0 to 100% of their modelled range in any given category (**Figure 2.6**). Thus, our dataset likely captured much of the actual variation in elevational niche that exists among American seed plants.

Notably, we found that alpine specialists (defined here as having at least 60% of their range confined to alpine areas) and species confined below the treeline (regardless of elevation) did not differ significantly in B_{TEMP}. Rather, a significant increase in B_{TEMP} was associated with entry into alpine habitat, as species below the treeline had significantly narrower B_{TEMP} than species with just 10% of their SDM in alpine areas. This difference remained significant for species with up to 60% of their range in alpine habitat but disappeared when we compared species found exclusively below the treeline to species with greater than 60% SDM above the treeline. Thus, species whose distribution was centered in alpine regions had approximately the same B_{TEMP} as those species found only below the treeline. Species which “bled over” into alpine regions, however, had greater B_{TEMP}.

These results might suggest that different strategies are needed by alpine generalists and specialists with respect to temperature responses, as the generalist species may encounter a wider range of temperatures than the specialists. However, we also acknowledge that our climatic analyses incorporated only macroscopic climatic conditions and did not address ways in which species at high elevations find and create microclimatic conditions to enhance survival (Körner, 2003). Although balancing such fine-scale considerations with larger, macroscale analyses is challenging, our results do allow for partitioning of species into generalist/ specialist categories which could, for instance, be used to address comparative questions about closely related organisms that differ in their degree of alpine specialization.

Although generalists and specialists might employ differing strategies, we found that genera with alpine *specialists* had a greater fraction of both alpine *generalists* and *montane* specialists (**Figure 2.S4**). This could indicate that alpine specialists derive mainly from alpine generalists and montane specialists (i.e., diversification has occurred as species encountered novel, higher-elevation habitat), consistent with a 'montane speciation model' (*sensu* Roy, 1997; but see Dagallier et al., 2020). It could also indicate that alpine specialists tend to arise within lineages having greater evolutionary potential to adapt to high-elevation conditions, even if the strategies involved differ among these lineages (Folk et al., 2020; Martínez-Padilla et al., 2017).

Species lists and regional species pools.

Combining SDMs with GMBA-defined mountain ranges, we generated potential species lists for all American mountains and mountain ranges, cross-referenced by country (**Online Supplement**). Although these species lists represent predictions of where species should occur, and not necessarily actual observations, they provide an excellent dataset from which to investigate the expected species pool for different montane regions within the Americas. These lists could be used to examine questions of expected co-occurrence (e.g., which species should be found in the Colombian Andes), or as climatically based predictions of species distributions from which to generate hypotheses about more local-scale process shaping biodiversity (e.g., by examining residual variation in species occurrences not explained by the SDMs).

Our dataset also highlighted areas in which increased sampling efforts would be useful. For example, **Figure 2.S1** shows relatively species-poor regions within the Amazon Basin surrounded by otherwise species-rich areas. This was likely due to sparse sampling in these regions and not indicative of dramatic shifts in species richness across the Basin at this larger spatial scale.

Conclusion.

By integrating georeferenced occurrence records with climate and phylogeny, we assembled a large dataset describing the climatic niches of ~70,000 American seed plants with respect to five climatic variables: MAT, MAP, ELV, $B_{TEMP.}$, and B_{PREC}. This dataset provided quantitative insight into American seed plant temperature and precipitation niches at a large, biogeographic scale, as well as characterizing the distribution of seed plants across elevational gradients. Our results supported phylogenetic niche conservatism as an important contributor to plant diversity patterns and suggested that plants may tend to track climatically favorable habitats (Donoghue, 2008). Specifically, we observed greater conservatism in precipitation niche and niche breadth than for temperature or elevation, and niche conservatism was greater at smaller phylogenetic scales.

Nevertheless, even among closely related species, variation in climatic niche was widespread. Thus, the role of niche shifts in shaping plant diversity patterns should also not be discounted. Further, clade-specific differences in the prevalence of niche shift were observed, indicating potential differences among lineages in their responses to changing environmental conditions. We summarized the extent of niche shift among different genera as one means of encouraging clade-specific questions on the importance of niche conservatism for different groups (with regards to those climate variables investigated here). Finally, our classification of American seed plant distributions along elevational gradients and compilation of expected species lists for individual mountains and mountain ranges provides an unprecedented first look at the expected species pool for montane and alpine habitats across the Americas.

REFERENCES

Antonelli, A., Kissling, W. D., Flantua, S. G. A., Bermúdez, M. A., Mulch, A., Muellner-Riehl, A. N., Kreft, H., Linder, H. P., Badgley, C., Fjeldså, J., Fritz, S. A., Rahbek, C., Herman, F., Hooghiemstra, H., & Hoorn, C. (2018). Geological and climatic influences on mountain biodiversity. *Nature Geoscience*, *11*(10), 718–725. https://doi.org/10.1038/s41561-018-0236-z

Antonelli, A., Nylander, J. A. A., Persson, C., & Sanmartin, I. (2009). Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proceedings of the National Academy of Sciences*, *106*(24), 9749–9754. https://doi.org/10.1073/pnas.0811421106

Bennett, K. D., Tzedakis, P. C., & Willis, K. J. (1991). Quaternary Refugia of North European Trees. *Journal of Biogeography*, *18*(1), 103–115. https://doi.org/10.2307/2845248

Brummitt, R.K. (2001) World geographical scheme for recording plant distributions, Ed. 2. *Biodiversity Information Standards* (TDWG) http://www.tdwg.org/standards/109.

Clarke, A., & Gaston, K. J. (2006). Climate, energy and diversity. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1599), 2257–2266. https://doi.org/10.1098/rspb.2006.3545

Craine, J. M., Ocheltree, T. W., Nippert, J. B., Towne, E. G., Skibbe, A. M., Kembel, S. W., & Fargione, J. E. (2013). Global diversity of drought tolerance and grassland climate-change resilience. *Nature Climate Change*, *3*(1), 63–67. https://doi.org/10.1038/nclimate1634

Crisp, M. D., Arroyo, M. T. K., Cook, L. G., Gandolfo, M. A., Jordan, G. J., McGlone, M. S., Weston, P. H., Westoby, M., Wilf, P., & Linder, H. P. (2009). Phylogenetic biome conservatism on a global scale. *Nature*, *458*(7239), 754–756. https://doi.org/10.1038/nature07764

Dagallier, L. P. M. J., Janssens, S. B., Dauby, G., Blach-Overgaard, A., Mackinder, B. A., Droissart, V., Svenning, J. C., Sosef, M. S. M., Stévart, T., Harris, D. J., Sonké, B., Wieringa, J. J., Hardy, O. J., & Couvreur, T. L. P. (2020). Cradles and museums of generic plant diversity across tropical Africa. *New Phytologist*, *225*(5), 2196–2213. https://doi.org/10.1111/nph.16293

Ding, W. N., Ree, R. H., Spicer, R. A., & Xing, Y. W. (2020). Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. *Science*, *369*(6503), 578–581. https://doi.org/10.1126/science.abb4484

Donoghue, M. J. (2008). A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences*, *105*(Supplement 1), 11549–11555. https://doi.org/10.1073/pnas.0801962105

Donoghue, M. J., & Edwards, E. J. (2014). Biome Shifts and Niche Evolution in Plants. *Annual Review of Ecology, Evolution, and Systematics*, *45*(1), 547–572. https://doi.org/10.1146/annurev-ecolsys-120213-091905

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Folk, R. A., Siniscalchi, C. M., & Soltis, D. E. (2020). Angiosperms at the edge: Extremity, diversity, and phylogeny. *Plant, Cell & Environment*, 1–23. https://doi.org/10.1111/pce.13887

Gagnon, E., Ringelberg, J. J., Bruneau, A., Lewis, G. P., & Hughes, C. E. (2019). Global Succulent Biome phylogenetic conservatism across the pantropical Caesalpinia Group (Leguminosae). *New Phytologist*, *222*(4), 1994–2008. https://doi.org/10.1111/nph.15633

GMBA. (2010). *The digital "Mountain Biodiversity Portal"*. http://www.mountainbiodiversity.org

Graham, A. (2011). The age and diversification of terrestrial New World ecosystems through Cretaceous and Cenozoic time. *American Journal of Botany*, *98*(3), 336–351. https://doi.org/10.3732/ajb.1000353

Harmon, L.J., Weir, J.T., Brock, C.D., Glor, R.E., & Challenger, W. (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics*, 24, 129-131.

Hawkins, B. A., Rodríguez, M. Á., & Weller, S. G. (2011). Global angiosperm family richness revisited: Linking ecology and evolution to climate: Angiosperm family richness. *Journal of Biogeography*, *38*(7), 1253–1266. https://doi.org/10.1111/j.1365-2699.2011.02490.x

Hawkins, B. A., Rueda, M., Rangel, T. F., Field, R., & Diniz-Filho, J. A. F. (2014). Community phylogenetics at the biogeographical scale: Cold tolerance, niche conservatism and the structure of North American forests. *Journal of Biogeography*, *41*(1), 23–38. https://doi.org/10.1111/jbi.12171

Hughes, C. E., & Atchison, G. W. (2015). The ubiquity of alpine plant radiations: From the Andes to the Hengduan Mountains. *New Phytologist*, *207*(2), 275–282. https://doi.org/10.1111/nph.13230

Hughes, C., & Eastwood, R. (2006). Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences*, *103*(27), 10334–10339. https://doi.org/10.1073/pnas.0601928103

Jamieson, M. A., Trowbridge, A. M., Raffa, K. F., & Lindroth, R. L. (2012). Consequences of Climate Warming and Altered Precipitation Patterns for Plant-Insect and Multitrophic Interactions. *Plant Physiology*, *160*(4), 1719–1727. https://doi.org/10.1104/pp.112.206524

Judd, W. S., Sanders, R., & Donoghue, M. J. (1994). Angiosperm family pairs: Preliminary phylogenetic analyses. *Harvard Papers in Botany*, *1*(5), 1–51.

Kerkhoff, A. J., Moriarty, P. E., & Weiser, M. D. (2014). The latitudinal species richness gradient in New World woody angiosperms is consistent with the tropical conservatism hypothesis. *Proceedings of the National Academy of Sciences*, *111*(22), 8125–8130. https://doi.org/10.1073/pnas.1308932111

Körner, C. (2003). *Alpine plant life: Functional plant ecology of high mountain ecosystems.* (2nd ed.). Springer, Berlin, Heidelberg.

Körner, C., Jetz, W., Paulsen, J., Payne, D., Rudmann-Maurer, K., & Spehn, E. M. (2017). A global inventory of mountains for bio-geographical applications. *Alpine Botany*, *127*(1), 1–15. https://doi.org/10.1007/s00035-016-0182-6

Körner, C., & Paulsen, J. (2004). A world-wide study of high altitude treeline temperatures: Study of high altitude treeline temperatures. *Journal of Biogeography*, *31*(5), 713–732. https://doi.org/10.1111/j.1365-2699.2003.01043.x

Körner, C., Paulsen, J., & Spehn, E. M. (2011). A definition of mountains and their bioclimatic belts for global comparisons of biodiversity data. *Alpine Botany*, *121*(2), 73–78. https://doi.org/10.1007/s00035-011-0094-4

Levins, R. (1968). *Evolution in changing environments.* Princeton University Press.

Loveys, B. R., Scheurwater, I., Pons, T. L., Fitter, A. H., & Atkin, O. K. (2002). Growth temperature influences the underlying components of relative growth rate: An investigation using inherently fast- and slow-growing plant species: Growth temperature and effects on carbon economy. *Plant, Cell & Environment*, *25*(8), 975–988. https://doi.org/10.1046/j.1365-3040.2002.00879.x

Macarthur, R. H. (1972). *Geographical ecology: Patterns in the distribution of species.* Princeton University Press.

Martínez-Padilla, J., Estrada, A., Early, R., & Garcia-Gonzalez, F. (2017). Evolvability meets biogeography: Evolutionary potential decreases at high and low environmental favourability. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1856), 20170516. https://doi.org/10.1098/rspb.2017.0516

Moles, A. T., Perkins, S. E., Laffan, S. W., Flores-Moreno, H., Awasthy, M., Tindall, M. L., Sack, L., Pitman, A., Kattge, J., Aarssen, L. W., Anand, M., Bahn, M., Blonder, B., Cavender-Bares, J., Cornelissen, J. H. C., Cornwell, W. K., Díaz, S., Dickie, J. B., Freschet, G. T., … Bonser, S. P. (2014). Which is a better predictor of plant traits: Temperature or precipitation? *Journal of Vegetation Science*, *25*(5), 1167–1180. https://doi.org/10.1111/jvs.12190

Muellner-Riehl, A. N., Schnitzler, J., Kissling, W. D., Mosbrugger, V., Rijsdijk, K. F., Seijmonsbergen, A. C., Versteegh, H., & Favre, A. (2019). Origins of global mountain plant biodiversity: Testing the 'mountain-geobiodiversity hypothesis.' *Journal of Biogeography*, *46*(12), 2826–2838. https://doi.org/10.1111/jbi.13715

Nie, J., Garzione, C., Su, Q., Liu, Q., Zhang, R., Heslop, D., Necula, C., Zhang, S., Song, Y., & Luo, Z. (2017). Dominant 100,000-year precipitation cyclicity in a late Miocene lake from northeast Tibet. *Science Advances*, e1600762.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3–4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Phillips, S. J., Dudík, M., & Schapire, R. E. (2004). A maximum entropy approach to species distribution modeling. *Twenty-First International Conference on Machine Learning - ICML '04*, 655–662. https://doi.org/10.1145/1015330.1015412

Qian, H. (2017). Relationship between clade age and temperature for angiosperm tree species in forest communities along an elevational gradient in tropical Asia. *Journal of Plant Ecology*, *10*(4), 618–625. https://doi.org/10.1093/jpe/rtw074

Qian, H., & Chen, S. (2016). Ecological determinants of mean family age of angiosperm trees in forest communities in China. *Scientific Reports*, *6*(1), 28662. https://doi.org/10.1038/srep28662

R Development Core Team. 2020. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. (n.d.).

Rahbek, C., Borregaard, M. K., Antonelli, A., Colwell, R. K., Holt, B. G., Nogues-Bravo, D., Rasmussen, C. M. Ø., Richardson, K., Rosing, M. T., Whittaker, R. J., & Fjeldså, J. (2019). Building mountain biodiversity: Geological and evolutionary processes. *Science*, *365*(6458), 1114–1119. https://doi.org/10.1126/science.aax0151

Rahbek, C., Borregaard, M. K., Colwell, R. K., Dalsgaard, B., Holt, B. G., Morueta-Holme, N., Nogues-Bravo, D., Whittaker, R. J., & Fjeldså, J. (2019). Humboldt's enigma: What causes global patterns of mountain biodiversity? *Science*, *365*(6458), 1108–1113.

Roy, M. S. (1997). Recent diversification in African greenbuls (Pycnonotidae: Andropadus) supports a montane speciation model. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *264*, 1337–1344.

Segovia, R. A., Pennington, R. T., Baker, T. R., de Souza, F. C., Neves, D. M., Davis, C. C., Armesto, J. J., Olivera-Filho, A. T., & Dexter, K. G. (2020). Freezing and water availability structure the evolutionary diversity of trees across the Americas. *Science Advances*, 9.

Smith, S. A., & Brown, J. W. (2018). Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany*, *105*(3), 302–314. https://doi.org/10.1002/ajb2.1019

Soltis, D. E., & Soltis, P. S. (2016). Mobilizing and integrating big data in studies of spatial and phylogenetic patterns of biodiversity. *Plant Diversity*, *38*(6), 264–270. https://doi.org/10.1016/j.pld.2016.12.001

Sun, M., Folk, R. A., Gitzendanner, M. A., Soltis, P. S., Chen, Z., Soltis, D. E., & Guralnick, R. P. (2020). Recent accelerated diversification in rosids occurred outside the tropics. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-17116-5

Wang, L. G., Lam, T. T. Y., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C. W., Jones, B. R., Bradley, T., Zhu, H., Guan, Y., Jiang, Y., & Yu, G. (2020). treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. *Molecular Biology and Evolution*, *37*(2), 599–603. https://doi.org/10.1093/molbev/msz240

Webb, C. O. (2000). Exploring the Phylogenetic Structure of Ecological Communities: An Example for Rain Forest Trees. *The American Naturalist*, *156*(2), 145–155. https://doi.org/10.1086/303378

Wen, J., Zhang, J. Q., Nie, Z. L., Zhong, Y., & Sun, H. (2014). Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Frontiers in Genetics*, *5*, 1–16. https://doi.org/10.3389/fgene.2014.00004

Whittaker, R. H. (1970). *Communities and ecosystems.* Macmillan.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag.

Wiens, J. J., Ackerly, D. D., Allen, A. P., Anacker, B. L., Buckley, L. B., Cornell, H. V., Damschen, E. I., Davies, T. J., Grytnes, J. A., Harrison, S. P., Hawkins, B. A., Holt, R. D., McCain, C. M., &

Stephens, P. R. (2010). Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters*, *13*(10), 1310–1324. https://doi.org/10.1111/j.1461-0248.2010.01515.x

Xie, Y., Wang, X., & Silander, J. A. (2015). Deciduous forest responses to temperature, precipitation, and drought imply complex climate change impacts. *Proceedings of the National Academy of Sciences*, *112*(44), 13585–13590. https://doi.org/10.1073/pnas.1509991112

Xing, Y., & Ree, R. H. (2017). Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. *Proceedings of the National Academy of Sciences*, *114*(17), E3444–E3451. https://doi.org/10.1073/pnas.1616063114

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36. https://doi.org/10.1111/2041-210X.12628

Zachos, J., Pagani, M., Sloan, L., Thomas, E., & Billups, K. (2001). Trends, Rhythms, and Aberrations in Global Climate 65 Ma to Present. *Science*, *292*(5517), 686–693. https://doi.org/10.1126/science.1059412

Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. A., FitzJohn, R. G., McGlinn, D. J., O'Meara, B. C., Moles, A. T., Reich, P. B., Royer, D. L., Soltis, D. E., Stevens, P. F., Westoby, M., Wright, I. J., Aarssen, L., Bertin, R. I., Calaminus, A., Govaerts, R., … Beaulieu, J. M. (2014). Three keys to the radiation of angiosperms into freezing environments. *Nature*, *506*(7486), 89–92. https://doi.org/10.1038/nature12872

SUPPLEMENTAL FIGURES & TABLES

***** ***** *****

Click here for a link to the files related to the Online Supplement.

***** ***** *****

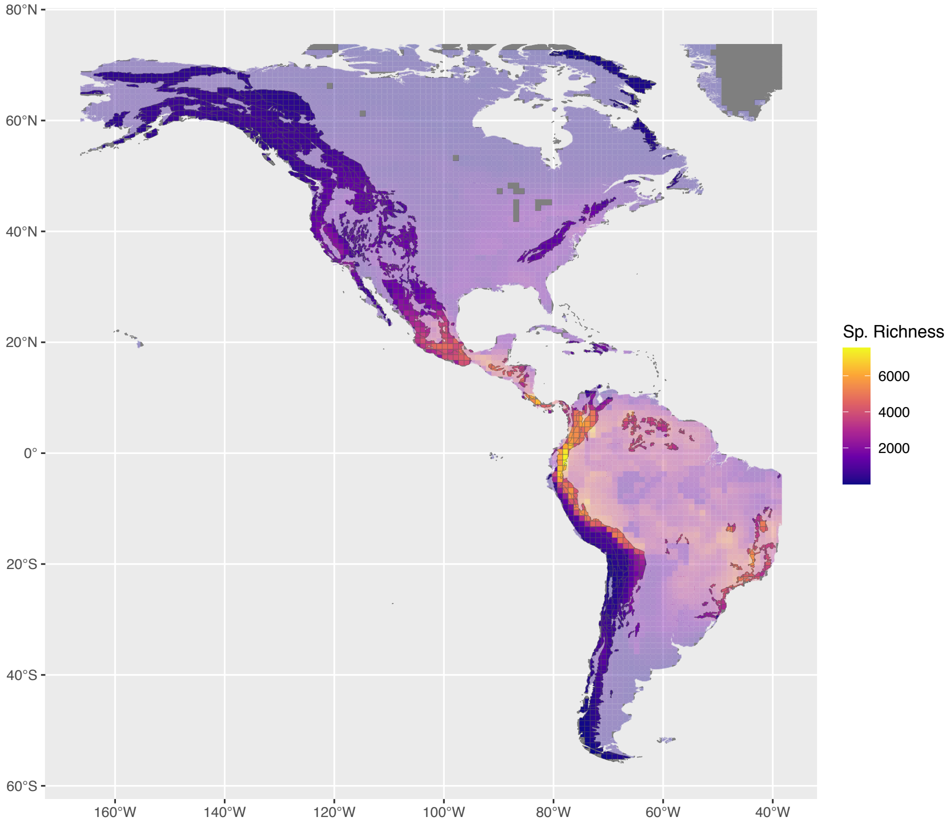


Figure 2.S1. Species richness and data set coverage. For each grid cell we tabulated the number of species predicted to occur there based on our species distribution models. Cooler (warmer) colors indicate lower (greater) species richness. *Foreground:* Mountain areas defined by the Global Mountain Biodiversity Assessment (GMBA). *Background:* non-mountainous areas in the sampling region. Grey cells indicate missing or insufficient data. In general, our data set covered almost all mountains and most of the Americas overall. Notable areas with missing data include the extreme Northern and Southern latitudes. Areas likely poorly sampled include the Amazon basin (blue regions within areas of otherwise high diversity). A latitudinal diversity gradient for both lowland and high elevation regions is apparent.

[LINK TO HIGH-RESOLUTION PHYLOGENIES WITH ASSOCIATED CLIMATIC NICHE INFORMATION]

Figure 2.S2. Phylogenetic distribution of climatic niches for Gymnosperms ($N_{species}$=301), Poales ($N_{species}$=6638), Ericales ($N_{species}$=2790), and Fabales ($N_{species}$=5579). *Tree panel*: phylogenetic relationships among all taxa. For *ELV* niches, colors indicate the fraction of that species' range falling into each of the 7 bioclimatic zones defined by the Global Mountain Biodiversity Assessment (GMBA). For *MAT* niches, colors indicate the fraction of that species' range falling within a 1° Celsius increment of temperature, with warmer (cooler) colors indicating hotter (colder) temperatures. For *MAP* niches, colors indicate the fraction of that species' range falling within a 250 mm increment of precipitation, with warmer (cooler) colors indicating less (more) rainfall. The continuously valued characterization of climatic niches in this way allows us to incorporate the biological reality that most species experience varied climatic conditions along their range. It also permits categorization of species, for example, as alpine specialists or generalists based on what habitat their range is centered upon. Statistical methods supported a model of phylogenetic niche conservatism for these four groups, and for all seed plants in the data set. Nevertheless, even closely related taxa often differ in climatic niche, demonstrating the importance of niche shift for understanding modern biodiversity patterns. For instance, among *Ephedra* and *Gnetum* (jointly, Gnetopyta, Gymnosperms panel bottom), there are species found in drier lowlands (e.g., *E. pedunculata*), drier highlands (e.g., *E. frustillata*), and wetter lowlands (e.g., *G. urens*), indicating at least two niche shifts may have occurred.

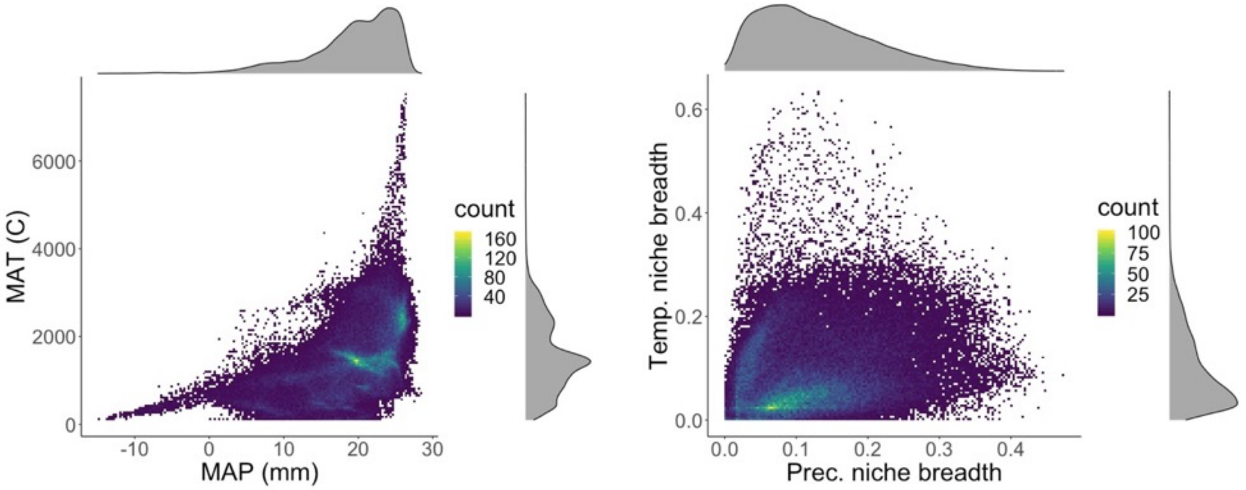


Figure 2.S3. (*left*) Distribution of mean annual temperature (MAT) and precipitation (MAP) values, and (*right*) distribution of niche breadth values for all 72,372 species in our data set. Colors indicate the number of species with SDMs that include that climatic value. Univariate histograms are included on the margins of each plot. The distribution of temperature and precipitation values here closely mirrors that found across global biomes (compare with Figure 1 of Donoghue & Edwards 2014). Species tended to have generally narrow temperature and precipitation niche breadths, though there is a long tail for each distribution, indicating some species with much larger niche breadths. Relatively few species having both a broad temperature and a broad precipitation niche.

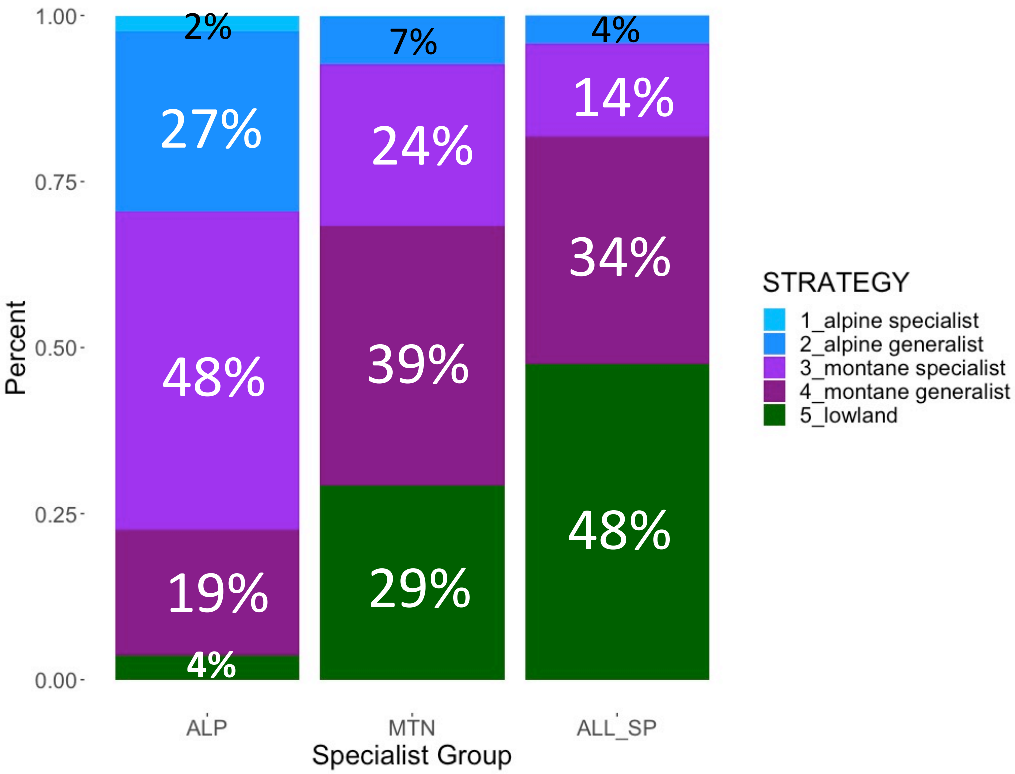


Figure 2.S4. Fraction of species with a given elevational specialization strategy, grouped by genera containing alpine specialists (*ALP*), montane specialists (*MTN*), or all genera (*ALL_SP*). Among genera with alpine specialists (*ALP, left column*; specialists for all categories here are defined as having at least 60% of their modelled range within a particular category), there are also proportionally more alpine generalist (27%) and montane specialist (48%) species, compared to those same fractions among all American seed plants (*ALL_SP, right column*; 4% and 14%, respectively). Thus, alpine specialists tend to fall within genera enriched in alpine *generalists* and *montane* specialists. Similarly, genera with montane specialists (*MTN, middle column*) are enriched in alpine generalist (7%) and montane generalist species (39%), compared to the full data set (4% and 34%, respectively), though the differences are not as extreme as for genera with alpine specialists. Thus, montane specialist species tend to fall within genera also enriched in montane generalists.

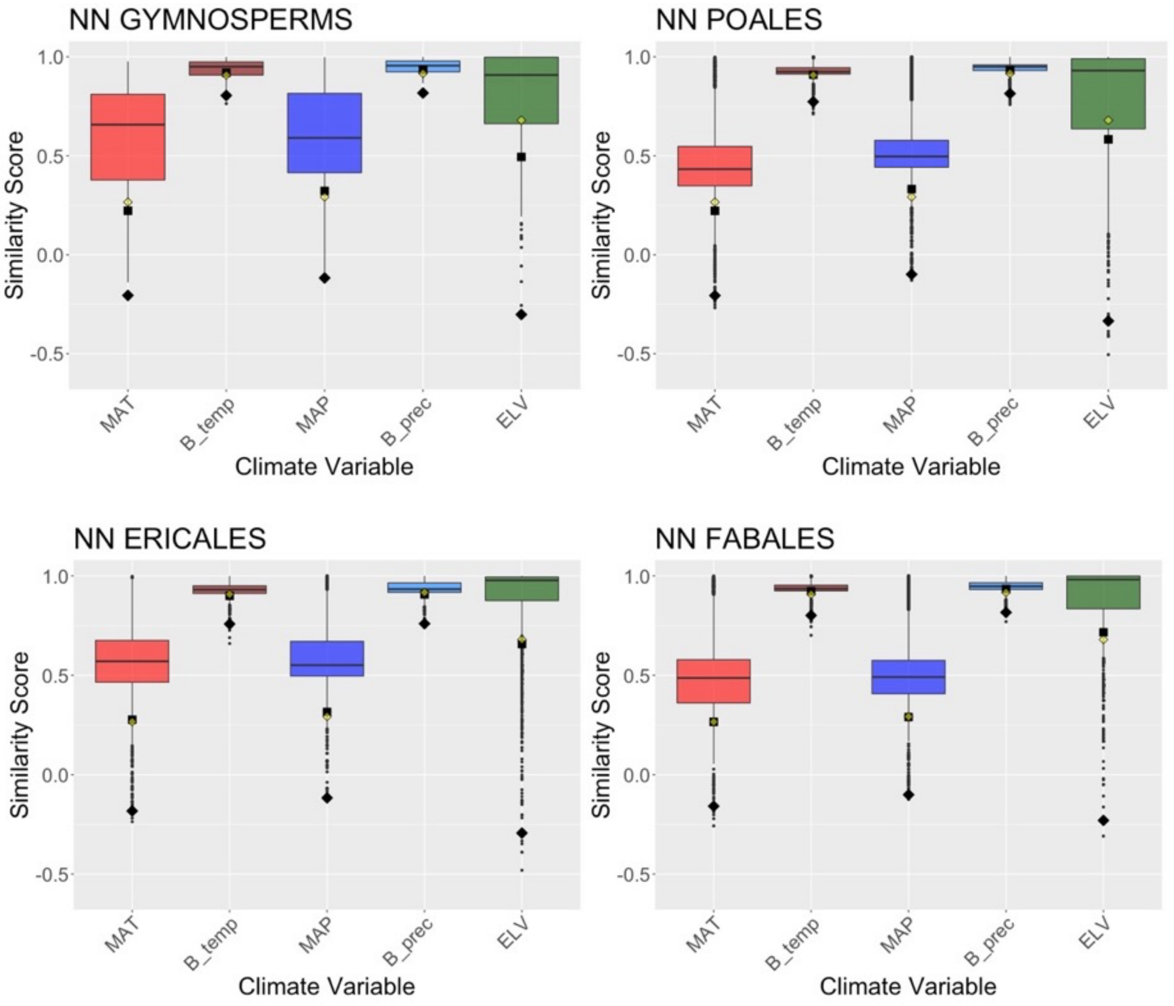


Figure 2.S5. Clade-specific variation in nearest-neighbor (*NN*) climatic niche similarity for temperature (*MAT*), precipitation (*MAP*), and elevation (*ELV*), as well as temperature and precipitation niche breadths (*B_temp*, *B_prec*, resp.). Boxplots show median similarity scores for each climatic variable. *Yellow diamonds* indicate the lower 5% quantile of expected niche similarity based on random pairwise comparisons drawn from among all seed plants. Points below yellow diamonds thus indicate nearest-neighbor pairs that are *significantly more different* (over-dispersed) than species selected at random from the entire seed plant phylogeny. Compared to this null, most species within each clade are more similar to their closest relative than expected by chance. At the same time, for each climatic variable, substantial fractions of species were actually less similar than expected by chance (i.e., values below yellow diamonds, see **Table 2.1** for fractions of species in these categories), indicating niche shifts had occurred. Dissimilarity in niche breadth among nearest neighbors was particularly prevalent. The distribution of within-clade similarity scores is also indicated in the plot: *Black diamonds* show the lower 5% quantile of niche similarity scores among all pairwise comparisons within the focal clade (e.g., only considering Ericales); *Black squares* indicate mean niche similarity score within the focal clade (e.g., average similarity among all Ericales). Notably, the variation in similarity scores among nearest neighbors equals the within-clade variation. In other words, for each clade, there are at least some closest relatives that differ in climatic niche by as much as the most disparate pair of species within that clade.

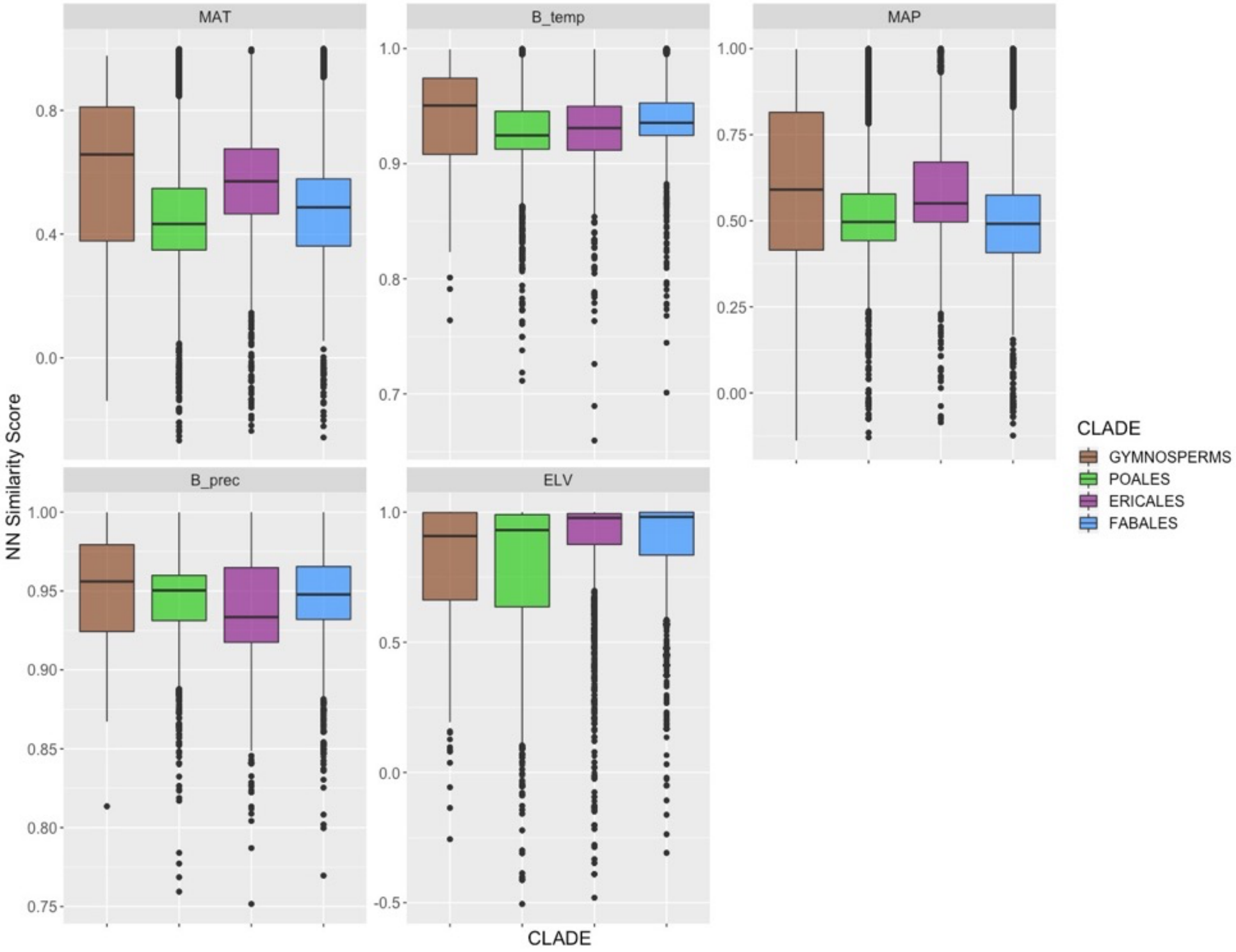


Figure 2.S6. Clade-specific variation in nearest-neighbor (*NN*) climatic niche similarity for temperature (*MAT*), precipitation (*MAP*), and elevation (*ELV*), as well as temperature and precipitation niche breadths (*B_temp*, *B_prec*, resp.). Boxplots show median similarity scores for each climatic variable for gymnosperms (brown), Poales (green), Ericales (purple), and Fabales (blue). Data here is the same as boxplots in **Figure 2.S5** above but rearranged by climatic variable in order to highlight niche differences between these four focal groups. Most differences between groups were significant (**Table 2.S2**). Notably, gymnosperms had higher within-clade niche similarity than any of the angiosperm groups for all niche variables except ELV.

Table 2.S1. Statistically significant correlations in the extent of niche shift among different climatic variables^{+}.

	MAT	*MAP*	*ELV*	*Temp. Breadth*	*Prec. Breadth*
MAT	-	-	-	-	-
MAP	0.24^{***}	-	-	-	-
ELV	0.48^{***}	0.09^{***}	-	-	-
Temp. Breadth	0.19^{**}	0.13^{***}	0.28^{***}	-	-
Prec. Breadth	-0.02^{NS}	0.21^{**}	-0.14^{NS}	0.06^{**}	-

*+Tabular values represent slopes from linear model fits. NS=not significant; *p<0.05; **p<0.01; ***p<0.001.*

Table 2.S2. Statistically significant differences in the prevalence of climatic niche shift among closest relatives.

	Gymnosperms	*Poales*	*Ericales*	*Fabales*
Gymnosperms	-	-	-	-
Poales	MAT, MAP, BTEMP, BPREC.	-	-	-
Ericales	ELV, BTEMP, BPREC.	MAT, MAP, ELV, BTEMP, BPREC.	-	-
Fabales	MAT, MAP, ELV, BPREC.	MAT, MAP, ELV, BTEMP, BPREC.	MAT, MAP, ELV, BTEMP, BPREC.	-

MAT=mean annual temperature; MAP=mean annual precipitation; ELV=elevation; B=niche breadth. Statistical significance was assessed using the Kruskal-Wallis test rather than ANOVA because samples had unequal variances and violated the normality assumption. Dunn's test was therefore used for post-hoc analyses to determine which comparisons were significantly different. Significance here means the adjusted p-value < 0.05.

Chapter Three

Contrasting Patterns of Phylogenetic Diversity and Alpine Specialization Across the Alpine Flora of the American Mountain Range System[††]

CHAPTER IN CONTEXT

In this chapter, I examine macroecological biodiversity patterns of alpine communities across the American mountain range system. I use the dataset assembled and described in **CHAPTER TWO** to quantify species richness, phylogenetic diversity, taxonomic structure, and degree of alpine specialization in assemblages across this study area. The results provide a detailed assessment of the current state of knowledge on alpine communities, based on data availability, and the study highlights both alpine regions in need of increased sampling as well as regions of high conservation priority based on their contribution to alpine diversity. Broad diversity patterns across all alpine lineages are presented, but specific clades contributing greatly to alpine richness are discussed in greater detail.

By examining diversity patterns of alpine communities at different phylogenetic scales, this study highlights the complex and scale-dependent interplay of how processes such as abiotic filtering and biogeographic history jointly interact to shape alpine floral diversity. For example, although the Solanaceae family was among the top five families contributing most to alpine species richness, this marked diversity in the alpine zone was restricted to the Andes. In this study, I discuss how the absence of Solanaceae from non-Andean mountain ranges might be due to ecological/ physiological limits rather than dispersal limitation, with modern ranges shaped by relatively strong niche conservatism of both cold and xeric intolerance. I suggest that modern

---

[††] This chapter was originally published in *Alpine Botany* (2021) 132: 107-122.

Solanaceae distributions thus reflect the influence of both abiotic filtering (limiting their latitudinal and elevational extents) as well as historical processes (substantially greater diversity in South America).

More broadly, alpine communities occurring at opposite ends of the American mountain range system, where macro-scale regional climatic conditions might be most comparable, showed marked differences in phylogenetic diversity patterns and degree of alpine specialization. Specifically, near-Arctic/ boreal alpine communities (above approx. 50° N) were typically phylogenetically clustered with a greater degree of alpine specialization, whereas southern Patagonian communities (below approx. 40° S) were more phylogenetically even with a lesser degree of specialization. These differences suggest an important role for historical factors in the assembly of their floras and indicate that abiotic filtering alone is unlikely to explain community assembly dynamics in these communities.

This chapter contributes to this dissertation by highlighting specific examples of how niche conservatism has helped to shape the modern American alpine flora. These examples further illustrate how community assembly processes act in scale-dependent manners and that different lineages show varying degrees of influence from processes such as abiotic filtering and biogeographic history. This framework, which attempts to take account of the (lineage-specific) contributions of multiple, interacting processes to community assembly, forms the main thesis of this dissertation.

ABSTRACT

Although mountainous habitats contribute substantially to global biodiversity, comparatively little is known about biogeographic patterns of distributions of alpine species across multiple mountain ranges. Here, we present a detailed analysis of the distributions and phylogenetic affinities of alpine seed plant lineages across North, Central, and South American mountain systems. Using a large dataset that characterizes the elevational niches of American seed plants in a continuously valued way, we related the proportion of alpine habitat occupied by plant lineages to their biogeographic distributions at a regional scale and place these results in a phylogenetic context. We found alpine species diversity to be greatest in the central Andes and western North America, and that assemblages with lower phylogenetic diversity contained species with a greater degree of alpine specialization. In particular, near-Arctic/ boreal alpine communities were characterized by low phylogenetic diversity and higher degrees of alpine specialization, whereas the opposite was observed for southern Patagonian communities. These results suggest that abiotic filtering alone in these climatically similar regions is unlikely to explain alpine community assembly. Nevertheless, the overall relative rarity of alpine specialists, and the tendency for such specialists to be most closely related to montane lineages, suggested that filtering was still an important factor in shaping alpine community structure. This work corroborates the importance of a nuanced and scale-dependent perspective on the 'history-filtering' debate axis, as both factors have likely contributed to modern biodiversity patterns observed in alpine plant communities across the Americas.

INTRODUCTION

Despite representing less than 25% of terrestrial surface (Körner et al. 2011), mountains harbor exceptional biodiversity (Myers et al. 2000; Körner 2003, 2004; Kreft and Jetz 2007; Kier et al. 2009; Körner et al. 2011; IPBES 2019; Rahbek et al. 2019a,b; Brummitt et al. 2021). High elevation habitats offer potential cooler-climate corridors for plant dispersal (Antonelli et al., 2009), and their topographic complexity provides myriad microclimatic niche space for plants to become established and survive (Körner 2003). Nevertheless, many questions remain unanswered regarding the biogeography (Rahbek et al. 2019a,b) and phylogenetic composition (González-Caro et al. 2020) of higher elevation assemblages. Additionally, historical and biogeographic processes such as climatic cycles and glaciations have fragmented and reconnected alpine "sky islands" (Marx et al. 2017). This intermittent connectivity across time has further promoted high levels of diversification and endemism in some of the hottest spots of diversity, such as the Andean mountain chain (Hughes and Eastwood 2006; Flantua et al. 2019) and the Tibet-Himalaya-Hengduan region (Hughes and Atchison 2015; Ding et al. 2020).

The complexity of mountain biodiversity is further evidenced by changes in species richness (Guo et al. 2013) and turnover (McFadden et al. 2019; Smithers et al. 2020) across elevational gradients. Such patterns have prompted researchers to delimit different elevational zones across the globe in a systematic way. In particular, Körner et al. (2011) defined seven life thermal belts based on bioclimatic and topographic characteristics: the nival (perpetual snowline), upper alpine, lower alpine (tree line estimate), upper montane, lower montane, remaining mountain area with frost, and remaining mountain area without frost. These thermal zones account for latitudinal differences in the absolute elevation of alpine and montane habitats (Körner et al. 2011) and provide a biologically meaningful and geographically robust assessment of different elevational belts. Further, these elevational categories allow for comparisons of biodiversity patterns across different mountain ranges at large biogeographic scales (Körner et al. 2011). Such a systematic approach is invaluable for studies seeking to draw general, macroecological and biogeographic inferences about processes structuring higher elevation assemblages across different mountain systems. This is especially important since, although it is clear that the combined history of orogeny, uplift, erosion, and climatic cycles have altered the landscape of mountains and mountain ranges in complex ways, the proximate sequence and timing of events that built the biodiversity we observe is still an open question.

Despite the importance of mountains for biodiversity, few studies have formally synthesized plant diversity patterns within and across mountain regions (Körner 1995, 2004). In particular, comparatively little is known about the biodiversity of the alpine life zone (including the nival, upper and lower alpine thermal belts). Alpine zones are characterized by extreme environments at the physiological limits of plant life (Körner 2003) and represent one end of the available niche spectrum in mountain regions. Alpine zones are also at the elevational limits of available terrestrial terrain (Elsen and Tingley 2015), and lineages specializing in this niche space are at risk for being extirpated as lowland species move up slope as a result of climate change and potentially outcompete alpine endemics or leave them with no further habitat upslope (Millar and Fagre 2007; Kelly and Goulden 2008; Morueta-Holme et al. 2015; Steinbauer et al. 2018; Moret et al. 2019).

Crucially, several knowledge gaps about general patterns of alpine plant diversity remain: (1) how alpine community richness varies across latitude and among different mountain regions; (2) whether alpine communities are phylogenetically clustered or overdispersed relative to the regional (e.g., montane) species pool (but see the recent study by Qian et al. 2021); (3) which lineages dominate in alpine communities and which lineages exhibit the highest turnover (beta-diversity) between mountain regions; and (4) how such lineage-specific patterns relate to processes of abiotic filtering, dispersal limitation, and historical (biogeographic) contingency.

Such macro-scale patterns in the composition and diversity of alpine communities likely arise from differing richness and phylogenetic diversity patterns among individual lineages comprising these assemblages. Therefore, an appreciation for the contributions that specific lineages make to total alpine diversity is also needed, as such contributions may reflect the confluence of historical processes, filtering, and ecological dynamics. For example, the *Astragalus/ Oxytropis* complex is a primary component of Fabaceae diversity across the Americas (Amiri et al. 2020), including alpine habitats, however the origins of the clade are likely Eurasian (Bagheri et al. 2017; Amini et al. 2019). It is therefore plausible that adaptations to cold or higher elevations arose in this lineage prior to the assembly of any one particular modern American alpine community. In contrast, Solanales diversity in the Americas is concentrated within South America, where the clade is believed to have originated (Olmstead 2013). The contribution Solanales make to American alpine assemblages might therefore be expected to reflect this distinct biogeographic history. Synthesizing how different biogeographic legacies of particular alpine lineages combine with lineage-specific physiological constraints and ecological filters to produce modern alpine assemblages would greatly enhance our current understanding of alpine biodiversity overall.

In addition to these questions regarding alpine community assembly and dynamics, the ability to discern macroecological patterns is also limited by data availability. Alpine floras, especially those in the Andes, are often composed of numerous range-limited endemics (Hughes and Eastwood 2006), sometimes known only from single mountain peaks. Data for such species can be equally scattered and diffuse (e.g., Sklenář and Balslev 2005; Al-Shehbaz 2018), creating significant challenges for addressing general questions regarding alpine plant ecology across different mountain ranges or through time. Defining the distributions of such narrowly distributed species and assessing the contributions of these lineages to the overall patterns of alpine biodiversity at larger spatial scales is an ongoing challenge.

In this study, we use georeferenced species occurrence records from digitized herbarium databases and climate data to build species distribution models (SDMs) and present a detailed investigation of alpine seed plant biodiversity across North, Central, and South American mountain ranges. We use this dataset to achieve two broad goals. First, we characterize general, macroecological and biogeographic diversity patterns among alpine communities such as latitudinal diversity gradients, phylogenetic diversity of alpine assemblages, and turnover in alpine community composition. Distribution models allowed us to continuously define the niche space of each species in the American mountains, distinguishing alpine specialists from generalists and mapping their diversity in order to quantify the degree of alpine specialization across different mountains. Second, we use our conservative data cleaning protocol to critically assess the quality of currently available occurrence data for such macroecological questions of alpine communities and address potential regional biases in data availability whose eventual resolution might greatly enhance the current state of knowledge on alpine community assembly.

METHODS

Species distributions and niche characterization.

For this study, we utilized a previously assembled dataset characterizing the elevation and climate niches of 72,372 seed plant species across the Americas using occurrence records and species distribution models (SDMs; Figueroa et al., *in review*). Briefly, this dataset was assembled by aggregating georeferenced occurrence records from the Global Biodiversity Information Facility (GBIF) and Integrated Digitized Biocollections (iDigBio). Occurrence records were initially cleaned separately using aggregator-specific flags detailed in Figueroa et al. (*in review*). The full dataset of occurrence records is accessible and fully described in the Dryad data repository at (https://doi.org/10.5061/dryad.9cnp5hqgx) and (https://doi.org/10.15468/dl.gtgtt5).

Aggregated GBIF and iDigBio records were further processed by excluding points with any of the following issues: (1) falling outside the study area (the Americas); (2) less than four decimal point precision (~11 m near the equator); (3) duplicate localities for a species (rarefaction); (4) falling outside polygons describing accepted species' distributions (as defined by Plants of the World Online, POWO; www.github.com/tdwg/wgsrpd; Brummitt 2001); and (5) species with fewer than twelve records (to build reliable models). Cleaned records were then passed to MaxEnt (version 3.1.4; www.github.com/mrmaxent/Maxent; Phillips et al. 2004, 2006) along with 2.5' resolution climate data from WorldClim (Fick and Hijmans 2017) in order to build SDMs for the range of each species across the Americas (full set of models available at https://doi.org/10.5061/dryad.9cnp5hqgx).

Having obtained SDMs for all 72,372 species in our dataset, we characterized elevational niches by parsing each SDM into different elevational categories. We first used the 2.5' shapefiles from the Global Mountain Biodiversity Assessment (GMBA 2010; Körner et al. 2011, 2017) to spatially define mountainous habitats (376 individual mountains, and 8 large-scale mountain ranges: Andes, Appalachia, Columbia, Cascades, Interior, Rockies, Sierra Madre, West Coast) in the Americas and then used the seven isothermal belt zones of (Körner and Paulsen 2004; Körner et al. 2011) to delimit elevational categories as montane, alpine, and lowland habitat areas within each mountain range. These zones integrate temperature, growing season length, and topographic information to provide a robust and biogeographically relevant assessment of the extent and type of montane and alpine habitat across the globe. They further permit an assignment of species distributions along elevational gradients despite latitudinal differences in what constitutes montane and alpine areas (see Table 2 in Körner et al. 2011 for precise climatic zone definitions). Because these elevational categories were built at 2.5' resolution, all of our analyses were matched to this for consistency.

To accommodate lineages that experience varied climate across their ranges, we characterized niches as continuously valued distributions and not as discrete scalar values. This allowed us to quantify the fraction of each species' range falling within either alpine, montane, or lowland habitat, rather than having to bin species arbitrarily in each category. This distinction is important for how biologists classify alpine species (Körner 2003) and allowed us to distinguish between species whose ranges were centered in alpine habitat (termed 'specialists'; defined below) from those whose ranges extended into the alpine zone but were centered elsewhere (termed 'generalists'; defined below and in **Figure 3.1**). Except when drawing comparisons with non-alpine species, the analyses presented here were performed only on species with elevational niches falling within the alpine (specialist or generalist) elevation category (**Figure 3.1**).

Assigning alpine specialist and generalist categories.

Using the SDMs and niche characterization described above, each of the 72,372 species in the dataset was assigned one of five, mutually exclusive, categories, as detailed in **Figure 3.1**: (1) alpine specialist; (2) alpine generalist-primarily montane; (3) alpine generalist-primarily lowland; (4) non-alpine primarily montane; or (5) non-alpine primarily lowland. Alpine specialists were defined as having at least 50% of their modelled range in alpine habitat. Alpine generalists had less than 50% but at least 5% of the modelled range in alpine areas. Non-alpine species had less than 5% of the modelled range within alpine regions. The sub-categories of primarily montane or lowland were assigned based on whether a majority of the remaining (i.e., non-alpine) range occurred in montane or lowland habitat, respectively. By categorizing the elevational strategy of every species in the dataset, we were also able to examine the distribution of elevational strategies for the closest relatives of each alpine species (phylogenetic relationships among species defined as described below), and to assess whether alpine specialists were more closely related to other specialists or species occurring primarily at lower elevations.

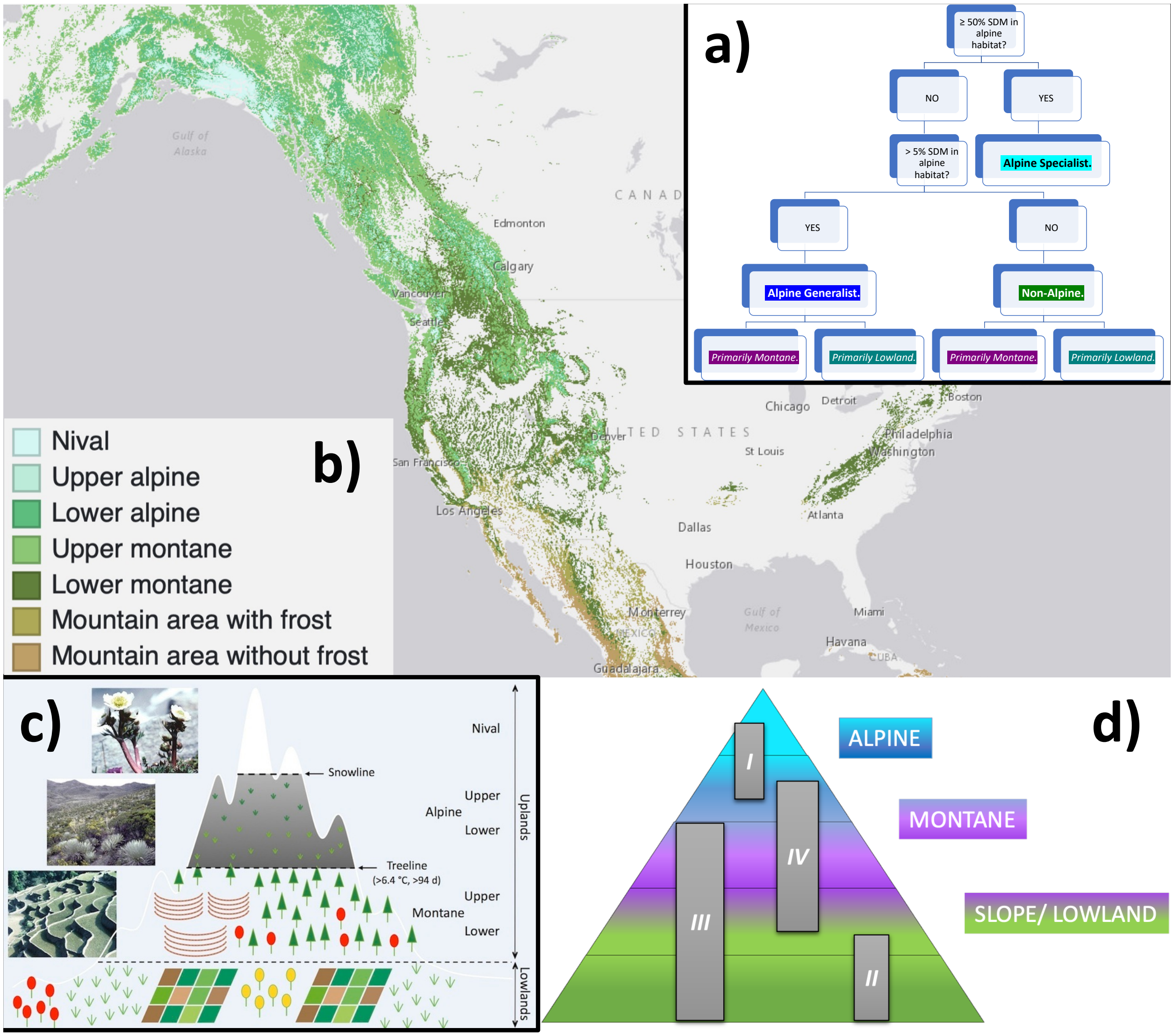


Figure 3.1. Conceptual illustration of elevational categories and species pool differences. (*a*) Decision tree used to categorize American seed plants as alpine specialists or generalists based on their species distribution model (SDM). The sub-categories of primarily montane or lowland were assigned based on where the majority of the non-

alpine range occurred. (*b*) The seven thermal life belts of Korner et al. (2017) shown in geographic context of North America as an example. (*c*) Schematic depiction of the seven thermal life belts of Korner et al. (2017) along a single mountain. The illustrations in (*b)* and (*c)* were adapted directly from (https://www.gmba.unibe.ch), which defined these categories. (*d*) Schematic illustrating how species (indicated by roman numerals) might be distributed along elevational gradients, with gray bars indicating the elevational niche breadth of each species. *Species I* has a range centered in the alpine belt and therefore would be classified as an alpine specialist. *Species II* has a range centered in lowlands that does not extend into the alpine region and would thus be a non-alpine and primarily lowland species. Both *Species III* and *IV* have ranges that extend into the alpine belt but are centered elsewhere and would therefore be classified as alpine generalists. The range of *Species IV* is centered in montane habitat, giving this species a sub-category of primarily montane.

Alpine species richness & biogeographic variation in specialization.

To investigate spatial diversity patterns of alpine species across the American mountain systems, we gridded North, Central, and South America (the ‘study area’) into 0.5° cells (termed ‘sites’). We then used a combination of approaches to determine species presence within a given site. First, we converted the SDMs describing species ranges into binary matrices indicating potential presence or absence (potential presence was defined as the SDM scoring at least 2% for that species in that site). We then masked these binary matrices by GMBA mountain shapefiles defining the geographic boundaries of all mountains and mountain ranges within the study area (Körner et al. 2017). These steps produced a potential species list for all sites within higher-elevation habitat. We then filtered potential lists by including only those species categorized as alpine specialists or generalists (**Figure 3.1**). Summarized alpine species lists (i.e., only including those species categorized as alpine specialist or generalist) based on these methods are deposited with Dryad (https://doi.org/10.5061/dryad.4qrfj6q8v), along with a taxonomy for these alpine species (**Online Resource 2**).

Species richness was obtained by counting the number of alpine species predicted to be present within each site. Each phylogenetic diversity metric was similarly calculated based on the alpine species predicted to be present at each site. We also provide a metric summarizing each site’s average degree of alpine habitat specialization. To assess this, we asked which species were present at each site as well as what fraction of each species’ total modelled range (i.e., across all American mountains) fell within alpine habitat. We then averaged this fraction of alpine habitat among all species present within each site.

Phylogeny, taxonomy, and phylogenetic diversity.

Phylogenetic relationships among American seed plants were defined by Smith and Brown (2018; ALLMB phylogeny), and polytomies were left unresolved. The backbone taxonomy from the World Checklist of Selected Plant Families (WCSP 2020) was used to match species to taxonomic families and orders as defined by the Angiosperm Phylogeny Group (APG IV 2016). Such taxonomic categories (e.g., in comparisons of relative occurrence) were used to highlight the distributions and prevalence of biologically relevant groups that are of interest to alpine botanists and not meant to imply that such rankings demarcate equivalent biological units. There were relatively few gymnosperms among the alpine species in our dataset, and none among alpine specialists. Among these, only *Lepidothamnus fonkii* and species of *Ephedra* were likely true alpine (i.e., non-tree) species. The remaining gymnosperms (*Pinus, Picea, Larix, Tsuga, Abies*, and *Thuja*) were tree species whose ranges extend to such high elevations that they sometimes fell within the GMBA lower alpine belt and were therefore classified as alpine generalists. As a result, we did not organize gymnosperm taxonomic families or orders as we did with angiosperms. However, because our broader dataset includes all American gymnosperms, these species are represented in the figures and phylogenetic diversity analyses where relevant.

Phylogenetic diversity was measured with the Phylogenetic Diversity Index (PDI), Mean Nearest Taxon Distance (MNTD), and Mean Pairwise Distance (MPD) using the 'PHYLOMEASURES' package (Tsirogiannis and Sandel 2017) in the R statistical computing environment (R Core Team 2020). This package provides functionality to standardize phylogenetic diversity metrics by species richness in order to overcome some of the non-independence between richness and phylogenetic diversity measures (Tsirogiannis and Sandel 2016; but see Sandel 2018). Sites with negative phylogenetic diversity values were interpreted as 'clustered', and sites with positive values as 'over-dispersed' (or phylogenetically 'even'). Although we acknowledge that quantitative comparisons of such diversity metrics across sites with very different richness and filtering/ selective processes should be taken with caution (Sandel 2018), here we have drawn general comparisons along the mountain systems of the Americas that do not depend on strict quantitative comparisons between individual sites. We assessed phylogenetic diversity among American alpine species (as defined in **Figure 3.1**), American high-elevation seed plant species (defined as having at least 5% of their SDM in montane or alpine habitat), and among all American seed plants in the larger dataset of Figueroa et al. (*in review*). Note that clustering of alpine sites here therefore indicates assemblages more closely related than expected in the context of American alpine lineages, whereas the assessment of phylodiversity in Qian et al. (2021) compares alpine phylodiversity to the surrounding communities at lower elevations.

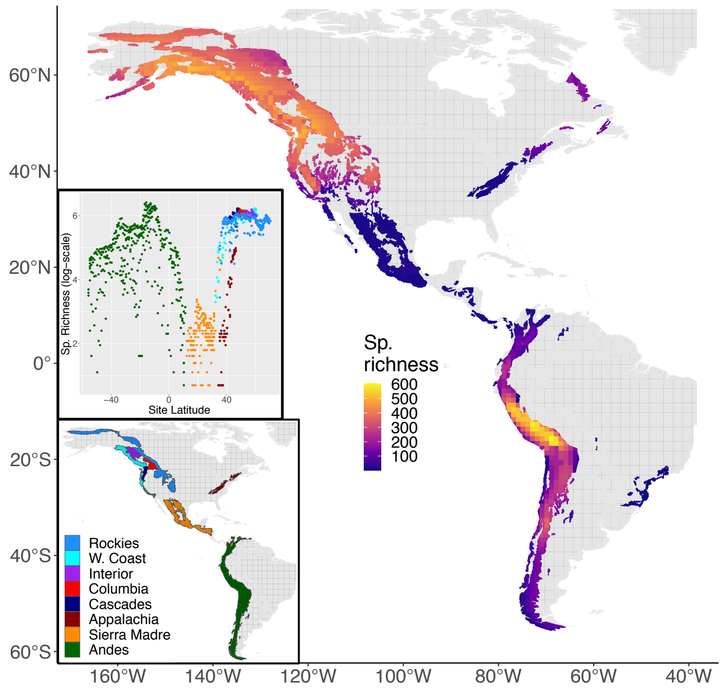


Figure 3.2. Heat map showing how alpine species richness across the Americas did not follow a simple latitudinal gradient. Richness was highest (warm tones) in both the central Andes and northern Rockies/ western North American ranges (in contrast, combined montane and alpine species richness peaked near the equator; see **Figure 3.S2**). Central American mountain ranges had lower numbers of alpine species (cool tones), which might indicate under sampling. *Scatter plot*: Alpine species richness vs. latitude across (equal area) sites within 8 major ranges emphasizing the bimodal aspect of alpine diversity. *Inset map*: Major mountain ranges described throughout this study (Appalachia, Andes, Cascades, Columbia, Interior, West Coast, and Rockies). *Color correspondences for these mountain ranges are consistent throughout.*

Dataset considerations.

Our data cleaning and filtering methods created a robust, biologically informed, and conservative estimate of the distributions of seed plants across the Americas well-suited for addressing general macro-scale questions of biodiversity patterns. However, because our methods involved excluding species with fewer than twelve occurrence records to have sufficient climate variability when building reliable models, there was the possibility of missing small-scale endemic species that might be of particular relevance to alpine biodiversity (discussed in Figueroa et al., *in review*). One type of bias that could occur with this method would be preferential loss of species from alpine communities. Examining the geographic distribution of seed plant records before and after the filtering step that excluded species with fewer than twelve points did not reveal any elevational biases based on the number of records available (**Figure 3.S1**). Geographic bias in species excluded by our method was another possibility, particularly for Andean alpine communities which are rich in endemic species (Hughes and Eastwood 2006). **Online Resource 3** provides a list of some Andean alpine taxa excluded by our conservative data cleaning and filtering protocol that demonstrate such a geographic bias in data availability (see Discussion).

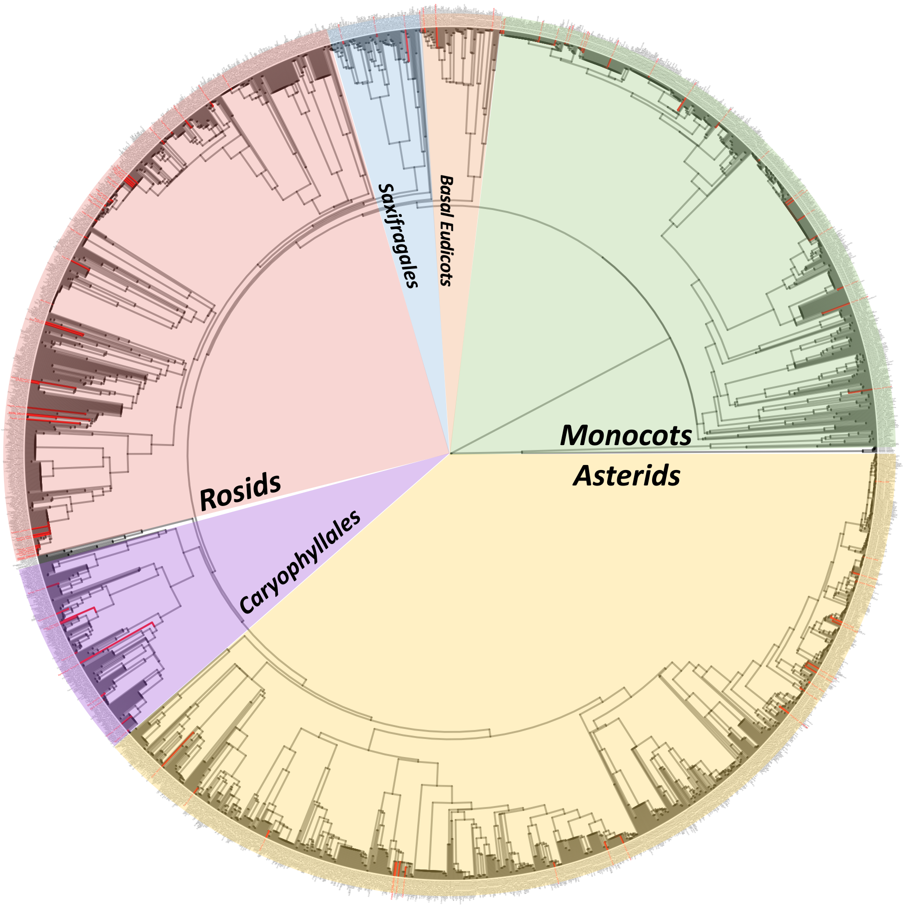


Figure 3.3. Phylogenetic distribution of all 2937 alpine species from North, Central, and South America in our dataset. Alpine specialist species (defined here as having at least 50% of the species' modelled range within alpine habitat) are marked in red. Most major plant groups (clades highlighted with colored panels) contain alpine specialists. ***Figure 3.S3*** *provides a phylogeny of just the alpine specialists in our dataset.* ***Table 3.S1*** *lists American angiosperm orders and families without alpine species.*

RESULTS

Dataset coverage and species richness.

Data cleaning and filtering methods resulted in a dataset of 2937 American alpine seed plants (~7% of all higher-elevation species, alpine or montane, N=35,952). Of these, 96 (representing 54 genera) were classified as alpine specialists, whose modelled range was centered (i.e., at least 50%) within alpine habitat. The remainder ($N_{species}$=2841; representing 709 genera) were classified as alpine generalists (i.e., between 5% and 50% of modelled range within alpine regions). Alpine species richness did not follow a unimodal latitudinal gradient (**Figure 3.2**; see **Figure 3.S2** for combined alpine and montane American seed plant species richness). Instead, richness was highest in both the northern Rockies/ western ranges of North America, and the central Andes of South America (i.e., species richness was latitudinally bimodal). Summarized species lists of alpine communities within each mountain and major mountain ranges were deposited with Dryad (https://doi.org/10.5061/dryad.4qrfj6q8v).

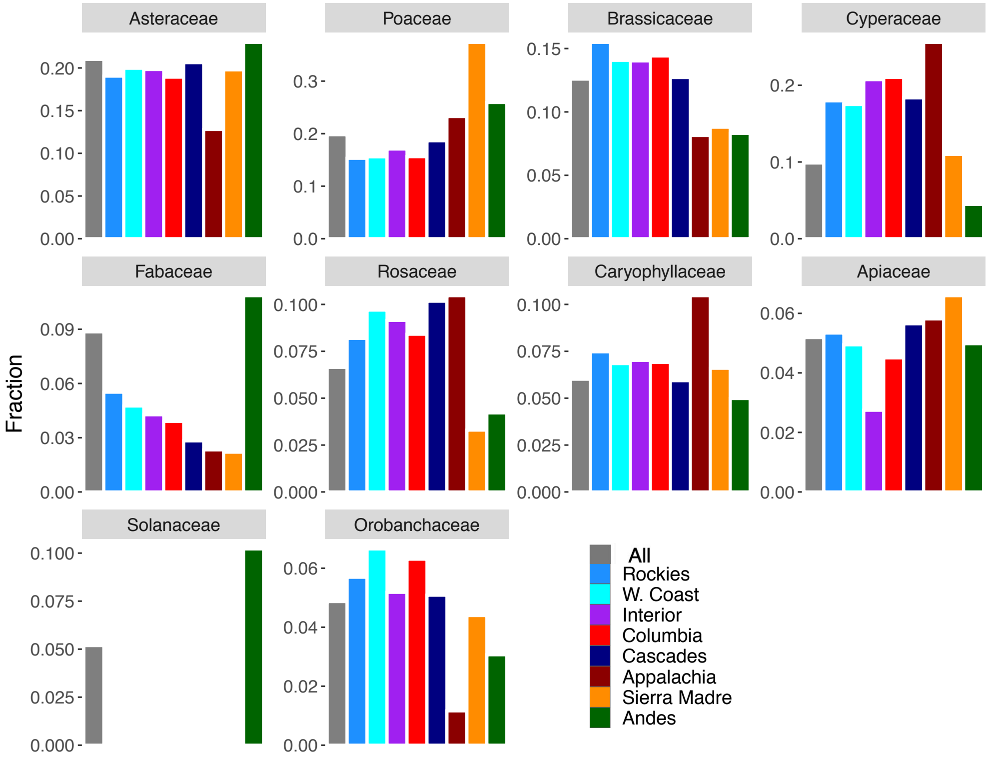


Figure 3.4. Distribution of the top ten most dominant taxonomic families with alpine species across eight major American mountain range systems (colors indicate Appalachia, Andes, Cascades, Columbia, Interior, West Coast, and Rockies). At this spatial scale, the relative frequencies of these families varied substantially between mountain ranges. For instance, Cyperaceae were dominant in all mountain ranges except the Andes and Sierra Madre, despite being only the fourth most frequent family when looking across all mountain ranges. Poaceae dominated in the Central American mountains to a greater extent than they did in other regions, and Solanaceae were absent from all major mountain range systems except the Andes. *Families are ordered by their frequency across the entire alpine species dataset (e.g., Asteraceae possess more alpine species than any other family), as indicated by gray bars (note each panel has an independent vertical axis).*

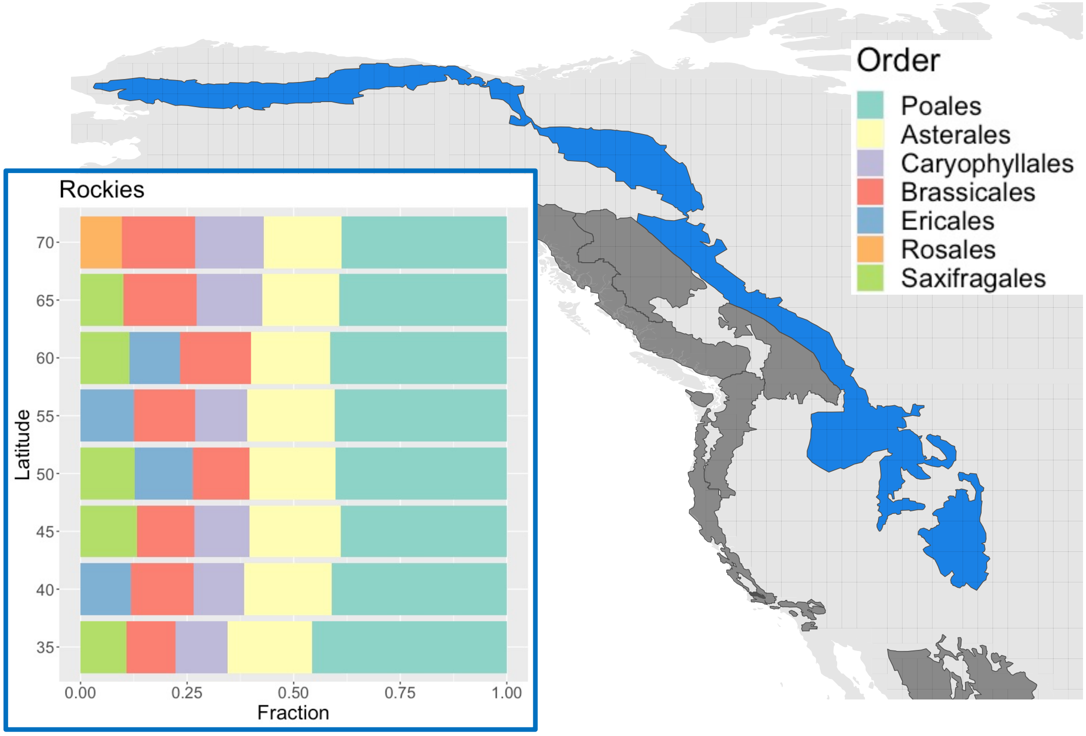
Phylogenetic distribution of alpine taxa.

Alpine species were distributed broadly across the American seed plant phylogeny, with specialists occurring in most major angiosperm clades (**Figure 3.3** and **Figure 3.S3**). Overall, 717 genera contained at least one alpine species (specialist or generalist), spread across 131 taxonomic families (41% of all families represented in our dataset). Nevertheless, alpine species were not distributed evenly across clades. Ten families (**Figure 3.4**) contained ~50% of all alpine species, and Asteraceae and Poaceae jointly accounted for just under 23%. Additionally, many families (~59%, N=187) represented in the American angiosperm phylogeny did not include a single alpine species (**Table 3.S1a,b**), and, in fact, entire orders were without alpine species in the Americas (~31%, or N=20, orders present in the dataset but absent from alpine communities). **Online Resource 2** provides a taxonomy of alpine species included in this study along with their familial and ordinal placements as defined by APG IV.

Spatial distribution of alpine taxa.

At the family level, alpine diversity varied among different American mountain ranges. Although ten families comprised almost half of all alpine species in our dataset, the relative frequencies of these ten families differed markedly between mountain ranges (**Figure 3.4**). For instance, Solanaceae in our dataset occurred exclusively in the Andes, and Poaceae dominated in the Sierra Madre. Although Cyperaceae ranked as only 4th overall among all alpine species, this family was either the first or second most species-rich family in all major mountain ranges except the Andes and Sierra Madre. Additionally, familial prevalence varied according to the mountain range (**Figure 3.S4**). For instance, Poaceae were more frequent in Central and South American mountains whereas Asteraceae occurred more commonly in North America.

Distributional biodiversity patterns at the level of plant orders mostly mirrored those of the overall bimodal diversity across all alpine species (**Figure 3.S5**). However, some orders, such as Fabales and Lamiales, were much more diverse in the Andes, whereas other orders, such as Ericales, Brassicales, and Saxifragales, had higher diversity in the northern Rockies and West Coast mountains. Regardless of spatial scale (individual mountains or mountain ranges), Poales were always the most dominant order. **Figure 3.5** highlights latitudinal turnover in alpine community composition for the Rockies and the Andes (see **Figure 3.S6** for other mountain range systems).


Rockies
Latitude
70
65
60
55
50
45
40
35
0.00
0.25
0.50
0.75
1.00
Fraction
Order
Poales
Asterales
Caryophyllales
Brassicales
Ericales
Rosales
Saxifragales

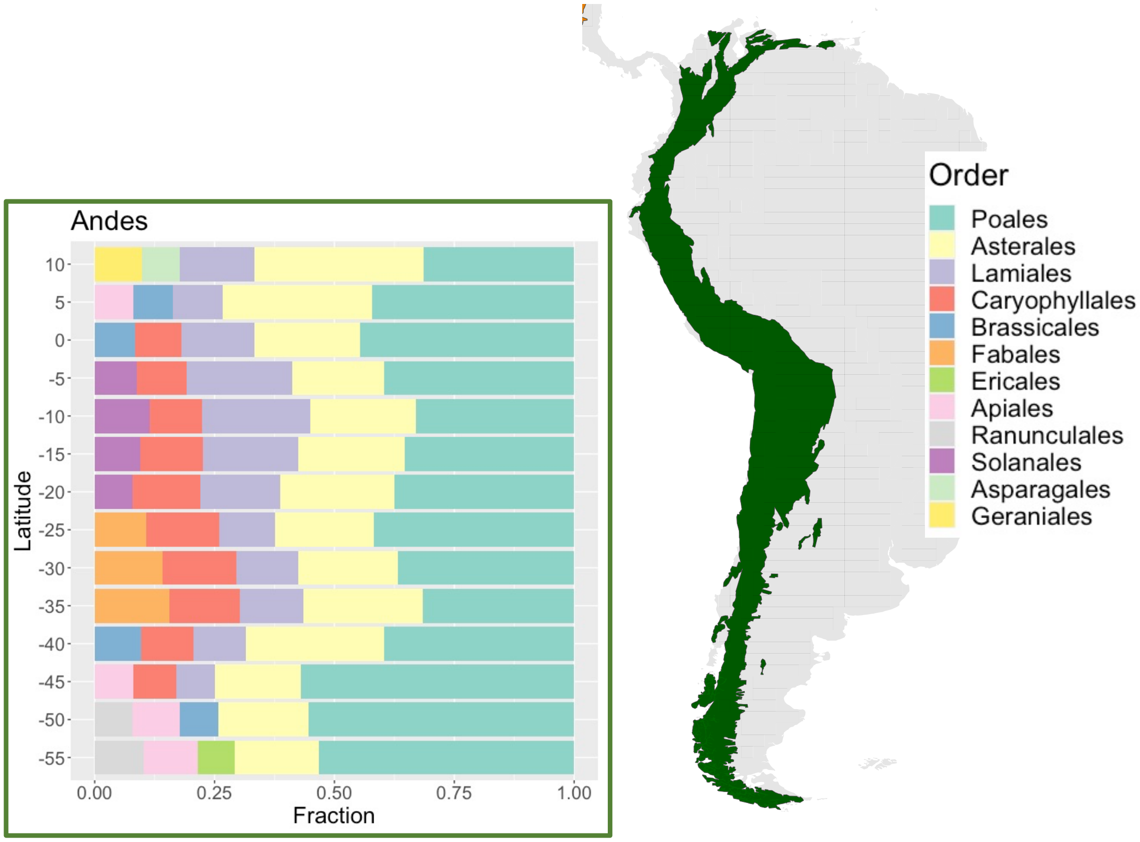

Andes
Latitude
10
5
0
-5
-10
-15
-20
-25
-30
-35
-40
-45
-50
-55
0.00
0.25
0.50
0.75
1.00
Fraction
Order
Poales
Asterales
Lamiales
Caryophyllales
Brassicales
Fabales
Ericales
Apiales
Ranunculales
Solanales
Asparagales
Geraniales

Figure 3.5. Patterns of ordinal-level alpine community turnover over the latitudinal extent of the Rockies (*top panel, blue*), and the Andes (*bottom panel, green*). Bar plots show the relative dominance of the top five most abundant orders among communities found within 5° latitudinal sections of each mountain range. Community composition varied between the Rockies and Andes; for example, Solanales were absent in alpine communities from the Rockies, and Saxifragales were never among the most dominant Andean orders. Dominance patterns also varied along the latitudinal extent of each mountain range. For example, the relative dominance of Poales increased moving southward along the Andes while that of Asterales decreased. Some orders, such as Caryophyllales and Saxifragales in the Rockies, varied in their relative frequency along latitudinal ranges and are not represented as top taxa in all zones.

Phylogenetic diversity and alpine specialization.

This montane and alpine dataset is nested within a larger dataset describing the climatic niches of 72,372 American seed plants (described elsewhere, Figueroa et al. *in review*). To examine the phylogenetic affinities of alpine specialists, we summarized the elevational strategies of each specialist species' phylogenetic nearest neighbor within the American seed plant phylogeny. Among alpine specialists in our dataset, ~62% (N=59) did not have another specialist among their closest relatives, and for 89% (N=85) the most common strategy among nearest phylogenetic neighbors was a primarily montane distribution.

We measured phylogenetic diversity among American alpine species within each range using species-richness standardized versions of the Phylogenetic Diversity Index (PDI), Mean Nearest Taxon Distance (MNTD), and Mean Pairwise Distance (MPD; **Figure 3.6**). The majority of the Rockies and Alaskan ranges were significantly more phylogenetically clustered than expected by chance (i.e., negative PDI values), whereas the Sierra Madre were phylogenetically overdispersed (i.e., positive PDI values). For the Andes, we observed mixed results, though sites tended to be more clustered; in particular, the central Andes showed significant phylogenetic clustering. MNTD was closely correlated with PDI ($R^2=0.94$, $p<0.001$) and showed similar spatial patterns. In contrast, however, MPD was significantly greater in the mid-Rockies/ western North American mountains. Overall, the Rockies were characterized by low phylogenetic diversity and MNTD but higher MPD. In the Andes, the central cordillera was characterized by low PDI, MNTD, and MPD, whereas the southern cordillera had moderately high values of all three metrics, indicating a general phylogenetic diversity gradient along the Andean mountains, increasing from north to south. These patterns were particular to alpine communities in our study area and were in contrast with the overall patterns of phylogenetic diversity among all angiosperms in the Americas as well as patterns observed when montane species were included (**Figure 3.S7** and **3.S8**; see Discussion).

Additionally, we examined spatial variation in the degree of alpine specialization (**Figure 3.6d**) by calculating the average amount of American alpine habitat occupied by all species found within sites (see Methods). The northern Rockies and Alaskan Ranges had greater proportions of alpine specialization, whereas the Sierra Madre region was more composed of generalists. In other words, the northern Rockies/Alaskan ranges were composed of a greater fraction of phylogenetically clustered alpine specialists and the Sierra Madre represented more phylogenetically overdispersed generalists. In contrast to the patterns of increased specialization near the Arctic (above approx. 50° N), sites in the extreme Southern Andes (below approx. 40° S) contained a greater fraction of generalists. Thus, the alpine flora of boreal/near-Arctic regions differed in phylogenetic diversity patterns and degree of alpine specialization from the Patagonian region of the Andes. In general, phylogenetic diversity was significantly and negatively correlated with the degree of alpine specialization (R^2=0.59 across all mountains; $p<0.001$). The strength of this relationship was markedly greater when considering only the major mountain ranges (**Figure 3.6d**; R^2=0.76; $p<0.001$).

DISCUSSION

Dataset considerations.

In this study, we offer a detailed investigation of alpine seed plant diversity across the Americas using occurrence records to build SDMs characterizing the extent of alpine habitat encountered by American seed plants. Our methods to build SDMs allowed us to overcome some of the limitations of incomplete sampling inherent in the direct use of georeferenced point occurrence records by providing broader geographic distributions of species (Barthlott et al. 2007; Meyer et al. 2015; Brummitt et al. 2021) and to describe the elevational niches of plants in our dataset in a graded and more biologically meaningful way. To build reliable climatic niche models, we employed a threshold minimum requirement of twelve records per species. Although this did not appear to introduce any elevational biases through the preferential exclusion of alpine taxa (**Figure 3.S1**), it certainly excluded numerous range-limited species, such as endemics. Further, this conservative occurrence record threshold is likely to disproportionately impact representation of Andean alpine lineages, where the degree of alpine endemism is substantial (Hughes and Eastwood 2006), and we offer a partial list of such potential exclusions in **Online Resource 3**. Thus, our dataset also represents an assessment of where focused collecting efforts and increased sampling of alpine plants could enhance our understanding of alpine biodiversity.

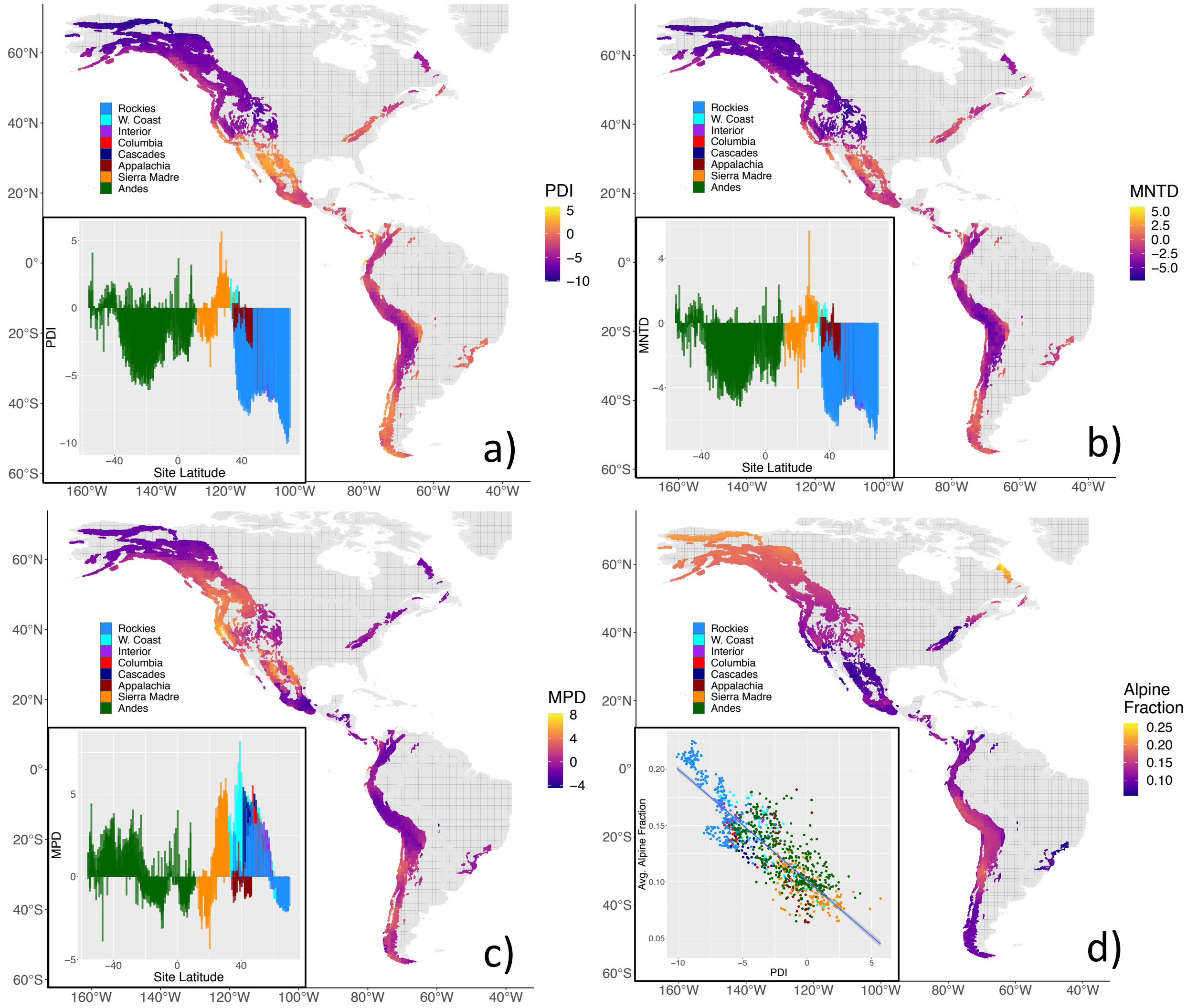


Figure 3.6. Phylogenetic diversity of American alpine species and the degree of alpine specialization across American mountain ranges. Each value was calculated among all species present within sites across the study area. *(a)* Phylogenetic Diversity Index (PDI) values indicate either phylogenetic clustering (PDI<0; observed in the Rockies) or overdispersion (PDI>0; observed in the Sierra Madre and Patagonian Andes). The inset plot shows PDI values at each site vs. latitude, with bars colored according to major mountain ranges. *(b)* Mean Nearest Taxon Distance (MNTD) generally mirrored the spatial patterns observed for PDI. *(c)* Mean Pairwise Distance (MPD), in contrast, was relatively high in the Rockies and western North America. Negative (positive) values of both MNTD and MPD were interpreted as indicating sites less (more) closely related than

expected by chance. Inset plots as in (*a*). *(d)* Community-averaged degree of alpine specialization. Inset plot shows that regions with low PDI (e.g., the northern Rockies) also had a greater degree of alpine-specialized species, and vice versa (R^2=0.76 for major mountain ranges, or R^2=0.59 across all mountains; p<0.001 in both cases). In this way, alpine species found in near-Arctic/ boreal regions (above approx. 50° N; typified by low PDI and higher degree of specialization) differed from southern Patagonian alpine species (below approx. 40° S; typified by higher PDI and a greater proportion of generalists). *Plots represent metrics only among alpine species (both generalists and specialists); see **Figure 3.S7** and **3.S8** for the same metrics across all American angiosperms (e.g., including lowland taxa and montane endemics) and combined montane and alpine species in our dataset.*

With these limitations in mind, we have attempted to draw broad, macroecological conclusions that are unlikely to be significantly impacted by lack of distribution models for range-limited endemics. Nevertheless, species excluded as a result of the conservative nature of our data cleaning protocol draw critical attention to the need for increased sampling of range-limited endemics to enhance the broader understanding of how such species impact macroecological patterns and conclusions. Further, although we have used our dataset to draw comparisons among alpine assemblages occurring across the entire American mountain range system, important differences between Northern and Southern Hemisphere alpine habitats, such as the degree of seasonality, growing season length, and extent and duration of snow cover, should be taken into account when forming conclusions from this macroecological dataset. Future studies should investigate how more local-scale processes (e.g., orogeny, climate, erosion) acting within individual mountain regions might be influencing the observed macroecological patterns we find here.

Biodiversity patterns.

Using the continuous extent of the SDMs we were able to distinguish between species whose ranges were centered in alpine habitat (termed “specialists” here) from those whose ranges extended above the treeline but were centered elsewhere (“generalists”). These distinctions have been important for how biologists classify alpine species (Körner 2003), which we have previously corroborated by showing statistically significant differences between the temperature niches of alpine seed plant specialists and generalists across the Americas (Figueroa et al. *in review*). Nevertheless, our SDM approach relied solely on American occurrence records. Incorporating global records might change the categorization of specialist and generalist species presented here, as some taxa might be generalists in the Americas but quite restricted to alpine habitat elsewhere or vice versa.

Generally, alpine seed plant diversity was latitudinally bimodal, with the greatest species richness occurring in sites within the central Andes and northern Rockies/western mountains of North America (**Figure 3.2**). These diversity patterns contrasted with those observed among seed plants generally or when combining alpine and montane species (**Figure 3.S2**), which in both cases were centered on the equator. It is possible that these differences were influenced by under-sampling in Central American alpine communities and the southwest region of North America, and we suggest that increased sampling in these areas might be of high value to alpine biologists. At the same time, these results have important implications for viewing western North American mountain ranges as important centers of alpine biodiversity in conservation efforts (Myers et al. 2000; IPBES 2019), particularly as North American alpine diversity stayed relatively high across a much greater latitudinal extent than observed in the Andes. This contrast between North and South alpine diversity patterns could reflect differing biogeographic histories or filtering effects operating in these two geographically disparate regions, and potentially suggests that dispersal between North and South American mountain ranges may be more limited for alpine plants than montane ones—a hypothesis which should be tested explicitly in future studies.

Lineage-specific patterns.

Although we found that most major American seed plant lineages contained at least one alpine species (generalist or specialist; **Figure 3.3**, **Figure 3.S3**), there was strong unevenness in the phylogenetic distribution of alpine species. Additionally, there were substantial differences in the taxonomic composition of different mountain ranges (**Figure 3.4**) and latitudinal turnover in richness patterns of alpine communities within individual mountain ranges (**Figure 3.5**, **Figure 3.S6**). Although grouping species by higher taxonomic ranks does not result in comparisons of equivalent biological units, by presenting results on specific clades, we highlight groups that may be of particular biological relevance to alpine botanists. For example, in the Andes, Fabaceae and Solanaceae were relatively more dominant than in any other mountain range (**Figure 3.4**). In fact, Solanaceae was absent from alpine communities outside the Andes, and yet was dominant enough within the Andes that this family still ranked among the top ten families contributing most to American alpine biodiversity overall, congruent with previous biogeographic studies in this family (Olmstead et al. 2008; Dupin et al. 2017).

Phylogenetic and biogeographic analyses of Solanaceae (reviewed in Olmstead 2013) suggest a South American origin for this family. The modern distributions of Solanaceae clades appear to be limited ecologically, rather than by dispersal, with modern ranges shaped by relatively strong niche conservatism of both cold and xeric intolerance (Olmstead 2013). Thus, the absence of Solanaceae from non-Andean mountain ranges might be due to ecological/ physiological limits rather than dispersal limitation. This might also explain why, despite their relatively higher frequency in the Andes, none of the Solanales in our dataset (N=84) are alpine specialists. The modern Solanaceae distributions thus reflect the influence of both abiotic filtering (limiting their latitudinal and elevational extents) as well as historical processes (substantially greater diversity in South America).

Excluding its higher diversity in the Andes, Fabaceae diversity otherwise declined moving southward along the American mountain ranges. In Northern Hemisphere sites, Fabaceae diversity was driven primarily by the *Astragalus/Oxytropis* complex, whose origins are likely broadly Eurasian (e.g., Bagheri et al. 2017; Amini et al. 2019). This could suggest the high diversity of *Astragalus/Oxytropis* species in the Americas is ultimately due to dispersal from Siberia and surrounding regions. Moving southward, Fabaceae diversity declined due to gradual loss of *Oxytropis* and *Hedysarum*. In contrast, the relative frequencies of *Astragalus* and *Lupinus* remained fairly high and consistent across the mountain ranges. In the Andes, the relatively higher Fabaceae prevalence was driven jointly by *Astragalus* and *Lupinus*, as well as several (N=10) Andean-specific genera, primarily *Adesmia* (the most species-rich Andean alpine Fabaceae genus in our dataset) but also less prevalent groups such as *Anarthophyllum* and *Desmodium*. These patterns highlight the scale-dependent interplay of how abiotic filtering and historical contingency shape species turnover among alpine communities.

Despite differences in taxonomic composition between mountain ranges, some broad similarities existed as well. Regardless of spatial scale (i.e., mountain ranges, individual mountains, or latitudinal sections within mountain ranges), Poales were always the most dominant order in American alpine communities, driven jointly by the presence of Poaceae and Cyperaceae in these sites. This result is perhaps not surprising given that Poales tend to dominate in open and dry habitats, and recent work has suggested this clade may have occupied such habitats since the Cretaceous (Bouchenak-Khelladi et al. 2014), which might have preadapted some lineages to success in alpine habitat. The only exception to this pattern was in the northernmost Andes (**Figure 3.5**), where Asterales prevailed. Here, Poales dominance increased moving north to south along the Andes, driven primarily by a general increase in the relative frequency of genera in the Poaceae moving southward, and not the total number of genera representing each order (for example, there were more Asteraceae than Poaceae genera in the Andes at all latitudes above 40° S) or abrupt changes in the genera present along the latitudinal extent of the Andes.

Phylogenetic diversity patterns.

Our phylogenetic diversity (PDI, MNTD, MPD) and alpine specialization metrics provided a nuanced picture of differences among alpine communities occurring in different American regions (**Figure 3.6**). Near-Arctic/ boreal alpine communities (above approx. 50° N) were typically phylogenetically clustered with a greater degree of alpine specialization, whereas southern Patagonian communities (below approx. 40° S) were more phylogenetically even with a lesser degree of specialization. These patterns differentiate the alpine communities occurring at opposite ends of the American mountain range system, despite broad climatic similarities in the two regions, and suggest an important role for historical factors in the assembly of their floras. Additionally, phylogenetic diversity in the Andes was generally lowest in the central region, and increased moving southward, where alpine specialization was the lowest (**Figure 3.6**). The alpine communities of the Andean mountain range were therefore characterized by their own latitudinal phylogenetic diversity gradient. Across all major mountain ranges, we also found a significant and strongly negative correlation ($R^2=0.76$; $p<0.001$) between a community's phylogenetic diversity and degree of alpine specialization (**Figure 3.6d**). Thus, sites with more phylogenetic clustering were also composed of species exhibiting a greater degree of alpine specialization.

A general trend of increasingly older divergence time estimates moving southward along the Andes, consistent with Andean uplift occurring from south to north, has been previously reported (Luebert and Weigend 2014). Our findings of greater phylogenetic diversity in southern Andean alpine communities are consistent with this reported general trend as well as some clade-specific specific phylogeographic studies on diversification along the Andes (Picard et al. 2008).

When alpine and montane species were combined, the southern Andes still showed higher levels of phylogenetic diversity compared to other Andean sites (**Figure 3.S7**); however, the trend was not as strong.

Historical constraint or abiotic filtering?

A key question concerning alpine community assembly is whether abiotic filtering or historical contingency (e.g., phylogenetic and biogeographic history) plays a greater role (e.g., Hughes and Eastwood 2006; Flantua et al. 2019). Invariably, the answers to such dichotomous questions are that both factors contribute. Nevertheless, the relatively low number of alpine specialists (~3% of all alpine species in our dataset) might suggest that abiotic filtering dominates over shorter evolutionary timescales, preventing larger numbers of species from having their range centered in the alpine belt. This might also explain why sites with greater alpine specialization were also more phylogenetically clustered (**Figure 3.6d**)—if there are strong abiotic filters to alpine specialization, then it may be that only certain lineages are able to pass through the filter (Marx et al. 2017). Notably, however, diversity metrics are influenced by both the spatial and taxonomic scale of the observations (e.g., Swenson et al. 2006; Marx et al. 2019). By comparing phylogenetic diversity among solely alpine communities (**Figure 3.5**) to both combined alpine-montane communities and across all seed plants (**Figure 3.S7**), we were able to gain a better understanding for how alpine biodiversity differs in phylogenetic structure along elevational gradients and among spatially proximate communities.

Consistent with the idea of strong filtering shaping alpine communities, we found that most alpine specialists (~62%, N=59) did not have another specialist among their closest relatives. The most common strategy among the closest relatives of alpine specialists, instead, was a primarily montane distribution (89%, N=85). This was in contrast to alpine generalists, for which ~20% were most closely related to primarily lowland taxa (data not shown). This suggests that lowland lineages might contribute proportionally more to alpine generalist lineages, whereas alpine specialists have a greater phylogenetic affinity with lineages that are already occupying montane habitats. Additionally, the discussion on whether and in what ways mountain uplift has helped shaped higher-elevation biodiversity patterns is often hampered by our ignorance of the time lag between orogenic effects and species' ability to colonize and survive in novel environments (e.g., Luebert and Weigend 2014). The extent of phylogenetic niche conservatism observed here could be helpful to this discussion insofar as it suggests that, despite older lineages being present in many habitats, the actual number of alpine species is quite low (~7% of higher-elevation species) and therefore the lag times could be considerable.

Given the strong asymmetries in how lineages are distributed geographically, the species composition found in a given mountain range will also surely be strongly influenced by phylogenetic history in that region. For instance, Poales dominated all mountain ranges at the regional level (**Figure 3.5**, **Figure 3.S6**). On the one hand, this could suggest a general ability of Poales to be successful in alpine habitats, consistent with the notion of abiotic filtering driving the composition of alpine communities. However, the individual genera giving rise to the dominance of Poales differed among ranges (**Figure 3.S4**), with *Carex* (Cyperaceae) being most dominant in the Rockies, *Poa* (Poaceae) in the Andes, and jointly *Carex/Agrostis* (Cyperaceae/Poaceae, respectively) in the Sierra Madre. On the other hand, these results do not negate the notion of a general ability of Poales taxa to disperse to alpine communities. For instance, the dominance of certain species-rich Poalean families, such as Poaceae and Cyperaceae, in open and dry habitats has been attributed to general physiological mechanisms of CO_2 concentration that enhance survival in these ecosystems (Bouchenak-Khelladi et al. 2014). However, these findings do suggest a more nuanced and scale-dependent perspective of the interplay between phylogeny, historical contingency, and abiotic filtering in the assembly of alpine communities (rather than a strictly dichotomous viewpoint).

Conclusions.

We have provided a detailed appraisal of American alpine seed plant diversity and major biogeographic patterns given the context of our current sampling of these notoriously remote regions. Employing a conservative and robust data cleaning protocol (Soltis and Soltis 2016; Figueroa et al., *in review*), we highlighted how species distribution models can enhance our understanding of alpine biodiversity. Our results also call attention to key issues in current data availability, such as the need for increased sampling of both range-limited Andean endemics and alpine communities in and around Central America, as well as the conservation importance of the western North American cordillera for alpine biodiversity. In general, we found alpine diversity to be greatest in the central Andes and western North America, and that sites with lower phylogenetic diversity were composed of species with a greater degree of alpine specialization.

Additionally, contrasting patterns of phylogenetic diversity may indicate different community assembly processes acting in otherwise climatically similar Arctic/boreal alpine communities and Patagonian ones, such that that abiotic filtering alone cannot explain these biodiversity patterns. At the same time, the relative rarity of alpine specialists and the tendency for alpine species to be most closely related to montane lineages suggest that filtering is still an important determinant of alpine community structure. This work corroborates the importance of a nuanced and scale-dependent perspective on this ‘history-filtering’ debate axis, as both factors have likely contributed to modern biodiversity patterns observed in alpine seed plant communities across the Americas.

REFERENCES

Al-Shehbaz, I.A. (2018) A monograph of the South American species of Draba (Brassicaceae). *Annals of the Missouri Botanical Garden* 103(4): 463-590.

Amini E, Kazempour-Osaloo S, Maassoumi AA, Zare-Maivan H (2019) Phylogeny, biogeography and divergence times of Astragalus section Incani DC. (Fabaceae) inferred from nrDNA ITS and plastid rpl 32 -trn L (UAG) sequences. *Nord J Bot* 37: https://doi.org/10.1111/njb.02059.

Amiri, M. S., Joharchi, M. R., & Nadaf, M. (2020) Ethnobotanical knowledge of Astragalus spp.: The world's largest genus of vascular plants. *Avicenna J Phytomed.* 10(2): 128-142.

Antonelli A, Daniel Kissling W, Flantua SGA, et al (2018) Geological and climatic influences on mountain biodiversity. *Nat Geosci* 11:718–725.

Antonelli, A., Nylander, J. A. A., Persson, C., & Sanmartín, I. (2009) Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proceedings of the National Academy of Sciences*, 106(24): 9749–9754. https://doi.org/10.1073/pnas.0811421106.

Bagheri A, Maassoumi AA, Rahiminejad MR, et al (2017) Molecular phylogeny and divergence times of Astragalus section Hymenostegis: An analysis of a rapidly diversifying species group in Fabaceae. *Sci Rep* 7:14033.

Barthlott W, Hostert A, Kier G, et al (2007) Geographic Patterns of Vascular Plant Diversity at Continental to Global Scales (Geographische Muster der Gefäßpflanzenvielfalt im kontinentalen und globalen Maßstab). *Erdkunde* 61:305–315.

Bouchenak-Khelladi Y, Muasya AM, Linder HP (2014) A revised evolutionary history of Poales: origins and diversification: Evolutionary History of Poales. *Bot J Linn Soc* 175:4–16.

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

Brummitt RK (2001) World geographical scheme for recording plant distributions, Ed. 2. Biodiversity Information Standards (TDWG).

Ding W-N, Ree RH, Spicer RA, Xing Y-W (2020) Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. *Science* 369:578–581.

Dupin J, Matzke NJ, Särkinen T, et al (2017) Bayesian estimation of the global biogeographical history of the Solanaceae. *J Biogeogr* 44:887–899.

Elsen PR, Tingley MW (2015) Global mountain topography and the fate of montane species under climate change. *Nat Clim Chang* 5:772–776.

Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas: NEW CLIMATE SURFACES FOR GLOBAL LAND AREAS. *Int J Climatol* 37:4302–4315.

Flantua SGA, O'Dea A, Onstein RE, et al (2019) The flickering connectivity system of the north Andean páramos. *J Biogeogr* 46:1808–1825.

GMBA (2010) The digital "Mountain Biodiversity Portal" (https://www.gmba.unibe.ch).

González-Caro, S., Duque, Á., Feeley, K. J., Cabrera, E., Phillips, J., Ramirez, S., & Yepes, A. (2020). The legacy of biogeographic history on the composition and structure of Andean forests. *Ecology*, 101(10). https://doi.org/10.1002/ecy.3131.

Guo Q, Kelt DA, Sun Z, et al (2013) Global variation in elevational diversity patterns. *Sci Rep* 3:3007.

Hughes C, Eastwood R (2006) Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proc Natl Acad Sci USA* 103:10334–10339.

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

IPBES (2019) Global assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. 1753.

Kelly AE, Goulden ML (2008) Rapid shifts in plant distribution with recent climate change. *Proc Natl Acad Sci USA* 105:11823–11826.

Kier G, Kreft H, Lee TM, et al (2009) A global assessment of endemism and species richness across island and mainland regions. *Proc Natl Acad Sci USA* 106:9322–9327.

Körner C (1995) Alpine Plant Diversity: A Global Survey and Functional Interpretations. In: Chapin FS, Körner C (eds) *Arctic and Alpine Biodiversity: Patterns, Causes and Ecosystem Consequences*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 45–62.

Körner C (2004) Mountain biodiversity, its causes and function. *Ambio Spec No* 13:11–17.

Körner C (2003) *Alpine Plant Life: Functional Plant Ecology of High Mountain Ecosystems*, 2nd edn. Springer-Verlag Berlin Heidelberg.

Körner C, Jetz W, Paulsen J, et al (2017) A global inventory of mountains for bio-geographical applications. *Alpine Botany* 127:1–15.

Körner C, Paulsen J (2004) A world-wide study of high altitude treeline temperatures: Study of high altitude treeline temperatures. *J Biogeogr* 31:713–732.

Körner C, Paulsen J, Spehn EM (2011) A definition of mountains and their bioclimatic belts for global comparisons of biodiversity data. *Alpine Botany* 121:73-78.

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

Luebert F, Weigend M (2014) Phylogenetic insights into Andean plant diversification. *Frontiers in Ecology and Evolution* 2:27. https://doi.org/10.3389/fevo.2014.00027.

Marx, H. E., Dentant, C., Renaud, J., Delunel, R., Tank, D. C., & Lavergne, S. (2017) Riders in the sky (islands): Using a mega-phylogenetic approach to understand plant species distribution and coexistence at the altitudinal limits of angiosperm plant life. *J Biogeogr* 44(11), 2618–2630. https://doi.org/10.1111/jbi.13073.

Marx, H. E., Richards, M., Johnson, G. M., & Tank, D. C. (2019) Increasing phylogenetic stochasticity at high elevations on summits across a remote North American wilderness. *American Journal of Botany* 106(7): 958-970.

McFadden IR, Sandel B, Tsirogiannis C, et al (2019) Temperature shapes opposing latitudinal gradients of plant taxonomic and phylogenetic β diversity. *Ecol Lett* 22:1126–1135.

Meyer C, Kreft H, Guralnick R, Jetz W (2015) Global priorities for an effective information basis of biodiversity distributions. *Nat Commun* 6:8221.

Millar C, Fagre DB (2007) Monitoring alpine plants for climate change: The North American GLORIA Project. Mountain Views 1.

Moret P, Muriel P, Jaramillo R, Dangles O (2019) Humboldt's Tableau Physique revisited. *Proc Natl Acad Sci USA* 116:12889–12894.

Morueta-Holme N, Engemann K, Sandoval-Acuña P, et al (2015) Strong upslope shifts in Chimborazo's vegetation over two centuries since Humboldt. *Proc Natl Acad Sci USA* 112:12741–12745.

Muellner-Riehl AN, Schnitzler J, Kissling WD, et al (2019) Origins of global mountain plant biodiversity: Testing the "mountain-geobiodiversity hypothesis." *J Biogeogr* 46:2826–2838.

Myers N, Mittermeier RA, Mittermeier CG, et al (2000) Biodiversity hotspots for conservation priorities. *Nature* 403:853–858.

Olmstead RG (2013) Phylogeny and biogeography in Solanaceae, Verbenaceae and Bignoniaceae: a comparison of continental and intercontinental diversification patterns: Solanaceae, Verbenaceae and Bignoniaceae. *Bot J Linn Soc* 171:80–102.

Olmstead RG, Bohs L, Migid HA, et al (2008) A molecular phylogeny of the Solanaceae. *Taxon* 57:1159–1181.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecol Modell* 190:231–259.

Phillips SJ, Dudík M, Schapire RE (2004) A maximum entropy approach to species distribution modeling. In: Proceedings of the twenty-first international conference on Machine learning. Association for Computing Machinery, New York, NY, USA, p 83.

Picard, D., Sempere, T., & Plantard, O. (2008) Direction and timing of uplift propagation in the Peruvian Andes deduced from molecular phylogenetics of highland biotaxa. *Earth and Planetary Science Letters*, 11.

Qian, H., Ricklefs, R. E., & Thuiller, W. (2021). Evolutionary assembly of flowering plants into sky islands. *Nature Ecology & Evolution*. https://doi.org/10.1038/s41559-021-01423-1.

Rahbek C, Borregaard MK, Antonelli A, et al (2019a) Building mountain biodiversity: Geological and evolutionary processes. *Science* 365:1114–1119.

Rahbek C, Borregaard MK, Colwell RK, et al (2019b) Humboldt's enigma: What causes global patterns of mountain biodiversity? *Science* 365:1108–1113.

R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Sandel B (2018) Richness-dependence of phylogenetic diversity indices. *Ecography* 41:837–844.

Sklenář, P. and Balslev, H. (2005) Superpáramo plant species diversity and phytogeography in Ecuador. *Flora-Morphology, Distribution, Functional Ecology of Plants* 200(5): 416-433.

Smithers BV, Oldfather MF, Koontz MJ, et al (2020) Community turnover by composition and climatic affinity across scales in an alpine system. *Am J Bot* 107:239–249.

Smith SA, Brown JW (2018) Constructing a broadly inclusive seed plant phylogeny. *Am J Bot* 105:302–314.

Soltis DE, Soltis PS (2016) Mobilizing and integrating big data in studies of spatial and phylogenetic patterns of biodiversity. *Plant Divers* 38:264–270.

Steinbauer MJ, Grytnes J-A, Jurasinski G, et al (2018) Accelerated increase in plant species richness on mountain summits is linked to warming. *Nature* 556:231–234.

Swenson NG, Enquist BJ, Pither J, et al (2006) The problem and promise of scale dependency in community phylogenetics. *Ecology* 87:2418–2424.

The Angiosperm Phylogeny Group, Chase MW, Christenhusz MJM, 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.

Tsirogiannis C, Sandel B (2017) PhyloMeasures-package: PhyloMeasures: Fast Computations of Phylogenetic Biodiversity.

Tsirogiannis C, Sandel B (2016) PhyloMeasures: a package for computing phylogenetic biodiversity measures and their statistical moments. *Ecography* 39:709–714.

WCSP (2020) World Checklist of Selected Plant Families. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet

SUPPLEMENTAL FIGURES & TABLES

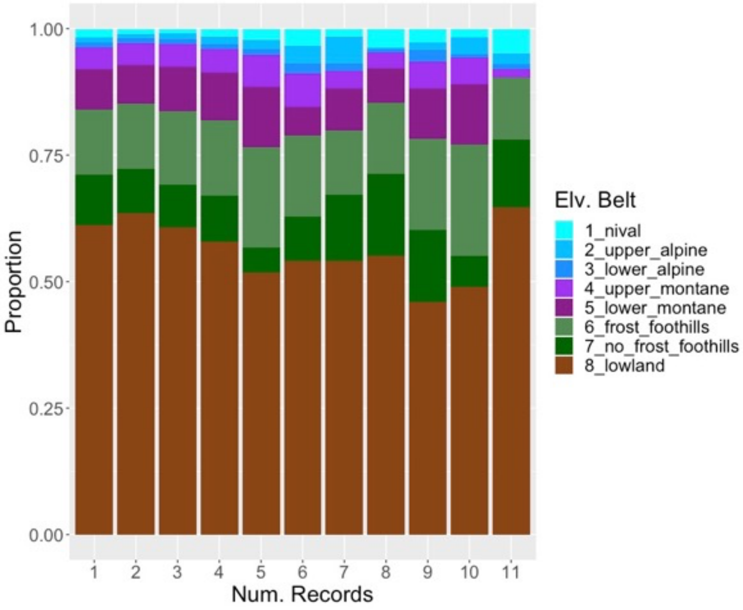


Figure 3.S1. Elevational distribution of occurrence records from the Global Biodiversity Integrated Facility (GBIF) falling within American montane habitat (as defined by the Global Mountain Biodiversity Assessment; GMBA). Bars are colored based on GMBA elevational zones and sorted according to the number of records available for each species. In the main text, we describe a method for building species distribution models that excludes species with fewer than 12 records. Here we show that species thus excluded are not preferentially biased into any particular elevational category.

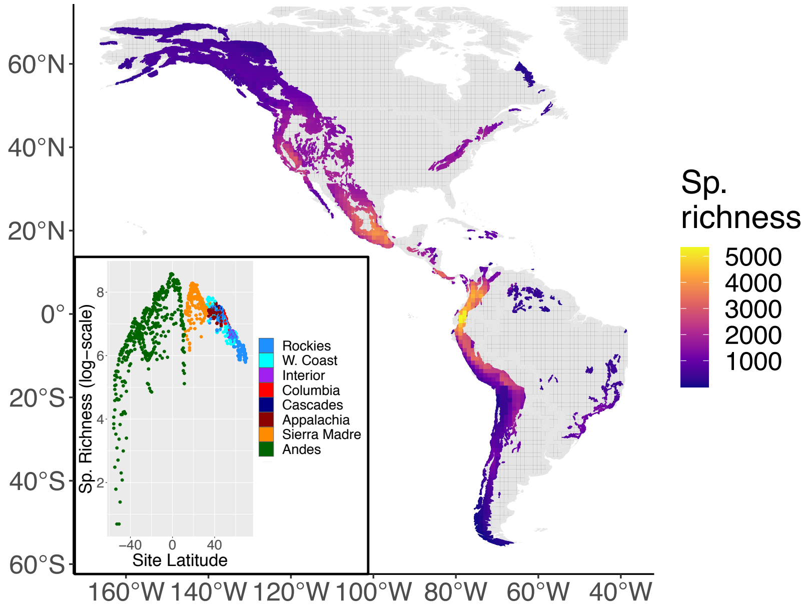


Figure 3.S2. Species richness among both montane and alpine taxa (defined as those species with at least 5% of their modelled range in montane or alpine habitat, respectively, as delimited by the Global Mountain Biodiversity Assessment). Unlike when looking solely at alpine species richness (**Figure 3.2**), these higher elevation communities follow a more typical latitudinal diversity gradient, with peak richness around the equator. *Inset plot* shows logged species richness against latitude for each (equal area) site in the study region, with points colored by the major mountain range

where that site occurs (Andes, Appalachia, Cascades, Columbia, Interior, Rockies, West Coast, Sierra Madre).

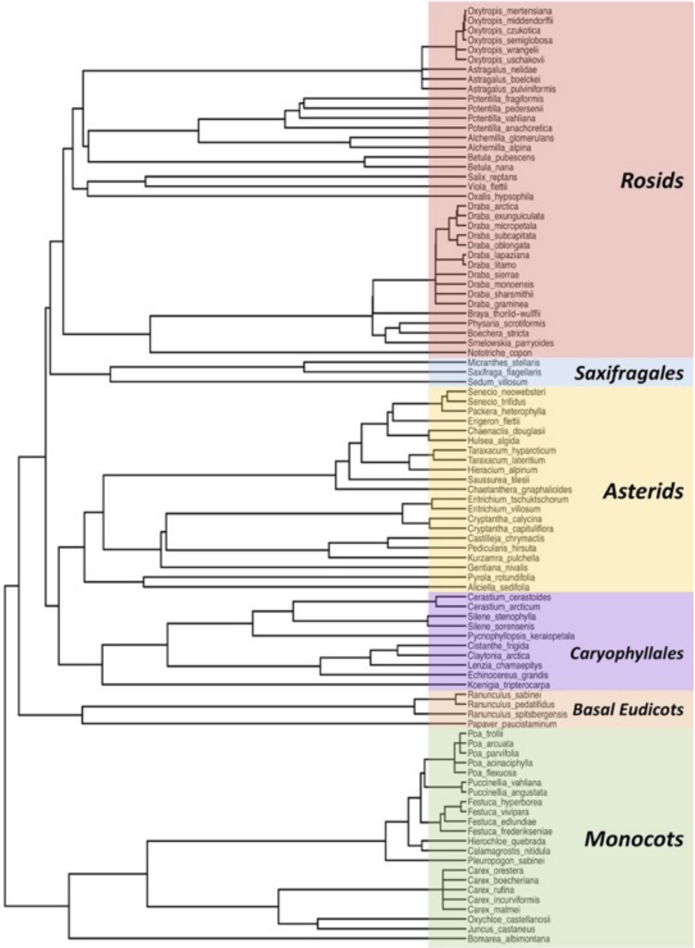


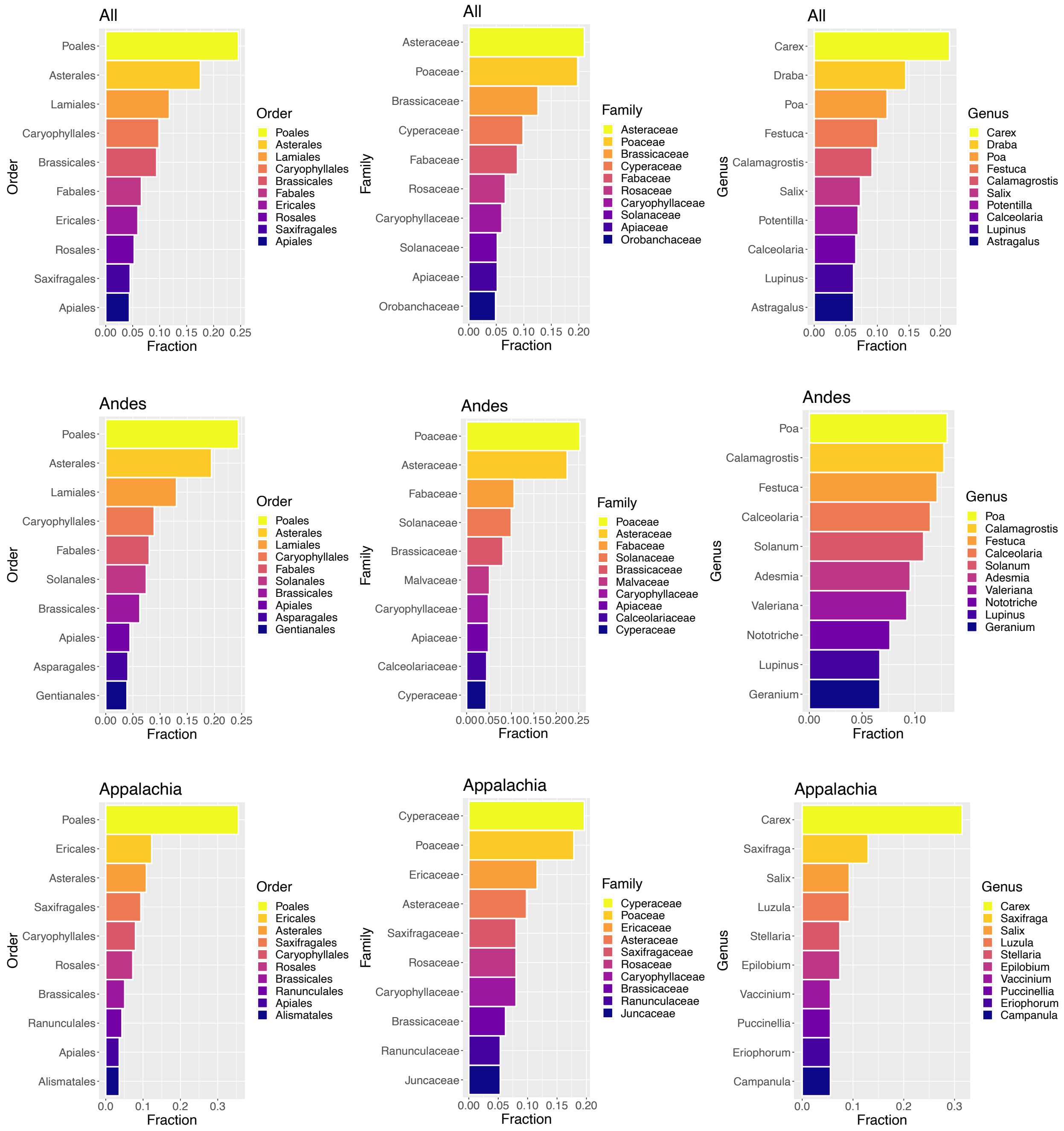
Figure 3.S3. Phylogenetic distribution of alpine specialists (defined as having at least 50% of their modelled range within alpine habitat) in our dataset, with major angiosperm groups highlighted (colors as in **Figure 3.3**). Most major clades contained at least one alpine specialist; however, Poaceae and Asteraceae jointly accounted for almost a quarter of all alpine species (specialists and generalists), and 10 taxonomic families (**Figure 3.4**) comprised ~50% of all alpine taxa.

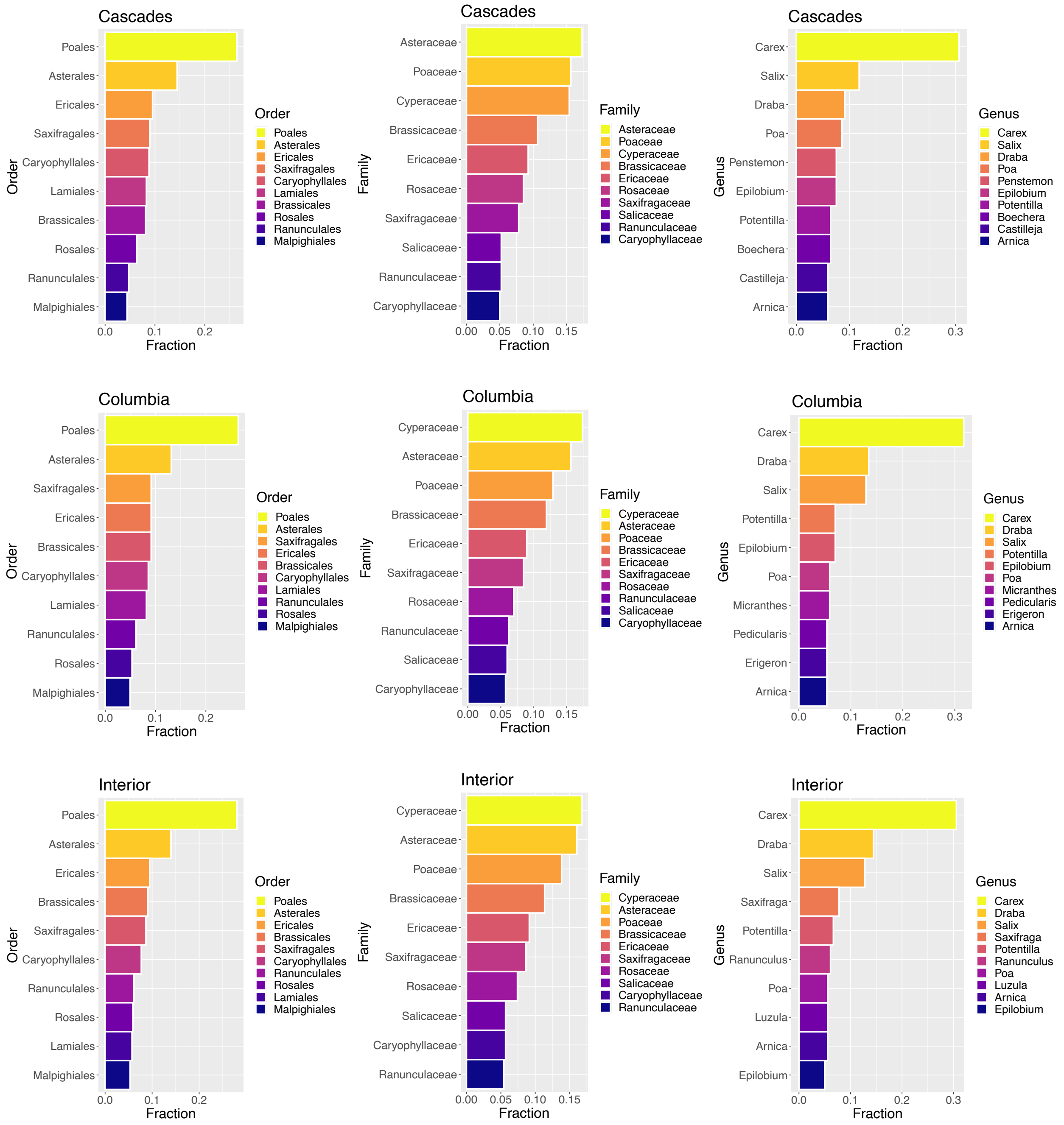
Table 3.S1a. American seed plant orders without alpine species.

Acorales	Arecales	Austrobaileyales	Berberidopsidales
Boraginaceae	Buxales	Ceratophyllales	Chloranthales
Crossosomatales	Cycadales	Dilleniales	Garryales
Huerteales	Icacinales	Magnoliales	Metteniusales
Nymphaeales	Pandanales	Picramniales	Zygophyllales

Table 3.S1b. American seed plant families without alpine species

Acanthaceae
Achariaceae
Achatocarpaceae
Acoraceae
Aextoxicaceae
Aizoaceae
Alismataceae
Altingiaceae
Alzateaceae
Anacampserotaceae
Anisophylleaceae
Annonaceae
Araucariaceae
Arecaceae
Asphodelaceae
Atherospermataceae
Balanophoraceae
Balsaminaceae
Bataceae
Berberidopsidaceae
Bixaceae
Bonnetiaceae
Brunelliaceae
Burmanniaceae
Burseraceae
Buxaceae
Cabombaceae
Calophyllaceae
Calycanthaceae
Canellaceae
Cannabaceae
Capparaceae
Cardiopteridaceae
Caricaceae
Caryocaraceae
Casuarinaceae
Ceratophyllaceae
Chloranthaceae
Chrysobalanaceae
Clethraceae
Colchicaceae
Combretaceae
Connaraceae
Costaceae
Crossosomataceae
Cucurbitaceae
Cyclanthaceae
Cymodoceaceae
Cyrillaceae
Daphniphyllaceae
Datiscaceae
Dichapetalaceae
Dilleniaceae
Dipentodontaceae
Dipterocarpaceae
Ebenaceae
Elaeocarpaceae
Eriocaulaceae
Erythroxylaceae
Euphroniaceae
Flagellariaceae
Fouquieriaceae
Francoaaceae
Garryaceae
Gelsemiaceae
Gesneriaceae
Gisekiaceae
Gnetaceae
Gomortegaceae
Goodeniaceae
Goupiaceae
Griseliniaceae
Guamatelaceae
Haemodoraceae
Halophytaceae
Hamamelidaceae
Heliconiaceae
Heliotropiaceae
Hernandiaceae
Humiriaceae
Hydrangeaceae
Hydroleaceae
Hypoxidaceae
Icacinaceae
Iteaceae
Ixonanthaceae
Joinvilleaceae
Juglandaceae
Koeberliniaceae
Krameriaceae
Lacistemataceae
Lardizabalaceae
Lecythidaceae
Lepidobotryaceae
Linderniaceae
Loganiaceae
Lythraceae
Magnoliaceae
Malpighiaceae
Marantaceae
Marcgraviaceae
Martyniaceae
Mayacaceae
Meliaceae
Menispermaceae
Metteniusaceae
Microteaceae
Molluginaceae
Monimiaceae
Moraceae
Muntingiaceae
Myricaceae
Myristicaceae
Myrtaceae
Nartheciaceae
Nelumbonaceae
Nymphaeaceae
Nyssaceae
Ochnaceae
Olacaceae
Opiliaceae
Paeoniaceae
Pandanaceae
Pedaliaceae
Pentaphylacaceae
Penthoraceae
Peraceae
Peridiscaceae
Petenaeaceae
Philesiaceae
Phyllanthaceae
Phyllonomaceae
Picramniaceae
Picrodendraceae
Pittosporaceae
Platanaceae
Plocospermataceae
Podostemaceae
Pontederiaceae
Proteaceae
Putranjivaceae
Quillajaceae
Rapateaceae
Resedaceae
Rhabdodendraceae
Rhizophoraceae
Ruppiaceae
Rutaceae
Sapotaceae
Sarcobataceae
Sarraceniaceae
Saururaceae
Scheuchzeriaceae
Schisandraceae
Schlegeliaceae
Setchellanthaceae
Simaroubaceae
Simmondsiaceae
Siparunaceae
Smilacaceae
Stachyuraceae
Staphyleaceae
Stegnospermataceae
Stemonaceae
Stemonuraceae
Strelitziaceae
Styracaceae
Surianaceae
Talinaceae
Tapisciaceae
Taxaceae
Tecophilaeaceae
Tetrachondraceae
Theaceae
Thurniaceae
Thymelaeaceae
Ticodendraceae
Tovariaceae
Trigoniaceae
Triuridaceae
Ulmaceae
Velloziaceae
Vochysiaceae
Xyridaceae
Zamiaceae
Zingiberaceae
Zygophyllaceae


All
Order
Poales
Asterales
Lamiales
Caryophyllales
Brassicales
Fabales
Ericales
Rosales
Saxifragales
Apiales
0.00 0.05 0.10 0.15 0.20 0.25
Fraction
All
Family
Asteraceae
Poaceae
Brassicaceae
Cyperaceae
Fabaceae
Rosaceae
Caryophyllaceae
Solanaceae
Apiaceae
Orobanchaceae
0.00 0.05 0.10 0.15 0.20
Fraction
All
Genus
Carex
Draba
Poa
Festuca
Calamagrostis
Salix
Potentilla
Calceolaria
Lupinus
Astragalus
0.00 0.05 0.10 0.15 0.20
Fraction
Andes
Order
Poales
Asterales
Lamiales
Caryophyllales
Fabales
Solanales
Brassicales
Apiales
Asparagales
Gentianales
0.00 0.05 0.10 0.15 0.20 0.25
Fraction
Andes
Family
Poaceae
Asteraceae
Fabaceae
Solanaceae
Brassicaceae
Malvaceae
Caryophyllaceae
Apiaceae
Calceolariaceae
Cyperaceae
0.000.050.100.150.200.25
Fraction
Andes
Genus
Poa
Calamagrostis
Festuca
Calceolaria
Solanum
Adesmia
Valeriana
Nototriche
Lupinus
Geranium
0.00 0.05 0.10
Fraction
Appalachia
Order
Poales
Ericales
Asterales
Saxifragales
Caryophyllales
Rosales
Brassicales
Ranunculales
Apiales
Alismatales
0.0 0.1 0.2 0.3
Fraction
Appalachia
Family
Cyperaceae
Poaceae
Ericaceae
Asteraceae
Saxifragaceae
Rosaceae
Caryophyllaceae
Brassicaceae
Ranunculaceae
Juncaceae
0.00 0.05 0.10 0.15 0.20
Fraction
Appalachia
Genus
Carex
Saxifraga
Salix
Luzula
Stellaria
Epilobium
Vaccinium
Puccinellia
Eriophorum
Campanula
0.0 0.1 0.2 0.3
Fraction


Cascades
Order
Poales
Asterales
Ericales
Saxifragales
Caryophyllales
Lamiales
Brassicales
Rosales
Ranunculales
Malpighiales
0.0
0.1
0.2
Fraction
Cascades
Family
Asteraceae
Poaceae
Cyperaceae
Brassicaceae
Ericaceae
Rosaceae
Saxifragaceae
Salicaceae
Ranunculaceae
Caryophyllaceae
0.00
0.05
0.10
0.15
Fraction
Cascades
Genus
Carex
Salix
Draba
Poa
Penstemon
Epilobium
Potentilla
Boechera
Castilleja
Arnica
0.0
0.1
0.2
0.3
Fraction
Columbia
Order
Poales
Asterales
Saxifragales
Ericales
Brassicales
Caryophyllales
Lamiales
Ranunculales
Rosales
Malpighiales
0.0
0.1
0.2
Fraction
Columbia
Family
Cyperaceae
Asteraceae
Poaceae
Brassicaceae
Ericaceae
Saxifragaceae
Rosaceae
Ranunculaceae
Salicaceae
Caryophyllaceae
0.00
0.05
0.10
0.15
Fraction
Columbia
Genus
Carex
Draba
Salix
Potentilla
Epilobium
Poa
Micranthes
Pedicularis
Erigeron
Arnica
0.0
0.1
0.2
0.3
Fraction
Interior
Order
Poales
Asterales
Ericales
Brassicales
Saxifragales
Caryophyllales
Ranunculales
Rosales
Lamiales
Malpighiales
0.0
0.1
0.2
Fraction
Interior
Family
Cyperaceae
Asteraceae
Poaceae
Brassicaceae
Ericaceae
Saxifragaceae
Rosaceae
Salicaceae
Caryophyllaceae
Ranunculaceae
0.00
0.05
0.10
0.15
Fraction
Interior
Genus
Carex
Draba
Salix
Saxifraga
Potentilla
Ranunculus
Poa
Luzula
Arnica
Epilobium
0.0
0.1
0.2
0.3
Fraction

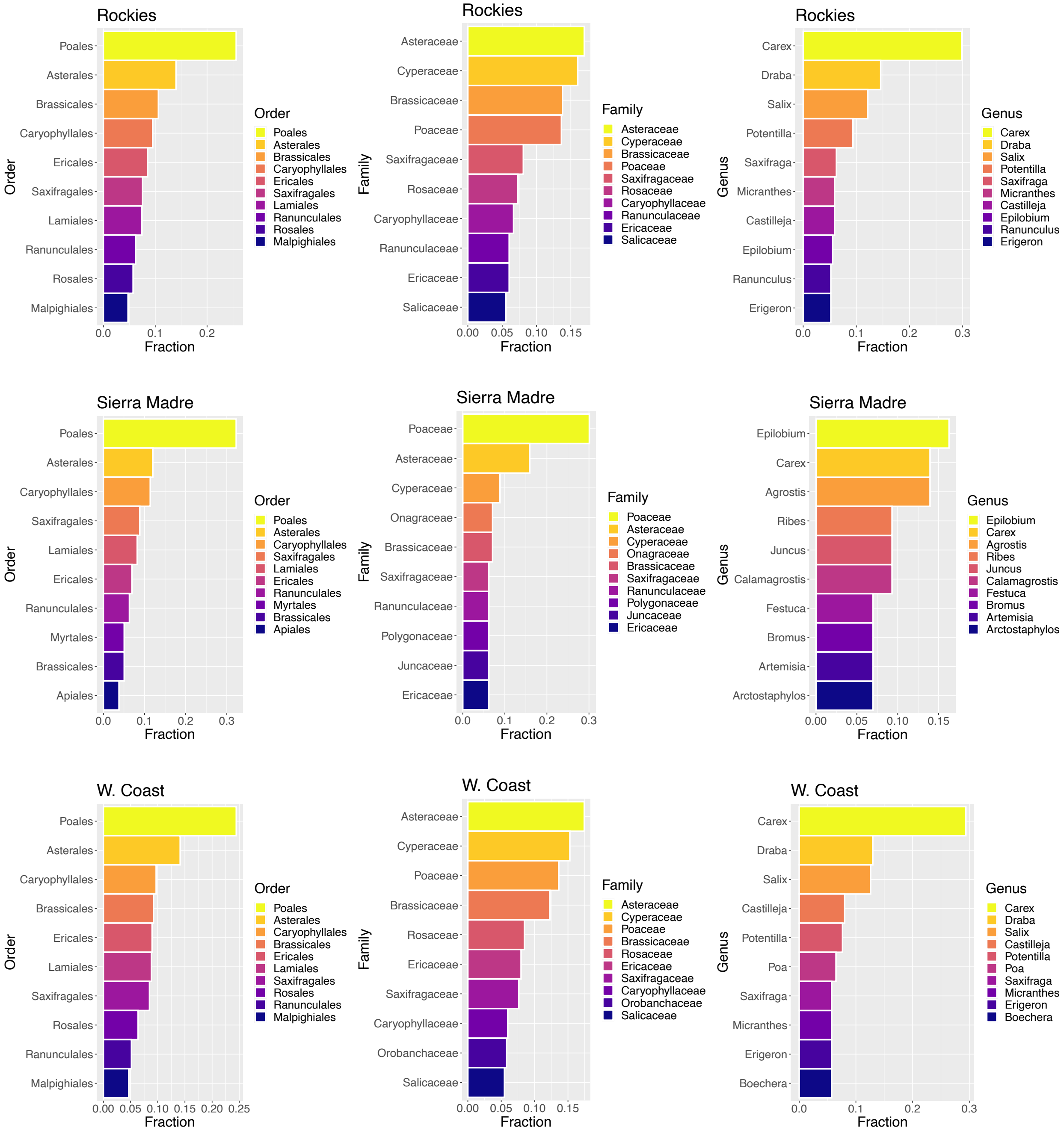
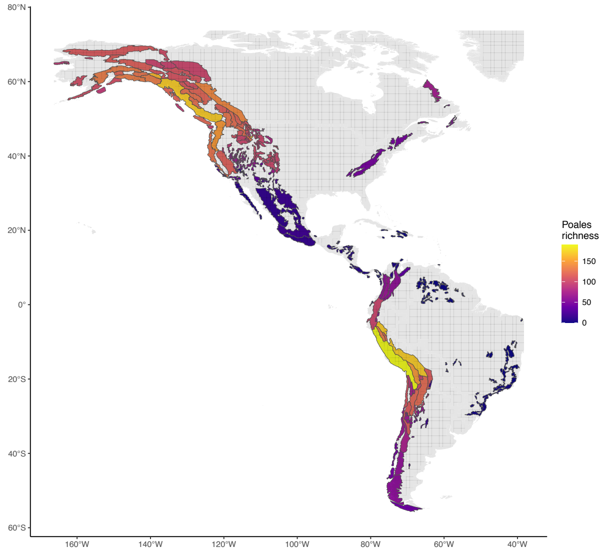
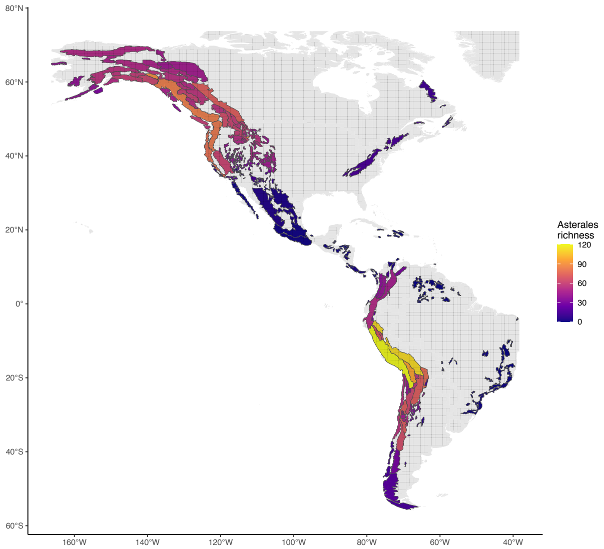
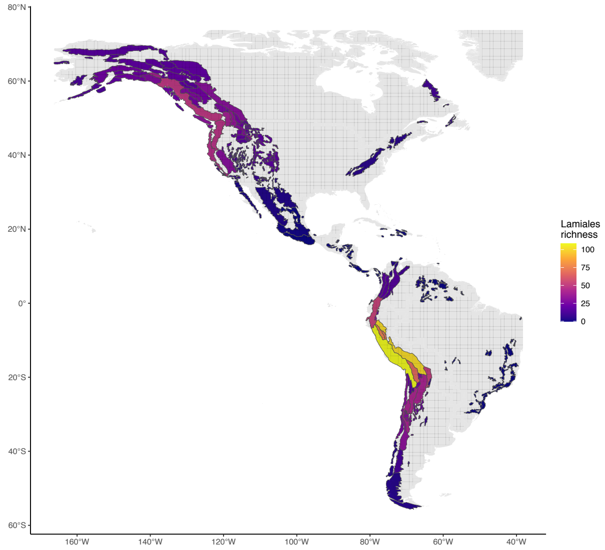
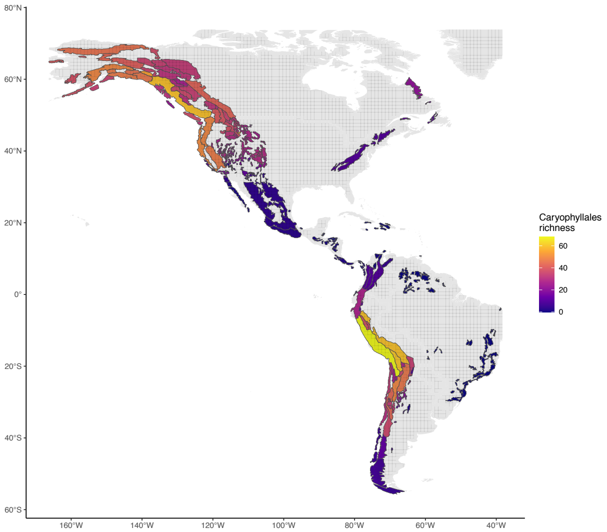


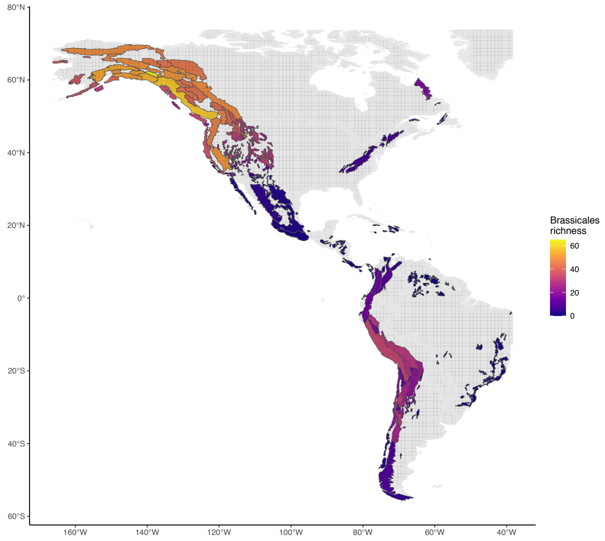
Figure 3.S4. Comparisons of the top ten most dominant Orders (*left column*), Families (*center column*), and Genera (*right column*) within each of the eight major mountain ranges discussed in the main text (Andes, Appalachia, Cascades, Columbia, Interior, Rockies, Sierra Madre, and West Coast; names provided atop each plot). See main text for discussion of differences in mountain composition. *Note that each plot has a separate and independent color legend.*

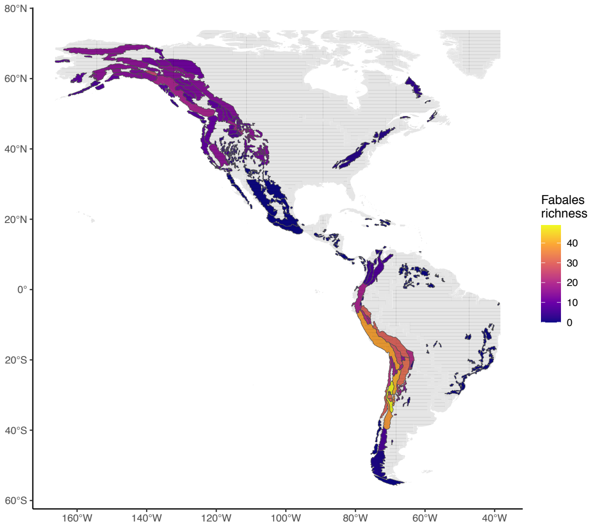

80°N
60°N
40°N
20°N
0°
20°S
40°S
60°S
160°W
140°W
120°W
100°W
80°W
60°W
40°W
Poales richness
150
100
50
0


80°N
60°N
40°N
20°N
0°
20°S
40°S
60°S
160°W
140°W
120°W
100°W
80°W
60°W
40°W
Asterales richness
120
90
60
30
0


80°N
60°N
40°N
20°N
0°
20°S
40°S
60°S
160°W
140°W
120°W
100°W
80°W
60°W
40°W
Lamiales richness
100
75
50
25
0


80°N
60°N
40°N
20°N
0°
20°S
40°S
60°S
160°W
140°W
120°W
100°W
80°W
60°W
40°W
Caryophyllales richness
60
40
20
0


80°N
60°N
40°N
20°N
0°
20°S
40°S
60°S
160°W
140°W
120°W
100°W
80°W
60°W
40°W
Brassicales richness
60
40
20
0


80°N
60°N
40°N
20°N
0°
20°S
40°S
60°S
160°W
140°W
120°W
100°W
80°W
60°W
40°W
Fabales richness
40
30
20
10
0

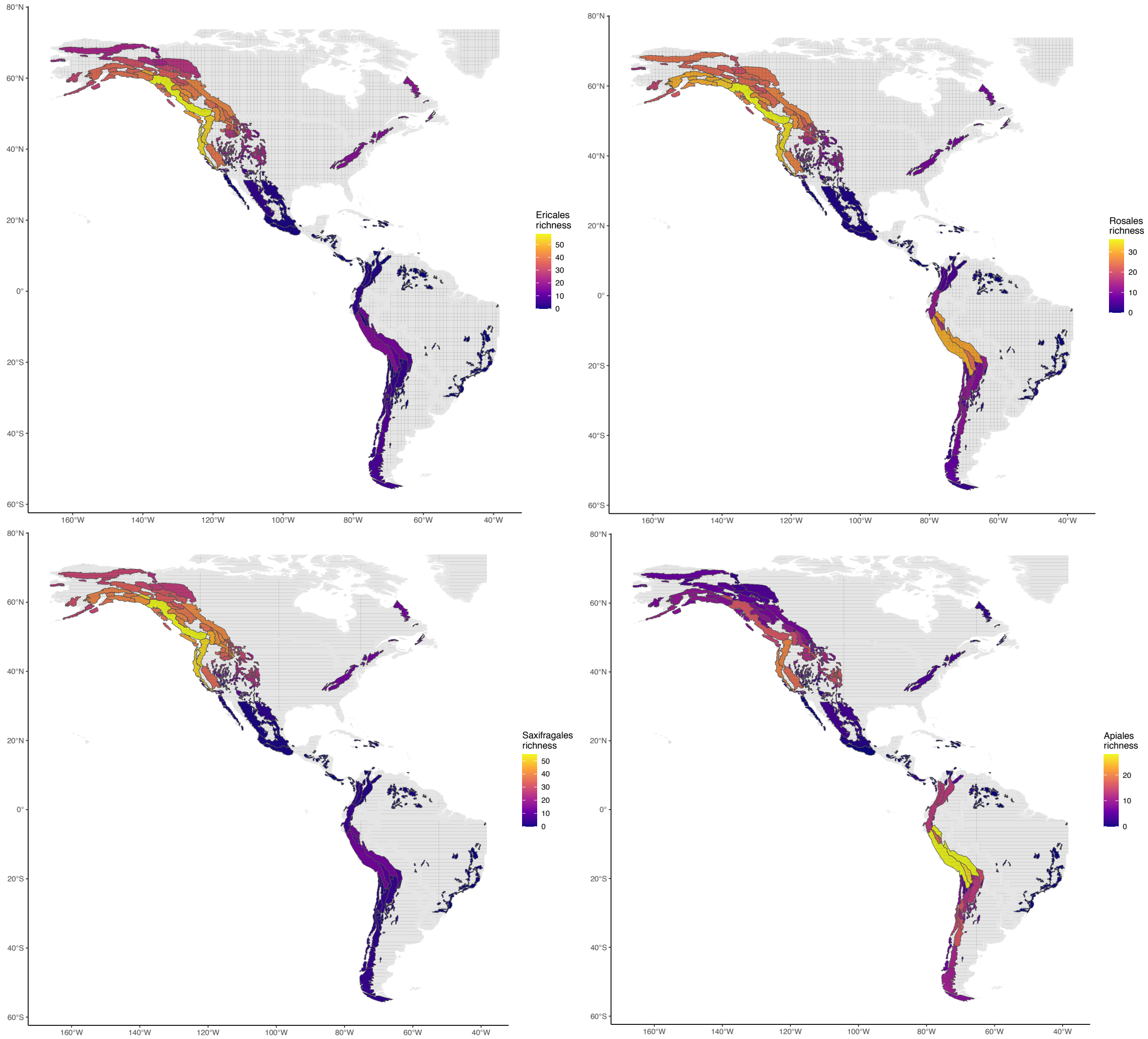


Figure 3.S5. Diversity patterns of alpine species richness for the ten angiosperm orders contributing most to alpine diversity in the Americas. Some groups, such as Ericales had higher diversity in North America, whereas other groups, such as Fabales had higher diversity in the central Andes. Overall, however, at this taxonomic scale, biodiversity patterns generally mirrored the bimodal diversity pattern observed among all alpine species.

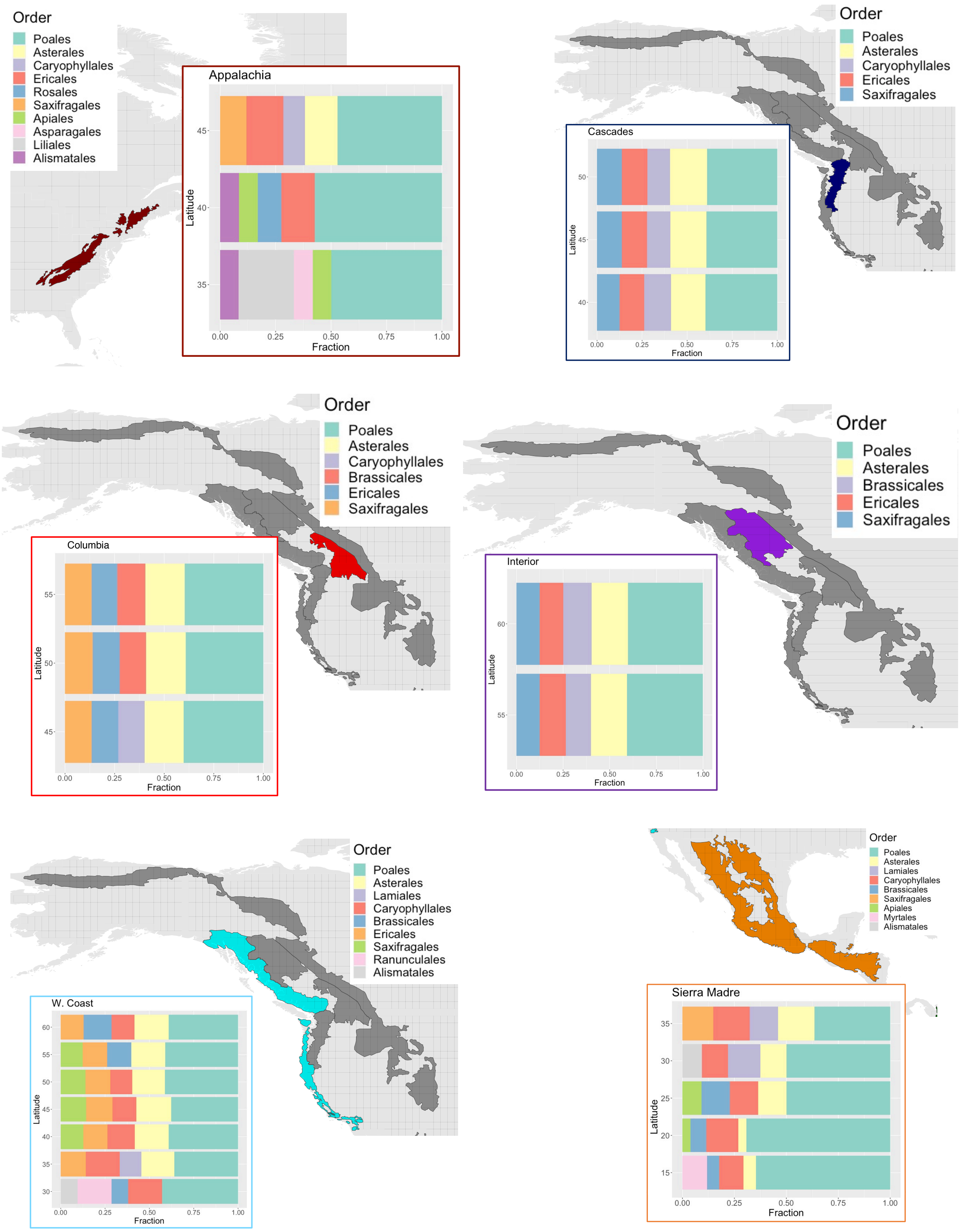


Figure 3.S6. Alpine community turnover in the relative dominance (i.e., species richness) of angiosperm orders for six major American mountain ranges along the latitudinal extent

of each range (see **Figure 3.5** for the Andes and Rockies). Bar plots show the relative dominance of only the top five most dominant plant orders in 5° latitudinal sections along each mountain range system. In all cases, Poales was always the most dominant order. However, individual ranges differed in their most prevalent plant orders as well as their relative rankings. For instance, Poales dominated the Sierra Madre to a greater extent than in any other mountain range. Additionally, alpine community dominance patterns changed with latitude for many mountain ranges. For example, Saxifragales tended to dominate only above ~40° N latitude, and Caryophyllales tended to show only intermittent dominance along mountain ranges. (*Note that each plot assigns colors to orders independently; please refer to figure legends for precise ordinal color assignments*).

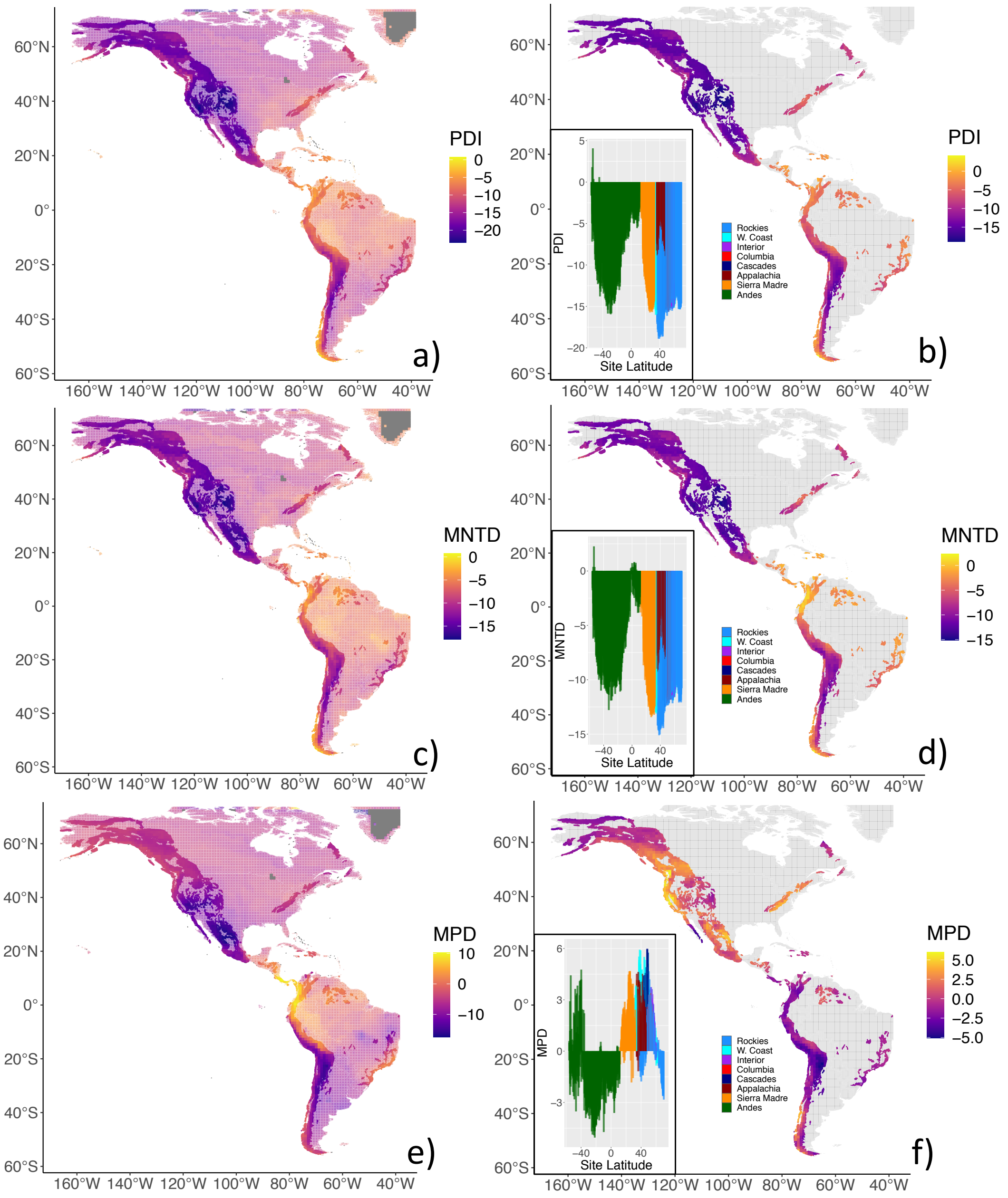


Figure 3.S7. Phylodiversity metrics. (*Left Column*), phylogenetic diversity metrics across all American angiosperm species in our data set (Figueroa et al. *in review*; see Methods). The foreground in each panel shows mountainous regions as defined by the Global Mountain Biodiversity Assessment (GMBA), background is non-mountainous habitat within the study area (gray cells indicate sites with insufficient data to calculate metrics). Patterns of phylogenetic diversity among all American angiosperms differed significantly from those observed for just alpine taxa (compare with **Figure 3.6**). (*Right Column*), same as in left column but only including montane and alpine species (defined here as those species having at least 5% of their modelled range in montane or alpine habitat, respectively). Patterns here also differed substantially from those for only alpine communities. Inset plots (in *b, d, f*) show the value of each metric at each site in the study area, with colors indicating the major mountain range in which the site occurs. (*a*) In the study area, the Phylogenetic Diversity Index (PDI) was generally always less than zero for all angiosperms, indicating that all sites were more phylogenetically clustered than expected by chance. (*b*) Among combined montane-alpine communities, PDI was also usually negative, though more sites had some overdispersion. These patterns are in contrast to the varying patterns of both clustering and over-dispersion observed among alpine communities, particularly in the southern Andes and Sierra Madre. *(c, d)* Spatial patterns of Mean Nearest Taxon Distance (MNTD) were very similar to those for PDI. Among all angiosperms and when looking only at combined montane-alpine communities, values tended to always indicate communities more closely related than expected by chance, whereas when examining only alpine communities (**Figure 3.6**) some sites comprised species less closely related than expected by chance. (*e*) Mean Pairwise Distance (MPD) among all angiosperms showed more variation across sites than PDI or MNTD, with communities showing both more and less relatedness than expected by chance. However, the patterns still differed significantly from those observed solely among alpine species. In particular, the alpine communities of the Northern Alaskan ranges and nearby areas showed the lowest MPD values, whereas when looking across all angiosperms, these regions had among the highest values. MPD for angiosperms was highest near the equator, whereas these sites had some of the lowest values among alpine communities. (*f*), When looking at combined montane-alpine communities, the patterns were qualitatively more similar to those for the alpine communities, though MPD values for alpine communities were still generally higher. The central and southern Andes, in particular, had much higher MPD values when considering only alpine species. The significance of differences in phylogenetic diversity patterns between alpine and non-alpine communities is discussed in the main text.

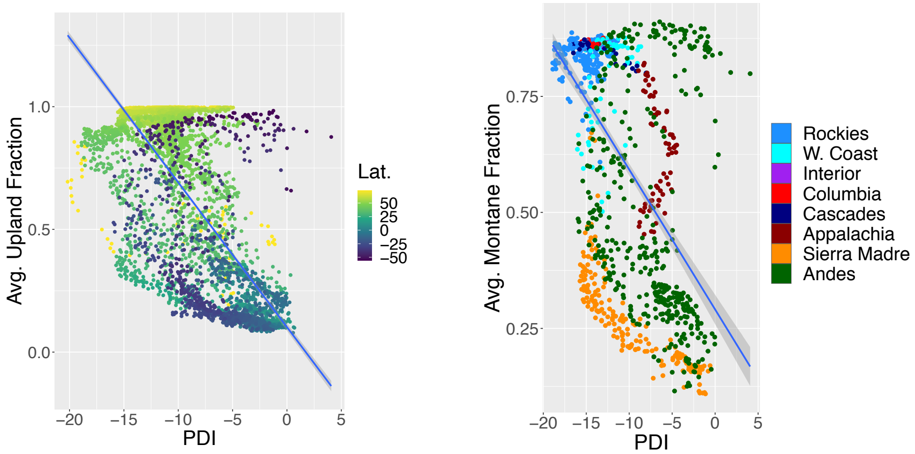


Figure 3.S8. Relationship between Phylogenetic Diversity (PDI) and the degree of higher elevation specialization (i.e., including both montane and alpine species as one, higher elevation, group) among species at each site in the study area. Unlike the scenario when considering only alpine species (Fig. 6d in the main text), in which a strong negative relationship was observed, here we combine alpine and montane species and the correlation between the two variables is lessened and is no longer visually apparent. (Note that there is still a statistically significant trend here, however: $R^2=0.25$; $p<0.001$). (*Left*), all sites in the study area, colored by latitude. (*Right*), only those sites falling within major mountain range systems, as indicated.

Supporting information statement

Additional supporting information (Online Resources 1-3) can be found in the online published version of this article (see Supplementary Material) contains the following: ESM_1—supplementary figures and tables referred to in the main text; ESM_2—CSV file containing a taxonomy of the angiosperm alpine species included in this study; ESM_3—List of likely range-limited endemic Andean alpine species excluded by our data cleaning protocol for building reliable niche models.

Chapter IV

Reconstructing the Joint Evolution of Temperature, Precipitation, and Elevational Niches for American Seed Plants

CHAPTER IN CONTEXT

In this chapter, I use continuous character ancestral state reconstruction to infer the past temperature, precipitation, and elevation niches of American seed plants, with the goal of enhancing our understanding of the evolutionary origins of alpine angiosperms. This chapter also draws upon the niche model dataset assembled and described in **CHAPTER TWO**, and the associated time-calibrated phylogeny. Reconstructions were performed on each taxonomic order with at least 100 species present in the niche model dataset. This allowed for comparisons between clades with and without contemporary alpine species. The inclusion of temperature and precipitation niche reconstructions further permitted an assessment of how changes in different niche axes are related. General patterns in the evolutionary timing of the origins of alpine lineages are discussed in a comparative framework.

One of the primary questions this study addresses is, how recently have alpine lineages evolved? Although substantial work has been done to investigate the ages of specific alpine *assemblages*, there is less known about the timing of the evolutionary origins of specific *lineages* that occupy those assemblages. For example, members of the *Astragalus/ Oxytropis* complex are an important component of Fabaceae alpine diversity across the American mountain ranges (**CHAPTER THREE**), however the origins of the clade are likely Eurasian. Therefore, alpine adaptations or characteristics may have arisen in this lineage prior to the assembly of any one particular modern alpine community.

The age of the alpine biome is unknown but presumed to be relatively young. Consistent with this expectation, I find that most reconstructed nodes with significant alpine character are predicted to have arisen relatively recently, mostly over the last 10 MY. At the same time, however, some clades departed from this trend and had alpine nodes extending from 10-50 MY, most notably Rosales, Ericales, Fagales, and Saxifragales. These clades appear to have acquired alpine character prior to other lineages that currently occupy alpine habitat, which could be evidence of pre-adaptation within those lineages. Interestingly, Poales, which is the most species-rich component of American alpine diversity in our dataset, did not show any alpine nodes older than 10 MY, and in general only three Poalean clades within our dataset have managed to radiate into the alpine belt.

By also looking at the reconstructed temperature and precipitation niches of clades within our dataset, I was able to begin addressing the question of how shifts into alpine or montane habitat were associated with shifts in temperature or precipitation. There was a tight association between temperature and elevation niche shifts for all clades. Most crucially, clades devoid of alpine species also showed restricted exploration of temperature niche-space over evolutionary time, supporting the notion that tolerance to cold or freezing temperatures might be a primary abiotic filter associated with entry into the alpine belt. In contrast, changes in precipitation niche did not correlate with elevation niche shifts, and clades with or without alpine species often occupied the same region of precipitation niche-space. Although precipitation changes with elevation, average annual differences in precipitation do not appear to present an abiotic filter for entry into the alpine belt.

This study contributes to the dissertation by providing a window into the evolutionary origins of alpine seed plants, and thus the timing and extent of niche conservatism and niche shifts. The results illustrate how phylogenetic niche conservatism and niche shift act together to produce the mosaic of distributional patterns observed within and across different clades. Using continuous character reconstructions also allowed for a more biologically meaningful assessment of the degree of niche shift and to visualize how changes in one niche axis were connected to changes in another. The ability to integrate niche evolution across multiple, ecologically relevant, niche axes is crucial for a better appreciation of how niche conservatism has helped shape the modern distributions of American seed plants.

ABSTRACT

Mountains contribute disproportionally to the Earth's biodiversity, yet relatively little is known about the evolutionary history of the lineages that comprise these higher elevation assemblages. In particular, tracing the evolutionary origins of alpine plant communities and lineages is an active research area. Alpine habitats are believed to be geologically relatively young, and survival in the alpine belt is intimately tied to tolerance of freezing conditions. Here, we use a large macroecological niche model dataset describing the realized climatic and elevational niche space of American seed plants to begin addressing questions about the evolutionary origins of alpine lineages. We employ continuous character ancestral state reconstruction to infer how the temperature, precipitation, and elevational niches of seed plants have changed over evolutionary time. Specifically, we ask (1) how old are alpine lineages; (2) Does the emergence of increased alpine character coincide with known climate or geologic events? (3) Are shifts in temperature and precipitation niches associated with shifts in elevation across the phylogenies of these groups? (4) How do clades without alpine species differ in their reconstructed climatic and elevation niche history? We find that only clades which explore sufficiently low temperature niche space have been able to enter higher elevations, whereas no clear association with niche shifts in precipitation were observed. Across all clades, alpine lineages were found to always be younger than 50 MY and generally below 10 MY, consistent with the hypothesis that alpine habitats are relatively young. Clades with contemporary alpine lineages showed few transitions into the alpine belt, suggesting strong abiotic filters and niche conservatism. Nevertheless, clades giving rise to alpine lineages were much older than the ages of the alpine species themselves, potentially indicating early preadaptation in specific clades. These results underscore how both niche conservatism and niche shift combine to help shape contemporary biodiversity patterns.

INTRODUCTION

Despite representing less than 25% of terrestrial surface (Hoorn et al., 2018; Körner et al., 2011), mountains are exceptionally diverse, for example, harboring more than 80% of vertebrate species (Rahbek, Borregaard, Colwell, et al., 2019) and disproportionally high densities of plant diversity (Barthlott et al., 2007). Further, high elevation habitats house numerous biodiversity hotspots (e.g., Ding et al., 2020; Flantua et al., 2019; Hughes & Atchison, 2015; Hughes & Eastwood, 2006), offer potential cooler-climate corridors for plant dispersal (Antonelli et al., 2009), and their topographic complexity provides myriad microclimatic niche space for plants to become established and survive (Körner, 2003). Despite their clear importance for biodiversity patterns, however, there are still numerous unanswered questions regarding the assembly, dynamics, and origins of higher elevation plant communities (Rahbek, Borregaard, Antonelli, et al., 2019; Rahbek, Borregaard, Colwell, et al., 2019).

At the regional scale, higher elevation communities can be described as occupying seven biologically relevant thermal belts (Körner et al., 2011, 2017; Körner & Paulsen, 2004): the nival (perpetual snowline), upper and lower alpine (treeline estimate), upper and lower montane, frost-exposed foothills, and remaining frost-free areas. Importantly, these classifications were delimited by growing season length, climate, and topography, and they account for latitudinal differences in the absolute elevation at which alpine or montane habitats occur (Körner et al., 2011).

Using these classifications, it has been shown that alpine plant communities, in particular, both across the American mountain range system and globally, show unique diversity patterns and depart from a traditional latitudinal diversity gradient (Qian et al., 2021; Testolin et al., 2021; Figueroa et al., 2021; **CHAPTER THREE**; **APPENDIX D**). These communities also show biogeographically distinct patterns of phylogenetic diversity and community assembly that differentiate them from montane and lowland communities (Figueroa et al., 2021; González-Caro et al., 2020; Qian et al., 2021). Alpine habitats are often at the potential physiological limits of where plants can survive (Körner, 2003) and therefore their composition and assembly processes might be strongly influenced by abiotic filters. And, indeed, the relatively small number of American plants with ranges centered in alpine areas (Figueroa et al. 2021) as well as their greater phylogenetic affinity with species already occupying the montane belt (Figueroa et al. 2021; González-Caro et al., 2020) are suggestive of strong abiotic filtering.

At the same time, we have previously shown that alpine communities occupying otherwise climatically similar sites at opposing ends of the American mountain range system differ in taxonomic composition, phylogenetic diversity, and degree of alpine specialization (Figueroa et al. 2021). These findings suggest an important role for historical processes, such as biogeographic and phylogenetic history, in assembling these floras. Taken together, these studies underscore the complex interplay of factors that define plant diversity at the extreme elevational limits of the alpine belt (Körner, 2003) and reiterate the challenges of defining alpine communities.

In addition to occupying habitats at the elevational limits of available terrestrial terrain (Elsen & Tingley, 2015), alpine areas are also geologically younger than lowlands (Ding et al., 2020), and alpine communities implicitly must be cold tolerant. Therefore, the origins of alpine taxa are expected to be inherently tied to both climatic and geologic events, such as mountain uplift to sufficient elevations and trends of global cooling. Recently, Ding et al. (2020) were able to link assembly of alpine communities in the Tibet-Himalaya-Hengduan (THH) region to both local orogeny and the intensification of the Asian monsoon, and their results suggested an Oligocene origin for alpine communities in that region. This estimate for alpine community age in the THH region contrasts with estimates of alpine community assembly for American communities (Hughes & Atchison, 2015; Madriñán et al., 2013; Muellner-Riehl et al., 2019; Wolfe, 1987). For instance, the modern alpine flora of the Rocky Mountains may have been assembled over the past 10 MY (Hughes & Atchison, 2015; Muellner-Riehl et al., 2019; Wolfe, 1987) and the Andean páramos communities could be as young as ~3.5 MY (Hughes & Atchison, 2015; Madriñán et al., 2013). The historical factors giving rise to modern alpine communities might therefore vary substantially for different mountain ranges.

Apart from determining the age of specific assemblages, biologists would also benefit from knowing the age of alpine lineages themselves, and when in the evolutionary history of particular clades alpine adaptations might have arisen. For example, the *Astragalus/ Oxytropis* complex is a primary component of Fabaceae alpine diversity across the American mountain ranges (Figueroa et al. 2021; Amiri et al., 2020), however the origins of the clade are likely Eurasian (Amini et al., 2019; Bagheri et al., 2017). It is therefore plausible that alpine adaptations or characteristics arose in this lineage prior to the assembly of any one particular modern alpine community. In contrast, Solanales alpine diversity in the Americas is concentrated within the Andes (Figueroa et al. 2021), consistent with studies suggesting a South American origin for this clade (Olmstead, 2013). The age of alpine characteristics for Solanales might therefore be expected to better align with Andean uplift or changes in South American climate specifically.

In this study, we adopt a macroecological approach across the Americas to investigate questions regarding the origins of alpine diversity within multiple, large, and broadly distributed clades. We utilize an ancestral state reconstruction methodology and draw comparisons across numerous clades simultaneously in order to ask the following questions regarding American alpine seed plants: (1) How old are alpine lineages, given the phylogenetic context of the American seed plant phylogeny? (2) Does the emergence of increased alpine character coincide with known climate or geologic events? (3) Are shifts in temperature and precipitation niches associated with shifts in elevation across the phylogenies of these groups? (4) How do clades without alpine species differ in their reconstructed climatic and elevation niche history?

METHODS

Dataset assembly and niche models.

This study utilized a previously assembled dataset characterizing the temperature, precipitation, and elevation niches of 72,372 seed plant species across the Americas using occurrence records and species distribution models (SDMs; **CHAPTER TWO**). Briefly, the dataset was assembled as follows:

We used the dated seed plant phylogeny of Smith & Brown (2018) as an estimate of phylogenetic relationships and to generate a potential species list. This phylogeny contains over 350,000 seed plant species and was assembled based on the Open Tree of Life (Hinchliff et al., 2015); backbone relationships were inferred based on DNA sequence data from GenBank, and taxa for which no such data were available were added based on taxonomic placements (version 'ALLOTB'; see Smith & Brown, 2018).

Occurrence data for species from this phylogeny falling within the study area (i.e., the Americas) were retrieved from herbarium specimen records in iDigBio (https://www.idigbio.org/portal, 2020) and GBIF (GBIF.org, 2020; https:// www.gbif.org/), the two largest online databases for georeferenced occurrence data. To define the study area, we used bounding boxes to delimit the biogeographical regions to be included in the analyses, and occurrence records were downloaded from within these bounding boxes. We defined WGS84 bounding boxes [min x, min y, max x, max y – values reflecting latitude (x) and longitude (y)] corresponding to the entire continent of America (-166.3209, -55.71303, -38.58075, 73.74967).

We performed data cleaning to ensure quality of records, a crucial step when dealing with large datasets (Sosef et al., 2017). Cleaning scripts (available at https:// github.com/lifemapper/lmpy/) ensured occurrence data were retained only for accepted species (324,009) in the phylogeny of Smith & Brown (2018) occurring within the limits of the bounding boxes described above with further filtering steps as follows. First, occurrence points were required to include at least four decimal places (~11 m near the equator). For a species to be included, it had to be represented at 12 or more distinct localities (to build reliable models). Then, we removed points with GBIF or iDigBio quality flags (invalid, mismatched, and suspect taxonomy, datum missing and errors, geocode errors, points at the origin). Duplicate collections of a plant specimen are commonly made and distributed to multiple herbaria, resulting in the same locality being counted redundantly in data aggregators. Duplication can also derive from overlap between our two main occurrence sources, iDigBio and GBIF, as they contain many of the same records. Thus, we removed duplicate localities and constrained occurrence data to the limits of each bounding box described above.

We used species range data from the Plants of the World Online (POWO, 2019; http://powo.science. kew.org/), accessed via Kew Python APIs, to retain only localities thought to represent the native range of the study species. To do this, we intersected the occurrence data retrieved from iDigBio and GBIF with regions specified as level 3 geographical descriptors from the World Geographical Schema for Recording Plant Distributions (WGSRPD; https://www.tdwg.org/ standards/wgsrpd/). These regions define localities where the taxa are found based on expert opinion; POWO also distinguishes between native and anthropogenic areas of species ranges. Therefore, if a locality for a given taxon was situated outside of its expert-assessed geographical range, we removed it from our records. We included in the analysis all occurrence points from taxa for which there was no information available on POWO on expert-assessed geographical range.

Contemporary environmental data for building niche models were retrieved from three different databases: WorldClim (Fick & Hijmans, 2017; https://worldclim.org/data/bioclim. html), Soil Grids (https://www.isric.org/explore/ soilgrids) and Global 30 Arc-Second Elevation (GTOPO30) (https://www.usgs.gov/centers/eros/ science/usgs-eros-archive-digital-elevation-global- 30-arc-second-elevation-gtopo30?qt-science_center_ objects=0#qt-science_center_objects). From WorldClim, we used Bioclim 1 (*annual mean temperature*), Bioclim 7 (*temperature annual range*), Bioclim 12 (*annual precipitation*) and Bioclim 17 (*precipitation of driest quarter*). From Soil Grids, we used *sand content*, *soil organic carbon content*, *soil pH in water* and *coarse fragment volumetric percent*. From GTOPO30 (https://doi.org/10.5066/F7DF6PQS), we used *elevation*. Only species with locality records that satisfied the filters described above were used to produce species distribution models (SDMs) in Maxent (Phillips et al., 2006, 2004) using the Lifemapper platform (Beach et al., 2020). SDMs expand the use of locality data by providing broader geographical distributions that help alleviate incomplete sampling biases (Barthlott et al., 2007).

The analyses described above were conducted using the workflow available through Lifemapper lmpy, at https://github.com/lifemapper/lmpy/, https://github.com/biotaphy/BiotaPhyPy/ and https://github. com/biotaphy/projects/. The data cleaning and filtering methods produced a dataset of 50,002,722 georeferenced occurrence records spanning 72,372 seed plant species, of which 68,241 (5397 genera; ~18% of total estimated seed plant lineages; Stevens, 2001) could be matched back to our phylogeny (Smith & Brown, 2018).

Quantification of realized niche space.

SDMs both delimit the predicted range of a species and describe the predicted likelihood of observing a species at each site within that range (i.e., the SDMs are scored). We used these two properties of niche models to describe the (predicted) realized niche space of each species in our dataset with regards to temperature, precipitation, and elevation (these methods are described more fully in **CHAPTERS TWO** and **THREE**). This niche characterization permits one to describe a species niche, for example, as 80% below freezing and 20% above freezing, and bound between 100- and 800-mm MAP. More specifically, we quantified the weighted proportion of each SDM falling within a predefined increment of either temperature (1° C) or precipitation (250 mm). The values obtained are not novel quantities derived from the niche models, but rather statistical descriptions of the data within the niche models. Put another way, this method is analogous to the statement that ~68% of a normal distribution lies within one standard deviation of the mean—it is simply a way of describing the niche model. This avoids any perceived circularity that might arise since the niche models are built using climatic information such as temperature and precipitation.

To define the elevational niches of each species, we calculated the fraction of the SDM falling within each of the seven bioclimatic zones defined by the Global Mountain Biodiversity Assessment (GMBA, 2010; Körner et al., 2011, 2017; Körner & Paulsen, 2004). These zones integrate temperature, growing season length, and topographic information to provide a robust and biogeographically relevant assessment of the extent and type of montane and alpine habitat across the globe. They further permit an assignment of species distributions along elevational gradients despite latitudinal differences in what constitutes montane and alpine areas. The seven GMBA bioclimatic zones used to define species' ELV niche were: 1—nival (perpetual snow); 2—upper alpine; 3—lower alpine; 4—upper montane; 5—lower montane; 6—mountain slope with frost; 7—mountain slope without frost or lowland. We chose to combine GMBA zone 7 with lowland areas for our study; thus, our assessment of lowland distributions was fundamentally linked with exposure to freezing temperatures.

These methods produced a description of the temperature, precipitation, and elevation niches of each species in our dataset. **Figure 2.1** provides an illustrative example of the information obtained from this type of niche characterization, and the reader is referred to that chapter for a more complete description of the calculations involved. By incorporating information from the SDMs on the distributions of climatic conditions experienced by species along their ranges, we have assembled a more biologically meaningful description of the predicted realized niche space of American seed plants.

Ancestral state reconstructions.

Here, we attempt to reconstruct ancestral abiotic niches through the use of continuous character reconstruction. This is a more appropriate method for our dataset since the "trait" data derived from the SDMs are distributions of climatic values (such as a range centered in montane habitat with 40% below freezing and 80% below 800 mm MAP), rather than discrete scalar quantities (such as a height of 5 m, a shrub growth form, or an 'alpine plant' designation). In addition to being a more appropriate method, continuous character reconstructions here also provide potentially useful insight into the timing of the emergence of particular niche combinations. For example, say the range of species *A* is 40% alpine and its sister species (*B*) is 60% alpine. A discrete approach might suggest that the ancestor of these two species was alpine. However, our methods would account for the proportion of alpine habitat encountered by each species and conclude that the range of their common ancestor likely encompassed ~50% alpine habitat—thus indicating a species both alpine and montane, as well as providing amore biologically realistic interpretation.

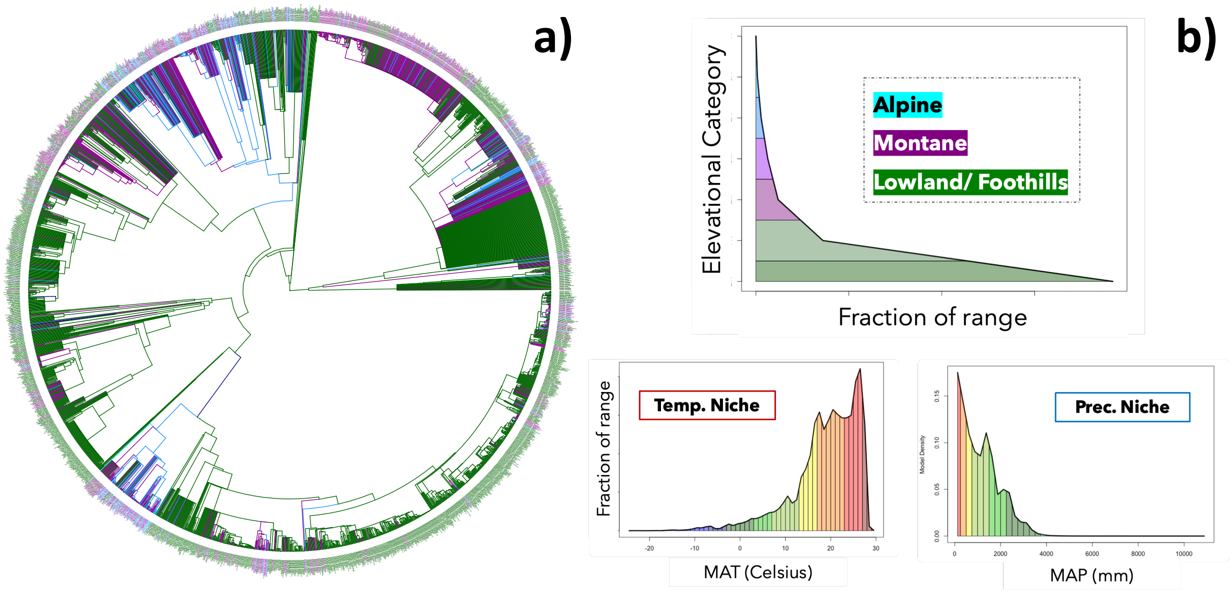


Figure 4.1. Ancestral state reconstruction of abiotic niches. (*a*) Simplified illustration of continuous character ancestral state reconstruction of elevational niches for Caryophyllales. Tips are colored according to habitat type at the center of each species' range (*alpine: blue; montane: purple; lowland: green*). Branches are colored according to the reconstructed majority habitat type using the same color scheme. For Caryophyllales, multiple transitions into alpine habitats are observed across the phylogeny. Although for illustrative purposes only a single color (habitat type) is shown, in reality the reconstructions produced a distribution of values as illustrated in (*b*) for the predicted elevation, temperature, and precipitation niches at the root state. In this case, the root state for Caryophyllales describes a range centered around 20° C MAT and ~1500 mm MAP, occurring in primarily frost-free lowlands. Taken together, these conditions are consistent with a low-elevation tropical to subtropical/ temperate rainforest with little to no frost exposure.

We performed reconstructions only on those taxonomic orders with at least 100 species represented in our dataset in order to reduce the impact of extreme (outlier) phenotypes on the estimation of ancestral states (CITE?). The clades chosen included both those with and without contemporary alpine plants, permitting a comparative approach. The use of the time-calibrated phylogeny from Smith and Brown (2018) also allowed us to quantify the timing of the emergence of alpine character (defined here as a node predicted to have at least 20% of the range within the alpine belt) across different groups. The backbone taxonomy from the World

Checklist of Selected Plant Families (WCSP 2020) was used to match species to taxonomic families and orders.

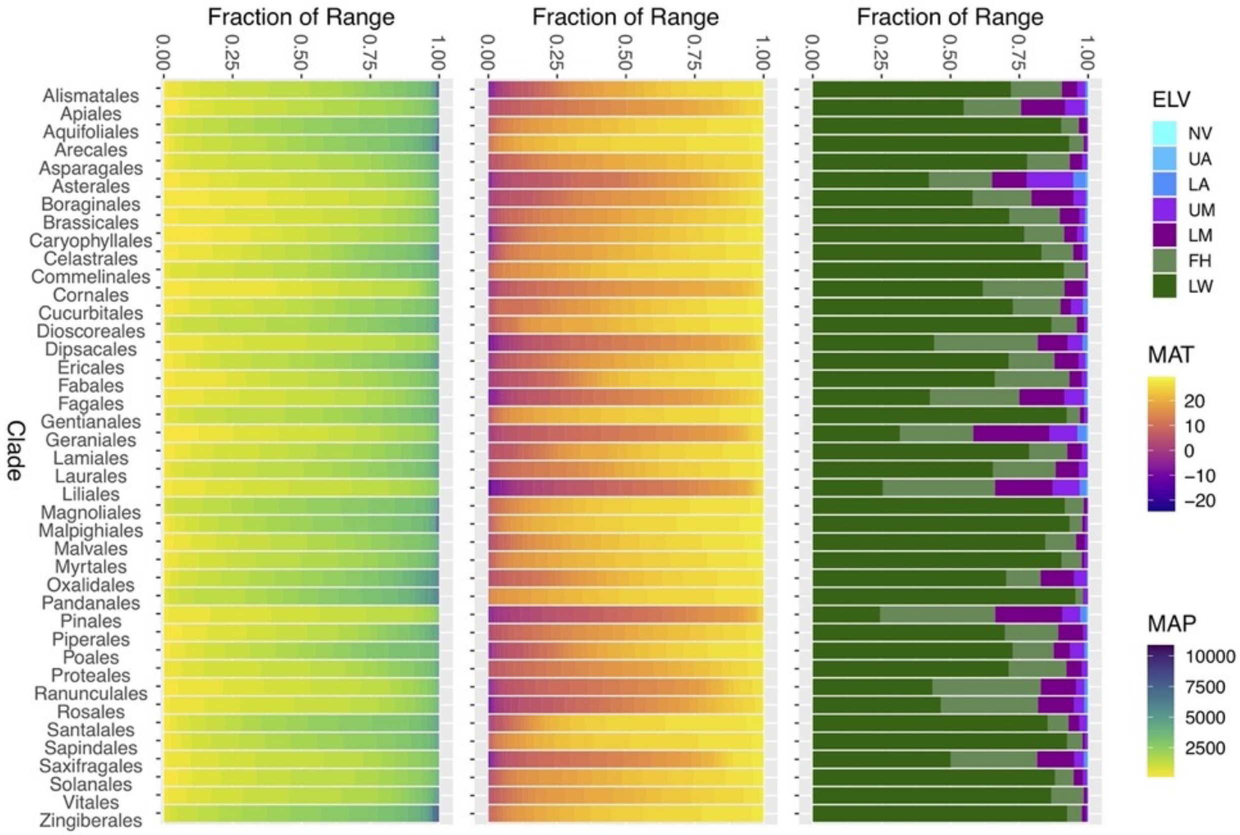


Figure 4.2. Reconstructed root state niches for precipitation (*MAP*), temperature (*MAT*), and elevation (*ELV*) for all taxonomic orders (*Clade*) with at least 100 contemporary species. Bars represent the distribution of climatic conditions predicted for the MRCA of all taxa in that clade, with the length of each color indicating the fraction of the range predicted to encompass that value of the abiotic variable. Most root states described almost exclusively frost-free lowland conditions centered above 18° C MAT with between 845-2500 mm MAP, consistent with a low elevation tropical to subtropical rainforest. Some clades, such as Rosales and Saxifragales showed comparatively lower predicted temperatures at the root, whereas clades devoid of contemporary alpine species—such as Arecales or Magnoliales—had root states restricted to much warmer temperatures. This was not true for all clades with contemporary alpine species, however. Caryophyllales, for example, was predicted to have a relatively warm root state.

RESULTS

Continuous character reconstructions provided insight into the potential distribution of abiotic conditions at internal nodes of each clade's phylogeny. **Figure 4.1** illustrates a representative example of abiotic niche continuous character reconstruction using data from

Caryophyllales. In this case, multiple transitions into both montane and alpine habitat are discernible among extant taxa and along internal branches. The reconstructed root-state niche for this group described a primarily (~77%) lowland habitat, with potential frost-exposed low elevation foothills, centered at 19° C MAT and ~1000 mm MAP. Additional reconstructions for individual clades are provided in **Figure 4.S1** and tabular values of reconstructed niches for each clade are available in the **Online Supplement**.

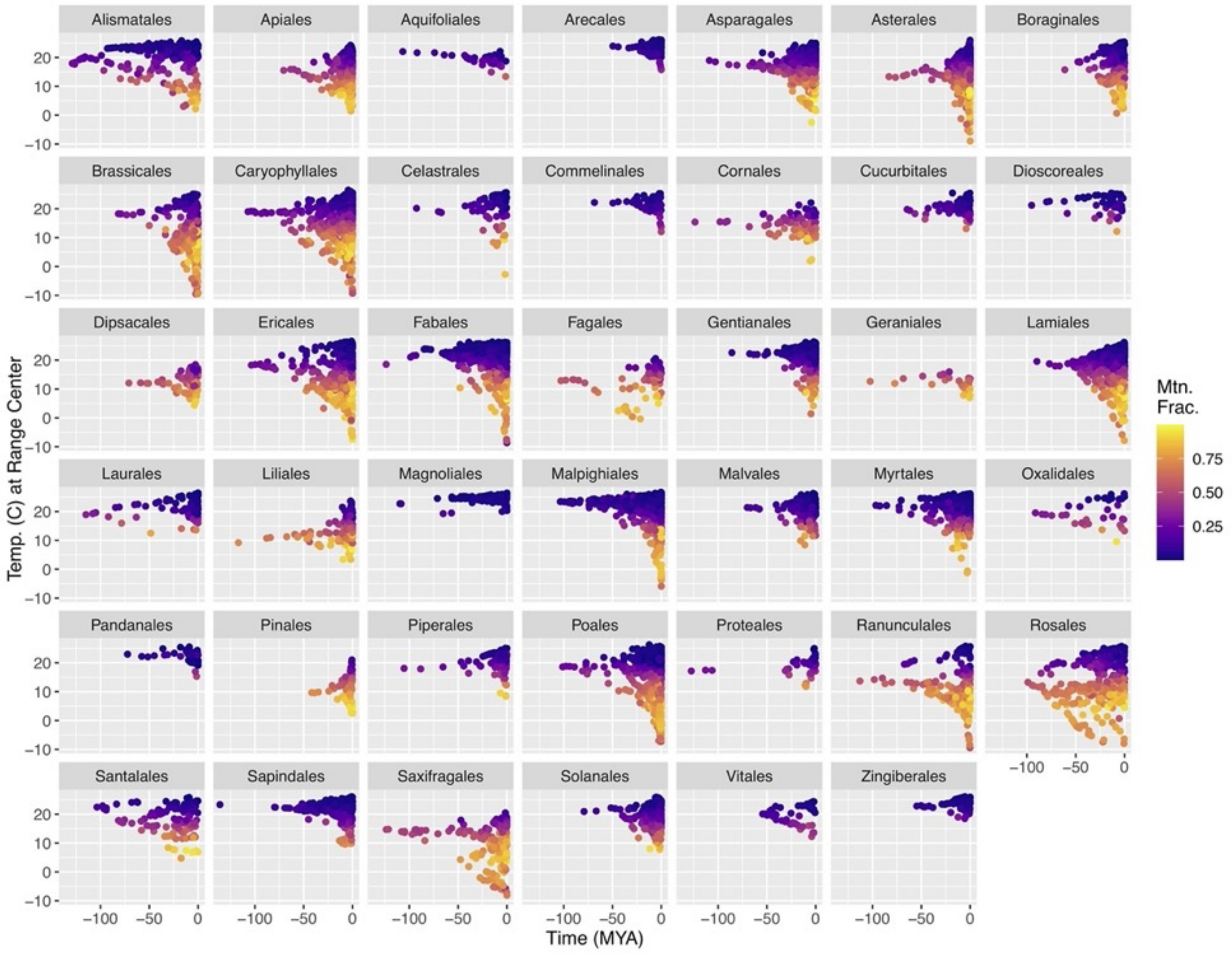


Figure 4.3. Exploration of temperature and elevation niche space over evolutionary time for each taxonomic order with at least 100 contemporary species. Each dot represents a single reconstructed node. Dots are colored based on the fraction of high-elevation habitat predicted to be occupied by species at that node. Vertical axis indicates the temperature (*°C*) at the center of species' ranges. The horizontal axis is time before the present (*MYA*). Clades that do not explore temperatures lower than ~10° C also do not enter higher elevation environments, consistent with the hypothesis that tolerance to freezing is a prerequisite for survival at higher elevations.

Reconstructed root-state niches for all clades were generally similar (**Figure 4.2**). In particular, elevation niches almost exclusively described primarily frost-free lowland conditions. Greater variation occurred for temperature and precipitation root-state niches, though most were centered above 18° C MAT with between 845-2500 mm MAP. Some differences between clades with and without contemporary alpine species were observed even at the reconstructed root state. For instance, clades such as Rosales and Saxifragales showed the suggestion of shifts to lower temperatures at the root, whereas clades devoid of contemporary alpine species—such as Arecales or Magnoliales—had root states restricted to much warmer temperatures. This was not true for all clades with contemporary alpine species, however. Caryophyllales, for example, was predicted to have a relatively warm root state.

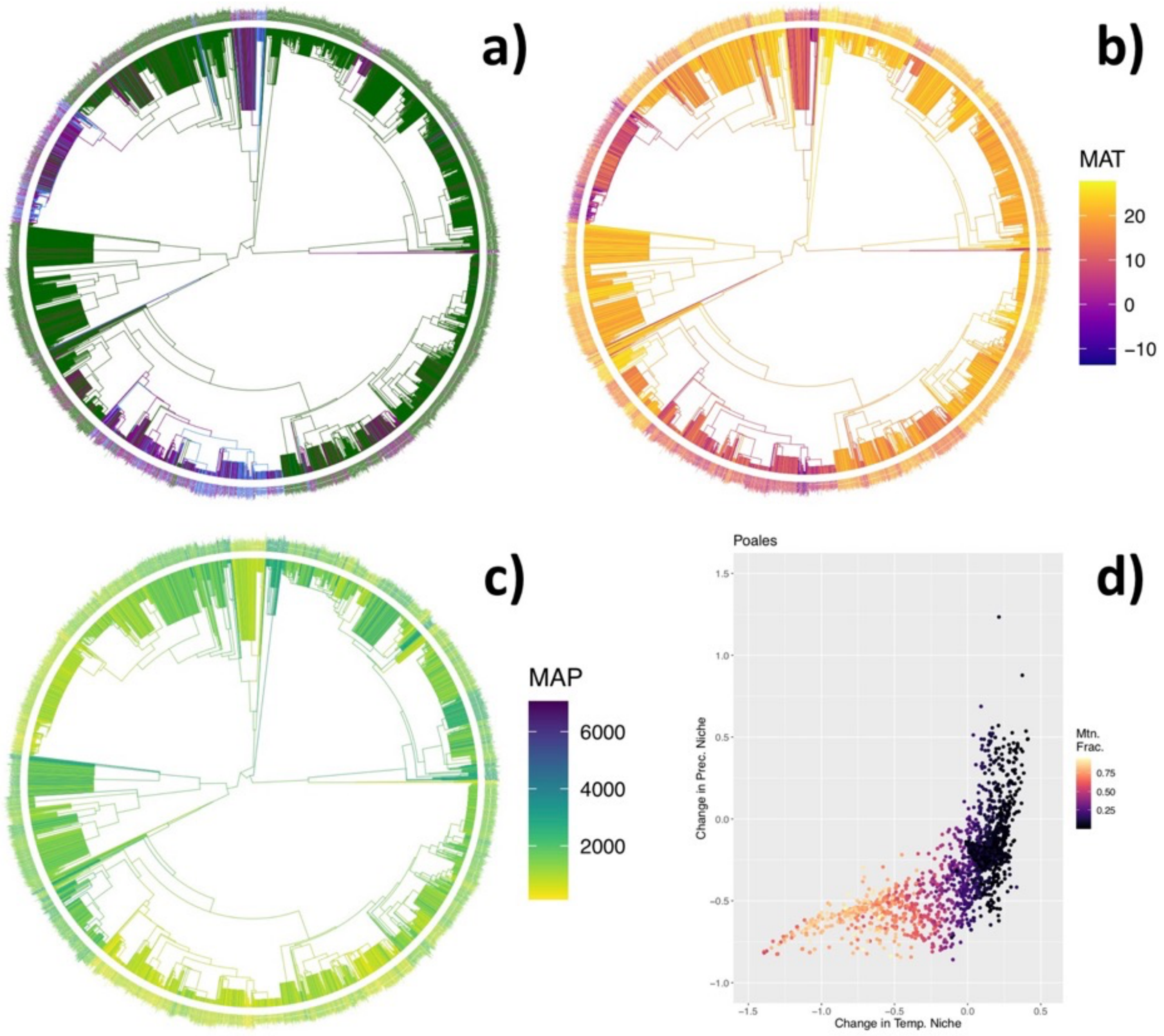


Figure 4.4. Comparison of the joint evolution of elevation (*a*), temperature (*b*), and precipitation (*c*) niches for Poales. This clade shows three major transitions into higher elevation regions (shaded regions in panel *a*), which correspond well with shifts to lower temperatures (*b*), but less so to shifts in precipitation (*c*). (*d*) Scatter plot showing the change in precipitation (vertical axis) and temperature (horizontal axis) from the root state (origin) for each reconstructed node (colored points). Nodes are colored based on the fraction of high-elevation habitat predicted to be occupied by species at that node. Shifts toward higher elevations are associated with shifts toward both colder and drier conditions, however the association with precipitation shifts is less pronounced.

Despite the general similarity of root-state niches, clades showed unique changes in temperature (**Figure 4.3**) and precipitation (**Figure 4.S2**) niches over time. Some clades (e.g., Arecales' and Magnoliales' temperature niches) explored relatively small portions of niche space whereas others (e.g., Ericales and Lamiales) diverged widely over time. Changes in occupied temperature and precipitation niche space were related to changes in elevation (**Figure 4.4** and **Figure 4.S3**), in particular shifts toward higher elevation were associated with decreased temperature and clades devoid of contemporary alpine species showed little to no exploration of low-temperature niche space.

We also used our reconstructions of elevation niches to investigate the distribution of ages for significantly alpine nodes (**Figure 4.5**-**4.6**). We found that nodes indicating at least 20% alpine habitat were no older than 50 MY, and the vast majority were younger than 10 MY. Further only 36.5% (N=15) of taxonomic orders showed any nodes indicating such a high fraction of alpine habitat (see *Discussion*). Clade-specific variation in the distribution of alpine node ages was also observed, with Rosales having the oldest nodes with alpine character. Ericales and Fagales showed a relative increase in the fraction of alpine nodes at ~30 MY not observed for other clades, and the majority of alpine nodes between 10 and 30 MYA occurred within the Saxifragales. Associations between alpine node age and climatic events were equivocal, however, the proportion of nodes with alpine character increased sharply and abruptly at ~50 MY (see *Discussion*). In **Figure 4.7** the ancestral state reconstruction for Fabales illustrates some of the scale-dependent complexities of interpreting reconstructions. In particular, although the phylogenetic distribution of contemporary alpine species suggests the potential for preadaptation at deeper timescales, the reconstructed timing of the origin of alpine character in specific evolutionary branches remains quite recent (see *Discussion*).

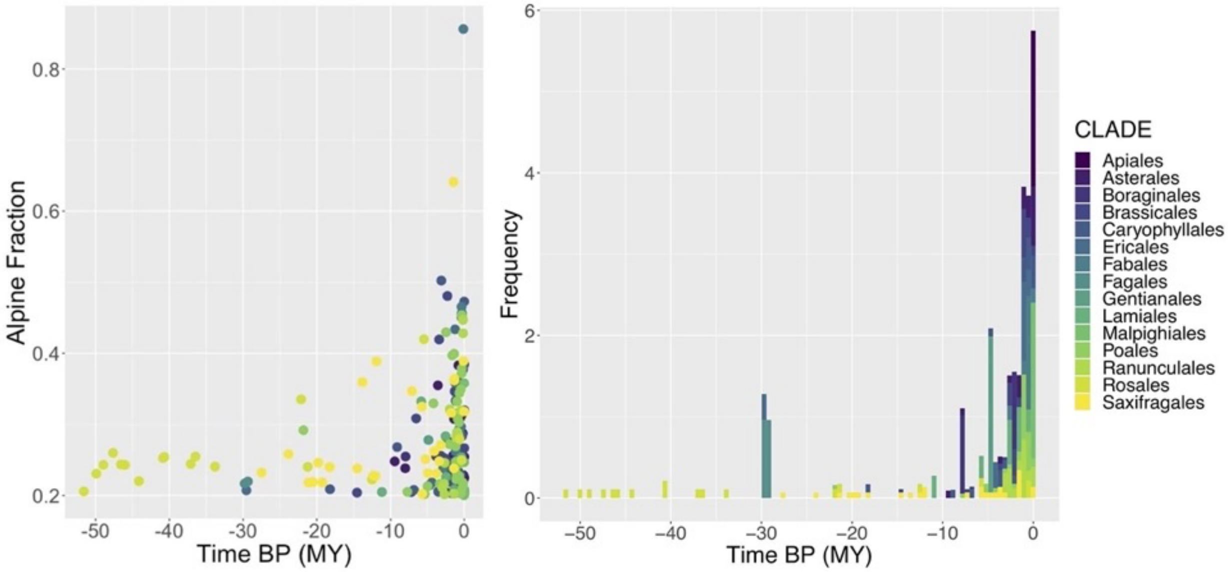


Figure 4.5. Age distributions of nodes with significantly alpine character. Significance was defined here as at least 20% of the predicted range falling within alpine habitat. Such nodes were no older than 50 MY, and the vast majority were younger than 10 MY, consistent with recent evolutionary adaptation to alpine habitat for seed plants. (*a*) Here, each dot represents a single node, colored by clade, and the vertical axis shows the fraction of alpine habitat predicted for that node. Nodes with increasing alpine character are also increasingly younger. (*b*) Frequency histogram for the age distribution of alpine nodes within each taxonomic order. Note that clade frequencies are calculated separately here to prevent the signal from larger clades masking that of smaller ones. This permits easier visualization of clade-specific differences. For instance, all nodes older than ~30 MY are from Rosales and the increase at ~30 MY is driven solely by Ericales and Fagales.

DISCUSSION

Comparison of root state niches.

We performed continuous character ancestral state reconstructions of temperature, precipitation, and elevation niches for all taxonomic orders in our dataset with at least 100 contemporary species present in our phylogeny. For the angiosperms, this represented ~140 MY of evolutionary history—the taxonomic organization of gymnosperms remains contentious (Lu et al. 2014), likely influenced strongly by sampling, and is not a primary focus of these analyses. The composite niche for the MRCA of each separate clade (**Figure 4.2**) was generally consistent with a low-elevation tropical to subtropical/ temperate rainforest with little to no frost exposure. Recognizable angiosperm lineages are generally recognized as having arisen during the Cretaceous (Friis et al., 2011; Taylor & Hickey, 1996), during a time of significantly warmer global temperatures, though ecological dominance might have been delayed considerably

(Ramírez-Barahona et al., 2020). Evidence for the presence of stable polar icecaps during this time remains equivocal (Friis et al., 2011; Graham, 2011), however it appears likely that angiosperms generally diversified under much warmer conditions than present today, consistent with our result for reconstructed temperature niches. Throughout the Cretaceous, mid-latitude environments oscillated between humid and arid conditions (Friis et al., 2011), which might explain the greater degree of variation in reconstructed precipitation niches observed here.

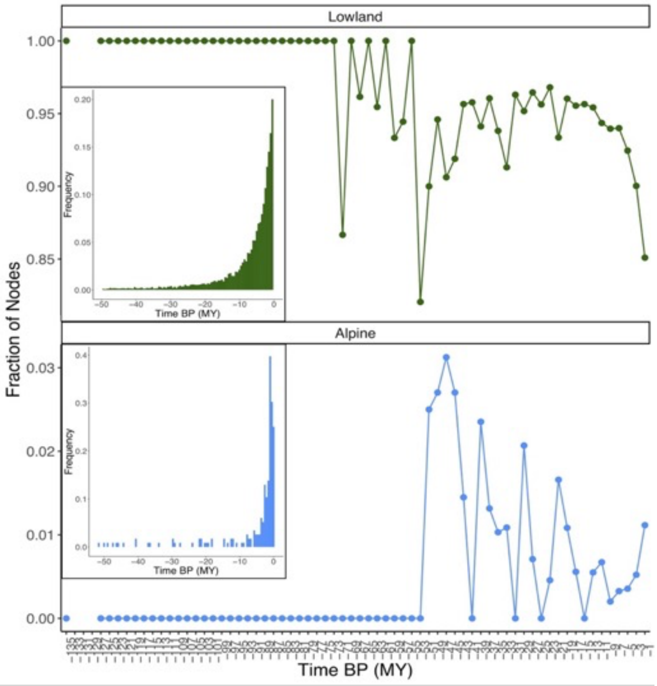


Figure 4.6. Graphical representation of the fraction of nodes with significant (defined as at least 20%) lowland (green) or alpine (blue) character across time for all seed plants in the dataset. At deeper time, almost all nodes are predicted to be lowland. However, beginning at ~50 MYA, a sharp and abrupt increase in the fraction of alpine nodes occurs, suggesting this marks a time period of increased transitions into alpine habitat for seed plants. Frequency histograms for the age distribution of nodes in either category are shown as inset plots. (*Note that because we define significance as being at least 20% of the range falling within a particular habitat, it is possible for a species to have a significant fraction of both alpine and lowland habitat across its range*.)

Age of alpine nodes.

Paleoelevation reconstructions remain a challenging are of intense research (Botsyun & Ehlers, 2021); nevertheless, the tectonically active Cretaceous, during which angiosperms arose, certainly resulted in significant volcanism, erosion, and uplift. At the same time, the results here suggest that lineages currently occupying alpine environments are relatively young (< 50 MY). We examined the emergence of significant alpine character for both specific lineages (**Figure 4.5**), and across all clades (**Figure 4.6**). We used the proportion of nodes at a given timeframe, rather than the absolute number of nodes, as there will always be a greater number of recent (and lowland) nodes than temporally distant (or alpine) ones. Using this method, we observed a sharp increase in the fraction of nodes with alpine character at ~50 MYA. This is generally consistent with other estimates in the literature. For instance, the current Rocky Mountain flora is estimated to be approximately 10 MY (Hughes & Atchison, 2015; Muellner-Riehl et al., 2019; Wolfe, 1987), and that of the Andean páramos could be as young as ~3.5 MY (Hughes & Atchison, 2015; Madriñán et al., 2013). Outside the Americas, Ding et al. (2020) recently traced the evolutionary origins of modern alpine Himalayan flora to the Oligocene (~20 MY). While our reconstructions do push the dates for the emergence of alpine character slightly farther back in time for some lineages (particularly Rosales) than these previous estimates, this is likely due to methodological differences. Here, we incorporated the fraction of each species range falling within a particular habitat type, and so our reconstructions are not binary 'alpine/ non-alpine' designations, but rather descriptions of when alpine character is estimated to have emerged.

Nevertheless, if high-elevation alpine floras existed at deeper timescales (>50 MYA), the signature of such assemblages appears quite elusive. This could indicate that elevational niches are somewhat labile over evolutionary timescales, and that lineages easily respond to the uplift and erosion of mountain ranges by moving up and down elevations as conditions change (Donoghue, 2008; Donoghue & Edwards, 2014). However, this is somewhat in conflict with the strong degree of observed niche conservatism for elevational niches (**CHAPTER TWO**). The absence of older age estimates for alpine lineages and assemblages might also indicate that alpine habitats simply are relatively young. This would be consistent with the relatively smaller number of species able to inhabit the alpine belt across a wide range of taxa (e.g., Perrigo et al., 2020).

Joint evolution of abiotic niches.

Niche evolution does not occur in a vacuum, and changes in one niche axis might be accompanied by changes in another. Recent analyses have suggested links between cold and arid adaptations across the evolutionary history of the angiosperms (Folk et al., 2020), and survival in the alpine belt is inherently tied to tolerance of freezing conditions (Körner, 2003). We have previously shown that Poales is a primary component of alpine assemblages across the Americas (Figueroa et al. 2021; **CHAPTER THREE**), and their success in such environments could be tied to adaptations to arid habitats dating back to the Cretaceous (Bouchenak-Khelladi et al., 2014). By simultaneously reconstructing the temperature, precipitation, and elevational niche history of different clades, we hoped to gain a better understanding of the joint evolution of niche shifts.

Temperature and elevation niche evolution were intimately related (**Figures 4.3**, **4.S2**, **4.4**, and **4.S3**), whereas connections to precipitation were less apparent. The biggest difference in our analyses between clades containing contemporary high elevation species and those without was the exploration of sufficiently low temperature niche space (**Figure 4.4**). This is consistent with a growing body of research suggesting that tolerance to freezing temperatures is a primary determinant of contemporary species distributions (e.g., (Segovia et al., 2020; Zanne et al., 2014), and suggests a tight link between the ability to tolerate freezing temperatures and survival at high elevations. A potential caveat between this link, however, is that the elevational categories used in this study are fundamentally tied to temperature and growing season length (Körner et al., 2011; Körner & Paulsen, 2004). This is due to the strong link between temperature conditions and the elevation at which the treeline occurs (Körner & Paulsen, 2004). Rather than seeing this as a critique, we suggest that this fact simply further underscores the tight link between temperature and elevation niches.

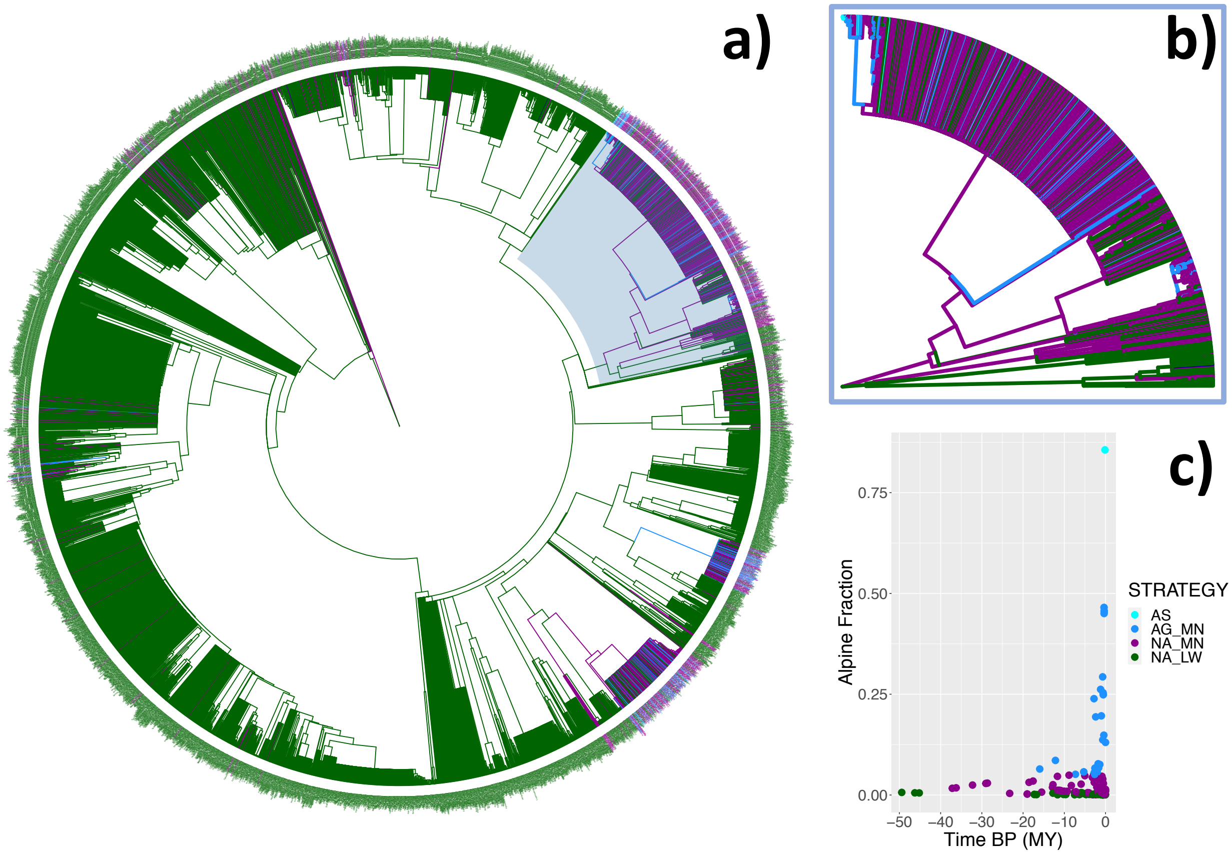


Figure 4.7. Reconstructed Fabales elevation niches with age distribution. (*a*) Ancestral state reconstruction of Fabales elevational niche, using the same simplified visualization

procedure as in **Figure 4.1** (*alpine: blue; montane: purple; lowland: green*). For this clade, there is a single group that constitutes a majority of transitions into higher elevation habitat (*shaded box*). (*b*) Magnification of this high-elevation Fabales clade, illustrating the spectrum of elevational niches. Genera within this clade include *Astragalus, Oxytropis, Hedysarum,* and *Lupinus*—all known from mountainous and frost-exposed habitats. (*c*) Age distribution of nodes circumscribed by this clade, colored by elevational strategy (*AS: alpine specialist, centered in alpine habitat; AG_MN: alpine generalist, centered in montane habitat; NA_MN: non-alpine species centered in montane habitat; NA_LW: non-alpine species centered in lowland habitat*). These results predict this high-elevation clade to be ~50 MY old, with transitions into alpine habitat occurring around 20 MYA. Nevertheless, tip ages for alpine species here are much younger (<2 MY).

Conclusions and further considerations.

For taxonomic orders with contemporary alpine species, it was often the case that a small number of distinct clades accounted for most, if not all, of their alpine taxa. For instance, among Poales (**Figure 4.4**), three clades comprise most of the alpine species in this group. This is remarkable considering that Poales was the most speciose component of alpine diversity across the Americas in our dataset (Figueroa et al. 2021; **CHAPTER THREE**). Similarly, among Fabales (**Figure 4.7**), a single clade contained most of the alpine species, with two, distinct, smaller clades comprising the rest. When these observations are taken together with the strong degree of elevational niche conservatism observed (**CHAPTER TWO**), they suggest that transitions into higher elevations are relatively rare within the angiosperms. For the Fabales higher-elevation clade, the prediction age for the MRCA of species comprising this clade was ~50 MYA, however the tip ages among contemporary alpine taxa were much younger (~2 MY). This could suggest a role for pre-adaptations to higher elevations that began to arise in this lineage relatively early.

These observations beg the question of whether there are strong connections between the ages of alpine lineages and the age of geologic events that might have created alpine habitat. Such interdisciplinary research questions would be of great interest to alpine biologists and should be considered a high priority for growing research programs. Other than the general trend of decreasing global temperatures beginning around 50 MYA (Graham 2011), the age distributions of alpine nodes (**Figures 4.5**-**4.6**) do not closely align with any major geologic events (though see Ding et al., 2020, for some possible links between climate, geology, and community assembly for the alpine Himalayas). Continued collaboration between alpine biologists, geologists, and paleo-climatologists should be encouraged in order to gain a better appreciation of when and where in Earth's history alpine habitats might have arisen and how such events might be linked to the phylogenetic and biogeographic history of the lineages currently occupying these remote habitats.

The results offered here represent a first pass and proof-of-concept for how continuous character reconstructions might be applied to a large dataset to gain a better appreciation for the history of niche conservatism among angiosperms. Nevertheless, the dataset used here was limited to American lineages. An important next step, therefore, would be to expand this niche model dataset to all angiosperms, as questions regarding the evolutionary origins of alpine character might be better addressed at the global scale. Although there has been some recent work in addressing questions of global patterns of alpine community assembly in relation to climate (e.g., Qian et al., 2021; Testolin et al., 2021), the analysis of a global niche model dataset with the methodologies described here would represent a novel undertaking—especially if done in concert with an assessment of the ages of paleoaltitudes estimated by geologists.

REFERENCES

Amini, E., Kazempour-Osaloo, S., Maassoumi, A. A., & Zare-Maivan, H. (2019). Phylogeny, biogeography and divergence times of *Astragalus* section *Incani* DC. (Fabaceae) inferred from nrDNA ITS and plastid *rpl* 32 *-trn* L $_{\text{(UAG)}}$ sequences. *Nordic Journal of Botany*, *37*(2), njb.02059. https://doi.org/10.1111/njb.02059

Amiri, M. S., Joharchi, M. R., & Nadaf, M. (2020). Ethnobotanical knowledge of Astragalus spp.: The world's largest genus of vascular plants. *Avicenna J Phytomed*, *10*(2), 128–142.

Antonelli, A., Nylander, J. A. A., Persson, C., & Sanmartín, I. (2009). Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proceedings of the National Academy of Sciences*, *106*(24), 9749–9754. https://doi.org/10.1073/pnas.0811421106

Bagheri, A., Maassoumi, A. A., Rahiminejad, M. R., Brassac, J., & Blattner, F. R. (2017). Molecular phylogeny and divergence times of Astragalus section Hymenostegis: An analysis of a rapidly diversifying species group in Fabaceae. *Scientific Reports*, *7*(1), 14033. https://doi.org/10.1038/s41598-017-14614-3

Barthlott, W., Hostert, A., Kier, G., Küper, W., Kreft, H., Mutke, J., Rafiqpoor, M. D., & Sommer, J. H. (2007). Geographic Patterns of Vascular Plant Diversity at Continental to Global Scales (Geographische Muster der Gefäßpflanzenvielfalt im kontinentalen und globalen Maßstab). *Erdkunde*, *61*(4), 305–315.

Beach, J., Stewart, A., & Grady, C. (2020). *Lifemapper web services.* available at: http://lifemapper.org.

Botsyun, S., & Ehlers, T. A. (2021). How Can Climate Models Be Used in Paleoelevation Reconstructions? *Frontiers in Earth Science*, *9*, 624542. https://doi.org/10.3389/feart.2021.624542

Bouchenak-Khelladi, Y., Muasya, A. M., & Linder, H. P. (2014). A revised evolutionary history of Poales: Origins and diversification: Evolutionary History of Poales. *Botanical Journal of the Linnean Society*, *175*(1), 4–16. https://doi.org/10.1111/boj.12160

Ding, W.-N., Ree, R. H., Spicer, R. A., & Xing, Y.-W. (2020). Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. *Science*, *369*(6503), 578–581. https://doi.org/10.1126/science.abb4484

Donoghue, M. J. (2008). A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences*, *105*(Supplement 1), 11549–11555. https://doi.org/10.1073/pnas.0801962105

Donoghue, M. J., & Edwards, E. J. (2014). Biome Shifts and Niche Evolution in Plants. *Annual Review of Ecology, Evolution, and Systematics*, *45*(1), 547–572. https://doi.org/10.1146/annurev-ecolsys-120213-091905

Elsen, P. R., & Tingley, M. W. (2015). Global mountain topography and the fate of montane species under climate change. *Nature Climate Change*, *5*(8), 772–776. https://doi.org/10.1038/nclimate2656

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Figueroa, H., Marx, H. E., De Souza Cortez, M. B., Grady, C. J., Engle-Wrye, N. J., Beach, J., Stewart, A., Folk, R. A., Soltis, D. E., Soltis, P. S., & Smith, S. A. (2021). Contrasting patterns of phylogenetic diversity and alpine specialization across the alpine flora of the American mountain range system. *Alpine Botany*, *In press*. https://doi.org/10.1007/s00035-021-00261-y

Flantua, S. G. A., O'Dea, A., Onstein, R. E., Giraldo, C., & Hooghiemstra, H. (2019). The flickering connectivity system of the north Andean páramos. *Journal of Biogeography*, *46*(8), 1808–1825. https://doi.org/10.1111/jbi.13607

Folk, R. A., Siniscalchi, C. M., & Soltis, D. E. (2020). Angiosperms at the edge: Extremity, diversity, and phylogeny. *Plant, Cell & Environment*, *43*(12), 2871–2893. https://doi.org/10.1111/pce.13887

Friis, E. M., Crane, P., & Pedersen, K. R. (2011). *Early Flowers and Angiosperm Evolution.* (1st ed.). Cambridge University Press.

González-Caro, S., Duque, Á., Feeley, K. J., Cabrera, E., Phillips, J., Ramirez, S., & Yepes, A. (2020). The legacy of biogeographic history on the composition and structure of Andean forests. *Ecology*, *101*(10). https://doi.org/10.1002/ecy.3131

Graham, A. (2011). The age and diversification of terrestrial New World ecosystems through Cretaceous and Cenozoic time. *American Journal of Botany*, *98*(3), 336–351. https://doi.org/10.3732/ajb.1000353

Hinchliff, C. E., Smith, S. A., Allman, J. F., Burleigh, J. G., Chaudhary, R., Coghill, L. M., Crandall, K. A., Deng, J., Drew, B. T., Gazis, R., Gude, K., Hibbett, D. S., Katz, L. A., Laughinghouse, H. D., McTavish, E. J., Midford, P. E., Owen, C. L., Ree, R. H., Rees, J. A., … Cranston, K. A. (2015). Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proceedings of the National Academy of Sciences*, *112*(41), 12764–12769. https://doi.org/10.1073/pnas.1423041112

Hoorn, C., Perrigo, A., & Antonelli, A. (2018). Mountains, Climate, and Biodiversity: An Introduction. In *Mountains, Climate, and Biodiversity* (1st ed., p. 14). John Wiley & Sons, Ltd.

Hughes, C., & Atchison, G. W. (2015). The ubiquity of alpine plant radiations: From the Andes to the Hengduan Mountains. *New Phytologist*, *207*(2), 275–282. https://doi.org/10.1111/nph.13230

Hughes, C., & Eastwood, R. (2006). Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences*, *103*(27), 10334–10339. https://doi.org/10.1073/pnas.0601928103

Körner, C. (2003). *Alpine plant life: Functional plant ecology of high mountain ecosystems.* (2nd ed.). Springer.

Körner, C., Jetz, W., Paulsen, J., Payne, D., Rudmann-Maurer, K., & M. Spehn, E. (2017). A global inventory of mountains for bio-geographical applications. *Alpine Botany*, *127*(1), 1–15. https://doi.org/10.1007/s00035-016-0182-6

Körner, C., & Paulsen, J. (2004). A world-wide study of high altitude treeline temperatures. *Journal of Biogeography*, *31*(5), 713–732. https://doi.org/10.1111/j.1365-2699.2003.01043.x

Körner, C., Paulsen, J., & Spehn, E. M. (2011). A definition of mountains and their bioclimatic belts for global comparisons of biodiversity data. *Alpine Botany*, *121*(2), 73–78. https://doi.org/10.1007/s00035-011-0094-4

Lu, Y., Ran, J.-H., Guo, D.-M., Yang, Z.-Y., & Wang, X.-Q. (2014). Phylogeny and Divergence Times of Gymnosperms Inferred from Single-Copy Nuclear Genes. *PLoS ONE*, *9*(9), e107679. https://doi.org/10.1371/journal.pone.0107679

Madriñán, S., Cortés, A. J., & Richardson, J. E. (2013). Páramo is the world's fastest evolving and coolest biodiversity hotspot. *Frontiers in Genetics*, *4*, 1–7. https://doi.org/10.3389/fgene.2013.00192

Muellner-Riehl, A. N., Schnitzler, J., Kissling, W. D., Mosbrugger, V., Rijsdijk, K. F., Seijmonsbergen, A. C., Versteegh, H., & Favre, A. (2019). Origins of global mountain plant biodiversity: Testing the 'mountain-geobiodiversity hypothesis.' *Journal of Biogeography*, *46*(12), 2826–2838. https://doi.org/10.1111/jbi.13715

Olmstead, R. G. (2013). Phylogeny and biogeography in Solanaceae, Verbenaceae and Bignoniaceae: A comparison of continental and intercontinental diversification patterns. *Botanical Journal of the Linnean Society*, *171*(1), 80–102. https://doi.org/10.1111/j.1095-8339.2012.01306.x

Perrigo, A., Hoorn, C., & Antonelli, A. (2020). Why mountains matter for biodiversity. *Journal of Biogeography*, *47*(2), 315–325. https://doi.org/10.1111/jbi.13731

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Phillips, S. J., Dudík, M., & Schapire, R. E. (2004). A maximum entropy approach to species distribution modeling. *Proceedings of the Twenty-First International Conference on Machine Learning*, 83. https://doi.org/10.1145/1015330.1015412

Qian, H., Ricklefs, R. E., & Thuiller, W. (2021). Evolutionary assembly of flowering plants into sky islands. *Nature Ecology & Evolution*, *5*, 640–646. https://doi.org/10.1038/s41559-021-01423-1

Rahbek, C., Borregaard, M. K., Antonelli, A., Colwell, R. K., Holt, B. G., Nogues-Bravo, D., Rasmussen, C. M. Ø., Richardson, K., Rosing, M. T., Whittaker, R. J., & Fjeldså, J. (2019). Building mountain biodiversity: Geological and evolutionary processes. *Science*, *365*(6458), 1114–1119. https://doi.org/10.1126/science.aax0151

Rahbek, C., Borregaard, M. K., Colwell, R. K., Dalsgaard, B., Holt, B. G., Morueta-Holme, N., Nogues-Bravo, D., Whittaker, R. J., & Fjeldså, J. (2019). Humboldt's enigma: What causes global patterns of mountain biodiversity? *Science*, *365*(6458), 1108–1113. https://doi.org/10.1126/science.aax0149

Ramírez-Barahona, S., Sauquet, H., & Magallón, S. (2020). The delayed and geographically heterogeneous diversification of flowering plant families. *Nature Ecology & Evolution*, *4*(9), 1232–1238. https://doi.org/10.1038/s41559-020-1241-3

Segovia, R. A., Pennington, R. T., Baker, T. R., Souza, F. C. de, Neves, D. M., Davis, C. C., Armesto, J. J., Olivera-Filho, A. T., & Dexter, K. G. (2020). Freezing and water availability structure the evolutionary diversity of trees across the Americas. *Science Advances*, *6*(19), eaaz5373. https://doi.org/10.1126/sciadv.aaz5373

Smith, S. A., & Brown, J. W. (2018). Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany*, *105*(3), 302–314. https://doi.org/10.1002/ajb2.1019

Sosef, M. S. M., Dauby, G., Blach-Overgaard, A., van der Burgt, X., Catarino, L., Damen, T., Deblauwe, V., Dessein, S., Dransfield, J., Droissart, V., Duarte, M. C., Engledow, H., Fadeur, G., Figueira, R., Gereau, R. E., Hardy, O. J., Harris, D. J., de Heij, J., Janssens, S., … Couvreur, T. L. P. (2017). Exploring the floristic diversity of tropical Africa. *BMC Biology*, *15*(15), 1–23. https://doi.org/10.1186/s12915-017-0356-8

Stevens, P. F. (2001). *Angiosperm Phylogeny Website. Version 14, July 2017 [and more or less continuously updated since].*

Taylor, D. W., & Hickey, L. (1996). *Flowering Plant Origin, Evolution & Phylogeny*. Chapman & Hall.

Testolin, R., Attorre, F., Borchardt, P., Brand, R. F., Bruelheide, H., Chytrý, M., De Sanctis, M., Dolezal, J., Finckh, M., Haider, S., Hemp, A., Jandt, U., Kessler, M., Korolyuk, A. Y., Lenoir, J., Makunina, N., Malanson, G. P., Montesinos-Tubée, D. B., Noroozi, J., … Jiménez-Alfaro, B. (2021). Global patterns and drivers of alpine plant species richness. *Global Ecology and Biogeography*, *30*(6), 1218–1231. https://doi.org/10.1111/geb.13297

Wolfe, J. A. (1987). An Overview of the Origins of the Modern Vegetation and Flora of the Northern Rocky Mountains. *Annals of the Missouri Botanical Garden*, *74*(4), 785–803. https://doi.org/10.2307/2399450

Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. A., FitzJohn, R. G., McGlinn, D. J., O'Meara, B. C., Moles, A. T., Reich, P. B., Royer, D. L., Soltis, D. E., Stevens, P. F., Westoby, M., Wright, I. J., Aarssen, L., Bertin, R. I., Calaminus, A., Govaerts, R., … Beaulieu, J. M. (2014). Three keys to the radiation of angiosperms into freezing environments. *Nature*, *506*(7486), 89–92. https://doi.org/10.1038/nature12872

SUPPLEMENTAL FIGURES & TABLES

****

The **Online Supplement** to this chapter contains the following:

- **Figure 4.S1**—ancestral state reconstructions for all 41 clades.
- Folder containing tabular values for all reconstructed niches.
- **Figure 4.S2**—Visualization of precipitation niche shifts over evolutionary time.
- **Figure 4.S3**—scatter plots showing the joint niche evolution of temperature, precipitation, and elevation for all 41 clades.

****

CONCLUSION

"It's always both."

-Dr. Deborah Goldberg, *Lecture on Scientific Dichotomies* (2017).

The notion that all life is related through evolutionary processes and shared common ancestry is both simple and profound in its far-reaching implications. This dissertation has examined one of the implications of this framework—namely, that close evolutionary relatives should share more in common than do more distantly related organisms—in order to better understand how contemporary flowering plant lineages are distributed across the Americas. Shared evolutionary history among angiosperm lineages can mean that closely related plants might exhibit similar traits and ecological strategies (**CHAPTER ONE**), share physiological and climatic tolerances (**CHAPTER TWO**), and may influence the phylogenetic diversity of plants along their overlapping ranges (**CHAPTER THREE**). The modern distributions of plants also give scientists a window into the past and might provide clues about the history of how niches have been retained or changed over evolutionary time (**CHAPTER FOUR**). Each chapter in this dissertation has investigated a different aspect of the niches of American angiosperms in relation to phylogenetic niche conservatism (PNC, as defined and discussed in the **INTRODUCTION**). Here, I will attempt to bring together the results from each study and show how the chapters present a guiding inferential framework with which to view how phylogenetic history has helped to shape the contemporary distributions of American angiosperms at the regional scale.

The appropriate phylogenetic scale of niche conservatism.

Scale-dependence is a near ubiquitous feature of biological systems (Swenson et al., 2006), and phylogenetic niche conservatism is no exception. For example, a field of ericaceous shrubs might all exhibit similar height—a functional trait important for light capture (**CHAPTER ONE**; **APPENDIX A**; D'Andrea et al., 2020; Figueroa & Smith, 2021)—despite representing numerous distinct genera. From this one might conclude that phylogeny has little bearing on plant height. At the same time, however, all such shrubs come from the same broader family (Ericaceae), which could indicate a tendency for conservation of plant height at that phylogenetic scale. Similarly, a community of alpine plants in the Rocky Mountains might contain representatives from *Astragalus* (milkvetch; Fabaceae), *Carex* (sedges; Cyperaceae), *Epilobium* (willowherbs; Onagraceae), and *Saxifraga* (rockfoils; Saxifragaceae), which jointly span a great fraction of the total known phylogenetic history of the angiosperms. This could indicate that there is no association between phylogeny and the ability to enter the alpine belt in this region. However, a closer look at each of the families from which these lineages have arisen would show that only certain clades in each branch of the angiosperm tree has been able to adapt to life at high elevations (**CHAPTER FOUR**). Thus, while at the broad phylogenetic scale of the entire community, no niche conservatism is observed, a more focused look at the evolutionary context of each lineage reveals a strong signal of phylogenetic and biogeographic history (**CHAPTERS THREE & FOUR**).

In **CHAPTER ONE**, I showed that functional traits related to resource acquisition (e.g., height, seed mass, growth form, and wood density) tended to be most similar among congeners and that variation in the trait syndrome represented by these traits was best observed by examining con-ordinal species from different families. At the same time, in **APPENDIX A**, I discuss work demonstrating that traits related to chemical defense can vary substantially among congeners, creating limited relation to phylogeny for some study systems. Conservation of climatic niches seems to be intermediate, with congeners tending toward similar climatic tolerances but with statistically significant amounts of niche shift still being present (**CHAPTER TWO**). Therefore, although phylogenetic scale is important in each case, which phylogenetic scale is most important can vary for different aspects of an organism's niche. This is an important consideration for studies employing genus-level phylogenies to investigate niche conservatism (e.g., Segovia et al., 2020), as such methods implicitly assume that within-genus variation is negligible. In this dissertation, I show that such an assumption must be explicitly tested for whichever niche axes are being investigated.

The results of this dissertation work also raise the question of whether comparisons between angiosperms from different taxonomic orders have much utility. The use of taxonomic ranks can be contentious, and it is generally recognized that such ranks do not define biologically equivalent units (Brummitt, 1997; Pleijel & Rouse, 2003). Nevertheless, taxonomic ranks are important here for two reasons: (1) they delimit groups of interest to researchers in the biological sciences[‡‡]; and (2) they might circumscribe clades of relatively independent evolutionary history. As an example, consider how direct comparisons of traits between a spider monkey (order: Primates), a leopard (Carnivora), and a bison (Artiodactyla) might be of limited use except in cases of great generality—such as the relationship between metabolic rate and body mass or other similar scaling relationships (Schmidt-Nielsen, 1984). Similarities among species in this example would generally be ascribed to deep shared ancestry (all are mammals) or convergent evolution (such as the forward-facing eyes with bony support of both cats and primates), and each represents a relatively independent evolutionary trajectory. For mammals, therefore, in practice, taxonomic orders often demarcate an evolutionarily relevant phylogenetic scale.

Does such an analogous phylogenetic scale exist for seed plants? Certainly, comparisons between gymnosperms and angiosperms might be of limited value, given the great amount of time since their last common ancestor. Measures of phylogenetic diversity and evolutionary novelty can be significantly impacted by grouping these two clades together (Moles et al., 2005; **CHAPTER THREE**). Further, within the angiosperms, in practice, monocots are often separated from eudicots in many analyses. For example, in the work summarized in **APPENDIX A** (D'Andrea et al., 2020), we excluded monocots, specifically palms, from trait-based analyses. This was not done because we believed monocots were somehow excluded from niche-based coexistence mechanisms, or that competition for light is somehow less important for palms, but rather due to the substantial physiological differences between them and true woody trees, which made comparisons between the manifestation of their functional traits inappropriate. Within angiosperms overall, the work presented in **CHAPTERS ONE** and **TWO** suggests that similarities in functional traits and climatic niches between members of different taxonomic orders might more likely represent convergent evolution and that species within such clades show as much variation as that seen across angiosperms as a whole.

The research offered in this dissertation suggests that for angiosperms, taxonomic orders might represent a functional phylogenetic scale at which meaningful comparisons can take place.

[‡‡] Some researchers might adopt titles that are implicitly agnostic to taxonomy, such as 'mountain ecologists.' In practice, however, even these researchers tend to focus on a particular group such as angiosperms, mammals, or birds. For example, a mountain ecologist focusing on fungi would likely need to make such a specialization apparent to their colleagues.

Future research should explicitly investigate this suggestion. Nevertheless, I believe there is a tendency among ecologists, myself included, to group all plants (especially angiosperms) together. However, the relatively small degree of shared ancestry between groups as diverged as Asterids and Rosids might preclude a valid comparison of many of their traits and their occupied niche space, and, therefore, by extension their ecological strategies and responses to community assembly dynamics and coexistence mechanisms.

***Filtering, biogeography, or ecological dynamics*?**

The composition of ecological communities can be shaped by abiotic filters, biogeographic and phylogenetic history, and by ecological dynamics. Although some researchers attempt to frame community assembly as being primarily driven by just one of these processes (e.g., Crisp et al., 2009; Segovia et al., 2020), in reality each process undoubtedly contributes to different extents among different lineages in different ways throughout their evolutionary history. The importance of such lineage-specific variation in these assembly processes for alpine communities was discussed in **CHAPTER THREE**. At a broader, conceptual level, however, phylogenetic niche conservatism (PNC) is relevant for each of these potential assembly mechanisms as evolutionary relationships could constrain physiology (thereby controlling the strength and extent of abiotic filtering) and traits relating to ecological interactions (thus impacting ecological dynamics such as competition), as well as impact distributional patterns and dispersal ability (thus influencing the contribution from biogeographic history). In this dissertation, I have shown some of the ways in which PNC acts at each of these levels and attempted to frame the results in such a way as to suggest that all of these processes are likely contributing to community assembly. Where possible, I have tried to highlight clade-specific variation in the contribution each process makes, and to analyze these questions at a variety of phylogenetic scales.

One group that highlights the complex and scale-dependent interplay of multiple processes acting to shape contemporary distributions are the Solanaceae. Olmstead (2013) used phylogenetic and biogeographic analyses to suggest a South American origin for this family. Contemporary Solanaceae are widespread throughout the Americas but exhibit strong niche conservatism of both cold and xeric intolerance (Olmstead, 2013). Species distribution models of Solanaceae showed that alpine members of this family are generally restricted to the Andes and tend to be alpine generalists and not specialists (**CHAPTER THREE**, Figueroa et al., *in press*). Taken together, these findings suggest that the absence of Solanaceae from non-Andean ranges might be due to ecological/ physiological limits rather than dispersal limitation. Contemporary Solanaceae distributions therefore reflect the influence of both abiotic filtering—which limits their latitudinal and elevational extents—as well as historical processes—substantially greater diversity in South America as a center of diversification for this clade.

To enhance the discussion of community assembly in the alpine belt, I have presented a summary comparison in **APPENDIX D** between alpine species and non-alpine taxa that otherwise inhabit the same temperature and precipitation niche space (termed Climatically Similar Non-Alpines; CSNA). These species provide a possible first approximation of which non-alpine American seed plant species might otherwise be able to inhabit the alpine belt and help address a key question of whether abiotic factors or historical processes (e.g., phylogenetic and biogeographic history) plays a larger role in alpine community assembly (e.g., Hughes & Eastwood, 2006). The geographically broad distribution of CSNA species, along with their apparent ability to tolerate the macro-scale climatic conditions of the alpine belt, and the fact that CSNA species richness was highest in regions of lowest alpine richness, could suggest that ecological factors such as competitive dynamics, rather than general dispersal limitation or broad physiological tolerances, constrains the modern ranges of these species. Such hypotheses should be tested explicitly by future studies. Nevertheless, these results highlight how a comparative approach incorporating multiple niche dimensions can provide a solid framework for looking at community assembly questions. Further, a taxonomic comparison of alpine and CSNA species revealed only ~28% overlap at the generic level, and ten taxonomic orders devoid of true alpines.

The observation that entire angiosperm orders seem to be devoid of alpine species in the Americas raises the question of what differentiates these groups. Ancestral state reconstructions of temperature and precipitation niches (**CHAPTER FOUR**) suggested that expansion into sufficiently cold temperature niche space is a requirement for expansion into higher elevations, whereas there was no similar requirement for how a clade explored precipitation niche space over evolutionary time. These results build on a growing body of literature suggesting that tolerance to cold or freezing temperatures is a primary determinant of the phylogenetic structure of communities at large spatial scales (Segovia et al., 2020; Zanne et al., 2014). Within a region, however, more local scale mesoclimatic gradients and biogeographic history might dominate (Testolin et al., 2021).

The use and abuse of big data to see the big picture.

Increasingly, biologists have access to large-scale datasets for their research questions. Examples of such growing and widely used datasets include global repositories of plant traits, such as TRY (Kattge et al., 2020) and BIEN (Enquist et al., 2016), and occurrence-record aggregators such as GBIF (www.gbif.org) and iDigBio (www.idigbio.org). Networks of forest inventory plots, such as those being compiled and sustained globally by ForestGEO (www.forestgeo.si.edu) further provide biologists with deep knowledge of local community structure within small plots that are censused and maintained using consistent practices across the globe. The increased availability of molecular and taxonomic data for classifying evolutionary relationships has led to the production and publication of large-scale phylogenies through projects such as the Open Tree of Life (www.opentreeoflife.github.io), as well as enhanced availability of software to build phylogenies based on existing backbone structures. Additionally, efforts are being made across the globe to provide scientists with detailed information on abiotic/climatic conditions at a variety of spatial scales, through organizations such as WorldClim (Fick & Hijmans, 2017), which provides information on climate variables and GMBA (www.gmba.unibe.ch), which integrates climate data with topography to provide a biologically meaningful assessment of the distribution of terrestrial elevational habitats.

These logistic, technological, and communication advancements are certainly a great boon and provide wonderful resources for developing engaging research questions among scientists working across the globe as well as providing information on biological questions at larger spatial scales. At the same time, it is a counterintuitive finding that adding more data does not always lead to clearer answers. This can be seen in phylogenetic studies quite readily, where adding additional gene sequences can actually increase difficulties in identifying the correct evolutionary relationships. It is important, therefore, that researchers question how best to use the vast and growing amounts of readily available information in order to engage in productive and meaningful research programs.

Throughout this dissertation, I have attempted to use the resources available in the modern age of “big data” and apply these to questions of community assembly and dynamics among angiosperms across the Americas. Much of this work has also involved assessing how best to employ large scale datasets to arrive at productive research questions. For example, in **CHAPTER ONE**, I integrated trait, occurrence record, and phylogenetic information but did so for only select clades (namely species within the taxonomic orders of Ericales and Fabales). This method of delimiting the taxonomic scope of which data would be used allowed me to demonstrate the phylogenetic scale over which a resource-acquisition trait syndrome could be observed. In contrast, other studies have examined a single trait and aggregated available data for all seed plants (e.g., Moles et al., 2005, 2009). This taxonomically broad approach can obscure the role that phylogeny and biogeography play in shaping contemporary trait distributions. Additionally, by examining multiple traits (instead of traits in isolation), I was able to ask how these traits might have evolved together to form an ecologically relevant trait syndrome.

Big data can also be used to highlight knowledge gaps or biases in our current understanding of angiosperm distributions. In **CHAPTER THREE**, I used a large dataset of over 72,000 seed plants to illustrate the current state of knowledge on alpine plants across the Americas. The results obtained helped to showcase the limited knowledge of alpine endemics, particularly in the Andes, and how that might impact our understanding of their evolution, as well as pointing out the need for increased sampling of alpine regions in the mountains of Central America and the Southwest of the United States. The dataset assembled and used for **CHAPTERS TWO**, **THREE**, and **FOUR** helps to demonstrate how continuous (re)evaluation of where the data we have are coming from is necessary to ensure that scientists are basing their research programs and questions on truly representative samples. Despite these potential limitations, however, harnessing big data can be extremely useful in comparative approaches to address broad evolutionary questions. For instance, in **CHAPTER FOUR**, I demonstrate that clades which do not explore sufficiently cold temperature niche space also never enter higher elevations, whereas no such relationship exists for precipitation. Here, big data was successfully applied to gain a "big picture" appreciation for the ways in which the realized niche space along different niche axes can be interdependent even among disparately related clades.

***Concluding thoughts*.**

The work presented in this dissertation represents a deep dive into how large-scale dataset mining can be used to assess how phylogenetic niche conservatism has helped to shape the modern distributions of American angiosperms. However, angiosperms do not exist in a vacuum, and their ecology and evolution are intimately tied to both the land in which they live and the other organisms co-occurring with them. A crucial next step, therefore, is to integrate the questions of angiosperm phylogenetic niche conservatism with both the geologic history of the Americas as well as the evolutionary and ecological history of other co-occurring organisms, such as soil microbes and terrestrial animals. For example, although we can trace the evolutionary origins of alpine character within different angiosperm groups (**CHAPTER FOUR**), it remains to be seen how the timing of such evolutionary characters corresponds with the timing of specific geologic events, such as mountain uplift or erosion. Integrating the evolution of alpine lineages with both alpine community assembly in specific areas and the geologic events of those areas remains an ongoing and active research area (but see Ding et al., 2020 for a recent exploration of the link between community assembly and geologic history). Additionally, although researchers have compiled large niche model datasets for both plant (e.g., the BiotaPhy platform) and animal (e.g., MapOfLife project) distributions across large regions, these data have yet to be combined in order to investigate how animal and plant distributions jointly interact to produce the ecosystem dynamics we observe. I believe that such integrations could be a productive endeavor for future research, both in terms of potential conservation implications as well as enhancing our understanding of how organisms from multiple trophic levels interact to produce the extraordinary complexity and diversity of the natural world.

REFERENCES

Brummitt, R. K. (1997). Taxonomy versus cladonomy, a fundamental controversy in biological systematics. *TAXON*, *46*(4), 723–734. https://doi.org/10.2307/1224478

Crisp, M. D., Arroyo, M. T. K., Cook, L. G., Gandolfo, M. A., Jordan, G. J., McGlone, M. S., Weston, P. H., Westoby, M., Wilf, P., & Linder, H. P. (2009). Phylogenetic biome conservatism on a global scale. *Nature*, *458*(7239), 754–756. https://doi.org/10.1038/nature07764

D'Andrea, R., Guittar, J., O'Dwyer, J. P., Figueroa, H., Wright, S. J., Condit, R., & Ostling, A. (2020). Counting niches: Abundance-by-trait patterns reveal niche partitioning in a Neotropical forest. *Ecology*, *101*(6). https://doi.org/10.1002/ecy.3019

Ding, W.-N., Ree, R. H., Spicer, R. A., & Xing, Y.-W. (2020). Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. *Science*, *369*(6503), 578–581. https://doi.org/10.1126/science.abb4484

Enquist, B. J., Condit, R., Peet, R. K., Schildhauer, M., & Thiers, B. M. (2016). *Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity* [Preprint]. PeerJ Preprints. https://doi.org/10.7287/peerj.preprints.2615v2

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Figueroa, H., & Smith, S. A. (2021). A targeted phylogenetic approach helps explain New World functional diversity patterns of two eudicot lineages. *Journal of Biogeography*, *48*(1), 202–215. https://doi.org/10.1111/jbi.13993

Hughes, C., & Eastwood, R. (2006). Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences*, *103*(27), 10334–10339. https://doi.org/10.1073/pnas.0601928103

Kattge, J., Str, H. K. … et al. (2020). TRY plant trait database – enhanced coverage and open access. *Global Change Biology*, *26*(1), 119–188. https://doi.org/doi:10.1111/gcb.14904

Moles, A. T., Ackerly, D., Webb, C., Tweddle, J., Dickie, J., & Westoby, M. (2005). A Brief History of Seed Size. *Science*, *307*(5709), 576–580. https://doi.org/10.1126/science.1104863

Moles, A. T., Warton, D. I., Warman, L., Swenson, N. G., Laffan, S. W., Zanne, A. E., Pitman, A., Hemmings, F. A., & Leishman, M. R. (2009). Global patterns in plant height. *Journal of Ecology*, *97*(5), 923–932. https://doi.org/10.1111/j.1365-2745.2009.01526.x

Olmstead, R. G. (2013). Phylogeny and biogeography in Solanaceae, Verbenaceae and Bignoniaceae: A comparison of continental and intercontinental diversification patterns. *Botanical Journal of the Linnean Society*, *171*(1), 80–102. https://doi.org/10.1111/j.1095-8339.2012.01306.x

Pleijel, F., & Rouse, G. W. (2003). Ceci n'est pas une pipe: Names, clades and phylogenetic nomenclature. *Journal of Zoological Systematics and Evolutionary Research*, *41*(3), 162–174. https://doi.org/10.1046/j.1439-0469.2003.00236.x

Schmidt-Nielsen, K. (1984). *Scaling. Why is animal size so important?* Cambridge University Press.

Segovia, R. A., Pennington, R. T., Baker, T. R., Souza, F. C. de, Neves, D. M., Davis, C. C., Armesto, J. J., Olivera-Filho, A. T., & Dexter, K. G. (2020). Freezing and water availability structure the evolutionary diversity of trees across the Americas. *Science Advances*, *6*(19), eaaz5373. https://doi.org/10.1126/sciadv.aaz5373

Swenson, N. G., Enquist, B. J., Pither, J., Thompson, J., & Zimmerman, J. K. (2006). The problem and promise of scale dependency in community phylogenetics. *Ecology*, *87*(10), 2418–2424. https://doi.org/10.1890/0012-9658(2006)87[2418:TPAPOS]2.0.CO;2

Testolin, R., Attorre, F., Borchardt, P., Brand, R. F., Bruelheide, H., Chytrý, M., De Sanctis, M., Dolezal, J., Finckh, M., Haider, S., Hemp, A., Jandt, U., Kessler, M., Korolyuk, A. Y., Lenoir, J., Makunina, N., Malanson, G. P., Montesinos-Tubée, D. B., Noroozi, J., … Jiménez-Alfaro, B.

(2021). Global patterns and drivers of alpine plant species richness. *Global Ecology and Biogeography*, *30*(6), 1218–1231. https://doi.org/10.1111/geb.13297

Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. A., FitzJohn, R. G., McGlinn, D. J., O'Meara, B. C., Moles, A. T., Reich, P. B., Royer, D. L., Soltis, D. E., Stevens, P. F., Westoby, M., Wright, I. J., Aarssen, L., Bertin, R. I., Calaminus, A., Govaerts, R., … Beaulieu, J. M. (2014). Three keys to the radiation of angiosperms into freezing environments. *Nature*, *506*(7486), 89–92. https://doi.org/10.1038/nature12872

APPENDIX A—Trait Clustering Work on the Barro Colorado Island 50 ha Plot.

In **CHAPTER ONE**, I examined variation in functional traits within two large and ecologically important clades—Ericales (Asterids) and Fabales (Rosids)—across the Americas. The goal with that chapter was to assess the degree of niche conservatism observed for ecologically relevant traits, such as seed mass, plant height, wood density, growth form, and leaf mass per area, in order to gain a better appreciation of how phylogeny influences the distributions of such traits within the broader evolutionary context of the lineage. I found that variation in these traits was best observed among species from different con-ordinal families, as congeners tended toward very similar trait values, and species from different taxonomic orders could either (1) converge to similar trait values, or (2) be influenced by clade-specific quantitative differences in trait values obtainable. My approach was inherently a macroecological one, and therefore the results might be expected to more likely capture biogeographical processes (Webb et al., 2002).

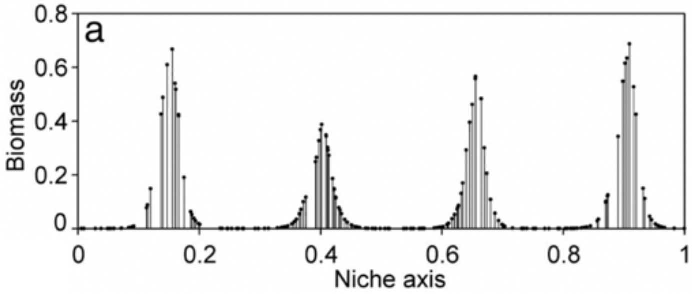


Figure A1. Advances in ecological theory suggest niche differences leading to stable coexistence at the local scale should produce species clusters exhibiting similar trait values, with maximal abundance (biomass) for species at cluster centers. *(Adapted from Figure 2 in Scheffer & van Nes 2006.)*

However, for traits relating to resource acquisition or other competitive dynamics, it is possible that the relevant spatial scale is much smaller, and that the important ecology of how such traits are distributed would be best captured within the local community. Revisions (Scheffer and van Nes, 2006) to classical (e.g., Macarthur and Levins, 1967) ecological theory posits that niche differences leading to stable coexistence at the local scale should manifest in clusters of species exhibiting similar trait values (**Figure A1**). Recent work by D'Andrea et al. (2020) was able to demonstrate the existence of such trait-clusters for wood density and plant height within a Neotropical forest community on the 50 ha Barro Colorado Island (BCI) plot. In this study, we were able to statistically validate the presence of trait clusters in this local community and interpreted these as evidence of niche partitioning structuring the trait-abundance relationships within the assemblage. In particular, the four clusters of maximum tree height we found were well correlated with previous suggestions (e.g., Terborgh, 1985) that tropical rainforests should have four (vertical) zones of maximal light capture (**Figure A2**).

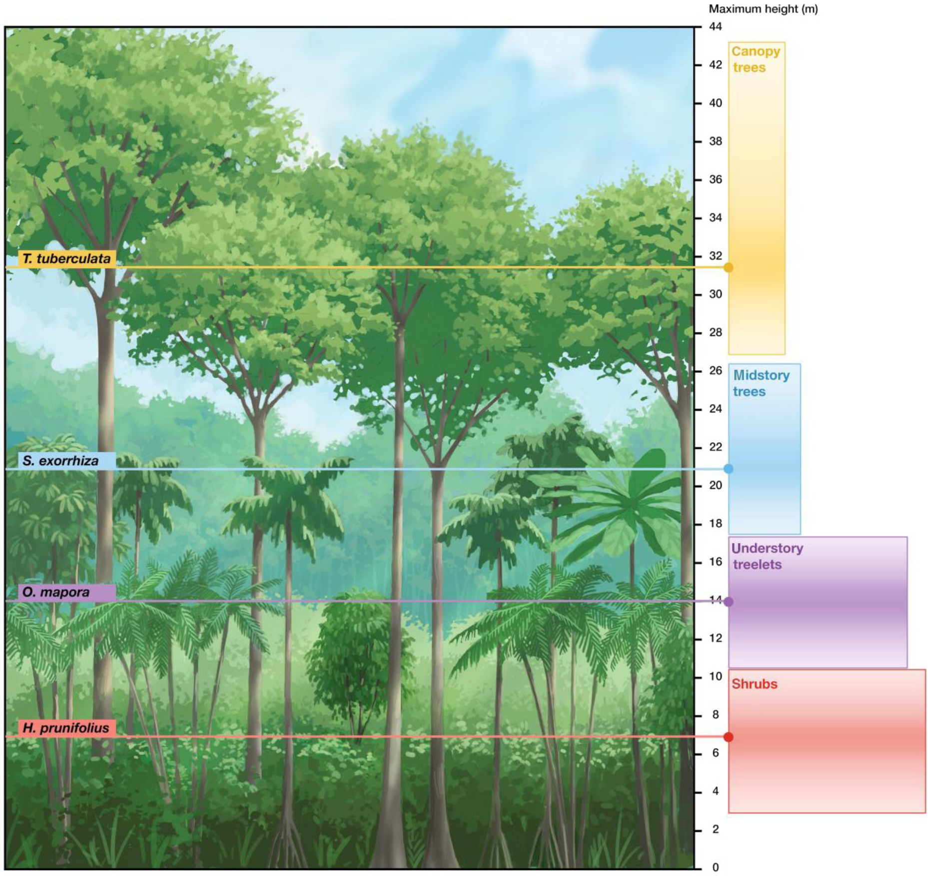


Figure A2. In the 50 ha Barro Colorado Island, Panama forest inventory plot, species showed evidence of niche partitioning into four clusters of maximal height (colored bars). The boundaries of these four groups (shrubs, understory, midstory, and canopy trees) aligned well with previous expectations of four zones of maximal light capture in tropical forests. The horizontal bars indicate the maximum height of the most abundant species within each cluster. (*Adapted from Figure 3 in D'Andrea et al. 2020.)*

For the purposes of this dissertation, this study provided me the opportunity to work with local-scale data for a well-studied and censused community. This is particularly important as local scale processes are more likely to be reflected in the abundances of species than in simple presence/ absence records; for example, abundances along a species' range might reflect the outcome of differing competitive interactions with neighboring species. However, at a regional scale, such as the entire Americas, such fine-scale abundance data is often unavailable. By incorporating abundance data into the analyses of trait distributions, it was possible to avoid detecting trait distributions that simply mirrored the biogeography and phylogenetic history of the regional species pool. These insights will be expanded in **APPENDICES B** and **C**.

In addition to looking for trait clustering among traditional functional traits such as wood density and seed mass, I have also participated in an ongoing investigation of whether traits related to secondary metabolism show evidence of niche partitioning on BCI. Specifically, Brian Sedio and colleagues have previously shown that otherwise ecologically similar and spatially proximate congeners, such as members of *Inga* and *Psychotria*, show widely divergent secondary metabolism (Sedio et al., 2018). Such research might indicate that coexistence in these species-rich tropical genera is being facilitated by niche differences related to specialist enemies (i.e., Janzen-Connell effects, *sensu latu*).

To that end, I have collaborated with Masters student Sarah Orth in Dr. Annette Ostling's lab to investigate spatial and phylogenetic patterning of chemical traits across the BCI plot. In contrast to typical functional traits, such as plant height, which have defined ordinations and dimensionality, chemical traits are multi-dimensional and without a clearly defined ordination. For example, consider whether benzoic acid should be ranked more similar to benzene or to benzyl acetate? Additionally, the chemical repertoire of plants contains many unidentified compounds which can, so far, only be identified to general chemical class, such as terpenes or phenolics. Chemical trait data therefore requires the development of novel classification techniques and ordination schemes (Sedio et al., 2018) to allow comparisons between species. Despite these challenges, preliminary results suggest that species do exhibit statistically significant spatial clustering patterns across the BCI plot (**Figure A3**).

In the context of this dissertation, the work we have done so far on chemical defense traits has given me a greater appreciation for how appropriate choice of both spatial and taxonomic scale is crucial. Janzen-Connell type effects are likely to operate at some of the smallest spatial scales, and thus stand in stark contrast to traits whose variation is often described along the latitudinal extent of entire continents (e.g., Moles et al., 2009). Additionally, only certain genera and study areas seem to exhibit evidence of niche differences based in these chemical defense traits. For example, Sedio et al. (2018) found that genera in the temperate forest SERC plot of Maryland did not exhibit the same degree of chemical disparity observed among tropical genera on BCI. Thus, an appreciation for the ecological impact of such mechanism requires careful accounting for spatial scale and taxonomic focus.

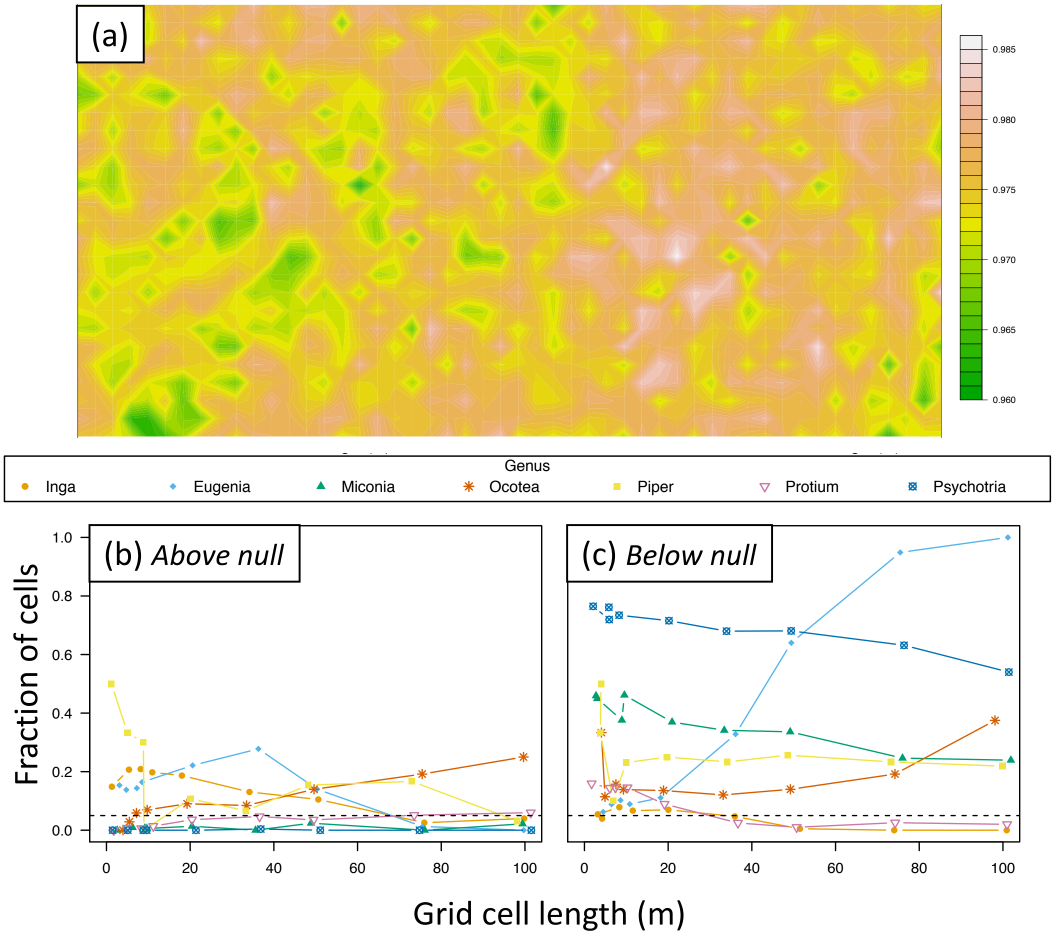


Figure A3. Spatial and phylogenetic patterning of traits related to secondary metabolism for species in the 50 ha Barro Colorado Island (BCI), Panama forest census plot. (*a*) Within 20 m quadrats, spatial variation in the degree of chemical trait similarity was observed across the BCI plot. Colors indicate the average chemical disparity among all species within that grid cell, with higher values indicating more divergent chemical repertoires. (*b-c*) Scatterplots assess statistical significance of average chemical disparity within select genera (colored points) across a range of spatial scales (grid cell length). We report the fraction of grid cells showing statistically significant deviations from a null expectation, separating deviations by those which are more similar than expected by chance (*b*) and those which are more divergent than expected by chance (*c*). Across most spatial scales, there was a tendency for species to more similar in chemical traits than expected by chance, counter to the predictions of a strict Janzen-Connell model. *(Adapted with permission from Sarah Orth's master's thesis.)*

APPENDIX B—How Regional Scale Abundance Data Changes Neutral Expectations of Community Structure.

Throughout this dissertation, I have tended toward a macroecological approach that examines patterns of phylogenetic niche conservatism at the regional scale. Implicitly, this approach is better suited to capturing biogeographical processes (Webb et al., 2002) and could be interpreted as assuming that smaller-scale processes acting within local communities are of little biological relevance (Ricklefs, 2008). In **APPENDIX A**, I attempted to highlight some of the collaborative work I have done examining how trait variation at smaller spatial scales might be linked to important ecological processes occurring within local communities. Here, I further the discussion of how regional and local scale processes might be linked by examining how incorporating regional scale abundance data impacts our expectations of community structure under a neutral model.

Neutral Biodiversity Theory (NBT; Hubbell, 2001) posits that species differences do not contribute to community structure and that assemblages are shaped primarily by 'ecological drift'—i.e., stochastic births and deaths with immigration into the local community from a neutral regional species pool. Although few biologists would claim that all communities are shaped solely by such stochastic processes, it is likely that communities differ in the degree to which neutral- and niche-based processes underly their dynamics (Marx et al., 2017), and that both processes contribute to varying extents across space and time. Further, NBT provides an excellent process-based null model by which to compare deviations in local community structure, and therefore offers a quantitative framework with which to form testable hypotheses about departures from neutral expectations at the community, species, or population level (Hubbell, 2001).

The thoroughly censused and well-studied 50 ha forest plot on Barro Colorado Island (BCI), Panama has been used repeatedly across the decades since its initial 1982 census as a site with which to study NBT predictions. In general, good agreement between theory and data for most census years has emerged (e.g., Volkov et al., 2003; but see McGill et al., 2006, for a contrasting viewpoint). However, it is unclear how neutral expectations for the local community would change if data from the regional pool were included. To begin addressing this, I have used abundance data from plots in and around the Panama Canal region to inform the regional species abundance distribution. This data-informed species set then serves as the regional pool from which neutral communities are drawn (**Figure B1**).

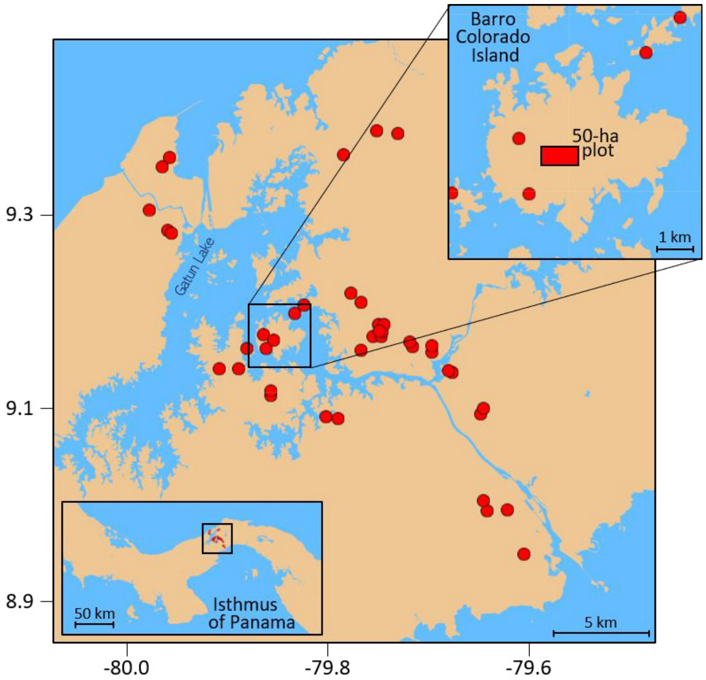


Figure B1. The 50-ha forest census plot (*red rectangle*) on Barro Colorado Island (BCI), Panama in relation to other CTFS forest census plots in the region (*red circles*). *Adapted from Figure 1 in D' Andrea et al. (2020)*.

In general, incorporating regional abundance data severely altered neutral expectations (**Figure B2**). Fixing immigration rates to those traditionally predicted by previous research resulted in overestimates of species richness, whereas allowing the immigration parameter to find a local optimum based on the data resulted in unreasonably low immigrations rates indicative of an unrealistically isolated local community. Examination of species-level departures indicated that in all cases the data-informed model resulted in statistically-significantly high fractions of species departing from neutral expectations.

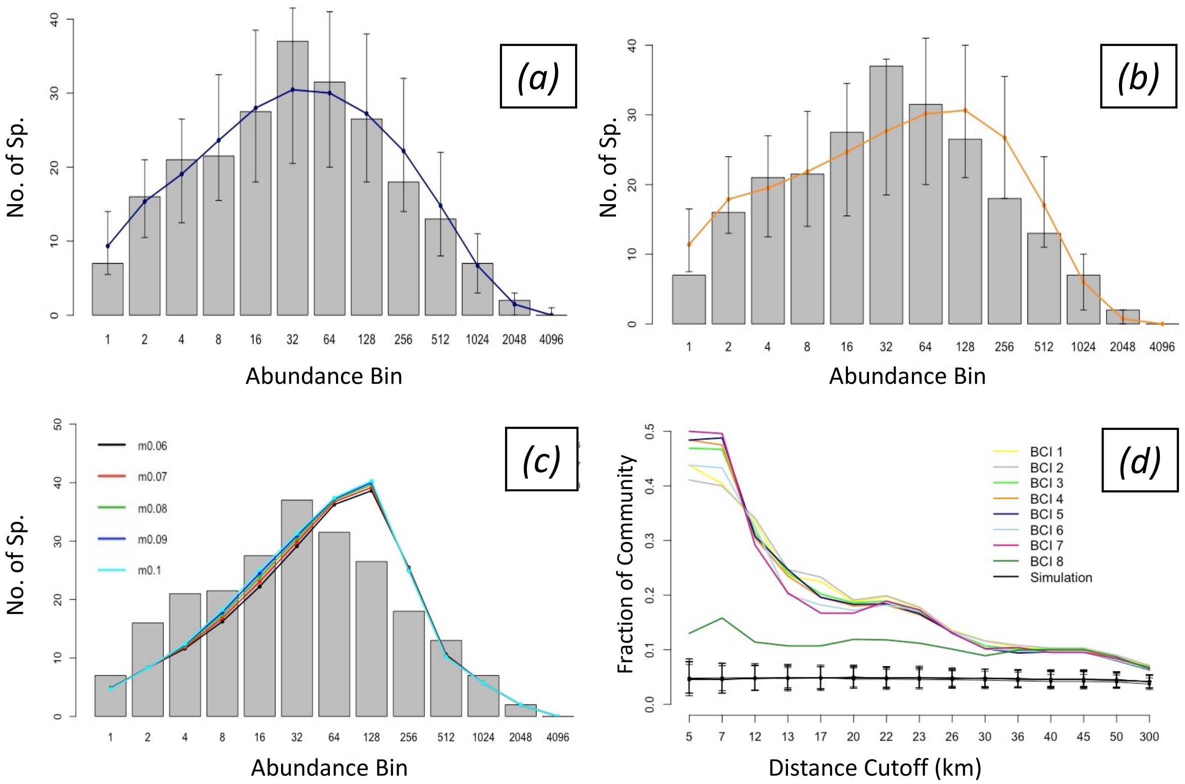


Figure B2. Comparison of the observed species abundance distribution (SAD) on BCI compared to neutral theory predictions. (*a*) SAD fit of standard neutral model (blue lines) to observed community (grey bars), which does not incorporate information on the regional species pool. The fit is considered good and not statistically significantly different from the observed values. (*b*) Comparison of the same, but this time using the data-informed neutral model which incarnates abundance information from the surrounding regional species pool. The fit is poorer, but still falls within stochastic variation predicted by simulations. However, to obtain this fit required tuning immigration parameters to unrealistically low values, indicating an effectively isolated local community. (*c*) Comparison of observed BCI abundance distribution to data-informed model when immigration parameter is fixed to the literature values (between 0.6-1.0). The fit is very poor and falls outside the expected statistical variation (i.e., the model is rejected). (*d*) Plot showing the fraction of species within the local BCI community whose abundance falls outside of neutral predictions based on the data-informed model. In all cases and for all census years, a statistically significant fraction of species (ranging from 10% to 50%) fell outside of neutral predictions. The lower grey lines indicate the expected 5% departure level for each census and regional pool combination. *For panels a-c, only the 5th BCI census is shown, as this census year shows the best agreement between theory and observation in the literature. For panels b-c, only results for the regional species pool comprising plots within 7 km of BCI is shown as this regional pool provided the best fit to the observed local community. Panel d shows species level departures for all combinations of census year and regional pool composition.*

In the context of this dissertation, this work is important because it highlights the link between regional composition and local community structure, even under the simplifying assumption of a neutral model. For me, these results illustrated how an understanding and appreciation of the composition of the regional species pool is crucial for accurately predicting local-scale dynamics. Further, the study highlighted the challenges of appropriately defining what is the species pool for a given local community. For instance, one of the data subsets we examined showed excellent potential as the regional species pool for BCI. This data subset matched parameter expectations from previous work on neutral theory (Hubbell, 2001), and departures from neutral theory were linked to the under-sampling of rare species—a reasonable and ongoing challenge in tropical forest ecology. Does the match between this potential regional pool and neutral theory predictions mean that we have accurately circumscribed the regional pool for BCI? It seems unlikely. Thus, even in the best-case scenario where data and theory match, it remains challenging to feel confident that the extent of the regional species pool can actually be described.

APPENDIX C—Simulation Framework for Accounting for Phylogenetic Signal in Local Competitive Dynamics.

The research summarized in the previous two Appendices illustrated how our understanding of community assembly processes acting within the local community can be enhanced by incorporating abundance data in conjunction with trait data, or by looking at how regional-scale abundances inform our expectations of local community structure. However, neither work has addressed the question of what we expect should in terms of local community phylogenetic structure. To begin addressing this question, I have built a simulation framework (**Figure C1**) that allows users to incorporate phylogenetic signal into ecological simulations of trait-based competitive dynamics. This modelling framework permits an assessment of how phylogenetic signal of traits within the regional species pool might impact the local community phylogenetic structure in the context of trait-mediated competition.

Briefly, regional species pools are generated randomly (or can be supplied by the user), and one or more traits are evolved along the phylogeny of the regional species pool. This approach is valid because, at the regional scale, functional traits often display phylogenetic niche conservatism (**CHAPTER ONE**), even if such phylogenetic signal is not present within a given local community (**APPENDIX A**). The abundances of the regional pool are drawn from a neutral distribution, though other distributions (or data) could be used instead. From the regional species pool, local communities are drawn based on user-defined parameters such as immigration rates or an abiotic (trait-based) filter. Once initialized, these local communities then undergo simulations of ecological dynamics, such as neutral ecological drift or Lotka-Volterra based trait competition. The strength of trait-based competition can be varied, and the traits mediating competition do not need to be the same ones as used for any abiotic filtering. Simulations proceed until a local stochastic equilibrium is reached.

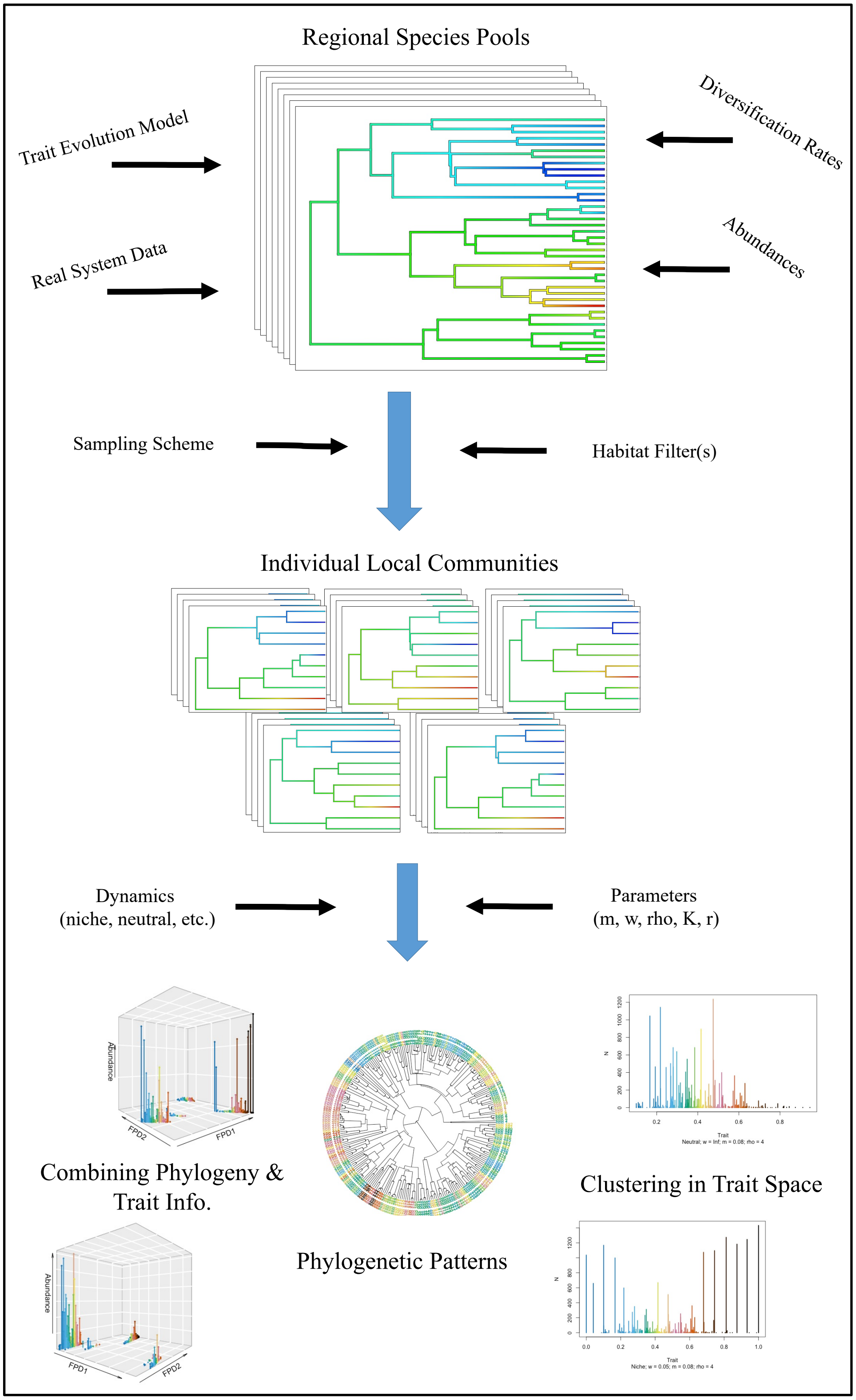


Figure C1. Concept map of the simulation framework to incorporate phylogenetic signal into ecological simulations of trait-based competitive dynamics. Details are outlined in the text of **APPENDIX C**.

The resulting local community composition can be analyzed in terms of the species abundance distribution (SAD), trait-abundance patterning (i.e., trait clustering), and phylogenetic community structure (i.e., strength of phylogenetic signal and clustered vs. over-dispersed). These values are also compared to the same parameters in the regional pool. For instance, this framework allows one to ask whether the phylogenetic signal of a particular trait at the local level is expected to be statistically significantly different from the extent of phylogenetic signal at the regional scale (i.e., how spatial scale and competitive dynamics interact to influence the scale-dependence of phylogenetic niche conservatism). We can also investigate more theoretical questions about how the relative strengths of abiotic filtering and phylogenetic signal interact to shape local-scale community assembly patterns. This simulation framework can also help address the question of what spatial scale of sampling is required to be able to interpret phylogenetic community patterns.

APPENDIX D—Climatically-Similar Non-Alpine Species[§§].

Climatically Similar Non-Alpine (CSNA) species, defined here as non-alpine species whose ranges otherwise overlapped with the climatic conditions of alpine taxa (**Figure D1**) tended to cluster geographically in the southwestern US and Mexico (**Figure D1c**), and therefore also departed from a traditional latitudinal diversity gradient as their greatest richness occurred at ~20° N. Although regions with higher CSNA richness harbored the lowest alpine diversity (**Figure D1b,c**), richness between these groups at these areas was strongly and positively correlated ($R^2=0.68$, $p<0.001$; Figure D1e). CSNA species ranges (**Figure D1d**) contained a significantly greater fraction of frost-exposed mountain foothills ($p<0.01$) and lower fraction of frost-free lowland ($p<0.01$), compared to other non-alpine species. To a lesser extent, they also had a greater proportion of their ranges in the lower montane belt than other non-alpine taxa. At the phylogenetic scale of the American seed plant flora, there was no clear separation between alpine and CSNA species (**Figure D1a**). However, there were substantial differences in the taxonomic composition of these groups at finer phylogenetic scales (**Table D1**), with several orders containing CSNA but not alpine species, and only ~28% overlap in genera between CSNA and alpines ($N_{alpine\_genera}=717$; $N_{CSNA\_genera}=1421$; $N_{shared\_genera}=395$).

Among non-alpine species, the subset whose ranges occupied similar temperature, precipitation, and niche-breadth space as those found in the alpine belt (CSNA species, as defined above) provided a possible first approximation of which non-alpine American seed plant species might otherwise be able to inhabit the alpine belt and help address a key question of whether abiotic factors or historical processes (e.g., phylogenetic and biogeographic history) plays a larger role in alpine community assembly (e.g., Hughes & Eastwood, 2006).

CSNA species were distributed broadly across the Americas but concentrated in the southern US and in Mexico and tended to have ranges with a greater proportion in frost-exposed foothills and lower montane habitat. This might suggest that ecological and physiological factors,

---

[§§] This appendix was adapted with permission of the author(s) from a submission to the journal, *Ecology and Evolution*.

rather than general dispersal limitation, constrains the ability of these species to enter the alpine belt. It is particularly interesting to note that CSNA species had their greatest richness in the same areas where alpine species richness was lowest, namely in and around Mexican mountains. This might suggest a role for ecological factors, such as competitive dynamics, in separating the distributions of these species along elevation, however finer scale studies are needed to specifically address such hypotheses since macro-scale analyses capture mostly biogeographical processes (Webb et al., 2002).

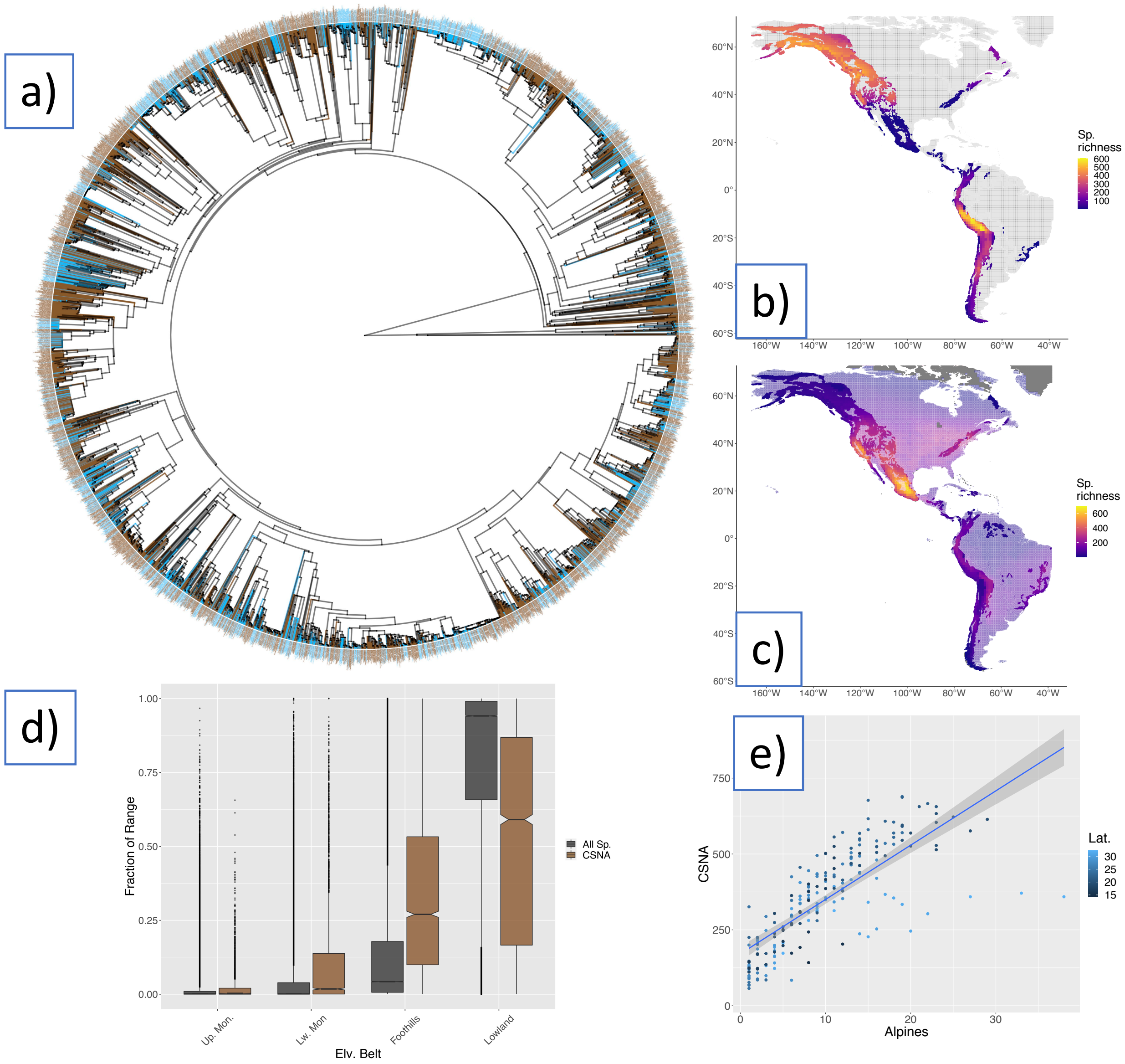


Figure D1. Summary of climatic niches and elevational distributions of Climatically-Similar Non-Alpine (CSNA) species, defined here as non-alpine species whose ranges encompassed the same temperature and precipitation conditions as alpine taxa. (*a*)

Phylogeny of alpine (*blue tips*) and CSNA (*brown tips*) species. At this broad scale, no clear phylogenetic separation of the two groups exists, and both groups are distributed broadly across the seed plant phylogeny. However, at the genus level, substantial taxonomic differences were observed, and these differences persisted at both the familial and ordinal levels (see **Table D1**). (*b-c*) Species richness for alpine taxa (*b*) and CSNA taxa (*c*), showing how both depart from traditional latitudinal diversity gradients, and highlighting how CSNA richness is greatest where alpine richness is lowest. (*d*) Elevational distribution of CSNA (*brown bars*), compared to the elevational distribution of all species in our dataset ("All Sp.", *grey bars*). Compared with the American seed plant species pool overall, CSNA species tended to have ranges that included a higher proportion of mountain foothills (with frost exposure) and, to a lesser extent, lower montane, habitat. Additionally, a lower proportion of their ranges were found in exclusively frost-free or lowland areas. (*Up. Mon: upper montane; Lw. Mon.: lower montane; alpine elevational categories omitted here for clarity.*) (*e*) Scatterplot showing a strong ($R^2=0.68$) and significant ($p<0.001$) positive correlation between species richness at Mexican sites where alpine richness was lowest and CSNA richness was highest, indicating that sites which support higher richness for one group likely do so for the other group was well.

Table D1. Comparison of taxonomic orders among alpines[1], CSNA[2], and non-alpine[3] groups.

Alpines	*CSNA*	*Non-Alpine*
Poales	Arecales	Acorales
Asterales	Chloranthales	Austrobaileyales
Lamiales	Crossosomatales	Berberidopsidales
Caryophyllales	Cycadales	Boraginaceae
Brassicales	Garryales	Buxales
Fabales	Magnoliales	Ceratophyllales
Ericales	Nymphaeales	Dilleniales
Rosales	Pandanales	Huerteales
Saxifragales	Picramniaales	Icacinales
Apiales	Zygophyllales	Metteniusales

[1]Alpines clades here are the top 10 (in order) contributing most to alpine diversity in this study; [2]CSNA are orders containing Climatically-Similar Non-Alpine species but devoid of true alpines here; [3]Non-alpine orders here are those devoid of alpine species across the entire dataset, excluding those already in the CSNA column.

Apart from ecological interactions, physiological limitations could also contribute to why these CSNA species have not entered the alpine belt. Alpine habitats are often at the physiological limits of what plants can tolerate (Körner, 2003). This strong abiotic filter might impose significant challenges and therefore prevent a larger number of lineages from entering the alpine environment. Only ~28% of CSNA genera overlapped with those of actual alpine species, and ten taxonomic orders differed between these groups (**Table D1**). Such disparity in the taxonomic composition of these groups might be consistent with strong abiotic filtering that only approximately ¼ of CSNA genera have been able to overcome at this time. Although the ‘competition-filtering’ dichotomy presents a reasonable framework with which to form testable hypotheses concerning the separation of alpine and CSNA species, these are not the only possible explanations for why certain species have not entered the alpine belt. As noted above for montane communities, soil conditions, and pathogen or dispersal-agent distributions, as well as changes to growth form (Sklenář et al., 2016) are all additional, non-mutually exclusive factors that could differ between these groups and their ranges. Another interesting and untested possibility is that the American alpine belt is at carrying capacity.

APPENDIX REFERENCE LIST

D'Andrea, R., Guittar, J., O'Dwyer, J. P., Figueroa, H., Wright, S. J., Condit, R., & Ostling, A. (2020). Counting niches: Abundance-by-trait patterns reveal niche partitioning in a Neotropical forest. *Ecology*, *101*(6). https://doi.org/10.1002/ecy.3019

Hubbell, S. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography.* (1st ed.). Princeton University Press.

Hughes, C., & Eastwood, R. (2006). Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences*, *103*(27), 10334–10339. https://doi.org/10.1073/pnas.0601928103

Körner, C. (2003). *Alpine plant life: Functional plant ecology of high mountain ecosystems.* (2nd ed.). Springer.

Macarthur, R., & Levins, R. (1967). The Limiting Similarity, Convergence, and Divergence of Coexisting Species. *The American Naturalist*, *101*(921), 377–385. https://doi.org/10.1086/282505

Marx, H. E., Dentant, C., Renaud, J., Delunel, R., Tank, D. C., & Lavergne, S. (2017). Riders in the sky (islands): Using a mega-phylogenetic approach to understand plant species distribution and coexistence at the altitudinal limits of angiosperm plant life. *Journal of Biogeography*, *44*(11), 2618–2630. https://doi.org/10.1111/jbi.13073

McGill, B. J., Maurer, B. A., & Weiser, M. D. (2006). Empirical evaluation of neutral theory. *Ecology*, *87*(6), 1411–1423. https://doi.org/10.1890/0012-9658(2006)87[1411:EEONT]2.0.CO;2

Moles, A. T., Warton, D. I., Warman, L., Swenson, N. G., Laffan, S. W., Zanne, A. E., Pitman, A., Hemmings, F. A., & Leishman, M. R. (2009). Global patterns in plant height. *Journal of Ecology*, *97*(5), 923–932. https://doi.org/10.1111/j.1365-2745.2009.01526.x

Ricklefs, R. E. (2008). Disintegration of the Ecological Community: American Society of Naturalists Sewall Wright Award Winner Address. *The American Naturalist*, *172*(6), 741–750. https://doi.org/10.1086/593002

Scheffer, M., & van Nes, E. H. (2006). Self-organized similarity, the evolutionary emergence of groups of similar species. *Proceedings of the National Academy of Sciences*, *103*(16), 6230–6235. https://doi.org/10.1073/pnas.0508024103

Sedio, B. E., Parker, J. D., McMahon, S. M., & Wright, S. J. (2018). Comparative foliar metabolomics of a tropical and a temperate forest community. *Ecology*, *99*(12), 2647–2653. https://doi.org/10.1002/ecy.2533

Sklenář, P., Kučerová, A., Macková, J., & Romoleroux, K. (2016). Temperature Microclimates of Plants in a Tropical Alpine Environment: How Much does Growth Form Matter? *Arctic, Antarctic, and Alpine Research*, *48*(1), 61–78. https://doi.org/10.1657/AAAR0014-084

Terborgh, J. (1985). The Vertical Component of Plant Species Diversity in Temperate and Tropical Forests. *The American Naturalist*, *126*(6), 760–776. https://doi.org/10.1086/284452

Volkov, I., Banavar, J., Hubbell, S., & Maritan, A. (2003). Neutral theory and relative species abundance in ecology. *Nature*, *424*, 1035–1037.

Webb, C. O., Ackerly, D. D., McPeek, M. A., & Donoghue, M. J. (2002). Phylogenies and Community Ecology. *Annual Review of Ecology and Systematics*, *33*(1), 475–505. https://doi.org/10.1146/annurev.ecolsys.33.010802.150448