Title

Life in the leaves: diversity and distribution of foliar fungal endophytes

Permalink

https://escholarship.org/uc/item/8ft5k6gp

Author

Apigo, Austen

Publication Date

2022

Supplemental Material

https://escholarship.org/uc/item/8ft5k6gp#supplemental

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Santa Barbara

Life in the leaves: diversity and distribution of foliar fungal endophytes

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Ecology, Evolution, and Marine Biology

by

Austen Arthur Lim Apigo

Committee in charge:

Professor Ryoko Oono, Chair

Professor Holly V. Moeller

Professor Todd Oakley

Professor Elizabeth G. Wilbanks

December 2022

The dissertation of Austen Arthur Lim Apigo is approved.

_____

Holly V. Moeller

_____

Todd Oakley

_____

Elizabeth G. Wilbanks

_____

Ryoko Oono, Committee Chair

September 2022

Life in the leaves: diversity and distribution of foliar fungal endophytes

Copyright © 2022

by

Austen Arthur Lim Apigo

ACKNOWLEDGEMENTS

Thank you to my family, partner, friends, and the academic community at UCSB for supporting and cheering me along throughout this journey.

To Sean Burns: We have come quite a way since the broken car battery in a Trader Joe's parking lot. You are the absolute best partner to walk through this life with. Thank you for the way in which you listen and communicate, your ability to make me laugh every day, and encouraging me along each step of this journey. I love you!

To Erolyn Apigo and Alex Apigo: In late October of 1990, you fell in love with a baby on the other side of the world and decided to make him a part of your family. Thank you for all your love and support over the years – I could not have asked for better parents.

To Adam Apigo, Ahren Apigo, Bryle Apigo, Roanne Gulpan, Ana Hess, Angelica Lim-Nguyen, Bryan Lim, Kayleen Aralar, Flormina De la Cruz, Glenn Hess, Shirvada Charlie-Apigo, and Kurumi Apigo: Thank you for being such a supportive family, regardless of distance and time apart. You have cheered me on through every stage of my life and I am so appreciative of each of you.

To Keilani Apigo, Kaleo Apigo, Hayden Walters, Trey Walters, Kaito Apigo, Hiroto Apigo, and Rianne Gulpan: While many of you may not be aware of it, you have brought so much joy to my life and being your uncle is one of the proudest titles I will ever have. I dedicate my dissertation to each of you.

To Shelley Bennett, An Bui, Jake Sarver, Jasmine Childress, Ana Miller-ter Kuile, Kelsey Dowdy, Shannon Haggerty, and Randy Long: Thank you creating so many moments

where we were laughing so hard, I could barely breathe. Your friendship kept me going throughout this entire journey and I am grateful for the time we had and will have together.

To Ryoko Oono: I still remember the first time I was in your office chatting about endophytes! Thank you for your ever-present support and advice throughout this journey. Your mentorship has helped me grow into a better scientist, communicator, teacher, and mentor.

To Holly Moeller, Todd Oakley, and Lizzy Wilbanks: Thank you for being a supportive committee. Holly, thank you for advice over the years, I will look back fondly on teaching ecological modeling together.

To the many funding sources and collaborators that have supported my work over the years: I would like to acknowledge the staff associated with the Bonanza Creek LTER, Denali State Park, Eklunta Lake State Park, Garibaldi Provincial Park, Kalamalka Provincial Park, Quesnel River Research Center, Clatsop State Forest, Tillamook State Forest, HJ Andrews Experimental Forest LTER, Santa Cruz Island UC Natural Reserve, James San Jacinto UC Natural Reserve, Reyes Peak State Park, La Malinche National Park, Parque Ecológico El Tecúan, and the Smithsonian Tropical Institute for facilitating my fieldwork in the most beautiful places. Dr. Rodolfo Salas-Lizana, thank you for your support and advice as a collaborator over the years. Many thanks to Ruben Montes Montiel for facilitating field sampling in Mexico. Many thanks to Dr. Jen Smith for performing the sequencing aspects of this work.

To the many undergraduates who have supported this work, thank you! This work would not have been possible without your hard work and dedication. Many thanks to Helen Chang, Abby Camenisch, Alecsa Burwell, Dulce Simental, Karen Cardenas, Timothy

v

VITA OF AUSTEN ARTHUR LIM APIGO
September 2022

EDUCATION

Bachelor of Science in Ecology, Evolution, and Biodiversity, University of California, Davis, June 2013

Doctor of Philosophy in Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, December 2022 (expected)

PUBLICATIONS

Apigo A and Oono R. 2022. Plant abundance, but not plant evolutionary relationships, shape host specificity in foliar fungal endophytes. Ecosphere 3(1): e03879.

Sarver J, Schultz E, **Apigo A**, Gernandt DS, Salas-Lizana R, Oono R. 2022. Deep sequencing across multiple host species tests pine-endophyte specificity. American Journal of Botany 109(1):83-98.

Miller-ter Kuile A, **Apigo A**, Bui A, DiFiore B, Forbes E, et al. 2022. Predator-prey interactions of terrestrial invertebrates are determined by predator body size and species identity. Ecology 103(5): e3634.

Miller-ter Kuile A, **Apigo A**, Young HS. 2021. Effects of consumer surface sterilization on diet DNA metabarcoding data of terrestrial invertebrates in natural environments and feeding trials. Ecology and Evolution 11(17): 12025–12034.

Oono R, Black D, Sickler B, Strom A, Slessarev E and **Apigo A**. 2020. Species diversity of fungal symbionts in stressed host environments. New Phytologist 228: 210-225.

Apigo A and Oono R. 2018. Dimensions of Host Specificity in Foliar Fungal Endophytes. In: Pirttilä A, Frank A (eds). Endophytes of Forest Trees. Forestry Sciences, vol 86. Springer, Cham.

AWARDS, FELLOWSHIPS, & GRANTS

2022   National Science Foundation Postdoctoral Research Fellowship in Biology
2021   University of California President's Dissertation Year Fellowship
         Academic Senate Grant, UC Santa Barbara
         Mycological Society of America Graduate Fellowship
         Sonoma County Mycological Association Graduate Fellowship

EEMB Departmental Block Grant, UC Santa Barbara

2020   Olivia Long Converse Fellowship, UC Santa Barbara

2019   Charles A. Storke Fellowship, UC Santa Barbara

Ellen Schamberg Burley Award, UC Santa Barbara

2018   Worster Award, UC Santa Barbara

Travel Grant, Graduate Student Association, UC Santa Barbara

2017   Smithsonian Tropical Research Institute Short Term Fellowship

EEMB Departmental Block Grant, UC Santa Barbara

2016   Olivia Long Converse Fellowship, UC Santa Barbara

EEMB Departmental Block Grant, UC Santa Barbara

2015   EEMB Departmental Block Grant, UC Santa Barbara

PROFESSIONAL EMPLOYMENT

2021 – 2022   UC President's Dissertation Year Fellow

2020 – 2021   Oliva Long Converse Fellow

2019, 2020   Teaching Assistant, Ecological Modeling, Winter Quarter 2019, Winter Quarter 2020, UC Santa Barbara

2019   Teaching Assistant, Introductory Biology Lab, Ecology and Evolution, Fall Quarter 2019, UC Santa Barbara

2017 – 2018   Teaching Assistant, Introductory Biology Lab, The Diversity of Life, Spring Quarters 2017, 2018, UC Santa Barbara

2017   Teaching Assistant, Introductory Biology Lab, Biochemistry, Cell Biology and Development, and Genetics, Fall Quarter 2017, UC Santa Barbara

2016 – 2018   Teaching Assistant, Introductory Biology Lab, Ecology, Evolution, Plant and Animal Physiology, Winter Quarters 2016, 2017, 2018, UC Santa Barbara

ABSTRACT

Life in the leaves: diversity and distribution of foliar fungal endophytes

by

Austen Arthur Lim Apigo

Understanding why species vary from place to place is a long-standing question in biogeography and ecology. To answer this, biologists have proposed "rules" to explain patterns of species diversity and distribution in plants and animals. Yet whether microbes play by the same set of rules as macroorganisms requires further investigation. In this dissertation, I tested whether biogeographic and community assembly rules could predict the diversity and distribution of foliar fungal endophytes (fungi that live inside plant leaves) across spatial and host scales.

In Chapter 1, I tested whether endophytes displayed a latitudinal diversity gradient, a biogeographic pattern where species richness increases towards the equator, across a 55° latitudinal gradient from Alaska to Panama. I found that endophyte species richness (alpha diversity) generally increased towards the equator but varied bimodally as a function of latitude. Endophyte richness was greatest in Panama and was also pronounced, to lesser degree, in Canada and Oregon, where plant communities received high annual precipitation, experienced low precipitation seasonality, and had high photosynthetic biomass. On the other hand, endophyte richness was lowest in California, where plant communities received low annual precipitation, experienced high precipitation seasonality, and had low photosynthetic biomass. This suggests endophytes responded to a suite of factors that were not necessarily correlated with latitude and this shaped their bimodal pattern of richness across latitude.

While endophyte alpha diversity increased towards the tropics, differences in endophyte species composition (beta diversity) among host species increased towards temperate regions. Additionally, endophytes in temperate plant communities occupied a lower proportion of their host communities, or had greater host specificity, than endophytes in subtropical and tropical plant communities. Beta diversity and host specificity were correlated and suggest that greater compositional dissimilarity among endophyte communities was driven by endophytes that were host-specific.

I propose that host specificity is a major driver of contrasting patterns in endophyte alpha and beta diversity as a function of latitude. I hypothesize that species-rich host communities are heterogenous landscapes for endophytes and barriers to the evolution of host specificity because specialist endophytes cannot reliably colonize compatible hosts via passive dispersal. Therefore, species-rich host communities in the tropics should select for host generalism while temperate communities with fewer host species should select for host specificity, driving patterns of latitudinal beta diversity. I also hypothesize that because host specialists are adapted to their host's chemistry or physiology, they competitively exclude other endophytes and reduce endophyte richness in hosts. Alternatively, host generalism may come at the cost of competitively ability and facilitates greater coexistence among endophytes. This allows hosts to accumulate endophyte species and shapes patterns of latitudinal alpha diversity.

Findings from Chapter 1 indicate that host communities can determine the community composition of endophytes. I expanded upon this in Chapter 2 by reviewing different ways endophytes can vary in their specificity among host species depending on host sampling, host phylogenetic, and host spatial scale. I then explored the methodological effect of rare

endophytes on metrics of host specificity because their low sequencing abundance is often correlated with high host specificity. I found that removing rare endophytes from the community made observed measures of host specificity indistinguishable from values expected by random community assembly. Thus, their inclusion can be important in detecting non-random community structure. Yet how rare endophytes could be fairly compared to their more abundant counterparts was still a remaining challenge.

I addressed this in Chapter 3 by developing a method that allows one to consider the host specificity of rare endophytes while reducing the bias caused by their low sequencing abundances. I used this method to test whether two community assembly rules for host-associated microbial communities, the common host hypothesis (the effect of host abundance) or phylosymbiosis (the effect of host evolutionary history), could explain variation in endophyte distributions in a single plant community. I found that more abundant plant species harbored endophytes that were more host-specific. These endophytes occupied fewer plant species and were consistently found in the same plant species across the landscape, supportive of my hypothesis in Chapter 1 that low host density can be a barrier to the evolution of host-specific interactions. Host phylogenetic distance was not predictive of host specificity.

In the conclusion of this dissertation, I frame patterns of endophyte host specificity in Chapter 3 in the context of the latitudinal survey in Chapter 1. I discuss why host-specific interactions may explain contrasting patterns of endophyte alpha and beta diversity as a function of latitude and how studies that extrapolate global fungal biodiversity can better incorporate fungal biogeography and ecology into their estimates.

TABLE OF CONTENTS

LIST OF FIGURES

I. Introduction

Understanding why species vary from place to place is a longstanding question in biogeography and ecology. To address this, biologists have proposed "rules" that attempt to predict patterns of species diversity and distribution. For example, the species-area relationship is one of the clearest rules that predicts larger areas will always contain more species than smaller areas (Arrhenius 1921). Such rules were initially developed from observations of plants and animals, yet whether microbes play by the same rules as macroorganisms is still an open question. Advances in sequencing technologies have allowed the scientific community to characterize any microbial community directly from the environment and test the applicability of these rules at an unprecedented scale. As a result, some rules such as the species-area relationship have acquired broad support across microbial kingdoms (100% of 29 studies support; Dickey et al. 2021). Yet for other rules, microbes display biogeographic patterns that rarely resemble those of plants and animals and suggest that some rules are not generalizable across the tree of life.

One of the most widely recognized rules that predicts how numbers of species are distributed on the planet is the latitudinal diversity gradient (LDG), the biogeographic pattern where species richness increases towards the equator (Hillebrand 2004). Current evidence suggests that microbial communities rarely follow this rule (32% of 78 studies confirm LDG; Dickey et al. 2021). Microbial communities may not conform to LDG because they respond to environmental conditions that are not correlated with latitude. For example, in a meta-analysis of 325 soil microbial communities, there was no relationship between air temperature, which is correlated with latitude, and soil bacterial and fungal diversity. On the other hand, there was a weakly positive relationship between soil pH,

which is not correlated with latitude (Slessarev et al. 2016), and belowground diversity (Hendershot et al. 2017), consistent with other studies (Fierer and Jackson 2006, Rousk et al. 2010). However, a simpler explanation for the lack of adherence to LDG among microbes is that latitudinal patterns for microbes are not as well documented compared to plants and animals. For example, Hillebrand (2004) synthesized 581 latitudinal gradients of plants and animals and found support for LDG across hemispheres, trophic levels, and dispersal modes. Therefore, more studies are needed to understand whether most microbial groups are truly exceptions to LDG or if we simply have not surveyed enough of those that do follow LDG.

Studies of microbial LDG tend to be skewed towards free-living microbes that live in belowground or aquatic environments (83% of 78 studies for LDG; Dickey et al. 2021). Additionally, studies of LDG for microbes that inhabit living leaves, or the phyllosphere, are particularly lacking (only two studies; Arnold and Lutzoni 2007, Wang et al. 2022). This is a significant knowledge gap considering the phyllosphere is an expansive environment for microbes and comprises about 60% of living biomass on Earth (Koskella 2020). The microbes that have evolved to live in the phyllosphere can influence the growth and survival of their hosts and regulate biogeochemical cycles (Sivakumar et al. 2020), yet we still have a limited understanding of what drives their diversity and distribution across broad spatial scales, particularly latitudinal gradients. Consequently, this also limits our ability to estimate phyllosphere-associated microbial diversity at a global scale and forecast how these communities will respond to global change.

In this dissertation, I explored the biogeography of foliar fungal endophytes, a major component of the microbial community in the phyllosphere. Foliar fungal endophytes (Class 3 endophytes sensu Rodriguez et al. 2009), hereafter "endophytes", are a species-rich and

phylogenetically diverse guild of microfungi that live asymptomatically within leaf tissues for some portion of their lifestyle but can also function as mutualists, pathogens, and latent decomposers (e.g., Carroll 1988, Arnold et al. 2003, Osono 2006, Busby et al. 2016). Endophytes have been isolated from a multitude of plant species spanning all major lineages of land plants and have diversified across every terrestrial ecosystem that supports plant life (Arnold 2007, Harrison et al. 2020). Thus, their spatial prevalence and associations across the plant kingdom lends itself as a useful system to test how geography, the environment, and hosts influence the assembly and structure of host-associated microbial communities.

One of the most comprehensive studies of endophyte diversity found support for LDG among endophytes in North America by intensively culturing thousands of endophytes from 21 plant species (Arnold and Lutzoni 2007). Endophytes may display an LDG because they respond to environmental variation or biotic interactions that trend with latitude. Ecological theory defines these sources of variation that organisms encounter as "filters" that permit or exclude the establishment of species and shape how biological communities assemble (Vellend 2010, Kraft et al. 2015). For example, differences in the species composition of endophyte communities often reflect geographic distance (Oono et al. 2017, Whitaker et al. 2018, Bowman and Arnold 2021), environmental dissimilarity (Zimmerman and Vitousek 2012, Darcy et al. 2020, Oita et al. 2021b), or host identity (Zhang and Yao 2015, Whitaker et al. 2020, U'Ren et al. 2019). Yet the effect of these factors on endophyte richness across latitudinal gradients is still unclear. I addressed this in Chapter 1 by performing the most comprehensive culture-free survey for endophytes to date across a 55° latitudinal gradient in North America by sampling 20 sites comprised of 1,657 plant hosts spanning 443 host genera.

For endophytes, hosts impose a significant biotic filter that can shape their distributions. All host-associated microbes can vary in their distribution among hosts, or their degree of host specificity, in ways that depend on host sampling, host spatial, and host phylogenetic scale. For example, an endophyte that is associated with three host species in a plant community would generally be categorized as more host-specific than another endophyte that associated with ten host species. Alternatively, such a categorization would likely be reversed if one found out that the former endophyte associated with three out of four host species in a plant community (75% of available host species) and the latter endophyte associated with 10 out of 100 host species in a different plant community (10% of available host species).

Host specificity can also reflect the phylogenetic breadth of host associations. An endophyte that associates with ten plant species of the same genus is more phylogenetically constrained to hosts than an endophyte that associates with ten plant species that are each of a different genus. Phylogenetically constrained associations by endophytes to hosts can lead to differences in the species composition of endophyte communities that are positively correlated with host phylogenetic distance, a relationship termed phylosymbiosis (Brooks et al. 2016, Lim and Bordenstein 2020; e.g., Liu et al. 2019, Sarver et al. 2022). Host-specific interactions can also vary with respect to host ranges such that the spatial distribution of plants dictates the spatial distribution of endophytes. For example, Vincent et al. (2016) showed that in the absence of strong environmental gradients, plants hosted highly similar endophyte communities across hundreds of kilometers in a New Guinean tropical forest, perhaps because geographically widespread host species reduced dispersal limitation for endophytes.

In Chapter 2, I reviewed metrics that quantify different ways endophytes can vary in their host specificity and their relevance to studies that explored how endophyte distributions varied across host species and space. I then explored the effect of rare endophytes on these metrics because they are highly common in high-throughput sequencing datasets and will tend to be categorized as host specific based on their low sequencing abundances. I compared observed metrics of host specificity to those from randomized communities to understand whether rare endophytes differentiated observed values of host specificity from those expected by random chance.

In Chapter 3, I expanded upon these host specificity metrics by developing a method that reduces the bias of low sequencing abundances on metrics of host specificity. This method standardizes observed measures of host specificity to their expected values under random community assembly. I argue this allows one to compare endophytes of different read abundances more fairly than not accounting for read abundance. I used this method to test two community assembly rules that predict the distribution of endophyte species: the common host hypothesis, which predicts low host density is a barrier to host specificity, and the phylosymbiosis hypothesis, which predicts that host evolutionary relatedness determines host specificity. Although host associations have been shown to structure endophyte communities in a variety of contexts (Vincent et al. 2016, Li et al. 2019, U'Ren et al. 2019, Whitaker et al. 2020), the effect of host abundance or host evolutionary relatedness on host specificity had yet to be tested with a study that considered all potential host species in a plant community.

In this dissertation, I combined ecological theory and modern molecular techniques to understand whether biogeographic and community assembly rules are useful guides to

understanding the diversity and distribution of foliar fungal endophytes. I used high-throughput sequencing to survey endophyte communities at broad spatial and host phylogenetic scales. In Chapter 1, I explored how endophyte community structure (alpha diversity, beta diversity, and host specificity) varied across a latitudinal gradient. In Chapter 2, I reviewed different ways endophytes can display host specificity and examined methodological considerations related to the quantification of host specificity. In Chapter 3, I tested two community assembly rules for endophytes. In the conclusion, I explained how patterns of host specificity from Chapter 3 can help us understand patterns of endophyte community structure in Chapter 1.

II. Chapter 1: Contrasting patterns of species richness and host specificity in foliar fungal endophytes across a North American latitudinal gradient

A. Abstract

Microorganisms have lived in symbiosis with plants and animals for millions of years. Despite the prevalence and persistence of these relationships over evolutionary time, our understanding of what drives host-associated microbial diversity at broad spatial scales is still limited. Here, we explored how the species richness and composition of foliar fungal endophytes, a diverse guild of fungi that live within photosynthetic tissues, varied across a broad latitudinal gradient (9°N-64°N) in North America that spanned 20 sites and 443 plant genera (n = 1,657 host individuals). We asked whether the species richness (alpha diversity) of endophyte communities increased towards the equator in accordance with the latitudinal diversity gradient (LDG). We also explored how endophytes varied in species composition across latitude and among host species (beta diversity and host specificity). We correlated factors related to climate, the host community, and geography to endophyte species richness, endophyte beta diversity, and endophyte host specificity to understand the extent to which these factors could explain variation in endophyte community structure. Endophyte richness varied bimodally as a function of latitude being greatest in the tropics, consistent with LDG, and was also pronounced, to a lesser degree, at mid-latitudes in temperate conifer forests. Temperature, precipitation, and the leaf environment (productivity, growing degree days, enhanced vegetation index, evapotranspiration) explained the presence of these two peaks in endophyte richness. Across latitudes, climatic dissimilarity and geographic distance were the

most influential factors that differentiated the species composition of endophyte communities. Within sites, differences in endophyte species composition among host species increased towards temperate regions. Endophytes in temperate plant communities occupied a lower proportion of their host communities, or had greater host specificity, than endophytes in sub-tropical and tropical plant communities, such that the latitudinal gradient in endophyte alpha diversity was inversely related to those of beta diversity and host specificity. Beta diversity was correlated with host specificity and suggests that generalist endophytes in the tropics reduced compositional differences among endophyte communities, while specialist endophytes in temperate plant communities differentiated endophyte communities. We hypothesize that species-rich host communities in the tropics select for host generalism because heterogenous host landscapes are barriers to host specific interactions. On the other hand, less diverse host communities in temperate regions select for host specificity because passively dispersed endophytes can reliably colonize the same host species across the landscape. We predict that because host specialists are adapted to their host's leaf chemistry or physiology, they competitively exclude other endophytes and decrease endophyte richness in temperate regions. Alternatively, host generalism may come at the cost of competitively ability, facilitating greater coexistence among endophytes and this allows hosts to accumulate endophyte species in the tropics. While the alpha diversity of endophytes may respond to climate and the leaf environment, we propose that the beta diversity of endophytes reflects host specificity and plays a major role in driving contrasting patterns of endophyte community structure as a function of latitude.

B. Introduction

Understanding why some places support more species than others is an important and puzzling question in ecology. At a global scale, the latitudinal diversity gradient (LDG) describes a biogeographic pattern where species richness increases towards the equator (Jablonski et al. 2006). This general rule in macroecology has been well-studied and supported across many plant and animal groups (Rohde 1992, Gaston 2000, Hillebrand 2004, Kreft and Jetz 2007, but see Kindlmann et al. 2007). However, microbial communities exhibit highly variable diversity patterns with respect to latitude (only 32% of studies confirm LDG; Dickey et al. 2021) and its abiotic covariates (Fierer and Jackson 2006, Hendershot et al. 2017), calling into question whether LDG is a useful paradigm for microorganisms. For example, communities of soil (Wang et al. 2019, Větrovský et al. 2019; but see Tedersoo et al. 2014,), ectomycorrhizal (Tedersoo et al. 2012, Shi et al. 2013, Liu et al. 2020), and leaf litter fungi (Jabiol et al. 2013, Duarte et al. 2016, Seena et al. 2019) are most species-rich in temperate or boreal environments. These contradictions to LDG have advanced our understanding of fungal biogeography (Peay et al. 2016) and guide efforts to forecast how fungal communities, which regulate nutrient cycling and plant productivity globally, will respond to climate change (Mucha et al. 2018, Větrovský et al. 2019, Morera et al. 2022). Thus, while significant progress has been made related to LDG and its implications for soil-dwelling microbial communities, other globally distributed and functionally diverse microbial guilds, especially those that live cryptically within plants, have been less explored and warrant further attention.

Foliar fungal endophytes, hereafter "endophytes", are a guild of fungi that live asymptomatically within photosynthetic plant tissues for some portion of their lifecycle but

can also function as mutualists, pathogens, or latent decomposers (Arnold 2007, Rodriguez et al. 2009, Wolfe and Ballhorn 2020). Their cryptic lifestyle and prevalence across all major lineages of land plants have led to suggestions that endophytes are a major source of undiscovered fungal diversity (Dreyfuss and Chapela 1994, Saikkonen et al. 1998, Fröhlich and Hyde 1999, Hawksworth 2001), spurring research to identify where endophytes are most diverse. For example, Arnold and Lutzoni (2007) intensively cultured endophytes from 21 species of angiosperms, conifers, and ferns across North America and found that endophyte diversity increased towards the tropics, consistent with LDG. Since Arnold and Lutzoni's foundational study, the advent of high-throughput sequencing has accelerated the rate and resolution at which endophyte communities have been characterized (e.g., Oita et al. 2021a), allowing studies to consider the richness of fungi that are rare, unknown, or unculturable. Yet no study to date has leveraged high-throughput methods to conduct a more comprehensive test of LDG for endophytes, even though much of the plant kingdom (70% of families) and the tropics remain unexplored for them (Harrison and Griffin 2020). Here, we used culture-free techniques to understand the factors that govern endophyte diversity and distribution across a latitudinal gradient in North America and revisited the question, "does endophyte species richness increase towards the tropics?"

Whether climate, a major axis of variation along latitudinal gradients, can explain latitudinal patterns in endophyte richness is still an open question. For soil-dwelling fungi, temperature and precipitation directly affect the soil environment through changes in pH or plant productivity and have been shown to predict the richness of these guilds across latitudinal gradients (Tedersoo et al. 2014, Zhou et al. 2016, Liu et al. 2020). For endophytes, studies that test the effect of climate on endophyte richness are difficult to

generalize. Studies have found that endophyte richness was positively correlated with annual precipitation (Lau et al. 2013, U'Ren et al. 2019), whereas other studies have found no relationship (Arnold and Lutzoni 2007, U'Ren et al. 2012, Bowman and Arnold 2021). In addition, patterns of endophyte richness across elevational gradients have offered limited insight on the effects of temperature and precipitation on endophyte richness. For example, Zimmerman and Vitousek (2012) and Cobain et al. (2019) found limited or no effect of elevation, respectively, on endophyte richness within *Metrosideros polymoropha* on Hawai'i. Studies of temperate trees found that endophyte richness was greater at higher elevations associated with increased precipitation and decreased temperature (Yang et al. 2016, Bowman and Arnold 2021), whereas another study in temperate grasses found no relationship (Kivlin et al. 2022). This mixture of results could reflect differences in host species, environment, and the range of each climate gradient. Alternatively, fungi that live inside plants may be more buffered from the direct effects of climate than soil fungi and instead respond more strongly to the indirect effects of climate via changes in plant productivity or the plant community.

Community assembly theory predicts that the spatial distribution of endophyte species should be determined by their compatibility to the environment and host species as well as dispersal limitation (Vellend 2010, Kraft et al. 2015, Peay et al. 2016). Indeed, studies using culture-independent methods have repeatedly shown that climate (Zimmerman and Vitousek 2012, Darcy et al. 2020, Oita et al. 2021b), hosts (Zhang and Yao 2015, Liu et al. 2019, U'Ren et al. 2019), geography (Oono et al. 2017, Whitaker et al. 2018, Bowman and Arnold 2021), or combinations of these factors (Hoffman and Arnold 2008, Christian et al. 2016, Barge et al. 2019, Whitaker et al. 2020) dictate the species composition, or beta diversity, of

endophyte communities. For example, studies within single host species have found that endophyte communities became increasingly different with increasing geographic distance or environmental dissimilarity (Oono et al. 2017, Whitaker et al. 2018, Bowman and Arnold 2021), consistent with patterns of dispersal or environmental limitation in other fungal guilds (Peay et al. 2012, Meiser et al. 2014, Talbot et al. 2014). On the other hand, studies at broader spatial and host phylogenetic scales have found that host identity or host phylogenetic distance explained more variation in endophyte composition than spatial distance or climate (Vincent et al. 2016, U'Ren et al. 2019, Darcy et al. 2020). Yet the extent to which climate, hosts, and geography explain differences in the composition of endophyte communities across latitudinal gradients is still unclear.

Biotic filtering by hosts may be of great relevance for fungi that live within plants. A recurring hypothesis in the literature predicts that diverse host communities reduce the likelihood that specialist endophytes encounter suitable hosts through horizontal transmission and thus host generalism is a dominant strategy in the tropics (May 1991, Kindlmann et al. 2007, Griffin et al. 2019, Oita et al. 2021b). In support of this hypothesis, endophytes in the tropics often have broad host or geographic ranges (Arnold and Lutzoni 2007, Higgins et al. 2011, 2014, Vincent et al. 2016, Suryanarayanan et al. 2018, Vaz et al. 2018). Studies have also found that endophyte communities are highly host-specific, or restricted to certain host plants or taxonomic groups, in arctic or boreal environments (Zhang and Yao 2015, U'Ren et al. 2019). Broad or narrow host associations have direct effects on endophyte beta diversity. For example, greater numbers of generalist endophytes in a plant community are expected to drive endophyte communities towards a common composition. However, studies have yet to measure host specificity considering all potential

host species in a given location (but see Apigo and Oono 2022) to systematically test how host specificity varies as a function of latitude. Here, we broadly sampled host communities to examine the relationships between host community structure, endophyte host specificity, and endophyte beta diversity.

In this study, we sampled plant communities across a 55º latitudinal gradient in North America to address the following questions: (1) How does endophyte community structure vary across a latitudinal gradient? (2) To what extent does climate, the host community, or geography explain these biogeographic patterns? Specifically, we sampled all co-occurring host species within five 50 m^2 quadrats in each of 20 sites spanning Panama to Alaska (9º-64ºN) and characterized the endophyte community with high-throughput sequencing. Within each quadrat, we measured endophyte richness in each host species (alpha diversity), differences in the species composition of endophyte communities among host species (beta diversity), and the breadth of host associations by endophytes (host specificity). We then compared how each measure of endophyte community structure varied with respect to latitude, climate, or host community structure to understand the relative importance of these factors to endophyte biogeography.

C. Methods

Site Characteristics and Sampling Design

Plants were collected from 20 sites that ranged in latitude from 9ºN to 64ºN, mean annual precipitation (MAP) from 422.6 kg m^{-2} year^{-1} to 3,305.2 kg m^{-2} year^{-1}, and mean annual temperature (MAT) -1.55ºC to 26.75ºC during 2015 to 2017 (Figure 1; Table 1; Appendix 2: Figure S1). At most sites (n = 15), conifers (e.g., *Abies*, *Calocedrus*, *Picea*, *Pinus, Pseudotsuga, Tsuga* spp.) composed the entirety of the forest canopy. Three sites in

13

California and Mexico were mixed stands of conifers (e.g., *Juniperus, Pinus* spp.) along with angiosperm trees and shrubs. Two sites were in lowland tropical rainforest in Panama and the forest canopy was composed of various angiosperms (e.g., *Fridericia, Machaerium, Psychotria, Piper* spp.). Within each site, we haphazardly identified a general area that included all dominant tree species and sampled from five 50 m^2 quadrats using a random number table with directional headings (1-360º) and number of steps (1-100) to the first corner of a given quadrat. Within each quadrat, we haphazardly sampled ten leaves with a healthy appearance from one individual of every plant species. For evergreen plants (e.g., conifers), we sampled leaves less than one-year-old to collect the cohort of fungi representative of that year's growing season. Plants were kept on ice during transport for 1-5 days until they were surface-sterilized and frozen at -80°C at the University of California, Santa Barbara. Findings from Oita et al. (2021a) showed that storage period (1-10 days) at 4°C was not a significant predictor of variation in endophyte richness or community composition within a fern, conifer, and angiosperm host assessed with a DNA metabarcoding approach. Plants collected from the two Panamanian sites were surface-sterilized and frozen the day after sampling at the Smithsonian Tropical Research Institute (Gamboa, Panama).
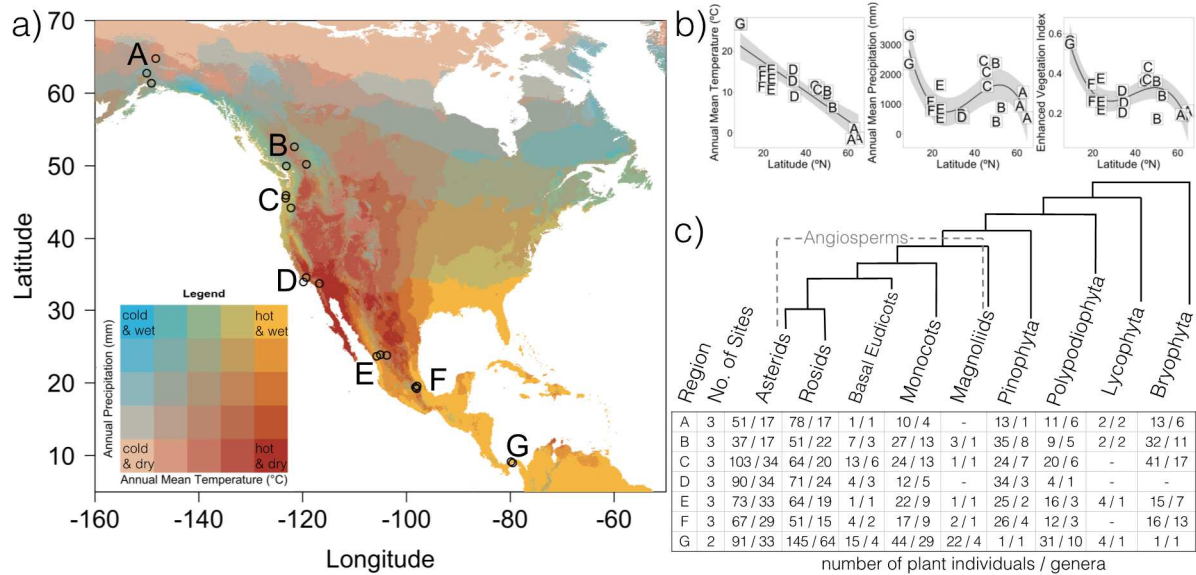
Figure 1. Endophyte communities were sampled across a 55° latitudinal gradient in North America. (a) Endophytes were sampled within five 50 m² quadrats (n = 100) from each site (n = 20; open circles) across seven sampling regions (A = Alaska, USA; B = British Columbia, Canada; C = Oregon, USA; D = California, USA; E = Durango, Mexico; F = Tlaxcala, Mexico; G = Panama, Panama). The map is overlain with mean annual temperature and precipitation. (b) Variation in temperature, precipitation, or Enhanced Vegetation Index (EVI) with respect to latitude with letters referring to sampling regions. Temperature was significantly correlated to precipitation (Pearson's r: 0.45; p = 0.049) and EVI (r: 0.68; p < 0.001). Precipitation was correlated to EVI (r: 0.84; p < 0.001). Shaded regions refer to 95% confidence intervals. (c) Table with the number of sites, host individuals (n = 1,657), and host genera (n = 443) per sampling region in the final dataset. The cladogram of major host groups is based on systematics from Sessa et al. (2014) and Li et al. (2021).

Molecular and Bioinformatic Methods

Plant leaves were surface-sterilized by sequential immersion in 10% commercial bleach for 2 minutes, UV-treated deionized water for 30 seconds, and 70% ethanol for 2 minutes. Leaves were air-dried in a sterile laminar flow hood, frozen at -80°C, and then homogenized by mortar and pestle with liquid nitrogen. DNA from 80 mg of plant leaf tissue was extracted with a modified 2% CTAB method (Doyle and Doyle 1987). The internal transcribed spacer 1 (ITS1) region was amplified from 10 ng of genomic DNA with fungal-specific ITS1F-KYO1 and ITS2-KYO1 (Toju et al. 2012) primers modified with Illumina

overhang adapters. Each endophyte community was sequenced twice on the Illumina MiSeq platform (n = 14 sequencing runs) with the same overhang adapter combination to measure variation in sequencing depths (Song et al. 2018) between replicate endophyte communities. Each run was sequenced with 250 paired-end reads and a 15% spike-in of PhiX at California NanoSystems Institute at the University of California, Santa Barbara. Details of PCR conditions and library preparation can be found in Appendix 1.

We included PCR4-TOPO TA vectors (Invitrogen, Carlsbad, CA, USA) containing the ITS1 region from Ascomycota and Basidiomycota endophytes (see Appendix 1; Table 2) as positive controls on each sequencing run to identify sequencing reads that were not barcoded with their expected i5 and i7 primer combination due to Illumina 'index hopping' (Huson et al. 2007, van der Valk et al. 2020). Sequencing methods and results of positive controls can be found in Appendix 1 and Table 2.

Sequenced reads were trimmed of primers with cutadapt v3.5 (Martin 2011) and then merged, filtered, and clustered into 97% Operational Taxonomic Units (OTUs) with de novo chimera identification using USEARCH v11.0.667 (Edgar and Flyvbjerg 2015) and VSEARCH v2.21.1 (Rognes et al. 2016). OTUs were queried against the GenBank database (Altschul et al. 1990) with BLAST+ v2.12.0 (Camacho et al. 2009). USEARCH, VSEARCH, and BLAST+ analyses were performed on the Knot computer cluster at the University of California, Santa Barbara. OTUs not identified within the kingdom Fungi were removed with MEGAN v6.22.2 (Huson et al. 2007). Details of bioinformatic analysis can be found in Appendix 1. All further analyses were performed in R v4.1.3 (R Core Team 2022).

Endophyte communities (n = 1,898 pairs of sequencing libraries) that were amplified from the same genomic DNA, but sequenced on different MiSeq runs, were combined by

16

summing read abundances. OTUs that did not occur in each replicate endophyte community were removed. Sample completeness, an estimate of how well an endophyte community was characterized by sequencing, was assessed with the "iNEXT" function in the *iNEXT* package (Hsieh et al. 2016). Endophyte communities were rarefied for alpha diversity and network analysis to a sequencing depth of 3,500 reads with the "rrarefy" function in *vegan*. This threshold was chosen based on curves that model sample completeness as a function of sequencing depth (Appendix 2: Figure S2). Singleton OTUs were removed after rarefying endophyte communities. For beta diversity, endophyte communities were normalized using cumulative sum scaling with the "cumNormStatFast", "cumNorm", and "normFactors" functions in the *metagenomeSeq* package (Paulson et al. 2013).

Plant Host Identification and Diversity

To identify plant hosts, the ribulose-1,5-bisphosphate carboxylase oxygenase (rbcL; forward: rbcLa-F, reverse: rbcLajf634-R; Kress and Erickson 2007, Fazekas et al. 2008) or internal transcribed spacer 2 (ITS2; forward: ITS-u3; reverse: ITS-u4; Cheng et al. 2016) regions were sequenced with Sanger or Illumina MiSeq chemistry, respectively. Top BLAST hits were cross-referenced with regional plant species lists and iNaturalist (inaturalist.org) observations to identify plants to the genus level.

Taxonomic ranks for each plant were standardized to the Integrated Taxonomic Information System (itis.gov) using the "get_tsn" function in the package *taxize* (Chamberlain and Szöcs 2013). Plants with multiple or unaccepted taxonomic ranks (n = 170) were manually standardized to NCBI taxonomy (ncbi.nlm.nih.gov/taxonomy). Plants were grouped into "major groups" based on angiosperm systematics from (Sessa et al. 2014,

Li et al. 2021) as asterids or superasterids (149 genera; n = 560 individuals), rosids or superrosids (134 genera; n = 552 individuals), basal eudicots (14 genera; n = 48 individuals), monocots (67 genera; n = 165 individuals), and Magnoliids (6 genera; n = 29 individuals) along with Pinophyta (10 genera; n = 165 individuals), Polypodiophyta (22 genera; n = 105 individuals), Lycophyta (4 genera; n = 14 individuals) and Bryophyta (44 genera; n = 129 individuals) for a total of 1,767 plant host identifications spanning 450 plant genera (Figure 1).

An ultrametric phylogenetic tree was pruned from a phylogeny composed of all extant vascular plant families in North America with the "phylo.maker" function in the *V.PhyloMaker* package (Jin and Qian 2019). Bryophytes were excluded from any analysis involving plant phylogenetic distance because *V.Phylomaker* only includes vascular plant families. Plant diversity was estimated as the absolute number or the mean pairwise phylogenetic distance, using the "mpd" function in the *picante* package (Kembel et al. 2010), of all unique plant genera at each site.

Climate and Vegetation Data

Climatologies at high resolution for the Earth's land surface areas (CHELSA v2.1; chelsa-climate.org) were extracted from the Geophysical Fluid Dynamics Laboratory Earth System Model 4 (Shared Socioeconomic Pathway 1; Representative Concentration Pathway 2.6) projections for 2011-2040 (Karger et al. 2017, 2020). We chose to use CHELSA v2.1 over WorldClim v2 (Fick and Hijmans 2017) because WorldClim reports past climate averages (1970-2000) outside of our sampling period and may underestimate precipitation in areas where we sampled (e.g., western Canada and southern Alaska; Beck et al. 2020). We

considered 25 variables from CHELSA that represent averages or variations in precipitation, temperature, or growing season for our analyses.

Factors relevant to the leaf environment, which included enhanced vegetation index (EVI), evapotranspiration, and net primary productivity (NPP), were measured by the Moderate Resolution Imaging Spectroradiometer (MODIS; modis.gsfc.nasa.gov; Table 1) and extracted using AρρEEARS v6.1 (19ppears.earthdatacloud.nasa.gov) for the years 2015 to 2017. EVI (MOD13A3.061) was measured monthly at a 1-kilometer resolution while total evapotranspiration (MOD16A3GF.061) and NPP (MOD17A3HGF.061) were measured at a 500-meter resolution on a yearly basis.

Endophyte Alpha Diversity, Beta diversity, and Network Metrics

Diversity and network metrics were quantified for each endophyte community that originated from a host individual's sampled leaves. While leaf masses did differ among plant species, endophyte communities were amplified from a standardized amount (10 ng) of DNA. Endophyte alpha diversity was measured as the richness of 97% endophyte OTUs with the "specnumber" function in the *vegan* package (Oksanen et al. 2020). Endophyte beta diversity was measured as the dispersion of endophyte communities grouped by quadrat in principal coordinates (PcoA) space with the "betadisper" function in the *vegan* package. We tested the effect of sampling region, site, quadrat, or host major group on endophyte community composition using permutational analysis of variance (PERMANOVA) with the "adonis2" function in the *vegan* package.

Specialization was measured for the host and endophyte community within each quadrat as network specialization (endophyte specialization to hosts and vice versa; H2) with the

"H2fun" function and ranges from 0 (most generalized) to 1 (most specialized). Host specificity by endophytes was measured as the proportion of hosts occupied by each 97% OTU in a given quadrat and was negated such that greater values correspond to greater host specificity. Endophyte associations with respect to host major group or sampling region were standardized by sampling effort (e.g., number of reads across asterids / number of asterid individuals sampled) and visualized with the "plotweb" function in the *bipartite* package (Dormann et al. 2009). Measures of endophyte community structure were regressed as a function of latitude. These analyses were repeated with the number or the phylogenetic diversity of host genera sampled per site as covariates to test whether inconsistencies in host sampling biased observed latitudinal patterns.

Effect of Climate and Host Diversity on Endophyte Alpha Diversity

We tested whether metrics related to climate or the host community could explain variation in endophyte richness using a mixed-effects multiple regression model. Metrics of absolute temperature (n = 7), absolute precipitation (n = 7), temperature seasonality (n = 4), leaf environment (net primary productivity, evapotranspiration, enhanced vegetation index; n = 3), and growing degree days (n = 6; Table 1) were scaled to means of zero and variances of one and reduced to their first principal components (PC1) with the "prcomp" function in the *stats* package (Appendix 2: Figure S3). Absolute temperature PC1, absolute precipitation PC1, leaf environment PC1, and growing degree days PC1 were negated such that greater PC1 values corresponded to greater values for a given predictor variable. Precipitation seasonality (coefficient of variation) was represented by one variable in the CHELSA climate dataset. The number of host genera and host phylogenetic diversity at a site were

20

used as measures of host diversity. Thus, eight predictors were screened to include in a mixed-effects model as fixed effects: absolute temperature PC1 (73.4% variation), absolute precipitation PC1 (67.5% variation), temperature seasonality PC1 (65.9% variation), leaf environment PC1 (76.4% variation), growing degree days PC1 (68.1%), precipitation seasonality, host species diversity, and host phylogenetic diversity. Quadrat (n = 100) was included as a random effect.

Before fitting the multiple regression model, we excluded temperature seasonality PC1, leaf environment PC1, and growing degree days PC1 because they were significantly correlated with absolute temperature PC1 (Pearson's r: -0.85; p < 0.001, Pearson's r: 0.65; p < 0.01, Pearson's r: 0.97; p < 0.001, respectively). With the five remaining predictors, we fit a linear mixed-effects model predicting endophyte 97% OTU richness with the "lmer" function in the *lme4* package (Bates et al. 2015) and checked model assumptions with the "check_model" function in the *performance* package (Lüdecke et al. 2021). The number of host genera at a given site was removed as a predictor because it had a high variance inflation factor (VIF) of 6.70, indicating collinearity with other predictors. Absolute temperature PC1, absolute precipitation PC1, precipitation seasonality, and host phylogenetic diversity had a VIF < 4 and were fit as predictors (n = 4) in the final linear mixed-effects model. The partial residuals of endophyte richness were extracted with the "partialize" function in the *jtools* package (Long 2022) and regressed as a function of the four remaining predictor variables. The coefficient for each of the four predictors in the final mixed-effects model corresponded to the slope of the regression line in each partial residual plot. Thus, partial residual plots display the effect of one predictor on endophyte richness

while accounting for the other three predictors as fixed effects and quadrat as a random effect.

Endophyte Phylogenetic Diversity

The ITS1 regions of 97% OTUs in the Pezizomycotina (43.62% of 20,800 OTUs and 56.43% of 5,796,623 reads) were placed within a Pezizomycotina reference tree using the Tree-Based Alignment Selector Toolkit v.2.1 (https://tbas.hpc.ncsu.edu/) with RaxML and the following settings: EPA with likelihood weights, a GTRCAT rate heterogeneity model, no genetic distance cutoff, and the provided outgroups selected. The phylogenetic diversity of Pezizomycotina endophytes was measured as the standardized effect size of the mean pairwise phylogenetic distance per quadrat with the "ses.mpd" function in the *picante* package using the "taxa.labels" null model and 999 cophenetic distance matrix randomizations.

Model Selection and Output

For each regression analysis, models were constructed with linear and quadratic terms with "lme" function in the *nlme* package (Pinheiro and Bates 2000) because measures of community structure for fungi have been shown to be non-linear with respect to various abiotic variables (Tedersoo et al. 2012, Cobian et al. 2019). Models were fitted with maximum likelihood and ranked by Akaike Information Criterion with a correction for small sample sizes (AICc) with the "model.sel" function in the *MuMIn* package (Barton 2009). Models with the lowest AICc were refit with restricted maximum likelihood and the significance of each linear or polynomial term was tested with the "anova.lme" function in

the *nlme* package. The marginal coefficient of determination (R^2), the amount of variance explained by the fixed effect(s), and the conditional R^2, the amount of variance explained by the fixed and random effects, were calculated with the "r.squaredGLMM" function in the *MuMIn* package (Nakagawa and Schielzeth 2013, Johnson 2014, Nakagawa et al. 2017). We tested for the presence of spatial autocorrelation, which can increase Type I errors (Legendre et al. 2002), after specifying quadrat or site as a random effect for each mixed-effects model by quantifying Moran's I with the "testSpatialAutocorrelation" function in the *DHARMa* package (Hartig 2020). We also performed regressions with only temperate sites, excluding the two tropical sites in Panama, which often represented outlier values in climate and the leaf environment (Appendix 2: Figure S1).

Generalized Dissimilarity Modeling

We used generalized dissimilarity modeling (GDM) to test whether climate, the leaf environment, host diversity, host phylogeny, or geographic distance could explain compositional differences among endophyte communities. GDM is a non-linear form of matrix regression that fits I-spline functions with maximum likelihood (Ferrier et al. 2007). GDM models the effect of each predictor variable while holding all others constant, thus the height of the I-spline represents how well a variable predicts compositional differences among endophyte communities and this value is relative to all other predictors. The slope of the I-spline corresponds to the rate of endophyte compositional turnover across the range of each predictor variable (Mokany et al. 2022).

Endophyte beta diversity was modeled with the same aforementioned predictors – climate PC1s, leaf environment PC1, and host diversity variables – with quadrat as a

categorical variable. The phylogenetic distances among hosts and the geographic distances among sites were included as distance matrices. Host phylogenetic distance was calculated with the "cophenetic.phylo" function in the *ape* package (Paradis and Schliep 2019). Geographic distance was measured as the Great Circle (WGS84 ellipsoid) distance with the "spDists" function in the *sp* package (Bivand et al. 2013). Endophyte community dissimilarity was measured as Bray-Curtis dissimilarity with the "vegdist" function in the *vegan* package.

We fit a full model with all predictors and then used backward elimination to remove variables that were not predictive of endophyte turnover with the "gdm.varImp" function in the *gdm* package (Ferrier et al. 2007). From this process, the number of host genera at a site and quadrat were eliminated as predictor variables. The final GDM model tested the effect of absolute temperature PC1, absolute precipitation PC1, temperature seasonality PC1, precipitation seasonality, leaf environment PC1, host phylogenetic diversity, host phylogenetic distance, and geographic distance on endophyte Bray-Curtis dissimilarities with the "gdm" function in the *gdm* package. We also performed linear matrix regressions (i.e., Mantel tests) among the same dissimilarity or distance matrices with the "mantel" or "mantel.partial" functions in the *vegan* package.

### D. Results

Sequencing Results

High-throughput sequencing of 1,898 pairs of endophyte communities across 14 Illumina MiSeq runs yielded 103,002,175 reads that passed quality control filters. After excluding endophyte communities for which the fungal or plant DNA did not amplify by PCR, removing endophyte OTUs that did not appear in replicate sequencing runs, and

rarefying to 3,500 reads per endophyte community, the final dataset contained 5,796,623

reads spanning 1,657 endophyte communities composed of 20,800 non-singleton 97%

OTUs in 443 plant genera. In general, replicate endophyte communities sequenced on

different Illumina MiSeq runs had very similar compositions of sequences. On average,

94.92 OTUs (0.35% of OTUs per endophyte community) appeared in only one of the two

replicate communities and these OTUs had an average read abundance of 2.86 (n = 1,898

pairs of endophyte communities). The vast majority of reads (mean = 99.84% of reads) were

correctly assigned to each positive control (n = 27; Table 2), suggesting that barcode

switching and contaminations were minimal in this dataset.

Endophyte Alpha Diversity

Endophyte 97% OTU richness increased towards the tropics but varied as a cubic

function (R^2_m = 0.69; R^2_c = 0.85; Figure 2a), with a local peak at 44-52ºN in Oregon and

British Columbia. Plant communities in Panama (9ºN; mean OTU richness per host

individual = 300.27 ± 6.50 SE) hosted the greatest richness of endophytes while plant

communities in California (33-34ºN; mean OTU richness = 51.33 ± 2.65) hosted the lowest

richness (Figure 2a). Trends remained after accounting for differences in host sampling

among sites (R^2_m = 0.65, R^2_c = 0.86; Wald test of linear and polynomial latitude terms: p <

0.05, number of host genera: p = 0.43, host phylogenetic diversity: p = 0.86). Endophyte

communities within angiosperms (asterids, rosids, basal eudicots, monocots, Magnoliids)

and ferns were more species-rich in the tropics than in temperate regions (Appendix 2:

Figure S4). Endophyte communities within Pinophyta, and Bryophyta did not vary

significantly as a function of latitude (Appendix 2: Figure S4). Only 12 endophyte

25

communities within the Lycophyta were sequenced and therefore, trends were not analyzed.

Differences in endophyte richness between the tropical and temperate sites were not

sensitive to the sampling depth of host genera (Appendix 2: Figure S5). Square-root

transformed read abundances (e.g., Tedersoo et al. 2014; U'Ren et al. 2019) yielded similar

latitudinal trends in endophyte richness as rarefying (Appendix 2: Figure S6). Endophyte

richness was not spatially autocorrelated (Moran's I spatial autocorrelation test: observed = -

0.18, expected = -0.053; p = 0.44).
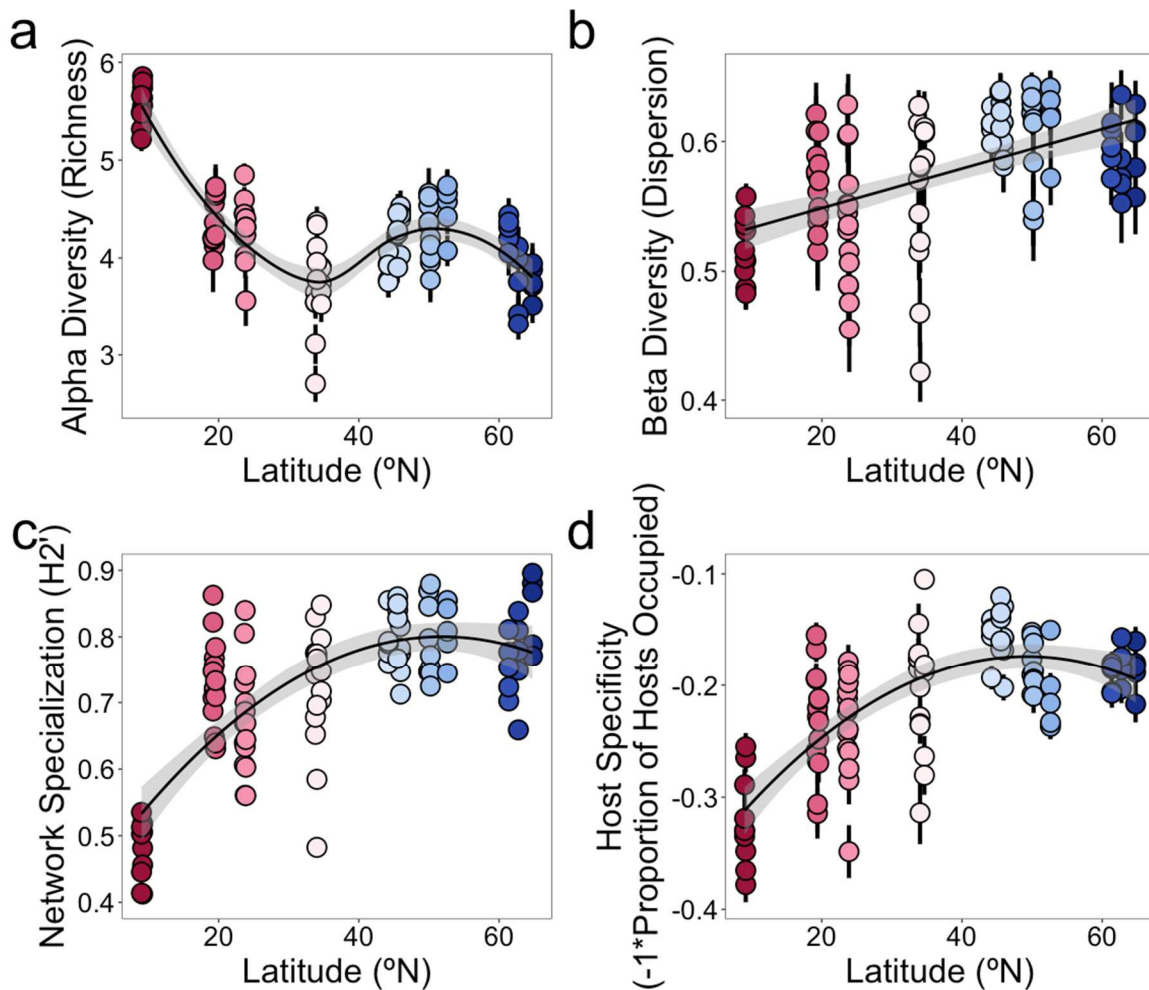
Figure 2. Patterns of endophyte (a) alpha diversity, (b) beta diversity, (c) network specialization, and (d) host specificity with respect to latitude. (a, b, d) Each point represents the mean value for diversity or host specialization per quadrat (n = 100) with bars referring to standard error. Shaded regions refer to 95% confidence intervals. (a) Alpha

26

diversity was measured as the richness of 97% endophyte OTUs within each host individual (n = 1,657) and is shown as natural log-transformed and averaged per quadrat. Richness varied as a cubic function with respect to latitude ($R^2_m = 0.69$; $R^2_c = 0.85$). (b) Beta diversity was measured for each quadrat as the dispersion of endophyte communities in PcoA space and varied as a linear function with respect to latitude ($R^2_m = 0.29$; $R^2_c = 0.56$). (c) Network specialization (endophyte specialization to hosts and vice versa) was measured as a single value per quadrat as the H2' index and varied as a quadratic function with respect to latitude ($R^2_m = 0.55$; $R^2_c = 0.67$). Greater values correspond to greater specialization. (d) Host specificity of endophytes was measured as -1 * the proportion of hosts occupied by each 97% OTU in a given quadrat and varied as a quadratic function with respect to latitude ($R^2_m = 0.54$; $R^2_c = 0.65$). Greater values correspond to greater host specificity.

Factors related to climate (Figure 3a-c) and host diversity (Figure 3d) were associated with endophyte richness (full model: $R^2_m = 0.37$; $R^2_c = 0.53$). Absolute temperature PC1 explained the most variation in endophyte richness (with Panama: $R^2_{adj} = 0.19$, $p < 0.001$; without Panama: $R^2_{adj} = 0.05$, $p < 0.001$), followed by absolute precipitation PC1 (with Panama: $R^2_{adj} = 0.15$, $p < 0.001$; without Panama: $R^2_{adj} = 0.08$, $p < 0.001$), precipitation seasonality (with Panama: $R^2_{adj} = 0.07$, $p < 0.001$; without Panama: $R^2_{adj} = 0.07$, $p < 0.001$), and host phylogenetic diversity (with Panama: $R^2_{adj} = 0.06$, $p < 0.001$; without Panama: $R^2_{adj} = 0.06$, $p < 0.001$). The effect of spatial autocorrelation was not significant (Moran's I: observed = 0.15, expected = -0.053; $p = 0.23$). Trends tended to be consistent among host major groups (Appendix 2: Figure S7).
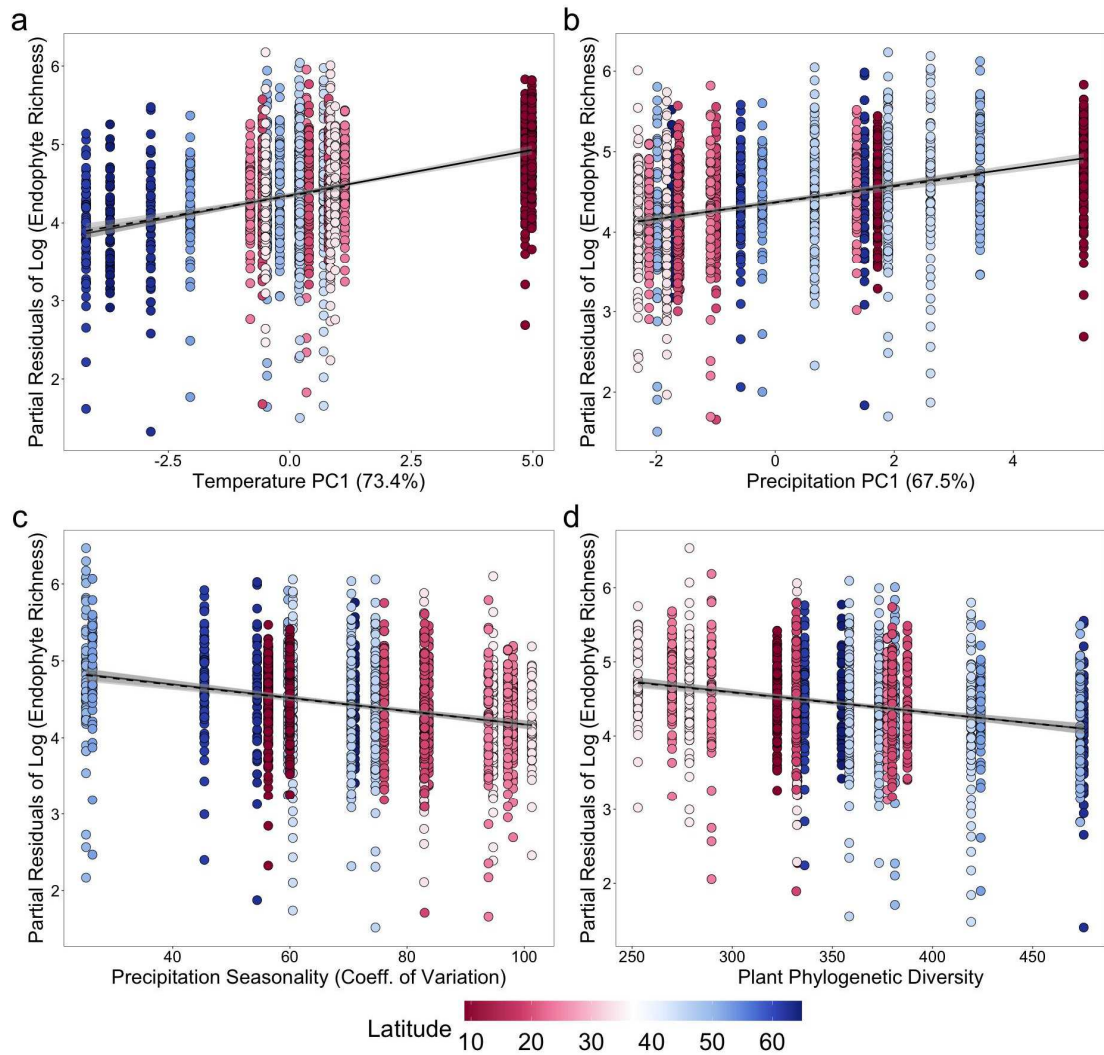
Figure 3. Climate (a-c) and host diversity (d) explain variation in endophyte richness. Partial residual plots derived from a linear mixed-effects model predicting log-transformed 97% OTU endophyte richness while accounting for temperature, precipitation, precipitation seasonality, and plant phylogenetic diversity as fixed effects and quadrat (n = 100) as a random effect (full model: $R^2_m = 0.37$; $R^2_c = 0.53$; Moran's I spatial autocorrelation test: p = 0.23). Each point represents the endophyte community within one host individual (n = 1,657). Each partial residual plot shows the effect of one predictor on endophyte richness when all other predictors are accounted for. Greater principal component (PC) values (a-b) represent greater values for a given predictor variable. Temperature seasonality PC1, leaf environment PC1, growing season PC1, and the number of plant genera at a given site were excluded from the model because they were highly correlated to other predictors (see Methods). Regressions included (solid lines) or excluded (dashed lines) the two Panamanian sites (darkest red points) to test whether the significance of each predictor variable was dependent on the inclusion of tropical ecosystems. Shaded regions refer to 95% confidence intervals. (a) Partial residuals of endophyte richness with respect to temperature PC1 (with Panama: $R^2_{adj} = 0.19$, p < 0.001; without Panama: $R^2_{adj} = 0.05$, p < 0.001). (b) Partial

residuals of endophyte richness with respect to precipitation PC1 (with Panama: $R^2_{adj} = 0.15$, $p < 0.001$; without Panama: $R^2_{adj} = 0.08$, $p < 0.001$). (c) Partial residuals of endophyte richness with respect to precipitation seasonality (with Panama: $R^2_{adj} = 0.07$, $p < 0.001$; without Panama: $R^2_{adj} = 0.07$, $p < 0.001$). (d) Partial residuals of endophyte richness with respect to host phylogenetic diversity measured as the mean pairwise phylogenetic distance among all host genera at a given site (with Panama: $R^2_{adj} = 0.06$, $p < 0.001$; without Panama: $R^2_{adj} = 0.06$, $p < 0.001$). Trends tended to be consistent among host major groups (Appendix 2: Figure S7).

Phylogenetic diversity of endophytes within the Pezizomycotina decreased as a function of latitude ($R^2_m = 0.46$; $R^2_c = 0.93$; Appendix 2: Figure S11). Endophyte communities in Panama had the lowest phylogenetic diversity (SES MPD: -2.83 ± 0.45) while communities in Oregon had the greatest (SES MPD: 0.13 ± 0.20). Trends remained with only angiosperm hosts ($R^2_m = 0.46$; $R^2_c = 0.93$; Appendix 2: Figure S11). Patterns of phylogenetic diversity were not spatially autocorrelated (Moran's I: observed $= -0.10$, expected $= -0.053$, $p = 0.75$).

Endophyte Beta Diversity

Endophyte beta diversity, or the differences in endophyte composition among host species within a quadrat, decreased towards the tropics ($R^2_m = 0.29$, $R^2_c = 0.56$; Figure 2b). The two plant communities in Panama (9°N) hosted the most similar communities of endophytes among quadrats (mean dispersion $= 0.52 \pm 0.0042$ SE) while plant communities in Oregon (44 – 45°N) had the most distinct communities of endophytes (mean dispersion $= 0.61 \pm 0.0044$; Figure 2b). Trends remained after accounting for differences in host sampling among sites ($R^2_m = 0.34$, $R^2_c = 0.60$; Wald test of linear latitude term: $p < 0.05$, number of host genera: $p = 0.96$, host phylogenetic diversity: $p = 0.18$). Endophyte beta diversity was greater in plant communities composed of fewer plant genera ($R^2_m = 0.11$; $R^2_c = 0.56$; Figure 4a) that were more distantly related ($R^2_m = 0.16$; $R^2_c = 0.57$; Figure 4d).

Endophyte beta diversity was not spatially autocorrelated (Moran's I: observed = -0.053, expected = -0.053; p = 0.9996).
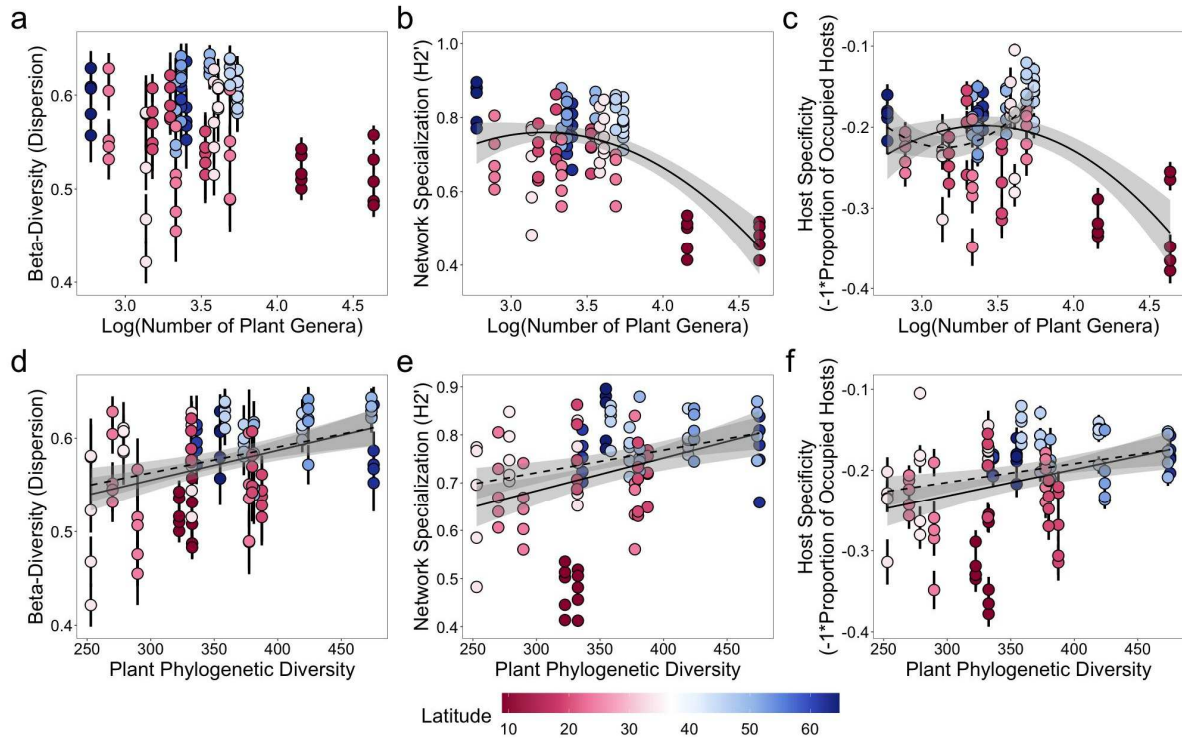
Figure 4. Host diversity explains variation in endophyte beta diversity, network specialization, and host specificity. (a, c, d, f) Each point represents the mean value for beta diversity or host specificity per quadrat (n = 100) with bars referring to standard error. Regressions included (solid lines) or excluded (dashed lines) the two Panamanian sites (darkest red points) to test whether the significance of each predictor variable was dependent on the inclusion of tropical ecosystems. The absence of a regression line indicates the parameter estimate for the linear or quadratic term was not statistically significant. (a-c) Endophyte beta diversity ($R^2_m = 0.11$; $R^2_c = 0.56$), network specialization ($R^2_m = 0.38$; $R^2_c = 0.66$), and host specificity (with Panama: $R^2_m = 0.27$, $R^2_c = 0.65$; without Panama: $R^2_m = 0.20$, $R^2_c = 0.48$) were higher in plant communities composed of fewer host genera. (d-f) Endophyte beta diversity (with Panama: $R^2_m = 0.16$, $R^2_c = 0.57$; without Panama: $R^2_m = 0.14$, $R^2_c = 0.50$), network specialization (with Panama: $R^2_m = 0.13$, $R^2_c = 0.67$; without Panama: $R^2_m = 0.13$, $R^2_c = 0.32$), and host specificity (with Panama: $R^2_m = 0.12$, $R^2_c = 0.65$; without Panama: $R^2_m = 0.10$, $R^2_c = 0.48$) were higher in plant communities composed of more distantly-related host genera.

Network Specialization and Host Specificity

Network specialization ($R^2_m = 0.55$, $R^2_c = 0.67$; Figure 2c) and host specificity ($R^2_m =$ 0.54, $R^2_c = 0.65$; Figure 2d) decreased towards the tropics. Plants and endophytes in Panama (9ºN; mean H2' = 0.48 ± 0.014 SE) were the most generalized in their associations while plants and endophytes in Oregon (44 - 45ºN; mean H2' = 0.80 ± 0.011) were the most specialized (Figure 2c). Similarly, endophyte species in Panama had the greatest average host range (average proportion of hosts occupied in a quadrat = 0.32 ± 0.041 SE) while endophyte species in Oregon had the narrowest average host range (average proportion of hosts occupied in a quadrat = 0.15 ± 0.022 SE; Figure 2d). Network specialization (H2') and host specificity were greater in temperate plant communities composed of fewer plant genera (H2': $R^2_m = 0.38$, $R^2_c = 0.66$; host specificity: $R^2_m = 0.27$, $R^2_c = 0.65$; Figure 4b-c) that were more distantly related (H2': $R^2_m = 0.13$, $R^2_c = 0.67$; host specificity: $R^2_m = 0.12$, $R^2_c = 0.65$; Figure 4e-f).

For network specialization, trends remained after accounting for differences in host sampling among sites ($R^2_m = 0.59$, $R^2_c = 0.68$; Wald test of linear and quadratic latitude terms: p < 0.05), even though number of host genera (p = 0.019) and host phylogenetic diversity (p = 0.075) were significant. Network specialization was spatially autocorrelated (Moran's I spatial: observed = 0.33, expected = -0.053; p = 0.017). Adding a Gaussian spatial correlation structure to the quadratic model neither improved model fit (delta AICc = 2.92) nor changed the significance of latitude as a predictor (linear and quadratic terms were still p < 0.05), indicating that the effect of latitude on network specialization was still significant after accounting for spatial autocorrelation. For host specificity, trends remained after accounting for differences in host sampling among sites ($R^2_m = 0.43$, $R^2_c = 0.65$; Wald

31

test of linear and quadratic latitude terms: $p < 0.05$, number of host genera: $p = 0.75$, host phylogenetic diversity: $p = 0.25$). Host specificity was not spatially autocorrelated (Moran's I: observed = -0.015, expected = -0.053; $p = 0.82$).

Endophyte Community Composition

Region ($R^2 = 0.16$; $p = 0.001$), site ($R^2 = 0.23$; $p = 0.001$), quadrat ($R^2 = 0.31$; $p = 0.001$), and host major group ($R^2 = 0.028$; $p = 0.001$) explained differences in endophyte community composition using PERMANOVA (n = 999; Figure 5a). Factors related to climate, the host community, and geography explained 58.0% of the variation among endophyte communities using generalized dissimilarity modeling (Figure 5b). Absolute temperature PC1 (sum of I-spline coefficients; greater values indicate more predictive power: 2.00) explained the most dissimilarity among endophyte communities followed by geographic distance (1.16), host phylogenetic distance (0.54), absolute precipitation PC1 (0.39), host phylogenetic diversity (0.21), precipitation seasonality (0.12), temperature seasonality PC1 (0.029), and leaf environment PC1 (0.015). Using linear matrix regression, geographic distance was the most correlated with endophyte Bray-Curtis dissimilarity (Mantel r: 0.54; $p = 0.001$), followed by climate dissimilarity (temperature and precipitation variables; Mantel r: 0.51; $p = 0.001$), vegetation dissimilarity (evapotranspiration, enhanced vegetation index, net primary productivity; Mantel r: 0.42; $p = 0.001$), dissimilarity in host species and phylogenetic diversity (Mantel r: 0.30; $p = 0.001$), and host phylogenetic distance (Mantel r: 0.14; $p = 0.001$).
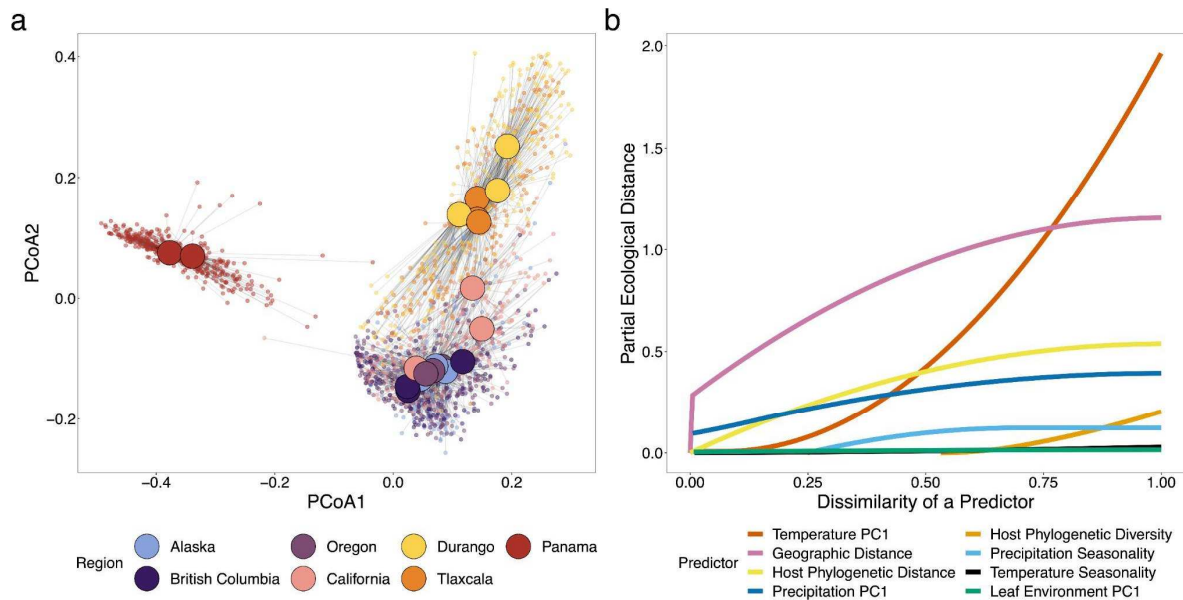
Figure 5. The composition of endophyte communities is structured by geographic distance, temperature, climate, and the host community. (a) Differences in endophyte community composition (Bray-Curtis dissimilarity) in principal coordinate space reflect differences in geography. Small circles refer to the endophyte community within each host individual (n = 1,657), large circles refer to the centroid among endophyte communities within a given site (n = 20), and colors refer to sampling regions (n = 7). Region (R^2 = 0.16; p = 0.001), site (R^2 = 0.23; p = 0.001), quadrat (R^2 = 0.31; p = 0.001), and host major group (R^2 = 0.028; p = 0.001) explained differences in endophyte community composition using PERMANOVA (n = 999). (b) The generalized dissimilarity model suggests temperature PC1, geographic distance, host phylogenetic distance, precipitation PC1, host community diversity, and precipitation seasonality structure endophyte communities. The maximum height of each fitted spline corresponds to the relative importance of that variable predicting differences in endophyte composition. The shape of each fitted spline corresponds to the rate of endophyte turnover across the range of each predictor variable. Values for each predictor variable were divided by their maximum value (e.g., each temperature value / max temperature value) to scale all predictors along a common x-axis. Host phylogenetic diversity and precipitation seasonality do not have values at the origin because all values are greater than zero. Host phylogenetic distance is the phylogenetic distance between any pair of plant hosts while host phylogenetic diversity refers to the mean pairwise phylogenetic distance among all host genera at a given site.

Fungal Taxonomic Associations Among Hosts and Sites

Abundant fungal classes of endophytes varied by host group and geographic region.

Dothideomycetes, Sordariomycetes, and Leotiomycetes were the most abundant classes of

endophytes by OTU number and read abundance (Figure 6). Each of these three dominant

classes had associations with most major groups and could be found in many sites but did

display varying degrees of host clade and geographic specificity. For example,

Dothideomycete and Leotiomycete endophytes were abundant in temperate regions and

broadly associated with most plant groups. On the other hand, Sordariomycete endophytes

were abundant in the tropics and mostly associated with ferns, magnoliids, and monocots.

Although most of the endophytes were in the Ascomycota, endophytes in the

Basidiomycota, particularly the class Agaricomycetes associated with bryophytes in

northern and southern latitudes (Figure 6b).
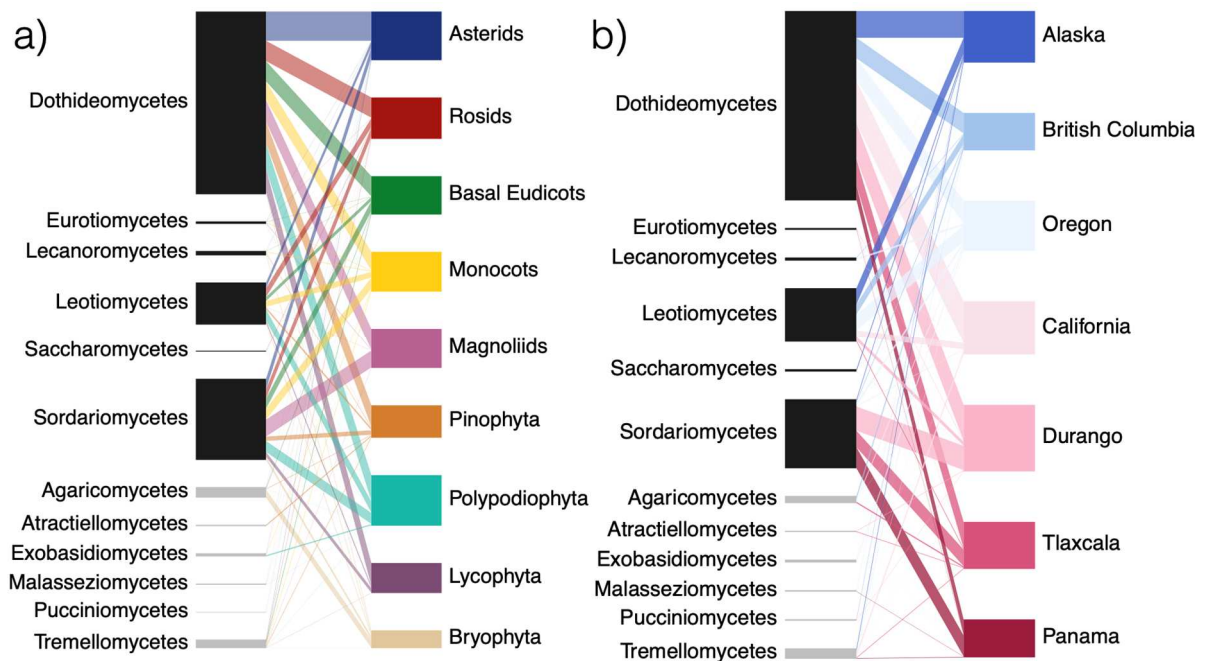
Figure 6. Endophytes vary in distribution by host group and geographic region. (a) Endophyte 97% OTUs with a relative read abundance >0.1% (n = 154 OTUs) grouped by taxonomic class (left bars; n = 12) and their associations to major host groups (right bars; n = 9). (b) Endophyte 97% OTUs with a relative read abundance >0.1% (n = 154 OTUs) grouped by taxonomic class (left bars; n = 12) and their distributions among sampling regions (right bars; n = 7). The width of each bar or pairwise connection was standardized by dividing the number of reads between each host group or sampling region by the number of host individuals sampled within that given group (e.g., number of reads in asterids / number of asterid individuals sampled). Fungal classes are color-coded by phylum (black = Ascomycota; grey = Basidiomycota). Results were similar with a more inclusive relative abundance threshold (Appendix 2: Figure S10).

E. Discussion

Endophyte alpha diversity

Endophyte species richness (alpha diversity) generally increased towards the equator (Figure 2; Appendix 2: Figure S5), consistent with LDG (Hillebrand 2004). Our findings agree with Arnold and Lutzoni (2007) who found that the culturable diversity of endophytes in North America increased towards the tropics. Arnold and Lutzoni (2007) also found that latitudinal patterns in endophyte diversity did not vary among angiosperms, conifers, and ferns. However, in this study, endophyte richness in angiosperms and ferns increased towards the tropics while endophyte richness in conifers and bryophytes did not significantly vary as a function of latitude (Appendix 2: Figure S4). Thus, host identity was a factor that biased the measured relationship between endophyte richness and latitude. On the other hand, sampling effort was not. The two tropical plant communities in Panama hosted the greatest richness of endophytes at varying levels of host grouping: individuals, quadrats, and sites (Figure 2; Appendix 2: Figure S4; Figure S5).

Although endophyte richness was highest in the tropics, richness was also pronounced, or was hump-shaped, at mid-latitudes along the west coast of North America (44-52ºN; Figure 2a). Arnold and Lutzoni (2007) found that certain species of conifers (*Picea* and *Pinus* spp.) hosted the greatest diversity of cultured endophytes in temperate regions along the east coast of North America (35-50ºN). However, in this study, angiosperms, conifers, ferns, and bryophytes had similarly high richness at mid-latitudes (Appendix 2: Figure S4) and we could not attribute the hump-shaped pattern to particular host groups. This hump-shaped pattern in temperate regions parallels biogeographic patterns in soil, litter, and ectomycorrhizal fungi (e.g., Tedersoo et al. 2012, Seena et al. 2019, Zhang et al. 2020),

where fungal diversity tends to be highest in temperate or boreal ecosystems. These fungal groups may depend on environmental conditions or host associations that do not trend with latitude. For example, the hump-shaped pattern of ectomycorrhizal richness as a function of latitude is likely shaped by a greater abundance of ectomycorrhizal hosts in temperate regions, the negative effect of very high or low temperatures, and the negative effect of high precipitation that may lead to low soil oxygen (Tedersoo et al. 2012). While the mechanisms underlying reversed or hump-shaped fungal LDGs may vary among fungal guilds, these results indicate that temperate environments host significant free-living and host-associated fungal biodiversity.

Patterns of fungal diversity often reflect differences in climate (Tedersoo et al. 2014) which could explain why endophyte richness was greatest in the tropics and peaked, to a lesser extent, at mid-latitudes (Figure 2a). Sites with higher absolute temperature and precipitation had greater endophyte richness than colder or drier sites (Figure 3a-b), similar to many soil fungal guilds (Tedersoo et al. 2014, Glynou et al. 2016, Zhou et al. 2016, Liu et al. 2020). Additionally, sites in Panama, Oregon, and British Columbia received high amounts of annual precipitation, experienced low precipitation seasonality, and had high photosynthetic biomass (Appendix 2: Figure S1). Endophytes have been cultured from a greater proportion of host tissues in environments that receive more precipitation (Arnold and Lutzoni 2007, U'Ren et al. 2012). Greater moisture availability could mobilize fungi on the leaf surface to indiscriminately colonize the leaf interior, resulting in a positive relationship between precipitation and endophyte richness (e.g., U'Ren et al. 2019; but see U'Ren et al. 2012).

In addition, Oita et al. (2021b) found that tropical forests in Panama that experienced lower temperature and precipitation seasonality hosted a greater richness of endophytes than more seasonal forests, consistent with our finding that endophyte richness was greater in areas with lower precipitation seasonality (Figure 3c). The latitudinal-niche breadth (LNB) hypothesis (Vázquez et al. 2004) predicts that lower seasonality in the tropics reduces fluctuations in resources and endophyte populations, facilitating greater specialization and resource partitioning (e.g., host specificity). This in turn results in greater coexistence among endophytes and greater endophyte richness in the tropics than more seasonal temperate environments. For example, Gilbert and Webb (2007) found that fungal pathogens in the tropics were limited by the phylogenetic breadth of hosts they could infect and Tedersoo et al. (2014) found that fungal pathogens were most species-rich in the tropics. This suggests greater pathogen richness could be a consequence of greater host phylogenetic specificity in the tropics and is supportive of LNB.

For endophytes, we found greater host specificity in environments with greater temperature seasonality (slope = 0.32, p < 0.0001; R^2_m = 0.50, R^2_c = 0.65), contrary to the expectations of LNB. While lower precipitation seasonality in Panama, Oregon, and British Columbia was associated with greater endophyte richness (Figure 3c), precipitation seasonality did not explain variation in host specificity (slope = -0.00028; p = 0.60; R^2_m = 0.011, R^2_c = 0.65), the driving mechanism for greater endophyte richness via LNB. One explanation for this contradiction is that species-rich environments with low precipitation seasonality also had high photosynthetic biomass. Greater photosynthetic biomass, the habitat for endophytes, could support greater numbers of endophyte species than sites with

less photosynthetic biomass and such a relationship could operate independently of host specificity and LNB.

Climate could also have an indirect effect on endophyte richness by influencing the leaf environment. Plant communities in the tropics that experienced greater precipitation and temperature also had greater evapotranspiration, photosynthetic biomass, plant species diversity, and longer growing seasons compared to temperate plant communities (Figure S1). Each of these factors was positively associated with endophyte richness, explaining 38-60% of the variation in endophyte richness (Appendix 2: Figure S9). Thus, greater endophyte richness in the tropics could be due to the direct effect of climate, the effect of climate on plant communities, or both. High temperature, high precipitation, and long growing seasons likely explain high photosynthetic biomass and net primary productivity in Panama, where endophyte richness was greatest. Because the tropics comprise such an expansive habitat for fungi, endophytes may not need to have constrained niches or be host-specific to coexist, as proposed by LNB. Other factors such as evapotranspiration may simply be correlated with climatic or plant factors (air temperature, soil water content, stomatal conductance; Zha et al. 2013) that lead to greater photosynthetic biomass. While we are unable to isolate the effects of climate while controlling for leaf environment, and vice versa, these results support a large body of literature relating greater temperature (Allen et al. 2002), greater net primary productivity (Gillman et al. 2015), lower seasonality (Cirtwill et al. 2015), and greater host diversity ('diversity begets diversity'; Whittaker 1972, Calcagno et al. 2017, Madi et al. 2020) to greater species richness in the tropics.

We found that the phylogenetic diversity of Pezizomycotina endophytes decreased towards the tropics (Figure S11). These findings agree with Arnold and Lutzoni's (2007)

38

results that endophyte communities of angiosperms in a boreal forest spanned twice as many taxonomic classes than in a Panamanian tropical forest. However, the Pezizomycotina only composed about half of the endophyte community, by OTU number and read count, across all host individuals. Patterns of endophyte phylogenetic diversity should be re-examined with future reference trees that include the Saccharomycotina, Taphrinomycotina, and Basidiomycota endophytes or with high-throughput platforms that output longer reads that are appropriate for phylogenetic analysis.

Endophyte Beta Diversity and Host Specificity

Within quadrats, endophyte communities were more similar to one another in the tropics than in temperate regions (Figure 2b), resulting in a latitudinal beta diversity gradient opposite to that of endophyte alpha diversity. I hypothesize that greater compositional dissimilarity among endophyte communities in temperate regions is driven by host specific endophytes. In support of this, endophytes in temperate plant communities occupied a lower proportion of their host communities, or had greater host specificity, than endophytes in sub-tropical and tropical plant communities and this measure was positively correlated with endophyte beta diversity. For example, endophytes in Panama associated with an average of 32% of host species in a given quadrat and this metric of host specificity decreased as a function of latitude to 15% for endophytes in Alaska (Figure 2d).

More heterogeneous environments may impose environmental filters that increase differences in the composition of the endophyte community. For example, greater beta diversity of soil fungi across a 35° latitudinal gradient in China was associated with climatic regimes and soil environments that were more variable (Zhang et al. 2020). For endophytes,

greater diversity in the species or phylogenetic composition of a host community may reflect a more heterogeneous environment, decreasing the likelihood that specialist endophytes encounter suitable host individuals in the community via horizontal transmission (May 1991, Kindlmann et al. 2007, Griffin et al. 2019, Oita et al. 2021b). A greater proportion of specialist endophytes in high latitude plant communities (Figure 2c-d) explains greater differences in endophyte composition among host species (Figure 2b; Figure 7). Conversely, more generalists in the tropics (Figure 2c-d) explains increased compositional similarity among endophyte communities (Figure 2b; Figure 7). Our findings complement studies that show endophytes often display broad host associations in the tropics (Arnold and Lutzoni 2007, Higgins et al. 2011, 2014) while endophytes in temperate or arctic environments are more host-specific (Zhang and Yao 2015, U'Ren et al. 2019).

We also found that host generalism was greater in the tropics where host genera were greater in number and more closely related to each other compared to temperate plant communities (Figure 4). Greater host phylogenetic diversity was associated with greater host specificity (Figure 4f), suggesting that distantly related hosts facilitate host specificity. This complements findings from the GDM model showing that host phylogenetic distance was as significant factor influencing differences among endophyte communities. Phylogenetic distance among host species can be correlated with differences in host traits that impose biotic filters and shape endophyte community assembly, such that host species more closely related to each other select for similar species of endophytes. This pattern of phylosymbiosis (Brooks et al. 2016, Kohl 2020) has been observed for endophytes at varying host phylogenetic and spatial scales (Liu et al. 2019, U'Ren et al. 2019, Darcy et al. 2020).
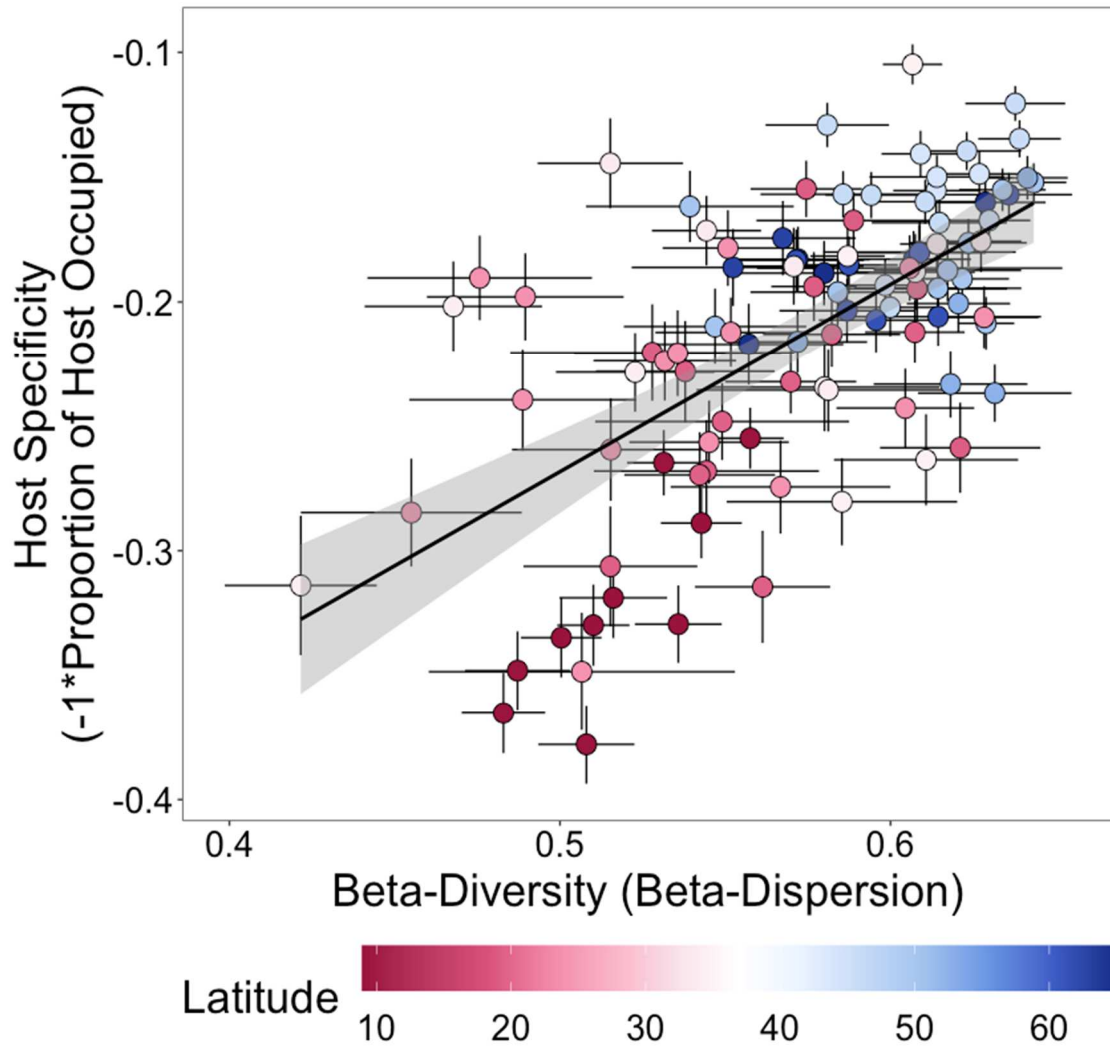
Figure 7. Endophyte beta diversity and host specificity are correlated. Measures of beta diversity and host specificity were correlated (Pearson's r: 0.60; p < 0.01) suggesting that greater compositional dissimilarity among endophyte communities was driven by endophytes that occupied a lower proportion of host species in a plant community. Each point represents the mean value for beta-dispersion or host specificity per quadrat (n = 100) with bars referring to standard error. Beta diversity was measured as the dispersion of endophyte communities within a quadrat in PCoA space. Host specificity was measured as -1 * the proportion of hosts occupied by each 97% OTU in a given quadrat.

While host diversity predicted endophyte turnover within quadrats and sites (Figure 4),

we found that temperature and geography were strongly associated with differences in

endophyte composition across latitudes (Figure 5). This finding is consistent with a large

body of evidence showing that dissimilarity among endophyte communities is often

associated with greater environmental dissimilarity or geographic distance (Zimmerman and Vitousek 2012, Higgins et al. 2014, Oono et al. 2017, Whitaker et al. 2018, Barge et al. 2019, Oita et al. 2021b; but see Vincent et al. 2016; U'Ren et al. 2019). Taken together, our findings indicate that environmental dissimilarity and geographic distance were the most influential factors that differentiated the species composition of endophyte communities across North America (Figure 5) while within sites, host diversity played an important role (Figure 4).

Relationships between Endophyte Alpha Diversity, Beta Diversity, and Host Specificity

Endophyte host specificity could explain why endophyte species richness (alpha diversity) increased towards the tropics while differences in endophyte species composition (beta diversity) and endophyte distribution among hosts (host specificity) increased towards temperate regions. Host specialists may be well-adapted to their hosts or the nutrients within them such that they acquire resources or grow rapidly and competitively exclude other endophytes (Büchi et al. 2014). On the other hand, host generalism may come at the cost of competitive ability. Taken together, we hypothesize that lower host species richness in temperate plant communities selects for host specialists that dominate host tissues, thereby decreasing endophyte richness in hosts. On the other hand, we hypothesize that greater host species richness in the tropics selects for host generalists and their inability to outcompete other endophytes facilitates coexistence, increasing endophyte richness in hosts. In support of this, endophyte richness was negatively correlated with host specificity (Spearman's rho: -0.52, $p < 0.001$). We have proposed three related arguments that can explain contrasting patterns of endophyte alpha and beta diversity as a function of latitude: (1) host specificity

drives endophyte beta diversity (Figure 7), (2) greater host diversity selects for greater host generalism (Figure 4), and (3) the specialist or generalist lifestyle determines endophyte richness within any given host. In the absence of environmental gradients, these proposed mechanisms should produce contrasting patterns of endophyte alpha and beta diversity as a function of plant diversity.

F. Conclusions

Our survey across a 55º latitudinal gradient in North America provides new perspectives on the biogeography of endophytes and their associations among 443 plant host genera. Tropical plant communities hosted the greatest richness of endophytes and suggest that the endophyte communities surveyed in this study conform to the latitudinal diversity gradient. We also found that temperate conifer forests hosted significant endophyte biodiversity across most plant lineages and that temperature, precipitation, and factors related to the leaf environment explained these two peaks in endophyte richness. Across latitudes, endophyte communities were highly differentiated by regions (Panama *vs.* Mexico *vs.* United States and Canada) highlighting the importance of climatic dissimilarity and geographic distance on endophyte species composition. Within sites, endophyte communities were more similar to one another in tropical forests where endophytes were more generalist, whereas endophyte communities in temperate forests were more dissimilar to one another where endophytes were more host-specific. Thus, I propose that host specificity, along with climate and the leaf environment, is a major driver of contrasting patterns of endophyte alpha and beta diversity as a function of latitude. Much of what is known of endophyte diversity and distribution at broad spatial scales has been shaped by surveys in temperate and boreal conifer forests (e.g., Arnold and Lutzoni 2007; U'Ren et al. 2012, 2019; this study; but see

Vincent et al. 2016; Darcy et al. 2020, Oita et al. 2021). Future studies should consider

broadening biogeographic studies to encompass angiosperm-dominated forests and

grasslands as well as incorporating other variables at the scale of leaves, such as nutrient

composition and leaf traits, that will further enrich our understanding of the factors that

govern the diversity and composition of endophyte communities.

III. Chapter 2: Dimensions of host specificity in foliar fungal endophytes

This article is published as Apigo A and Oono R. 2018. Dimensions of Host Specificity in Foliar Fungal Endophytes. In: Pirttilä A, Frank A (eds). Endophytes of Forest Trees. Forestry Sciences, vol 86. Springer, Cham.

A. Abstract

Foliar fungal endophytes colonized the phyllosphere at least 400 million years ago and have since diversified across every terrestrial ecosystem that supports plant life. Understanding how these complex symbiotic associations are generated, distributed and maintained is a challenging task that requires an understanding of host specificity. We propose a conceptual framework that outlines four 'dimensions' of host specificity that account for the geographic, phylogenetic or sampling scale under consideration. These 'dimensions' quantify endophyte abundance and evenness (structural specificity), interaction strength (network specificity), evolutionary relationships (phylogenetic specificity) and the spatial or temporal consistency of the interaction (beta specificity). We present one case study that quantifies and compares structural, network and phylogenetic specificity across endophyte communities partitioned by taxonomy (Ascomycota *vs*. Basidiomycota). We focus on the effects of rare endophyte species, approximated as Operational Taxonomic Units (OTUs), as a key methodological consideration for communities surveyed with next-generation sequencing (NGS) because the statistical nature of rarity confounds the quantification of host specificity. The exclusion of rare endophyte OTUs consistently changed ecological inference by decreasing host specificity averages and increasing

variances within endophyte phyla. To evaluate the degree to which rare endophyte OTUs affect statistical power, we compared our empirical community to that of randomized communities. Excluding rare endophyte OTUs (>10% of total sequences in the case community removed) may lead to spurious host specificity metrics that are not statistically significant from that of randomized communities. Therefore, rare endophyte OTU removal should be done with explicit rationale. We propose conceptualizing endophyte host specificity with a multidimensional framework that will allow future studies to use quantitative, comparable and theory-driven metrics that can scale towards more meaningful estimates of global fungal biodiversity.

B. Introduction

Foliar fungal endophytes (endophyte; Class 3 endophytes sensu Rodriguez et al. 2009) asymptomatically inhabit the aboveground, photosynthetic tissues of all currently described plant phyla. They are a species-rich and phylogenetically diverse guild distributed worldwide across every terrestrial biome (Bacon and White 2000), ranging from arctic to tropical ecosystems (Arnold 2007, Higgins et al. 2007). The ubiquity and apparent 'hyperdiversity' of endophyte (Arnold et al. 2000) has elicited decades of research documenting patterns of their diversity and distribution (e.g., Carroll and Carroll 1978, Saikkonen et al. 1998, Faeth and Fagan et al. 2002, U'Ren et al. 2012, Zhang and Yao 2015) with ultimately minimal consensus regarding the measurement of one factor intrinsic to the study of their biodiversity and relationships within the plant kingdom – their **host specificity**. Host specificity is one example of a biological process more broadly referred to in the literature as ecological specialization (Poisot et al. 2012). This classic explanation for species coexistence, as a result of resource partitioning (Walter 1991), describes a process

where the realized niche of an organism narrows in trade-off for higher fitness on a smaller fraction of resources (Futuyma and Moreno 1988, Devictor et al. 2010). For endophytes, we propose defining host specificity as the state of occupying a restricted proportion of hosts in a plant community.

Surveys of endophyte biodiversity frequently investigate various aspects of endophyte host specificity (e.g., endophyte community composition as a function of plant host identity), but rarely define the type of specificity under consideration. Terminology can vary across studies, such as host affinity (Higgins et al. 2007), host association (Vincent et al. 2016), host preference (Cannon and Simmons 2002) and host range (Arnold and Lutzoni 2007). Comparisons across different endophyte studies are also challenging because host specificity can be evaluated in different ways depending on the scale of the study. The majority of endophyte studies utilize multivariate ordination approaches to qualitatively compare endophyte compositional differences among plant host species or environmental factors (e.g., Sun et al. 2012, Zhang and Yao 2015). These approaches can assess the relative influences of host specificity, environmental filtering or stochastic assembly on community composition (Brooks et al. 2016). However, ordination approaches can vary significantly by direct or indirect gradient analyses (e.g., canonical correspondence analysis *vs.* nonmetric multidimensional scaling), distance or dissimilarity metrics employed (e.g., Euclidean *vs.* Bray-Curtis) and the number of dimensions or axes considered in multidimensional space (Minchin 1987). Additionally, studies may explore different 'dimensions' of host specificity that cannot be directly compared to one another (e.g., the number of hosts *vs.* the phylogenetic breadth of hosts an endophyte occupies). To transition future studies to more quantitative and comparable approaches, we propose a common framework of four

47

'dimensions' of host specificity (adapted and modified from Poulin et al. 2011) that address

different sampling, spatial and phylogenetic scales for which endophyte are studied in a

plant community:

> (1) Structural specificity (endophyte abundance and evenness)

> (2) Network specificity (interaction strength)

> (3) Phylogenetic specificity (evolutionary relationships)

> (4) Beta specificity (spatial or temporal turnover)

For each of these metrics, a narrower host breadth indicates higher host specificity.

Structural specificity quantifies the most fundamental 'dimension' of host specificity, the

sum and evenness of abundance among hosts (Poulin et al. 2011). **Network specificity**

quantifies the strength of plant-endophyte interactions by accounting for all potential hosts

an endophyte could occupy in a plant community. **Phylogenetic specificity** quantifies host

specificity relative to the phylogenetic scale of the plant hosts in a community, or the mean

phylogenetic distance among occupied hosts (Webb et al. 2008). Structural, network and

phylogenetic specificity quantify the degree of host specificity within a single locality,

termed alpha-specificity (Figure 8; Poulin et al. 2011). Analogous to alpha diversity

(Whittaker 1972), these three host specificity metrics do not account for spatiotemporal

variation of the interaction. **Beta specificity**, however, quantifies the degree to which a

given endophyte displays consistent host specificity across a range of contexts.

Among the three alpha 'dimensions' of host specificity, we explored the influence of (1)

rare endophyte species in the community, which tend to inflate host specificity, (2) data

structure (presence-absence *vs*. abundance-weighted metrics) and (3) correlations among

each of these metrics with a case study endophyte community surveyed with next-generation

sequencing (NGS). We then compared the empirical host specificity values to those of randomized communities to understand the influence of rare endophyte species and data structure on the statistical power of these metrics. Throughout this review, we describe and quantify host specificity among co-occurring plant species in a community context (e.g., Figure 8). However, these 'dimensions' of host specificity could also be adapted to quantify specificity within host individuals (e.g., tissue specificity) or among host populations (e.g., genotype specificity), depending on the context of the study. We will use the term 'plant host' as a general descriptor that encompasses these various partitions of the host pool.
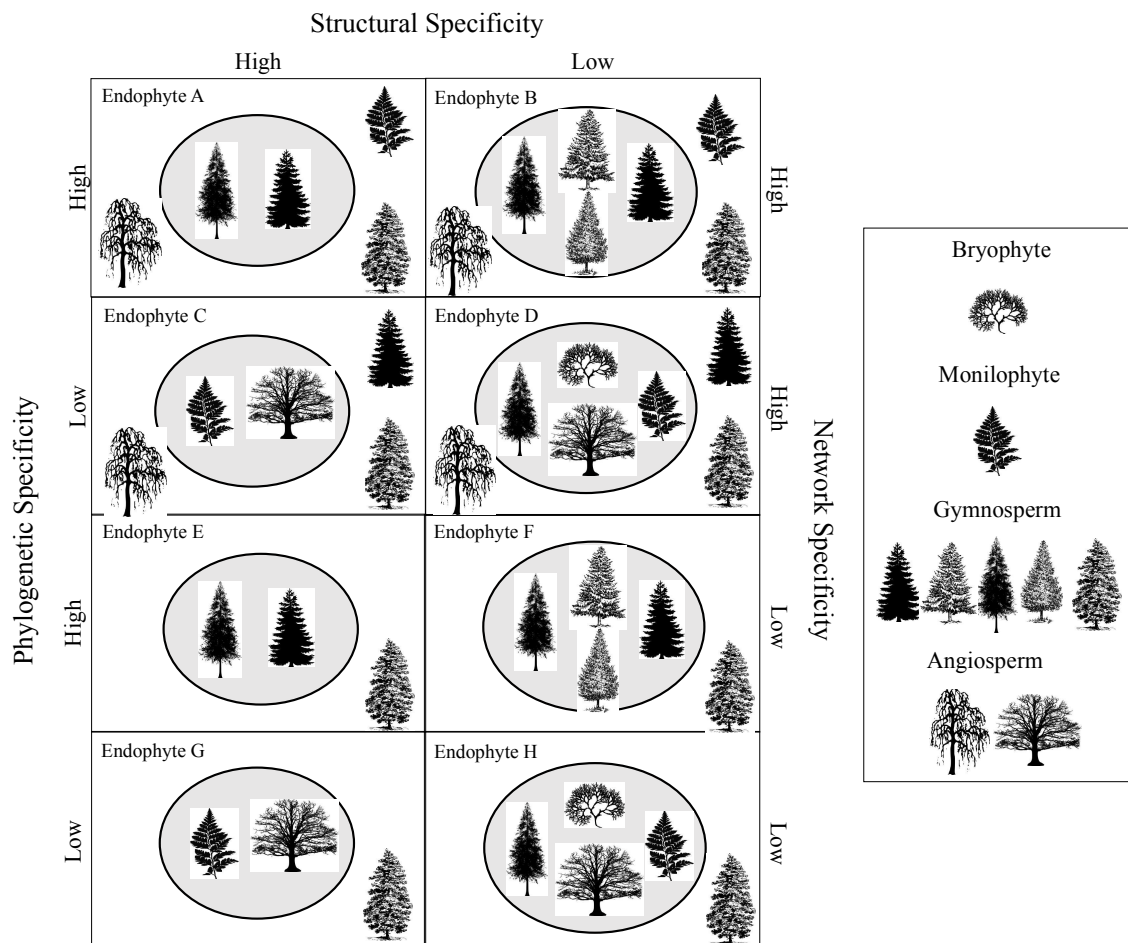
Figure 8. Conceptual diagram of structural, network and phylogenetic specificity. Plant hosts within the grey circles represent the realized ecological niche of each endophyte. endophyte 'A' has high structural specificity (two occupied hosts), phylogenetic specificity (restricted to gymnosperm hosts) and network specificity (three unoccupied hosts).

endophyte 'H' has low structural specificity (four occupied hosts), phylogenetic specificity (bryophyte, monilophyte, gymnosperm and angiosperm hosts) and network specificity (one unoccupied host). Plant species images were procured from a public domain illustration repository, https://openclipart.org.

C. Structural Specificity

Structural specificity of a given endophyte describes the number of occupied hosts, also referred to as 'basic specificity' (Poulin et al. 2011), or its differential abundance among them. Host richness (equation 1) quantifies all hosts an endophyte occupies, regardless of endophyte abundance in each host, and has a presence-absence data structure. In equation 1, r is the number of plant hosts occupied by one endophyte species. A given endophyte will have higher structural specificity if it occupies fewer plant hosts (Figure 8). Shannon's H diversity index (Shannon and Weaver 1948), a commonly used abundance-weighted metric in ecological studies, quantifies host richness and endophyte evenness among occupied hosts. In equation 2, r is the number of plant hosts occupied by one endophyte species and p_i is the relative read abundance of that endophyte species in the ith host. For abundance-weighted Shannon's H, an endophyte will have higher structural specificity if its abundance is unevenly distributed relative to another endophyte, given that host richness is equal between them. In this chapter, host richness and Shannon's H have been negated (multiplied by -1), such that lower host richness indicates higher structural specificity (Figures 8-13).

$$Host\ richness\ (presence - absence) = r \quad (1)$$

$$Shannon's\ H\ (abundance - weighted) = -\sum_{i=1}^{r} p_i \ln p_i \quad (2)$$

Structural specificity provides a straightforward metric to compare sites within a single study with a standardized sampling design. However, the number of hosts in which a given endophyte is found or its abundance within a particular host is entirely dependent on the

scale of each study (Levin 1992, Cavender-Bares et al. 2009). Sampling can also vary at the level of the community (e.g., number of host species sampled), population (e.g., number of host conspecifics sampled) or individual (e.g., number of leaves sampled). Consequently, structural specificity is challenging to compare across different studies.

D. Network Specificity

Network specificity quantifies the number and 'strength' of interactions by accounting for all potential hosts a given endophyte could occupy (Figure 8). The Resource Range Index (RRI; Schoener 1989; equation 3) is a presence-absence metric that normalizes the number of unoccupied hosts ($R - r$) by the total number of possible hosts in the community minus one ($R - 1$). In equation 3, R is the total number of hosts in the community and r is host richness. The abundance-weighted Paired Difference Index (PDI; Poisot et al. 2012; equation 4) compares the relative link strengths (based on endophyte abundance) of all occupied and unoccupied hosts to the highest link strength normalized by the total number of possible hosts in the community minus one ($R - 1$; see Poisot et al. (2012), which shows PDI outperforms other network indices for discriminating generalists and specialists). In equation 4, P_1 is the read abundance in the host with the highest link strength (maximum read abundance), P_i is the link strength (read abundance) in the ith host and R is the total number of plant hosts. RRI and the PDI are scaled such that 0 indicates a perfect generalist (occurs in all plant hosts for RR; occurs in all plant hosts in equal abundance for PDI) and 1 indicates a perfect specialist (occurs in only one plant host for RRI and PDI).

$$Resource\ Range\ Index\ (presence-absence) = \frac{R - r}{R - 1} \quad (3)$$

$$Paired\ Difference\ Index\ (abundance - weighted) = \frac{1}{R-1}\sum_{i=2}^{R}\left(\frac{P_1 - P_i}{P_1}\right) \quad (4)$$

Network metrics are reported as scale-independent measures that are robust to variability in sampling and network structure (Blüthgen et al. 2006, Poisot et al. 2012), lending themselves as a potential common 'unit' across future endophyte studies. Network specificity can be quantified not only at the species-level (e.g., RRI or PDI), but also at the group-level (e.g., plants or endophyte) or network-level (whole network architecture values; Dormann et al. 2009). Certain types of ecological associations have been suggested to correlate with network-level structure (Bascompte 2010, Thébault and Fontaine 2010). For example, obligate, mutualistic associations are thought to display higher modularity, or subgroups of strong species interactions (Wardhaugh et al. 2015), relative to facultative, mutualistic associations. Modularity may promote network stability by localizing the effects of a perturbation within a module of tightly associated species (Olesen et al. 2007). Alternatively, more facultative, mutualistic relationships have been thought to display higher nestedness, or reciprocal specialization, where specialists of one class (e.g., plants or endophyte) strongly interact with generalists of the other class. Nestedness may also promote network stability by reducing competition among organisms of one class (Saavedra et al. 2011).

Cultured endophytes were found to have lower reciprocal specialization (nestedness), a lower number of realized interactions to their plant hosts (connectance) and distributed in more compartmentalized interacting sub-groups (modularity), relative to cultured endolichenic fungi (fungal endophytes living within lichen thalli; Chagnon et al. 2016). The consistency of these patterns across a diverse array of host species and environments (Chagnon et al. 2016) suggests cultured endophyte are more intimately associated with their

hosts relative to cultured endolichenic fungi. Although these network-level metrics that consider all members of a community will correlate with the species-level metrics (RRI and PDI), these network-level metrics are useful for comparing communities that may differ in composition and structure, such as endophytic and endolichenic fungi. Alternatively, species-level metrics, such as RRI or PDI, measured per endophyte could reveal specific fungal or host taxa that predominately contribute to observed patterns among communities or across studies.

E. Phylogenetic Specificity

Structural and network specificity do not consider the phylogenetic diversity of the plant community, which may strongly influence symbiont community structure (Webb et al. 2002, Cavender-Bares et al. 2009, Brooks et al. 2016). **Phylogenetic specificity** quantifies the Mean Pairwise Phylogenetic Distance (MPD; Webb 2000; equation 5). among all hosts occupied by an endophyte and can be used to compare host specificity across different phylogenetic scales (Figure 8). For presence-absence MPD, a given endophyte will have higher phylogenetic specificity if it occupies hosts that are more closely related. This metric can also be abundance-weighted ('structural phylogenetic specificity' in Poulin et al. 2011). For abundance-weighted MPD, a given endophyte will have higher phylogenetic specificity if its abundance is unevenly distributed relative to another endophyte, given that MPD is equal between them. In equations 5 and 6, r is host richness, $\delta_{i,j}$ are the phylogenetic distances between hosts i and j. For abundance-weighted phylogenetic specificity, f is the endophyte abundance in hosts i and j. In this chapter, presence-absence and abundance-weighted MPD have been negated (multiplied by -1), such that lower MPD indicates higher phylogenetic specificity (Figures 8-13).

$$Mean\ Pairwise\ Distance\ (presence - absence) = \frac{\sum_i^r \sum_j^r \delta_{i,j}}{r}, i \neq j \quad (5)$$

$$Mean\ Pairwise\ Distance\ (abundance - weighted) = \frac{\sum_i^r \sum_j^r \delta_{i,j} f_i f_j}{\sum_i^r \sum_j^r f_i f_j}, i \neq j \quad (6)$$

If phylogenetically conserved traits of the host influence the community composition of symbionts, decreasing host phylogenetic distance is predicted to be associated with increasing similarity in endophyte community composition (Webb et al. 2002; Brooks et al. 2016). The relative roles of host evolutionary history (Webb et al. 2002), environmental filtering (Kraft et al. 2014) and drift (Hubbell 2001) can be inferred from how endophyte community composition is distributed relative to the phylogenetic distance, environmental dissimilarity and geographic distance among hosts. For example, if endophyte community assembly is strongly influenced by host evolutionary relationships, an endophyte cluster dendrogram or other metric for compositional similarity (e.g., non-metric multidimensional scaling; NMDS) will show topological congruence in endophyte community composition as a function of host phylogenetic distance (i.e., phylosymbiosis; Brooks et al. 2016). Alternatively, if endophyte community assembly is strongly influenced by environmental filtering, endophyte compositional dissimilarity will have a positive relationship with increasing distance or increasing environmental dissimilarity among sites.

There seems to be mixed support for the influence of phylogenetic specificity in endophyte systems depending on the host phylogenetic scale and the ecosystem. In temperate forests, there is empirical evidence for host specificity at coarse host taxonomic levels (e.g., gymnosperms *vs*. angiosperms; Higgins et al. 2007, U'Ren et al. 2012) for cultured endophytes. Intriguingly, U'Ren et al. (2012)'s findings of host specificity patterns in temperate environments may correlate with the divergence of the fungal orders

Diaporthales and Helotiales with angiosperms and gymnosperms, respectively, approximately 300 million years ago (Sieber 2007). However, within finer host taxonomic levels (e.g., endophyte among plant congeners), geography may be a stronger predictor of endophyte community composition relative to host evolutionary distance (Davis and Shaw 2008, Davey et al. 2013). In tropical forests, phylogenetic specificity has not been strongly supported at either coarse or fine phylogenetic scales (Pandey et al. 2003, Arnold and Lutzoni 2007, Gilbert et al. 2007, Vincent et al. 2016). For example, the composition of epi- or endophytic fungi in tropical rainforests did not significantly vary with plant species relatedness (Gilbert et al. 2007, Vincent et al. 2016). This suggests that the tropics have widespread endophyte phylogenetic generalism compared to temperate regions, which may be due to higher plant species diversity at lower latitudes (Mittelbach et al. 2007, Kerkhoff et al. 2014). A greater diversity of plant hosts could constrain the ability of specialist endophyte to transition among hosts due to increasing functional differences in host life history, physiology or immunology (Walker et al. 2013). Host diversity, therefore, may be a key determinant for specialist endophyte population dynamics as endophyte generalists would not be constrained by similar adaptive barriers (Poisot et al. 2011).

Host population genetic effects have not been a common focus of endophyte host specificity studies, but Ahlholm et al. (2002) demonstrated that the infection frequency and genotypic diversity of one endophytic species, *Venturia ditricha*, was associated with specific genotypes of birch trees (*Betula pubescens* ssp. *czerepanovii*). Birch genotypes with low *V. ditricha* incidence had higher *V. ditricha* genotypic diversity and vice versa, although this pattern depended on environment (Ahlholm et al. 2002), highlighting variation in host specificity even among host conspecifics. To understand variation in endophyte community

composition within the same plant hosts in different environments or different studies, a common framework for host specificity is needed.

F. Biological and methodological considerations: a comparative case study for structural, network, and phylogenetic specificity

Endophytes are horizontally transmitted through the environment, as opposed to vertical transmission from parent to progeny in graminoid-endophyte symbioses (Class 1 *vs.* Class 3 endophytes; sensu Rodriguez et al. 2009). The degree to which plants can preferentially admit specific endophyte species at the epi- to endofoliar interface is unknown. Therefore, 'accidental tourism', or observations in peripheral hosts (Moran and Southwood 1982; Vega et al. 2010) may be prevalent in endophyte communities. These 'tourist' endophyte may be in low biological abundance within any one sample (Vega 2008; Vega et al. 2010), as they could be poorly adapted to host tissues or outcompeted by the locally adapted microbial community. However, today's NGS-based data are typically highly asymmetric and dominated by rare OTUs. Rare OTUs could be derived from biological (e.g., 'accidental tourism') or methodological (e.g., sequencing error or sequencing depth; Patin et al. 2013) sources. This presents a challenge to quantifying host specificity because rarity highly correlates with specificity, particularly for presence-absence metrics. For example, singleton OTUs will always have the lowest host richness and thus highest structural specificity.

We demonstrate how the three alpha 'dimensions' of host specificity vary as a function of rare OTU exclusion and metric type (presence-absence *vs.* abundance-weighted) using a case study endophyte community taxonomically partitioned into two groups (Ascomycota *vs.* Basidiomycota). The endophyte community was surveyed using NGS (see methodology) where one individual of every co-occurring plant species in five 50 m^2 quadrats were

sampled. Comparisons of host specificity among coarse endophyte taxonomic groups (e.g., phylum, class or order) may reveal key traits that distinguish their association with distantly related host species. For example, cultured Ascomycete endophyte in the classes Dothideomycetes, Sordariomycetes and Leotiomycetes are abundant within the plant families Cupressaceae, Fagaceae and Pinaceae, respectively (Arnold 2007). We chose to compare host specificity between fungal phyla because the scale of our sampling encompassed all co-occurring plant species in a community; however, other taxonomic partitions could also be used (e.g., class-level). Comparisons among lower taxonomic groups may reveal key traits for association with a narrower phylogenetic subset of plant hosts in the community. For example, endophyte within the genus *Lophodermium* (Rhytismataceae, Leotiomycetes) are prevalent within the family Pinaceae (*Pinus*, *Abies* and *Picea* spp.; Stone and Petrini 1997, Ortiz-García et al. 2003), but not other gymnosperm families (e.g., Cupressaceae; Arnold 2007).

Ecological differences between Ascomycota and Basidiomycotan endophyte may lead to predictable differences in their host specificity. For example, endophyte typically have higher representation within the phylum Ascomycota (Arnold 2007; Rodriguez et al. 2009) than Basidiomycota, possibly because Basidiomycete endophytes prefer woody over foliar tissues (Arnold 2007). If rare Basidiomycete endophyte tend to be 'accidental tourists' and are indiscriminately distributed across fewer host species, they should tend to have higher structural and network specificity, but lower phylogenetic specificity relative to Ascomycete endophyte.
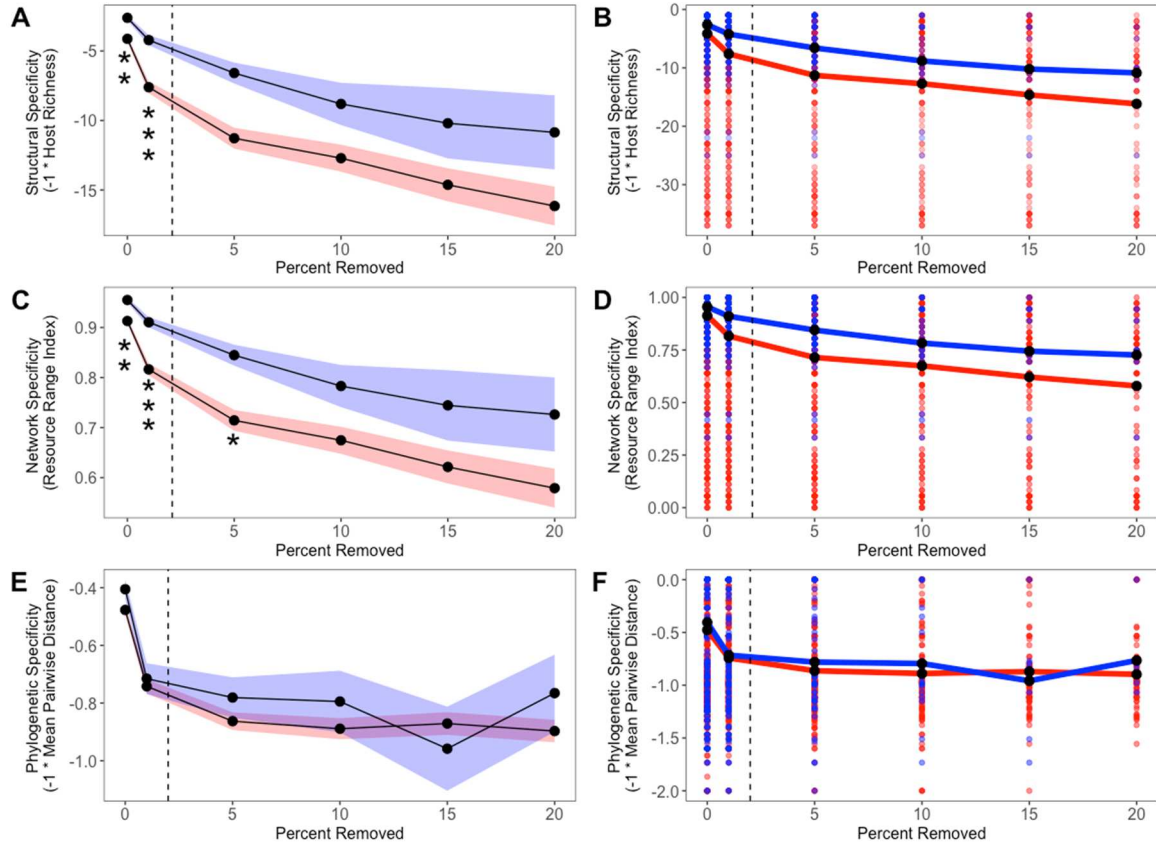
Figure 9. Structural (a-b), network (c-d), and phylogenetic specificity (e-f) measured as a function of rare OTU removal with presence-absence data. (a-b) Host Richness (HR) quantifies the number of hosts per endophyte (equation 1). **(c-d)** The Resource Range Index (RRI) quantifies the host niche per endophyte and is scaled from 0 (perfect generalist) to 1 (perfect specialist; equation 3). **(e-f)** The Mean Pairwise Phylogenetic Distance (MPD) quantifies the mean phylogenetic distance of occupied hosts per endophyte (equation 5). OTUs (structural and network specificity: n = 1,388; phylogenetic specificity: n = 1,342) were rank ordered according to read abundance with reads sequentially removed from the rarest OTUs in 1% and then 5% intervals thereafter. HR and MPD were multiplied by -1, such that more positive values indicate higher host specificity. Host specificity measurements of Ascomycete and Basidiomycete OTUs are color-coded by red or blue, respectively, with shaded regions indicating standard error (panels A, C and E). Colored points represent host specificity measurements for each OTU (panels B, D and F). The vertical dashed line references a commonly used 'rare' OTU threshold of 0.01% relative read abundance (Liu et al. 2015). For panels A, C and E, asterisks correspond to statistical significance (***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$) between the host specificities of Ascomycete and Basidiomycete endophyte OTUs using a Wilcoxon rank sum test.
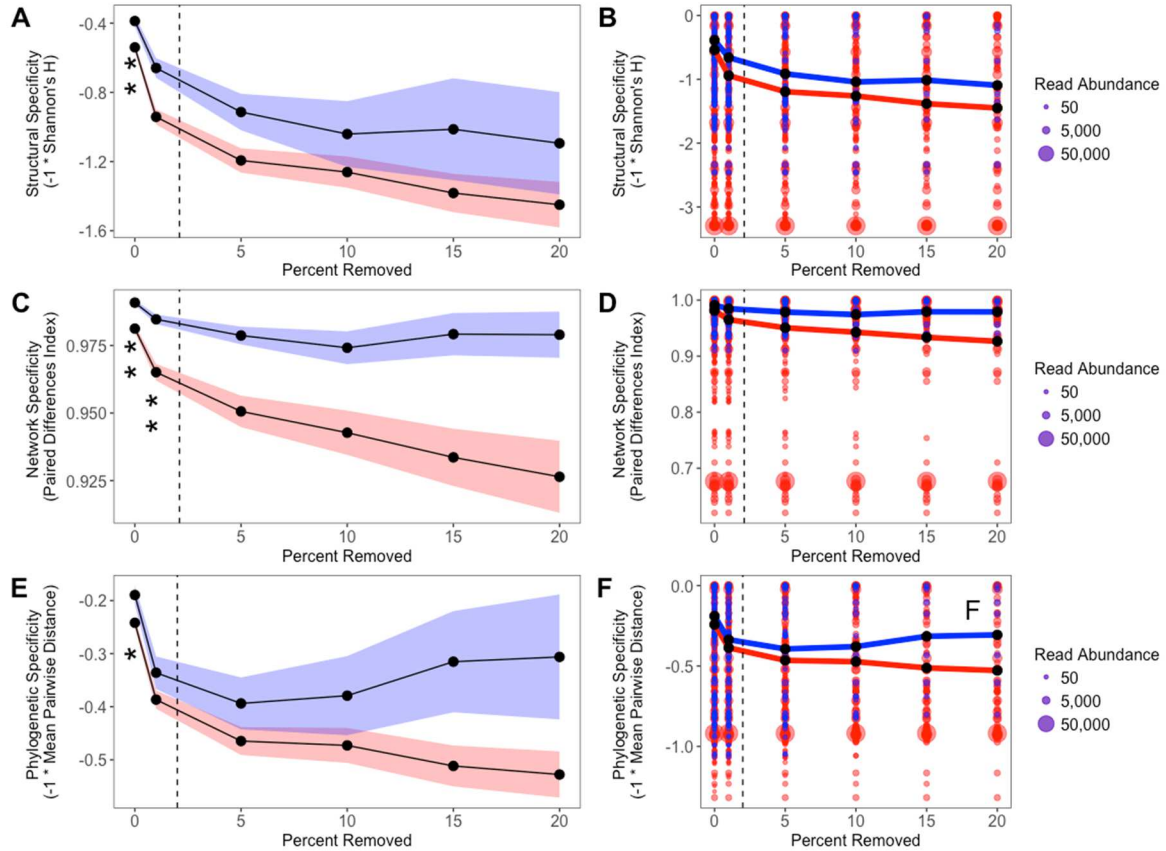
Figure 10. Structural (a-b), network (c-d) and phylogenetic specificity (e-f) measured as a function of rare OTU removal with abundance-weighted data. (a-b) Shannon's *H* (SH) quantifies host richness and endophyte evenness within hosts per endophyte (equation 2). **(c-d)** The Paired Differences Index (PDI) quantifies differential abundance among all hosts in the community and is scaled from 0 (perfect generalist) to 1 (perfect specialist; equation 4). **(e-f)** The abundance-weighted variant of the Mean Pairwise Phylogenetic Distance (MPD) quantifies the mean phylogenetic distance of occupied hosts per endophyte with greater weight given to hosts where a given endophyte is more abundant (equation 6). OTUs (structural and network specificity: n = 1,388; phylogenetic specificity: n = 1,342) were rank ordered according to read abundance with reads sequentially removed from the rarest OTUs in 1% and then 5% intervals thereafter. SH and MPD were multiplied by -1, such that more positive values indicate higher host specificity. Host specificity measurements of Ascomycete and Basidiomycete OTUs are color-coded by red or blue, respectively, with shaded regions indicating standard error (panels A, C and E). Circles represent host specificity measurements for each OTU and are size-scaled by total read abundance of the OTU (panels B, D and F). The vertical dashed line indicates a commonly used 'rare' OTU threshold of 0.01% relative read abundance (Liu et al. 2015) for reference. For panels A, C and E, asterisks correspond to statistical significance (***: p < 0.001, **: p< 0.01, *: p < 0.05) between the host specificities of Ascomycete and Basidiomycete endophyte OTUs using a Wilcoxon rank sum test.

Our case community showed similar structure to other endophyte communities where Basidiomycete endophyte OTUs occurred less frequently compared to Ascomycete endophyte OTUs (structural and network specificity: 301 *vs*. 1087 OTUs, 21.7% *vs*. 78.3%; phylogenetic specificity: 289 *vs*. 1053 OTUs, 21.5% *vs*. 78.5%) and a had lower mean read abundance per OTU (structural and network specificity: 106.37 ± 15.12 *vs*. 423.66 ± 89.03; phylogenetic specificity: 90.29 ± 13.81 *vs*. 354.42 ± 77.36; mean ± standard error). The complete Basidiomycete endophyte community consistently displayed higher structural, network and phylogenetic specificities across presence-absence and abundance-weighted metrics, relative to the complete Ascomycete endophyte community (host specificity for 0% removed values in Figures 9 and 10). This suggests that, on average, Basidiomycete endophyte OTUs are more asymmetrically distributed (Figures 10a-d) across fewer hosts (Figures 9a-d) that comprise a narrower phylogenetic breadth (Figure 9e-f and 10e-f) than Ascomycete endophytes. Even though Basidiomycete endophytes tend to be rarer, these OTUs are not necessarily more likely to be 'accidental tourists' than Ascomycete endophytes since they occupy a relatively narrower host phylogenetic breadth (Figures 10e-f).

For all 'dimensions' of host specificity, sequence exclusion from rare endophyte OTUs decreased average host specificities and increased variances within phyla. The inclusion of rare OTUs made the differences between the host specificities of the two phyla more pronounced (Figure 9a-f and 10a-f). However, the statistical differences between these groups cannot define an exclusion threshold for rare endophyte because the proportion of rare taxa that are biologically informative is unknown. For example, in our case study (Figures 9-10), if many rare OTUs were products of sequencing error, their inclusion would

incorrectly indicate that the host specificities of Ascomycota and Basidiomycota are significantly different (type I error). Alternatively, if the majority of rare OTUs represented highly host-specific endophyte, their exclusion would incorrectly suggest that their host specificities are not significantly different (type II error). We do not outline an explicit rare OTU exclusion threshold as every dataset can vary depending on the biological community, sequencing platform, bioinformatic processing and normalization method, for example. However, our case study highlights how balancing the exclusion of sequencing artifacts with the inclusion of biologically informative sequences will affect ecological inference. To investigate the probability of observed host specificity occurring by random chance, we compared our empirical results to that of a model of random community assembly (see section 7).

Each 'dimension' of host specificity is useful for understanding different ecological or evolutionary aspects of plant-endophyte assemblages. In this case study, we provided examples of structural, network and phylogenetic specificity from the endophyte perspective. Alternatively, from the plant perspective, host specificity can vary depending on the composition of the endophyte community among plant species or functional groups. For example, endophyte communities could also be partitioned by plant functional traits (woody *vs*. herbaceous plants), plant taxonomy (angiosperms *vs*. gymnosperms) or life history traits (annual *vs*. perennial). When comparing different plant communities, we suggest randomly and uniformly sub-sampling each community (e.g., 10 hosts per group), generating a distribution of values (e.g., 50 values of host specificity) and comparing statistical significance among these simulated distributions because differences in the number of plant hosts considered can affect observed host specificity.

61

G. Relationships among host specificity dimensions

We evaluated rank correlations (Spearman's rho) between each of the three alpha 'dimensions' of host specificity from our case community to better demonstrate each of their unique and complementary features. We then evaluated how host specificity varied as a function of read abundance (log-scaled; Figures 11-12).
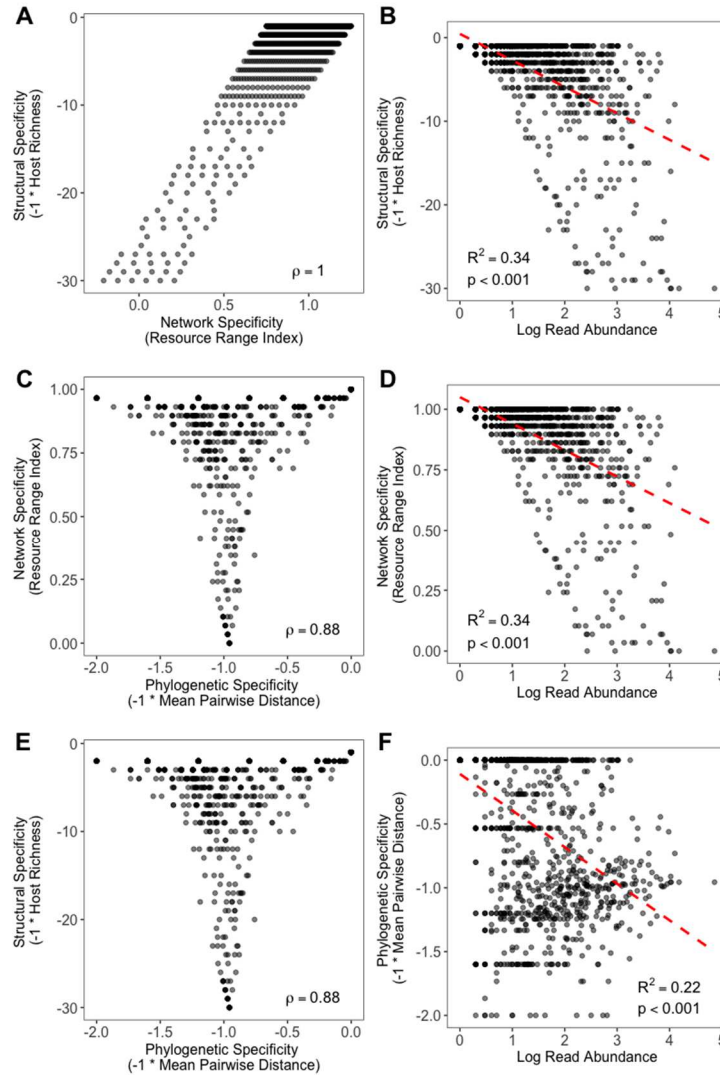
Figure 11. Correlations among alpha-host specificities with presence-absence data (a, c, and e) and between host specificities and read abundances per OTU (b, d, and f). Panels a, c, and e compare the three alpha-host specificities per endophyte OTU (n = 1,360) and report their rank correlation coefficients, Spearman's rho (ρ). Panels b, d, and f compare host specificity values to read abundance (log-scaled) per endophyte OTU (n = 1,360) and

report the coefficients of determination and p-values. The points in panel A are jittered for clarity.
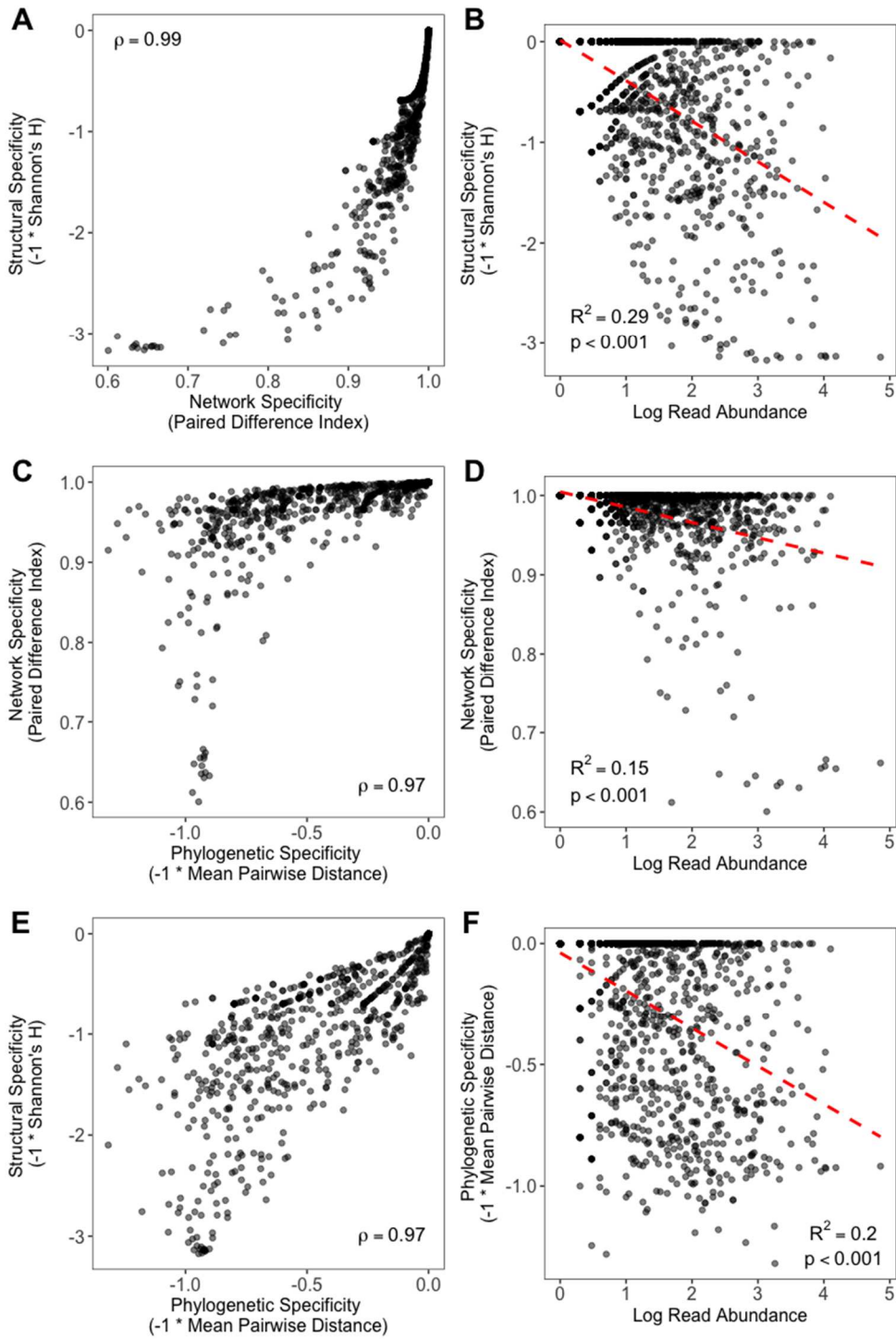
Figure 12. Correlations among alpha-host specificities with abundance-weighted data (a, c, and e) and between host specificities and read abundances per OTU (b, d, and f). Panels a, c, and e compare the three alpha-host specificities per endophyte OTU (n = 1,360)

and report their rank correlation coefficients, Spearman's rho (ρ). Panels b, d, and f compare the host specificity values to read abundance (log-scaled) per endophyte OTU (n = 1,360) and report the coefficients of determination and p-values.

Structural and network specificity were perfectly (Host Richness, HR *vs.* Resource Range Index, RRI; ρ = 1; Figure 11a) or highly correlated in rank (Shannon's *H*, SH *vs.* Paired Difference Index, PDI; ρ = 0.99; Figure 12a) based on presence-absence or abundance-weighted metrics, respectively. These two 'dimensions' of host specificity were highly correlated because the number of occupied hosts is always inversely related to the number of unoccupied hosts per endophyte OTU. Rarer OTUs tended to have higher structural and network specificity than more abundant OTUs, but the variances in SH and PDI were explained less by the read abundance compared to the variances in HR or RRI (Figures 11b *vs.* 12b and 11d *vs.* 12d).

The read abundance of an OTU was less likely to explain the variance in phylogenetic specificity (Mean Pairwise Phylogenetic Distance; MPD) compared to the variance in structural or network specificity for presence-absence metrics (Figures 11b and 11d *vs.* 11f). The variance in presence-absence MPD may be the least explained by read abundance because rare endophyte could still occupy distantly related hosts. For abundance-weighted metrics, the variance in network specificity was less likely to be explained by read abundance compared to structural or phylogenetic specificity (Figures 12b and 12f *vs.* 12d), which could be due to the properties of PDI (equation 4). Although host specificity and read abundance were more negatively correlated for SH and MPD (slope of regression lines; Figure 12b and 12f), even OTUs with relatively moderate read abundances tended to be highly host-specific for PDI because they were unevenly distributed within their hosts.

Positive correlations between phylogenetic specificity and structural or network specificity (Figures 11c, 11e, 12c and 12e) suggest that endophyte OTUs that occupy a broader host phylogenetic breadth tended to be more evenly distributed across more hosts, whereas endophyte OTUs that are restricted within a narrower host phylogenetic breadth tended to be more asymmetric in abundance in fewer hosts. This pattern could occur due to methodological or biological reasons. For example, when an endophyte occupies only one host (i.e., highest HR and RRI values) or is sequenced so rarely that it can only be found in one host (i.e., high SH and PDI values), its phylogenetic specificity is the highest possible value (i.e., zero). Host richness has been shown to be less correlated with MPD relative to other phylogenetic distance metrics, such as Faith's phylogenetic distance (Swenson 2014). However, phylogenetic specificity is still a scale-dependent metric because the variance of MPD decreases predictably with increasing HR (Figure 11e).

One possible solution to address this scale-dependence is to calculate the standardized effect size of the mean pairwise host phylogenetic distance (SES.MPD) per endophyte OTU. SES.MPD is calculated by taking the difference between the observed MPD and the mean MPD of a set of randomized communities and dividing this difference by the standard deviation of the random set of MPDs (Swenson 2014). The random set of MPDs is calculated from a randomized set of ultrametric phylogenetic trees at a given level of HR (Swenson 2014) that are generated by either shuffling tip labels, abundances within samples or abundances within species (see the "ses.mpd" function in the *picante* R package for more details; Kembel et al. 2010). However, technical constraints of a randomization method can preclude its use. For presence-absence data, if any endophyte OTU occurs in all sampled hosts (e.g., the most abundant OTU in our case community), a randomization method that

randomizes tree tip labels or OTUs within samples will have the same MPD for all randomized ultrametric trees and therefore, have a standard deviation of zero. Alternatively, for abundance-weighted data, the randomized phylogenies will always have a standard deviation greater than zero because all randomization methods change how the phylogenetic distance between hosts i and j are multiplied with respect to the abundance in any given cell (equation 6). We suggest utilizing SES.MPD to account for the scale-dependence of phylogenetic specificity only if the randomized distributions of the most cosmopolitan OTUs remain statistically meaningful.

H. Comparisons to a random community assembly model

NGS-based community surveys with high numbers of rare OTUs could be particularly vulnerable to biased ecological inferences of these 'dimensions' of host specificity since host specificity is highly correlated with the number of rare OTUs. The technical constraints of each metric make it necessary to evaluate the probability that the observed patterns arose due to random chance. We calculated the probability that our observed host specificity values were different from that of randomized communities (see Methods). We compared the host specificity values per endophyte OTU for the full dataset of the case community (i.e., not partitioned by fungal phyla) to the averaged host specificity per OTU across 1,000 randomized tables. We evaluated statistical significance between empirical and randomized communities (each with an equal number of host specificity values and OTUs) with a two-sample test using the non-parametric bias-corrected and accelerated bootstrap method (Efron 1987) with 9,999 bootstrap replicates. We repeated the randomization and bootstrap tests as we incrementally excluded rare OTUs, as in previous analyses.
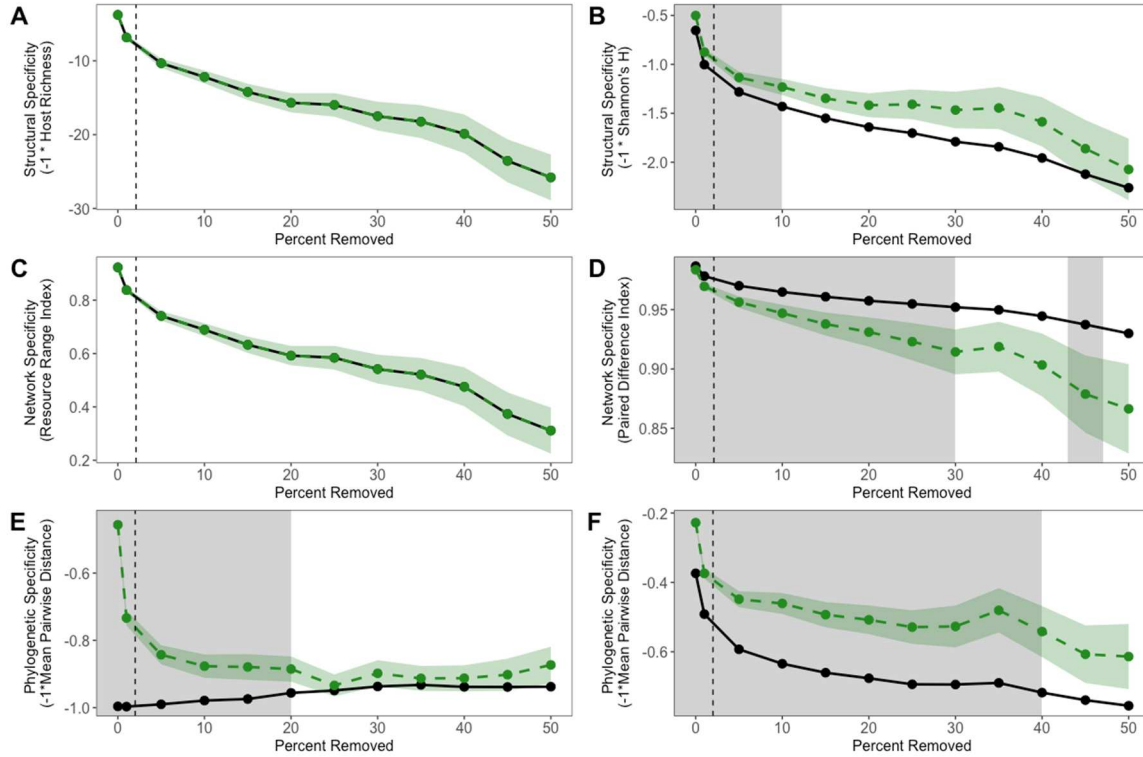
Figure 13. Structural (a-b), network (c-d), and phylogenetic specificity (e-f) measured as a function of rare OTU removal with presence-absence (a, c, e) and abundance-weighted (b, d, f) data for empirical and randomized communities. OTUs (structural and network specificity: n = 1,406; phylogenetic specificity: n = 1,360) were rank ordered according to read abundance with reads sequentially removed from the rarest OTUs in 1% and then 5% intervals thereafter. The green dotted and black solid lines represent host specificity averages for the empirical and randomized communities, respectively. For the randomized communities, the mean host specificity value per OTU was averaged across 1,000 randomized OTU tables. Shaded regions indicate standard error (minimal for OTUs from the randomized datasets). The vertical dashed line indicates a commonly used 'rare' OTU threshold of 0.01% relative read abundance (Liu et al. 2015) for reference. Shaded and white backgrounds correspond to statistical significance at $p < 0.05$ and $p > 0.05$, respectively, between the empirical and randomized datasets using the non-parametric bias-corrected and accelerated bootstrap method (Efron 1987) with 9,999 bootstrap replicates.

Excluding rare endophyte OTUs (e.g., >10% of total sequences; Figure 13a-f) resulted in specificity values that were not statistically significant from those of randomized communities. Thresholds for statistical power as a function of rare OTU removal varied among the host specificity 'dimensions' with abundance-weighted structural and phylogenetic specificity having the lowest and highest thresholds for statistical power

(Figure 13b and 13f), respectively. Abundance-weighted data has a greater capacity to be structurally distinct from randomized communities (i.e., there are more possible ways to restructure the community), relative to presence-absence data. This could explain why abundance-weighted phylogenetic specificity was less sensitive to rare OTU removal compared to presence-absence phylogenetic specificity (Figure 13e *vs*. 13f). The empirical distributions for presence-absence structural and network specificity completely overlapped with randomized distributions (Figure 13a and 13c) because connectance (i.e., the proportion of zeros to counts, or the number of realized interactions) was constrained during randomization. The lack of statistical significance at any removal threshold is a technical constraint of the randomization method which resulted in identical host specificity averages between the empirical and randomized communities (Figure 13a and 13c). Utilizing a different randomization method that either adds or subtracts counts ad hoc (i.e., altered connectance) could result in statistically meaningful randomized communities.

OTUs with low read abundance are common among NGS data (Brown et al. 2015), such that low thresholds for OTU removal based on relative abundance results in a disproportionately large number of OTUs to be removed. In Figure 13, the removal of reads with less than 0.01% relative abundance, which is a commonly utilized threshold, removed 70.7% of OTUs (994 of 1,406) for structural and network specificity and 68.5% of OTUs (932 of 1,360) for phylogenetic specificity. It is unlikely that commonly accepted filtering techniques, such as rarefying or relative abundance thresholds, would remove enough sequences that would prevent distinction from a randomized community (white areas; p > 0.05; Figure 13). However, comparisons with randomized communities highlight how statistical power can vary widely depending on data-structure and the host specificity metric.

68

We only applied one randomization method here but suggest comparing the structure of the empirical community with communities from multiple randomization methods because the type of randomization (e.g., constrained marginal totals or connectance) can critically alter the structure of a simulated community.

I. Beta specificity

Structural, network and phylogenetic specificity describe host specificity at the local scale, analogous to the scale of alpha-diversity (Whittaker 1972), whereas beta specificity describes host specificity at a scale analogous to the regional scale of beta diversity (Poulin et al. 2011). Beta specificity, or turnover in host specificity, can quantify the degree to which endophyte host specificity varies across space (e.g., the geographic range of a given plant host) or time (e.g., disturbance or host ontogeny; Krasnov et al. 2011). For example, a given endophyte would display higher beta specificity if it was found in the same plant host at different localities relative to another endophyte that occupied completely different plant hosts at different localities (Figure 14). A measure of host specificity across multiple scales is important to consider as the abundance of the host and environmental context can be highly variable at regional scales, but homogenous at local scales. Significant endophyte community turnover across the geographic range or developmental stages of a given host could reveal important features of endophyte biology such as dispersal limitation or adaptation to particular host life stages. We outline a beta specificity metric across replicate sampling sites as an example (Figure 14), but this metric could be applied across other contexts, such as sites that vary in plant community structure or multiple stages during plant development.

69
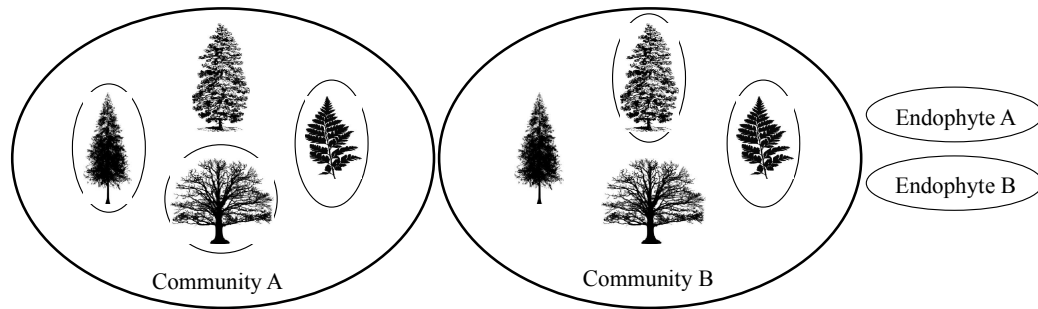
Figure 14. Conceptual diagram of beta specificity. The realized ecological niche of each endophyte within two replicate plant communities is represented by colored circles around a plant host. Endophyte 'A' occupies completely different plant hosts in the two communities, such that beta specificity equals zero. Endophyte 'B' occupies the same plant host in both communities, such that beta specificity equals one. Plant species images were procured from a public domain illustration repository, https://openclipart.org.

We propose a multiple-site similarity measure recommended by Poulin et al. (2011), which was derived from the Sørensen similarity index, to quantify turnover in host specificity. Instead of using an averaged similarity among all pairwise comparisons of communities (Ricotta and Pavoine 2015), this multiple-site similarity metric quantifies the consistency of the interaction by preserving the identity of hosts across two or more localities (Diserud and Odegaard 2007). This presence-absence multiple-site similarity measure is scaled from 0 to 1, such that 1 indicates an endophyte that occupies the same plant hosts in all considered localities (high beta specificity) and 0 indicates an endophyte that occupies completely different plant hosts in all considered localities (low beta specificity). We recommend the original beta specificity equation from Diserud and Odegaard (2007), such that more positive values indicate higher beta specificity (but see the modification in Poulin et al. 2011). T is the number of localities under consideration, R is the total number of plant hosts occupied by a given endophyte across all T localities and r is the total number of plant hosts occupied at locality i (Diserud and Odegaard 2007).

$$\text{Multiple site similarity} = \frac{T}{T-1}\left(1 - \frac{R}{\sum_{i=1}^{T} r_i}\right) \quad (7)$$

J. Conclusions

Any 'dimension' of endophyte host specificity depends on methodological approaches (e.g., sampling, bioinformatic pipeline) as well as the scale of the study (e.g., phylogenetic, spatial). The removal of rare OTUs is often recommended before proceeding with ecological analyses because it may be statistically inappropriate to assess specificity from low abundance OTUs (Nguyen et al. 2014, Brown et al. 2015). The exclusion of biologically uninformative sequences that are confounded with specificity could be removed relative to a read abundance threshold, a positive control (e.g., Nguyen et al. 2014) or a uniform sequencing depth per sample (but see McMurdie and Holmes 2014). We do not outline an explicit rare OTU exclusion threshold as every NGS dataset can vary in structure due to different library preparation methods (e.g., sample collection and primers) or sequencing platforms (e.g., errors and depths). Rare OTU exclusion should be done with explicit rationale because removal can have significant effects on ecological inference (Figures 9 and 10) and statistical power (Figure 13).

Structural specificity metrics are useful within individual studies with uniform sampling designs. However, endophyte with similar structural specificities in different localities could have significantly different network specificity depending on plant community diversity (Figure 8). Network specificity metrics are more comparable across studies because they quantify how the distribution of an endophyte species varies among all sampled hosts in a plant community. endophyte could also occupy similar proportions of the plant community but differ in the phylogenetic composition of hosts they occupy. Phylogenetic specificity can

71

quantify the distribution of an endophyte species relative to the evolutionary relationships among occupied hosts. Finally, the biogeographical approach of beta specificity has important implications for global fungal biodiversity measurements.

Endophytes have been suggested as an indicator group for global fungal biodiversity due to their prevalence (Suryanarayanan 2011). Estimates of endophyte diversity from Dreyfuss and Chapela (1994) and others (Fröhlich and Hyde 1999, Arnold and Lutzoni 2007) suggest that the commonly referenced 1.5 million species estimate, based on a uniform plant-to-fungal ratio (e.g., 1:6; Hawksworth 2001), for global fungal biodiversity is a vast underestimate. Accounting for spatial variability in endophyte host specificity can provide more accurate fungal biodiversity estimates through the use of correction factors. For example, high endophyte host specificity has been suggested to characterize boreal forest biomes (Arnold and Lutzoni 2007). This would produce high endophyte beta diversity among plant host species. Higher heterogeneity in endophyte community composition among plant hosts could increase regional diversity (i.e., gamma diversity) depending on the spatial scale under consideration and within host diversity (i.e., alpha diversity). Alternatively, in tropical ecosystems, generalism may play a predominant role in structuring diversity (Arnold and Lutzoni 2007; Scheuling et al. 2012). Low host specificity would produce lower beta diversity in endophyte community composition among plant species dampening gamma diversity at larger regional scales. Hence, a uniform plant-to-fungal ratio underestimates biodiversity in areas of high host specificity, but overestimates biodiversity in areas of low host specificity. These predictions demonstrate how understanding alpha and beta 'dimensions' of host specificity in foliar fungal endophytes can have significant effects on future assessments of global fungal biodiversity.

K. Methods

Sampling Design

Plant samples were collected from the James San Jacinto Reserve (University of California, Natural Reserve System; 33° 48' 29", -116° 46' 36") in July of 2016 from five randomly placed 50 m^2 quadrats within a 3 km^2 sampling range. The canopy of this forest was dominated by *Pinus ponderosa* and *Quercus kelloggii*. Sampling was designed to maximize plant breadth within quadrats with more abundant hosts sampled across quadrats. Ten leaves or 5 shoots from one individual of every co-occurring plant species (n = 79 plant samples, n = 37 plant species) were collected and surface sterilized as in Oono et al. (2015).

Molecular Methods and Sequencing

DNA was extracted with a modified 2% CTAB method (Branco et al. 2015) and the internal transcribed spacer 1 of the ribosomal DNA (ITS1 rDNA) was amplified with ITS1F-KYO1 and ITS2-KYO1 (Toju et al. 2012) primers modified with Illumina overhang adapters. Samples were sequenced on the Illumina MiSeq (Genomics Center, Institute for Genomic Medicine, UC San Diego) with 250 paired-end reads for 500-cycles. Reads were processed according to the UNOISE pipeline (Edgar 2016) and clustered at 97% sequence similarity with USEARCH (version 9.2.64). Sequences were deposited in GenBank with SRA BioProject Accession Number: PRJNA356423. OTUs were taxonomically assigned with BLAST+ (Camacho et al. 2009) and non-fungal OTUs were removed with MEGAN (Huson et al. 2007) and QIIME (Caporaso et al. 2010). Reads from replicate plant species across quadrats were summed and sequencing depth was normalized by rarefying to 13,322 reads per plant species. We considered only Ascomycete and Basidiomycete OTUs for Figures 9-10 (structural and network specificity: n = 1,388; phylogenetic specificity: n =

1,342), only OTUs shared among the structural, network and phylogenetic specificity

datasets for Figures 11-12 (n = 1,360) and all OTUs for Figure 13 [structural and network

specificity: total OTUs (n = 1,406), Ascomycota (n = 1,087), Basidiomycota (n = 301),

Chytridiomycota (n = 6), Mucoromycota (n = 6), Neocallimastigomycota (n = 5) and

Zoopagomycota (n = 1); phylogenetic specificity: total OTUs (n = 1,360), Ascomycota (n =

1,053), Basidiomycota (n = 289), Chytridiomycota (n = 6), Mucoromycota (n = 6),

Neocallimastigomycota (n = 5) and Zoopagomycota (n = 1)].

Host Species Phylogeny

Sequences from host conspecifics or congeners were procured from GenBank for

phylogenetic analyses (rbcl accession numbers: AB063374.1, KJ773371.1, AY300097.1,

KX582009.1, GU135146.1, AF297134.1, JX258357.1, KM980628.1, KM372993.1,

Z37457.1, JN033544.1, KC237117.1, HM024269.1, KF683137.1, KT178128.1,

JN847834.1, AY497224.1, KC482774.1, HQ600457.1, KX679216.1, AB029648.1,

JF940720.1, GU176649.1, JF940720.1, FJ548255.1, D88906.1, KM003101.1, KX371919.1,

KJ841515.1, JF944117.1, JN681689.1; matk accession numbers: AB080924.1, KC539612,

FR865060.1, HQ593182.1, GU134983.1, KJ772764.1, JN895143.1, HQ593309.1,

KM372683.1, JX981412.1, JN033545.1, KJ028426.1, AF152178.1, LM652873.1,

KT176610.1, JN585004.1, EF546716.1, KC474725.1, HQ600036.1, AY386910.1,

APC92700.1, KC474077.1, KC473972.1, FJ548086.1, HM850737.1, KM002235.1,

EU628517.1, KC290085.1, KJ840981.1, JF956157.1, JF729129.1). Host species that did not

have rbcl and matk sequences available in GenBank were excluded from phylogenetic

specificity analyses (7 of 37 host species). Sequences were aligned in MAFFT (Katoh et al.

2002) and trimmed with trimAl (Capella-Gutiérrez et al. 2009) using the phyloGenerator

platform (Pearse and Purvis 2013). The phylogenetic tree was constructed in RAxML (Stamatakis 2014) with bootstrapped nodal support (1,000 permutations) with default parameters.

Sequence Removal

Truncated data sets were generated with Multivariate Cutoff Level Analysis (MultiCoLA; Gobet et al. 2010). OTUs (structural and network specificity: n = 1,406; phylogenetic specificity: n = 1,360) were rank ordered according to read abundance with reads sequentially removed from the rarest OTUs in 1% and then 5% intervals thereafter. One percent sequence removal corresponded to 4,948 of 492,914 reads for structural and network specificity and 4,010 of 399,660 for phylogenetic specificity. Five percent removal corresponded to 24,781 of 492,914 reads for structural and network specificity and 20,002 of 399,660 reads for phylogenetic specificity. The 0.01% relative abundance threshold corresponded to 10,023 of 492,914 reads or 2.0% sequence removal (994 of 1,406 OTUs; 70.7% OTUs removed) for structural and network specificity and 8,000 of 399,660 reads or 2.0% sequence removal (932 of 1,360 OTUs; 68.5% OTUs removed) for phylogenetic specificity. For Figures 9 and 10, only Ascomycete (structural and network specificity: n = 1,087; phylogenetic specificity: n = 1,053) and Basidiomycete (structural and network specificity: n = 301; phylogenetic specificity: n = 289) OTUs were included in the analysis. For Figure 13, all OTUs (structural and network specificity: n = 1,406; phylogenetic specificity: n = 1,360) were included in the analysis.

Host Specificity Metrics

All host specificity metrics were calculated in R (version 3.4.2). Structural specificity was calculated using the 'specnumber' and 'diversity' functions in the vegan package

(Oksanen et al. 2017). Network specificity was calculated using the 'PDI' function in the

bipartite package (Dormann et al. 2008; Dormann 2011). Phylogenetic specificity was

calculated using the 'mpd' function in the picante package (Kembel et al. 2010). Structural

and phylogenetic specificity were negated (multiplied by -1), such that more positive values

indicated higher host specificity. For structural and network specificity, we chose the

randomization method 'shuffle.web' within the 'nullmodel' function in the bipartite package

(Dormann et al. 2009; Dormann 2011). This randomization method redistributed all

abundance data among OTUs and hosts, thereby changing marginal totals, but maintained

connectance (i.e., the OTU table has same number of zeros and values). For phylogenetic

specificity, we used the randomization method 'frequency' within the 'ses.mpd' function in

the picante package (Webb et al. 2008), which randomized abundances within OTUs.

Rank Correlations between Structural, Network and Phylogenetic Specificity

For Figures 11 and 12, we used 1,360 OTUs that were shared among the structural,

network and phylogenetic specificity datasets. We calculated the rank correlation

coefficient, Spearman's rho (ρ), among the three types of alpha specificity and two types of

data structure (presence-absence *vs*. abundance-weighted). We then regressed observed host

specificity as a function of read abundance (log-scaled).

Comparisons to a Random Community Assembly Model

Using the above functions, we generated 1,000 randomized community matrices for the

two types of data structure (presence-absence *vs*. abundance) and three alpha 'dimensions'

of host specificity. Because host specificity was quantified per endophyte and not as one

total community metric (unlike for network-level metrics), we calculated host specificity per

endophyte OTU across each of the 1,000 randomized tables and then averaged the 1,000

76

host specificity measurements per endophyte OTU. We then compared the randomized and empirical communities (n = 1,406 OTUs each) with a non-parametric two-sample test using the bias-corrected and accelerated bootstrap method (Efron 1987) implemented with the 'boot.two.bca' function in the package wBoot (Weiss 2016) with 9,999 bootstrap replicates. Resampling occurred with replacement and resampling size was consistent with the size of the original population.

Data Availability

Supplementary material and reproducible code for bioinformatic processing, sequence removal and host specificity measurements can be accessed on Github at https://github.com/austenapigo.

III. Chapter 3: Plant abundance, not plant evolutionary history, shapes patterns of host specificity in foliar fungal endophytes

This article is published as Apigo A and Oono R. 2022. Plant abundance, but not plant evolutionary relationships, shape host specificity in foliar fungal endophytes. Ecosphere 3(1): e03879.

A. Abstract

Understanding the origins and maintenance of host-specificity, or why horizontally acquired symbionts associate with some hosts but not others, remains elusive. In this study, we explored whether patterns of host specificity in foliar fungal endophytes, a guild of highly diverse fungi that occur within the photosynthetic tissues of all plant lineages, were related to characteristics of the plant community. We comprehensively sampled all plant host species within a single community and tested the relationship between plant abundance or plant evolutionary relatedness and metrics of endophyte host specificity. We quantified host specificity with methods that considered the total endophyte community per plant host (i.e., multivariate methods) along with species-based methods (i.e., univariate metrics) that considered host specificity from the perspective of each endophyte. Univariate host specificity metrics quantified plant alpha diversity (structural specificity), plant beta diversity (beta specificity) and plant phylogenetic diversity (phylogenetic specificity) per endophyte. We standardized the effect sizes of univariate host specificity metrics to randomized distributions to avoid spurious correlations between host specificity metrics and endophyte abundance. We found that more abundant plant species harbored endophytes that occupied fewer plant species (higher structural specificity) and were consistently found in the same plant species across the landscape (higher beta specificity). There was no

relationship between plant phylogenetic distance and endophyte community dissimilarity. We still found that endophyte community composition significantly varied among plant species, families and major groups supporting a plant identity effect. In particular, endophytes in angiosperm lineages associated with narrower phylogenetic breadths of plants (higher phylogenetic specificity) compared to endophytes within conifer and fern lineages. We identified an effect of plant abundance, not plant evolutionary history, on patterns of endophyte host specificity that may help explain why horizontally transmitted endophytes are not necessarily fixed within particular host lineages.

B. Introduction

Microbial symbionts are hypothesized to be associated with all plants and animals (Rosenberg and Zilber-Rosenberg 2018) and their host specificity, or the degree to which they are restricted within a host community (Christian et al. 2015, Apigo and Oono 2018), is a defining feature of these relationships. High-throughput sequencing has increased the rate and resolution at which microbial symbionts have been characterized, often identifying associations between hosts and symbionts that are new to the scientific community (e.g., Lian et al. 2019, Vohník et al. 2019). As a result, biologists are increasingly faced with the challenge of interpreting datasets that usually lack basic ecological and evolutionary information about host-symbiont relationships (Peay 2014, Christian et al. 2015). Quantifying measures of host specificity provides a reasonable first step to identifying the ecological and evolutionary constraints that limit symbiont distribution. In this study, we explored the relationships between plants and their cryptic fungal symbionts to understand whether features of the host community could explain host-specific patterns.

Foliar fungal endophytes (hereafter referred to as 'endophytes') are a commonly found, but cryptic and understudied, guild of Fungi. They occur asymptomatically within the photosynthetic tissues of all plant lineages and are primarily acquired by horizontal transmission (Arnold 2007, Rodriguez et al. 2009) where spores are dispersed by the environment, land on leaf surfaces and cause localized infections within leaves (Herre et al. 2007, Wang et al. 2009). Studies of their distribution among plants have repeatedly shown that endophyte community composition varies by geography (Zimmerman and Vitousek 2012, Langenfeld et al. 2013, Oono et al. 2017, Whitaker et al. 2018), plant identity (Zhang and Yao 2015, Vincent et al. 2016, Liu et al. 2019, U'Ren et al. 2019) or both (Hoffman and Arnold 2008, Christian et al. 2016, Whitaker et al. 2020). However, these foundational observations have yet to be explored with sampling designs that explicitly account for host community structure.

Plant species abundance remains understudied in the context of endophyte host specificity. Many endophyte studies restrict sampling to the most dominant plant taxa that occur across a study range as an experimental control for host identity across environmental gradients or geographic distance (e.g., Zimmerman and Vitousek 2012, Zhang and Yao 2015) and consequently, do not sample rare plant species. U'Ren et al. (2019) surveyed the boreal biome and concluded that the similarity of endophyte communities was limited more by the availability of hosts endophytes could colonize than by geographic distance. This suggests that host abundance could predict whether a host-specific interaction is more likely to occur. Low host population density as a barrier to host specificity has been described as the Resource Fragmentation Hypothesis in host-parasite systems (Janzen and Pond 1975) or the Common Host Hypothesis (CHH; Kindlmann et al. 2007) for host-specific interactions. For endophytes, these hypotheses predict that endophytes are more host-specific to plant

species that occur abundantly across the landscape because passively dispersed symbionts can reliably recolonize them and evolve host-specific interactions.

Furthermore, few studies have evaluated how endophytes vary in host specificity among closely- and distantly related plant species (but see Arnold and Lutzoni 2007, Vincent et al. 2016, U'Ren et al. 2019). Differences in endophyte composition among distantly related plant species may occur if phylogenetically conserved plant traits act as a biotic filter to endophyte colonization and establishment. However, a plant identity effect on endophyte community composition could be due to filtration by leaf traits not correlated with host evolutionary history. For example, a study by González□Teuber et al. (2020) demonstrated that chemical and physical leaf traits independently distributed with respect to the evolutionary relatedness of tropical tree species in Southern Chile were significantly correlated with endophyte community composition.

A common way to test for an effect of host evolutionary history is to quantify the correlation between symbiont compositional dissimilarity and host phylogenetic distance, or their degree of *phylosymbiosis* (Brooks et al. 2016, Lim and Bordenstein 2020, Kohl 2020, O'Brien et al. 2020). For example, Liu et al. (2019) and U'Ren et al. (2019) found evidence for phylosymbiosis demonstrating that more closely related plants harbored more similar endophyte communities at local to intercontinental scales. Alternatively, plant phylogenetic distance may not be predictive of endophyte community composition, but endophyte associations can still be non-randomly distributed with respect to the plant phylogeny. *Phylospecificity* (Poulin et al. 2011, Cooper et al. 2012, Huang et al. 2014, Clark and Clegg 2017), or the tendency of symbionts to occupy more closely related hosts than expected by chance, shapes differences in endophyte communities that may not be correlated to plant phylogenetic distance. For example, Whitaker et al. (2020) did not find evidence for

phylosymbiosis because the genetic distances among 18 Asteraceae hosts were not correlated with endophyte community dissimilarity but did find evidence for phylospecificity, or a plant identity effect, where conspecific hosts tended to have more similar endophyte communities than heterospecific ones. Here, we tested for both phylosymbiosis, or the effect of phylogenetic distance on endophyte community dissimilarity, and phylospecificity, the non-random distribution of endophyte associations with respect to the plant phylogeny, to understand how endophyte host specificity varies with plant phylogenetic distance.

In this study, we asked the following questions: (1) Do common plant species harbor more host-specific endophytes? (2) Are patterns of endophyte host specificity reflective of plant phylogenetic relationships? To address these questions, we designed a field survey that sampled and recorded the abundance of every co-occurring plant species within a plant community (3 km^2 sampling area) and characterized the endophyte community with culture-independent techniques. Current microbial ecology studies almost exclusively infer host specificity by the degree of compositional similarity of symbiont communities per host in multivariate ordination space (e.g., Chaib De Mares et al. 2017, Maurice et al. 2021, Sierra et al. 2020). Specialist symbionts that are consistently found within the same host species, but not others, are expected to increase compositional similarity among hosts of the same species and increase dissimilarity among different host species. Few studies utilize univariate approaches measured per symbiont (e.g., the number of host species a symbiont occupies), but we propose that their usage will allow us to understand how individual endophytes vary in their distributions across a plant community. We measured three univariate metrics (structural, beta- and phylogenetic specificity) that quantify the distribution of individual endophytes with respect to plant alpha diversity, plant beta

diversity and plant phylogenetic diversity. We tested the relationship between univariate and multivariate measures of host specificity to either host abundance or host phylogenetic identity.

C. Methods

Site Characteristics and Sampling Design

Plants were collected from the James San Jacinto Mountains Reserve in Idyllwild, California ($33° 48' 29''$, $−116° 46' 36''$) during July 2016. This mixed conifer forest has an elevation of 1 623 m, a mean annual precipitation of 665 mm, warm summers (average August temperature: 10.5ºC to 28.8ºC) and snow-packed winters (average January temperature: -2ºC to 12.2ºC). The canopy was dominated by *Calocedrus decurrens*, *Pinus ponderosa, Pinus lambertiana* and *Quercus kelloggii* with much of the understory composed of herbaceous angiosperms along with one abundant fern species, *Pteridium aquilinum*. We haphazardly identified general areas that included all dominant tree species and sampled from five 50 m^2 quadrats by using a random number table for directional headings (1-360º) and number of steps (1-100) to the first corner of a given quadrat. Within each quadrat, we haphazardly sampled 10 leaves with healthy appearance from one individual of every plant species. For perennial plants (e.g., conifers), we collected leaves less than one-year-old to standardize collections within that year's growing season across all plant species. Plant species were identified by observation (28 plant species) or by amplifying plant genetic markers (10 plant species) for a total of 38 plant species across 81 plant individuals from five quadrats. Plant species occurrence ranged from one to five individuals across quadrats while plant species relative abundance was estimated in the field per sampling quadrat with the Braun-Blanquet cover-abundance scale (Mueller-Dombois and Ellenberg 1974) and

averaged per plant species across the five quadrats. Plant species occurrence was not correlated with plant species relative abundance (Spearman's rho: 0.079, $P = 0.48$; Appendix 4: Fig. S1). Plants were kept at 4°C until they were surface-sterilized and frozen at -80°C within 48 hours of sampling.

Molecular and Bioinformatic Methods

Plant material was surface-sterilized by sequential immersion in 10% commercial bleach for 2 minutes, UV-treated deionized water for 30 seconds and 70% ethanol for 2 minutes. Plant tissue was left to air dry in a sterile laminar flow hood, frozen at -80°C then homogenized by mortar and pestle with liquid nitrogen. DNA from 80-100 mg of plant leaf tissue was extracted with a 2% CTAB method (Branco et al. 2015). The internal transcribed spacer 1 (ITS1) region was amplified with fungal-specific ITS1F-KYO1 and ITS2-KYO1 (Toju et al. 2012) primers modified with Illumina overhang adapters. Samples were sequenced on the Illumina MiSeq platform with 250 paired-end reads for 500 cycles and a 15% spike-in of PhiX (see Open Research Statement).

Sequenced reads were processed with a modified UNOISE pipeline (Edgar 2016) and clustered into amplicon sequence variants (ASVs) with USEARCH (version 11.0.667; Edgar 2010). ASVs were queried against a GenBank database (Altschul et al. 1990) with BLAST+ (Camacho et al. 2009) and non-fungal ASVs were removed with MEGAN (Huson et al. 2007; see Appendix 4: Supplementary Methods). One sample with low sequencing depth (JSJ2U9; 34 reads) was excluded from subsequent analyses. We included two PCR4-TOPO TA vectors (Invitrogen, Carlsbad, CA, USA) containing the ITS1 region from two fungal species (one Ascomycete and one Basidiomycete endophyte; see Open Research Statement) as positive controls to identify sequencing reads that were potential contaminants due to read

misassignment from Illumina 'index hopping' (van der Valk et al. 2019). Descriptions of PCR conditions, library preparation and bioinformatic processing of sequencing results can be found in Appendix 4: Supplementary Methods.

DNA from plant species that we were unable to identify by keying in the field were amplified with plant-specific ribulose-1,5-bisphosphate carboxylase oxygenase (rbcL; forward: rbcLa-F, reverse: rbcLajf634-R; Kress and Erickson 2007, Fazekas et al. 2008) or internal transcribed spacer 2 (ITS2; forward: ITS-u3; reverse: ITS-u4; Cheng et al. 2016) primers and Sanger sequenced (see Open Research Statement). We cross-referenced the top BLAST hit with a plant species list curated by the James Reserve to identify sequenced plant species.

Demultiplexed, raw sequence files (.fastq) can be downloaded from GenBank with Sequence Read Archive BioProject Accession Number: PRJNA356423. Positive MiSeq controls (GenBank Accessions: MG840195 and MG840196) and amplified plant rbcL or ITS2 regions (GenBank Accessions: MW308177-MW308178; MW321512-MW321521) were sequenced at the University of California, Berkeley DNA Sequencing Facility. Datasets and code can be accessed on figshare (doi.org/10.6084/m9.figshare.13404884) and are provided in Data S1. See *lotus* (https://github.com/austenapigo/lotus), for reproducible calculations for absolute and relative structural, phylogenetic and beta specificity.

Rarefaction Analysis and Community Data

Sequenced endophyte communities from each plant individual were rarefied to a sequencing depth of 15 075 reads (Appendix 4: Fig. S2) across 80 plant individuals and 1 117 endophyte ASVs. Sampling completeness, an estimate for how well an endophyte community was characterized by sequencing, was assessed in iNEXT (Hsieh et al. 2016) by

interpolation and extrapolation of rarefaction curves (>99% sampling completeness for all samples; Appendix 4: Fig. S2). All analyses were performed in R (version 3.6.1).

For all analyses, we compared endophyte communities per plant individual, or the endophyte community that originated from an individual's 10 sampled leaves. Quadrat was a significant predictor of endophyte community composition by PERMANOVA ($R^2 = 0.16$; $P = 0.001$; Appendix 4: Fig. S3) using the "adonis" function in the *vegan* package (Oksanen et al. 2020). This indicated that plant individuals within the same quadrat, each of a different plant species, tended to be more similar in their endophyte communities than plant individuals from the same plant species in different quadrats. Hence, for all analyses we either included quadrat as a categorical covariate or constrained permutations per quadrat to test for statistical significance such that plant species represented by multiple plant individuals only contributed once to each quadrat-level relationship. Plant communities within each quadrat did not significantly differ in average plant species relative abundance (Kruskal-Wallis rank sum test: $P = 0.89$) or plant species phylogenetic diversity based on mean pairwise plant phylogenetic distance (Kruskal-Wallis rank sum test: $P = 0.33$).

Univariate Host Specificity Per Endophyte

An alternative approach to multivariate metrics that quantify host specificity per endophyte community are univariate metrics that quantify host specificity per endophyte species or another taxonomic unit. We have leveraged a host specificity framework (Apigo and Oono 2018) proposed in host-parasite systems by Poulin et al. (2011) used to quantify various ways parasites can display host specificity (Poulin and Mouillot 2003, 2005,

Krasnov et al. 2011, Cooper et al. 2012). We quantified three univariate metrics for host specificity: structural specificity, phylogenetic specificity, and beta specificity (Figure 15).
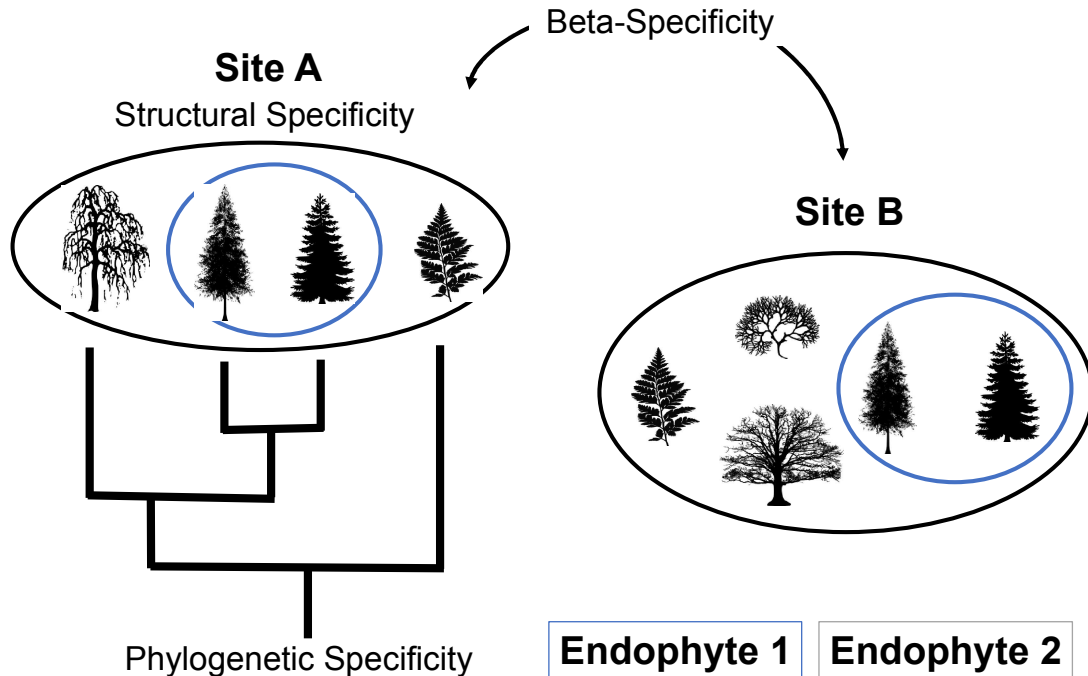
Figure 15. Conceptual diagram of univariate host specificity metrics quantified per endophyte. Two endophytes (blue and grey shaded areas) occupy varying plant species (bryophyte, monilophyte, gymnosperm and angiosperm) across two sites (Sites A and B). Structural specificity quantifies how endophytes vary in presence or evenness among host species, the most fundamental feature of host-symbiont interactions. For example, Endophyte 1 (blue shade) occupies fewer plant species in Sites A and B (two plant species per site) relative to Endophyte 2 (grey shade; four plant species in Site A and five hosts in Site B) and has higher structural specificity. Phylogenetic specificity (cladogram) quantifies host specificity in the context of plant evolutionary relatedness. Endophyte 1 occupies a narrower phylogenetic breadth of plant species in Sites A and B (gymnosperms only) relative to Endophyte 2 (angiosperm, bryophyte, gymnosperm and monilophyte) and has higher phylogenetic specificity. In contrast to structural and phylogenetic specificity, beta specificity (arrow between Sites A and B) quantifies how endophytes vary in the consistency of their specificity to plant species by geography or time (not depicted). Endophyte 1 occupies the same set of hosts in Sites A and B and has higher beta specificity than Endophyte 2 which occupies a more variable set of plant species across Sites A and B. If Endophyte 1 or 2 was not found in a third site, Site C (not depicted), they would only have lower beta specificity if a plant species they associated with in Sites A or B was also present in Site C. These host specificity metrics were originally described by Poulin et al. (2011).

Structural specificity was measured as either the richness or diversity of host species that an endophyte occupies (Poulin et al. 2011). We measured host species richness with the "specnumber" function in the *vegan* package (Oksanen et al. 2020) and host species diversity with Shannon's H diversity index (Shannon and Weaver 1948) using the "diversity" function in the *vegan* package (Oksanen et al. 2020). Host species richness and host species diversity specificity metrics were negated (i.e., multiplied by negative one) such that more positive values corresponded to a narrower distribution among host species. For example, an endophyte that occupied one plant species would have higher structural specificity than another that occupied 10 plant species (Figure 15). Similarly, an endophyte that was more unevenly (i.e., more narrowly) distributed in abundance among plant species relative to another endophyte with more even distribution (but equal host species richness) would have higher structural specificity.

Phylogenetic specificity, a measure of evenness and breadth across plant lineages occupied per endophyte, was measured as the Mean Pairwise Phylogenetic Distance (MPD; Webb 2000) with the "mpd" function in the *picante* package (Kembel et al. 2010). An ultrametric phylogenetic tree was pruned from a backbone phylogeny of 74 533 vascular plant species representing all extant vascular plant families in North America with the 'phylo.maker' function in the *V.PhyloMaker* package (Appendix 4: Fig. S4; Jin and Qian 2019). We grouped plant individuals at multiple plant phylogenetic scales, including major groups, families, and species, since inference could vary depending on the phylogenetic scale (Graham et al. 2018, Ladau and Eloe-Fadrosh 2019). Angiosperms were grouped into major groups based on nomenclature and systematics from Jansen et al. (2007) as asterids (14 species; n = 28 individuals), rosids (12 species; n = 22 individuals), basal eudicots

(family Ranunculaceae; two species; n = 3 individuals) and commelinids (five species; n = 8 individuals) along with conifers as Pinophyta (four species; n = 15 individuals) and ferns as Polypodiophyta (one species; n = 4 individuals). MPD was negated (i.e., multiplied by negative one) such that more positive values corresponded to a narrower phylogenetic distribution among host species. For example, an endophyte that occupied more closely related plant species relative to another endophyte would be more narrowly distributed with respect to plant phylogenetic relationships and thus have greater phylogenetic specificity (Figure 15).

Beta specificity, a measure of how consistent a given endophyte is associated with host species across the landscape, was measured as the presence-absence Sørensen (Diserud and Odegaard 2007) or abundance-weighted Morisita-Horn (Chao et al. 2008, Jost et al. 2011) Multiple-Assemblage Overlap Measure. A beta specificity of 0 indicates an endophyte never occurs in the same plant species across sampling quadrats. A beta specificity of 1 indicates that an endophyte is found in equal abundance among all individuals of a single or multiple plant species. If an endophyte occurred in a subset of individuals that made up a given plant species range, this endophyte would have a lower beta specificity relative to an endophyte that occurred in all individuals of a plant species range (Figure 15).

For all univariate analyses, we quantified host specificity per plant species. For example, if an endophyte appeared in multiple plant individuals of the same species, but not others, this endophyte would have a host species richness of one. The host specificity value for each endophyte ASV was based on its presence or evenness across the five quadrats (i.e., the entire plant community). Hence, each endophyte ASV contributed the same value to the host specificity average for all plant individuals or plant species in which it was found. All

89

singletons, or ASVs that appeared only once in this dataset, were removed for all analyses after rarefaction. In addition to ASV singletons, we excluded any endophyte ASVs that only appeared in one plant species, no matter their occurrence across quadrats, from our analyses because these ASVs cannot be compared in a meaningful way to a null model (more details in next section).

Null Host Specificity Models

Host specificity analyses could be vulnerable to biased ecological inferences by not accounting for spurious relationships between host specificity and endophyte abundance. Rarer endophytes have a lower theoretical maximum of the number of plants they could occupy and were more host-specific relative to abundant endophytes (Appendix 4: Fig. S5a). The relationship between an endophyte's host specificity and abundance is an extension of abundance-occupancy relationships, the widely documented biogeographic pattern in microbial and macroecological systems that predicts a positive relationship between a species abundance and its occurrence across sites (Gaston et al. 2000, Shade et al. 2018). Since more abundant endophytes can occupy more plants, there is expected to be a positive relationship between an endophyte's abundance and its occupancy across plants, its structural specificity or host species richness. While endophyte read abundance may not accurately represent biological abundance due to the constraints associated with high-throughput sequencing (Harrison et al. 2021), read abundances will ultimately bias the quantification of host specificity because rarer endophyte ASVs will still be found in fewer plant individuals and categorized as more host-specific relative to more abundant endophyte ASVs.

To account for this bias, studies typically analyze rare and abundant taxa separately at a variety of relative abundance thresholds (e.g., 0.1% - 1%; Logares et al. 2014, Zhang et al. 2018a, 2018b, Xue et al. 2018). Instead, we compared observed values of endophyte host specificity to null expectations based on randomized endophyte read counts among plant individuals. We term an endophyte's observed host specificity as "absolute host specificity" and standardized each value against an ASV-specific null distribution that we term "relative host specificity". An individual endophyte ASV's absolute host specificity was only compared to the null distribution of itself among the randomized communities with its read abundance held constant within each randomization.

We calculated relative host specificity as a standardized effectsize, or the difference between an endophyte's absolute host specificity and its mean null host specificity value divided by the standard deviation of its null host specificity values as in Cooper et al. 2012 and Svensson-Coelho et al. 2013. A relative host specificity value greater than zero indicates that an endophyte had a higher host specificity value than itself within randomized communities. Here, we report relative host specificities because this method classifies host-symbiont interactions depending on their expectation to occur by chance rather than being biased towards classifying endophytes as host-specific simply because they are rare (Appendix 4: Fig. S5a). Absolute host specificities are provided in Appendix 4: Fig. S9 and S16.

For structural and beta specificity, we performed a constrained randomization method (Swenson and Weiser 2014, sensu Swenson 2014) with the "quasiswapcount" randomization algorithm using the "permatfull" function in the *vegan* package that varied the number of plant species associated with any endophyte ASV. This method preserved total endophyte

read abundance per plant individual (i.e., row sums) and in the whole plant community (i.e., column sums) as well as the proportion of associations between endophyte ASVs and plant individuals across the entire community (i.e., connectance). Randomizations were restricted within each quadrat with the "strata" parameter to avoid plant-endophyte interactions that were unlikely to occur due to the significant effect of geography on endophyte community composition (Appendix 4: Fig. S3). We also repeated the analysis with a less constrained randomization method (using "shuffle.web" parameter in the "nullmodel" function in the *bipartite* package) that redistributed all read counts among plant individuals and endophyte ASVs, thereby changing row sums, column sums and endophyte occupancy across plant species but maintained connectance.

For phylogenetic specificity, we used two constrained randomization methods that permuted the plant taxonomic labels of the distance matrix, while keeping all other properties of the community matrix constant, because endophytes that occupied rare distantly-related lineages within the plant community (e.g., the one fern species) were more likely to have greater pairwise phylogenetic distances relative to endophytes that occurred in common plant lineages (e.g., one of the 33 angiosperm species). We used the "taxa.labels" algorithm to randomize the labels of the plant species phylogenetic distance matrix with the "ses.mpd" function in the *picante* package (Kembel et al. 2010). We conducted another analysis with the "phylogeny.pool" algorithm that randomized the phylogenetic distance matrix labels by drawing plant species with equal probability.

Linear Models

We assessed the relationships between relative structural or beta specificity and plant abundance using linear models with the "lm" function in the *stats* package. We also compared relative phylogenetic specificities among plant phylogenetic groups with the "Anova" function in the *car* package. Presence-absence host specificity metrics (e.g., host species richness) were regressed as a function of presence-absence plant species occurrence. Abundance-weighted host specificity metrics (e.g., host species diversity) were regressed as a function of abundance-weighted plant species relative abundance. We quantified presence-absence and abundance-measures of endophyte host specificity for linear models with a categorical predictor variable (e.g., plant major group). We kept quadrat as a covariate for all linear models such that a plant species represented by multiple plant individuals contributed only once to each quadrat-level relationship between host specificity and plant abundance or plant phylogenetic group. We tested all linear models with additional categorical covariates that included plant species along with interaction terms and assessed the best-fit model with the Akaike Information Criterion with a correction for small sample sizes (AICc) with the "model.sel" function in the *MuMIn* package (Bartón 2009). We assessed normality of model residuals with the "simulateResiduals" function in the *DHARMa* (Hartig 2020) package. Color palettes used to discriminate quadrats and plant phylogenetic groups were chosen with the *calecopal* (Bui et al. 2020) and *RColorBrewer* (Neuwirth 2014) packages.

We assessed occupancy-abundance relationships between the absolute host specificity metrics and log-transformed endophyte read abundance with the "cor.test" function in the *stats* package. We quantified correlations among the three relative host specificity metrics to

understand if endophytes that occupied fewer host species were also more likely to display host-specific patterns in other ways by occupying more closely related host species or consistently occurring among the same host species across quadrats. We also tested whether the mean absolute host specificity value of the observed endophyte community was more or less host-specific than expected by random chance with a one-sample t-test (Appendix 4: Fig. S6) with the "t.test" function in the *stats* package.

Multivariate Analyses

Bray-Curtis dissimilarities of endophyte communities were represented in ordination space by Principal Coordinates Analysis (PCoA; Gower 1966) with the "cmdscale" function in the *stats* package. We analyzed the effect of plant phylogenetic distance on endophyte community Bray-Curtis dissimilarity (Mantel r; Legendre and Legendre 1998) while controlling for geography using the "mantel.partial" function in the *vegan* (permutations = 999; Oksanen et al. 2020) package. The geographic distance matrix was composed of Euclidean distances from quadrat latitude and longitude coordinates (see Open Research Statement). A positive relationship between endophyte community dissimilarity and plant phylogenetic distance supports phylosymbiosis, indicating that host specificity is predicted by plant evolutionary relatedness. We tested for phylospecificity by analyzing the effect of plant identity on endophyte community composition, using PERMANOVA with a permutation test for statistical significance (permutations = 999) constrained per quadrat. We grouped plant individuals at varying plant phylogenetic scales (plant species, plant family, plant major group) as with univariate phylogenetic specificity. All multivariate analyses were repeated with the presence-absence Jaccard Index.

94

D. Results

Endophyte Abundance Biases Univariate Host Specificity Metrics

Absolute structural or beta specificity values were negatively correlated with log endophyte read abundance (Pearson's r: Shannon's H = -0.22; Sørensen Multiple-Site Overlap = -0.33; Morisita-Horn Multiple-Site Overlap = -0.20; $P < 0.05$), with host species richness (i.e., structural specificity using presence-absence data) being the most negatively correlated (Pearson's r: -0.89, $P < 0.01$), confirming positive occupancy-abundance relationships (Appendix 4: Fig. S5). On the other hand, absolute phylogenetic specificity was not correlated with log endophyte read abundance (Pearson's r: presence-absence MPD = 0.0059; abundance-weighted MPD = -0.031; $P > 0.05$) but displayed decreasing variance as the number of plant species occupied by a given endophyte increased, as expected (Appendix 4: Fig. S7; Appendix 4: Results; Swenson 2014).

Univariate Structural Specificity

Plant species that occurred in more quadrats tended to harbor endophytes that occupied fewer plant species (i.e., higher relative structural specificity, $R^2 = 0.36$, $P < 0.001$; Figure 16a). However, plant species with higher average relative abundance within quadrats tended to harbor endophytes that occupied more plant species (i.e., lower relative structural specificity, $R^2 = 0.12$, $P = 0.036$; Appendix 4: Fig. S8a). Absolute structural specificity and the alternative community randomization method showed similar trends (Appendix 4: Fig. S9a-b; S10a-b).
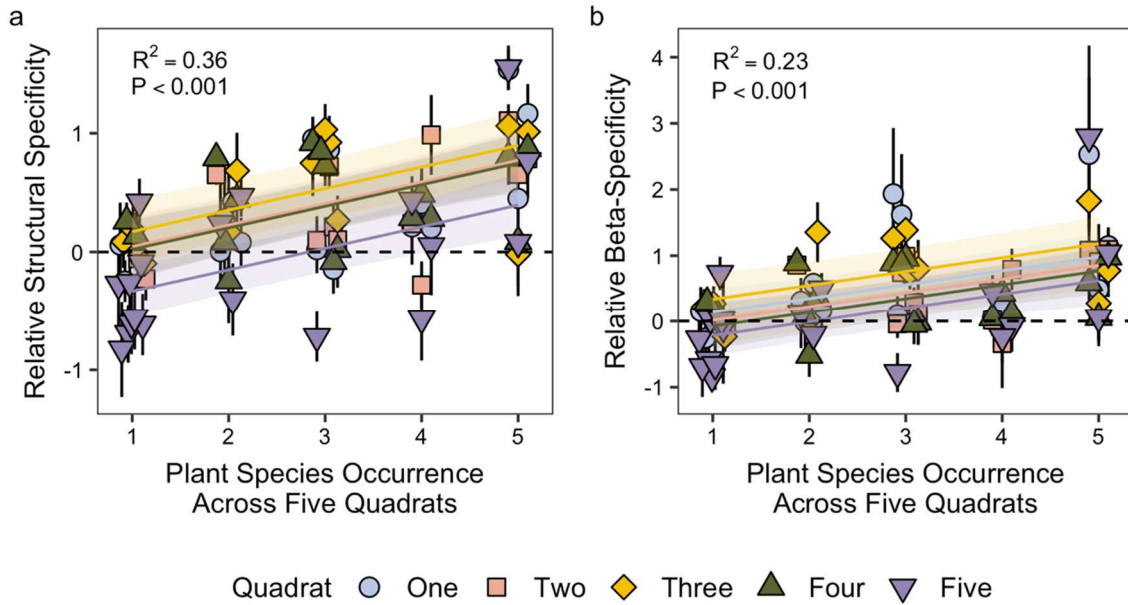
Figure 16. Plant species more abundant across the landscape harbor endophytes that occupy fewer plant species (a) and are more likely to be found in the same plant species across quadrats (b). (a) Regression of relative structural specificity measured as the standardized effect size of -1 * host species richness ($R^2 = 0.36$, $P < 0.001$) as a function of plant species occurrence. (b) Regression of relative beta specificity measured as the standardized effect size of the Sørensen Multiple-Assemblage Overlap Measure ($R^2 = 0.23$, $P < 0.001$) as a function of plant species occurrence. Each point represents the average relative structural or beta specificity value among all endophyte ASVs of one plant individual (n = 80 individuals) with standard error bars. Positive relative host specificities (above the dashed horizontal lines) indicate higher host specificity than expected by chance. Plant individuals are color- and shape-coded by quadrat and jittered along the x-axis by plant species. Linear regressions included quadrat as a covariate such that plant species represented by multiple plant individuals only contributed one value to each quadrat-specific regression analysis. Shaded regions refer to 95% confidence intervals.

Univariate Beta Specificity

Similarly, plant species present in more quadrats, but not found more abundantly within quadrats, harbored endophytes that were more consistent in their occupation of the same plant species across quadrats (i.e., higher relative beta specificity, $R^2 = 0.23$, $P < 0.001$, Fig. 16b; $R^2 = 0.086$, $P = 0.60$, Appendix 4: Fig. S8b). Absolute beta specificity and alternative community randomization methods showed similar trends (Appendix 4: Fig. S9c-d; S10c-d).

*Multivariate and Univariate Phylogenetic Specificity*

Multivariate analyses showed differences in endophyte community dissimilarity among major plant groups but not across plant phylogenetic distance (Figure 17; Appendix 4: Fig. S11). Major plant group (PERMANOVA $R^2 = 0.079$; $P = 0.002$), plant family ($R^2 = 0.33$; $P = 0.001$) and plant species ($R^2 = 0.51$; $P = 0.001$) were significant predictors of endophyte community composition (Figure 17; Appendix 4: Fig. S12-S13). However, relative distances between endophyte communities in ordination space (PCoA) were not indicative of host specificity by plant phylogenetic distance. For example, endophyte communities of asterids, rosids and commelinids were not more similar to the exclusion of all other plant groups (Figure 17a; Appendix 4: Fig. S11a). Similarly, endophyte community dissimilarities and plant phylogenetic distances were not significantly correlated (Partial Mantel r = -0.17; $P = 1$; Figure 17b) while controlling for geographic distance. Partial Mantel tests within angiosperms only (r = -0.068; $P = 0.98$) and conifers only (r = -0.011; $P = 0.51$) were also non-significant. Mantel tests with the Jaccard Index showed similar trends (Appendix 4: Fig. S11b).
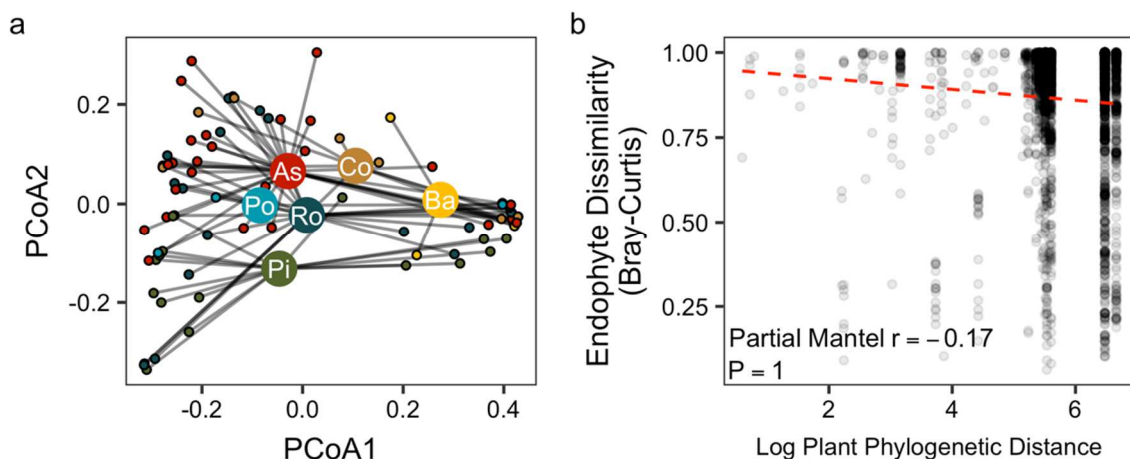
Figure 17. Endophyte communities significantly vary in composition among major plant groups (a) but not as a function of plant phylogenetic distance (b). (a) Principal coordinates analysis of endophyte Bray-Curtis dissimilarities among major plant groups.

Smaller circles represent the endophyte community of a given plant individual (n = 80 individuals). Larger circles with corresponding labels (As = Asterids, Ba = Basal Eudicots, Co = Commelinids, Pi = Pinophyta, Po = Polypodiophyta, Ro = Rosids) refer to the centroid of plant individuals within each major plant group. PERMANOVA among major plant groups with permutations (n = 999) constrained per quadrat was significant (R^2 = 0.079; P = 0.002). PERMANOVA were significant at the plant family (R^2 = 0.33; P = 0.001) and plant species level (R^2 = 0.51; P = 0.001; Appendix 4: Fig. S12-S13). (b) Correlation of endophyte Bray-Curtis dissimilarities and log-transformed plant phylogenetic distance controlling for geographic distance (Partial Mantel r = -0.17; P = 1). The partial Mantel test with the Jaccard Index was also not significant (Partial Mantel r = 0.04; P = 0.12; Appendix 4: Fig. S11).

Endophytes found in angiosperms (asterids, rosids, basal eudicots and commelinids) and conifers (Pinophyta) generally occupied narrower breadths of the plant phylogeny (i.e., higher relative phylogenetic specificity; ANOVA $F_{(5,70)}$ = 17.58; P < 0.001; Figure 18a) than endophytes found in the one fern species (Polypodiophyta) that were significantly more dispersed in plant phylogenetic breadth (Appendix 4: Table S1). Relative phylogenetic specificity based on presence-absence community data, relative phylogenetic specificity with an alternative community randomization method and absolute phylogenetic specificity showed similar trends (Appendix 4: Fig. S14-S18; Appendix 4: Table S2).
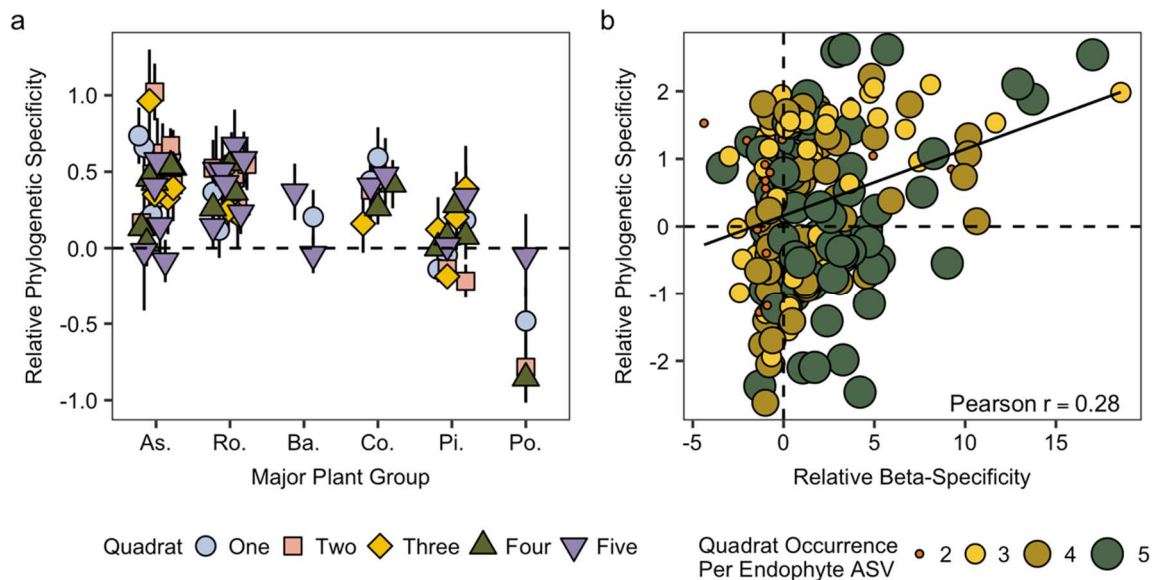
Figure 18. Endophytes within angiosperms (As = Asterids, Ro = Rosids, Ba = Basal Eudicots, Co = Commelinids) and conifers (Pi = Pinophyta) associate with narrower

phylogenetic breadths of plants (a) and are more consistent in their interactions to those plant species across the landscape (b) than endophytes within the fern species (Po = Polypodiophyta). Positive relative host specificities indicate higher host specificity than expected by chance (above or to the right of the dashed lines). (a) Relative phylogenetic specificity measured as the standardized effect size of -1* mean pairwise phylogenetic distance across major plant groups with quadrat as a covariate (ANOVA: $F_{(5,70)}$ = 17.58; P < 0.001). Each point represents the average relative phylogenetic specificity value among all endophyte ASVs of one plant individual (n = 80 individuals) with standard error bars. Plant individuals are color- and shape-coded by quadrat and jittered along the x-axis by plant species. Pairwise comparisons were evaluated with Tukey's HSD (Appendix 4: Table S1). (b) Correlation between relative phylogenetic and relative beta specificity per endophyte ASV (Pearson's r: 0.28; P < 0.001). Each point represents an endophyte ASV and is colored-coded by an endophyte ASV's presence across quadrats. Beta specificity was measured as the standardized effect size of the Morisita-Horn Multiple-Assemblage Overlap Measure. Analyses were repeated with presence-absence indices (Appendix 4: Fig. S14). Endophytes that appeared in one quadrat never had higher absolute beta specificity relative to the null model because occurring in a single quadrat precluded their ability to display consistency in plant interactions across quadrats and thus were excluded from the analysis.

Relationships between Structural, Phylogenetic and Beta specificity

Endophytes that were more consistent in their occupation of certain plant species across quadrats (higher relative beta specificity) tended to occupy a narrower phylogenetic breadth of plants (higher relative phylogenetic specificity; Pearson's r: 0.28; P < 0.001; Figure 18b). Endophytes that occupied fewer plant species (higher relative structural specificity) tended to occupy the same plant species across quadrats (Pearson's r: 0.46; P < 0.001) and occupy a narrower phylogenetic breadth of plant species (Pearson's r: 0.32; P < 0.001; Appendix 4: Fig. S19-S20).

E. Discussion

Plant Abundance and Patterns of Endophyte Host Specificity

Endophytes found in more commonly occurring plant species across the landscape tended to occupy fewer plant species and were also found more consistently within those same plant species across quadrats (Figure 16). This finding is supportive of the Common

Host Hypothesis that proposes a plant population must be common enough for passively dispersed endophytes to reliably recolonize the same plant species and evolve host specificity. In contrast, we found no relationship between relative structural or relative beta specificity and the average relative abundance of a plant species within quadrats suggesting relative abundance of hosts within a 50 m^2 plot may not represent landscape-level availability relevant for endophytes in our study.

We found a distance-decay relationship where endophyte communities became increasingly dissimilar with geographic distance (Mantel r: 0.17; $P = 0.001$) across the entire plant community. However, there was no distance-decay relationship between endophyte similarity and geographic distance for plant species that occurred across all quadrats (Mantel r: 0.079; $P = 0.185$). This suggests that differences in the relative abundance of plant species could explain distance-decay patterns described in past studies in the absence of strong environmental gradients. For example, in Vincent et al. (2016), a distance-decay pattern may not have been found because they sampled the most abundant plant species (e.g., *Ficus*) across a wide geographic range (300 km) which tended to host the same endophytes (i.e., high relative beta specificity). We also found that rarer endophytes were less likely to occupy the same plant species across quadrats (i.e., low relative beta specificity; Figure 18b), supporting findings in Oono et al. (2017). Further work may establish the positive relationships between endophyte host specificity and plant abundance as the underlying mechanism for distance-decay patterns in endophyte communities.

Just as the spatial abundance of a host influences substrate availability for endophytes, the temporal abundance, such as perennial life histories or evergreen leaves, could also be an important host factor for the evolution of host specificity. For example, a study by Mastura

and Fukuda (2013) showed that evergreen or deciduous life histories in broadleaved plants and evergreen conifers were predictors of cultured endophyte community composition across three geographically distant sites in Japan. In our study, plant species that were more common across the landscape were also more likely to be evergreen than deciduous (3.36 *vs*. 2.46 quadrat occurrence). Plant species that were more common across the landscape still harbored endophytes that occurred in fewer plant species ($R^2 = 0.46$, $P < 0.001$) and more consistently within them across quadrats ($R^2 = 0.38$, $P = 0.001$; Figure 19) regardless of leaf phenology (evergreen *vs*. deciduous). Endophytes in evergreen plant species tended to be found more consistently in fewer plant species across quadrats compared to those found in deciduous plants (Figure 19). Hence, spatial as well as temporal prevalence could both play significant roles in driving patterns of endophyte host specificity in plant species.
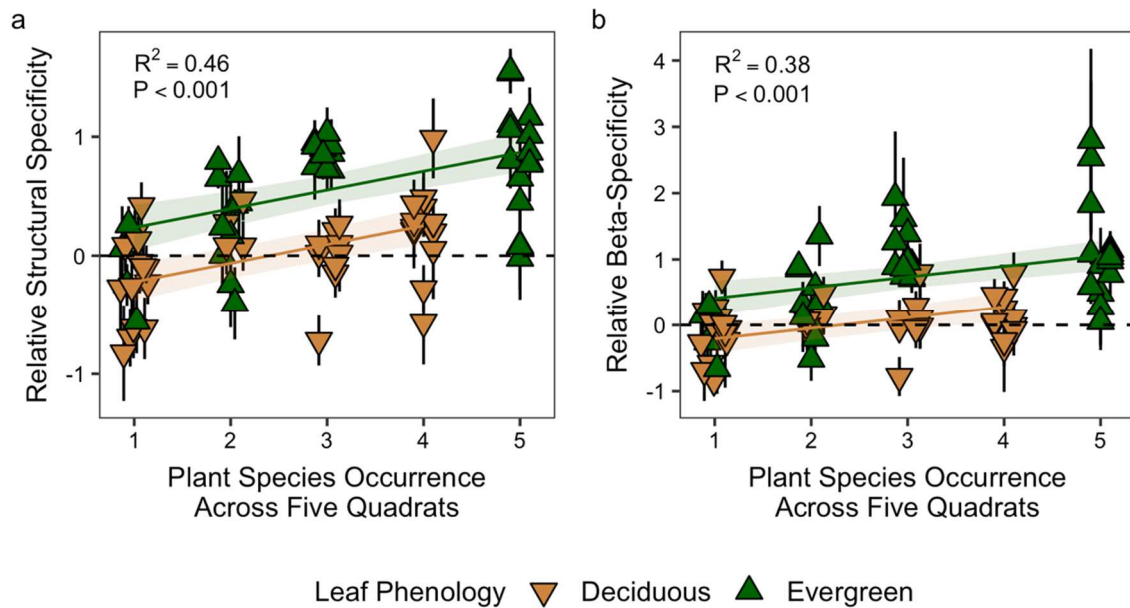
Figure 19. Endophytes in evergreen plant species are found in fewer plant species (a) and more consistently in those same plant species across quadrats (b) compared to endophytes found in deciduous plants. (a) Regression of relative structural specificity measured as the standardized effect size of -1 * host species richness ($R^2 = 0.46$; $P < 0.001$) as a function of plant species occurrence. (b) Regression of relative beta specificity measured as the standardized effect size of the Sørensen Multiple-Assemblage Overlap Measure ($R^2 = 0.38$, $P = 0.001$) as a function of plant species occurrence. Each point represents the average relative structural or beta specificity value among all the endophyte

ASVs of one plant individual (n = 80 individuals) with standard error bars. Positive relative host specificities indicate higher host specificity than expected by chance (above the dashed lines). Plant samples are color- and shape-coded by quadrat and jittered along the x-axis by plant species. Linear regressions include leaf phenology as a covariate with shaded regions that refer to 95% confidence intervals.

Plant Evolutionary Relatedness and Patterns of Endophyte Host Specificity

We did not find an effect of plant phylogenetic distance on endophyte community similarity, or phylosymbiosis, but did find evidence for phylospecificity. Endophyte communities were structured with respect to plant species, families and major groups, similar to past studies (Vincent et al. 2016, Whitaker et al. 2020). In particular, we found that endophytes within major groups of angiosperms (asterids, rosids, basal eudicots and commelinids) associated with narrower phylogenetic breadths of plant species compared to endophytes within conifers and the single fern species.

Endophyte species associated with conifers, the most common plant group at this site by quadrat occurrence, were found in many other plant groups as well (i.e., low phylogenetic specificity) even though we expected high host specificity, especially among Leotiomycete endophytes (Arnold 2007, Sieber 2007, U'ren et al. 2010, Moler and Aho 2018). Leotiomycete endophytes did not make up the majority of endophyte reads in *Pinus lambertiana* (two plant individuals) and *Pinus ponderosa* (five plant individuals) and were also commonly found in 36 non-conifer plant species (72 plant individuals), suggesting that previous pine-Leotiomycete associations may have been biased by the sampled plant community (see Appendix 4: Supplementary Results). However, at a finer phylogenetic resolution, endophytes within the genus *Lophodermium* (Rhytismataceae, Leotiomycetes) were highly restricted within the *Pinus* genus (see Appendix 4: Supplementary Results) supporting findings from previous studies (Ortiz-García et al. 2003, Salas-Lizana and Oono

2018). Hence, phylogenetic specificity from prior culture-based studies could be found within our next-generation sequencing data, but these associations may be exceptions in communities dominated by unculturable endophytes with general host associations among distantly related plant species.

Endophytes within the single fern species, *Pteridium aquilinum*, were the broadest in their host associations with respect to plant phylogenetic relationships (Figure 18a; Appendix 4: Supplementary Results). This supports a "co-opting hypothesis", where ferns may have co-opted endophytes of angiosperms during diversification (Schuettpelz and Pryer 2009, Olmo-Ruiz and Arnold 2017). For example, in three tropical plant communities, Olmo-Ruiz and Arnold (2017) concluded that culturable endophytes of ferns rarely displayed host specificity when comparing endophyte occurrence among ferns and angiosperms.

Endophyte Life History

Although the majority of endophytes living asymptomatically within leaves are hypothesized to be horizontally transmitted through the environment (Herre et al. 2007, Rodriguez et al. 2009), we cannot rule out the possibility that endophytes in this community may be vertically-transmitted. Vertically transmitted foliar endophytes have been documented to occur in grass (Saikkonen et al. 2016) and non-grass plant species (e.g., forbs; Hodgson et al. 2014) and would be expected to occur consistently in plants of the same species (i.e., high beta specificity), but not others, because they are directly propagated to the next plant generation by colonizing plant seeds. If plant species that were more common across the landscape were also more likely to harbor host-specific vertically

transmitted endophytes than rarer plant species, this could explain the positive relationship between relative beta specificity and plant species occurrence (Figure 16b). For example, if individuals of a more abundant plant species harbored a higher proportion of vertically transmitted endophytes that were not found in other plant species, these plant individuals would share a more similar endophyte community and be more closely clustered in PCoA space than individuals from a rarer plant species. However, there was no significant relationship between the average distance of plant individuals of the same species (R^2 = - 0.019; P = 0.43) or family (R^2 = 0.18; P = 0.072) to their group centroid in PCoA space and plant group occurrence across quadrats (Appendix 4: Fig. S12, S13, S21). This suggests that although common plant species often harbored endophytes that were found more consistently across the landscape in the same hosts (Figure 16b), these endophytes were also found in many other plant species in the community. Hence, it is unlikely that common plant species also hosted more vertically transmitted endophytes than rarer plant species, unless these vertically transmitted endophytes were host-specific across multiple plant species. Future work could identify potential vertically transmitted endophytes with the beta specificity metric to investigate their prevalence in a system that commonly assumes horizontal transmission as the primary mode of endophyte dispersal (Herre et al. 2007, Rodriguez et al. 2009).

F. Conclusions

In this study, plant species that were more common across the landscape tended to harbor endophytes that occupied fewer plant species and were more likely to be found in the same plant species. We did not find evidence for phylosymbiosis, or the effect of plant phylogenetic distance on endophyte community composition, but did find an effect of plant

104

identity, known as phylospecificity. These correlations support the presence of host-specific patterns in endophyte communities with respect to host abundance but more definitive tests with culture-based methods (e.g., inoculation trials on hosts) are needed to confirm endophyte host specificity. Furthermore, there are many other reasons endophytes could display host-specific patterns to plants outside of plant abundance and plant evolutionary history. Endophyte communities have been shown to be strongly structured by climate (e.g., temperature and rainfall; Zimmerman and Vitousek 2012) and host physiology (Valkama et al. 2005, Van Bael et al. 2017, e.g., González□Teuber et al. 2020) not explored in this study. Incorporating climatic and other leaf-related predictors, beyond leaf phenology, into models of host specificity will better shape our understanding of the relative importance of plant abundance and phylogenetic relationships in the context of environmental gradients and leaf traits.

Data Availability

Demultiplexed, raw sequence files (.fastq) can be downloaded from GenBank with Sequence Read Archive BioProject Accession Number: PRJNA356423. Positive MiSeq controls (GenBank Accessions: MG840195 and MG840196) and amplified plant rbcL or ITS2 regions (GenBank Accessions: MW308177–MW308178; MW321512–MW321521) were sequenced at the University of California, Berkeley DNA Sequencing Facility. Data sets and code can be accessed on Figshare (https://doi.org/10.6084/m9.figshare.13404884) and are provided in Data S1. A. Apigo wrote an R package, called lotus (https://github.com/austenapigo/lotus), to reproducibly offer calculations for absolute and relative structural, phylogenetic, and beta-specificity.

IV. Conclusions

In this dissertation, I showed that endophyte species richness (alpha diversity) was greatest in the tropics, supportive of LDG, and this result remained consistent across different plant groups and extrapolative measurements of richness. Interestingly, endophyte richness did not increase incrementally towards the tropics but reflected variation in climate and the leaf environment. Endophyte richness was greatest in the tropical forests of Panama and was also pronounced, to a lesser degree, in temperate conifer forests of Canada and Oregon. Endophyte richness was lowest in temperate conifer forests and a shrubland in California. Differences in temperature, precipitation, precipitation seasonality, host diversity, and the leaf environment were correlated with endophyte richness and explained why patterns of endophyte richness had a bimodal distribution as a function of latitude.

Plant communities in California experienced high precipitation seasonality, received low precipitation, and had low amounts of photosynthetic biomass. These plant communities may have hosted the fewest endophyte species because these environments were the most limited in water and photosynthetic biomass, two key resources for endophytes. However, plant communities in Mexico had similar levels of water and photosynthetic biomass but hosted similar amounts of endophyte species as wet environments in Oregon and Canada with high photosynthetic biomass. While climate and the leaf environment explained a significant amount of variation in endophyte richness, host interactions also played a significant role in structuring endophyte communities.

I found that differences in endophyte composition among host species (beta diversity) were greater in temperate forests than sub-tropical and tropical forests. Host species in temperate conifer forests harbored more distinct endophyte communities, while host species

106

in sub-tropical and tropical forests harbored more similar endophyte communities. Host specificity was also greatest in temperate plant communities, consistent with my expectation that greater compositional dissimilarity among endophyte communities was driven by endophytes that were more host specific.

Endophyte host specificity could explain why endophyte species richness (alpha diversity) increased towards the tropics while differences in endophyte species composition (beta diversity) increased towards temperate regions. I hypothesize that species-rich host communities can be a barrier to host-specific interactions in the tropics because host specialists cannot reliably disperse to compatible host individuals in such heterogenous environments (May 1991, Kindlmann et al. 2007, Griffin et al. 2019, Oita et al. 2021b). This mechanism may explain why I found greater host generalism in the species-rich plant communities of Panama, while host specificity was greater in temperate conifer forests that had fewer host species.

Furthermore, host specialists may be well-adapted to their hosts and competitively exclude other endophytes (Büchi et al. 2014) while host generalists are "jacks of all trades but masters of none" and their ability to colonize many host species comes at the cost of competitive ability. Therefore, I propose that low host species richness in temperate plant communities selects for host specialists that dominate host tissues, thereby decreasing endophyte richness in hosts. On the other hand, high host species richness in the tropics selects for host generalists and facilitates coexistence among endophytes, allowing hosts to accumulate endophyte species. In the absence of environmental gradients, this proposed mechanism should produce contrasting patterns of endophyte alpha and beta diversity as a function of plant diversity, which follows LDG.

My hypothesis assumes the driver of host specificity is not necessarily host species richness but rather how increasing host species richness decreases the relative abundance of any given host species and increases dispersal limitation for host specific endophytes. Within a single plant community, I found evidence that suggested host abundance can preclude or facilitate host-specific interactions. Specifically, I found that more abundant plant species harbored endophytes that were more host specific. These endophytes occupied fewer plant species and were consistently found in the same plant species across the landscape, supporting the hypothesis that low host density can be a barrier to host specific interactions. This relationship between host specificity and host abundance could explain why host generalists are abundant in diverse tropical plant communities; it is because the relative abundance of any given host in the tropics is low and host specialists are often dispersal limited.

My hypothesis relating host specificity to latitudinal patterns of endophyte species richness and composition hinges on the following assumptions that may not always hold. First, hosts within any plant community are well-mixed across the landscape. Plant species clustered in space could be locally abundant and reduce dispersal limitation for specialist endophytes. Second, hosts in species-rich host communities always have low relative abundance. Many foundational tree species on Barro Colorado Island, Panama are consistently found across the island and across years (89% of 308 tree species present in seven censuses spanning 1982–2010; Condit et al. 2012). Whether host abundance has a significant relationship with endophyte host specificity in the tropics is still unanswered but future work will examine this. Third, host controls on endophyte colonization are minimal such that host abundance is the only major barrier to host specificity. It has been shown that host morphological, chemical, or physiological traits are correlated with differences in

endophyte communities (González☐Teuber et al. 2020, Tellez et al. 2022) but whether hosts can tightly control the colonization of endophytes is still untested. Future work should consider how patchiness in host ranges across the landscape, how variation in the effect of host abundance across plant communities, and how plant physiology shapes host specific interactions.

Future work could also explore the relationships between host diversity, host specificity and endophyte species richness with common garden designs. For example, Laforest-Lapointe et al. (2017) and Griffin et al. (2019) examined how varying tree species and phylogenetic diversity in common gardens influenced foliar bacterial or fungal community structure, respectively. Griffin et al. (2019) found that greater host diversity decreased endophyte diversity, counter to my hypothesis and findings from Lapointe et al. (2017) who found a positive relationship between host diversity and phyllosphere bacterial diversity. Their study design manipulated tree species richness to a maximum 12 host trees species in 35 m^2 plots. However, these plots may not present a heterogenous enough host landscape to select for generalists or be large enough at the plot-level to buffer the endophyte community from neighboring sources of fungal inoculum. More manipulative experiments and surveys that test the relationships between host diversity, endophyte diversity, and host specificity without the confounding influence of climate, as in Chapter 1, will clarify how host community structure influences symbiont communities.

Recent studies estimate that global fungal biodiversity ranges from 2.2–3.8 million (Hawksworth and Luecking 2017), using scaling ratios that extrapolate fungal richness from plant richness (1 plant-to-9.8 fungi), or 11.7–13.2 million species (Wu et al. 2019), using scaling ratios that extrapolate culture-free fungal richness from cultured fungal richness (1

cultured fungus-to-8.8 uncultured fungi). Such variable estimates beg the question of whether uniform scaling ratios are useful tools for global estimates of diversity. Here, I show that endophytes vary considerably in their species richness and host specificity across environments and host species. Estimates of species richness could be used to generate plant-to-fungal scaling ratios specific to environments and host groups that have been sampled. Estimates of host specificity could then be used to impose correction factors on such scaling ratios. For example, observed plant-to-endophyte scaling ratios in the tropics should be lowered to avoid overestimating endophyte richness because many endophyte species are shared among host species, and vice versa for temperate plant and endophyte communities. While the tropics are hypothesized to be a 'hotspot' for fungal biodiversity, a growing body of evidence has shown that soil, litter-associated, and ectomycorrhizal fungal guilds do not follow LDG and are most species-rich in the temperate environments (Tedersoo et al. 2012, Jabiol et al. 2013, Shi et al. 2013, Duarte et al. 2016, Seena et al. 2019, Wang et al. 2019, Větrovský et al. 2019, Liu et al. 2020). Considering variation in fungal species richness and host or environmental specificity across fungal guilds will help us produce estimates of global fungal biodiversity that account for the biogeography and ecology of fungi.

V. References

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

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. Journal of Molecular Biology 215:403–410.

Arrhenius, O. 1921. Species and Area. Journal of Ecology 9:95–99.

Apigo, A., and R. Oono. 2018. Dimensions of Host Specificity in Foliar Fungal Endophytes. Pages 15–42 in A. M. Pirttilä and A. C. Frank, editors. Endophytes of Forest Trees. Springer, Cham, Switzerland.

Apigo, A., and R. Oono. 2022. Plant abundance, but not plant evolutionary history, shapes patterns of host specificity in foliar fungal endophytes. Ecosphere 13(1): e03879.

Arnold, A. E. 2007. Understanding the diversity of foliar endophytic fungi: progress, challenges, and frontiers. Fungal Biology Reviews 21:51–66.

Arnold, A. E., and F. Lutzoni. 2007. Diversity and host range of foliar fungal endophytes: Are tropical leaves biodiversity hotspots? Ecology 88:541–549.

Baas-Becking, L. G. M. 1934. Geobiologie of Inleiding Tot de Milieukunde. Van Stockkum & Zoon, The Hague.

Barge, E. G., D. R. Leopold, K. G. Peay, G. Newcombe, and P. E. Busby. 2019. Differentiating spatial from environmental effects on foliar fungal communities of *Populus trichocarpa*. Journal of Biogeography 46(9):2001-2011.

Barton, K. 2009. MuMIn: multi-model inference. http://r-forge. r-project. org/projects/mumin/.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 67(1):1–48.

Beck, H. E., E. F. Wood, T. R. McVicar, M. Zambrano-Bigiarini, C. Alvarez-Garreton, O. M. Baez-Villanueva, J. Sheffield, and D. N. Karger. 2020. Bias Correction of Global High-Resolution Precipitation Climatologies Using Streamflow Observations from 9372 Catchments. Journal of Climate 33(4):1299–1315.

Bivand, R. S., E. Pebesma, and V. Gómez-Rubio. 2013. Classes for Spatial Data in R. Pages 21–57. Applied Spatial Data Analysis with R, 2nd Edition. Springer, New York, New York.

Bowman, E. A., and A. E. Arnold. 2021. Drivers and implications of distance decay differ for ectomycorrhizal and foliar endophytic fungi across an anciently fragmented landscape. The ISME Journal 15:3437–3454.

Branco, S., P. Gladieux, C. E. Ellison, A. Kuo, K. LaButti, A. Lipzen, I. V. Grigoriev, H.-L. Liao, R. Vilgalys, K. G. Peay, J. W. Taylor, and T. D. Bruns. 2015. Genetic isolation between two recently diverged populations of a symbiotic fungus. Molecular Ecology 24:2747–2758.

Brooks, A. W., K. D. Kohl, R. M. Brucker, E. J. van Opstal, and S. R. Bordenstein. 2016. Phylosymbiosis: relationships and functional effects of microbial communities across host evolutionary history. PLoS Biology 14:e2000225.

Bui, A., H. Lowman, A. S. Guerra, and A. M.-T. Kuile. 2020. calecopal: A California-inspired Package of Color Palettes. R package version 0.1.0. https://github.com/an-bui/calecopal.

Calcagno, V., P. Jarne, M. Loreau, N Mouquet, and P. David. 2017. Diversity spurs diversification in ecological communities. Nature Communications 8:15810.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: architecture and applications. BMC bioinformatics 10:421.

Chaib De Mares, M., D. Sipkema, S. Huang, B. Bunk, J. Overmann, and J. D. van Elsas. 2017. Host Specificity for Bacterial, Archaeal and Fungal Communities Determined for High- and Low-Microbial Abundance Sponge Species in Two Genera. Frontiers in Microbiology 8:2560.

Chamberlain, S. A., and E. Szöcs. 2013. taxize: taxonomic search and retrieval in R. F1000 Research 2:191.

Chao, A., L. Jost, S. C. Chiang, Y.-H. Jiang, and R. L. Chazdon. 2008. A two-stage probabilistic approach to multiple-community similarity indices. Biometrics 64:1178–1186.

Cheng, T., C. Xu, L. Lei, C. Li, Y. Zhang, and S. Zhou. 2016. Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity. Molecular Ecology Resources 16:138–149.

Christian, N., B. K. Whitaker, and K. Clay. 2015. Microbiomes: unifying animal and plant systems through the lens of community ecology theory. Frontiers in Microbiology 6:869.

Christian, N., C. Sullivan, N. D. Visser, and K. Clay. 2016. Plant host and geographic location drive endophyte community composition in the face of perturbation. Microbial Ecology 72:621–632.

Cirtwill, A. R., D. B. Stouffer, and T. N. Romanuk. 2015. Latitudinal gradients in biotic niche breadth vary across ecosystem types. Proceedings of the Royal Society B 282:20151589.

Clark, N. J., and S. M. Clegg. 2017. Integrating phylogenetic and ecological distances reveals new insights into parasite host specificity. Molecular Ecology 26:3074–3086.

Cobian, G. M., C. P. Egan, and A. S. Amend. 2019. Plant–microbe specificity varies as a function of elevation. The ISME Journal 13(11):2778–2788.

Condit, R., R. A. Chisholm, and S. P. Hubbell. 2012. Thirty Years of Forest Census at Barro Colorado and the Importance of Immigration in Maintaining Diversity. PLoS ONE 7(11): e49826.

Cooper, N., R. Griffin, M. Franz, M. Omotayo, C. L. Nunn, and J. Fryxell. 2012. Phylogenetic host specificity and understanding parasite sharing in primates. Ecology Letters 15:1370–1377.

Darcy, J. L., S. O. I. Swift, G. M. Cobian, G. L. Zahn, B. A. Perry, and A. S. Amend. 2020. Fungal communities living within leaves of native Hawaiian dicots are structured by landscape-scale variables as well as by host plants. Molecular Ecology 29:3103–3116.

Dickey, J. R., R. A. Swenie, S. C. Turner, C. C. Winfrey, D. Yaffar, A. Padukone, K. K. Beals, K. S. Sheldon, and S. N. Kivlin. 2021. The Utility of Macroecological Rules for Microbial Biogeography. Frontiers in Ecology and Evolution 9:633155.

Diserud, O. H., and F. Odegaard. 2007. A multiple-site similarity measure. Biology Letters 3:20–22.

Dormann, C. F., J. Frund, N. Bluthgen, and B. Gruber. 2009. Indices, Graphs and Null Models: Analyzing Bipartite Ecological Networks. The Open Ecology Journal 2:7-24.

Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19(1):11-15.

Dreyfuss, M. M., and I. H. Chapela. 1994. Potential of fungi in the discovery of novel, low-molecular weight pharmaceuticals. Biotechnology 26:49–80.

Duarte, S., F. Bärlocher, C. Pascoal, and F. Cássio. 2016. Biogeography of aquatic hyphomycetes: Current knowledge and future perspectives. Fungal Ecology 19:169-181.

Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461.

Edgar, R. C. 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. bioRxiv https://doi.org/10.1101/081257.

Edgar, R. C., and H. Flyvbjerg. 2015. Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics 31:3476–3482.

Fazekas, A. J., K. S. Burgess, P. R. Kesanakurti, S. W. Graham, S. G. Newmaster, B. C. Husband, D. M. Percy, M. Hajibabaei, and S. C. H. Barrett. 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS One 3:e2802.

Ferrier, S., G. Manion, J. Elith, and K. Richardson. 2007. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. Diversity and Distributions 13(3):252-264.

Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: new 1□km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37(12):4302-4315.

Fierer, N., and R. B. Jackson. 2006. The diversity and biogeography of soil bacterial communities. Proceedings of the National Academy of Sciences 103(3):626-631.

Fisher, P. J. 1996. Survival and spread of the endophyte *Stagonospora pteridiicola* in *Pteridium aquilinum*, other ferns and some flowering plants. New Phytologist 132:119–122.

Fisher, P. J., and E. Punithalingam. 1993. *Stagonospora pteridiicola* sp. nov., a new endophytic coelomycete in *Pteridium aquilinum*. Mycological Research 97:661–664.

Frank, J., P. W. Crous, J. Z. Groenewald, B. Oertel, K. D. Hyde, P. Phengsintham, and H.-J. Schroers. 2010. *Microcyclospora* and *Microcyclosporella*: novel genera accommodating epiphytic fungi causing sooty blotch on apple. Persoonia 24:93–105.

Fröhlich, J., and K. Hyde. 1999. Biodiversity of palm fungi in the tropics: are global fungal diversity estimates realistic? Biodiversity and conservation 8:977–1004.

Gaston, K. J. 2000. Global patterns in biodiversity. Nature 405:220–227.

Gaston, K. J., T. M. Blackburn, J. J. D. Greenwood, R. D. Gregory, R. M. Quinn, and J. H. Lawton. 2000. Abundance-occupancy relationships. Journal of Applied Ecology 37:39–59.

Gillman, L. N., S. D. Wright, J. Cusens, P. D. McBride, Y. Malhi, and R. J. Whittaker. 2015. Latitude, productivity and species richness. Global Ecology and Biogeography 24(1):107-117.

Gleason, M. L., R. Zhang, J. C. Batzer, and G. Sun. 2019. Stealth Pathogens: The Sooty Blotch and Flyspeck Fungal Complex. Annual Review of Phytopathology 57:135–164.

Glynou, K., T. Ali, A.-K. Buch, S. H. Kia, S. Ploch, X. Xia, A. Çelik, M. Thines, and J. G. Maciá-Vicente. 2016. The local environment determines the assembly of root endophytic fungi at a continental scale. Environmental Microbiology 18(8):2418-34.

González⬚Teuber, M., C. Vilo, M. J. Guevara⬚Araya, C. Salgado⬚Luarte, and E. Gianoli. 2020. Leaf resistance traits influence endophytic fungi colonization and community composition in a South American temperate rainforest. Journal of Ecology 108:1019–1029.

Gower, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–338.

Graham, C. H., D. Storch, and A. Machac. 2018. Phylogenetic scale in ecology and evolution. Global Ecology and Biogeography 27:175–187.

Griffin, E. A., J. G. Harrison, M. K. McCormick, K. T. Burghardt, and J. D. Parker. 2019. Tree Diversity Reduces Fungal Endophyte Richness and Diversity in a Large-Scale Temperate Forest Experiment. Diversity 11:234.

Harrison, J. G., and E. A. Griffin. 2020. The diversity and distribution of endophytes across biomes, plant phylogeny and host tissues: how far have we come and where do we go from here? Environmental Microbiology 22:2107–2123.

Harrison, J. G., W. John Calder, B. Shuman, and C. Alex Buerkle. 2021. The quest for absolute abundance: The use of internal standards for DNA-based community ecology. Molecular Ecology Resources 21:30–43.

Hartig, F. 2020. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package. https://cran.r-project.org/web/packages/DHARMa/index.html

Hawksworth, D. L. 2001. The magnitude of fungal diversity: the 1.5 million species estimate revisited. Mycological Research 105(12):1422–1432.

Hawksworth DL, Luecking R. 2017. Fungal diversity revisited: 2.2 to 3.8 million species. Microbiol Spectrum 5(4).

Hendershot, J. N., Q. D. Read, J. A. Henning, N. J. Sanders, and A. T. Classen. 2017. Consistently inconsistent drivers of microbial diversity and abundance at macroecological scales. Ecology 98:1757–1763.

Herre, E. A., L. C. Mejía, D. A. Kyllo, E. Rojas, Z. Maynard, A. Butler, and S. A. Van Bael. 2007. Ecological implications of anti-pathogen effects of tropical fungal endophytes and mycorrhizae. Ecology 88:550–558.

Higgins, K. L., K. Lindsay Higgins, A. Elizabeth Arnold, P. D. Coley, and T. A. Kursar. 2014. Communities of fungal endophytes in tropical forest grasses: highly diverse host- and habitat generalists characterized by strong spatial structure. Fungal Ecology 8(1):1-11.

Higgins, K. L., K. Lindsay Higgins, P. D. Coley, T. A. Kursar, and A. Elizabeth Arnold. 2011. Culturing and direct PCR suggest prevalent host generalism among diverse fungal endophytes of tropical forest grasses. Mycologia 103(2):247–260.

Hillebrand, H. 2004. On the Generality of the Latitudinal Diversity Gradient. American Naturalist 163(2):192-211.

Hodgson, S., C. de Cates, J. Hodgson, N. J. Morley, B. C. Sutton, and A. C. Gange. 2014. Vertical transmission of fungal endophytes is widespread in forbs. Ecology and Evolution 4:1199–1208.

Hoffman, M. T., and A. E. Arnold. 2008. Geographic locality and host identity shape fungal endophyte communities in Cupressaceous trees. Mycological Research 112:331–344.

Hsieh, T. C., K. H. Ma, and A. Chao. 2016. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). Methods in Ecology and Evolution 7:1451–1456.

Huang, S., O. R. P. Bininda-Emonds, P. R. Stephens, J. L. Gittleman, and S. Altizer. 2014. Phylogenetically related and ecologically similar carnivores harbour similar parasite assemblages. Journal of Animal Ecology 83:671–680.

Huson, D. H., A. F. Auch, J. Qi, and S. C. Schuster. 2007. MEGAN analysis of metagenomic data. Genome Research 17:377–386.

Jabiol, J., A. Bruder, M. O. Gessner, M. Makkonen, B. G. McKie, E. T. H. Peeters, V. C. A. Vos, and E. Chauvet. 2013. Diversity patterns of leaf-associated aquatic hyphomycetes along a broad latitudinal gradient. Fungal Ecology 6(5):439-448.

Jablonski, D., K. Roy, and J. W. Valentine. 2006. Out of the tropics: evolutionary dynamics of the latitudinal diversity gradient. Science 314:102–106.

Jansen, R. K., Z. Cai, L. A. Raubeson, H. Daniell, C. W. Depamphilis, J. Leebens-Mack, K. F. Müller, M. Guisinger-Bellian, R. C. Haberle, A. K. Hansen, T. W. Chumley, S.-B. Lee, R. Peery, J. R. McNeal, J. V. Kuehl, and J. L. Boore. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proceedings of the National Academy of Sciences 104:19369–19374.

Janzen, D. H., and C. M. Pond. 1975. A comparison, by sweep sampling, of the arthropod fauna of secondary vegetation in Michigan, England and Costa Rica. Transactions of the Royal Entomological Society of London 127:33–50.

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

Johnson, P. C. D. 2014. Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. Methods in Ecology and Evolution 5(9):944-946.

Jost, L., A. Chao, Chazdon, and R. L. 2011. Compositional similarity and β (beta) diversity. Pages 66–84 in A. E. Magurran and B. J. McGill, editors. Biological Diversity: Frontiers in Measurement and Assessment. Oxford University Press, New York, United States of America.

Karger, D. N., D. R. Schmatz, G. Dettling, and N. E. Zimmermann. 2020. High-resolution monthly precipitation and temperature time series from 2006 to 2100. Scientific Data 7:248.

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

Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26:1463–1464.

Kindlmann, P., I. Schödelbauerová, and A. F. G. Dixon. 2007. Inverse latitudinal gradients in species diversity. Pages 246–257 in D. Storch, P. A. Marquet, and J. H. Brown, editors. Scaling Biodiversity. Cambridge University Press, New York, United States of America.

Kivlin, S. N., M. A. Mann, J. S. Lynn, M. R. Kazenel, D. Lee Taylor, and J. A. Rudgers. 2022. Grass species identity shapes communities of root and leaf fungi more than elevation. ISME Communications 2:25.

Kohl, K. D. 2020. Ecological and evolutionary mechanisms underlying patterns of phylosymbiosis in host-associated microbial communities. Philosophical Transactions of the Royal Society of London. Series B, Biological sciences 375:20190251.

Koskella, B. 2020. The phyllosphere. Current Biology 30(19):R1143-R1146.

Kraft, N. J. B., P. B. Adler, O. Godoy, E. C. James, S. Fuller, and J. M. Levine. 2015. Community assembly, coexistence and the environmental filtering metaphor. Functional Ecology 29(5):592-599.

Krasnov, B. R., D. Mouillot, G. I. Shenbrot, I. S. Khokhlova, and R. Poulin. 2011. Beta specificity: the turnover of host species in space and another way to measure host specificity. International Journal for Parasitology 41:33–41.

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

Kress, W. J., and D. L. Erickson. 2007. A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. PloS One 2:e508.

Ladau, J., and E. A. Eloe-Fadrosh. 2019. Spatial, Temporal, and Phylogenetic Scales of Microbial Ecology. Trends in Microbiology 27:662–669.

Langenfeld, A., S. Prado, B. Nay, C. Cruaud, S. Lacoste, E. Bury, F. Hachette, T. Hosoya, and J. Dupont. 2013. Geographic locality greatly influences fungal endophyte communities in *Cephalotaxus harringtonia*. Fungal Biology 117:124–136.

Lau, M. K., A. Elizabeth Arnold, and N. C. Johnson. 2013. Factors influencing communities of foliar fungal endophytes in riparian woody plants. Fungal Ecology 6(5):365-378.

Legendre, P., and L. F. J. Legendre. 1998. Numerical Ecology 2nd Edition. Elsevier. Amsterdam.

Legendre, P., M. R. T. Dale, M.-J. Fortin, J. Gurevitch, M. Hohn, and D. Myers. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. Ecography 25(5):601-615.

Li, H.-T., Y. Luo, L. Gan, P.-F. Ma, L.-M. Gao, J.-B. Yang, J. Cai, M. A. Gitzendanner, P. W. Fritsch, T. Zhang, J.-J. Jin, C.-X. Zeng, H. Wang, W.-B. Yu, R. Zhang, M. van der Bank, R. G. Olmstead, P. M. Hollingsworth, M. W. Chase, D. E. Soltis, P. S. Soltis, T.-S. Yi, and D.-Z. Li. 2021. Plastid phylogenomic insights into relationships of all flowering plant families. BMC Biology 19:232.

Lian, C.-A., G.-Y. Yan, J.-M. Huang, A. Danchin, Y. Wang, and L.-S. He. 2019. Genomic Characterization of a Novel Gut Symbiont From the Hadal Snailfish. Frontiers in Microbiology 10:2978.

Lim, S. J., and S. R. Bordenstein. 2020. An introduction to phylosymbiosis. Proceedings of the Royal Society B: Biological Sciences 287:20192900.

Liu, J., J. Zhao, G. Wang, and J. Chen. 2019. Host identity and phylogeny shape the foliar endophytic fungal assemblages of Ficus. Ecology and Evolution 9:10472–10482.

Liu, S., H. Wang, P. Tian, X. Yao, H. Sun, Q. Wang, and M. Delgado-Baquerizo. 2020. Decoupled diversity patterns in bacteria and fungi across continental forest ecosystems.

Logares, R., S. Audic, D. Bass, L. Bittner, C. Boutte, R. Christen, J.-M. Claverie, J. Decelle, J. R. Dolan, M. Dunthorn, B. Edvardsen, A. Gobet, W. H. C. F. Kooistra, F. Mahé, F. Not, H. Ogata, J. Pawlowski, M. C. Pernice, S. Romac, K. Shalchian-Tabrizi, N. Simon, T. Stoeck, S. Santini, R. Siano, P. Wincker, A. Zingone, T. A. Richards, C. de Vargas, and R. Massana. 2014. Patterns of rare and abundant marine microbial eukaryotes. Current Biology 24:813–821.

Long, J. A. 2022. jtools: Analysis and Presentation of Social Scientific Data. https://cran.r-project.org/web/packages/jtools/jtools.pdf

Lüdecke, D., M. Ben-Shachar, I. Patil, P. Waggoner, and D. Makowski. 2021. performance: An R Package for Assessment, Comparison and Testing of Statistical Models. Journal of Open Source Software 6(60):3139.

Naïma, M., M. Vos, C. L. Murall, P. Legendre, and B. J. Shapiro. 2020. Does diversity beget diversity in microbiomes? eLife 9:e58999.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17(1).

Matsumura, E., and K. Fukuda. 2013. A comparison of fungal endophytic community diversity in tree leaves of rural and urban temperate forests of Kanto district, eastern Japan. Fungal Biology 117:191–201.

Maurice, S., G. Arnault, J. Nordén, S. S. Botnen, O. Miettinen, and H. Kauserud. 2021. Fungal sporocarps house diverse and host-specific communities of fungicolous fungi. The ISME Journal 15:1445–1457.

May, R. M. 1991. A fondness for fungi. Nature 352:475–476.

Meiser, A., M. Bálint, and I. Schmitt. 2014. Meta-analysis of deep-sequenced fungal communities indicates limited taxon sharing between studies and the presence of biogeographic patterns. The New Phytologist 201:623–635.

Minter, D. W. 1981. Lophodermium on pines. Mycological Papers 147:1–54.
Mirzwa-Mróz, E., M. Wińska-Krysiak, J. Marcinkowska, and M. L. Gleason. 2011. First Report of *Microcyclosporella mali* Causing Sooty Blotch and Flyspeck Disease on Plum in Poland. The American Phytopathological Society 95(4).

Mokany, K., C. Ware, S. N. C. Woolley, S. Ferrier, and M. C. Fitzpatrick. 2022. A working guide to harnessing generalized dissimilarity modelling for biodiversity analysis and conservation assessment. Global Ecology and Biogeography 31(4):802-821.

Moler, E. R. V., and K. Aho. 2018. Whitebark pine foliar fungal endophyte communities in the southern Cascade Range, USA: host mycobiomes and white pine blister rust. Fungal Ecology 33:104–114.

Morera, A., J. M. de Aragón, M. De Cáceres, J. A. Bonet, and S. de-Miguel. 2022. Historical and future spatially-explicit climate change impacts on mycorrhizal and saprotrophic macrofungal productivity in Mediterranean pine forests. Agricultural and Forest Meteorology 319:108918.

Mucha, J., K. G. Peay, D. P. Smith, P. B. Reich, A. Stefański, and S. E. Hobbie. 2018. Effect of Simulated Climate Warming on the Ectomycorrhizal Fungal Community of Boreal and Temperate Host Species Growing Near Their Shared Ecotonal Range Limits. Microbial Ecology 75:348–363.

Mueller-Dombois, D., and H. Ellenberg. 1974. Aims and Methods of Vegetation Ecology. John Wiley & Sons, New York, United States of America.

Nakagawa, S., and H. Schielzeth. 2013. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. Methods in Ecology and Evolution 4(2):133-142.

Nakagawa, S., P. C. D. Johnson, and H. Schielzeth. 2017. The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. Journal of the Royal Society of Interface 14:20170213.

Neuwirth, E. 2014. RColorBrewer: ColorBrewer Palettes. R package. https://cran.r-project.org/web/packages/RColorBrewer/index.html

O'Brien, P. A., S. Tan, C. Yang, P. R. Frade, N. Andreakis, H. A. Smith, D. J. Miller, N. S. Webster, G. Zhang, and D. G. Bourne. 2020. Diverse coral reef invertebrates exhibit patterns of phylosymbiosis. The ISME Journal 14:2211–2222.

Oita, S., J. Carey, I. Kline, A. Ibáñez, N. Yang, E. F. Y. Hom, I. Carbone, J. M. U'Ren, and A. Elizabeth Arnold. 2021a. Methodological Approaches Frame Insights into Endophyte Richness and Community Composition. Microbial Ecology 82:21–34.

Oita, S., A. Ibáñez, F. Lutzoni, J. Miadlikowska, J. Geml, L. A. Lewis, E. F. Y. Hom, I. Carbone, J. M. U'Ren, and A. Elizabeth Arnold. 2021b. Climate and seasonality drive the richness and composition of tropical fungal endophytes at a landscape scale. Communications Biology 4:313.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner. 2022. vegan: Community Ecology Package. https://cran.r-project.org/web/packages/vegan/index.html

Olmo-Ruiz, M. D., and E. A. Arnold. 2017. Community structure of fern-affiliated endophytes in three neotropical forests. Journal of Tropical Ecology 33:60–73.

Oono, R., A. Rasmussen, and E. Lefèvre. 2017. Distance decay relationships in foliar fungal endophytes are driven by rare taxa. Environmental Microbiology 19:2794–2805.

Ortiz-García, S., D. S. Gernandt, J. K. Stone, P. R. Johnston, I. H. Chapela, R. Salas-Lizana, and E. R. Alvarez-Buylla. 2003. Phylogenetics of *Lophodermium* from pine. Mycologia 95:846–859.

Osono, T., and D. Hirose. 2011. Colonization and lignin decomposition of pine needle litter by *Lophodermium pinastri*. Forest Pathology 41:156–162.

Paradis, E., and K. Schliep. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35(3):526-528.

Paulson, J. N., O. Colin Stine, H. C. Bravo, and M. Pop. 2013. Differential abundance analysis for microbial marker-gene surveys. Nature Methods 10(12):1200-2.

Peay, K. G. 2014. Back to the future: natural history and the way forward in modern fungal ecology. Fungal Ecology 12:4–9.

Peay, K. G., M. G. Schubert, N. H. Nguyen, and T. D. Bruns. 2012. Measuring ectomycorrhizal fungal dispersal: macroecological patterns driven by microscopic propagules. Molecular Ecology 21:4122–4136.

Peay, K. G., P. G. Kennedy, and J. M. Talbot. 2016. Dimensions of biodiversity in the Earth mycobiome. Nature Reviews. Microbiology 14:434–447.

Petrini, O., P. J. Fisher, and L. E. Petrini. 1992. Fungal endophytes of bracken (*Pteridium aquilinum*), with some reflections on their use in biological control. Sydowia 44:282-293.

Pinheiro, J. C., and D. M. Bates. 2000. Mixed-Effects Models in Sand S-PLUS. Springer New York, NY.

Poulin, R., and D. Mouillot. 2003. Parasite specialization from a phylogenetic perspective: a new index of host specificity. Parasitology 126:473–480.

Poulin, R., and D. Mouillot. 2005. Combining phylogenetic and ecological information into a new index of host specificity. Journal of Parasitology 91:511–514.

Poulin, R., B. R. Krasnov, and D. Mouillot. 2011. Host specificity in phylogenetic and geographic space. Trends in Parasitology 27:355–361.

R Core Team. 2022. R: A language and environment for statistical computing.

Rodriguez, R. J., J. F. White Jr, A. E. Arnold, and R. S. Redman. 2009. Fungal endophytes: diversity and functional roles. New Phytologist 182:314–330.

Rodriguez, R. J., J. F. White Jr, A. E. Arnold, and R. S. Redman. 2009. Fungal endophytes: diversity and functional roles. The New Phytologist 182:314–330.

Rognes, T., T. Flouri, B. Nichols, C. Quince, and F. Mahé. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584.

Rohde, K. 1992. Latitudinal Gradients in Species Diversity: The Search for the Primary Cause. Oikos 65:514-527.

Rosenberg, E., and I. Zilber-Rosenberg. 2018. The hologenome concept of evolution after 10 years. Microbiome 6:78.

Rousk, J., E. Bååth, P. C. Brookes, C. L. Lauber, C. Lozupone, J. G. Caporaso, R. Knight, and N. Fierer. Soil bacterial and fungal communities across a pH gradient in an arable soil. The ISME Journal 4:1340-1351.

Saikkonen, K., C. A. Young, M. Helander, and C. L. Schardl. 2016. Endophytic Epichloë species and their grass hosts: from evolution to applications. Plant Molecular Biology 90:665–675.

Saikkonen, K., S. H. Faeth, M. Helander, and T. J. Sullivan. 1998. Fungal Endophytes: A Continuum of Interactions with Host Plants. Annual Review of Ecology and Systematics 29:319-343.

Salas-Lizana, R., and R. Oono. 2018. A comparative analysis of *Lophodermium fissuratum*, sp. nov., found in haploxylon pine needles in the Pacific Northwest, and other *Lophodermium* endophytes. Mycologia 110:797–810.

Schuettpelz, E., and K. M. Pryer. 2009. Evidence for a Cenozoic radiation of ferns in an angiosperm-dominated canopy. Proceedings of the National Academy of Sciences 106:11200–11205.

Seena, S., F. Bärlocher, O. Sobral, M. O. Gessner, D. Dudgeon, B. G. McKie, E. Chauvet, L. Boyero, V. Ferreira, A. Frainer, A. Bruder, C. D. Matthaei, S. Fenoglio, K. R. Sridhar, R. J. Albariño, M. M. Douglas, A. C. Encalada, E. Garcia, S. D. Ghate, D. P. Giling, V. Gonçalves, T. Iwata, A. Landeira-Dabarca, D. McMaster, A. O. Medeiros, J. Naggea, J. Pozo, P. M. Raposeiro, C. M. Swan, N. S. D. Tenkiano, C. M. Yule, and M. A. S. Graça. 2019. Biodiversity of leaf litter fungi in streams along a latitudinal gradient. The Science of the Total Environment 661(15):306–315.

Slessarev, E. W., Y. Lin, N. L. Bingham, J. E. Johnson, Y. Dai, J. P. Schimel, and O. A. Chadwick. 2016. Water balance creates a threshold in soil pH at the global scale Nature 540:567–569.

Sessa, E. B., J. A. Banks, M. S. Barker, J. P. Der, A. M. Duffy, S. W. Graham, M. Hasebe, J. Langdale, F.-W. Li, D. B. Marchant, K. M. Pryer, C. J. Rothfels, S. J. Roux, M. L. Salmi, E. M. Sigel, D. E. Soltis, P. S. Soltis, D. W. Stevenson, and P. G. Wolf. 2014. Between two fern genomes. GigaScience 3:15.

Shade, A., R. R. Dunn, S. A. Blowes, P. Keil, B. J. M. Bohannan, M. Herrmann, K. Küsel, J. T. Lennon, N. J. Sanders, D. Storch, and J. Chase. 2018. Macroecology to unite all life, large and small. Trends in Ecology & Evolution 33:731–744.

Shannon, C. E., and W. Weaver. 1948. The Mathematical Theory of Communication. University of Illinois Press.

Shi, L.-L., P. E. Mortimer, J. W. Ferry Slik, X.-M. Zou, J. Xu, W.-T. Feng, and L. Qiao. 2013. Variation in forest soil fungal diversity along a latitudinal gradient. Fungal Diversity.

Sieber, T. N. 2007. Endophytic fungi in forest trees: are they mutualists? Fungal Biology Reviews 21:75–89.

Sierra, M. A., D. C. Danko, T. A. Sandoval, G. Pishchany, B. Moncada, R. Kolter, C. E. Mason, and M. M. Zambrano. 2020. The Microbiomes of Seven Lichen Genera Reveal Host Specificity, a Reduced Core Community and Potential as Source of Antimicrobials. Frontiers in Microbiology 11:398.

Sivakumar, N., R. Sathishkumar, G. Selvakumar, R. Shyamkumar, and K. Arjunekumar. Phyllospheric Microbiomes: Diversity, Ecological Significance, and Biotechnological Applications. Plant Microbiomes for Sustainable Agriculture 25: 113–172.

Sokolski, S., Y. Piche, and J. A. Berube. 2004. *Lophodermium macci* sp. nov., a new species on senesced foliage of five-needle pines. Mycologia 96:1261.

Song, Z., D. Schlatter, D. M. Gohl, and L. L. Kinkel. 2018. Run-to-Run Sequencing Variation Can Introduce Taxon-Specific Bias in the Evaluation of Fungal Microbiomes.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312–1313.

Stenström, E., and K. Ihrmark. 1997. Detection of *Lophodermium seditiosum* from pine needles using PCR techniques. Developments in Plant Pathology:223–226.

Stone J., and O. Petrini. 1997. Endophytes of Forest Trees: A Model for Fungus-Plant Interactions. In Plant Relationships Part B. Springer, Berlin, Heidelberg.

Sun X., Q. Ding, K. D. Hyde, L. D. Guo. 2012. Community structure and preference of endophytic fungi of three woody plants in a mixed forest. Fungal Ecology 5: 624–632.

Suryanarayanan T.S. 2011. Diversity of Fungal Endophytes in Tropical Trees. Pages 67–80 in A. M. Pirttilä and A. C. Frank, editors. Endophytes of Forest Trees. Springer, Cham, Switzerland.

Suryanarayanan, T. S., P. T. Devarajan, K. P. Girivasan, M. B. Govindarajulu, V. Kumaresan, T. S. Murali, T. Rajamani, N. Thirunavukkarasu, and G. Venkatesan. 2018. The Host Range of Multi-Host Endophytic Fungi. Current Science 115(10).

Svensson-Coelho, M., J. G. Blake, B. A. Loiselle, A. S. Penrose, P. G. Parker, and R. E. Ricklefs. 2013. Diversity, Prevalence, and Host Specificity of Avian Plasmodium and Haemoproteus in a Western Amazon Assemblage. Ornithological Monographs 76:1–47.

Swenson, N. G. 2014. Functional and Phylogenetic Ecology in R. Springer Science & Business Media, New York, United States of America.

Swenson, N. G., and M. D. Weiser. 2014. On the packing and filling of functional space in eastern North American tree assemblages. Ecography 37:1056–1062.

Talbot, J. M., T. D. Bruns, J. W. Taylor, D. P. Smith, S. Branco, S. I. Glassman, S. Erlandson, R. Vilgalys, H.-L. Liao, M. E. Smith, and K. G. Peay. 2014. Endemism and functional convergence across the North American soil mycobiome. Proceedings of the National Academy of Sciences 111:6341–6346.

Tanney, J. B., and K. A. Seifert. 2017. *Lophodermium resinosum* sp. nov. from red pine (*Pinus resinosa*) in Eastern Canada. Botany 95:773–784.

Tedersoo, L., M. Bahram, M. Toots, A. G. Diédhiou, T. W. Henkel, R. Kjøller, M. H. Morris, K. Nara, E. Nouhra, K. G. Peay, S. Põlme, M. Ryberg, M. E. Smith, and U. Kõljalg. 2012. Towards global patterns in the diversity and community structure of ectomycorrhizal fungi. Molecular ecology 21:4160–4170.

Tedersoo, L., M. Bahram, S. Põlme, U. Kõljalg, N. S. Yorou, R. Wijesundera, L. Villarreal Ruiz, A. M. Vasco-Palacios, P. Q. Thu, A. Suija, M. E. Smith, C. Sharp, E. Saluveer, A. Saitta, M. Rosas, T. Riit, D. Ratkowsky, K. Pritsch, K. Põldmaa, M. Piepenbring, C. Phosri, M. Peterson, K. Parts, K. Pärtel, E. Otsing, E. Nouhra, A. L. Njouonkou, R. H. Nilsson, L. N. Morgado, J. Mayor, T. W. May, L. Majuakim, D. J. Lodge, S. S. Lee, K.-H. Larsson, P. Kohout, K. Hosaka, I. Hiiesalu, T. W. Henkel, H. Harend, L.-D. Guo, A. Greslebin, G. Grelet, J. Geml, G. Gates, W. Dunstan, C. Dunk, R. Drenkhan, J. Dearnaley, A. De Kesel, T. Dang, X. Chen, F. Buegger, F. Q. Brearley, G. Bonito, S. Anslan, S. Abell, and K. Abarenkov. 2014. Fungal biogeography. Global diversity and geography of soil fungi. Science 346:1256688.

Tellez P. H., A. E. Arnold, A. B. Leo, K. Kitajima, and S. A. Van Bael. 2022. Traits along the leaf economics spectrum are associated with communities of foliar endophytic symbionts. Frontiers in Microbiology 13:927780.

Thébault E., and C. Fontaine. 2010. Stability of ecological communities and the architecture of mutualistic and trophic networks. Science 329: 853–856

Toju, H., A. S. Tanabe, S. Yamamoto, and H. Sato. 2012. High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. PloS one 7:e40863.

U'ren, J. M., F. Lutzoni, J. Miadlikowska, and A. E. Arnold. 2010. Community analysis reveals close affinities between endophytic and endolichenic fungi in mosses and lichens. Microbial Ecology 60:340–353.

U'Ren, J. M., F. Lutzoni, J. Miadlikowska, A. D. Laetsch, and A. Elizabeth Arnold. 2012. Host and geographic structure of endophytic and endolichenic fungi at a continental scale. American Journal of Botany 99(5):898-914.

U'Ren, J. M., F. Lutzoni, J. Miadlikowska, N. B. Zimmerman, I. Carbone, G. May, and A. E. Arnold. 2019. Host availability drives distributions of fungal endophytes in the imperilled boreal realm. Nature Ecology & Evolution 3:1430–1437.

Valkama, E., J. Koricheva, J.-P. Salminen, M. Helander, I. Saloniemi, K. Saikkonen, and K. Pihlaja. 2005. Leaf surface traits: overlooked determinants of birch resistance to herbivores and foliar micro-fungi? Trees 19:191–197.

Van Bael, S., C. Estrada, and A. Elizabeth Arnold. 2017. Foliar Endophyte Communities and Leaf Traits in Tropical Trees. Pages 79–94 in J. Dighton and J. F. White, editors. The Fungal Community Its Organization and Role in the Ecosystem. CRC Press, Taylor & Francis Group, Boca Raton, Florida.

van der Valk, T., F. Vezzi, M. Ormestad, L. Dalén, and K. Guschanski. 2019. Index hopping on the Illumina HiseqX platform and its consequences for ancient DNA studies. Molecular Ecology Resources 20:1171–1181.

Vaz, A. B. M., P. L. C. Fonseca, F. Badotti, D. Skaltsas, L. M. R. Tomé, A. C. Silva, M. C. Cunha, M. A. Soares, V. L. Santos, G. Oliveira, P. Chaverri, and A. Góes-Neto. 2018. A multiscale study of fungal endophyte communities of the foliar endosphere of native rubber trees in Eastern Amazon. Scientific Reports 8:16151.

Vázquez, D. P., and R. D. Stevens. 2004. The latitudinal gradient in niche breadth: concepts and evidence. American Naturalist 164:E1–E19.

Vega, F. E. 2008. Insect pathology and fungal endophytes. Journal of Invertebrate Pathology 98: 277–279.

Vega, F. E., A. Simpkins, M. C. Aime, F. Posada, S. W. Peterson, S. A. Rehner, F. Infante, A. Castillo, A. E. Arnold. 2010. Fungal endophyte diversity in coffee plants from Colombia, Hawai'i, Mexico and Puerto Rico. Fungal Ecology 3:122–138.

Vellend, M. 2010. Conceptual synthesis in community ecology. The Quarterly review of Biology 85:183–206.

Větrovský, T., P. Kohout, M. Kopecký, A. Machac, M. Man, B. D. Bahnmann, V. Brabcová, J. Choi, L. Meszárošová, Z. R. Human, C. Lepinay, S. Lladó, R. López-Mondéjar, T. Martinović, T. Mašínová, D. Morais, D. Navrátilová, I. Odriozola, M.

Štursová, K. Švec, V. Tláskal, M. Urbanová, J. Wan, L. Žifčáková, A. Howe, J. Ladau, K. G. Peay, D. Storch, J. Wild, and P. Baldrian. 2019. A meta-analysis of global fungal distribution reveals climate-driven patterns. Nature Communications 10:5142.

Vincent, J. B., G. D. Weiblen, and G. May. 2016. Host associations and beta diversity of fungal endophyte communities in New Guinea rainforest trees. Molecular Ecology 25:825–841.

Vohník, M., O. Borovec, Z. Kolaříková, R. Sudová, and M. Réblová. 2019. Extensive sampling and high-throughput sequencing reveal *Posidoniomyces atricolor* gen. et sp. nov. (Aigialaceae, Pleosporales) as the dominant root mycobiont of the dominant Mediterranean seagrass *Posidonia oceanica*. MycoKeys 55:59–86.

Walker, D. M., L. A. Castlebury, A. Y. Rossman, L. Struwe. 2013. Host conservatism or host specialization? Patterns of fungal diversification are influenced by host plant specificity in Ophiognomonia (Gnomoniaceae: Diaporthales). Biological Journal of the Linnean Society 111: 1–16.

Walter, G. H. 1991. What is resource partitioning? Journal of Theoretical Biology 150: 137–143.

Wang, P., Y. Chen, Y. Sun, S. Tan, S. Zhang, Z. Wang, J. Zhou, G. Zhang, W. Shu, C. Luo, and J. Kuang. 2019. Distinct Biogeography of Different Fungal Guilds and Their Associations With Plant Species Richness in Forest Ecosystems. Frontiers in Ecology and Evolution 7:216.

Wang, Z., P. R. Johnston, Z. L. Yang, and J. P. Townsend. 2009. Evolution of reproductive morphology in leaf endophytes. PloS One 4:e4246.

Wang, Z., Y. Jiang, M. Zhang, C. Chu, Y. Chen, S. Fang, G. Jin, M. Jiang, J. Lian, Y. Li, Y. Liu, K. Ma, X. Mi, X. Qiao, X. Wang, X. Wang, H. Xu, W. Ye, L. Zhu, Y. Zhu, F. He, and S. W. Kembel1. 2022. Diversity and biogeography of plant leaf bacteria along a latitudinal gradient. bioRxiv https://doi.org/10.1101/2022.06.23.497395

Wardhaugh, C. W., W. Edwards, N. E. Stork. 2015. The specialization and structure of antagonistic and mutualist networks of beetles on rainforest canopy trees. Biological Journal of the Linnean Society 114: 287–295.

Webb, C.O. 2000. Exploring the Phylogenetic Structure of Ecological Communities: An Example for Rain Forest Trees. American Naturalist 156:145–155.

Webb, C.O., D.D. Ackerly, and S.W. Kembel. 2008. Phylocom: software for the analysis of phylogenetic community structure and trait evolution. Bioinformatics 24:2098–2100

Webb, C.O., D. D. Ackerly, M. A. McPeek, M. J. Donoghue. 2002. Phylogenies and Community Ecology. Annual Review of Ecology, Evolution, and Systematics 33: 475–505.

Webb, C. O. 2000. Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. The American Naturalist 156:145–155.

Weiss, N.A. 2016. wBoot: Bootstrap Methods. R package version 1.0.3. https://CRAN.R-project.org/package=wBoot

Whitaker, B. K., H. L. Reynolds, and K. Clay. 2018. Foliar fungal endophyte communities are structured by environment but not host ecotype in *Panicum virgatum* (switchgrass). Ecology 99:2703–2711.

Whitaker, B. K., N. Christian, Q. Chai, and K. Clay. 2020. Foliar fungal endophyte community structure is independent of phylogenetic relatedness in an Asteraceae common garden. Ecology and Evolution 10:13895–13912.

Whittaker, R. H. 1972. Evolution and Measurement of Species Diversity. Taxon 21: 213.

Wolfe, E. R., and D. J. Ballhorn. 2020. Do Foliar Endophytes Matter in Litter Decomposition? Microorganisms 8.

Wu B., M. Hussain, W. Zhang, M. Stadler, X. Liu, and M. Xiang. 2019. Current insights into fungal species diversity and perspective on naming the environmental DNA sequences of fungi. Mycology 10(3):127-140.

Xue, Y., H. Chen, J. R. Yang, M. Liu, B. Huang, and J. Yang. 2018. Distinct patterns and processes of abundant and rare eukaryotic plankton communities following a reservoir cyanobacterial bloom. The ISME Journal 12:2263–2277.

Yang, H. L., G. Y. Sun, J. C. Batzer, P. W. Crous, J. Z. Groenewald, and M. L. Gleason. 2010. Novel fungal genera and species associated with the sooty blotch and flyspeck complex on apple in China and the USA. Persoonia 24:29–37.

Yang, T., P. Weisenhorn, J. A. Gilbert, Y. Ni, R. Sun, Y. Shi, and H. Chu. 2016. Carbon constrains fungal endophyte assemblages along the timberline. Environmental microbiology 18:2455–2469.

Younginger, B. S., and D. J. Ballhorn. 2017. Fungal endophyte communities in the temperate fern Polystichum munitum show early colonization and extensive temporal turnover. American Journal of Botany 104:1188–1194.

Zha, T., C. Li, S. Kellomäki, H. Peltola, K.-Y. Wang, and Y. Zhang. 2013. Controls of Evapotranspiration and CO_2 Fluxes from Scots Pine by Surface Conductance and Abiotic Factors. PLoS One. 2013; 8(7): e69027.

Zhang, T., and Y.-F. Yao. 2015. Endophytic fungal communities associated with vascular plants in the high arctic zone are highly diverse and host-plant specific. PloS One 10:e0130051.

Zhang, W., Y. Pan, J. Yang, H. Chen, B. Holohan, J. Vaudrey, S. Lin, and G. B. McManus. 2018a. The diversity and biogeography of abundant and rare intertidal marine microeukaryotes explained by environment and dispersal limitation. Environmental Microbiology 20:462–476.

Zhang, X., S. Liu, J. Wang, Y. Huang, Z. Freedman, S. Fu, K. Liu, H. Wang, X. Li, M. Yao, X. Liu, and J. Schuler. 2020. Local community assembly mechanisms shape soil bacterial β diversity patterns along a latitudinal gradient. Nature Communications 11:5428.

Zhang, Y., G. Wu, H. Jiang, J. Yang, W. She, I. Khan, and W. Li. 2018b. Abundant and rare microbial biospheres respond differently to environmental and spatial factors in Tibetan hot springs. Frontiers in Microbiology 9:2096.

Zhou, J., Y. Deng, L. Shen, C. Wen, Q. Yan, D. Ning, Y. Qin, K. Xue, L. Wu, Z. He, J. W. Voordeckers, J. D. Van Nostrand, V. Buzzard, S. T. Michaletz, B. J. Enquist, M. D. Weiser, M. Kaspari, R. Waide, Y. Yang, and J. H. Brown. 2016. Temperature mediates continental-scale diversity of microbes in forest soils. Nature Communications 7:12083.

Zimmerman, N. B., and P. M. Vitousek. 2012. Fungal endophyte communities reflect environmental structuring across a Hawaiian landscape. Proceedings of the National Academy of Sciences 109:13022–13027.

VI. Appendices

Appendix 1: Library preparation, sequencing, bioinformatic pipeline for Chapters 1 – 3.

Appendix 2: Supplementary Figures 1-11 and Tables 1-3 for Chapter 1.

Appendix 3: Supplementary Tables 1-6 for Chapter 2.

Appendix 4: Supplementary Figures 1-21 for Chapter 3.

Appendix 1: Library preparation, sequencing, and bioinformatic pipeline for Chapters 1 – 3.

Library Preparation

Genomic DNA was extracted from two duplicate subsamples (each with 80 mg of plant tissue) with a modified 2% CTAB method (Doyle and Doyle 1987) and eluted in 20 μl TE buffer. DNA quantification was performed on a Qubit 2.0 fluorometer using the Qubit dsDNA High Sensitivity Assay Kit (Invitrogen, Carlsbad, CA, US). The first round of PCR amplified the internal transcribed spacer 1 of the ribosomal DNA with ITS1F-KYO1 and ITS2-KYO1 (Toju et al. 2012) primers modified with Illumina transposase adapters in 25 μl PCR reactions. Volumes and initial concentrations of each PCR component were as follows: 1 μl of genomic DNA (10 ng/μl), 2.5 μl of BSA (1 mg/ml), 2.5 μl of 10x PCR buffer containing 15 mM $MgCl_2$, 2.5 μl of 8 mM dNTPs, 0.15 μl of Choice-Taq DNA polymerase (5 units/μl; Denville Scientific Inc., Holliston, MA, USA), 1.25 μl of each primer (0.5 μM; Integrated DNA Technologies, Coralville, IA, USA) and 13.85 μl of nuclease-free water. PCR conditions were 3 minutes at 95°C, followed by 35 cycles of 30 seconds at 95°C, 30 seconds at 47°C, 30 seconds at 72°C, and a final elongation of 5 min at 72°C. Duplicate PCR reactions were run per sample and 5 μl of each PCR product was used as a template in

the second round of PCR using primers designed after Illumina's Nextera XT Index Kit v2 (i5 and i7 indices). The second PCR conditions were 3 minutes at 95°C, followed by 10 cycles of 30 seconds at 95°C, 30 seconds at 55°C, 30 seconds at 72°C, and a final elongation for 5 min at 72°C. Amplicons were purified with Agencourt AMPure XP SPRI magnetic beads (Beckman Coulter, Brea, CA, USA) using a 1:1 bead-to-amplicon ratio. Purified amplicons were quantified with the Qubit 2.0 fluorometer and diluted to 4 nM in 10 mM TRIS before sequencing.

One Ascomycete and one basidiomycete clone library were prepared with the TOPO TA Cloning Kit for Sequencing with pCR4-TOPO vector and One Shot TOP10 Chemically Competent *E. coli* (Invitrogen, Carlsbad, CA, US) according to manufacturer's protocol. These clone libraries were included within each Illumina MiSeq run as positive controls to assess and account for barcode 'index-hopping' (van der Valk et al. 2020) and were also Sanger sequenced (University of California, Berkeley DNA Sequencing Facility).

Volumes and initial concentrations of each PCR to amplify plant ribulose-1,5-bisphosphate carboxylase oxygenase (rbcL; forward: rbcLa-F, reverse: rbcLajf634-R; Kress and Erickson 2007, Fazekas et al. 2008) or internal transcribed spacer 2 (ITS2; forward: ITS-u3; reverse: ITS-u4; Cheng et al. 2016) were as follows: 1 μl of genomic DNA (10 ng/μl), 1.25 μl of each primer (0.5 μM; Integrated DNA Technologies, Coralville, IA, USA), 12.5 μl of GoTaq Green Master Mix (Promega Corporation, USA) and 9 μl of nuclease-free water. PCR conditions were 3 minutes at 95°C, followed by 30 cycles of 30 seconds at 95°C, 30 seconds at 55°C, 30 s at 72°C, and a final elongation of 5 min at 72°C. Amplicons were purified with Exo-AP (New England Biolabs, Ipswitch, MA) according to the manufacturer's protocol.

Sequencing of Endophyte and Plant Communities

Samples were sequenced on the Illumina MiSeq platform with 250 paired-end reads for 500 cycles with a 15% spike-in of PhiX (Institute for Genomic Medicine, University of California, San Diego or the California NanoSystems Institute, University of California, Santa Barbara). Each endophyte community (n = XXXX) was sequenced twice on two separate Illumina MiSeq runs. Replicated libraries across the two MiSeq runs originated from the same purified PCR product and thus had the same i5 and i7 barcode combinations across the two runs. Plant rbcL regions were Sanger sequenced (University of California, Berkeley DNA Sequencing Facility) and plant ITS2 regions were multiplexed for Illumina MiSeq sequencing (California NanoSystems Institute, University of California, Santa Barbara).

Sequence Processing

Forward and reverse reads for each endophyte community were merged with the -*fastq_mergepairs* command in USEARCH version 11.0.667 (Edgar and Flyvbjerg 2015). Merged reads shorter than 150 base pairs were presumed to be chimeric or residual adapter dimers and were excluded. The minimum percent of nucleotide matches required in each alignment was decreased to 65% (from 90%) with the -*fastq_pctid* parameter but still resulted in the mean number of expected errors per merged read being less than one across all reads in a given sequencing run. The maximum number of mismatches allowed in each alignment was increased to 300 with the -*fastq_maxdiffs* parameter such that the -*fastq_mergepairs* command defaulted to the -*fastq_pctid* parameter. Read pairs that were successfully merged were relabeled with their sample ID and numeric identifiers with the -

relabel @ parameter. For Chapter 1, forward and reverse primers were removed with cutadapt version 3.5 (Martin 2011) as linked adapters and merged reads that were untrimmed from either the 5' or 3' end were discarded. For Chapters 2 and 3, forward and reverse primers were removed from merged reads with *-fastx_truncate* command. Merged reads were filtered with a maximum expected error score of 1.0 (*-fastq_maxee* parameter) with the *-fastq_filter* command in USEARCH. Filtered reads across all sequencing runs were concatenated into one file with the *cat* command in terminal. Filtered sequences were "deprelicated" to isolate only unique sequences (i.e., 100% similarity) with the -*fastx_unqiues* command and cluster size annotations were added to each read label with the -*sizeout* parameter in VSEARCH version 2.21.1 (Rognes et al. 2016). Dereplicated sequences were sorted by read abundance and filtered into amplicon sequence variants (ASVs) with the --*cluster_unoise* command in VSEARCH. ASVs that had less than eight replicate sequences (default parameter) were excluded.

For Chapter 1, ASVs that were identified as chimeric were removed with the --*uchime3_denovo* command in VSEARCH with an abundance skew of 16 (default parameter). ASVs were sorted by length with the *-sortbylength* command and clustered into 97% Operational Taxonomic Units (OTUs) with the *-cluster_smallmem* command in USEARCH. For Chapters 2 and 3, cluster size annotations were added to each read label with the *-sizeout* parameter. Unique sequences were sorted by read abundance with the -*sortbysize* command. Unique sequences were run against the UNOISE algorithm (Edgar 2016) to generate 97% OTUs (Chapter 2) or amplicon sequence variants (ASVs; Chapter 3) with the *-unoise3* command. ASVs that did not pass 'denoising' or had less than eight replicate sequences (default parameter) were excluded.

Merged reads trimmed of primers across all sequencing runs were concatenated into a single file and aligned to 97% OTUs with the *--usearch_global* command in USEARCH or VSEARCH, searching from the 5' end of each read with the *--strand plus* parameter and matching at 97% identity measured as the number of matching columns divided by the alignment length with the *--iddef 1* parameter. Bioinformatic scripts were run on the Knot high-performance computing cluster at the University of California, Santa Barbara.

Taxonomic Assignment with BLAST+

ASVs and 97% OTUs were queried against the entire GenBank nucleotide database (January 2022 for Chapter 1; November 2017 for Chapter 2; March 2020 for Chapter 3) with BLAST+ (version 2.12.0 for Chapter 1; version 2.07.1 for Chapters 2 and 3; ref. Camacho et al. 2009) with an e-value of 0.001 on the Knot high-performance computing cluster at the University of California, Santa Barbara.

Taxonomic Filtering with MEGAN6

ASVs or 97% OTUs queried against the GenBank nucleotide database were parsed through MEGAN6 (version 6.22.2 for Chapter 1; version 6.17.0 for Chapter 3; Huson et al. 2007) with default parameters. ASVs or 97% OTUs identified within the kingdom Fungi were used to create a plant-by-endophyte community matrix. ASVs 97% OTUs in the "No hits" or the "Not assigned" categories were manually parsed through BLAST to confirm taxonomic assignments.

Sequencing Positive Controls

Clone libraries were Sanger sequenced at the University of California, Berkeley DNA Sequencing Facility. The top GenBank hit for C1AH with taxonomic information was

Endoconidioma populi (Accession number: KX611029.1; E-value: 3e-130). The top

GenBank hit for C4BT with taxonomic information was *Tremellales* sp. clone (Accession

number: MK282065.1; E-value: 4e-78). Clone Sanger sequences were used to identify the

correct ASV or 97% OTU within each clone library.

C1AH Sanger sequence (GenBank Accession Number MG840195):

ggtcatttagaggaagtaaaagtcgtaacaaggtttccgtaggtgaacctgcggaaggatcattaaagagttagggtccta

gtggcccaacctccaaccctctgtcgttaaaactaccttgttgctttggcggggaccgttcggtctccgagcgcgttaaccttc

gggttggcgagcgcccgccagagtccaaccaaactcttgtattaaaccagtcgtctgagtataaaattttaattaaattaaaa

ctttcaacaacggatctcttggttctcgcaccgatgaagaa

C4BT Sanger sequence (GenBank Accession Number MG840196):

ggtcatttagaggaagtaaaagtcgtaacaaggtttccgtaggtgaacctgcggaaggatcattagtgattggcctcacgg

ccgttctcaaatccacacacaccgtgaaccgttccctcgtcctccggggcgtcggactaccaaccaaacattgtctacaga

acgtgaaatgtataaacgaaacaaaactttcaacaacggatctcttggctctcgca

The read abundance of all other ASVs or 97% OTUs that were not identified as the

C1AH or C4BT clone was presumed to be a result of barcode 'index-hopping' and

subtracted from all samples. For example, if an ASV or 97% OTU was present in a clone

library, but was not C1AH or C4BT, its read abundance was subtracted from all samples

across the respective MiSeq run with a minimum possible read abundance of zero from a

subtraction event.

Library Averaging Across Runs

Sequencing samples that were replicated across the two MiSeq runs were combined by

summing read abundances before rarefying.
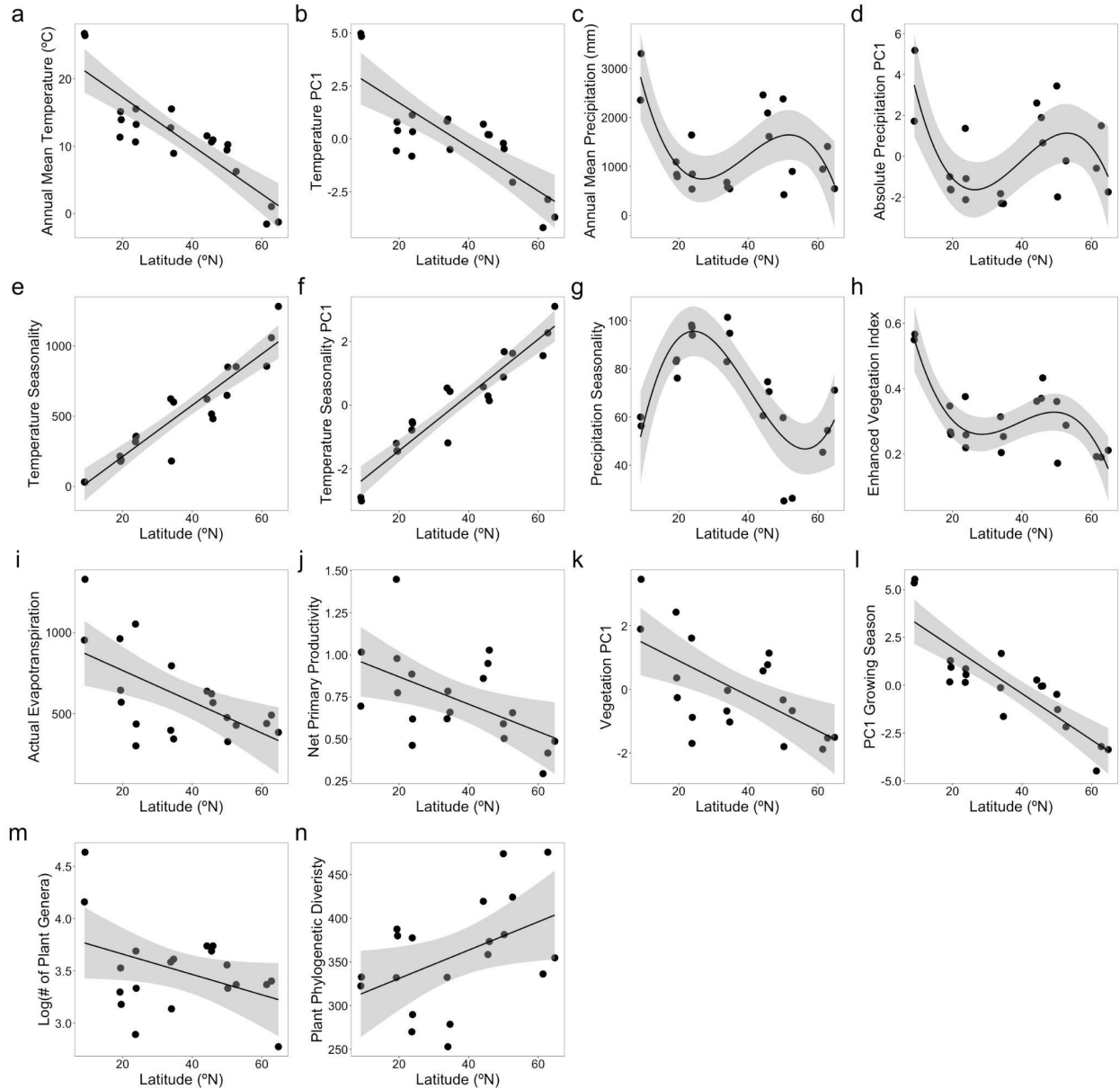
Figure S1. Factors related to climate, the leaf environment, or host diversity with respect to latitude. Each point represents one site (n = 20). Shaded regions refer to 95% confidence intervals.
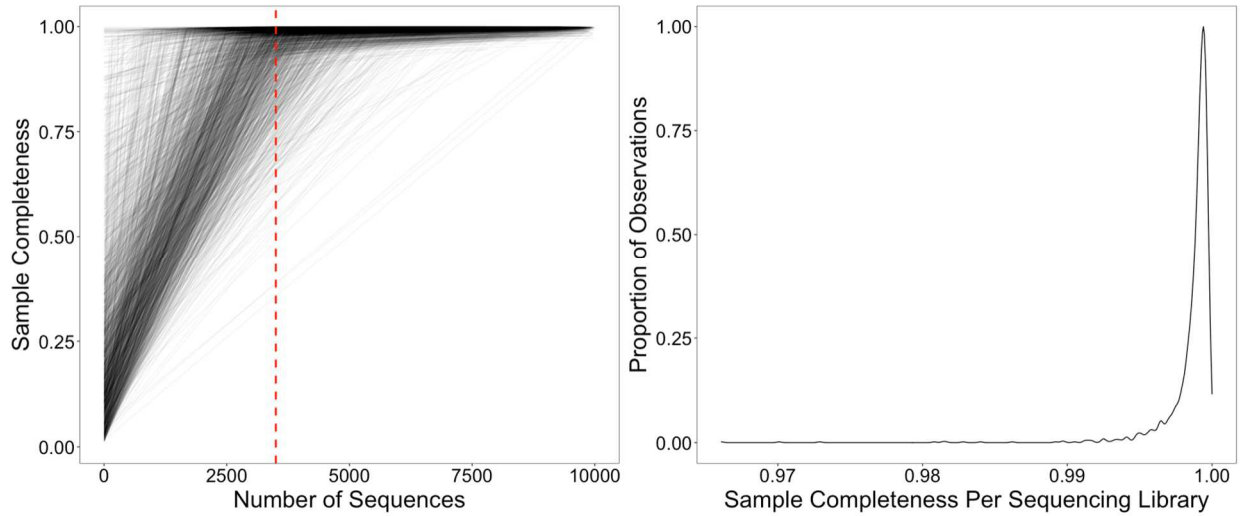
Figure S2. Sample completeness, a metric for how well an endophyte community was characterized by sequencing, in relation to sequencing depth (a) or as a histogram (b). (a) Endophyte communities (n = 1,657; each black line) were rarefied to a sequencing depth of 3500 reads (vertical red dashed line) because sample completeness tended to asymptote near that threshold. (b) The majority of the endophyte communities had sample completeness >99%.
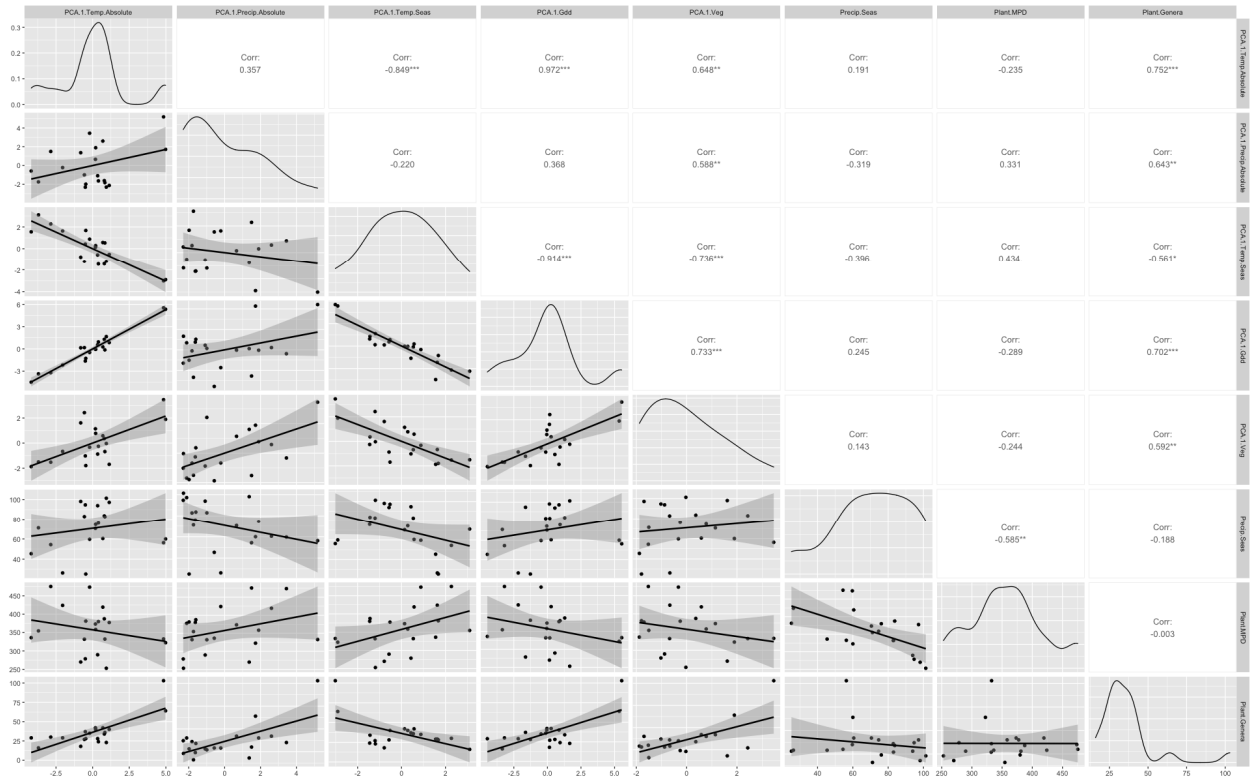
Figure S3. Correlations between factors used for mixed-effects models. Each point represents one site (n = 20). Correlations were measured as Pearson's correlation coefficient. Asterisks refer to statistical significance (* = p < 0.05; ** = p < 0.01; *** = p < 0.001).
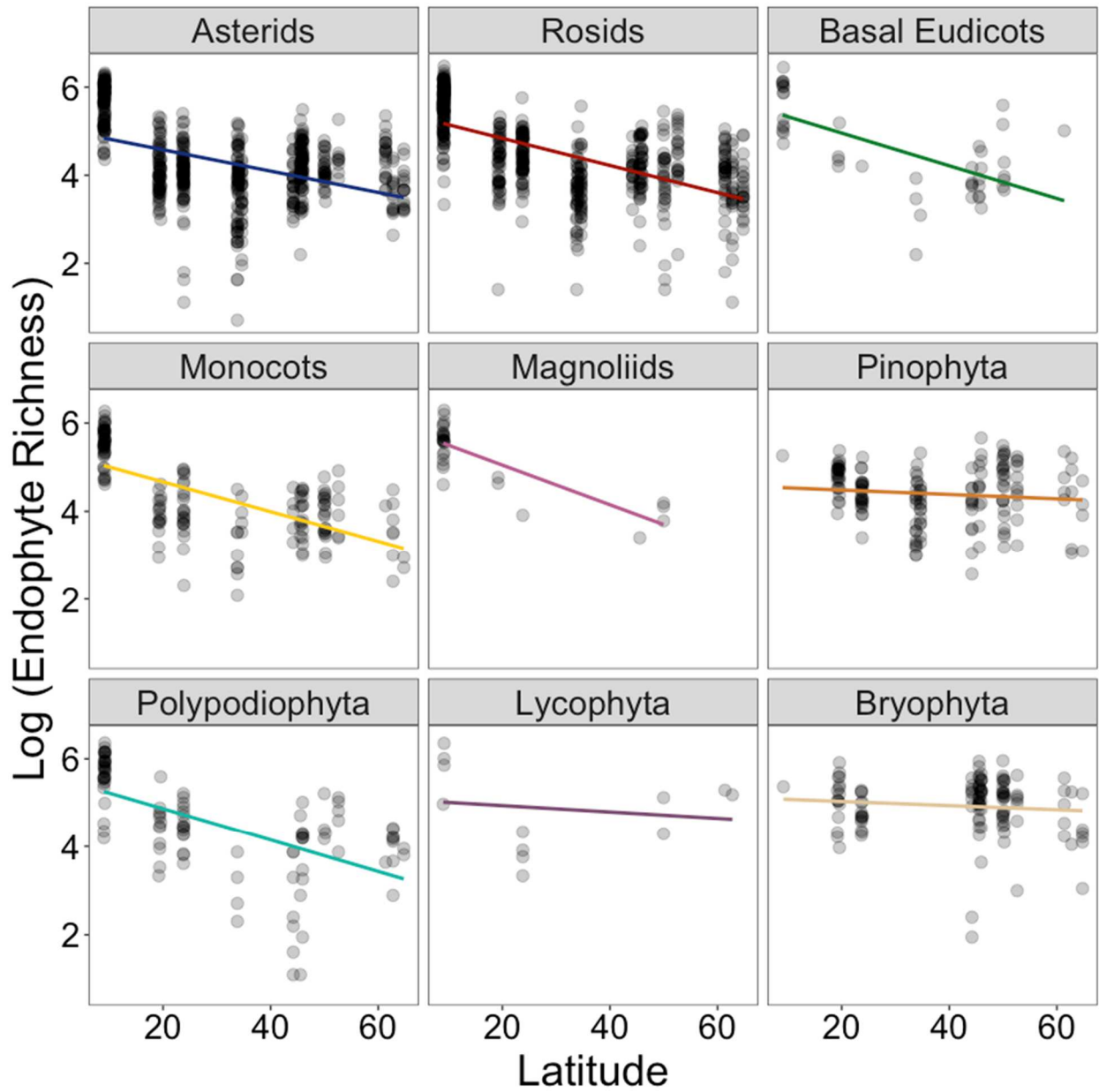
Figure S4. Endophyte richness as a function of latitude per major host group. Each point represents the 97% OTU richness of endophytes within a given host individual. Trends among major host groups are color-coded.
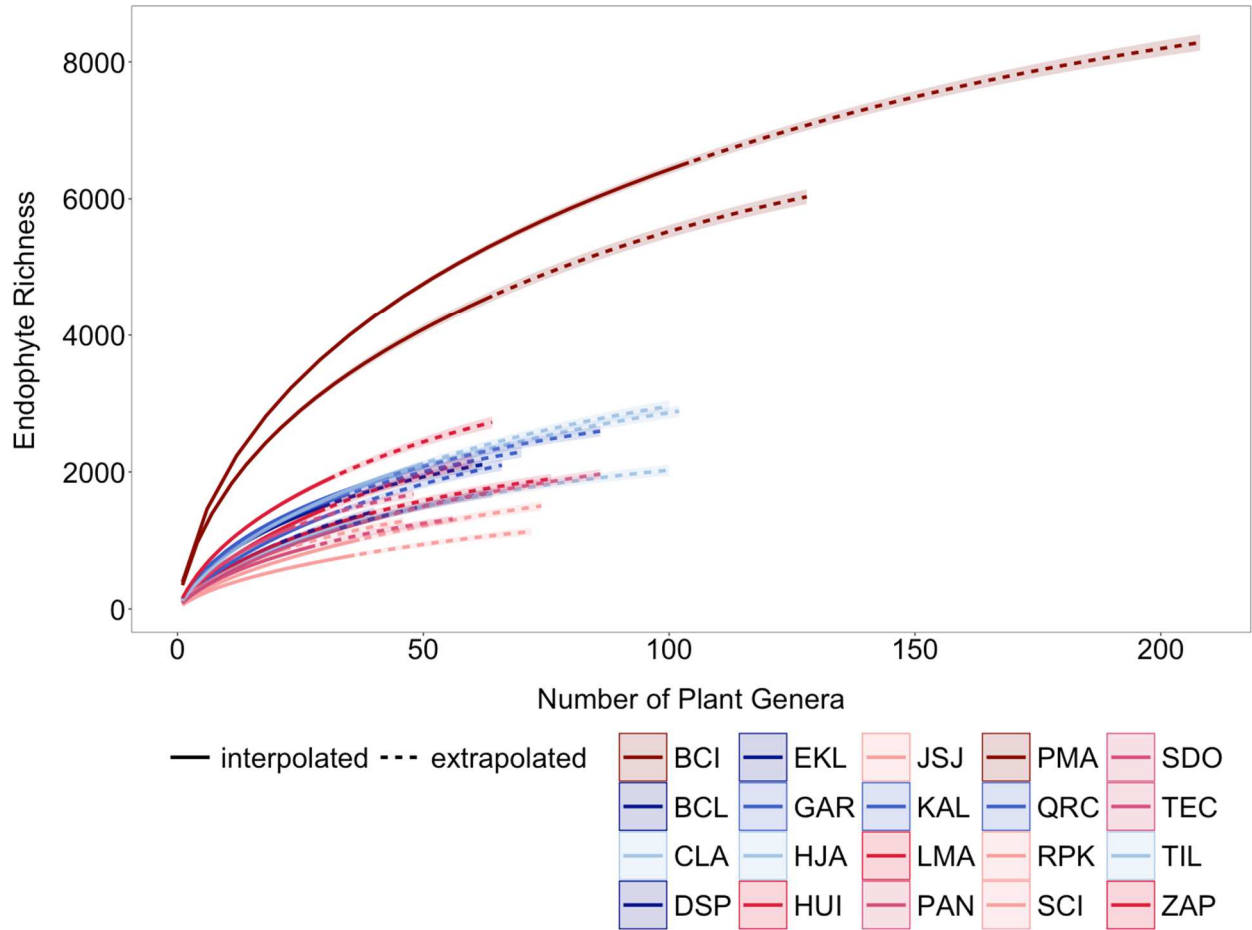
Figure S5. Endophyte communities in tropical ecosystems have greater species richness than endophyte communities in temperate or sub-tropical ecosystems. (a) Each accumulation curve (n = 20 sites) represents the relationship between endophyte 97% OTU richness and sampling of host genera. Solid lines represent interpolated richness values while dotted lines represented extrapolated values with shaded regions referring to 95% confidence intervals. Curves are color-coded by each of the seven sampling regions with darker red shades referring to sites at lower latitudes (Alaska = BCL, DSP, EKL; British Columbia = GAR, KAL, QRC; Oregon = CLA, HJA, TIL; California = JSJ, RPK, SCI; Durango = PAN, SDO, TEC; Tlaxcala = LMA, HUI, ZAP; Panama = BCI, PMA). Within each site, endophyte communities associated with the same host genera were combined and rarefied to 3,500 reads.
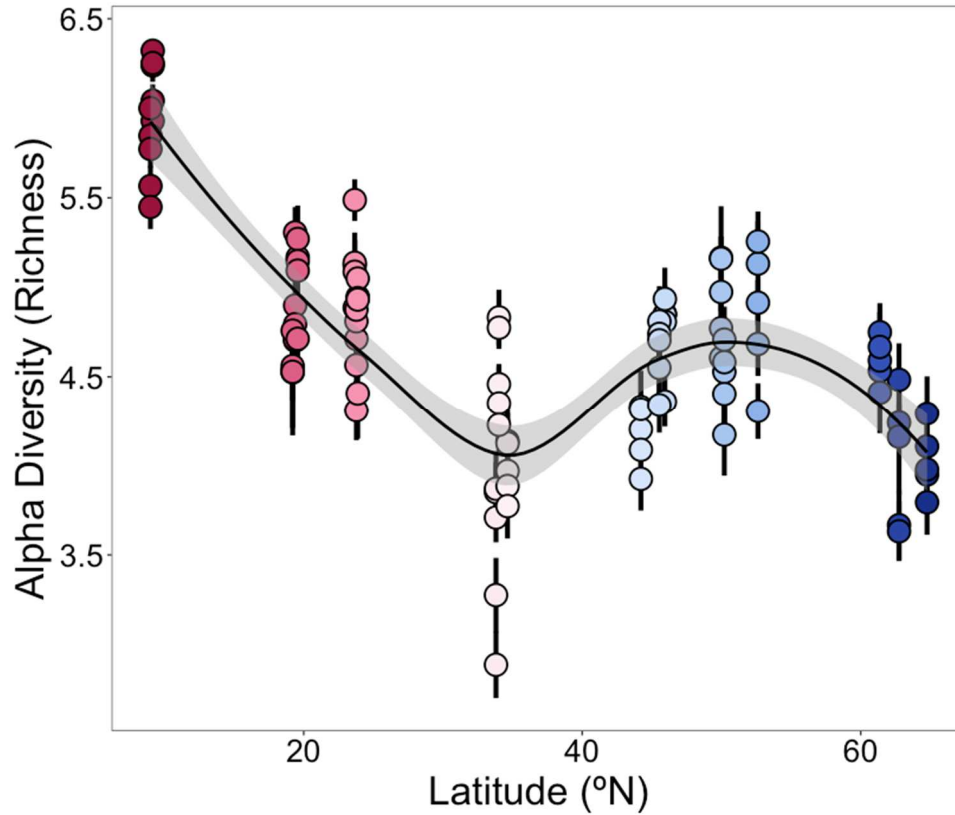
Figure S6. Square-root transformation on endophyte read abundances yields similar patterns as endophyte communities subsampled to 3,500 reads. Each point represents the mean value for endophyte richness (n = 100) with bars referring to standard error. Shaded regions refer to 95% confidence intervals.
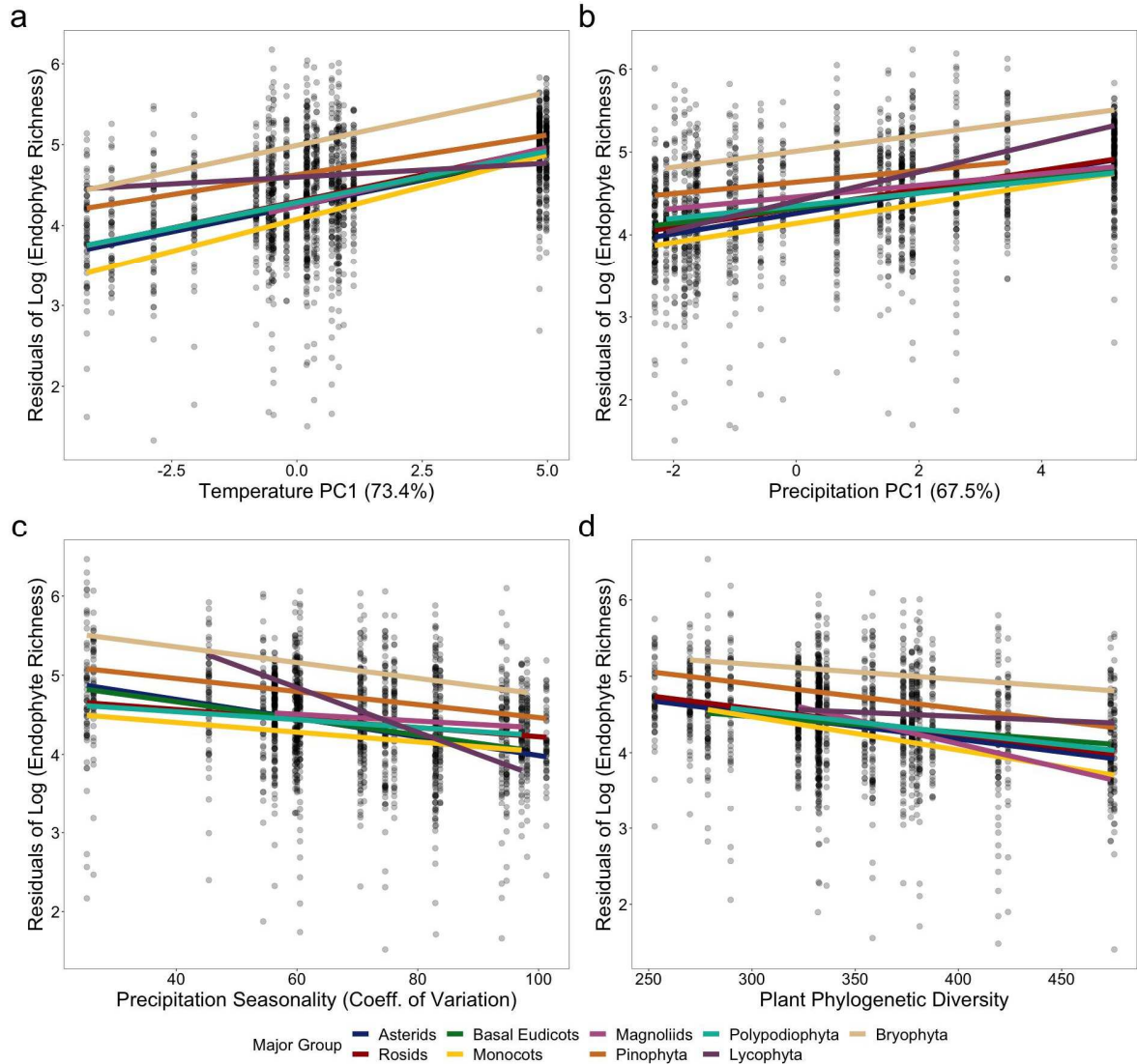
Figure S7. Partial residuals of endophyte richness as a function of (a) temperature PC1, (b) precipitation PC1, (c) precipitation seasonality PC1, or (d) plant phylogenetic diversity visualized per host major group. Partial residual plots derived from a linear mixed-effects model predicting log-transformed 97% OTU endophyte richness while accounting for temperature, precipitation, precipitation seasonality, and plant phylogenetic diversity as fixed effects and quadrat (n = 100) as a random effect (full model: $R^2_m = 0.37$; $R^2_c = 0.53$; Moran's I spatial autocorrelation test: p = 0.23). Each point represents the endophyte community within one host individual (n = 1,657). Each partial residual plot shows the effect of one predictor on endophyte richness when all other predictors are accounted for. Greater principal component (PC) values (a-b) represent greater values for a given predictor variable. Temperature seasonality PC1, leaf environment PC1, growing season PC1, and the number of plant genera at a given site were excluded from the model because they were highly correlated to other predictors (see Methods). Regressions are color-coded by host major group.
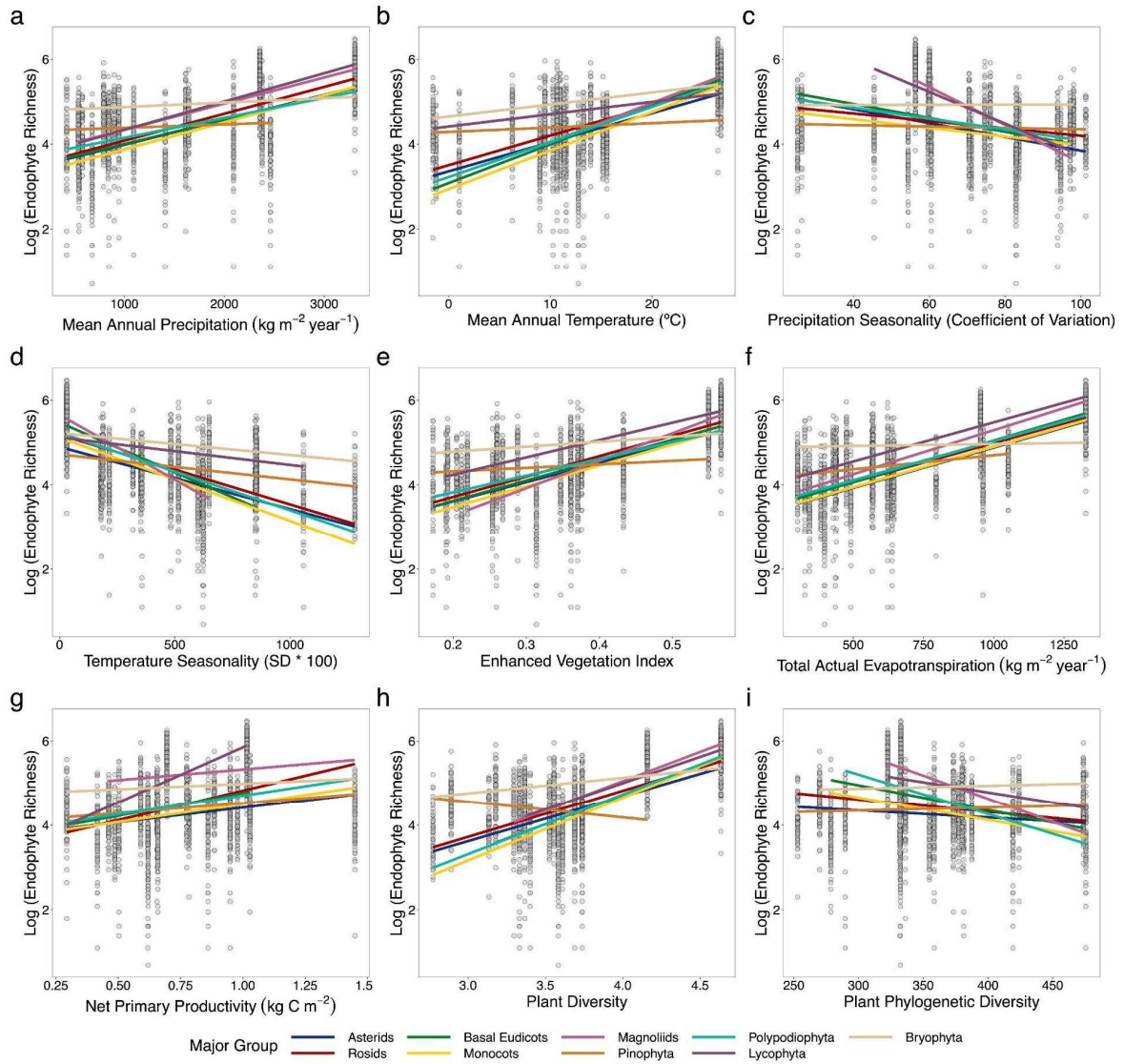
Figure S8. Factors relating to climate (a-d), the leaf environment (e-g), and host diversity (h-i) explain variation in endophyte richness for communities within angiosperm and fern hosts. Each point represents the 97% OTU richness for endophytes within a given host individual. Trends among major host groups are color-coded.
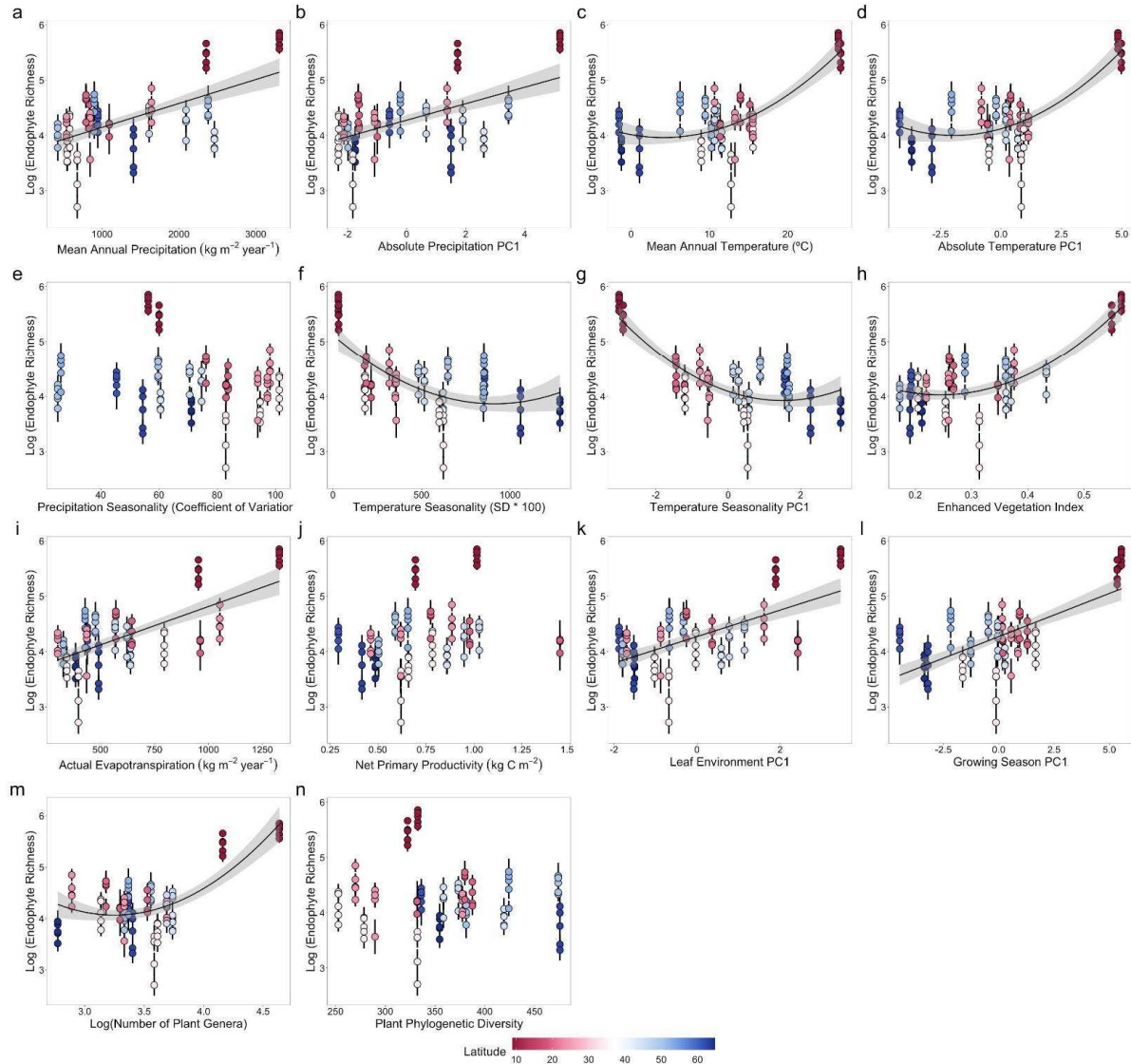
Figure S9. Factors relating to climate (a-d), the leaf environment (e-g), and host diversity (h-i) explain variation in endophyte richness but these trends are contingent on the inclusion of tropical ecosystems. Each point represents the mean endophyte 97% OTU richness per quadrat (n = 100) with bars referring to standard error. Regressions included (solid lines) or excluded (dashed lines) the two Panamanian sites (darkest red points) to test whether the significance of each predictor variable was dependent on the inclusion of tropical ecosystems. The absence of a regression line indicates the parameter estimate for the linear or quadratic term was not statistically significant. Each mixed-effects model considered the 97% OTU richness of endophytes within each host individual (n = 1,657) with site (n = 20) as a random effect. Shaded regions refer to 95% confidence intervals. (a) Endophyte richness with respect to MAP ($R^2_m = 0.37$; $R^2_c = 0.85$). (b) Endophyte richness with respect to absolute precipitation PC1 ($R^2_m = 0.31$; $R^2_c = 0.85$). (c) Endophyte richness with respect to MAT ($R^2_m = 0.46$; $R^2_c = 0.85$). (d) Endophyte richness with respect to absolute temperature PC1 ($R^2_m = 0.46$; $R^2_c = 0.93$). (e) Endophyte richness with respect to precipitation seasonality ($R^2_m = 0.039$; $R^2_c = 0.85$). (f) Endophyte richness

143

with respect to temperature seasonality ($R^2_m = 0.43$; $R^2_c = 0.85$). (g) Endophyte richness with respect to temperature seasonality PC1 ($R^2_m = 0.57$; $R^2_c = 0.85$). (h) Endophyte richness with respect to Enhanced Vegetation Index ($R^2_m = 0.60$; $R^2_c = 0.85$). (i) Endophyte richness with respect to evapotranspiration ($R^2_m = 0.43$; $R^2_c = 0.85$). (j) Endophyte richness with respect to net primary productivity ($R^2_m = 0.071$; $R^2_c = 0.85$). (k) Leaf environment PC1 ($R^2_m = 0.38$; $R^2_c = 0.85$). (l) Endophyte richness with respect to leaf environment PC1 ($R^2_m = 0.60$; $R^2_c = 0.85$). (m) Endophyte richness with respect to the log-transformed number of host genera at a given site ($R^2_m = 0.48$; $R^2_c = 0.85$). (n) Endophyte richness with respect to host phylogenetic diversity measured as the mean pairwise phylogenetic distance among all host genera at a given site ($R^2_m = 0.0054$; $R^2_c = 0.85$).
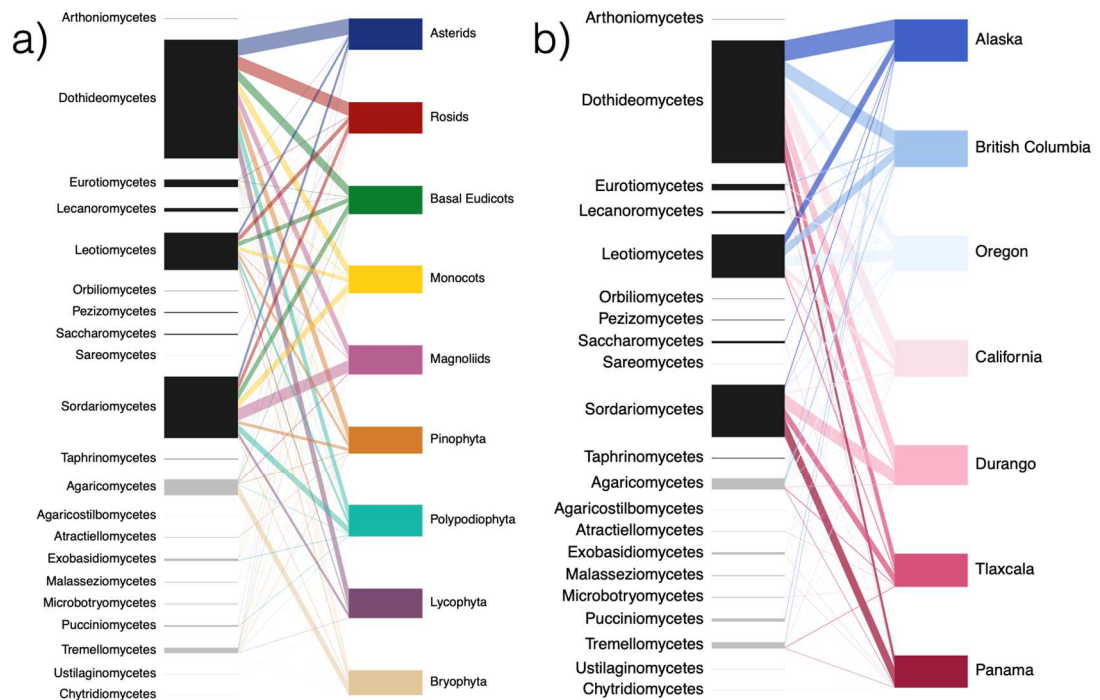
Figure S10. Endophytes varied in distribution depending on host group and geographic region. (a) Endophyte 97% OTUs with a relative read abundance >0.01% (n = 1,309 OTUs) grouped by taxonomic class (left bars; n = 21) and their associations to major host groups (right bars; n = 9). (b) Endophyte 97% OTUs with a relative read abundance >0.01% (n = 1,309 OTUs) grouped by taxonomic class (left bars; n = 21) and their distributions among sampling regions (right bars; n = 7). The width of each bar or pairwise connection was standardized by dividing the number of reads between each host group or sampling region by the number of host individuals sampled within that given group (e.g., number of reads in asterids / number of asterid individuals sampled). Fungal classes are color-coded by phylum (black = Ascomycota; grey = Basidiomycota and Chytridiomycota).
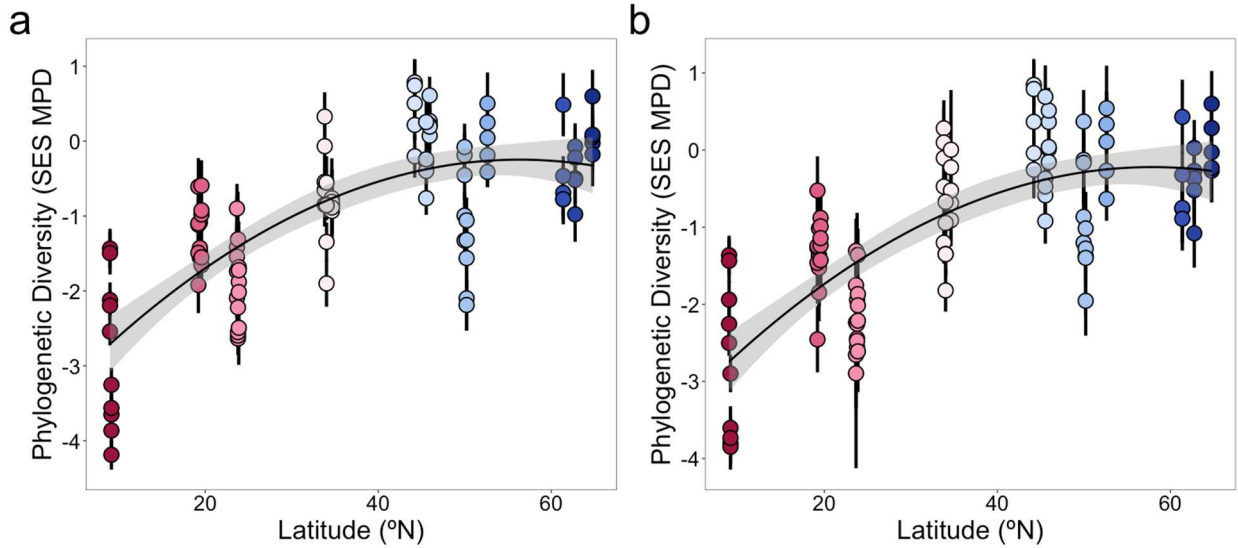
Figure S11. Phylogenetic diversity of Pezizomycotina endophytes decreases towards the tropics. Phylogenetic diversity was measured as the standardized effect size of the mean pairwise phylogenetic distance for endophytes within (a) all hosts or (b) angiosperms only. Each point is the mean phylogenetic diversity for Pezizomycotina endophytes in a given quadrat. Lower values indicate greater phylogenetic clustering of endophyte communities than expected by chance.

Appendix 3: Supplementary Tables 1-6 for Chapter 2.

Structural Specificity: Host Richness

Percent Removed	Phylum	N (OTUs)	Mean	Standard Error	W statistic	p-value
0	Ascomycota	1087	-4.129	0.202	147200	0.004
	Basidiomycota	301	-2.618	0.182		
1	Ascomycota	458	-7.611	0.424	24332	0.001
	Basidiomycota	132	-4.220	0.363		
5	Ascomycota	208	-11.279	0.750	4204.5	0.022
	Basidiomycota	51	-6.588	0.755		
10	Ascomycota	135	-12.711	0.975	1210.5	0.282
	Basidiomycota	21	-8.810	1.519		
15	Ascomycota	97	-14.629	1.188	398	0.351
	Basidiomycota	10	-10.200	2.533		
20	Ascomycota	71	-16.155	1.401	198.5	0.381
	Basidiomycota	7	-10.857	2.668		

Table S1. P-values for Figure 9a-b with presence-absence data using a two-sample Wilcoxon Rank Sum test.

Network Specificity: Resource Range Index

Percent Removed	Phylum	N (OTUs)	Mean	Standard Error	W statistic	p-value
0	Ascomycota	1087	0.913	0.006	147200	0.004
	Basidiomycota	301	0.955	0.088		
1	Ascomycota	458	0.816	0.012	24332	0.001
	Basidiomycota	132	0.911	0.116		
5	Ascomycota	208	0.714	0.021	4204.5	0.022
	Basidiomycota	51	0.845	0.150		
10	Ascomycota	135	0.675	0.027	1210.5	0.282
	Basidiomycota	21	0.783	0.193		
15	Ascomycota	97	0.621	0.033	398	0.351
	Basidiomycota	10	0.744	0.223		
20	Ascomycota	71	0.579	0.039	198.5	0.381
	Basidiomycota	7	0.726	0.196		

**Table S2.** P-values for Figure 9c-d with presence-absence data using a two-sample Wilcoxon Rank Sum test.

Phylogenetic Specificity: Presence-absence Mean Pairwise Phylogenetic Distance

Percent Removed	Phylum	N (OTUs)	Mean	Standard Error	W statistic	p-value
0	Ascomycota	1053	-0.477	0.018	141920	0.057
	Basidiomycota	289	-0.405	0.565		
1	Ascomycota	456	-0.742	0.024	28242	0.669
	Basidiomycota	127	-0.715	0.605		
5	Ascomycota	206	-0.863	0.031	4777.5	0.161
	Basidiomycota	53	-0.781	0.511		
10	Ascomycota	134	-0.889	0.037	1166	0.349
	Basidiomycota	20	-0.795	0.478		
15	Ascomycota	96	-0.871	0.040	488.5	0.927
	Basidiomycota	10	-0.958	0.462		
20	Ascomycota	71	-0.897	0.039	176.5	0.208
	Basidiomycota	7	-0.765	0.354		

Table S3. P-values for Figure 9e-f with presence-absence data using a two-sample Wilcoxon Rank Sum test.

Structural Specificity: Shannon's *H*

Percent Removed	Phylum	N (OTUs)	Mean	Standard Error	W statistic	p-value
0	Ascomycota	1087	-0.540	0.024	148610	0.009
	Basidiomycota	301	-0.388	0.561		
1	Ascomycota	458	-0.942	0.044	25560	0.006
	Basidiomycota	132	-0.659	0.654		
5	Ascomycota	208	-1.194	0.071	4590	0.136
	Basidiomycota	51	-0.914	0.754		
10	Ascomycota	135	-1.261	0.091	1267.5	0.436
	Basidiomycota	21	-1.041	0.872		
15	Ascomycota	97	-1.382	0.111	395.5	0.338
	Basidiomycota	10	-1.013	0.931		
20	Ascomycota	71	-1.450	0.131	211	0.512
	Basidiomycota	7	-1.094	0.783		

Table S4. P-values for Figure 10a-b with abundance-weighted data using a two-sample Wilcoxon Rank Sum test.

Network Specificity: Paired Difference Index

Percent Removed	Phylum	N (OTUs)	Mean	Standard Error	W statistic	p-value
0	Ascomycota	1087	0.981	0.001	147960	0.006
	Basidiomycota	301	0.991	0.016		
1	Ascomycota	458	0.965	0.003	25562	0.007
	Basidiomycota	132	0.985	0.020		
5	Ascomycota	208	0.951	0.006	4506	0.096
	Basidiomycota	51	0.979	0.023		
10	Ascomycota	135	0.943	0.008	1251.5	0.389
	Basidiomycota	21	0.974	0.028		
15	Ascomycota	97	0.934	0.011	372.5	0.229
	Basidiomycota	10	0.979	0.025		
20	Ascomycota	71	0.926	0.013	202	0.416
	Basidiomycota	7	0.979	0.023		

Table S5. P-values for Figure 10c-d with abundance-weighted data using a two-sample Wilcoxon Rank Sum test.

Phylogenetic Specificity: Abundance-weighted Mean Pairwise Phylogenetic Distance

Percent Removed	Phylum	N (OTUs)	Mean	Standard Error	W statistic	p-value
0	Ascomycota	1053	-0.242	0.010	139870	0.023
	Basidiomycota	289	-0.189	0.300		
1	Ascomycota	456	-0.387	0.017	26289	0.110
	Basidiomycota	127	-0.336	0.346		
5	Ascomycota	206	-0.465	0.026	4864.5	0.221
	Basidiomycota	53	-0.394	0.357		
10	Ascomycota	134	-0.473	0.033	1126	0.250
	Basidiomycota	20	-0.379	0.333		
15	Ascomycota	96	-0.512	0.039	340.5	0.132
	Basidiomycota	10	-0.315	0.301		
20	Ascomycota	71	-0.528	0.044	162.5	0.133
	Basidiomycota	7	-0.306	0.312		

Table S6. P-values for Figure 10e-f with abundance-weighted data using a two-sample Wilcoxon Rank Sum test.
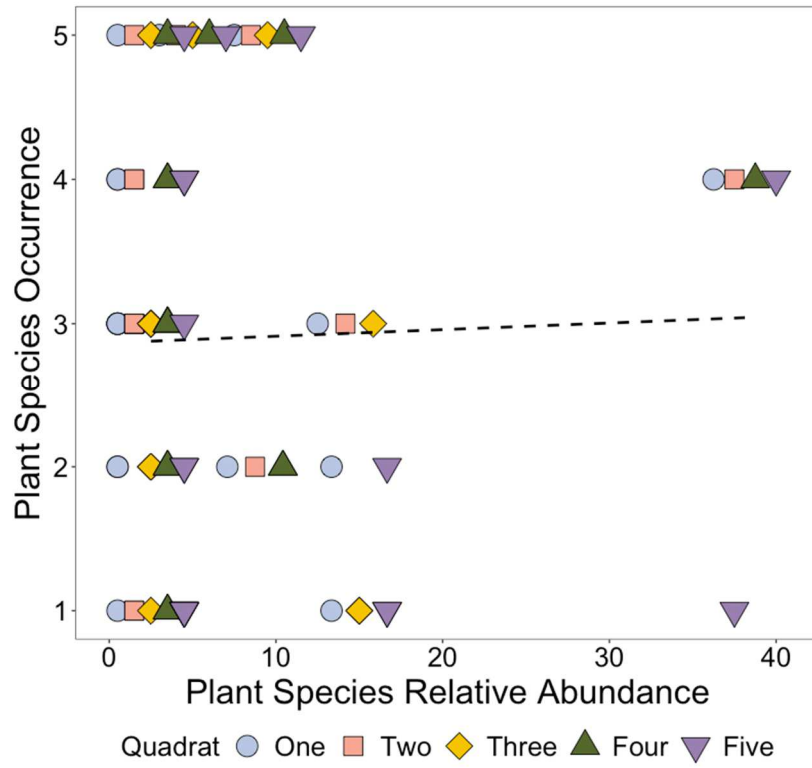
Figure S1. Correlation between plant species occurrence across five quadrats and average plant species relative abundance within quadrats. Plant species occurrence across quadrats was not correlated with average relative abundance within quadrats (Spearman's rho: 0.079, *P* = 0.48). Each point represents one plant individual (n = 80 individuals). Plant individuals are colored- and shape-coded by quadrat and jittered along the x-axis for clarity.
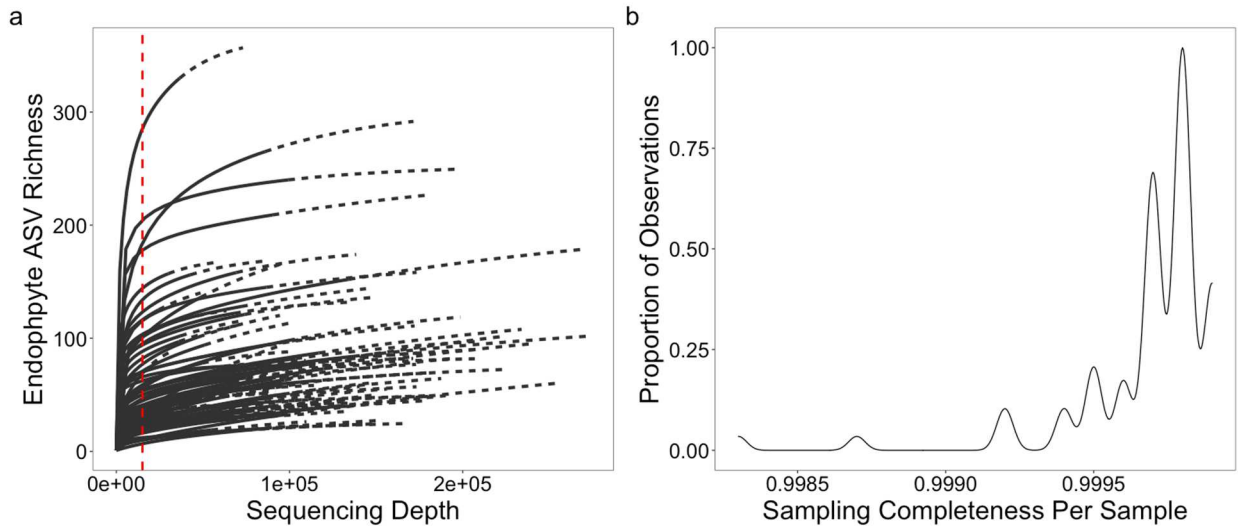
Figure S2. Sampling depth and completeness per plant individual. (a) Rarefaction curves were evaluated per plant individual. Solid and dashed black lines represent interpolated and extrapolated rarefaction curves, respectively. The red dashed line at 15 075 reads indicates the threshold plant individuals were rarefied to. (b) Histogram of sampling completeness per sample calculated from interpolated and extrapolated rarefaction curves. Sampling completeness is an estimate for how well a sample was characterized by sequencing with a maximum of one. All plant individuals were sequenced to greater than 99% sampling completeness.
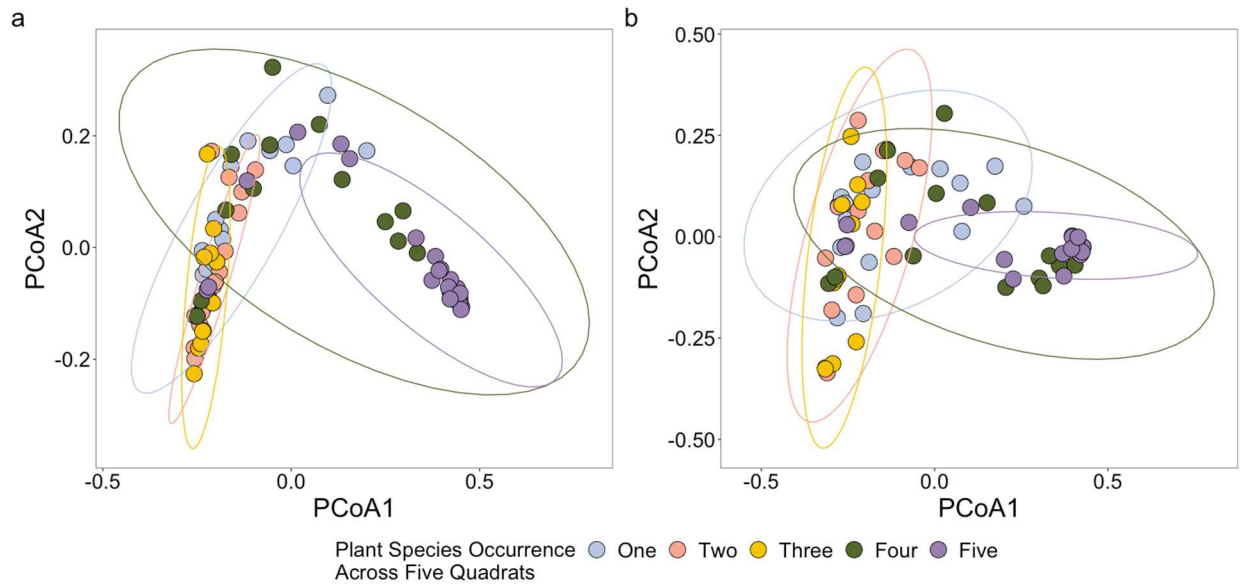
Figure S3. Principal coordinate analysis of endophyte communities grouped by quadrat. Quadrat was a significant predictor of endophyte composition by PERMANOVA with (a) Jaccard ($R^2 = 0.14$; $P = 0.001$) and (b) Bray-Curtis ($R^2 = 0.16$; $P = 0.001$) indices. Each point represents the endophyte community of one plant individual (n = 80 individuals).
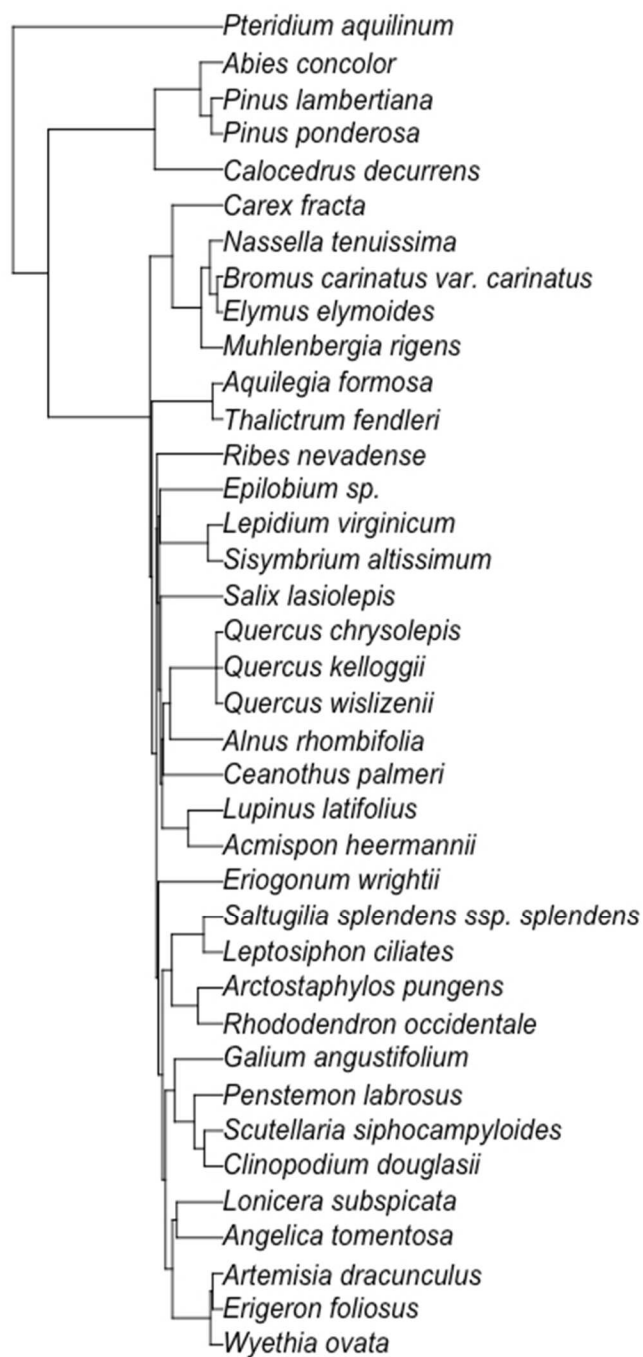
Figure S4. Phylogenetic tree of plant species in this study used to calculate phylogenetic specificity. The phylogenetic tree used to calculate phylogenetic specificity was pruned from a backbone phylogeny representing all extant vascular plant families in North America with the 'phylo.maker' function in the V.PhyloMaker package (Jin and Qian 2019).
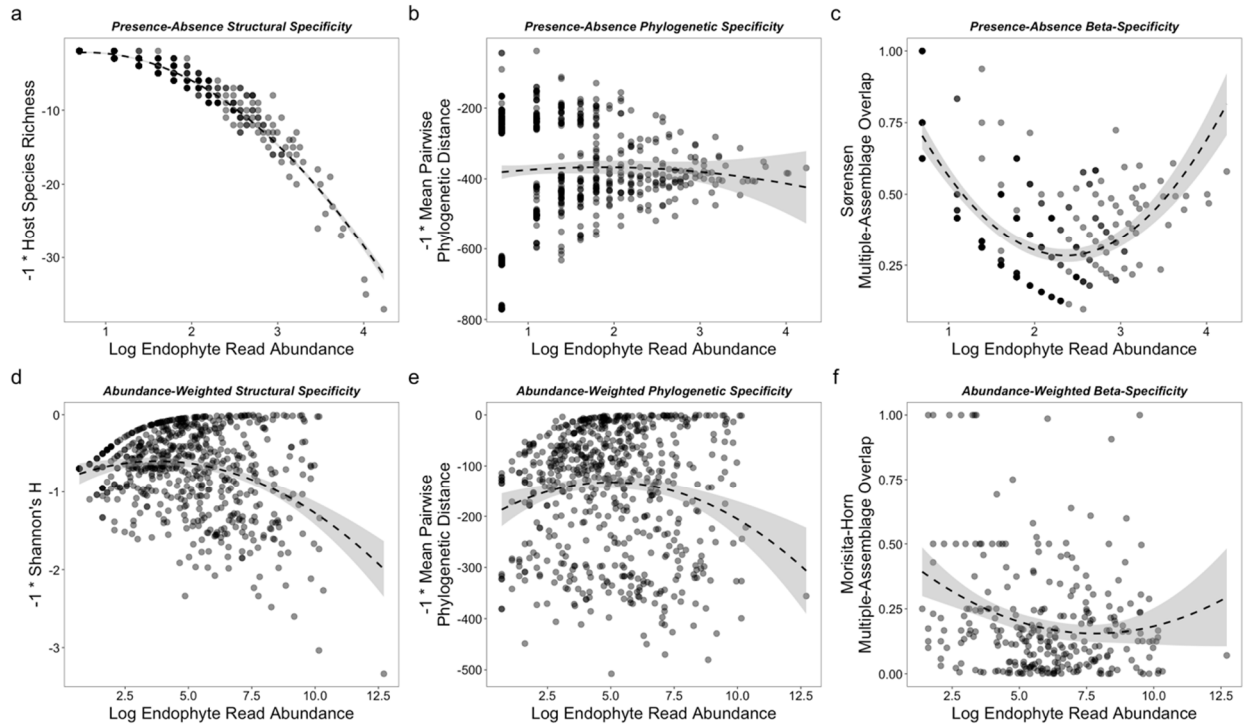
Figure S5. Relationships between absolute structural, phylogenetic or beta specificity and log endophyte read abundance. Each point represents one endophyte ASV. Log-transformed endophyte read abundances represent an endophyte's read abundance across the entire plant community after rarefying to 15 075 reads per sample. Dashed lines represent quadratic regressions and shaded areas represent 95% confidence intervals. (a) Relationship between presence-absence structural specificity measured as -1 * host species richness and log endophyte read abundance. (b) Relationship between presence-absence phylogenetic specificity measured as -1 * Mean Pairwise Phylogenetic Distance and log endophyte read abundance. (c) Relationship between presence-absence beta specificity measured as Sørensen Multiple-Assemblage Overlap and log endophyte read abundance. (d) Relationship between abundance-weighted structural specificity measured as -1 * Shannon's H and log endophyte read abundance. (e) Relationship between abundance-weighted phylogenetic specificity measured as -1 * Mean Pairwise Phylogenetic Distance and log endophyte read abundance. (f) Relationship between abundance-weighted beta specificity measured as Morisita-Horn Multiple-Assemblage Overlap and log endophyte read abundance.
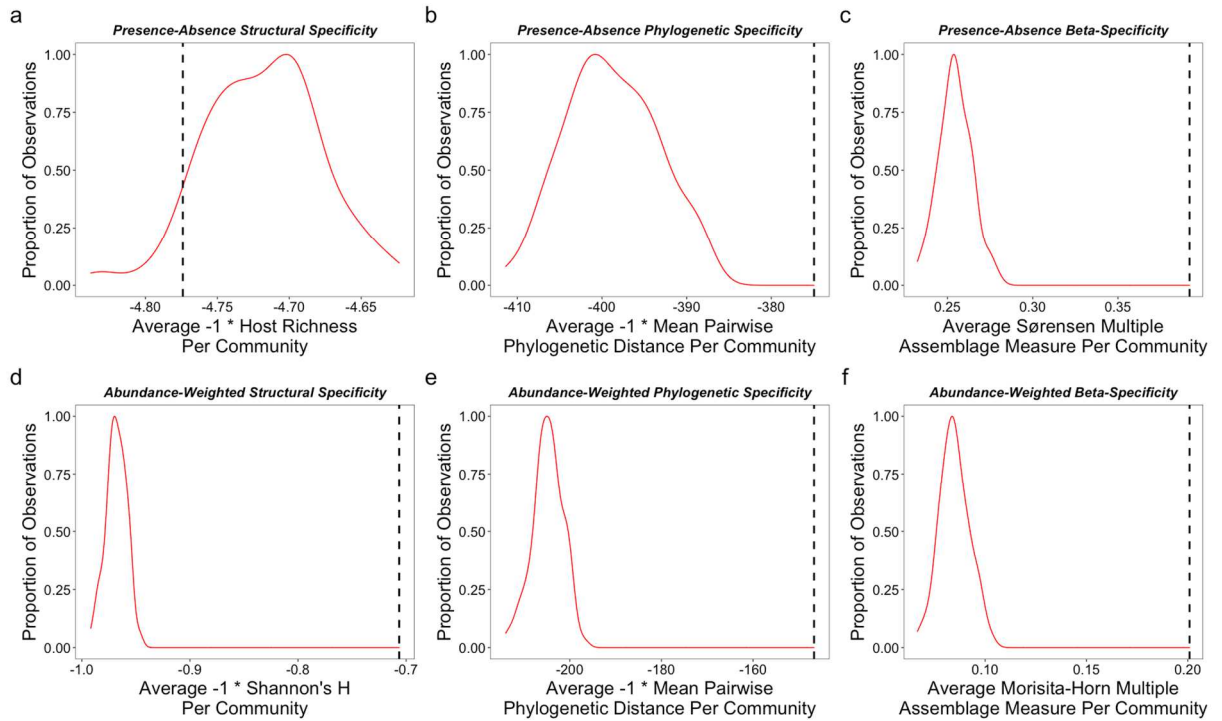
Figure S6. Distributions of the average absolute host specificity of randomized communities relative to the average absolute host specificity of the observed community. The solid red line represents the distribution of average host specificity from each of the 100 randomized communities. The dashed black line represents the absolute host specificity value averaged across all endophyte ASVs from the observed community. (a) Presence-absence structural specificity measured as -1 * host species richness. (b) Presence-absence phylogenetic specificity measured as -1 * Mean Pairwise Phylogenetic Distance. (c) Presence-absence beta specificity measured as Sørensen Multiple-Assemblage Overlap. (d) Abundance-weighted structural specificity measured as -1 * Shannon's H. (e) Abundance-weighted phylogenetic specificity measured as -1 * Mean Pairwise Phylogenetic Distance. (f) Abundance-weighted beta specificity measured as Morisita-Horn Multiple-Assemblage Overlap. Average host specificities of randomized communities were significantly different from the average host specificity of the observed community for all host specificity metrics with a one-sample t-test ($P < 0.05$). Normality was assessed with a Shapiro-Wilk test ($P > 0.05$).
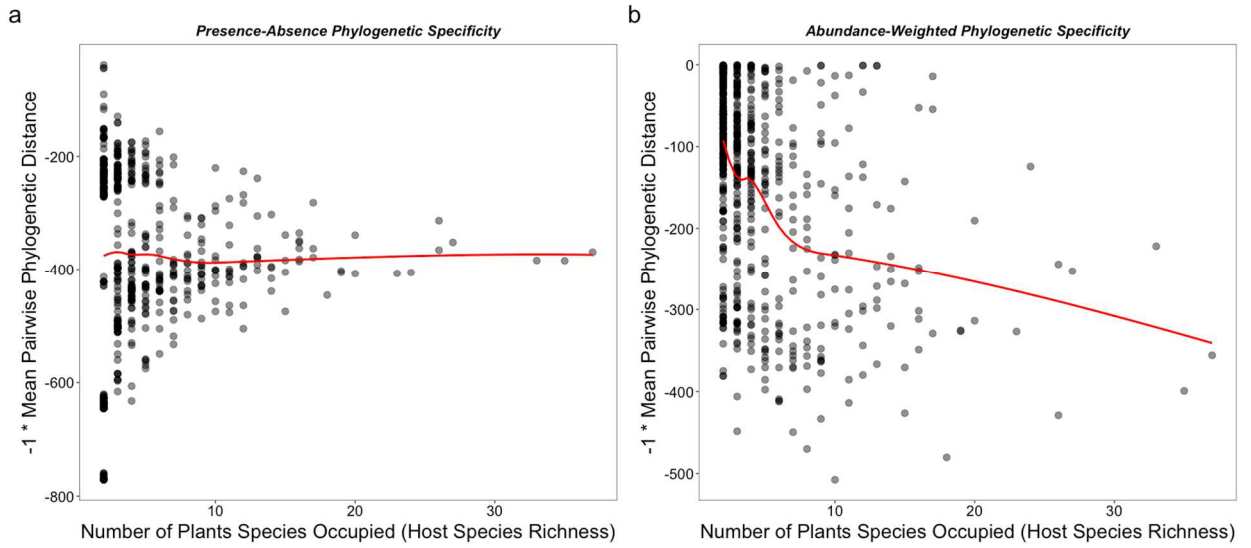
Figure S7. Relationships between absolute phylogenetic specificity and endophyte host species richness. Each point represents one endophyte ASV. Correlations between (a) presence-absence and (b) abundance-weighted phylogenetic specificity measured as -1 * Mean Pairwise Phylogenetic Distance and endophyte host species richness, or structural specificity. Variance in mean pairwise phylogenetic distance decreases as the number of plants species an endophyte occupies increases.
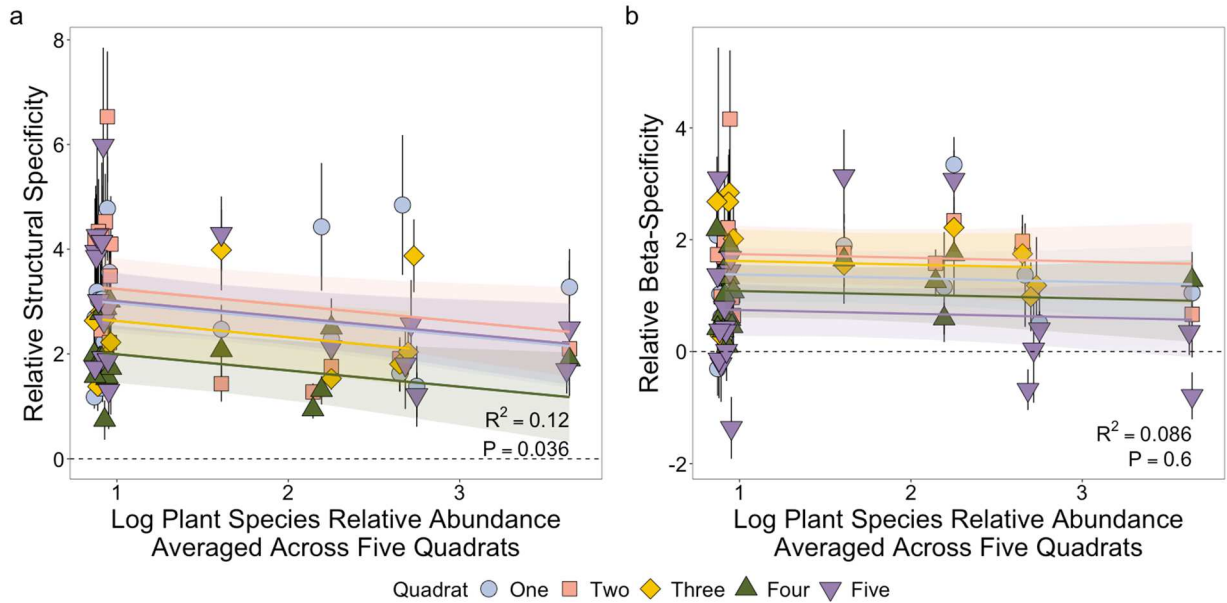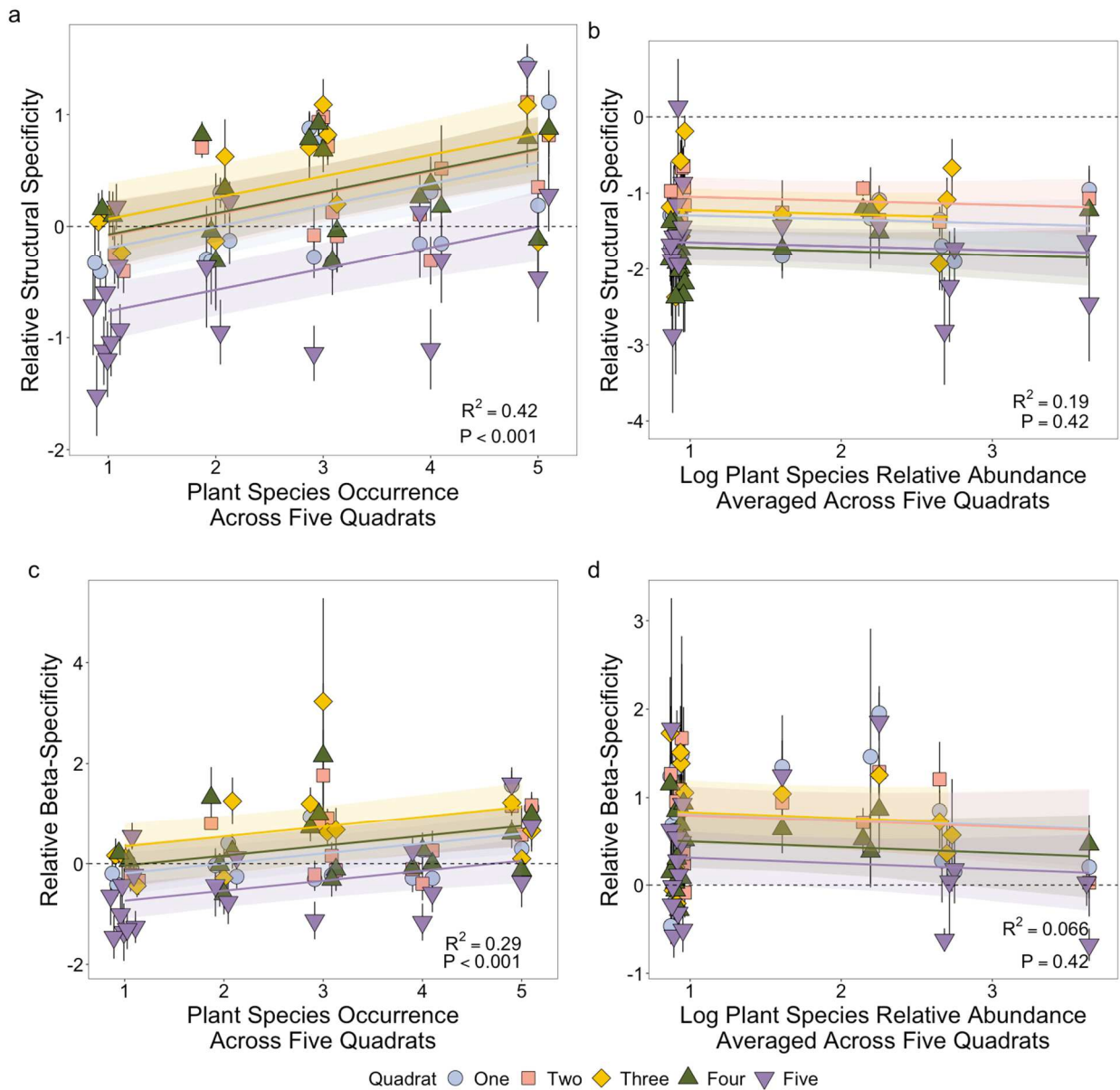
Figure S8. Relative structural or beta specificity and their relationships with average plant species relative abundance within quadrats. Regressions of (a) relative structural specificity measured as the standardized effect size of -1 * Shannon's H ($R^2 = 0.12$, $P = 0.036$) and (b) relative beta specificity measured as Morisita-Horn Multiple-Assemblage Overlap Measure ($R^2 = 0.086$, $P = 0.60$) as a function of log-transformed plant species relative abundance. Each point represents the average relative structural or beta specificity value among all endophyte ASVs of one plant individual (n = 80 individuals) with standard error bars. Positive relative host specificities (above the dashed horizontal lines) indicate higher host specificity than expected by chance. Plant individuals are color- and shape-coded by quadrat location and jittered along the x-axis by plant species. Linear regressions include quadrat as a covariate with shaded regions that refer to 95% confidence intervals.
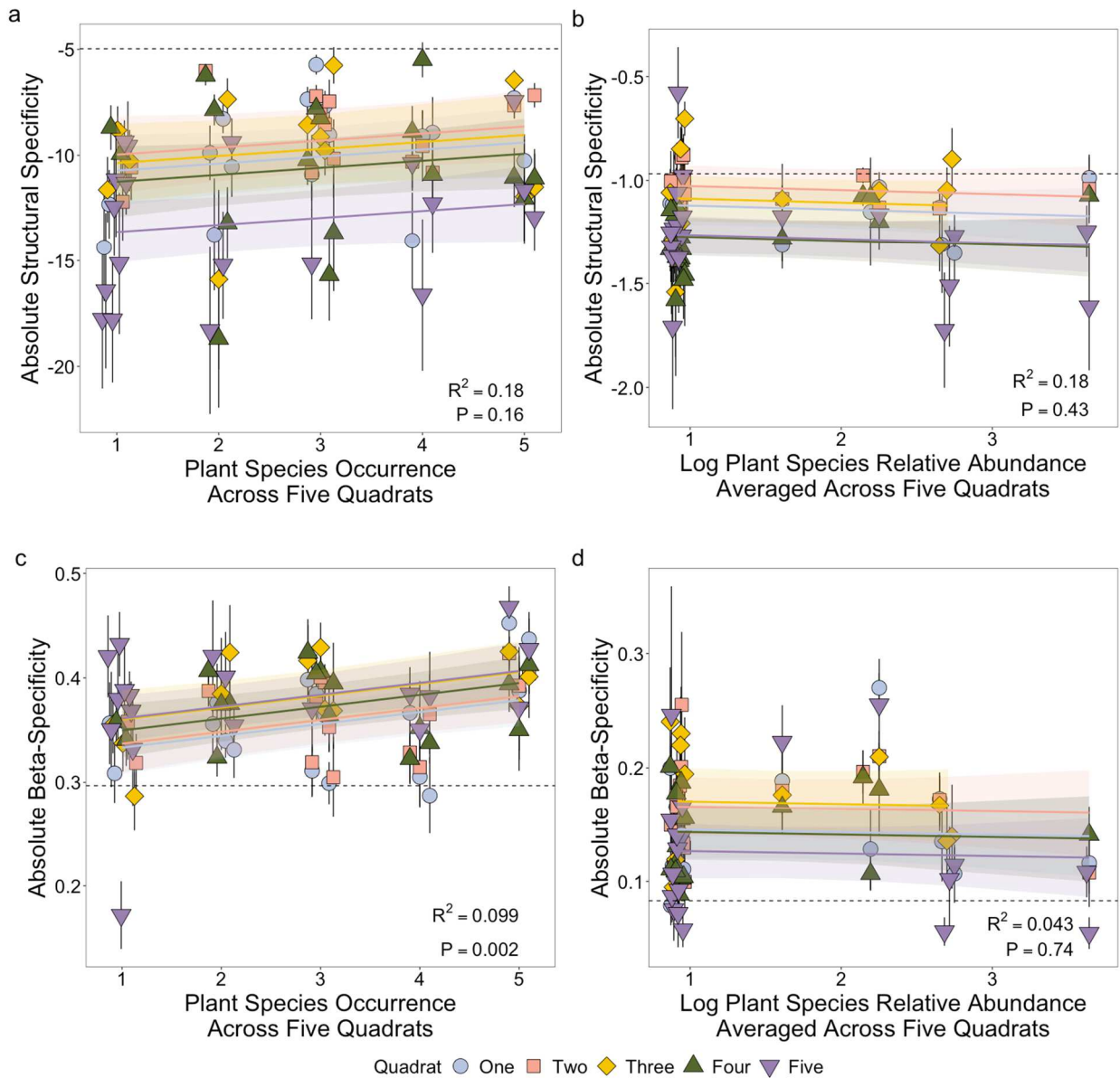
Figure S9. Absolute structural or beta specificity and their relationships with average plant species relative abundance within quadrats. (a) Regression of absolute structural specificity measured as -1 * host species richness as a function of plant species occurrence ($R^2 = 0.18$, $P = 0.16$). (b) Regression of absolute structural specificity measured as -1 * Shannon's H ($R^2 = 0.18$, $P = 0.43$) as a function of log-transformed plant species relative abundance. (c) Regression of absolute beta specificity measured as Sørensen Multiple-Assemblage Overlap Measure as a function of plant species occurrence ($R^2 = 0.099$, $P = 0.002$). (d) Regression of absolute beta specificity measured as -1 * Shannon's H ($R^2 = 0.043$, $P = 0.74$) as a function of log-transformed plant species relative abundance. Each point represents the average absolute structural or beta specificity value among all endophyte ASVs of one plant individual (n = 80 individuals) with standard error bars. Positive relative host specificities (above the dashed horizontal lines) indicate higher host

159

specificity than expected by chance. Plant individuals are color- and shape-coded by quadrat location and jittered along the x-axis by plant species. Linear regressions include quadrat as a covariate with shaded regions that refer to 95% confidence intervals.

Figure S10. Relative structural specificity or relative beta specificity measured with an alternative null model and their relationships with plant species abundance. (a) Regression of relative structural specificity measured as the standardized effect size of -1 * host species richness as a function of plant species occurrence ($R^2 = 0.42$, $P < 0.001$). (b) Regression of relative structural specificity measured as -1 * Shannon's H ($R^2 = 0.19$, $P = 0.42$) as a function of log-transformed plant species relative abundance. (c) Regression of relative beta specificity measured as Sørensen Multiple-Assemblage Overlap Measure as a function of plant species occurrence ($R^2 = 0.29$, $P < 0.001$). (d) Regression of relative beta specificity measured as -1 * Shannon's H ($R^2 = 0.066$, $P = 0.42$) as a function of log-transformed plant species relative abundance. Each point represents the average relative structural or beta specificity value among all endophyte ASVs of one plant individual (n = 80 individuals) with standard error bars. Positive relative host specificities (above the dashed

161

horizontal lines) indicate higher host specificity than expected by chance. Plant individuals are color- and shape-coded by quadrat location and jittered along the x-axis by plant species. Linear regressions include quadrat as a covariate with shaded regions that refer to 95% confidence intervals. The plant-endophyte community was randomized with the "shuffle.web" (abundance-weighted) algorithm using the "nullmodel" function in the *bipartite* package.
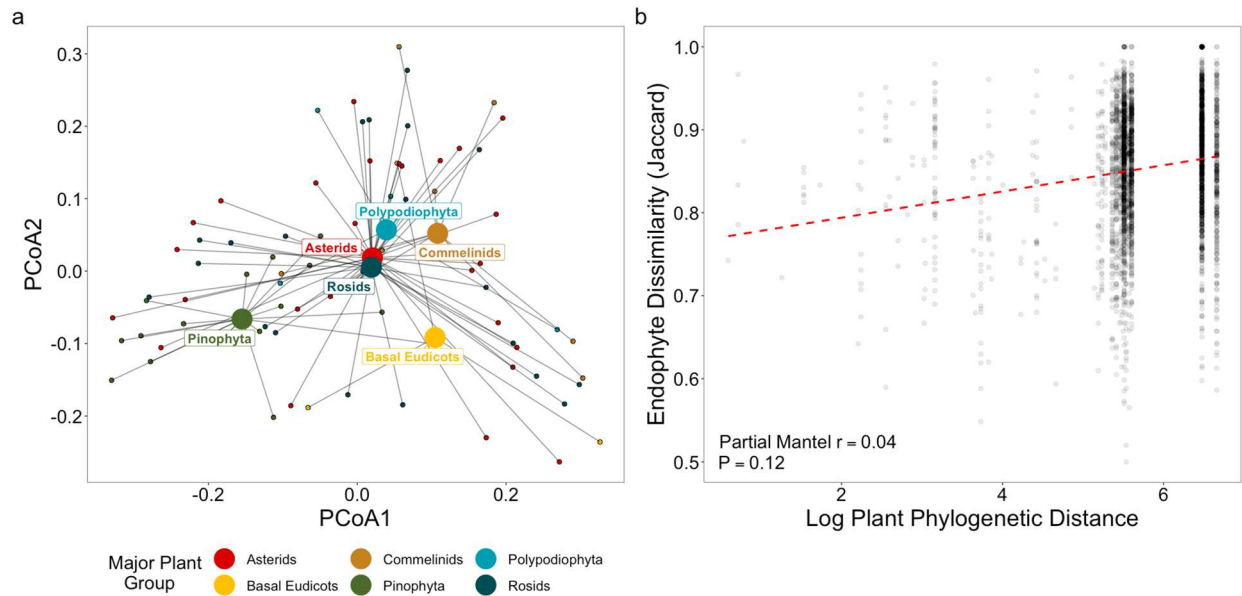
Figure S11. Endophyte communities and their relationships among major plant groups or phylogenetic distances. (a) Principal coordinates analysis of endophyte Jaccard dissimilarities among major plant groups. Smaller circles represent the endophyte community of a given plant individual (n = 80 individuals). Larger circles with corresponding labels refer to the centroid of plant individuals within each major plant group. PERMANOVA among major plant groups with permutations (n = 999) constrained per quadrat was significant (PERMANOVA R^2 = 0.082; P = 0.001). (b) Correlation between endophyte Jaccard dissimilarities and log-transformed plant phylogenetic distance controlling for geographic distance was not significant (Partial Mantel r = 0.04; P = 0.12).
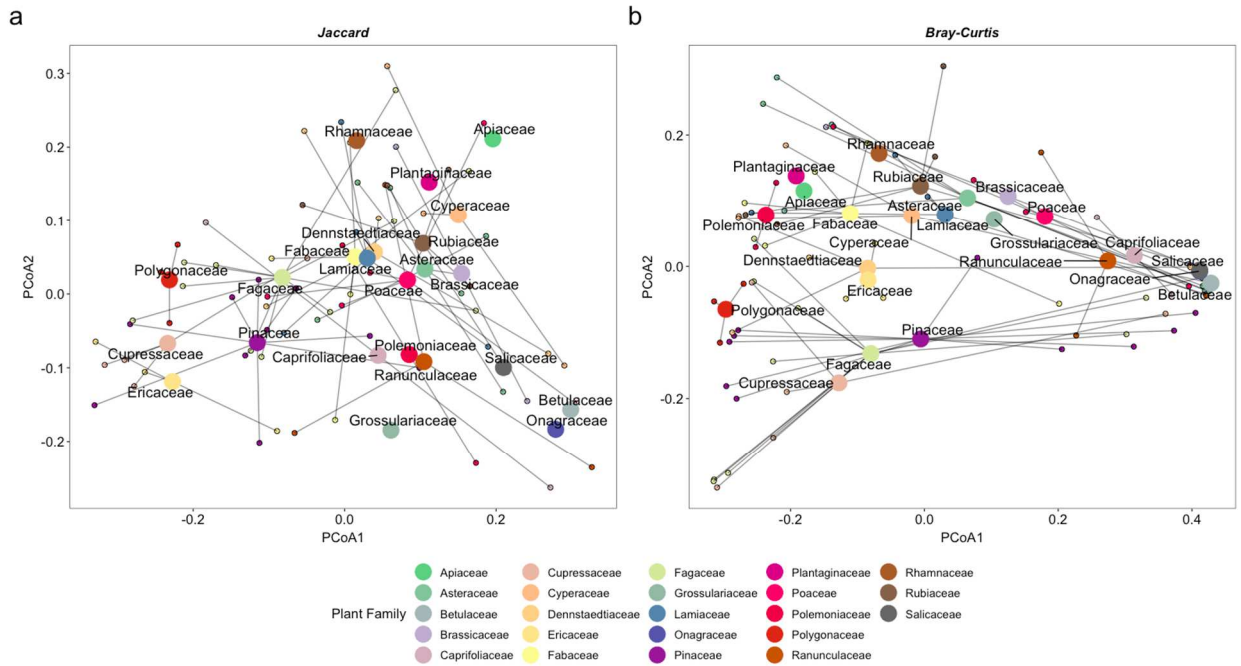
Figure S12. Principal coordinate analysis of endophyte communities grouped by plant family. Plant family was a significant predictor of endophyte composition by PERMANOVA with (a) Jaccard ($R^2 = 0.32$; $P = 0.001$) and (b) Bray-Curtis ($R^2 = 0.33$; $P = 0.001$) indices. PERMANOVA permutations (n = 999) were constrained per quadrat. Smaller circles represent the endophyte community of a given plant individual (n = 80 individuals). Larger circles with corresponding labels refer to the centroid of plant individuals within each plant family.
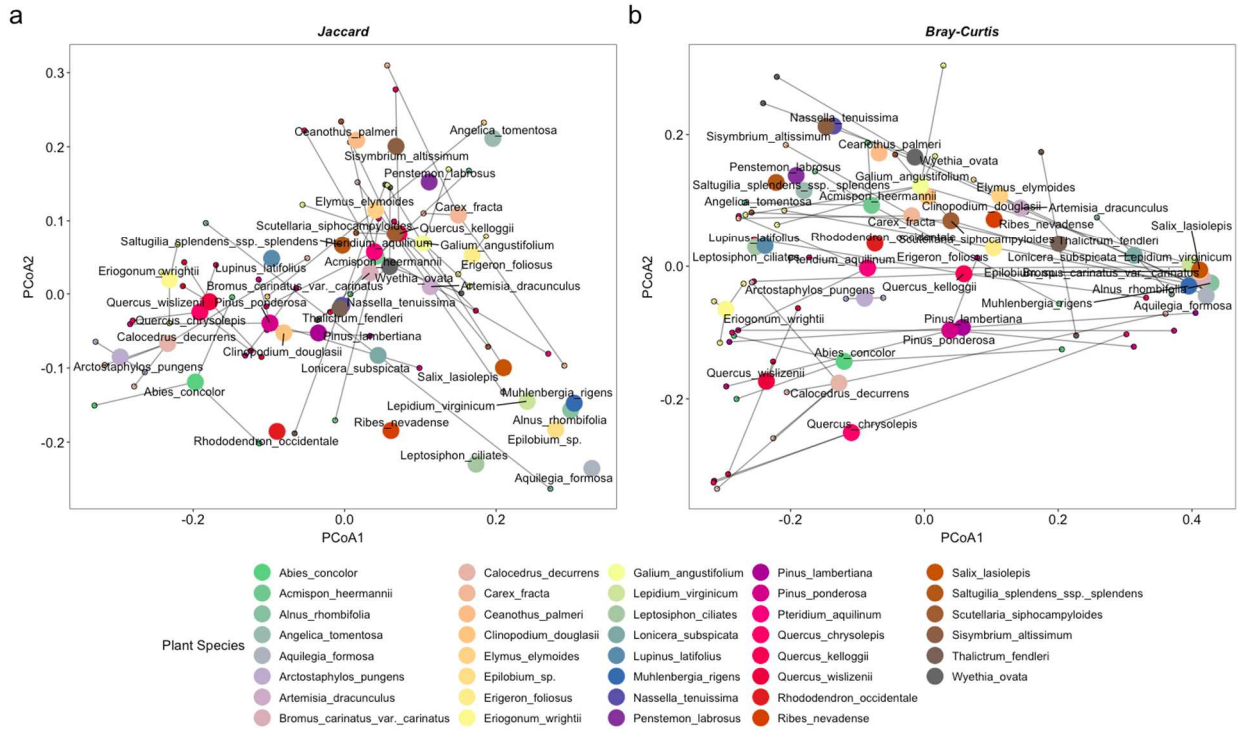
Figure S13. Principal coordinate analysis of endophyte communities grouped by plant species. Plant family was a significant predictor of endophyte composition by PERMANOVA with (a) Jaccard ($R^2 = 0.51$; $P = 0.001$) and (b) Bray-Curtis ($R^2 = 0.51$; $P = 0.001$) indices. PERMANOVA permutations (n = 999) were constrained per quadrat. Smaller circles represent the endophyte community of a given plant individual (n = 80 individuals). Larger circles with corresponding labels refer to the centroid of plant individuals within each plant species.
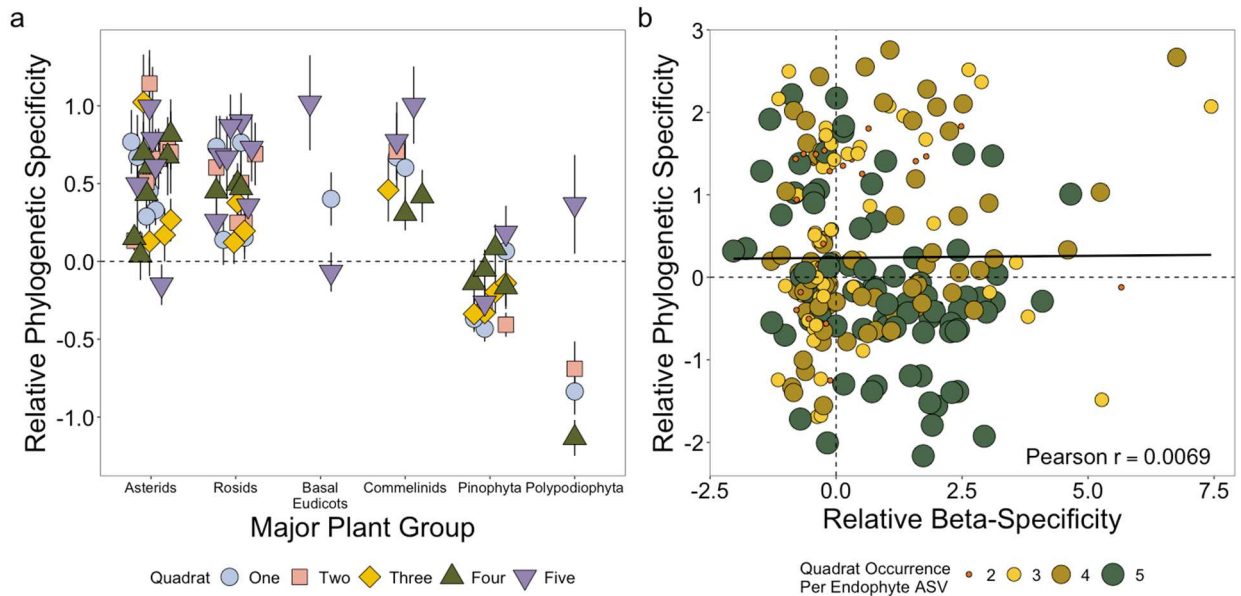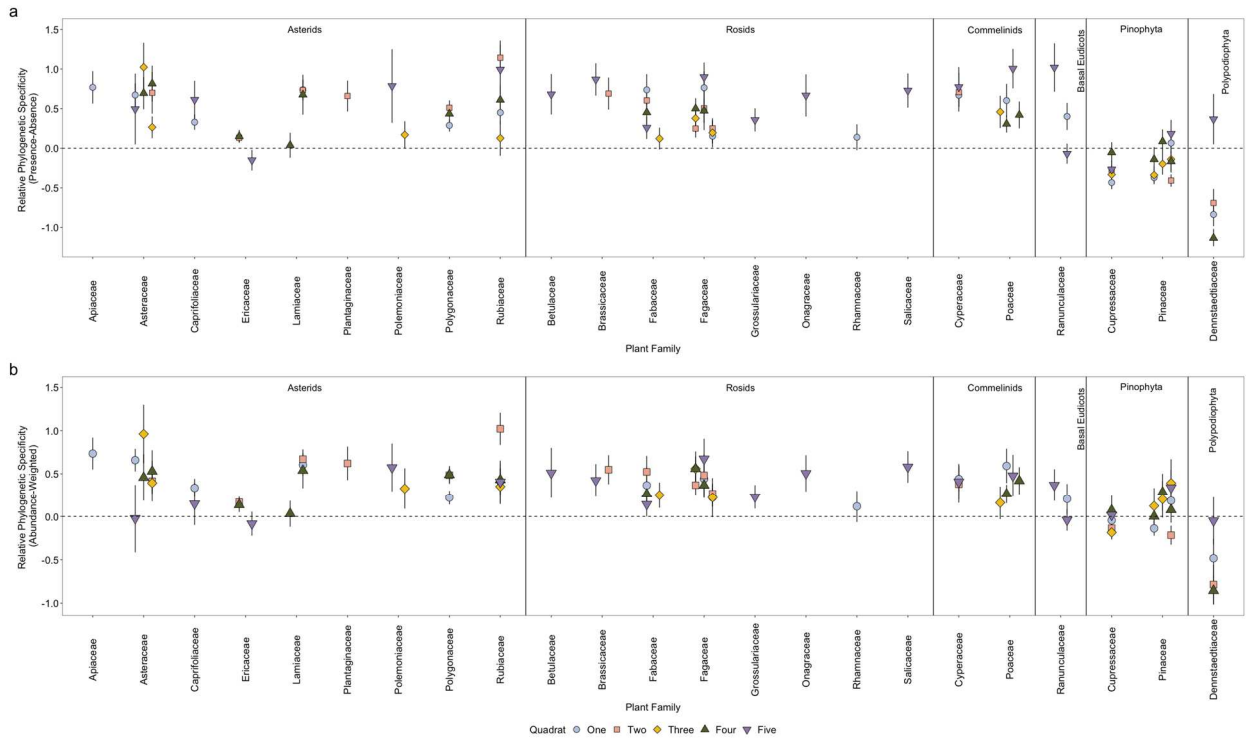
Figure S14. Presence-absence relative phylogenetic specificity among major plant groups and its relationship to beta specificity. Positive relative host specificities indicate higher host specificity than expected by chance (above or to the right of the dashed lines). (a) Endophyte relative phylogenetic specificity measured as -1 * standardized effect size of the mean pairwise phylogenetic distance across major plant groups with quadrat as a covariate (ANOVA: $F_{(5,70)} = 21.56$; $P < 0.001$). Each point represents the average relative phylogenetic specificity value among all endophyte ASVs of one plant individual (n = 80 individuals) with standard error bars. Plant individuals are color- and shape-coded by quadrat and jittered along the x-axis by plant species. Pairwise comparisons were evaluated with Tukey's HSD (Appendix 4: Table S2). (b) Correlation between relative phylogenetic and relative beta specificity per endophyte ASV (Pearson's r: 0.0069; $P = 0.91$). Each point represents an endophyte ASV and is colored-coded by an endophyte ASV's presence across quadrats. Beta specificity was measured as the standardized effect size of the Sørensen Multiple-Assemblage Overlap Measure. Endophytes that appeared in one quadrat never had higher absolute beta specificity relative to the null model because occurring in a single quadrat precluded their ability to display consistency in plant interactions across quadrats and thus were excluded from the analysis.
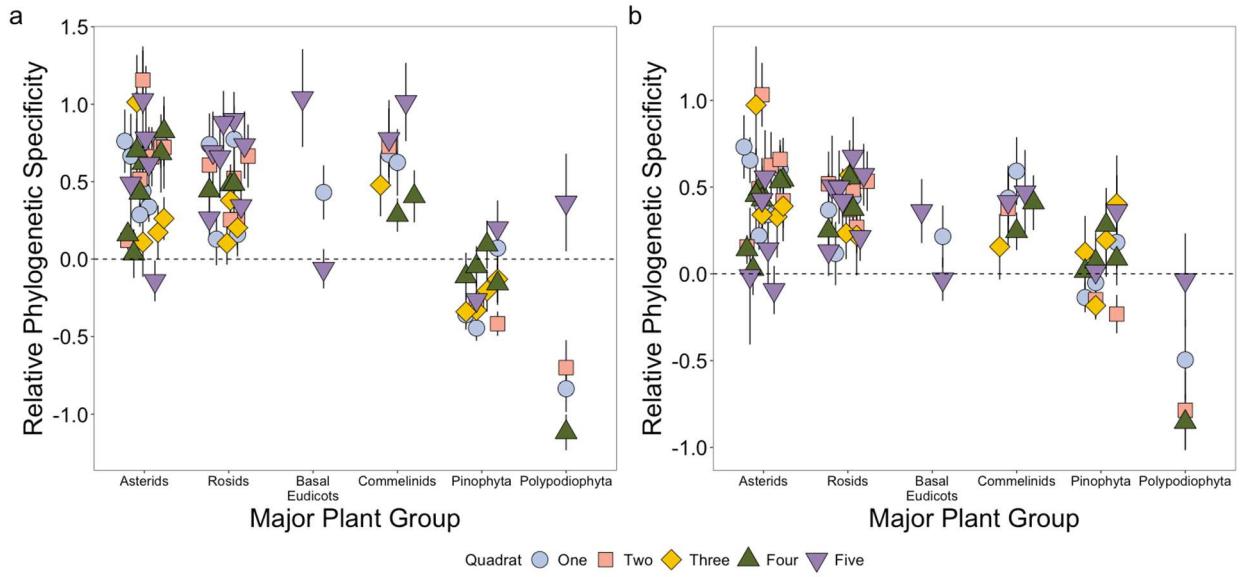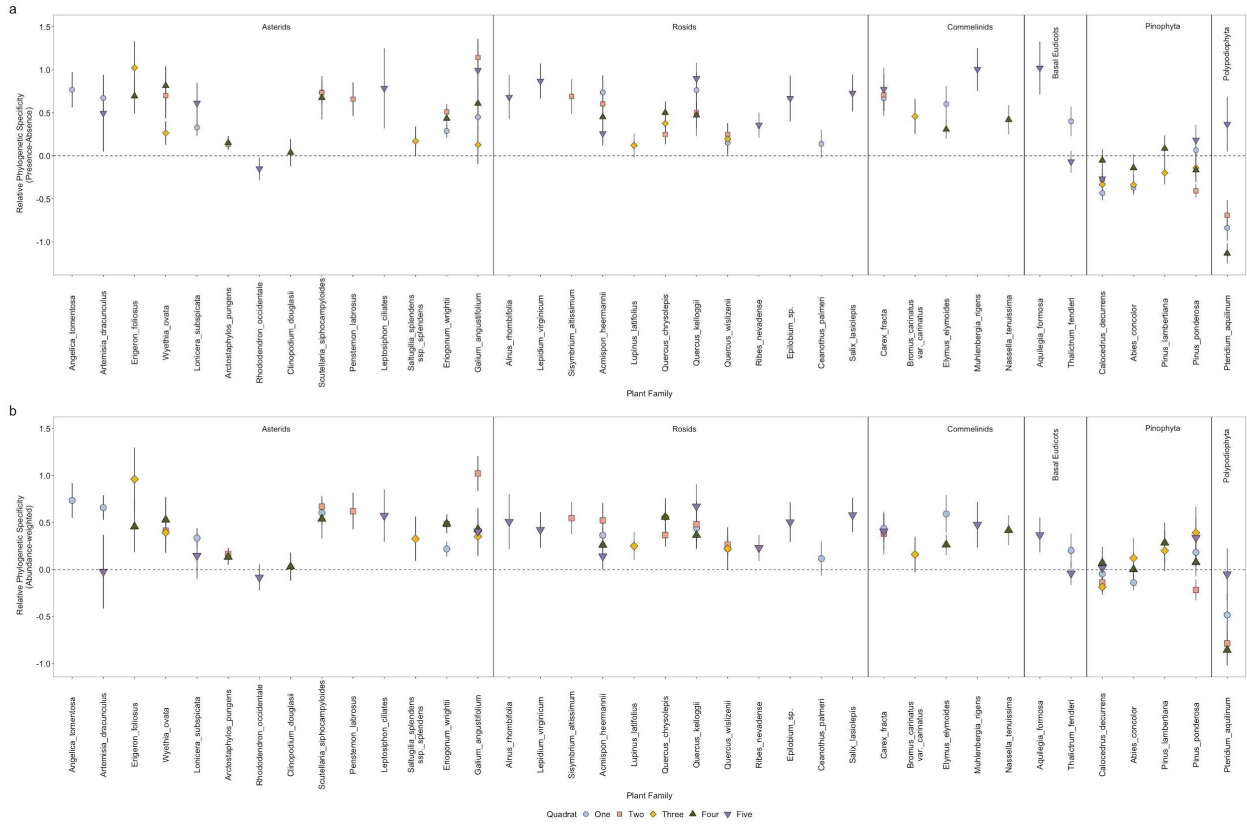
Figure S15. Relative phylogenetic specificity values measured with an alternative null model. Endophyte phylogenetic specificity measured as the standardized effect size of (a) presence-absence (ANOVA $F_{(5,70)}$ = 21.11; P < 0.001) or (b) abundance-weighted (ANOVA $F_{(5,70)}$ = 17.07; P < 0.001) -1 * mean pairwise phylogenetic distance across major plant groups with quadrat as a covariate. Each point represents the average relative phylogenetic specificity value among all endophyte ASVs of one plant individual (n = 80 individuals) with standard error bars. Plant individuals are color- and shape-coded by quadrat and jittered along the x-axis by plant species. Positive relative host specificities indicate higher host specificity than expected by chance (above the dashed lines). The phylogenetic distance matrix was randomized with the "phylogeny.pool" algorithm using the "ses.mpd" function in the *picante* package.
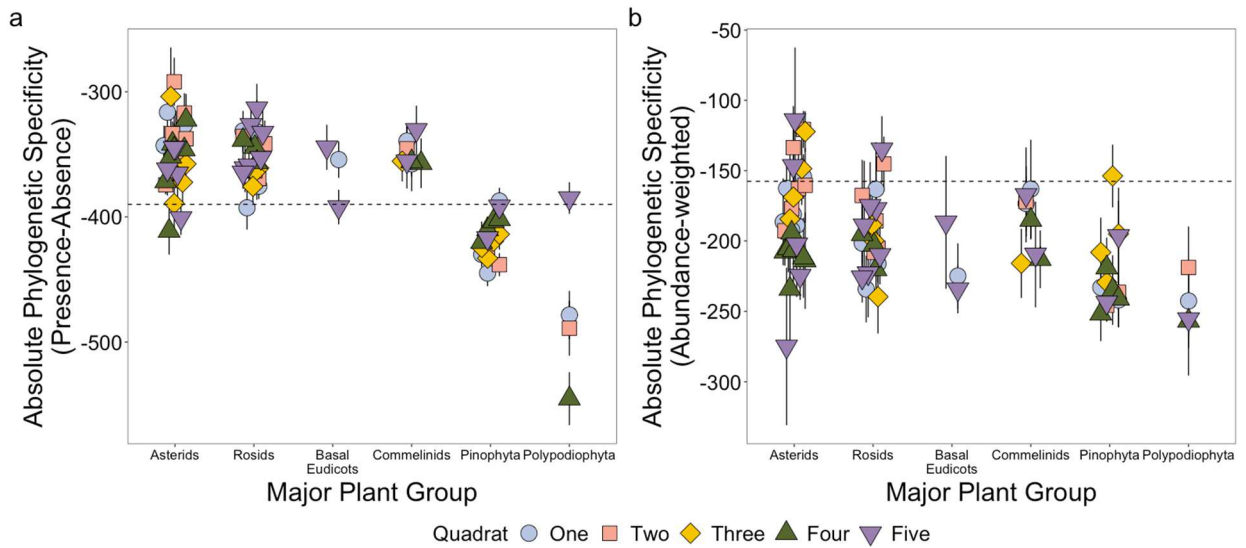
Figure S16. Absolute phylogenetic specificity among major plant groups. Endophyte relative phylogenetic specificity measured as (a) presence-absence (ANOVA $F_{(5,70)}$ = 30.74; P < 0.001) or abundance-weighted (ANOVA $F_{(5,70)}$ = 6.92; P < 0.001) -1 * mean pairwise phylogenetic distance across major plant groups. Each point represents the average absolute phylogenetic specificity value among all endophyte ASVs of one plant individual (n = 80 individuals) with standard error bars. Plant individuals are color- and shape-coded by quadrat location and jittered along the x-axis by plant species. Positive relative host specificities indicate higher host specificity than expected by chance (above the dashed lines).

Figure S17. Relative phylogenetic specificity among plant families. Endophyte relative phylogenetic specificity measured as the standardized effect size of (a) presence-absence (ANOVA $F_{(22,53)}$ = 6.25; P < 0.001) or (b) abundance-weighted (ANOVA $F_{(22,53)}$ = 5.14; P < 0.001) -1 * mean pairwise phylogenetic distance across 23 plant families. Each point represents the average relative phylogenetic specificity value among all endophyte ASVs of one plant individual (n = 80 individuals) with standard error bars. Plant individuals are color- and shape-coded by quadrat and jittered along the x-axis by plant species. Positive relative host specificities indicate higher host specificity than expected by chance (above the dashed lines).
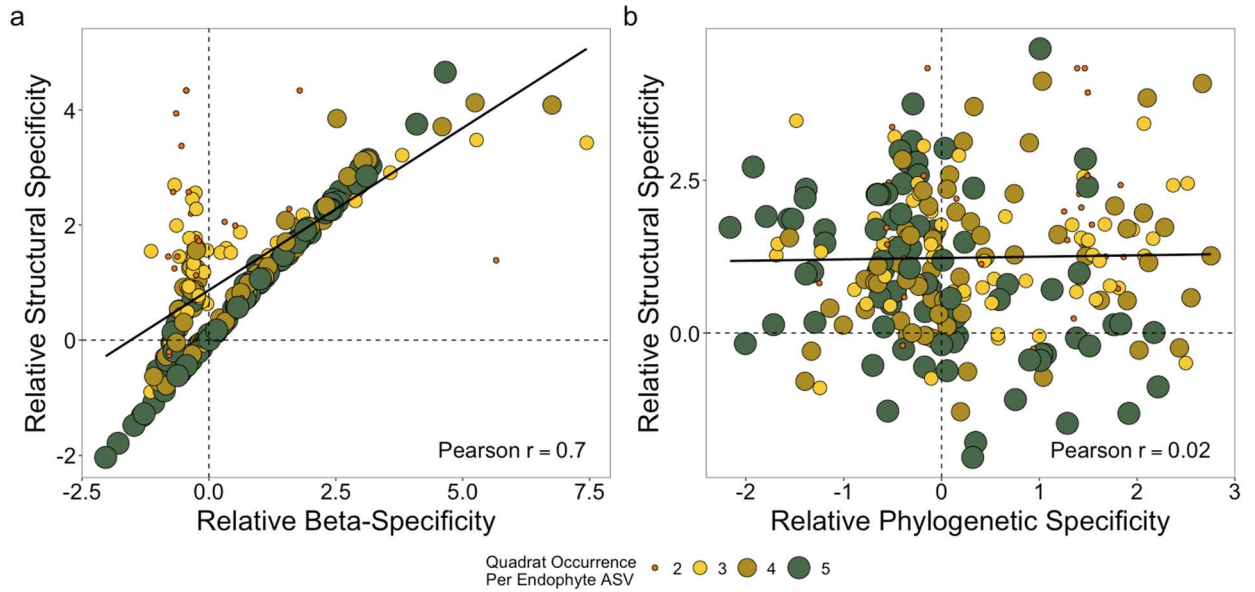
Figure S18. Relative phylogenetic specificity among plant species. Endophyte phylogenetic specificity measured as the standardized effect size of (a) presence-absence (ANOVA $F_{(37,38)}$ = 4.85; P < 0.001) or (b) abundance-weighted (ANOVA $F_{(37,38)}$ = 3.87; P < 0.001) -1 * mean pairwise phylogenetic distance across 38 plant species. Each point represents the average relative phylogenetic specificity value among all endophyte ASVs of one plant individual (n = 80 individuals) with standard error bars. Plant individuals are color- and shape-coded by quadrat and jittered along the x-axis by plant species. Positive relative host specificities indicate higher host specificity than expected by chance (above the dashed lines).

170
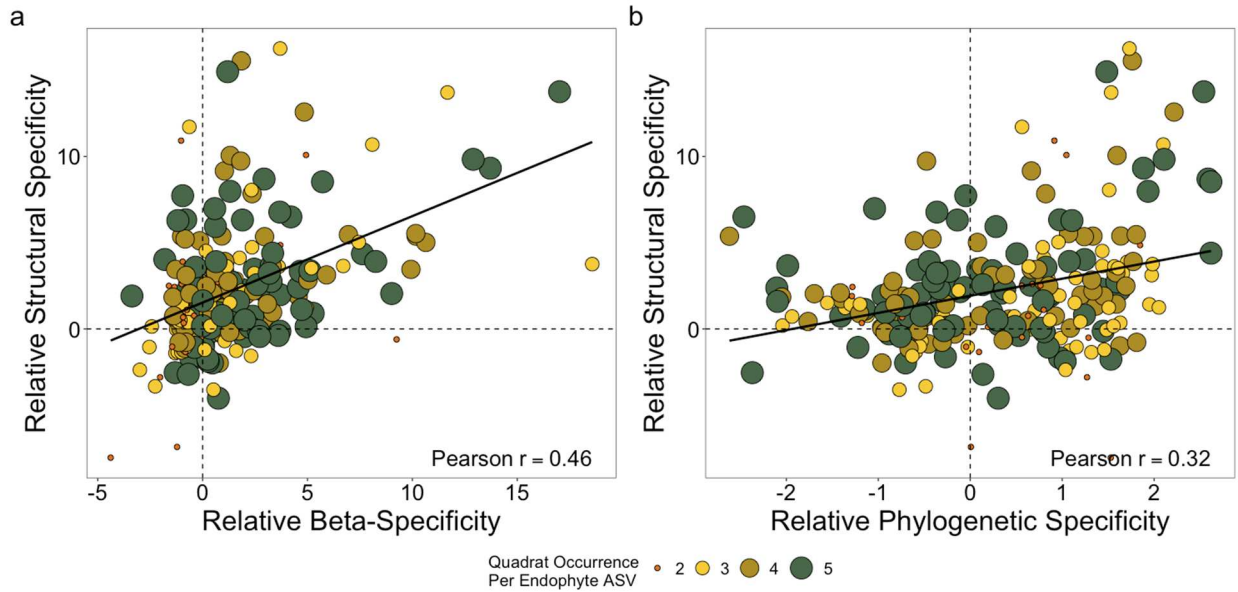
Figure S19. Relationships between abundance-weighted relative structural, phylogenetic and beta specificity. Each point represents an endophyte ASV and is color-coded by its occurrence across the sampling quadrats. (a) Correlation between relative structural and relative beta specificity (Pearson's r: 0.46; $P < 0.001$). (b) Correlation between relative structural and relative phylogenetic specificity (Pearson's r: 0.32; $P < 0.001$). Positive relative host specificities indicate higher host specificity than expected by chance (above or to the right of the dashed lines).

Figure S20. Relationships between presence-absence relative structural, phylogenetic and beta specificity. Each point represents an endophyte ASV and is color-coded by its occurrence across the sampling quadrats. (a) Correlation between relative structural and relative beta specificity (Pearson's r: 0.70; $P < 0.001$). (b) Correlation between relative structural and phylogenetic (Pearson's r: 0.02; $P = 0.76$). Positive relative host specificities indicate higher host specificity than expected by chance (above or to the right of the dashed lines).
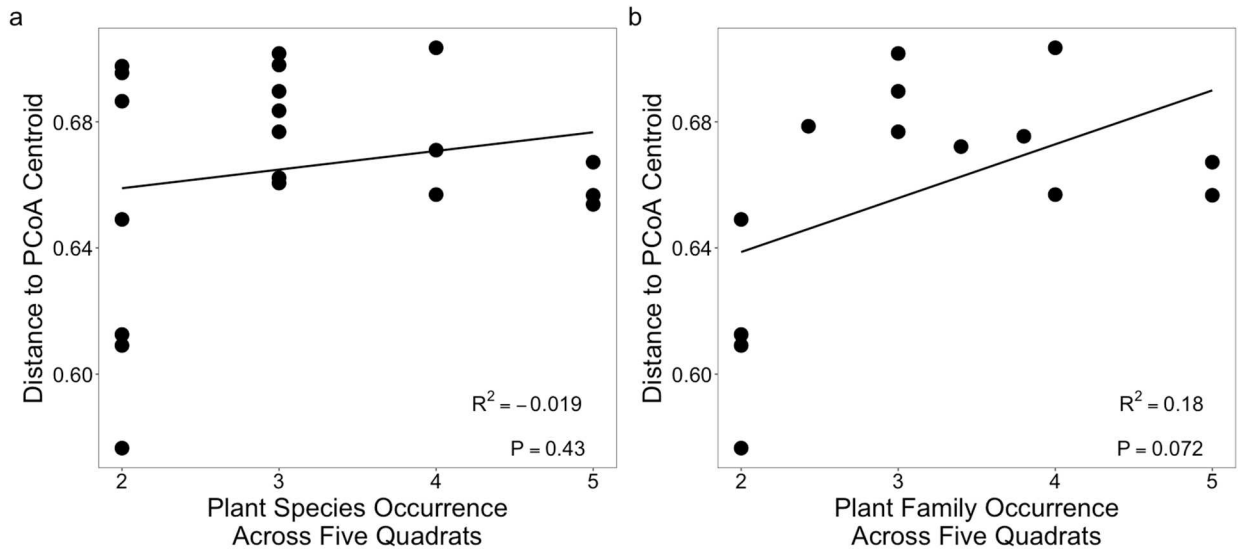
Figure S21. Relationships between the average distance of plant individuals of the same (a) plant species or (b) plant family to their group centroid in PCoA space and plant species occurrence across quadrats. Every point represents a plant *species* rather than a plant individual (all other analyses). Each point is the average distance of plant individuals to their group centroid in PCoA space when grouped by (a) plant species or (b) plant family. Plant species or families represented by only one plant individual were excluded. (a) Plant species more abundant across the landscape were not more similar in their endophyte communities relative to rarer plant species ($R^2 = -0.019$; *P* = 0.43; see Appendix 4: Fig. S12 for PCoA). (b) Plant families more abundant across the landscape were not more similar in their endophyte communities relative to rarer plant families ($R^2 = 0.18$; *P* = 0.072; see Appendix 4: Fig. S13 for PCoA).

Tables S1 - S2

Table S1. Tukey HSD for abundance-weighted relative phylogenetic specificity among major plant groups. Statistically significant results are in bold.

Pairwise Comparison	Difference in Mean Relative Phylogenetic Specificity	Lower 95% CI	Upper 95% CI	Adjusted p-value
Rosids-Asterids	-0.037	-0.224	0.150	0.992
Basal Eudicots-Asterids	-0.252	-0.650	0.146	0.440
Commelinids-Asterids	-0.037	-0.299	0.226	0.998
Pinophyta-Asterids	**-0.364**	**-0.574**	**-0.155**	**<0.001**
Polypodiophyta-Asterids	**-0.972**	**-1.322**	**-0.622**	**<0.001**
Basal Eudicots -Rosids	-0.214	-0.618	0.189	0.628
Commelinids-Rosids	0.001	-0.270	0.271	1.000
Pinophyta-Rosids	**-0.327**	**-0.546**	**-0.108**	**0.001**
Polypodiophyta-Rosids	**-0.935**	**-1.291**	**-0.579**	**<0.001**
Commelinids- Basal Eudicots	0.215	-0.228	0.659	0.715
Pinophyta-Basal Eudicots	-0.113	-0.527	0.302	0.967
Polypodiophyta-Basal Eudicots	**-0.720**	**-1.221**	**-0.220**	**0.001**
Pinophyta-Commelinids	**-0.328**	**-0.614**	**-0.041**	**0.016**
Polypodiophyta-Commelinids	**-0.935**	**-1.336**	**-0.534**	**<0.001**
Polypodiophyta-Pinophyta	**-0.608**	**-0.976**	**-0.239**	**<0.001**

Table S2. Tukey HSD for presence-absence relative phylogenetic specificity among major plant groups. Statistically significant results are in bold.

Pairwise Comparison	Difference in Mean Relative Phylogenetic Specificity	Lower 95% CI	Upper 95% CI	Adjusted p-value
Rosids-Asterids	-0.046	-0.288	0.196	0.993
Basal Eudicots -Asterids	-0.080	-0.597	0.436	0.997
Commelinids-Asterids	0.087	-0.254	0.428	0.975
Pinophyta-Asterids	**-0.716**	**-0.988**	**-0.444**	**<0.001**
Polypodiophyta -Asterids	**-1.104**	**-1.558**	**-0.650**	**<0.001**
Basal Eudicots-Rosids	-0.034	-0.557	0.489	1.000
Commelinids-Rosids	0.133	-0.218	0.484	0.875
Pinophyta-Rosids	**-0.670**	**-0.955**	**-0.386**	**<0.001**
Polypodiophyta -Rosids	**-1.058**	**-1.520**	**-0.596**	**<0.001**
Commelinids-Basal Eudicots	0.167	-0.408	0.743	0.957
Pinophyta-Basal Eudicots	**-0.636**	**-1.174**	**-0.099**	**0.011**
Polypodiophyta -Basal Eudicots	**-1.024**	**-1.673**	**-0.375**	**<0.001**
Pinophyta-Commelinids	**-0.803**	**-1.175**	**-0.431**	**<0.001**
Polypodiophyta-Commelinids	**-1.191**	**-1.711**	**-0.670**	**<0.001**
Pteridophyte-Pinophyta	-0.388	-0.866	0.091	0.180

Supplementary Results

Comparisons to Null Models of Host Specificity

Host specificity distributions from randomized communities were significantly different from the mean absolute structural, phylogenetic and beta specificity of the observed endophyte community (Appendix 4: Fig. S6; t-test: $P < 0.05$).

Univariate Phylogenetic Specificity

As the number of plant species an endophyte occupies approached the total number of plant species in the community, the range of possible phylogenetic specificity values converged to a single value, or the Mean Pairwise Distance (MPD) of the entire plant community, consistent with expectations for the MPD metric (Appendix 4: Fig. S7; Swenson 2014).

Differences in relative phylogenetic specificity among major plant groups were significant by ANOVA ($F_{(5,70)} = 17.58$; $P < 0.001$; Figure 17a). Pinophyta were significantly lower in mean relative phylogenetic specificity compared to asterids, rosids or commelinids (Appendix 4: Table S1). Polypodiophyta were significantly lower in mean relative phylogenetic specificity compared to all other groups (Appendix 4: Table S1). Relative phylogenetic specificity at the plant family and species-level reflected findings by major plant group (Appendix 4: Fig. S17-S18).

Relationships between Structural, Phylogenetic and Beta specificity

From a methodological perspective, it is not surprising that endophytes that were more consistent in their occupation of the same plant species (i.e., highly beta-specific) also occupied a narrower phylogenetic breadth of plants (i.e., highly phylogenetically specific).

For example, the endophyte with the highest relative beta specificity in this study (endophyte ASV 48) was found almost exclusively in one plant species (99.92% of 13 067 reads) and thus also had a high relative phylogenetic-specificity value. However, the positive trend between relative phylogenetic and relative beta specificity was not necessarily because endophytes were always restricted within single plant species. Fifty-seven endophyte ASVs had higher relative phylogenetic- and beta specificity than expected by chance and occurred, on average, within 6.89 plant species.

Focal Endophyte Analysis of Leotiomycetes

We explored associations between endophytes and plants and compared our findings to host-specific relationships previously documented in the literature. Specifically, we identified the host associations for Leotiomycetes endophytes for which there are numerous instances of their relationships with coniferous plants. For example, Arnold (2007) and U'ren et al. (2010) documented high Leotiomycete abundance in Pinaceae hosts (*Pinus* and *Pseudotsuga* spp.) but few Leotiomycetes in other co-occurring plant families Fagaceae (*Quercus* spp.) and Cupressaceae (*Cupressus* and *Platycladus* spp.) with similar findings by Sieber (2007) and Moler and Aho (2018). *Pinus lambertiana* (two samples) and *Pinus ponderosa* (five samples) harbored 29 Leotiomycete endophytes that were the second most abundant class in pines (35.66% reads of 105 521 reads) to the Dothideomycetes (64 ASVs; 52.91% of 105 521 reads) but more abundant than the Eurotiomycetes (11 ASVs; 3.82% of 105 521 reads). Leotiomycetes had lower mean relative phylogenetic specificity (-0.011) compared to Dothideomycetes (-0.0075) and Eurotiomycetes (0.041) after correcting for ASV number per class.

At a finer phylogenetic resolution, endophytes within the genus *Lophodermium* (Rhytismataceae, Leotiomycetes) are commonly found in Pinaceae hosts (Minter 1981, Stenström and Ihrmark 1997, Tanney and Seifert 2017, Salas-Lizana and Oono 2018) and may be host-specific within this host group as a latent saprotroph (Sokolski et al. 2004, Osono and Hirose 2011). Four of the five *Lophodermium* endophytes we identified had high relative phylogenetic specificity to pines (*L. pinastri, L. nitens, L. fissuratum* and unidentified congener reads occurred within pines) supporting findings from previous studies (Ortiz-García et al. 2003, Salas-Lizana and Oono 2018). The majority of *Lophodermium* endophyte reads (81.53% of 25 723 reads) belonged to *Pinus lambertiana* or *Pinus ponderosa* samples, but their reads were rarely consistent between the same plant species (four endophytes) or were found in only one plant individual (one endophyte). *Lophodermium* endophytes had higher relative phylogenetic- and higher relative beta specificity compared to null models (genus average: relative phylogenetic specificity = 1.64; relative beta specificity = 1.62).

Endophytes within Pteridium aquilinum

Endophytes within the genus *Stagonospora* have been observed within *P. aquilinum* (Fisher and Punithalingam 1993, Fisher 1996). We were unable to find previously documented endophytes found in *P. aquilinum* in the genus *Stagonospora* (Fisher and Punithalingam 1993, Fisher 1996). Three *Neostagonospora* ASVs and four other endophyte ASVs within the *Phaeosphaeriaceae* family occurred within our dataset, but none of these endophytes appeared in our *P. aquilinum* samples perhaps due to very different sampling regions (N. America *vs*. Europe).

The most abundant endophyte in *Pteridium aquilinum* was assigned to the genus

Alternaria (Dothideomycetes; ASV 1) and was the most abundant ASV in the community

(33 2754 reads) that occurred in the highest number of plants species (37 plant species) with

lower phylogenetic (-0.87) and beta specificity (-3.37) than expected by chance (i.e., more

generalist). No single endophyte ASV appeared in all four samples of *P. aquilinum* and only

13.97% of 60 295 reads (28.68% of 136 ASVs) were unique to *P. aquilinum* samples. These

results seem to be consistent with a previous study by Petrini et al. (1992) that found

completely different assemblages of fungal genera within *P. aquilinum* that depended on

season along with similar findings in another temperate fern (Younginger and Ballhorn

2017). These findings support the low relative phylogenetic specificity we observed in

endophytes of *P. aquilinum* even if they were more likely to have greater pairwise

phylogenetic distances relative to endophytes within more common angiosperm lineages.

Other Focal Endophytes

Interestingly, many of the most beta-specific endophytes (endophyte ASVs 48, 132, 269,

256, 1855, 51, 647, 537, 116, 297, 90, 229, 111, 957, 123, 578, 984, 523) were associated

with *Calocedrus decurrens*, a common tree species in the western United States and at this

site. Almost half of the reads from these endophytes (49.05% of 47 590 reads) were only

associated with *C. decurrens*, but many of these endophytes have yet to be documented as

host-specific to *C. decurrens* or even as endophytes. For example, the most beta-specific

endophyte in this study was endophyte ASV 48 and taxonomically assigned within the

genus *Microcyclospora* sp. (Dothideomycetes, Capnodiales, Teratosphaeriaceae; Gleason et

al. 2019). This *Microcyclospora* sp. was the most abundant within this group of endophytes

with 99.92% of 13 067 reads belonging to all five *Calocedrus decurrens* samples. This fungus has only been associated with blemishes on the surfaces of apples and plums (Yang et al. 2010, Frank et al. 2010, Mirzwa-Mróz et al. 2011) suggesting that *Microcyclospora* sp. may have a varied life history strategy as an endophyte and pathogen.