

UC Santa Barbara

UC Santa Barbara Electronic Theses and Dissertations

Title

Species abundance and host phylogenetic breadth influence detection of phyllosymbiosis in foliar fungal endophyte communities across North America

Permalink

<https://escholarship.org/uc/item/57b7r8zn>

Author

Dighe, Abhishek

Publication Date

2024

Supplemental Material

<https://escholarship.org/uc/item/57b7r8zn#supplemental>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Santa Barbara

Species abundance and host phylogenetic breadth influence detection of phylosymbiosis in
foliar fungal endophyte communities across North America

A Thesis submitted in partial satisfaction of the
requirements for the degree Master of Science
in Ecology, Evolution, and Marine Biology

by

Abhishek Dighe

Committee in charge:

Professor Ryoko Oono, Chair

Professor Cherie Briggs

Professor Rodolfo Salas-Lizana

December 2024

The thesis of Abhishek Dighe is approved.

Cherie Briggs

Rodolfo Salas-Lizana

Ryoko Oono, Committee Chair

September 2024

ABSTRACT

Species abundance and host phylogenetic breadth influence detection of phylosymbiosis in
foliar fungal endophyte communities across North America

by

Abhishek Dighe

Foliar fungal endophyte (FFE) assemblages are shaped by evolutionary and ecological processes. However, it remains unresolved how the evolutionary history and phylogenetic relatedness of host plants and endophytes influence the assembly process. Testing for phylosymbiosis by measuring correlation between microbial community dissimilarity and host phylogeny in plant-FFE systems can reveal the extent to which these phylogenetic factors influence FFE assemblage. In this paper, I searched for patterns of phylosymbiosis by testing the relationship between phylogenetic distance separating plant hosts and the phylogenetic dissimilarity of their FFE communities from a phylogenetically diverse set of plants collected from 20 sites across seven latitudinal zones. Phylogenetic dissimilarity between FFE communities was defined by three phylogenetic beta diversity (PBD) metrics: UniFrac, phylogenetic community dissimilarity (PCD), and beta mean nearest taxon distance (bMNTD). Phylosymbiosis was not a prevalent pattern across sites, however, plant and FFE communities of the seven sites that showed the strongest indicators of phylosymbiosis revealed that the detection of phylosymbiosis was sensitive to the

presence of rare FFE species and the phylogenetic breadth of hosts being tested. Additionally, in some cases, compositional differences between communities were driving phylogenetic community dissimilarity rather than the phylogenetic relationships between FFE species. Phyllosymbiosis was not detected when testing the relationship between phylogenetic distance separating endophytes and the phylogenetic dissimilarity of their host communities using PCD and bMNTD as PBD metrics.

A. Introduction

Symbioses are abundant across the ecological landscape, pervade every level of biological organization, and have a rich history dating back hundreds of millions of years to the microscopic predecessors of the cells that make up organisms today. The most ubiquitous and ancient symbioses occur at the microscopic scale between microbes and multicellular organisms (Raina et al. 2018), such as between plants and fungi. Fungi are common in the microbiomes associated with the plant holobiont (Vandenkoornhuyse et al. 2015) and have been found living in and around plants of all phyla studied. Foliar fungal endophytes (FfEs) are a hyper-diverse group of fungi that asymptotically reside within the photosynthetic tissue of plants (Wilson 1995). FfEs are ubiquitous (Arnold 2007; Krings et al. 2007) and involved in a wide range of interactions with their plant hosts, including defense against pathogens (Arnold et al. 2003; Grabka et al. 2022; Busby et al. 2016) and herbivores (Omacini et al. 2001), enhanced nitrogen uptake (Christian et al. 2019), pathogenesis (Bacon et al. 1977), and saprotrophism (David et al. 2023; Weatherhead et al. 2022). This incredible functional range of FfEs is particularly interesting because of its potential applications in agricultural and ecological management. Just as we manipulate microbiomes of animals for medical purposes (Omer et al. 2022), there is a growing interest in utilizing FfEs and their diverse functions to our advantage, such as inoculating crops with fungal species that promote host defense against pathogens and pests (Grabka et al. 2022). Transplanting and engineering fungi is deeply complex as there are a multitude of factors that determine host associations. Understanding the factors that determine FfE assemblage and distribution in plants is the first step to advance us towards applying FfEs for the benefit of society.

Numerous evolutionary and ecological factors have been observed to influence FFE assemblage. Abiotic factors, such as temperature, water availability, and UV exposure, are the first habitat filters on FFE communities (Saunders et al. 2010). Plant hosts impose a second habitat filter on endophyte communities based on their phenotype (Kolattukudy 1985; Saunders and Kohn 2009). Interspecific competition among endophytes inside of a leaf contributes to more filtering (Saunders et al. 2010). Since these biological interactions are built upon plant host and fungal symbiont phenotypes that may be evolutionarily conserved, studying phylogenetic history as an evolutionary factor of assemblage has garnered interest (Saunders et al. 2010). However, the relative influence of host evolutionary history and endophyte evolutionary history, and their position in the hierarchy of FFE assemblage factors remains unresolved.

Previous studies exploring phylogenetic signals in plant-fungal systems investigated the effects of host phylogeny on microbiome composition and searched for *phylosymbiosis*, or the quantitative correlation between microbial community dissimilarity and host phylogeny (Lim and Bordenstein 2020). If a host-microbe system is experiencing phylosymbiosis, then the dissimilarities between microbial communities resemble the host phylogenetic relationships, with closely related host species sharing more similar microbial communities. Selection is hypothesized to drive phylosymbiosis as host filtering excludes certain microbe species (Lin et al. 2024). Studies on plant-microbe interactions have reported positive results supporting phylosymbiosis (Wehner et al. 2013; Yeoh et al. 2017). For example, in an analysis of the root microbiomes of 31 plant species representing six phyla, Yeoh et al. (2017) found strong effects of host phylogeny on determining bacterial root microbiome composition. Wehner et al. (2013) also found that host phylogeny was the

most significant factor in explaining community composition, with closely related plants harboring more similar communities than expected among the root-associated fungal communities of 25 species of *Asteraceae*.

Phylogenetic signals within plant-FFE systems have been variable. Among 46 species of *Ficus* collected from the Xishuangbanna Tropical Botanical Garden, a significant correlation was found between host phylogenetic distance and the Bray-Curtis dissimilarity of their FFE assemblages (Liu et al. 2019) in support of phylosymbiosis. Likewise, increasing Bray-Curtis dissimilarity of FFE assemblages with increasing host phylogenetic distance was observed in six species of pine hosts sampled from northeastern United States (Sarver et al. 2021). Some other studies, like Whitaker et al. (2020)'s study among 18 species of *Asteraceae* planted in a common garden experiment, found significant effect of host species identity on assemblage composition, but no evidence for phylosymbiosis. Similarly, a study of FFE dispersal in New Guinea rainforest trees found no significant correlation between FFE assemblage dissimilarity and host phylogeny across five plant families (*Moraceae*, *Euphorbiaceae*, *Rubiaceae*, *Myrtaceae*, and *Gnetaceae*) (Vincent et al. 2016). Detecting phylosymbiosis may be sensitive to the phylogenetic breadth of the study, as Liu et al. (2019) and Sarver et al. (2021) tested within one plant genus and found evidence of phylosymbiosis, while Whitaker et al. (2020) and Vincent et al. (2016) tested within and across families, respectively, to find no phylosymbiosis. Testing phylosymbiosis across different phylogenetic breadths needs to be further explored to determine if the range of detection is within-genus or extends further across higher taxonomic levels.

All of these past studies use beta diversity to compare FFE assemblages across host phylogenies. Beta diversity reveals compositional differences between assemblages, but

does not address the phylogenetic relatedness of individuals within those assemblages. For example, if two endophytes that are sister species live within two different plants, beta diversity metrics would not indicate that these plants have similar assemblages since they do not share the same species of endophyte. This form of measurement ignores the phylogenetic relationship of the endophytes. Phylogenetic beta diversity (PBD) solves this by comparing phylogenetic relationships between taxa of communities, and can reveal the associations of fungal lineages, rather than individual species, to certain plant hosts. Nevertheless, PBD is not commonly used in studying fungal assemblages due to the challenges of building phylogenies solely using the ITS amplicon region, which is commonly used in metabarcoding studies (Schoch et al. 2012).

In this paper, I used a novel phylogenetic placement program, the Tree-Based Alignment Selector (T-BAS) toolkit (Carbone et al. 2019), to generate a phylogenetic tree of FFEs by placing ITS1 amplicon sequences onto a curated reference tree. This allowed me to implement the PBD metrics UniFrac (Lozupone and Knight 2005), phylogenetic community dissimilarity (PCD) (Ives and Helmus 2010), and beta mean nearest taxon distance (bMNTD) (Webb et al. 2008). Each metric provides different information about the phylogenetic relatedness between communities and has its own benefits and limitations in the context of this sampling design (discussed in Methods; Table 1). UniFrac broadly assesses PBD by calculating the ratio of shared and unshared evolutionary branches occupied by two communities on a phylogenetic tree. UniFrac can also consider relative abundance of the species by weighing occupied branches depending on their read abundance. A limitation of branch-based PBD metrics, such as UniFrac, is that they are influenced by differences in species richness and phylogenetic diversity (Ives and Helmus

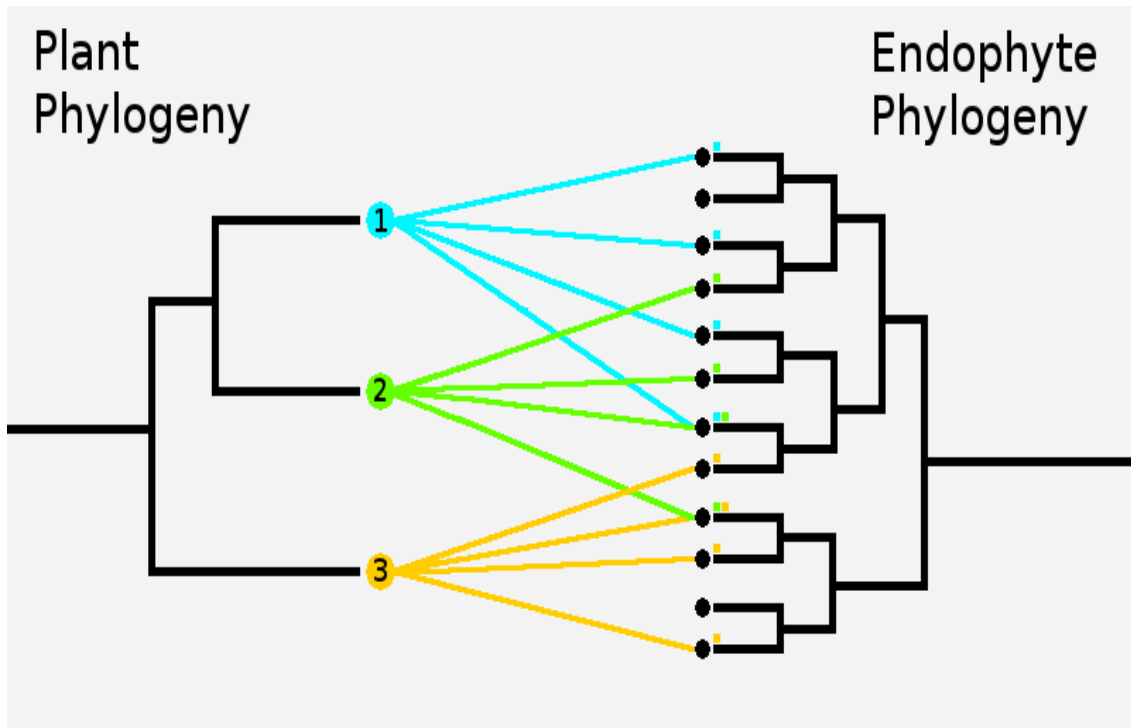
2010). If two communities differ substantially in species richness or phylogenetic diversity, even if the smaller community is completely nested within the other, the number of unshared branches can inflate the dissimilarity value. PCD addresses the limitations of branch-based dissimilarity by computing a phylogenetic component PCD_p, which measures the relatedness of unshared species, and a non-phylogenetic component PCD_c, which measures compositional similarities between communities. bMNTD gives a quantitative measure of the average phylogenetic distance separating the species of one community from the species of another. I incorporated the information from multiple PBD metrics to paint a more holistic picture of the phylogenetic structure of the plants and FFE assemblages targeted in this study.

I explored the phylogenetic relationship between plant hosts and the assemblage of their FFE communities using a data set that contains FFEs from a phylogenetically diverse set of plants collected across seven latitudinal zones (Apigo thesis). Testing across a set of geographically diverse sites permits investigation of phylogenetic signals in the presence of other non-phylogenetic assemblage factors and across a more diverse set of plants and FFE (endophyte) communities. I measured phylogenetic dissimilarity using phylogenetic beta diversity metrics to quantify the relatedness of plant hosts and the relatedness of their endophyte communities, in order to address the question:

Do closely related plants share more closely related endophyte communities as compared to distantly related plants?

I approached this question from both the plant and endophyte perspectives and used the framework of phylogenetically conserved traits underlying host-specificity in endophytes (Gilbert and Webb 2007) to form my hypotheses. From the perspective of the plant host, I hypothesized that closely related plants will have more phylogenetically similar endophyte communities as compared to distantly related plants (Figure 1-A), such that the phylogenetic dissimilarity of endophyte communities will increase as plant hosts become more distantly related (Figure 1-B). From the perspective of the endophyte, I hypothesized that closely related endophytes will have more phylogenetically related host communities as compared to distantly related endophytes (Figure 1-B), with closely related endophytes appearing in the same or closely related hosts.

(A)



(B)

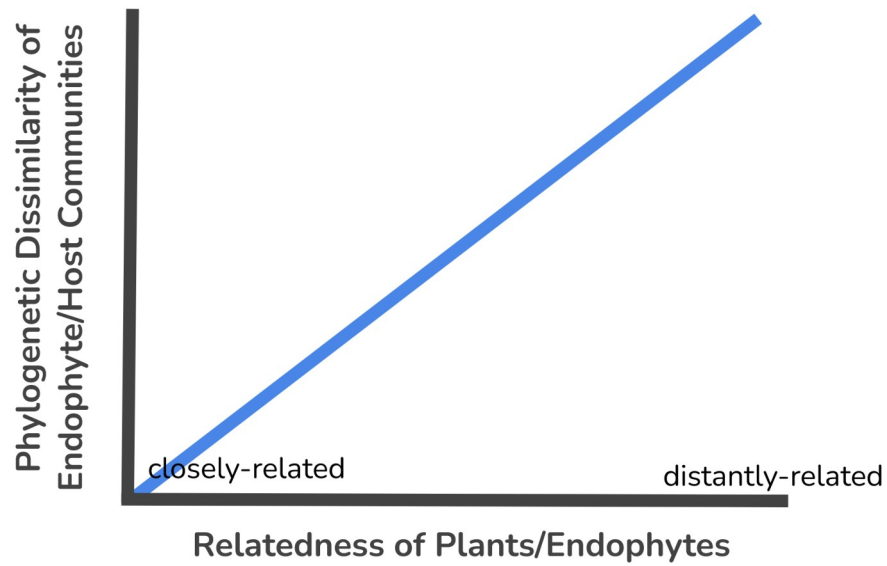


Figure 1. Diagram of study hypotheses. (A) Lines connecting endophyte taxa to plant taxa represent presence of that endophyte in the plant's endophyte community. Closely related plants, such as plants 1 and 2, do not necessarily share the same endophyte species but do share phylogenetically related species. As plants become more distantly related, such as plants 1 and 3, the endophytes that compose their communities also become more distantly related. (B) As the phylogenetic distance separating plants/endophytes increases, their endophyte/host communities become more phylogenetically dissimilar.

B. Methods

Sampling and Bioinformatics

Plant tissues were collected from 20 sites across seven latitudinal zones (Panama, Alaska, Oregon, British Columbia, South Mexico, North Mexico, and California) for an ongoing latitudinal gradient study (Fig. 1; see Apigo et al. *in prep.* for detailed sampling method, Apigo thesis). Five 50 m² quadrats were sampled from each site. Ten leaves were collected from one plant of each species found within each quadrat. The collected tissue samples were surface-sterilized and preserved at -80°C until DNA extraction. DNA was extracted from 80-100 mg of each plant sample that was cooled with liquid nitrogen and homogenized with mortar and pestle. A modified 2% CTAB protocol (Doyle and Doyle 1987) was used to extract DNA. The internal transcribed spacer 1 (ITS1) region was amplified using the fungal-specific ITS1F-KYO1 and ITS2-KYO1 primers (Toju et al. 2012), barcoded, and pooled for sequencing on the Illumina MiSeq platform. Endophyte sequences were clustered into 97% operational taxonomic units (OTUs), which approximately represents fungal species (Blaalid et al. 2013). The *rbcL* or ITS2 plant genetic markers from the tissue samples were sequenced to identify host plants to the genus level. In order to reduce the overrepresentation of species that were present in multiple quadrats, samples across all five quadrats were merged by genus using “merge” from *phyloseq* (McMurdie and Holmes 2013). Sampling effort was standardized by rarefying the abundance data using “rarefy_even_depth” from *phyloseq* (Appendix - Table 1.1).

Phylogenetic Trees

Fungal Phylogeny

Tree-Based Alignment Selector (T-BAS) toolkit (Carbone et al. 2019) was used to generate a phylogenetic tree for endophytes from ITS1 sequences using the EPA-NG algorithm (Barbera et al. 2019). The T-BAS parameters used are listed in Appendix - Document 1. All analyses involving phylogenetic relationships of endophytes include only those placed by T-BAS.

Plant Phylogeny

Plant taxonomic ranks were standardized and then utilized to prune a phylogenetic tree from the extant vascular plant phylogenetic tree using the “phylo.maker” function from *V.PhyloMaker2* (Jin and Qian 2022). Bryophytes were excluded from the data set since only the vascular plant phylogeny was available (Apigo et al. *in prep*, Apigo thesis).

Statistical Analyses

All statistical analyses were conducted in R 4.2.1 (R Core Team 2022) unless stated otherwise. The 20 sites were analyzed independently due to differences in non-phylogenetic factors that may be present between sites, such as climate.

Phylogenetic Beta Diversity - Plant Perspective

To investigate if closely related plants are more likely to have phylogenetically similar communities than distantly related plants, the phylogenetic relatedness of plants was tested against the phylogenetic relatedness of their endophyte communities. The relatedness of plant samples was quantified as the phylogenetic distance, or the sum of the lengths of

evolutionary branches separating samples on the vascular plant phylogenetic tree. The function “cophenetic.phylo” from *ape* (Paradis and Schliep 2019) was used to generate a distance matrix containing pairwise phylogenetic distances between hosts at each site.

The relatedness of endophyte communities was quantified as phylogenetic beta dissimilarity as defined by pairwise PBD metrics (Table 1). Pairwise PBD comparisons incorporate the relatedness of species in pairs of communities, and sometimes compositional abundance, to assess how phylogenetically similar they are to each other. Three PBD metrics were employed: UniFrac, phylogenetic community dissimilarity (PCD), and beta mean nearest taxon distance (bMNTD). Compositional beta diversity was measured as Bray-Curtis dissimilarity to compare results with those of other studies. Bray-Curtis dissimilarity was implemented using the function “vegdist” from *vegan* (Oksanen et al. 2022).

UniFrac is a measure of how many evolutionary branches of a phylogenetic tree are shared and unshared between two communities (Lozupone and Knight 2005) and ranges from 0 (i.e., identical) to 1 (sharing minimum evolutionary branches). The function “UniFrac” from *phyloseq* (McMurdie and Holmes 2013) was used to generate dissimilarity matrices containing pairwise weighted (i.e., species abundance data) and unweighted (i.e., presence-absence data) UniFrac comparisons between samples. The ability of UniFrac to fairly compare communities is reduced when comparing communities of vastly different sampling depths (Lozupone et al. 2011) or species richness (Ives and Helmus 2010). Two communities composed of endophytes belonging to the same clade should be considered phylogenetically similar. However, a larger community will naturally cover more evolutionary branches in that clade as compared to a smaller community. Therefore, when calculating UniFrac, the number of unshared branches between the two will increase the

PBD value even if their species are closely related. Since the species richness of the endophyte communities of each plant host within a site was highly variable (Appendix - Table 1.1), a fairer comparison was needed to compute PBD.

Phylogenetic community dissimilarity (PCD) is a PBD metric that can more fairly evaluate the phylogenetic dissimilarity of two communities regardless of differences in species richness by dividing itself into a phylogenetic (PCDp) and non-phylogenetic component (PCDc) (Ives and Helmus 2010). PCDp is dependent on the phylogenetic relatedness of unshared species, while PCDc measures compositional similarity based only on shared species. PCD is equal to the product of PCDc and PCDp; the phylogenetic component reduces the effect of species richness on dissimilarity, allowing for fairer comparisons between communities of different sizes. PCD uses presence-absence data and was computed with the function “pcd” from *phyr* (Ives et al. 2020). The T-BAS generated fungal phylogenetic tree lacked finer resolution at lower taxonomic levels and resorted to polytomies with branch lengths equal to 0 to place very closely related OTUs. Because PCD cannot compute dissimilarity when branch lengths are unresolved or equal to 0, a value of 0.000001 (i.e., 1% of the shortest branch length) was added to all branches on the tree to resolve zero length branches while retaining relative branch lengths.

Beta mean nearest taxon distance (bMNTD) measures the phylogenetic distance between a taxon in one community and its nearest phylogenetic neighbor in another community (Fine and Kembel 2011). bMNTD is most effective at revealing phylogenetic patterns occurring at tips of the phylogenetic tree. Abundance-weighted bMNTD was computed using the function “comdistnt” from *picante* (Kembel et al. 2010).

PBD Metric	Data Type	Pros	Cons
Weighted UniFrac	Abundance-weighted	Reveals phylogenetic trends among high abundance species	Does not account for differences in phylogenetic diversity Overlooks phylogenetic patterns among rare species
Unweighted UniFrac	Presence-absence	Reveals broader phylogenetic trends independent of abundance Emphasizes rare species	Does not account for differences in phylogenetic diversity Overlooks important abundance information that may be ecologically significant
PCD	Presence-absence	More fairly compares communities that differ in species richness	May be influenced by differences in phylogenetic diversity
bMNTD	Abundance-weighted	Gives a quantitative value of phylogenetic relatedness in units of phylogenetic distance Reveals phylogenetic patterns at lower taxonomic levels	Relies on phylogenetic trees being finely resolved Overlooks broader phylogenetic patterns

Table 1. A summary of the pros and cons observed in this study of each PBD metric.

Simple Mantel tests were conducted to evaluate the correlation (Spearman's ρ) between the phylogenetic distance separating pairs of plant hosts and the PBDs of their endophyte communities using the function “mantel” from *vegan* with 1000 permutations. Mantel tests were conducted for the entire plant host data set and subsets of plants grouped by class, in order to test for correlation across different ranges of phylogenetic breadth. In the whole plant set, phylogenetic distance between hosts ranged from 0 to 800 nucleotide substitutions per site. The subsets were grouped by class (Magnoliopsida, Polypodiopsida, Lycopodiopsida, and Pinopsida) with the maximum phylogenetic distance between hosts set at 400 nucleotide substitutions per site. The largest subset and most abundant class in every

site was Magnoliopsida. All other subsets lacked enough samples to conduct a Mantel test and were not tested. A Benjamini-Hochberg (BH) test was performed, using the function “p.adjust” from *stats* (R Core Team 2022), on the significance values of the Mantel tests to account for multiple-hypothesis testing across the 20 sites. Correlation was visualized for highly significant sites after BH testing as ranked scatter plots with linear models plotted using the function “ggplot” from *ggplot2* (Wickham 2016). Linear correlation between values of different PBD metrics was tested using the function “cor.test” from *stats*.

PERMANOVA tests were performed on PBD values with plants grouped at three taxonomic levels (class, major group, and order) to determine if endophyte communities of plants within the same taxonomic groups were more or less similar than those of plants belonging to different taxonomic groups. PERMANOVA was computed with 999 permutations using the function “adonis2” from *vegan*.

PBD values were grouped into three comparison categories (Magnolid to Magnolid, Magnolid to Pinopsid, and Pinopsid to Pinopsid). Since an overwhelming majority of the plants at each site belonged to Magnolids (Appendix - Table 1.1), there were fewer PBD comparisons between the few Pinopsids and many Magnolids as compared to the Magnolids to themselves. Thus, 152 comparisons, the maximum number of Pinopsid to Pinopsid comparisons, were randomly subsampled per site from the other comparison groups in order to compare their value distributions. An ANOVA test was performed to evaluate differences in the means of comparison groups, using the function “aov” from *stats*. Tukey’s HSD test was performed using the function “TukeyHSD” from *stats*, to determine which groups differed significantly.

Phylogenetic Beta Diversity - Endophyte Perspective

To investigate if closely related endophytes are more likely to share phylogenetically similar hosts, the correlation between the phylogenetic distance separating pairs of OTUs and PBDs of their host communities were conducted with simple Mantel tests for each site. The analyses are identical to the previously stated except the variables are reversed with OTUs as samples and their observed set of hosts as communities. Host communities were defined by the presence of the OTU within plant genera representing samples based on non-rarefied abundance data. Since host richness varied significantly among OTUs (Appendix - Table 1.1), PCD and bMNTD were used to measure host community PBD, and not UniFrac. OTUs found in only one host were analyzed separately, instead correlating phylogenetic distance separating OTUs to the phylogenetic distance between their single hosts. Significance values were corrected using the BH test.

I also explored specific focal endophyte species to investigate if phylogenetic signals were unique to certain types of endophytes. Three types of focal species (generalists, specialists, and indicator species) were selected based on their ratio of total read abundance to number of unique hosts and from an indicator species analysis. Generalist species were identified as the OTUs with the largest total abundance and largest number of unique hosts. Specialist species were identified as the OTUs with largest total abundance and lowest number of unique hosts. An indicator species analysis was performed using the function “multipatt” from *indicspecies* (De Cáceres and Legendre 2009) to determine which OTUs were significantly associated with plant orders and families at each site. At least five generalist, five specialist, and one indicator species per plant family and order were tested using the function “cor.test” from *stats*. Generalist species were expected to show little to no

correlation since they occupy multiple hosts and other host communities were likely to be nested in theirs. Specialist and indicator species were expected to share phylogenetically similar host communities with their closest relatives, with host communities becoming more dissimilar as phylogenetic distance between species increased. T-tests were performed on Pearson's r slopes of generalists and specialist species using "t.test" from *stats* to determine if their means differed significantly and if they deviated from the null expectation of no correlation ($r = 0$). T-tests were similarly performed on indicator species for each site individually and together by averaging Pearson's r slopes of each site to determine significant deviation from the null expectation.

C. Results

Phylogenetic Trees

T-BAS placed 17,040 fungal ITS1 OTUs (out of 35,965 total), spanning 56 fungal taxonomic classes (Appendix - Table 2.1), with 941 polytomies total across the generated tree. The host phylogenetic tree generated by *V.PhyloMaker2* contained 401 vascular plant genera, spanning four classes, eight major groups, 45 orders and 123 families (Appendix - Table 1.3). The most closely related plant genera were separated by 2.71 nucleotide substitutions per site and the most distantly related genera were separated by 801.57 nucleotide substitutions per site. The majority of plants (73-92%) in each site belonged to class Magnoliopsida, followed by Pinopsida, Polypodiopsida, and Lycopodiopsida (Appendix - Table 1.1).

Phylogenetic Beta Diversity - Plant Perspective

Correlations between host phylogenetic distance and PBDs (UniFrac, PCD, and bMNTD) of endophyte communities for all sites are reported below with examples of sites to illustrate phylosymbiosis.

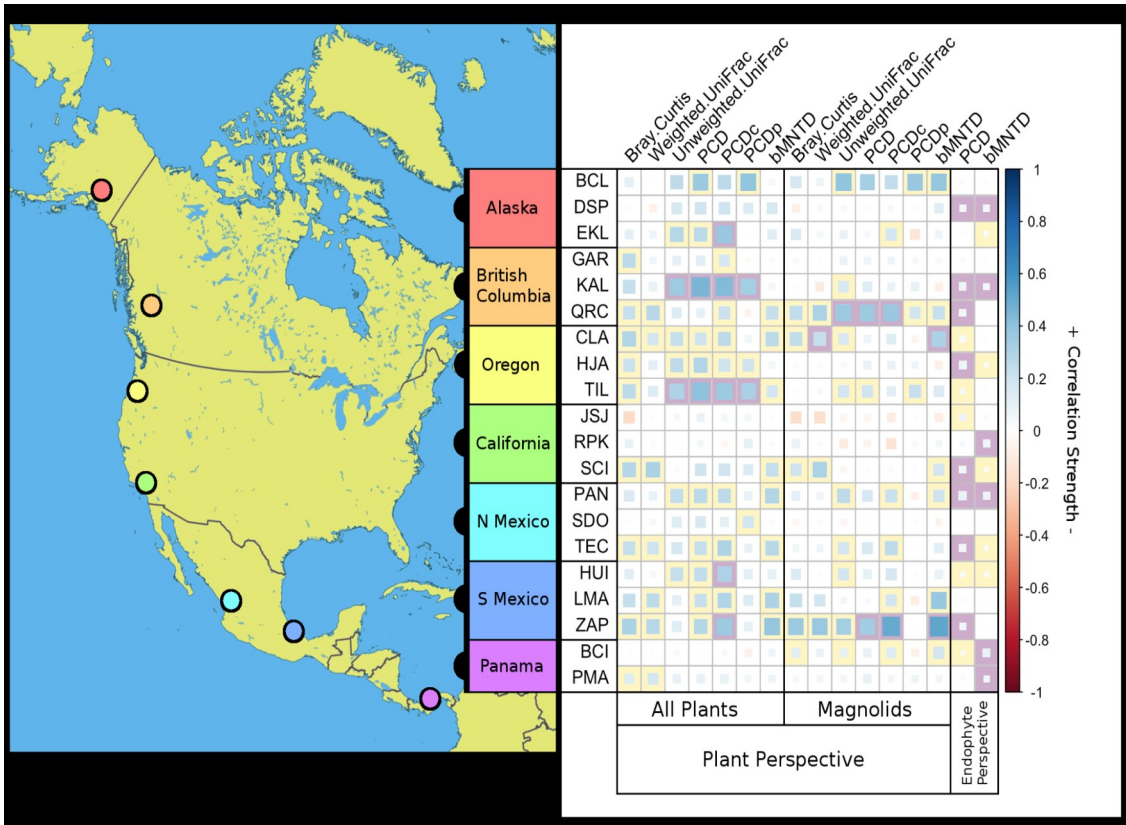


Figure 2. Summary of Mantel test results correlating host or endophyte phylogenetic distance with the phylogenetic dissimilarity of endophyte communities or host communities, respectively. Regions are color coded to their location on the map of North America. Size and color of the inner squares represent the strength and direction of Spearman’s ρ . Correlations with p-values less than 0.05 are outlined by yellow boxes. Significant correlations after correcting p-values with BH tests are outlined by purple boxes. Latitude and longitude coordinates for each site are listed in Appendix - Table 1.2. Spearman correlation and significance values for the plant perspective are listed in Appendix - Table 1.4, and for the endophyte perspective in Appendix - Table 1.7.

Bray-Curtis

Closely related plants shared more similar endophyte species compositions and levels of species abundance as compared to distantly related plants at nine of 20 sites (Figure 2, average $\rho_{BC} = 0.249$), but none were significant after adjusting p-values for multiple testing with BH. Five sites also displayed similar positive correlation values for the Magnolid subset of plants, none of which were significant after BH testing.

UniFrac

Endophyte communities of closely related plants shared more evolutionary branches as compared to distantly related plants at some sites, but this relationship often depended on whether the branches were weighed by the relative abundance of endophytes or not. Weighted UniFrac (wUF) dissimilarities of endophyte communities were significantly positively correlated with the phylogenetic distance of the hosts at seven of 20 sites (Figure 2). After BH testing, no sites were significant. Unweighted UniFrac (uUF) dissimilarities were also positively correlated with host phylogenetic distance at seven of 20 sites, but only one of these (CLA) was also significant with wUF. After the BH test, correlations at sites KAL ($\rho_{uUF} = 0.353$; Figure 3) and TIL ($\rho_{uUF} = 0.309$) remained significant.

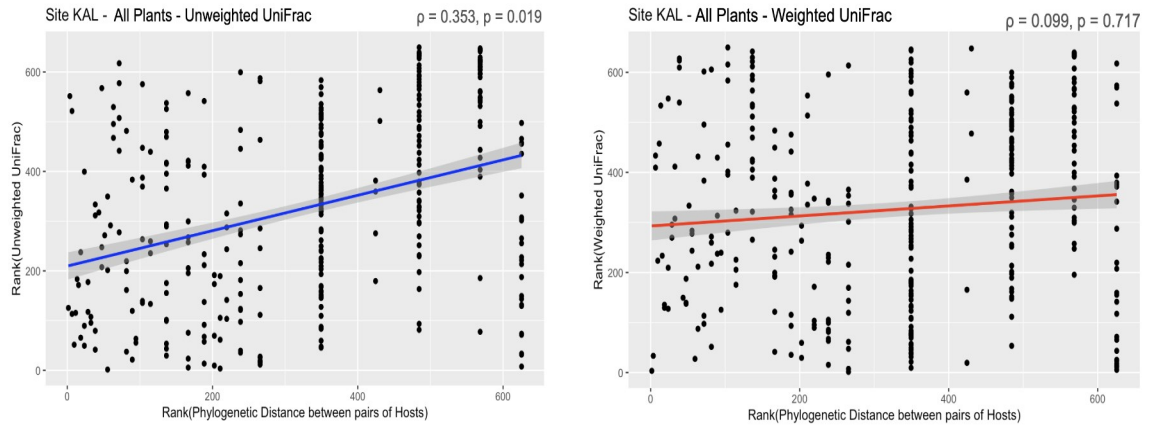


Figure 3. Examples of relationships between Spearman ranks of unweighted and weighted UniFrac dissimilarities of endophyte communities and Spearman ranks of phylogenetic distances between pairs of plant hosts at site KAL (British Columbia). On the left, ranks of phylogenetic distances between hosts were positively correlated with ranks of the unweighted UniFrac dissimilarities of their endophyte communities. On the right, ranks of phylogenetic distance of hosts were not significantly correlated with ranks of weighted UniFrac dissimilarities of their endophyte communities. The phylogenetic signal was lost after incorporating species abundance.

The detection of a phylogenetic signal was sensitive to the phylogenetic breadth of hosts being tested. The correlation between host phylogenetic distance and the uUF dissimilarities of their endophyte communities became insignificant at sites KAL and TIL when only the Magnolid subset was analyzed (Figure 2). On the other hand, the correlation became positively significant at sites CLA and QRC for host phylogenetic distance and wUF (CLA, $\rho_{wUF} = 0.237$) or uUF (QRC, $\rho_{uUF} = 0.367$) after BH correction once the host dataset was pruned to just the Magnolids subset (Figures 2 and 4).

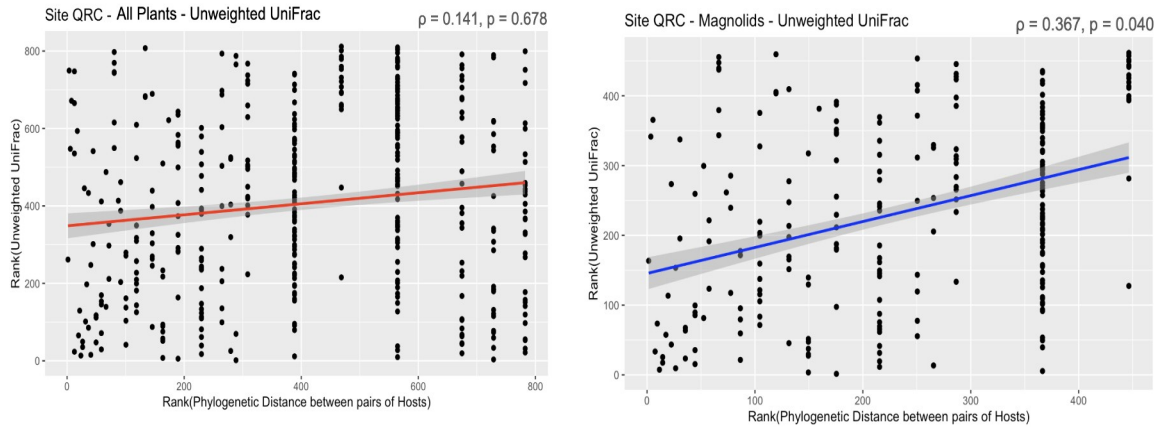


Figure 4. Examples of relationships between Spearman ranks of unweighted UniFrac dissimilarities of endophyte communities and Spearman ranks of phylogenetic distances between pairs of host plants at site QRC (British Columbia). On the left, ranks of phylogenetic distances between all hosts were not significantly correlated with ranks of the unweighted UniFrac dissimilarities of their endophyte communities. On the right, ranks of phylogenetic distance of Magnolid hosts were significantly positively correlated with ranks of unweighted UniFrac dissimilarities of their endophyte communities. In this case, significant correlation was lost when testing across a wider phylogenetic breadth of hosts.

UniFrac Examples

Some sites showed significant correlations between host phylogenetic distance and weighted UniFrac but insignificant correlations when UniFrac metric was unweighted (wUF+ | uUF-). This trend occurs when closely related plants share their most abundant endophyte species, while the less abundant species are not more or less related between closely and distantly related plants. Site SCI in California was an example of such a pattern (Figure 5-A) where closely related hosts, *Quercus* and *Ceanothus*, were both dominated by endophyte species from sister classes of Sordariomycetes and Leotiomyces while hosts distantly related to them, *Corethrogyne*, *Solidago*, and *Erigeron*, were dominated by

endophytes in the Dothideomycetes class. However, when abundance data was removed, all communities appeared more similar regardless of host phylogenetic distance.

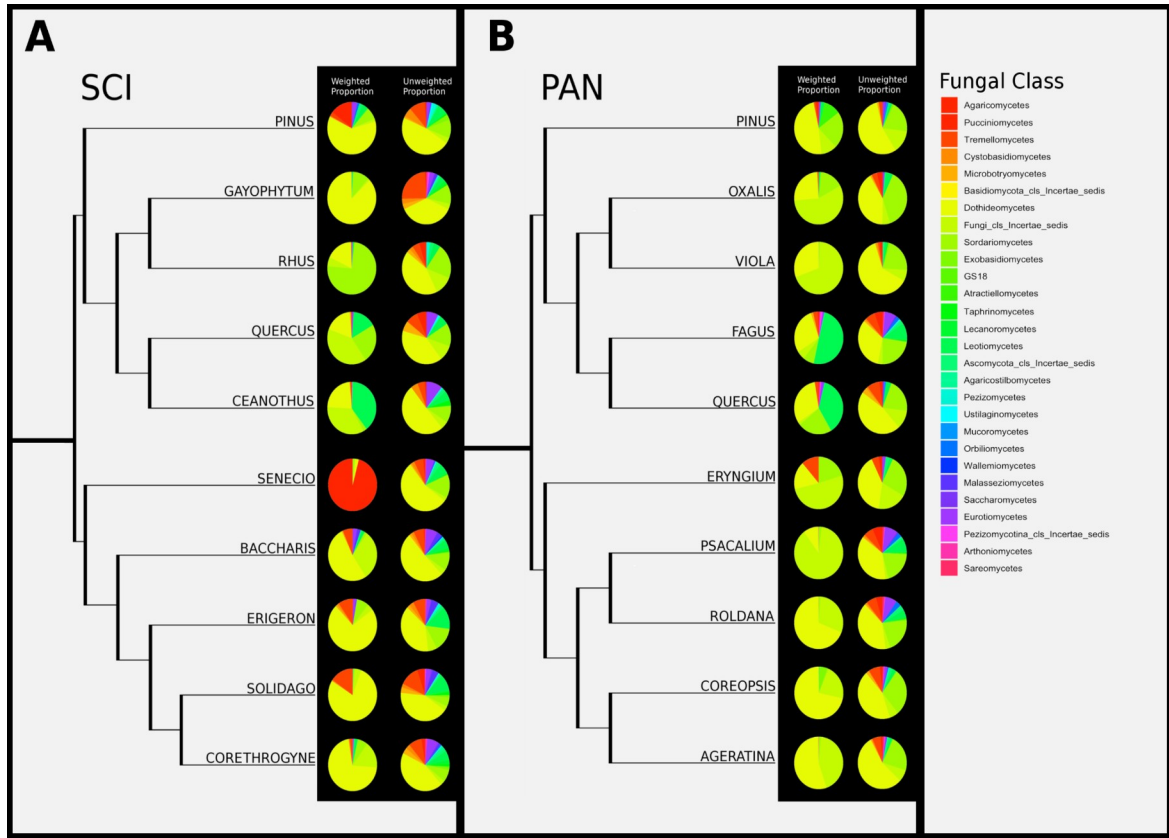


Figure 5. Phylogenies of the plants at site SCI (A) and PAN (B) pruned for the two most distantly related clades and their endophyte community composition represented as abundance-weighted and unweighted pie charts. Pie slices represent the fungal classes of endophytes that compose a host's community. Fungal classes are color coded so the phylogenetic relatedness of classes follows a rainbow gradient. Weighted pie charts have pie slices proportional to the read abundance of each fungal class. Unweighted pie charts are based on presence-absence data; pie slices are proportional to the number of unique endophytes within each class.

On the other hand, some sites showed the reverse trend with significant correlations between host phylogenetic distance and unweighted UniFrac but insignificant correlations for weighted UniFrac (wUF- | uUF+). This pattern can occur when closely related plants share more phylogenetically similar rare species, but the most abundant species are either distantly related or ubiquitous in every host plant. For example, at site PAN in northern Mexico (Figure 5-B), the weighted community of *Psacalium* was more phylogenetically similar to the weighted community of a distant relative, *Oxalis* (wUF = 0.169), as compared to its closest relative in the phylogeny, *Roldana* (wUF = 0.420). When the UniFrac metric is unweighted, the *Psacalium* community was more similar to that of *Roldana* (uUF = 0.618) than to that of *Oxalis* (uUF = 0.640).

Phylogenetic Community Dissimilarity (PCD)

Endophyte communities of closely related plants were more phylogenetically similar than expected by chance at 11 of 20 sites with PCD (Figure 2). However, their dissimilarity depended heavily on shared species compositions (PCDc) rather than the phylogenetic structure of unshared species (PCDp). After BH correction, KAL ($\rho_{PCD} = 0.453$, $\rho_{PCDc} = 0.434$, $\rho_{PCDp} = 0.337$) and TIL ($\rho_{PCD} = 0.400$, $\rho_{PCDc} = 0.362$, $\rho_{PCDp} = 0.327$) showed significant correlations for PCD, PCDc, and PCDp. For both sites, the correlation strength for PCDc was greater than PCDp ($\rho_{PCD} > \rho_{PCDc} > \rho_{PCDp}$), indicating that as the phylogenetic distance between hosts increased, the increasing dissimilarity between their communities was driven more by differences in species composition (PCDc) than the increasing phylogenetic distances of their unshared species (PCDp).

For three sites, PCDC was significantly positively correlated with host phylogenetic distance (EKL, $\rho_{PCDC} = 0.373$; HUI, $\rho_{PCDC} = 0.301$; ZAP, $\rho_{PCDC} = 0.369$; Figure 2) after BH correction, but PCDp was not, resulting in insignificant correlations between host phylogenetic distance and the overall PCD of communities at these sites. Since PCDC acts as a compositional beta diversity metric, in these sites communities of closely related plants were more likely to share the same endophyte species (Figure 6-A), but their unshared species were not likely to be phylogenetically similar (Figure 6-B). Their unshared species were phylogenetically dissimilar enough to deem the communities of closely related plants dissimilar (Figure 6-C).

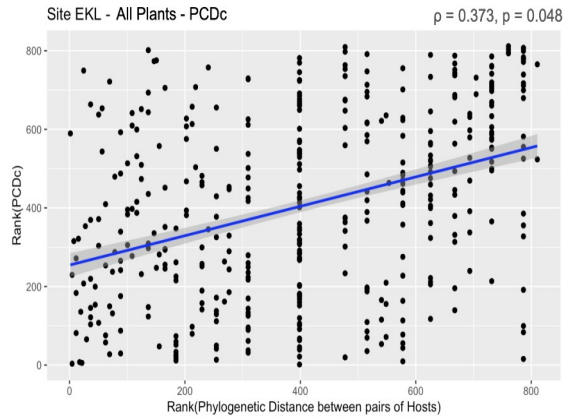
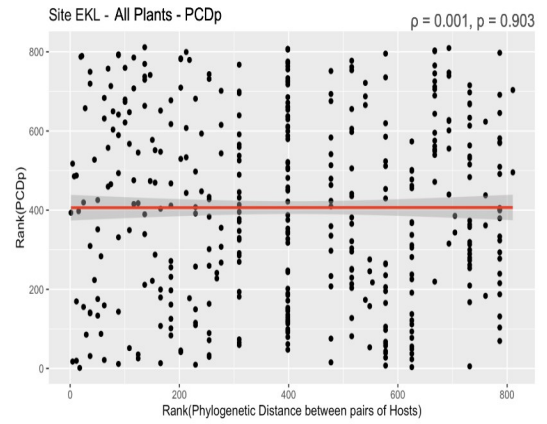
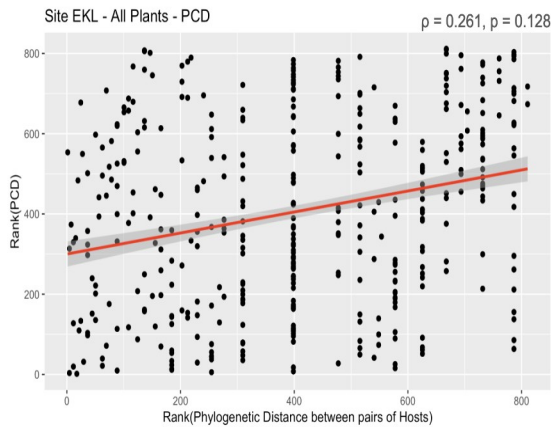
(A)**(B)****(C)**

Figure 6. Examples of relationships between Spearman ranks of PCD, PCDc and PCDp of endophyte communities and Spearman ranks of phylogenetic distances between pairs of host plants at site EKL (Alaska). Ranks of phylogenetic distances between hosts were significantly positively correlated with ranks of (A) PCDc of their endophyte communities. Ranks of phylogenetic

distance of hosts were not significantly correlated with ranks of (B) PCDp and (C) PCD of their endophyte communities.

Similar to trends seen with UniFrac, testing across different plant phylogenetic breadths affected the detection of significant correlations between host phylogenetic distance and PCD. There were more positive correlations when testing included all plants as compared to within the Magnolid subset (11 vs three sites before BH correction). Sites with significant correlations between PCD and host phylogenetic distance across the entire plant set after BH correction (KAL and TIL) showed no significant correlations for PCD within their Magnolids (Figure 7). On the other hand, sites QRC and ZAP showed significant correlation for both PCD (QRC, $\rho_{\text{PCD}} = 0.355$, $p = 0.02$; ZAP, $\rho_{\text{PCD}} = 0.341$, $p = 0.038$) and PCDc (QRC, $\rho_{\text{PCDc}} = 0.375$, $p = 0.019$; ZAP, $\rho_{\text{PCDc}} = 0.506$, $p = 0.019$) after BH correction for their Magnolid subset but not for all plants. Site QRC and ZAP showed stronger PCDc correlation than their correlation for PCD ($\rho_{\text{PCDc}} > \rho_{\text{PCD}}$), following the trend of closely related plants being more compositionally similar than phylogenetically similar.

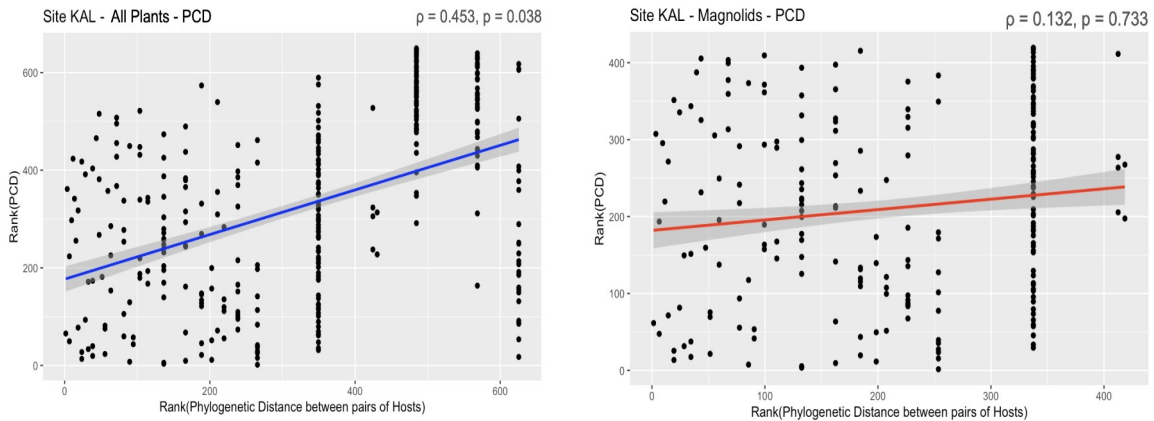


Figure 7. Relationships between Spearman ranks of PCD of endophyte communities and Spearman ranks of phylogenetic distances between pairs of all host plants or Magnolid hosts for site KAL (British Columbia). On the left, ranks of phylogenetic distances between all plants are positively correlated with ranks of the PCD of their endophyte communities. On the right, ranks of phylogenetic distance of Magnolid hosts

were not significantly correlated with ranks of PCD of their endophyte communities. In this case, significant correlation was lost when testing across a narrower phylogenetic breadth of hosts.

Beta Mean Nearest Taxon Distance (bMNTD)

As compared to the other PBD metrics, correlation between host phylogenetic distance and bMNTD of endophyte communities was the least common among sites (Figure 2). Only eight of 20 sites showed significant positive correlations for the entire plant set with none remaining significant after BH correction. Significant positive correlation between host phylogenetic distance and bMNTD of endophyte communities in Magnolids was observed for eight of 20 sites, with CLA ($\rho_{\text{bMNTD}} = 0.315$) and ZAP ($\rho_{\text{bMNTD}} = 0.520$) (Figure 8) remaining significant after BH testing. Sites CLA and ZAP did not show significance for this bMNTD correlation among all their plants, suggesting that phyllosymbiosis was only detectable when removing comparisons against communities of distantly related plants.

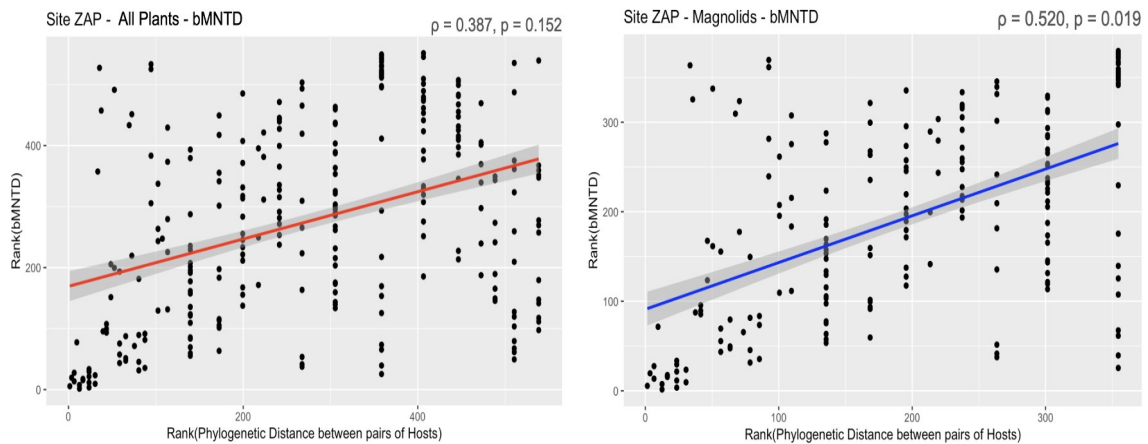


Figure 8. Examples of relationships between Spearman ranks of bMNTD of endophyte communities and Spearman ranks of phylogenetic distances between pairs of host plants and pairs of Magnolids at site ZAP (S Mexico). On the left, ranks of phylogenetic distances between all plant hosts were not significantly correlated with ranks of the bMNTD of their endophyte communities. On the right, ranks of

phylogenetic distance of Magnolid hosts were significantly positively correlated with ranks of bMNTD of their endophyte communities. Significant correlation was lost when testing across a wider phylogenetic breadth of hosts.

Correlations between PBD Metrics

Significant correlations were found between some PBD metrics (Figure 9). For example, all seven sites that showed significant correlation for unweighted UniFrac also did for PCD and PCDC (Figure 2). Values of unweighted UniFrac were highly correlated with those of PCD ($r_{uUF-PCD} = 0.822$) and PCDC ($r_{uUF-PCDC} = 0.834$), indicating that the ratio of branches shared and unshared between communities (uUF) reflected similar levels of dissimilarity as the likelihood of shared species between communities (PCDC). PCDC was more highly correlated with PCD ($r_{PCD-PCDC} = 0.834$) than PCDp ($r_{PCD-PCDp} = 0.647$), further suggesting that PCD between endophyte communities was more influenced by compositional similarities than by phylogenetic relationships of the unshared endophyte species (Figure 6).

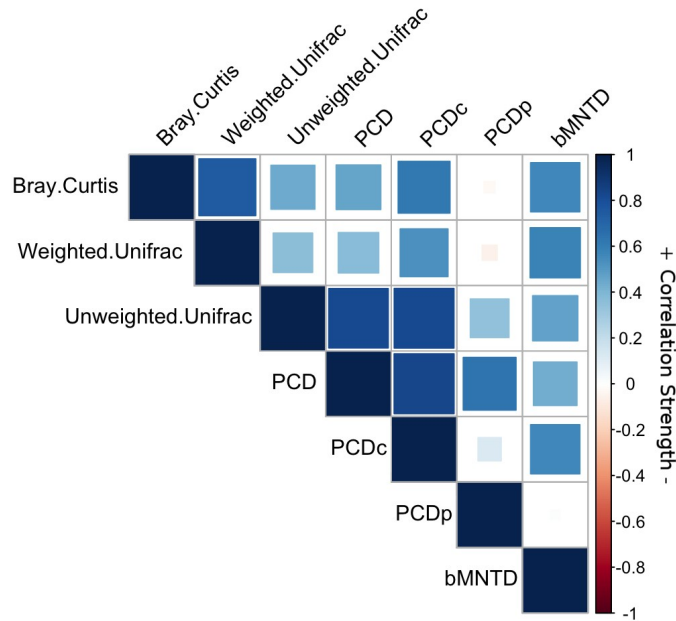


Figure 9. Summary of Pearson's correlations between PBD values of endophyte communities across all sites. Size and color of squares represent correlation strength and direction. All correlations were statistically significant ($p < 0.0001$) (Appendix - Table 1.9).

PERMANOVA

The phylogenetic dissimilarities among endophyte communities tended to be more significantly different between plants of different classes as compared to plants of different major groups and orders (Figure 10). The number of sites with significantly dissimilar endophyte communities across plant class substantially dropped when comparing plants by major group and order for unweighted UniFrac (Figure 10-B) and PCD (Figure 10-C). This concurs with the positive correlation between host phylogenetic distance and PBDs of communities across larger phylogenetic breadths that becomes insignificant when comparing only within the Magnolids (Figures 2 and 7). Sites that did not test significantly for weighted UniFrac (Figure 10-A) tested significantly when abundance was removed in unweighted UniFrac (Figure 10-B), suggesting that dissimilarity was reduced across class

due to the most abundant portions of the endophyte communities being similar between distantly related plants (Figure 3). This trend was common across sites as more sites significantly tested for the presence-absence PBD metrics (uUF, PCD) than the abundance-weighted metric (wUF, bMNTD) across plant class (16/20 vs 9/20).

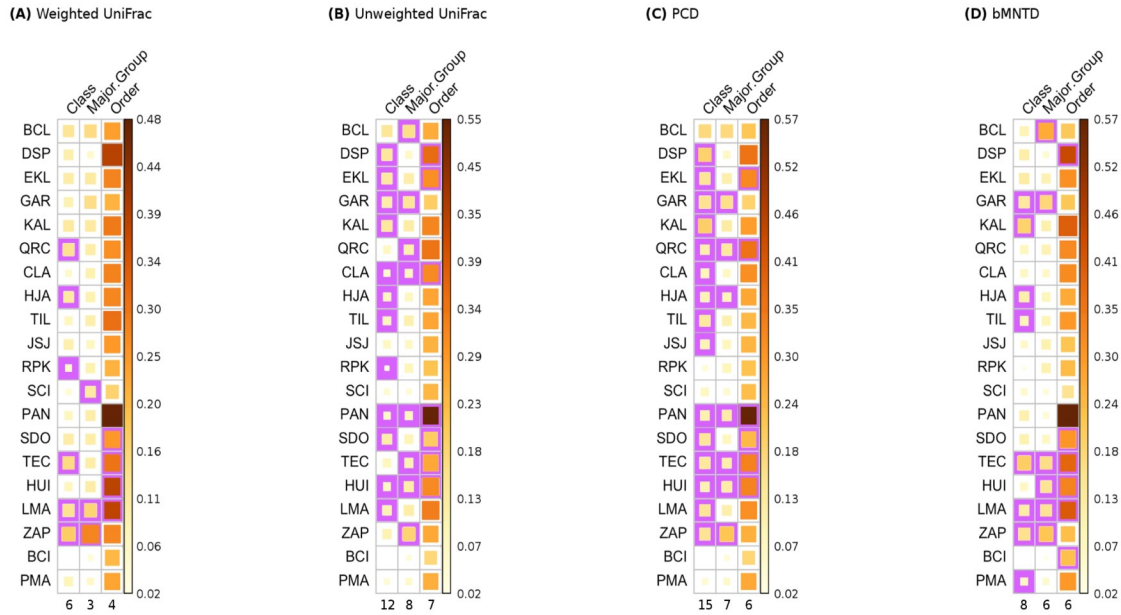


Figure 10. Summary of PERMANOVA results testing for differences between PBDs of endophyte communities belonging to plants of different taxonomic groups (class, major group, order). Size and color of boxes represent the R^2 value of the test. Results with p-values less than 0.05 are boxed in purple. Number of sites that showed significant results per taxonomic group tested are listed at the bottom of each column. PERMANOVA results are listed in Appendix - Table 1.5.

Phylogenetic Beta Diversity - Comparing between Plant Classes

PBDs between endophyte communities of Magnolid and Pinopsid host pairs were significantly greater than PBDs between pairs of hosts within the same class (Figure 11). For unweighted UniFrac (Figure 11-B) and PCD (Figure 11-C), communities of plants belonging to hosts of different classes were more phylogenetically dissimilar than

communities of plants belonging to the same class. For weighted UniFrac (Figure 11-A) and bMNTD (Figure 11-D), communities of Pinopsids were equally dissimilar to each other as they were to communities of Magnolids. Average values of unweighted UniFrac were greater than those of weighted UniFrac values suggesting that communities were more phylogenetically similar when species abundance was incorporated.

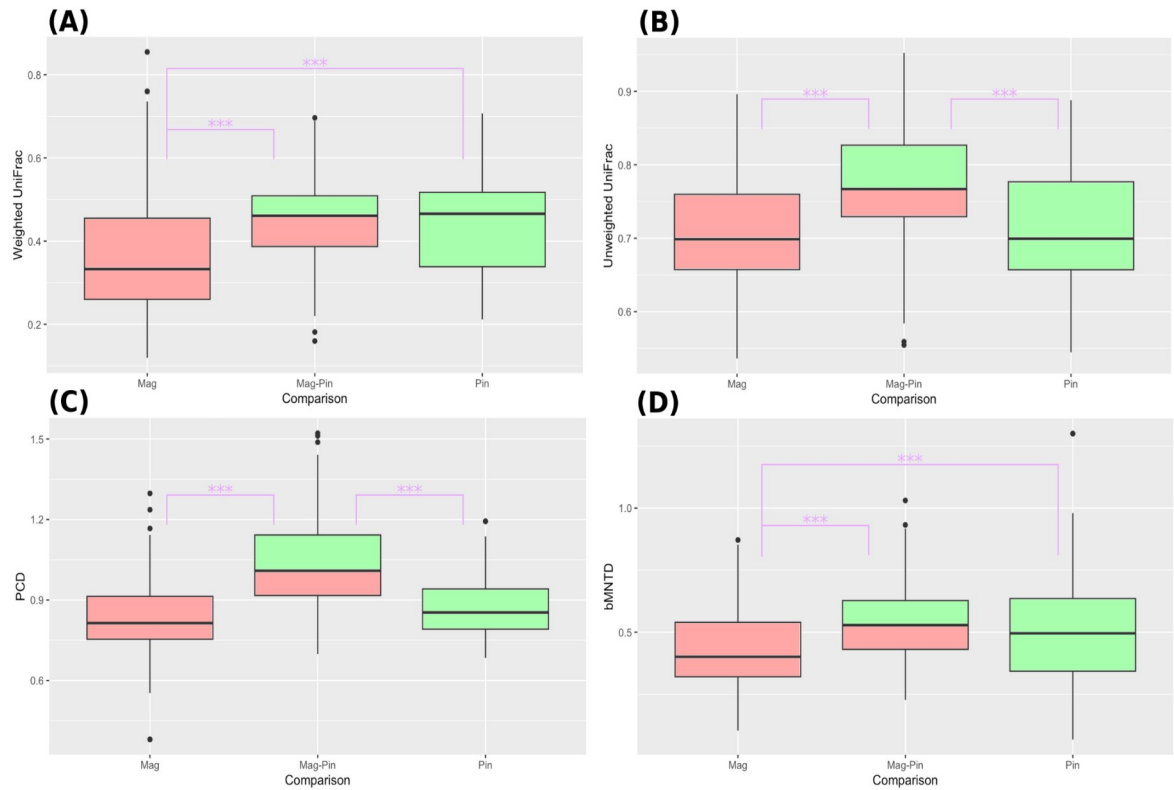


Figure 11. Box and whisker plots of 152 randomly selected PBD values from three categories of comparisons of plants. PBD values are categorized by the classes of the hosts being compared (Mag = Magnolid to Magnolid, Mag-Pin = Magnolid to Pinopsid, and Pin = Pinopsid to Pinopsid). Significantly different groups by Tukey's HSD are marked by purple asterisks ($p < 0.003$). Results of ANOVA and Tukey's HSD are listed in Appendix - Table 1.6.

Phylogenetic Beta Diversity - Endophyte Perspective

Host communities of closely related endophytes were not substantially more or less phylogenetically similar as compared to those of distantly related endophytes. Thirteen of 20 sites (eight after BH correction) showed significant but weak positive correlation (ρ_{PCD} range: 0.056 - 0.076) between the phylogenetic distance separating endophyte OTUs and the PCD of their host communities (Figure 2). Eleven of 20 sites (6 after BH correction) showed significant weak positive correlation (ρ_{PCD} range: 0.050 - 0.086) between phylogenetic distance separating OTUs and the bMNTD of their communities (Figure 2). There was no significant correlation found between the phylogenetic distance separating endophytes appearing in only one host and the phylogenetic distance separating their hosts (Appendix - Table 1.8).

The host communities of focal species (generalist, specialist, and indicator species) were not substantially different between closely related endophytes as compared to between distantly related endophytes (mean Pearson's $r \sim 0$; Figure 12) across both PBD metrics, reflecting the results of the Mantel tests. However, for generalist species correlation was slightly more negative than that of specialist species, and significantly deviated from the null expectation of no correlation (mean $r_{PCD} = -0.012$, $p = 0.031$; mean $r_{bMNTD} = -0.038$, $p = 0.031$) (Figure 12) (Appendix - Table 2.24). Significant positive correlations were identified among some indicator species, but this was specific to certain OTUs and was not a prevalent pattern in the larger data set (mean $r_{PCD} = 0.001$, mean $r_{bMNTD} = -0.010$).

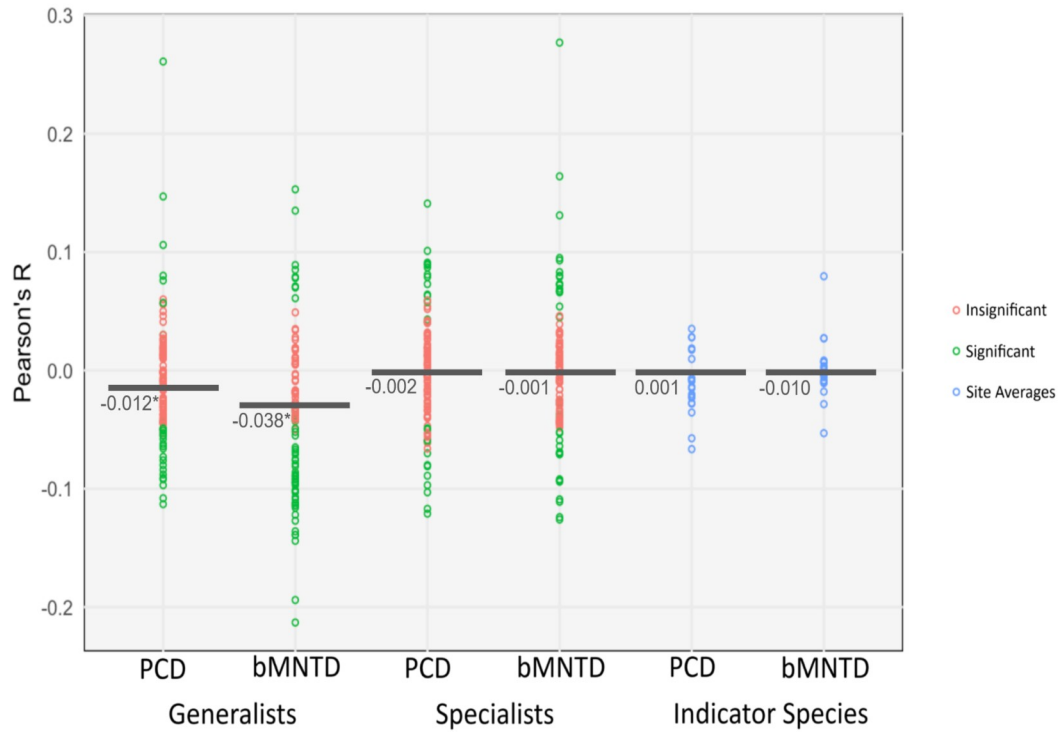


Figure 12. Pearson's r coefficients for generalist and specialist endophyte species. For generalists and specialists tested, green points represent Pearson's r correlation coefficients for species that showed significant correlation ($p < 0.05$), while orange points represent coefficients from insignificant correlations. Blue points represent average Pearson's r values of indicator species for each site. Dark gray bars represent the mean Pearson's r for each set of points. Means that significantly deviated from $r = 0$ are noted with asterisks. Pearson's r and significance values, and t-test results are listed in Appendix - Table 2.2-25.

D. Discussion

Plant Perspective

Overall, closely related plants did not share more closely related endophyte communities as compared to distantly related plants. Phyllosymbiosis was not a prevalent

pattern across sites or regions when considering the compositional beta diversity and phylogenetic beta dissimilarity of the endophyte communities. Thirteen of 20 sites showed no significant correlation between host phylogenetic distance and endophyte community dissimilarity after BH correction (Figure 2), similar to the findings of Vincent et al. (2016) and Whitaker et al. (2020). Communities in these sites may have been homogenized so closely and distantly related plants shared phylogenetically similar communities. Endophytes in the tropics have been observed to occupy broader host ranges (Arnold and Lutzoni 2007) as compared to endophytes in more temperate zones (Zhang and Yao 2015). Sites located in tropical climates, such as BCI and PMA in Panama, did not show significant correlation for any of the PBD metrics as expected, possibly due to the homogeneity of endophyte communities (Figure 10) driven by broad host associations. Spatial distance was not tested in these analyses but may explain this result, as horizontal transmission of endophytes (Rodriguez et al. 2009) and cross-colonization from neighboring plants can contribute to community homogenization between closely located plants. For other cases, perhaps there was no discernible or linear pattern in assemblage dissimilarity along host phylogenetic distance as a result of non-phylogenetic, stochastic processes.

Seven of 20 sites showed highly significant correlation (i.e., significant after BH correction) between host phylogenetic distance and endophyte community dissimilarity, and displayed different types of phylosymbiotic patterns for different PBD metrics and subsets of plants (Figure 2). These patterns revealed that endophyte species abundance, phylogenetic breadth of hosts, and compositional differences between communities were relevant factors in detecting phylosymbioses at these sites.

Effect of Endophyte Species Abundance

Phylosymbiosis was occurring among rarer species and not the most abundant members of endophyte communities. Sites that showed significant correlations between host phylogenetic distance and presence-absence PBD metrics did not show significant correlations for abundance-weighted PBD metrics (Figures 2 and 3). Endophyte communities were more phylogenetically similar among the most abundant portions of their communities as weighing abundance decreased the UniFrac dissimilarity between communities of both closely and distantly related plants ($wUF < uUF$) (Figure 11). This was also true for compositional beta diversity as sites significant for correlation between host phylogenetic distance and PCDC (a presence-absence metric) were not significantly correlated for Bray-Curtis (an abundance weighted metric), suggesting that closely related plants were more likely to share the same endophyte species but not in similar abundances. My findings contrast with Liu et al. (2019), who found phylosymbiosis among the communities of *Ficus* using Bray-Curtis dissimilarity. Rather, I found that discounting abundance revealed phylosymbiosis among rarer species.

Sensitivity to Phylogenetic Breadth of Plant Hosts

The detection of phylosymbiosis was sensitive to the phylogenetic breadth of plant hosts being tested. Sites that showed significant correlation between host phylogenetic distance and PBDs of endophyte communities across all their plants were not significant for the Magnolid subset (Figure 2 and 7). Simultaneously, there were also sites with significant correlation between host phylogenetic distance and PBDs of endophyte communities for the Magnolid subset and insignificant correlation for the entire set of plants (Figures 2, 4, and

8). However, this was not a prevalent pattern, as without conservative hypothesis testing (i.e. before BH correction), sites displayed significant correlations for the whole plant set and lost significance with the Magnolid subset. There was stronger indication of the largest PBD occurring between communities belonging to plants of different classes (Figure 10 and 11), suggesting that decreasing the phylogenetic breadth of hosts being tested led to the loss of a phylogenetic signal. The significant positive correlation displayed by sites for all their plants was driven by the PBD comparisons between plants separated by the largest phylogenetic distances, such as Magnolid hosts against Pinopsid hosts.

Effect of Compositional Differences between Endophyte Communities

Dissimilarity appeared to be driven by compositional differences over phylogenetic relatedness. PCDC was the more influential component in determining PCD over PCDp between endophyte communities (Figures 6 and 9), meaning that compositional differences between communities were driving dissimilarity rather than the phylogenetic unrelatedness between unshared species. I hypothesized that phylogenetic trait conservatism underlying host-specificity in endophyte species (Gilbert and Webb 2007) would lead closely related endophyte species to appear in closely related hosts. However, it appeared that closely related plants were more likely to share the same endophyte species but did not necessarily share more phylogenetically related species as compared to distantly related plants. This may be explained by closely related plants imposing similar host filters which select for certain endophyte lineages (Lin et al. 2024), and species from different lineages having similar capability to persist through these filters. Furthermore, coexistence of sister species

from the same lineage may be discouraged by other processes, such as interspecific competition (Saunders et al. 2010).

Endophyte Perspective

Phylosymbiosis was not detected from the endophyte perspective. Host communities between closely related endophytes were only slightly more dissimilar as compared to those of distantly related endophytes; however, this correlation was very weak. Generalists that were defined by greater host breadth and larger read abundances, appeared to share slightly more phylogenetically similar hosts with their distant relatives as compared to their close relatives (Figure 12). Abundance has been hypothesized to be correlated with endophyte growth rate and competitive ability (Huang 2020), possibly explaining how generalist species with large abundances may be outcompeting their closest relatives. Host communities were also defined using non-rarefied presence-absence data, which may have overestimated the ecologically relevant host communities of endophytes and weakened any signals of phylosymbiosis. Overall, these exploratory analyses from the endophyte perspective did not detect phylosymbiosis as observed from the plant perspective.

Potential Biases and Caveats

Branch-Based PBD Metrics

Since branch-based PBD metrics, such as UniFrac, can be biased when communities have variable species richness (SR) or phylogenetic diversity (PD) (Leprieur et al. 2012), such as in our case (Appendix - Table 1.1), we focused much of our analyses on PCD. However, randomly sub-sampling the endophyte communities to equal richness may offer a

solution. Another method would be to perform partial Mantel tests between host phylogenetic distance and branch-based PBD while controlling for differences in PD or SR.

Phylogenetic Trees

Lack of phylogenetic resolution intrinsically compromises accurate computations of PBDs of endophyte communities. Some phylogenetic resolution was lost in deeper clades as the generated endophyte phylogenetic tree contained 941 polytomies, which most likely diluted phylogenetic distance measurements between OTUs belonging to these polytomies. There is an ongoing conversation about the robustness of PBD metrics in dealing with low resolution phylogenies and polytomies, and about the pros and cons of building phylogenies from sequence data for phylogenetic analyses (Li et al. 2019). Since the basal clades were more resolved, these PBD metrics were able to capture phylogenetic dissimilarity on a broad scale. I expect that the expansion of fungal species databases and updated phylogenetic placement will create more finely resolved trees, allowing PBD metrics to more accurately represent phylogenetic relatedness between endophyte communities.

Sampling Design

In this sampling design, the majority of plants sampled at each site belonged to the class Magnoliopsida (Appendix - Table 1.1). More representation of other classes would have provided further insight into phylogenetic signals within and across classes. It would be most informative to sample a large and equal number of plant taxa across a broad phylogenetic breadth. Lastly, I acknowledge that the methods used to extract endophyte

DNA from plant tissue are prone to error and may not capture all species richness (Hyde and Soyong 2008).

Future Avenues

More Phylogenetic Analyses

I designed my phylogenetic approach around PBD metrics that followed different conceptual frameworks in order to compare them and gain unique insights about this data set. There were PBD metrics unused in this paper that could be used for future studies. For example, PhyloSor (Bryant et al. 2008) is another widely used branch-based PBD metric that only uses shared evolutionary branches to compute dissimilarity. PhyloSor uses presence-absence data, so UniFrac was used in this paper instead as it offers both a presence-absence and an abundance-weighted version which can be compared to assess the relevance of species abundance in dissimilarity. Net relatedness index (NRI) and nearest taxon index (NTI) (Webb et al. 2002) are alpha diversity metrics that measure how phylogenetically clustered species are within a community. NRI and NTI can be modified into beta diversity metrics to measure how phylogenetically clustered species are between pairs of communities (González-Caro et al. 2014). It would be beneficial to compare the results from various phylogenetic analyses in order to determine the appropriate metrics for high-throughput compositional community datasets and inspire new thoughts to develop improvements and novel metrics.

Given the ancient history and hyper-diversity of fungi (Naranjo-Ortiz and Gabaldón 2019), phylogenetic signals may not be consistently present across all lineages or phylogenetic scales. It is possible that certain fungal clades display phylosymbiosis and not

others. Therefore, pruning the fungal phylogenetic tree for subsets of endophytes may identify clade-specific phylosymbiotic signals, although none of the focal endophytes studied in this paper displayed patterns of phylosymbiosis. Focused investigations of host-specific fungal clades would be a next step to understanding FFE assemblage.

Common Garden Experiments

Phylogenetic signals were undetected across most of the 20 sites using this sampling design, but this does not mean that host phylogeny is not a factor of FFE assemblage. Perhaps host evolutionary history is a weaker factor in a natural environment when placed in a complicated web of interactions between abiotic and biotic factors. The role of host phylogeny would be better tested if the effects of these other factors were reduced by exerting control over the study area (Brooks et al. 2016). Common garden experiments, such as the ones conducted by Liu et al. (2019) and Whitaker et al. (2020), can address the unpredictable challenges posed by stochastic processes. Allowing endophyte communities to grow and establish across a wide variety of plants in the absence or reduced presence of non-phylogenetic assemblage factors could better reveal how evolutionary histories of hosts and fungi sort endophyte communities. An accessible garden can also be revisited to obtain diversity measurements over time, allowing for the long-term assessment of phylogenetic turnover and the dynamics of assemblage.

The Phylosymbiotic Model

The ultimate method to determine the true host range of an endophyte and collect conclusive evidence for phylosymbiosis would be to manually infect every single plant

species and observe an endophyte's ability to colonize them. In terms of time, resource requirements, and capability to exert control over abiotic and biotic factors, such a thorough inoculation experiment is impractical and may be too artificial to resemble natural infections. If host specificity is phylogenetically conserved and consistent patterns of phylosymbiosis are identified, then phylogenetic relationships can be used to develop a model to predict host-endophyte associations. To develop such a model is a daunting task and will require expansion of our current knowledge of the functional traits that drive symbioses between endophytes and plants, the relationship between these functional traits and plant-fungal phylogenetic history, and the extent to which phylosymbiosis occurs in plant-fungal systems. Applications of this model will be incredibly useful in agriculture, genetic engineering, and ecological manipulation to benefit our purposes, and will also serve to advance our understanding of other symbiotic systems.

References

1. Apigo, A. (2022). Life in the leaves: diversity and distribution of foliar fungal endophytes. *UC Santa Barbara*. ProQuest ID: Apigo_ucsb_0035D_15731. Merritt ID: ark:/13030/m5909jp5. Retrieved from <https://escholarship.org/uc/item/8ft5k6gp>.
2. Arnold EA, Mejía LC, Kyllo D, Rojas EI, Maynard Z, Robbins N, Herre EA. Fungal endophytes limit pathogen damage in tropical tree. *PNAS*. 2003; 100(26): 15649-15654. <https://doi.org/10.1073/pnas.2533483100>.
3. Arnold, AE. Understanding the diversity of foliar endophytic fungi: progress, challenges, and frontiers. *Fungal Biology Reviews*. 2007 May-Aug; 21(2-3): 51-66. <https://doi.org/10.1016/j.fbr.2007.05.003>.
4. Arnold AE, Lutzoni F. Diversity and host range of foliar fungal endophytes: are tropical leaves biodiversity hotspots? *Ecology*. 2007 Mar; 88(3): 541-9. doi: 10.1890/05-1459.
5. Bacon CW, Porter JK, Robbins JD, Luttrell ES. *Epichloë typhina* from toxic tall fescue grasses. *Appl Environ Microbiol*. 1977 Nov; 34(5): 576-81. doi: 10.1128/aem.34.5.576-581.1977.
6. Barbera P, Kozlov AM, Czech L, Morel B, Darriba D, Flouri T, Stamatakis A. EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences. *Syst Biol*. 2019 Mar 1; 68(2): 365-369. doi: 10.1093/sysbio/syy054.
7. Blaalid R, Kumar S, Nilsson RH, Abarenkov K, Kirk PM, Kauserud H. ITS1 versus ITS2 as DNA metabarcodes for fungi. *Mol Ecol Resour*. 2013 Mar; 13(2): 218-24. doi: 10.1111/1755-0998.12065.
8. Brooks AW, Kohl KD, Brucker RM, van Opstal EJ, Bordenstein SR. Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. *PLoS Biol*. 2016 Nov 18; 14(11): e2000225. doi: 10.1371/journal.pbio.2000225.

9. Bryant JA, Lamanna C, Morlon H, Kerkhoff AJ, Enquist BJ, Green JL. Colloquium paper: microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *Proc Natl Acad Sci U S A*. 2008 Aug 12;105 Suppl 1(Suppl 1):11505-11. doi: 10.1073/pnas.0801920105.
10. Busby PE, Ridout M, Newcombe G. Fungal endophytes: modifiers of plant disease. *Plant Mol Biol*. 2016 Apr; 90(6): 645-55. doi: 10.1007/s11103-015-0412-0.
11. Carbone I, White JB, Miadlikowska J, Arnold AE, Miller MA, Magain N, U'Ren JM, Lutzoni F. T-BAS version 2.1: Tree-Based Alignment Selector toolkit for evolutionary placement of DNA sequences and viewing alignments and specimen metadata on curated and custom trees. 2019. doi: 10.1128/mra.00328-19.
12. Christian N, Herre EA, Clay K. Foliar endophytic fungi alter patterns of nitrogen uptake and distribution in *Theobroma cacao*. *New Phytol*. 2019 May; 222(3): 1573-1583. doi: 10.1111/nph.15693.
13. David EL, Weatherhead E, Koide RT. The potential saprotrophic capacity of foliar endophytic fungi from *Quercus gambelii*. *Fungal Ecology*. 2023 April; 62. <https://doi.org/10.1016/j.funeco.2022.101221>.
14. De Cáceres M, Legendre P. Associations between species and groups of sites: indices and statistical inference. *Ecology*. 2009 Dec; 90(12): 3566-74. doi: 10.1890/08-1823.1.
15. Doyle JJ, and JL Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19(1):11-15.
16. Fine PVA, Kembel SW. Phylogenetic community structure and phylogenetic turnover across space and edaphic gradients in western Amazonian tree communities. *Ecography*. 2011 Aug; 34(4): 552–565. <https://doi.org/10.1111/j.1600-0587.2010.06548.x>.
17. Gilbert GS, Webb CO. Phylogenetic signal in plant pathogen-host range. *Proc Natl Acad Sci U S A*. 2007 Mar 20; 104(12): 4979-83. doi: 10.1073/pnas.0607968104.

18. González-Caro S, Umaña MN, Álvarez E, Stevenson PR, Swenson NG. Phylogenetic alpha and beta diversity in tropical tree assemblages along regional-scale environmental gradients in northwest South America. *Journal of Plant Ecology*. 2014 April; 7(2): 145–153. <https://doi.org/10.1093/jpe/rtt076>.
19. Grabka R, d'Entremont TW, Adams SJ, Walker AK, Tanney JB, Abbasi PA, Ali S. Fungal Endophytes and Their Role in Agricultural Plant Protection against Pests and Pathogens. *Plants (Basel)*. 2022 Jan 30; 11(3): 384. doi: 10.3390/plants11030384.
20. Huang YL. Effect of Host, Environment and Fungal Growth on Fungal Leaf Endophyte Communities in Taiwan. *J Fungi (Basel)*. 2020 Oct 23;6(4):244. doi: 10.3390/jof6040244.
21. Hyde KD, Soyong, K. The fungal endophyte dilemma. *Fungal Diversity*. 2008 Nov 30. 33:163–173.
22. Ives AR, Helmus MR. Phylogenetic metrics of community similarity. *Am Nat*. 2010 Nov; 176(5): E128-42. doi: 10.1086/656486.
23. Ives A, Dinnage R., Nell LA, Helmus M, Li D (2020). `_phyr: Model Based Phylogenetic Analysis_`. R package version 1.1.0, <<https://CRAN.R-project.org/package=phyr>>.
24. Jin Y, Qian H. V.PhyloMaker2: An updated and enlarged R package that can generate very large phylogenies for vascular plants. *Plant Divers*. 2022 May 27; 44(4): 335-339. doi: 10.1016/j.pld.2022.05.005.
25. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*. 2010 Jun; 26: 1463-1464. <https://doi.org/10.1093/bioinformatics/btq166>.

26. Kolattukudy PE. Enzymatic penetration of the plant cuticle by fungal pathogens. *Annual Review of Phytopathology*. 1985 Sep; 23: 223-250. <https://doi.org/10.1146/annurev.py.23.090185.001255>.
27. Krings M, Taylor TN, Hass H, Kerp H, Dotzler N, Hermsen EJ. Fungal endophytes in a 400-million-yr-old land plant: infection pathways, spatial distribution, and host responses. *New Phytol*. 2007; 174(3): 648-657. doi: 10.1111/j.1469-8137.2007.02008.x.
28. Leprieur F, Albouy C, De Bortoli J, Cowman PF, Bellwood DR, Mouillot D. Quantifying phylogenetic beta diversity: distinguishing between 'true' turnover of lineages and phylogenetic diversity gradients. *PLoS One*. 2012; 7(8): e42760. doi: 10.1371/journal.pone.0042760.
29. Li D, Trotta L, Marx HE, Allen JM, Sun M, Soltis DE, Soltis PS, Guralnick RP, Baiser B. For common community phylogenetic analyses, go ahead and use synthesis phylogenies. *Ecology*. 2019 Sep; 100(9): e02788. doi: 10.1002/ecy.2788.
30. Lim SJ, Bordenstein SR. An introduction to phylosymbiosis. *Proc Biol Sci*. 2020 Mar 11; 287(1922): 20192900. doi: 10.1098/rspb.2019.2900.
31. Lin L, Tembrock LR, Wang L. Prevalence and underlying mechanisms of phylosymbiosis in land plants. *Journal of Plant Ecology*. Jun 2024; 17(6). doi: 10.1093/jpe/rtae051.
32. Liu J, Zhao J, Wang G, Chen J. Host identity and phylogeny shape the foliar endophytic fungal assemblages of *Ficus*. *Ecol Evol*. 2019 Aug 13; 9(18): 10472-10482. doi: 10.1002/ece3.5568.
33. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol*. 2005 Dec; 71(12): 8228-35. doi: 10.1128/AEM.71.12.8228-8235.2005.

34. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. *ISME J.* 2011 Feb; 5(2): 169-72. doi: 10.1038/ismej.2010.133.

35. McMurdie P, and Holmes S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE.* 2013; 8(4): e61217. <https://doi.org/10.1371/journal.pone.0061217>.

36. Naranjo-Ortiz MA, Gabaldón T. Fungal evolution: diversity, taxonomy and phylogeny of the Fungi. *Biol Rev Camb Philos Soc.* 2019 Dec; 94(6):2101-2137. doi: 10.1111/brv.12550.

37. Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Solymos P, Stevens M, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, Evangelista H, FitzJohn R, Friendly M, Furneaux B, Hannigan G, Hill M, Lahti L, McGlinn D, Ouellette M, Ribeiro Cunha E, Smith T, Stier A, Ter Braak C, Weedon J (2022). *_vegan: Community Ecology Package_*. R package version 2.6-4, <<https://CRAN.R-project.org/package=vegan>>.

38. Omacini M, Chaneton EJ, Ghera CM, Müller CB. Symbiotic fungal endophytes control insect host-parasite interaction webs. *Nature.* 2001 Jan 4; 409(6816): 78-81. doi: 10.1038/35051070.

39. Omer R, Mohsin MZ, Mohsin A, Mushtaq BS, Huang X, Guo M, Zhuang Y, Huang J. Engineered Bacteria-Based Living Materials for Biotherapeutic Applications. *Front Bioeng Biotechnol.* 2022 Apr 28; 10: 870675. doi: 10.3389/fbioe.2022.870675.

40. Paradis E, Schliep K. ape 5.7.1: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics.* 2019; 35:526-528. doi:10.1093/bioinformatics/bty633.

41. R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

42. Raina JB, Eme L, Pollock FJ, Spang A, Archibald JM, Williams TA. Symbiosis in the microbial world: from ecology to genome evolution. *Biol Open*. 2018 Feb 22; 7(2): bio032524. doi: 10.1242/bio.032524.
43. Rodriguez RJ, White JF Jr, Arnold AE, Redman RS. Fungal endophytes: diversity and functional roles. *New Phytol*. 2009; 182(2): 314-330. doi: 10.1111/j.1469-8137.2009.02773.x.
44. Saunders M, Kohn LM. Evidence for alteration of fungal endophyte community assembly by host defense compounds. *New Phytol*. 2009; 182(1): 229-238. doi: 10.1111/j.1469-8137.2008.02746.x.
45. Saunders M, Glenn AE, Kohn LM. Exploring the evolutionary ecology of fungal endophytes in agricultural systems: using functional traits to reveal mechanisms in community processes. *Evol Appl*. 2010 Sep; 3(5-6): 525-37. doi: 10.1111/j.1752-4571.2010.00141.x.
46. Sarver J, Schultz E, Apigo A, Gernandt D, Salas-Lizana R, Oono R. Deep sequencing across multiple host species tests pine-endophyte specificity. *American Journal of Botany*. 25 October 2021; 109(1):83-98. <https://doi.org/10.1002/ajb2.1792>.
47. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W; Fungal Barcoding Consortium; Fungal Barcoding Consortium Author List. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc Natl Acad Sci U S A*. 2012 Apr 17; 109(16): 6241-6. doi: 10.1073/pnas.1117018109.
48. Toju H, Tanabe AS, Yamamoto S, Sato H. High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. *PLoS One*. 2012; 7(7): e40863. doi: 10.1371/journal.pone.0040863.

49. Vandenkoornhuyse P, Quaiser A, Duhamel M, Le Van A, Dufresne A. The importance of the microbiome of the plant holobiont. *New Phytol.* 2015 Jun; 206(4): 1196-206. doi: 10.1111/nph.13312.
50. Vincent JB, Weiblen GD, May G. Host associations and beta diversity of fungal endophyte communities in New Guinea rainforest trees. *Mol Ecol.* 2016 Feb; 25(3): 825-41. doi: 10.1111/mec.13510.
51. Weatherhead E, Davis EL, Koide RT. Many foliar endophytic fungi of *Quercus gambelii* are capable of psychrotolerant saprotrophic growth. *PLoS One.* 2022 Oct 12; 17(10): e0275845. doi: 10.1371/journal.pone.0275845.
52. Webb CO, Ackerly DD, McPeck MA, Donoghue MJ. Phylogenies and Community Ecology. *Annual Review of Ecology, Evolution, and Systematics.* 2002 Aug 14; 33: 475-505. <https://doi.org/10.1146/annurev.ecolsys.33.010802.150448>.
53. Webb CO, Ackerly DD, Kembel SW. Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics.* 2008 Sep 15; 24(18): 2098-100. doi: 10.1093/bioinformatics/btn358.
54. Wehner J, Powell JR, Muller LAH, Caruso T, Veresoglou SD, Hempel S, Rillig MC. Determinants of root-associated fungal communities within Asteraceae in a semi-arid grassland. *Journal of Ecology.* 2013 December 11; 102(2): 425-436. <https://doi.org/10.1111/1365-2745.12197>.
55. Wickham H. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. 2016.
56. Wilson D. Endophyte: The Evolution of a Term, and Clarification of Its Use and Definition. *Oikos.* 1995 Jun; 73(2), 274–276. <https://doi.org/10.2307/3545919>.
57. Whitaker BK, Christian N, Chai Q, Clay K. Foliar fungal endophyte community structure is independent of phylogenetic relatedness in an Asteraceae common garden. *Ecology and Evolution.* 2020 Nov 26; 10(24): 13895-13912. <https://doi.org/10.1002/ece3.6983>.

58. Yeoh Y, Dennis PG, Paungfoo-Lonhienne C, Weber L, Brackin R, Ragan MA, Schmidt S, Hugenholtz P. Evolutionary conservation of a core root microbiome across plant phyla along a tropical soil chronosequence. *Natural Communications*. 2017 Aug 9; 8(215). <https://doi.org/10.1038/s41467-017-00262-8>
59. Zhang T, Yao YF. Endophytic Fungal Communities Associated with Vascular Plants in the High Arctic Zone Are Highly Diverse and Host-Plant Specific. *PLoS One*. 2015 Jun 12; 10(6): e0130051. doi: 10.1371/journal.pone.0130051.

Appendix

Refer to Supplemental Material for Document 1 and Tables 1 & 2.