Dimensions of Host Specificity in Foliar Fungal Endophytes

Austen Apigo and Ryoko Oono

Abstract Foliar fungal endophytes (FFE) 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 FFE 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 FFE communities partitioned by taxonomy (Ascomycota *vs*. Basidiomycota). We focus on the effects of rare FFE 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 FFE OTUs consistently changed ecological inference by decreasing host specificity averages and increasing variances within FFE phyla. To evaluate the degree to which rare FFE OTUs affect statistical power, we compared our empirical community to that of randomized communities. Excluding rare FFE 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 FFE OTU removal should be done with explicit rationale. We propose conceptualizing FFE host specificity with a multidimensional framework that will allow

A. Apigo (✉) · R. Oono
Department of Ecology, Evolution and Marine Biology,
University of California Santa Barbara, Santa Barbara, CA, USA
e-mail: aapigo@ucsb.edu

A. M. Pirttilä and A. C. Frank (eds.), *Endophytes of Forest Trees*,
Forestry Sciences 86, https://doi.org/10.1007/978-3-319-89833-9_2

future studies to use quantitative, comparable and theory-driven metrics that can scale towards more meaningful estimates of global fungal biodiversity.

Abbreviations

FFE Foliar fungal endophytes
NGS Next-generation sequencing
OTU Operational taxonomic unit

1 Introduction

Foliar fungal endophytes (FFE; 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 FFE (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 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 FFE, we propose defining host specificity as the state of occupying a restricted proportion of hosts in a plant community.

Surveys of FFE biodiversity frequently investigate various aspects of FFE host specificity (e.g., FFE 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 FFE studies are also challenging because host specificity can be evaluated in different ways depending on the scale of the study. The majority of FFE studies utilize multivariate ordination approaches to qualitatively compare FFE 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 a FFE occupies). To transition future studies to more quantitative and comparable approaches, we propose a common framework of four 'dimensions' of host specificity (adapted and modified from Poulin et al. 2011) that address different sampling, spatial and phylogenetic scales for which FFE are studied in a plant community:

(1) Structural specificity (FFE 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-FFE interactions by accounting for all potential hosts a FFE 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 (Fig. 1; 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 FFE displays consistent host specificity across a range of contexts.

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

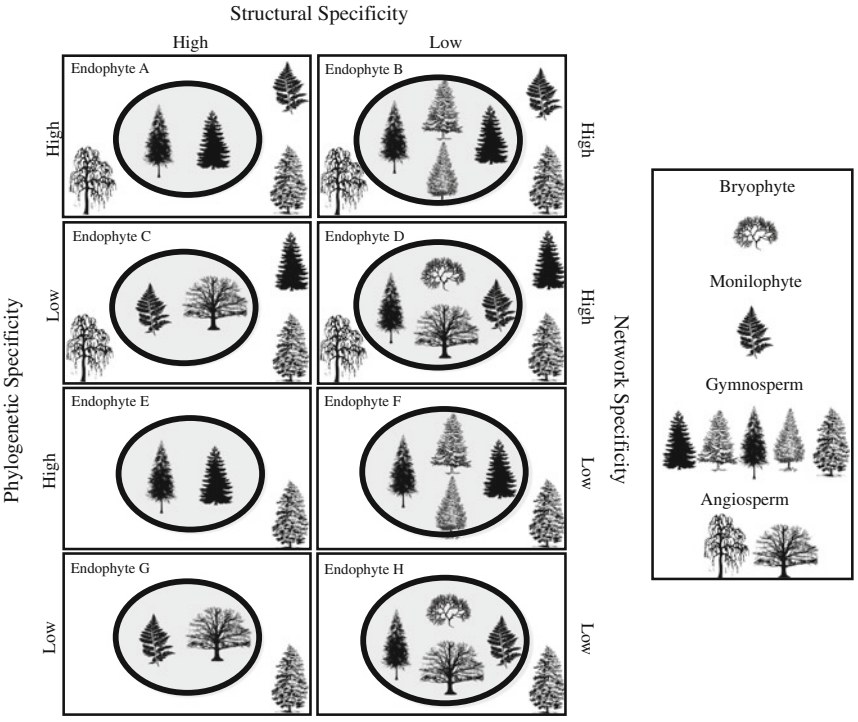

Fig. 1 Conceptual diagram of structural, network and phylogenetic specificity. Plant hosts within the grey circles represent the realized ecological niche of each FFE. FFE 'A' has high structural specificity (two occupied hosts), phylogenetic specificity (restricted to gymnosperm hosts) and network specificity (three unoccupied hosts). FFE '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

2 Structural Specificity

Structural specificity of a given FFE describes the number of occupied hosts, also referred to as 'basic specificity' (Poulin et al. 2011), or its differential abundance among them. Host richness (Eq. 1) quantifies all hosts a FFE occupies, regardless of FFE abundance in each host, and has a presence-absence data structure. In Eq. 1, r is the number of plant hosts occupied by one FFE species. A given FFE will have higher structural specificity if it occupies fewer plant hosts (Fig. 1). Shannon's H diversity index (Shannon and Weaver 1948), a commonly used abundance-weighted metric in ecological studies, quantifies host richness and FFE evenness among occupied hosts. In Eq. 2, r is the number of plant hosts occupied

by one FFE species and p_i is the relative read abundance of that FFE species in the ith host. For abundance-weighted Shannon's H, a FFE will have higher structural specificity if its abundance is unevenly distributed relative to another FFE, 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 (Figs. 1, 2, 3, 4, 5 and 6).

$$Host\ richness\ (presence - absence) = r \tag{1}$$

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

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 FFE 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.

3 Network Specificity

Network specificity quantifies the number and 'strength' of interactions by accounting for all potential hosts a given FFE could occupy (Fig. 1). The Resource Range Index (RRI; Schoener 1989; Eq. 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 Eq. 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; Eq. 4) compares the relative link strengths (based on FFE 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 Eq. 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 FFE 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 FFE) 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 FFE) 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 FFE 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 FFE 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 FFE could reveal specific fungal or host taxa that predominately contribute to observed patterns among communities or across studies.

4 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;

Eq. 5). Among all hosts occupied by a FFE and can be used to compare host specificity across different phylogenetic scales (Fig. 1). For presence-absence MPD, a given FFE will have higher phylogenetic specificity if it occupies hosts that are more closely related to one another. This metric can also be abundance-weighted ('structural phylogenetic specificity' in Poulin et al. 2011). For abundance-weighted MPD, a given FFE will have higher phylogenetic specificity if its abundance is unevenly distributed relative to another FFE, given that MPD is equal between them. In Eqs. 5 and 6, r is host richness, $_{i,j}$ are the phylogenetic distances between hosts i and j. For abundance-weighted phylogenetic specificity, f is the FFE 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 (Figs. 1, 2, 3, 4, 5 and 6).

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

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

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 FFE 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 FFE community composition is distributed relative to the phylogenetic distance, environmental dissimilarity and geographic distance among hosts. For example, if FFE community assembly is strongly influenced by host evolutionary relationships, a FFE cluster dendrogram or other metric for compositional similarity (e.g., non-metric multidimensional scaling; NMDS) will show topological congruence in FFE community composition as a function of host phylogenetic distance (i.e., phylosymbiosis; Brooks et al. 2016). Alternatively, if FFE community assembly is strongly influenced by environmental filtering, FFE 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 FFE 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 FFEs. 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., FFE among plant congeners), geography may be a stronger predictor of FFE 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 FFE 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 FFE 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 FFE population dynamics as FFE generalists would not be constrained by similar adaptive barriers (Poisot et al. 2011).

Host population genetic effects have not been a common focus of FFE 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 FFE community composition within the same plant hosts in different environments or different studies, a common framework for host specificity is needed.

5 Biological and Methodological Considerations: A Comparative Case Study for Structural, Network and Phylogenetic Specificity

FFE are horizontally transmitted through the environment, as opposed to vertical transmission from parent to progeny in graminoid-FFE symbioses (Class 1 *vs*. Class 3 endophytes; sensu Rodriguez et al. 2009). The degree to which plants can preferentially admit specific FFE 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 FFE communities. These 'tourist' FFE 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 FFE community taxonomically partitioned into two groups (Ascomycota *vs*. Basidiomycota). The FFE 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 FFE 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 FFE 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, FFE 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 Basidiomycota FFE may lead to predictable differences in their host specificity. For example, FFE 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 FFE 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 FFE.

Our case community showed similar structure to other FFE communities where Basidiomycete FFE OTUs occurred less frequently compared to Ascomycete FFE 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 FFE community consistently displayed higher structural, network and phylogenetic specificities across presence-absence and abundance-weighted metrics, relative to the complete Ascomycete FFE community (host specificity for 0% removed values in Figs. 2 and 3). This suggests that, on average, Basidiomycete FFE OTUs are more asymmetrically distributed (Fig. 3a–d) across fewer hosts (Fig. 2a–d) that comprise a narrower phylogenetic breadth (Figs. 2e, f and 3e, f) than Ascomycete FFEs. Even though Basidiomycete FFEs tend to be rarer, these OTUs are not necessarily more

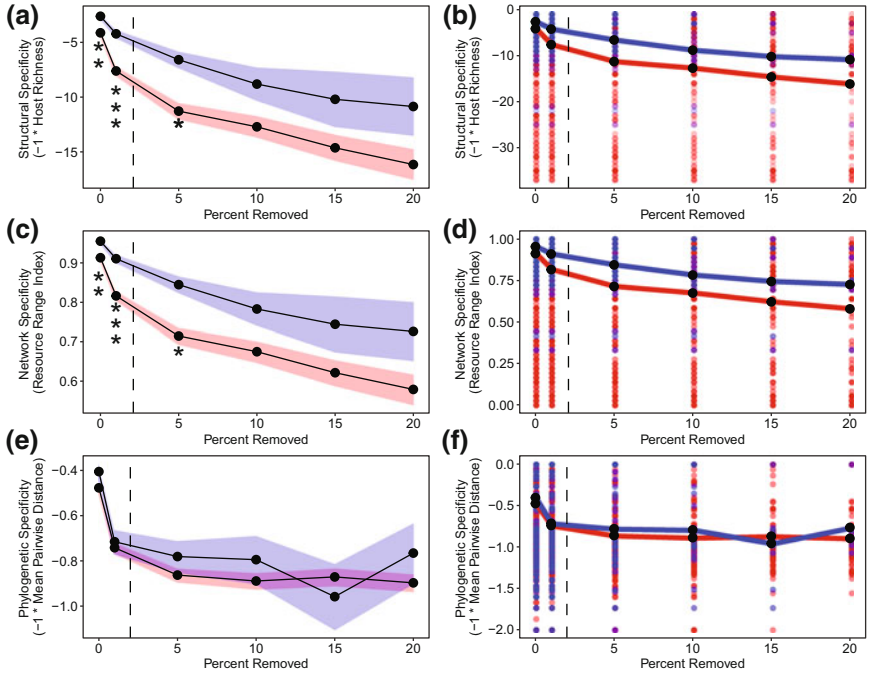

Fig. 2 Structural (**a**, **b**), network (**c**, **d**) and phylogenetic specificity (**e**, **f**) measured as a function of rare OTU removal with presence-absence data. **a**, **b** Host Richness (HR) quantifies the number of hosts per FFE (Eq. 1). **c**, **d** The Resource Range Index (RRI) quantifies the host niche per FFE and is scaled from 0 (perfect generalist) to 1 (perfect specialist; Eq. 3). **e**, **f** The Mean Pairwise Phylogenetic Distance (MPD) quantifies the mean phylogenetic distance of occupied hosts per FFE (Eq. 5). OTUs (structural and network specificity: n = 1388; phylogenetic specificity: n = 1342) 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 FFE OTUs using a Wilcoxon rank sum test

likely to be 'accidental tourists' than Ascomycete FFEs since they occupy a relatively narrower host phylogenetic breadth (Fig. 3e, f).

For all 'dimensions' of host specificity, sequence exclusion from rare FFE 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 (Figs. 2a–f and 3a–f). However, the statistical differences between these groups cannot define an exclusion threshold for rare FFE

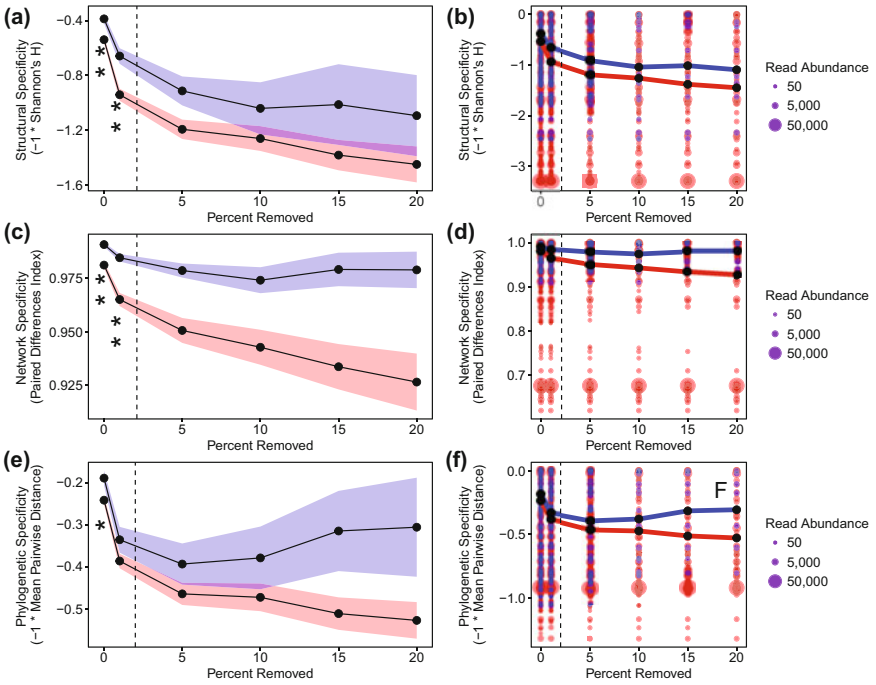

Fig. 3 Structural (**a**, **b**), network (**c**, **d**) and phylogenetic specificity (**e**, **f**) measured as a function of rare OTU removal with abundance-weighted data. **a**, **b** Shannon's *H* (SH) quantifies host richness and FFE evenness within hosts per FFE (Eq. 2). **c**, **d** The Paired Differences Index (PDI) quantifies differential abundance among all hosts in the community and is scaled from 0 (perfect generalist) to 1 (perfect specialist; Eq. 4). **e**, **f** The abundance-weighted variant of the Mean Pairwise Phylogenetic Distance (MPD) quantifies the mean phylogenetic distance of occupied hosts per FFE with greater weight given to hosts where a given FFE is more abundant (Eq. 6). OTUs (structural and network specificity: n = 1388; phylogenetic specificity: n = 1342) 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 FFE OTUs using a Wilcoxon rank sum test

because the proportion of rare taxa that are biologically informative is unknown. For example, in our case study (Figs. 2 and 3), 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 FFE, 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 Sect. 7).

Each 'dimension' of host specificity is useful for understanding different ecological or evolutionary aspects of plant-FFE assemblages. In this case study, we provided examples of structural, network and phylogenetic specificity from the FFE perspective. Alternatively, from the plant perspective, host specificity can vary depending on the composition of the FFE community among plant species or functional groups. For example, FFE 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.

6 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; Figs. 4 and 5).

Structural and network specificity were perfectly (Host Richness, HR *vs*. Resource Range Index, RRI; $\rho = 1$; Fig. 4a) or highly correlated in rank (Shannon's *H*, SH *vs*. Paired Difference Index, PDI; $\rho = 0.99$; Fig. 5a) 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 FFE 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 (Figs. 4b *vs*. 5b and 4d *vs*. 5d).

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 (Fig. 4b, d *vs*. f). The variance in presence-absence MPD may be the least explained by read abundance because rare FFE 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 (Fig. 5b, f *vs.* d), which could be due to the properties of PDI (Eq. 4). Although host specificity and read abundance were more negatively correlated for SH and MPD (slope of regression lines; Fig. 5b, f), 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 (Figs. 4c, e and 5c, e) suggest that FFE OTUs that occupy a broader host phylogenetic breadth tended to be more evenly distributed across more hosts, whereas FFE 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 a FFE 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 (Fig. 4e).

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 FFE 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 ses.mpd{picante} in R for more details; Kembel et al. 2010). However, technical constraints of a randomization method can preclude its use. For presence-absence data, if any FFE 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 (Eq. 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.

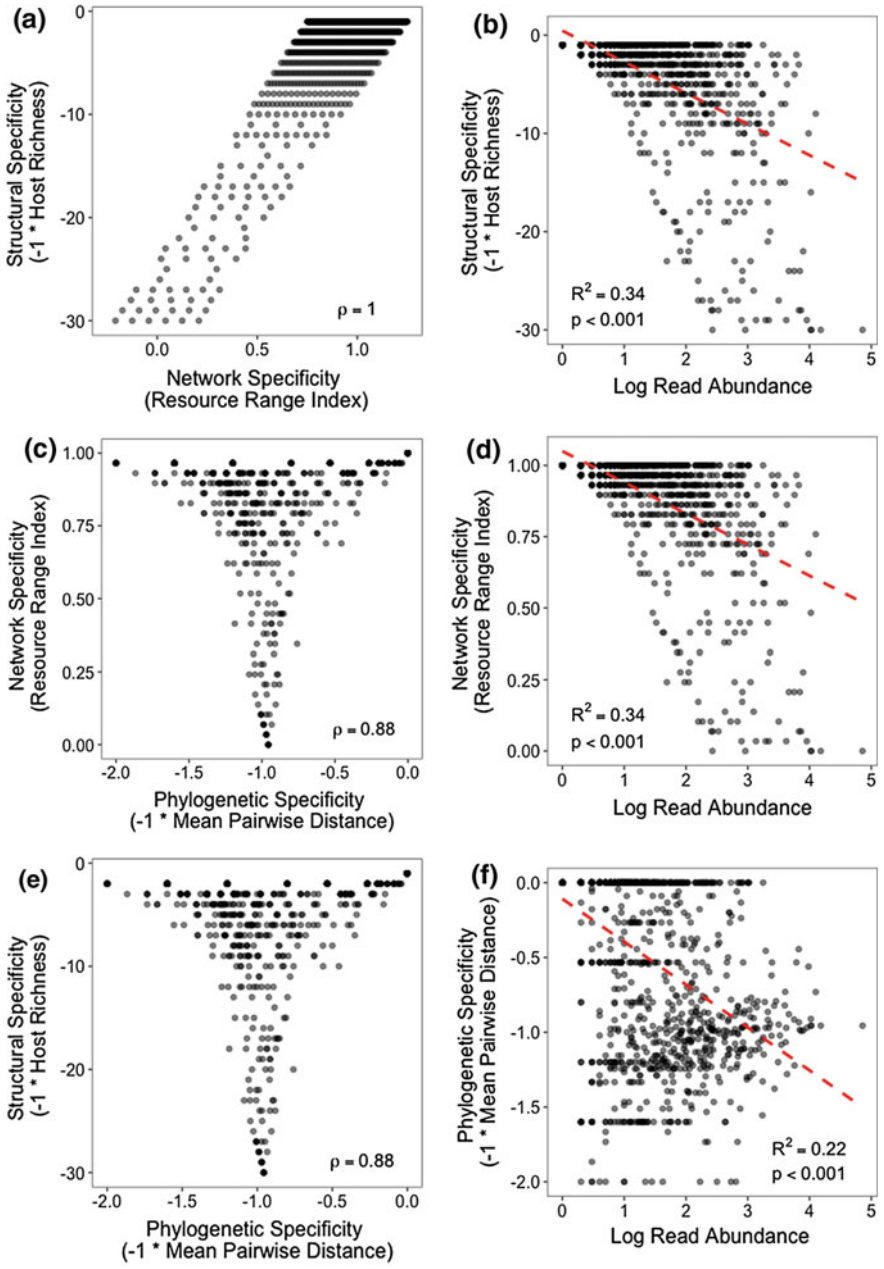

Fig. 4 Correlations among alpha-host specificities with presence-absence data (**a**, **c** and **e**) and between host specificities and read abundances per OTU (**b**, **d** and **f**). Panels **a**, **c** and **e** compare the three alpha-host specificities per FFE OTU (n = 1360) and report their rank correlation coefficients, Spearman's rho (ρ). Panels **b**, **d** and **f** compare host specificity values to read abundance (log-scaled) per FFE OTU (n = 1360) and report the coefficients of determination and p-values. The points in panel **a** are jittered for clarity

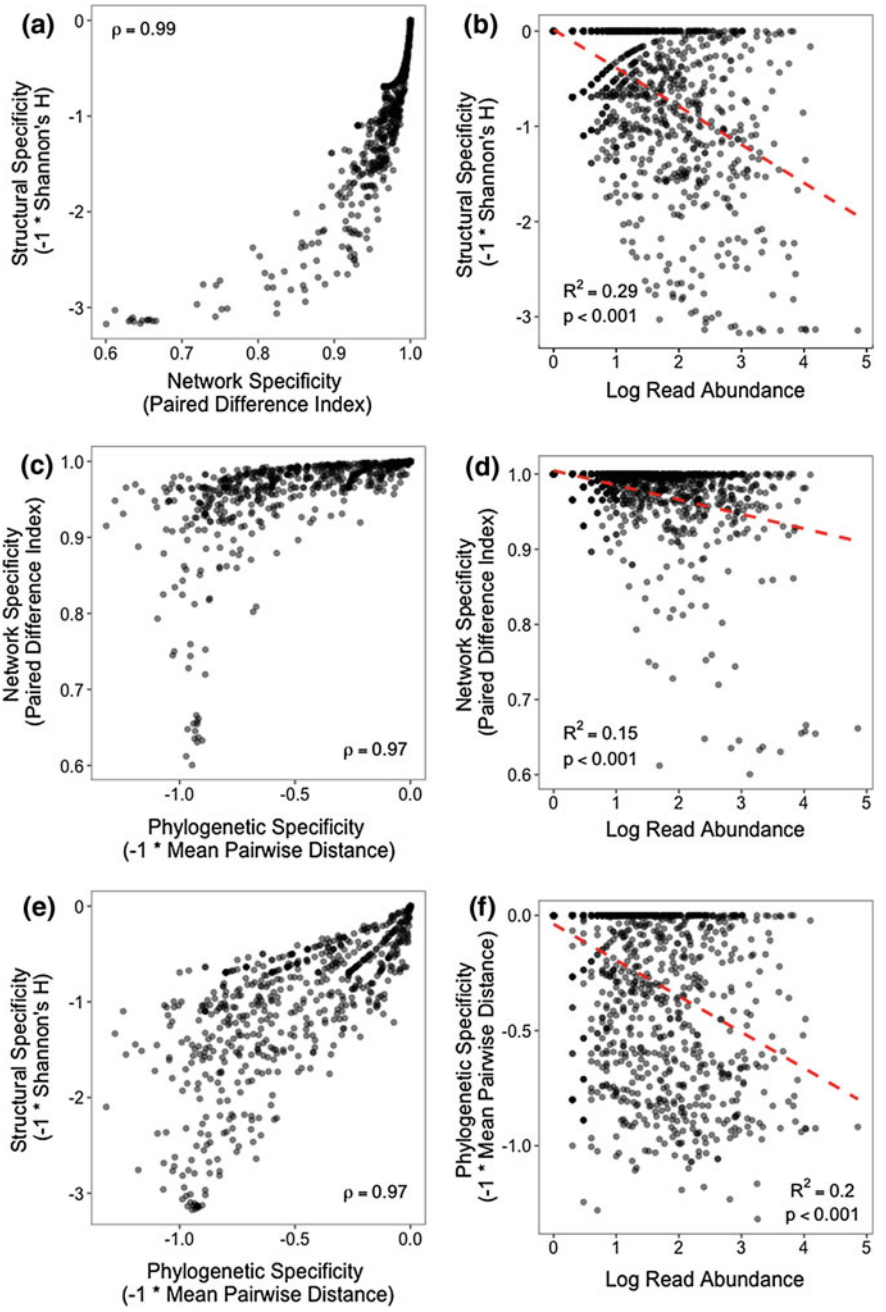

Fig. 5 Correlations among alpha-host specificities with abundance-weighted data (**a**, **c** and **e**) and between host specificities and read abundances per OTU (**b**, **d** and **f**). Panels **a**, **c** and **e** compare the three alpha-host specificities per FFE OTU (n = 1360) and report their rank correlation coefficients, Spearman's rho (ρ). Panels **b**, **d** and **f** compare the host specificity values to read abundance (log-scaled) per FFE OTU (n = 1360) and report the coefficients of determination and *p*-values

7 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 methodology). We compared the host specificity values per FFE OTU for the full dataset of the case community (i.e., not partitioned by fungal phyla) to the averaged host specificity per OTU across 1000 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 9999 bootstrap replicates. We repeated the randomization and bootstrap tests as we incrementally excluded rare OTUs, as in previous analyses.

Excluding rare FFE OTUs (e.g., >10% of total sequences; Fig. 6a–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 (Fig. 6b, f), 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 (Fig. 6e *vs.* f). The empirical distributions for presence-absence structural and network specificity completely overlapped with randomized distributions (Fig. 6a, c) 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 (Fig. 6a, c). 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 Fig. 6, the removal of reads with less than 0.01% relative abundance, which is a commonly utilized threshold, removed 70.7% of OTUs (994 of 1406) for structural and network specificity and 68.5% of OTUs (932 of 1360) 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

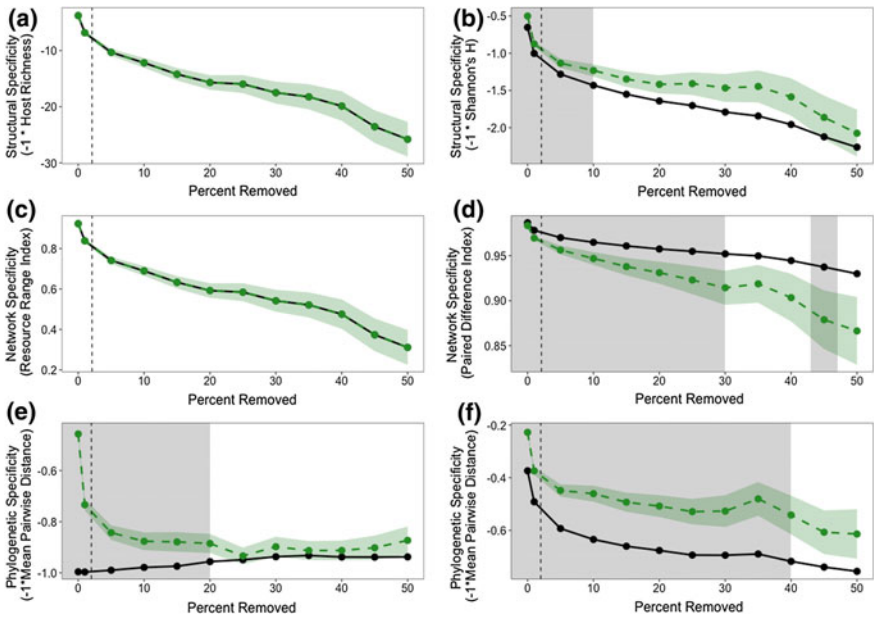

Fig. 6 Structural (**a**, **b**), network (**c**, **d**) and phylogenetic specificity (**e**, **f**) measured as a function of rare OTU removal with presence-absence (**a**, **c** and **e**) and abundance-weighted (**b**, **d** and **f**) data for empirical and randomized communities. OTUs (structural and network specificity: n = 1406; phylogenetic specificity: n = 1360) 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 1000 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 9999 bootstrap replicates

distinction from a randomized community (white areas; $p > 0.05$; Fig. 6). However, comparisons with randomized communities highlight how statistical power can vary widely depending on data-structure and the host specificity metric. 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.

8 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 FFE 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 FFE would display higher beta-specificity if it was found in the same plant host at different localities relative to another FFE that occupied completely different plant hosts at different localities (Fig. 7). 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 FFE community turnover across the geographic range or developmental stages of a given host could reveal important features of FFE 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 (Fig. 7), but this metric could be applied across other contexts, such as sites that vary in plant community structure or multiple stages during plant development.

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 a FFE that occupies the same plant hosts in all considered localities

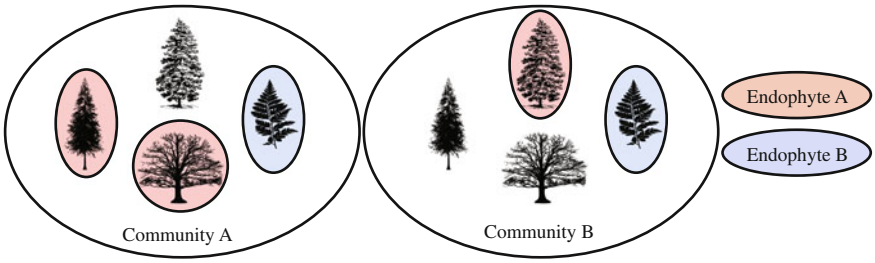

Fig. 7 Conceptual diagram of beta-specificity. The realized ecological niche of each FFE within two replicate plant communities is represented by colored circles around a plant host. Endophyte 'A' occupies completely different plant hosts in the two communities, such that beta-specificity equals zero. 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

(high beta-specificity) and 0 indicates a FFE 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 FFE across all T localities and r is the total number of plant hosts occupied at locality i (Diserud and Odegaard 2007).

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

9 Conclusion

Any 'dimension' of FFE 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 (Figs. 2 and 3) and statistical power (Fig. 6).

Structural specificity metrics are useful within individual studies with uniform sampling designs. However, FFE with similar structural specificities in different localities could have significantly different network specificity depending on plant community diversity (Fig. 1). Network specificity metrics are more comparable across studies because they quantify how the distribution of a FFE species varies among all sampled hosts in a plant community. FFE could also occupy similar proportions of the plant community, but differ in the phylogenetic composition of hosts they occupy. Phylogenetic specificity can quantify the distribution of a FFE species relative to the evolutionary relationships among occupied hosts. Finally, the biogeographical approach of beta-specificity has important implications for global fungal biodiversity measurements.

FFE have been suggested as an indicator group for global fungal biodiversity due to their prevalence (Suryanarayanan 2011). Estimates of FFE 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 FFE host specificity can provide more accurate fungal biodiversity estimates through the use of correction factors. For example, high FFE host specificity has been suggested to characterize boreal forest biomes (Arnold and Lutzoni 2007). This would produce high FFE beta-diversity among plant host species. Higher heterogeneity in FFE 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; Schleuning et al. 2012). Low host specificity would produce lower beta-diversity in FFE 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.

10 Methodology

10.1 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).

10.2 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 Figs. 2 and 3 (structural and network specificity: n = 1388; phylogenetic specificity: n = 1342), only OTUs shared among the structural, network and phylogenetic specificity datasets for Figs. 4 and 5 (n = 1360) and all OTUs for Fig. 6 [structural and network specificity: total OTUs (n = 1406), Ascomycota (n = 1087), Basidiomycota (n = 301), Chytridiomycota (n = 6), Mucoromycota (n = 6), Neocallimastigomycota (n = 5) and Zoopagomycota (n = 1); phylogenetic specificity: total OTUs (n = 1360), Ascomycota (n = 1053), Basidiomycota (n = 289), Chytridiomycota (n = 6), Mucoromycota (n = 6), Neocallimastigomycota (n = 5) and Zoopagomycota (n = 1)].

10.3 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; see Appendix D). 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 (1000 permutations) with default parameters.

10.4 Sequence Removal

Truncated data sets were generated with Multivariate Cutoff Level Analysis (MultiCoLA; Gobet et al. 2010). OTUs (structural and network specificity: n = 1406; phylogenetic specificity: n = 1360) 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 4948 of 492,914 reads for structural and network specificity and 4010 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 1406 OTUs; 70.7% OTUs removed) for structural and network specificity and 8000 of 399,660 reads or 2.0% sequence removal (932 of 1360 OTUs; 68.5% OTUs removed) for phylogenetic specificity. For Figs. 2 and 3, only Ascomycete (structural and network specificity: n = 1087; phylogenetic specificity: n = 1053) and Basidiomycete (structural and network specificity: n = 301; phylogenetic specificity: n = 289) OTUs were included in the analysis. For Fig. 6, all OTUs (structural and network specificity: n = 1406; phylogenetic specificity: n = 1360) were included in the analysis.

10.5 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.

10.6 Rank Correlations Between Structural, Network and Phylogenetic Specificity

For Figs. 4 and 5, we used 1360 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).

10.7 Comparisons to a Random Community Assembly Model

Using the above functions, we generated 1000 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 FFE and not as one total community metric (unlike for network-level metrics), we calculated host specificity per FFE OTU across each of the 1000 randomized tables and then averaged the 1000 host specificity measurements per FFE OTU. We then compared the randomized and empirical communities (n = 1406 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 9999 bootstrap replicates. Resampling occurred with replacement and resampling size was consistent with the size of the original population (see Appendix E).

Acknowledgements We thank Cindy Quach and Kevin Feller for their assistance in the field at the University of California, James San Jacinto Reserve. We thank Kiana Lee, Vanessa Greenman, Angelina Hyunh and Theodore Kwan for their assistance in the lab. We thank Professor Holly Moeller for reviewing previous drafts and helpful advice that improved the quality of this chapter.

Appendices

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

Appendix A

Bioinformatic pipeline using USEARCH (version 9.2.64; Edgar 2010), BLAST+ (Camacho et al. 2009), MEGAN (Huson et al. 2007) and QIIME (Caporaso et al. 2010).

Appendix B

OTU tables used for structural, network and phylogenetic specificity with fungal taxonomic annotations.

Appendix C

R code for (1) sequence removal using Multivariate Cutoff Level Analysis (MultiCoLA; Gobet et al. 2010), (2) structural (vegan; Oksanen et al. 2017), network (bipartite; Dormann et al. 2008; Dormann 2011) and phylogenetic (picante; Kembel et al. 2010) specificity, and (3) randomized community analysis using the 'nullmodel' function (bipartite; Dormann et al. 2008) for structural and network specificity and 'ses.mpd' function (picante; Kembel et al. 2010) for phylogenetic specificity.

Appendix D

Host species phylogenetic tree used to calculate phylogenetic specificity in Newick format.

Appendix E

Table of *p*-values for Figs. 2, 3 and 6.

References

Ahlholm JU, Helander M, Henriksson J, Metzler M, Saikkonen K (2002) Environmental conditions and host genotype direct genetic diversity of *Venturia ditricha*, a fungal endophyte of birch trees. Evolution 56:1566–1573

Arnold AE (2007) Understanding the diversity of foliar endophytic fungi: progress, challenges, and frontiers. Fungal Biol Rev 21:51–66

Arnold AE, Lutzoni F (2007) Diversity and host range of foliar fungal endophytes: are tropical leaves biodiversity hotspots? Ecology 88:541–549

Arnold AE, Maynard Z, Gilbert GS, Coley PD, Kursar TA (2000) Are tropical fungal endophytes hyperdiverse? Ecol Lett 3:267–274

Bacon CW, White J (2000) Microbial endophytes. CRC Press, Boca Raton

Bascompte J (2010) Structure and dynamics of ecological networks. Science 329:765–766

Blüthgen N, Menzel F, Blüthgen N (2006) Measuring specialization in species interaction networks. BMC Ecol 6:9

Branco S, Gladieux P, Ellison CE, Kuo A, LaButti K, Lipzen A et al (2015) Genetic isolation between two recently diverged populations of a symbiotic fungus. Mol Ecol 24:2747–2758

Brooks AW, Kohl KD, Brucker RM, van Opstal EJ, Bordenstein SR (2016) Phylosymbiosis: relationships and functional effects of microbial communities across host evolutionary history. PLoS Biol 14:e2000225

Brown SP, Veach AM, Rigdon-Huss AR, Grond K, Lickteig SK, Lothamer K et al (2015) Scraping the bottom of the barrel: are rare high throughput sequences artifacts? Fungal Ecol 13:221–225

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinform 10:421

Cannon PF, Simmons CM (2002) Diversity and host preference of leaf endophytic fungi in the Iwokrama Forest Reserve, Guyana. Mycologia 94:210–220

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK et al (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336

Carroll GC, Carroll FE (1978) Studies on the incidence of coniferous needle endophytes in the Pacific Northwest. Can J Bot 56:3034–3043

Cavender-Bares J, Kozak KH, Fine PVA, Kembel SW (2009) The merging of community ecology and phylogenetic biology. Ecol Lett 12:693–715

Chagnon P-L, U'Ren JM, Miadlikowska J, Lutzoni F, Arnold AE (2016) Interaction type influences ecological network structure more than local abiotic conditions: evidence from endophytic and endolichenic fungi at a continental scale. Oecologia 180:181–191

Davey ML, Heimdal R, Ohlson M, Kauserud H (2013) Host- and tissue-specificity of moss-associated Galerina and Mycena determined from amplicon pyrosequencing data. Fungal Ecol 6:179–186

Davis EC, Shaw AJ (2008) Biogeographic and phylogenetic patterns in diversity of liverwort-associated endophytes. Am J Bot 95:914–924

Devictor V, Clavel J, Julliard R, Lavergne S, Mouillot D, Thuiller W et al (2010) Defining and measuring ecological specialization. J Appl Ecol 47:15–25

Diserud OH, Odegaard F (2007) A multiple-site similarity measure. Biol Lett 3:20–22

Dormann CF (2011) How to be a specialist? Quantifying specialisation in pollination networks. Netw Biol 1:1–20

Dormann CF, Gruber B, Fruend J (2008) Introducing the bipartite package: analysing ecological networks. R News 8:8–11

Dormann CF, Fründ J, Blüthgen N, Gruber B (2009) Indices, graphs and null models: analyzing bipartite ecological networks. Open Ecol J 2:7–24

Dreyfuss MM, Chapela IH (1994) Potential of fungi in the discovery of novel, low-molecular weight pharmaceuticals. Biotechnology 26:49–80

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461

Edgar RC (2016) UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing [Internet]. https://doi.org/10.1101/081257

Efron B (1987) Better bootstrap confidence intervals. J Am Stat Assoc 82(397):171–185

Faeth SH, Fagan W (2002) Fungal endophytes: common host plant symbionts but uncommon mutualists. Integr and Comp Biol 42:360–368

Fröhlich J, Hyde KD (1999) Biodiversity of palm fungi in the tropics: are global fungal diversity estimates realistic? Biodiv and Conserv 8:977–1004

Futuyma DJ, Moreno G (1988) The evolution of ecological specialization. Annu Rev Ecol Syst 19:207–233

Gilbert GS, Reynolds DR, Bethancourt A (2007) The patchiness of epifoliar fungi in tropical forests: host range, host abundance, and environment. Ecology 88:575–581
Gobet A, Quince C, Ramette A (2010) Multivariate cutoff level analysis (MultiCoLA) of large community data sets. Nucleic Acids Res 38:e155
Hawksworth DL (2001) The magnitude of fungal diversity: the 1.5 million species estimate revisited. Mycol Res 105:1422–1432
Higgins KL, Arnold AE, Miadlikowska J, Sarvate SD, Lutzoni F (2007) Phylogenetic relationships, host affinity, and geographic structure of boreal and arctic endophytes from three major plant lineages. Mol Phylogenet Evol 42:543–555
Hubbell SP (2001) The unified neutral theory of biodiversity and biogeography (MPB-32). Princeton University Press, Princeton
Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. Genome Res 17:377–386
Katoh K, Misawa K, Kuma K-I, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066
Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD et al (2010) Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26:1463–1464
Kerkhoff AJ, Moriarty PE, Weiser MD (2014) The latitudinal species richness gradient in New World woody angiosperms is consistent with the tropical conservatism hypothesis. Proc Natl Acad Sci USA 111:8125–8130
Kraft NJB, Adler PB, Godoy O, James EC, Fuller S, Levine JM (2014) Community assembly, coexistence and the environmental filtering metaphor. Funct Ecol 29:592–599
Krasnov BR, Mouillot D, Shenbrot GI, Khokhlova IS, Poulin R (2011) Beta-specificity: the turnover of host species in space and another way to measure host specificity. Int J Parasitol 41:33–41
Levin SA (1992) The problem of pattern and scale in ecology. Ecological Time Series
Liu L, Yang J, Yu Z, Wilkinson DM (2015) The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China. ISME J 9:2068–2077
McMurdie PJ, Holmes S (2014) Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comput Biol 10:e1003531
Minchin PR (1987) An evaluation of the relative robustness of techniques for ecological ordination. Vegetatio 69:89–107
Mittelbach GG, Schemske DW, Cornell HV, Allen AP, Brown JM, Bush MB et al (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. Ecol Lett 10:315–331
Moran VC, Southwood TRE (1982) The guild composition of arthropod communities in trees. J Animal Ecol 51:289–306
Nguyen NH, Smith D, Peay K, Kennedy P (2014) Parsing ecological signal from noise in next generation amplicon sequencing. New Phytol 205:1389–1393
Oksanen JF, Blanchet FG, Friendly M, Kindt R, Legendre P, et al (2017) Vegan: community ecology
Olesen JM, Bascompte J, Dupont YL, Jordano P (2007) The modularity of pollination networks. Proc Natl Acad Sci USA 104:19891–19896
Oono R, Lefèvre E, Simha A, Lutzoni F (2015) A comparison of the community diversity of foliar fungal endophytes between seedling and adult loblolly pines (Pinus taeda). Fungal Biol 119:917–928
Ortiz-García S, Gernandt DS, Stone JK, Johnston PR, Chapela IH, Salas-Lizana R et al (2003) Phylogenetics of Lophodermium from pine. Mycologia 95:846–859
Package. R package version 2.4-4. https://CRAN.R-project.org/package=vegan
Pandey AK, Reddy MS, Suryanarayanan TS (2003) ITS-RFLP and ITS sequence analysis of a foliar endophytic Phyllosticta from different tropical trees. Mycol Res 107:439–444
Patin NV, Kunin V, Lidström U, Ashby MN (2013) Effects of OTU clustering and PCR artifacts on microbial diversity estimates. Microb Ecol 65:709–719

Pearse WD, Purvis A (2013) phyloGenerator: an automated phylogeny generation tool for ecologists. Methods Ecol Evol 4:692–698
Poisot T, Bever JD, Nemri A, Thrall PH, Hochberg ME (2011) A conceptual framework for the evolution of ecological specialisation. Ecol Lett 14:841–851
Poisot T, Canard E, Mouquet N, Hochberg ME (2012) A comparative study of ecological specialization estimators. Methods Ecol Evol 3:537–544
Poulin R, Krasnov BR, Mouillot D (2011) Host specificity in phylogenetic and geographic space. Trends Parasitol 27:355–361
Ricotta C, Pavoine S (2015) A multiple-site dissimilarity measure for species presence/absence data and its relationship with nestedness and turnover. Ecol Indic 54:203–206
Rodriguez RJ, White JF Jr, Arnold AE, Redman RS (2009) Fungal endophytes: diversity and functional roles. New Phytol 182:314–330
Saavedra S, Stouffer DB, Uzzi B, Bascompte J (2011) Strong contributors to network persistence are the most vulnerable to extinction. Nature 478:233–235
Saikkonen K, Faeth SH, Helander M, Sullivan TJ (1998) Fungal endophytes: a continuum of interactions with host plants. Ann Rev Ecol Syst 29:319–343
Schleuning M, Fründ J, Klein A-M, Abrahamczyk S, Alarcón R, Albrecht M et al (2012) Specialization of mutualistic interaction networks decreases toward tropical latitudes. Curr Biol 22:1925–1931
Schoener TW (1989) Food webs from the small to the large: The Robert H. MacArthur award lecture. Ecology 70:1559–1589
Shannon CE, Weaver W (1948) The mathematical theory of communication. University of Illinois Press, Urbana
Sieber TN (2007) Endophytic fungi in forest trees: are they mutualists? Fungal Biol Rev 21:75–89
Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313
Stone J, Petrini O (1997) Endophytes of forest trees: a model for fungus-plant interactions. In Plant Relationships Part B. Springer, Berlin, Heidelberg
Sun X, Ding Q, Hyde KD, Guo LD (2012) Community structure and preference of endophytic fungi of three woody plants in a mixed forest. Fungal Ecol 5:624–632
Suryanarayanan TS (2011) Diversity of fungal endophytes in tropical trees. For Sci
Swenson NG (2014) Functional and phylogenetic ecology in R. Springer, New York, NY
Thébault E, Fontaine C (2010) Stability of ecological communities and the architecture of mutualistic and trophic networks. Science 329:853–856
Toju H, Tanabe AS, Yamamoto S, Sato H (2012) High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. PLoS ONE 7: e40863
U'Ren JM, Lutzoni F, Miadlikowska J, Laetsch AD, Arnold AE (2012) Host and geographic structure of endophytic and endolichenic fungi at a continental scale. Am J Bot 99:898–914
Vega FE (2008) Insect pathology and fungal endophytes. J Invertebr Pathol 98:277–279
Vega FE, Simpkins A, Catherine Aime M, Posada F, Peterson SW, Rehner SA et al (2010) Fungal endophyte diversity in coffee plants from Colombia, Hawai'i, Mexico and Puerto Rico. Fungal Ecol 3:122–138
Vincent JB, Weiblen GD, May G (2016) Host associations and beta diversity of fungal endophyte communities in New Guinea rainforest trees. Mol Ecol 25:825–841
Walker DM, Castlebury LA, Rossman AY, Struwe L (2013) Host conservatism or host specialization? Patterns of fungal diversification are influenced by host plant specificity in Ophiognomonia (Gnomoniaceae: Diaporthales). Biol J Linn Soc Lond 111:1–16
Walter GH (1991) What is resource partitioning? J Theor Biol 150:137–143
Wardhaugh CW, Edwards W, Stork NE (2015) The specialization and structure of antagonistic and mutualist networks of beetles on rainforest canopy trees. Biol J Linn Soc 114:287–295
Webb CO (2000) Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. Am Nat 156:145–155

Webb CO, Ackerly DD, McPeek MA, Donoghue MJ (2002) Phylogenies and community ecology. Ann Rev Ecol Syst 33:475–505

Webb CO, Ackerly DD, Kembel SW (2008) Phylocom: software for the analysis of phylogenetic community structure and trait evolution. Bioinformatics 24:2098–2100

Weiss NA (2016) wBoot: bootstrap methods. R package version 1.0.3. https://CRAN.R-project.org/package=wBoot

Whittaker RH (1972) Evolution and measurement of species diversity. Taxon 21:213

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