**TITLE:**

Fungal communities living within leaves of native Hawaiian plants are structured by landscape-scale variables as well as by host plants

**AUTHORS:**

Darcy JL, Cobian G, Swift S, Zahn G, Perry BA, Amend AS

**ABSTRACT:**

*AIM:*A phylogenetically diverse array of fungi live within healthy leaf tissue. Many studies have examined these endophytes within a single plant species and/or at small spatial scales, but landscape-scale variables that determine their community composition are not well understood, either across geographic space, across climatic conditions, or in the context of host plant phylogeny. Here, we evaluate the contributions of these variables to endophyte community composition using our survey of foliar fungal endophytes in native Hawaiian plants sampled across the Hawaiian archipelago.

*LOCATION:* Hawai’i.

*METHODS:* The Hawaiian archipelago offers a uniquely tractable system to study biogeography of foliar fungal endophytes, because the islands harbor a wide array of climatic conditions, but native plant species distributions are often orthogonal to climate and geography. Thus, we can disentangle the effects of host plant identity, geography, and of local climate on endophyte communities. We used Illumina technology to sequence fungal ITS1 amplicons in order to characterize foliar endophyte communities in the leaves of 896 plants across 5 islands and 80 host plant genera. Using Generalized Dissimilarity Modeling (GDM) we tested the effect of landscape-scale variables on observed differences in foliar endophyte communities. Bipartite network analysis was used to examine the extent to which each island harbored specialized or cosmopolitan foliar endophytes.

*RESULTS*: Foliar fungal endophyte communities in the Hawaiian archipelago are structured most strongly by evapotranspiration, elevation, vegetation/habitat type, and by the phylogeny of host plants. Whereas previous studies determined that rainfall is a significant predictor of FEF community composition this variable was not significant in our analysis. The five islands we sampled each harbored significantly specialized endophyte communities as well.

*MAIN CONCLUSIONS*: Factors that structure foliar endophyte communities at small geographic and narrow host phylogenetic scales are broadly generalizable to the larger scales we studied here, although not universally. Evapotranspiration, a variable with resolution 250 m2, was the most robust predictor of endophyte community dissimilarity in our study, although it had not previously been considered an important determinant of FEF communities, and rainfall may not be an important variable when larger spatial and host phylogenetic scales are considered.

**1. INTRODUCTION:**

Less than two out of every thousand fungal species thought to exist on Earth have been described (Blackwell, 2011). Of those species awaiting discovery, a large percentage are presumed to live cryptic lifestyles in association with plant or animal hosts (Hawksworth & Rossman, 1997; Blackwell, 2011). Foliar endophytic fungi (FEF), defined here as all fungi living within leaf tissue but not causing any outward signs of disease (*sensu* (Stone et al., 2000)) are effectively invisible and represent a “hotspot” of undescribed fungal diversity (Arnold & Lutzoni, 2007; Porras-Alfaro & Bayman, 2011). Most commonly, tropical FEF are nested throughout the non-lichen forming classes of the subphylum Pezizomycotina (Blackwell et al., 2006; Arnold et al., 2009; Rodriguez et al., 2009), but due to their cryptic lifestyles and high species richness, many questions remain about which factors determine how FEF are distributed throughout nature.

Although all available evidence suggests that most eudicot FEF are horizontally transmitted and not inherited via seed (Bayman et al., 1998), very little is known about which factors structure FEF community composition, and how those factors differ in their relative importance. The few studies that have examined FEF communities in this ecological and biogeographic context have noted several different drivers of FEF community composition and biogeography. Temperature (Zimmerman & Vitousek, 2012; Coince et al., 2014), geographic distance (U’Ren et al., 2012), elevation and rainfall (Zimmerman & Vitousek, 2012), vegetation density/urbanization (Jumpponen & Jones, 2009, 2010), and host plant specificity (Unterseher et al., 2012; Massimo et al., 2015) have all been identified as putatitvely important variables for FEF community composition. This diversity of results arises from individual studies examining a narrow range of hypotheses for the drivers of FEF community composition. For example, studies finding a significant effect of rainfall and elevation (Zimmerman & Vitousek, 2012) may focus on a narrow phylogenetic breadth of host plant species, and therefore not find that different host plants recruit unique FEF communities. Thus, only a significant effect of rainfall may be reported, even though host selection could be present as well.

In fact, the bulk of FEF research is represented by studies focusing on a specific host plant (Jumpponen & Jones, 2010; Zimmerman & Vitousek, 2012; Mejía et al., 2014; Saucedo-García et al., 2014; Oono et al., 2015; Polonio et al., 2015; Felber et al., 2016; González-Teuber, 2016; Kato et al., 2017). The few studies that have surveyed FEF across multiple plant species have found significant host effects (Unterseher et al., 2012; Kembel & Mueller, 2014; Massimo et al., 2015; Huang et al., 2016; U’Ren & Arnold, 2016), suggesting that host identity likely interacts with the abiotic environment to structure FEF community composition. Furthermore, most previous studies used culture-dependent methods to characterize FEF communities, which may not account for the large proportion of microorganisms (fungi included) that are difficult or impossible to isolate on artificial media.

Here, we use the results of previous FEF studies to inform our hypotheses regarding the spatial, climactic, and host phylogenetic drivers of FEF community composition in native Hawaiian plants. Specifically, we test the hypotheses that elevation, rainfall, geographic distance, and host plants are the strongest predictors of FEF community composition of native plants across the Hawaiian archipelago (Figure 1). We include more potential explanatory variables in our analysis as well, including variables that have been shown to be important for other non-foliar fungal endophytes such as evapotranspiration (Lewis et al., 1997; Kivlin et al., 2011) and geographic space (Higgins et al., 2014). Unlike previous studies of FEF biogeography that have focused on narrow geographic and host phylogenetic scales, our study spans 80 genera of host plants, as well as a large and ecologically diverse geographic area (Hawai’i). We used Illumina sequencing of the fungal ITS1 region (White et al., 1990) to characterize FEF communities of native plants across the Hawaiian archipelago, in order to test our hypotheses about plant host specificity of FEF, the effect of climate on FEF communities, and whether there are significant geographic patterns in FEF community structure.

The Hawaiian archipelago is an ideal setting in which to investigate how geography structures FEF community composition, because a large body of theory exists describing how biodiversity can be distributed across islands (MacArthur & Wilson, 1967; Hubbell, 2001), such as those studied here (Figure 1). If FEF community composition is the result of dispersal limitation either between islands or simply over large geographic distances, geographically proximate FEF communities are expected to be compositionally similar, while geographically distant communities are more compositionally different (Nekola & White, 1999). Because the Hawaiian archipelago is an island chain (Figure 1), this hypothesis can be tested over a fairly large geographic range (approx. 582 km), with each island as a discrete intermediate unit. In this system, the effect of island specificity can be tested separately from the effect of geographic distance. Additionally, the relatively young geological age and isolation of the islands also allows for a small but endemic native flora, many of which encompass unusually wide niches and elevational distributions (Raich et al., 1997; Wagner, 1999). This orthogonality between plant species distribution and environment allows us to disentangled statistically the effects of host specificity and environment on FEF community composition.

Although geographic distance was a very weak (but significant) variable in the GDM analysis, we found that each of the five Hawaiian islands we sampled harbored significantly specific (i.e. non-cosmopolitan) fungi, commensurate with the high local and regional diversity of FEF. Furthermore, we found that evapotranspiration, elevation, and vegetation type (NDVI) were significant contributors to FEF community structure at the landscape scale, although phylogenetic relationships among host plants were important as well.

**2. MATERIALS AND METHODS:**

*2.1. Sample Collection*

Sites were selected to maximize habitat, phylogenetic and spatial diversity. At each site, a single individual of each native dicot species was haphazardly selected for sampling. Only seemingly healthy, mature, naturally recruited individuals were selected. Leaves were collected such that when combined, they covered a surface area roughly equivalent to two adult-sized hands. The location of each plant was recorded with a GPS and plants were positively identified in the field and/or vouchered for subsequent identification (vouchers deposited at Joseph F. Rock Herbarium at the University of Hawaii, Manoa; HAW). Leaves were refrigerated until subsequent processing (within 72 hours of collection).

We surface sterilized leaves to exclude fungi present on leaf surfaces. After rinsing in water, forty leaf disks were extracted per individual host by punching leaves with a sterile standard paper single hole punch (approximately 0.5 cm diameter). Leaf disks were then placed into loose-leaf tea bags that were subsequently stapled shut, submerged in 1% NaClO for 2 minutes, then 70% EtOH for 2 minutes, followed by two rinses with sterile water for 2 minutes each. Rinse water was included in extraction controls to verify sterility of surface water.

*2.2. DNA Isolation*

Ten leaf disks per DNA extraction were placed in MP Biomedical Lysing Matrix A tubes (MP Biomedical, Santa Ana, CA, USA) containing DNA isolation solutions from the MoBio PowerPlant Pro DNA Isolation kit (Solution PD1, Solution PD2, Phenolic Separation Solution, and RNase A Solution; MO Bio, Carlsbad, CA, USA). Leaf disks were homogenized using a Mini-Beadbeater 24 (BioSpecs Inc. OK) at 3,000 oscillations per min for 2 minutes. Lysate was centrifuged at 13,000 RPM for 2 minutes and transferred to individual wells of a MoBio PowerPlant Pro DNA 96-well Isolation kit for subsequent extraction following the manufacturer’s protocol.

*2.3. PCR Amplification and Illumina Library Preparation*

We amplified the ITS1 region of the ribosomal cistron using fungal specific primers ITS1f and ITS2, along with Illumina adaptors and Golay barcodes incorporating a dual indexing approach, using previously published thermal cycling parameters (Smith and Peay, 2014). PCRs were carried out in 25 μl reactions using the KAPA3G Plant PCR kit (KAPA Biosystems, Wilmington, MA, USA), 9ul of DNA extraction and 0.2 μM each of the forward and reverse primers. Negative PCR and extraction controls were included. PCR products were purified and normalized using just-a-plate 96 PCR Purification and Normalization Kit (Charm Biotech, San Diego, California, USA). Normalized PCR products were pooled and concentrated using a streptavidin magnetic bead solution. Pooled PCR products were sequenced on five separate reactions using the 2 x 300 paired-end (PE) sequencing protocol on an Illumina MiSeq sequencing platform (Illumina Inc., Dan Diego, CA, USA).

*2.4. DNA Sequence data processing and bioinformatics*

QIIME (Caporaso et al., 2010) was used to demultiplex raw DNA sequence data into individual fastq files for each original sample. Although paired-end sequencing was used, only the R1 read (corresponding to primer ITS1f) was used for downstream analysis, since sequencing quality of reverse reads was generally poor. Vsearch (Rognes et al., 2016) was used to discard reads with an average quality score below 25 (illumina Q+33 format), then ITSx (Bengtsson-Palme et al., 2013) was used to extract the ITS1 region from quality-filtered files.

To cluster ITS1 sequences using the unoise3 algorithm (Edgar, 2016), sequences were first de-replicated at 100% identity using vsearch (Rognes et al., 2016), then zOTU centroid sequences were picked and chimeric sequences were removed using unoise3 (Edgar, 2016). Then, all sequences were mapped onto zOTU seeds to create a zOTU table (species x site contingency table) using vsearch. zOTU stands for “zero-radius operational taxonomic unit” (Edgar, 2016). Unlike *de novo* OTUs clustered at user-determined identity cutoffs like 0.97 or 0.95, zOTUs are exact sequence variants (ESVs), which are better able to detect novel diversity while simultaneously filtering out artificial diversity caused by sequencing and PCR error (Callahan et al., 2017).

Taxonomy was assigned to each zOTU using the UNITE database (v 7)(Nilsson, 2011) and QIIME’s assign\_taxonomy.py script (Caporaso et al., 2010) with the BLAST method (Altschul et al., 1990). zOTUs within the pezizomycotina (Blackwell et al., 2006) were retained in the zOTU table to the exclusion of all others, because this group of fungi is known to be largely made up of Class 2 endophytes, which have documented life histories of horizontal transmission, asymptomatic residence within leaf tissue, and post-senescent sporulation (Blackwell et al., 2006; Arnold et al., 2009; Rodriguez et al., 2009). The zOTU table was then rarefied (i.e. randomly downsampled) to 1500 sequences per sample (Supplemental Figure 1). Samples with fewer than 4 zOTUs detected post-rarefaction were discarded because these samples represented outliers in a distribution of zOTU richnesses (Supplemental Figure 2).

GhostTree (Fouquier et al., 2016) was used to construct a phylogenetic tree for the remaining pezizomycotina phylotypes. Briefly, GhostTree allows phylogenetic trees to be made from ITS1 sequence data, which are often un-alignable across families. This is done using a backbone tree created with the 18S rRNA gene, then ITS1 sequences are used to refine the tree at a phylogenetic scale where those sequences can be meaningfully aligned (e.g. genus level). A new GhostTree was made using the SILVA database (v 128) for the 18S backbone, and the UNITE database (v 7). Tips of the GhostTree were renamed with zOTU identifiers where zOTUs were assigned taxonomy to a UNITE entry in the GhostTree. In cases where multiple zOTUs were assigned to the same UNITE entry, a polytomy was created to fit those zOTUs into the tree.

The tree was used with weighted UniFrac (Lozupone & Knight, 2005)(hereafter referred to as “UniFrac”) to construct a beta-diversity matrix for the samples. UniFrac was used because it can better account for fungal diversity than non-phylogenetic methods (Fouquier et al., 2016). Furthermore, UniFrac collapses “spurious” alpha diversity contributed by slight intragenomic variation among tandem repeats. Even if an individual fungus contains several zOTUs, each will only contribute a negligible amount of branch lengths to a sample, versus non-phylogenetic metrics (*e.g.* Bray-Curtis), which would consider those zOTUs as different as any other pair. Because UniFrac community dissimilarity considers the shared phylogenetic branch-lengths between two communities, it is robust to the case where a zOTUs is only found within one sample, and is similarly robust to the case where samples do not share any zOTUs, which can be problematic for other community dissimilarity metrics. This is important for our analysis of FEF communities, because many previous studies have shown that FEF are “hyperdiverse” even at local scales (Jumpponen & Jones, 2009, 2010; Rodriguez et al., 2009; Zimmerman & Vitousek, 2012). We suspected that this large amount of diversity would result in many pairs of samples that shared few or zero zOTUs, which would result in an inflation of 1-values (maximum dissimilarity) when using non-phylogenetic beta-diversity metrics such as Bray-Curtis or Jaccard community dissimilarity. We confirmed that this was indeed the case, and that UniFrac distance values were normally distributed but Bray-Curtis distances were severely 1-inflated (Supplemental Figure 3). For this reason, we used the UniFrac beta-diversity matrix for the remainder of our analysis.

*2.5. Geographic data*

Using sample geographic coordinates, environmental and climatic data for each sample were extracted from GIS layers using the R packages raster (Hijmans et al., 2014) and rgdal (Pebesma et al., 2012). Table 1 shows the sources of each GIS layer. These explanatory variables were chosen either because previous studies of FEF had identified them as important (air temperature, elevation, rainfall), or because they were easily obtained (slope, aspect), or because they were easily available and made intuitive sense in the context of fungi that live within leaves (solar radiation, transpiration, evapotranspiration, leaf area index, NDVI). Slope and aspect of each sampling location were calculated from elevation raster data using the terrain function in the raster package (Hijmans et al., 2014). Data for aspect (the direction a sampling site faces) were converted into a distance matrix using the smallest arc-difference between any two given aspects. This was done because Euclidean distance is unsuitable for a measurement like aspect, where 355° is closer to 1° than it is to 340°. All variables are mean annual values, except for NDVI, which are mean values from December 2017 (the most recent values at the time of analysis). NDVI, or Normalized Difference Vegitation Index, is an index calculated from the amount of infra-red light reflected by plants, which is normalized using multiple wavelengths of visible light. This allows for discrimination between habitat types that are differentially vegetated. Similarly, leaf area index is a measure of surface area of leaves (one-sided) per unit area of ground, and while it does not discriminate between different types of vegetation as does NDVI, unlike NDVI it measures the density of leaf surface area (habitat for FEF).

*2.6. Host plant phylogeny*

A distance matrix of host plant phylogenetic distances was created using the angiosperm phylogeny of Qian and Jin (Qian & Jin, 2016). This distance matrix was made because the modeling approach we use here (GDM, below) can accommodate distance matrices as explanatory variables, allowing for a phylogenetic distance matrix of hosts to be used instead of a simplified data structure such as a principal components vector or an array of taxon identities. For each pair-wise comparison between two samples, pair-wise host plant phylogenetic distance was calculated as the mean cophenetic (branch-length) distance between members of the plant genera that were sampled. In cases where host plant genera were not included in the phylogeny, the genus was substituted for the most closely-related genus that was present. Four genera were substituted in this way out of 80 total genera: *Labordia* → *Logania*, *Touchardia* → *Urtica*, *Waltheria* → *Hermannia*, *Nothocestrum* → *Withania*.

*2.7. GDM analyses*

Generalized dissimilarity modeling (GDM)(Ferrier et al., 2007) was used to model FEF beta diversity based on climatic factors (Table 1), geography, and host plant phylogeny. GDM is a form of non-linear matrix regression that is well-suited to statistical questions involving dissimilarity matrices (e.g. our beta-diversity matrix, host plant cophenetic distance matrix, and aspect arc-difference matrix). Unlike pair-wise Mantel tests or PerMANOVA/ADONIS, which make use of similar data (Oksanen et al., 2016), GDM can quantify the relative importance of environmental and geographic variables on community dissimilarity, even when the functional relationship between community dissimilarity and the environment is nonlinear (Fitzpatrick et al., 2013; Warren et al., 2014). Furthermore, GDM is effective because it can accommodate explanatory variables in the form of distance matrices, column vectors, or geographic coordinates, simultaneously.

Prior to fitting GDM to the data, redundant (i.e. highly correlated) explanatory variables were discarded. Cloud frequency was discarded because it was highly correlated with both rainfall and with solar radiation, air temperature was discarded because it was highly correlated with elevation, relative humidity was discarded because it was highly correlated with solar radiation, and had a clear functional relationship with elevation/air temperature. Evapotranspiration, transpiration, and leaf area index were all correlated with each other, but since we had no strong *a priori* reason to choose between those variables, we performed Mantel correlations (not tests) for each against the FEF beta-diversity matrix, and the variable with the highest absolute *rM* value was chosen (evapotranspiration) and the others discarded. Plots and correlation coefficients of explanatory variables can be seen in Supplemental Figure 4.

We used backward elimination as implemented in the GDM package (Ferrier et al., 2007) to build a model, and then to simplify the model by removing minimally predictive variables. We began this process with the full model (excluding redundant variables; see above), and then tested each variable within the model for significance using a permutation test. During this iterative process, the variable with the highest *P*-value was eliminated, and then the model was recalculated. This process was repeated until all remaining variables were statistically significant (*P* < 0.05). The full model (before backward elimination) modeled fungal beta-diversity using NDVI, evapotranspiration, rainfall, solar radiation, elevation, and slope as column vectors. Additionally, host plant phylogenetic distance and aspect were included in the model as distance matrices, and the geographic coordinates of sample sites were included as geographic information.

*2.8. Island FEF specificity analysis*

Bipartite network analysis was used to test the extent to which each island (Figure 1) harbored specific FEF. The *d'* (“d prime”) statistic was calculated for each island using the zOTU table using the Bipartite package in R (Dormann et al., 2008). *d’* is a measure of network specialization that ranges from 0 to 1, where 0 is perfect cosmopolitanism (all species are evenly shared among islands) and 1 is perfect specialization (each species is specific to only one island). *d’* is calculated using a contingency matrix where each row is a unique lower-level group (island) and each column is a unique higher-level group (zOTU), but in our table each island contains multiple samples. To remedy this, we calculated *d’* by aggregating all samples from the same island into one large sample (column sums), rarefied this aggregated table using the same depth that samples were rarefied to above (1500 observations), then calculated *d’* values. This procedure was repeated 1000 times to obtain a bootstrap distribution of empirical *d’* values for each island. A null model for this analysis was created by randomizing the identities (and therefore islands) of each sample before the calculation of *d’*, and repeating this process 1000 times to generate a null distribution of *d’* values. Statistical significance for *d’* values for each island was tested using Welch’s unequal variance *t*-test. This test was 2-tailed, since *d’* could be significantly lower than the null distribution indicating cosmopolitanism, or significantly higher indicating specificity.

**3. RESULTS**

*3.1. Sample sites and variables*

Samples were collected across a wide range of climatic conditions (Figure 2), which also reflect the distributions of those conditions for Hawai’i. Many sites had a northeastern aspect, because of one large transect on Hawai’i island (Figure 1), however the distribution of aspects in the data set still spanned all directions.

*3.2. Sequence data*

Our data set comprised 896 samples that passed quality-filtering and ITS1 extraction, consisting of 7482 zOTUs. After zOTUs were filtered out that were not in the pezizomycotina (*sensu* (Blackwell et al., 2006)), samples were rarefied to 1500 sequences per sample, and samples with extremely low richness (<4 zOTUs) were discarded, 722 samples remained, containing 4786 zOTUs. Mean richness (number of zOTUs observed) per sample was 32.6 with a standard deviation of 21.3 (Supplementary Figure 2).

*3.2. GDM model simplification*

We eliminated the variables cloud frequency, air temperature, leaf area index, transpiration, and relative humidity because they were confounded with other variables in our analysis (Supplementary Figure 4). Mantel correlations suggested that evapotranspiration (*rM* = 0.104) was a stronger correlate of FEF beta diversity than transpiration (*rM* = 0.040) or leaf area index (*rM* = 0.068). The remaining explanatory variables (NDVI, evapotranspiration, rainfall, solar radiation, elevation, slope, host plant phylogenetic distance matrix, aspect distance matrix, and geographic distance) were used in GDM to explain dissimilarity in the FEF UniFrac beta-diversity matrix, with backward elimination to iteratively remove extraneous variables with *P*-values from GDM’s permutation test were higher than our alpha of 0.05. Variables discarded this way were slope (*P*=1), aspect (*P*=1), rainfall (*P*=0.14), and solar radiation (*P*=0.74).

*3.4. GDM results*

In the final GDM, evapotranspiration explained the most compositional dissimilarity in FEF communities, as given by its GDM coefficient (the maximum height of its spline)(Ferrier et al., 2007; Fitzpatrick et al., 2013), which was 0.176 (Figure 3). This value can be interpreted as evapotranspiration explaining 17.6% of the observed differences in FEF communities when all other variables in the model are held constant. Elevation had a coefficient of 0.158, NDVI had a coefficient of 0.059, plant phylogenetic distance had a coefficient of 0.066, and geographic distance had a coefficient of 0.004 (Figure 3).

*3.5. Bipartite network analysis results*

Each of the 5 islands we sampled showed a statistically significant pattern of FEF specialization (Figure 4), since the *d’* values for each island were higher than those generated using our null model (null *d’* values centered around *d’*=0.4 for each island). Specialization in this case means that each island harbors more zOTUs that are unique to that island than would be expected by chance.

**4. DISCUSSION**

The most striking pattern we found in our analysis of FEF communities in native Hawaiian plants was that evapotranspiration, a variable with spatial resolution of 250 m2 (Table 1), is an meaningful variable for the community composition of microscopic fungi living within plant leaves. Evapotranspiration was the most important variable in our analysis in terms of FEF community composition (Figure 3), even moreso than elevation which was measured at a much finer spatial scale (Table 1). In our GDM model, when all other explanatory variables were held constant, evapotranspiration explained 17.6% of differences in FEF community structure. However, this result makes the interpretation of our hypothesis regarding the effect of climate on FEF community composition challenging, since previous studies (Zimmerman & Vitousek, 2012; Coince et al., 2014) suggested that temperature, elevation, and rainfall would be the most important factors structuring FEF community composition instead. While elevation (tightly correlated with temperature; supplementary figure 4) was a significant explanatory variable in our GDM analysis, its effect was smaller than that of evapotranspiration (Figure 3).

Very few studies measure fungal community response to evapotranspiration, and to our knowledge none yet have included FEF. In a meta-analysis of arbuscular mycorrhizal fungal (AMF) community composition at the global scale, Kivlin et al. (Kivlin et al., 2011) found a significant relationship between evapotranspiration and AMF beta diversity. In that study, evapotranspiration significantly explained a very small proportion of AMF beta diversity (R2 = 0.022, PerMANOVA), although even their most robust explanatory variable had a small effect size, too (Latitude, R2 = 0.030). Either because soils are more variable than the insides of leaves, or because that study was a meta-analysis that combined multiple data sets, or because GDM can better account for nonlinear patterns in beta-diversity data than can PerMANOVA, the effect of evapotranspiration on FEF communities we observed here is much more predictive. In a grass system spanning 15 European countries, the response of endophytic fungi to a transpiration gradient was substantial (Lewis et al., 1997), although in that system the endophytes are vertically transmitted, unlike the horizontal transmission that occurs in the dicots we sampled here. The endophytic *Acremonium* *spp.* were significantly more abundant when evapotranspiration was high, suggesting that an interaction between evapotranspiration and fungal community assembly can exist. Furthermore, in *Theobroma cacao*, native FEF have been experimentally shown to almost double the rate of water loss from leaves during maximum stomatal closure (Arnold & Engelbrecht, 2007), suggesting that there is a mechanistic basis for the interaction between FEF and transpiration as well. Evapotranspiration could also drive FEF community structure by changing the leaf interior habitat, and thereby select for different FEF communities at high vs. low evapotranspiration. Indeed, evapotranspiration is strongly related to the moisture content of leaves (Lambers et al., 2008). Evapotranspiration encompasses both plant transpiration and the evaporation of water from soil and other surfaces, and both soil water content and stomatal conductance affect leaf interior moisture (Tardieu et al., 1996; Lambers et al., 2008). In light of these previous studies, our finding that evapotranspiration is a significant predictor of differences between FEF communities makes sense, although the mechanisms by which evapotranspiration affects or is affected by FEF are still not clear.

The other significant drivers of FEF community composition (Figure 3) were mostly expected, particularly elevation which explained 15.8% of FEF community dissimilarity when all other variables in our analysis were held constant. This value is similar to a result reported by Zimmerman and Vitousek (2012), who used PerMANOVA to test the effects of rainfall, elevation, and substrate age on FEF beta-diversity patterns in *Metrosideros polymorpha* (O’hia) trees. They found that elevation explained roughly 17% of compositional dissimilarity between FEF communities, varying slightly depending on which dissimilarity metric was used. Since that study also took place in Hawai’i, and the area sampled overlaps partially with the area of Hawai’i island that we sampled (Figure 1), this result is not in disagreement with previous work. However, Zimmerman and Vitousek (2012) found a significant effect of rainfall, which we did not observe. In fact, rainfall was eliminated from our GDM model via backwards elimination. It may be that rainfall effects *M. polymorpha* FEF communities more strongly than other native Hawaiian plants, since our study encompasses 80 genera compared to the one genus sampled by Zimmerman and Vitousek (2012). Rainfall has also been shown to be a significant driver of FEF community structure in grasses (Giauque & Hawkes, 2013), but in a larger continental-scale analysis of cultured FEF isolates, rainfall was not a strong predictor of FEF diversity (U’Ren et al., 2012).

Unlike elevation, rainfall, and evapotranspiration, which have each been used to model FEF community dissimilarity by only a handful of studies, patterns of host-specificity of FEF communities have been thoroughly documented (Unterseher et al., 2012; Kembel & Mueller, 2014; Massimo et al., 2015; Huang et al., 2016; Kato et al., 2017). Thus, it is not surprising that phylogenetic difference among host plants was a statistically significant predictor of FEF community dissimilarity in our analysis (Figure 3), even though the proportion of dissimilarity explained was only 6.6%. Unlike previous studies that found host specificity of FEF, we used the phylogeny of host plants as an explanatory variable in place of their identity, meaning that under our hypothesis of host specificity, more phylogenetically similar plants are expected to harbor similar FEF communities, and conversely, phylogenetically distant plants are expected to harbor more different FEF communities. We observed this pattern in our data as mentioned above, but our results may also mean that there is a degree of niche conservatism (Wiens et al., 2010) in either FEF or in host plants. For example, FEF community preference may be phylogenetically conserved among closely related plant species, or perhaps host preference is conserved among closely related FEF. In our analysis, host plant phylogeny may have been a more robust predictor of FEF community dissimilarity if our host plants had been classified to species level instead of genus level, but this would have made the use of an existing phylogeny (Qian & Jin, 2016) more difficult, although future studies may sequence host DNA to construct a *de novo* phylogeny. Nevertheless, previous studies of broad-scale host specificity for FEF have shown that host plant specificity occurs at the order or family level (Kembel & Mueller, 2014). Thus, our observation of a significant relationship between host plant phylogenetic distance and FEF community dissimilarity is not surprising, even with only genus-level resolution for the host phylogeny, and our hypothesis of that FEF are host-plant specific is supported.

NDVI (normalized difference vegetation index) was also a significant predictor of FEF community dissimilarity, suggesting that areas that are differentially vegetated harbor different communities of FEF. This difference may be related to the total percent land cover of vegetation, which is a component of NDVI (Purevdorj et al., 1998), or related to the type of plant cover, *i.e.* different plant community compositions (Lunetta et al., 2006), which is also addressed by NDVI. Indeed, FEF communities have been shown to potentially respond to both land cover and habitat type in the host plant *Quercus macrocarpa* (Jumpponen & Jones, 2009, 2010). The pixel size for the MODIS NDVI data we used was 250 m2 (Table 1), meaning that the value at any given location is an aggregate value for a large plant community. Thus, the 5.9% of FEF community dissimilarity that was significantly explained by NDVI in our analysis is related to the density and composition of plant communities. In addition to supporting our hypothesis that FEF communities of native Hawaiian plants would show significant patterns of host specificity, this result suggests that FEF communities may respond to landscape-scale plant community structure and/or density, which been suggested by others studies as well (Kato et al., 2017).

We also hypothesized that FEF communities across the five islands we sampled (Figure 1) would exhibit significant geographic patterns, but the results of our GDM analysis do not strongly support this idea except for a large effect of elevation, discussed above. Although geographic distance was a significant term in the model, it explained only a tiny percentage of FEF community dissimilarity (Figure 3). However, we also used a bipartite network analysis to investigate the extent to which each island had unique FEF zOTUs, and this analysis revealed that each island harbors a significant share of FEF that are regionally specific (Figure 4). This result is expected, given the generally high diversity of FEF in other systems (Jumpponen & Jones, 2009; Rodriguez et al., 2009; Zimmerman & Vitousek, 2012). This specificity does not directly translate to spatial patterns in FEF beta-diversity, though, because we did not observe a significant effect of geographic space in our GDM analysis. Nevertheless, our finding of significant island-specificity of FEF supports our hypothesis of significant spatial patterns in FEF, just not at the community level.

In summary, our analysis highlights the various factors contributing to FEF community structure across the Hawaiian archipelago. It is also the first study of FEF to analyze these important plant symbionts across a large geographic scale (Figure 1) and across a large host phylogenetic scale (80 plant genera), using high-throughput sequencing to thoroughly inventory FEF community composition. We tested leading hypotheses about the effects of climate, geography, and host identity using this system, and for the most part, found them to be strong predictors of differences in FEF communities between samples – even when the measurements were taken at large spatial scales (Table 1). We found that elevation (Zimmerman & Vitousek, 2012), host plant specificity (Unterseher et al., 2012; Kembel & Mueller, 2014; Massimo et al., 2015; Huang et al., 2016), spatial (U’Ren et al., 2012), and habitat type (Jumpponen & Jones, 2009, 2010; Kato et al., 2017) hypotheses all held true to varying extents in our analysis (Figure 3). The hypothesis that rainfall significantly structures FEF communities (U’Ren et al., 2012; Zimmerman & Vitousek, 2012) was not supported by our analysis, but we instead found that evapotranspiration, which had not been previously considered as an important variable, was the strongest predictor or difference in FEF communities across the Hawaiian archipelago.

**5. ACKNOWLEDGEMENTS**

This work was supported by NSF award #1255972 to Amend, NSF award #XXXXX to Perry and an NSF GRFP award to Cobian. The authors thank Don Hemmes, Jesse Adams, Vincent Costello, Erin Datlof, Seana Walsh, Tim Flynn, Adam Williams, Richard O’Rorke, Steve Perlman, Pat Bily and Kristin Coelho for help in the field, Leah Tooman for assistance in the lab. The authors declare that no conflict of interest exists.

**6. FIGURE CAPTIONS**

Figure 1: Map of sample sites across the Hawaiian Archipelago. Samples were collected from the 5 major islands in the Hawaiian archipelago: Hawai’i, Maui, Moloka’i, O’ahu, and Kauai. Sampling was most dense on Hawai’i and on O’ahu islands, where accessibility was most feasible. Our sampling strategy was to use elevational transects where possible, in order to capture elevational and climatic variation. This is visible in the transects on Hawai’i, Moloka’i, Kauai, and O’ahu, which run orthogonally to the topographic lines (white). Transects are less pronounced on Maui because of limited accessibility.

Figure 2: Ranges and distributions of explanatory variables. Each variable included in our analysis covered a wide range of environmental heterogeneity. Some variables had skewed distributions either because of sampling bias or because that variable’s distribution across the Hawaiian archipelago is naturally skewed. For example, our distribution of aspects is skewed toward a northeastern direction, because of large transects on Hawai’i island, however the full range of aspects (full 360 degrees) is still included in the distribution.

Figure 3: Model fit and coefficients for GDM model of FEF community dissimilarity. The observed community dissimilarity (UniFrac distance) between pairwise samples exhibited a linear but noisy relationship with the community dissimilarity predicted by the GDM model (top), which roughly corresponded to a 1:1 line (dashed line). Evapotranspiration was the strongest predictor variable in our analysis (green curve, bottom), although most of the explanatory power of evapotranspiration was at the lower 40% of its range (see Figure 2). Elevation (blue) was the second strongest explanatory variable, and unlike evapotranspiration it significantly explained FEF community dissimilarity across the entire range of values (see Figure 2). Host plant phylogenetic distance (purple) and NDVI (pink) were both statistically significant drivers of FEF community dissimilarity as well, with total explained community dissimilarity of 6.6% and 5.9%, respectively. Geographic distance (brown) was statistically significant, but explained less than 1% of community dissimilarity.

Figure 4: FEF specialization of each island. *d’* is a measure of how unique or cosmopolitan the zOTUs found on an island are. In this violin plot, distributions shown in black are bootstrapped empirical *d’* values for each island, and distributions shown in gray are null models where the identity of islands was randomized before *d’* was calculated. Welch’s unequal variance *t-*tests show that each island’s FEF community is significantly specialized, meaning that FEF zOTUs are more unique to their island of origin than expected by chance.

**7. REFERENCES**

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., & Lipman, D.J. (1990) Basic local alignment search tool. *Journal of molecular biology*, **215**, 403–10.

Arnold, A.E. & Engelbrecht, B.M.J. (2007) Fungal endophytes nearly double minimum leaf conductance in seedlings of a neotropical tree species. *Journal of Tropical Ecology*, **23**, 369–372.

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

Arnold, A.E., Miadlikowska, J., Higgins, K.L., Sarvate, S.D., Gugger, P., Way, A., Hofstetter, V., Kauff, F., & Lutzoni, F. (2009) A phylogenetic estimation of trophic transition networks for ascomycetous Fungi: Are lichens cradles of symbiotrophic Fungal diversification? *Systematic Biology*, **58**, 283–297.

Bayman, P., Angulo-Sandoval, P., Báez-Ortiz, Z., & Lodge, D.J. (1998) Distribution and dispersal of Xylaria endophytes in two tree species in Puerto Rico. *Mycological Research*, **102**, 944–948.

Bengtsson-Palme, J., Ryberg, M., Hartmann, M., et al. (2013) Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in Ecology and Evolution*, **4**, 914–919.

Blackwell, M. (2011) The fungi: 1, 2, 3 ... 5.1 million species? *American Journal of Botany*, **98**, 426–438.

Blackwell, M., Hibbett, D.S., Taylor, J.W., & Spatafora, J.W. (2006) Research Coordination Networks: a phylogeny for kingdom Fungi (Deep Hypha). *Mycologia*, **98**, 829–837.

Callahan, B.J., McMurdie, P.J., & Holmes, S.P. (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, .

Caporaso, J.G., Kuczynski, J., Stombaugh, J., et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature methods*, **7**, 335–6.

Coince, A., Cordier, T., Lengell??, J., Defossez, E., Vacher, C., Robin, C., Bu??e, M., & Mar??ais, B. (2014) Leaf and root-associated fungal assemblages do not follow similar elevational diversity patterns. *PLoS ONE*, **9**, .

Dormann, C.F., Gruber, B., & Fruend, J. (2008) Introducing the bipartite Package: Analysing Ecological Networks. *R news*, **8**, 8–11.

Edgar, R.C. (2016) UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv*, .

Felber, A.C., Orlandelli, R.C., Rhoden, S.A., Garcia, A., Costa, A.T., Azevedo, J.L., & Pamphile, J.A. (2016) Bioprospecting foliar endophytic fungi of Vitis labrusca Linnaeus, Bordô and Concord cv. *Annals of Microbiology*, **66**, 765–775.

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

Fitzpatrick, M.C., Sanders, N.J., Normand, S., Svenning, J.-C., Ferrier, S., Gove, A.D., & Dunn, R.R. (2013) Environmental and historical imprints on beta diversity: insights from variation in rates of species turnover along gradients. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20131201–20131201.

Fouquier, J., Rideout, J.R., Bolyen, E., Chase, J., Shiffer, A., McDonald, D., Knight, R., Caporaso, J.G., & Kelley, S.T. (2016) Ghost-tree: Creating hybrid-gene phylogenetic trees for diversity analyses. *Microbiome*, **4**, .

Giauque, H. & Hawkes, C. V. (2013) Climate affects symbiotic fungal endophyte diversity and performance. *American Journal of Botany*, **100**, 1435–1444.

González-Teuber, M. (2016) The defensive role of foliar endophytic fungi for a South American tree. *AoB Plants*, **8**, plw050.

Hawksworth, D.L. & Rossman, A.Y. (1997) Where Are All the Undescribed Fungi? *Phytopathology*, **87**, 888–891.

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

Hijmans, R.J., Etten, J. van, Mattiuzzi, M., Sumner, M., Greenberg, J.A., Lamigueiro, O.P., Bevan, A., Racine, E.B., & Shortridge, A. (2014) Package “raster.” *R*, 1–27.

Huang, Y.-L., Devan, M.M.N., U’Ren, J.M., Furr, S.H., & Arnold, A.E. (2016) Pervasive Effects of Wildfire on Foliar Endophyte Communities in Montane Forest Trees. *Microbial Ecology*, **71**, 452–468.

Hubbell, S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32).* Princeton University Press,

Jumpponen, A. & Jones, K.L. (2009) Massively parallel 454 sequencing indicates hyperdiverse fungal communities in temperate *Quercus macrocarpa* phyllosphere. *New Phytologist*, **184**, 438–448.

Jumpponen, A. & Jones, K.L. (2010) Seasonally dynamic fungal communities in the Quercus macrocarpa phyllosphere differ between urban and nonurban environments. *New Phytologist*, **186**, 496–513.

Kato, S., Fukasawa, Y., & Seiwa, K. (2017) Canopy tree species and openness affect foliar endophytic fungal communities of understory seedlings. *Ecological Research*, **32**, 157–162.

Kembel, S.W. & Mueller, R.C. (2014) Plant traits and taxonomy drive host associations in tropical phyllosphere fungal communities. *Botany*, **92**, 303–311.

Kivlin, S.N., Hawkes, C. V., & Treseder, K.K. (2011) Global diversity and distribution of arbuscular mycorrhizal fungi. *Soil Biology and Biochemistry*, **43**, 2294–2303.

Lambers, H., Chapin, F.S., & Pons, T.L. (2008) Plant Water Relations. *Plant Physiological Ecology* pp. 163–223. Springer New York, New York, NY.

Lewis, G.C., Ravel, C., Naffaa, W., Astier, C., & Charmet, G. (1997) Occurrence of Acremonium endophytes in wild populations of Lolium spp. in European countries and a relationship between level of infection and climate in France. *Annals of Applied Biology*, **130**, 227–238.

Lozupone, C. & Knight, R. (2005) UniFrac : a New Phylogenetic Method for Comparing Microbial Communities UniFrac : a New Phylogenetic Method for Comparing Microbial Communities. *Applied and environmental microbiology*, **71**, 8228–8235.

Lunetta, R.S., Knight, J.F., Ediriwickrema, J., Lyon, J.G., & Worthy, L.D. (2006) Land-cover change detection using multi-temporal MODIS NDVI data. *Remote Sensing of Environment*, **105**, 142–154.

MacArthur, R.H. & Wilson, E.O. (1967) *Theory of Island Biogeography.* Princeton University Press,

Massimo, N.C., Nandi Devan, M.M., Arendt, K.R., Wilch, M.H., Riddle, J.M., Furr, S.H., Steen, C., U’Ren, J.M., Sandberg, D.C., & Arnold, A.E. (2015) Fungal endophytes in aboveground tissues of desert plants: infrequent in culture, but highly diverse and distinctive symbionts. *Microbial ecology*, **70**, 61–76.

Mejía, L.C., Herre, E.A., Sparks, J.P., Winter, K., García, M.N., Van Bael, S.A., Stitt, J., Shi, Z., Zhang, Y., Guiltinan, M.J., & Maximova, S.N. (2014) Pervasive effects of a dominant foliar endophytic fungus on host genetic and phenotypic expression in a tropical tree. *Frontiers in microbiology*, **5**, 479.

Nekola, J.C. & White, P.S. (1999) The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, **26**, 867–878.

Nilsson, R.H. (2011) Molecular Identification of Fungi: Rationale, Philosophical Concerns, and the UNITE Database. *The Open Applied Informatics Journal*, **5**, 81–86.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGinn, D., Minchin, P.R., O’Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., & Wagner, H. (2016) vegan: Community Ecology Package. .

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 Biology*, **119**, 917–928.

Pebesma, E., Rowlingson, B., & Rogerbivandnhhno, M.R.B. (2012) Package “ rgdal .” *R-CRAN*, **4**, 1–41.

Polonio, J.C., Almeida, T.T., Garcia, A., Mariucci, G.E.G., Azevedo, J.L., Rhoden, S.A., & Pamphile, J.A. (2015) Biotechnological prospecting of foliar endophytic fungi of guaco (Mikania glomerata Spreng.) with antibacterial and antagonistic activity against phytopathogens. *Genetics and Molecular Research*, **14**, 7297–7309.

Porras-Alfaro, A. & Bayman, P. (2011) Hidden Fungi, Emergent Properties: Endophytes and Microbiomes. *Annual Review of Phytopathology*, **49**, 291–315.

Purevdorj, T.S., Tateishi, R., Ishiyama, T., & Honda, Y. (1998) Relationships between percent vegetation cover and vegetation indices. *International Journal of Remote Sensing*, **19**, 3519–3535.

Qian, H. & Jin, Y. (2016) An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure. *Journal of Plant Ecology*, **9**, 233–239.

Raich, J.W., Russell, A.E., & Vitousek, P.M. (1997) Primary productivity and ecosystem development along an elevational gradient on Mauna Loa, Hawai’i. *Ecology*, **78**, 707–721.

Rodriguez, R.J., White, J.F., Arnold, a E., & Redman, R.S. (2009) Fungal endophytes: diversity and functional roles. *The New phytologist*, **182**, 314–330.

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ Preprints*, **4**, e2409v1.

Saucedo-García, A., Anaya, A.L., Espinosa-García, F.J., & González, M.C. (2014) Diversity and Communities of Foliar Endophytic Fungi from Different Agroecosystems of Coffea arabica L. in Two Regions of Veracruz, Mexico. *PLoS ONE*, **9**, e98454.

Stone, J.K., Bacon, C.W., & White, J.F. (2000) An overview of endophytic microbes: endophytism defined. *Microbial Endophytes*, 3–29.

Tardieu, F., Lafarge, T., & Simonneau, T. (1996) Stomatal control by fed or endogenous xylem ABA in sunflower: interpretation of correlations between leaf water potential and stomatal conductance in anisohydric species. *Plant, Cell and Environment*, **19**, 75–84.

U’Ren, J.M. & Arnold, A.E. (2016) Diversity, taxonomic composition, and functional aspects of fungal communities in living, senesced, and fallen leaves at five sites across North America. *PeerJ*, **4**, e2768.

U’Ren, J.M., Lutzoni, F., Miadlikowska, J., Laetsch, A.D., & Arnold, A.E. (2012) Host and geographic structure of endophytic and endolichenic fungi at a continental scale. *American Journal of Botany*, **99**, 898–914.

Unterseher, M., Petzold, A., & Schnittler, M. (2012) Xerotolerant foliar endophytic fungi of Populus euphratica from the Tarim River basin, Central China are conspecific to endophytic ITS phylotypes of Populus tremula from temperate Europe. *Fungal Diversity*, **54**, 133–142.

Wagner, W.L. (1999) *Manual of the flowering plants of Hawaiʻi.* University of Hawai’i Press: Bishop Museum Press, Honolulu.

Warren, D.L., Cardillo, M., Rosauer, D.F., & Bolnick, D.I. (2014) Mistaking geography for biology: Inferring processes from species distributions. *Trends in Ecology and Evolution*, **29**, 572–580.

White, T.J., Bruns, T., Lee, S., & Taylor, J. (1990) AMPLIFICATION AND DIRECT SEQUENCING OF FUNGAL RIBOSOMAL RNA GENES FOR PHYLOGENETICS. *PCR Protocols* pp. 315–322.

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

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