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

RUNNING TITLE
Hawaiian leaf fungi at the landscape scale

AUTHORS
John L. Darcy[1], Gerald M. Cobian[1], Sean O. I. Swift[1], Geoffrey L. Zahn[2], Brian A. Perry[3], Anthony S. Amend[1]

[1] Department of Botany, University of Hawai'i Mānoa, Hawai'i, USA
[2] Department of Biology, Utah Valley University, Utah, USA
[3] Department of Biological Sciences, California State University East Bay, California, USA

AUTHOR CONTRIBUTIONS
JLD wrote the paper, performed bioinformatic and statistical analyses, and created figures. ASA and BAP conceived of the experiment. SOS, GC, and GLZ collected data and performed laboratory work. All authors contributed to the preparation of the manuscript.

ACKNOWLEDGEMENTS

ABSTRACT

AIM: A phylogenetically diverse array of fungi live within healthy leaf tissue of dicotyledonous plants. 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 endophytic fungi in native Hawaiian dicots sampled across the Hawaiian archipelago.

LOCATION: Hawai'i.

METHODS: The Hawaiian archipelago offers a uniquely tractable system to study biogeography of foliar endophytic fungi, because the islands harbor a wide array of climatic conditions, and native plant species are often found across wide elevational and climactic ranges. 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: Communities of foliar endophytic fungi in the Hawaiian archipelago are structured most strongly by evapotranspiration, elevation, vegetation/habitat type, and by the phylogeny of host plants. 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 m^2, 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.

KEYWORDS: Fungi, Endophytes, Dicots, Pezizomycotina, High-throughput sequencing, Island Biogeography

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), it is unclear which factors structure FEF community composition, and how those factors differ in their relative importance. Previous 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; Vincent et al., 2016) have all been identified as putatively 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, or studying those hypotheses within a narrow host phylogenetic framework.

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). Studies that have surveyed FEF or other leaf-associated fungi 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; Vincent et al., 2016), suggesting that host identity likely interacts with the abiotic environment to structure FEF community composition. Furthermore, many 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 plant phylogeny 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 evapotranspiration (Lewis et al., 1997; Kivlin et al., 2011) and geographic distance (Higgins et al., 2014): variables shown to be important for other non-foliar fungal endophytes. 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 the effect of host plant phylogeny on FEF, the effect of climate on FEF communities, and whether there are significant geospatial patterns in FEF community structure.

The Hawaiian archipelago is an appropriate 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 should be compositionally similar, while geographically distant communities should be compositionally distinct (Nekola & White, 1999). Additionally, the relatively young geological age and isolation of the islands results in a small but mostly endemic native flora, and many of these species encompass unusually wide niches and elevational distributions (Raich et al., 1997; Wagner, 1999). This, combined with a phylogeny of the native plants studied, enables us to analyze the effect host plant phylogenetic dissimilarity has on FEF community composition while accounting for the effects of other predictor variables using Generalized Dissimilarity Modeling (GDM).

Although geographic distance was a very weak (but significant) variable in our 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 NDVI (normalized difference vegetation index) were significant contributors to FEF community structure at the landscape scale (across the Hawaiian archipelago), although phylogenetic relationships among host plants were important as well.

MATERIALS AND METHODS

Sample Collection

General sampling locations were selected to maximize habitat, phylogenetic, and spatial diversity. We selected locations to prioritize access, a reasonable permitting process, and the known presence of ten or more native plant species. Access was the most limiting factor in comparatively fewer sampling locations on Molokai and Maui Islands. We chose not to collect plants that are federally listed as threatened or endangered. At each location, the first occurrence of an apparently healthy, naturally recruited individual was selected for sampling. Only native plants were selected this way. We limited our sampling to single individuals, because we did not wish to confound our data with uneven species density and resulting spatial heterogeneity among samples. Mature, sun leaves were collected such that when combined, they covered a surface area roughly equivalent to two adult-sized hands. Life form and stature differed markedly across our sampled plants, so it was difficult to precisely standardize collection location on plants. When possible leaves from trees and shrubs were collected at eye level and from at least four aspects of the tree canopy. We focused on dicots because they are easily identifiable, accessible, and have horizontally transmitted FEF (compared to many vertically transmitted FEF in some monocots). Although not all host plant genera were collected at every sample location, the three most common genera we collected (*Metrosideros*, n=119; *Leptecophylla*, n=89; and *Vaccinium, n*=147) were each collected across elevations ranging from sea level to over 2000 meters above sea level. 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). A total of 1099 samples were collected in this way, but not all were used in the analyses presented here (see results section).

We surface sterilized leaves to exclude fungi present on leaf surfaces. After rinsing in water, forty leaf discs were extracted per individual host by punching leaves with a sterile standard paper single hole punch (approximately 0.5 cm diameter). For plants with very small leaves, entire leaves were used such that the area of those leaves was the same as the area of leaf discs. Leaf discs (or aforementioned collections of small leaves) were then placed into loose-leaf tea bags that were subsequently stapled shut, submerged in 1% NaOCl 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.

DNA Isolation

Ten leaf discs 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 discs were homogenized using a Mini-Beadbeater 24 (BioSpec 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.

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 & 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 (concentration not measured) 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 in order of processing completion, using the 2 x 300 paired-end (PE) sequencing protocol on an Illumina MiSeq sequencing platform (Illumina Inc., Dan Diego, CA, USA). We did not include controls for sequencing run variation, and we acknowledge this as a limitation of our study. Negative control samples were discarded due to extremely low read count.

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 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 sample 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 using the default maximum e-value of 0.001 (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 3 endophytes (Rodriguez et al., 2009), 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). Other FEF may exhibit these features as well, although less universally, and it is not our view that other FEF are ecologically unimportant. Furthermore, Pezizomycotina were chosen for analysis because they were disproportionately represented within our data set: 90% of our samples were over 50% composed of Pezizomycotina, and 50% of our samples were over 90% composed of Pezizomycotina. The zOTU table was then rarefied (i.e. randomly downsampled) to 1500 sequences per sample (Figure S1), discarding samples with fewer than 1500 sequences or samples for which host plants could not be satisfactorily identified.

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, and then ITS1 sequences are used to refine the tree at a phylogenetic scale where those sequences can be meaningfully aligned (e.g. genus level). When used in an analysis of real and simulated ITS1 data, the GhostTree+UniFrac approach resulted in higher amounts of variance explained than non-phylogenetic metrics (Fouquier et al., 2016), and this approach was also successfully used to study differences in fungal community composition in the human infant microbiome (Ward et al., 2018). 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. This approach may exclude novel taxa that are not present in the UNITE reference database, and this is a limitation of all studies using such a database, including this one.

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 is not sensitive to "spurious" extra OTUs 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 length 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 (e.g. within

one host or within one hundred meters) (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. Indeed, this same issue has been discussed by previous users of GDM as a saturation of maximal dissimilarities, and was solved by using phylogenetic dissimilarty metrics (Rosauer et al., 2014). We confirmed that this was indeed the case, and that UniFrac distance values were nicely distributed but Bray-Curtis dissimilarities were severely 1-inflated (Figure S2). For this reason, we used the UniFrac beta-diversity matrix for the remainder of our analysis. The methods described above were additionally used to create a second UniFrac distance matrix for all FEF, not just Pezizomycotina.

Spatiotemporal data

Using sample geographic coordinates and collection dates, 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). Data were extracted from rasters generated for the same month that samples were collected (except for elevational data). Table 1 shows the sources of each GIS layer. Many explanatory variables were obtained from the Rainfall of Hawaiʻi and Evapotranspiration of Hawaiʻi websites (Giambelluca et al., 2013, 2014), and elevational data were obtained using the USGS EarthExplorer online tool (http://earthexplorer.usgs.gov), courtesy of NASA EOSDIS Land Processes Distributed Active Archive Center and the United States Geological Survey's Earth Resources Observation and Science Center. 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 location 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 monthly values. NDVI (normalized difference vegetation 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).

Host plant phylogeny

A distance matrix of host plant phylogenetic distances was created using the angiosperm phylogeny of Qian and 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 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*.

GDM analysis

Generalized dissimilarity modeling (GDM)(Ferrier et al., 2007) was used to model FEF beta diversity based on climatic factors (Table 1), geographic distance, 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, geographic 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 models community dissimilarity associated with a given predictor variable while holding all other predictor variables constant (Fitzpatrick et al., 2013; Landesman et al., 2014). For example, this enables GDM to model the effect of elevation while accounting for the effect of host plant phylogenetic distance.

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 including all explanatory variables mentioned above except for air temperature, which was excluded because it was highly correlated with elevation ($r = 0.99$). Geographic distance between samples, host plant phylogenetic distance, aspect arc-difference matrix, and UniFrac beta-diversity matrix were included as matrices in the GDM; all other variables were supplied as column vectors. GDM 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 ($P < 0.05$). Leaf area index, transpiration, slope, wet canopy evaporation, and aspect were eliminated this way.

The GDM modeling procedure described above was also carried out using Pezizomycotina community distance matrices created using Bray-Curtis dissimilarity for both zOTUs and OTUs clustered at 95% sequence similarity using vsearch (Rognes et al., 2016), to verify that the severe 1-inflation observed in those distance matrices resulted in poor model fitting. Additionally, this procedure was carried out using the UniFrac distance matrix generated for all FEF.

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, so the number of observations per island is not consistent. 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. Two null models were used to ensure that structural biases in our data were not responsible for the observed patterns of specificity. The Vázquez null model is a fixed-connectance null model for bipartite networks (Vázquez et al., 2007), which is implemented as the vaznull function of the R package bipartite (Dormann et al., 2008). Vázquez was run on the aggregated table, and then a null d' was calculated using the result. A second null model ("island null") was created by randomly shuffling the identities of samples before they were aggregated by island, so that aggregated samples were mixtures of multiple islands, and then a null d' was calculated using the result. This procedure of aggregating, calculating an empirical d', calculating a Vázquez null d', and calculating an island null d' was repeated 1000 times in order to obtain bootstrapped distributions of empirical, Vázquez null, and island null d' values for each island. Statistical significance for d' values for each island was tested using Welch's unequal variance t-test to compare empirical and island null d' values. This test was 2-tailed, since d' could be significantly lower than the null distribution indicating cosmopolitanism, or significantly higher indicating specificity. A similar procedure was applied to host plant genera in place of fungal zOTUs. Since host is only observed once per sample, only the Vázquez null model was used and d' was not bootstrapped. This was done in order to test the extent to which our selection of host plants was significantly specialized to islands.

RESULTS

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.

Sequence data

Our data set comprised 896 samples that passed quality-filtering and ITS1 extraction, consisting of 7482 zOTUs. After non Pezizomycotina (*sensu* (Blackwell et al., 2006)) zOTUs were discarded, samples were rarefied to 1500 sequences per sample, resulting in 5239 zOTUs across 720 samples. Of those samples, 399 were from Hawai'i, 80 from Kaua'i, 51 from Maui, 67 from Moloka'i, and 123 from

O'ahu. Mean richness (number of zOTUs observed) per sample was 30.8 with a standard deviation of 21.5.

GDM results

In the final GDM, evapotranspiration and NDVI explained the most compositional dissimilarity in FEF communities, as given by their GDM coefficient (the maximum height of their splines)(Ferrier et al., 2007; Fitzpatrick et al., 2013), which were 0.0991 and 0.1042, respectively (Figure 3). This can be interpreted as evapotranspiration explaining 10% of the observed difference in FEF communities when all other variables in the model are held constant. Elevation had a coefficient of 0.0813, host plant phylogenetic distance had a coefficient of 0.0644, and Julian date had a coefficient of 0.0238 (Figure 3). All other statistically significant variables had coefficients less than 0.020: cloud frequency = 0.0175, rainfall = 0.0174, relative humidity = 0.0173, solar radiation = 0.0171, geographic distance = 0.0021. In contrast to the GDM fit using the UniFrac distance matrix (Figure 3), GDM run with Bray-Curtis dissimilarity matrices (zOTU and 95% OTUs) fit poorly as a result of severe 1-inflation of distances (Figures S2 and S3). Even so, both UniFrac and Bray-Curtis GDM analyses found statistically significant effects of geographic distance, host plant phylogeny, NDVI, evapotranspiration, and elevation. Evapotranspiration was not a significant variable in the GDM analysis of all FEF (instead of only Pezizomycotina above), although NDVI (0.1129), host plant phylogenetic distance (0.0660), elevation (0.0540) and Julian date (0.0311) had similar GDM coefficients to the Pezizomycotina model (Figure S4).

Bipartite network analysis

Each of the 5 islands we sampled showed a statistically significant ($P < 0.001$, Welch's unequal variance t-test) 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 associated with that island than would be expected by chance. We also found that our selection of host plant genera was significantly specialized to island as well (Figure S5, Table S1), however host-island d' values were not significantly related to zOTU-island d' values (linear regression, $P = 0.45$), which we interpret to mean that FEF island specificity is not explained by island-biased plant sampling.

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 m^2 (Table 1), is a meaningful variable for the community composition of microscopic fungi living within plant leaves. Evapotranspiration was the second-most important variable in our analysis in terms of FEF community composition (Figure 3), even more so than elevation, which was measured at a much finer spatial resolution (Table 1). In our GDM model, when all other explanatory variables were held constant, evapotranspiration explained 10% of differences in FEF community structure. This result is both novel and surprising, since we

expected that temperature, elevation, and rainfall would be the most important factors structuring FEF community composition as has been found in previous studies (Zimmerman & Vitousek, 2012; Coince et al., 2014) whereas evapotranspiration has not been previously considered as an important variable for FEF. This result may also partly due to our focus on Pezizomycotina, because evapotranspiration was not a significant variable in our analysis when all FEF were considered (Figure S4). While elevation (tightly correlated with temperature), host plant phylogeny, and Julian date were significant explanatory variables in our GDM analyses of Pezizomycotina and of all FEF, each of their effects were smaller than that of evapotranspiration for Pezizomycotina (Figure 3).

Very few studies measure fungal community response to evapotranspiration, and to our knowledge none yet have included FEF. 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 fungus *Acremonium spp.* was significantly more abundant when evapotranspiration was high, suggesting that an interaction between evapotranspiration and fungal community composition is possible. Furthermore, in *Theobroma cacao*, addition of native FEF can 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. Kivlin et al. (Kivlin et al., 2011) conducted a global-scale meta-analysis of arbuscular mycorrhizal fungal (AMF) community composition, across multiple host plants, and found a weak but statistically significant relationship between evapotranspiration and AMF beta diversity ($R^2 = 0.022$, PERMANOVA). However, even the most robust explanatory variable in that study had a small effect size, too (Latitude, $R^2 = 0.030$).

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 previous studies suggesting a link between fungal endophytes and evapotranspiration (Lewis et al., 1997; Arnold & Engelbrecht, 2007; Kivlin et al., 2011), 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 8% of FEF community dissimilarity when all other variables in our analysis were held constant. This 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 Hawaiʻi, and the area sampled overlaps partially with the area of Hawaiʻi island that we sampled (Figure 1), this result is not in disagreement with previous work. However, Zimmerman and Vitousek (2012) found a large effect of rainfall, which was much smaller (but still significant) in our data set. It may be that rainfall effects *M. polymorpha* FEF communities more strongly than those of other native Hawaiian plants, since our study encompasses 80 genera compared to the single species sampled by Zimmerman and Vitousek (2012). Rainfall can 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 association in FEF communities have been thoroughly documented (Unterseher et al., 2012; Kembel & Mueller, 2014; Massimo et al., 2015; Huang et al., 2016; Vincent 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). Unlike most previous studies that found host associations of FEF, we used the phylogeny of host plants as an explanatory variable in place of their identity, meaning that under our hypothesis that FEF communities are structured by host phylogenetic difference, more phylogenetically similar plants are expected to harbor similar FEF communities, and conversely, phylogenetically distant plants are expected to harbor more different FEF communities. Although we observed a significant relationship between host phylogenetic structure and FEF community dissimilarity in our data, our results could either mean that FEF, host plants, or both exhibit a degree of niche conservatism (Wiens et al., 2010). 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) impossible, although future studies may sequence host DNA to construct a *de novo* phylogeny. Nevertheless, previous studies of broad-scale host association for foliar fungal epiphytes have shown that host plant association 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 community composition is related to host-plant phylogeny is supported.

NDVI (normalized difference vegetation index) was the most robust predictor of FEF community dissimilarity in our data set (although essentially tied with evapotranspiration), 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 NDVI data we used was 250 m^2 (Table 1), meaning that the value at any given location is an aggregate value for a large plant community. Thus, the 10% of FEF community dissimilarity that was significantly explained by NDVI in our analysis is related to the density and composition of plant communities, 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 strong geospatial 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. However, we also used a bipartite network analysis to investigate the extent to which each island had specialized Pezizomycotina zOTUs, and this analysis revealed that each island harbors a significant share of zOTUs that are island-specific (Figure 4). Distinct zOTUs are not phylogenetically weighted by similarity within the bipartite analysis like they are in the GDM, and this may explain why the bipartite analysis detected significant specificity of Pezizomycotina zOTUs to islands while geographic distance was a significant but very weak term in the GDM model. Furthermore, specificity of a zOTU to an island does not necessarily cause a biogeographic pattern where adjacent islands harbor similar zOTUs, which is essentially the pattern tested in the GDM with geographic distance.

Since our beta-diversity analysis revealed a significant effect of host on Pezizomycotina community composition, it may have been that instead of FEF being island-specific, host plants were instead island-specific within the context of our experiment. In other words, our experimental design could have introduced FEF specificity by sampling host plant specificity. A bipartite analysis of host-island specificity (Figure S5) indicated that in our experimental design, hosts are indeed specific to islands. However, we expected that if host-island specificity bias within our experimental design was driving the zOTU-island specificity pattern we observed, the two would strongly correlate. There was no significant relationship between host-island specificity and zOTU-island specificity, and host-island *d'* values were much lower than those observed for zOTUs. So while part of the zOTU specificity we observed may be due to bias in our sampling design, there is likely a significant pattern of zOTU island specificity as well. Nonetheless, 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 island specialization is yet another result showing that FEF are distributed across space at the landscape scale (across the Hawaiian archipelago), with each island likely harboring many FEF that are significantly specialized to it.

In summary, our analysis highlights that the various factors contributing to Hawaiian FEF community structure do so at the landscape scale. Ours is also the first study of FEF to analyze these important plant symbionts across both a large geographic scale (Figure 1) and across a large host phylogenetic scale (80 plant genera), while using high-throughput sequencing to thoroughly inventory FEF community composition. We tested leading hypotheses about the effects of climate,

geographic distance, 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 relatively coarse resolution (Table 1). We found that elevation (Zimmerman & Vitousek, 2012), host plant phylogenetic difference (Unterseher et al., 2012; Kembel & Mueller, 2014; Massimo et al., 2015; Huang et al., 2016; Vincent 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). We found limited (but significant) support for the hypothesis that rainfall structures FEF communities (U'Ren et al., 2012; Zimmerman & Vitousek, 2012), and we also found that evapotranspiration, which had not been previously considered as an important variable, was a significant predictor of difference in FEF communities across the Hawaiian archipelago.

TABLES

Table 1: Sources for GIS data.

Data type	Units	Pixel size	Source	Source URL
Rainfall	mm*y^{-1}	250 m^2	Rainfall atlas of Hawai'i	rainfall.geography.hawaii.edu
Evapotranspiration	mm*y^{-1}	250 m^2	Evapotranspiration of Hawai'i	evapotranspiration.geography.hawaii.edu
Leaf area index	index	250 m^2	Evapotranspiration of Hawai'i	evapotranspiration.geography.hawaii.edu
Wet Canopy Evaporation	mm*y^{-1}	250 m^2	Evapotranspiration of Hawai'i	evapotranspiration.geography.hawaii.edu
Transpiration	mm*y^{-1}	250 m^2	Evapotranspiration of Hawai'i	evapotranspiration.geography.hawaii.edu
Solar radiation	W/m^2	250 m^2	Evapotranspiration of Hawai'i	evapotranspiration.geography.hawaii.edu
Relative humidity	%	250 m^2	Evapotranspiration of Hawai'i	evapotranspiration.geography.hawaii.edu
Air temperature	°C	250 m^2	Evapotranspiration of Hawai'i	evapotranspiration.geography.hawaii.edu
Cloud frequency	ratio	250 m^2	Evapotranspiration of Hawai'i	evapotranspiration.geography.hawaii.edu
Elevation	m	10.3 m^2	National elevation dataset	lta.cr.usgs.gov/NED
Slope	Degrees	10.3 m^2	Calculated from elevation	N/A (calculated from elevation)
Aspect	Degrees	10.3 m^2	Calculated from elevation	N/A (calculated from elevation)
NDVI	index	250 m^2	Evapotranspiration of Hawai'i	evapotranspiration.geography.hawaii.edu

FIGURES

Figure 1: Map of sample locations across the Hawaiian Archipelago. Samples were collected from the 5 major islands in the Hawaiian archipelago: Hawaiʻi, Maui, Molokaʻi, Oʻahu, and Kauaʻi. Sampling was most dense on Hawaiʻi and on Oʻahu islands, where accessibility was easiest. Our sampling strategy was to use elevational transects where possible, in order to capture elevational and climatic variation. Elevations in figure legend are meters above sea level. This is visible in the transects on Hawaiʻi, Molokaʻi, Kauaʻi, and Oʻahu, which run orthogonally to the topographic lines (white). Transects are less pronounced on Maui because of limited accessibility.
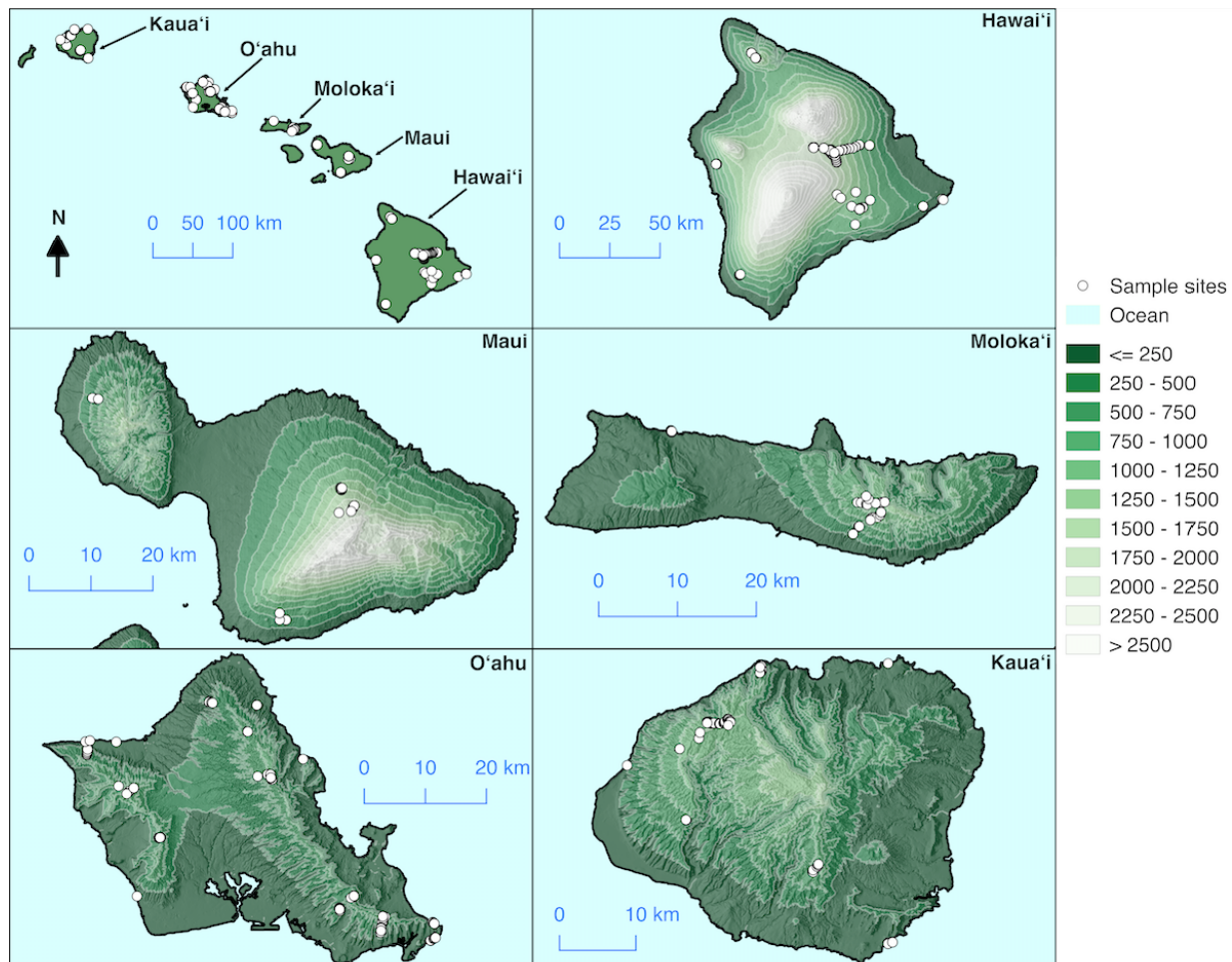
Figure 2: Ranges and distributions of explanatory variables. In this figure, each variable's distribution across 722 samples is shown as a smoothed density curve between its range (numbers to left and right of curves, height of curve is relative frequency of observation). Units for range values are shown in Table 1. Each variable included in our analysis covered a wide range of environmental heterogeneity.
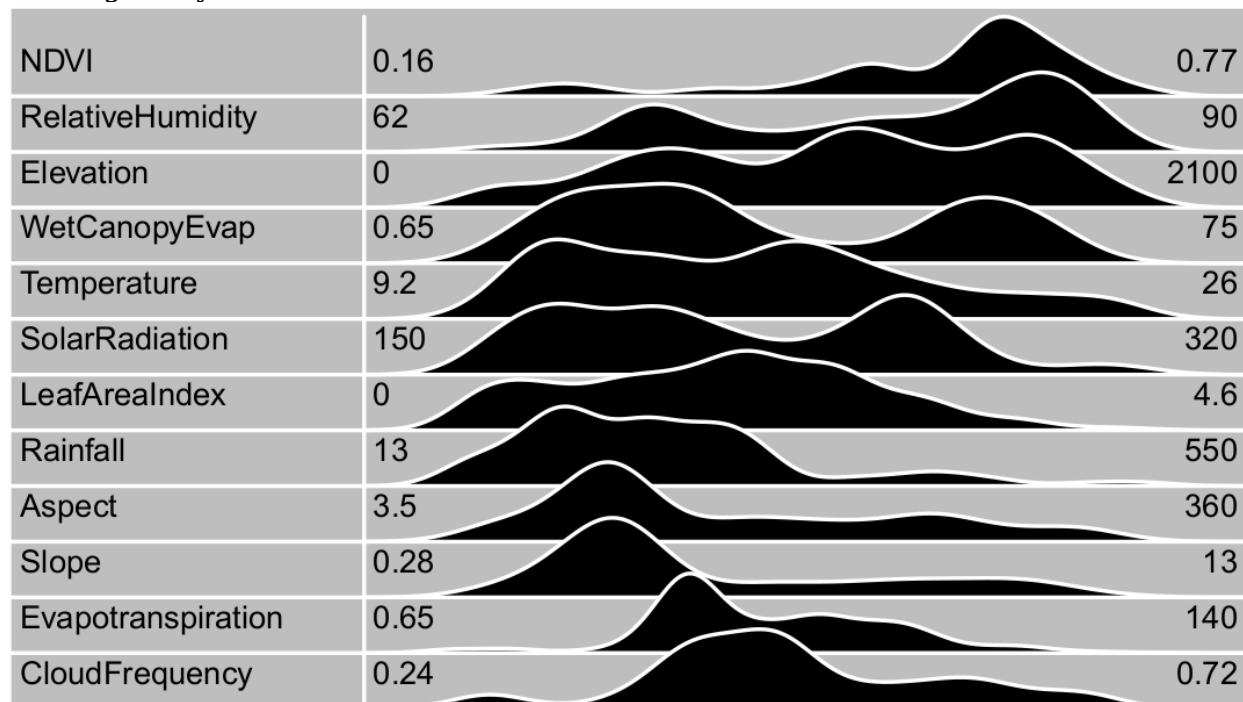
Figure 3: Model fit and coefficients for GDM model of FEF community dissimilarity. The observed community dissimilarity (UniFrac distance) between pairwise samples exhibited a linear but noisy relationship with the community dissimilarity predicted by the GDM model (top), which roughly corresponded to a 1:1 line (dashed line). In the bottom plot, GDM i-splines are shown for statistically significant variables that explained over 2% of community dissimilarity. Spline height is the amount of cumulative community dissimilarity explained by its predictor variable, and the spline's slope corresponds to the rate of change in compositional dissimilarity over the range of pairwise dissimilarities within the variable (Ferrier et al., 2007; Landesman et al., 2014). NDVI and Evapotranspiration were the strongest predictor variables in our analysis (green and blue), although most of the explanatory power of evapotranspiration was at the lower 40% of its range. Elevation (red) significantly explained FEF community dissimilarity across the entire range of elevation (see Figure 2). Host plant phylogenetic distance (purple) was a significant driver of community dissimilarity across a deep phylogenetic breadth, and the significance of julian date (brown) indicates as small but important temporal trend that is strongest at short temporal scales (less than a year). Other significant variables in this analysis were cloud frequency, rainfall, relative humidity, solar radiation, and geographic distance, but they all explained very little community dissimilarity and overlapped to too large a degree to show in the figure.
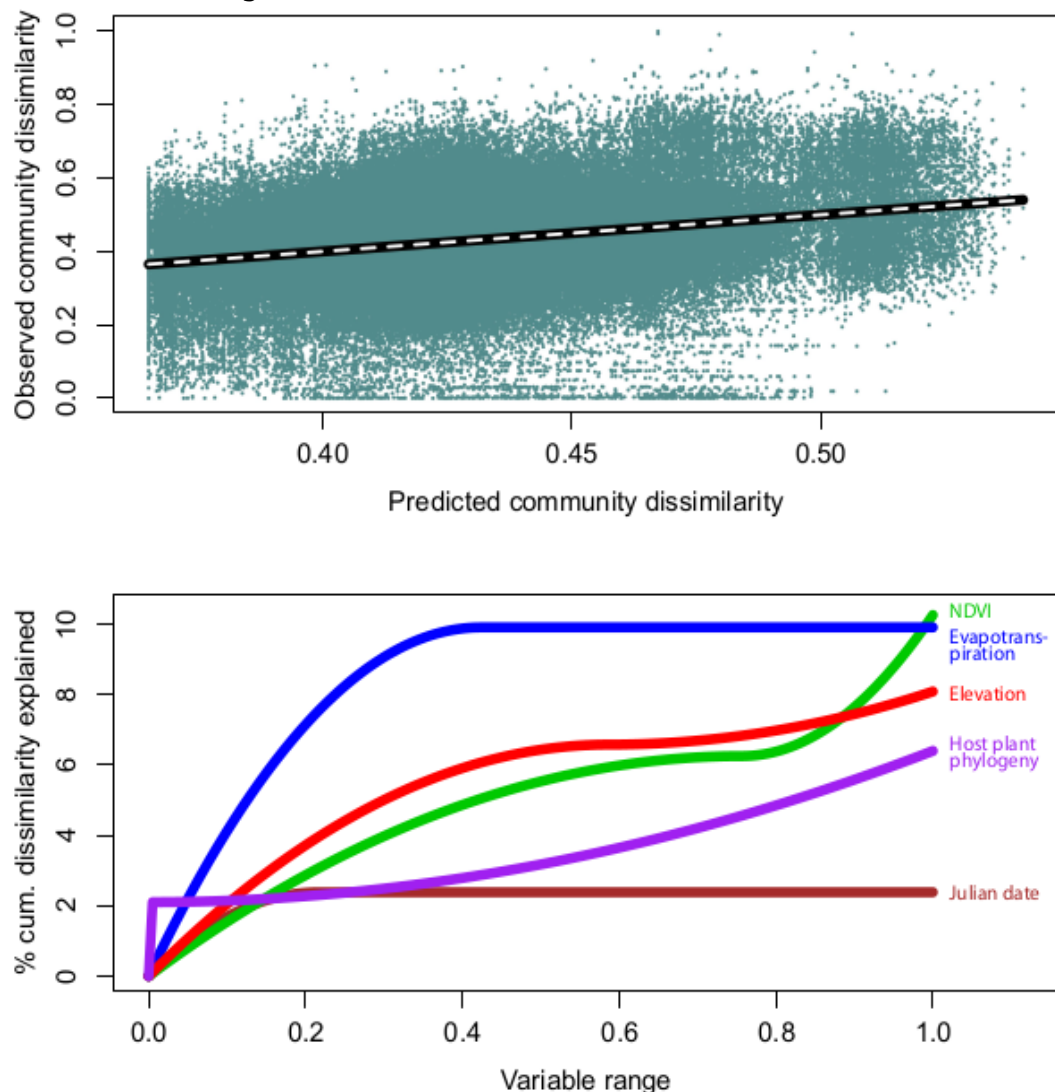
Figure 4: FEF specialization of each island. *d'* is a measure of how unique or cosmopolitan are the zOTUs found on an island. In this violin plot, distributions shown in as filled violins with black borders are bootstrapped empirical *d'* values for each island. Unfilled distributions with a solid gray border are from our island null model, where the island datum of each sample was randomly permuted before samples were aggregated and *d'* was calculated. Unfilled distributions with black dashed borders are from the Vázquez null model for bipartite networks (Vázquez et al., 2007), which is a fixed-connectance null model. Welch's unequal variance *t*-tests show that each island's FEF community is significantly specialized as compared to the island null, meaning that FEF zOTUs are more specific to their island of origin than expected by chance.
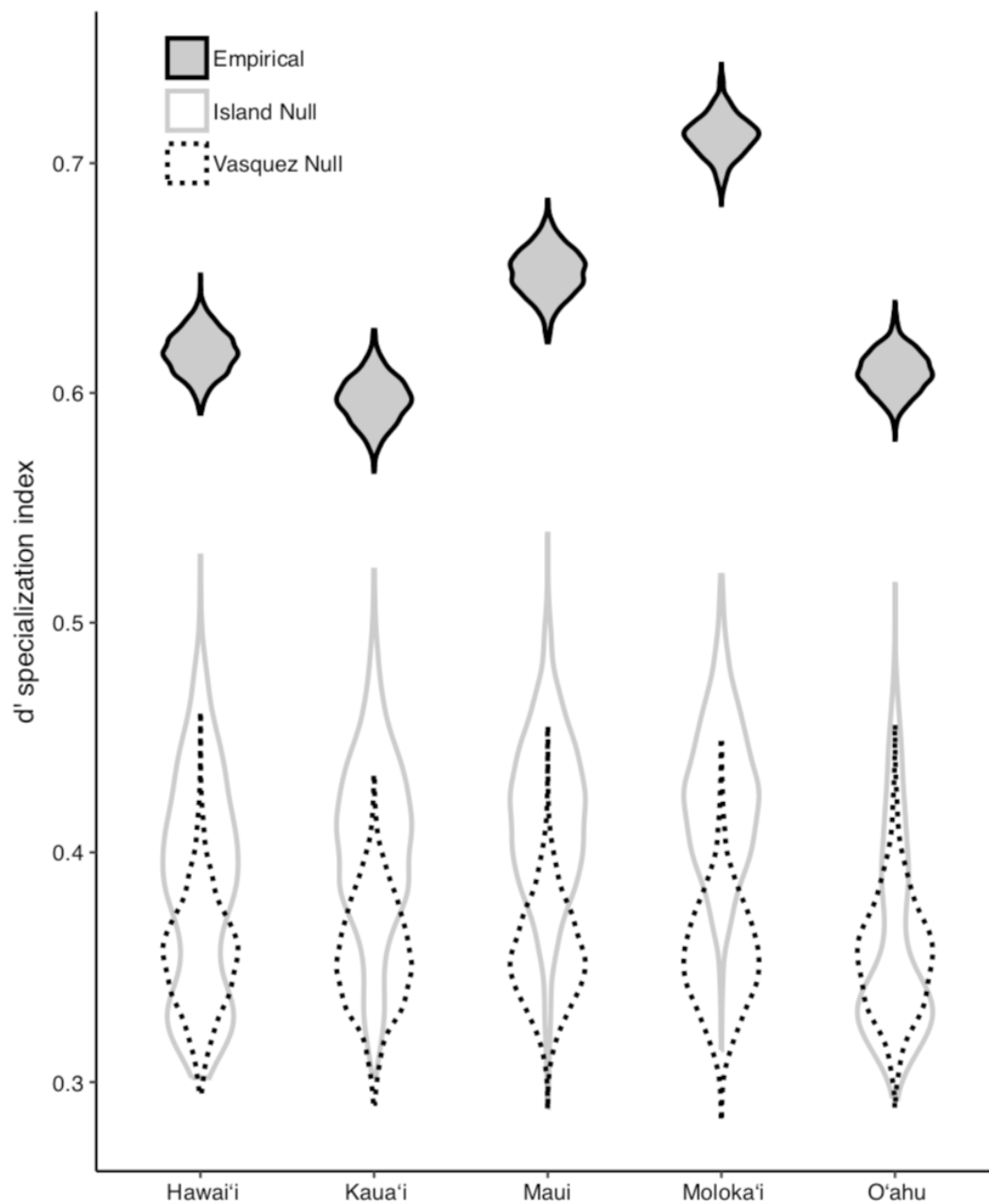
REFERENCES

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*, **11**, 2639-2643.

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., Lengellé, J., Defossez, E., Vacher, C., Robin, C., Buée, M., & Marçais, B. (2014) Leaf and root-associated fungal assemblages do not follow similar elevational diversity patterns. *PLoS ONE*, **9**, e100668.

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*, 081257.

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.

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**, 11.

Giambelluca, T.W., Chen, Q., Frazier, A.G., Price, J.P., Chen, Y.L., Chu, P.S., Eischeid, J.K., & Delparte, D.M. (2013) Online rainfall atlas of Hawai'i. *Bulletin of the American Meteorological Society*, **94**, 313–316.

Giambelluca, T.W., Shuai, X., Barnes, M.L., Alliss, R.J., Longman, R.J., Miura, T., Chen, Q., Frazier, A.G., Mudd, R.G., Cuo, L., & Businger, A.D. (2014) Evapotranspiration of Hawai'i. Final report submitted to the U.S. Army Corps of Engineers—Honolulu District, and the Commission on Water Resource Management, State of Hawaii.

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.

Landesman, W.J., Nelson, D.M., & Fitzpatrick, M.C. (2014) Soil properties and tree species drive ß-diversity of soil bacterial communities. *Soil Biology and Biochemistry*, **76**, 201–209.

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.

Rosauer, D. F., Ferrier, S. , Williams, K. J., Manion, G. , Keogh, J. S. and Laffan, S. W. (2014), Phylogenetic generalised dissimilarity modelling: a new approach to analysing and predicting spatial turnover in the phylogenetic composition of communities. Ecography, **37**, 21-32.

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.

Smith, D.P. & Peay, K.G. (2014) Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. *PLoS ONE*, **9**, e90234.

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.

Vázquez, D.P., Melián, C.J., Williams, N.M., Blüthgen, N., Krasnov, B.R., & Poulin, R. (2007) Species abundance and asymmetric interaction strength in ecological networks. *Oikos*, **116**, 1120–1127.

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

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

Ward, T.L., Dominguez-Bello, M.G., Heisel, T., Al-Ghalith, G., Knights, D., & Gale, C.A. (2018) Development of the Human Mycobiome over the First Month of Life and across Body Sites. *mSystems*, **3**, e00140-17.

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.

BIOSKETCH:

John L. Darcy is a postdoctoral researcher at the University of Hawai'i, Mānoa, in Anthony S. Amend's lab. The lab's work focuses on diversity and evolution of fungal endophytes in native Hawaiian plants and biogeography of microorganisms (www.amendlab.com).

DATA ACCESSIBILITY:

DNA sequence data have been submitted to the NCBI Sequence Read Archive (SRA) database under BioProject accession number PRJNA470970. Computer code and input files that can be used to replicate this analysis can be obtained by contacting the authors, and will be made publicly available on FigShare following publication. Additionally, a simple summary of the four most commonly sampled plant genera for each island is available as Table S1. DNA sequence data are available from both NCBI SRA and FigShare repositories.