Journal of Ecology
July 2020, Volume 108, Issue 4, Pages 1605-1616
https://doi.org/10.1111/1365-2745.13344
https://archimer.ifremer.fr/doc/00606/71808/

A test of community assembly rules using foliar endophytes from a tropical forest canopy

Donald Julian [1, *], Roy Mélanie [1], Suescun Uxue [1], Iribar Amaia [1], Manzi Sophie [1], Péllissier Léonie [2], Gaucher Philippe [3], Chave Jérôme [1, *]

[1] Université Toulouse 3 Paul Sabatier, CNRS IRD, UMR 5174 Evolution et Diversité Biologique (EDB) 118 route de NarbonneF-31062 Toulouse , France
[2] School of Pharmaceutical Sciences EPGL, University of Geneva, University of Lausanne Rue Michel Servet 1CH-1211Geneva , Switzerland
[3] CNRS, Université de Guyane, UMRS LEEISA Campus IRD de Cayenne 97300, Cayenne

* Corresponding authors : Julian Donald, email address : julesdonald@gmail.com ; Jérôme Chave, jerome.chave@univ-tlse3.fr

Abstract :

1.Community assembly theory assumes that ecological communities are spatially delimited into patches. Within these patches, coexistence results from environmental filtering, competition, and immigration. Truly delineated communities exist in laboratory studies of microbial cultures in Petri dishes, yet empirical tests conducted in continuous environments often use patches defined by convention as opposed to realised boundaries.

2.Here we perform a test of ecological community assembly rules using foliar endophyte communities from a tropical rainforest, where leaves are considered as patches for both fungal and bacterial communities. We determined the diversity of fungal and bacterial endophytes using environmental DNA sequencing of 365 top-canopy leaves, collected from 38 host trees belonging to 22 different species across a 4-hectare research plot. Three leaves were collected from three or more branches within each tree crown. We tested the effect of host tree species and their level of phylogenetic relatedness on community composition as well as the contribution of geographic distance between leaves to endophyte community diversity.

3.Endophyte diversity significantly differed across host tree species, as did community composition. Within certain endophytic orders (Xylariales, Rhizobiales) species assemblages significantly differed across host tree species, but this trend was weaker or non-existent in other orders known to contain pathogens and saprotrophs (Polyporales, Solirubrobacterales). Phylogenetically related host tree species displayed more similar endophyte communities than expected by chance, but geographically close trees did not. Consistent with the finding of host-specificity, nearby leaves tended to host more similar communities than distantly positioned ones.

4.These findings demonstrate that foliar endophytes are structured by dispersal across small spatial scales, but at the scale of the canopy they display patterns of neutral filtering, with only a small part of variation described by host tree differences. Endophyte communities thus act as a model system in evoking the rules predicted by theoretical community ecology.

Keywords : Bacteria, Distribution, French Guiana, Fungi, Metabarcoding, Plant–pathogen interactions, Rainforest

Introduction

Understanding the mechanisms which allow for the emergence and maintenance of species diversity, and in particular the coexistence of competing species, is a challenge for community ecology (Ricklefs 2004). Theory suggests that local species diversity results from a balance between local constraints, including environmental conditions and competition, and immigration from external sources through dispersal (Vandermeer 1972, Leibold 1995, Hubbell 2001, Condit et al. 2002, Chase et al. 2005). Studies aiming to contrast the relative weight of these mechanisms require spatially discrete communities of identifiable species at a single trophic level, organized in discrete patches, and existing in large enough numbers to allow for replication (Ellwood et al. 2009). Whilst such conditions are achievable for microbiologists conducting laboratory-based experiments on communities grown in Petri dishes (e.g. Maynard et al. 2017), these fail to represent the levels of variability experienced by natural communities (Crowther et al. 2018). Field tests of community assembly rules have often been conducted using trees in tropical rainforests forests, habitats characterized by high levels of species diversity, with large levels of species turnover from one community to the next (Hubbell 2001, Fine et al. 2004). In these studies, communities of trees are defined by convention as opposed to realised boundaries, and as such inferences of structural patterns may be biased, masking the true processes structuring ecological communities.

Foliar endophytes inhabit the intercellular layers of leaves, without provoking a symptomatic response from the host plant, although certain species may be latent saprobes or pathogens (de Bary 1886; Carroll 1988; Porras-Alfaro & Bayman 2011). Within leaves, multiple species coexist in spite of likely competition for the resources available. Because they are constrained to live within leaves, in communities more delineated than other microbial habitats such as soils, they represent a model example of a metacommunity. Colonisation by endophytes is thought to occur soon after the flushing of new leaves, when spores deposit, germinate and subsequently penetrate the leaf cuticle without provoking an antipathogenic response (Arnold and Herre 2003, Griffin & Carson 2015). Initial leaf colonists are likely to maintain their dominance in relation to future colonists, resulting in an effective host defence against pathogens and invertebrate herbivores (Herre et al. 2007; Van Bael et al. 2018), albeit at a cost to the host plant in terms of nitrogen metabolism and photosynthetic rate (Mejia et al. 2014, Christian et al. 2019).

Foliar endophytes have been found in all sampled plant hosts (U'Ren et al. 2012), with over 8,000 species of fungi, and 7,000 species of bacteria described as occurring only within leaves according to a recent survey (Griffin & Carson 2018). Tropical regions have shown to support much higher levels of endophyte diversity than temperate regions (Arnold et al. 2000, Arnold & Engelbrecht 2007, Arnold & Lutzoni 2007), although the factors responsible for generating such high levels of diversity remain elusive, as well as the mechanisms that structure communities within leaves. Studies of bacterial endophytes are for the most part limited to their occurrence in agricultural crops or trees in temperate regions (Griffin & Carson 2015, Miyambo et al. 2016, Carrell & Carper 2016). The roles of host tree species (Lodge 1997, Arnold & Herre 2003), dispersal among endophyte communities (Cordier et al. 2012, Izuno et al. 2016), and abiotic conditions (Suryanararanan et al. 2002, 2003, Zimmerman & Vitousek 2012) have all been explored. In these studies, only part of community variation was explained, with the idea of community composition as shaped by neutral processes largely ignored. Indeed, these studies have focused on leaf samples collected from the forest understory, potentially masking the greater levels of wind-driven dispersal in the upper canopy (Gilbert & Reynolds 2005), and restricting the sampling to leaves that contribute little to the overall net primary production of tropical forest ecosystems. Furthermore, the majority of previous studies have relied on techniques which represent endophyte communities based on fungi and bacteria detected following tissue culturing, although it is known that many species do not grow easily on standard culture media (Hyde & Soytong 2008). Environmental DNA (eDNA) sequencing potentially captures a greater breadth of the diversity associated with environmental samples by identifying of observable taxonomic units (OTUs) as proxies for species, and thus is a promising technique for the exploration of endophyte diversity within leaf samples (e.g. Vincent et al. 2016, Del Olmo-Ruiz & Arnold 2017, Haruna et al. 2017).

This study aims to elucidate the factors structuring ecological communities, by characterising the diversity of upper-canopy fungal and bacterial endophytes in a primary rainforest, and shifts in their community composition across phylogenetic and geographical gradients. Because endophytes are restricted to leaf habitats, this system represents an ideal setup to test community ecology theory. Our study makes use of an innovative canopy-access system to sample upper crown leaves from multiple host trees across a fully inventoried 4-ha permanent tree plot. If foliar endophyte communities are structured to a greater extent by their niche, then we would expect to see a strong correlation between the composition of endophyte

communities and their host plants. This is to be expected if closely allied host species are more similar with respect to their foliar functional traits than unrelated species (i.e. the traits are phylogenetically conserved as shown by Baraloto et al. (2012) and Courtois et al. (2016)). If dispersal limitation is important in structuring communities, then we would expect to see proximal communities within the canopy following an isolation-by-distance pattern, from leaves sharing the same branch, to those associated with separate tree crowns across the study plot.

We hypothesize that at the scale of the 4-ha plot, canopy endophyte assemblages will display patterns of neutral structuration as a result of dispersal limitation. However, when controlling for the scale at which communities assemble, we expect that host tree species will determine endophyte colonization success, and that closely allied host species will match this by having more similar communities than distantly related species. Similarly, we expect that within crowns, proximate leaves will be more similar than distant leaves. In particular, we hypothesize that the response of endophyte occurrence within certain host trees will depend on their level of host specificity, detected when focusing on endophyte community responses within specific fungal and bacterial orders.

Materials and Methods

Sample Site

Data collection took place at the Nouragues Ecological Research Station in French Guiana, within an area of pristine lowland tropical rainforest protected since 1996. This area is characterised by a high level of floral diversity, with over 1700 recorded angiosperm species (Sabatier, 1990, van der Meer and Bongers, 1996). The reserve experiences mid-day temperatures of 26 °C and rainfall of 2861 mm per year (Réjou-Méchain et al. 2015). It experiences two dry seasons, a longer dry season between September and mid-November, and a shorter dry season in March.

Fieldwork described here took place in August 2017. Sampling was conducted in a 4-ha permanent sampling plot adjacent to the Pararé research station (4°02’ N, 52°41’ W) located on the Arataye tributary of the Approuague River. Access to the canopy was made possible by the ‘Canopy Operational Permanent Access System’ (COPAS), consisting of a trio of 45 m high pylons set up at the vertices of an equilateral triangle with edges 180 m in

length, equipped with a mechanically operated harness which allows for the displacement of an individual anywhere within the upper canopy across the research plot (Fig. 1).

Sample design and preparation

Leaves were collected from the upper crown of 38 trees, of 22 different species, selected to represent a phylogenetically diverse range of species (Fig. 2 & S.I. 1). A stratified sampling design was conducted for each tree: three leaves were collected from each of a minimum of three branches positioned across the crown. For those trees with a large crown, 4 or 5 branches were selected accordingly, so that sampling accurately represented the total spread of the crown. In total this resulted in 365 leaves.

Leaves were immediately transported to a field laboratory where they were surface sterilised in order to ensure that sampling consisted of only endophytes. This sterilisation included several steps, first mechanical cleaning and scraping with toothbrushes. Secondly, following from the protocol detailed by Arnold et al. (2000), leaves were soaked in diluted bleach (0,525%) for 2 min before transfer into ethanol (70%) for a further 2 min. This method is the standard in eDNA extraction for endophyte research (e.g. Izuno et al. 2016, Haruna et al. 2018), proving efficient in removing the maximum amount of epiphytic fungi and bacteria from the leaf surface. Finally, leaves were dried with tissue paper, and stored in Ziploc bags containing silica gel before transfer to the lab for DNA extraction.

DNA Isolation, PCR amplification and sequencing

Two 0.5 cm^2 pieces of each leaf were placed into a 96 well plate, and a 3mm aluminium bead was added to each well. To facilitate the grinding step, the sealed plates were placed in a -80 °C freezer for at least an hour before grinding using a Tissue Lyser (Qiagen, Germany) for 45 sec at 30Hz, followed by a centrifugation for 30 seconds at 6000 x g. DNA was extracted from samples using a NucleoSpin Plant II (96) kit (Macherey-Nagel, Düren, Germany) according to the manufacturer protocol modifying only cell lysis time to 1 h and elution volume to 100µL.

PCRs were performed using two primers; ITS1 nuclear rDNA primers to target fungi (Fwd: ITS5 GGAAGTAAAAGTCGTAACAAGG (Epp et al. 2012) and a modified version of Rev: 5.8S_Fungi CAAGAGATCCGTTGTTGAAAGTK, Taberlet et al. 2018), and 16S rDNA (V5-V6) primers to target bacteria (Bact01 primers - Fwd:

GGATTAGATACCCTGGTAGT and Rev: CACGACACGAGCTGACG (Fliegerova et al. 2014)). To discriminate samples after sequencing, forward and reverse primers were synthetized with a combination of two-different 8-nucleotide tags per sample, following a double-indexing strategy (Binladen et al. 2007). As such, each PCR was amplified with a unique combination of tagged primers. Each PCR reaction was performed in a total volume of 20 µl and comprised 10 µl of AmpliTaq GoldMaster Mix (Life Technologies, Carlsbad, CA, USA), 5.84 µL of nuclease-free water Ambion™ (Invitrogen, Waltham, Massachusetts, USA), 0.25 µM of each primer, 3.2µg of BSA (Roche Diagnostic, Basel, Switzerland), and 2 µl of DNA template. Two PCRs per sample were performed under the following conditions: Polymerase reactivation 10 min at 95°C, followed by 35 (fungi) or 30 (bacteria) cycles of 30 sec at 95°C, 30 sec at 55°C (fungi) or 57°C (bacteria) and 1 min at 72°C; followed by a final step of 7 min at 72°C. Two wells in each plate were filled with water, to act as PCR negative controls. Eight wells positioned randomly across each plate were left empty (with no PCR products) to act as sequencing controls (non-used tag combinations).

Amplicons were pooled and libraries were prepared with the TruSeqNano PCR free Illumina kit and were sequenced on Paired-end (2X250 bp) in the Illumina Miseq platform (Illumina, San Diego, CA, USA) at the INRA Genotoul-GetPlaGe core facility (Toulouse, France) using the Paired-end MiSeq Reagent Kit V3 (Illumina, San Diego, CA, USA), following the manufacturer's instructions.

Sequence data curation

The generated reads were subject to a data-curation pipeline using scripts in R (R Development Core Team, 2013), and the OBITools package (Boyer et al. 2016). Firstly, 'Illuminapairedend' was used for paired-end read assembly where an exact alignment algorithm assigns a quality score for each nucleotide position, generating a score for each read based on the number of mismatches, followed by 'ngsfilter' which removes primer and tag sequences and assigns reads to samples. Here, reads with any primers with more than two mismatches, and any tags with 1 or more mismatches are removed. 'obiuniq' then dereplicates only those sequences which are strictly identical. 'obigreb' removes sequences of low quality (containing Ns or with a pair alignment score below 50). Finally sequences with only one read across the global dataset (singletons) were removed, since these represent potential sequencing artefacts.'sumaclust' was then used to generate Operational Taxonomic Units (OTUs), where pairwise sequence dissimilarity using the raw number of mismatches

was calculated, and where those sequences ≥ 97% or more were assigned to the same OTU, a threshold commonly applied to delimitate proxies for during molecular analysis of fungi in particular (Nilsson et al., 2008; Coissac 2012). For each OTU, the sequence with the highest number of reads in each cluster was assigned as the seed sequence, used for subsequent taxonomic assignment.

Taxonomic assignment of OTUs was conducted using reference databases with the 'ecotag' function, which uses global alignment of sequences against full-length references. For fungi the reference database was obtained by running an in-silico PCR with the ecoPCR program (Ficetola et al., 2010) on Genbank (release 197; ftp://ftp.ncbi.nlm.nih.gov/genbank) using the primer pairs employed here. Taxonomic assignment yielding the highest similarity score was kept, with similarity scores for fungi ranging from 0.5 to 1 with sequences in ncbi. For Bacteria, taxonomic assignment was conducted using the SILVA database (Quast et al. 2012), with similarity parameters kept at 0.97, higher than fungi since bacterial sequence databases are much more complete than for fungi (Truong et al., 2017). This approach allowed for the elimination of a certain number of chimeras which are likely to have formed during PCR. In particular, taxonomic identification facilitated later analysis, by distinguishing between orders known to contain OTUs corresponding to endophytic fungi (Xylariales, Capnodiales, Chaetothyriales, Phyllachorales, Diaporthales, Sordariales, Pleosporales, Hypocreales) against orders which are known to contain generalist saprotrophs (Polyporales, Eurotiales, Corticiales). Similarly, bacterial endophyte orders which have strict relationships with plants (e.g. Rhizobiales, Acetobacterales, Sphingomonadales) could be compared with orders known to contain more generalist bacteria (e.g. Solirubrobacterales, Pseudonocardiales, Acidobacteriales).

Although initial processing of sequences using OBITools removes a large number of erroneous sequences introduced to the dataset during sequencing, further processing based on methodological controls can remove additional erroneous sequences and contaminants. OTUs which could represent contaminants arising during extraction and during PCR (present within the PCR negative controls) were removed from the dataset by comparing occurrence in controls with that in samples. Potential contaminants were detected by calculating the mean sum of reads for each OTU in PCR negative controls, and sequencing controls (blanks) divided by the mean sum of reads across both these controls and samples. For each OTU, the acceptable level of OTU abundance present in PCR negative controls was set to be 5% or less

than the total number of reads. Above this threshold, an OTU was more likely to be a contaminant, and as such it was excluded. OTUs which only occur in negative controls were also removed. Finally, all sequence data associated with blank controls was removed from the datasets.

All OTUs which were identified as non-fungal following taxonomic assignment were removed from the fungal database (S.I. 2). Bacterial sequences which were identified as originating from ribosomes or chloroplasts were removed from the bacterial dataset, since these likely represented plant host genetic material rather than that of bacterial endophytes (S.I. 3). It is noted that future studies of bacterial endophytes should use primers more adapted to reduce this effect (e.g. Beckers et al. 2016). PCR replicates were then pooled. We excluded fungal samples with less than 100 reads from subsequent analysis. These low-quality samples likely result from PCR amplification failure. This filter was chosen as a result of the majority of samples averaging a much higher read count. This resulted in the removal of 14 samples, with subsequent analysis for the fungal communities performed on a dataset of 351 leaves. In comparison, bacterial read counts per sample were dominated by ribosome and chloroplast sequences. Once these were removed, all samples containing a minimum of 1 read were kept in the dataset, resulting in the removal of 8 leaves, with subsequent analysis for the bacterial communities performed on a dataset of 357 leaves.

Following clustering of sequence reads with a 97% similarity cut-off, 11,026 OTUs (67,338,264 reads) were obtained for fungi, and 2,859 OTUs for bacteria (12,746,769 reads). Removing OTUs occurring disproportionately in the negative controls and blank PCR wells brought the number of OTUs down to 10,303 for fungi, and to 2,438 for bacteria. Removal of non-fungal OTUs brought the total number of OTUs down to 10,160 (S.I. 2), and 1,953 OTUs for bacteria (S.I 3). Subsequent analysis was performed on these reduced datasets.

Statistical analysis

In order to establish the relative effects of host tree niche creation and dispersal limitation, leaf sample endophyte diversity was classified and community composition at multiple spatial and phylogenetic scales was determined. All statistical analysis was performed using Rstudio version 1.1.442 (R Development Core Team, 2013).

306 To quantify endophyte diversity, rarefication curves for each PCR based on the
307 number of reads were compiled using the 'rarefy' function in the vegan package (Oksanen et
308 al. 2013). These curves demonstrated whether sequencing depth was sufficient in representing
309 the diversity of selected host tree endophyte diversity, defined by whether curves reached
310 asymptote. Species accumulation curves were also prepared using the 'specaccum' function,
311 which determine whether the number of leaves sampled for each species was sufficient in
312 capturing endophyte diversity. Subsequent analysis was conducted on a standardised dataset,
313 whereby the observed dataset was resampled so as to have the same relative abundance of
314 OTUs, but for read count per sample to be equal, using the first quantile of the dataset as the
315 new total read count per sample. Endophyte alpha diversity for each of the datasets (351
316 leaves for fungi, 357 for bacteria) was characterized by calculating a Shannon Index for each
317 sample, selected to neither over nor under represent rare OTUs. This was converted to
318 effective number of species, to reflect the number of equally-common species required to give
319 a particular value (Jost 2006).

320 The effect of host tree species in influencing endophyte alpha diversity was
321 determined using ANOVA, with Tukey Post Hoc tests revealing which species-by-species
322 comparisons were significant. To further understand within communities which OTU groups
323 differed significantly from one host tree species to the next, data were converted to presence
324 absence, and aggregated at the order level, chosen since taxonomic assignment for fungi in
325 particular was best at this level, and to allow for enough replication to statistically test for
326 patterns of variation, using ANOVA.

327 To determine host species effects on community composition and beta diversity,
328 variability in endophyte community composition was examined across all 351 leaf samples.
329 Bray-Curtis similarity measures were used to construct a community distance matrix based on
330 the observed endophyte communities within each sample. The variance in distance between
331 groups of interest was compared with a nonparametric null distribution based on 999
332 permutations of the observed data. The effect of host tree species on endophyte community
333 composition was tested on the complete dataset (n = 351 for fungi, 357 for bacteria). The
334 effect of individual host tree, and branch within each tree on endophyte community
335 composition was tested on a reduced similarity matrix composed of samples collected from
336 only 3 branches per tree (n = 327 for fungi, 331 for bacteria). These were evaluated with
337 permutational multivariate analysis of variance (PERMANOVA; Anderson 2013), using the
338 'adonis' function in the vegan package (Oksanen et al. 2013). To visualise the variation

explained by host tree individual, host tree species, and branch identity, partitioning of variance was conducted using the 'varpart' function in vegan.

To explore whether observed variance was common across all fungal endophytes, or whether it differed across OTUs within different orders PERMANOVA were performed on two subsets of the data; one containing only OTUs identified as belonging to the order Xylariales (well documented endophytes, see Rodriguez et al., 2009), the other containing OTUs identified to order Polyporales (order known to contain pathogens). Similarly, for bacteria, analysis was rerun on a subset of the data, comprising those OTUs identified to the order Rhizobiales (an order known to contain plant associated bacteria) and another data set comprised of OTUs identified to the order Solirubrobacterales (an order known to contain generalist bacteria).

To determine the influence of phylogenetic and geographic distance of the host tree species on the endophytic communities, OTU reads for each sample from the same tree were summed, reducing the full datasets for both fungi and bacteria down to 38 samples. DNA read counts were normalised across all trees so as to have a balanced dataset. Bray-Curtis similarity of each tree endophyte community was determined and compiled in a community similarity matrix. A phylogenetic tree representing all of the species sampled here was prepared using the ape package in R (Fig. 3; Paradis et al. 2004), by pruning a species-level phylogenetic hypothesis published elsewhere (Baraloto et al. 2012). Terminal nodes representing members of the same species, i.e. where more than one tree was sampled, were included as polytomies with an intraspecific edge distance of 0.5 between all, so as to be very low in comparison to the interspecific edge distances elsewhere in the phylogeny. The resulting edge distances were then used to prepare a phylogenetic distance matrix.

A second distance matrix was prepared comparing the geographic distance between each of the 38 trees across the 4-ha sampling plot in metres. Both matrices were directly compared with the endophyte community similarity matrices by performing partial Mantel tests based on Pearson's product-moment correlation, using the function mantel.partial in the ecodist package (Goslee & Urban 2007). In each case, the effect of one distance metric was controlled for when determining the effect of the other.

In order to determine whether the presence of OTUs within the leaves of a host tree represented true host specificity, and to avoid issues associated with low sample size (e.g.

Cordier et al. 2012, Novotny et al. 2002), a metric was calculated to determine the probability that OTU presence within a single host species occurred by chance. Using the bipartite package (Dormann et al. 2008), the degree of each OTU (i.e. the number of leaf samples within which they occur) was calculated. Host tree name was replaced by a numerical value (1-22 for each sampled host), before a random sample of host tree for each of these degrees was determined. The standard deviation of the randomly generated host tree numbers was then calculated. Each time this equalled zero (i.e. the degrees were sampled as the same species), a count of 1 was assigned. This process was reiterated 1000 times, with the count tally of iterations divided by 1000. The calculated metric thus ranged from 0-1, with 0 representing no chance that an OTU occurred in sampled host by chance, and 1 representing certainty that OTUs occurred in sampled host by chance. For OTUs with a degree of 1, such a calculation was not possible. To determine the importance of rare OTUs in generating patterns detected during analysis, the percentage of all OTUs only occurring in each host tree was calculated.

Results

Endophyte α-diversity

Average OTU richness per leaf was calculated to be 168 for fungi, ranging from 15 to 1018, and 46.3 for bacteria, ranging from 1 to 504. Fungal OTU identification rate was 45%, 36%, and 19% at the order, family, and genus level, respectively. Bacterial OTU identification was 98%, 88%, and 59% at the order, family, and genus level, respectively. OTU accumulation curves indicated that sampling was representative of the total endophyte community for fungi, but to a lesser extent for bacteria (S.I.4). Sequencing depth varied across samples (see S.I. 5), justifying standardisation approaches when calculating alpha diversity metrics.

Per-leaf endophyte diversity was dependent on the host tree with above average numbers of OTUs only occurring in specific host trees (Fig. 2). Significant differences in host tree Shannon diversity indices were detected for both fungi (ANOVA $F_{21,335}= 4.42$, $p = <0.001$), and to a lesser extent for bacteria (ANOVA $F_{21,336}= 1.82$ $p = 0.016$. Tukey HSD revealed that leaves collected from Carapa surinamensis, Dussia discolor, Inga huberi and Virola kwatae had significantly higher diversity indices than those collected from other trees for fungi, whilst for bacteria, only Virola kwatae was found to be significantly higher in

diversity than Eperua rubiginosa, with all other host species not differing significantly in their bacterial diversity (Fig. 2).

The relative occurrence of certain fungal orders (using OTU occurrence aggregated for each order) differed significantly from one host tree to the next (Fig. 3; S.I. 6).

Host-tree effects on β-diversity

When examining the effects of host tree identity on leaf endophyte community composition, PERMANOVA revealed the significant contribution of both species (Fungi: $F = 2.65$, $p = 0.001$; Bacteria: $F = 1.97$, $p = 0.001$) and nested individual trees (Fungi: $F = 1.58$, $p = 0.001$; Bacteria: $F = 1.3$, $p = 0.001$) in contributing to variation (Table 1).

In addition to the importance of individual host tree (Fungi: $F = 2.25$, $p = 0.001$; Bacteria: $F = 1.70$, $p = 0.001$), PERMANOVA also demonstrated a significant effect of nested branch (Fungi: $F = 1.08$, $p = 0.001$; Bacteria: $F = 1.12$, $p = 0.001$) in determining endophyte community composition (Table 2). Although host tree is a significant contributor, it only explains a part of the variation amongst communities, as revealed by the associated R2 values (Table 1). Partitioning of variance demonstrated the cumulative effects of species, host individual and branch only accounted for 12% of variation for fungi, and 10% of variation for bacteria (Fig. 4)

Endophyte order-specific effects on β-diversity

When determining the effects of host tree species on specific orders of fungi, PERMANOVA demonstrated a significant effect for Xylariales ($F = 1.78$, $p = 0.001$; Table 1) but not for Polyporales ($F = 1.01$, $p = 0.434$; Table 1). When determining the effects of host tree species on specific orders of bacteria, PERMANOVA demonstrated a significant effect for both Rhizobiales ($F = 1.92$, $p = 0.001$; Table 1) and Solirubrobacterales ($F = 1.5$, $p = 0.007$; Table 1).

Phylogenetic and geographic effects on β-diversity

Mantel tests revealed a significant correlation between endophyte community composition and phylogenetic distance of the host tree species for both fungi ($r = 0.1251$, $p = 0.044$), and bacteria ($r = 0.227$, $p = 0.009$). Both groups displayed no significant correlation between tree

endophyte community composition and geographic distance (Fungi: r = 0.091, p = 0.064; Bacteria: r = -0.062, p = 0.783).

OTU host specificity

Out of the 10158 detected fungal OTUs, 5542 occurred in only one host tree. Of these, 4982 only occurred in one leaf and as such host specificity could not be determined. For the remaining 560 OTUs, when accounting for the probability that their occurrence within a host was determined by chance with a cut off of 0.05, only 81 OTUs demonstrated true host specificity. This usually corresponded to OTUs occurring in 3 or more leaves of host trees of the same species, encompassing a broad taxonomy of fungal species (S. I. 7).

Out of the 1954 detected bacterial OTUs, 901 occurred in only one host tree. Of these, 737 only occurred in one leaf and as such host specificity could not be determined. For the remaining 164 OTUs, when accounting for the probability that their occurrence within a host was determined by chance with a cut off of 0.05, only 14 OTUs demonstrated true host specificity. This usually corresponded to OTUs occurring in 3 or more leaves of host trees of the same species, encompassing a broad taxonomy of bacterial species (S. I. 7).

Conversely, no single OTU occurred across all sampled leaves, and only 31 fungal OTUs and 3 bacterial OTUs occurred in more than half of the total leaf number.

Discussion

Tropical foliar endophytes represent an ideal model system to explore the factors shaping ecological communities. The high levels of both α- and β-diversity found in this study reveal the relative importance of host tree contributions to endophyte niche and dispersal limitation across spatial scales. In spite of acknowledged biases introduced by PCR methods and clustering thresholds (Dickie 2010, Lindner & Banik 2011), and a potentially low level of bacteria detection as a result of a primer which also amplified chloroplast and mitochondrial sequences, eDNA methods used here reveal levels of endophyte diversity an order of magnitude higher than culture-based methods, with OTU taxonomic identification and relative abundance consistent with previous studies of tropical canopy microbial communities (e.g. Izuno et al. 2016, Haruna et al. 2018).

Differences in endophyte diversity and community composition were detected at the host tree level for fungi. Given this trend was consistent with tree phylogenetic relatedness, it is likely that endophyte colonization success is dependent on shifts in leaf traits across host tree species, such as cuticle thickness or tannin levels (Arnold and Herre 2003). Carapa surinamensis, Dussia discolor, Inga huberi and Virola kwatae trees displayed higher levels of fungal endophyte diversity than other species present under similar conditions, and given the known role of certain tropical endophytes as symbionts (e.g. Arnold et al. 2003; Bae et al. 2009; Van Bael et al. 2009; Christian et al. 2019), such variation in the endophytes present could have functional consequences on host plant health for canopy leaves.

When observing the diversity of endophytes associated with host tree species at the order level, clear differences in the relative abundance of certain groups are apparent. Certain orders of endophytes differed significantly across different host species (e.g. Xylariales, for fungi; Rhizobiales for bacteria), whilst orders characterized by OTUs known to occur across a wide range of habitat types (e.g. Polyporales for fungi; Solirubrobacterales for bacteria) did not differ significantly in their occurrence among host tree species. Such a finding mirrors that of Peay et al. (2013), which showed that soil fungal communities with lineages containing pathogens and mycorrhizal fungi correlated with tree species richness of tropical forest plots, whilst saprotrophs did not. Nevertheless, at the 4-ha plot scale, although significant, host tree only explained a small proportion of the variation in endophyte community composition, suggesting the contribution of other factors.

Limitations in dispersal ability are known to have consequences in structuring spore dependent ecological communities (Peay et al. 2012). Here, distance between leaves was found to result in significant differences in endophyte community assembly, although this was scale dependent. When aggregating samples at the host tree individual level, no significant correlation was found between tree endophyte community composition and their proximity to one another within the 4-ha plot. At the scale of a single tree crown however, leaves positioned on the same branch contained endophyte communities more similar to one another than those found on an adjacent branch. This result could reflect differences in the age of leaves, given that leaves occurring on the same branch will be more similar in age, with consequences for leaf chemistry and community establishment. The result could also suggest an effect of dispersal limitation in determining endophyte community composition within a tropical forest upper canopy, albeit only detectable at scales of less than 10 m. Beyond this, in absence of the physical limitations to dispersal such as in the understory (e.g. Arnold et al.

2003, Izuno et al. 2016), of a strong environmental gradients (e.g. Suryanarayanan et al. 2002, Zimmerman & Vitousek 2012), and because of high rates of spore dispersal (e.g. Peay et al. 2012) tropical forest canopies are likely to display communities characterised by high turn-over. This contrasts with the findings of low species turnover at both local (10-100m) and regional scales (100km) in a tropical forest understory for fungal endophytes (Vincent et al. 2016). Here, authors used culturing to detect fungal OTUs and detected high numbers of generalist OTUs across their sampled host trees and sites. In comparison, our use of metabarcoding fails to detect high numbers of generalists, and instead reveals the importance of large numbers of rare OTUs.

Partitioning of the variance in endophyte community composition revealed that at the canopy scale, only a small part of total variation in endophyte assemblage could be explained by the summed effects of branch, host tree, and host species (12% for fungi, 10% for bacteria). Thus, although significant, each of these factors are limited in explaining the high levels of variation in the system. Instead, leaf endophyte assemblages are largely characterized by rare OTUs which occur sporadically within samples, the sum of which represents a large part of the total diversity within the canopy. Indeed, when accounting for the probability of chance in determining the presence of an OTU within a host tree, for both fungi and bacteria only 0.7% of detected OTUs could truly be considered host specific. Conversely, no OTUs were so generalist that they occurred across all samples, and only a small proportion occurred in more than half of all sampled leaves. In spite of the relatively small sample size here, this finding counters the idea of a ‘core microbiome’ for endophytes (e.g. Kembel et al. 2014), at least in the context of this study. Thus, the detected variation across host species is likely to be driven by the differences in the relative abundance of a small number of generalist OTUs and the numerous rare OTUs.

Due to their presence in well-defined sampling units, foliar endophytes represent an ideal example of community assembly within which to test assembly rules for betadiversity in tropical habitats. Communities associated with leaves collected in the upper canopy are known to be distinct from those found at lower levels (Izuno et al. 2016), but the findings presented here demonstrate the scale dependant factors contributing to their composition. When carefully controlling for community scale, and sampling intensity, community assembly becomes clearly non-neutral in leaf-associated fungi and bacteria, but not in opportunistic species which may tend to be rare and found across a range of samples (Magurran & Henderson 2003). Given the emerging knowledge surrounding their role as

symbionts, the community structure of endophytes within the canopy layer is likely to play an important role in the non-negligible exchanges of energy and matter which define this habitat.

Acknowledgements – The Nouragues Ecological Research Station is supported by CNRS, and is part of the Nouragues Natural Reserve. Fieldwork was supported by Patrick Châtelet and Marie-Loréa Tapie-Petit. Valuable contributions to the manuscript were provided by Lucie Zinger and other participants of the 2018 Metabarcoding School. We are grateful to the Genotoul bioinformatics platform (Toulouse, Midi-Pyrénées) for providing computing resources. This work was supported by a joint Agence Nationale de la Recherche-Swiss National Science Foundation (ANR-SNF) grant (SECIL, ref ANR-15-CE21-0016 and SNF N° 310030E-164289) and from an "Investissement d'Avenir" grants also managed by Agence Nationale de la Recherche (CEBA, ref. ANR-10-LABX-25-01; ANAEE-France: ANR-11-INBS-0001, TULIP: ANR-10-LABX-0041).

Author contributions – All authors contributed to manuscript preparation; US, LP, PG and JC conducted field work collection of samples; MR, US, AI and SM performed lab-based data preparation; JD, MR and JC performed data analysis.

Data Archive – Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.XXXXXX

References

Anderson, M. J. and Walsh, D. C. 2013. PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: what null hypothesis are you testing? – Ecological Monographs 83: 557-574.

Arnold, A. E. and Engelbrecht, B. M. 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. and Herre, E. A. 2003. Canopy cover and leaf age affect colonization by tropical fungal endophytes: ecological pattern and process in Theobroma cacao (Malvaceae). – Mycologia 95: 388-398.

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

554 Arnold, A. E. et al. 2000. Are tropical fungal endophytes hyperdiverse? – Ecology Letters 3:
555 267-274.

556 Arnold, A. E. et al. 2003. Fungal endophytes limit pathogen damage in a tropical tree. –
557 PNAS 100: 15649-15654.

558 Bae, H. et al. 2009. The beneficial endophyte Trichoderma hamatum isolate DIS 219b
559 promotes growth and delays the onset of the drought response in Theobroma cacao. – J. Exp.
560 Bot. 60: 3279-3295.

561 Baraloto, C. et al. 2012. Using functional traits and phylogenetic trees to examine the
562 assembly of tropical tree communities. – J. Ecology 100: 690-701.

563 Binladen J, et al. 2007. The use of coded PCR primers enables high-throughput sequencing of
564 multiple homolog amplification products by 454 parallel sequencing. PLoS ONE, 2: e197.

565 Boyer, F. et al. 2016. obitools: a unix-inspired software package for DNA metabarcoding. –
566 Mol Ecol Resour. 16: 176-182.

567 Carrell, A. and Carper, D. 2016. Subalpine conifers in different geo- graphical locations host
568 highly similar foliar bacterial endophyte communities. – FEMS Microbiol. Ecol. 92:fiw124

569 Carroll, G. 1988. Fungal endophytes in stems and leaves: from latent pathogen to mutualistic
570 symbiont. – Ecology 69: 2-9.

571 Chase, J. M. et al. 2005. Competing theories for competitive metacommunities. – In:
572 Holyoak, M., Leibold, M. A. and Holt, R. D. (eds), Metacommunities: spatial dynamics and
573 ecological communities. University of Chicago Press, pp. 335-354.

574 Christian, N. et al. 2019. Foliar endophytic fungi alter patterns of nitrogen uptake and
575 distribution in Theobroma cacao. – New Phytol. https://doi.org/10.1111/nph.15693

576 Coissac, E. et al. 2012. Bioinformatic challenges for DNA metabarcoding of plants and
577 animals. – Mol Ecol. 21: 1834-1847.

578 Condit, R. et al. 2002. Beta-diversity in tropical forest trees. – Science. 295: 666-669.

579 Cordier, T. et al. 2012. The composition of phyllosphere fungal assemblages of European
580 beech (Fagus sylvatica) varies significantly along an elevation gradient. – New Phytol. 196:
581 510-519.

582 Courtois, E. A. et al. 2016. Evolutionary patterns of volatile terpene emissions across 202
583 tropical tree species. – Ecol. Evol 6: 2854-2864.

584 Crowther, T. W. et al. 2018. The use of artificial media in fungal ecology. – Fung. Ecol. 32:
585 87-91.

586 Del Olmo-Ruiz, M. and Arnold, A. E. 2017. Community structure of fern-affiliated
587 endophytes in three neotropical forests. – J Trop Ecol. 33: 60-73.

588 Dickie, I. A. 2010. Insidious effects of sequencing errors on perceived diversity in molecular
589 surveys. – New Phytol. 188: 916-918.

590 Dormann, C. F. et al. 2008. Introducing the bipartite package: analysing ecological
591 networks. – Interaction, 8: 8-11.

592 Ellwood, M. D. F. et al. 2009. Stochastic and deterministic processes jointly structure tropical
593 arthropod communities. – Ecol Lett. 12: 277-84.

594 Epp, L. S. et al. 2012. New environmental metabarcodes for analysing soil DNA: potential for
595 studying past and present ecosystems. – Mol. Ecol. 21: 1821-1833.

596 Ficetola, G. F. et al. 2010. An in silico approach for the evaluation of DNA barcodes. – BMC
597 genomics. 11: 434.

598 Fine, P. V. et al. 2004. Herbivores promote habitat specialization by trees in Amazonian
599 forests. – Science. 305: 663-665.

600 Fliegerova, K. et al. 2014. Effect of DNA extraction and sample preservation method on
601 rumen bacterial population. – Anaerobe, 29: 80-84.

602 Gilbert, G. S. and Reynolds, D. R. 2005. Nocturnal fungi: Airborne spores in the canopy and
603 understory of a tropical rain forest. – Biotropica. 37: 462-464.

604 Goslee, S. C. and Urban, D. L. 2007. The ecodist package for dissimilarity-based analysis of
605 ecological data. – J. Stat. Softw. 22: 1-19.

606 Griffin, E. A. and Carson, W. P. 2015. The ecology and natural history of foliar bacteria with
607 a focus on tropical forests and agroecosystems. – Bot Rev. 81: 105-149.

608 Griffin, E. A. & Carson, W. P. 2018. Tree Endophytes: Cryptic Drivers of Tropical Forest
609 Diversity. – In: Pirttilä, A. M. and Frank, C. (eds), Endophytes of Forest Trees: Biology and
610 Applications. Springer, pp. 63-103.

611 Haruna, E. et al. 2018. Extensive overlap of tropical rainforest bacterial endophytes between
612 soil, plant parts, and plant species. – Microb ecol. 75: 88-103.

613 Herre, E.A. et al. 2007. Ecological implications of anti-pathogen effects of tropical fungal
614 endophytes and mycorrhizae. – Ecology. 3: 550-558.

615 Hubbell, S.P. 2001. The unified neutral theory of species abundance and diversity. –
616 Princeton Univ. Press.

617 Hyde, K. and Soytong, K. 2008. The fungal endophyte dilemma. – Fungal Divers. 33: e173.

618 Izuno, A. et al. 2016. Vertical Structure of Phyllosphere Fungal Communities in a Tropical
619 Forest in Thailand Uncovered by High-Throughput Sequencing. – PloS one. 11: e0166669.

620 Jost, L. 2006. Entropy and diversity. – Oikos. 113: 363-375.

621 Kembel, S. W. et al. 2014. Relationships between phyllosphere bacterial communities and
622 plant functional traits in a neotropical forest. PNAS. 111: 13715-13720.

623 Leibold, M. A. 1995. The niche concept revisited: mechanistic models and community
624 context. – Ecology. 76: 1371-1382.

625 Lindner, D. L. and Banik, M. T. 2011. Intragenomic variation in the ITS rDNA region
626 obscures phylogenetic relationships and inflates estimates of operational taxonomic units in
627 genus Laetiporus. – Mycologia. 103: 731-740.

628 Lodge, D. J. 1997. Factors related to diversity of decomposer fungi in tropical forests. –
629 Biodivers. Conserv. 6: 681-688.

630 Magurran, A. E. and Henderson, P. A. 2003. Explaining the excess of rare species in natural
631 species abundance distributions. – Nature. 422: 714-716.

632 Maynard, D. S. et al. 2017. Fungal interactions reduce carbon use efficiency. – Ecol Lett. 20:
633 1034-1042.

634 Mejía, L.C. et al. 2014. Pervasive effects of a dominant foliar endophytic fungus on host
635 genetic and phenotypic expression in a tropical tree. – Front. Microbiol. 5: 479.

636 Miyambo T et al. 2016. Plants of the fynbos biome harbour host species-specific bacterial
637 communities. – FEMS Microbiol. Lett. 363:fnw122.

638 Nilsson, R.H., et al. 2008. Intraspecific ITS variability in the kingdom Fungi as expressed in
639 the international sequence databases and its implications for molecular species identification.
640 – Evolutionary bioinformatics. 4: EBO-S653.

641 Novotny, V. et al. 2002. Low host specificity of herbivorous insects in a tropical forest. –
642 Nature 416: 841–844.

643 Oksanen, J. et al. 2013. Package 'vegan', Community ecology package, version, vol. 2, no. 9.

644 Paradis, E. et al. 2004. APE: analyses of phylogenetics and evolution in R language. –
645 Bioinformatics. 20: 289-290.

646 Peay, K. G. et al. 2013. Strong coupling of plant and fungal community structure across
647 western Amazonian rainforests. – ISME. 7: no. 9, pp. 1852.

648 Peay, K. G. et al. 2012. Measuring ectomycorrhizal fungal dispersal: macroecological patterns
649 driven by microscopic propagules. – Mol Ecol. 21: 4122-4136.

650 Porras-Alfaro, A. and Bayman, P. 2011. Hidden Fungi, Emergent Properties: Endophytes and
651 Microbiomes. – Annu. Rev. Phytopathol. 49:291-315.

652 Quast, C. et al. 2012. The SILVA ribosomal RNA gene database project: improved data
653 processing and web-based tools. – Nucleic Acids Res. 41: 590-596.

654 R Core Team. 2013. R: A language and environment for statistical computing, Vienna,
655 Austria.

656 Réjou-Méchain, M. et al. 2015. Using repeated small-footprint LiDAR acquisitions to infer
657 spatial and temporal variations of a high-biomass Neotropical forest. – Remote Sens Environ.
658 169: 93-101.

659 Ricklefs, R. E. 2004. A comprehensive framework for global patterns in biodiversity. – Ecol
660 Lett. 7: 1-15.

661 Rodriguez, R.J. et al. 2009. Fungal endophytes: diversity and functional roles. – New phytol.
662 182: 314–330.

663 Sabatier, D. 1990. Variations du peuplement forestier à l'échelle stationnelle: le cas de la
664 station des Nouragues en Guyane Francaise. – Actes de l'Atelier MAB-UNESCO sur
665 l'Aménagement et la Conservation de l'Ecosystème Forestier Humide.

666 Suryanarayanan, T. et al. 2002. Occurrence and distribution of fungal endophytes in tropical
667 forests across a rainfall gradient. – Can J Bot. 80: 818-826.

668 Suryanarayanan, T. et al. 2003. Endophytic fungal communities in leaves of tropical forest
669 trees: diversity and distribution patterns. – Curr. Sci. 85: 489-493.

670 Taberlet, P. et al. 2018. Environmental DNA: For Biodiversity Research and Monitoring. –
671 Oxford University Press.

672 Truong, C. et al. 2017. How to know the fungi: combining field inventories and
673 DNA-barcoding to document fungal diversity. - New Phytol. 214: 913-919.

674 U'Ren, J. M. et al. 2012. Host and geographic structure of endophytic and endolichenic fungi
675 at a continental scale. – Am. J. Bot. 99: 898-914.

676 Van Bael, S. et al. 2017. Foliar endophyte communities and leaf traits in tropical trees. In:
677 The fungal community: its organization and role in the ecosystem. CRC Press, Boca Raton,
678 pp.79-94.

679 van der Meer, P. J. and Bongers, F. 1996. Patterns of tree-fall and branch-fall in a tropical rain
680 forest in French Guiana. – J. Ecology 84:19-29.

681 van der Meer, J. H. 1972. Niche theory. – Annu. Rev. Ecol. Evol. Syst. 3:107-132.

682 Vincent, J. B. et al. 2016. Host associations and beta diversity of fungal endophyte
683 communities in New Guinea rainforest trees. – Mol. Ecol. 25: 825-841.

684 Zimmerman, N. B. and Vitousek, P. M. 2012. Fungal endophyte communities reflect
685 environmental structuring across a Hawaiian landscape. – PNAS. 109: 13022-13027.

686 Zinger, L. et al. 2016. Extracellular DNA extraction is a fast, cheap and reliable alternative for
687 multi-taxa surveys based on soil DNA. – Soil Biol. Biochem. 96: 16-1

Table 1: PERMANOVA analysis to determine the effect of host tree species and nested host tree individual in determining fungal and bacterial endophyte community composition. Analysis was performed on community dissimilarity matrixes constructed from six separate datasets; one containing all fungal OTUs detected in the analysis of 351 leaves; one containing only OTUs identified as Xylariales; one containing only OTUs identified as Polyporales; one containing all bacterial OTUs detected in the analysis of 358 leaves; one containing only OTUs identified as Rhizobiales and one containing only OTUs identified as Solirubrobacterales.

	Fungi	351 leaf data				Xylariales data				Polyporales data				Bacteria	358 leaf data				Rhizobiales data				Solirubrobacterales			
		df	F	R2	p	df	F	R2	p	df	F	R2	p		df	F	R2	p	df	F	R2	p	df	F	R2	p
Sp		21	2.65	0.14	0.001	21	1.78	0.11	0.001	21	1.01	0.12	0.434		21	1.97	0.11	0.001	21	1.92	0.18	0.001	15	1.5	0.35	0.007
Sp: Tree		16	1.58	0.06	0.001	16	1.32	0.06	0.001	16	0.91	0.09	0.855		16	1.3	0.05	0.001	15	1.28	0.09	0.001	9	1.09	0.15	0.336
Residuals		313		0.79		287		0.83		134		0.79			320		0.83		166		0.73		31		0.5	
Total		350		1		324		1		171		1			357		1		202		1		55		1	

Table 2: PERMANOVA analysis to determine the effect of host tree individual and nested branch in determining endophyte community composition. Analysis was performed on community dissimilarity matrixes constructed from a dataset comprised of 327 leaves for fungi, 331 for bacteria.

	Fungi				Bacteria			
	df	F	R2	p	df	F	R2	p
Tree	37	2.25	0.22	.001	37	1.70	0.17	.001
Tree: Branch	75	1.08	0.21	.001	75	1.12	0.23	.001
Residuals	214		0.57		218		0.60	
Total	326		1		330		1	

Figure 1: The Canopy Operational Permanent Access (COPAS) system, allowing access for unrestricted access to the upper canopy. This was used to sample leaves from which endophyte communities were extracted.

Figure 2: The effective Shannon diversity index of each sampled tree species, based on the of OTUs within each leaf sample. The fungal dataset is based on 351 leaves and the bacterial dataset is based on 357 leaves. Trees are arranged by phylogenetic relatedness.

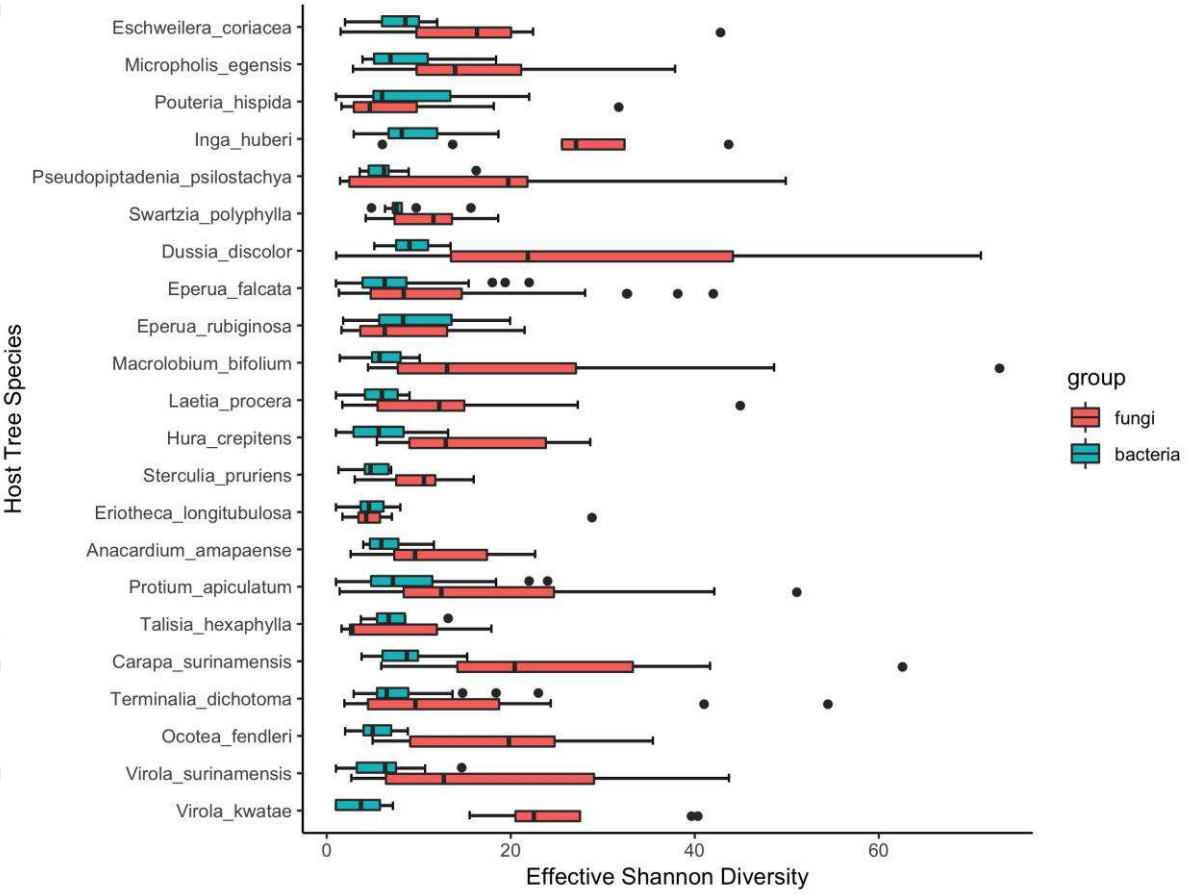

Figure 3: The relative occurrence of identified fungal OTUs in host trees when classed by order for a) fungi and b) bacteria. Non-identified OTUs are not included here, but represent on average 50% of all detected fungal OTUs for each sample. Host tree species are arranged by order of phylogenetic relatedness, modified from Baraloto et al. (2012). Orders are arranged in order to represent their decreasing relative occurrence from left to right, indicated by a shift from red to yellow.

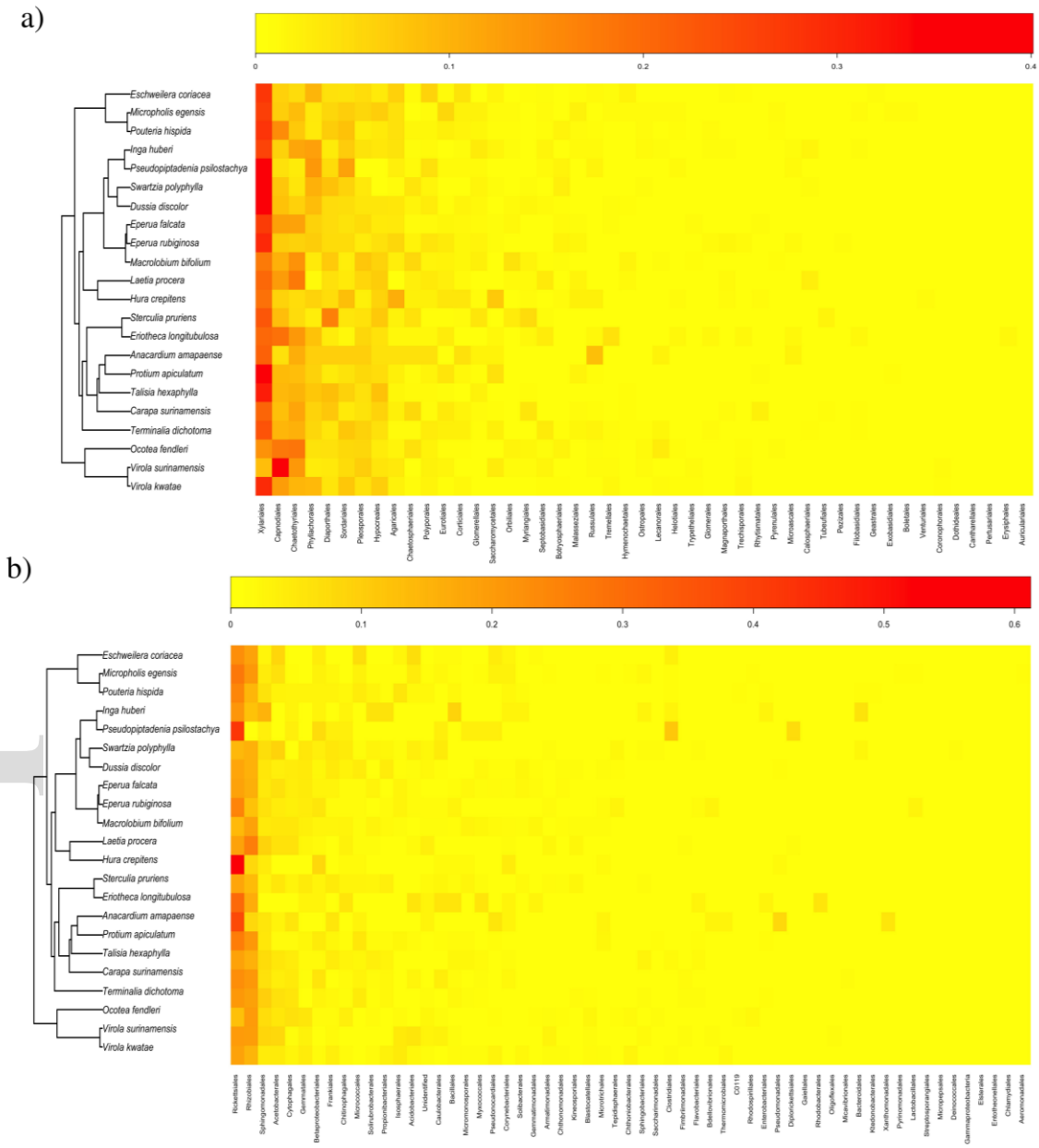

Figure 4: Partitioning of variance of endophyte community composition of leaves explained by host individual, host species and host branch for a) fungi and b) bacteria. Residuals indicate the amount of total variation unexplained by these factors.