UNDERSTANDING THE ROLES OF INTERKINGDOM MICROBIAL INTERACTIONS, MICROBIAL TRAITS, AND HOST FACTORS IN THE ASSEMBLY OF PLANT MICROBIOMES

By

Julian Aaron Liber

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Plant Biology – Master of Science

2021

PUBLIC ABSTRACT

UNDERSTANDING THE ROLES OF INTERKINGDOM MICROBIAL INTERACTIONS, MICROBIAL TRAITS, AND HOST FACTORS IN THE ASSEMBLY OF PLANT MICROBIOMES

By

Julian Aaron Liber

Multicellular life forms, from humans to plants, are not singular organisms but instead consist of a community of microbes on and often within a host. These microbes, such as bacteria and fungi, interact with the host, in mutually beneficial interactions as well as parasitic relationships. Each member of this community has traits, which include the types of nutrition it can consume, the metabolites it excretes, and how it grows in response to environmental conditions. I studied microbial communities of plants, which interact with the host via nutrition, signaling, disease, or immunity. Because an organism's community is important in the health of the host, understanding what matters in the formation of the microbial community could help reduce disease and increase resilience for the host. First, I examine how the interactions of bacteria and fungi alone and in plant roots contribute to host traits, and how viral disease of the host changes the communities of microbes infecting roots. I present a tool developed to assign meaningful names to organisms detected in complex communities to understand their traits and communicate about the characteristics of a given community. Finally, I use this tool and report on an observational experiment which identifies important factors contributing to the makeup of fungal communities in leaves, leaf litter, and soil to forest fungal communities, and to improve methods for sampling these types of communities. The complex nature of microbial communities provides challenges for understanding how they work, but I attempt to bridge scales, from simple to complex, to describe how microbial communities assemble and function.

ABSTRACT

UNDERSTANDING THE ROLES OF INTERKINGDOM MICROBIAL INTERACTIONS, MICROBIAL TRAITS, AND HOST FACTORS IN THE ASSEMBLY OF PLANT MICROBIOMES

By

Julian Aaron Liber

The community of organisms that associate with plants are vital to both the survival of the host plant but also the diseases which may kill it. The processes by which this community, called the microbiome, assemble and function can contribute to the traits of the host, including plants that humans rely on for food, resources, and ecosystems services. This thesis focuses on understanding the assembly of microbiomes at the scale of microbe-microbe interactions and traits of individual microbes, as well as how characters of the host may change this process. I first address this by examining the *in vitro* and *in planta* interactions within small synthetic communities of root-inhabiting bacteria and fungi and with the plant host and viral disease of the host. While intermicrobial interactions *in vitro* were not predictive of *in planta* interactions, adding host disease or additional organisms to the system altered the assembly process. I then show the development and applications of the CONSTAX2 classifier, a taxonomic assignment tool for metabarcoding studies, which offers improved accuracy and ease of use for conducting metabarcoding studies exploring the diversity and structure of microbial communities. Last, I present a study testing which factors affected the composition of forest fungal communities to understand the ecology of litter-inhabiting fungi and improve methodologies for sampling leaf-associated fungal communities. The factors affecting the assembly of plant microbiomes are complex and varied but connecting individual interactions to community composition and ultimately function may improve our abilities to predict and manage microbiome processes.

To my love, Molly. You taught me to find resilience, determination, and joy to bring a better tomorrow for all.

ACKNOWLEDGEMENTS

I thank my ever helpful and enthusiastic advisor Dr. Gregory Bonito, as well as Dr. Frances Trail and Dr. Marjorie Weber on my committee, for facilitating and supporting my research and professional goals during this degree with feedback and encouragement.

The projects submitted as parts of this thesis were contributed to by the hard work of many collaborators over the past few years and were enhanced by the expertise and support of friends, family, and colleagues. Chapter 1 started as a project of Michigan State's 2018 iGEM Team, the "Endophyte Club", including Erin Uhelski, Jordan Lee, Sophia Viola, Casper Gate, Ashley Del Rose, Jessica Schultz, and Sarah Caldwell, and advised by Dr. Björn Hamberger, Dr. Michaela TerAvest, Dr. Tim Whitehead, Dr. Danny Ducat, and Davis Mathieu each of whom contributed to the development, execution, outreach, and presentation of our project. Since then, Natalie Golematis, Morgane Chretien, Ellen Cole, Kota Nakasato, Dr. Carolyn Malmstrom, Dr. Natalie Vande Pol, Dr. Pedro Beschoren da Costa, Dr. Zachary Noel, Dr. Gian Maria Niccolò Benucci, Bryan Rennick, Judson Van Wyk, Reid Longley, Ashley Barlow, and Dr. Gregory Bonito have contributed to the experimental design, lab work, plant maintenance, data collection, analysis, and understanding of this work. This work would have not been possible without the kind efforts of Dr. Alessandro Desirò, Dr. Anne-Sophie Bohrer, Dr. Katerina Lay-Pruitt, Dr. Eva Farré, Dr. Zhi-Yan Du, Abigail Bryson, Izzie Gall, and those mentioned earlier to train me in laboratory skills and techniques.

Chapter 2 was an extension of the original CONSTAX classifier, which Dr. Natalie Vande Pol helped transition over to the new version. I also thank Dr. Zachary Noel, Reid Longley, Dr. Acer VanWallendael, and Shay Shemanski for helping test the software. It was co-

authored by Dr. Gregory Bonito and Dr. Gian Maria Niccolò Benucci. Chapter 3 originated as a project for Dr. Gregory Bonito's Advanced Mycology course, and was co-authored by Douglas Minier, Anna Stouffer-Hopkins, Judson Van Wyk, Reid Longley, and Dr. Gregory Bonito. I acknowledge financial support for research and teaching from US National Science Foundation DEB 1737898, US Department of Agriculture NIFA MICL02416, and the departments of Plant Biology and Plant, Soil, and Microbial Sciences. Beyond research help, I appreciate the great instructors I have had at my time at MSU, the friends from lab, my cohort, and elsewhere who have been supportive and enjoyable, my parents, grandparents, and siblings for their constant encouragement, and for Molly McSween, who has always been there with love, support, kindness, and friendship.

TABLE OF CONTENTS

LIST OF TABLES viii

LIST OF FIGURES ix

LIST OF ALGORITHMS xi

KEY TO ABBREVIATIONS xii

INTRODUCTION 1
Research Goals 2
Interactions of plant host, virus, fungi, and bacteria 2
The CONSTAX2 classifier 2
Leaf litter communities 3

CHAPTER 1: FOUR-WAY INTERACTIONS OF HOST, VIRUS, FUNGI, AND BACTERIA IN THE PLANT MICROBIOME 5
Abstract 5
Introduction 6
Methods 9
Isolation of endophytic bacteria and fungi 9
Transformation of bacterial endophyte with GFP 10
Volatile interactions of bacterial and fungal endophytes 11
Growth of bacteria, fungi, and Brachypodium in axenic plate cultures 11
Quantification of endophyte infection of roots 12
Infection of Brachypodium with endophytes and BYDV-PAV-MI 13
Assessment of plant traits and endophyte infection 15
Microscopy to evaluate endophyte presence in roots 16
Results 17
In-vitro bacterial-fungal interactions 17
Plant biomass and root morphology 17
Bacterial infection 19
Effect of viral infection on plant traits and endophyte communities 22
Discussion 27
Inter-Kingdom microbial and endophyte-host interactions 27
Plant trait changes from endophyte and virus inoculation 29
Root endophyte presence in response to viral inoculation 30
Implications for understanding viral plant disease 31
Limitations and future directions 32
Conclusion 33

CHAPTER 2: CONSTAX2 – IMPROVED TAXONOMIC CLASSIFICATION OF ENVIRONMENTAL DNA MARKERS 34

Abstract 34
Introduction 34
Methods 35
Data and code availability 35
CONSTAX2 algorithm 35
Clade partition cross-validation 36
Classification Counts 37
Algorithm speed 37
Definition of classification metrics 38
Plotting and analysis 39
Results 40
Algorithm speed and memory usage 40
Algorithm performance 41
Implementation 47
Conclusion 49

CHAPTER 3: LEAF LITTER FUNGAL COMMUNITIES REFLECT PRE-SENESCENT LEAF COMMUNITIES IN A TEMPERATE FOREST ECOSYSTEM 50
Abstract 50
Introduction 51
Methods 53
Site 53
Sample collection 55
Sample preparation and DNA extraction 56
Sequence analysis 57
Ecological analysis 58
Swab material comparisons 59
Results 60
Samples and sequencing 60
Comparison of substrates, host species, and site 61
Comparison of swab materials 72
Discussion 75
Comparing the phyllospheres of maple and hickory 76
Fungal communities in maple and hickory leaf litter 77
No differences found between swab types on epiphytic fungal diversity indices 78
Limitations and future directions 79
Conclusion 81

CONCLUSION 82

REFERENCES 83

LIST OF TABLES

Table 1.1 Fungal endophyte strains used in interaction experiments. 10

Table 1.2 ANOVA tables of the observed effects of endophyte and virus inoculants and the soil depth covariate on plant traits. 24

Table 2.1 Classification performance of each classifier, for each database, region, and partition level. 43

Table 3.1 Distances between sample sites. 55

Table 3.2 PERMANOVA tables of substrate, host species, site, and interactions effects on fungal communities both all factor and pairwise. 62

Table 3.3 ANOVA tables of substrate, host species, and site effects of within-sample diversity. 64

Table 3.4 Substrate and cross-substrate indicator OTUs. 65

Table 3.5 Indicator taxa for combined substrate and host. 67

Table 3.6 Total, unique, and shared OTUs by substrate. 69

LIST OF FIGURES

Figure 1.1 Volatile interactions of bacteria on fungal colony radii. 17

Figure 1.2 Dry root and shoot biomass of Brachypodium seedlings. 18

Figure 1.3 Root hair zone diameter in response to bacterial inoculation. 19

Figure 1.4 qPCR quantification of *E. ludwigii* FCP2-01 bacterial infection of plant roots by fungal co-inoculant. 20

Figure 1.5 *R. oryzae* MC20 hyphae and *E. ludwigii* FCP2-01 in Brachypodium roots. 21

Figure 1.6 *R. oryzae* MC20 and *E. ludwigii* FCP2-01 on surface of Brachypodium root. 22

Figure 1.7 Trait values of Brachypodium in response to inoculation with endophytes and/or virus 25

Figure 1.8 Principal components analysis ordination of plant traits. 26

Figure 1.9 Relative abundances of bacterial and fungal endophytes determined by qPCR in roots of Brachypodium. 27

Figure 2.1 Performance of the CONSTAX algorithm. 42

Figure 2.2 Effects of max hits and confidence threshold parameters on UNITE classification. .. 45

Figure 2.3 Effect of max hits and confidence threshold parameters on SILVA classification. 46

Figure 2.4 Classification counts for each classifier and the CONSTAX classification. 47

Figure 3.1 Soil types in the sampling quadrat. 54

Figure 3.2 Rarefaction curves. 61

Figure 3.3 Within-sample diversity of samples by substrate and host species. 63

Figure 3.4 Heatmap of OTU abundance by substrate. 66

Figure 3.5 Pearson correlation heatmap of samples. 68

Figure 3.6 Venn diagram of shared OTUs by substrate. 70

Figure 3.7 Comparison of community composition by substrate, host species, site, and swab material. 71

Figure 3.8 Read counts by swab material. 74

Figure 3.9 Richness and within-sample diversity of fungal communities by swab material. 75

LIST OF ALGORITHMS

Algorithm 2.1 Sensitivity 38

Algorithm 2.2 Misclassification rate 38

Algorithm 2.3 Over-classification rate 38

Algorithm 2.4 Errors per query 38

Algorithm 2.5 Binomial model of classification errors 39

KEY TO ABBREVIATIONS

ANOVA - Analysis of Variance

BLAST - Basic Local Alignment Search Tool

BLASTn - BLAST nucleotide

BSA - Bovine Serum Albumin

BYDV - Barley Yellow Dwarf Virus

CB - CONSTAX2 with BLAST

CBC - CONSTAX2 with BLAST and Conservative rule

CI - Confidence Interval

CLSM – Confocal Laser-Scanning Microscopy/Microscope

CONSTAX - Consensus Taxonomy classifier

CU - CONSTAX2 with UTAX

CUC - CONSTAX2 with UTAX and Conservative rule

DNA - Deoxyribose Nucleic Acid

EPQ - Errors per Query

ES - Extraction Solution

FDR - False Discovery Rate

GFP - Green Fluorescent Protein

HSD - Honest Significant Difference

ITS - Internal Transcribed Spacer

LB - Lysogeny Broth

MC - Misclassification

NCBI - National Center for Biotechnological Information

NMDS - Non-metric Multidimensional Scaling

OC - Over-classification

OTU - Operation Taxonomic Unit

PAR - Photosynthetically Active Radiation

PCR - Polymerase Chain Reaction

PERMANOVA - Permutational Analysis of Variance

PERMDISP – Permutational test of multivariate Dispersions

qPCR - quantitative PCR

RAM – Random-Access Memory

rDNA - ribosomal DNA

RDP - Ribosomal Database Project

SRA - Sequence Read Archive

SSU – Small ribosomal Subunit

USDA - United States Department of Agriculture

UV - Ultraviolet

WSL - Windows Subsystem for Linux

INTRODUCTION

What makes a plant a plant? While some plants can survive alone, without a community, most require an assemblage of diverse microbes and other organisms to exists in the environments that they inhabit (Vandenkoornhuyse *et al.*, 2015). Members of this community interact with the host plant – in mutualistic and parasitic ecologies (Martin *et al.*, 2017; Brader *et al.*, 2017), but also interact with the other microbes (Regalado *et al.*, 2020) and components of the community (Bray and Wickings, 2019). As we seek to understand the assembly, diversity, and function of plant microbiomes, we can take a reductionistic approach – microbial traits, pairwise interactions, culturable representatives – or an ecosystem approach – whole community inoculation, manipulating large taxonomic groupings, and/or culture-independent analyses. Each of these methodologies has benefits and drawbacks. Reductionistic approaches fail to capture emergent behavior of large communities, or the diversity of traits and combinations of traits which each organism may possess but are generally more tractable and can establish ecological rules for how microbes interact with their host. The ecosystem approach can struggle to apply predicable rules to complex systems with poorly understood components yet offers demonstrations of methodologies in realistic settings and observes organisms how they typically exist. In this thesis I work to study both sides of microbial ecology: connecting the interactions of a small number of components (Chapter 1) to occurrence patterns in real communities (Chapter 3), whereby taxonomic assignment (Chapter 2) can provide the insights of individual traits at community scales. As plant microbiome science develops, studies will have to bridge scales to be generalizable, make specific and falsifiable predictions, and translate fundamental principles to useful applications.

Research Goals

Interactions of plant host, virus, fungi, and bacteria

In Chapter 1, I set out to understand the mechanism of how endophytic bacteria and fungi colonize plant roots, and how interactions between these microbes and the plant host influence and are influenced by the infection process. I later included a plant pathogenic virus in the bacteria, fungi, and plant interaction system which allowed for investigating the potential of endophytes in biocontrol of viral disease and for observing how changes in the host's physiology (nutrition and immunity statuses) reflected on its microbiome. I approached these experiments with the following hypotheses: 1) pairwise *in vitro* interactions of microbes would predict their interactions in plant roots, 2) microbial inoculants would have impacts on plant traits, 3) co-inoculation of bacteria and fungi would alter their abundance compared to single inoculation via facilitation or competition, and 4) viral disease would alter plant traits and the abundances of root endophytes.

The CONSTAX2 classifier

Chapter 2 focuses on the development and validation of the CONSTAX2 classifier, a bioinformatics tool designed to assist in assigning accurate and meaningful taxonomic labels to DNA sequences obtained from metabarcoding studies. These types of studies sequence a community of organism by amplifying a single genetic marker, often part of the rDNA operon, and seek to understand the processes and factors which contribute to the structure and function of microbial communities. The classification of operational taxonomic units (OTUs), a stand-in for species in the context of metabarcoding surveys, facilitates the assignment of functional attributes, such as trophic guild in fungi (Nguyen et al., 2016). We can understand a taxon name and relate it to known characters about that group from the literature. While it is no substitute for

functional analysis with metatranscriptomics or other meta-omics approaches (Santoferrara *et al.*, 2020; Elferink *et al.*, 2020), taxonomic assignment offers ways to communicate about and understand more deeply community analysis than simply patterns in the occurrence of OTUs.

While several tools already exist to perform taxonomic assignment in metabarcoding studies, CONSTAX2 was designed with the following goals: 1) provide a simple to install, well documented, and easy to customize tool with non-legacy software dependencies, 2) improve accuracy to be equal to or better than existing tools, 3) allow for use with any desired group of organisms, and 4) maintain useable time and memory requirements. CONSTAX2 was applied to describe the communities of fungi associated with maple and hickory leaf, litter, and soil communities in Chapter 3.

Leaf litter communities

Chapter 3 developed as an education project to learn the tools and approaches for using metabarcoding in community ecology. Our original plan in the project was to trace the associations of fungal communities on (epiphytes) and in (endophytes) leaves, fallen leaf litter, and soil in a temperate forest to assess the sources of litter fungal communities. Litter decomposition is an important process for nutrient cycling (Parton et al., 2007) , and the assembly of this community is likely important to its function. However, due to the single timepoint sampling and lack of functional characterization of community members, we refocused on examining the factors contributing to the structure of these communities. The samples were derived from two tree host species, from five sites within a forest plot, and from the four substrates or compartments: epiphytes, endophytes, litter, and soil. Additionally, we tested two swab types for epiphyte community sampling to improve the cost effectiveness of sampling. Our hypotheses for this experiment were: 1) the factors of host species, site, and substrate contribute

to the structure of fungal communities, 2) epiphyte and endophyte communities would be more similar to litter communities than soil, and 3) swab types would have differential performance. Our efforts to characterize these fungal communities adds to our understanding of the distribution and traits of the organisms detected, suggests mechanisms of dispersal across substrates, and provides evidence for the effectiveness of sampling methodologies.

CHAPTER 1: FOUR-WAY INTERACTIONS OF HOST, VIRUS, FUNGI, AND BACTERIA IN THE PLANT MICROBIOME

Abstract

Complex, multi-domain communities exist in, on, and around plants. We aimed to understand how intermicrobial and host-microbe interactions lead to the assembly of a plant microbiome to better create beneficial communities capable of maintaining health under biotic and abiotic stresses. Bacteria and fungi are common plant endophytes, organisms which grow within plant tissues without causing notable disease. We examined how bacterial and fungal root-inhabiting endophytes interacted *in vitro* and *in planta* in the annual grass *Brachypodium distachyon*. We further tested the effects of barley yellow dwarf virus (BYDV), a plant-pathogenic virus, on plant traits and endophyte communities when co-inoculated with the bacterium *Enterobacter ludwigii* and the fungus *Fusarium falciforme*. Antagonistic volatile interactions were observed between *E. ludwigii* and *F. oxysporum*, and *E. ludwigii* had a negative effect on shoot biomass without any effect or interaction with fungal co-inoculants. The addition of fungi to plants pre-inoculated with *E. ludwigii* reduced abundance of the bacteria in roots, as did the infection of the plants with BYDV. Endophyte and virus inoculations both had effects on plant traits but were not interacting. Fungal tissue abundance in roots was unaffected by BYDV infection. Intermicrobial interactions and host disease were observed to alter the assembly of plant microbiomes, which suggests potential for design of microbial inoculants based on characterization of intermicrobial interactions and managing viral disease by reducing impacts on microbiome functions.

Introduction

Plants do not grow as an isolated entity. Rather, they exist in a community composed of microbes – viruses (Koskella and Taylor, 2015), bacteria (Ulbrich *et al.*, 2021), archaea (Moissl-Eichinger *et al.*, 2018), fungi (Materatski *et al.*, 2019), oomycetes (Maciá-Vicente *et al.*, 2020), protists (Sapp *et al.*, 2018) – and larger organisms, such as invertebrate animals (Bray and Wickings, 2019) and other plants (Kong *et al.*, 2021). Each organism in this community can occupy one or more distinct or overlapping niches: saprotrophs (Kuo *et al.*, 2014), biotrophs (Tisserant *et al.*, 2013), parasites and hyperparasites (Vandermeer *et al.*, 2009; Jeffries, 1995), pathogens (Grünwald *et al.*, 2016), autotrophs (Priya *et al.*, 2015), mycorrhizae (Cope *et al.*, 2019), and nitrogen fixers (diCenzo *et al.*, 2020). In each of these roles the organisms can affect both one another and the plant host. Some produce or consume plant hormones (Nassar *et al.*, 2005; Gravel *et al.*, 2007; Waqas *et al.*, 2012; Saikia *et al.*, 2018) or provide critical nutrients such as phosphorus (Thingstrup *et al.*, 2000) or nitrogen (Ryu *et al.*, 2020). Plants often have closely associated microbes on the surface of root and shoot tissue (rhizoplane and phylloplane epiphytes) and within these same tissues (root or shoot endophytes; Gopal and Gupta, 2016). The composition of these communities have previously been associated with plant health, especially with suppressing disease (Carrión *et al.*, 2019), tolerating stresses (Vujanovic *et al.*, 2019), or even resistance to herbivory.

Plant pathogenic viruses may be the simplest component of the plant microbiome in terms of size alone. Yet, plant pathogenic viruses, especially those of crops, cause serious threats to global food security (Jones and Naidu, 2019). In cereal crops including barley, wheat, and oats (Miller *et al.*, 2002), barley yellow dwarf virus (BYDV) can reduce yields by 0.34-0.55% for each percent increase in virus infection, which was determined as 24.4% ± 2.5% infected and

47.8 ± 2.5% in the naturally infected plots assessed by Perry et al. (2000) in Illinois and Indiana in 1997. This ssRNA(+) virus is transmitted by aphids, which feed from the phloem where the virus replicates (Choudhury *et al.*, 2017; Miller *et al.*, 2002). In turn, the virus results in stunted growth of the host due to compromised vascular transport (Liu and Buchenauer, 2005), associated with symptoms such as reduced root length, seed mass, seed yield, tiller number, tiller height, and shoot mass (Riedell *et al.*, 2003). This pathogen is typically treated by suppressing aphid populations with insecticides (McKirdy *et al.*, 2002). This viral pathogen is a useful system for exploring plant-pathogen-microbiome interactions because it can be transferred to new plants in a controlled manner by transfer of aphids from an infected host to a non-infected host, is a consequential disease, and is likely altering both immune and nutrient status of the host in ways that may affect the microbial community.

Bacteria and fungi, as abundant members of rhizosphere and endosphere communities, are commonly interacting through a variety of mechanisms (Deveau *et al.*, 2018). Space and nutrients are limited in soil and root tissues, which suggests that resource competition may dictate the assembly and dynamics of these communities (Bulgarelli *et al.*, 2013). While fungi and bacteria can exclude one another from root colonization (Mousa *et al.*, 2016), fungi and bacteria may alternatively show mutual benefit in motility and nutrient acquisition (Zhang *et al.*, 2016; Jiang *et al.*, 2021). Furthermore, the host plant has multiple types of responses to microbes, which may trigger immune responses that alter the community (Teixeira *et al.*, 2019). Host plants defective in pattern-triggered immunity can assemble unhealthy communities of microbes, termed dysbiosis (Chen *et al.*, 2020). Community assembly theory suggests that a series of “filters” may dictate the organisms present in the community, whereby organismal traits are critical to survive selection by environmental conditions, including biotic and abiotic factors

(Keddy, 1992). The composition of the communities in and around the roots of a plant could be determined by interspecific interactions with other organisms present, host factors, or abiotic factors (nutrients, soil type, climate).

Biocontrol of crop diseases has been considered as a management approach to improve sustainability, avoid agrochemical applications, and overcome disease pressure for many years (New and Kerr, 1972; Lewis and Papavizas, 1991; Stenberg *et al.*, 2015; Ellis, 2017; Syed Ab Rahman *et al.*, 2018; De Silva *et al.*, 2019). This approach employs a beneficial inoculant, often a bacterium or fungus, to reduce disease effects on the crop (Syed Ab Rahman *et al.*, 2018). The biocontrol agents are often single, specialized strains with known capabilities (New and Kerr, 1972; Howell *et al.*, 1993; Kamensky *et al.*, 2003), but more recently communities of microbes have been developed for improved effectiveness (Gross *et al.*, 2018; De Vrieze *et al.*, 2018; Niu *et al.*, 2020). As with more general interactions of microbes, biocontrol of diseases can be facilitated by the actions of special metabolites (Christiansen *et al.*, 2020) or enzymes (Kappel *et al.*, 2020; Sorokan *et al.*, 2020) exported by the biocontrol agent, resource competition (Wei *et al.*, 2017), or mediated through host immunity (Van Wees *et al.*, 1997; de Lamo *et al.*, 2018).

The host plant *Brachypodium distachyon* (L.) P.Beauv., hereafter “Brachypodium” or *B. distachyon*, is a C3 annual grass, and an established model organism for cereals, including wheat and barley (Draper *et al.*, 2001). *B. distachyon* has a small diploid genome, small footprint, fast life cycle, extensive experimental resources, associates with bacterial and fungal endophytes (Gagné-Bourque *et al.*, 2015; Penner and Sapir, 2021), and is susceptible to BYDV (Tao *et al.*, 2016) making it an appropriate system for studying the interactions of microbial symbionts and viral disease. Furthermore, it has previous been characterized as a non-host model for

switchgrass (*Panicum virgatum*; Gill *et al.*, 2015), from which a large collection of fungal isolates had been cultured by our research group.

As interest is growing in using microbial inoculations to improve the sustainability and productivity of agriculture, practitioners face a challenge in understanding the context and conditions that effect the efficacy of these approaches. Moreover, microbial inoculations may have additional, unexplored benefits for treating or preventing crop diseases. We aim to describe how interactions between microbes, the host, and viral disease result in alteration to the microbial community and impacts on the host plant. Because microbes often compete for space or nutrients from the host, and viral disease alters the supply of these resources, we expected that inoculating additional microbes on the plant will change the abundances of present members, and that viral disease will alter the host's microbial community. Characterizing the effects of these complex, multi-member interactions may help improve the predictability and application of agricultural microbial inoculants.

Methods

Isolation of endophytic bacteria and fungi

Switchgrass roots obtained from a fertilized Great Lakes Bioenergy Research Center plot at the Kellogg Biological Station were washed until no visible soil was present, then sterilized in ethanol and bleach according to Bashan et al. (1993). Roots were finely chopped with sterilized scissors, then shaken at 30°C for 48 hr in 0.85% sterile NaCl solution. After incubation, the solution was serially diluted in three series, down to 10^{-4}, and spread (100 μL) or streaked on King's B agar (Bashan *et al.*, 1993) and incubated at 30°C overnight. After colonies were visible, morphotypes were restreaked on new plates and roughly identified with 16S PCR and Sanger sequencing using 27F and 1492R primers. Isolate FCP2-01 was later genotyped using HSP60

primers and identified as *Enterobacter ludwigii* (Hoffmann *et al.*, 2005) with BLASTn against the NCBI Nucleotide Database (HSP60 sequence: MW988544). Fungal isolates were obtained from surface sterilized roots by Morgane Chretien, and later identified with sequencing of the ITS1 locus and identification with BLASTn (Table 1.1). MC67 was later identified with RPB2 (MW988545, MW988546) and EF1α (MW988543) locus sequencing with primers 5F2/11aR (O'Donnell *et al.*, 2007) and EF1/EF2 (O'Donnell *et al.*, 1998), respectively.

Strain Name	Species	*In vitro*	*In planta*	ITS Accession
MC20	*Rhizopus oryzae*	X	X	MW980743
MC50	*Alternaria* sp.	X		MW980744
MC54	*Mucor circinelloides*	X	X	MW981501
MC64	*Talaromyces* sp.	X		MW980745
MC67	*Fusarium falciforme*	X	X	MW974615
MC68	*Fusarium oxysporum*	X		

Table 1.1 Fungal endophyte strains used in interaction experiments.
Strains were identified with sequencing of the ITS1 and/or additional loci. Each "X" denotes that a given strain was used for *in vitro* or *in planta* experiments.

Transformation of bacterial endophyte with GFP

A transformation protocol was developed for FCP2-01 due to its growth characteristics and susceptibility to chloramphenicol and streptomycin selection. The plasmid BBa_K608011 which constitutively expressed GFP was used to quickly determine transformants, and later to visualize bacteria on and within roots. FCP2-01 was grown in 50 mL of lysogeny broth (LB) overnight (15-24 hr) at 225 rpm at 30 °C, then transferred to a 50 mL conical tube and centrifuged at 4000 x g for 10 min. The supernatant was poured off and replaced with 25 mL sterile 18 MΩ·cm water. The cells were resuspended, and washed centrifuge again with the same conditions. Cells were resuspended in 25 mL of 10% glycerol in water. This suspension was frozen at -80°C until used, whereupon cells were thawed on ice prior to the addition of plasmid

DNA. Cells were transformed by adding 100 µL of suspension and 100 ng plasmid into an ice-chilled 1 mm electroporation cuvette, which was then used to electroporate cells at 1.8 kV. 1 mL of warmed (37°C) SOC was then added to the cuvette, and then as much cell suspension as possible was pipetted out of the cuvette and into a 1.5 mL microcentrifuge tube and shaken at 30°C at 175 rpm for 1 hour. 100 µL of cell suspension was spread plated on LB agar supplemented with 30 $\mu g \cdot mL^{-1}$ chloramphenicol, then plates were sealed with parafilm and incubated at 30°C until colonies were visible, about 24-48 hr.

Volatile interactions of bacterial and fungal endophytes

Prior to plant inoculation, fungal partners of the bacteria were evaluated *in vitro*. Volatile interactions were evaluated on 4-way divided 100 mm plates, with fungal and bacteria colonies inoculated on MEA and LB agar, respectively, but sharing headspace. 10 µL of *E. ludwigii* FCP2-01, grown 24 hr at 30°C were placed in opposite quadrants. Fungal plugs were place in adjacent quadrants. Fungal growth was measured using by averaging three radii, and time of measurement was dependent on growth rate. Comparisons were made to plates with fungi but absent bacterial inoculation.

Growth of bacteria, fungi, and Brachypodium in axenic plate cultures

Brachypodium distachyon Bd21-3 (genotyped at gene Bradi1g34430) seeds were sterilized prior to plate growth experiments by shaking seeds in a solution of 0.6% NaOCl and 0.1% Tween 20 for 7 minutes, followed by 5 washes in sterile water. Seeds were kept in a foil-covered 15 mL conical tube at 4°C for 72 hr. *Enterobacter ludwigii* FCP2-01[BBa_K608011] was grown in 16 mm x 100 mm culture tubes at 30°C for about 18 hr, reaching OD_{600} (1/10) of ~0.4. Seeds were transferred to separate 15 mL conical tubes with 6 mL of LB or bacterial culture, then placed on a 2D rocker at 4°C for 15 min. Seeds were then placed onto LS/2 plates

(Linsmaier and Skoog, 1965) without sucrose, covered in aluminum foil, and placed in a growth chamber set at 23°C, 16 hr/8 hr day/night cycle, covered in foil for the first 24 hr to facilitate germination. Root hair zone images were acquired at 7 days with a Leica S9D stereo microscope (Leica Microsystems, Bannockburn, IL, USA), then the root hair zone was measured with ImageJ (Schneider *et al.*, 2012) by an evaluator who was unaware the applied treatments. After 8 days in the growth chamber, 1 cm square blocks of fungi-colonized malt extract agar (10 $g \cdot L^{-1}$ malt extract, 1 $g \cdot L^{-1}$ yeast extract, and 10 $g \cdot L^{-1}$ Bacto agar) were placed onto each plant plate. Plants were harvested after 10 days.

Quantification of endophyte infection of roots

Root tissue was collected from plates using sterilized forceps, then placed into a 50 mL conical tube to which 10% bleach (0.6% NaOCl) and 0.1% Tween 20 was added, then shaken for 30 s followed by 1 rinse with sterile 18 MΩ·cm water. Shoots and roots were split by cutting at the green to brown transition with a scalpel and the seed coat was removed. Roots were then placed into pre-massed 2 mL round bottom tubes, and frozen at -80°C. Frozen root samples were then lyophilized for about 2 days until fully dry. Shoot tissue was also harvested, placing into pre-massed 2 mL tubes and lyophilizing, but without rinsing. Root and shoot biomass were determined after drying. Two sterilized 5/32” steel beads were added to each root tissue tube, and roots were pulverized at 30 Hz for 1 min in a TissueLyzer II (Qiagen, Carlsbad, CA, USA). Pulverized root tissue was extracted with the Mag-Bind® Plant DNA Kit (Omega Bio-tek, Norcross, GA, USA).

Quantitative PCR was performed with these DNA extracts as template. The primers Bd_UBC18_F1 (5’-CTAGTTCCACCTTGCATCTCTAC-3’), Bd_UBC18_R1 (5’-TCAGTTGCTCTGGCTTCTAAC-3’), EntLud_rodA_F2 (5’-

GGTATGGTGAGTGGTATTCTGC- 3'), and EntLud_rodA_R2 (5'-TCGACATCACGATACCAAACC- 3') were designed with IDT PrimerQuest to target single-copy genes for *B. distachyon* (BdiBd21-3.4G0006300.1) and *E. ludwigii* (ECWSU1_RS05980), respectively. qPCRs were 15 μL per reaction, with 7.5 μL Power SYBR® Green Master Mix (ThermoFisher Scientific, USA) or Luna® Universal qPCR Master Mix (New England Biolabs, Ipswich, MA, USA), 0.25 μL of each 10 μM primer, 5.5 μL nuclease-free water, and 1.5 μL template. Reactions were performed on an Eppendorf MasterCycler® RealPlex 2 (Eppendorf, Hauppauge, NY, USA), consisting of 10 min at 95°C, then 40 cycles of 95°C for 15 s and 60°C for 1 min, with a fluorescence read during anneal/elongation for Power SYBR® Green Master Mix, or with a modified 2 min initial denaturation and 30 s elongation for Luna® Universal qPCR Master Mix. A melting curve of 95°C for 15 s, 60°C for 15 s, then up to 95°C over 20 min with fluorescence reading was performed to assess amplification specificity. C_t values derived from the automated threshold were then used to calculate relative bacterial cells per plant cell with the Pfaffl method (2001), accounting for experimentally determined primer efficiency.

Infection of Brachypodium with endophytes and BYDV-PAV-MI

Following axenic plate experiments, we incorporated the plant virus BYDV-PAV (*Luteovirus*) strain "MI" to examine the how a plant pathogen interacts with inoculated endophytes. Potting mix weas prepared by wetting Arabidopsis mix from the Michigan State University Plant Research Laboratory Growth Chamber Facility until wet but not dripping, and autoclaving twice for 45 minutes at pressure, separated by 24 hr. The mix was rinsed with sterile distilled water after the second autoclave run and packed into pots. Seeds were sterilized as before, then stored at 4°C for 15 days before planting in soil. Plants were watered every third day with distilled water from the bottom. Plants were grown in an indoor temperature-controlled

room at 26°C with 16 hr/8hr light/dark cycle under fluorescent lights at ~136 $\mu mol \cdot m^{-2} \cdot s^{-1}$ PAR at soil level.

Infection with BYDV-PAV-MI was performed 14 days after plating. Aphids () were transferred from infected oat shoot tissue by cutting a piece of oat leaf or stem with ~5 aphids and moving it into the pot with the Brachypodium and covering with a cage consisting of a clear tube capped by fine mesh to prevent escape. Negative control plants were mock infected with aphids feeding from non-infected oat plants. After 2 days, the plants were drenched in Astro®, a permethrin-based insecticide, to kill any aphids on the plant.

Fusarium falciforme MC67 was grown for 28 days at room temperature under a 50 $\mu mol \cdot m^{-2} \cdot s^{-1}$ PAR 12hr/12hr light/dark cycle on PDA/2 +YE (12 $g \cdot L^{-1}$ potato dextrose broth, 1 $g \cdot L^{-1}$ yeast extract, 10 $g \cdot L^{-1}$ Bacto agar). These plates were sliced with a scalpel and placed into a 50 mL conical tube with sterile water and glass beads, then vortexed and filtered through Miracloth fabric (MilliporeSigma, Burlington, MA, USA) to obtain spores. Spore concentration was determined with a counting camber, then diluted with water to final inoculum concentration. *E. ludwigii* FCP2-01[BBa_K608011] was grown at 30°C for 22 hr in a 250 mL baffled flask containing 50 mL LB, shaking at 200 rpm. This obtained an OD_{600} of 7.24. The cells were centrifuged at 4000 x g for 10 min at room temperature, then diluted with water using CFU estimates from a standard curve of dilutions. The combined fungal/bacterial inoculum contained $\sim 3 \cdot 10^8$ bacteria cells and 10^3 conidial spores per mL. Plants were inoculated with either 1 mL of the combined fungal/bacterial inoculum or an equal amount of sterile water at 20 days post planting.

Assessment of plant traits and endophyte infection

Plants were harvested 41 days after seeds were plated. Leaf area was determined for all above-ground parts of the plant by scanning with a LI-COR LI-3100 Area Meter (LI-COR Biosciences, Lincoln, NE, USA). Plant height was measured from the potting mix surface to the tallest spike. Number of tillers were recorded at the number of distinct, flower-bearing stems with leaves present. Depth to potting mix was recorded as a covariate. Shoot tissue was placed in Whirl-Pak bags (Whirl-Pak, Madison, WI, USA), which were pierced and lyophilized. Roots were removed from potting mix with forceps and water, sterilized with bleach and Tween 20 as before, then placed into pre-massed 2 mL microcentrifuge tubes with two 5/32" steel beads and lyophilized. Dried shoots and roots were massed, then roots were pulverized and DNA was extracted as before. DNA was diluted to 20x and 50x with nuclease-free water, then used as qPCR template with Power SYBR® Green Master Mix reaction mix and condition from before, and the BD_UBC18, EntLud_rodA, and a primer set for the fungal inoculum, FusSol_EF1a_F2 (5'-AGCGTGAGCGTGGTATC-3') and FusSol_EF1a_R2 (5'-ACATACCAATGACGGTGAC-3'), targeting a conserved region of the EF1α locus of strain MC67 in the *Fusarium solani* species complex, which was later identified as *F. falciforme*. For the *E. ludwigii* and *F. falciforme* primer sets, the C_t value was compared to the C_t value for the *B. distachyon* primer set at equal dilution, to account for effects of PCR inhibitors. C_t values were taken as the mean of at least two technical replicates. If the two replicates had a standard deviation greater than 1, another replicate was performed and the mean C_t values was calculated from the closest two replicates. Relative inoculum cells per plant cells were determined with $2^{\Delta Ct}$. Endophyte abundance distributions determined by qPCR were highly non-normal and were therefore analyzed with non-parametric tests.

Plant trait effects were modeled with a linear model ("lm" function in base R; R Core Team, 2019) using endophyte and virus inoculation and their interaction as random effects and soil depth as a covariate. Each trait value was transformed if needed to approximate a normal distribution. Estimated marginal means and pairwise differences were calculated with emmeans 1.3.5 (Lenth, 2020) and multcomp 1.4-13 (Hothorn *et al.*, 2008). Analysis in R and plot creation were performed with tidyverse 1.3.0 (Wickham, Averick, *et al.*, 2019), including tibble 3.0.5 (Müller and Wickham, 2019), tidyr 1.1.2 (Wickham and Henry, 2020), dplyr 1.0.3 (Wickham, François, *et al.*, 2019), ggplot2 3.2.1 (Wickham, 2016), patchwork 1.0.0 (Pedersen, 2019), maditr 0.7.4 (Demin, 2020), ggfortify 0.4.11 (Horikoshi and Tang, 2018; Tang *et al.*, 2016), and ggpubr 0.2.3 (Kassambara, 2019).

Microscopy to evaluate endophyte presence in roots

Plant roots were derived from seedlings grown identically to those used for biomass and qPCR experiments, with the exception of only planting 3 seedlings per plate and light present for 24 hr per day. The roots were harvested for confocal microscopy by breaking up the agar medium and removal of roots, then placement into 2 mL microcentrifuge tubes. 500 μL HPSS (pH 9.1) was added to each tube, in addition to 100 μL of 1 $mg \cdot mL^{-1}$ Calcofluor white M2R (360/430 nm, $\lambda_{ex}/\lambda_{em}$) and 20 μL of 100 $\mu g \cdot mL^{-1}$ Wheat Germ Agluttinin - 640R (642/662 nm, Biotium, Fremont, CA, USA). The samples were then vacuum infiltrated 3 times and viewed using an Olympus Fluoview FV10i Confocal Laser Scanning Microscope (referred to as CLSM, Olympus Life Science Solutions, Waltham, MA, USA). Bacterial cells were visualized using GFP fluorescence (488/507 nm).

Results

In-vitro bacterial-fungal interactions

In-vitro volatile interactions were observed pairwise between MC68 (*Fusarium oxysporum*) with bacteria present, showing decreased growth (FDR = 0.0080). All other pairwise comparisons were not significant (FDR > 0.05; Figure 1.1).

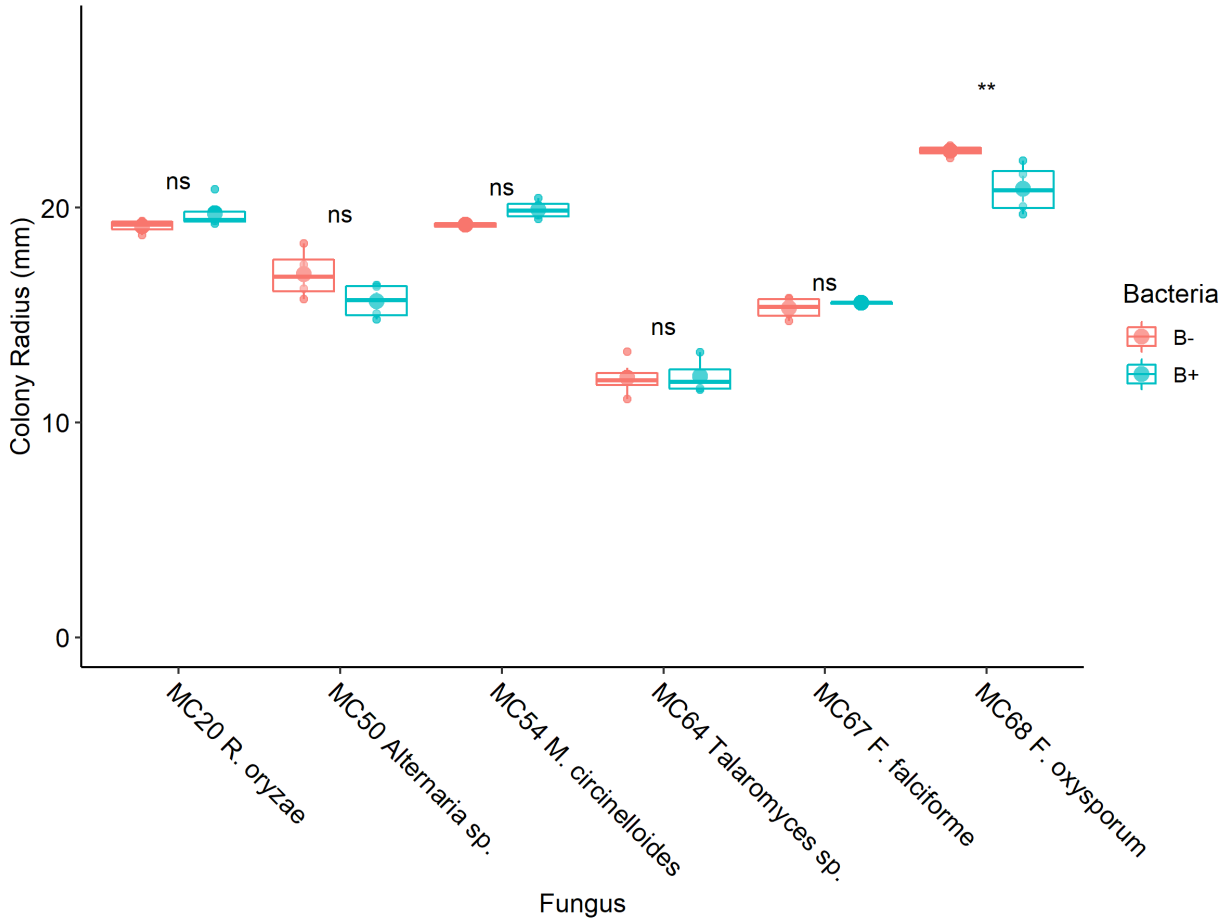

Figure 1.1 Volatile interactions of bacteria on fungal colony radii.
Radii of fungal endophyte colonies exposed to volatile-containing headspace of the bacterium *E. ludwigii* FCP2-01 (blue, "B+") and without bacterial volatiles (orange, "B-"). Pairwise comparisons are Tukey's post-hoc test, double asterisks indicate significance at FDR < 0.01.

Plant biomass and root morphology

A significant negative effect of bacteria inoculation was observed on dry shoot biomass ($F = 5.075$, $p = 0.028$), but not on root biomass ($F = 2.375$, $p = 0.13$; Figure 1.2). Additionally,

the effects of fungus overall or of fungus-bacteria interaction were not observed as significant for either shoots or roots ($p > 0.05$). Post-hoc tests did not show any significant pairwise comparisons for root or shoot biomass (FDR > 0.05). Measurements of the root hair zone diameters of 6-day-old Brachypodium seedlings showed inoculated seedlings having significantly larger root hair zones (t = 7.1972, $p = 7.88 \cdot 10^{-9}$, Figure 1.3).

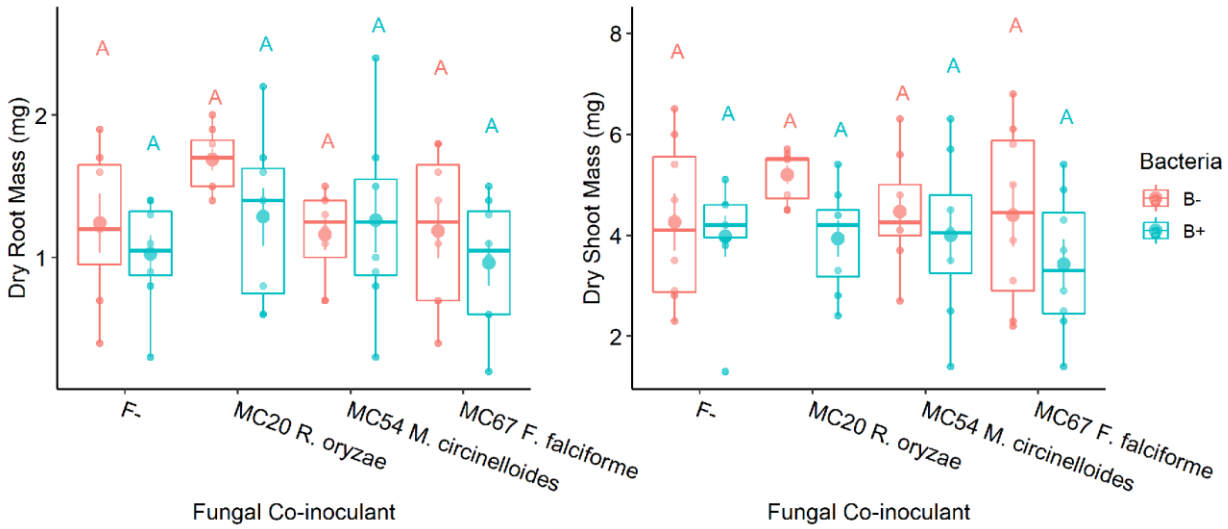

Figure 1.2 Dry root and shoot biomass of Brachypodium seedlings.
Mean values are shown with larger solid circles, with ±1 standard error displayed as vertical bars above and below the mean. Individual values are shown as smaller semi-transparent circles. Bacteria absent ("B-") and bacteria present ("B+") treatments are shown with orange and blue colors, respectively, and each fungal co-inoculant or negative control ("F-") are shown along the x-axis. Different letters indicate significant differences at FDR < 0.05 (Tukey's post-hoc test). A significant negative effective was observed for the bacterial inoculant on shoot biomass (F = 5.075, $p = 0.0282$) but not for root biomass (v).

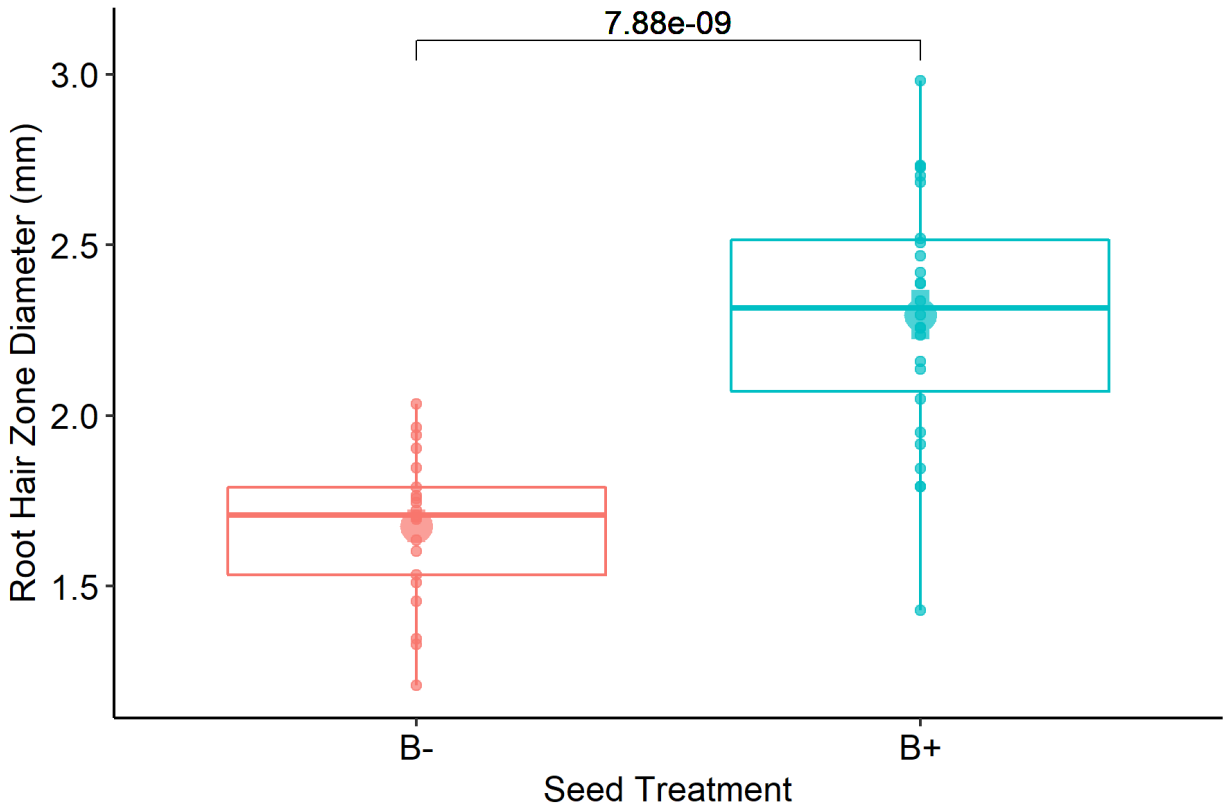

Figure 1.3 Root hair zone diameter in response to bacterial inoculation.
Mean values are shown with larger solid circles, with ±1 standard error displayed as vertical bars above and below the mean. Seeds which received a bacterial seed coating ("B+") demonstrated significantly longer root hairs than without bacteria ("B-"; $p = 7.88 \cdot 10^{-9}$).

Bacterial infection

Quantification of bacterial infection of the roots showed a significant effect fungal inoculant (ANOVA, F = 8.977, $p = 0.0003$), with the control condition exhibiting the highest infection (Figure 1.4). The control was significantly different from all three fungal-inoculated

conditions, but these did not vary between each other (Tukey's post-hoc test, FDR < 0.05).

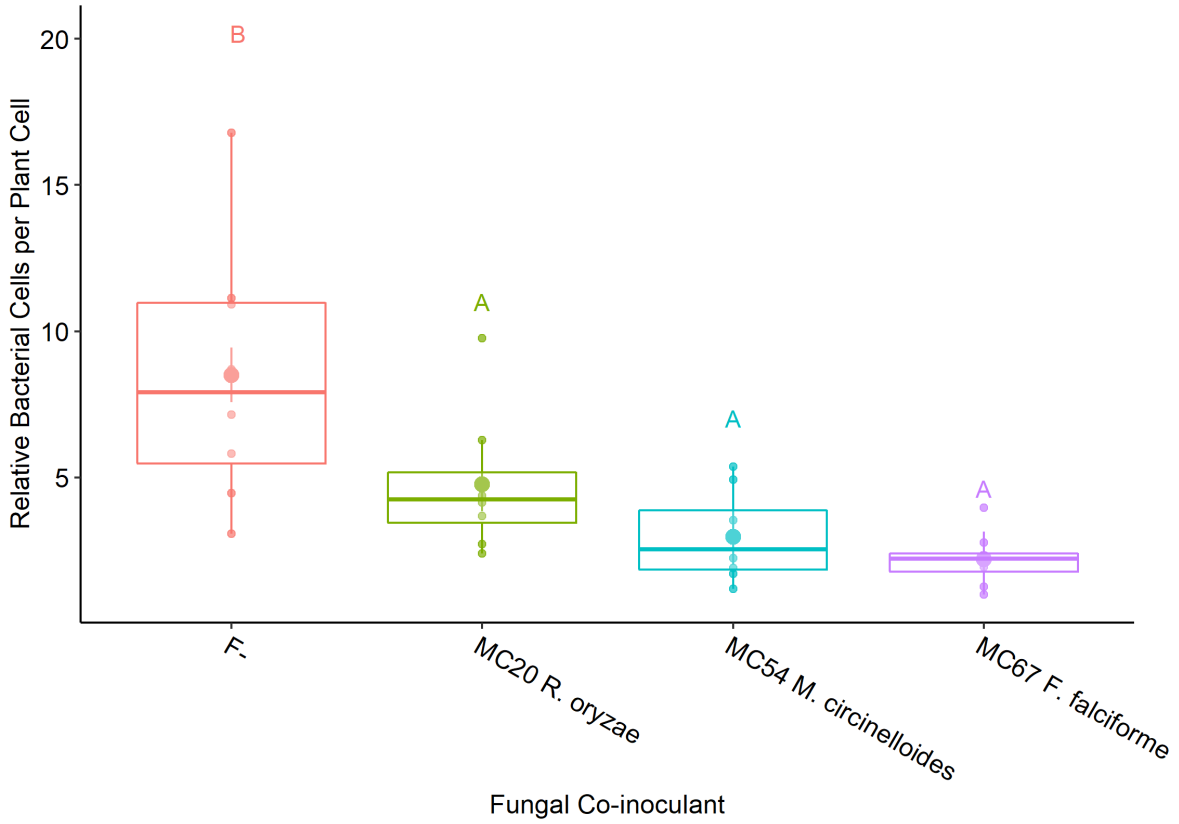

Figure 1.4 qPCR quantification of *E. ludwigii* FCP2-01 bacterial infection of plant roots by fungal co-inoculant.
Fungal co-inoculants are shown along the x-axis, with "F-" representing no fungal inoculant. Mean values are shown with larger solid circles, with ±1 standard error displayed as vertical bars above and below the mean, with individual values as semi-transparent smaller circles. Different letters indicate significant differences between fungal inoculant conditions (Tukey's post-hoc test, FDR < 0.05). Relative number of bacterial cells per plant cell, determined by qPCR copy-number ratio, was significantly associated with fungal co-inoculant ($F = 8.977$, $p = 0.0003$).

CLSM of Bacterial, Fungal, and Plant Interactions

Bacterial cells of FCP2-01 and fungal hyphae of *R. oryzae* MC20 were visualized inside roots of Brachypodium seedlings. Image planes of the inside of roots show both hyphae and bacteria cells within the apoplast of the plant roots (Figure 1.5). There are some locations were both microbes are present (red arrow), however these areas may be on the surface and therefore

less exclusive. At higher magnification (Figure 1.6), bacterial cells are located on the surface of the root, but not primarily on the hyphae.

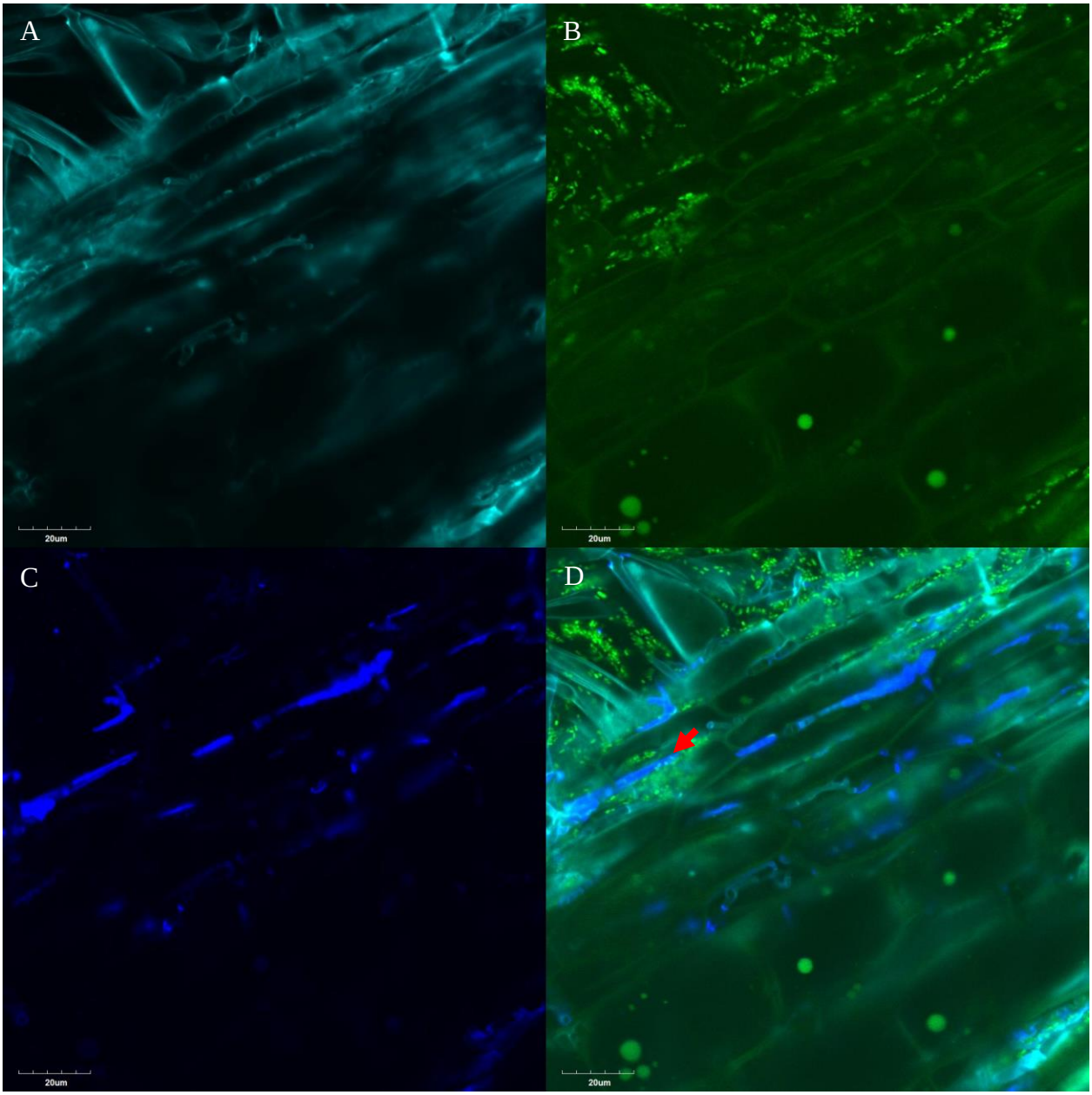

Figure 1.5 *R. oryzae* MC20 hyphae and *E. ludwigii* FCP2-01 in Brachypodium roots.
A) Cellulose and chitin stained with Calcofluor White M2R for staining of plant and fungal tissue, B) GFP signal from *E. ludwigii* FCP2-01[BBa_K608011], C) fungal tissue stained with WGA, and D) merged image. Scale bars represent 20 μm. Bacterial cells and fungal hyphae do not appear to co-localize, but instead exclude each other from root apoplast spaces, except for the site indicated by the red arrow appeared to be on the root surface.

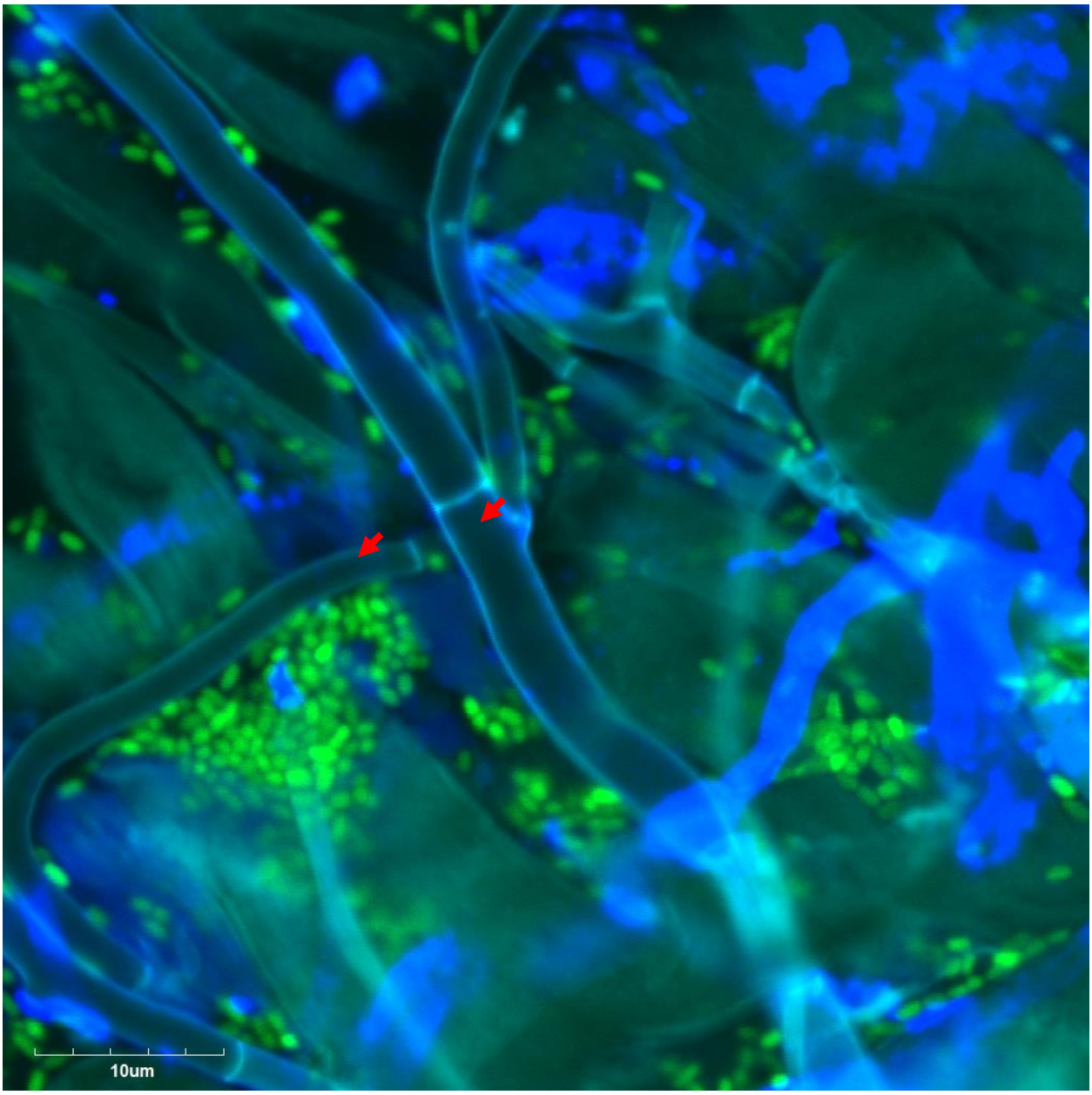

Figure 1.6 *R. oryzae* MC20 and *E. ludwigii* FCP2-01 on surface of Brachypodium root. Cells of FCP2-01 appear to occupy spaces independent of the hyphae (red arrows) and are not in high abundance on the surface of the hyphae.

Effect of viral infection on plant traits and endophyte communities

Several plant traits were assessed for four conditions: negative control (no inoculum), endophyte only (*E. ludwigii* FCP2-01 and *F. falciforme* MC67), virus only, and endophyte and virus. While number of tillers was recorded for each plant, this was not included in the analysis

because only 2 plants had more than a single tiller. For each trait, either endophyte inoculation ("Endophyte" – leaf area, area to height ratio) or virus infection ("Virus" – height, dry root mass, dry shoot mass, root to shoot ratio) had significant effects ($p < 0.05$; Table 1.2). However, the interaction of endophyte inoculation and virus infection ("Endophyte:Virus") was not significant for any plant trait, nor was the soil depth covariate. Pairwise differences were observed between treatments for each of the phenotypes assessed (Figure 1.7). Plants inoculated with endophytes had significantly smaller leaf area and area to height ratio, while virus infected plants had decreased height, dry shoot mass, dry root mass, and root to shoot ratio. Additionally, ordination by the four primary traits (height, leaf area, dry root mass, and dry shoot mass) showed separation by treatment (Figure 1.8).

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
Leaf area	Endophyte	1	1.1401	1.1401	12.4275	0.0012
	Virus	1	0.3662	0.3662	3.9919	0.0535
	Soil depth	1	0.1253	0.1253	1.3663	0.2503
	Endophyte:Virus	1	0.0966	0.0966	1.0524	0.3120
	Residuals	35	3.2110	0.0917		
Height	Endophyte	1	0.8702	0.8702	0.0653	0.7999
	Virus	1	181.9023	181.9023	13.6389	0.0008
	Soil depth	1	0.6875	0.6875	0.0515	0.8217
	Endophyte:Virus	1	9.0995	9.0995	0.6823	0.4144
	Residuals	35	466.7942	13.3370		
Dry root mass	Endophyte	1	0.0020	0.0020	1.7858	0.1901
	Virus	1	0.0282	0.0282	25.1191	1.55E-05
	Soil depth	1	0.0002	0.0002	0.1974	0.6596
	Endophyte:Virus	1	0.0020	0.0020	1.7609	0.1931
	Residuals	35	0.0392	0.0011		
Dry shoot mass	Endophyte	1	0.0212	0.0212	0.3420	0.5624
	Virus	1	0.8432	0.8432	13.5775	0.0008
	Soil depth	1	0.1368	0.1368	2.2027	0.1467
	Endophyte:Virus	1	0.0431	0.0431	0.6938	0.4105
	Residuals	35	2.1735	0.0621		
Root:shoot ratio	Endophyte	1	0.0157	0.0157	1.1619	0.2884
	Virus	1	0.2462	0.2462	18.2040	0.0001
	Soil depth	1	0.0025	0.0025	0.1852	0.6696
	Endophyte:Virus	1	0.0107	0.0107	0.7883	0.3807
	Residuals	35	0.4734	0.0135		
Area:height ratio	Endophyte	1	1.4630	1.4630	10.8411	0.0023
	Virus	1	0.1322	0.1322	0.9797	0.3291
	Soil depth	1	0.2140	0.2140	1.5855	0.2163
	Endophyte:Virus	1	0.1061	0.1061	0.7861	0.3813
	Residuals	35	4.7231	0.1349		

Table 1.2 ANOVA tables of the observed effects of endophyte and virus inoculants and the soil depth covariate on plant traits.

Endophyte inoculation consisted of a combined suspension of *E. ludwigii* FCP2-01 cells and *F. falciforme* MC67 conidia. Barley yellow dwarf virus strain PAV-MI was used as the viral inoculant and vectored into plants via *Rhopalosiphum padi* feeding.

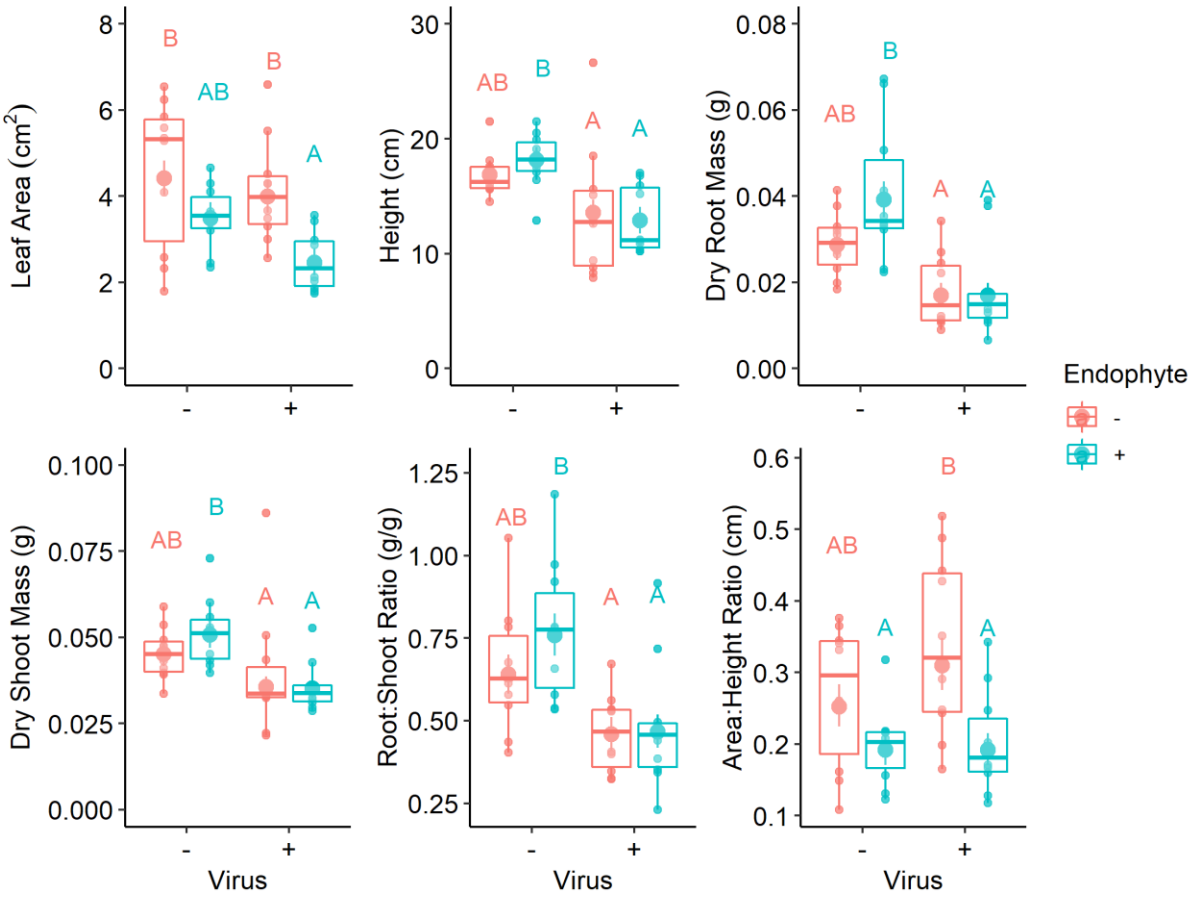

Figure 1.7 Trait values of Brachypodium in response to inoculation with endophytes and/or virus.

Observed trait values are on the y-axis, with BYDV-PAV-MI ("Virus") infection treatment on the x-axis (infected "+" and uninfected "-") and color representing *E. ludwigii* FCP2-01 and *F. falciforme* MC67 treatment ("Endophyte") as inoculated (blue, "+") or uninoculated (orange, "-"). Mean values are shown with larger solid circles, with ±1 standard error displayed as vertical bars above and below the mean. Values for individual plants are plotted with smaller, semi-transparent circles. Different letters indicate significant pairwise differences at FDR < 0.05 (Tukey's post-hoc test).

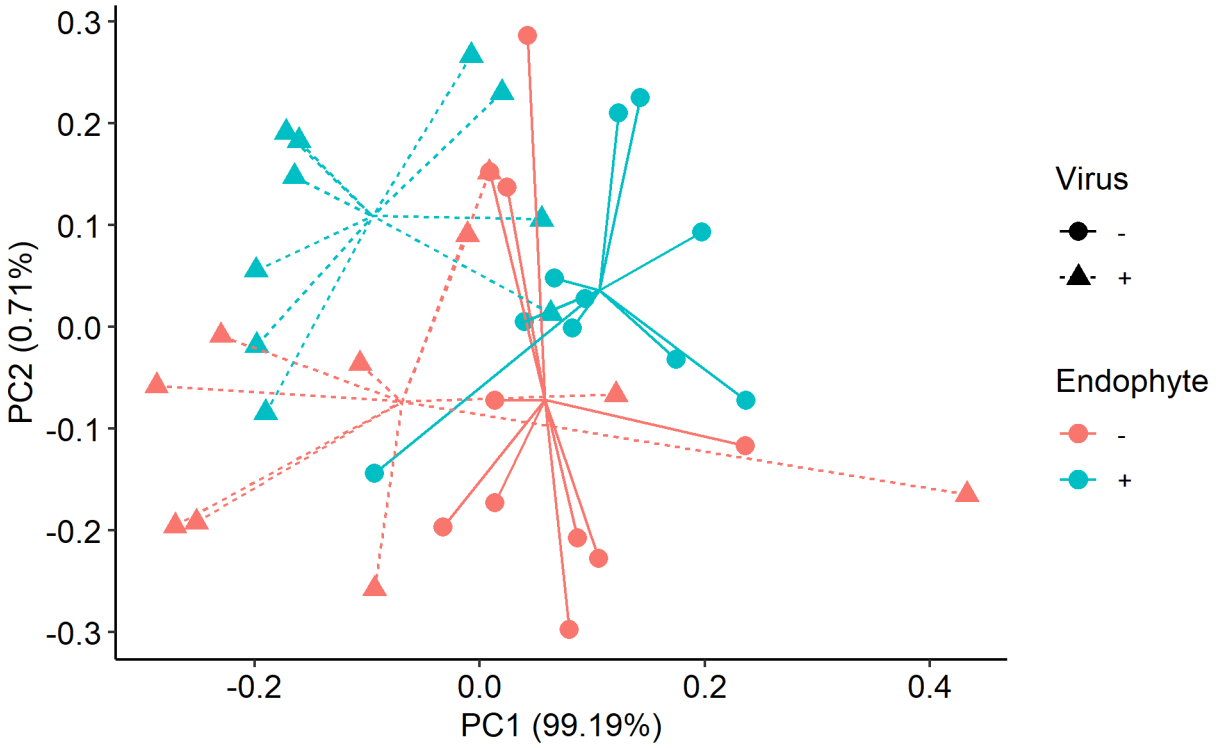

Figure 1.8 Principal components analysis ordination of plant traits.
Virus (BYDV-PAV-MI) and endophyte (combined *E. ludwigii* FCP2-01 and *F. falciforme* MC67) inoculations are represented by shape and color, respectively, with line segments connecting points in the same treatment groups to the mean coordinates of each group. Percentages on axes labels are how much of the variation in the dataset is accounted for by each principal component.

Observed endophyte communities in roots, as determined by qPCR for single copy genes for *E. ludwigii* FCP2-01 and *F. falciforme* MC67 normalized to a single-copy Brachypodium gene, showed effects of virus infection only for bacteria, not fungi. A lower normalized bacterial abundance was observed in the surface-sterilized roots from virus-infected plants compared to the virus-uninfected (Wilcoxon, W = 81, $p = 0.0185$; Figure 1.9). However, no differences were observed between virus-infected and uninfected condition for fungal abundance (W = 38, $p = 0.888$) or the ratio of bacterial and fungal abundance (W = 52, $p = 0.139$).

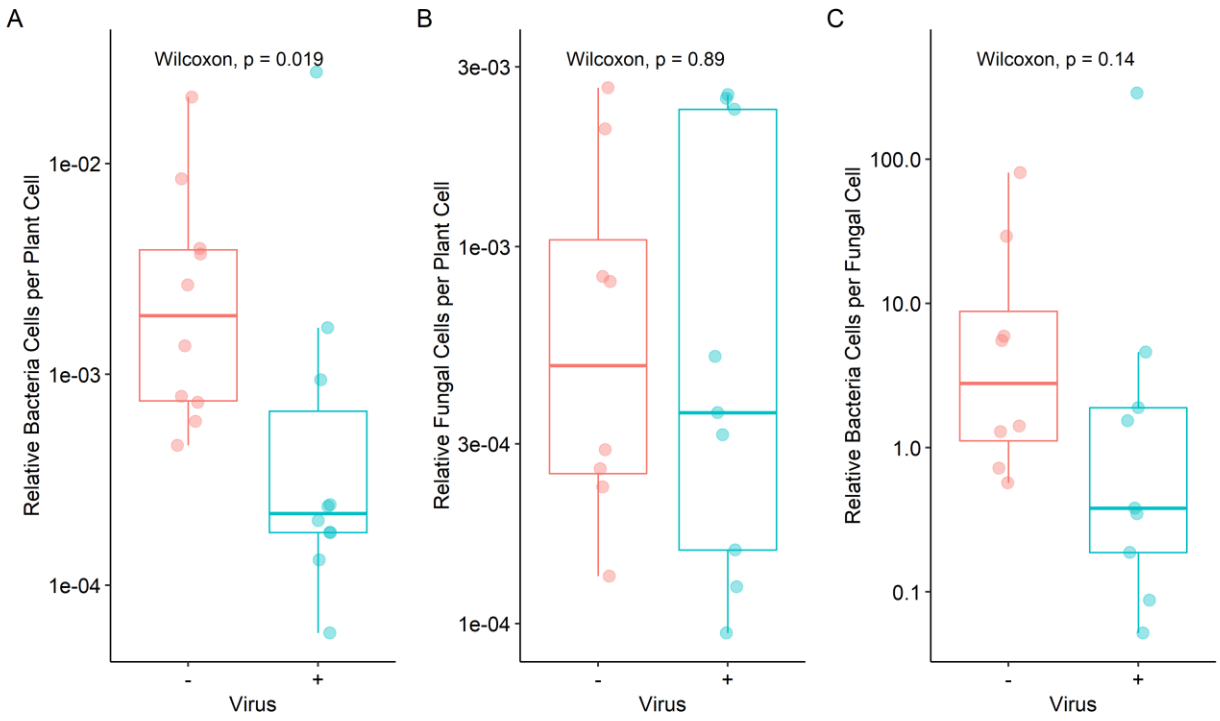

Figure 1.9 Relative abundances of bacterial and fungal endophytes determined by qPCR in roots of Brachypodium.
Normalized relative abundance by copy-number ratio of A) *E. ludwigii* FCP2-01 bacterial cells per plant cell, B) *F. falciforme* MC67 fungal cells per plant cell, and C) bacterial cells per fungal cell. Wilcoxon test *p*-values are displayed above the bars, comparing medians between virus-uninfected ("-") and virus-infected ("+") plants. Individual values for each plant are displayed as semi-transparent circles, and the y-axis is log base 10 transformed.

Discussion

Inter-Kingdom microbial and endophyte-host interactions

This research involved the characterization of multiple aspects of the interaction of bacterial and fungal endophytes with each other, their plant host, and a host-pathogenic virus. The endophyte partners were observed interacting *in-vitro* in potentially competitive ways, which is expected given the numerous niches these organisms co-occupy and previous observed interactions (Kamensky *et al.*, 2003; Mousa *et al.*, 2016).

The effects on host biomass by the bacterial endophyte suggests this bacterium may be parasitic on the plant host, rather than acting as a mutualist, given the reductions in host biomass.

However, the consequences of reduced shoot biomass may not detriment host fitness (viable seed production), which was not measured in this experiment. This endophyte may provide a comparison to known plant-growth promoting endophytic bacteria (Ait Barka *et al.*, 2006), given its negative impact on plant biomass, allowing for comparisons in infection mechanism and potential host selectivity. Contrary to the effect on biomass, the observed effect on root hair length may be important to adaptation to drought stress (He *et al.*, 2015; Brown *et al.*, 2012). Notably, *E. ludwigii* FCP2-01 was selected for these experiments due to its favorable characteristics as a synthetic biology chassis (growth requirements, antibiotic susceptibility, and transformability) and known ability to infect and remain within Brachypodium roots. Selecting a bacterium based on its known effects on plant traits may have resulted in different observed interactions.

The results on bacterial infection in the plate growth experiment were contrary to the initial prediction and is incompatible with the "fungal highway" mechanism of host infection. The plants in this experiment were inoculated with the bacteria by seed coating, while the fungi were added 6 days after germination as plugs of colonized agar on the surface of the plant media. It appears likely that fungal infection of the roots competes with bacteria, and if the bacteria are already present, the hyphae may cause displacement. The lack of co-localization seen with CLSM supports the potential displacement of bacterial endophytes already present. An additional trial to consider is adding bacteria to the roots at the same time as the fungi, to test if together they more effectively colonize the available apoplast. However, this approach may be less comparable to in-field conditions, for which a prospective beneficial bacterial endophyte would likely be applied as a seed coating.

The root apoplast provides a limited niche for colonization by endophytes, and the chronology of infection may influence the root endophyte communities' final composition through priority effects (Fukami *et al.,* 2010). Competing root colonizers have been known to reduce overall abundance of microbial biomass in mixed inoculums relative to monocultures (Engelmoer *et al.,* 2014). Manipulating the timing of inoculation of may alter final abundances of microbes as well as the host's characteristics.

Measured *in-vitro* interactions were expected to be predictive of interactions in the plant, such as enhancement of fungal growth by the bacteria translating into greater infection of the host plant by the bacteria when co-inoculated with the given fungus. However, these effects were not demonstrated in the expected direction in a two-microbe axenic system. In a more complicated system resembling a natural community, such *in-vitro* interactions may determine which partners are successful in colonizing the root community.

Plant trait changes from endophyte and virus inoculation

The infection of Brachypodium with BDYV-PAV-MI showed expected effects on host plants – reduced height and reduction in biomass of both root and shoot organs (Riedell *et al.*, 2003), as well as altered biomass allocation with a lower root to shoot biomass ratio (Hoffman and Kolb, 1997). Because of the vascular impedance and ultimate nutrient deprivation of roots, these trait changes are to be expected, but root and shoot biomass effects of the virus are only sometimes negative, depending on cultivar (Hoffman and Kolb, 1997). Alternately, the reduced leaf area and area to height ratio associated with endophyte inoculation does not have a clear explanation. Production of phytohormones, such as gibberellic acids, by *E. ludwigii* FCP2-01 or *F. falciforme* MC67 could alter host architecture (Paleg, 1965; Waqas *et al.*, 2012), as could interaction with immune response pathways (Ma *et al.*, 2013; Li *et al.*, 2014).

Root endophyte presence in response to viral inoculation

The observed changes in the abundance of endophytic bacteria (specifically *E. ludwigii*) in the root tissue as a response to viral infection supports the importance of the host in structuring associated microbial communities, however the mechanism by which reduced bacterial infection occur is not apparent in this experiment. Because a smaller amount of photosynthate is being translocated into root tissues of plants infected by the, it is possible that this reduces the ability of bacteria to grow within root tissue. Yet, endophytic fungi (*F. falciforme*) grew at a similar abundance in roots of virus-infected and uninfected plants. An important caveat is that the ratio of bacteria and fungi (copy-number ratio of *E. ludwigii* rodA to *F. falciforme* EF1α) was not significantly different between BYDV-PAV-MI infected and non-infected plants, and thus we cannot clearly ascertain that differential impacts on bacterial and fungal endophytes occurred. However, differential impacts remain possible and could be explained by plausible mechanisms. In this case, the bacteria and fungi may be consuming different nutrient sources within the root, a case of niche partitioning, which would relate reduced availability of a specific nutrient or resource with reduced abundance of the dependent organism (Stomp *et al.*, 2004). Similar to nutrient-based niche partitioning, spatial partitioning (Schoener, 1974) could explain differential impacts, especially if some root tissues are disproportionately impacted by viral infection. A further investigation which examined the metabolic requirements and preferences of these endophytes, paired with metabolic sampling and characterization of virus-infected and uninfected plant roots, could support a nutrient-based niche partitioning model of root endophyte competition and coexistence, while further microscopic observations of the spatial organization of endophytes in the root and the impacts of BYDV-PAV-MI on root tissue could support a spatial-based niche partitioning model.

An alternate model to niche partitioning is the role of the host's immunity as a selective filter, which allows some organisms to grow in the root based on those organisms' compatibility with host immune responses. The application of one stress, biotic or abiotic, can change the host susceptibility to other stresses (Prasch and Sonnewald, 2013), and could affect host genetic pathways such as those implicated in mediation of microbiome assembly (Stringlis *et al.*, 2018; Chen *et al.*, 2020). In humans, host genetic and immune factors are known to be associated with microbiome variation of specific taxa (Tang *et al.*, 2020), while in plants, genetic variants have similarly been implicated in microbiome variation (VanWallendael *et al.*, 2021). Viral infections in Brachypodium have been associated with alternative mRNA splicing for defense-related genes, which could alter immune responses to other microbial infections (Mandadi and Scholthof, 2015). Conducting gene expression analysis and examining known variants associated with microbiome interactions may indicate that specific immune response pathways or factors are responsible for the observed microbiome changes. A follow-up experiment based on the data presented assess the effect of inoculation of the same bacteria, fungi, and/or virus on expression of genes in host immune response pathways, but inoculants did not infect as expected and no differential expression was observed.

Implications for understanding viral plant disease

Alterations in the host microbiome in response to BYDV infection allow for the possibility that the disease's symptoms extend beyond the host's tissue and include the community of microbes associated with the plant. The potential for feedback between the microbial community and host exists, whereby host disease leads to loss (or gain) of services provided by the microbial community. Feedbacks between plants and rhizospheric microbes can effect plant traits in later generations (Hu *et al.*, 2018). If changes in microbial communities are

responsible for some symptoms of BYDV, microbial inoculations could attempt to complement lost services to reduce disease impacts. Yet, interaction effects were not observed for endophyte and viral inoculations in our experiments, which does not support the use of these specific inoculants as biocontrol for BYDV. Examining changes in complex natural communities may suggests specific traits lost in disease-associated microbiomes, and better inform the selection of biocontrol strains for managing BYDV.

Limitations and future directions

The set of experiments presented here attempt to replicate complex microbiome assembly processes with a small synthetic community of endophytic fungi and bacteria and a plant-pathogenic virus. Microbial traits are important to the assembly and function of a community (Wood *et al.*, 2018), and thus the single strains that we used do not represent the diversity of traits or combinations of traits relevant to community assembly and function. Furthermore, in a complex 185-member synthetic community Finkel et al. (2020) showed that the addition of only a single bacterial genus (*Variovorax*) is largely responsible for the direction of effects on host traits. Without having assessed a more taxonomically and functionally diverse set of bacterial and fungal isolates, we cannot be confident that the interactions observed are common in plant microbiomes. Subsequent experiments should test a wider range of *in vitro* interactions and test how the direction and type (volatile, water film transport, antimicrobial compound secretion) may affect interactions within plant roots.

Additionally, the limited number of traits assessed, the absence of bacteria-only and fungi-only treatments, a single inoculation schedule, and the unblocked experimental design could all be improved in later experiments to better understand the effects of each inoculant and account for environmental factors on observed traits. Looking forward, an understanding of the

mechanism of how viral disease impacts host microbiomes will require detailed assessment of the immune responses of the host, metabolic changes, and structural effects of viral disease, which can be linked to the immune recognition, metabolism, and spatial niche of the microbiome community members, respectively.

Conclusion

This study helps describe the organization of bacterial and fungal communities in roots as dependent on the interactions of the microbial partners. Bacterial endophytes are subject to displacement or inhibition by fungal partners later introduced to the system, suggesting that the root niche is restricted to a limited biomass of microbes and that competition is more likely than facilitation. While bacteria can move on fungal hyphae, the mechanism of bacteria on the rhizoplane entering roots by swimming on hyphal surfaces was not supported by these experiments. Both endophyte and viral inoculation of the Brachypodium host were associated with changes in plant traits, and the viral infection was associated with an altered microbial community occupying root tissue within reduced bacterial colonization. While the mechanism by which viral infection reduced endophytic bacterial presence but not fungal presence in the root is unknown. Connecting host disease to microbiome community assembly offers strategies to improve management of viral plant disease through the microbiome.

CHAPTER 2: CONSTAX2 – IMPROVED TAXONOMIC CLASSIFICATION OF ENVIRONMENTAL DNA MARKERS

Abstract

CONSTAX - the CONSensus TAXonomy classifier - was developed for accurate and reproducible taxonomic annotation of fungal rDNA amplicon sequences and is based upon a consensus approach of RDP, SINTAX, and UTAX algorithms. CONSTAX2 extends these features to classify prokaryotes as well as eukaryotes and incorporates BLAST-based classifiers to reduce classification errors. Additionally, CONSTAX2 implements a conda-installable, command line tool with improved classification metrics, faster training, multithreading support, capacity to incorporate external taxonomic databases, new isolate matching and high-level taxonomy tools, replete with documentation and example tutorials.

Introduction

High-throughput sequencing has revolutionized metagenomics and microbiome sciences (Di Bella *et al.*, 2013). These culture-independent methods have revealed previously unrecognized microbial diversity and has allowed researchers to detect organisms occurring at extremely low abundances (Brown *et al.*, 2015). Amplicon-based sequencing, which relies on amplification and sequencing of genetic markers such as the rRNA operon or protein-coding genes, is an extremely popular technique for studying microbiomes and microbial communities. Following sequencing, quality control, and demultiplexing, amplicon reads are clustered and representative sequences are classified taxonomically. Many algorithms have been developed to conduct the task of assigning taxonomy to environmental sequences. Some of the most popular include BLAST-based tools (Altschul *et al.*, 1997; Bokulich *et al.*, 2018), the Ribosomal

Database Project (RDP) naive Bayesian classifier (Wang *et al.*, 2007), and the USEARCH algorithms SINTAX (Edgar, 2016) and UTAX (Edgar, 2013).

While each of these tools can be implemented independently to assign taxonomy, a consensus-based approach was demonstrated to increase the accuracy and number of sequences with taxonomic assignments (Gdanetz *et al.*, 2017). Since the original release of the CONSTAX classifier, we have found the need for improved ease of use, updated software compatibility, simpler installation, improved accuracy and adaptability, and application to bacteria or other organisms. To address these needs, an updated version, CONSTAX2, has been developed.

Methods

Data and code availability

Shell, Python, and R code for running experiments and analyses, as well as data files for recreating figures are available at https://github.com/liberjul/CONSTAXv2_ms_code.

CONSTAX2 algorithm

CONSTAX2, known hereafter as "CONSTAX", begins by taking an input database file, formatted as one downloaded from UNITE (Nilsson *et al.*, 2019) or SILVA (Glöckner *et al.*, 2017) databases, and creating the necessary files for training the classifiers. SILVA-formatted databases have arbitrary ranks, which do not necessarily apply across all domains of life. To address this arbitrary ranking, SILVA taxonomy is assigned Rank 1 (equivalent to domain) to Rank *n* (lowest assigned rank). It is recommended to filter the SILVA database to a given domain (Bacteria, Archaea, Eukaryota) to preserve the meaning of assigned ranks, which can be performed with the "--select_by_keyword" option.

Classification is completed with SINTAX, UTAX, and RDP without the "-b, --blast" flag, or with SINTAX, BLAST, and RDP with the "-b, --blast" flag. The BLAST search

implementation is comparable to that described in Bokulich et al. (2018). Each input sequence is searched against a BLAST database generated from the database file using the blastn algorithm. A maximum number of hits is returned according to "-m, --mhits", which have an e-value equal to or below "-e, --evalue" and a proportion identity equal to or above "-p, --p_iden". A confidence score is generated based on the greatest proportion of hits which agree at the given rank. SINTAX, UTAX, and RDP are already conventional classifiers, so no special rules are required.

Returned taxonomy assignments from each classification method are reformatted to be consistent. Taxonomy assignments are then filtered according to the confidence threshold and combined to create a consensus with the following rules: 1) if no classifications are above threshold, no taxa is assigned; 2) if two or three classifications are above threshold and agree, the majority taxa is assigned; 3) if only one classification is above threshold, that taxa is assigned unless the "--conservative" flag is used, whereby no taxa is assigned; 4) if two or three classifications are above threshold and each is unique, the highest confidence taxa is assigned.

Clade partition cross-validation

We employed the approach used to validate the SINTAX classifier (Edgar, 2016), clade partition cross-validation (CPX), as a means to assess the ability for CONSTAX to classify both known and novel taxa. At both the family and genus ranks, records within sub-taxa (genera and species, respectively), were randomly partitioned to reference or query groupings. Singletons, families or genera with only one sub-taxon, were assigned to the query group as novel taxa. Sensitivity, misclassification rates, over-classification rates, and errors per query were calculated according to (Edgar, 2016), for UNITE and SILVA databases. Classification performance was assessed on 5 replicates for each partition (family and genus rank) and for UNITE fungal

representative sequences and SILVA bacterial ‘SSURef’ sequences. The same partitions were assessed with standard and conservative voting rules, and for commonly used regions of each marker. These regions were ITS1 and ITS2 from UNITE fungal sequences, extracted using ITSx (Bengtsson-Palme *et al.*, 2013), and the V3-4 and V4 hypervariable regions from SILVA bacterial sequences, extracted using in-silico PCR with primer sets 357wF-785R (Van Der Pol *et al.*, 2019) and 515f-806R (Parada *et al.*, 2016; Apprill *et al.*, 2015) allowing for 3 mismatched bases. For the UNITE database, classification was implemented with UTAX and BLAST implementations, with individual and consensus assignments compared for both implementations. However, given the size of the SILVA SSURef database, training time for the UTAX implementation would exceed 100 hours per replicate. Therefore, only the performance of the BLAST implementation was assessed for the SILVA database. Both UNITE and SILVA datasets were compared to the qiime2-Naive-Bayes feature classifier (Bokulich *et al.*, 2018), the mothur Wang classifier, the mothur k-nearest neighbors classifier with knn=3 (Schloss *et al.*, 2009), and Kraken 2 (Wood *et al.*, 2019), while for UNITE the SPINGO (Allard *et al.*, 2015) classifier was tested.

Classification Counts

Representative bacterial and fungal OTU sequences from Benucci et al. (2020) were classified with the BLAST CONSTAX implementation at recommended settings with the suggested UNITE and SILVA databases. The conservative voting rule was applied for the bacterial library, but not for the fungal library, given the results observed with CPX trials.

Algorithm speed

Runtime was determined for both training and classification steps using printed timestamps 1) before calling the CONSTAX executable, 2) after training completion within the

CONSTAX executable, written to STDOUT, and 3) after implementation of the CONSTAX executable. Training was performed on a single core on an Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz processor with 32 GB of requested memory. Each training database consisted of 500, 1000, 2000, 4000, 8000, or 16,000 sequence records sampled from the reference databases of the SILVA CPX test sets. Classification was performed with 1, 4, 8, 16, 32, 64, and 96 cores on an Intel® Xeon® CPU E7-8867 v4 @ 2.40GHz processor with 16 GB of requested memory, using 1000, 2000, or 4000 sequence records sampled from bacterial sequences in SILVA SSURef release 138. Training and classification were each performed with the default UTAX implementation or the "-b,--blast" BLAST implementation.

Definition of classification metrics

The classification performance framework from Edgar (2016) included the following classification performance metrics for clade-partition cross validation:

(2.1) $$Sensitivity = TP/N_{known}$$

(2.2) $$Misclassification\ rate = FP_{mis}/N_{known}$$

(2.3) $$Over-classification\ rate = FP_{over}/N_{novel}$$

(2.4) $$Errors\ per\ query = \frac{FP_{mis} + FP_{over}}{N}$$

Where $N_{known}$and N_{novel} are the number of queries known (at a rank above or equal to the partition level) and novel (at a rank below the partition level), $TP$is the true positive predictions of known queries, FP_{mis} is the number of false positive predictions of known queries, and FP_{over} is the number of false positive predictions of novel queries. $N$is the total number of queries and the sum of N_{known} and N_{novel}.

Plotting and analysis

Data generated via CONSTAX testing runs were parsed and reorganized with Python scripts, and uploaded into R 3.6.1 (R Core Team, 2019) for analysis. Plotting and preparation of tables were performed with tidyverse 1.3.0 (Wickham, Averick, *et al.*, 2019), including tibble 3.0.5 (Müller and Wickham, 2019), tidyr 1.1.2 (Wickham and Henry, 2020), dplyr 1.0.3 (Wickham, François, *et al.*, 2019), and forcats 0.5.0 (Wickham, 2020), and ggplot2 3.2.1 (Wickham, 2016). Patchwork 1.0.0 (Pedersen, 2019) and maditr 0.7.4 (Demin, 2020) were used for figure preparation. Classification performance metrics were compared between classifiers at each region, partition level, and database using a generalized mixed effects model with *glmer* function in lme4 1.1-21 (Bates *et al.*, 2015), in which classifier, partition level, and region are random effects and partition iteration is a fixed effect, and the metrics are modeled according to the binomial distribution.

(2.5) $$binomial(E_{ijk}, N_{ijk}) \sim classifier_i + level_j + region_k + (1|partition)$$

Where $binomial(E_{ijk}, N_{ijk})$ is the logit-transformed proportion of errors (total errors, misclassifications, and over-classification) to their respective number of queries (total, known taxa, and novel taxa), $classifier_i$ is the classification tool implemented, $level_j$ is the partition level (genus or family), $region_k$ is the region of sequence classified (Full, ITS1, or ITS2 for UNITE and Full, V3-4, or V4 for SILVA), and $(1|partition)$ is the pair of query-reference sequence sets (0-4, same sets for each classifier, level, and region).

Pairwise comparisons were performed with emmeans 1.3.5 (Lenth, 2020) and multcomp 1.4-13 (Hothorn *et al.*, 2008). Several scripts involved the Python packages pandas (The pandas development team, 2020; McKinney, 2010), numpy (Harris *et al.*, 2020), and xlsxwriter (McNamara, 2021).

Results

Algorithm speed and memory usage

The implementation of the BLAST algorithm as a third classifier and replacement of UTAX provides crucial speedup of the training step (Figure 2.1A), facilitating the use of the much larger SILVA database. For 16,000 sequences randomly sampled from the SILVA database, the BLAST implementation (including SINTAX, RDP, and BLAST) trained 370±32.1 sequences·s^{-1} (mean±SD), while the UTAX implementation (including SINTAX, RDP, and UTAX) trained 41.9±0.911 sequences·s^{-1}, an approximately 9-fold improvement. Furthermore, the BLAST implementation trains faster per sequence at larger database sizes.

Although the BLAST implementation is faster for training, classification is faster with the UTAX implementation (Figure 2.1B). The maximum classification speed was achieved at 32 threads for the BLAST implementation and between 4 and 8 threads for the UTAX implementation, depending on the number of query sequences classified, which minorly affected per-sequence rates. At 4000 query sequences, the BLAST implementation classified at a speed of 16.3±0.298 sequences·s^{-1} on 32 threads, while the UTAX implementation classified at a speed of 34.4±0.611 sequences·s^{-1} on 4 threads.

Training with bacterial records in the SILVA 138 SSU release (1,983,818 sequences, 2.8 Gb) with the BLAST implementation used 102.96 GB of RAM, while the fungal UNITE database (95,481 sequences, 60 Mb) used 15.24 GB for BLAST and 12.72 GB for UTAX implementations. Classification with the SILVA database with 16 threads used 28.16 GB for 500 sequences and 30.88 GB for 1000 sequences, while the UNITE database used 6-7 GB, regardless of implementation, threads, or number of query sequences.

Algorithm performance

Clade partitioned cross-validation and classification metrics from SINTAX (Edgar, 2016) were used (Supplementary Information) on each of the classifiers and consensus taxonomy assignments were compared for genus and family-level partitions as well as for full length ITS1-5.8S-ITS2 or 16S regions (accounting for the commonly used subregions ITS1, ITS2, V4, V3-4) with errors per query (sum of false negative and false positive rates), over-classification (false positive rate of unknown taxa), and misclassification (false positive rate of known taxa), for 5 query-reference paired datasets (Figure 2.1C-D, Table 2.1). The popular mothur knn and Wang classifiers (Schloss *et al.*, 2009), qiime q2-feature-classifier plugin (Bokulich *et al.*, 2018), Kraken 2 (Wood *et al.*, 2019), and SPINGO (Allard *et al.*, 2015) classifiers were compared using the same protocol. CONSTAX with the non-conservative consensus with BLAST had the fewest errors per query (EPQ) for any classifier (0.236-0.248, 95% CI for all regions and partition levels), or tied for fewest with the UTAX consensus, across the UNITE dataset. Alternatively, CONSTAX with the conservative consensus with BLAST had the fewest errors for all classifications in the SILVA dataset (EPQ=0.214-0.259). The BLAST implementation was valuable in decreasing misclassifications for the UNITE dataset compared to UTAX, but this was generally associated with increased (erroneous) over-classifications. The effect of the "-c,--conf" and "-m,--mhits" parameters on classification performance in shown for UNITE (Figure 2.2) and SILVA (Figure 2.3) datasets. The number of classifying representative sequences (OTUs) varied between classifiers and datasets, with more bacterial sequences classified at high ranks but fewer at low ranks compared to UNITE (Figure 2.4).

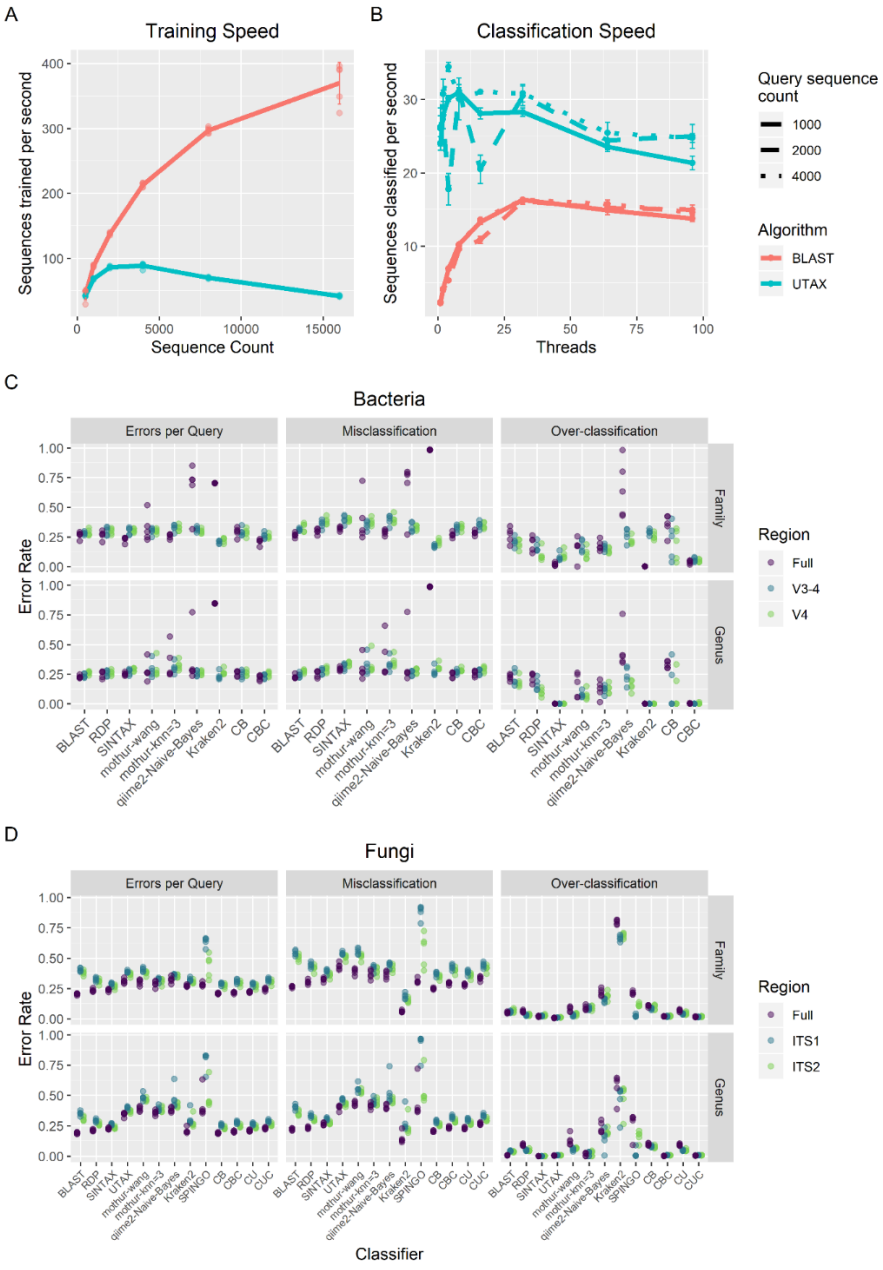

Figure 2.1 Performance of the CONSTAX algorithm.
A) Reference sequences parsed per second for training of the CONSTAX implementation with BLAST and UTAX, as a function of the size of the training set. B) Sequences classified per second with BLAST and UTAX implementations, as a function of query set size and threads used for parallelization. C-D) Classification performance resulting from clade-partition cross-validation, at genus and family partition ranks, for full and extracted regions, corresponding to each CONSTAX classifier and other common classification tools, for C) Bacteria in the SILVA SSURef release 138 dataset and D) Fungi in the UNITE RepS Feb 4 2020 general release. CB - CONSTAX with BLAST, CBC - CONSTAX with BLAST and conservative rule, CU - CONSTAX with UTAX, CUC - CONSTAX with UTAX and conservative rule.

Partition level			Family												
		Classifier	BLAST	RDP	SINTAX	UTAX	mothur-wang	mothur-knn=3	qiime2-Naive-Bayes	Kraken2	SPINGO	CB	CBC	CU	CUC
UNITE	Full	EPQ	20.3±0.7 (G)	23.4±1.2 (D)	23.7±1 (B)	31.1±2 (I)	30.2±2.4 (J)	28.7±2.6 (F)	32.2±2.6 (H)	27.1±0.8 (E)	28.3±1.5 (K)	20.8±0.5 (A)	21.4±0.8 (C)	22±0.6 (A)	24.3±1.2 (D)
		MC	26.4±0.8 (J)	30.3±1.7 (F)	32.3±1.5 (D)	43.2±2.8 (L)	39±2.6 (K)	36.6±3.6 (G)	36.3±2.9 (I)	6.1±0.8 (A)	31.1±2 (M)	24.8±0.6 (B)	29.2±1.1 (E)	28.1±0.9 (C)	33.4±1.8 (H)
		OC	5.1±0.6 (H)	6.2±0.8 (E)	2.1±0.4 (D)	0.7±0.1 (A)	7.9±2 (F)	9±1.9 (I)	21.8±2.9 (L)	79.6±1.9 (M)	21.3±1.4 (K)	10.7±0.8 (J)	2±0.2 (C)	6.8±0.9 (G)	1.6±0.3 (B)
	ITS1	EPQ	40.3±1.4 (G)	32.4±1.6 (D)	28.5±1.4 (B)	38.7±0.9 (I)	40.4±1.7 (J)	32.3±1.8 (F)	36.1±1.2 (H)	32.3±1.5 (E)	63.7±3.7 (K)	28.8±1.1 (A)	31±1.6 (C)	28.5±1.4 (A)	32.1±1.5 (D)
		MC	54.2±2.1 (J)	44.2±2.3 (F)	39±1.9 (D)	53.8±1.3 (L)	55.5±2.4 (K)	41.7±2.6 (G)	44.5±2.2 (I)	18.6±2.4 (A)	88.2±5.6 (M)	36.9±1.7 (B)	42.6±2.2 (E)	38.4±1.9 (C)	44.3±2.2 (H)
		OC	5.7±0.8 (H)	2.9±0.5 (E)	2.3±0.3 (D)	1±0.1 (A)	2.7±0.8 (F)	8.6±1.1 (I)	15.3±1.8 (L)	66.4±2.3 (M)	2.4±1.2 (K)	8.4±0.7 (J)	1.9±0.4 (C)	4±0.5 (G)	1.6±0.3 (B)
	ITS2	EPQ	37.9±1.9 (G)	30.7±2.1 (D)	26.3±1.7 (B)	36.4±1.6 (I)	37.6±1.7 (J)	30.6±2.5 (F)	35.1±1.2 (H)	30.4±1.4 (E)	43.7±9.5 (K)	27.2±1.8 (A)	29.1±1.8 (C)	26.6±1.8 (A)	30.2±2 (D)
		MC	50.4±2.6 (J)	41.6±2.8 (F)	35.7±2.1 (D)	50.4±2.2 (L)	51.1±2.7 (K)	39.3±3.2 (G)	41.6±3 (I)	15.2±1.9 (A)	56.5±13.7 (M)	34.3±1.9 (B)	39.8±2.4 (E)	35.4±2.1 (C)	41.5±2.7 (H)
		OC	6.7±1.5 (H)	3.2±0.7 (E)	2.7±0.8 (D)	1.2±0.2 (A)	3.7±1 (F)	8.9±1.8 (I)	18.9±3.5 (L)	68.5±2.2 (M)	11.5±1.4 (K)	9.6±2 (J)	2.2±0.5 (C)	4.5±1.2 (G)	1.8±0.4 (B)
SILVA	Full	EPQ	26.3±2.9 (B)	26.5±3.6 (C)	22.8±2.3 (B)		32.5±11.6 (D)	25.6±1.9 (D)	66.3±20.4 (F)	70.3±0.1 (E)		29.1±3.9 (C)	20.9±2.5 (A)		
		MC	26.4±1.8 (A)	29.7±2.3 (C)	31.9±1.6 (E)		39.3±19.5 (E)	28.8±2.3 (D)	66.6±22.4 (F)	98.3±0.2 (G)		26.7±2.2 (A)	28.3±2 (B)		
		OC	26.5±6.5 (F)	19.3±6.2 (CD)	1.8±1.3 (B)		15.6±9.4 (C)	17.7±4.3 (D)	65.6±23.8 (H)	0.1±0.1 (E)		35.2±8.5 (G)	3.7±1.5 (A)		
	V3-4	EPQ	27.9±1.2 (B)	29.8±3.5 (C)	29.9±3 (B)		29.7±2.9 (D)	32.6±2.9 (D)	31.3±2.1 (F)	20.5±1.3 (E)		29.1±3.4 (C)	26±2.5 (A)		
		MC	31.2±1 (A)	35.7±3.4 (C)	39.6±3.4 (E)		34.6±3.6 (E)	39.2±4.2 (D)	33.2±4 (F)	17.2±1.2 (G)		32.2±2.7 (A)	34.9±3 (B)		
		OC	20.8±4 (F)	16.4±4.1 (CD)	7.9±3.4 (B)		17.4±4.9 (C)	16±2.6 (D)	26.6±5.6 (H)	29±2.3 (E)		21.6±15.3 (G)	5.6±1.7 (A)		
	V4	EPQ	28.6±2.7 (B)	29.6±2.4 (C)	30.4±1.6 (B)		30±1.9 (D)	32.3±2.6 (D)	29.6±2 (F)	22.1±2.1 (E)		28.5±3.3 (C)	25.3±1.8 (A)		
		MC	33.1±3.5 (A)	38.3±3.1 (C)	39.4±1.4 (E)		37.6±3.1 (E)	39.8±3.6 (D)	32.7±2 (F)	20.8±2.3 (G)		32.8±2.5 (A)	34.4±2.6 (B)		
		OC	18.8±4.1 (F)	9.8±5.5 (CD)	9.8±3.7 (B)		10.8±5 (C)	13.5±2.3 (D)	21.8±3.4 (H)	25.4±2.9 (E)		18.7±13 (G)	4.8±1.3 (A)		

Table 2.1 Classification performance of each classifier, for each database, region, and partition level.

Values are percentages: mean±SD, with entries sharing letters are not significantly different at FDR < 0.01 for a given database, region, and partition level, as determined by a generalized linear mixed model using a binomial distribution, with region and classifier as random effects and partition iterations as a blocking effect. Performance metrics are defined in Supplementary Information. CB - CONSTAX with BLAST, CBC - CONSTAX with BLAST and conservative rule, CU - CONSTAX with UTAX, CUC - CONSTAX with UTAX and conservative rule.

Table 2.1 (cont'd)

Partition level			Genus												
		Classifier	BLAST	RDP	SINTAX	UTAX	mothur-wang	mothur-knn=3	qiime2-Naive-Bayes	Kraken2	SPINGO	CB	CBC	CU	CUC
UNITE	Full	EPQ	19±0.8 (G)	21.2±0.8 (D)	22.6±0.9 (C)	34.3±1.8 (H)	38.9±1.9 (K)	35.7±1.9 (I)	38.1±1.9 (J)	20.8±2.4 (F)	42.1±11.8 (L)	18.9±0.7 (A)	20.2±0.8 (C)	20.9±0.8 (B)	22.9±0.8 (E)
		MC	22±1 (G)	23.2±0.8 (E)	26.3±1 (D)	39.9±2.1 (H)	43.2±1.8 (K)	41.4±2.4 (I)	40.5±1.9 (J)	14.8±4.5 (A)	44.7±15.4 (L)	20.4±0.7 (B)	23.5±0.9 (D)	22.8±0.8 (C)	26.6±1 (F)
		OC	0.6±0.2 (C)	9.2±0.8 (D)	0±0 (ABCDEFGH)	0.4±0.1 (A)	13±4.5 (E)	1.5±1.3 (B)	23.5±4.7 (G)	56.9±10.6 (H)	26.7±9.9 (F)	9.6±0.9 (E)	0.1±0 (A)	9.3±0.8 (D)	0.4±0.1 (A)
	ITS1	EPQ	35.1±1.8 (G)	29.2±1.1 (D)	26.2±0.8 (C)	39.7±1.2 (H)	48.5±2.9 (K)	39±2.2 (I)	47.8±9 (J)	31±6.1 (F)	78.9±7.8 (L)	25.6±1 (A)	27.8±1.1 (C)	26.4±0.9 (B)	28.8±1.1 (E)
		MC	40.3±2 (G)	33.3±1.3 (E)	30.6±1 (D)	46.3±1.4 (H)	55.7±3.5 (K)	45.2±2.8 (I)	53.4±11.8 (J)	28.5±9.3 (A)	91.8±9.7 (L)	28.5±1.1 (B)	32.3±1.3 (D)	30±1 (C)	33.5±1.3 (F)
		OC	4.2±0.4 (C)	4.7±0.5 (D)	0±0 (ABCDEFGH)	0.7±0.2 (A)	5.4±1.2 (E)	2±1.8 (B)	14.1±8.6 (G)	46.3±13.1 (H)	2±3.9 (F)	8±0.8 (E)	0.8±0.1 (A)	4.7±0.5 (D)	0.7±0.1 (A)
	ITS2	EPQ	30.7±1.9 (G)	26.2±1.3 (D)	23.4±1 (C)	36.6±1.1 (H)	46.5±1.4 (K)	38.5±2.3 (I)	42±2 (J)	27.9±5.1 (F)	48.9±11.4 (L)	23±1.1 (A)	24.7±1.2 (C)	23.7±1.1 (B)	25.8±1.2 (E)
		MC	35.3±2.2 (G)	29.8±1.3 (E)	27.2±1.1 (D)	42.6±1.3 (H)	53.1±1.7 (K)	44.5±2.8 (I)	45.7±2.2 (J)	24.7±7.9 (A)	54.3±14 (L)	25.6±1.1 (B)	28.7±1.4 (D)	26.8±1.1 (C)	30±1.4 (F)
		OC	3.5±0.4 (C)	4.7±1 (D)	0±0 (ABCDEFGH)	0.7±0.2 (A)	6.5±0.5 (E)	2.2±2 (B)	19.8±2.5 (G)	47.5±12.3 (H)	16.3±4.5 (F)	7.6±1.1 (E)	0.6±0.1 (A)	4.7±1 (D)	0.6±0.2 (A)
SILVA	Full	EPQ	22.6±1.2 (B)	24.7±3 (C)	25.1±1.3 (D)		27.7±8.3 (E)	34.3±13.8 (G)	37±22.6 (F)	84.6±0.2 (H)		25.5±2.3 (C)	21.9±2.1 (A)		
		MC	22.7±1.7 (A)	25.1±2.8 (BC)	29.3±1.5 (E)		29.7±9.3 (E)	38±17.2 (F)	35.6±23.5 (D)	98.7±0.2 (G)		24.3±2.4 (B)	25.5±2.4 (C)		
		OC	22.4±2.6 (F)	22.2±4 (D)	0±0 (ABCDEFG)		16.1±10 (B)	12.1±7.1 (C)	45.5±17.2 (G)	0±0 (ABCDEFG)		32.8±2.8 (E)	0.3±0.2 (A)		
	V3-4	EPQ	23.7±1.7 (B)	25.7±2.5 (C)	26.9±2.4 (D)		29.1±6.8 (E)	30.3±4.6 (G)	25.6±2.4 (F)	23.3±3.3 (H)		24.7±3.2 (C)	23.2±2 (A)		
		MC	24±1.9 (A)	27.2±2.5 (BC)	31.3±2.8 (E)		32.7±8 (E)	33.4±5.7 (F)	26±2.2 (D)	27.2±3.8 (G)		26.7±1.8 (B)	27.1±2.3 (C)		
		OC	21.9±6.1 (F)	16.2±4.9 (D)	0±0 (ABCDEFG)		7.7±2.6 (B)	11.8±3.4 (C)	23.7±7.3 (G)	0±0 (ABCDEFG)		13.2±19 (E)	0±0 (A)		
	V4	EPQ	25.9±1.4 (B)	26.1±2.2 (C)	28.8±1.4 (D)		29.5±7.6 (E)	32.3±4 (G)	26.1±1.8 (F)	26.6±2.6 (H)		25.8±2.3 (C)	24.9±2 (A)		
		MC	27.2±1.4 (A)	28.9±2.4 (BC)	33.7±1.6 (E)		33.1±8.8 (E)	35.6±4.8 (F)	27.8±2.2 (D)	31±3 (G)		28.3±1.6 (B)	29±2.3 (C)		
		OC	18.4±4.5 (F)	9.5±3.3 (D)	0±0 (ABCDEFG)		7.5±4.3 (B)	12.6±4 (C)	15.4±4.5 (G)	0±0 (ABCDEFG)		10.5±15.2 (E)	0.3±0.6 (A)		

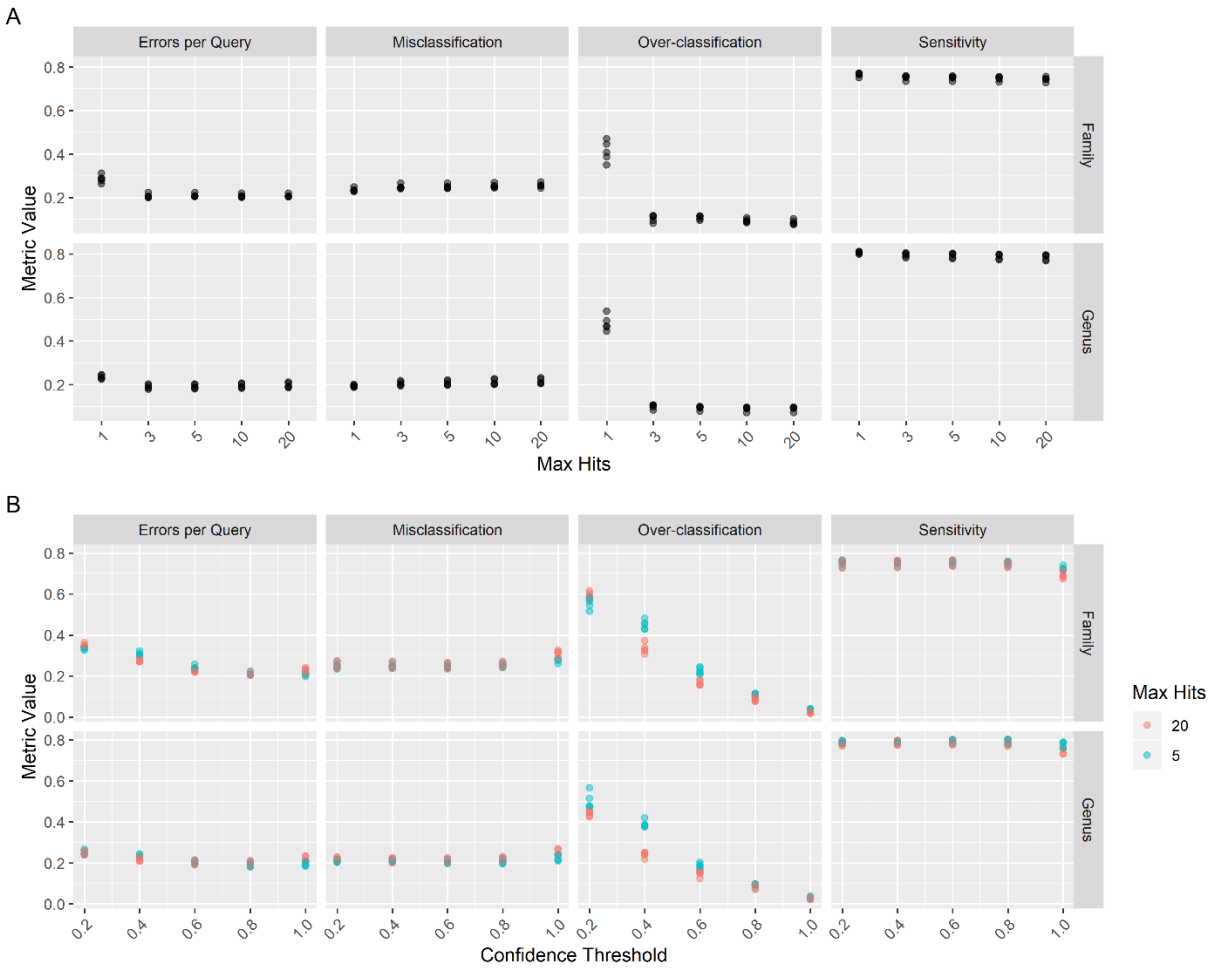

Figure 2.2 Effects of max hits and confidence threshold parameters on UNITE classification.
Errors per Query, Misclassification, Over-classification, and Sensitivity were determined using Clade-Partition Cross Validation while varying the "--mhits" (A) or "--conf" (B) parameters on 1000 query sequences from the UNITE Fungi database. Confidence threshold effects were compared at both 5 and 20 max hits.

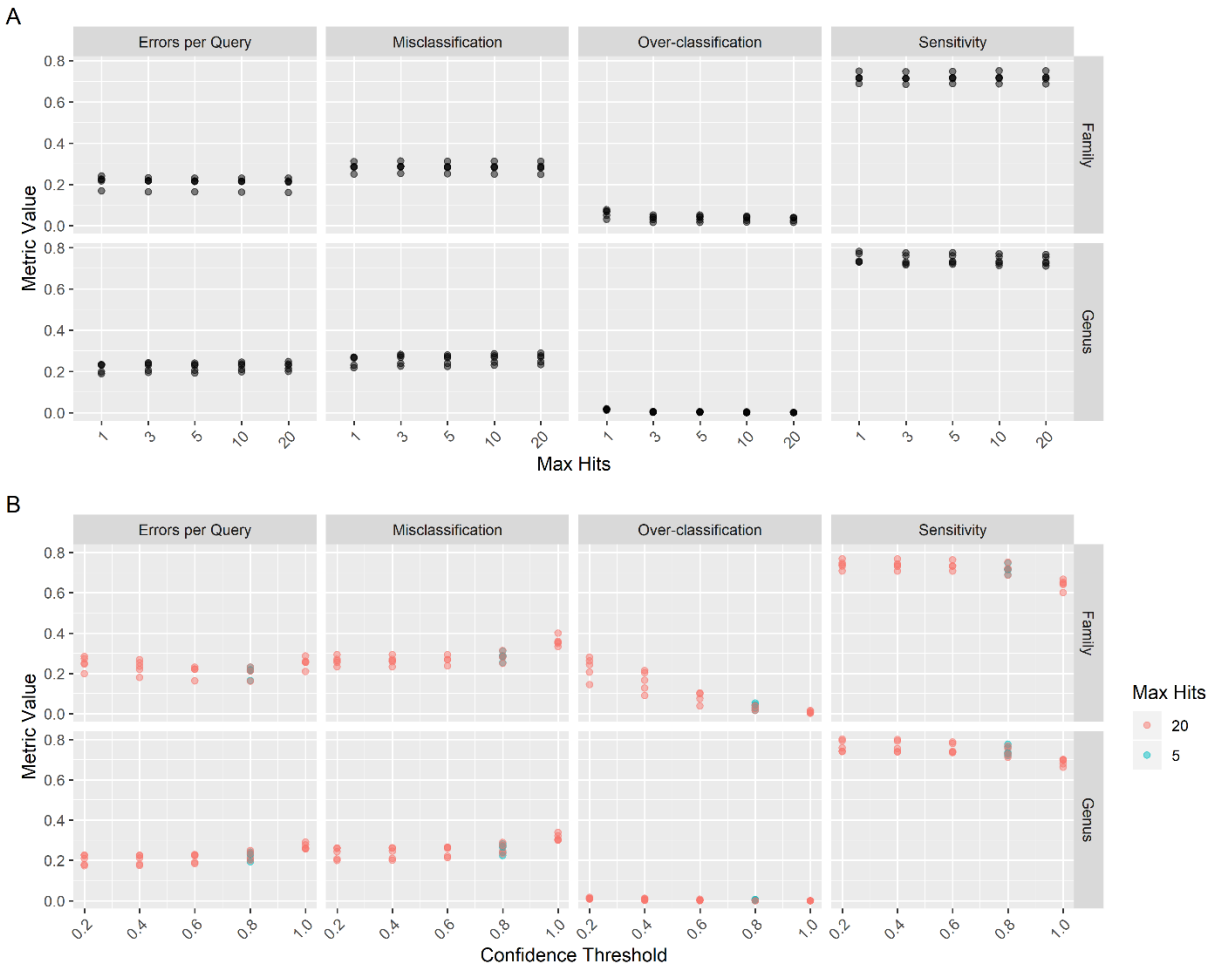

Figure 2.3 Effect of max hits and confidence threshold parameters on SILVA classification. Errors per Query, Misclassification, Over-classification, and Sensitivity were determined using Clade-Partition Cross Validation while varying the "--mhits" (A) or "--conf" (B) parameters on 1000 query sequences from the SILVA SSURef release 138.

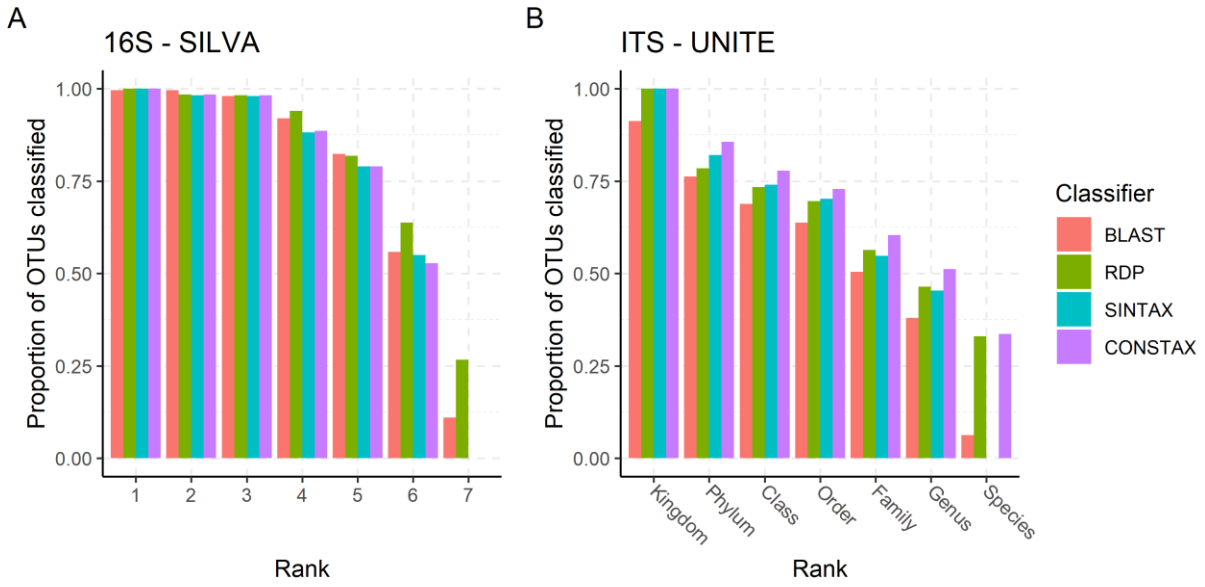

Figure 2.4 Classification counts for each classifier and the CONSTAX classification. OTUs from Benucci et al. 2020 (Benucci *et al.*, 2020), 500 each from bacterial and fungal libraries, which have been classified using databases for bacteria and fungi at recommended settings. Counts indicate the number of OTUs which had a taxon assigned at or above the confidence threshold of 0.8 at each rank. For bacteria, rank 1 corresponds to domain and decreases with higher rank numbers.

Implementation

CONSTAX is released as a conda-installable command-line tool, available from the bioconda installation channel (Grüning *et al.*, 2018) for LinuxOS, MacOS, and WSL systems. It is installed with the command "conda install -c bioconda constax", see https://github.com/liberjul/CONSTAXv2. CONSTAX requires two files: 1) "-d, --db" a reference database file in FASTA format (Pearson and Lipman, 1988) with header lines containing taxonomy of the sequences in SILVA (Glöckner *et al.*, 2017) or UNITE (Nilsson *et al.*, 2019) style, and 2) "-i, --input" an input file of user-submitted query sequences in FASTA format. This version implements a BLAST classification algorithm instead of the legacy UTAX classifier if the "-b, --blast" flag is used.

The user may designate several additional parameters, including confidence threshold for assignment ("-c, --conf"), BLAST classifier parameters, and whether to use a conservative

consensus rule ("--conservative"), which requires agreement of two (instead of one) non-null assignments to assign a taxonomy at the given rank. CONSTAX offers multithreaded classification with the argument, "-n, --num_threads".

CONSTAX generates three directories while running: 1) training files directory ("-f, --trainfile"), 2) taxonomy assignments directory ("-x, --tax"), and 3) an output directory ("-o, --output"). Prior to classifying sequences, training must be performed on any newly used database file with the "-t, --train" flag. After initial training, generated training files can be used in any later run by designating the same training files directory. When training is performed, CONSTAX will automatically generate formatted database files required by each classifier, as long as the supplied database has SILVA or UNITE header formatting. Following training, the classification or search command is performed for each classifier, and files are output to the taxonomic assignments directory. Finally, each classification output is reformatted and used to generate a consensus hierarchical taxonomy, then each classifier's result and the consensus result are stored in the output directory as tab-delimited value files with each row corresponding to a query sequence and values as the hierarchical taxonomy assigned to each query.

CONSTAX2 offers two additional features: 1) the ability to match input sequences to isolates using the "--isolates" option; and 2) the ability to determine higher-level taxonomy using representative databases with the "--high_level_db" option. Both approaches implement the BLAST algorithm to associate input sequences with hits from the respective databases, returning a single best hit with a default threshold of query cover $\geq 75\%$ and E value ≤ 10. Cutoffs for query coverage and percent identity can be specified. Isolate matching is designed to find best matches to sequenced organisms in pure culture, which may streamline culture-dependent and culture-independent analyses, and can also be used to implicate potential contamination by

association with known isolates previously worked with in the laboratory or sequencing facility where the samples were processed. Higher-level taxonomy designations are also useful in filtering host, organelles, or non-target taxa, which may show up in rDNA surveys. For 16S rDNA prokaryote datasets the latest SILVA SSURef NR99 database is recommended, while the latest UNITE Eukaryotes database is recommended for ITS studies of Fungi.

Conclusion

The newest implementation of CONSTAX offers improvement over its predecessor by ease of use, and improved applicability and accuracy. Hierarchical taxonomy classification accuracy by a consensus approach in CONSTAX2 is demonstrated to outperform commonly used classifiers while remaining computationally feasible.

CHAPTER 3: LEAF LITTER FUNGAL COMMUNITIES REFLECT PRE-SENESCENT LEAF COMMUNITIES IN A TEMPERATE FOREST ECOSYSTEM

Abstract

Fungal communities are known to contribute to living plant microbiomes and affect vital ecosystem services, such as pathogen resistance and nutrient cycling. Yet, factors that drive structure and function of phyllosphere mycobiomes and their fate in leaf litter require further research. We sought to determine the factors contributing to the composition of communities in temperate forest substrates, with culture-independent amplicon sequencing of fungal communities of pre-senescent leaf surfaces, internal tissues, leaf litter, underlying humus soil of co-occurring red maple (*Acer rubrum*) and shagbark hickory (*Carya ovata*). Paired samples were taken at five sites within a temperate forest in southern Michigan, USA. Fungal communities were differentiable based on substrate, host species, and site with significant variable interactions. Ordination and co-occurrence of taxa indicate that soil communities are unique from both phyllosphere and leaf litter communities. Correspondence of endophyte, epiphyte, and litter communities suggests dispersal of fungal taxa between these niches. Fungal communities are known to contribute to living plant microbiomes and affect vital ecosystem services, such as pathogen resistance and nutrient cycling. Yet, factors that drive structure and function of phyllosphere mycobiomes and their fate in leaf litter require further research. We sought to determine the factors contributing to the composition of communities in temperate forest substrates, with culture-independent amplicon sequencing of fungal communities of pre-senescent leaf surfaces, internal tissues, leaf litter, and underlying humus soil. Two host plant species were sampled at five sites within a temperate forest in southern Michigan, USA. Fungal communities were differentiable based on substrate, host species, and site with significant

variable interactions. Ordination and co-occurrence of taxa indicate that soil communities are unique from both phyllosphere and leaf litter communities. Correspondence of endophyte, epiphyte, and litter communities suggests dispersal of fungal taxa between these niches.

Introduction

Plant and soil microbiomes contribute to critical ecosystem functions (Wagg *et al.*, 2019; Delgado-Baquerizo *et al.*, 2016) and complex interactions within communities (Regalado *et al.*, 2020). Fungi are important members of these microbial communities, and collectively they inhabit leaf and twig surfaces, rhizosphere, soil, and dead plant tissues. The community of fungi present in the plant microbiome can colonize healthy roots, leaves, stems, and seeds (Porras-Alfaro and Bayman, 2011; Floc'h *et al.*, 2020) and have been implicated in disease susceptibility (Gu *et al.*, 2020), nutrient acquisition and cycling (Herzog *et al.*, 2019), and stress tolerance or resilience (Waller *et al.*, 2005; Márquez *et al.*, 2007). Yet, fungi also occupy several critical roles in the plant and soil microbial communities, functioning as plant pathogens (Brader *et al.*, 2017), hyperparasites (Falk *et al.*, 1995; Vandermeer *et al.*, 2009), mycorrhizal associates (Soudzilovskaia *et al.*, 2019), saprobes (Zhou and Hyde, 2001), and specialized litter (Osono, 2007) or wood degraders (Schilling *et al.*, 2020). Studying the distribution of fungal taxa within environments and between living and non-living substrates may elucidate the natural history and functional ecology of these organisms (Peay, 2014).

Fungal endophytes are defined as living asymptomatically within plant tissues (Arnold *et al.*, 2000). The diversity and composition of endophytes can be distinguished from epiphytic communities, the microbes observed on surfaces of plant tissues, but some overlapping taxa are common (Gomes *et al.*, 2018; Yao *et al.*, 2019). In fact, in many cases epiphyte communities are presumed to give rise to endophyte communities, as they penetrate openings, such as stomata or

damaged tissues (Porras-Alfaro and Bayman, 2011). Some authors have hypothesized potential ecological roles of endophytic fungi, including as plant pathogens, latent saprotrophs, or mutualists (Veneault-Fourrey and Martin, 2011; Chen *et al.*, 2018; Brader *et al.*, 2017). Following senescence of above-ground tissues, some of the endophytes and epiphytes (collectively the phyllosphere community) may switch nutritional modes to become active saprotrophs (Promputtha *et al.*, 2007; He *et al.*, 2012; Gundel *et al.*, 2017) in forest floor leaf litter communities.

Examinations of plant-associated fungal communities have implicated multiple drivers in community composition and structure. Rhizosphere communities have shown to be structured in-part by plant host species, location, and land-use effects (Schöps *et al.*, 2020; Bonito *et al.*, 2019). Similarly, aboveground tree endophytes showed significant discrimination based on plant tissue, host species, and site in a hemiboreal forest (Küngas *et al.*, 2020). Similar patterns have been observed for leaf endophytes, which may be associated with leaf secondary metabolites in some cases (Christian *et al.*, 2020).

In the current study, we ask whether endophytic and epiphytic fungal phyllosphere communities of forest tree species differ, and whether phyllosphere fungal taxa persist in leaf litter or soil communities. To address these questions, fungal communities of pre-senescent leaves, leaf litter and soils were assessed with high-throughput amplicon sequencing and compared. We compared fungal communities of red maple (*Acer rubrum* L.) and shagbark hickory (*Carya ovata* (Mill.) K.Koch) host tree species at five sites within a forest ecosystem in southern Michigan where the species co-occur. Based on previous studies of fungal endophytes, we hypothesized that phyllosphere fungal communities between the two plant species would be similar. We also expected that many detected taxa would persist in leaf litter samples, and less so

in soils. To determine how these niches structured communities, we sampled across four forest substrates: 1) the surface of pre-senescent leaves, 2) the internal tissues of pre-senescent leaves, 3) leaf litter, and 4) the soil in direct contact with the leaf litter.

Sampling protocols and downstream sample processing can have overlooked effects on the microbial community analyses (Hallmaier-Wacker *et al.*, 2018). The methods and materials chosen can bias the community observed through DNA extraction and PCR/primer decisions (Brooks *et al.*, 2015). While much research has focused on comparison of kit or DNA extraction methods (Angebault *et al.*, 2020; Brooks *et al.*, 2015; Hallmaier-Wacker *et al.*, 2018; Vo and Jedlicka, 2014), we did not find literature assessing swab material type for observation of phyllosphere communities. A second aim of this study was to compare the efficacy of swab types composed of two different materials in assessing epiphytic leaf communities with the goal of promoting cost-efficient yet consistent and thorough environmental sampling.

Methods

Site

Samples were collected within a 9-hectare quadrat in a semi-homogeneous deciduous forest site within Dansville State Game Area, Dansville, Michigan, USA (N 42.5171, W 84.3260, altitude 291 m). Soil types in the quadrat included loam, sandy loam, and loamy sand, and fall under the following USDA soil taxonomy classifications: Oxyaquic Glossudalfs, Typic Endoaquolls, Aquollic Hapludalfs, Aquic Arenic Hapludalfs (Figure 3.1). Mean annual climate observed at Jackson County Airport, 30.7 km SSW of the site, was 800 mm precipitation, 9.1°C temperature, and 69% relative humidity. Altitude measurements were performed using Google

Earth Pro (ver. 7.3.2). Distances between sites were determined using the “Measure distance” function in Google Maps.

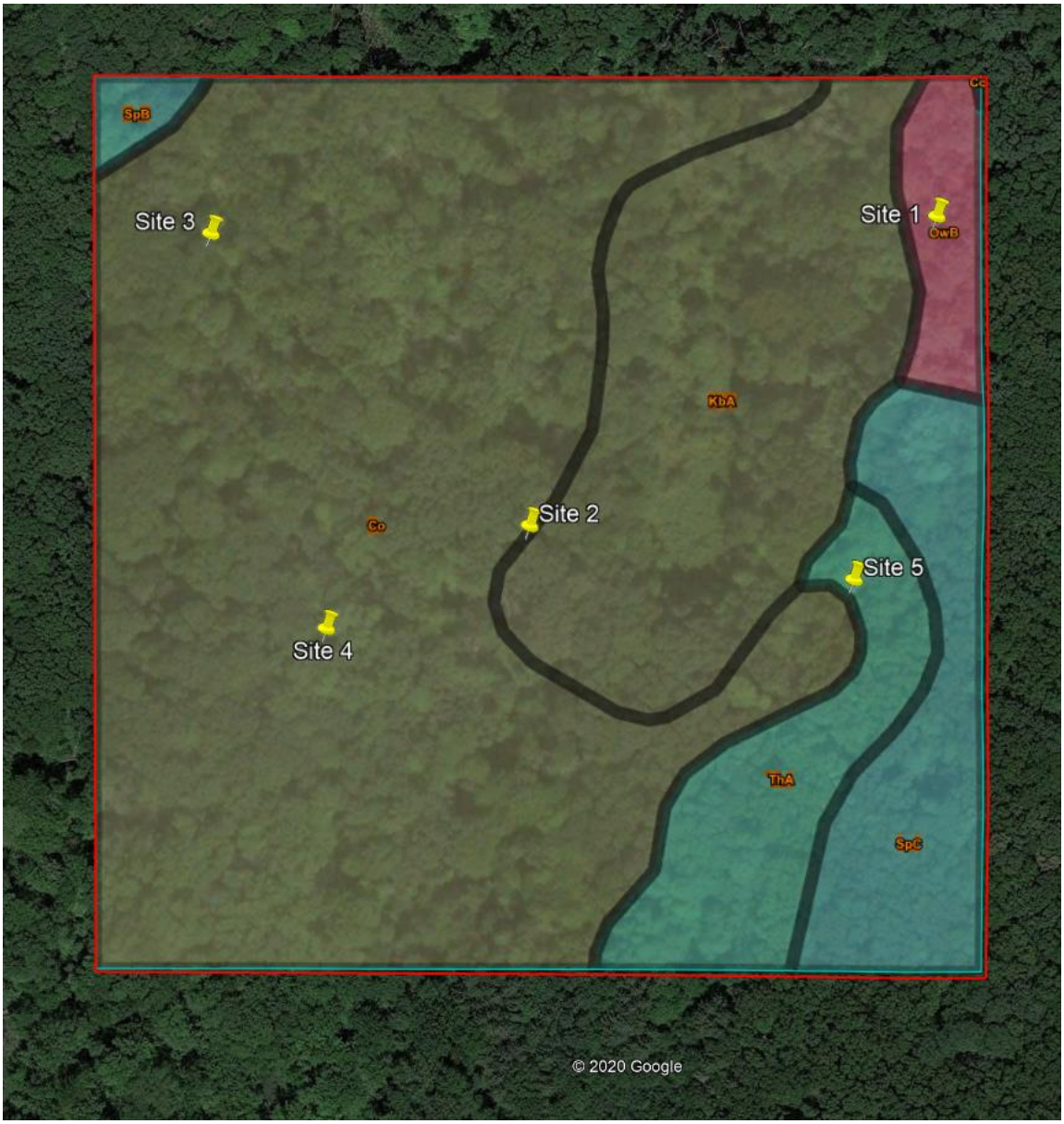

Figure 3.1 Soil types in the sampling quadrat.
(Co) Colwood-Brookston loams; (KbA) Kibbie loam, 0 to 3 percent slopes; (OwB) Owosso-Marlette sandy loams, 2 to 6 percent slopes; (SpB) Spinks loamy sand, 0 to 6 percent slopes; (SpC) Spinks loamy sand, 6 to 12 percent slopes, (ThA) Thetford loamy sand, 0 to 3 percent slopes.

Sample collection

Samples were collected on September 16, 2018 from five sites 76-246m apart (Table 3.1) selected within the study area described above. These include one site each in the northeast, northwest, southeast, southwest, and center areas on the quadrat. At each site at least one shagbark hickory tree (*Carya ovata*), one red maple tree (*Acer rubrum*) and a suitable amount of host leaf litter beneath the canopy of sampled trees were present. We selected these trees as they were consistently found across the forest, providing the opportunity to test site effects. The litter collected was intact leaves of the target host, that were presumed to have fallen the previous year. Ten fresh leaves (*Acer*) or leaflets (*Carya*) per site and tree were collected axenically from branches within reaching or jumping height (<3 m from ground) and swabbed in the field, then transported to the laboratory for further processing. The entire surface of the leaf or leaflet, both top and bottom, was swabbed for those used. For each site and tree, three leaves or leaflets of litter were collected (separately for each host species), and the topsoil (mull humus) immediately below the leaves was collected with a sterilized metal scoop. All collected leaf and soil materials were stored in sealed plastic bags, placed in an insulated cooler, brought directly back to the lab, and stored at 4°C until processing (over the following week). All samples of a given substrate were processed on the same day.

	Site 2	**Site 3**	**Site 4**	**Site 5**
Site 1	170.43	242.77	246.20	123.66
Site 2	-	145.00	76.29	109.95
Site 3	-	-	137.90	244.50
Site 4	-	-	-	177.50

Table 3.1 Distances between sample sites.
Straight-line distances in meters are between row-column pairs of sites.

Ten leaves or leaflets of each tree at each site were swabbed with two swabs that were dipped in Extraction Solution (ES - 100 mM Tris, 250 mM KCl, 10 mM disodium EDTA,

adjusted to pH 9.5-10) directly prior to sampling, producing two extractions per tree and site. Leaves were swabbed with both cotton-tipped applicators and polymer-tipped PurFlock ULTRA® applicators (Puritan Medical Products, Guilford, Maine, USA) referred to as "cotton" and "synthetic", respectively. Swab heads were broken off into individual 2 mL microcentrifuge tubes and frozen at -20°C until processing. A sterile head of each swab type was placed into ES, exposed to air briefly (30 sec), and placed into the ES to serve as negative controls.

Sample preparation and DNA extraction

Prior to extraction, 500 µL ES was added to each tube containing a swab head. Tubes were heated to 95°C for 10 minutes to lyse cells. 500 µL of 3% bovine serum albumin (BSA) was then added to the tubes to stabilize the reaction. This product was mixed, spun down, and the supernatant was used as template for subsequent PCR reactions (Bonito *et al.*, 2017). To clean and surface sanitize leaves, collected leaf materials were first placed in a solution composed of 10% bleach (0.6% active sodium hypochlorite) and 0.1% Tween 20 and agitated for 7 minutes, followed by rinsing in sterile water and drying with sterile filter paper. Surface-sanitized leaves were placed in wax paper bags and lyophilized. Leaf litter was also lyophilized. Leaf litter was not treated to the same epiphyte/endophyte sampling given its non-living and fragile state. Leaves and leaf litter were pooled by host species and site. These composite samples were ground with 6 mm ceramic beads in 50 mL centrifuge tubes for 5 - 10 minutes using a modified paint shaker (DC-1-C, Miracle Paint Rejuvenator Co, Grove Heights, MN; modified by the Michigan State University Physics Shop).

DNA was extracted from lyophilized plant tissue using the Mag-Bind® Plant DNA Kit (Omega Bio-tek, Norcross, GA, USA), with the recommended ~15 mg of dry tissue. Soil samples were dried with silica gel beads and homogenized, then DNA was extracted from ~0.5 g

of processed soils with the PowerMag® Soil DNA Isolation Kit (Qiagen, Carlsbad, CA, USA) following manufacturer's recommendations.

Fungal amplicon libraries were generated with ITS1f-ITS4 primers (Gardes and Bruns, 1993; White *et al.*, 1990) and DreamTaq Green DNA Polymerase (ThermoFisher Scientific, USA) following previously described protocols (Lundberg *et al.*, 2013; Chen *et al.*, 2018). Six PCR no-template negative controls were included in library preparation and sequencing. PCR products were visualized under UV light on an ethidium bromide-stained 0.9% agarose gel after separation by electrophoresis. DNA concentrations of samples were normalized with a SequalPrep™ Normalization Plate Kit (ThermoFisher Scientific, USA) and samples were then pooled into a single library. Amplicons were then concentrated 20:1 with Amicon® Ultra 0.5 mL 50K filters (EMDmillipore, Germany) and purified with Agencourt AMPure XP magnetic beads (Beckman Coulter, USA). A synthetic mock community with 12 taxa and 4 negative (no DNA added) controls were included to assess sequencing quality (Palmer *et al.*, 2018). Amplicons were sequenced on an Illumina MiSeq analyzer using the v3 600 cycles kit (Illumina, USA).

Sequence analysis

Read quality was assessed with FastQC (Andrews, 2010). Sequences were then demultiplexed in QIIME (Caporaso *et al.*, 2010). Primers were trimmed with Cutadapt 1.18 (Martin, 2011), and the conserved regions trimmed with usearch -fastq_filter (Edgar, 2010). Filtered and trimmed reads were then clustered into operational taxonomic units (OTUs) based on 97% sequence similarity with the uparse pipeline (Edgar, 2013). Fungal OTU classification was determined with the CONSTAX2 consensus technique comparing RDP Classifier, BLAST, and SINTAX algorithms trained on the UNITE fungal general release dataset from Feb 04, 2020 (Abarenkov *et al.*, 2020) with 80% confidence threshold and recommended settings (Liber *et al.*,

2021). Statistical analyses and plots were prepared in R version 3.6.1 (R Core Team, 2019) using the following packages for plotting and data handling: ggplot2 3.2.1 (Wickham, 2016), patchwork 1.0.0 (Pedersen, 2019), tidyr 1.0.2 (Wickham and Henry, 2020), dplyr 0.8.3 (Wickham, François, *et al.*, 2019), purrr 0.3.2 (Henry and Wickham, 2019), ggpubr 0.2.3 (Kassambara, 2019), gplots 3.0.3 (Warnes *et al.*, 2020). OTU data, tables, and scripts for statistical analysis and plotting are available at: https://github.com/liberjul/Leaf_litter_communities.

Further identification of indicator OTUs that were poorly classified was completed based upon BLASTn searches (Altschul *et al.*, 1990) against the Fungal RefSeq ITS nucleotide database with default search settings. Taxa were classified based the following prioritized rules: 1) If 1 taxa was greater than 99.5% identity and 100% query cover, the species rank was assigned, 2) if multiple taxa were greater than 95% identity and 100% query cover, the lowest rank in common was assigned, 3) if at least five taxa had greater than 80% identity and 100% query cover, the lowest rank in common was assigned, 4) if fewer than five taxa with 80% identity and 100% query cover, no taxa was assigned.

Ecological analysis

A rarefaction curve (rarecurve function in vegan 2.5-6) (Oksanen *et al.*, 2019) and OTU table were generated by rarefying to a minimum acceptable depth (function rrarefy from vegan). Within-sample (alpha) diversity was estimated with phyloseq 1.26.0 (McMurdie and Holmes, 2013) with both Shannon (Shannon, 1948) and Simpson (Simpson, 1949) diversity indices. Both indices are shown to be minimally biased, but the Shannon index weights rare species more heavily compared to the Simpson index (Mouillot and Leprêtre, 1999). Effects of substrate, host species, and site on within-sample diversity were examined with linear models. Site was used as

a variable to account for variation due to host genotype, edaphic effects, and microenvironment not captured by substrate or host species effects. The fit of these models was compared with AICctab from bbmle 1.0.23.1(Anderson and Burnham, 2002; Bolker and R Development Core Team, 2020). Fungal community differences within and between substrates were visualized based on Bray-Curtis dissimilarities (Bray and Curtis, 1957). A Venn diagram and a table of shared OTUs were constructed in VennDiagram 1.6.20 (Chen, 2018) with OTUs having at least 0.01% weighted abundance. Indicator taxa analysis was performed with rarefied OTU counts with the multipatt function in indicspecies 1.7.6 (De Cáceres and Legendre, 2009).

Ordinations were performed first with PCoA (function cmdscale in vegan) with Bray-Curtis dissimilarity, Sørensen-Dice similarity (Dice, 1945; Sørensen, 1948), and Jaccard similarity (Jaccard, 1901). Based on variance explained, Bray-Curtis dissimilarity was used for non-metric multidimensional scaling (NMDS) ordination (metaMDS in vegan). Substrate community compositions were compared with Bray-Curtis dissimilarity by PERMANOVA, for centroids (adonis2 function in vegan for full model, pairwise.perm.manova in RVAideMemoire 0.9-75 for pairwise comparisons) (Hervé, 2020), and by PERMDISP, for dispersion (betadisper function in vegan).

Swab material comparisons

To compare the communities sampled by each swab material, PERMANOVA and PERMDISP were performed as described above for substrate, host species, and site comparisons. Swab materials were further compared by the number of reads sequenced per sample. The number of reads sequenced were modeled with a negative binomial mixed model with "glmer.nb" from package lme4 1.1-21 (Bates *et al.*, 2015), with swab material as a fixed effect and leaf sampled as a random intercept effect.

Results

Samples and sequencing

In this study, a total of 50 samples were collected for fungal community analysis. These were derived from each of four substrates: leaf surface (20), leaf tissue (10), leaf litter (10), or soil (10), referred to hereafter as “epiphyte”, “endophyte”, “litter”, and “soil”. From these samples, amplicon sequencing generated 8,184,524 forward reads, which were reduced to 5,859,924 after quality filtering and trimming. Raw reads have been deposited in NCBI SRA under BioProject PRJNA632843.

Clustering OTUs at 97% sequence similarity resulted in 3402 OTUs. Any OTUs for which more than half of reads occurred in negative controls were labelled as possible cross-contamination and excluded from further analysis, leaving 3366 OTUs in our sample matrix. A rarefaction curve was performed (Figure 3.2), which informed the rarefaction to 5926 reads per sample. After sample selection and rarefaction, 18 epiphyte, 10 endophyte, 8 leaf litter, and 10 soil samples remained.

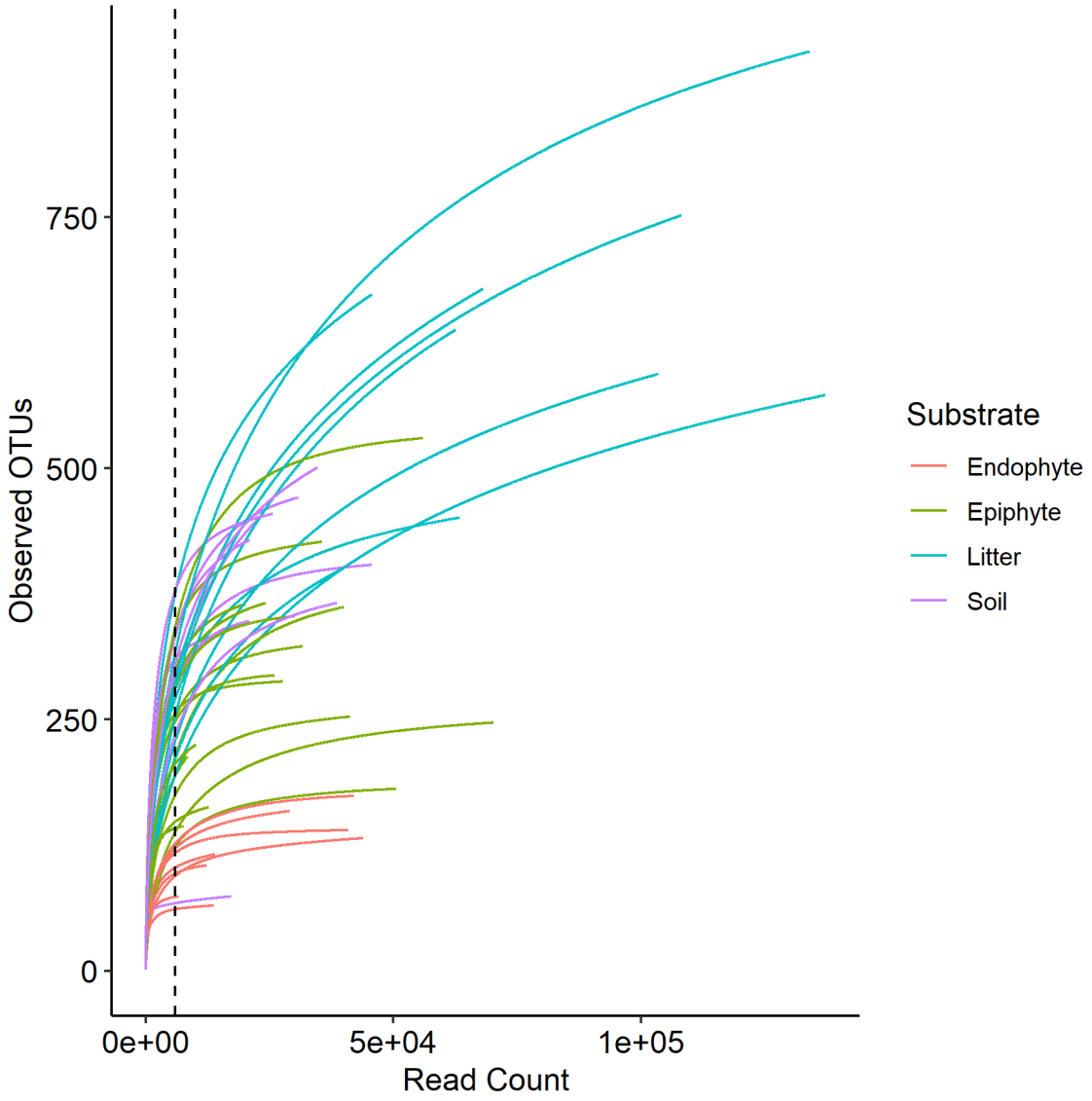

Figure 3.2 Rarefaction curves.
Curves were determined using "rrarefy" in vegan to assess rarefaction depth to use for downstream analyses. A vertical line is positioned at the applied rarefaction depth.

Comparison of substrates, host species, and site

Rarefied OTU counts were ordinated through non-metric multidimensional scaling (NMDS) based upon Bray-Curtis dissimilarity. Bray-Curtis dissimilarity explained the most variance (34.6%) in the first two axes when ordinated with PCoA, compared to abundance-based Sørensen-Dice (32.6%) or incidence-based Jaccard (25.7%) indices. Dispersion was significant

between all substrates ($p < 0.0001$) and pairwise (Tukey's HSD test, $p < 0.05$) between all pairs of substrates except Epiphyte-Endophyte ($p = 0.99$), and Soil-Litter ($p = 1.00$).

Centroids, compared with Bray-Curtis dissimilarity between rarefied OTU counts, were significantly different between all substrates ($p < 0.001$) and pairwise ($p < 0.005$) comparisons based upon PERMANOVA. Substrate was the largest contributing variable to community composition ($R^2 = 0.308$, $p < 0.001$), while host species ($R^2 = 0.075$, $p < 0.001$) and site ($R^2 = 0.086$, $p < 0.001$) were also significant but contributed less (Table 3.2). All interactions were significant ($p < 0.001$) with substrate x site ($R^2 = 0.206$), substrate x host species x site ($R^2 = 0.122$), and substrate x host species ($R^2 = 0.106$) accounting for a substantial part of the variability. Mean within-sample (alpha) diversity was significantly different by substrate (ANOVA, $p < 0.01$) and by host species ($p < 0.05$) for both Shannon and Inverse Simpson indices, but notably not by site ($p > 0.10$) (Figure 3.3, Table 3.3).

All-factor PERMANOVA test statistics			
Factor	R^2	F	Pr(>F)
Substrate	0.3079	19.62	0.001
Host species	0.0751	14.37	0.001
Site	0.0864	4.13	0.001
Substrate:Host species	0.1057	6.74	0.001
Substrate:Site	0.2063	3.29	0.001
Host species:Site	0.0551	2.63	0.002
Substrate:Host species:Site	0.1216	2.32	0.001
Residual	0.0418		
Total	1		

Pairwise PERMANOVA comparisons by substrate			
	Epiphytes	**Litter**	**Soil**
Endophytes	0.0036	0.004	0.0015
Epiphytes	-	0.0015	0.0015
Litter	-	-	0.0015

Table 3.2 PERMANOVA tables of substrate, host species, site, and interactions effects on fungal communities both all factor and pairwise.
Included are both the full model comparison and *p*-values from permuted pairwise tests.

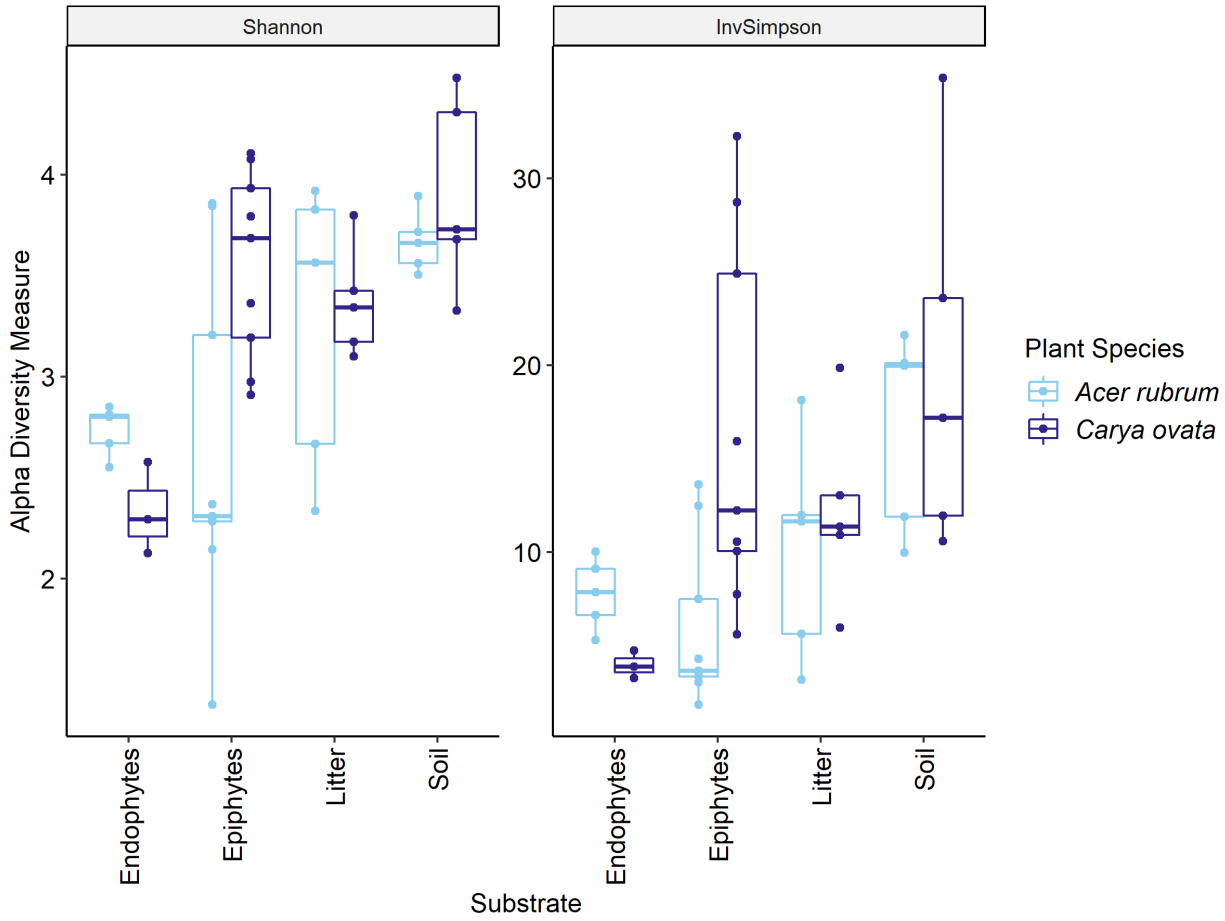

Figure 3.3 Within-sample diversity of samples by substrate and host species.
Shannon and Inverse Simpson estimates of within-sample (alpha) diversity for all samples, grouped by substrate and host species. Diversity estimates were determined using the phyloseq package.

Shannon Index					
Factor	**Df**	**Sum Sq**	**Mean Sq**	**F value**	**Pr(>F)**
Substrate	3	7.06	2.35	7.67	0.0004
Host species	1	1.52	1.52	4.97	0.0319
Site	4	2.46	0.62	2.01	0.1133
Residuals	37	11.34	0.31		
Inverse Simpson Index					
Factor	**Df**	**Sum Sq**	**Mean Sq**	**F value**	**Pr(>F)**
Substrate	3	670.35	223.45	4.73	0.0068
Host species	1	230.87	230.87	4.89	0.0333
Site	4	212.16	53.04	1.12	0.3605
Residuals	37	1747.35	47.23		

Table 3.3 ANOVA tables of substrate, host species, and site effects of within-sample diversity.
Diversity was estimated for Shannon and Inverse Simpson indices using diversity function in vegan.

Indicator taxa were determined for each of the substrates and substrate groupings, which shows that several leaf-associated taxa were unique to pre-senescent leaves and litter (Table 3.4). Several indicator taxa are significantly associated with phyllosphere and litter communities, occurring at high abundance across endophyte, epiphyte, and litter (Figure 3.4). Several taxa are also significantly associated with substrate-host combinations (Table 3.5). A Pearson correlation heatmap illustrates similarity between phyllosphere and some litter communities at the OTU-level (Figure 3.5).

Substrate	OTU	FDR	CONSTAX2 Result	BLAST Result
Epiphyte	OTU_150	0.003	Fungi sp.	*Ceramothyrium* sp.
	OTU_56	0.003	*Golubevia pallescens*	Unknown
	OTU_58	0.003	Fungi sp.	Unknown
	OTU_108	0.003	Mycosphaerellaceae sp.	*Acrodontium* sp.
	OTU_191	0.003	*Taphrina vestergrenii*	Unknown
Endophyte	OTU_48	0.01	*Phyllosticta minima*	*Phyllosticta* sp.
	OTU_83	0.012	*Seimatosporium* sp.	Seimatosporium sp.
	OTU_34	0.022	*Rhytisma* sp.	*Rhytismataceae sp.*
	OTU_683	0.033	*Zygophiala tardicrescens*	*Schizothyrium* sp.
	OTU_1104	0.034	*Hypoxylon carneum*	Unknown
Litter	OTU_23	0.003	*Colletotrichum* sp.	*Colletotrichum* sp.
	OTU_32	0.003	Xylariales sp.	Xylariales sp.
	OTU_40	0.003	*Codinaea lambertiae*	*Codinaea lambertiae*
	OTU_52	0.003	Ascomycota sp.	*Coleophoma* sp.
	OTU_112	0.003	*Lophiostoma* sp.	*Amorocoelophoma* sp.
Soil	OTU_19	0.003	*Saitozyma podzolica*	*Saitozyma* sp.
	OTU_41	0.003	*Archaeorhizomyces* sp.	*Archaeorhizomyces* sp.
	OTU_85	0.003	*Solicoccozyma terricola*	*Solicoccozyma terricola*
	OTU_93	0.003	*Trichoderma hamatum*	*Trichoderma* sp.
	OTU_785	0.003	*Pseudogymnoascus roseus*	*Pseudogymnoascus* sp.
Epiphyte + Endophyte	OTU_189	0.003	*Ramularia nyssicola*	*Ramularia nyssicola*
	OTU_479	0.003	*Ramularia* sp.	*Ramularia* sp.
	OTU_251	0.007	Ascomycota sp.	*Ramularia* sp.
	OTU_3326	0.009	*Ramularia* sp.	*Ramularia* sp.
	OTU_2	0.012	*Ampelomyces* sp.	Unknown
Epiphyte + Litter	OTU_66	0.003	*Bulleribasidium* sp.	*Bulleribasidium* sp.
	OTU_68	0.003	*Epicoccum* sp.	*Epicoccum* sp.
	OTU_82	0.003	Capnodiales sp.	Dothideomycetes sp.
	OTU_165	0.003	*Dioszegia athyri*	*Dioszegia* sp.
	OTU_226	0.003	Agaricales sp.	Unknown
Endophyte + Litter	OTU_87	0.033	Ascomycota sp.	*Paraconiothyrium* sp.
	OTU_1	0.033	Glomerellaceae sp.	*Colletotrichum* sp.
	OTU_127	0.035	Diaporthales sp.	*Diaporthe* sp.
Litter + Soil	OTU_380	0.003	Ascomycota sp.	Helotiales sp.
	OTU_282	0.013	*Striatibotrys eucylindrospora*	*Striatibotrys* sp.
	OTU_193	0.02	*Cylindrocladium peruvianum*	*Cylindrocladiella* sp.
	OTU_121	0.023	*Chalara* sp.	Leotiomycetes sp.
Epiphyte + Endophyte + Litter	OTU_4	0.003	*Ramularia* sp.	*Ramularia* sp.
	OTU_5	0.003	*Ramularia pratensis*	*Ramularia* sp.
	OTU_8	0.003	Dothideomycetes sp.	*Cladosporium* sp.
	OTU_10	0.003	Didymellaceae sp.	Didymellaceae sp.
	OTU_11	0.003	Pleosporales sp.	*Alternaria* sp.

Table 3.4 Substrate and cross-substrate indicator OTUs.
Indicator OTUs were determined using the multipatt function the indicspecies package. Substrate groupings with no significant indicator OTUs are not displayed, and only those with a false discovery rate (FDR) < 0.05 are included (maximum of 5 each). Classification against the UNITE with and Fungal RefSeq ITS databases with CONSTAX2 and BLASTn determined the closest known taxa.

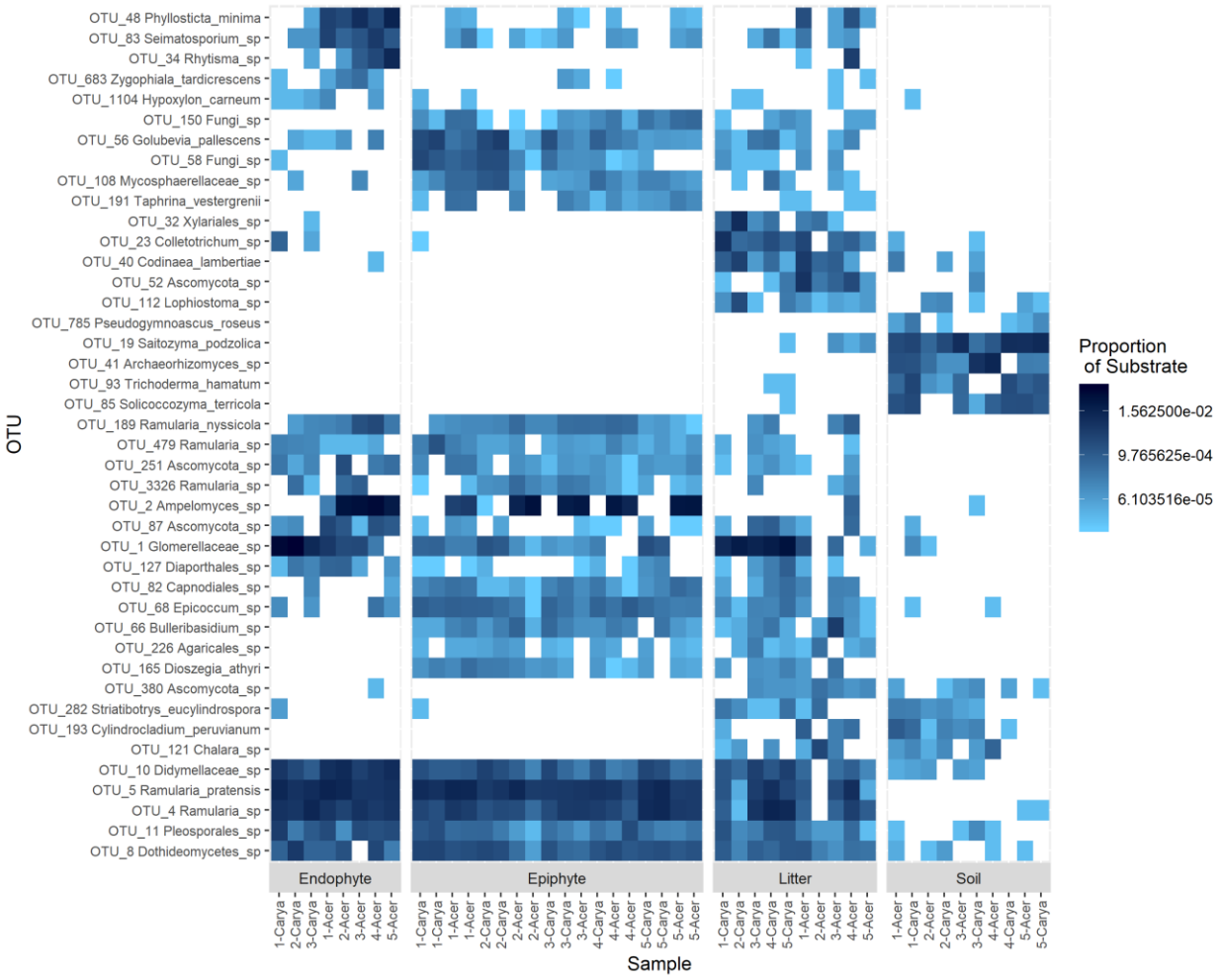

Figure 3.4 Heatmap of OTU abundance by substrate.
A total of 42 OTUs were included in the heatmap, all which were significant indicator taxa as determined using the multipatt function in the indicspecies package. Taxa names were determined using the CONSTAX2 classifier. Sample names are included beneath each column, with host species as either *Acer rubrum* (Acer) and *Carya ovata* (Carya) and site number designated in the name.

Substrate	Host	OTU	FDR	CONSTAX2 Result
Epiphyte	*Acer rubrum*	OTU_2	0.011	*Ampelomyces* sp.
		OTU_281	0.011	*Ramularia* sp.
		OTU_191	0.017	*Taphrina vestergrenii*
		OTU_149	0.017	Pseudeurotiaceae sp.
		OTU_381	0.017	*Taphrina* sp.
	Carya ovata	OTU_56	0.011	*Golubevia pallescens*
		OTU_12	0.011	*Erysiphe* sp.
		OTU_29	0.011	*Exobasidium* sp.
		OTU_68	0.011	*Epicoccum* sp.
		OTU_75	0.011	Fungi sp.
Endophyte	*Acer rubrum*	OTU_48	0.011	*Phyllosticta minima*
		OTU_83	0.011	*Seimatosporium* sp.
		OTU_20	0.011	*Plagiostoma* sp.
		OTU_116	0.011	*Angustimassarina acerina*
		OTU_461	0.017	*Venturia* sp.
	Carya ovata	OTU_332	0.011	*Sphaerulina* sp.
		OTU_270	0.023	Ascomycota sp.
		OTU_3279	0.039	*Sphaerulina* sp.
		OTU_359	0.04	Mycosphaerellaceae sp.
		OTU_105	0.042	*Sphaerulina* sp.
Litter	*Carya ovata*	OTU_50	0.011	Hypocreales sp.
Soil	*Carya ovata*	OTU_2619	0.011	*Trichoderma* sp.
		OTU_2124	0.03	Fungi sp.
		OTU_54	0.04	*Mortierella minutissima*
		OTU_893	0.048	*Penicillium* sp.

Table 3.5 Indicator taxa for combined substrate and host.
Up to 5 indicator taxa with FDR < 0.05 were determined with multipatt for each substrate-host combination. These taxa were identified with the CONSTAX2 classifier against the UNITE database.

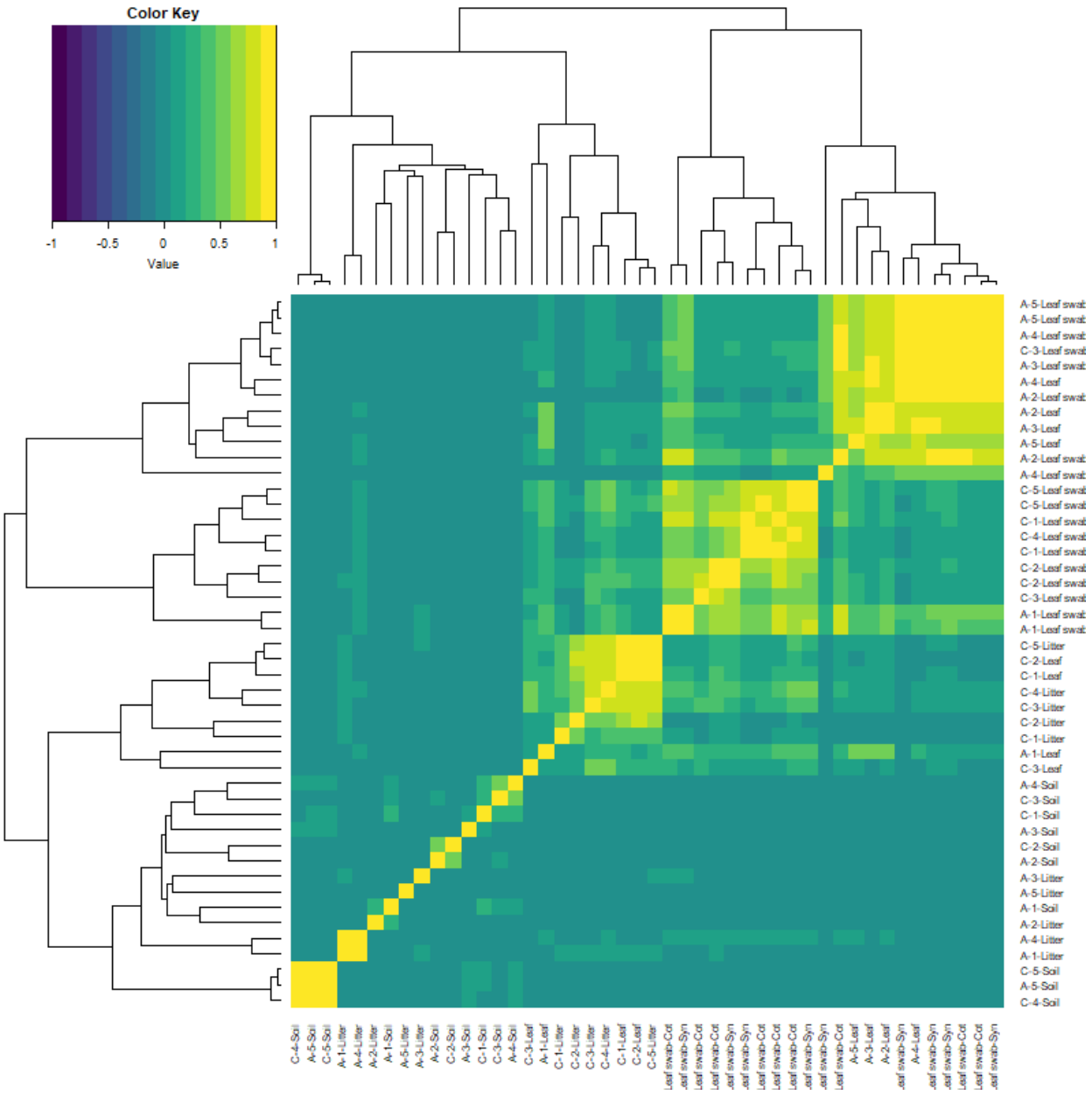

Figure 3.5 Pearson correlation heatmap of samples.
Heatmap coloration shows the Pearson r coefficient between samples using rarefied OTU counts.

A total of 657 OTUs were identified as having individual weighted abundances greater than 0.01%. The soil environment had the highest number of unique OTUs (212) and the lowest number of shared OTUs (152). The litter had the highest total number of OTUs (421) and the highest number of shared OTUs (357), although not the highest proportion of shared OTUs (Figure 3.6, Table 3.6). Community composition based upon the 30 most abundant genera across

all samples is consistent with expectations of substrate and host species affinity (Figure 3.7A). *Archaeorhizomycetes*, *Tuber*, *Inocybe* and other root-associated or ectomycorrhizal fungi (Tedersoo *et al.*, 2010; Menkis *et al.*, 2014) were limited to soil samples, while *Ramularia* and *Ampelomyces*, a leaf pathogen and leaf-occuring hyperparasite, respectively (Videira *et al.*, 2016; Falk *et al.*, 1995) were common in the phyllosphere.

	Epiphyte	**Endophyte**	**Litter**	**Soil**
Total OTUs	268	202	421	364
Unique OTUs	5 (2%)	2 (1%)	64 (15%)	212 (58%)
Shared with Epiphytes	-	180 (89%)	247 (59%)	57 (16%)
Shared with Endophytes	180 (67%)	-	187 (44%)	43 (12%)
Shared with Litter	247 (92%)	187 (93%)	-	145 (40%)
Total Shared	263 (98%)	200 (99%)	357 (85%)	152 (42%)

Table 3.6 Total, unique, and shared OTUs by substrate.
Comparison of the number of OTUs in each of four substrates. Shown are the total number of OTUs, the number of unique OTUs in each substrate and the number of OTUs shared between substrates. Total number of shared OTUs is the number of unique OTUs subtracted from the total number. OTUs were filtered to exclude those with a substrate-weighted abundance of less than 0.01%.

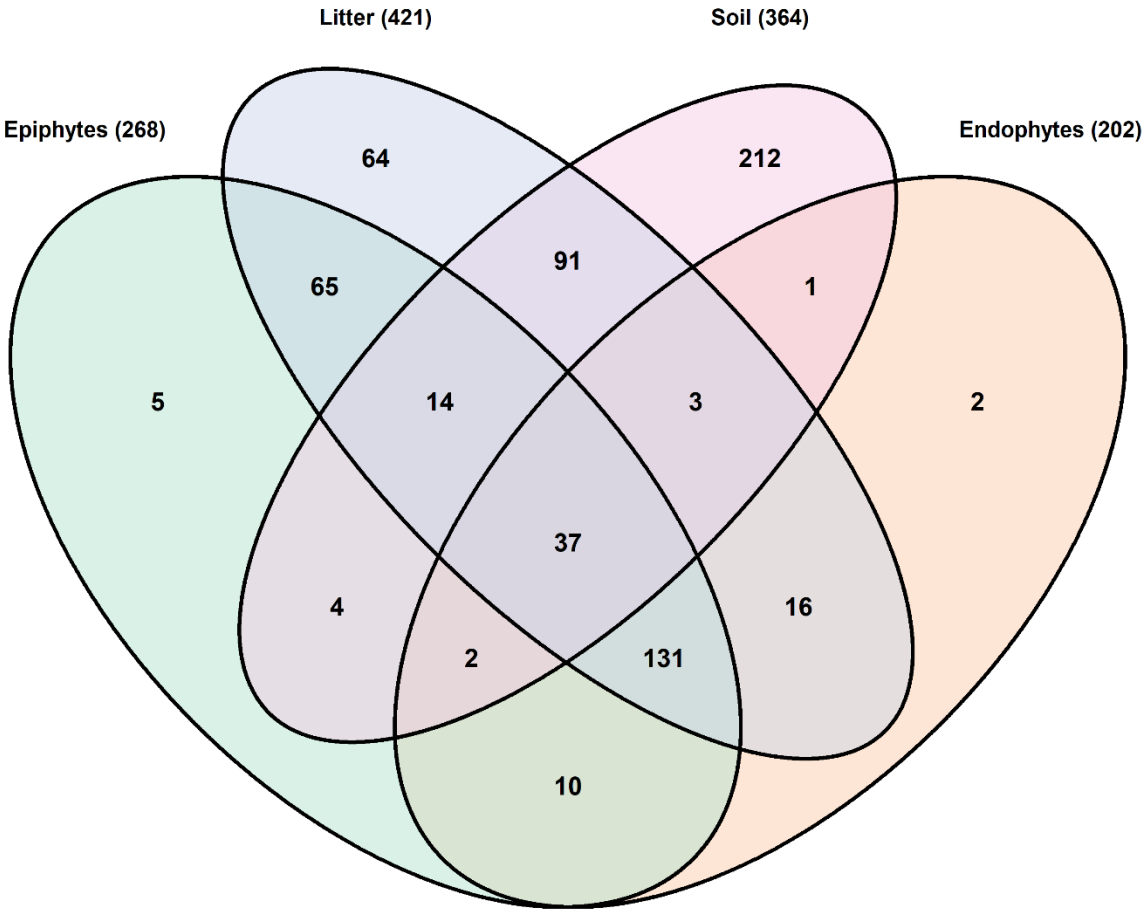

Figure 3.6 Venn diagram of shared OTUs by substrate.
657 OTUs were included which had substrate-weighted abundance greater than 0.01%. Overlapping regions indicate OTUs with at least one read in each of the substrates.

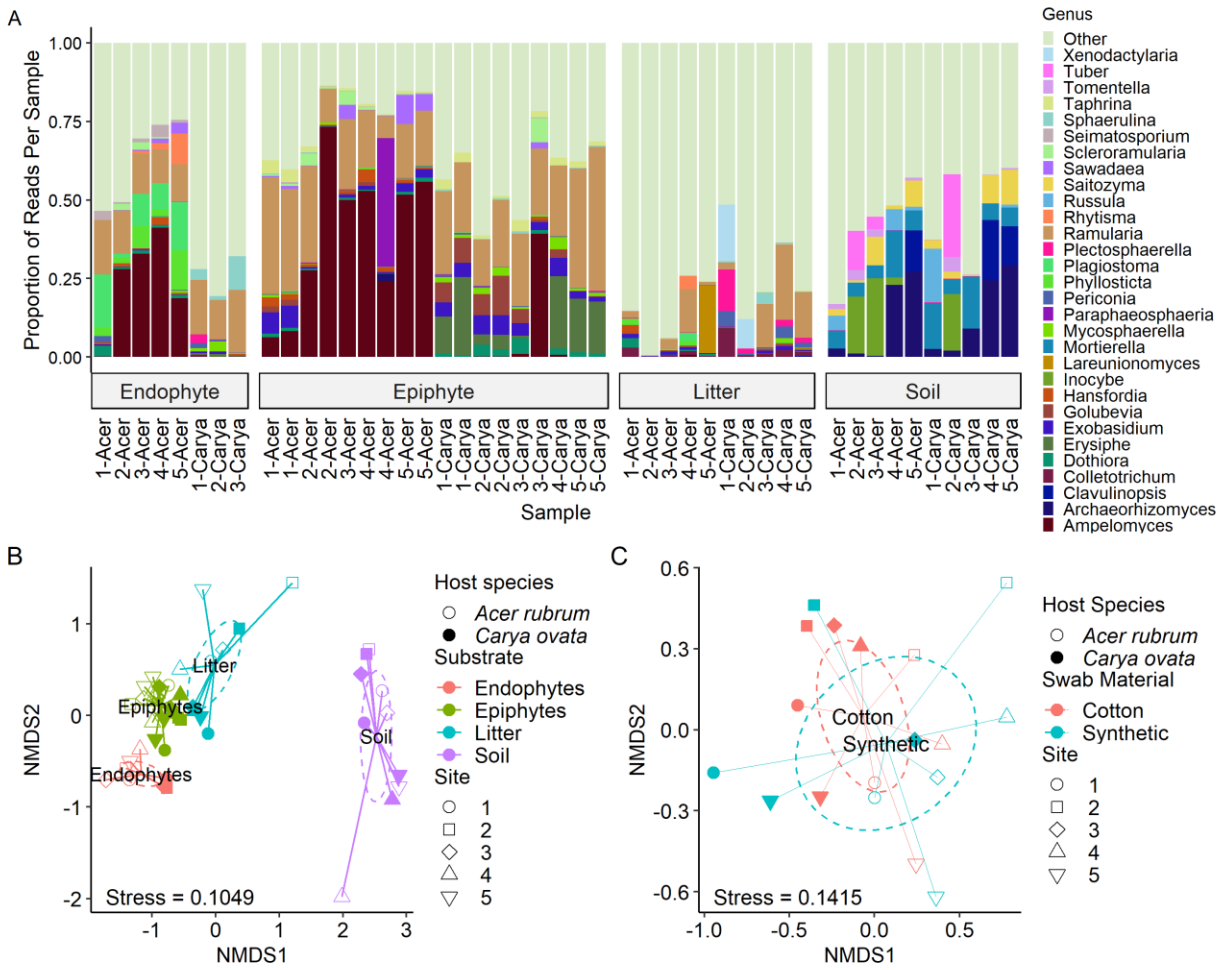

Figure 3.7 Comparison of community composition by substrate, host species, site, and swab material.
A) The most abundant 30 genera, determined by substrate-weighted abundance and CONSTAX2 classification, are displayed by the proportion of the community composed of OTUs classified within each genus. "Other" includes all OTUs not in the top 30, regardless of abundance. Samples are labeled by host species (genus name) and site with each substrate. B) Communities were ordinated using non-metric multidimensional scaling (NMDS) and Bray-Curtis distance, with ellipses representing 97% confidence interval estimates of centroids. C) Epiphyte communities were re-ordinated separately from the remaining samples. Ellipses show 97% confidence interval estimates of centroids of each swab material.

Within each substrate, effects of host species were notable. Endophyte communities of hickory were enriched for *Sphaerulina* spp. (5.13% vs 0.030% in maple) while depleted for *Phyllosticta* spp. (0.011% vs 5.30%) and *Seimatosporium* spp. (0.045% vs 1.88%). Epiphyte communities of hickory were uniquely characterized by *Erysiphe* (11.4% vs 0.019%) and

Golubevia spp. (4.64% vs 0.70%), with a much lower abundance of *Ampelomyces* spp. (4.54% vs 38.9%). OTUs within these genera were significant indicator taxa for their substrate and host species. In leaf litter, some genera were much more abundant in maple litter, including *Lareunionomyces* (4.31% vs. 0.010%), *Dothiora* (0.84% vs 0.10%), *Ampelomyces* (0.14% vs 0.030%), and *Saitozyma* (0.047% vs 0.0034%), or in hickory such as *Periconia* (2.14% vs 0.051%), *Plectosphaerella* (3.73% vs 0.10%), and *Mycosphaerella* (0.40% vs 0.030%). *Ramularia* (3.99% in maple vs 10.97% in hickory), *Taphrina* (0.061% vs 0.14%) and *Scleroramularia* (0.020% vs 0.0034%) were between 2- and 6-fold different between host species.

Despite differences in communities based on host species, some core taxa were observed at similar abundance between host species within a substrate. Endophyte communities had taxa such as *Ramularia* spp. (13.3% in *A. rubrum* vs 16.6% in *C. ovata*), while in epiphyte communities *Ramularia* spp, (21.5% vs 26.2%), *Exobasidium* spp. (2.47% vs 4.11%), *Taphrina* spp. (1.51% vs 2.11%), and *Scleroramularia* spp. (1.30% vs 0.93%) had similar relative abundances between hosts. Similarly abundant genera in leaf litter were *Exobasidium* (0.73% in *Acer rubrum* vs 0.50% in *Carya ovata*), *Seimatosporium* (0.091% vs 0.084%), *Inocybe* (0.017% in both), *Mortierella* (0.010% in both), and *Russula* (0.0033% in both). Ordination with non-metric multidimensional scaling (NMDS) and Bray-Curtis dissimilarity show a broad pattern of fungal communities segregating by substrate, host species, and site (Figure 3.7B).

Comparison of swab materials

In this study we compared cotton-tipped and synthetic polymer-tipped swabs for collecting epiphytic samples. Total read counts were modeled using a negative binomial mixed model, with a fixed effect of swab material and a random intercept effect of leaf sampled. Read

counts per sample were not significantly different between paired cotton and synthetic swab samples (Figure 3.8) (p = 0.736, z = -0.338, incidence rate ratio for synthetic : cotton = 0.886, 95% CI 0.432 – 1.799). PERMANOVA was used to compare similarity using rarefied OTU counts and Bray-Curtis distance, and showed no significant difference (p = 0.7) between swab types. Dispersion was also not significantly different (p = 0.36). NMDS ordination of swab (epiphyte) samples demonstrated this overlap of centroids and dispersion between material types (Figure 3.7C). Comparison of within-sample diversity showed no significant difference between mean diversity for cotton and synthetic samples (Figure 3.9), with observed richness (p = 0.31, F = 1.17, 95% CI difference = -30.76 – 88.83), Shannon diversity (p = 0.61, F = 0.2865, 95% CI difference = -0.51 – 0.84), or Inverse Simpson (p = 0.68, F = 0.18, CI difference = -6.26 – 9.49). However, samples collected with the synthetic swab material had somewhat higher diversity for each metric examined.

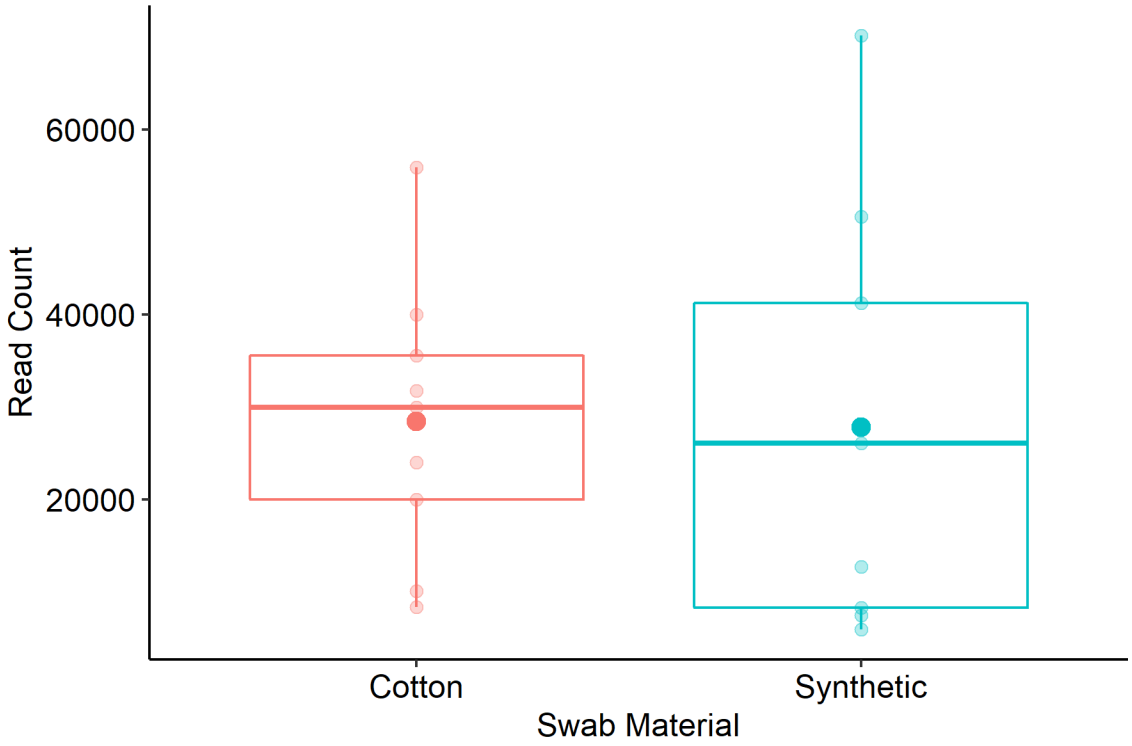

Figure 3.8 Read counts by swab material.
Read counts prior to rarefaction were compared between swab material types. Means are displayed with a large solid circle.

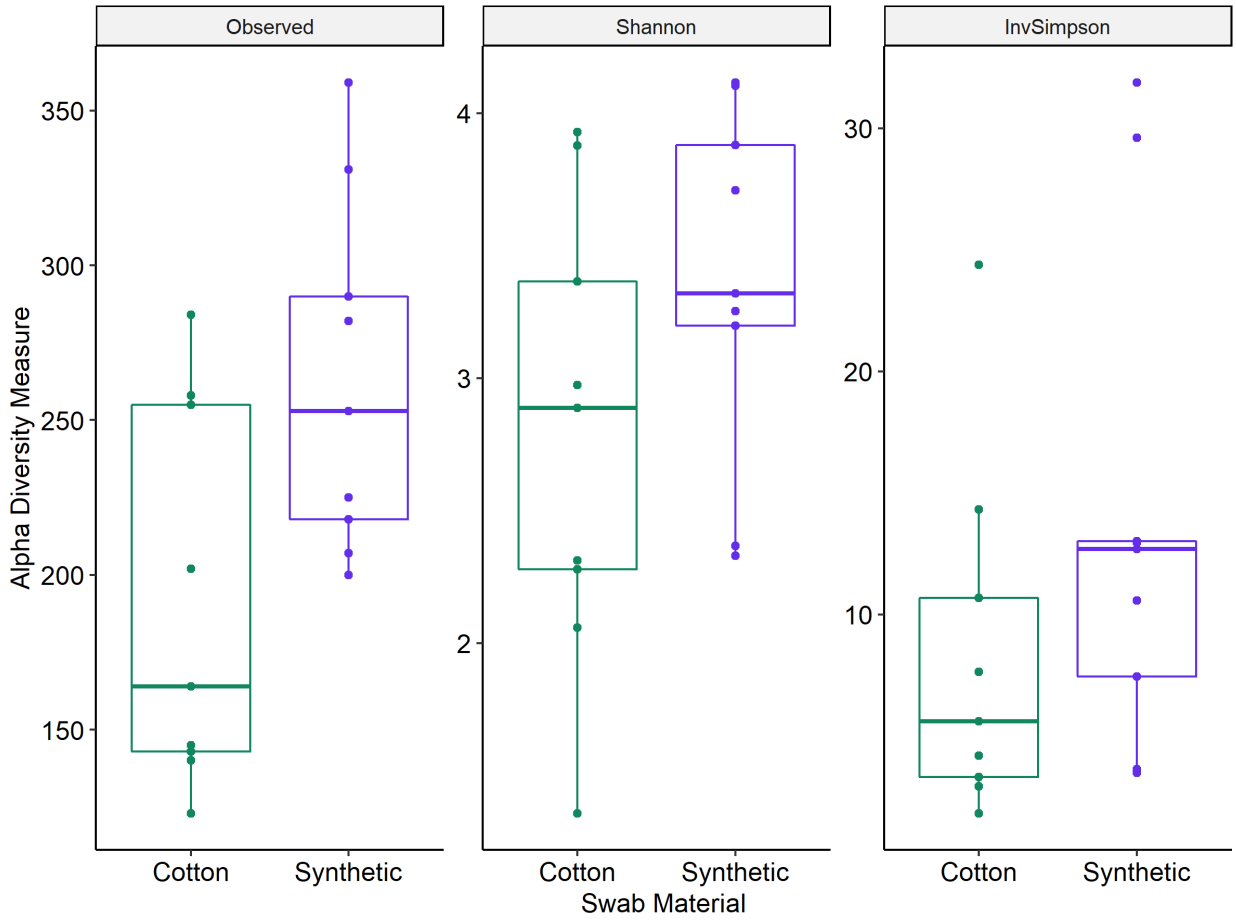

Figure 3.9 Richness and within-sample diversity of fungal communities by swab material. Epiphyte communities sampled with swabs were compared by observed richness, Shannon, and Inverse Simpson estimates of within-sample (alpha) diversity. Diversity estimates were determined using the phyloseq package.

Discussion

In the current study, we set out to determine how host species, substrate, and site structured fungal communities in a temperate forest ecosystem. We investigated this question with high-throughput ITS amplicon sequencing of fungal communities from pre-senescent leaves, litter, and soil, across five sites and between two host species, *A. rubrum* and *C. ovata*. The plant-associated fungal communities we observed show an effect of each of the factors characterized. We observed similar dispersion between endophytes and epiphytes, as well as soil and litter, while the litter communities had more similar composition to pre-senescent leaf

communities than soil communities. Within-sample diversity varied by substrate and host species, but not site. As reported in other studies, soil fungal communities were more diverse than litter and phyllosphere communities (Kong *et al.*, 2020). These data indicate discrete mechanisms of community assembly and maintenance that differ between substrates, and lead to unique spatially explicit fungal communities based on substrate type. Host species, substrate and site contributed most to community structure in our study, have also been deemed significant in other studies (Abdelfattah *et al.*, 2019; Küngas *et al.*, 2020; Christian *et al.*, 2020).

Comparing the phyllospheres of maple and hickory

In this study, we compared shagbark hickory (*Carya ovata*) and red maple (*Acer rubrum*). The phyllosphere microbial communities of *A. rubrum* have previously been examined, but only with culture-based fungal surveys (U'Ren *et al.*, 2012) and amplicon-based bacterial surveys (Laforest-Lapointe *et al.*, 2016a). U'Ren et al. (2012) isolated endophytic fungi, which were primarily classified as *Pestalotiopsis*, *Phyllosticta*, *Colletotrichum*, *Plagiostoma*, and *Ramularia* spp. These genera accounted for about 29% of reads in our *A. rubrum* endophyte samples. *Seimatosporium* spp. composed an additional 1.88% of these samples. *Ampelomyces* spp. contributed substantially (38.9%) to *A. rubrum* epiphyte communities. While we find consistency across these studies, our culture-independent method allowed for greater sampling depth and likely sequenced taxa that are difficult to culture. Laforest-Lapointe et al. (2016a) showed an effect of intra- and inter-individual variability on leaf endophyte communities, which aligns with the changes we see between sampled hosts.

We are unaware of any phyllosphere leaf fungal community studies in *C. ovata*. Despite growing in the same habitat, our analyses show significant differences in phyllosphere (epiphytic and

endophytic) fungal communities between host species, with *C. ovata* phyllosphere communities showing greatly increased *Sphaerulina*, *Erysiphe,* and *Golubevia* spp. compared to *A. rubrum*.

Fungal communities in maple and hickory leaf litter

As leaves senesce and become litter, priority effects and leaf traits are expected to contribute to the communities observed (Veen *et al.*, 2019; Bhatnagar *et al.*, 2018). We observed that host species was a significant determinant of phyllosphere communities, and expected that combined priority effects and leaf trait effects would carry forward community differences between the hosts. While most of the taxa detected in leaf litter (421 OTUs, Figure 3.9) were also present in the phyllosphere (63%, 266 OTUs) or soil (34%, 145 OTUs), around 15% of the fungal community in the litter was unique to that niche (30% for maple communities, 11% for hickory, of which 17% of the litter unique OTUs were shared between species). Despite a small number of shared OTUs between host species, only a single indicator taxon, annotated as Hypocreales sp. (OTU_50), found to be significantly associated with hickory litter. Clustering by NMDS further supports differences between host species but not as strongly as for phyllosphere taxa. Out of the core litter taxa shared between host species, the ecologies of *Inocybe*, Mortierellaceae, and *Russula* as soil-inhabiting saprobic and ectomycorrhizal fungi (Seress *et al.*, 2016; Vandepol *et al.*, 2020; Geml *et al.*, 2010) and their low abundance suggests that these taxa may have been detected as the result of soil clinging to sampled litter. However, for other litter core taxa, abundance patterns existing in phyllosphere pre-senescence appear to continue in litter for *Exobasidium* spp., or reflect a loss of host specificity in the case of *Seimatosporium* spp.

Fungi detected in the leaf litter are not necessarily derived from either soil or phyllosphere sources. Because the litter we sampled was most likely one year old, it may be possible that OTUs are persistent in the litter, or they may turnover or propagate throughout the

season or from year to year. Some fungal taxa are known to persist in temperate forest litter for periods greater than 1 year (Purahong *et al.*, 2016). These taxa, such as specialized basidiomycete litter degraders, may be directly colonizing new litter after it settles on the forest floor.

How do soil and litter fungal communities differ under maple and hickory?

Soils are known to be hyper-diverse environments (Hu *et al.*, 2019). In this study the soil fungal community had the highest number of unique OTUs (364) and lowest proportion of shared OTUs (42%) of any sample type. There were 145 OTUs shared between the litter and the soil, 54 of which were also shared with pre-senescent leaves (19/78 for maple, 37/87 for hickory). While these OTUs may have originated in pre-senescent leaves then dispersed to the litter then soil, dispersal could have alternately occurred whereby soil particles carrying fungi reached leaves and litter by wind or rain splash. Soil communities may be reservoirs of phyllosphere diversity (Zarraonaindia *et al.*, 2015), but we cannot assess the direction of dispersal in this experiment. Only 22% (91 of 421) of the OTUs found in the soil were exclusively shared between soil and litter, which indicates a limited contribution of the soil community to the litter community. Overall, community composition of the soil differed significantly from the other environments (Figure 3.7A).

No differences found between swab types on epiphytic fungal diversity indices

A second research question in this study was whether the swab material used to sample epiphytic communities affected the measured diversity. We found no significant differences between synthetic and cotton swabs in any diversity measurement, including richness, similarity, and read counts. The cost per unit of synthetic swabs are approximately 10 times higher than the cost per unit of cotton swabs, which motivated this research question. These data indicate that

future studies can use sterile cotton swabs for sampling without any loss in sampling effectiveness, a more cost-effective strategy.

Limitations and future directions

Our sampling was limited by sampling at a single time point and comparison of only two host plants. Time-series data would allow for measurements of community assembly, stability, and turnover. Vořišková and Baldrian (2013) proposed that some of the early fungal diversity in leaf litter may come from the living leaves. It is possible these taxa were missed by the time we collected samples (late summer). Conducting the experiment earlier in the season, along a temporal transect with more time-discrete sampling, or controlling for this effect by using spore traps (Abdelfattah *et al.*, 2019) would address this issue, and provide a clearer view of the successional changes in fungal communities during leaf senescence and decay. Variability in plant microbiomes exists across scales, from intra-individual (Laforest-Lapointe *et al.*, 2016a; Osono and Mori, 2004), to vertically within a canopy (Izuno *et al.*, 2016), geographically within a forest (Cordier *et al.*, 2012), between species (Laforest-Lapointe *et al.*, 2016b), and temporally throughout the year (Materatski *et al.*, 2019). We cannot differentiate causes of inter-individual variation in this study as due to within-species genotype variability or site environment variability, because only a single tree was sampled for each species at each site. At the individual/vertical scale, we only sampled leaves within 3 m of the ground. Since smaller trees have lower branches, our sampling was biased towards younger trees that were more accessible, and this may have biased the observed phyllosphere communities. Yet, our results are significant, and we were able to explain much of the observed patterns of fungal diversity.

rDNA amplicon-based surveys of fungal communities include biases, such as detecting only DNA, which fails to discriminate between active and dormant community members, which

a transcriptomic approach could (Morrison *et al.*, 2019). Such an approach could identify if the taxa identified in each substrate are metabolically active, and better demonstrate which shared taxa are active in more than one substrate. In using rarefied read counts to assess community dissimilarity (Bray-Curtis) and diversity (Shannon and Simpson indexes), it is possible that inter-organism rDNA copy number variation, primer bias, non-linear amplification, or other biases could alter our observations of community structure (Nilsson *et al.*, 2019). However, Bray-Curtis dissimilarity best accounted for variability between samples, and presence-absence and intra-taxa-based analyses (OTU overlap, relative taxa proportions) generally offered similar conclusions.

Our study dissected factors that affect fungal community structure in forest ecosystems, and could be expanded to address several unanswered questions. Conducting the experiment earlier in the season, along a temporal transect with more time-discrete sampling, or controlling for aerial dispersal with spore traps could further assess sources and sinks of fungal propagules, and provide a clearer view of the successional changes in fungal communities during leaf senescence and decay. At a larger scale, broader geographic sampling would likely show some interesting differences of fungal communities, even on trees of the same species, assuming large enough differences in climate, phenology, host genotype, or proximity to other individuals of the same or different species. A more extensive sampling approach incorporating metatranscriptomics, amplification-free metagenomics, and/or genome-based accounting of rDNA copy number could reduce the potential biases inherent to our methods, but also provide increased detail of the functional roles of the organisms in the community, distinguishing dormant or commensal organisms from active saprobes, pathogens, hyperparasites, or mutualists. Finally, to relate fungal communities to ecosystem functions and biochemical processes, surveys

of community composition should be combined with host plant phenotypes and measurements of nutrients in litter and soil, leading to insights about host fitness and biogeochemical cycling.

Conclusion

Our sampling of leaf, leaf litter, and soil microbiomes shows that the factors of substrate, host species, site, and their interactions account for a substantial amount of variation in the fungal community. These factors are indicative of specific influences affecting the community, such as substrate chemistry and local environments. Fungal phyllosphere communities differ by host species, an effect which largely persists in the litter community following leaf senescence. We found that pre-senescent leaf communities overlap substantially with litter community composition, and less so with the soil beneath. The assembly and function of litter communities is complex, thus further studies are needed to address spatial and temporal variability, activity of community members, and their effects on ecosystem processes.

CONCLUSION

Plant microbiomes are complex, in both the diversity of their components (microorganisms) as well as the processes which dictate their assembly. In Chapter 1, I relate the traits of the microbes in the communities of plant roots to their interactions with each other and with their host plant, demonstrating that competitive interactions can occur within plant roots and that changes to the host from viral disease can result in alterations to the microbial community. While the interactions that were observed cannot necessarily be generalized to all microbes nor all host plants, these interactions suggest that lower diversity of endophyte communities may result from resource limitation and that physiological changes of host plants may cause feedbacks with a plant's microbiome. Chapter 2 describes the CONSTAX2 classifier, a tool designed to aid in the understanding of metabarcoding studies by accurately assigning taxonomy to detected DNA sequences. The classifier performs as well or better than any currently competing taxonomic assignment tool, and has improved usability, functionality, and computational requirements compared to its predecessor. This type of tool allows for understanding DNA-based surveys in the context of ecological literature, by which the assignment of an informative name connects an observed taxon to knowledge about its occurrence in other studies, its physiology, and its interactions with other organisms. Lastly, Chapter 3 describes the factors affecting fungal community composition in maple and hickory leaves, leaf litter, and soil, assesses the similarity of these communities across substrates, demonstrates the utility of taxonomy assignment, and provides evidence for the effectiveness of epiphyte sampling techniques. In advancing our knowledge of what matters to plant microbiome assembly, I hope that this work contributes to the sustainability and livability of this planet for all its people.

REFERENCES

REFERENCES

Abarenkov,K. *et al.* (2020) UNITE general FASTA release for Fungi.

Abdelfattah,A. *et al.* (2019) Revealing Cues for Fungal Interplay in the Plant–Air Interface in Vineyards. *Front. Plant Sci.*, **10**, 922.

Ait Barka,E. *et al.* (2006) Enhancement of chilling resistance of inoculated grapevine plantlets with a plant growth-promoting rhizobacterium, *Burkholderia phytofirmans* strain PsJN. *Appl. Environ. Microbiol.*, **72**, 7246–7252.

Allard,G. *et al.* (2015) SPINGO: a rapid species-classifier for microbial amplicon sequences. *BMC Bioinformatics*, **16**, 324.

Altschul,S.F. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

Altschul,S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Anderson,D.R. and Burnham,K.P. (2002) Avoiding pitfalls when using information-theoretic methods. *J. Wildl. Manage.*, 912–918.

Andrews,S. (2010) FastQC: a quality control tool for high throughput sequence data.

Angebault,C. *et al.* (2020) Combined bacterial and fungal targeted amplicon sequencing of respiratory samples: Does the DNA extraction method matter? *PLoS One*, **15**, e0232215.

Apprill,A. *et al.* (2015) Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.*, **75**, 129–137.

Arnold,A.E. *et al.* (2000) Are tropical fungal endophytes hyperdiverse? *Ecol. Lett.*, **3**, 267–274.

Bashan,Y. *et al.* (1993) Isolation and Characterization of Plant Growth-Promoting Rhizobacteria. *Methods Plant Mol. Biol. Biotechnol.*, 331–345.

Bates,D. *et al.* (2015) Fitting Linear Mixed-Effects Models Using {lme4}. *J. Stat. Softw.*, **67**, 1–48.

Di Bella,J.M. *et al.* (2013) High throughput sequencing methods and analysis for microbiome research. *J. Microbiol. Methods*, **95**, 401–414.

Bengtsson-Palme,J. *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 Ecol. Evol.*, **4**, 914–919.

Benucci,G.M.N. *et al.* (2020) Patient propagules: Do soil archives preserve the legacy of fungal and prokaryotic communities? *PLoS One*, **15**, e0237368.

Bhatnagar,J.M. *et al.* (2018) Litter chemistry influences decomposition through activity of specific microbial functional guilds. *Ecol. Monogr.*, **88**, 429–444.

Bokulich,N.A. *et al.* (2018) Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*, **6**, 90.

Bolker,B. and R Development Core Team (2020) bbmle: Tools for General Maximum Likelihood Estimation.

Bonito,G. *et al.* (2017) *Atractiella rhizophila* , sp. nov., an endorrhizal fungus isolated from the *Populus* root microbiome. *Mycologia*, **109**, 18–26.

Bonito,G. *et al.* (2019) Fungal-bacterial networks in the *Populus* rhizobiome are impacted by soil properties and host genotype. *Front. Microbiol.*, **10**, 1–21.

Brader,G. *et al.* (2017) Ecology and Genomic Insights into Plant-Pathogenic and Plant-Nonpathogenic Endophytes. *Annu. Rev. Phytopathol.*, **55**, 61–83.

Bray,J.R. and Curtis,J.T. (1957) An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol. Monogr.*, **27**, 325–349.

Bray,N. and Wickings,K. (2019) The roles of invertebrates in the urban soil microbiome. *Front. Ecol. Evol.*, **7**.

Brooks,J.P. *et al.* (2015) The truth about metagenomics: Quantifying and counteracting bias in 16S rRNA studies Ecological and evolutionary microbiology. *BMC Microbiol.*, **15**, 66.

Brown,L.K. *et al.* (2012) What are the implications of variation in root hair length on tolerance to phosphorus deficiency in combination with water stress in barley (*Hordeum vulgare*)? *Ann. Bot.*, **110**, 319–328.

Brown,S.P. *et al.* (2015) Scraping the bottom of the barrel: are rare high throughput sequences artifacts? *Fungal Ecol.*, **13**, 221–225.

Bulgarelli,D. *et al.* (2013) Structure and Functions of the Bacterial Microbiota of Plants. *Annu. Rev. Plant Biol.*, **64**, 807–838.

De Cáceres,M. and Legendre,P. (2009) Associations between species and groups of sites: Indices and statistical inference. *Ecology*, **90**, 3566–3574.

Caporaso,J.G. *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods*, **7**, 335–336.

Carrión,V.J. *et al.* (2019) Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. *Science (80-.).*, **366**, 606–612.

Chen,H. (2018) VennDiagram: Generate High-Resolution Venn and Euler Plots.

Chen,K.H. *et al.* (2018) RNA-based analyses reveal fungal communities structured by a

senescence gradient in the moss *Dicranum scoparium* and the presence of putative multi-trophic fungi. *New Phytol.*, **218**, 1597–1611.

Chen,T. *et al.* (2020) A plant genetic network for preventing dysbiosis in the phyllosphere. *Nature*, **580**, 653–657.

Choudhury,S. *et al.* (2017) Barley yellow dwarf viruses: infection mechanisms and breeding strategies. *Euphytica*, **213**, 1–22.

Christian,N. *et al.* (2020) Host affinity of endophytic fungi and the potential for reciprocal interactions involving host secondary chemistry. *Am. J. Bot.*, **107**, 219–228.

Christiansen,L. *et al.* (2020) Fungal-Associated Molecules Induce Key Genes Involved in the Biosynthesis of the Antifungal Secondary Metabolites Nunamycin and Nunapeptin in the Biocontrol Strain *Pseudomonas fluorescens* In5. *Appl. Environ. Microbiol.*, **86**.

Cope,K.R. *et al.* (2019) The Ectomycorrhizal Fungus *Laccaria bicolor* Produces Lipochitooligosaccharides and Uses the Common Symbiosis Pathway to Colonize *Populus* Roots. *Plant Cell*, **31**, 2386–2410.

Cordier,T. *et al.* (2012) Spatial variability of phyllosphere fungal assemblages: Genetic distance predominates over geographic distance in a European beech stand (*Fagus sylvatica*). *Fungal Ecol.*, **5**, 509–520.

Demin,G. (2020) maditr: Fast Data Aggregation, Modification, and Filtering with Pipes and 'data.table'.

Deveau,A. *et al.* (2018) Bacterial-fungal interactions: Ecology, mechanisms and challenges. *FEMS Microbiol. Rev.*, **42**, 335–352.

Dice,L.R. (1945) Measures of the Amount of Ecologic Association Between Species. *Ecology*, **26**, 297–302.

diCenzo,G.C. *et al.* (2020) Genome-scale metabolic reconstruction of the symbiosis between a leguminous plant and a nitrogen-fixing bacterium. *Nat. Commun.*, **11**, 2574.

Draper,J. *et al.* (2001) *Brachypodium distachyon* . A New Model System for Functional Genomics in Grasses. *Plant Physiol.*, **127**, 1539–1555.

Edgar,R. (2016) SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences. *bioRxiv*, 074161.

Edgar,R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.

Edgar,R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods*, **10**, 996–998.

Elferink,S. *et al.* (2020) Comparative Metabarcoding and Metatranscriptomic Analysis of Microeukaryotes Within Coastal Surface Waters of West Greenland and Northwest Iceland. *Front. Mar. Sci.*, **7**.

Ellis,J.G. (2017) Can plant microbiome studies lead to effective biocontrol of plant diseases? *Mol. Plant-Microbe Interact.*, **30**, 190–193.

Engelmoer,D.J.P. *et al.* (2014) Intense competition between arbuscular mycorrhizal mutualists in an in vitro root microbiome negatively affects total fungal abundance. *Mol. Ecol.*, **23**, 1584–1593.

Falk,S.P. *et al.* (1995) Parasitism of *Uncinula necator* cleistothecia by the mycoparasite *Ampelomyces quisqualis*. *Phytopathology*, **85**, 794–800.

Finkel,O.M. *et al.* (2020) A single bacterial genus maintains root growth in a complex microbiome. *Nature*.

Floc'h,J.-B. *et al.* (2020) Fungal Communities of the Canola Rhizosphere: Keystone Species and Substantial Between-Year Variation of the Rhizosphere Microbiome. *Microb. Ecol.*, **80**, 762–777.

Fukami,T. *et al.* (2010) Assembly history dictates ecosystem functioning: Evidence from wood decomposer communities. *Ecol. Lett.*, **13**, 675–684.

Gagné-Bourque,F. *et al.* (2015) Accelerated growth rate and increased drought stress resilience of the model grass *Brachypodium distachyon* colonized by *Bacillus subtilis* B26. *PLoS One*, **10**, 1–23.

Gardes,M. and Bruns,T.D. (1993) ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizae and rusts. *Mol. Ecol.*, **2**, 113–118.

Gdanetz,K. *et al.* (2017) CONSTAX: A tool for improved taxonomic resolution of environmental fungal ITS sequences. *BMC Bioinformatics*, **18**, 1–9.

Geml,J. *et al.* (2010) Phylogenetic and ecological analyses of soil and sporocarp DNA sequences reveal high diversity and strong habitat partitioning in the boreal ectomycorrhizal genus *Russula* (Russulales; Basidiomycota). *New Phytol.*, 14.

Gill,U.S. *et al.* (2015) Characterization of *Brachypodium distachyon* as a nonhost model against switchgrass rust pathogen *Puccinia emaculata*. *BMC Plant Biol.*, **15**, 113.

Glöckner,F.O. *et al.* (2017) 25 years of serving the community with ribosomal RNA gene reference databases and tools. *J. Biotechnol.*, **261**, 169–176.

Gomes,T. *et al.* (2018) Endophytic and Epiphytic Phyllosphere Fungal Communities Are Shaped by Different Environmental Factors in a Mediterranean Ecosystem. *Microb. Ecol.*, **76**, 668–679.

Gopal,M. and Gupta,A. (2016) Microbiome selection could spur next-generation plant breeding strategies. *Front. Microbiol.*, **7**, 1–10.

Gravel,V. *et al.* (2007) Growth stimulation and fruit yield improvement of greenhouse tomato plants by inoculation with *Pseudomonas putida* or *Trichoderma atroviride*: Possible role of indole acetic acid (IAA). *Soil Biol. Biochem.*, **39**, 1968–1977.

Gross,S. *et al.* (2018) Characterization of antagonistic yeasts for biocontrol applications on apples or in soil by quantitative analyses of synthetic yeast communities. *Yeast*, **35**, 559–566.

Grüning,B. *et al.* (2018) Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nat. Methods*, **15**, 475–476.

Grünwald,N.J. *et al.* (2016) Population Genomics of Fungal and Oomycete Pathogens. *Annu. Rev. Phytopathol.*, **54**, 323–346.

Gu,Z. *et al.* (2020) Nitrate Stabilizes the Rhizospheric Fungal Community to Suppress Fusarium Wilt Disease in Cucumber. *Mol. Plant-Microbe Interact.*, **33**, 590–599.

Gundel,P.E. *et al.* (2017) Direct and indirect effects of the fungal endophyte *Epichloë uncinatum* on litter decomposition of the host grass, *Schedonorus pratensis*. *Plant Ecol.*, **218**, 1107–1115.

Hallmaier-Wacker,L.K. *et al.* (2018) The impact of storage buffer, DNA extraction method, and polymerase on microbial analysis. *Sci. Rep.*, **8**, 6292.

Harris,C.R. *et al.* (2020) Array programming with NumPy. *Nature*, **585**, 357–362.

He,X. *et al.* (2012) Diversity and decomposition potential of endophytes in leaves of a *Cinnamomum camphora* plantation in China. *Ecol. Res.*, **27**, 273–284.

He,X. *et al.* (2015) HvEXPB7, a novel β-expansin gene revealed by the root hair transcriptome of Tibetan wild barley, improves root hair growth under drought stress. *J. Exp. Bot.*, **66**, 7405–7419.

Henry,L. and Wickham,H. (2019) purrr: Functional Programming Tools.

Hervé,M. (2020) RVAideMemoire: Testing and Plotting Procedures for Biostatistics.

Herzog,C. *et al.* (2019) Microbial succession on decomposing root litter in a drought-prone Scots pine forest. *ISME J.*, **13**, 2346–2362.

Hoffman,T.K. and Kolb,F.L. (1997) Effects of barley yellow dwarf virus on root and shoot growth of winter wheat seedlings grown in aeroponic culture. *Plant Dis.*, **81**, 497–500.

Hoffmann,H. *et al.* (2005) Description of *Enterobacter ludwigii* sp. nov., a novel *Enterobacter* species of clinical relevance. *Syst. Appl. Microbiol.*, **28**, 206–212.

Horikoshi,M. and Tang,Y. (2018) ggfortify: Data Visualization Tools for Statistical Analysis Results.

Hothorn,T. *et al.* (2008) Simultaneous Inference in General Parametric Models. *Biometrical J.*, **50**, 346–363.

Howell,C.R. *et al.* (1993) Antibiotic Production by Strains of *Gliocladium virens* and its Relation to the Biocontrol of Cotton Seedling Diseases. *Biocontrol Sci. Technol.*, **3**, 435–441.

Hu,L. *et al.* (2018) Root exudate metabolites drive plant-soil feedbacks on growth and defense by shaping the rhizosphere microbiota. *Nat. Commun.*, **9**, 2738.

Hu,Y. *et al.* (2019) Contrasting latitudinal diversity and co-occurrence patterns of soil fungi and plants in forest ecosystems. *Soil Biol. Biochem.*, **131**, 100–110.

Izuno,A. *et al.* (2016) Vertical Structure of Phyllosphere Fungal Communities in a Tropical Forest in Thailand Uncovered by High-Throughput Sequencing. *PLoS One*, **11**, e0166669.

Jaccard,P. (1901) Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull Soc Vaudoise Sci Nat*, **37**, 547–579.

Jeffries,P. (1995) Biology and ecology of mycoparasitism. *Can. J. Bot.*, **73**, 1284–1290.

Jiang,F. *et al.* (2021) Arbuscular mycorrhizal fungi enhance mineralisation of organic phosphorus by carrying bacteria along their extraradical hyphae. *New Phytol.*, **230**, 304–315.

Jones,R.A.C. and Naidu,R.A. (2019) Global Dimensions of Plant Virus Diseases: Current Status and Future Perspectives. *Annu. Rev. Virol.*, **6**, 387–409.

Kamensky,M. *et al.* (2003) Soil-borne strain IC14 of *Serratia plymuthica* with multiple mechanisms of antifungal activity provides biocontrol of *Botrytis cinerea* and *Sclerotinia sclerotiorum* diseases. *Soil Biol. Biochem.*, **35**, 323–331.

Kappel,L. *et al.* (2020) Chitin and chitosan remodeling defines vegetative development and *Trichoderma* biocontrol.

Kassambara,A. (2019) ggpubr: 'ggplot2' Based Publication Ready Plots.

Keddy,P.A. (1992) Assembly and response rules: two goals for predictive community ecology. *J. Veg. Sci.*, **3**, 157–164.

Kong,H.G. *et al.* (2021) Achieving similar root microbiota composition in neighbouring plants through airborne signalling. *ISME J.*, **15**, 397–408.

Kong,X. *et al.* (2020) Maize (*Zea mays* L. Sp.) varieties significantly influence bacterial and fungal community in bulk soil, rhizosphere soil and phyllosphere. *FEMS Microbiol. Ecol.*, **96**.

Koskella,B. and Taylor,T.B. (2015) The Potential Role of Bacteriophages in Shaping Plant-Bacterial Interactions. *Bact. Interact. Adv. Res. Futur. Trends*, 199–220.

Küngas,K. *et al.* (2020) Host tree organ is the primary driver of endophytic fungal community structure in a hemiboreal forest. *FEMS Microbiol. Ecol.*, **96**, fiz199.

Kuo,H.C. *et al.* (2014) Secret lifestyles of *Neurospora crassa*. *Sci. Rep.*, **4**, 1–6.

Laforest-Lapointe,I. *et al.* (2016a) Host species identity, site and time drive temperate tree phyllosphere bacterial community structure. *Microbiome*, **4**, 1–10.

Laforest-Lapointe,I. *et al.* (2016b) Tree phyllosphere bacterial communities: Exploring the magnitude of intra- and inter-individual variation among host species. *PeerJ*, **2016**, e2367.

de Lamo,F.J. *et al.* (2018) Xylem Sap Proteomics Reveals Distinct Differences Between R Gene- and Endophyte-Mediated Resistance Against Fusarium Wilt Disease in Tomato. *Front. Microbiol.*, **9**, 1–13.

Lenth,R. V (2020) emmeans: Estimated Marginal Means, aka Least-Squares Means.

Lewis,J.A. and Papavizas,G.C. (1991) Biocontrol of plant diseases: the approach for tomorrow. *Crop Prot.*, **10**, 95–105.

Li,M. *et al.* (2014) Proline isomerization of the immune receptor-interacting protein RIN4 by a cyclophilin inhibits effector-triggered immunity in Arabidopsis. *Cell Host Microbe*, **16**, 473–483.

Liber,J. *et al.* (2021) CONSTAX2: Improved taxonomic classification of environmental DNA markers. *bioRxiv*.

Linsmaier,E.M. and Skoog,F. (1965) Organic growth factor requirements of tobacco tissue cultures. *Physiol. Plant*, **18**.

Liu,Y. and Buchenauer,H. (2005) Effect of infections with barley yellow dwarf virus and *Fusarium* spp. on assimilation of $^{14}CO_2$ by flag leaves and translocation of photosynthates in wheat. *Zeitschrift fur Pflanzenkrankheiten und Pflanzenschutz*, **112**, 529–543.

Lundberg,D.S. *et al.* (2013) Practical innovations for high-throughput amplicon sequencing. *Nat. Methods*, **10**, 999–1002.

Ma,X. *et al.* (2013) A gain-of-function mutation in the ROC1 gene alters plant architecture in Arabidopsis. *New Phytol.*, **197**, 751–762.

Maciá-Vicente,J.G. *et al.* (2020) Root filtering, rather than host identity or age, determines the composition of root-associated fungi and oomycetes in three naturally co-occurring Brassicaceae. *Soil Biol. Biochem.*, **146**, 107806.

Mandadi,K.K. and Scholthof,K.-B.G. (2015) Genome-Wide Analysis of Alternative Splicing Landscapes Modulated during Plant-Virus Interactions in *Brachypodium distachyon*. *Plant Cell*, **27**, 71–85.

Márquez,L.M. *et al.* (2007) A Virus in a Fungus in a Plant: Three-Way Symbiosis Required for Thermal Tolerance. *Science (80-.).*, **315**, 513–515.

Martin,F.M. *et al.* (2017) Ancestral alliances: Plant mutualistic symbioses with fungi and bacteria. *Science (80-.).*, **356**.

Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10–12.

Materatski,P. *et al.* (2019) Spatial and temporal variation of fungal endophytic richness and diversity associated to the phyllosphere of olive cultivars. *Fungal Biol.*, **123**, 66–76.

McKinney,W. (2010) Data Structures for Statistical Computing in Python. In, Walt,S. van der and Millman,J. (eds)., pp. 56–61.

McKirdy,S.J. *et al.* (2002) Quantification of yield losses caused by Barley yellow dwarf virus in wheat and oats. *Plant Dis.*, **86**, 769–773.

McMurdie,P.J. and Holmes,S. (2013) Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One*, **8**.

McNamara,J. (2021) Creating Excel files with Python and XlsxWriter.

Menkis,A. *et al.* (2014) *Archaeorhizomyces borealis* sp. nov. and a sequence-based classification of related soil fungal species. *Fungal Biol.*, **118**, 943–955.

Miller,W.A. *et al.* (2002) Barley yellow dwarf virus: Luteoviridae or Tombusviridae? *Mol. Plant Pathol.*, **3**, 177–183.

Moissl-Eichinger,C. *et al.* (2018) Archaea Are Interactive Components of Complex Microbiomes. *Trends Microbiol.*, **26**, 70–85.

Morrison,E.W. *et al.* (2019) Warming alters fungal communities and litter chemistry with implications for soil carbon stocks. *Soil Biol. Biochem.*, **132**, 120–130.

Mouillot,D. and Leprêtre,A. (1999) A comparison of species diversity estimators. *Res. Popul. Ecol. (Kyoto).*, **41**, 203–215.

Mousa,W.K. *et al.* (2016) Root-hair endophyte stacking in finger millet creates a physicochemical barrier to trap the fungal pathogen *Fusarium graminearum*. *Nat. Microbiol.*, **1**, 16167.

Müller,K. and Wickham,H. (2019) tibble: Simple Data Frames.

Nassar,A.H. *et al.* (2005) Promotion of plant growth by an auxin-producing isolate of the yeast *Williopsis saturnus* endophytic in maize (*Zea mays* L.) roots. *Biol. Fertil. Soils*, **42**, 97–108.

New,P.B. and Kerr,A. (1972) Biological control of crown gall: Field measurements and glasshouse experiments. *J. Appl. Bacteriol.*, **35**, 279–287.

Nguyen,N.H. *et al.* (2016) FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.*, **20**, 241–248.

Nilsson,R Henrik, Anslan,S., *et al.* (2019) Mycobiome diversity: high-throughput sequencing and identification of fungi. *Nat. Rev. Microbiol.*, **17**, 95–109.

Nilsson,R Henrik, Larsson,K.H., *et al.* (2019) The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.*, **47**, D259–D264.

Niu,B. *et al.* (2020) Microbial Interactions Within Multiple-Strain Biological Control Agents Impact Soil-Borne Plant Disease. *Front. Microbiol.*, **11**, 1–16.

O'Donnell,K. *et al.* (1998) Multiple evolutionary origins of the fungus causing panama disease of banana: Concordant evidence from nuclear and mitochondrial gene genealogies. *Proc. Natl. Acad. Sci. U. S. A.*, **95**, 2044–2049.

O'Donnell,K. *et al.* (2007) Phylogenetic diversity and microsphere array-based genotyping of human pathogenic fusaria, including isolates from the multistate contact lens-associated U.S. keratitis outbreaks of 2005 and 2006. *J. Clin. Microbiol.*, **45**, 2235–2248.

Oksanen,J. *et al.* (2019) vegan: Community Ecology Package.

Osono,T. (2007) Ecology of ligninolytic fungi associated with leaf litter decomposition. *Ecol. Res.*, **22**, 955–974.

Osono,T. and Mori,A. (2004) Distribution of phyllosphere fungi within the canopy of giant dogwood. *Mycoscience*, **45**, 161–168.

Paleg,L.G. (1965) Physiological effects of gibberellins. *Annu. Rev. Plant Physiol.*, **16**, 291–322.

Palmer,J.M. *et al.* (2018) Non-biological synthetic spike-in controls and the AMPtk software pipeline improve mycobiome data. *PeerJ*, **2018**, e4925.

Parada,A.E. *et al.* (2016) Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.*, **18**, 1403–1414.

Parton,W. *et al.* (2007) Global-scale similarities in nitrogen release patterns during long-term decomposition. *Science (80-.).*, **315**, 361–364.

Peay,K.G. (2014) Back to the future: natural history and the way forward in modern fungal

ecology. *Fungal Ecol.*, **12**, 4–9.

Pedersen,T.L. (2019) patchwork: The Composer of Plots. *R Packag. version 1.0.0*.

Penner,S. and Sapir,Y. (2021) Foliar Endophytic Fungi Inhabiting an Annual Grass Along an Aridity Gradient. *Curr. Microbiol.*, **1**, 3.

Perry,K.L. *et al.* (2000) Yield effects of Barley yellow dwarf virus in soft red winter wheat. *Phytopathology*, **90**, 1043–1048.

Pfaffl,M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.*, **29**, 45e – 45.

Van Der Pol,W.J. *et al.* (2019) In Silico and Experimental Evaluation of Primer Sets for Species-Level Resolution of the Vaginal Microbiota Using 16S Ribosomal RNA Gene Sequencing. *J. Infect. Dis.*, **219**, 305–314.

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

Prasch,C.M. and Sonnewald,U. (2013) Simultaneous application of heat, drought, and virus to Arabidopsis plants reveals significant shifts in signaling networks. *Plant Physiol.*, **162**, 1849–1866.

Priya,H. *et al.* (2015) Influence of cyanobacterial inoculation on the culturable microbiome and growth of rice. *Microbiol. Res.*, **171**, 78–89.

Promputtha,I. *et al.* (2007) A phylogenetic evaluation of whether endophytes become saprotrophs at host senescence. *Microb. Ecol.*, **53**, 579–590.

Purahong,W. *et al.* (2016) Life in leaf litter: novel insights into community dynamics of bacteria and fungi during litter decomposition. *Mol. Ecol.*, **25**, 4059–4074.

R Core Team (2019) R: A Language and Environment for Statistical Computing.

Regalado,J. *et al.* (2020) Combining whole-genome shotgun sequencing and rRNA gene amplicon analyses to improve detection of microbe – microbe interaction networks in plant leaves. *ISME J.*, 2116–2130.

Riedell,W.E. *et al.* (2003) Root and shoot responses to bird cherry-oat aphids and Barley yellow dwarf virus in spring wheat. *Crop Sci.*, **43**, 1380–1386.

Ryu,M.H. *et al.* (2020) Control of nitrogen fixation in bacteria that associate with cereals. *Nat. Microbiol.*, **5**, 314–330.

Saikia,J. *et al.* (2018) Alleviation of drought stress in pulse crops with ACC deaminase producing rhizobacteria isolated from acidic soil of Northeast India. *Sci. Rep.*, **8**, 1–16.

Santoferrara,L. *et al.* (2020) Perspectives from Ten Years of Protist Studies by High-Throughput

Metabarcoding. *J. Eukaryot. Microbiol.*, **67**, 612–622.

Sapp,M. *et al.* (2018) Protists are an integral part of the *Arabidopsis thaliana* microbiome. *Environ. Microbiol.*, **20**, 30–43.

Schilling,J.S. *et al.* (2020) Using Wood Rot Phenotypes to Illuminate the "Gray" Among Decomposer Fungi. *Front. Microbiol.*, **11**, 1288.

Schloss,P.D. *et al.* (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl. Environ. Microbiol.*, **75**, 7537–7541.

Schneider,C.A. *et al.* (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*, **9**, 671–675.

Schoener,T.W. (1974) Resource partitioning in ecological communities. *Science (80-.).*, **185**, 27–39.

Schöps,R. *et al.* (2020) Resident and phytometer plants host comparable rhizosphere fungal communities in managed grassland ecosystems. *Sci. Rep.*, **10**, 1–11.

Seress,D. *et al.* (2016) Characterisation of seven *Inocybe* ectomycorrhizal morphotypes from a semiarid woody steppe. *Mycorrhiza*, **26**, 215–225.

Shannon,C.E. (1948) A Mathematical Theory of Communication. *Bell Syst. Tech. J.*, **27**, 379–423.

De Silva,N.I. *et al.* (2019) Use of endophytes as biocontrol agents. *Fungal Biol. Rev.*, **33**, 133–148.

Simpson,E.H. (1949) Measurement of diversity. *Nature*, **163**, 688.

Sørensen,T.A. (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biol. Skar.*, **5**, 1–34.

Sorokan,A. *et al.* (2020) Endophytic *Bacillus* spp. as a Prospective Biological Tool for Control of Viral Diseases and Non-vector *Leptinotarsa decemlineata* Say. in *Solanum tuberosum* L. *Front. Microbiol.*, **11**, 1–13.

Soudzilovskaia,N.A. *et al.* (2019) Global mycorrhizal plant distribution linked to terrestrial carbon stocks. *Nat. Commun.*, **10**, 1–10.

Stenberg,J.A. *et al.* (2015) Optimizing Crops for Biocontrol of Pests and Disease. *Trends Plant Sci.*, **20**, 698–712.

Stomp,M. *et al.* (2004) Adaptive divergence in pigment composition promotes phytoplankton biodiversity. *Nature*, **432**, 104–107.

Stringlis,I.A. *et al.* (2018) MYB72-dependent coumarin exudation shapes root microbiome assembly to promote plant health. *Proc. Natl. Acad. Sci. U. S. A.*, **115**, E5213–E5222.

Syed Ab Rahman,S.F. *et al.* (2018) Emerging microbial biocontrol strategies for plant pathogens. *Plant Sci.*, **267**, 102–111.

Tang,J. *et al.* (2020) GIMICA: host genetic and immune factors shaping human microbiota. *Nucleic Acids Res.*, **49**, 715–722.

Tang,Y. *et al.* (2016) ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages. *R J.*, **8**.

Tao,Y. *et al.* (2016) *Brachypodium distachyon* is a suitable host plant for study of Barley yellow dwarf virus. *Virus Genes*, **52**, 299–302.

Tedersoo,L. *et al.* (2010) Ectomycorrhizal lifestyle in fungi: Global diversity, distribution, and evolution of phylogenetic lineages. *Mycorrhiza*, **20**, 217–263.

Teixeira,P.J.P. *et al.* (2019) Beyond pathogens: microbiota interactions with the plant immune system. *Curr. Opin. Microbiol.*, **49**, 7–17.

The pandas development team (2020) pandas-dev/pandas: Pandas.

Thingstrup,I. *et al.* (2000) Phosphate transport by hyphae of field communities of arbuscular mycorrhizal fungi at two levels of P fertilization. *Plant Soil*, **221**, 181–187.

Tisserant,E. *et al.* (2013) Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proc. Natl. Acad. Sci.*, **110**, 20117–20122.

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

Ulbrich,T.C. *et al.* (2021) Intraspecific Variability in Root Traits and Edaphic Conditions Influence Soil Microbiomes Across 12 Switchgrass Cultivars. *Phytobiomes J.*, PBIOMES-12-19-0.

Vandenkoornhuyse,P. *et al.* (2015) The importance of the microbiome of the plant holobiont. *New Phytol.*, **206**, 1196–1206.

Vandepol,N. *et al.* (2020) Resolving the Mortierellaceae phylogeny through synthesis of multi-gene phylogenetics and phylogenomics. *Fungal Divers.*, **104**, 267–289.

Vandermeer,J. *et al.* (2009) Evidence for hyperparasitism of coffee rust (*Hemileia vastatrix*) by the entomogenous fungus, *Lecanicillium lecanii*, through a complex ecological web. *Plant Pathol.*, **58**, 636–641.

VanWallendael,A. *et al.* (2021) Host genetic control of succession in the switchgrass leaf fungal microbiome. *bioRxiv*.

Veen,G.F. (Ciska. *et al.* (2019) Relationships between fungal community composition in decomposing leaf litter and home-field advantage effects. *Funct. Ecol.*, **33**, 1524–1535.

Veneault-Fourrey,C. and Martin,F. (2011) Mutualistic interactions on a knife-edge between saprotrophy and pathogenesis. *Curr. Opin. Plant Biol.*, **14**, 444–450.

Videira,S.I.R. *et al.* (2016) All that glitters is not Ramularia. *Stud. Mycol.*

Vo,A.T.E. and Jedlicka,J.A. (2014) Protocols for metagenomic DNA extraction and Illumina amplicon library preparation for faecal and swab samples. *Mol. Ecol. Resour.*, **14**, 1183–1197.

Voříšková,J. and Baldrian,P. (2013) Fungal community on decomposing leaf litter undergoes rapid successional changes. *ISME J.*, **7**, 477–486.

De Vrieze,M. *et al.* (2018) Combining Different Potato-Associated *Pseudomonas* Strains for Improved Biocontrol of *Phytophthora infestans*. *Front. Microbiol.*, **9**, 1–13.

Vujanovic,V. *et al.* (2019) Transgenerational role of seed mycobiome – an endosymbiotic fungal composition as a prerequisite to stress resilience and adaptive phenotypes in *Triticum*. *Sci. Rep.*, **9**, 1–13.

Waller,F. *et al.* (2005) The endophytic fungus *Piriformospora indica* reprograms barley to salt-stress tolerance, disease resistance, and higher yield. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 13386–13391.

Wang,Q. *et al.* (2007) Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl. Environ. Microbiol.*, **73**, 5261–5267.

Waqas,M. *et al.* (2012) Endophytic fungi produce gibberellins and indoleacetic acid and promotes host-plant growth during stress. *Molecules*, **17**, 10754–10773.

Warnes,G.R. *et al.* (2020) gplots: Various R Programming Tools for Plotting Data.

Van Wees,S.C.M. *et al.* (1997) Differential induction of systemic resistance in Arabidopsis by biocontrol bacteria. *Mol. Plant-Microbe Interact.*, **10**, 716–724.

Wei,Z. *et al.* (2017) Seasonal variation in the biocontrol efficiency of bacterial wilt is driven by temperature-mediated changes in bacterial competitive interactions. *J. Appl. Ecol.*, **54**, 1440–1448.

White,T.J. *et al.* (1990) Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics. *PCR Protoc.*, **18**, 315–322.

Wickham,H., François,R., *et al.* (2019) dplyr: A Grammar of Data Manipulation.

Wickham,H. (2020) forcats: Tools for Working with Categorical Variables (Factors).

Wickham,H. (2016) ggplot2: Elegant Graphics for Data Analysis Springer-Verlag New York.

Wickham,H., Averick,M., *et al.* (2019) Welcome to the tidyverse. *J. Open Source Softw.*, **4**, 1686.

Wickham,H. and Henry,L. (2020) tidyr: Tidy Messy Data.

Wood,D.E. *et al.* (2019) Improved metagenomic analysis with Kraken 2. *Genome Biol.*, **20**, 257.

Wood,J.L. *et al.* (2018) Competitive traits are more important than stress-tolerance traits in a cadmium-contaminated rhizosphere: A role for trait theory in microbial ecology. *Front. Microbiol.*, **9**, 121.

Yao,H. *et al.* (2019) Phyllosphere epiphytic and endophytic fungal community and network structures differ in a tropical mangrove ecosystem. *Microbiome*, **7**, 57.

Zarraonaindia,I. *et al.* (2015) The soil microbiome influences grapevine-associated microbiota. *MBio*, **6**, e02527--14.

Zhang,L. *et al.* (2016) Carbon and phosphorus exchange may enable cooperation between an arbuscular mycorrhizal fungus and a phosphate-solubilizing bacterium. *New Phytol.*, **210**, 1022–1032.

Zhou,D. and Hyde,K.D. (2001) Host-specificity, host-exclusivity, and host-recurrence in saprobic fungi. *Mycol. Res.*, **105**, 1449–1457.