EXPLORING AND CAPITALIZING ON GENOTYPIC VARIATION OF THE OAT MICROBIOME

By

Zachary D. Zalewski

A thesis submitted in partial fulfillment of the

requirements for the degree of

Doctorate of Philosophy

(Plant Pathology)

At the

UNIVERSITY OF WISCONSIN-MADISON

2023

Acknowledgements

First and foremost, I want to thank my mom. She died during the completion of this project and left a cold, empty space in a great many hearts. The depth of our grief is a testimony to the beauty of having you and Sorority in our lives. I can easily imagine how over-the-top excited you would have been to call me “Dr. Zalewski.” Thank you for your unconditional support, it sustains me even in your absence. My father too, deserves endless gratitude. Your unwavering faith in me has always helped me see a better version of myself.

Of course, thank you to my future wife Dr. Shannon Goes. If you hadn’t been here to check in on me, I probably would have withered away and died in front of my computer while writing this thesis. It’s absurd how much better my life is with you in it.

Thank you to all the rest of my friends and family for their love and support. Thank you to my wonderful crazy siblings for never letting me get bored. Thank you to Kris and Bill (how lucky am I to have a second set of amazing parents?!). Thank you to Aunt Jeanne, Uncle Pat, Uncle Dan, Joel, Jodi and all of my cousins. Thank you to my lovely bunch of friends who made sure I made time for fun over these last several years. Seth, Kyle, Brandon, Bri, Neve, Ethan, Vicky, Helena, Shannon, and so many others. I am endlessly grateful to know each and every one of you.

Finally, thank you to the Lankau Lab and the Plant Pathology department in general. The plant pathology community at UW-Madison is nothing short of special. Thank you to my lab mates, especially Katie Thompson my grad school comrade through thick and thin. Thank you to my committee for their expertise, and a special thank you to Rick Lankau for your mentorship as both a scientist and an educator.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 4

Chapter 1

Introduction: Exploring and capitalizing on genotypic variation of the oat microbiome

Abstract

With both a changing climate and increasing demand for more sustainable agroecosystems (Altieri et al., 2017; Delonge et al., 2016), it is imperative that we continue to innovate and refine our agricultural systems. Crop rotations are an ancient strategy for improving crop health and productivity through several biotic and abiotic changes to the soil known as soil legacy effects (Jernigan et al., 2020). These soil legacy effects have been shown to vary not only by crop species, but also by the cultivars within a given species (Ellouze et al. 2013). These variations indicate a certain amount of host genetic control over microbial community assembly. This thesis focuses on the soil legacy effects of various oat (*Avena sativa*) genotypes and the benefits or detriments these legacy effects confer to subsequently planted oat and corn crops (*Zea mays*). This focus was explored through three main goals: 1) Determining how the genotype dependent changes in soil legacy affect corn traits 2) Using statistical modeling tools to use microbial community data to predict outcomes in oats and next-season corn while also identifying candidate mutualists/parasites 3) Associate genomic regions within elite oat genotypes that control the assembly of the rhizosphere community in multiple environmental contexts. The results of this project could inform future crop rotation programs with the goal of optimizing soil legacies for the yield of subsequent crops and provide information for the potential crop breeding programs focused on the selection of specific, desirable members of the microbial community.

Introduction

Growers have engaged in plant breeding efforts since the advent of agriculture, and centuries of breeding has resulted in relatively hardy, disease-tolerant crops optimized for growth in an agricultural setting. Modern breeding strategies have streamlined this process, selecting for desirable plant alleles with stark efficiency. Breeding regimens have resulted in an equally efficient loss of undesirable or neglected alleles, resulting in a severe reduction of genetic variation within some crop species (Zhang et al. 2015). While this phenomenon has wide ranging implications, one inadvertent effect is an apparent reduction in crop dependency on the surrounding microbial communities (Kiers et al., 2007; Kim et al., 2020, Porter and Sachs 2020) Despite this potential trend, crops still exhibit varying degrees of genetic variation in microbial community assembly, as evidenced by the variation in microbial community structure among crop cultivars (Qiao et al., 2017; Marques et al., 2014). This genotypic variation in host-driven, microbial community assembly indicates the potential for breeding programs to focus on host microbial selection.

With all of the known benefits the plant associated microbial community (i.e. plant microbiome) can confer to a host (Bender et al., 2016), it is no wonder that those with a vested interest in agriculture are increasingly interested in leveraging these communities to increase plant productivity in the field. In fact, microbiome manipulation is considered by some to be the key to the next green revolution (Jez et al., 2016). Whether this lofty goal is ever achieved is partially dependent on our ability as researchers to overcome significant hurdles in the study and manipulation of plant microbiomes.

It is important to note that the leveraging of soil communities to our benefit is not a new idea that requires new technologies. Agriculturalists have long understood that crop rotations can

improve maintenance of crop health and yield. Many of the benefits of crop rotation can be attributed to an increase in diversity and a decrease of host pathogens within the microbial community of a given field (Wagg et al., 2019; Paulitz et al., 2010). In addition, crop rotations have been shown to alter microbial community structure (Song et al., 2018). Given the short-term adaptability of many microbes and the overall sensitivity of microbial ecosystems in general, it is not unrealistic to say any management decision a grower makes in the field can have an impact on microbial populations within that field. While this statement seems reasonable, the difficulties associated with culturing and sequencing microbes at a community level have made it historically hard to challenge. It is only with the recent advent of next generation sequencing that agricultural researchers have been able to deeply explore variation within microbial communities.

While our ability to discover the taxonomic make-up of a given community has vastly increased, our understanding of how that community will affect a host crop lags far behind. Further behind is our ability to make nuanced and specific alterations to a microbial community to achieve predictable outcomes. Agriculturists are just beginning to wrestle with these challenges, but plants have evolved for millennia to overcome the recalcitrant nature of microbial communities and thus secrete a bevy of carbon compounds and secondary metabolites into the soil to manipulate community structure to their benefit (Moe 2013; Weston and Methesius 2013). Given the plant capacity for microbiome alteration there is potential for relying on plant hosts to make the microbial adjustments desired by growers. In addition, the variation in microbial community assembly observed between the genotypes of some crops suggests the potential to direct these host driven community adjustments through plant breeding (Gopal and Gupta 2016).

Admittedly, breeding for microbiome control and sensitivity may come with undesirable trade-offs. In fact, breeders might have unwittingly bred microbial sensitivity out of many elite crops. The high-nutrient context in which crop selections are frequently made is the very context in which plants have the least need for microbial help. That being said, there is increasing desire, in both consumer and grower populations, to move towards lower-input systems (Altieri et al., 2017; Delonge et al., 2016). This shift requires further understanding of the degree to which plant hosts can be bred to manage their microbiome as well as the consequential trade-offs.

Fitness trade-offs might make breeding a cash crop for a specific community assembly a difficult task. However, we may be able to bypass this problem by breeding a low-value rotation or cover crop specifically for community assembly traits that would benefit the subsequent high-value crop. Depending on the system, this could allow breeders to focus on yield traits in a high-value crop while still hopefully reaping the benefits of an optimally crafted microbiome. A system like this would not come without potential pitfalls. Pathogens tend to be more specialized than mutualists, but two crops related enough to share important mutualists could potentially share pathogens as well. In addition, the legacy effect of the community cultivating crop would have to exist for a long enough duration to benefit the cash crop when it is planted. These issues cannot be evaluated or overcome without further study.

Aside from the potential economic value, there is much to be learned from exploring the relationship between plant genetics and plant-associated microbes. By screening for plant genomic regions that significantly correlate with measurements of microbial community structure, we can provide evidence for putative mutualists/parasites, generate hypotheses about how the plant interacts with these microbes, and start to build an understanding of whether the genetic interaction a plant exhibits with a particular microbial taxon is a qualitative relationship

or a quantitative multigenic relationship. By using GWAS we can explore two main ideas that provide insight into the genetic architecture of relationships between the oat host and its microbial symbionts. 1) The heritability of a given microbial population and 2) microbial association with single-nucleotide polymorphisms in the oat genome. In other words, this first idea answers the question “do certain genotypes exhibit differential cultivation of certain microbes?” and the second idea answers the question “is a genotype associated microbial trait controlled by a few genes of large effect or by many genes of small effect?” By comparing these two pieces of information we can identify important oat associates and form hypotheses about their relationship to the host plant. In general, based on our understanding of the gene-for-gene model of plant/pathogen interactions, we would predict that pathogenic microbes would be more likely to associate with oat genome SNPs (i.e. their populations are influenced by a few genes of large effect). Mycorrhizal mutualists, on the other hand, are typically quite general in their host range and we would predict their populations to exhibit low heritability and lack significant association with SNPs (Dowarah et al., 2022). Community summaries such as diversity measurements would likely be affected by many genes and therefore would lack association with any SNPs.

In this thesis I explore the genetic relationships between oats and their associated microbial communities, as well as the impact those relationships have on the oat soil legacy. I studied oat-associated rhizosphere communities and their impacts on oat/corn traits in both a greenhouse and field setting with the main goal determining whether we can exploit genotypic variation in oats to cultivate a microbial community that improves productivity of subsequently planted corn crops. The results of this project are outlined in the following chapters.

Literature Cited.

Altieri, M. A., Nicholls, C. I., & Montalba, R. (2017). Technological approaches to sustainable agriculture at a crossroads: An agroecological perspective. *Sustainability (Switzerland)*, *9*(3), 1–13. https://doi.org/10.3390/su9030349

Bender, S. F., Wagg, C., & Van Der Heijden, M. G. A. (2016). An Underground Revolution: Biodiversity and Soil Ecological Engineering for Agricultural Sustainability. *Trends in Ecology & Evolution*, *31*(6), 440–452. https://doi.org/10.1016/j.tree.2016.02.016

DeLonge, M. S., Miles, A., & Carlisle, L. (2016). Investing in the transition to sustainable agriculture. *Environmental Science & Policy*, *55*, 266–273. https://doi.org/10.1016/J.ENVSCI.2015.09.013

Dowarah, B., Gill, S.S. & Agarwala, N. Arbuscular Mycorrhizal Fungi in Conferring Tolerance to Biotic Stresses in Plants. *J Plant Growth Regul* **41**, 1429–1444 (2022). https://doi.org/10.1007/s00344-021-10392-5

Ellouze, W., Hamel, C., Vujanovic, V., Gan, Y., Bouzid, S., & St-Arnaud, M. (2013). Chickpea genotypes shape the soil microbiome and affect the establishment of the subsequent durum wheat crop in the semiarid North American Great Plains. *Soil Biology and Biochemistry*, *63*, 129–141. https://doi.org/10.1016/j.soilbio.2013.04.001

Gopal, M., & Gupta, A. (2016). Microbiome selection could spur next-generation plant breeding strategies. *Frontiers in Microbiology*, *7*(DEC), 1–10. https://doi.org/10.3389/fmicb.2016.01971

Jernigan, A. B., Wickings, K., Mohler, C. L., Caldwell, B. A., Pelzer, C. J., Wayman, S., & Ryan, M. R. (2020). Legacy effects of contrasting organic grain cropping systems on soil health indicators, soil invertebrates, weeds, and crop yield. *Agricultural Systems*, *177*, 102719. https://doi.org/10.1016/J.AGSY.2019.102719

Jez, J. M., Lee, S. G., & Sherp, A. M. (2016). The next green movement: Plant biology for the environment and sustainability. *Pl. trans. biol. 353*(6305).

Kiers E.T., Hutton J.G., and Denison R.F. 2007. Human selection and the relaxation of legume defences against ineffective rhizobia. Proc. R. Soc. B.2743119–3126

Kim, H., Lee, K. K., Jeon, J., Harris, W. A., & Lee, Y. H. (2020). Domestication of Oryza species eco-evolutionarily shapes bacterial and fungal communities in rice seed. *Microbiome*, *8*(1), 1–17. https://doi.org/10.1186/s40168-020-00805-0

Marques, J. M., da Silva, T. F., Vollu, R. E., Blank, A. F., Ding, G. C., Seldin, L., & Smalla, K. (2014). Plant age and genotype affect the bacterial community composition in the tuber rhizosphere of field-grown sweet potato plants. *FEMS Microbiology Ecology*, *88*(2), 424–435. https://doi.org/10.1111/1574-6941.12313

Moe, L. A. (2013). Amino acids in the rhizosphere: From plants to microbes. *American Journal of Botany*, *100*(9), 1692–1705. https://doi.org/10.3732/ajb.1300033

Paulitz, T. C., Schroeder, K. L., & Schillinger, W. F. (2010). Soilborne pathogens of cereals in an irrigated cropping system: Effects of tillage, residue management, and crop rotation. *Plant Disease*, *94*(1), 61–68. https://doi.org/10.1094/PDIS-94-1-0061

Porter, S. S., and J. L. Sachs. 2020. Agriculture and the Disruption of Plant-Microbial Symbiosis. Trends in Ecology & Evolution **35**:426-439.

Qiao, Q., Wang, F., Zhang, J., Chen, Y., Zhang, C., Liu, G., Zhang, H., Ma, C., & Zhang, J. (2017). The Variation in the Rhizosphere Microbiome of Cotton with Soil Type, Genotype and Developmental Stage. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-04213-7

Song, X., Tao, B., Guo, J., Li, J., & Chen, G. (2018). Changes in the microbial community structure and soil chemical properties of vertisols under different cropping systems in northern China. *Frontiers in Environmental Science*, *6*(NOV), 1–14. https://doi.org/10.3389/fenvs.2018.00132

Wagg, C., Schlaeppi, K., Banerjee, S., Kuramae, E. E., & van der Heijden, M. G. A. (2019). Fungal-bacterial diversity and microbiome complexity predict ecosystem functioning. *Nature Communications*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-12798-y

Weston, L. A., & Mathesius, U. (2013). Flavonoids: Their Structure, Biosynthesis and Role in the Rhizosphere, Including Allelopathy. *Journal of Chemical Ecology*, *39*(2), 283–297. https://doi.org/10.1007/s10886-013-0248-5

Zhang, H., Mittal, N., Leamy, L. J., Barazani, O., & Song, B. H. (2017). Back into the wild—Apply untapped genetic diversity of wild relatives for crop improvement. *Evolutionary Applications*, *10*(1), 5–24. https://doi.org/10.1111/eva.12434

Chapter 2

Assessing the Oat Soil Legacy Effect at the Genotype Level in a Greenhouse Setting

Abstract

Crop rotations have long been used to improve agricultural outcomes via plant soil legacy effects. Recently, soil legacy effects have been shown to differentiate even at the level of crop genotype. This genotypic variation in soil legacy offers the possibility to further optimize field rotations by selecting rotational crop genotypes with the most beneficial soil legacy effect for the next crop in rotation. In this study, we sequenced the soil microbial communities associated with eight different oat genotypes in a field setting. We then used soil collected from those field plots as inoculum in a greenhouse potting experiment to test the oat genotype specific soil legacy effect on subsequently planted oat and corn plants. We show evidence that oat genotype significantly impacts community fungal community composition as well as fungal and bacterial diversity. In addition, we show oat genotype specific soil legacies differentially affect the phenotypic outcomes of both subsequently planted oat and corn plants in high and/or low nutrient conditions. These data could be a preliminary step towards crop rotations that are refined and optimized to the genotype level.

Introduction

It has long been understood that crop rotations can increase soil health and crop yield. These benefits are, in part, conferred through soil legacy effects, by which the abiotic and biotic alterations made in the surrounding soil by one host plant have an impact on a subsequent host (Jernigan et al., 2020, (Mariotte et al. 2018). One of the most important biotic manipulations made through soil rotations is the depletion of host specific pathogens, and growers have been using this strategy to successfully reduce disease pressure for centuries. Less understood are the

many other biotic costs and benefits that may be conferred through a rotational system. The complexity of the microbial communities associated with host crops makes them especially difficult to study; however, with the advent of next-generation sequencing technology, researchers have been taking greater strides in elucidating these complex interactions.

Crop rotations can affect community structure and increase microbial diversity to varying degrees across several systems (Song et al., 2018; Jiang et al. 2016). Additionally, increased microbial diversity has been correlated with increased ecosystem multifunctionality and increased plant productivity (Wagg et al. 2019; Tautges et al. 2016). Many of the host benefits conferred by soil rotations may be attributable to an increase in functional traits available to the holobiont. Given how much the fitness of a host plant can be impacted by its associated microbial community, it should be no surprise that plants have evolved to manipulate the structure of these communities. By secreting a bevy of carbon compounds and secondary metabolites into the soil, plants are able to manipulate microbial community structure to their benefit (Moe 2013; Weston and Methesius 2013). However, plants are not in complete control of their associated microbial communities. In fact, in both agricultural and natural systems, the soil legacy of a plant species is overwhelmingly more likely to be negative for subsequent members of that plant species, compared to an unrelated plant species (Kulmatiski 2008).

Given the plant capacity for microbiome alteration there is potential for relying on plant hosts to make the agronomically desirably adjustments to a given soil community. In addition, the variation in microbial community assembly observed between the genotypes of some crops suggests the potential to direct these host driven community adjustments through plant breeding (Gopal and Gupta 2016). Before such a breeding program can exist, it is necessary to carefully

explore the genetic variation of soil legacies within a given crop species and to evaluate the potential outcomes these legacies may have on subsequently planted crops.

To that end, our study used a plant-soil feedback approach in a greenhouse setting, to test for differences in oat (*Avena sativa*) and corn (*Zea mays*) trait outcomes after inoculation with living soil cultivated by eight different oat genotypes. Plant-soil feedback experiments are commonly used by plant ecologists to study how one plant species impacts soil conditions, and the consequences of these impacts con- or heterospecific plant species (Klironomos 2002, Kulmatiski et al. 2008, Mariotte et al. 2018). These experiments involve a conditioning phase, in which different plant species (or genotypes) are grown with a common soil in either field or controlled conditions, followed by a testing phase, in which new plants of the same or different species are then grown in controlled conditions using the previously conditioned soils (Klironomos 2002). In our case, we used eight oat (*Avena sativa*) genotypes grown in a randomized complete block design field experiment, with three replicate field plots per genotype, to condition soil under realistic field conditions for one growing season, and then used these conditioned soils from the field plot as inoculum in a greenhouse experiment testing the growth of oat and corn plants.

The objectives of this study were (i) to evaluate whether oat (*Avena sativa*) genotypes exhibit detectable differences in host-associated, soil microbial communities, (ii) to determine if these differences in community structure affect the growth of subsequently planted oat and corn plants in high and low nutrient environments, and (iii) to analyze correlations between microbial community structure and plant host phenotypes. We hypothesized that (i) oat genotype would significantly predict soil microbial community structure, (ii) the microbiome inoculating oat

genotype would describe more variation in oat and corn outcomes in low nutrient conditions, and (iii) community traits such as diversity would significantly correlate with plant host phenotypes.

Materials and Methods

Inoculum Source: We collected live soils to use as inoculum for both oat and corn plants from a field experiment conducted at the West Madison Agricultural Research Station. This field experiment compared agronomic traits of 16 different oat varieties grown in three replicated plots in each of three spatial blocks. Eight of these oat genotypes were selected to provide inoculum soil for this experiment based on the preliminary data to achieve a wide variation in microbial community diversity. For each of the eight selected genotypes, we collected soil from each of the three replicate field plots, and kept these soils separate (for a total of 24 distinct soil collections). Inoculum soil was collected at heading time (>50% of plants headed). Soil collection was done using a garden trowel to remove the top layer of soil next to oat plants within a plot and harvesting soil 10 cm deep and within 7.5 cm of the oat root structure. The living soil inocula were stored under refrigerated conditions (4 ºC) for ~6 months before being used in this experiment to simulate overwintering.

Oat Greenhouse Experiment: To test the genotype-specific oat soil legacy effect on subsequently planted oat plants, we used a greenhouse pot experiment inoculating oat plants with living soil inocula collected from oat field plots. Eight varieties of oats were planted in this study. These corresponded to the same 8 varieties from which soil inocula were collected. Oat plants were grown in 1.6 L treepots (10.2 X 10.2 X 24.1 cm, Stuewe & Sons, Inc., Tangent, OR). These pots were filled with 95% (by volume) autoclaved background soil collected from the West Madison Agricultural Research Station and 5% living, inoculum soil collected from one of three replicate field plots from each of 8 oat genotypes. Each oat genotype used in the

greenhouse experiment was grown with microbial inocula from three different conditioning oat genotypes from the field experiment (see Table 1), with three replicate pots for each oat genotype X soil inocula source combination. In each replicate, the soil inocula came from a separate replicate field plot (three replicate field plots per oat conditioning genotype).

After soil inocula were added, the pots were left to sit for seven days (being watered every other day) to give the microbial communities time to respond to and colonize the autoclaved background soil. At the end of this period, oat seeds were planted into each pot at a depth of one inch. Each oat variety was subjected to four inoculation treatments (living inocula from three oat varieties and one sterile control, Table 1), and either a high or low nutrient treatment. High nutrient pots were fertilized a single time, once all plants had germinated, with 3g of Nutricote slow-release fertilizer (13-13-13, Arysta LifeSciences). Low nutrient pots received no additional nutrient input. The treated pots were arranged in a completely randomized design to distribute variation in greenhouse environment between the experimental units. Plants were carefully hand watered once every 2-3 days to limit variation water provided. The watering regimen was selected so water would not be a limiting resource for any plant.

Once grain was able to be harvested (~9 weeks post planting), oat plants were destructively sampled. The below ground portion of each plant was carefully washed to remove any soil clinging to the root structure. Both the above and below ground portions of each plant were dried at 65˚C for one week then weighed to determine biomass. After the above ground portions were weighed, the grain was removed and threshed using a small seed grain threshing machine. The weight and volume of the dried grain collected from each pot was then recorded and used to calculate grain density. Grain density was used as a proxy for test weight because the quantities of grain collected in this experiment were insufficient to measure test weight.

Corn Greenhouse Experiment: To test the genotype-specific oat soil legacy effect on subsequently planted corn, we also ran a greenhouse pot experiment inoculating corn plants with living soil inocula from the same eight different oat varieties. One elite, hybrid variety of corn was used in this study. Planter pots (11.3 L) were filled with living and autoclaved background soils at the same rates and using the same methods as described for oats. One corn kernel was planted into each pot at a depth of 4 cm. Corn plants were subjected to nine microbial inoculation treatments (living inocula from all eight oat varieties and one sterile control, Table 1) and either a high or low nutrient treatment. Again, there were three replicate pots for each soil inocula type and nutrient combination, and for each replicate the soil inocula came from a separate replicate field plot of the same oat genotype. High nutrient pots were fertilized a single time, once all plants had germinated, with 20g of Nutricote slow-release fertilizer (13-13-13, Arysta LifeSciences). Low nutrient pots received no additional nutrient input. The treated pots were arranged in a completely randomized design to distribute variation in the greenhouse environment between the experimental units. Corn plants were watered in the same manner described for oats. After 10 weeks of growth, corn plants were destructively sampled, dried, and weighed in the same manner described for oats. This timing was chosen to prevent the plants from becoming root bound in their planter pots and was insufficient to bring corn to harvest.

Amplicon preparation and sequencing: DNA from the initial soil inoculum samples were extracted using a Qiagen DNeasy PowerSoil Kit according to the manufacturer's instructions (Qiagen, Hilden, Germany). To characterize 16S and archaea communities, the prokaryotic 16S-V4-V5 region was amplified using 515F forward primer (5'-GTGYCAGCMGCCGCGGTAA, (Caporaso et al. 2011) and 926R reverse primer (5'-CCGYCAATTYMTTTRAGTTT, Parada et

al 2016). For fungal community characterization, the fungal ITS2 sequence was amplified using ITS3-KYO2 forward primer (5'-GATGAAGAACGYAGYRAA, (Toju et al. 2012)) and ITS4 reverse primer (5'-TCCTCCGCTTATTGATATGC, (White et al. 1990)). External fusion PCR primers contained a 14-bp overlap to the trailing end of internal primers with 12bp i7 index and P7 flow cell adaptor or an i5 index, 7-bp spacer and P5 adapter (Lankau and Keymer 2016). Amplicon library preparation was composed of two PCR steps. The first round of PCR amplified the ITS2 or 16S V4-5 region along with associated Nextera read primers. PCR was performed in 10 µl reactions using 0.2 µL of a hot-start, high fidelity polymerase (Clonetech Prime Star GLX, Fitchburg, WI) with 2 µL of its 5X buffer, 0.8 µL dNTPs (at 10 nM concentration), 0.25 µL of each primer (at 10 nM), 0.7 µg T4 gene 32 protein, and 10 ng of template DNA. The thermocycling program for the ITS2 region was a 5-minute hot start at 98˚C, 35 cycles of denaturing (98˚C, 0:30), annealing (50˚C, 0:45), and extension (68˚C, 1:00) and a final extension of 15 minutes at 68˚C. The thermocycling program for the 16S region was 5-minute hot start at 98˚C, 35 cycles of denaturing (98˚C, 0:45), annealing (50˚C, 0:45), and extension (68˚C, 1:00) and a final extension of 15 minutes at 68˚C. Successful amplification was verified using agarose gel electrophoresis. The second round of PCR added the P5 and P7 flowcell adapters to prepare the library for sequencing on an Illumina MiSeq, along with an external set of sample barcodes located between the flowcell adaptors and read primers. Fungal ITS2 and 16S amplicons were cleaned with the Omega BioTek E-Z 96 Cycle Pure kit. Purified products were quantified using a Qubit 2.0 fluorometer with the Qubit dsDNA HS assay and then pooled at equal concentrations (Thermo Scientific, Grand Island, NY). Amplicon products were then sequenced on Illumina Miseq using a 300 cycle Paired-End run at the University of Wisconsin-Madison Biotechnology Center.

Bioinformatics: Raw external sequences were initially trimmed at both the 5' and 3' ends using Cutadapt (version 1.18). The Qiime2 (v2017.12) pipeline was used to processed trimmed reads using DADA2. Samples were filtered and further trimmed by DADA2 using the following parameters (16S-V4 sequences: truncLen = (0, 280), p-max-ee = 4; ITS2 sequences: truncLen = (0, 260), p-max-ee = 4). We used the RDP Naïve Bayesian Classifier to assign taxonomy to bacterial and fungal amplicon sequence variants (AVS) using SILVA (version 13.8) and UNITE (version 8.0) reference databases for bacteria and fungi, respectively. Reads assigned to chloroplast and mitochondria were removed. Based on inspection of rarefaction curves, 16S samples with fewer than 500 reads were removed from the study, and it was unnecessary to remove any ITS samples. Out of the 24 soil inocula communities sampled for this study, 23 16S and 24 ITS remained for analysis. Bacterial and fungal ASV sequence counts were transformed into relative abundance values using the "vegan" package in R (Oksanen et al. 2015).

Statistical Analysis:

Obj. 1: Evaluate whether oat (Avena sativa) genotypes exhibit detectable differences in host-associated, soil microbial communities.

To test whether the microbial community structure of soil samples differed significantly between conditioning oat genotypes we used a permutational MANOVA (perMANOVA) using the adonis function within the "vegan" package. We created Principal Coordinates Analysis (PcoA) plots to visualize fungal and bacterial compositional variation, using the capscale function in the vegan package. These PcoA plots were created using Bray-Curtis distance matrices of the relative abundances of each fungal/bacterial genus. PcoA results were visualized using the R package "ggplot2" (Wickham 2016). A perMANOVA p-value threshold of less than 0.05 was

used to determine whether oat genotype had a significant effect on the microbial community composition.

To test whether soil conditioning oat genotype affects microbial community diversity, we calculated the Shannon-Weiner diversity index for both fungal and bacterial communities using the diversity function in the vegan package. We performed linear mixed models using the lmer package in R that included oat conditioning genotype as a random effect, to calculate the proportion of variation attributable to conditioning genotype (the conditioning genotype variance component). We also ran linear models that used oat conditioning genotype as a fixed effect, to calculate the statistical significance of the genotype effect. A P-value threshold of 0.05 for the F-test was used to determine whether genotype had a significant effect on microbial diversity.

We also tested whether soil conditioning oat genotype affected the relative abundances of individual clades or genera. To select candidates for genotype association we used a filtering process based on total relative abundance and frequency across samples. First, fungal and bacterial ASV's were collapsed by genus resulting in 98 fungal genera and 78 bacterial genera. Of these initial genera, those that showed an average relative abundance of >0.1% of the total fungal/bacterial abundance within a sample and were present in >50% of all samples were subjected to further analysis. This filtering step resulted in a remaining pool of 20 fungal genera and 10 bacterial genera. Then, we used linear mixed models and fixed effect models as described above to quantify how much of the variation in their relative abundances was attributable to the oat genotype from which inoculum soils were sampled. Genera for which microbiome cultivating oat genotype described greater than 10% of the variation in relative abundance and a fixed effect p-value of <0.10 highlighted as potential important taxa underlying the oat genotype on microbial community structure and function.

Obj. 2. Determine if these differences in community structure affect the growth of subsequently planted oat and corn plants in high and low nutrient environments.

We used linear mixed effect models using the "lme4" package in R to determine how much variation in each oat and corn trait of interest was attributable to the oat microbiome conditioning genotype. For oats, we analyzed total biomass (shoot, root, and grain dry mass), aboveground dry mass, belowground dry mass, the shoot:root ratio, and the grain density. Since the oat experiment used multiple oat genotypes in the experiment, as well as microbial inocula sourced from multiple oat genotypes in the field, we used linear mixed models with two crossed random effects – host oat genotype and microbiome conditioning oat genotype. Because host oat genotypes were not fully crossed with microbiome conditioning genotypes, we did not test for interactions between these two variables. As before, to determine statistical significance of the host genotype and microbiome conditioning genotype effects on oat phenotypes, we also ran linear models with both factors as fixed effects. These models were run separately for high and low nutrient conditions. Nutrient level was tested as a fixed effect to verify the treatment had a significant effect in the experiment.

For corn phenotypes, we used linear mixed models with oat microbiome conditioning genotype as the sole random effect since only one corn genotype was used in the experiment. Corn phenotypes included aboveground dry mass, belowground dry mass, total dry mass, and the shoot:root ratio. To estimate the variance component of the oat microbiome conditioning genotype, we ran separate linear mixed models for each corn phenotypic trait for high and low nutrient conditions. We also ran linear models with oat microbiome conditioning genotype as a fixed effect for each trait in both high and low nutrient conditions to determine statistical

significance. As with oats, these models were run separately for high and low nutrient conditions, and nutrient level was tested as a fixed effect to verify the success of this treatment.

Obj. 3. Analyze correlations between initial microbial community structure and plant host phenotypes. If the microbial community inocula sourced from field plots containing different oat genotypes had effects on subsequent oat and/or corn plant growth, this must be due to changes the structure of the microbial communities. To explore potential microbial mediators of these inocula effects, we performed a series of analyses correlating oat and corn phenotypes with microbial community diversity and structure of the initial inoculum. Using linear models we tested for correlations between Shannon-Weiner diversity index (in fungal and bacterial communities) and the oat and corn phenotypes described in objective 2. We used perMANOVA to correlate these crop phenotypes to bacterial or fungal community structure. These perMANOVA results were visualized in PcoA plots presented using ggplot2 as described in objective 1. To test whether correlations between response variables and community structure is occurring in a non-linear manner (i.e. plant response changes based on departure in any direction from some "ideal" or "worst-case" community) we applied a smooth curve analyses to these PcoA plots using the ordisurf() function in "vegan."

In total, we used 24 distinct microbial inocula in each nutrient condition for each crop species. Because we grew three different oat genotypes in each distinct microbial inocula, to correlate diversity indices to oat phenotypes of interest, we first used the "emmeans" package in R to calculate estimated means of phenotypic traits for the oats treated with each distinct microbial inocula, while controlling for oat host genotype effects (Searle et al. 1980). This was unnecessary for corn phenotypes, as only one corn variety was used in this study.

Results

Obj. 1: Evaluate whether oat (Avena sativa) genotypes exhibit detectable differences in host-associated, soil microbial communities.

PerMANOVA analysis showed strong evidence for the clustering of soil fungal communities by conditioning oat genotype, but little to no evidence for clustering of bacterial communities (Fig. 1; $p = 0.046$ and 0.158 respectively). For the fungal communities, this oat genotype association was also present when evaluating community composition as presence/absence of fungal species rather than relative abundances ($p = 0.0645$), however, this was not observed for the bacterial communities ($p = 0.1217$). The microbiome cultivating oat genotype showed strong control over Shannon's diversity index at the species level, describing 42.6% and 25.1% of the variation in fungal and bacterial communities respectively (Fig. 2). The soil cultivating oat genotypes exhibiting the highest and lowest diversity indices were relatively stable across fungal and bacterial communities, with Antigo and IL (some genotype names have been simplified to improve clarity) at the bottom and esker and Deon within the top three for both communities.

Twenty fungal and 10 bacterial genera that made it past the filtering process (average relative abundance >0.1% and present in >50% of samples). When tested as a random effect, microbiome cultivating oat genotype described >10% of variation in the relative abundances of 9/20 fungal genera and 5/10 bacterial; However, the fixed effect p-value of microbiome cultivating oat genotype was only below the significance threshold ($p < 0.10$) for 3/20 fungal genera and no bacterial genera (Tables 2 & 3). All three of these genotype associating fungal genera are known to contain important oat pathogens.

Obj. 2. Determine if these differences in community structure affect the growth of subsequently planted oat and corn plants in high and low nutrient environments.

Effects on oat phenotypes: The community cultivating oat genotype also showed strong control over trait outcomes in the inoculated host oat and corn plants (Table 4,5, Fig. 3-5). For oats, the inoculated host genotype generally described more of the total variation in phenotype in high nutrient conditions, while the effect of the microbiome cultivating (inoculum providing) oat genotype was dependent on the measured phenotype. In most cases, the non-inoculated control plants (planted in only autoclaved background soil) performed best, implying a net negative effect of the microbial community in these conditions. The microbiome cultivating oat genotype described 21.9% of the variation in the above ground biomass of the host plant in high nutrient conditions ($p = 0.1027$) and 11.81% of the variation in low nutrient conditions ($p = 0.147$, Table 4, Fig. 3). This trend is inverted when it comes to the below ground biomass of oats, where the microbiome cultivating oat genotype describes 21.8% of the variation in low nutrient conditions ($p = 0.0025$) and 3.2% of variation in high nutrient conditions ($p = 0.3499$, Table 4, Fig. 3). For both above and below ground oat biomass, the significance of and variation attributed to the genotype of the inoculated host plant is much higher in high nutrient rather than low nutrient conditions.

The ratio of above/below ground biomass can offer insight into how the microbial communities affected the resource allocation of the host plant. In low nutrient conditions the microbiome cultivating oat genotype described 21.7% of the variance in the ratio of above ground biomass to below ground biomass in oats ($p < 0.001$, Table 4, Fig. 3), while the inoculated host genotype described 6.2% of the variance. In high nutrient conditions the

microbiome cultivating oat genotype explained only 1.3% of the total variation in this trait ($p = 0.5598$, Table 4, Fig. 3).

In terms of grain yield, both inoculated host genotype and microbiome cultivating genotype displayed a significant effect under high nutrient conditions, describing 5.50% ($p = 0.00916$) and 19.8% ($p = 0.07744$) of the variation respectively (Table 4, Fig. 4). In low nutrient conditions inoculated host genotype described 0% of the variation in grain yield, while microbiome cultivating genotype described 6.6% ($p = 0.1263$, Table 4). Inoculated host genotype described 27.0% of variation in grain density, while microbiome conditioning genotype described 0.0% (Table 4, Fig. 4). In low nutrient conditions host and microbiome conditioning genotype described 2.7% and 2.4% respectively (Table 4, Fig. 4).

Effects on corn phenotypes: The microbiome conditioning oat genotype also showed strong influence over inoculated corn phenotypes. Only one variety of corn was tested, so inoculated host genotype is not a factor for corn growth in this study. As observed in oats, corn plants in non-inoculated pots tended to perform better than inoculated plants. This again implies a net-negative effect of the microbial community in these conditions; however, in the low nutrient treatments some microbiome conditioning genotypes (e.g. Esker, Deon, IL) were on par with the non-inoculated control with respect to corn total biomass. In general, the microbiome cultivating oat genotype only affected host corn phenotypes in high nutrient conditions, describing 26.3% of the variation in above ground biomass ($p = 0.01527$) and 25.2% of the variation in below ground biomass ($p = 0.09674$, Table 5, Fig. 5). The ratio of above to below ground biomass was an exception to this trend, with microbiome cultivating genotype describing 14.1% of the variation in the low nutrient treatment ($p = 0.2549$) and 10.4% of the variation in the high nutrient treatment (Table 5, Fig. 5). Corn biomass also displayed a strong interaction effect between

microbiome cultivating genotype and nutrient level, with the best performing microbiome genotypes in one condition sometimes performing the worst in the other (Fig. 5). In comparison, oat host phenotypes showed no interaction between these explanatory variables, and the ranking of best to worst microbiome genotypes did not shift from high to low nutrient conditions.

Obj. 3. Analyze correlations between initial microbial community structure and plant host phenotypes.

The Shannon-Weiner diversity indices of fungal and bacterial communities of the soil inocula were significantly positively correlated with various oat phenotypes, but not with any corn phenotypes. The significance of bacterial and fungal diversity on oat phenotypes closely mirrored the significance of the fixed effect of microbiome cultivating genotype on each of these phenotypes. The Shannon-Weiner diversity indices correlated positively with oat above ground biomass in high nutrient conditions for both fungal (Fig. 6; $p = 0.0205$) and bacterial (Fig. 7; $p = 0.0065$) communities, while below ground biomass showed significant correlations with diversity in low nutrient conditions for both fungal (Fig. 6; $p = 0.0047$) and bacterial (Fig. 7; $p = 0.0031$) communities. The ratio between above and below ground biomass in oats was only significantly described by microbial diversity in low nutrient conditions, again by both fungal (Fig. 6; $p = 0.0004$) and bacterial (Fig. 7; $p = 0.0005$) communities. Oat yield only significantly correlated with diversity in high nutrient conditions, showing nearly identical explanatory power for both fungal (Fig. 8; $p = 0.0031$) and bacterial (Fig. 8; $p = 0.0033$) diversity. Only bacterial diversity showed any influence over oat grain density, displaying significance in high nutrient conditions ($p = 0.0446$), but not in low nutrient conditions ($p = 0.7090$).

While microbiome cultivating oat genotype did describe significant variation in corn biomass under high nutrient conditions, fungal and bacterial species diversity did not significantly correlate with any of corn phenotypes in either nutrient condition. However, some corn phenotypes displayed correlations with the community composition of the inoculated fungal and bacterial communities, as evidenced by perMANOVA. In low nutrient conditions, the ratio of above to below ground biomass in corn showed a weak correlation with both fungal and bacterial community structure. For fungal communities this trend was most significant when analyzed using a community abundance matrix ($p = 0.0691$) while bacterial community structure most significantly described the biomass ratio when analyzed using a presence/absence matrix (Fig. 9; $p = 0.0654$). The overall biomass of corn in high nutrient conditions showed no evidence for correlation with fungal community structure ($p = 0.2523$) and weak evidence for a correlation with bacterial community structure ($p = 0.0638$); however, overlaying these PCoA plots with a smooth curve analysis shows evidence for a non-linear correlation between corn total biomass and community structure of both fungi ($p = 0.0164$) and bacteria (Fig. 9; $p = 0.00484$). In both cases, high corn biomass was associated with communities falling near the center of the ordination space, while low corn biomass was associated with communities spread around the edges of the ordination space.

Discussion

Soil legacies, in which one plant conditions the biotic and/or abiotic properties of soil in a way that affects the growth of subsequent plants in that soil, are ubiquitous in natural and managed ecosystems and are the driving force behind the benefits of crop rotation. However, soil legacies have been studied primarily between plant species. If genetic variation existed within a crop species that modulated the strength or direction of soil legacy effects, this could open a new avenue for plant breeders to target traits that not only optimize the agronomic value of the bred crop, but also leave beneficial legacies for future crops. This study confirmed the hypothesis that oat genotypes cultivate specific microbial soil legacies which, in turn, have variable outcomes on subsequently planted oat and corn hosts. However, the effects of a given microbial community differed across nutrient contexts, and across the two target crop species. Finally, microbial diversity of the conditioned soil was a strong predictor of oat phenotypes when grown with these communities, while corn growth was independent of diversity.

Comparing microbial communities in soils collected from replicate field plots sown to eight different oat genotypes demonstrated that oat genotype had a significant impact on fungal community composition, but little to no impact on bacterial community composition. On the other hand, oat genotypes had strong and consistent effects fungal diversity in soils, and weaker effects on bacterial diversity. The tendency to detect stronger oat genotype effects on soil fungi compared to bacteria may reflect the many fungal taxa that live in intimate association with plant roots, such as arbuscular mycorrhizal fungi, root endophytes, and pathogens. Indeed, when analyzing the effect of oat genotypic variation on relative abundance of specific microbial guilds and genera, no bacterial genera demonstrated statistically significant oat genotype effects. However, several fungal genera or guilds did vary substantially among oat genotypes, and these

came from known root mutualists (arbuscular mycorrhizal fungi) or pathogen/endophyte groups (Giberrela, Fusarium, Colletotrichum). This study did not investigate the specific aspects of oat genotypes that led to development of particular soil microbial communities, but these results suggest that these effects may arise from variation in specific interactions with intimate root colonizing fungi.

The differences observed in microbial community composition and diversity in soil conditioned by different oat genotypes translated into impacts on the growth of both oat and corn plants in our greenhouse experiment. Consistent with our original hypothesis, the total biomass of oats was most effected by microbiome cultivating genotype communities in low nutrient conditions. However, looking at above and below ground biomass individually showed this trend to be inconsistent, with above ground biomass being most affected in high nutrient conditions and below ground most affected in low nutrient conditions. Since the ratio of above to below ground biomass was also significantly influenced by microbiome cultivating genotype, these data speak to the general ability of the microbial community to manipulate host resource allocation in various nutrient contexts. In high nutrient contexts, taxa in some communities may allow more nutrients to travel to the above ground portion of the plants or differential parasite loads could have had a deleterious effect on root biomass. As predicted, oat plants invested relatively more to root systems in low compared to high nutrient conditions, reflecting the greater limitation by soil resources. Differential root biomass in oat plants inoculated with different microbial communities could again reflect access to mutualist root symbionts (like arbuscular mycorrhizal fungi) or microbial communities that differ in net rates of biogeochemical processes, such as nitrogen mineralization. Importantly, despite the subtle differences in oat allocation patterns, there was a remarkable consistency in the ranking of oat microbiome conditioning genotype

effects on subsequent oat growth across nutrient contexts. This suggests that the growth promoting or growth depressing aspects of the oat conditioned microbial communities were consistent across nutrient availabilities.

While no microbial taxon singularly explained the trends in oat growth observed in this study, there were several significantly associated with oat genotype and could be candidates for further study. Significant correlations with microbial diversity were observed in every oat phenotype which significantly associated with microbiome cultivating oat genotype. This is especially interesting as fungal diversity, especially, was strongly determined by the oat microbiome conditioning genotype. Microbial diversity was positively correlated for all growth phenotypes, which is consistent with many studies showing a positive relationship between microbial diversity and plant-host fitness. Positive microbiome diversity-host productivity relationships can stem from several potential mechanisms (Saleem et al. 2019). More diverse soil communities may have higher niche complementarity, in which more total niche space is filled resulting in greater efficiency of resource use (Tilman 1999, Loreau 2000). Higher diversity can increase the probability of key growth promoting taxa being present in a given community, known as sampling effects, or increase redundancy to allow more consistent levels of function despite fluctuations in individual populations (Allison and Martiny 2008, Miki et al. 2014). Finally, more diverse communities may increase the probability of synergistic activities among microbes with positive consequences for plant growth (Saleem et al. 2019).

Contrary to our expectations, corn phenotypes were only affected by oat genotype microbial communities in high, rather than low, nutrient conditions. Many studies show plant hosts exhibit a greater positive growth response to microbial communities in a resource poor context (Hoeksema et al. 2010, Johnson et al. 2015). In interpreting these results, it is important

to remember these are elite varieties of oats and corn that have been bred for optimized yield in high input agroecosystems. Domesticated plants may have reduced capacity to interact with beneficial members of their associated microbial communities (Perez-Jaramillo et al. 2016). Many studies show evidence that the reduced benefit from microbial partners in modern crop varieties compared to wild relatives stems from a reduction in the crop plants dependence on the microbial services (Porter and Sachs 2020). In other words, the elite crop varieties were able to maintain high growth in the absence of microbial symbionts, while wild relatives have greatly reduced growth without these symbionts (Sawers et al. 2010, Martin-Robles et al. 2018, Porter and Sachs 2020). It is possible the plants in our study simply lack the capability to interact as fully with the inoculated microbes, and/or are insensitive to the benefits provided due to an enhanced ability to acquire resources directly through roots. The variation in corn growth observed in the high nutrient treatments could have been due to microbial competition for the easily-accessible, mineralized NPK provided, rather than directly beneficial interactions. Alternatively, the variation in corn growth in high nutrients could have stemmed from differential pathogen/parasite loads across the different microbial inocula. While evidence suggests that plant breeders have inadvertently bred for a reduction in plant-microbe interactivity, it is also possible that there has been inadvertent breeding for the particular plant-microbe interactions that dominate in the high nutrient contexts typically seen in agroecosystems.

While this study took place in a greenhouse setting with a high degree of control over sources of variation, other studies have detected genotypic effects on microbial community structure in field conditions (Gopal and Gupta 2016). The data in this study raise the possibility of a soil management program that considers crop genotype as an extra level of control when it

comes to manipulating the microbial community in a given field. Interestingly, the microbial communities that performed the best for oat hosts did not shift from one nutrient context to another. The opposite phenomenon can be observed in corn where there was a strong interaction between microbiome cultivating genotype and nutrient treatment. This implies a more stable interaction between oat hosts and their own cultivated microbial communities than the interaction between oat cultivated communities and corn plants. Even if genetic variation in microbiome selection exists, given the inconsistency between best/worst performing microbiome communities in oat vs. corn hosts, it may be difficult to breed for a rotational crop that optimizes the soil legacies for subsequently planted crops without large trade-offs in productivity of the rotational crop.

Our study shows high agronomic potential in the exploration of genotype specific microbiome cultivation in crop plants, as well as the impacts these microbial alterations have on soil legacies. The genotypic variation in soil legacy observed in oats provides the possibility for crop rotation strategies that resolve to the level of crop genotype and even breeding programs that intentionally target plant host ability to cultivate beneficial microbial communities.

Literature Cited.

Allison, S. D., and J. B. H. Martiny. 2008. Resistance, resilience, and redundancy in microbial communities. Proceedings of the National Academy of Sciences of the United States of America **105**:11512-11519.

Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, C. A. Lozupone, P. J. Turnbaugh, N. Fierer, and R. Knight. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences **108**:4516-4522.

Gopal, M., & Gupta, A. (2016). Microbiome selection could spur next-generation plant breeding strategies. *Frontiers in Microbiology*, *7*(DEC), 1–10. https://doi.org/10.3389/fmicb.2016.01971

Hoeksema, J. D., V. B. Chaudhary, C. A. Gehring, N. C. Johnson, J. Karst, R. T. Koide, A. Pringle, C. Zabinski, J. D. Bever, J. C. Moore, G. W. T. Wilson, J. N. Klironomos, and J. Umbanhowar. 2010. A meta-analysis of context-dependency in plant response to inoculation with mycorrhizal fungi. Ecology Letters **13**:394-407.

Johnson, N. C., G. W. T. Wilson, J. A. Wilson, R. M. Miller, and M. A. Bowker. 2015. Mycorrhizal phenotypes and the Law of the Minimum. New Phytologist **205**:1473-1484.

Klironomos, J. N. 2002. Feedback with soil biota contributes to plant rarity and invasiveness in communities. Nature **417**:67-70.

Kulmatiski, A., K. H. Beard, J. R. Stevens, and S. M. Cobbold. 2008. Plant-soil feedbacks: a meta-analytical review. Ecology Letters **11**:980-992.

Lankau, R. A., and D. P. Keymer. 2016. Ectomycorrhizal richness declines toward the host species' range edge Molecular Ecology **25**:3224-2341.

Loreau, M. 2000. Biodiversity and ecosystem functioning: recent theoretical advances. Oikos **91**:3-17.

Mariotte, P., Z. Mehrabi, T. M. Bezemer, G. B. De Deyn, A. Kulmatiski, B. Drigo, G. F. Veen, M. G. A. van der Heijden, and P. Kardol. 2018. Plant-Soil Feedback: Bridging Natural and Agricultural Sciences. Trends in Ecology & Evolution **33**:129-142.

Martin-Robles, N., A. Lehmann, E. Seco, R. Aroca, M. C. Rillig, and R. Milla. 2018. Impacts of domestication on the arbuscular mycorrhizal symbiosis of 27 crop species. New Phytologist **218**:322-334.

Miki, T., T. Yokokawa, and K. Matsui. 2014. Biodiversity and multifunctionality in a microbial community: a novel theoretical approach to quantify functional redundancy. Proceedings of the Royal Society B-Biological Sciences **281**.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2015. vegan: Community Ecology Package.

Perez-Jaramillo, J. E., R. Mendes, and J. M. Raaijmakers. 2016. Impact of plant domestication on rhizosphere microbiome assembly and functions. Plant Molecular Biology **90**:635-644.

Porter, S. S., and J. L. Sachs. 2020. Agriculture and the Disruption of Plant-Microbial Symbiosis. Trends in Ecology & Evolution **35**:426-439.

Saleem et al. 2019. Annual review of Ecology, Evolution, and Systematics. "More than the sum of its parts: microbiome diversity as a driver of plant growth and soil health"

Sawers, R. J. H., M. N. Gebreselassie, D. P. Janos, and U. Paszkowski. 2010. Characterizing variation in mycorrhiza effect among diverse plant varieties. Theoretical and Applied Genetics **120**:1029-1039.

S. R. Searle, F. M. Speed & G. A. Milliken (1980) Population Marginal Means in the Linear Model: An Alternative to Least Squares Means, The American Statistician, 34:4, 216-221, DOI: 10.1080/00031305.1980.10483031

Tilman, D. 1999. The ecological consequences of changes in biodiversity: A search for general principles. Ecology **80**:1455-1474.

Toju, H., A. S. Tanabe, S. Yamamoto, and H. Sato. 2012. High-Coverage ITS Primers for the DNA-Based Identification of Ascomycetes and Basidiomycetes in Environmental Samples. Plos One **7**.

White, T. J., T. Bruns, S. Lee, and J. W. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pages 315-325 *in* M. A. Innis, D. H. Gelfand, J. J. Snisky, and T. J. White, editors. PCR Protocols: A Guide to Methods and Applications. Academic Press, Inc., New York.

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Table 1. Greenhouse Experiment Inoculation Scheme

	Micro A	Micro B	Micro C	Micro D	Micro E	Micro F	Micro G	Micro H
Oat A		AxB	AxC	AxD				
Oat B			BxC	BxD	BxE			
Oat C				CxD	CxE	CxF		
Oat D					DxE	DxF	DxG	
Oat E						ExF	ExG	ExH
Oat F	FxA						FxG	FxH
Oat G	GxA	GxB						GxH
Oat H	HxA	HxB	HxC					
Corn	CornxA	CornxB	CornxC	CornxD	CornxD	CornxF	CornxG	CornxH

Eight oat genotypes (represented here as A-H) and one corn genotype were grown in planter pots in a controlled greenhouse setting. Pots were filled with 95% autoclaved background soil and 5% living soil inoculum sourced from plots of eight different oat genotypes (A-H) replicated across three spatial blocks of a field experiment conducted at West Madison Agricultural Research Station. In the greenhouse each oat genotype planted was inoculated with three different genotype microbial communities (with no overlap between planted community and inoculating community). The one genotype of corn planted was inoculated with all eight oat genotype microbial communities. Each planted genotype x inoculating genotype treatment was replicated three times with each replicate inoculum being sourced from one of the three replicated plots in the original field study.

Table 2. Fungal Community Trait associations with oat conditioning genotype

Fungal Community Trait	**% Microbiome Genotype**	**Fixed Effect p-value**
Fungal Species Diversity*	42.59866623	0.0248
Arbuscular Mycorrhizal Fungi Abundance*	41.39840066	0.02819
Trichoderma Abundance	22.44505291	0.1423
Fusarium Abundance*	29.81136443	0.08223
Gibberella Abundance*	39.06934049	0.03579
Alternaria Abundance	25.23113905	0.1169
Solicoccozyma Abundance	17.2252305	0.1992
Acremonium Abundance	16.17660572	0.212
Lipomyces Abundance	21.06	0.1562
Metarhizium Abundance	20.24	0.1647
Colletotrichum Abundance*	27.38	0.09959

Results of random or fixed effects linear models on select aspects of fungal communities. “% Microbiome Genotype” shows the estimated variance component associated with the random effect of oat microbiome conditioning genotype in a random effects model (Variance of genotype effect divided by total variance). “Fixed effect P-value” = P value for the effect of microbiome conditioning oat genotype when modeled as a fixed effect in a linear model.

Table 3. Bacterial community trait associations with oat conditioning genotype

Bacterial Community Trait	**% Microbiome Genotype**	**Fixed Effect p-value**
Bacterial Species Diversity	25.0519442	0.1346
Arthrobacter Abundance	16.10608293	0.2374
Gemmata Abundance	2.509819778	0.4602
Nitrospira Abundance	14.01	0.229
Bacillus Abundance	28.02	0.1089
Methylobacterium Abundance	12.82	0.2827

Results of random or fixed effects linear models on select aspects of bacterial communities. "% Microbiome Genotype" shows the estimated variance component associated with the random effect of oat microbiome conditioning genotype in a random effects model (Variance of genotype effect divided by total variance). "Fixed effect P-value" = P value for the effect of microbiome conditioning oat genotype when modeled as a fixed effect in a linear model.

Table 4. Results of random and fixed effects models of oat phenotypes

Oats

Trait	**Test**	**High Nutrients**	**Low Nutrients**
Above Ground Biomass (g)	% Host Genotype	9.461	0.1888
	% Microbiome Genotype	21.91	11.81
	Host Genotype p-value	0.0004197	0.1188
	Microbiome Genotype p-value	0.1027	0.147
Below Ground Biomass (g)	% Host Genotype	13.43	5.216
	% Microbiome Genotype	3.242	21.77
	Host Genotype p-value	0.02554	0.06129
	Microbiome Genotype p-value	0.3499	0.002484
Total Biomass (g)	% Host Genotype	13.05	2.895
	% Microbiome Genotype	11.63	13.80
	Host Genotype p-value	0.002477	0.05532
	Microbiome Genotype p-value	0.2677	0.07592
Above/Below Ground Biomass (g)	% Host Genotype	3.03	6.23
	% Microbiome Genotype	1.33	21.68
	Host Genotype p-value	0.2733	0.2778
	Microbiome Genotype p-value	0.5598	0.0003035
Grain Yield (g)	% Host Genotype	5.497	0.0
	% Microbiome Genotype	19.80	6.623
	Host Genotype p-value	0.009161	0.4006
	Microbiome Genotype p-value	0.07744	0.1263
Grain Density (g/ml)	% Host Genotype	27.04	2.727
	% Microbiome Genotype	<0.0001	2.355
	Host Genotype p-value	0.0008815	0.2329
	Microbiome Genotype p-value	0.6852	0.03086

Results of linear random or fixed effects models on oat phenotypes. Each model included the genotype of the oat plant (host genotype) as well as the genotype of the oat that conditioned the microbial inocula (Microbiome genotype). For each oat phenotype, models were run with both terms as random effects, and the variance component presented (% Host Genotype, % Microbiome Genotype). A second model was run with the two terms modeled as fixed effects, and the P value reported (Host Genotype p-value, Microbiome Genotype p-value). Separate models were run for the high and low nutrient treatments.

Table 5. Results of random and fixed effects models of corn phenotypes

Corn

Trait	**Test**	**High Nutrients**	**Low Nutrients**
Above Ground Biomass (g)	% Microbiome Genotype	26.27	0
	p-value	0.01527	0.4822
Below Ground Biomass (g)	% Microbiome Genotype	25.19	0
	p-value	0.09674	0.5781
Total Biomass (g)	% Microbiome Genotype	25.013	0
	p-value	0.01826	0.4960
Above/Below Ground Biomass (g)	% Microbiome Genotype	10.44	14.087
	p-value	0.1793	0.2549

Results of linear random or fixed effects models on corn phenotypes. Each model included the genotype of the oat that conditioned the microbial inocula (Microbiome genotype). For each oat phenotype, models were the term as a random effect, and the variance component presented (% Microbiome Genotype). A second model was run with the term modeled as a fixed effect, and the P value reported (Microbiome Genotype p-value). Separate models were run for the high and low nutrient treatments.

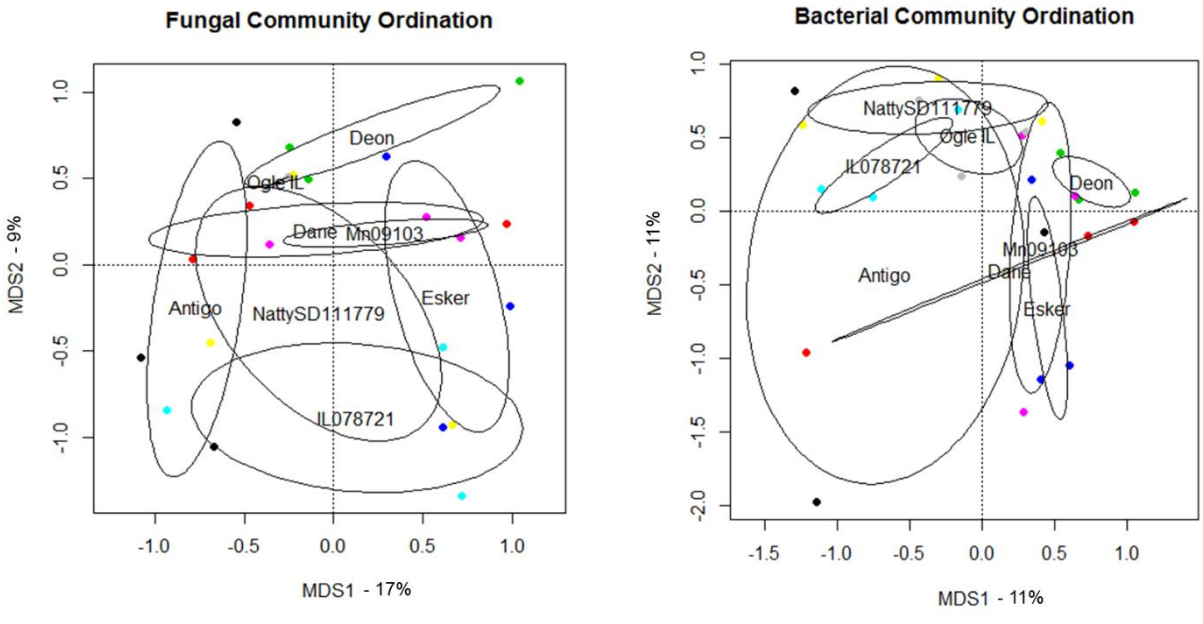

Figure 1. PCoA plots depicting clustering of fungal and bacterial communities by soil cultivating oat genotype. Soil was collected from field plots of eight different oat genotypes replicated across three spatial blocks in a field experiment conducted at West Madison Agricultural Research Station. Labels show centroids for the three replicate communities of each oat genotype.

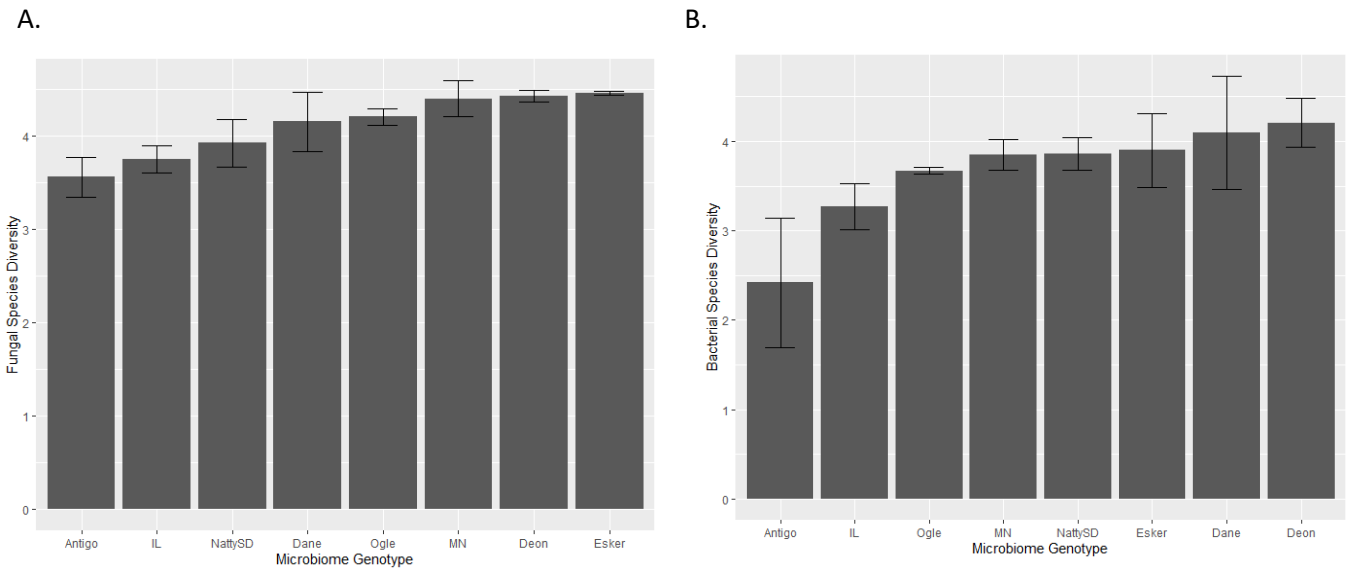

Figure 2. Shannon-Weaner diversity of fungal (A) and bacterial (B) communities in soil conditioned by eight different oat genotypes in the field (Microbiome Genotype). Each bar shows the mean and standard deviation of three replicated field plots per genotype.

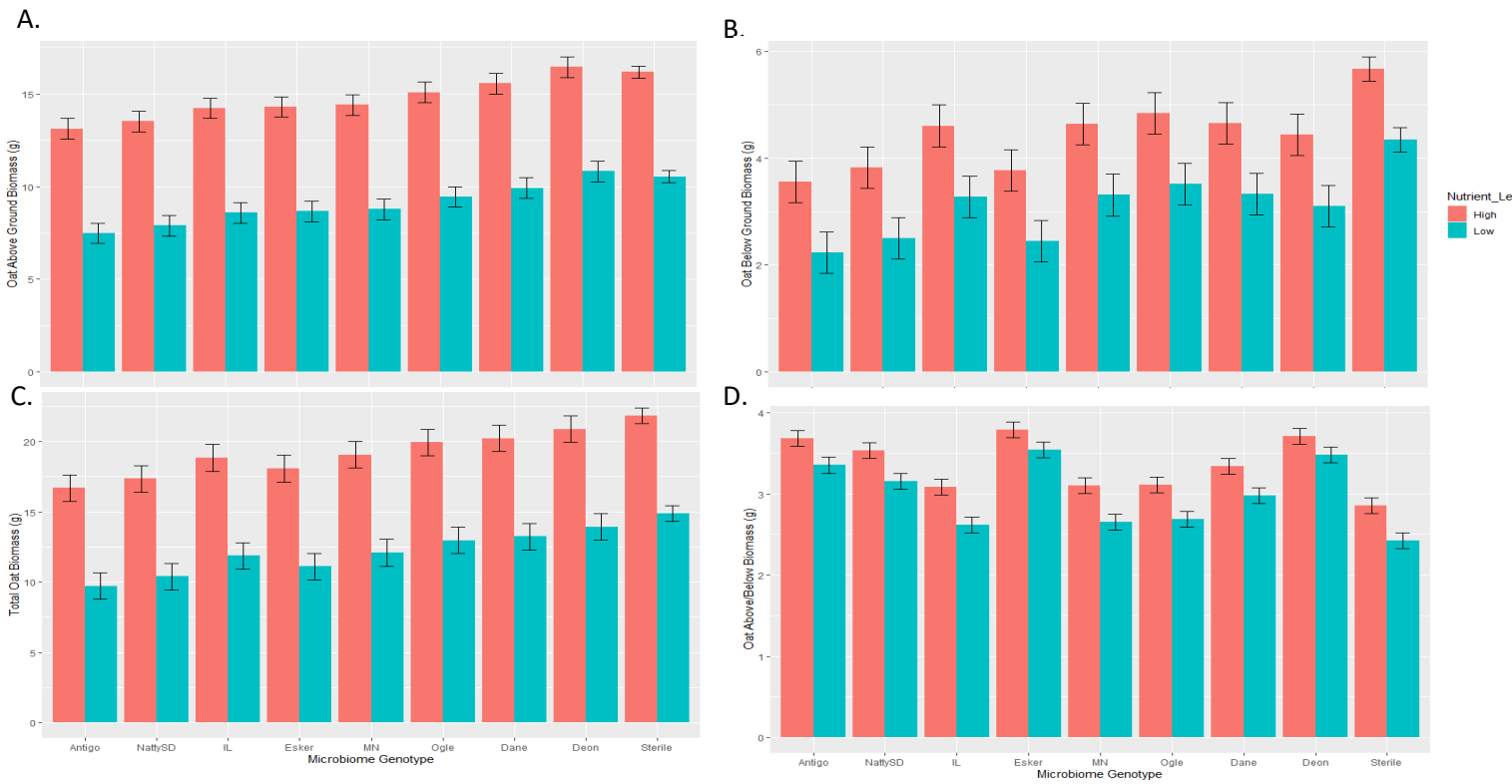

Figure 3. Mean (+ Standard Error) of oat biomass in the greenhouse experiment by the microbiome conditioning oat genotype, for high (red) and low (blue) nutrient conditions. Means are least square means averaging over the effect of host oat genotype.

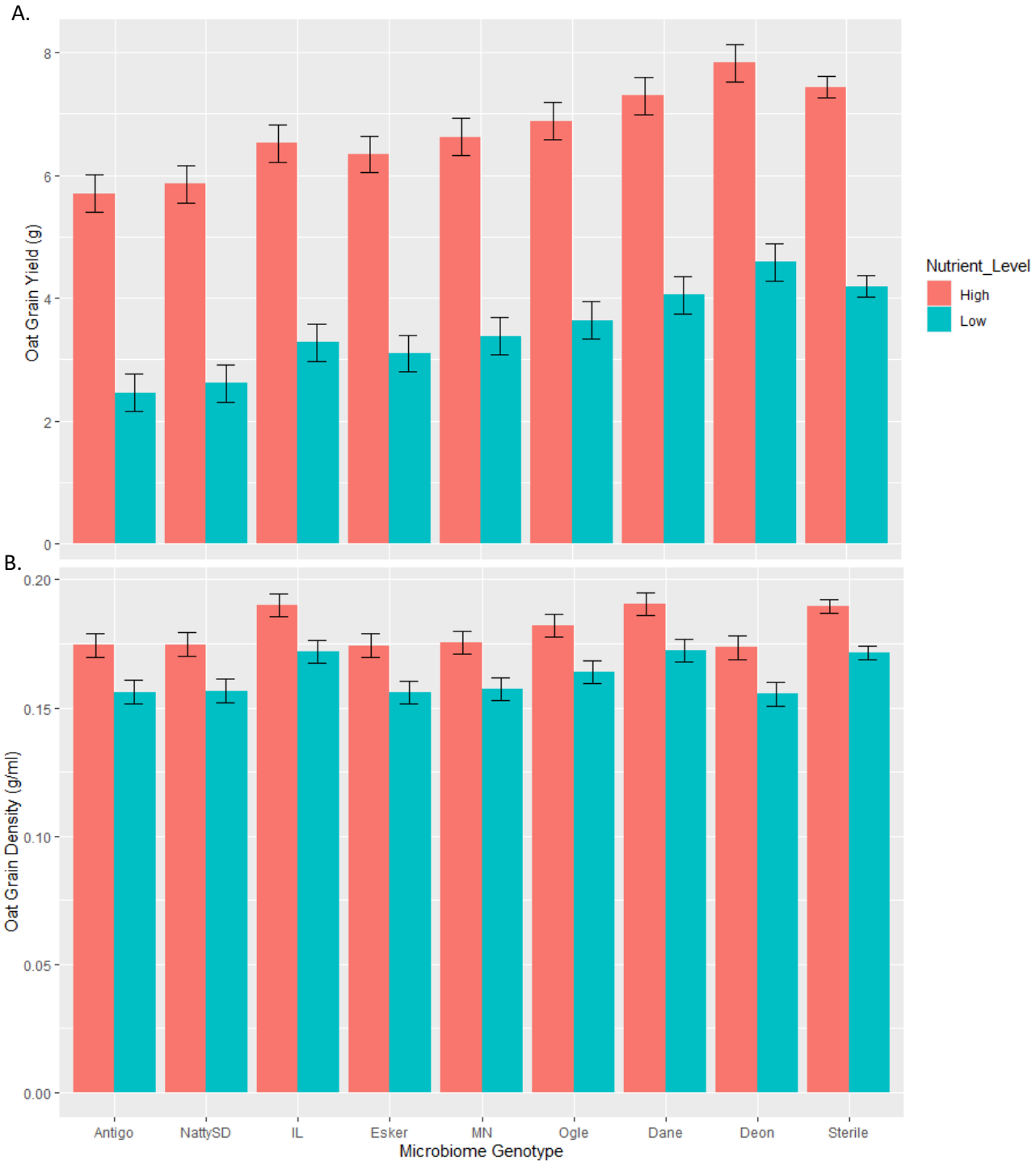

Figure 4. Mean (+ Standard Error) of oat grain yield and density in the greenhouse experiment by the microbiome conditioning oat genotype, for high (red) and low (blue) nutrient conditions. Means are least square means averaging over the effect of host oat genotype.

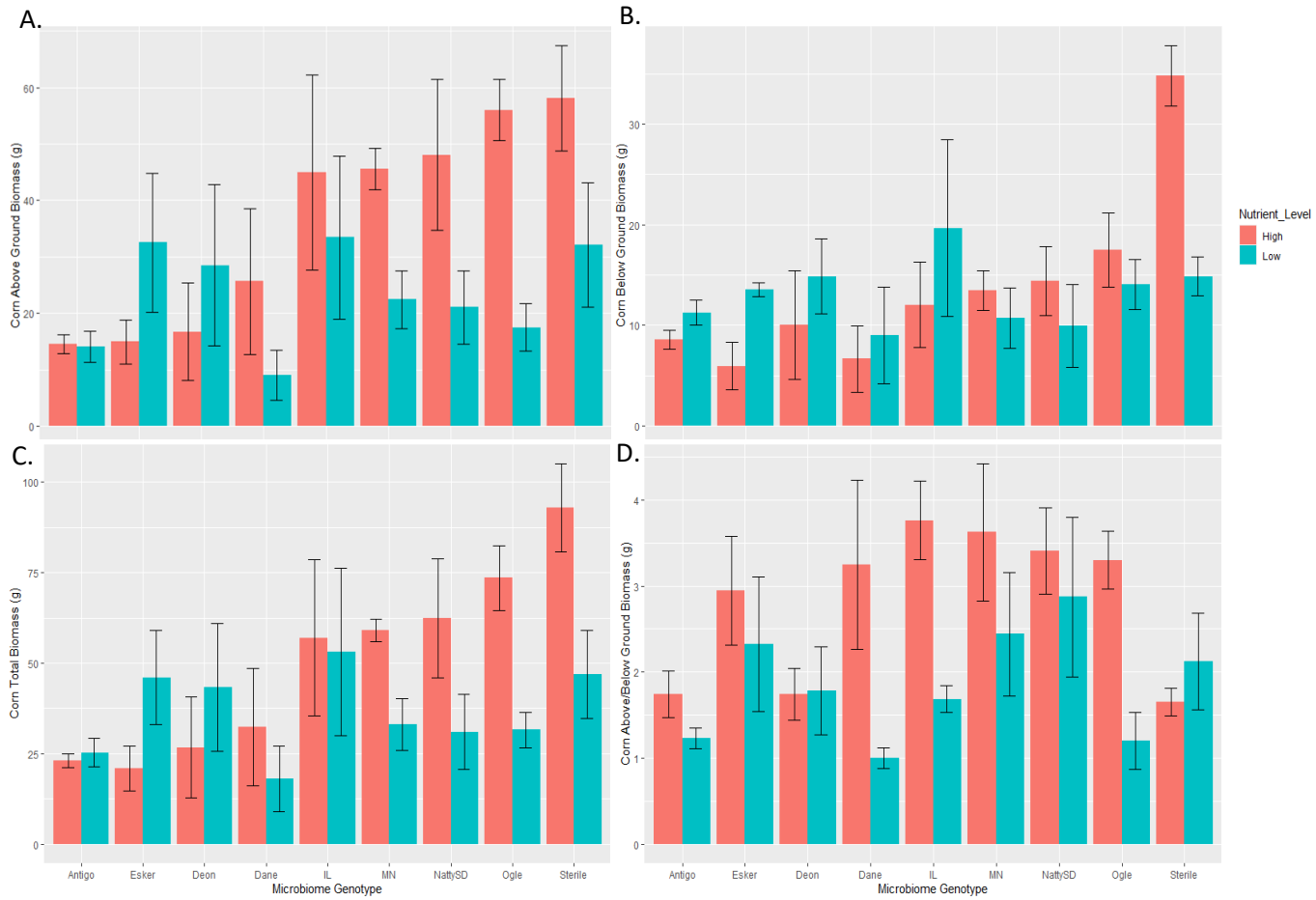

Figure 5. Mean (+ Standard Error) of corn biomass in the greenhouse experiment by the microbiome conditioning oat genotype, for high (red) and low (blue) nutrient conditions.

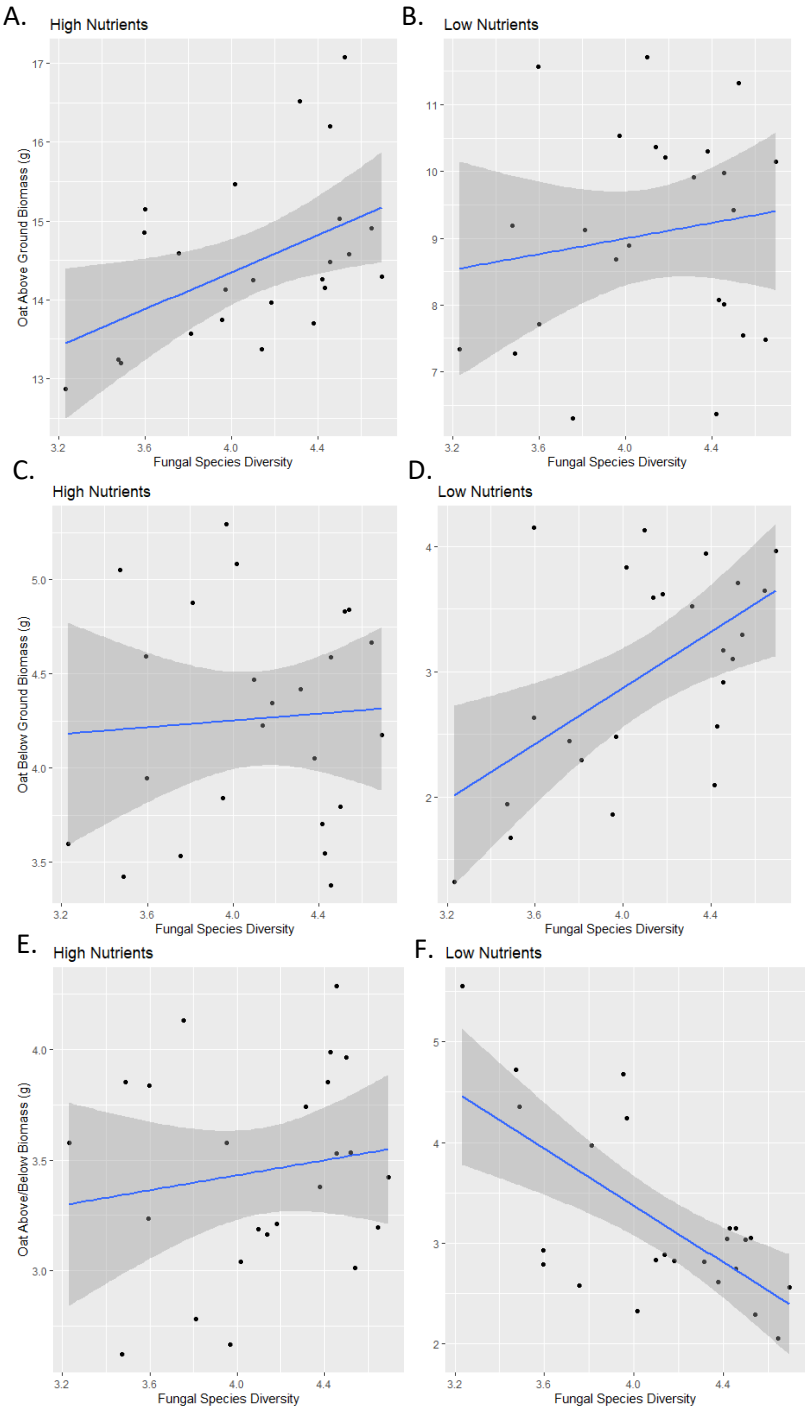

Figure 6. Oat biomass in the greenhouse experiment versus the Shannon-Weiner diversity of the fungal community in the soil inoculum. A), B) Aboveground biomass; C), D) Belowground biomass; E),F) Shoot:root ratio. A),C), E), High nutrient treatment, B), D), F), low nutrient treatment.

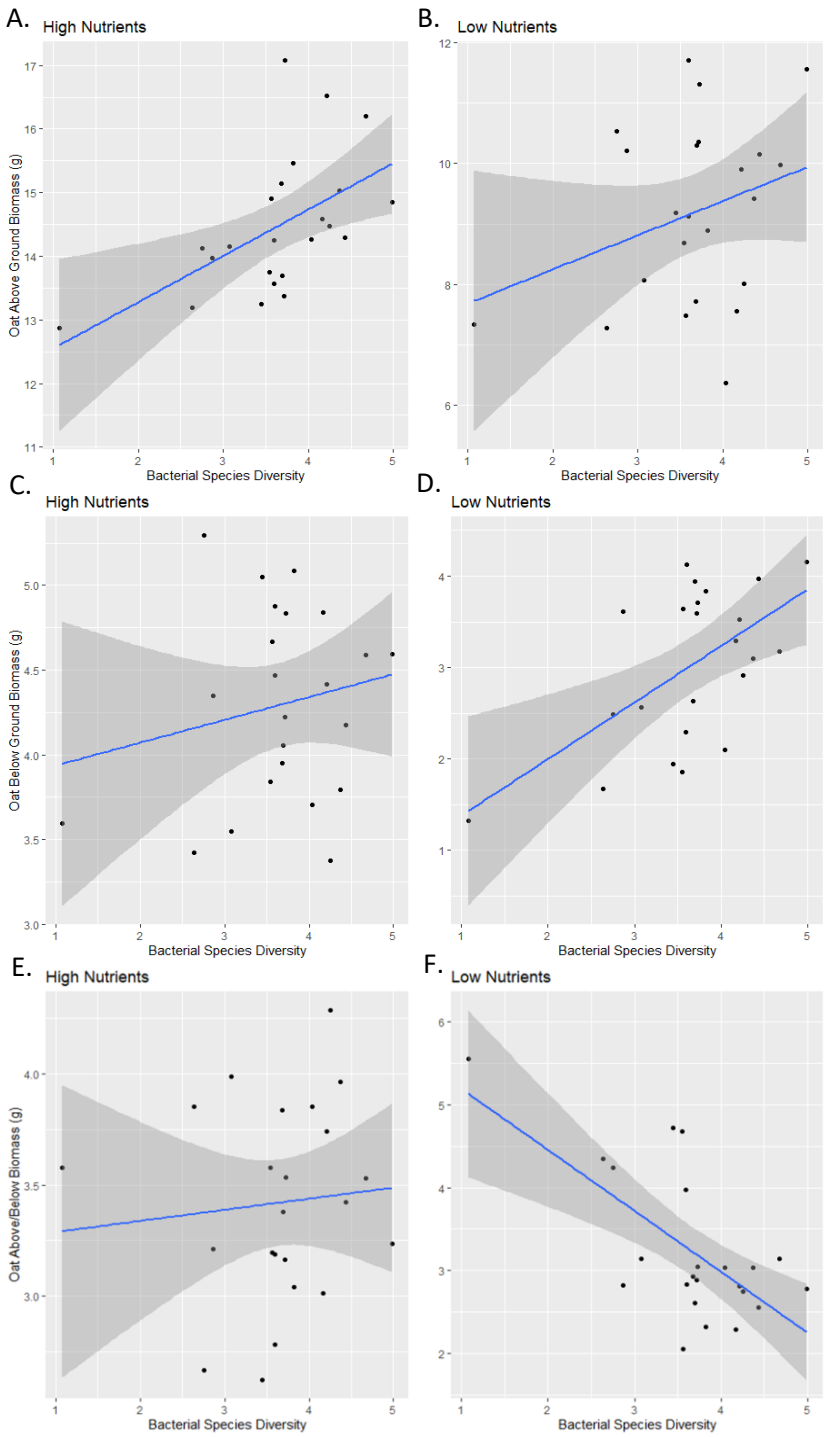

Figure 7. Oat biomass in the greenhouse experiment versus the Shannon-Weiner diversity of the bacterial community in the soil inoculum. A), B) Aboveground biomass; C), D) Belowground biomass; E),F) Shoot:root ratio. A),C), E), High nutrient treatment, B), D), F), low nutrient treatment.

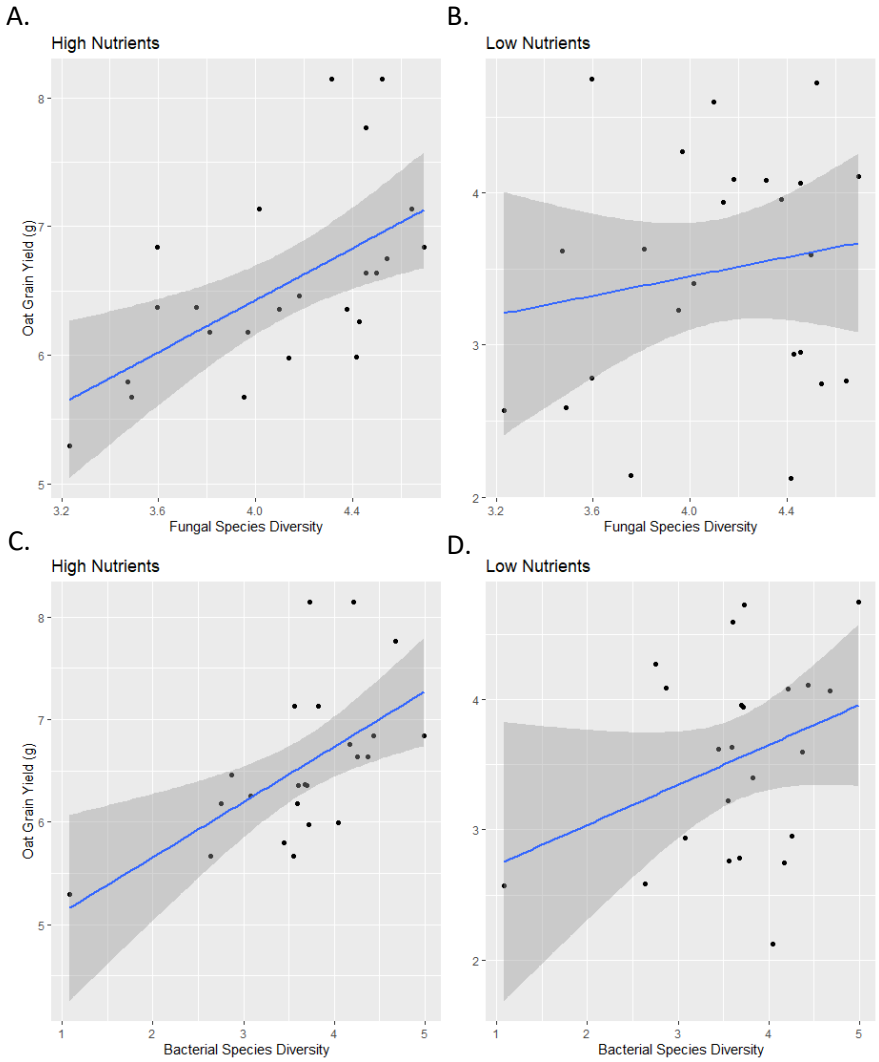

Figure 8. Oat grain yield in the greenhouse experiment versus the Shannon-Weiner diversity of the fungal (A, B) and bacterial (C,D) community in the soil inoculum. A),C) High nutrient treatment, B), D) low nutrient treatment.

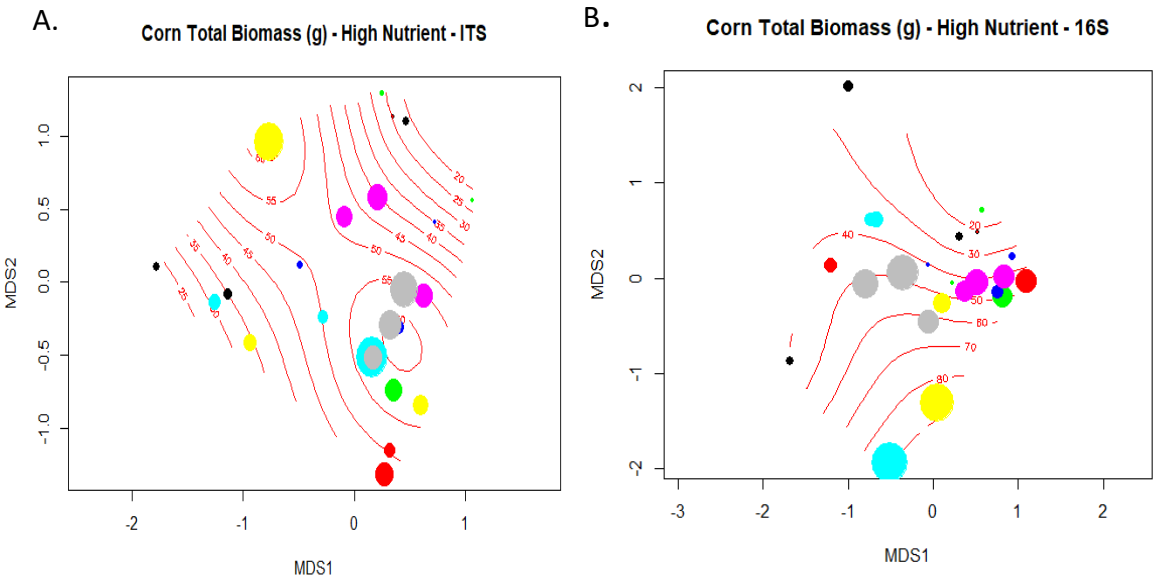

Figure 9. Principle Coordinates analysis (PCoA) plots of fungal communities (A) and bacterial communities (B) of the initial soil inocula. Symbol size reflects the total biomass of corn plants inoculated with each community in the high nutrient treatment. The surface represents the predicted values of a generalized additive model fitting the corn biomass as a non-linear function of the PCoA axes scores. For A), the GAM had a deviance explained of 49.1% and an overall model P=0.016. For B) the GAM had a deviance explained of 28.5% and an overall model P=0.0638.

Chapter 3

Using LASSO regression models to evaluate the effect of oat rhizosphere fungal community structure on oats and next-season corn

Abstract

As our ability to collect microbial community data improves, our ability to reliably analyze these data runs the risk of lagging behind. When correlating microbial populations to important agronomic outcomes, over-testing can result in false positives which have the potential to waste resources. Unfortunately, strategies that correct for multiple-testing are often quite stringent and can lead to missed leads in the form of false negatives. In this study we use LASSO (least absolute shrinkage and selection operator) regression as a tool to test for correlations between hundreds of fungal taxa and agronomic outcomes while mitigating many of the downsides associated with microbial community data analysis. We extracted fungal rhizosphere communities from a GWAS (genome-wide association study) containing 312 oat genotypes replicated across two locations. Additionally, one season after the oat GWAS was harvested, we planted corn into the fields and tracked the growth phenotypes of corn plants within the confines of the plots in which each oat genotype had been grown. By filtering the 1,239 unique fungal genera present in this data set based on relative abundance and frequency of detection and then applying LASSO regression we were able to create models that used fungal community structure to successfully predict both oat and next-season corn outcomes in a naïve data set at both the genotype and phenotype level. We were also able to highlight certain fungal taxa within these models as candidate associates that may be important to the productivity of oat and corn crops and merit further investigation.

Introduction

Due to their sessile nature, plants have been forced to develop survival strategies that enable them to overcome a variety of potential stresses. A plant seed must make do with the environment in which it is deposited, regardless of how favorable or unfavorable that environment may be. For this reason, plants have been exposed to various selective pressures during their evolutionary timeline, selecting for a diversified functional portfolio to increase their fitness in a variety of environmental contexts. But since no single organism can optimally perform every function needed to maintain fitness across multiple contexts, plants have been selected to depend on the surrounding microbiota for many of their functional capabilities.

Due to all of the known benefits the plant associated microbial community (i.e. plant microbiome) can confer to a host (Bender et al., 2016), agronomists are increasingly interested in leveraging these communities to increase plant productivity in the field (Schlaeppi and Bulgarelli 2015). Next-generation sequencing technologies (NGS) have revolutionized our ability to identify the taxonomic make-up of microbial communities across a huge range of environments. Using this technology, recent studies have suggested the existence of consistent plant-host associated microbial communities made up of microbial taxa that occur at high frequency in samples of a given host. This is often referred to as a "core" microbiome (Simonin et al. 2020, Singh et al. 2020, Singh and Goodwin 2022). Other studies have found that differences in host associated microbial community structure are resolved to the level of host genotype (Gopal and Gupta 2016). Studies like this demonstrate the usefulness of NGS and offer new ways to explore the interactions between agricultural plants and their associated microbial communities. Exploring microbial symbionts that consistently associate with crop phenotype can offer us insight into the types of communities that make for healthy soils. Additionally, exploring

microbial symbionts that associate with crop genotypes, even when averaged across environmental conditions, can highlight microbial relationships potentially responsive to crop breeding, offering new ways to improve crop success.

Unfortunately, the ability to gather microbial community data has outstripped the ability to strategically manipulate these communities in ways that lead to predictable, beneficial outcomes. This has led to an emphasis in microbiome research on documenting how microbial community structure, and/or individual microbial taxa, respond to environmental gradients, land management, etc. (Mendes et al. 2011, Edwards et al. 2015). However, in order for microbiome research to produce translational, applied value it is necessary to determine if, and how, these widely documented changes in microbial community structure translate into outcomes of interest (Bakker et al. 2012, Berendsen et al. 2012, Xu and Coleman-Derr 2019). For agronomists, the most important "outcome" for plant-associated microbiomes will often be crop yield.

Here, we investigated how the structure of oat root-associated fungal communities (i.e. mycobiome [the relative abundance of individual fungal taxa across oat plants]) affects the yield of oat plants as well as subsequently planted corn plants. The most direct way to establish this relationship would be to experimentally exclude individual fungal taxa from a plot and observe any resulting change in plant traits; unfortunately, we do not have the means to independently manipulate individual fungal taxa in soil under natural conditions. Instead, we rely on the alternative logic of causal reasoning (Pearl 2010)- if an outcome can be consistently *predicted* by the variation in fungal community structure, then this provides evidence that this variation may be causally related to the outcome. We can then take advantage of the design of our field study to further decompose the predictive ability of fungal communities into direct (more likely causal) and indirect (more likely non-causal) components.

As we mine large community sequence data sets to investigate functional consequences, we are often forced to make trade-offs. One approach has been to focus on populations of known associates such as pathogens or mycorrhizal mutualists which can exclude plant-microbe relationships that have yet to be explored. Alternatively, researchers could test for associations with all detected taxa, but risk mistaking noise (false positives) for signal when performing hundreds to thousands of statistical tests. Multiple testing corrections can help offset the risk of false-positives but can become excessively conservative with high numbers of tests (Benjamini and Hochberg 1995). Moreover, multiple testing corrections may fail to account for confounding and co-linear relationships among microbial taxa. In any case, microbial ecologists must balance leaving unexplored information behind with the possibility of wasting time on false leads.

In this study we showcase Least Absolute Shrinkage and Selection Operator (LASSO) regression, a statistical tool that offers another way to explore datasets derived from community sequencing. LASSO is a regularization technique that performs simultaneous model selection and parameter estimation, in which it reduces model complexity by applying coefficient shrinkage and dropping coefficients from the model if their magnitude reaches zero (Tibshirani 1996). We applied this tool to a rhizosphere fungal community dataset derived from a genome-wide association study (GWAS) of 312 oat genotypes replicated three times at each of two locations. The season following oat harvest, an elite variety of corn was planted over the previous oat plots to determine whether oat genotype specific oat soil legacies affect the productivity of next-season corn. By using LASSO regression, we were able to test the association of many fungal taxa to oat and corn phenotypes while balancing the goals of producing a highly predictive, but not overfit model. Using this approach allows us to achieve two goals simultaneously: 1) to determine how well variation in fungal community composition

can predict crop outcomes, as an indirect test of whether this commonly observed variation has functional consequences, and 2) provide information on individual fungal taxa of high consequence for crop outcomes.

Furthermore, we used this approach to determine relationships between root-associated fungal communities at two biological scales (phenotypic and genotypic) and two temporal scales (within season and across season). When testing how variation in fungal communities across individual field plots associates with crop yields in those plots, we ignore the factors that are creating the variation in fungal community structure. This is potentially problematic, as this could lead to statistical associations between fungal abundance and crop yields that are not causal, but instead reflect shared responses to an unmeasured environmental gradient. For example, low elevation field plots with higher soil moisture may have both higher abundance of a certain fungal taxa and higher crop yields, but this is not because the fungal taxa has a beneficial relationship with the crop plant. To address this possibility, we took two approaches. First, we compared the predictive ability of models that did or did not include information on spatial gradients in the field site along with the fungal community structure. If the predictive ability of the fungal community can be effectively replaced with spatial gradients, this suggests (but does not prove) that the prediction may occur because fungal communities are sensitive indicators of soil conditions, rather than directly causing crop yields to increase or decrease. Secondly, we took advantage of the replicated oat genotypes to test the predictive ability of models based on genotypic means of fungal taxa, averaged across multiple field plots. Here, we explicitly restrict our models to use only the variation in fungal community structure that derives from oat genotypic variation. By doing so, we eliminate the possibility of shared environmental

factors creating confounding, non-causal relationships between fungal taxa and crop yields (Rausher 1992)

We tested the ability of variation in root-associated fungal communities on oat roots to predict both oat yields within the same season, as well as corn yields in the following season. We focus on fungal communities because of the many known beneficial (mycorrhizal, endophytic) and detrimental (pathogenic and parasitic) relationships between fungal taxa and host plants. Within a season, the fungal communities on oat roots could predict the yield of those oat plants via myriad positive and negative ways, including direct pathogenesis, suppression of pathogenesis, nutrient acquisition, and manipulation of plant hormones. For fungal communities on oat roots to predict corn yields in the following season, the fungal taxa must be involved in a soil legacy effect in which the microbial community in a soil is shaped by one plant species in such a way as to affect the growth of subsequent plants in that soil (Kulmatiski et al. 2008, Mariotte et al. 2018). Thus, by building cross-season (and cross-crop species) models, we aim to investigate the underlying drivers of the soil legacy effects previously documented in this system (see Chapter 1).

The goals of this study were (i) to determine if a statistical model correlating relative abundance of fungal taxa to oat and corn traits could successfully predict these traits in a test data set, (ii) to determine the effectiveness of these models at both the oat genotype and plant phenotype (plot) level and, (iii) to identify candidate fungal taxa with strong beneficial or detrimental effects on oats and next-season corn. These goals are largely hypothesis generating; However, we did hypothesize that (i) models based on oat rhizosphere fungal community structure would more accurately and fully predict oat phenotypes than next-season corn

phenotypes and (ii) models would predict oat traits at both the genotype and phenotype (plot) level but would only predict corn traits at the phenotype level.

Materials and Methods

Field experimental design and sampling: This research was conducted through the use of a genome wide association study (GWAS). In the Spring of 2019, 312 unique oat genotypes were planted at each of two locations: The West Madison Agricultural Research Station (WMAD) and the Arlington Agricultural Research Station (ARL). The oat genotypes were grown in 1.52 x 3.04 m plots with a 1.5m spacer between them and a boundary row of oats around the edges of the field. Each genotype was grown in at least three replicated plots, with eight genotypes being grown in six replicated plots for a total of 960 plots at each location (1,920 plots total). The two field locations were organized in a randomized block design. Each location was organized into three randomized blocks, each containing one of the replicated field plots from each oat genotype. Each field location consisted of 30 columns and 32 rows (10 columns per replicated block). The oat fields did not receive any nutrient, water, or fungicide inputs throughout the season. Each location was treated once with an herbicide to mitigate competition from weeds

Oat root core samples were collected from each of the 960 plots within one week after the peak grain heading day at each location (>50% of oat plants heading). Root cores were stored at 4ºC for a maximum of 72 hours before processing. Root tissue was manually collected from each core by shaking off any excess soil before storing in 96 deep-well plates at -80ºC to await subsequent DNA extraction and sequencing. The oat plots were harvested in the Fall of 2019 and the plot weight, plot density, and test weight (a grain density measurement [Kg/hL]) of each genotype was recorded.

In the spring of 2019 an elite variety of corn (*Zea mays*) was freely seeded into the oat plots to observe potential oat genotype specific soil legacy effects. The fields at each location were measured and plotted out to identify the center of each unique oat genotype plot from the previous season. To understand the effect of microbial community throughout the corn lifecycle, corn plants were sampled from each location at the V6 and R1 stages as well as at harvest. At the V6 and R1 sampling periods, one meter of corn was hand cut from a row at the middle-most point in each oat plot marked from the previous season (Fig. 1). The plants from each unique plot were bundled together, labeled, and placed in industrial drying ovens at 65ºC for 14 days. After drying, the plants were counted and weighed to collect density and biomass data from each timepoint. At harvest, one meter of plants from the middle of each previous oat plot were counted before the ears were removed from each plant, shucked, and placed in a labeled canvas bag before drying at 65ºC for two weeks. After this drying period, the cobs from each plot were counted and threshed. The threshed grain was tested for percent moisture and then weighed. To summarize, the resulting phenotypic data from corn were V6 dry weight, R1 dry weight, threshed grain dry weight, and number of plants sampled at each stage. Biomass at each stage was divided by the plant number to get average biomass at each growth stage. Plant number from each stage was added together to achieve a proxy for within plot plant density, and this density was multiplied by the average grain yield per corn plant to achieve an extrapolated yield for the whole plot. The corn fields did not receive any fertilizer, water, or pesticide inputs throughout the growing season.

Amplicon preparation and sequencing: DNA from the 1,920 oat rhizosphere samples were extracted using Promega E-Z Plant DNA extraction kits (Promega, Madison, WI) according to the manufacturer's instructions. To characterize the fungal community, the fungal ITS2 region

was amplified in each sample using the ITS1f forward primer(5'-CTTGGTCATTTAGAGGAAGTAA , (Gardes and Bruns, 1993)) and ITS4 reverse primer (5'-TCCTCCGCTTATTGATATGC, (White et al., 1990)). External fusion PCR primers contained a 14-bp overlap to the trailing end of internal primers with 12bp i7 index and P7 flow cell adaptor or an i5 index, 7-bp spacer and P5 adapter (Lankau and Keymer 2016). Amplicon library preparation was completed in two rounds of PCR. The first round of PCR amplified the fungal ITS2 region as well as the Nextera read primers. PCR was performed in 10 µl reactions using 0.2 µL of a hot-start, high fidelity polymerase (Clonetech Prime Star GLX, Fitchburg, WI) with 2 µL of its 5X buffer, 0.8 µL dNTPs (at 10 nM concentration), 0.25 µL of each primer (at 10 nM), 0.7 µg T4 gene 32 protein, and 10 ng of template DNA. The thermocycling program for the ITS2 region was a 5-minute hot start at 98˚C, 35 cycles of denaturing (98˚C, 0:30), annealing (50˚C, 0:45), and extension (68˚C, 1:00) and a final extension of 15 minutes at 68˚C. The second round of PCR added the P5 and P7 flowcell adapters to prepare the library for sequencing on an Illumina MiSeq, along with an external set of sample barcodes located between the flowcell adaptors and read primers. The amplicons were cleaned with the Omega BioTek E-Z 96 Cycle Pure kit (Omega Bio-tek, Norcross, GA). Purified products were quantified using a Qubit 2.0 fluorometer with the Qubit dsDNA HS assay and then pooled at equal concentrations (Thermo Scientific, Grand Island, NY). Amplicon products were then sequenced on Illumina Miseq using a 300 cycle Paired-End run at the University of Wisconsin-Madison Biotechnology Center. To achieve sufficient sequence depth, the experiment was split across three separate runs, with one of the three replicates from each location being sequenced on each run.

Bioinformatics: Raw external sequences were initially trimmed at both the 5' and 3' ends using Cutadapt (version 1.18). The Qiime2 (v2017.12) pipeline was used to processed trimmed reads

using DADA2. Samples were filtered and further trimmed by DADA2 using the following parameters (ITS2 sequences: p-trunc-len-f = 0, p-trunc-len-r = 280, p-max-ee = 4). We used the RDP Naïve Bayesian Classifier to assign taxonomy to fungal amplicon sequence variants (AVS) using the UNITE (version 8.0) reference database. Reads assigned to plant taxa were removed. Samples with fewer than 2500 reads were removed based on inspection of rarefaction curves. In the end, this study collected data from 1,920 samples split evenly between two locations (ARL and WMAD). Of these original samples, a total of 1,801 were retained (899 from ARL and 902 from WMAD).

Statistical analysis: First, we transformed fungal ASV sequence counts into relative abundance values using the "vegan" package in R (Oksanen et al. 2018). To create the datasets for the comparison of oat genotype associated fungal communities to host phenotypes, we collapsed the Fungal ASV's by genus and sorted them into probable lifestyle guilds using the FungalTraits database (Põlme et al., 2020). This resulted in two data frames: one table consisting of 1,239 unique fungal genera across all samples and one table with these genera collapsed into probable guilds (e.g. plant pathogen, arbuscular mycorrhizal fungi, saprotrophs, etc…). We used the relative abundances of these fungal community genera and guilds as predictors in LASSO models for oat and corn traits at the phenotype (plot) level (see below).

In order to build models at the genotypic level, we need to create a dataset of average fungal relative abundance for each oat genotype (in each location). Because field location, even within spatial blocks, affected both oat and corn traits and fungal abundance, we used the "emmeans" package in R to find the estimated mean of the relative abundance of each genus and lifestyle guild for each oat genotype (Searle et al. 1980). The estimated means were extracted from a unique linear model for each location including blocking factors (e.g. row, column, block)

to control for as much environmental variation as possible. To focus this process on the fungal taxa most likely to produce robust statistical results, we screened the fungal community for taxa with sufficient information (presence and abundance) to allow for reliable estimates of genotypic averages using relative abundance and frequency of presence within samples to screen for taxa. To do this, we selected those fungal genera that showed an average of >0.1% of the total fungal abundance within a sample and were present in >33% of samples for further analysis. This filtering step resulted in 123 unique fungal genera. We then used these estimated relative abundance means as explanatory variables using LASSO regression to determine if oat genotype associated fungal community structure could accurately predict genotype-averaged outcomes in oat and next-season corn traits.

LASSO regression was used to create optimized models relating fungal community composition to oat and corn agronomic traits. To make the results easier to interpret, we used the function “scale()” in the R package “car” to center each phenotypic response variable around the mean (Fox and Weisberg 2019). We then used the function “slice_sample” within the R package “DPLYR” to randomly create a test data set consisting of 10% of the samples (dplyr ref). This resulted in a “test” data set consisting of 10% of the samples and a “train” data set consisting of the remaining 90% of samples. This was done separately at both the genotype level where 10% of the genotypes were randomly placed into the genotype “test” data set and at the phenotype level where 10% of the oat plots were randomly placed into the phenotype “test” data set. Next, we used the function “cv.glmnet()” within the R package “glmnet” to find the best coefficient constraining lambda value (Friedman et al. 2010). The regularization function of the LASSO model depends on the value of the lambda parameter- lower lambda values result in more permissive models that retain more variables, while higher lambda values result in more

conservative models with fewer retained variables. This function permutationally determines the lambda value that results in the lowest model error using cross-validation within the training data set. The lambda value resulting in the models with the lowest average error was then used within the function "glmnet()" to create the final model of the original training data set for each plant trait at the genotype and phenotype level. To determine the predictive ability of this model with novel data, we then tested the model against the original "test" data set by correlating the predicted values from the model against the observed values from the "test" data set. We used the "cor()" function from base R to calculate an R^2 value from this correlation as a basis for the success of the model in predicting novel data. Models whose predictions described greater than 5% of the variation in the test data observations were considered successful enough to extract coefficients for use in the identification of fungal taxa of potential importance to the plant host. Visualizations of these models were created using the package "ggplot2" in R (Wickham 2016).

To test if fungal community data had predictive power for plant host traits at the phenotype (plot) level, we compared three model types; a fungal only model using only fungal community data, a spatial model based only on the within field spatial factors "row", "column", and "block", and a full model including all fungal and spatial data. By comparing the success of each of these models we can draw conclusions about the power for fungal communities to predict plant host outcomes. If the fungal only model has predictive power, then mycobiome structure could be an important determinant in plant host outcomes for that trait. If the spatial model has significant predictive power, then variation along a north-south and/or east-west gradient in the field is an important determinant in plant host outcomes for that trait (regardless of the precise biological, chemical, or physical properties involved). Comparing the full model to the other two models provides further information about the importance of the fungal or spatial

data. For example, in a case where both the fungal and spatial models are predictive, the full model helps us determine if the fungal community variation provides information on crop outcomes that is independent of the spatial gradients in the field. If the full model does not display more predictive power than the spatial model, then the predictive power of the fungal data can be absorbed into spatial variation. If the full model does provide more predictive power than the spatial model, we conclude that some fungal populations important to plant health vary independently from spatial gradation.

To identify potentially important fungal taxa at the plant phenotype level, we extracted coefficients from the full model. In this way, fungal coefficients that are only significant because their populations follow significant environmental trends are dropped out of the model. However, fungal taxa that are both important to host outcomes and vary along spatial gradations can remain in the model and provide information based on their relationship to the spatial variables. The genotype level models inherently averaged out environmental blocking variables by averaging across genotype plots replicated within three blocks at each location; therefore, we did not include spatial gradients in the LASSO model, and we interpret the retained fungal coefficients directly.

Screening for candidate fungal taxa: Like any statistical method, individual results from LASSO regression models will include a mixture of false positive and false negative events. Because fungal communities are characterized by high unevenness (few common and many rare taxa in a single community) and high sparseness (many taxa present in only a few communities), estimating the effect of rare and infrequent fungal taxa on oat or corn traits can be inherently challenging. To identify candidate fungal taxa of most interest for future experimentation we

applied a strict screening procedure to the taxa that remained as coefficients in successful LASSO models. LASSO coefficients will be most reliable and repeatable for high abundance, high frequency taxa since those taxa have sufficient information in our dataset. Therefore, we filtered by the frequency of detection within samples, highlighting taxa that occurred in more than one-third of samples. This step was unnecessary in the genotype-level models as it had occurred previously when determining which taxa to feed into the models. Next, we further filtered by the average relative abundance in the samples in which these taxa were detected. We chose the detection dependent average to avoid filtering out taxa which are rare, but still impactful to the growth of the host. We filtered out taxa that did not meet or exceed an average relative abundance of one percent of the communities in which they were detected. This threshold was chosen based on the average per sample genus richness of 95 and 97 at WMAD and ARL respectively. We consider these taxa that remained in successful LASSO models while meeting both the sample frequency and relative abundance thresholds to be candidate taxa of highest potential importance to oats and/or corn. These candidate fungal taxa were then sorted based on their coefficient estimates within the LASSO models. Finally, for this subset of taxa we performed a simple linear regression of the oat/corn outcome against the relative abundance of the taxa alone. Additionally, taxa that appear in both genotype and phenotype models or in models for both host species were considered noteworthy even if they failed to meet the previously described criteria.

Results

Predicting oat phenotypes with oat rhizosphere fungal community data

At the phenotype level: To assess correlations occurring at the phenotype level, we compared fungal, spatial, and full LASSO regression models as described in the methods. This allowed us to account for the possibility that correlations between fungal taxa and host traits are driven by their shared correlation with environmental gradients. At the WMAD location, the spatial model was predictive of oat plot yield (correlation between observed and predicted outcomes, Table 1; $p = 0.0006$). Neither the fungal nor full models were predictive of oat traits at the phenotype level and so were unable to improve upon the spatial model.

At the ARL location, models created using fungal community data significantly predicted oat plot yield, but not oat plot density or oat test weight (Table 1). At the ARL location, the fungal model was able to describe 18.0% of the oat yield variation in the training data set and 16.2% of the data in the testing data set (Fig. 2; $p < 0.001$). In comparison, the spatial model at the ARL location was able to describe only 6.7% of the variation in the training data set and 4.8% in the testing data set ($p = 0.0373$). The full model did not improve over the fungal model, describing 19.7% of the variation in oat plot weight in the training data set and 15.4% in the testing data set ($p = 0.000134$).

At the genotype level*:* To determine which, if any, of the signals evident in our phenotypic models stemmed in part from genetic variation among oat varieties, in either their recruitment or response to fungal taxa, we refit LASSO models using genotype means for fungal taxa relative abundance and oat traits. At the WMAD location, models were able to significantly predict oat test weight, but not oat yield or plot density (Table 2). For oat test weight, the best model

described 10.5% of the variation in the training data set and 13.4% of variation in the testing data set ($p = 0.0396$).

At the ARL location, models using oat rhizosphere fungal community structure were able to successfully predict oat plot yield and density. Genotype level LASSO regression models explained 22.1% of the variation in oat plot yield in the training data set and 13.5% of the variation in the testing data set (Fig. 3; $p = 0.03826$). For oat plot density (plants/meter), the best model described 6.4% of the variation in the training data set, and this rose to 14.3% in the testing data set ($p = 0.03255$). Oat test weight was not significantly described by fungal community structure at the genotype level at the ARL location. The best model described 7.4% of variation in oat test weight in the training data set, 4.7% of variation described in the test data set, and the correlation between predicted and observed values was not significant ($p = 0.2353$).

Identifying candidate fungal taxa of importance to oat agronomic outcomes

Candidate fungal taxa of potential importance to oat plot yield were identified at both the genotype and phenotype level at the ARL location. At the ARL location, 19 fungal taxa were retained in the best-fitting model at the phenotype level, along with two guild level groupings: total plant pathogen and arbuscular mycorrhizal fungi (AMF) relative abundance. Nine of these taxa were estimated to negatively affect oat yield compared to the 11 with positive estimated effects. In the genotype level model, seven fungal taxa were retained as predictors of oat genotype average yield, along with the two guilds (plant pathogens and AMF). Five of the taxa were estimated to negatively impact oat yield and four were estimated to have a positive impact (Table 3).

Of the 19 fungal taxa present in the phenotype level LASSO model, three positively associating and three negatively associating taxa met the criteria to be considered promising candidates of importance to oat hosts (Table 3). In order of most to least impactful, the negatively associating genera were Periconia, Puccinia, and Coprinopsis. This trend held true when the relative abundances of these taxa were individually tested against oat plot yield in linear regression models, with Periconia being the most significant at a p-value of $6.27x10^{-5}$ and describing 1.7% of the variation. In order of most to least impactful, the positively associating genera were Funneliformis, Heydenia, and Rhizophagus. When individually regressed against oat plot yield all three of these genera significantly correlated with the response variable and described ~1.3-1.8% of the variation in oat yield. The plant pathogen guild was the least extreme negatively estimated coefficient (-15.34) while AMF were the least extreme positively estimated coefficient (226.30); however, the magnitude of the AMF estimate was much larger than that of the plant pathogens.

Of the seven fungal taxa present in the genotype level LASSO model of oat plot weight, three negatively associating taxa met the criteria for candidacy (Table 4). In order from the most to least impactful, these genera were Puccina, Plectosphaerella, and an unidentified genus in the class Sodariomycetes. When the genotype mean abundances of these genera were regressed individually against genotype mean oat plot yield, Puccinia and Plectosphaerella each described more than 3.0% of the variation in yield, while the Sordariomycetes genus described 1.5%. The guild level coefficients also significantly described oat plot yield in individual models. Plant pathogen abundance displayed an estimate of –203.68 and described 3.0% of the variation in genotype mean yield. AMF abundance displayed an estimate of 843.88 and described 6.1% of the variation in genotype mean yield.

Notably, certain taxa appeared in both the phenotype and genotype level models. Puccinia and Funneliformis each met the criteria for candidacy at the phenotype level; however, Funneliformis narrowly missed the threshold for detection dependent average relative abundance in the genotype level model. An unidentified genus in the Rozellomycotina family failed to meet the threshold for detection dependent relative abundance in both models yet was also the most extreme negative coefficient in each model. When individually regressed against oat plot yield, this Rozellomycotina genus performed higher than any other taxa in either model, describing 5.7% of the variation in oat yield at the phenotype level and 9.8% of the variation in oat yield at the genotype level. Discrepancies in average relative abundance for the same taxon in the genotype vs. phenotype level models are present because the relative abundances used in the genotype level models are mean relative abundances estimated while accounting for spatial variation.

Predicting corn phenotypes with previous year oat rhizosphere fungal community data

At the phenotype level: Phenotype level fungal community data only showed predictive power over corn traits at the WMAD location (Table 1). The fungal model was able to describe 9.9% of variation in corn V6 stage biomass in the training data set and 3.5% of variation in the testing data set (Fig. 4; $p = 0.0791$). The spatial model improved greatly over the fungal model, describing 40.0% of variation in V6 biomass within the testing data set ($p = 3.041 \times 10^{-11}$). The full model containing both fungal community and blocking factor data displayed similar predictive power to the spatial model, describing 42.6% of the variation in V6 biomass within the testing data set ($p = 4.167 \times 10^{-12}$). No variation in R1 stage biomass was described by fungal community structure in the LASSO model, but predictive power was restored in models correlating the rhizosphere mycobiome with corn plot yield. The fungal model described 14.5%

of variation in corn plot yield in the training data set and 14.9% of the variation in the testing data set (Fig. 5; $p = 0.000173$). The spatial model was more significantly predictive of corn plot yield, describing 30.0% of the variation in the training data set and 61.5% of the variation in the testing data set ($p < 0.001$). The full model did not display any improvement over the spatial model, describing 27.2% of the variation in the training data set and 56.8% of the variation in the testing data set ($p < 0.001$).

At the genotype level: Only V6 biomass at the WMAD location was significantly predicted by a LASSO regression model using previous years fungal community data at the oat genotype level (Table 2). In the training data set, 4.6% of the variation in V6 biomass was described while 9.3% of the variation was described in the test data set. The correlation between predicted and observed values was only weakly significant ($p = 0.08904$). This predictive power was lost as the corn continued into the R1 and harvest stages. Genotype level models at WMAD only predicted 0.005% and 0.004% of the variation in the test data sets for R1 and harvest, respectively ($p = 0.7001, 0.7415$).

Identifying candidate fungal taxa of importance to corn agronomic outcomes: Candidate fungal taxa of potential importance to corn plot yield were only identified at the phenotype level, while candidates affecting V6 stage dry weight were only identified at the genotype level. At the WMAD location the phenotype level LASSO model of corn plot yield contained 31 fungal taxa, and the plant pathogen guild. Nineteen of these taxa negatively associated with corn yield, and 12 positively associated with corn plot yield (Table 5). The genotype level LASSO model of genotype mean V6 stage dry weight only contained a single fungal genus which positively associated with V6 biomass.

Of the 31 fungal taxa in the phenotype level LASSO model of corn plot yield, four met the criteria to be considered candidate taxa of importance to legacy effects on corn traits (Table 5). Two of these, Cladosporium and another unidentified genus from the order Capnodiales, negatively associated with corn yield, and two more, Clohesomyces and Mortierella, positively associated with corn yield. When individually regressed against corn plot yield, neither the Capnodiales genus nor Clohesomyces were able to describe much, if any, variation in yield. Cladosporium and Mortierella were each able to describe 1.5 and 1.2% of the variation respectively (p-values 0.000172 and 0.000736). The plant pathogen guild described about 2% of the variation in corn plot yield when individually tested ($p < 0.001$). While no taxa met the criteria for candidacy in the phenotype level LASSO model for V6 stage dry weight, the soil saprotroph guild did stand out as a potentially important positive associate. The relative abundance of soil saprotrophs, when individually tested, described 3.1% of the variation in V6 stage dry weight.

An unidentified genus from the order Sordariales was the single taxon present in the genotype level LASSO model of V6 stage dry weight. This genus positively associated with genotype mean V6 dry weight and described 4.3% of the variation in the response variable ($p < 0.001$).

Discussion

Here we show that the oat rhizosphere mycobiome composition can be used to predict end-season oat host traits as well as next season corn traits, at least under some conditions. Successfully predicting plant outcomes from fungal community composition provides evidence

for the causal importance of mycobiome structure in the absence of manipulative experiments. We have identified candidate fungal taxa of potential importance to oat and corn hosts at both the oat genotype and phenotype level, suggesting potentially fruitful lines of future inquiry. The LASSO regression approach allows us to test the breadth of a microbial community data set with mitigated concern for covariance or overfitting a model.

These models show the potential for using fungal community structure at the genotype and/or phenotype level to predict crop outcomes. Models at both the phenotype and genotype levels predictive of oat yield, whereas only phenotype level models were successful in predicting next-season outcomes in corn yield. This is in agreement with our original hypothesis; however, it is interesting that neither location's mycobiome was predictive of both oat and corn traits at either the genotype or phenotype level. Given the known variability of microbial communities both between and within locations (Edwards et al. 2015), the location effect observed in this study is unsurprising, but understanding what environmental factors contribute to increasing/decreasing the impact of the microbial community on host plants will be important as we strive to manipulate the microbial community for agronomic gain (Berendsen et al. 2012, Schlaeppi and Bulgarelli 2015).

Can we accurately predict oat yield and quality from rhizosphere fungal communities?

Phenotype and genotype level models showed a location dependent ability to predict oat agronomic outcomes using mycobiome data. ARL based fungal models were predictive of oat yield and test weight whereas WMAD based models were not. In this case, the predictive power of a location's mycobiome seems to be inversely related to crop performance. Oats performed more poorly at ARL, and ARL models were able to successfully predict oat outcomes whereas WMAD models were not. Corn outcomes were poorer on average at WMAD and most

successfully described by WMAD mycobiome models. In oats, the phenotype level fungal models corroborate both possibilities. The ARL based oat yield model at the phenotype level was made up of a fairly even split between positive and negative fungal associates, which could imply oats outcomes were affected by both pathogen presence and interaction with beneficial microbes at ARL. It is possible that conditions were favorable enough for oats at WMAD that beneficial microbes were largely unnecessary for crop success and pathogens did not contribute enough of a yield drag to show up in the model.

To further tease apart these patterns at the phenotypic level we can compare the success of the full, fungal, and spatial models. Oat plot yield was successfully predicted by the spatial model at WMAD. This could either be a direct effect of spatial gradients within the field or an indirect effect of those spatial gradients mediated by a correlation between fungal populations and changes in field environment. Because the fungal model was not successful in predicting oat yield and the full model did not improve upon the spatial model, we conclude that a direct effect of spatial gradients on oat yield is the most probable scenario here. At the ARL location all three models (full, fungal, and spatial) were successful in predicting oat yield; however, the fungal model was the most successful and the full model did not improve upon it. This provides evidence that the mycobiome is the main differentiator of oat success at the ARL location, and correlations with spatial gradients may largely stem from a correlation between those gradients and populations of oat associating fungi.

The genotype level model at ARL was also successful in predicting oat yield. This implies a set of interactions between oat hosts and fungal associates that are tied to genetic variability in the oat population. Since spatial variance is randomly distributed throughout the genotype means used in this model, there is less ambiguity in evaluating the relationship between

this genotype associating mycobiome and oat yield. We can conclude a direct effect between these variables, but we must be cautious about assuming the directionality of this relationship. Something about variation in oat yield may affect these microbial populations or the variation in microbial populations may describe oat yield. The presence of AMF population as a coefficient in this model at least provides coherence for the influence of fungal population on oat yield. The success of this genotype level model also provides promise in terms of breeding oats to derive greater benefit from mycobiome interactions, both by mitigating interaction with negative associates and improving cooperation with positive associates.

Understanding soil legacy effects – can we predict corn productivity in the following season from rhizosphere fungal communities of the preceding oat crop?

A further objective of this study was to analyze the oat soil legacy effect on next-season corn at both the phenotype and genotype level. Interspecies soil legacy effects are common and important in both natural and agricultural systems, but beyond the impact of well-known pathogens, the microbial mechanism is often mysterious. Previous studies demonstrate intraspecific variability of soil legacy in some plant hosts (Schweitzer et al. 2008, Felker-Quinn et al. 2011). The data shown in Chapter 1 of this thesis provide evidence for this intraspecific variability in oats but given the lack of success of genotype level models in predicting corn outcomes that evidence has not persisted at the field level. Still, oat mycobiome data from the WMAD location was able to predict next-season corn yield at the phenotype level. The oat soil legacy effect could still be impactful on corn productivity, just not refined to the level of oat genotype. The corn yield spatial model is more successful than the fungal model by far, and the full model does not improve upon the spatial model. As described for the oat models, this could be due to two possibilities: 1) corn is responding to abiotic spatial gradients directly, and

correlation to fungal composition is just reflective of shared correlation with abiotic gradients. Or, 2) the spatial variation, in part, reflects the variation in fungal communities. The significance of the fungal model implies that, at least in part, some variation in corn yield is explained by the second possibility.

Candidate fungal taxa effecting oat agronomic outcomes.

The third objective of this study focused on the identification of known and unknown fungal contributors to oat and corn traits, and LASSO regression models provided further insights into the relationship between crop hosts and known associates (e.g. AMF) and allowed us to highlight candidate fungal associates that merit further observation.

AMF are well known plant mutualists, but their relative in agricultural systems seems to be highly dependent on both plant host and the environmental context. Some studies show AMF interaction can even decrease host yield in high-nutrient contexts (Hoeksema et al. 2010, Johnson et al. 2015). As AMF tend to be highly generalist in their host interactions, it is also unclear whether composition of the AMF community or just the general abundance of AMF are more important determinants of host health (Maherali and Klironomos 2007, Powell et al. 2009). In this study we found a genotypic signal for the interaction between AMF in oat hosts. Since AMF are so generalized we did not expect to find the AMF interaction to be a genotype-dependent trait. Several mechanisms exist to explain how this genetic variability could be influencing this interaction, including changes to root architecture, differences in the ability of the host to reward and/or sanction AMF mutualists, and differences in host ability to acquire nutrients independent of AMF cooperation (Sawers et al. 2010, Schultz et al. 2010, Sawers et al. 2018). The summed AMF population (as well as specific AMF genera) was found as a positive coefficient in models describing oat yield at both the phenotype and genotype level. So, AMF

interaction as a trait is not entirely compartmentalized within a given genotype, but the depth of interaction seems to be influenced by genetic variability within the oat population.

Plant pathogens, both as a guild and as individual genera, also made it through our candidate filtering process at both the genotype and phenotype level. The genotype and phenotype level association between oat yield and Puccinia is likely attributable to crown rust caused by *Puccinia coronata.* While this study uses mycobiome data from the oat rhizosphere, it is highly probably that crown rust spores fell to the soil and were incorporated into root cores during the sampling process. Variation in spore amount could still be observed in our samples and crown rust resistance is a well-documented genotype-dependent trait in oats (Park et al. 2022). The genus Periconia remained in the phenotype level model describing oat yield. This genus belongs to an ambiguous (non-taxonomic) grouping of grass-root associates known as dark septate endophytes (DSE). *Periconia macrospinosa* is a DSE that has been implicated in both positively and negatively affecting grass hosts depending on the study (Mandyam and Jumpponen 2014, Santos et al. 2021). This study shows a negative association between Periconia abundance and oat yield at the phenotype level. Even if a pathogen is well documented on a given host crop, LASSO regression can allow us to identify situations in which a plant pathogen/parasite is contributing to a yield drag that would be too subtle to identify with simpler modeling strategies or visible symptoms.

Other candidates observed in this study have no prior implication as oat associates. At the phenotype level both Coprinopsis and Heydenia were identified as a negative and positive oat associate respectively. At the genotype level, an unidentified genus from the order Sodariomycetes and the genus Plectosphaerella are both implicated in negatively affecting oat yield. Sordariomycetes contains several genera with well-known grass pathogens such as

Magnaporthe and Gaeumannomyces. The genus observed in this study seems to be differentiated enough from these genera to elude a more taxonomically resolved identification. Plectosphaerella, to our knowledge, does not contain any species known to be oat pathogens. Our observation of these taxa in our models marks them as potentially important to oat plants and provides justification for their further study.

Candidate fungal taxa underlying soil legacy effects on corn

Some oat rhizosphere associates in this study were also implicated as influencing corn outcomes. While the directionality of correlations between oat and fungal populations cannot be proven, the design of this study provides a fixed directionality between oat mycobiomes and corn outcomes. Still, there are multiple direct and indirect pathways by which oats and their associated rhizosphere communities could affect the outcomes of subsequently planted corn (Fig. 6). Mortierella, a phenotype level descriptor of corn yield, is a nearly ubiquitous saprotrophic soil microbe, and therefore could simply be affecting corn outcomes by correlating with spatial gradients in the same manner as corn plants in the field. Another genus in the phenotype level corn yield model is Clohesyomyces. This genus is typically mentioned in the literature as an endophyte of aquatic plants, so its presence here as a positive associate of corn is novel.

One community member, an unidentified genus from the order Sordariales, associated with oat genotype and had a positive effect on the biomass of corn at the V6 stage. This was the only fungal taxon at the genotype level that had any significant predictive power over outcomes in any corn trait.

This study shows oats have a strong genetic effect on populations of certain fungal taxa that could affect the productivity of the oat host, and at least one fungal taxon that could affect

early corn growth. It also shows the general usefulness of LASSO regression when it comes to understanding associations between microbial communities and a given response, as well as highlighting which specific community members may be most important to these associations. These data show how microbial communities can be used to predict host outcomes both within and between seasons. By continuing to model these microbial community associations we can improve our ability to understand what a generally health soil community looks like and how this changes based on location and host species. Through this study we have also highlighted several candidate taxa that could be used as breeding targets to improve the productivity of oats/corn or warrant further study to confirm their importance to these host crops.

Literature Cited.

Bakker, M. G., D. K. Manter, A. M. Sheflin, T. L. Weir, and J. M. Vivanco. 2012. Harnessing the rhizosphere microbiome through plant breeding and agricultural management. Plant and Soil **360**:1-13.

Benjamini, Y., and Y. Hochberg. 1995. CONTROLLING THE FALSE DISCOVERY RATE - A PRACTICAL AND POWERFUL APPROACH TO MULTIPLE TESTING. Journal of the Royal Statistical Society Series B-Methodological **57**:289-300.

Berendsen, R. L., C. M. J. Pieterse, and P. Bakker. 2012. The rhizosphere microbiome and plant health. Trends in Plant Science **17**:478-486.

Edwards, J., C. Johnson, C. Santos-Medellin, E. Lurie, N. K. Podishetty, S. Bhatnagar, J. A. Eisen, and V. Sundaresan. 2015. Structure, variation, and assembly of the root-associated microbiomes of rice. Proceedings of the National Academy of Sciences of the United States of America **112**:E911-E920.

Felker-Quinn, E., J. K. Bailey, and J. A. Schweitzer. 2011. Soil biota drive expression of genetic variation and development of population-specific feedbacks in an invasive plant. Ecology **92**:1208-1214.

Fox J, Weisberg S (2019). *An R Companion to Applied Regression*, Third edition. Sage, Thousand Oaks CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Friedman J, Hastie T, Tibshirani R (2010). "Regularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software*, **33**(1), 1–22. doi:10.18637/jss.v033.i01, https://www.jstatsoft.org/v33/i01/.

Gardes M, Bruns TD. ITS primers with enhanced specificity for basidiomycetes--application to the identification of mycorrhizae and rusts. Mol Ecol. 1993 Apr;2(2):113-8. doi: 10.1111/j.1365-294x.1993.tb00005.x. PMID: 8180733.

Gopal, M., & Gupta, A. (2016). Microbiome selection could spur next-generation plant breeding strategies. *Frontiers in Microbiology*, *7*(DEC), 1–10. https://doi.org/10.3389/fmicb.2016.01971

Hoeksema, J. D., V. B. Chaudhary, C. A. Gehring, N. C. Johnson, J. Karst, R. T. Koide, A. Pringle, C. Zabinski, J. D. Bever, J. C. Moore, G. W. T. Wilson, J. N. Klironomos, and J. Umbanhowar. 2010. A meta-analysis of context-dependency in plant response to inoculation with mycorrhizal fungi. Ecology Letters **13**:394-407.

Johnson, N. C., G. W. T. Wilson, J. A. Wilson, R. M. Miller, and M. A. Bowker. 2015. Mycorrhizal phenotypes and the Law of the Minimum. New Phytologist **205**:1473-1484.

Kulmatiski, A., K. H. Beard, J. R. Stevens, and S. M. Cobbold. 2008. Plant-soil feedbacks: a meta-analytical review. Ecology Letters **11**:980-992.

Lankau, R. A., and D. P. Keymer. 2016. Ectomycorrhizal richness declines toward the host species' range edge Molecular Ecology **25**:3224-2341.

Maherali, H., and J. N. Klironomos. 2007. Influence of Phylogeny on fungal community assembly and ecosystem functioning. Science **316**:1746-1748.

Mandyam, K. G., & Jumpponen, A. 2014. Mutualism-parasitism paradigm synthesized from results of root-endophyte models. *Frontiers in Microbiology*, *5*(DEC), 1–13. https://doi.org/10.3389/fmicb.2014.00776

Marie Simonin, Cindy Dasilva, Valeria Terzi, Eddy L M Ngonkeu, Diégane Diouf, Aboubacry Kane, Gilles Béna, Lionel Moulin. 2020. Influence of plant genotype and soil on the wheat rhizosphere microbiome: evidences for a core microbiome across eight African and European soils. *FEMS Microbiology Ecology*. 96(6): fiaa067. https://doi.org/10.1093/femsec/fiaa067

Mariotte, P., Z. Mehrabi, T. M. Bezemer, G. B. De Deyn, A. Kulmatiski, B. Drigo, G. F. Veen, M. G. A. van der Heijden, and P. Kardol. 2018. Plant-Soil Feedback: Bridging Natural and Agricultural Sciences. Trends in Ecology & Evolution **33**:129-142.

Mendes, R., M. Kruijt, I. de Bruijn, E. Dekkers, M. van der Voort, J. H. M. Schneider, Y. M. Piceno, T. Z. DeSantis, G. L. Andersen, P. A. H. M. Bakker, and J. M. Raaijmakers. 2011. Deciphering the Rhizosphere Microbiome for Disease-Suppressive Bacteria. Science **332**:1097-1100.

Oksanen, A. J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P. R., Hara, R. B. O., Simpson, G. L., Solymos, P., Stevens, M. H. H., & Szoecs, E. 2018. *Package ' vegan .' April*.

Park, R.F., Boshoff, W.H.P., Cabral, A.L. *et al.* 2022. Breeding oat for resistance to the crown rust pathogen *Puccinia coronata* f. sp. *avenae*: achievements and prospects. *Theor Appl Genet* **135**, 3709–3734. https://doi.org/10.1007/s00122-022-04121-z

Pearl, J. 2010. An Introduction to Causal Inference. International Journal of Biostatistics **6**:61.

Powell, J. R., J. L. Parrent, M. M. Hart, J. N. Klironomos, M. C. Rillig, and H. Maherali. 2009. Phylogenetic trait conservatism and the evolution of functional trade-offs in arbuscular mycorrhizal fungi. Proceedings of the Royal Society B-Biological Sciences **276**:4237-4245.

Rausher, M. D. 1992. The measurement of selection on quantitative traits: biases due to environmental covariances between traits and fitness. Evolution **46**:616-626.

Santos, Mila, Ignacio Cesanelli, Fernando Diánez, Brenda Sánchez-Montesinos, and Alejandro Moreno-Gavíra. 2021. Advances in the Role of Dark Septate Endophytes in the Plant Resistance to Abiotic and Biotic Stresses. *Journal of Fungi*. 7(11): 939. https://doi.org/10.3390/jof7110939

Sawers, R. J. H., M. N. Gebreselassie, D. P. Janos, and U. Paszkowski. 2010. Characterizing variation in mycorrhiza effect among diverse plant varieties. Theoretical and Applied Genetics **120**:1029-1039.

Sawers, R. J. H., M. R. Ramirez-Flores, V. Olalde-Portugal, and U. Paszkowski. 2018. The impact of domestication and crop improvement on arbuscular mycorrhizal symbiosis in cereals: insights from genetics and genomics. New Phytologist **220**:1135-1140.

Schlaeppi, K., and D. Bulgarelli. 2015. The Plant Microbiome at Work. Molecular Plant-Microbe Interactions **28**:212-217.

Schultz, C. J., L. V. Kochian, and M. J. Harrison. 2010. Genetic variation for root architecture, nutrient uptake and mycorrhizal colonisation in Medicago truncatula accessions. Plant and Soil **336**:113-128.

Schweitzer, J. A., J. K. Bailey, D. G. Fischer, C. J. Leroy, E. V. Lonsdorf, T. G. Whitham, and S. C. Hart. 2008. Plant-soil-microorganism interactions: Heritable relationship between plant genotype and associated soil microorganisms. Ecology **89**:773-781.

Searle, S. R. , Speed, F. M., & Milliken, G. A. 1980. Population Marginal Means in the Linear Model: An Alternative to Least Squares Means, The American Statistician, 34:4, 216-221, DOI: 10.1080/00031305.1980.10483031

Singh, B.K., Trivedi, P., Egidi, E. *et al.* 2020. Crop microbiome and sustainable agriculture. *Nat Rev Microbiol* **18**, 601–602. https://doi.org/10.1038/s41579-020-00446-y

Singh, R. and Goodwin S.B. 2022. Exploring the Corn Microbiome: A Detailed Review on Current Knowledge, Techniques, and Future Directions, PhytoFrontiers, 2-3:158-175.

Tibshirani, R. 1996. Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society Series B-Methodological **58**:267-288.

White, T. J., T. Bruns, S. Lee, and J. W. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pages 315-325 *in* M. A. Innis, D. H. Gelfand, J. J. Snisky, and T. J. White, editors. PCR Protocols: A Guide to Methods and Applications. Academic Press, Inc., New York.

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Xu, L., and D. Coleman-Derr. 2019. Causes and consequences of a conserved bacterial root microbiome response to drought stress. Current Opinion in Microbiology **49**:1-6.

Table 1. Summary of phenotype-level LASSO regression models

Location	Agronomic Trait	Full Train R^2	Full Test R^2	Full Test Cor p-value	Fungal Train R^2	Fungal Test R^2	Fungal Test Cor p-value	Spatial Train R^2	Spatial Test R^2	Spatial Test Cor p-value
WMAD	Oat Plot Yield	0.0823	0.0303	0.101	0.0595	0.00379	0.5644	0.173	0.125	0.000633
WMAD	Oat Plot Density	0	0	NA	0	0	NA	0.00623	0.00182	0.69
WMAD	Oat Test Weight	0.0988	0.00528	0.496	0	0	NA	0.0723	0.00103	0.924
WMAD	Corn V6 Weight	0.342	0.426	4.17×10^{-12}	0.0991	0.035	0.0791	0.307	0.4	3.04×10^{-11}
WMAD	Corn R1 Weight	0.276	0.384	3.28×10^{-10}	0	0	NA	0.3	0.39	2.2×10^{-10}
WMAD	Corn Extrapolated Plot Yield	0.272	0.568	2.2×10^{-16}	0.144	0.149	0.000173	0.299	0.615	2.2×10^{-16}
ARL	Oat Plot Yield	0.197	0.154	0.000134	0.18	0.162	8.53×10^{-5}	0.0667	0.0484	0.0373
ARL	Oat Plot Density	0	0	NA	0	0	NA	0	0	NA
ARL	Oat Test Weight	0.129	0.0306	0.0994	0.14	0.0247	0.139	0.0609	0.0483	0.0373
ARL	Corn V6 Weight	0.21	0.233	1.91×10^{-6}	0.0398	0.0242	0.148	0.187	0.23	2.34×10^{-6}
ARL	Corn R1 Weight	0.0905	0.0705	0.0159	0	0	NA	0.128	0.18	7.16×10^{-5}
ARL	Corn Extrapolated Plot Yield	0.0102	0.0592	0.0209	0.0303	0	NA	0.0102	0.0595	0.0205

Evaluation of the success of phenotype-level LASSO regression models in using oat mycobiome data to predict agronomic traits in a naïve test dataset. Outcomes of agronomic traits at each plot were each predicted with three models: "Full" incorporates both mycobiome community data and spatial blocking factors, "Fungal" uses only mycobiome community data, and "Spatial" uses only spatial blocking factors. "Train R^2" columns show the agronomic trait variation described by a model built using data from 90% of available data. "Test R^2" columns show amount of variation the trained model can describe in the remaining 10% of the data. "Cor p-value" columns show the p-value result of a statistical test determining significance of the correlation between values predicted by the model and those observed in the naïve 10% of withheld data.

Table 2. Summary of genotype-level LASSO regression models

Location	Trait	Train R2	Test R2	Test Cor p-value
WMAD	Oat Plot Yield	0.191	0.0458	0.239
WMAD	Oat Plot Density	0	0	NA
WMAD	Oat Test Weight	0.104	0.134	0.0396
WMAD	Corn V6 Weight	0.0456	0.0934	0.089
WMAD	Corn R1 Weight	0.0896	0.00502	0.7
WMAD	Corn Extrapolated Plot Yield	0.0276	0.00368	0.742
ARL	Oat Plot Yield	0.221	0.135	0.0383
ARL	Oat Plot Density	0.0644	0.143	0.0326
ARL	Oat Test Weight	0.0737	0.0466	0.236
ARL	Corn V6 Weight	0.032	0.0408	0.2674
ARL	Corn R1 Weight	0.0285	0.0332	0.318
ARL	Corn Extrapolated Plot Yield	0	0	NA

Evaluation of the success of genotype-level LASSO regression models in using oat mycobiome data to predict agronomic traits in a naïve test dataset. Estimated genotype means of fungal relative abundance were used to as variables in a LASSO regression model to predict genotype means of agronomic traits. "Train R^2" shows the agronomic trait variation described by a model built using 90% of available data. "Test R^2" shows amount of variation the trained model can describe in the remaining 10% of the data. "Cor p-value" shows the p-value result of a statistical test determining significance of the correlation between values predicted by the model and those observed in the naïve 10% of withheld data.

Table 3. Summary of fungal taxa coefficients in **phenotype-level** LASSO regression model describing oat yield at ARL

Fungal Coefficient	**LASSO Estimate**	**LM P-value**	**LM R^2**	**Sample Frequency (%)**	**Mean Relative Abundance When Detected (%)**
g__Betamyces	-60432.42	5.08E-08	0.0315	11.346	0.101
f__Stachybotryaceae	-58790.47	0.03915	0.00363	0.2224	0.137
g__Spiromastix	-47062.15	0.01693	0.00523	0.2225	0.045
c__Rozellomycotina	-15931.26	2.57E-13	0.0569	45.940	0.303
g__Periconia*	-814.13	6.27E-05	0.0166	86.541	1.289
g__Puccinia*	-506.29	0.000981	0.0109	97.553	2.526
g__Dioszegia	-498.30	0.076	0.00240	33.259	0.116
g__Coprinopsis*	-129.73	0.00749	0.00684	57.953	4.203
plant pathogens*	-15.34	0.000241	0.0138	99.889	16.034
Row	-0.19	NA	NA	NA	NA
Col	1.50	NA	NA	NA	NA
(Intercept)	73.13	NA	NA	NA	NA
arbuscular mycorrhizal*	226.30	7.48E-05	0.0163	96.997	4.908
g__Paraphoma	334.57	0.208	0.000656	31.702	0.246
g__Rhizophagus*	390.73	3.50E-05	0.0178	90.100	2.719
g__Heydenia*	397.58	2.83E-05	0.0183	80.534	1.927
o__Hypocreales	521.0	5.17E-05	0.0170	36.374	0.188
g__Funneliformis*	710.33	0.000302	0.0134	70.078	1.098
g__Tranzscheliella	767.13	0.0223	0.00470	0.444	0.029
g__Gaeumannomyces	772.43	0.0313	0.00405	3.115	0.298
g__Nectria	1110.48	0.00186	0.00964	72.747	0.404
g__Sporisorium	6558.53	0.0346	0.00386	0.334	0.190
g__Debaryomyces	28546.99	0.0293	0.00417	3.003	0.055
g__Perenniporia	285142.74	0.00134	0.0103	0.334	0.0549

Evaluation of fungal coefficients remaining in phenotype-level LASSO model. The relative abundances of these fungal taxa/guilds in each plot are predictive of oat yield outcomes in that plot. “Fungal Coefficient” shows the identify of each coefficient remaining in the LASSO model. Taxonomic identities are preceded by a letter denoting the level of identification (p = phylum, c = class, etc…). “LM p-value” and “LM R^2” show the results of a linear regression model testing the relative abundance of each individual taxon/guild against oat plot yield at ARL. “Sample Frequency (%)” denotes the percent of samples in which a given fungal taxon/guild was present, and “Mean Relative Abundance When Detected (%)” shows the percentage of the total fungal abundance is constituted by that taxon/guild across the samples in which it appeared. Asterisks denote taxa/guilds that meet criteria to be considered candidates of influence to oat/corn traits.

Table 4. Summary of fungal taxa coefficients in **genotype-level** LASSO regression model describing oat yield at ARL

Fungal Coefficient	**LASSO Estimate**	**LM P-value**	**LM R^2**	**Sample Frequency (%)**	**Mean Relative Abundance When Detected (%)**
c__Rozellomycotina	-20965.97	9.74E-09	0.09792	83.01	0.172
g__Puccinia*	-1943.49	0.0009976	0.03128	100	2.481
plant pathogens*	-203.68	0.001335	0.02959	100	16.020
g__Plectosphaerella*	-157.52	0.000674	0.03355	99.36	1.590
o__Sordariomycetes	-128.32	0.0167	0.01518	96.79	2.338
(Intercept)	39.99	NA	NA	NA	NA
g__Nectria	190.06	0.0144	0.01602	98.72	0.3005
arbuscular mycorrhizal*	843.88	5.66E-06	0.06137	99.04	4.854
g__Coniochaeta	1328.88	0.01154	0.01725	100	0.707
g__Funneliformis	2317.73	4.27E-05	0.04962	91.35	0.848

Evaluation of fungal coefficients remaining in genotype-level LASSO model. The genotype-level relative abundances of these fungal taxa/guilds in each plot are predictive of oat genotype yield outcomes. “Fungal Coefficient” shows the identify of each coefficient remaining in the LASSO model. Taxonomic identities are preceded by a letter denoting the level of identification (p = phylum, c = class, etc…). “LM p-value” and “LM R^2” show the results of a linear regression model testing the genotype-mean relative abundance of each individual taxon/guild against oat genotype yield at ARL. “Sample Frequency (%)” denotes the percent of samples in which a given fungal taxon/guild was present, and “Mean Relative Abundance When Detected (%)” shows the percentage of the total fungal abundance is constituted by that taxon/guild across the samples in which it appeared. Asterisks denote taxa/guilds that meet criteria to be considered candidates of influence to oat/corn traits.

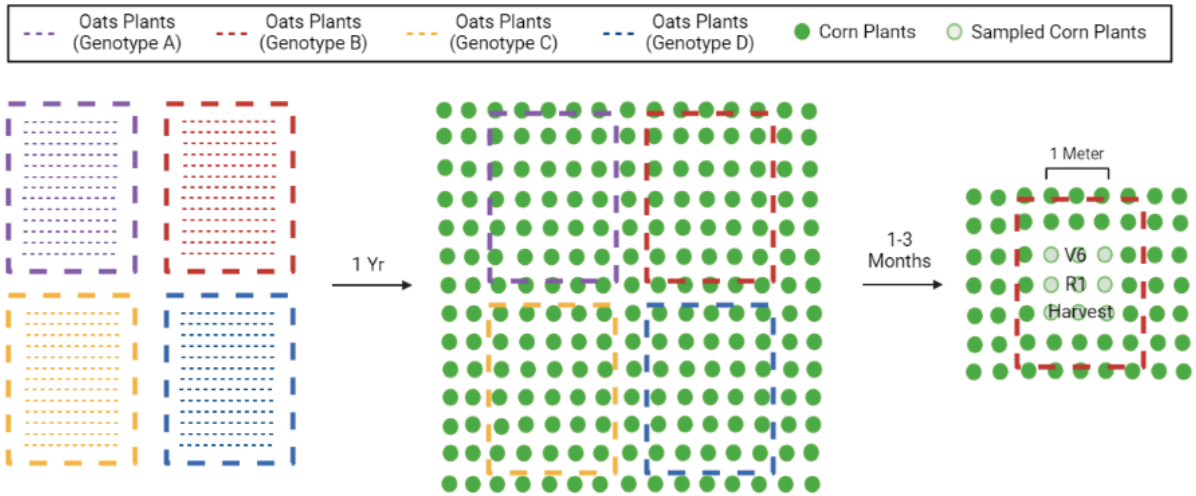

Figure 1. Experimental design of genome wide association study. 312 unique oat genotypes were planted at each of two locations: The West Madison Agricultural Research Station (WMAD) and the Arlington Agricultural Research Station (ARL). Each genotype was grown in at least three replicated plots, with eight genotypes being grown in six replicated plots for a total of 960 plots at each location (1,920 plots total). The following season, an elite variety of corn (*Zea mays*) was freely seeded into the oat plots to observe potential oat genotype specific soil legacy effects. To understand the effect of microbial community throughout the corn lifecycle, corn plants were sampled from each location at the V6 and R1 stages as well as at harvest. At the V6 and R1 sampling periods, one meter of corn was hand cut from a row at the middle-most point in each oat plot marked from the previous season.

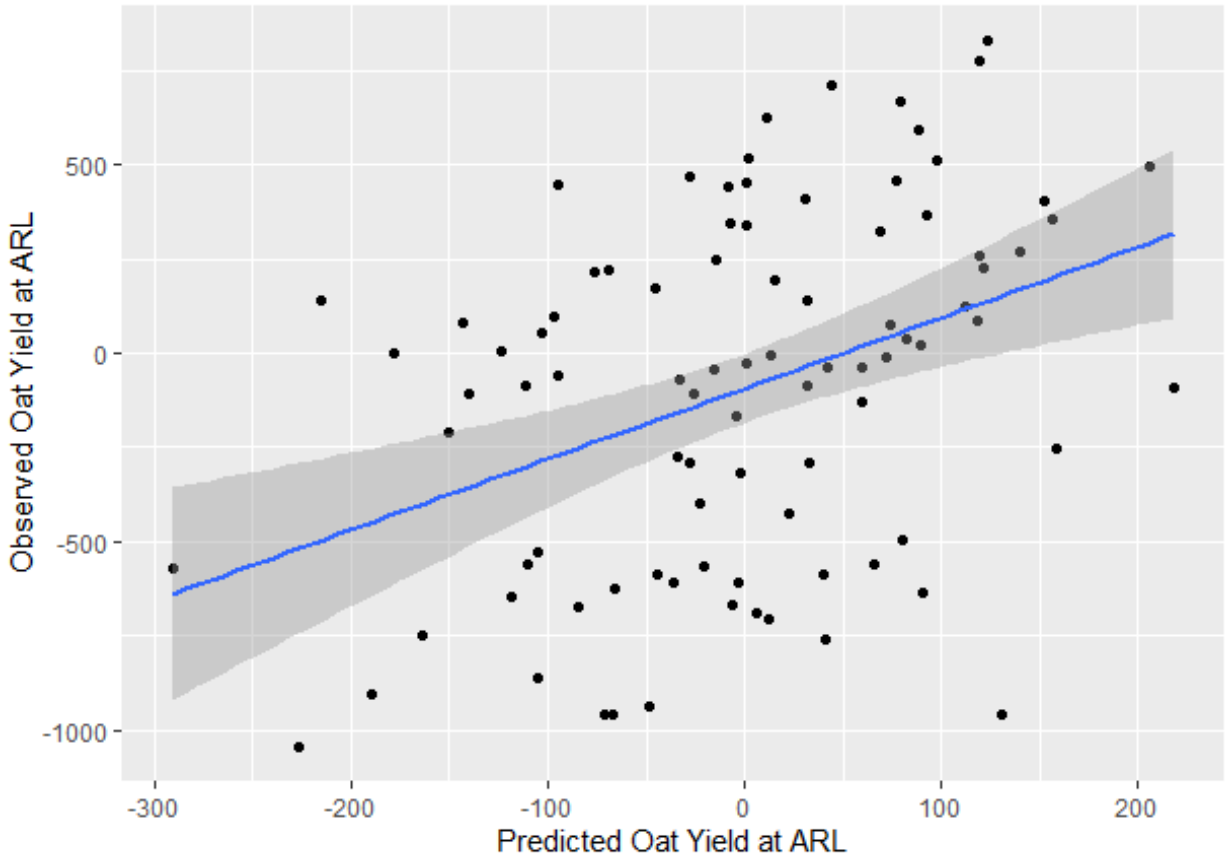

Figure 2. Regression of oat yield values predicted by **phenotype-level** LASSO regression model vs. values observed in a naïve subsection of the data. Predicted values are derived from a model using 90% of available plot mycobiome community data. Observed values are derived from the remaining 10% of naïve data that was withheld.

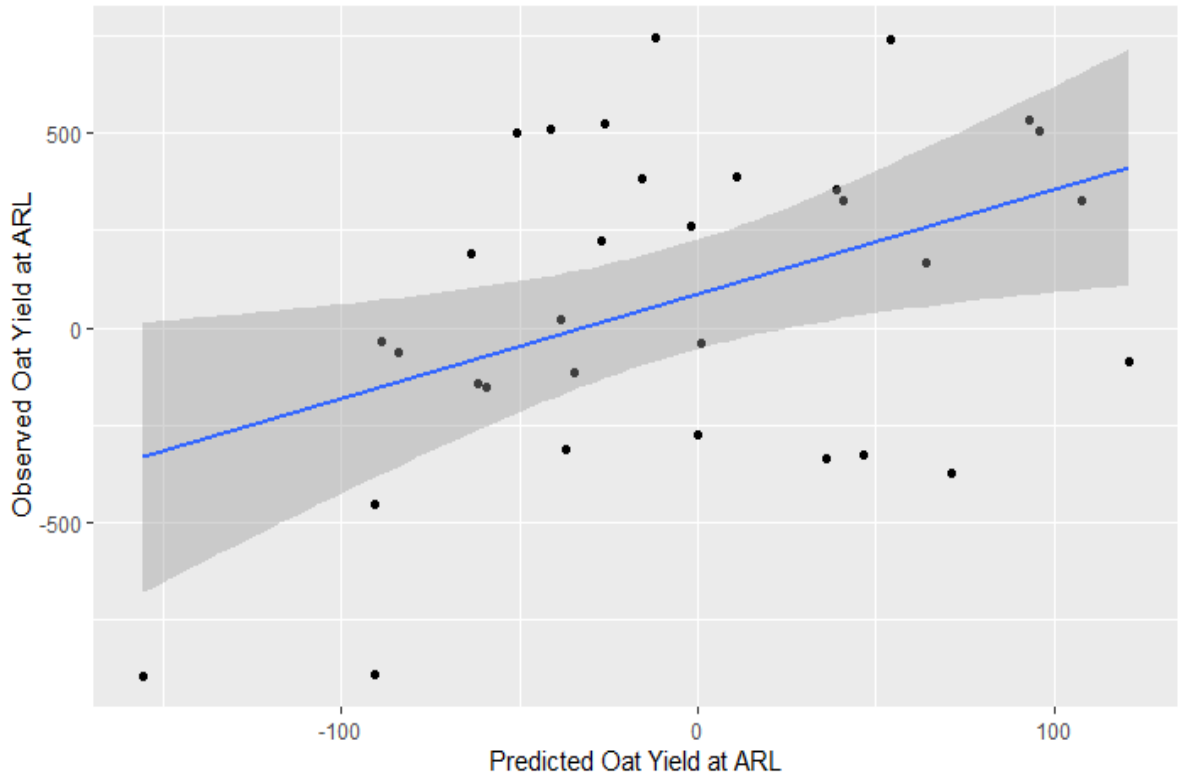

Figure 3. Regression of oat yield values predicted by **genotype-level** LASSO regression model vs. values observed in a naïve subsection of the data. Predicted values are derived from a model using 90% of available oat genotype mycobiome community data. Observed values are derived from the remaining 10% of naïve data that was withheld.

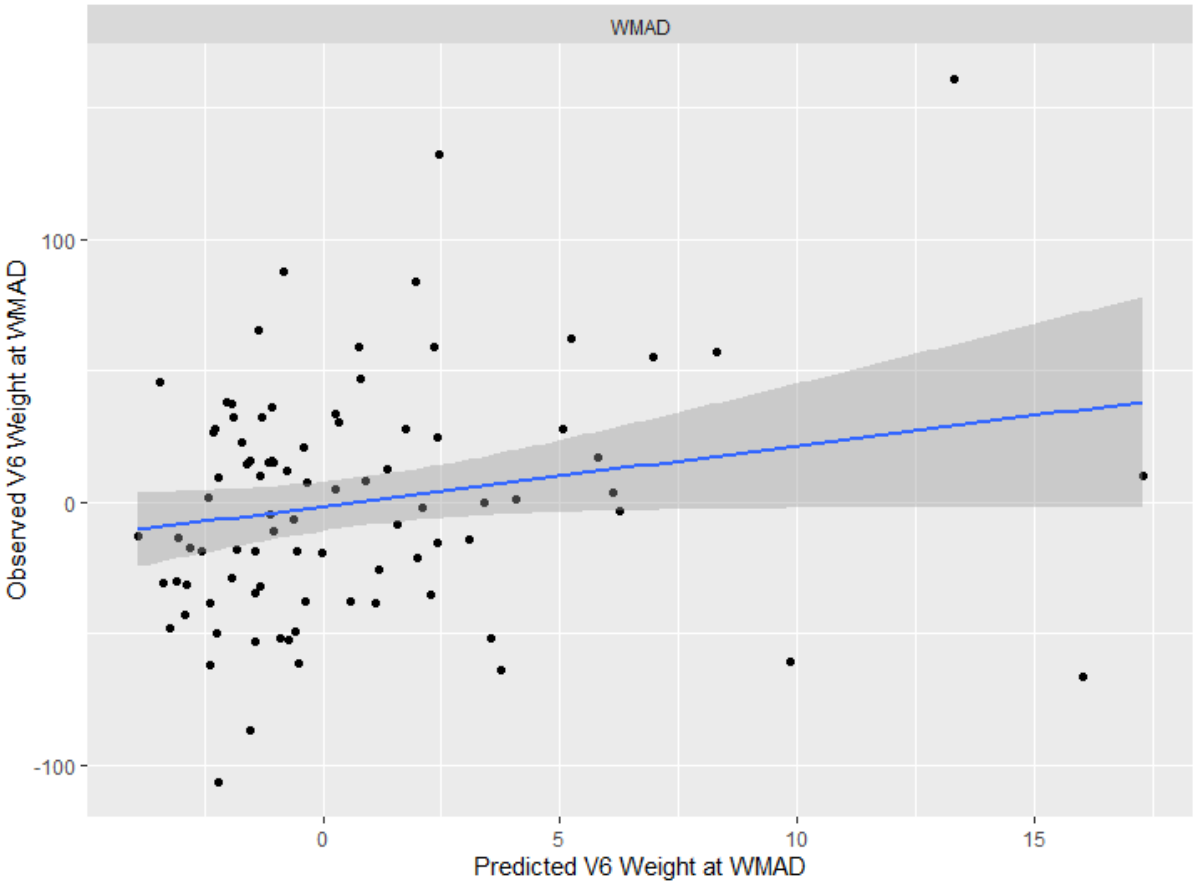

Figure 4. Regression of corn V6 stage biomass values predicted by **phenotype-level** LASSO regression model vs. values observed in a naïve subsection of the data. Predicted values are derived from a model using 90% of available plot mycobiome community data. Observed values are derived from the remaining 10% of naïve data that was withheld.

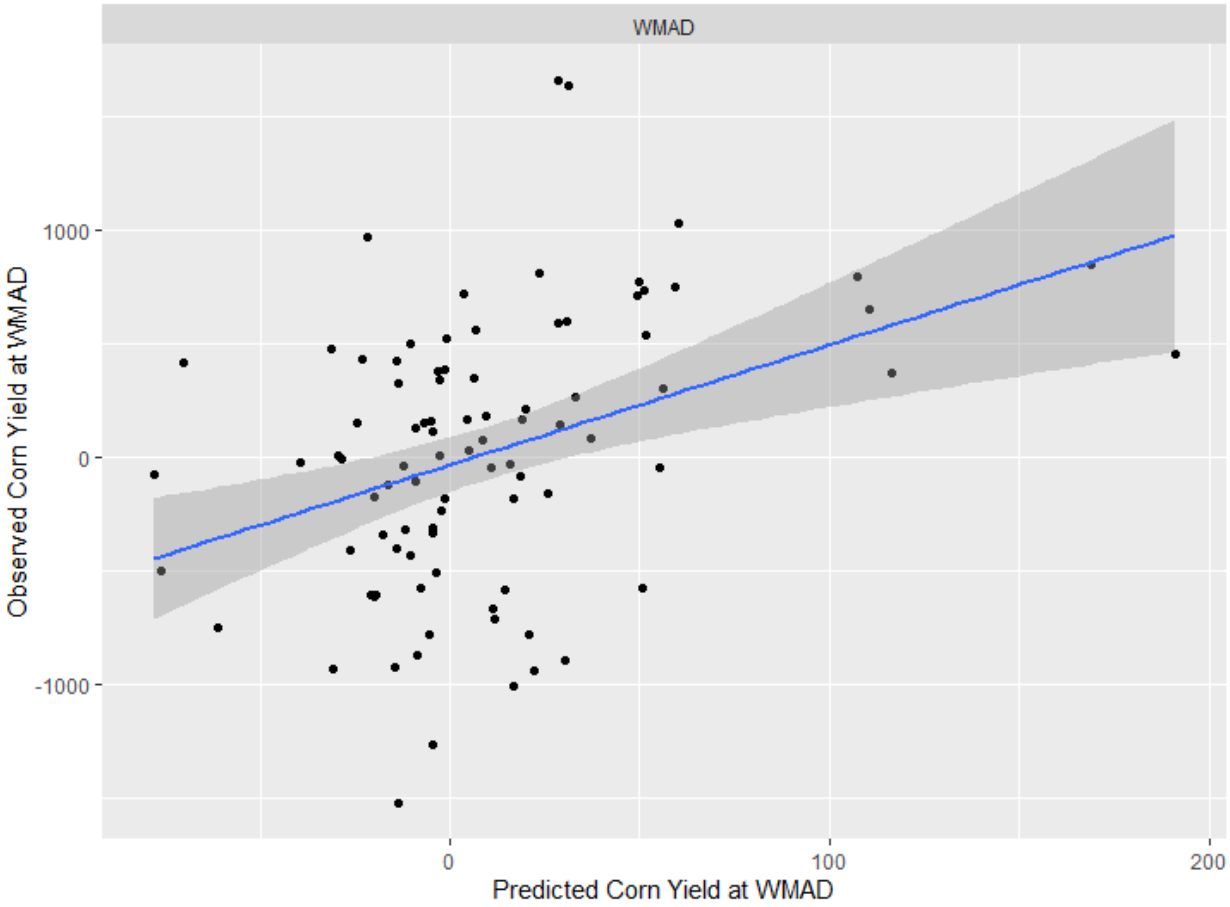

Figure 5. Regression of corn yield values predicted by **phenotype-level** LASSO regression model vs. values observed in a naïve subsection of the data. Predicted values are derived from a model using 90% of available plot mycobiome community data. Observed values are derived from the remaining 10% of naïve data that was withheld.

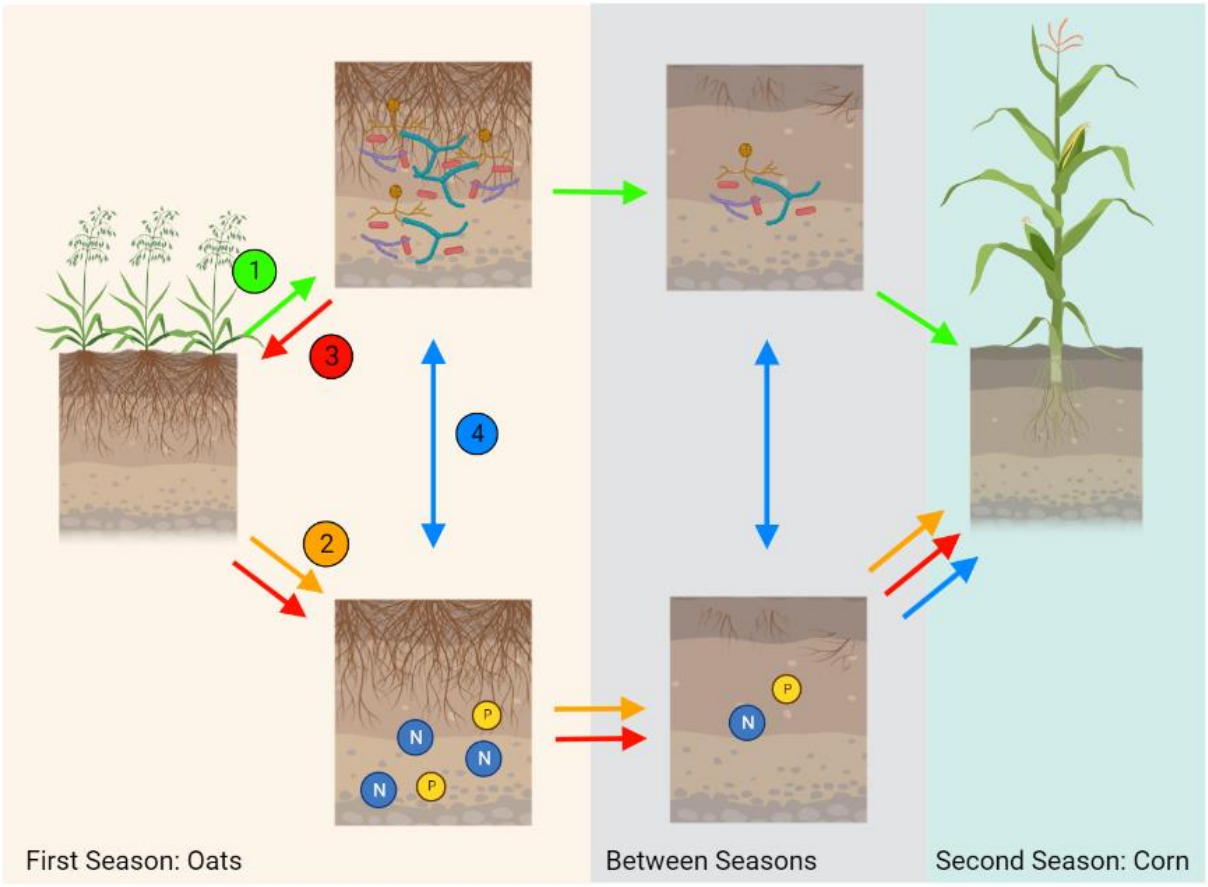

Figure 6. Model depicting potential pathways by which oat soil legacies can affect next-season corn. Oat genotypes can influence next-season corn outcomes by 1) manipulating the microbial soil communities, 2) directly altering the nutrient pool by differential uptake, 3) being altered by existing microbial communities in a way that increases or decreases nutrient uptake, and 4) indirectly altering the nutrient pool by changing microbial populations that contribute to nutrient uptake or cycling.

Chapter 4

Leveraging a genome wide association study to assess correlations between genetic variation in oats, fungal soil communities, and soil legacy effects.

Abstract

Interactions between soil microbes and host plants could be an important source of agricultural improvement. Genome wide association studies (GWAS) have the potential to be a powerful tool in associating genomic regions of crop plants with alterations to host-associated microbial communities. GWAS results could lead to breeding programs focused on improving the host's ability to cultivate a beneficial microbial community and, more presently, can provide rich hypothesis generating datasets exploring potentially novel plant-microbe interactions. In this study, we extracted fungal rhizosphere communities from a GWAS containing 312 oat genotypes replicated across two unique locations. Additionally, one season after the oat GWAS was harvested, we planted corn into the fields and tracked the agronomic outcomes of corn plants within the confines of the plots in which each oat genotype was grown. We identify 17 genera that significantly associate with SNPs from a marker dataset containing ~3,000 markers from across the oat genome. We also use linear modeling approaches to identify fungal rhizosphere community members that significantly associate with agronomic outcomes in both oats and subsequently planted corn.

Introduction

Sustainability is only becoming more important as the global population rises and the consequences of climate change become more severe. Plant associated microbial communities are known to grant many benefits to the plant host, including nutrient acquisition, drought tolerance, disease tolerance, and more (Bender et al., 2016). Agricultural strategies that rely more on microbial communities and less on inputs are one method by which sustainability could be improved in the field. Despite the potential for crop improvement, manipulating microbial soil communities for consistent benefit has proven difficult. Breeding elite crop cultivars for optimum microbiome cultivation and utilization could be a viable strategy for more predictably manipulating the plant-associated microbial community with the goal of maintaining crop yield in lower input, more sustainable systems.

For this breeding scheme to be feasible the crops involved must meet certain criteria. Firstly, they would have to retain meaningful variation in the traits responsible for mycobiome cultivation. The domestication hypothesis suggests that intensive breeding of a crop species correlates with a reduced dependency on the microbial community, therefore this crop would likely need to be one that has not been subjected to such breeding regiments (Kiers et al., 2007; Kim et al., 2020, Porter and Sachs 2020). Secondly, given the context dependency of many plant/microbe interactions, the crop involved should typically be grown with lower nutrient inputs. A lower nutrient context would allow for a situation in which genes responsible for cultivating relationships with beneficial microbes (e.g. mycorrhizae). In light of these specificities, this strategy could be a costly, time-consuming process that could result in undesirable agronomic trade-offs. Many cash crops have undergone decades of genetic

optimization to flourish in a high nutrient, agricultural setting. Another possible avenue exists in the breeding of rotational, cover crops to cultivate a beneficial microbial community for a subsequently planted crop. Certain cover crops fit the both previously stated criteria and have potential when it comes to breeding for microbial community cultivation.

This study focuses on a rotational system of oats (*Avena sativa*) as a cover crop and corn (*Zea mays*) as the subsequently planted cash crop. The genotypic variation in community selection has been shown to have a consequence on soil legacy in certain systems (Ellouze et al. 2013). Tapping into this genetic variability in microbiome selection could allow for more focused breeding programs in both the microbiome cultivating crop and a subsequently planted money crop, thus maintaining the benefits of a robust microbiome without the potential agronomic trade-offs. The first step towards realizing this goal lies in identifying the genetic control underlying the genotypic variation in microbiome selection. Studying the genes associated with microbial community assembly could potentially lead to the breeding of oat crops to select for specific, desirable microbial taxa within the soil. The approach of focusing on identifying these traits within a rotational crop rather than higher-value, cash crop could eventually give growers the freedom to choose a rotational crop species/genotype that conditions a healthier soil community without having to change their selection of which high-value crop species to plant.

In this study we used genome wide association to describe the genotypic variation in the structure of the oat-associated microbial community and correlate these differences in microbial assembly to specific genes within the collective oat genome. By sampling the rhizosphere of all oat plots in the GWAS, planting corn over those same plots in the subsequent

season and tracking the success of the corn plants within the confines of each plot, we were able to measure the genotypic variation on the soil fungal community and explore whether that variation leads to differential soil legacy effects on the subsequently planted corn.

Aside from the potential economic value, there is much to be learned from exploring the relationship between plant genetics and plant-associated microbes. By screening for plant genomic regions that significantly correlate with measurements of microbial community structure, we can provide evidence for putative mutualists/parasites, generate hypotheses about how the plant interacts with these microbes, and start to build an understanding of whether the genetic interaction a plant exhibits with a particular microbial taxon is a qualitative relationship Dor a quantitative multigenic relationship. By using GWAS we can explore two main ideas that provide insight into the genetic architecture of relationships between the oat host and its microbial symbionts. 1) The heritability of a given microbial population and 2) microbial association with single-nucleotide polymorphisms in the oat genome. In other words, this first idea answers the question “do certain genotypes exhibit differential cultivation of certain microbes?” and the second idea answers the question “is a genotype associated microbial trait controlled by a few genes of large effect or by many genes of small effect?” By comparing these two pieces of information we can identify important oat associates and form hypotheses about their relationship to the host plant. In general, based on our understanding of the gene-for-gene model of plant/pathogen interactions, we would predict that pathogenic microbes would be more likely to associate with oat genome SNPs (i.e. their populations are influenced by a few genes of large effect). Mycorrhizal mutualists, on the other hand, are typically quite general in their host range and we would predict their populations to exhibit low heritability and lack significant

association with SNPs (Dowarah et al., 2022). Community summaries such as diversity measurements would likely be affected by many genes and therefore would lack association with any SNPs.

It is important to note that any potential host gene interactions with a given microbe require that microbe to be present in the soil. So, unless a given microbe is ubiquitous, we expect any genotype level microbial associations to be strongly influenced by the environment. This genotype by environment interaction could be highly significant when it comes to soil legacy effects. If an oat genotype specific soil legacy effect on corn is due to an assortment of genes which influence microbial diversity in a beneficial manner then the environment effect might be less impactful on the rotational breeding strategy we have laid out. If this soil legacy effect is due to the cultivation or mitigation of a specific microbe then the success of this strategy could entirely hinge on the environmental context.

The objectives of this study were (i) assess the heritability of fungal taxa within the oat mycobiome, (ii) screen for oat genome SNPs that significantly associate with mycobiome cultivation, and (iii) determine if an oat soil legacy effect on next-season corn can be detected at the level of oat genotype or oat genome SNPs**.** In predicting outcomes of heritability, we hypothesized that (i) heritability for pathogens would be higher at the species level, whereas heritability for arbuscular mycorrhizal fungi (AMF) would be higher at the guild level. (ii) In general, broader scale community measurements (e.g. alpha diversity) would exhibit a higher heritability than single species abundance, and (iii) next-season corn traits would exhibit a detectable yet greatly diminished heritability compared to oat agronomic or mycobiome

cultivating traits. Our hypotheses concerning detection of significant SNPs were (i) SNPs will significantly associate with specific pathogens at a higher frequency than guilds with a more generalist lifestyle (e.g. AMF, saptrotrophs), and (ii) oat genome SNPs would not significantly associate with corn as soil legacy effect will be largely driven by generalist microbes.

Materials and Methods

Field experimental design and sampling: This research was conducted through the use of a genome wide association study (GWAS). In the Spring of 2019, 312 unique oat genotypes were planted at each of two locations: The West Madison Agricultural Research Station (WMAD) and the Arlington Agricultural Research Station (ARL). The oat genotypes were grown in 1.52 x 3.04 m plots with a 1.5m spacer between them and a boundary row of oats around the edges of the field. Each genotype was grown in at least three replicated plots, with eight genotypes being grown in six replicated plots for a total of 960 plots at each location (1,920 plots total). The two field locations were organized in a randomized block design. Each location was organized into three randomized blocks, each containing one of the replicated field plots from each oat genotype. Each field location consisted of 30 columns and 32 rows (10 columns per replicated block). The oat fields did not receive any nutrient, water, or fungicide inputs throughout the season. Each location was treated once with an herbicide to mitigate competition from weeds.

Oat root core samples were collected from each of the 960 plots within one week after the peak grain heading day at each location (>50% of oat plants heading). Root cores were stored at 4ºC for a maximum of 72 hours before processing. Root tissue was manually collected from each core by shaking off any excess soil before storing in 96 deep-well plates at -80ºC to await subsequent DNA extraction and sequencing. The oat plots were harvested in the Fall of 2019 and

the plot weight, plot density, and test weight (a grain density measurement [Kg/hL]) of each genotype was recorded.

In the spring of 2019, an elite variety of corn (*Zea mays*) was freely seeded into the oat plots to observe potential oat genotype specific soil legacy effects. The fields at each location were measured and plotted out to identify the center of each unique oat genotype plot from the previous season. To understand the effect of microbial community throughout the corn lifecycle, corn plants were sampled from each location at the V6 and R1 stages as well as at harvest. At the V6 and R1 sampling periods, one meter of corn was hand cut from a row at the middlemost point in each oat plot marked from the previous season (Fig. 1). The plants from each unique plot were bundled together, labeled, and placed in industrial drying ovens at 65ºC for 14 days. After drying, the plants were counted and weighed to collect density and biomass data from each timepoint. At harvest, one meter of plants from the middle of each previous oat plot were counted before the ears were removed from each plant, shucked, and placed in a labeled canvas bag before drying at 65ºC for two weeks. After this drying period, the cobs from each plot were counted and threshed. The threshed grain was tested for percent moisture and then weighed. To summarize, the resulting phenotypic data from corn were V6 dry weight, R1 dry weight, threshed grain dry weight, and number of plants sampled at each stage. Biomass at each stage was divided by the plant number to get average biomass at each growth stage. Plant number from each stage was added together to achieve a proxy for within plot plant density, and this density was multiplied by the average grain yield per corn plant to achieve an extrapolated yield for the whole plot. The corn fields did not receive any fertilizer, water, or pesticide inputs throughout the growing season.

***Amplicon preparation and sequencing*:** DNA from the 1,920 oat rhizosphere samples were extracted using Promega E-Z Plant DNA extraction kits (Promega, Madison, WI) according to the manufacturer's instructions. To characterize the fungal community, the fungal ITS2 region was amplified in each sample using the ITS1f forward primer (5' - CTTGGTCATTTAGAGGAAGTAA, (Gardes and Bruns, 1993)) and ITS4 reverse primer (5'- TCCTCCGCTTATTGATATGC, White et al., 1990)). External fusion PCR primers contained a 14-bp overlap to the trailing end of internal primers with 12bp i7 index and P7 flow cell adaptor or an i5 index, 7-bp spacer and P5 adapter (See Lankau and Keymer 2015). Amplicon library preparation was completed in two rounds of PCR. The first round of PCR amplified the fungal ITS2 region as well as the Nextera read primers. PCR was performed in 10 µl reactions using 0.2 µL of a hot-start, high fidelity polymerase (Clonetech Prime Star GLX, Fitchburg, WI) with 2 µL of its 5X buffer, 0.8 µL dNTPs (at 10 nM concentration), 0.25 µL of each primer (at 10 nM), 0.7 µg T4 gene 32 protein, and 10 ng of template DNA. The thermocycling program for the ITS2 region was a 5-minute hot start at 98˚C, 35 cycles of denaturing (98˚C, 0:30), annealing (50˚C, 0:45), and extension (68˚C, 1:00) and a final extension of 15 minutes at 68˚C. The second round of PCR added the P5 and P7 flowcell adapters to prepare the library for sequencing on an Illumina MiSeq, along with an external set of sample barcodes located between the flowcell adaptors and read primers. The amplicons were cleaned with the Omega BioTek E-Z 96 Cycle Pure kit (Omega Bio-tek, Norcross, GA). Purified products were quantified using a Qubit 2.0 fluorometer with the Qubit dsDNA HS assay and then pooled at equal concentrations (Thermo Scientific, Grand Island, NY). Amplicon products were then sequenced on Illumina Miseq using a 300 cycle Paired-End run at the University of Wisconsin-Madison Biotechnology Center. To

achieve sufficient sequence depth, the experiment was split across three separate runs, with one of the three replicates from each location being sequenced on each run.

Bioinformatics: Raw external sequences were initially trimmed at both the 5' and 3' ends using Cutadapt (version 1.18). The Qiime2 (v2017.12) pipeline was used to processed trimmed reads using DADA2. Samples were filtered and further trimmed by DADA2 using the following parameters (ITS2 sequences: p-trunc-len-f = 0, p-trunc-len-r = 280, p-max-ee = 4). We used the RDP Naïve Bayesian Classifier to assign taxonomy to fungal amplicon sequence variants (AVS) using the UNITE (version 8.0) reference database. Reads assigned to plant taxa were removed. Samples with fewer than 2500 reads were removed based on inspection of rarefaction curves. In the end, this study collected data from 1,920 samples split evenly between two locations (ARL and WMAD). Of these original samples, a total of 1,801 were retained (899 from ARL and 902 from WMAD). Once the fungal ASV sequence counts were converted into tabular form, they were transformed into relative abundance values using the "vegan" package in R (Oksanen et al. 2018).

Statistical Analysis:

Objective 1. Assess the heritability of fungal taxa within the oat mycobiome:

PERMANOVA analysis on the community structure at each location was run using the adonis function within the "vegan" package, and stacked bar charts visualizing abundance of fungal abundance were created using the R package "ggplot2" (Wickham, 2016). Rather than indiscriminately testing the genotypic association of all fungal ASVs, we employed a filtering process to narrow our selection. Fungal ASV's were collapsed by genus and sorted into probable

lifestyle guilds using the FungalTraits database (Põlme et al., 2020). This resulted in two data frames: one table consisting of 1,239 unique, fungal genera across all samples and one table with these genera collapsed into probably lifestyle guilds (e.g. plant pathogen, arbuscular mycorrhizal fungi, saprotrophs, etc…). Of these initial groupings, those that showed an average abundance of >0.1% of the total fungal abundance within a sample and were present in >33% of all samples were subjected to further analysis. This filtering step resulted in 123 unique fungal genera. These 123 genera were tested for oat genotype association using a mixed effect models created using the "lme4" package in R. These mixed effect models contained spatial factors (row, column, replicate, incomplete block) as random effects and included genotype as a fixed effect. Separate models were run for each location and both locations together. In models including both locations, spatial factors were nested within location and location was included as a mixed effect. Fungal traits (e.g. relative abundance, diversity) that significantly associated with oat genotype ($p < 0.100$) were then tested via mixed effect models (separately for each location and both locations together) to determine how much variation in their relative abundances is attributable to oat genotype (G), environment (E), and gene x environment (GxE) interactions (Bates et al., 2015). We then used the output of these random effects models to calculate heritability for each of these traits. Fungal traits that significantly associated with genotype as a fixed effect and exhibited a heritability >0.05 (i.e. greater than 5% of variation is attributable to host genetics) were concluded to be genotype level oat associates.

Objective 2. screen for oat genome SNPs that significantly associate with mycobiome cultivation.

The relative abundances of the 123 genera and 8 guilds that passed met the threshold for both sample frequency (>33% of samples) and average relative abundance (>0.1% of total ASVs) were analyzed for association with oat genome SNPs. A GWAS approach was used to map the microbial community assembly quantitative trait loci (QTL) by using the microbial community abundance and diversity indexes as oat phenotypes to be mapped. Standard GWAS mixed model approaches were used correcting for population structure (Q) and genetic relatedness (K) as necessary. The "GWASpoly" package in R was used to run the GWAS models and create Manhattan plots for each fungal community, corn, and oat trait. These Manhattan plots were created using a minimum allele frequency (MAF) of 10%, and a threshold value of significance for the $-\log10(p)$ values was set using the effective number of tests (M_{eff}) method (Gutierrez et al., 2011; Lipka et al. 2012; Asif et al., 2021). The oat GWAS included the 312 oat genotypes replicated in three plots at each of the two locations (WMAD and ARL), and the QTL mapping was done across a data set of ~3,000 high quality SNPs sequenced from the genomes of these genotypes. The fungal traits were then visualized and tested for SNP association via Manhattan plots within the "GWASpoly" package. The estimated means used in this analysis were acquired using the "emmeans" package in R (Searle et al., 1980). The estimated means of each fungal trait were extracted from a unique linear model for each location including blocking factors (e.g. row, column, block) to control for as much environmental variation as possible. In some cases, the mycobiome traits that significantly associated with oat genotype (heritability > 5% and/or fixed effect p-value < 0.10) and/or SNPs at either or both locations were tested for association between their relative abundances and oat agronomic traits using the estimated genotype means (of both the fungal and agronomic trait) and linear modeling. All visualizations of linear models were created using the R package "ggplot2" (Wickham, 2016).

Objective 3. Determine if an oat soil legacy effect on next-season corn can be detected at the level of oat genotype or oat genome SNPs: Using the same methods as described in objectives 1 and 2, oat genotype association, heritability, and SNP associations were analyzed for all of the corn phenotypes measured in this study. If a corn phenotype exhibits a significant association with oat genotype ($p < 0.100$), a heritability of greater than 0.05, or significantly associates with an oat genome SNP then we will have evidence for an oat soil legacy effect on next-season corn as well as information about the genetic architecture of that soil legacy effect. As for oats, the mycobiome traits that significantly associated with oat genotype (heritability > 5% and/or fixed effect p-value < 0.10) and/or SNPs at either or both locations were tested for association between their relative abundances and corn agronomic traits using the estimated genotype means (of both the fungal and agronomic trait) and linear modeling and visualizations of linear models were created using the R package "ggplot2."

Results

Variation between locations: The rhizosphere associated fungal community structure was significantly different between the WMAD and ARL locations (Fig. 2; permanova ($p < 0.001$)). The ARL location showed both a higher proportion of plant pathogens and a lower proportion of mycorrhizal species compared to the WMAD location. In terms of crop success, oats exhibited a higher average yield at the WMAD location (t-test; p-value <0.0001), while corn exhibited a higher average yield at the ARL location (t-test; p-value <0.0001). There was a significant inverse correlation between corn density (number of plants per meter) and average corn biomass. The strength of this correlation increased throughout the growing season, being the lowest at the V6 stage and highest at harvest (Fig. 3). Due to this phenomenon, we included both aggregated and average biomass measurements in any analysis concerning corn phenotypes.

Objective 1. Assess the heritability of fungal taxa within the oat mycobiome:

Genotypic variation in both agronomic and fungal community traits is observable at each location, however random modeling shows oat genotype describes zero percent or a very low percent of variation in many of these traits (Table 1). Genotype (G) and genotype by environment (GxE) effects were both subtly observed in the species diversity (Shannon-Weiner Index) when measured across both locations. G and GxE covariates each described 0.26% and 1.49% of the variation in species diversity respectively. When modeled within each individual location, heritability was 0.0 in WMAD and 0.13 in ARL. This location dependency was observed frequently throughout the modeled traits.

Out of the 123 fungal genera and eight fungal guilds that were tested, 24 genera and four guilds met the thresholds for genotype association ($p < 0.100$) and heritability (>0.05) necessary to be considered as an oat genotype influenced trait in this study. Ten of these genera were

identified at the ARL location and 14 were identified at the WMAD location (Table 1). Only one genus, Mortierella, met the thresholds for genotype association at both locations. At the guild level, arbuscular mycorrhizae significantly associated with oat genotype at ARL, while three different saprotrophic guilds (litter, soil, and unspecified) associated with oat genotype at WMAD (Table 1).

Objective 2. screen for oat genome SNPs that significantly associate with mycobiome cultivation.

Several single-nucleotide polymorphisms (SNPs) from the oat metagenome were significantly correlated with fungal community traits (Table 2) via Manhattan plots. Ten fungal genera showed a significant association between relative abundance and one or more SNPs within the oat metagenome. Eight of these SNP associated genera were observed in WMAD and two were observed in ARL. Two of these genera also associated with an SNP via the combined model of both locations, but no genus significantly associated with one or more SNP at each location individually. Only two of these genera (Ustilago and an unidentified genus within the class Rozellomycotina) met the thresholds for heritability and genotype association, although Funneliformis was very close with a heritability of 0.048 ($p = 0.1009$).

Objective 3. Determine if an oat soil legacy effect on next-season corn can be detected at the level of oat genotype or oat genome SNPs: None of the corn phenotypes tested associated with any oat SNPs above the semi-stringent significance threshold (Table 3). Despite this, variation in the Manhattan plots was evident. Especially at ARL, certain SNPs tested for association with yield traits approach or exceed a $-\log_{10}(p)$ score of 4.0 (Fig 4). These SNPs could be candidates for further testing in their importance to oat soil legacy and its impact on subsequently planted

crops. In contrast, all three of the oat agronomic traits tested significantly associated with oat SNPs at one or both locations (Fig. 5).

Overall we observed little to no evidence of an oat genotype specific soil legacy effect on next-season corn. Of the corn traits measured, R1 Weight (aggregated dry weight of corn harvested at the R1 growth stage) and the extrapolated plot yield (average grain yield of each corn plant multiplied by the total number of corn plants [density] sampled throughout the season) showed association with oat genotype. When R1 Weight analyzed across both locations the G and GxE effects were 0.20% and 0.0% respectively, which is coherent with the 0.0% of described by the G effect at ARL. At WMAD, however, the G effect was 2.01% with a heritability of 0.11 and genotype showed a very low amount of evidence for association with R1 Weight ($p=0.1404$). The extrapolated plot yield showed a similar trend, with low G and GxE effects across locations (0.34%, <0.01%), a low G effect of 0.32% at ARL ($p=0.451$), and a higher G effect of 1.12% and heritability of 0.069 at WMAD ($p=0.149$). The other corn traits all exhibited G or GxE effects of less than 1.0%, however, the G effect was generally higher at WMAD where corn performed more poorly in comparison to ARL. While the oat genotype effect on corn traits is being measured across a wide temporal gap, it can still be useful to compare to the G and GxE effects observed on the traits of the oats themselves. For example, when the aggregated grain yield of each plot was analyzed across both locations they displayed G, GxE, and E effects of 30.1%, 22.7%, and 16.3% respectively. The G effect was also strong within locations, describing 39.1% of variation in oat plot weight at WMAD ($p<0.001$) and 72.5% of variation at ARL ($p<0.001$).

Correlating oat mycobiome traits to oat and corn phenotypes: Fungal taxa or community level traits that displayed substantial heritabilities and/or significant associations with individual SNPs

were tested for their correlation with oat and corn phenotypes of interest. Funneliformis associated with a significant SNP in the combined models and at the ARL location (Fig. 6), but only showed weak evidence of an association with oat genotype, displaying heritability scores of 0.048 (p=0.101) when tested across both locations and scores of 0.0 (p=0.080) and 0.114 (p=0.138) at WMAD and ARL respectively. The estimated genotype means of Funneliformis relative abundance correlated significantly with those of oat and corn agronomic outcomes. Funneliformis abundance showed a positive relationship with oat plot weight (Fig. 7A; $p < 0.001$) and test weight (Fig. 7B; $p = 0.0046$). At WMAD, no association was seen between Funneliformis abundance and oat traits, however, a negative association was observed in the later stages of corn growth where Funneliformis abundance correlated with a distinct decrease in corn yield (Fig. 8; $p = 0.00575$).

Pyrenophora, a genus containing known foliar pathogens, showed the highest level of variation described by oat genotype (Table 1). When modeled across both locations, G and GxE described 1.98% and 1.57% of variation in relative Pyrenophora abundance respectively. In comparison, the environment effect (E) only described 0.082% of variation in Pyrenophora abundance. When modeled within each location individually, G described 7.56% of variation in Pyrenophora abundance at WMAD and 1.97% at ARL. When the fixed effect of genotype and location on Pyrenophora abundance were analyzed using a mixed modeling approach, genotype was significant (p=0.04848) while location was not (p=0.2463). The significance of genotype as a fixed effect term was maintained at WMAD (p=0.0233) but not at ARL (p=0.3649). Pyrenophora abundance showed a positive association with corn yield at ARL (Fig. 9A; p=0.019). This could be explained by a decrease of resource competition from first season oats if Pyrenophora decreased their biomass, however, the only correlation between Pyrenophora

abundance and an oat trait was a positive relationship observed with oat test weight at WMAD (Fig 9B; p=0.0472). As test weight is a density measurement, sometimes pathogen infection can increase this measurement by decreasing grain size, but we are unable to confirm this with our dataset.

The fungal genera that were tested due to their significant association with one or more oat SNP showed high variation in their genotype association and heritability scores (Table 2). The unidentified genus belonging to the Rozellomycotina class associated with one SNP at ARL (Fig. 10), where it displayed a relatively high heritability of 0.30, negatively associated with oat yield (Fig. 11A, $p < 0.001$) and positively associated with corn yield (Fig. 11B, $p = 0.066$). The genus within the order Hypocreales was another unidentified genus of potential import, showing a heritability of 0.137 across both locations and positively associating with V6 corn biomass at WMAD (p=0.0499) and with oat yield at ARL (p=0.0869). Some genera, such as Vishniacozyma, show zero heritability and no association with oat genotype despite associating with one or more oat SNP. Given the high levels of variation in microbial population both within and between locations, it is possible the increased number of replicates when testing for associations with oat markers allows for clarity in trends that cannot be observed when testing against genotype. It is also possible that some of these SNP association are due to random error. We have attempted to show multiple layers of evidence for associations between host and each microbe tested. We believe the SNP associating genera that do not associate with genotype and/or host phenotypes should be treated with a higher degree of skepticism than those that do.

Discussion

GWAS has proven a viable tool for the identification of genomic breeding targets to improve crop plant phenotypes. This study shows GWAS is also a viable tool for correlating genomic regions to host cultivation of specific rhizosphere communities. Using GWAS, we found several oat genomic markers that significantly associate with the relative abundance of important oat pathogens and mutualists as well as markers that associate with unresolved fungal taxa. While no oat SNPs significantly correlated with the success of subsequently planted corn, the Manhattan plots correlating corn traits to oat SNPs did show a high degree of variation among SNPs at the two individual locations, although the feasibility of breeding oats to affect next-season corn productivity seems low.

The percentages of variation in specific fungal taxa described by oat genotype were generally very small (<1.0%-10%) when they were existent at all, but even these slight associations with genotype can provide evidence for genetic relationships between plant host and associated microbe that could be explored further. Other recent studies using GWAS as a tool to ascribe QTLs to changes in the host associated microbial communities have also mentioned the challenge of the high environmental variation in these microbial traits (Deng et al., 2021; Walters et al., 2018). Given the extreme potential for variation in fungal communities both between and within field locations, it is improbable to observe the same levels of variation genotype described in plant encapsulated traits such as yield and test weight.

Of the 123 fungal genera that were tested, 24 (19.5%) showed a genotypic association. By guild, five of these genotype associating genera were plant pathogens, two were arbuscular mycorrhizal fungi, one was a mycoparasite, eight were saprotrophs of some kind, and eight could not be identified to one guild. Ten of the 123 genera (8.1%) associated with at least one SNP at either

location. Two of these were plant pathogens, one was an arbuscular mycorrhizal fungus, four were saprotrophs of some kind, and three could not be identified to the guild level. These findings do not support our hypotheses concerning pathogen and AMF association with oat genotype/SNPs.

For instance, the significant association of Funneliformis with two SNPs was a surprising find. Given the generalist nature of most arbuscular mycorrhizal (AM) species, we did not expect to find an interaction at the genotype level or association with specific host SNPs (Dowarah et al., 2022). Another study found specific AM associating SNPs in maize (Deng et al., 2020), so GWAS could prove a viable tool for elucidating specific AM genetic interaction on the host side, even if the fungi boast a large host range. Our data also show a distinctly negative association between Funneliformis abundance and corn yield outcomes at one location. This could be an example of context dependency in host/AM interactions, with corn paying carbon for a service it had no need for (Johnson et al., 2015) or could possibly be an example of AM showing a trend towards a more parasitic lifestyle when associating with a less-cooperative host (Martin-Robles et al., 2018). It also must be acknowledged that this association could simply be due to an association with greater oat success in the previous season. As Funneliformis abundance correlated with an increase in oat yield, this could have led to a depletion of nutrient resources that were then unavailable to next season corn. Whatever the reason, the association of Funneliformis abundance with oat/corn traits, as well as oat SNPs, make this an exciting candidate for further testing. However, the flip from a positive to negative relationship between oat and corn associations could make it a difficult breeding target with inevitable trade-offs between the two crops.

The other SNP associating genera did not significantly associate with oat genotype in the mixed effect models and/or had very low heritability scores. This could raise concerns as to the validity of these associations. When GWAS is applied to plant phenotypic traits such as yield and grain density, the variation in the trait is attributed to genotype, environment, and genotype by environment interactions. These plant encapsulated traits are controlled by plant genetics, changes in environmental conditions, and the way plant genes respond in different environmental conditions. When observing a plant phenotype based in manipulating a microbial community the perspective shifts. The structure of a fungal community varies vastly across environment, plant genetics can exert some control over this variation, and observing the effects of host genes on microbial taxa is highly dependent on the presence of those taxa in the first place.

Many of the fungal community traits evaluated in this study exhibited little to no heritability. Typically a trait with a very low heritability score would not be considered for further GWAS analysis, however, due to the large degree of variation present within and between locations and the low number of replications of each genotype, it may be unrealistic to expect significance in a genotypic effect for many fungal traits, regardless of their relationship to the host. In some cases the extra statistical power granted by tested for associations with genetic markers rather than genotypes could reveal important fungal relationships that are not resolved at the genotypic level. We argue that these SNP associations should still be considered as potentially relevant, especially those that are biologically coherent with the system. The relative abundance of Vishniacozyma, a yeast genus attributed to a saprotrophic lifestyle, was not significantly described by genotype but did significantly associate with two SNPs in the GWAS analysis. While this could be in error, a recent article describes this genus as being found in wheat kernels sampled from Canadian prairies and associating with increased kernel weight and resistance to

key pathogens (Vujanovic, 2021). While this by no means proves the significance of the SNPs associating with Vishniacozyma in this study, it does provide an argument that they should not be discounted outright due to the genus' non-significant association with oat genotype.

A major goal of this study was to test for an oat genotype specific soil legacy impact on next-season corn. While no corn phenotypes at either location associated with any oat SNPs above the semi-stringent significance threshold, variation in the Manhattan plots was evident. Especially at ARL, certain SNPs tested for association with yield traits approach or exceed a $-\log_{10}(p)$ score of 4.0. These SNPs could be candidates for further testing in their importance to oat soil legacy and its impact on subsequently planted crops. Despite the lack of SNP association, corn traits were significantly associated with the abundance of several fungal taxa in the estimated means models. As these models tested the estimated genotype means of fungal community traits (e.g. Shannon's diversity, abundances) against the oat genotype means of agronomic traits, their significance provides evidence for an oat soil legacy effect persisting even a year post oat harvest. While no single taxon generally described more than a few percent of the variation of a given corn phenotype, models containing several of the individually significant taxa reached R^2 values approaching 0.20.

While discussing the results of this study, it is important to concretely understand the temporal aspect of the experimental design. Because we sampled the oat rhizosphere community around the grain heading day, the fungal populations we observe could be influencing the growth of the oats and/or the success of the oats could be influencing the fungal communities. When it comes to the corn plant, however, the fungal populations we observe in our data set could not be influenced by the success of the corn plants as they were sampled one season prior to the planting of the corn. While much of the soil community likely changed in the interim time

between oat harvest and corn planting, the temporal aspect of this study allows us to have a clear understanding of the directionality of the relationships being tested. Still, there are multiple direct and indirect pathways by which oats and their associated rhizosphere communities could affect the growth of subsequently planted corn (Fig. 12). Many of the fungal taxa outlined in this study are either oat/corn pathogens or mutualists. It is possible oat genotypes could affect the productivity of next-season corn by contributing to an increase or decrease taxa that also have a detrimental or beneficial relationship with corn. Since no nutrients were added during either growing season, oat genotypes could also affect next-season corn by depleting nutrients, thereby leaving fewer resources for the corn plants to access in the future. Given these principals, it also follows the microbial community could affect the outcomes of future corn by affecting the nutrient pool in one of two ways: affecting the biomass of oat plants which could increase or decrease the amount of nutrients depleted during oat lifecycle, or directly depleting or adding to the nutrient pool at any point between oat rhizosphere and corn sampling through microbial processes such as replication and decomposition.

Taken together, the results of this study provide evidence for a heritable, genotype driven soil legacy effect in oats. One which is capable of influencing the agronomic traits of corn planted a full season after oat harvest. This study joins a short but growing list of studies showing the capabilities of GWAS in associating genomic regions with changes in host-associated microbial communities. As was observed in many of these studies, the inherent variation of microbial community populations posed a problem in the clarity of signals. Despite this GWAS provides a powerful tool in the exploration of significant plant-microbe interactions present in many crop plants, and future studies can continue to refine sampling methods, experimental design, and statistical models to reduce or account for this environmental variation.

Literature Cited

Asif, H., Alliey-Rodriguez, N., Keedy, S., Tamminga, C. A., Sweeney, J. A., Pearlson, G., Clementz, B. A., Keshavan, M. S., Buckley, P., Liu, C., Neale, B., & Gershon, E. S. (2021). GWAS significance thresholds for deep phenotyping studies can depend upon minor allele frequencies and sample size. *Molecular Psychiatry*, *26*(6), 2048–2055. https://doi.org/10.1038/s41380-020-0670-3

Bates D, Mächler M, Bolker B, Walker S (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, **67**(1), 1–48. doi:10.18637/jss.v067.i01.

Bender, S. F., Wagg, C., & Van Der Heijden, M. G. A. (2016). An Underground Revolution: Biodiversity and Soil Ecological Engineering for Agricultural Sustainability. *Trends in Ecology & Evolution*, *31*(6), 440–452. https://doi.org/10.1016/j.tree.2016.02.016

Deng, S., Caddell, D. F., Xu, G., Dahlen, L., Washington, L., Yang, J., & Coleman-Derr, D. (2021). Genome wide association study reveals plant loci controlling heritability of the rhizosphere microbiome. *ISME Journal*, *15*(11), 3181–3194. https://doi.org/10.1038/s41396-021-00993-z

Dowarah, B., Gill, S.S. & Agarwala, N. Arbuscular Mycorrhizal Fungi in Conferring Tolerance to Biotic Stresses in Plants. *J Plant Growth Regul* **41**, 1429–1444 (2022). https://doi.org/10.1007/s00344-021-10392-5

Ellouze, W., Hamel, C., Vujanovic, V., Gan, Y., Bouzid, S., & St-Arnaud, M. (2013). Chickpea genotypes shape the soil microbiome and affect the establishment of the subsequent durum wheat crop in the semiarid North American Great Plains. *Soil Biology and Biochemistry*, *63*, 129–141. https://doi.org/10.1016/j.soilbio.2013.04.001

Gardes M, Bruns TD. ITS primers with enhanced specificity for basidiomycetes--application to the identification of mycorrhizae and rusts. Mol Ecol. 1993 Apr;2(2):113-8. doi: 10.1111/j.1365-294x.1993.tb00005.x. PMID: 8180733.

Gutiérrez, L., Cuesta-Marcos, A., Castro, A. J., von Zitzewitz, J., Schmitt, M., & Hayes, P. M. (2011). Association Mapping of Malting Quality Quantitative Trait Loci in Winter Barley: Positive Signals from Small Germplasm Arrays. *The Plant Genome*, *4*(3), 256–272. https://doi.org/10.3835/plantgenome2011.07.0020

Johnson, N. C., Wilson, G. W. T., Wilson, J. A., Miller, R. M., & Bowker, M. A. (2015). Mycorrhizal phenotypes and the Law of the Minimum. *New Phytologist*, *205*(4), 1473–1484. https://doi.org/10.1111/nph.13172

Kiers E.T., Hutton J.G., and Denison R.F. 2007. Human selection and the relaxation of legume defences against ineffective rhizobia. Proc. R. Soc. B.2743119–3126

Kim, H., Lee, K. K., Jeon, J., Harris, W. A., & Lee, Y. H. (2020). Domestication of Oryza species eco-evolutionarily shapes bacterial and fungal communities in rice seed. *Microbiome*, *8*(1), 1–17. https://doi.org/10.1186/s40168-020-00805-0

Lankau, R. A., and D. P. Keymer. 2016. Ectomycorrhizal richness declines toward the host species' range edge Molecular Ecology **25**:3224-2341.

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S., & Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*, *28*(18), 2397–2399. https://doi.org/10.1093/bioinformatics/bts444

Martín-Robles, N., Lehmann, A., Seco, E., Aroca, R., Rillig, M. C., & Milla, R. (2018). Impacts of domestication on the arbuscular mycorrhizal symbiosis of 27 crop species. *New Phytologist*, *218*(1), 322–334. https://doi.org/10.1111/nph.14962

Oksanen, A. J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P. R., Hara, R. B. O., Simpson, G. L., Solymos, P., Stevens, M. H. H., & Szoecs, E. 2018. *Package ' vegan .' April*.

Porter, S. S., and J. L. Sachs. 2020. Agriculture and the Disruption of Plant-Microbial Symbiosis. Trends in Ecology & Evolution **35**:426-439.

S. R. Searle, F. M. Speed & G. A. Milliken (1980) Population Marginal Means in the Linear Model: An Alternative to Least Squares Means, The American Statistician, 34:4, 216-221, DOI: 10.1080/00031305.1980.10483031

Walters, W. A., Jin, Z., Youngblut, N., Wallace, J. G., Sutter, J., Zhang, W., González-Peña, A., Peiffer, J., Koren, O., Shi, Q., Knight, R., del Rio, T. G., Tringe, S. G., Buckler, E. S., Dangl, J. L., & Ley, R. E. (2018). Large-scale replicated field study of maize rhizosphere identifies heritable microbes. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(28), 7368–7373. https://doi.org/10.1073/pnas.1800918115

Wang, Y., Wang, X., Sun, S. *et al.* GWAS, MWAS and mGWAS provide insights into precision agriculture based on genotype-dependent microbial effects in foxtail millet. *Nat Commun* **13**, 5913 (2022). https://doi.org/10.1038/s41467-022-33238-4

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

White, T. J., T. Bruns, S. Lee, and J. W. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pages 315-325 *in* M. A. Innis, D. H. Gelfand, J. J. Snisky, and T. J. White, editors. PCR Protocols: A Guide to Methods and Applications. Academic Press, Inc., New York.

Vujanovic, V. (2021). Tremellomycetes yeasts in kernel ecological niche: Early indicators of enhanced competitiveness of endophytic and mycoparasitic symbionts against wheat pathobiota. *Plants*, *10*(5). https://doi.org/10.3390/plants10050905

Table 1. Heritabilities of core oat associates

Trait (Relative Abundance)	Level	Lifestyle Guild	Location	Geno (p)	Heritability
Rozellomycotina	Class	Unidentified	ARL	1.42E-06	0.3001
Sordariomycetes	Class	Unidentified	ARL	0.02020	0.1788
Chaetosphaeriaceae	Family	Unidentified	ARL	0.00833	0.2273
Lasiosphaeriaceae	Family	Unidentified	ARL	0.00224	0.1536
Ilyonectria	Genus	Plant Pathogen	ARL	0.00167	0.2835
Leptosphaeria	Genus	Plant Pathogen	ARL	0.02325	0.1567
Nectria	Genus	Plant Pathogen	ARL	0.03172	0.1899
Mortierella	Genus	Soil Saprotroph	ARL	0.04316	0.1419
Acremonium	Genus	Unspecified Saprotroph	ARL	0.03665	0.1718
Humicola	Genus	Wood Saprotroph	ARL	0.09004	0.1305
Glomeraceae	Family	Arbuscular Mycorrhizae	WMAD	0.00001	0.2777
Serendipitaceae	Family	Unidentified	WMAD	0.07252	0.1395
Podospora	Genus	Dung Saprotroph	WMAD	0.09288	0.0998
Trichoderma	Genus	Mycoparasite	WMAD	0.00002	0.1872
Pyrenophora	Genus	Plant Pathogen	WMAD	0.02326	0.1984
Ustilago	Genus	Plant Pathogen	WMAD	0.01873	0.1812
Coprinopsis	Genus	Soil Saprotroph	WMAD	0.00520	0.2173
Mortierella	Genus	Soil Saprotroph	WMAD	0.05722	0.1194
Coniochaeta	Genus	Unspecified Saprotroph	WMAD	0.08371	0.0902
Talaromyces	Genus	Unspecified Saprotroph	WMAD	0.00167	0.2881
Tremellodendropsidales	Order	Unidentified	WMAD	0.04511	0.0974
Xylariales	Order	Unidentified	WMAD	0.00004	0.1469
Glomeromycota	Phylum	Arbuscular Mycorrhizae	WMAD	0.01451	0.1678
Ascomycota	Phylum	Unidentified	WMAD	0.03578	0.1112
Arbuscular Mycorrhizae	Guild	Arbuscular Mycorrhizae	ARL	0.09977	0.1367
Litter Saprotrophs	Guild	Litter Saprotroph	WMAD	0.02188	0.1803
Soil Saprotrophs	Guild	Soil Saprotroph	WMAD	0.07921	0.1204
Unidentified	Guild	Unidentified	WMAD	0.02532	0.1880
Unspecified Saprotrophs	Guild	Unspecified Saprotroph	WMAD	0.06165	0.1419

Summary of statistical tests determining fungal association with oat genotype. The fungal taxa/guilds listed here were tested from a pool of 123 genera and 8 guilds that met criteria to be considered part of the “core” oat microbiome (presence in >33% of samples, and >0.1% average relative abundance). “Geno (p)” shows the p-value of oat genotype in a mixed effect model testing relative abundance against spatial variables as random effects and oat genotype as a fixed effect. “Heritability” is calculated from the output of a random effect model containing spatial values and oat genotype as random effects, and denotes the % of variation in the trait that can be ascribed to host genetics.

Table 2. Summary of fungal taxa that associated with one or more SNPs

Trait (Relative Abundance)	Level	Lifestyle Guild	Location	Loc (p)	Geno (p)	Heritability	%G	%E	%GxE	Significant SNPs
Hypocreales	Order	Unidentified	Both	0.0012	0.1065	0.1373	2.3055	2.8853	0.1085	1
			WMAD		0.3286	0.0826	2.6935			1
			ARL		0.1894	0.1228	3.4783			0
Vishniacozyma	Genus	Soil Saprotroph	Both	0.0176	0.7039	0.0000	0.0000	2.7102	0.0000	1
			WMAD		0.7303	0.0000	0.0000			1
			ARL		0.7399	0.0000	0.0000			0
Plenodomus	Genus	Plant Pathogen	Both	0.0193	0.4481	0.0023	0.0331	8.7412	0.5814	0
			WMAD		0.3393	0.0214	0.6819			1
			ARL		0.3760	0.0857	3.0165			0
Coprinellus	Genus	Soil Saprotroph	Both	0.3892	0.5487	0.0000	0.0000	0.0000	0.0000	0
			WMAD		0.5688	0.0000	0.0000			1
			ARL		0.0032	0.0000	0.0000			0
Udeniozyma	Genus	Unspecified Saprotroph	Both	0.0050	0.3854	0.0000	0.0000	11.0202	4.2295	0
			WMAD		0.2517	0.1376	4.7785			1
			ARL		0.9460	0.0000	0.0000			0
Orbiliaceae	Family	Unidentified	Both	4.44E-12	0.2605	0.0014	0.0198	16.2276	0.7987	0
			WMAD		0.0881	0.0304	1.0020			1
			ARL		0.7448	0.0000	0.0000			0
Ustilago	Genus	Plant Pathogen	Both	0.0238	0.0005	0.2375	4.7211	2.1452	1.1428	0
			WMAD		0.0187	0.1812	6.5415			1
			ARL		0.3527	0.0661	2.2121			0
Mrakia	Genus	Unspecified Saprotroph	Both	2.15E-08	0.3275	0.0000	0.0000	8.8194	0.0000	0
			WMAD		0.3856	0.0000	0.0000			3
			ARL		0.2761	0.0000	0.0000			0
Rozellomycotina	Class	Unidentified	Both	9.22E-10	0.0017	0.0168	0.3176	8.8592	10.8134	0
			WMAD		0.7109	0.0000	0.0000			0
			ARL		1.42E-06	0.3001	12.3336			1
Funneliformis	Genus	Arbuscular Mycorrhizae	Both	4.56E-08	0.1009	0.0482	0.7508	10.0974	2.1308	0
			WMAD		0.0799	0.0000	0.0000			0
			ARL		0.1375	0.1144	3.9146			2

Summary of statistical tests determining fungal association with oat genotype for the genera that successfully associated with one or more SNP at either location or across both locations. Summary of statistical tests determining fungal association with oat genotype. The fungal taxa/guilds listed here were tested from a pool of 123 genera and 8 guilds that met criteria to be considered part of the "core" oat microbiome (presence in >33% of samples, and >0.1% average relative abundance). "Geno (p)" shows the p-value of oat genotype in a mixed effect model testing relative abundance against spatial variables as random effects and oat genotype as a fixed effect. "Heritability" is calculated from the output of a random effect model containing spatial values and oat genotype as random effects, and denotes the % of variation in the trait that can be ascribed to host genetics. The output of the random effect model is shown in "%G", "%E", and "%GxE" which show the % of trait variation described by genotype, environment, and interaction between genotype and environment, respectively.

Table 3. Heritabilities and SNP association of corn outcomes and oat yield

Trait	**Location**	**Loc (p)**	**Geno (p)**	**Heritability**	**Significant SNPs**
Oat Plot Yield (g)	Both	0.011	2.00E-16	0.676	1
	WMAD		2.20E-16	0.826	2
	ARL		2.20E-16	0.913	1
Corn V6 Weight (g)	Both	0.623	0.530	0.000	0
	WMAD		0.495	0.000	0
	ARL		0.423	0.006	0
Corn R1 Weight (g)	Both	0.00347	0.145	0.078	0
	WMAD		0.140	0.107	0
	ARL		0.741	0.000	0
Corn Total Yield (g)	Both	0.0072	0.621	0.000	0
	WMAD		0.824	0.000	0
	ARL		0.828	0.000	0

Summary of statistical tests determining genotype and SNP association with corn phenotypes and oat plot yield (as a comparison point). “Geno (p)” shows the p-value of oat genotype in a mixed effect model testing relative abundance against spatial variables as random effects and oat genotype as a fixed effect. “Heritability” is calculated from the output of a random effect model containing spatial values and oat genotype as random effects, and denotes the % of variation in the trait that can be ascribed to host genetics.

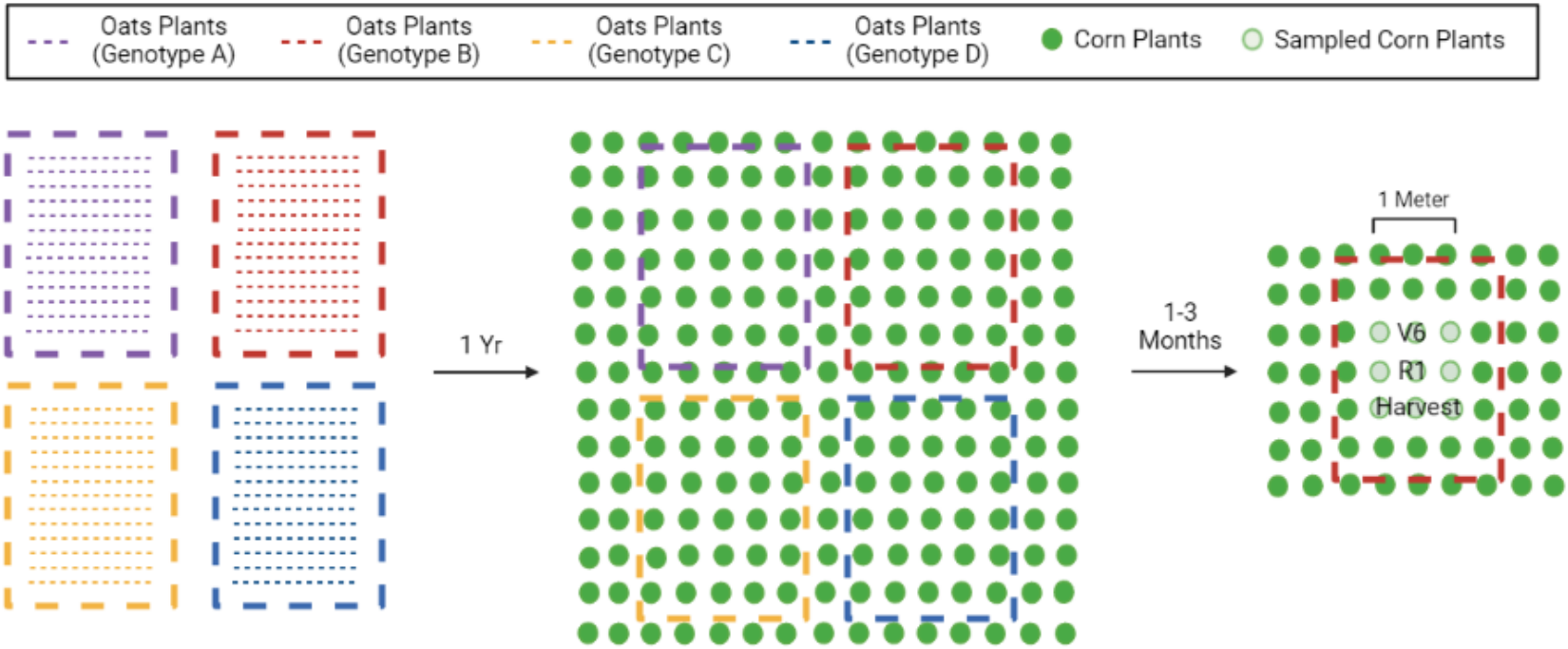

Figure 1.

Experimental design of genome wide association study. 312 unique oat genotypes were planted at each of two locations: The West Madison Agricultural Research Station (WMAD) and the Arlington Agricultural Research Station (ARL). Each genotype was grown in at least three replicated plots, with eight genotypes being grown in six replicated plots for a total of 960 plots at each location (1,920 plots total). The following season, an elite variety of corn (*Zea mays*) was freely seeded into the oat plots to observe potential oat genotype specific soil legacy effects. To understand the effect of microbial community throughout the corn lifecycle, corn plants were sampled from each location at the V6 and R1 stages as well as at harvest. At the V6 and R1 sampling periods, one meter of corn was hand cut from a row at the middle-most point in each oat plot marked from the previous season.

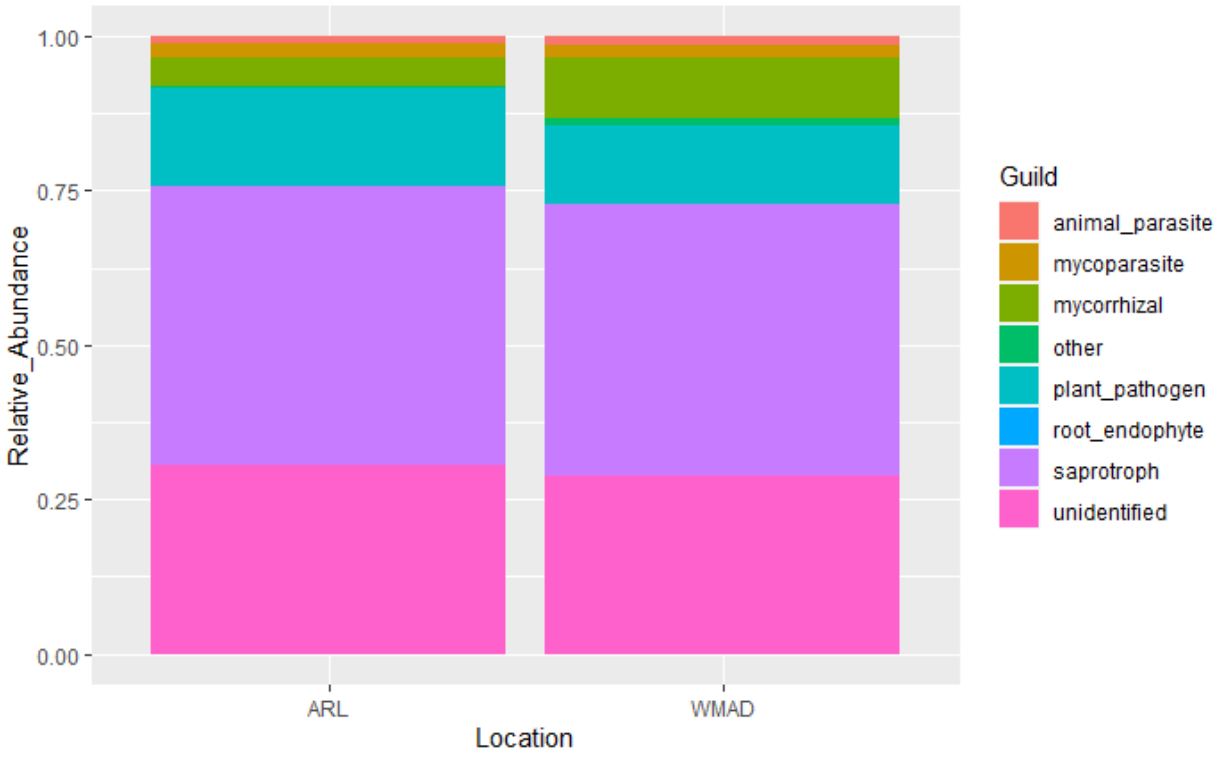

Figure 2. Stacked bar chart showing guild level variation between ARL and WMAD locations. Guild populations were compared using relative abundance of fungal ASVs.

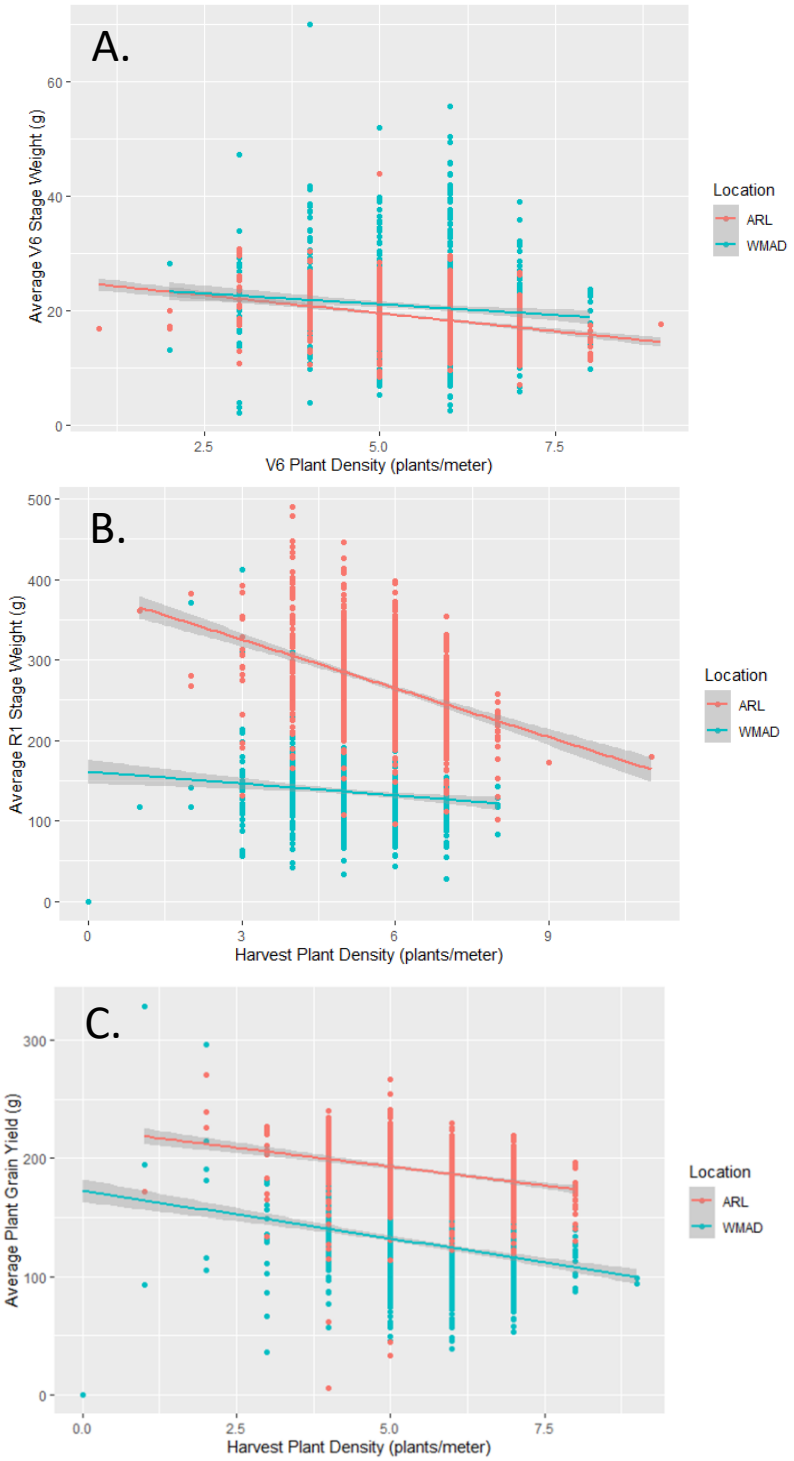

Figure 3. Correlation between plant density and average corn plant weight at three stages; A) V6, B) R1, and C) Harvest.

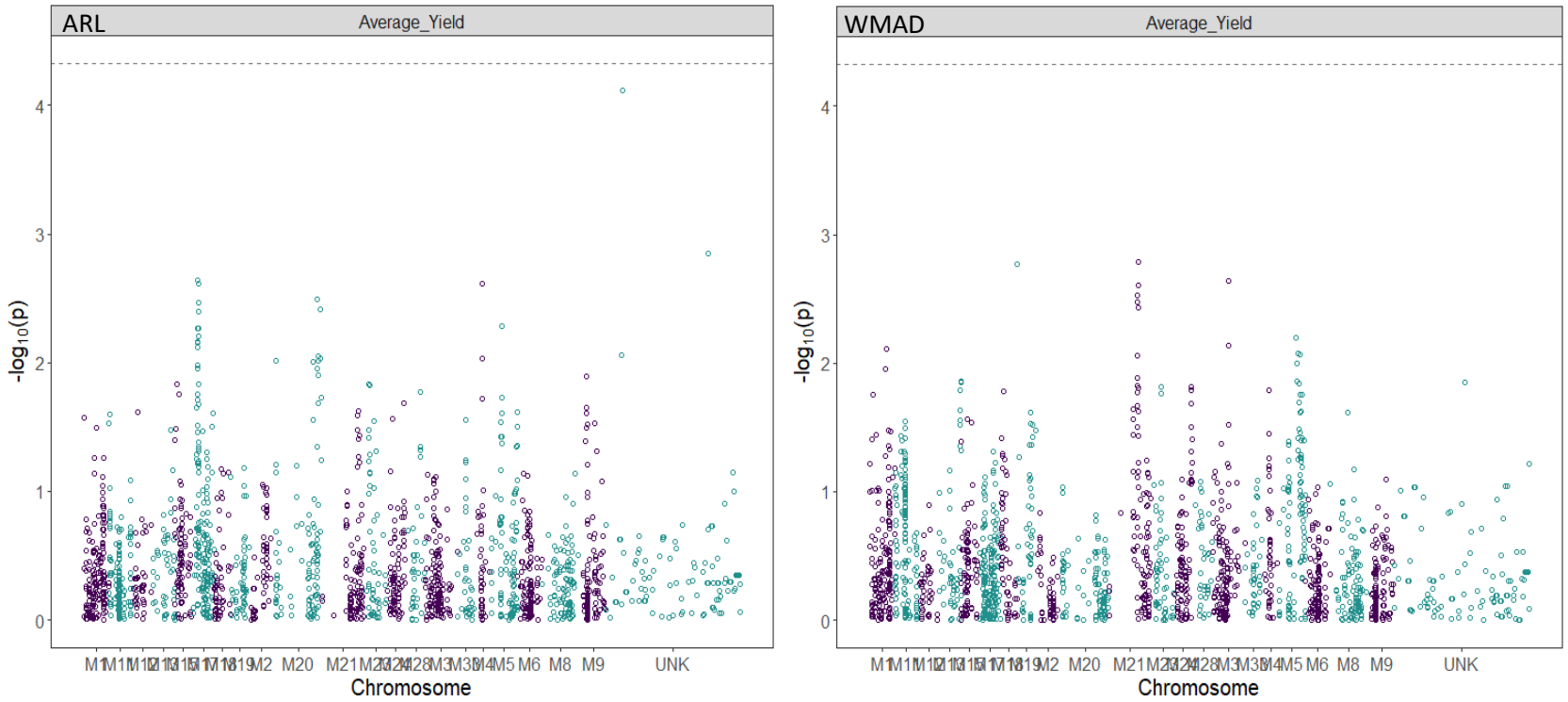

Figure 4. Manhattan plot showing SNP associations with average corn yield at ARL and WMAD. Each SNP is denoted by an individual dot on the figure. SNPs that appear above the dotted line have exceeded the significance threshold (shown on y-axis as the negative log p-value) and are considered to significantly associate with the tested trait.

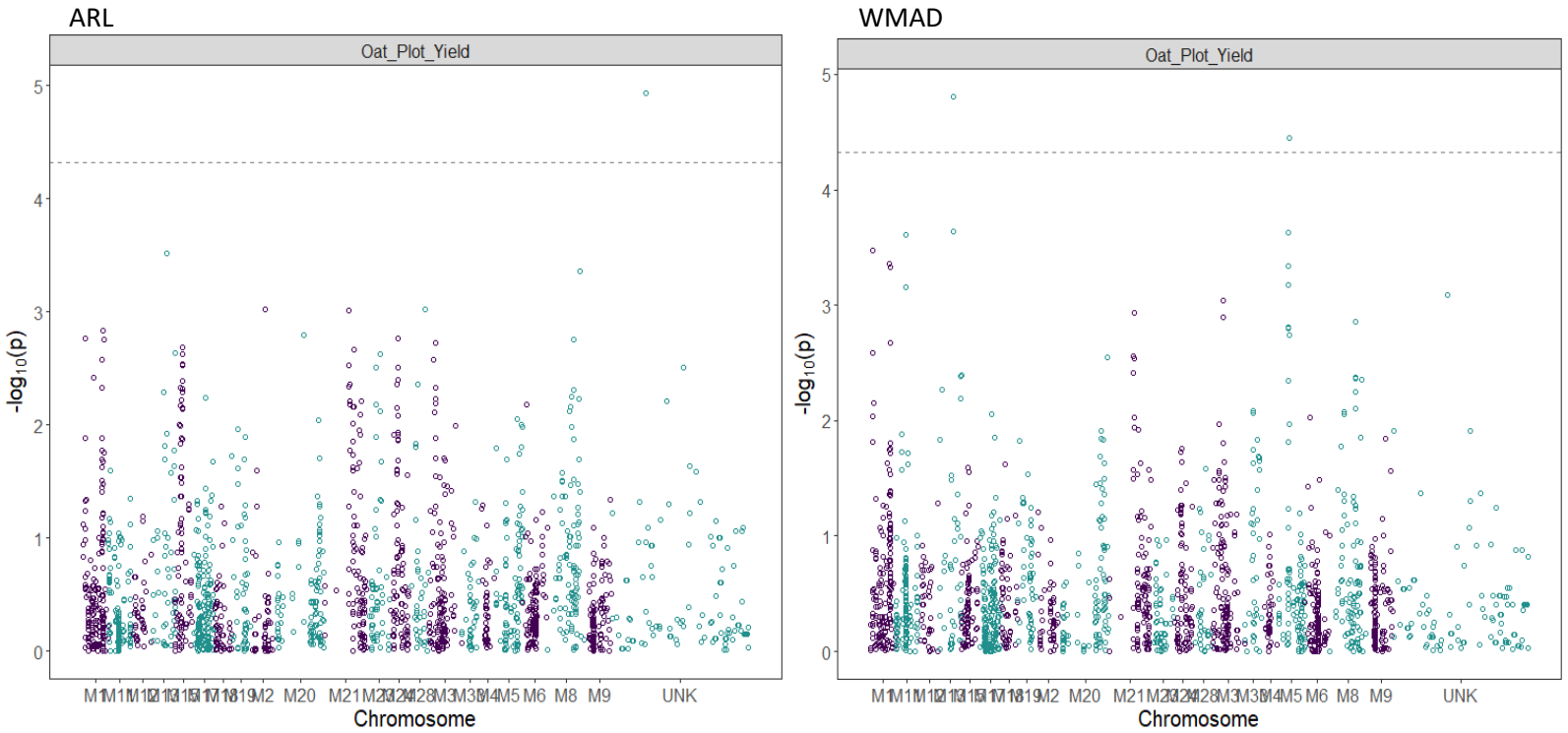

Figure 5. Manhattan plot showing SNP associations with oat plot yield at ARL and WMAD. Each SNP is denoted by an individual dot on the figure. SNPs that appear above the dotted line have exceeded the significance threshold (shown on y-axis as the negative log p-value) and are considered to significantly associate with the tested trait.

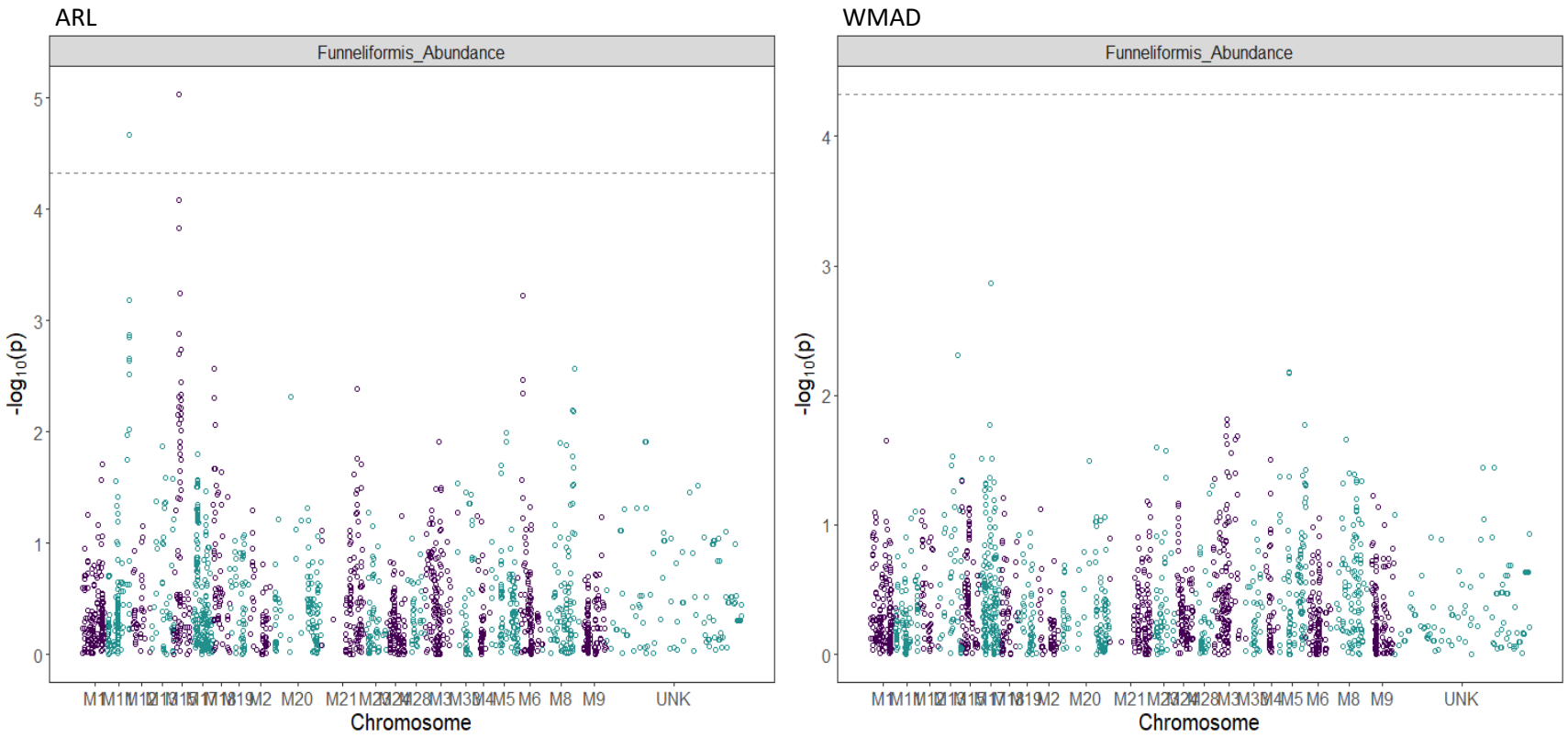

Figure 6. Manhattan plot showing SNP associations with Funneliformis relative abundance at ARL and WMAD. Each SNP is denoted by an individual dot on the figure. SNPs that appear above the dotted line have exceeded the significance threshold (shown on y-axis as the negative log p-value) and are considered to significantly associate with the tested trait.

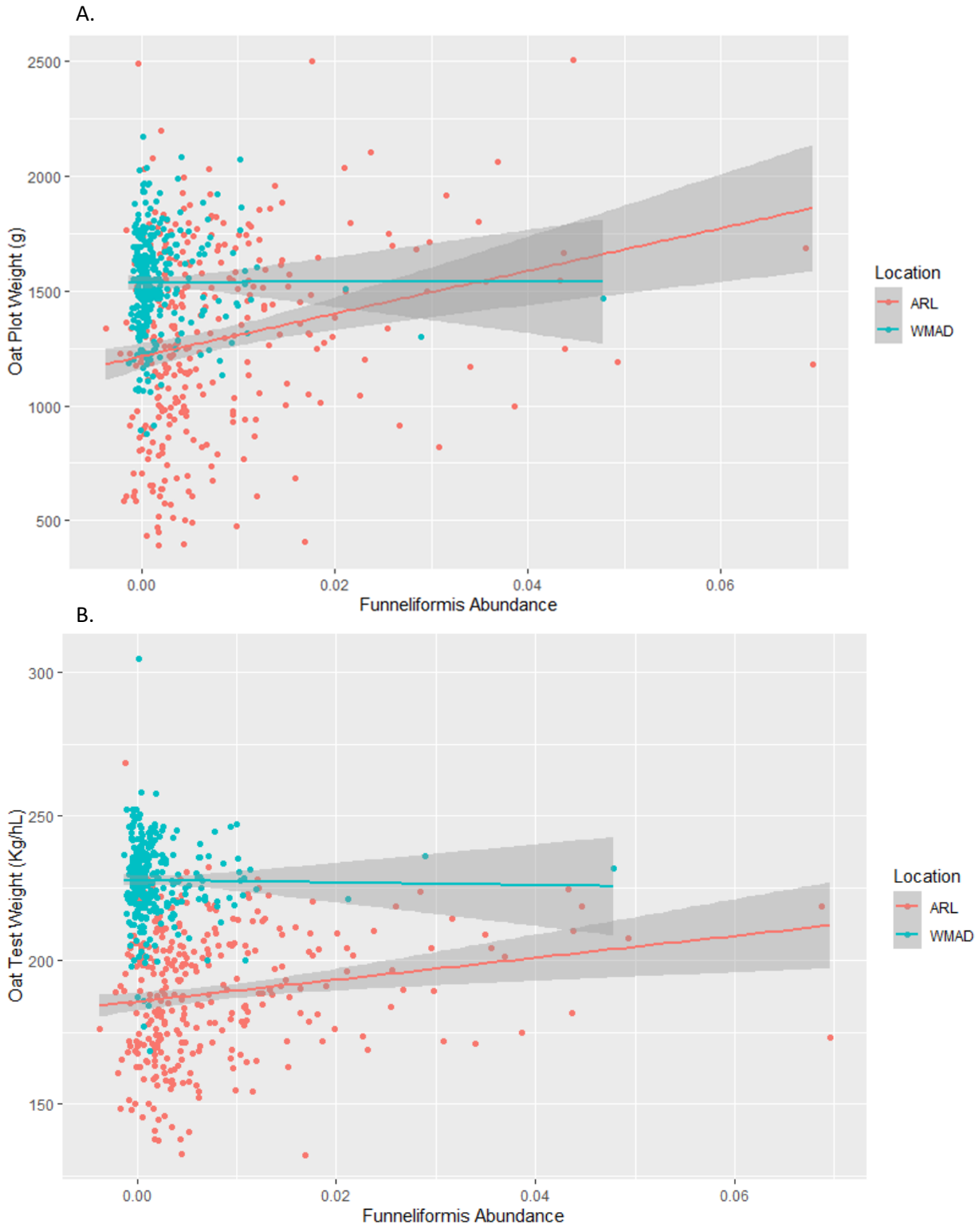

Figure 7. Scatterplot of genotype mean Funneliformis relative abundance vs. genotype means of A) Oat plot weight (g) and B) Oat Test Weight (Kg/hL).

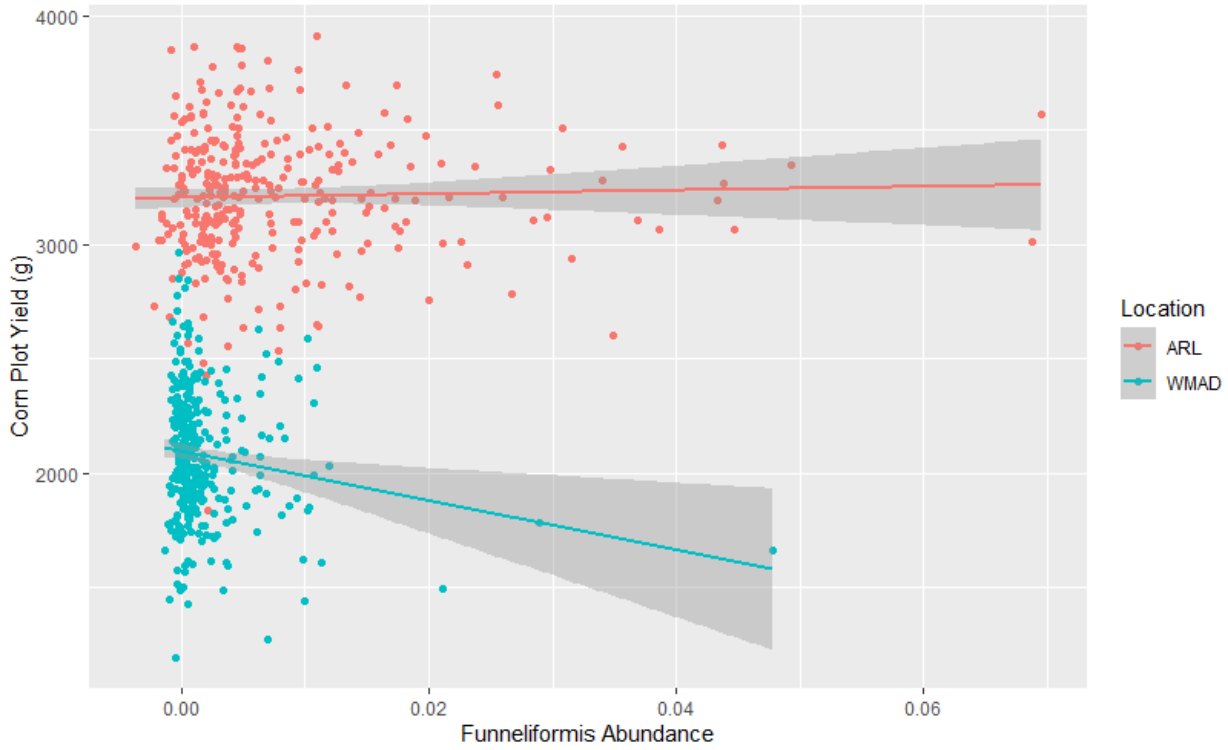

Figure 8. Scatterplot of genotype mean Funneliformis relative abundance vs. genotype mean of total corn plot yield (g).

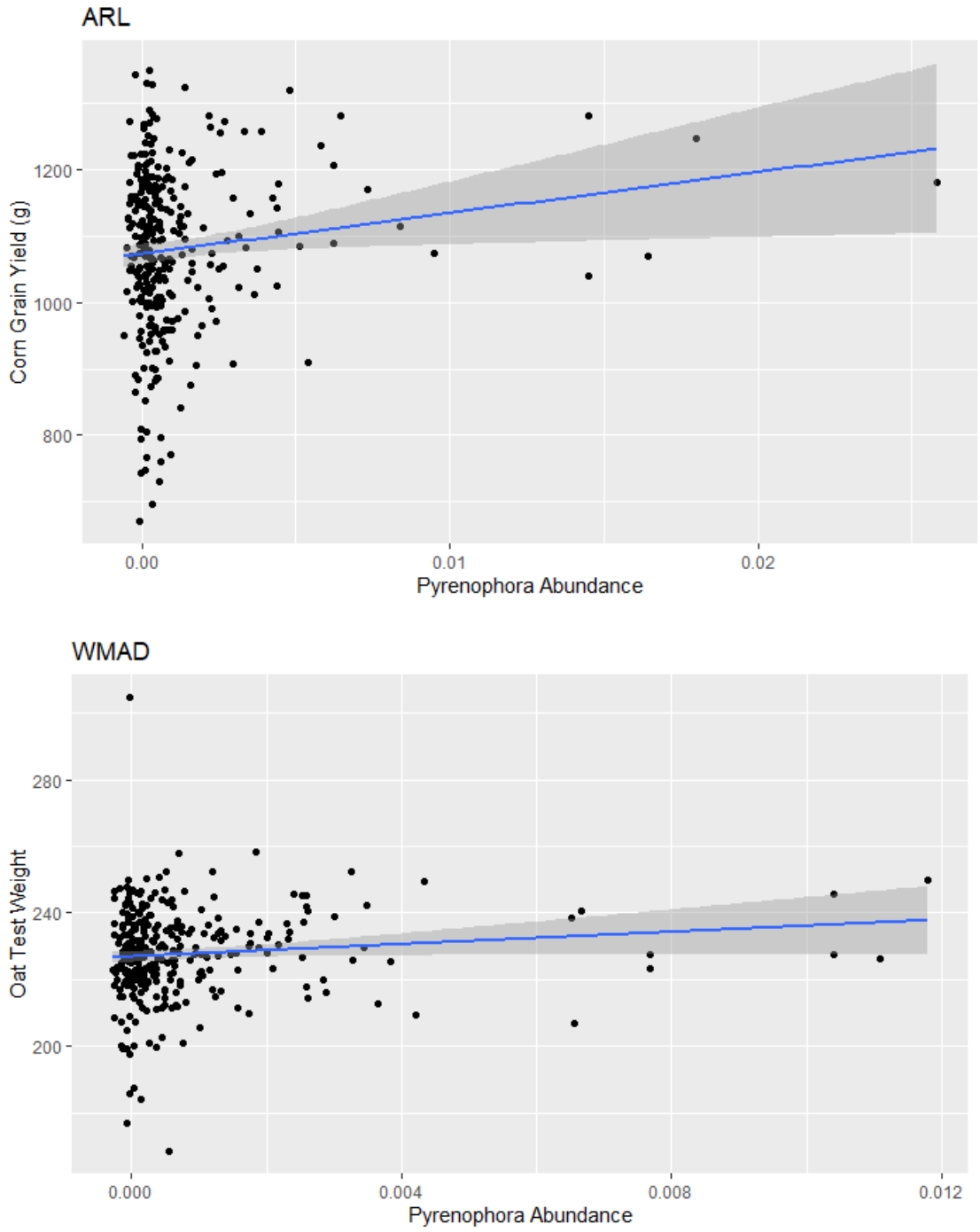

Figure 9. Scatterplot of genotype mean Pyrenophora relative abundance vs. genotype means of A) total corn plot yield (g) at ARL and B) oat test weight (Kg/hL) at WMAD.

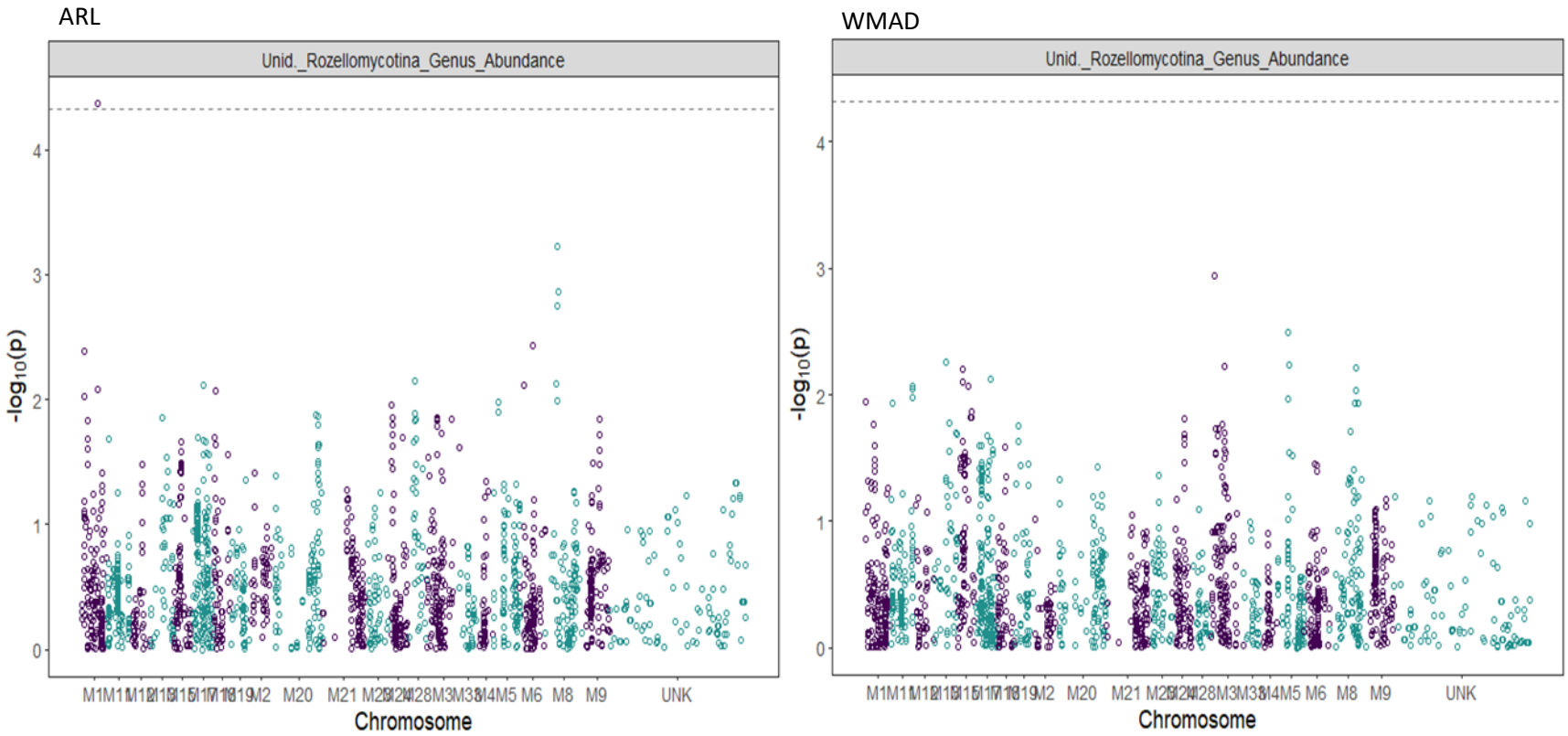

Figure 10. Manhattan plot showing SNP associations with the relative abundance of an identified Rozellomycotina genus at ARL and WMAD. Each SNP is denoted by an individual dot on the figure. SNPs that appear above the dotted line have exceeded the significance threshold (shown on y-axis as the negative log p-value) and are considered to significantly associate with the tested trait.

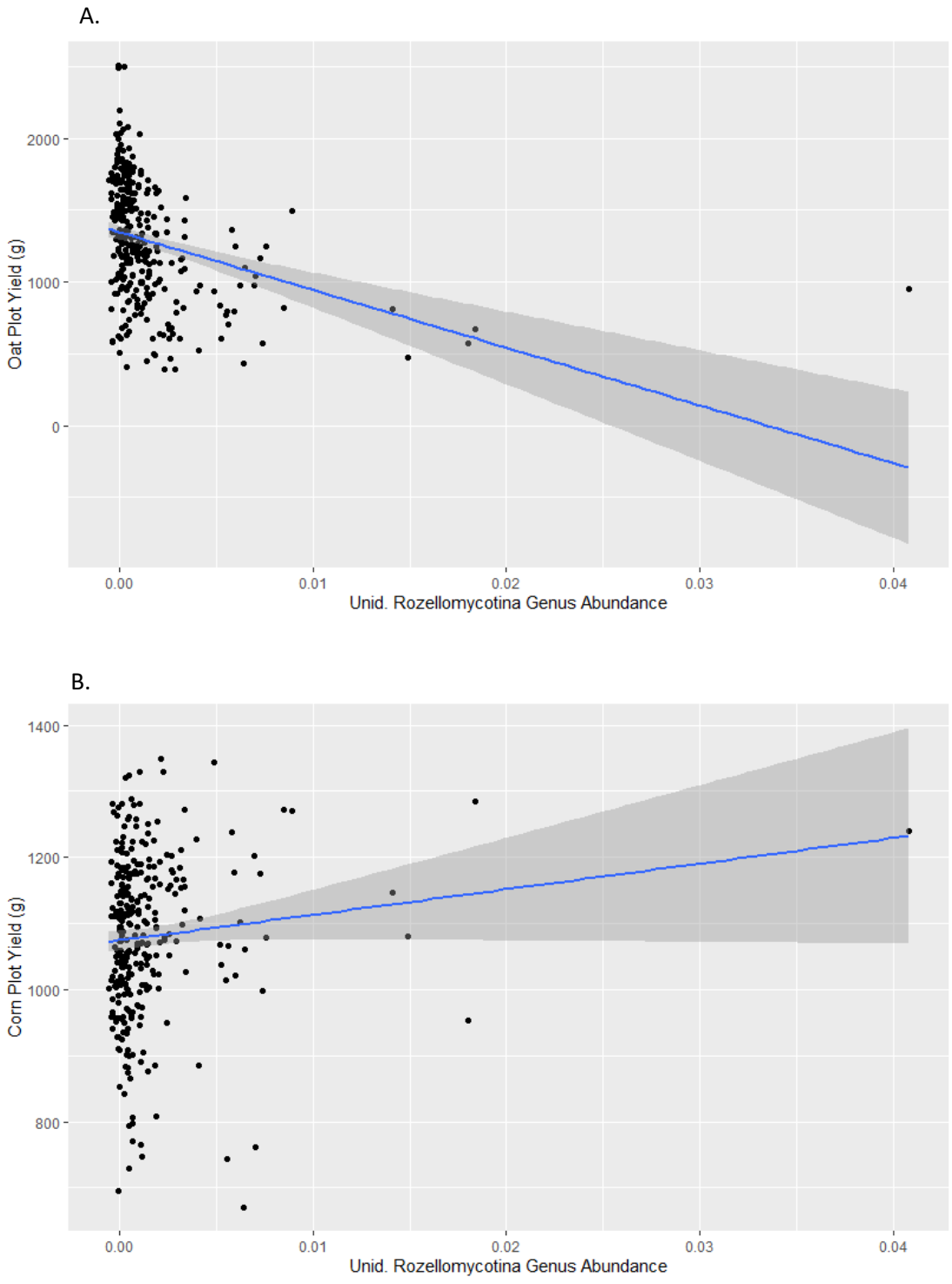

Figure 11. Scatterplot of genotype mean unidentified Rozellomycotina genus relative abundance vs. genotype means of A) oat plot weight (g) and B) total corn plot yield (g) at ARL.

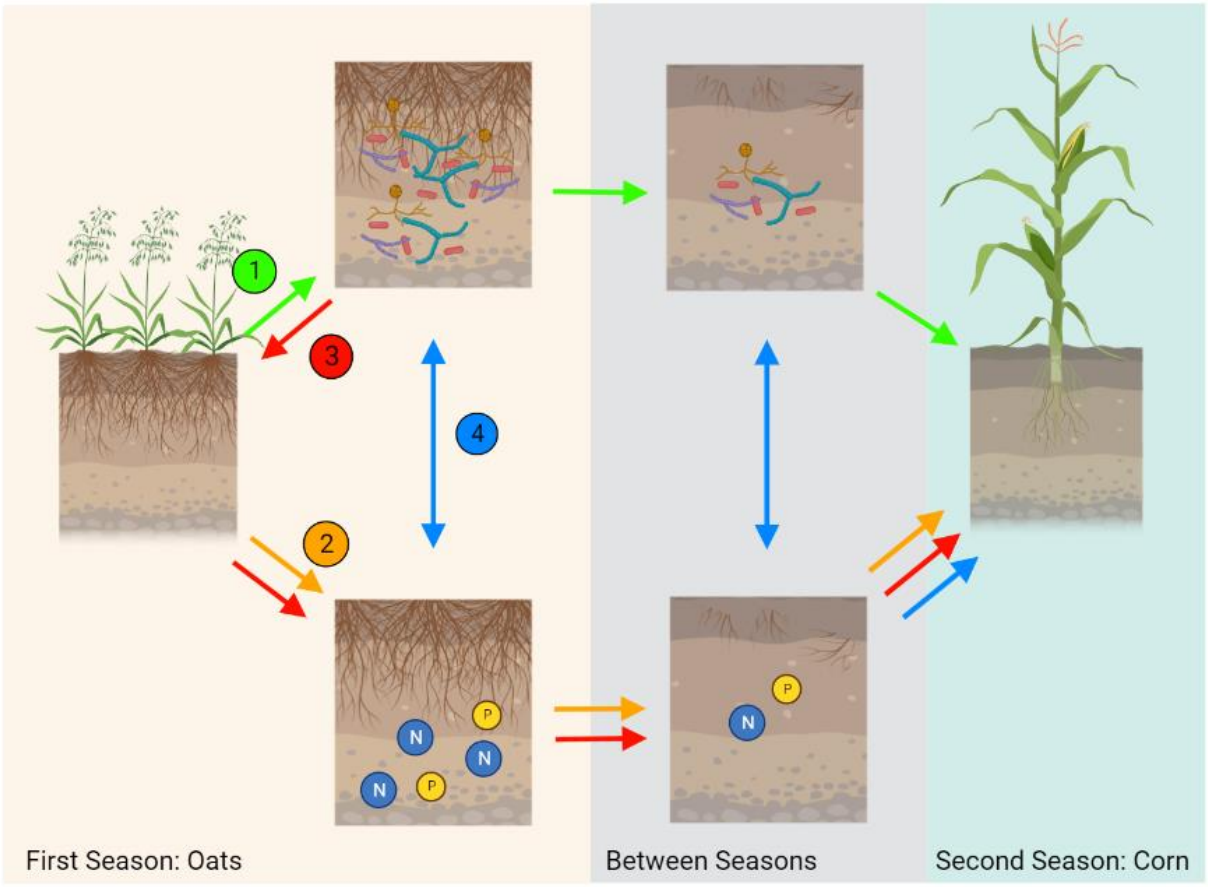

Figure 12. Model depicting potential pathways by which oat soil legacies can affect next-season corn. Oat genotypes can influence next-season corn traits by 1) manipulating the microbial soil communities, 2) directly altering the nutrient pool by differential uptake, 3) being altered by existing microbial communities in a way that increases or decreases nutrient uptake, and 4) indirectly altering the nutrient pool by changing microbial populations that contribute to nutrient uptake or cycling.

Chapter 5

Conclusions: Exploring and capitalizing on genotypic variation of the oat microbiome

Throughout this thesis I have outlined results that explore the impacts of the oat-associated soil microbial community on the agronomic outcomes of oat hosts and next-season corn. This oat soil legacy effect was studied in the greenhouse and at the field level.

In chapter 2, I described a study in which we sequenced soil from nearby oat roots of 8 different genotypes replicated across three different spatial blocks in a field experiment at West Madison Agricultural research station. We inoculated oat and corn plants under high and low nutrient conditions with these 24 soils in a greenhouse potting experiment. From this study we found that oat genotypes exhibit detectable differences in their associated soil-microbial communities. We also found that these different genotype associated microbial communities have differential impacts on the productivity of inoculated oat and corn hosts. Specifically, oat hosts were generally more impacted by genotype soil community under low nutrient conditions while corn hosts were more impacted under high nutrient conditions. We also found that oat genotype describes significant variation in the diversity of the associated microbial community and differences diversity correlates with oat above and below ground biomass. This study provided significant evidence for an genotype specific soil legacy effect for subsequently planted oats and corn.

In chapter 3, I described a study in which we studied this oat genotype-specific soil legacy at the field level by sequencing oat rhizosphere communities from every plot of a genome-wide association study consisting of 312 genotypes replicated at least three times at each of two locations. In this chapter I described our use of LASSO regression, a statistical modeling tool

that allowed us to use oat-associated fungal community data to make predictions about oat and corn agronomic outcomes. The results from this study showed that we could successfully use oat genotype associated fungal communities to predict oat yield outcomes at ARL, and we could use plot level fungal communities to predict corn yield outcomes at WMAD. These results also allowed us to identify some candidate fungal associates of oats and corn that merit further investigation in future studies.

In chapter 4, I describe a different facet of the same genome-wide association study. In this chapter I discuss the heritabilities of core oat rhizosphere community members and the oat genome SNPs that associate with the variation in relative abundances of some of these members. The results of this study allowed us to identify oat genomic regions that associate with known and unknown members of the oat rhizosphere community, and identify fungal community members that associate with oat genotype, but in a more quantitative multigenic fashion. This study opens up the possibility for breeding oats to increase association with certain fungal community members, but did not provide evidence for an oat genotype specific soil legacy effect on subsequently planted corn.

Overall, this study showed evidence of an oat genotype-specific soil legacy effect in corn at in a controlled greenhouse study, but this effect was lost at the field level. In the end, the goal of breeding oats to cultivate a better microbial community for next-season corn seems unobtainable. Both because of the lack of soil legacy effects at the field level and the lack of overlap between important oat and corn community members detected in the LASSO models. Be that as it may, this series of studies still provides fascinating insights into the genetic relationships between oats and oat rhizosphere communities. As interest in lower input agricultural systems increases, studies like this one can identify breeding targets for improving crop association with beneficial

microbes, and identify previously unknown microbes that are important to crop health. Surprising finds, like oat SNP association with arbuscular mycorrhizal fungi can change our understanding of these relationships and allow us to adjust our input strategies based on genotype and environmental context.