Title
Host Genetic Regulation of Root-Associated Bacterial Communities in Sorghum and A Case Study of Manipulating Agricultural Microbiomes with Biologicals

Author
Deng, Siwen

Publication Date
2019

Peer reviewed|Thesis/dissertation

Host Genetic Regulation of Root-Associated Bacterial Communities in Sorghum and A Case Study of Manipulating Agricultural Microbiomes with Biologicals

By

Siwen Deng

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Plant Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Devin Coleman-Derr, Co-Chair
Professor Sarah C. Hake, Co-Chair
Professor Frank G. Harmon
Professor Mary K. Firestone

Fall 2019

Host Genetic Regulation of Root-Associated Bacterial Communities in Sorghum and A Case Study of Manipulating Agricultural Microbiomes with Biologicals

Siwen Deng

Abstract

Host Genetic Regulation of Root-Associated Bacterial Communities in Sorghum And A Case Study of Manipulating Agricultural Microbiomes with Biologicals

by

Siwen Deng

Doctor of Philosophy in Plant Biology

University of California, Berkeley

Professor Devin Coleman-Derr, Co-Chair

Professor Sarah C. Hake, Co-Chair

Recent work suggests that plants have a plethora of previously underappreciated microbial allies that help them during their growth and development. The close interactions between plants and their complex microbial communities, otherwise known as the plant microbiome, play essential roles in plant health and productivity (Turner, James and Poole, 2013). Recent studies in various plant species have discovered multiple factors that shape the plant root microbiome, including geographical location, soil source, host genotype, and cultivation practice (Peiffer *et al.*, 2013; Edwards *et al.*, 2015; Naylor *et al.*, 2017; Xu *et al.*, 2018). Accumulating evidence suggests an interaction between host genetics and plant microbiome composition (Horton *et al.*, 2014; Naylor *et al.*, 2017; Fitzpatrick *et al.*, 2018; Walters *et al.*, 2018). However, identifying specific mechanisms driving microbiome acquisition and assembly, as well as the host genetic variants involved in these processes, has proved challenging. A few groups have delved into the impact of individual genes on microbiome composition (Lebeis *et al.*, 2015; Castrillo *et al.*, 2017). However, these studies were guided by *a priori* hypotheses of gene involvement. We believe that additional research is needed to better understand the host genetic regulation of plant-associated microbes within a naturally occurring ecological context.

Here I describe how we utilize recent advances in sequencing technology, quantitative genetics, statistical models, and data mining in conjunction with a successful sorghum diversity panel (Casa *et al.*, 2008) to perform a population-level microbiome study using large scale genetic approaches. In this experiment, we used 200 diverse sorghum ecotypes and planted them in a replicated field trial. Our experimental design allowed us to control the environmental effect on associated microbiome composition in order to better resolve the impacts of host genetics. After examining multiple sample types, including the leaf, root, and rhizosphere, we first demonstrated that for sorghum plants, the rhizosphere represents the most suitable fraction of the plant microbiome for studying the host genetic effect, as compared to the root and leaf. We calculated broad

sense heritability scores of the rhizosphere-associated microbial partners across all of the 200 sorghum diverse lines. We then identified which rhizosphere microbes exhibit reproducible associations with specific sorghum genotypes; these included members of the microbial orders *Verrucomicrobiales*, *Flavobacteriales*, *Planctomycetales*, and *Burkholderiales*. We also demonstrated that the heritable taxa in sorghum show a strong degree of overlap with heritable lineages previously identified in maize. To gain further insight into the host genetic regulation that is responsible for determining root microbiome variation, we employed a Genome-Wide Association Study (GWAS) approach, which allows us to rapidly scan genetic markers across the complete sets of DNA of many sorghum germplasm to find genetic variation associated with particular microbiome-related traits. We identified multiple plant loci that are associated with variation in the sorghum root microbiome. Furthermore, we demonstrated that GWAS could be used as a non-candidate approach to predict microbiome structure based solely on host genetic information. Collectively, this work demonstrates the utility of GWAS for analysis of host-mediated control of rhizosphere microbiome phenotypes and advances our knowledge of the relationship between the plant microbiome and host genetics control.

Finally, I describe our work on testing the effect of a commercially available soil amendment on agricultural soil and strawberry root bacterial microbiome. This final study highlights the main overarching significance of the above mentioned studies: understanding host genetic regulation of the plant root microbiome can provide insights into how to develop new and improved microbial formulations that are able to enhance crop productivity under a range of biotic and abiotic factors.

Content

List of Figures

List of Tables

Acknowledgments

This work would not have been possible without the support of many people. I want to take the opportunity here to express my sincere appreciation to all of them.

First and foremost, I would like to express my deepest gratitude to my advisor, Professor Devin Coleman-Derr, for being an amazing mentor, a wonderful teacher, a trustworthy colleague, and an inspiring friend. I'm extremely grateful to be your first student. Thank you for providing me extensive guidance and for teaching me a great deal about both scientific research and life in general. Thank you for your patience from which my scientific endeavor bore fruition and all of the encouragement that enabled me to discover my research interest in my pursuit of becoming a scientist.

I would also like to extend my gratitude to the professors at Berkeley who provided mentorship during my time in graduate school. Special thanks to Prof. Sarah Hake, Prof. Frank Harmon, and Prof. Mary Firestone for serving on my qualifying exam committee and dissertation committee, the support and guidance of whom helped me accomplish this work. They have modeled the qualities of good scientists. Thanks should also go to Prof. Louise Glass and Prof. Henrik Scheller for being PI mentors in my rotations, Prof. Jennifer Fletcher for being my professor for my first GSI assignment and support in my qualifying exam, and Prof. Patricia Zambryski for all of her help in my GSI stint and for nominating me as an outstanding GSI. Thanks should also go to Prof. Peggy Lemaux for all her support in my scientific communication with the CLEAR group and my future career goals.

I would also like to thank all my awesome lab mates Dr. Ling Xu, Dr. Daniel Caddell, Edi Wipf, Tuesday Simmons, Grady Pierroz, Alex Styer, Dr. Feng Yu, Dr. Zhaobin Dong, Lorenzo Washington, Laxmi Mishra, Dr. Tui Ray, and Dr. Dan Naylor for their generous support and encouragement. In particular, I would like to thank Edi Wipf for being a dear friend and companion, for giving me so many wonderful memories, for being supportive all the time, and for working with me on the VESTA project. I would like to extend my sincere gratitude to Dr. Daniel Caddell for his relentless support and invaluable contribution, for writing multiple papers together, for the humorous and thoughtful discussions that elevated the intellectual plane of the projects. Lastly, I would like to thank Dr. Ling Xu for her invaluable contribution to many projects, for her unrelenting support, and for becoming my sister.

I gratefully acknowledge all of those with whom I have had the pleasure to work with during my Ph.D. study — all of the staff, professors, postdocs, graduate students, undergraduates, and researchers. In particular, I would like to acknowledge my wonderful and talented undergraduates and high school interns, Lindsay Dahlen, Sruthi Veeragandham, Kristy Cheng, Yixing Cai, Janki Desai, Michelle Ngai, Annie Zell, and Ilsa Zhang, Mishi Vachev and Qinwen Deng for their hard work in the field, in the greenhouse, and in the lab. I would also like to thank Dr. Samuel Leiboff for his insightful suggestions to the GWAS project. Thanks also to Sabrina Soracco, the Director of Graduate Writing Center, for organizing writing groups and offering valuable advice to my dissertation. Most importantly, I would like to thank my collaborator, Prof. Jinliang Yang, for teaching me quantitative genetics step by step, for inviting me to visit his institution, and for his unrelenting support and edifying advice.

I would like to thank my wonderful cohort for their support and friendship. A particular thanks to Bailey Bonet for her dear friendship and companionship. I could not have survived grad school without her. Also, I would like to thank Brianna Haining for the constant companionship during my last semester, for our routine of exercising and writing dissertations together. I would also like to thank Dr. Ping-Hung Hsieh and Dr. Ouwei Wang for supporting me and giving me practical suggestions. Thanks also to my peer advisor Dr. Becky Mackelprang for helping me during my first year and introducing me to Bay Area Scientists in Schools (BASIS). I cannot leave PMB without mentioning my academic advisor, Rocio Sanchez, for keeping me on track and always being there when I needed help.

I would like to extend sincere thanks to acknowledge many many friends. In particular, my dear friends from my fellowships, Dr. Fang Huang, Dr. Hao Liu, Dr. Bo Li, Phoebe Wang, Shen Wang, Dr. Eric Wu, Jie Ma, Sarah Huang, Lucie Huang, Christie Ng, Dr. Qiongyu Zeng, Dr. Yi-Pei Li, Dr. Chenyu Wang, Rhea Li, Jia Wang, Zoë Wei, Karen Lin, Chengchun Gao, Sarah Meng, Kuan Gao, Bin Cao, Dr. Kai-Siang Chen, Kristin Kwok, Yihui Chen, Dr. Xiaoshuang Chen, Yang Xu, and Dr. Sheng Chen for their company and enduring love. Special thanks to Dr. Hao Liu and Dr. Eric Wu for their valuable advice on fellowship applications and this dissertation. I am also grateful to Prof. Zinmay Sung, Dr. Yuan Chen, Dr. Li Pu, and Dr. Congcong Hou for giving me the opportunity to work on promoting women in science and traveling together. Thanks should also go to Dr. Yi Xiong, Dr. Lina Qin, and Dr. Xin Li for their supervision during my lab rotations. Also, I'm deeply indebted to Prof. Ricardo San Martin for being a caring advisor, for his relentless support, and for being a source of unending inspiration. My deepest appreciation especially goes out to Jessica Schwabach, for everything, including not only starting an exciting adventure together but also being an abiding friend who is a continual source of unwavering support.

Nobody has been more important to me in the pursuit of this project than the members of my family. This dissertation is dedicated to my parents, Yingzhi Liu and Liwei Deng, for their unconditional love and endless support. Without my parents, I would not be able to pursue my dream and become a scientist. Special thanks to my little sister, Qinwen Deng, for the privilege of being a big sister and having to earn her respect. I am also grateful to my family in China for their love and support, especially my cousin, Siyi Li, cousin-in-law, Shuang Zheng, and my niece, Yunhan Zheng. And I would like to thank my dearest grandma for her selfless love and for the untold number of lessons I have learned from her.

Fiat lux

神说：「要有光。」就有了光。

The light has illuminated my understanding of science and beyond, allowing me the privilege to catch a glimpse of the transcendental beauty that is shrouded in the language of natural science.

Chapter 1

Environmental and Host-Specific Factors Shaping Root-Associated Microbiomes and an Evaluation of Quantitative Approaches for Dissecting Host Genetic Control of Microbiome Composition

1.1 Introduction

Microbes are fundamental to the function and maintenance of ecosystems, as they can influence many important ecosystem processes and can be found in virtually all environments, including soils, oceans, and the bodies of eukaryotic organisms (van der Heijden, Bardgett and van Straalen, 2008; Fierer and Lennon, 2011). The community of microorganisms, or microbes, living in a particular environment is known as a microbiome. For example, the human gut microbiome harbours a plethora of microorganisms having important functions relevant to human health (Clemente *et al.*, 2012), including increasing the host's ability to harvest energy and absorb nutrients from food (Krajmalnik-Brown and Ilhan, 2012), protecting the host from pathogen invasion and colonization (Kamada *et al.*, 2013) and modulating host immunity and metabolism (Zhang *et al.*, 2019). For many terrestrial endophytic communities, or communities residing within another host's body, the soil microbiome represents the primary reservoir of microbial diversity. Soil microbiomes have crucial roles in global nutrient cycling, the maintenance of soil health, and soil carbon sequestration, and can affect the health of plants and animals both directly and indirectly (Fierer, 2017; Saleem, Hu and Jousset, 2019). The soil microbiome plays a particularly influential role in shaping the plant root microbiome, which has been shown to enhance plant growth and health via mechanisms that include facilitating the solubilization and uptake of mineral nutrients, producing phytohormones to help plants withstand environmental stress, and eliminating phytotoxic compounds and soil-borne pathogens (Berendsen, Pieterse and Bakker, 2012).

1.1.1 Interactions between plants and soil microbes

Plants provide habitats, such as the phyllosphere (leaf), spermosphere (seed), and rhizosphere (root), for microbial colonization, and provide resources for microbial growth, as they are a major source of organic compounds (Turner, James and Poole, 2013). Rhizospheres, the narrow zone of soil adhering to the surface of the roots, are the primary interface between soil microbes and plant roots due to the organic carbon provided by the plant (Mendes, Garbeva and Raaijmakers, 2013). In fact, a large percentage of total organic carbon in the rhizosphere comes from sloughed plant root cells and tissues, as well as carbon-rich exudates, including carbohydrates, amino acids, organic acid ions, lipids, coumarins, flavonoids, and polysaccharide mucilage, which all serve as substrates for microbial metabolism (Reinhold-Hurek *et al.*, 2015). Studies using stable isotope probing, where ^{13}C was given to plants and fixed though photosynthesis, have identified a number of microbes actively metabolized plant-derived carbon (Lu *et al.*, 2006; Haichar *et al.*, 2008). In addition to carbohydrates, amino acids that are secreted by plants have also been discovered to drive the development of plant-specific microbial communities in the rhizosphere (Moe, 2013). Amino acid uptake by microbes has been traced using dual (^{13}C, ^{15}N) labelled glycine in wheat (Nasholm, Huss-Danell and Hogberg, 2001; Owen and Jones, 2001; Moe, 2013). In this, and similar experiments, amino acids were shown to provide microbes with not only a source of nitrogen, but carbon as well. Deamination or transamination of amino acids releases inorganic nitrogen, leaving carbon skeletons that can be used by microbes as a carbon source.

As mentioned above, the close interactions between plants and their complex microbial communities (the plant microbiome), can in turn play important roles in plant health and productivity (Turner, James and Poole, 2013). For example, one microbe that is widely reported as a plant growth promoter is *Trichoderma*, which has been reported to have the ability to enhance plant growth and yield, activate plant defenses against pathogens, promote nutrient uptake, improve seed germination, modulate carbohydrate metabolism and photosynthesis, and synthesize phytohormones to alter plant growth (Harman *et al.*, 2004; Harman, 2006; Stewart and Hill, 2014). Some *Trichoderma* species, such as *T. harzianum*, can directly interact with plants by colonizing the root surface and intercellular spaces in the epidermis and outer root cortex of the plants, where they are able to promote plant growth through this contact (Yedidia, Benhamou and Chet, 1999). Another recent study reported a *Trichoderma* species, *T. asperellum*, regulating ethylene production by decreasing its precursor 1-aminocyclopropane-1-carboxylate (ACC). As ethylene is a plant growth inhibitor, observed *Trichoderma* spp. with ACC deaminase activity are able to promote root elongation in *Brassica napus* (Viterbo *et al.*, 2010). Many *Trichoderma* bioinoculants are now commercially available with strain mixes becoming increasingly common, due to their greater consistency of performance. Another extensively studied plant growth-promoting bacteria is *Rhizobia* spp., whose members are capable of forming a symbiotic association with the roots of legumes and perform nitrogen fixation. Notably, they form specialized structures within the plant root called nodules, where nitrogen is converted to ammonia by *Rhizobia* in exchange for photosynthate (Masson-Boivin and Sachs, 2018). Mechanistically, these benefits are the result of a variety of complex exchanges with the host. For instance, the improvements to plant health due to *Rhizobia* are the result of improved nutrient uptake, which is, in large part, thanks to their ability to fix nitrogen from the surrounding environment. Other microbes improve yield by competing with pathogens; for example, *Pseudomonas* spp. suppresses plant diseases by producing antibiotics that inhibit pathogen colonization (And and Keel, 2003). Still others are capable of increasing plant host immunity without triggering yield penalties, priming plants for future attacks. As an example, some plant growth-promoting bacteria and fungi enhance plant defenses against pathogens and insects via an important mechanism called induced systemic resistance (ISR), by which selected microbes in the rhizosphere prime the whole plant body for enhanced defense against a broad range of pathogens and insect herbivores. Interestingly, a wide variety of root-associated mutualists, including *Pseudomonas*, *Bacillus*, *Trichoderma*, and mycorrhizal fungi, have been shown to promote the plant immune system for enhanced defense without directly activating costly defenses (Pieterse *et al.*, 2014); mechanisms behind these processes are still under investigation.

However, in addition to these beneficial interactions, there are also a host of detrimental ones, and the specific fitness outcome for a given plant is determined by the total composition and activity of its microbial community. As a result, much effort has been put into identifying the forces that contribute to plant microbiome development, and a wide range of host and environmental factors are now known to influence plant microbiome composition. Plant-associated microbial communities have been shown to vary according to different spatial compartments (Lundberg *et al.*, 2012; Peiffer *et al.*, 2013; Edwards *et al.*, 2015), with above and below ground and epiphytic and endophytic

communities often showing significant differences. In the case of the root microbiome, the community that colonizes the plant is a subset of the surrounding soil microbiome, and recent research has shown that this source soil microbiome is strongly influenced by a variety of environmental factors like temperature, soil pH, and seasonal variation (Fierer, 2017). Finally, the identity of the plant host and plant genetics has also been shown to be an important player in modulating plant-associated microbiomes (Agler *et al.*, 2016).

1.1.2 Host genotypic effects on plant associated microbial communities

The early papers that have explored host genotype's influence on the microbiome have shown that it seems to be small, and in some cases non-significant, relative to many of the other important factors. The very first plant microbiome paper that explored differences between soil type and genotype involved two accessions of the model plant *Arabidopsis thaliana* (Lundberg *et al.*, 2012). They observed that the impact of host genotype was secondary to most other influential factors, including sample type and soil origin. Another study used wild and domesticated accessions of barley (*Hordeum vulgare*) to investigate the structural and functional diversification among root-associated microbial communities and found that host genotype has a small but significant effect on the diversity of root-associated bacterial communities, which possibly represents a footprint of barley domestication (Bulgarelli *et al.*, 2015). Yet a third study investigated the rhizosphere bacterial diversity of 27 modern maize inbred lines to test the influence of maize host genotype on its rhizosphere microbial community across field environments. The maize rhizosphere also exhibited a small but significant fraction of variation in microbial diversity (both α- and β-diversity) across fields that could be attributed to host genetics (Peiffer *et al.*, 2013). A fourth study on three *Agave* species found that the microbial communities associated with agave plants are shaped by a number of environmental and host-related factors, in which the biogeography of the plant host species played the dominant role. Although both the prokaryotic and fungal data sets displayed clustering by sample type, geography and host species, additional comparison with other sympatric agave or non-agave species in the same arid environments would be necessary to fully understand the host genotype effect (Desgarennes *et al.*, 2014; Coleman-Derr *et al.*, 2016). Consistently in these studies, species effects tended to be nonsignificant (Schlaeppi *et al.*, 2014) or small (Lundberg *et al.*, 2012; Peiffer *et al.*, 2013; Fonseca-García *et al.*, 2016); however, many of these studies only considered a small subset of cultivars or species.

Later studies have delved deeper into this question of host influence by examining more genotypes or cultivars, and these found a consistent, significant, though small, effect of genotype. A recent study in *Arabidopsis thaliana* used 4 accessions for 600 plants total to show that the root rhizosphere and endophytic compartment microbiota of plants grown under controlled conditions in natural soils are sufficiently dependent on the host to remain consistent across different soil types and developmental stages, as well as sufficiently dependent on host genotype in varying between inbred *Arabidopsis thaliana* accessions. Their results provided a key step towards defining microbiome functional capacity and the host genes that potentially contribute to microbial association

phenotypes. Another recent study of 19 diverse cereal grasses found that host species were highly significant and explained ~19.4 and 23.4% of variance within the rhizosphere and root endosphere, respectively. The species effect size observed in this study, as compared with the relatively low values observed in previous studies, suggests that the influence of host species on the root microbiome should be considered in the context of the working phylogenetic framework, and a broader array of considered species will accordingly produce a greater host effect (Naylor *et al.*, 2017). Finally, a recent, large-scale longitudinal field study of the maize rhizosphere microbiome identified bacterial taxa of which relative abundances were partially explained by genetic differences between the maize lines, above and beyond the strong influences of field, plant age, and weather on the diversity of the rhizosphere microbiome (Walters *et al.*, 2018). Collectively, these prior studies demonstrate not only that genotype has the potential to influence community composition, but also that the strength of the impact of genotype can vary from sample type to sample type and study to study.

One exciting idea that emerges from these more recent studies is that comparisons of microbiomes across a broad range of hosts can be used to answer broader questions about microbial sensitivity to genotype. For instance, in a recent study of the maize microbiome, several key features of microbial sensitivity to genotype, or 'heritability', were revealed. First and foremost, while it is true that the heritability of individual rhizosphere operational taxonomic unit (OTUs) in maize plants was lower than for most traditional agronomic traits (0.15–0.25, whereas the heritability of plant yield is around 0.3 and flowering time can be 0.9 or higher), there is a noticeable range of effect size of plant genetics on the relative abundances across rhizosphere associated taxa. Despite strong environmental patterning, they identified close to 150 OTUs with significantly high heritability. These heritable OTUs were highly diverse, including 26 Alphaproteobacteria, 9 Betaproteobacteria, 12 Actinobacteria, 6 Verrucomicrobia, and 8 Bacteroidetes. Others belong to the bacterial classes WS3, Beta- and Gammaproteobacteria, Planctomycetes, Firmicutes, Chloroflexi, Acidobacteria, and Gemmatimonadetes, as well as, interestingly, the Archaeal phylum Crenarchaeota (candidatus Nitrososphaera). Furthermore, and more importantly, mapping the heritable OTUs onto a common phylogeny revealed some clusters of related taxa. This is not completely unexpected: since plant genotype selects on microbial phenotypes, heritable taxa likely encode functions that are phylogenetically restricted, with function manifested at the taxon level.

Another recent study of 3,024 rice (*Oryza sativa*) accessions also found heritable taxa in the leaf microbiome. This study investigated the community composition of bacteria across leaves of 3,024 rice accessions from field trials in China and the Philippines using metagenomics. Through this experimental approach, they identified rice genomic regions controlling the abundance of microbial network hubs, and showed that they were enriched for processes involved in stress responses and carbohydrate metabolism. Additionally, they found that the networks revealed key microbial groups that not only regulate the establishment of the community, but also appear to be controlled genetically by the host. Both this study and the previously discussed study of the maize microbiome have emerged in the past two years, and to date there are no

studies on heritability of microbial lineages in the microbiome of other plant tissues. For this reason, the extent to which heritability of microbial lineage is consistent across other plant tissue types, as well as across other plant hosts, remains to be tested.

1.1.3 Quantitative genetics approaches for dissecting host effects on plant microbiome composition

A second potentially important idea to emerge out of recent efforts to dissect the interplay between host genotype and microbiome composition is that this style of experiment can also be adopted for a traditional dissection of the host genetics controlling the microbiome through Genome-Wide Association Studies (GWAS) or quantitative trait loci (QTL) analysis. GWAS represent a powerful approach to determine aspects of the genetic architecture, which are associated with complex traits in large natural populations with high genetic diversity. Though pioneered by human geneticists, many more GWAS studies have been conducted in plants (Brachi, Morris and Borevitz, 2011), and it has become increasingly popular for studying the genetics of natural variation and traits of agricultural importance. When inbred lines are available, GWAS can be particularly useful because it has the advantage of repetitively using panels of diverse lines with known genotypes that can be phenotyped for any trait of interest at any given time, making it possible to study many different traits in multiple environments (Atwell *et al.*, 2010; Huang and Han, 2014).

The first large and well-designed GWAS study was conducted in humans to study the genetic basis of seven common human diseases (Burton, P., Clayton, D., Cardon, L. et al., 2007), and advanced the understanding of the genetic basis for many common phenotypes of biomedical importance (McCarthy *et al.*, 2008). In plants, the most common method for identifying genetic loci responsible for variation in complex traits was QTL mapping (Glazier, Nadeau and Aitman, 2002). However, QTL mapping has limitations associated with allelic diversity and genomic resolution (Borevitz and Nordborg, 2003). GWAS allows for the screening of markers across the genome (especially single-nucleotide polymorphisms, or SNPs) that are correlated with single or multiple scorable phenotypes. Mapping using GWAS in large populations draws on recombination events to identify the genetic loci underlying traits, and because there are millennia of such events to draw upon, resolution and mapping power are much higher in GWAS than with QTL mapping (Korte and Farlow, 2013). When combined with dense, genome wide marker coverage, GWAS can substantially improve genetic resolution relative to QTL mapping (Morrell, Buckler and Ross-Ibarra, 2011). The higher resolution provided by GWAS can even reach to the gene level and samples from previously well-studied populations can be used, where genetic variations can be associated with phenotypic variation (Brachi, Morris and Borevitz, 2011). After mapping has been completed, loci of interest may be correlated with candidate genes within the interval for further annotation and functional tests.

GWAS have been successfully carried out in more than 22 plant species and have lead to the discovery of new mechanisms and genes underlying variation in plant growth

and development. Studies have examined the basis of agronomic traits, including flowering time, plant height, seed quality and weight, fruit weight, leaf angle and leaf size, in rice, maize, barley, wheat, millet, and many more (Ogura and Busch, 2015). GWAS for seed-related traits have received the most attention due to their high relevance for crop yield and breeding (Linkies *et al.*, 2010). For example, a collection of inbred lines of rapeseed (*Brassica napus*) which originated from geographically diverse regions, were utilized to identify multiple loci associated with seed quality and seed weight-related traits via a GWAS approach (Li *et al.*, 2014). GWAS for seed-related or fruit-related phenotypes has also led to the identification of multiple novel candidate genes in many more species, such as maize, rice, barley, soybean, and tomato (Huang *et al.*, 2011, 2013; Pasam *et al.*, 2012; Xue *et al.*, 2013; Hwang *et al.*, 2014; Sauvage *et al.*, 2014; Yang *et al.*, 2014; Ogura and Busch, 2015). In addition, GWASs on leaf and root traits has proven being powerful in dissecting molecular processes that underlie the quantitative regulation of organ growth and development. For example, the *liguleless* genes, which regulate the leaf angle in maize and leads to more upright leaves in modern maize varieties, were mapped by GWAS in Nested Association Mapping (NAM) lines for maize leaf angle and size (Tian *et al.*, 2011). Over the last few years, large-scale image acquisition and automated trait quantification from these images have emerged to allow high-throughput phenotyping of root traits. Efficient phenotyping of root traits coupled with GWAS approaches have led to the uncovering of genes regulating the elongation rate of primary roots, such as CALCIUM SENSOR RECEPTOR (Slovak *et al.*, 2014) and F-box family gene KURZ UND KLEIN (Meijón *et al.*, 2014).

However, there are a number of potential challenges to the successful implementation of GWAS. Firstly, GWAS studies are adversely affected by crop population structure. Often traits will fail to have a significant GWAS result when family or population structure produces spurious associations, resulting in an inflated discovery rate and false positives (Han and Huang, 2013). Computational methods such as mixed linear, multilocus, and multitrait mixed models have been developed, improved, and optimized to address the problem. Secondly, GWAS has low power to detect rare alleles, which are a substantial proportion of natural variation. Additionally, in the interval of one GWAS locus, there may be hundreds of genes, only one of which might be the causal gene. As a result, followup analyses (e.g. gene annotation, expression profiles, functional analysis) are necessary. Another challenge is low coverage sequencing. Modern high throughput sequencing can provide low cost sequencing for many samples, at the expense of relatively low coverage. A downside of this is that sometimes there are missing data that must be inferred using imputation methods. These methods are mostly based on HiddenMarkov models and are easiest for plants with long-range linkage disequilibrium and inbreeding (such as sorghum) (Han and Huang, 2013). In practice, using imputation in conjunction with low coverage sequencing has been shown to greatly increase the power of GWAS (Pasaniuc *et al.*, 2012). Collectively, this suggests that there are a number of potential pitfalls that may be encountered when attempting GWAS, which need to be taken into account or addressed in order to successfully implement such an approach.

It is worth noting that GWAS has not been used with equal prevalence across all crops. For example, research on the important cereal, feedstock and bioenergy crop sorghum, which is a staple for populations across the globe, has not been extensively studied with GWAS to date. This being said, a small number of GWAS have been previously conducted in sorghum, including one for plant height (Morris, Ramu, *et al.*, 2013) and flavonoid pigmentation (Morris, Rhodes, *et al.*, 2013), demonstrating the efficacy of this approach for sorghum. In fact, sorghum represents an attractive model system for GWAS due to its low ploidy level (2n), which makes high throughput sequencing less prone to error, and thereby facilitates easier GWAS. As opposed to crops like wheat, which can be found in tetra- and hexaploid varieties, using diploid relatives such as sorghum may be a good baseline for future genome level research in polyploidy crops (Morrell, Buckler and Ross-Ibarra, 2011). Finally, selfing species such as barley, rice, and sorghum, tend to have individual loci with large additive effects, for example with flowering time (Lin, Schertz and Paterson, 1995). Further analysis of loci discovered through GWAS in a selfing species may be of more significance due to the potential higher effect of said loci. Given the applicability of GWAS for sorghum, its historical lack of use, and the importance of sorghum for global agriculture, we argue that sorghum is an attractive model system for GWAS-based dissection of host-control of microbiome composition.

More recently, GWAS has been tried using traits that are complex or multivariate, like the microbiome, rather than simple, single, quantifiable phenotypes. The very first genome-wide scan for variants that are associated with microbiome traits was performed in the gut microbiomes of a large (n = 645) mouse advanced intercross line, which identified 18 host QTLs associated with relative abundances of specific microbial taxa in the gut (Benson *et al.*, 2010). Later, several GWAS have been carried out in human microbiome. In the Hutterites, a founder population that lives and eats communally, researchers examined the association of ~200K host genotypes with the relative abundance of fecal bacterial taxa and identified an association between a taxon known to affect obesity (genus *Akkermansia*) and a variant near Phospholipases D1 (PLD1), a gene previously associated with body mass index (Davenport *et al.*, 2015).

A recent GWAS that used the leaf microbial community as the phenotypic trait in *Arabidopsis thaliana* suggests that plant loci responsible for defense and cell wall integrity affect microbial community variation (Horton *et al.*, 2014), and this result established GWAS as a potential tool for studying the role of plant host genetic heterogeneity in shaping microbial community. However, a later study conducted by (Wallace *et al.*, 2018) analyzed the metabolically active bacteria of maize leaves across 300 diverse maize lines growing in a common environment using GWAS and was not successful in identifying significant correlations with host genetic loci. Heritability analysis of 49 community diversity metrics, 380 bacterial clades, and 9,042 predicted metagenomic functions only yielded 2 diversity metrics, 5 bacterial clades, and 200 metabolic functions as being significantly heritable.

The mixed success of microbiome GWAS in plants could be due to one of the general limitations and challenges of GWAS described above, or perhaps to more

microbiome specific issues. For example, the sample type chosen may have profound effects on the ability to detect correlations. Use of GWAS in the context of root associated microbiome has yet to be extensively explored; here, selection of sample type (rhizosphere or endosphere) and host system may be critical factors that determine the success of such efforts. Previous work comparing the root microbiomes of a broad range of cereal crops has demonstrated the degree to which microbial communities correlate with host phylogenetic distance is strongest in the root endosphere (Naylor *et al.*, 2017), while another study of different rice cultivars found that host genotype correlated with microbial association most strongly in the rhizosphere (Edwards *et al.*, 2015). Collectively, these data suggest the sample type exhibiting the strongest correlation may be different for each host, and that an initial evaluation of the degree of correlation between genotype and microbiome phenotype across sample types may be helpful for successful implementation of GWAS. Alternatively, the selection of host may have a large impact. For instance, in the rhizosphere, microbiomes are heavily influenced by exudation. Hosts, which have strong, complex or varied exudation patterns, may also have greater impact on these particular communities.

1.1.4 Design considerations for implementing a successful study to identify host genetic effect on plant microbiome regulation

Finally, and perhaps most importantly, microbiome studies are often challenging because of the large number of often uncontrollable or unaccounted for confounding environmental, sample processing, and analysis related factors (Simmons *et al.*, 2018). For this reason, when sampling a plant microbiome, one must take care to limit the influence of these other factors. Take for example, studying the rhizosphere fraction of the root microbiome. Studying root-associated microbiomes presents unique challenges, due in part to the inherent difficulties in sampling from soil. Soils are highly variable in terms of physical and chemical properties, and different soil conditions can be separated by as little as a few millimeters. This can lead to the samples, which are collected from adjacent sampling sites, having considerably different microbial community compositions and activities (Fierer and Lennon, 2011; O'Brien *et al.*, 2016). Designing experiments that account for these potential positional and depth-related effects, by means of replication and careful sampling planning, are essential.

As with any experimental platform, amplicon-based profiling can also introduce a number of potential biases that should be considered during sample processing and data analysis. These include methods chosen for DNA extraction, selection of PCR primers, and how library preparation is performed. Different methods can significantly impact the amount of usable data generated, as well as hinder the efforts to meaningfully compare results between studies. One important factor is the use of a consistent depth when excavating roots. Using soil core collectors and shovels to maintain consistent sampling depths and homogenization prior to DNA extraction is essential to the reproducibility within root microbiome studies (Simmons *et al.*, 2018). One key factor that can negatively impact or disrupt sequencing results is bacterial contamination, which can come from many sources and is sometimes impossible to distinguish from the sampled

environmental bacteria (Salter *et al.*, 2014; Weiss *et al.*, 2014). For this reason, careful sterilization of sampling tools, experimental materials, and working environments are vital in order to avoid contamination.

After sample collection, it is also essential to efficiently separate the rhizosphere and root fractions; using a harsh method of root surface sterilization can potentially lyse endophytes within roots prior to DNA extraction, while a more conservative wash may not remove all microbes from the root surface (Richter-Heitmann *et al.*, 2016). Striking the appropriate balance is key. After sample processing, obtaining high-quality DNA is a high priority for successful downstream analyses. In our experience, DNA extraction from field-grown root samples through alternative methods, such as through CTAB-based extraction, often contain substantially greater quantities of humic acids and other compounds compared to rhizosphere and soil samples. These compounds can prevent the enzymatic activity of the DNA polymerase during PCR amplification, even at low concentrations (Sutlović *et al.*, 2005; Sutlovic *et al.*, 2008). Using DNA extraction kits designed for soils on root samples, as opposed to a CTAB extraction followed by a phenol chloroform clean-up, can effectively rid samples of humic acids and result in high-quality DNA (Aleklett *et al.*, 2015; Bogas *et al.*, 2015; Hiscox *et al.*, 2015; Zhang and Yao, 2015). Accordingly, we recommend using a commercially available DNA extraction kit for root samples as well. Secondly, thorough and consistent root grinding is important to break down the plant tissue and lyse the microbial cells to release microbial DNA without introducing bias between samples due to the variation in grinding pressure and time.

Third, one main source for problems during amplification is the contamination of plant tissues with plant endosymbionts (chloroplast and mitochondria). The amplification from chloroplast or mitochondria 16S rRNA sequences can generate >80% of the sequences in root samples, and more in leaf tissues, though the amount of contamination is dependent on the choice of primers (Ghyselinck *et al.*, 2013). Thus, the use of PNA clamps, peptide nucleic acid (PNA) oligos which block the amplification of host DNA, are necessary during the PCR step to suppress plant host chloroplast and mitochondrial 16S contamination (von Wintzingerode *et al.*, 2000; Lundberg *et al.*, 2013). However, different plant species can have variation in the chloroplast and mitochondrial 16S sequence (Lundberg *et al.*, 2013); therefore, it is important to check the compatibility of the universal PNAs with your specific plant system bioinformatically during experimental design in order to verify that they will block amplification of chloroplast and mitochondrial 16S genes. Following the amplification step, it is not apparent whether the PNAs successfully bound to mitochondrial and chloroplast templates; this is only revealed after sequencing **[Figure 1-1]**. To help ensure that the PNAs will effectively block contaminant amplification, an alignment of the PNA sequence to each chloroplast and mitochondrial 16S rRNA gene (there may be multiple copies) for the plant host being investigated should not reveal any mismatches. Even a single mismatch to the 13 bp PNA sequence, especially in the middle of the PNA clamp, can drastically reduce the effectiveness, as in the case of one of our studies, the provided chloroplast PNA sequence and the chloroplast 16S rRNA gene of *Lactuca sativa* (lettuce) **[Figure 1-1]** (Simmons *et al.*, 2018).

Additionally, there are nine "hypervariable regions" (V1 to V9) in 16S rRNA genes, which contain considerable species-specific sequence diversity among different bacteria (Van de Peer, Chapelle and De Wachter, 1996; Ashelford *et al.*, 2005). These hypervariable regions are flanked by regions of more conserved sequence, which allows the use of universal primers to amplify target sequences to identify and compare bacterial taxa (Baker, Smith and Cowan, 2003) **[Figure 1-2]**. To take advantage of the current high-throughput next generation sequencing technologies, paired-end Illumina MiSeq Platform has been adopted for providing a flexible and cost-effective sequencing option for 16S rRNA microbiome studies (Kozich *et al.*, 2013). Typically, short regions of the 16S rRNA gene are used rather than the full-length gene, allowing for an increased number of short reads per sample. Different results can be obtained from the same community depending upon which hypervariable region is amplified (Cruaud *et al.*, 2014). Previous studies have found the V4 region to be one of the most reliable for assigning taxonomy and it has been used for other extensive microbiome surveys (Yang, Wang and Qian, 2016; Thompson *et al.*, 2017). Recently, a dual-indexing approach for multiplexed 16S rRNA of the V3 and V4 regions (approximately 500 bp) was developed (Fadrosh *et al.*, 2014) **[Figure 1-2]**. Therefore, lengthening the target to the V3-V4 region is suggested here to increase variability and improve taxonomic resolution.

Given that host genotype's influence on the microbiome have shown to be small, and difficult to detect compared to many of the other important factors, the ability of identifying host genetic effects is heavily dependent on the degree of limiting the influence of these other environmental and experimental factors. Thus, experimental design is vital for a successful study on investigating the host genetic regulation on plant microbiome.

1.1.5 The role of biologicals in plant growth and development

Ultimately, a main goal of plant microbiome research is to find ways to harness the microbiome to improve crop performance. During a growing season, agricultural crops are frequently exposed to abiotic stresses, which hinder plant growth and reduce crop productivity. Abiotic stresses include extreme temperatures, drought, submergence, soil salinization, and nutrient imbalances, and many of these stresses are predicted to increase in frequency or severity in the coming century. Maintaining high levels of crop productivity under increasingly suboptimal growth environments will require the development of new agronomic tools. One new and promising strategy involves the utilization of microorganisms, which have been shown to be capable of mitigating abiotic stresses, and the plant microbiome has the potential of promoting growth and protecting the host through a variety of molecular mechanisms. While studies have begun to explore how specific members of the root microbiome act to enhance plant growth, we still lack a full understanding of the role of the broader root microbiome in shaping plant stress tolerance.

Biologicals are products that can be used to improve crop performance. There already exists a large body of research on microorganisms beneficial to plants that can be

used to guide the screening process, to establish best practices for validation, and to help identify some pitfalls that may prevent transfer of benefits for greenhouse to field. These microbes are frequently labeled "plant growth promoting microbes" or PGPM and are defined as microbes, which confer any of a wide range of benefits to plants, including those growing under both favorable and stressful conditions. One of the most well-studied plant growth-promoting bacteria is *Rhizobia* spp., whose members are capable of performing nitrogen fixation. Notably, they form specialized structures within the plant root called nodules, where nitrogen is converted to ammonia by *Rhizobia* in exchange for photosynthate (Masson-Boivin and Sachs, 2018). However, most other PGPM are unable to perform nitrogen fixation and instead offer other benefits to their hosts. While there are many unique mechanisms through which PGPM can benefit the plant host, the most well-studied direct benefits are those that affect the host through facilitation of acquisition of mineral nutrients such as phosphate or through the synthesis and modulation of plant hormones, such as auxin and ethylene (Glick, 2012; Kim *et al.*, 2012). Outside of direct benefits, other PGPM provide indirect benefits to the plant host, for instance, by acting as biocontrol agents, which may protect the plant when an abiotic stress might otherwise leave the plant with enhanced susceptibility to potential pathogens (Mendes *et al.*, 2011; van der Voort *et al.*, 2016).

Beneficial microbes impact the native root microbiome in a variety of ways. The regulation of phytohormones is one of the main mechanisms through which PGPM are known to alleviate plant stress and rescue normal plant growth phenotypes during environmental stresses, such as drought. Many independent studies have demonstrated that PGPM have the potential to regulate hormone levels and metabolism in plants. Several *Azospirillum* spp. strains produce IAA or IAA-inducing signaling molecules, such as nitric oxide. These molecules can benefit plant growth during drought by increasing root growth and promoting lateral root and root hair formation, which facilitate increased uptake of water and nutrients (German *et al.*, 2000; Creus *et al.*, 2005; Molina-Favero *et al.*, 2008; Arzanesh *et al.*, 2011). GA also contributes to plant growth during drought (Colebrook *et al.*, 2014). PGPM that are able to synthesize GA have beneficial effects on growth and yield of many crop plants (Bottini, Cassán and Piccoli, 2004). For example, the GA-secreting rhizobacterium *Pseudomonas putida* H-2-3 improved soybean growth under drought conditions (Kang *et al.*, 2014). Notably, *P. putida* H-2-3 also displayed the ability to modulate superoxide dismutase, flavonoids, and radical scavenging activity, suggesting that the beneficial effect of this bacterium is relatively complex (Kang *et al.*, 2014). Perhaps the most significant response to drought stress is mediated by ABA, which has an important role in regulating stomatal closure and drought-induced signal transduction. Many PGPM confer drought stress resistance by modifying ABA levels. For example, Arabidopsis plants inoculated with *A. brasilense* Sp245 display elevated levels of ABA compared to non-inoculated plants (Cohen, Bottini and Piccoli, 2008). Additionally, a PGPM isolated from the rhizosphere of *Brassica napus*, *Phyllobacterium brassicacearum* strain STM196, enhanced ABA content, leading to decreased leaf transpiration in Arabidopsis plants (Bresson *et al.*, 2013). However, connecting PGPM to plant hormone modulation must be done carefully, as the manipulation of one hormone can also impact others due to crosstalk between plant hormone signaling pathways. For example, *Platycladus orientalis* seedlings inoculated

with *Bacillus subtilis*, a CK producing PGPM, displayed an elevated concentration of ABA in shoots, increased stomatal conductance, and enhanced tolerance to drought (Liu *et al.*, 2013).

Another key mechanism employed by PGPM to facilitate plant growth is to lower plant ethylene levels by secreting the enzyme 1-aminocyclopropane-1-carboxylate (ACC) deaminase, which cleaves ACC, the immediate precursor of ethylene. Under drought conditions, ethylene endogenously regulates plant homeostasis and results in reduced root and shoot growth. However, PGPM that produce the enzyme ACC deaminase improve drought tolerance and rescue normal plant growth by degrading ACC, thereby decreasing the level of ethylene in the plant (Glick, 2005). For example, inoculation of *Pisum sativum* with ACC deaminase producing *Pseudomonas fluorescens* biotype G (ACC-5) induced longer roots, which led to an increased uptake of water from soil during drought (Zahir *et al.*, 2008). Another well-characterized endophytic PGPM, *Burkholderia phytofirmans* PsJN, also produces ACC deaminase to promote plant growth. Mutants of these bacteria that lack ACC deaminase activity are no longer able to promote canola seedling root elongation (Sessitsch *et al.*, 2005; Sun, Cheng and Glick, 2009). Inoculation of wheat plants with ACC deaminase producing PGPM in axenic studies showed better root development with increased root-shoot length, root-shoot mass, and lateral root number compared with control plants, which helped inoculated plants uptake more water and nutrients, resulting in improved growth and yield (Shakir, Bano and Arshad, 2012). Bacteria that have ACC deaminase activity are present in diverse bacteria lineages. In a study of wheat, 38 PGPM containing ACC deaminase were isolated. These PGPM included 12 distinct genera belonging to phylum Firmicutes, class Gammaproteobacteria and Betaproteobacteria, and genus Flavobacterium (Gontia-Mishra *et al.*, 2017). Notably, not all strains within a given bacterial species contain ACC deaminase. Therefore, confirmation of ACC deaminase activity is often confirmed either directly by culturing bacteria with media in which ACC is the sole source of nitrogen or indirectly by PCR validating the presence of the ACC deaminase gene (Saravanakumar and Samiyappan, 2007; Jalili *et al.*, 2009). Taken together, the studies highlighted above demonstrate the critical role of phytohormone regulation in endophytic manipulation of plant growth and the response to drought stress.

A challenge of using PGPM is that they are not always persistent and interaction with the existing soil or rhizosphere microbiome may limit their utility. In our study on the PGPM effects on the strawberry root bacterial microbiome, discussed below in Chapter 4, we observed that many of the shifts associated with amendment application are dependent on time. In the final time point, the treated samples are more similar to control samples for all sample types. Importantly, in the fourth time point, samples were collected nearly one month after the most recent amendment application, in contrast to the other time points, in which the amendment had been applied two weeks prior to collection. This result is also typical for studies conducted on the introduction of single organisms into complex soil systems. Several studies have found the general reduction of levels of individual inoculants introduced to soils(van Veen, van Overbeek and van Elsas, 1997; Matos, Kerkhof and Garland, 2005; Kröber *et al.*, 2014; Schreiter *et al.*, 2014), finding in as little as a week there was more than a 99% reduction in abundance. Another

study with pathogenic *Pseudomonas aeruginosa* showed a decline to below detectable levels 3–5 weeks post introduction under non-sterile microcosms, while the population was maintained at high abundance under sterilized microcosms (Deredjian *et al.*, 2014). Our data suggest that even in cases where amendments represent complex communities of microorganisms, soil and root communities both have a resilience that will lean towards eventual recapitulation of the native state following microbial inoculation. To what extent the activity of a biological depends on the colonization and persistence of the microbes in the product has yet to be thoroughly explored, and is likely dependent on the mode of action of the product, the complexity of the target community, and other environmental factors. Collectively, these observations have important implications for the successful use of such products in commercial agriculture, and suggest that repeated applications might be beneficial for the persistence of the plant growth promoting agents.

1.2 Conclusion

In this chapter, we have discussed the complex interactions between environment, plants, and microbes, which ultimately act to shape the composition of plant-associated microbiomes. Recent studies in various plant species have discovered multiple factors in shaping the plant root microbiome, including geographical location, soil source, host genotype, and cultivation practice (Peiffer *et al.*, 2013; Edwards *et al.*, 2015; Naylor *et al.*, 2017; Xu *et al.*, 2018). However, identifying specific mechanisms driving microbiome acquisition and assembly and the host genetic variants involved in these processes has proved challenging. In particular, additional research is needed to explore the host genetic regulation of plant-associated microbes within the naturally occurring ecological context. There are numerous limitations in experimental methodologies in identifying the relative small host effect on microbiome regulation. We proposed to utilize recent advances in sequencing technology, computer science, statistical models, and a successful diversity panel to perform a population-level microbiome study using a large scale genetic approach. One optimal approach could be GWAS, which allows us to rapidly scan markers across the complete sets of DNA of many sorghum germplasm to find genetic variation associated with particular microbial traits. Such studies will help refine our knowledge and explore what universality may exist in how plants regulate their microbiome – for instance, to what extent root microbiome composition is varied with host genetics, as well as what genes are shared across populations and hosts that control these variations. In the following chapters, work that expands our knowledge of the relationship between host genetics and root microbiome composition and the impact of biological products on root microbiome composition will be described.

Figure 1-1

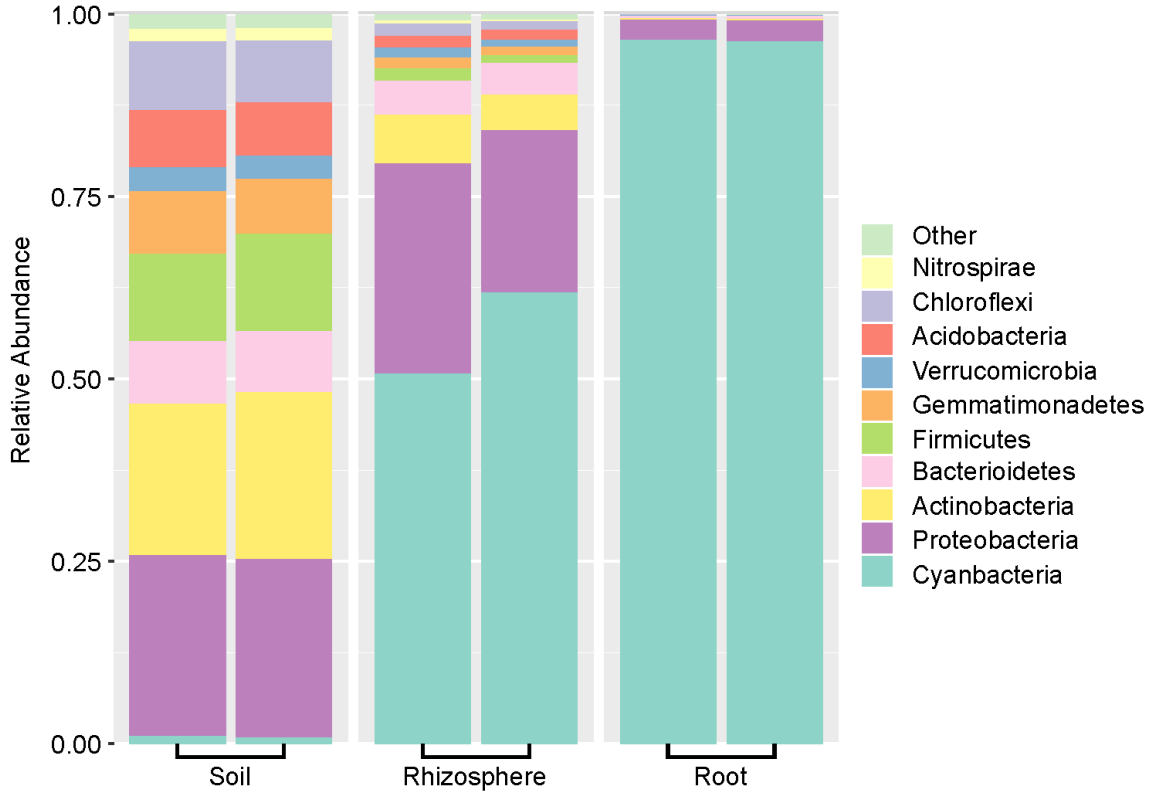

Figure 1- 1. Representative results of PNA inefficiency

Representative result from soil, rhizosphere, and root, soil samples from lettuce plants. The PNA sequence used to block chloroplast contamination of most plants is GGCTCAACCCTGGACAG (Lundberg *et al.,* 2013). However, lettuce contains a mismatch in the chloroplast 16S ribosomal RNA gene (GGCTCAACTCTGGACAG). This renders the PNA ineffective, resulting in a high relative abundance of reads that match to Cyanobacteria in rhizosphere and root samples.

Figure 1-2

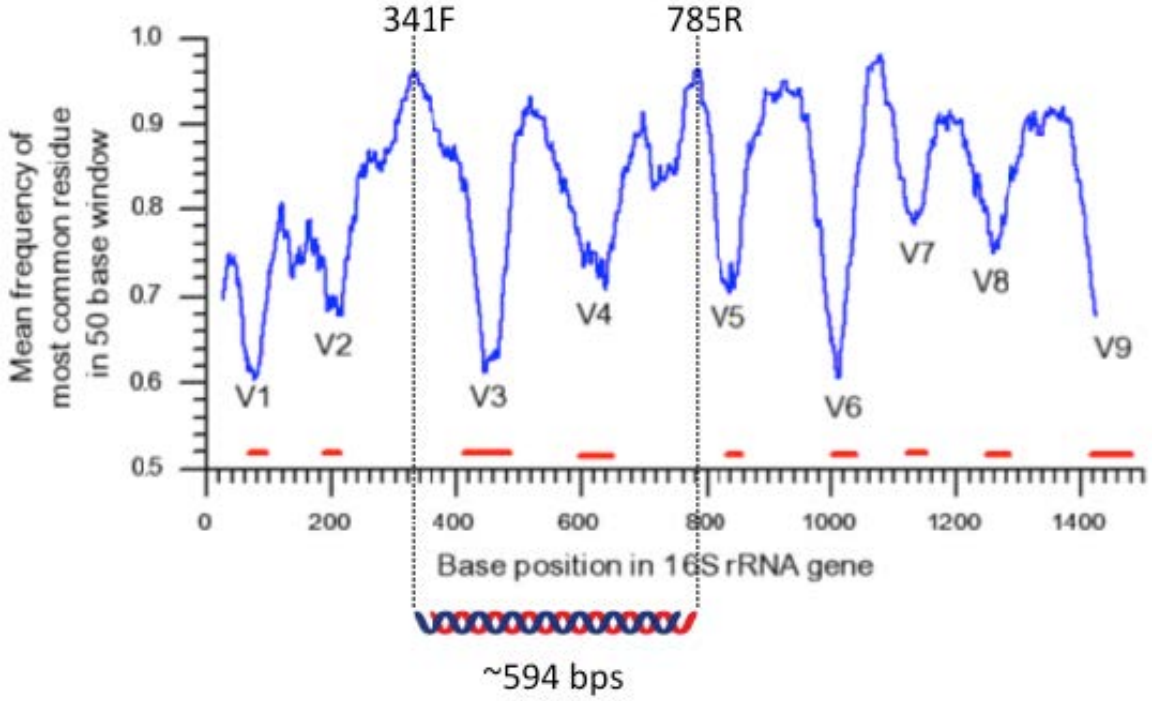

Figure 1- 2. Illustration of the variable regions within the 16S rRNA gene

The peaks are region of high conservation, while the valleys are regions of low conservation, and are labeled as hypervariable regions, with red bars on the x axis defining these regions as V1 to V9 (Modified from Ashelford *et al.,* 2005). There is an example of a set of primers designed in the conserved region to amplify V3-V4 region. The length of the product is around 594 bps, which can be sequenced using 300 bp paired-end MiSeq protocol.

Chapter 2

Identifying Heritable Microbes in the Sorghum Rhizosphere Microbiome

2.1 Abstract

Host genetics and the associated microbiome can influence both plant development and plant health. Despite much research, the rules governing microbial community recruitment to the host microbiome remain unknown. To improve our understanding of the relationship between the host genetics and plant microbiome, we first asked the questions of whether host genetic diversity results in differences of microbial community across a variety of sorghum cultivars. A recent study of 27 maize inbred line has identified that host genetic variance could partially explain the difference in the relative abundances of certain microbial taxa in maize rhizosphere, although environmental factors, such as soil source, plant age, and field conditions, have been shown to be more influential to the plant-associated microbiome. In our study, we conducted a population-level microbiome analysis of the rhizosphere microbiomes of 200 sorghum genotypes to dissect microbial sensitivity to host genotype. Using 16S rRNA amplicon datasets, we quantified the effect of plant genotype on both community-level traits and microbial lineage-level traits. We calculated broad sense heritabilities of the microbial partners of sorghum rhizosphere across all of these 200 sorghum diverse lines, and defined a fraction of the microbiome considered to be heritable. Interestingly, when we compared the heritable taxa within sorghum rhizosphere with the ones within maize rhizospheres, we observed a significant overlap in the heritable fraction of the rhizosphere microbiome of these two different cereal crops. Collectively, the results suggest that some rhizosphere microbes are heritable and they are phylogenetically clustered and may be similar across different hosts.

2.2 Introduction

Recent work has shown that root-associated microbial communities are in part shaped by host genetics (Peiffer *et al.*, 2013; Schlaeppi *et al.*, 2014; Edwards *et al.*, 2015; Naylor *et al.*, 2017). A study comparing the root microbiomes of a broad range of cereal crops has demonstrated a strong correlation between host genetic differences and microbiome differences (Naylor *et al.*, 2017), suggesting that a subset of the plant microbiome may be sensitive to the influence of host genotype across a range of plant hosts. In maize, these genotype-sensitive, or 'heritable', microbes have been shown to be phylogenetically clustered within specific taxonomic lineages (Walters *et al.*, 2018); however, it remains unclear whether this increased genotype sensitivity in these microorganisms is unique to maize or is a common feature amongst other plant hosts as well.

In the context of the root microbiome, we propose *Sorghum bicolor* (L.) as an excellent plant system for dissection of host-genetic control of microbiome composition. Sorghum is a heavy producer of root exudates, and the sorghum microbiome has been shown to house an unusually large number of host-specific microbes (Naylor *et al.*, 2017). Additionally, there is a wide range of natural adaptation in traditional sorghum varieties from across Africa and Asia, and a collection of breeding lines generated from U.S. sorghum breeding programs, both of which provide a rich source of phenotypic and genotypic variation (Casa *et al.*, 2008). Several genome sequences of sorghum varieties have been completed, and variation in nucleotide diversity, linkage disequilibrium, and recombination rates across the genome have been quantified (Morris, Ramu, *et al.*, 2013), providing an understanding of the genomic patterns of diversification in sorghum. (Morris, Ramu, *et al.*, 2013; Chopra *et al.*, 2017; Cuevas *et al.*, 2019; Zhou *et al.*, 2019) has also shown that sorghum is suitable for Genome-Wide Association Study (GWAS) analysis. Finally, sorghum is an important cereal crop grown widely throughout the world as a food, feedstock, and biofuel. Studying host genetic regulation of microbiome in an agronomically relevant cereal might prove useful for similar studies in other related crops, including maize and barley, and help guide microbiome-based efforts for crop improvement.

In this chapter, we show a model system that we used to dissect the host-genetic control of bacterial microbiome composition in sorghum rhizosphere. Using 16S rRNA sequencing, we have profiled the microbiome of a panel of 200 diverse genotypes of field-grown sorghum. First, we demonstrate that for sorghum, the rhizosphere represents a more suitable fraction of the plant microbiome for studying the host genetic effects than either root or leaf. Next, we show that a large fraction of the rhizosphere microbiome responds to host genotype. To quantify the response, we calculated broad sense heritabilities, which is defined as the proportion of phenotypic variance explained by the effects of host genotypes (Lynch, Walsh and Others, 1998) of rhizosphere related traits. We include both community-level traits, such as the overall community diversity and eigenvectors produced by ordination analysis (principal component analysis; PCA), and microbial lineage-level traits, which are the abundances of specific taxa comprising the microbial community. Furthermore, we define the fraction of the microbial taxa

considered to be heritable and find that this fraction is biased towards specific phylogenetic clades of often high abundance organisms. Finally, we combined our dataset of heritable taxa within sorghum rhizosphere with two previously published datasets of heritable taxa within maize rhizospheres, and performed a comparative analysis (Walters *et al.*, 2018). Within the rhizosphere, we demonstrate that the heritable taxa in sorghum show a strong degree of overlap with heritable lineages identified in maize, spanning fifteen different bacterial orders (Walters *et al.*, 2018). Collectively, these observations suggest that some rhizosphere microbes are heritable and these are phylogenetically clustered and may be similar across hosts. Host-mediated microbiome engineering provides a potential new direction of microbial applications.

2.3 Results

2.3.1 Selection of germplasm and sample type for quantitative genetic analyses

In this study, the relationship between host genotype and microbiome composition was explored through a field experiment involving 200 genotypes selected from the Sorghum Association Panel (SAP) germplasm collection **[Table 2-1]** (Casa *et al.*, 2008). To ensure a uniform starting inoculum for all sorghum seedlings and to control their planting density, seeds were first sown into a thoroughly homogenized field soil mix and grown in the greenhouse, followed by transplanting to the experimental field site in Albany, California. The field consisted of three replicate blocks, with each block containing plots of three replicate plants for each of the 200 selected genotypes. At nine weeks post germination, leaf, root and rhizosphere samples were collected from one plant per plot.

As prior studies suggest that the strength of the correlation between host genotype and microbiome composition may vary by sample type in a host species dependent manner, we first sought to determine whether leaf, root or rhizosphere samples were most suitable for downstream GWAS analyses in sorghum. The soil adhering to the surface of the roots, referred to as the rhizosphere, is a critical interface influenced by root, whereas the roots represents the inside of the root tissue. Using a subset of 24 genotypes from our collection of 200 **[Figure 2-1]**, the microbiome composition of leaf, root, and rhizosphere sample types was analyzed using paired-end sequencing of the V3–V4 region of the ribosomal 16S rRNA on the Illumina MiSeq platform (Illumina Inc., San Diego, CA, USA). The resulting dataset demonstrated comparatively high levels or microbial diversity within both root and rhizosphere samples and strong clustering of above and below ground sample types **[Figure 2-2a; Figure 2-2b]**. Three independent Mantel's tests were used to evaluate the degree of correlation between host genotypic distance and microbiome composition for leaf, root and rhizosphere sample types **[Figure 2-2c]**; of the three compartments, only rhizosphere exhibited a significant Mantel's correlation ($R2 = 0.12999624$, p value = 0.0150). Based on these results, subsequent investigation of the microbiomes of the full panel of 200 lines, including heritability and GWAS analyses, was performed using rhizosphere samples.

2.3.2 Heritable principal components and individual taxa within sorghum rhizospheres

To investigate host genotype dependent variation in the sorghum rhizosphere microbiome, the rhizospheres of 598 field grown plants (including three replicates of each of 200 genotypes) were profiled using V3-V4 16S rRNA amplicon sequencing. Compositional analysis of the resulting microbiome dataset (1189 OTUs) exhibited profiles consistent with recent microbiome studies involving the sorghum rhizosphere

(Naylor *et al.*, 2017; Xu *et al.*, 2018) from a variety of field sites, with Proteobacteria, Actinobacteria and Acidobacteria comprising the top three dominant phyla **[Figure 2-3]**.

To quantify the response, we calculated broad sense heritabilities, which is defined as the proportion of phenotypic variance explained by the effects of host genotypes rather than environmental effects (Lynch, Walsh and Others, 1998), of rhizosphere related traits. We include both community-level traits, such as the overall community diversity and eigenvectors produced by ordination analysis (principal component analysis; PCA), and microbial lineage-level traits, which are the abundances of specific taxa comprising the microbial community.

To explore whether microbes with high heritability in the sorghum dataset are phylogenetically clustered, we partitioned the 1189 OTUs into heritable (n=347) and non-heritable fractions (n=842) using an H2 cutoff score of 0.15 **[Figure 2-4a; Table 2-2]**. Several bacterial orders, including *Verrucomicrobiales, Flavobacteriales, Planctomycetales,* and *Burkholderiales*, were observed to have significantly greater numbers of OTUs within the heritable as compared to non-heritable OTU fraction (Fisher's test, $P<0.05$) **[Figure 2-4a; Table 2-2]**. Notably, all 6 Flavobacteriales OTUs were present in the heritable fraction **[Figure 2-4b]**. By comparison, the order Bacillalles contained significantly fewer OTUs in the heritable fraction, but the percentage of read counts attributable to this order was significantly greater for the heritable fraction, suggesting that its heritable members are abundant organisms within the rhizosphere **[Figure 2-4b; Table 2-2]**. Collectively, these data demonstrate that specific bacterial lineages are enriched for microorganisms that are susceptible to host genotypic selection.

2.3.3 Comparative analysis of heritable taxa within sorghum and maize rhizospheres

A recent study of two separate maize microbiomes suggests that specific bacterial lineages are more sensitive to the effect of host genotype than others (Walters *et al.*, 2018). To determine if a bacterial lineage's responsiveness to host genetics is a trait conserved across different plant hosts that diverged more than 11 million years ago (Swigonova *et al.*, 2004), the broad sense heritability (H2) of individual OTUs in our sorghum dataset was evaluated. Broad-sense heritabilities for individual OTUs ranged from 0 to 66%. By comparison, the variance attributable to genotype for individual OTUs in the first of two experiments across 27 inbred maize lines had a maximum of 23% (performed in 2010) , while the second exhibited a maximum variance attributable to genotype of 54% (performed in 2015) (Walters *et al.*, 2018).

We hypothesized that despite the considerable evolutionary distance between maize and sorghum, the bacterial lineages containing OTUs most responsive to host genotypic effects in maize would also contain OTUs exhibiting such susceptibility within sorghum. To test this, we compared the top 100 most heritable OTUs from both maize datasets (referred to as NAM 2010 and NAM 2015) and the sorghum dataset described above. Of the 65 bacterial orders represented by these 300 heritable OTUs, 26 orders contained heritable OTUs in at least two of the datasets, while a total of 15 orders were shared across all three datasets **[Figure 2-5]**. Of the remaining orders, most were unique

to maize; only a few orders were present in the maize heritable fraction (1 in NAM 2010, 10 in NAM 2015, and 1 shared between both) that were only present in the non-heritable fraction of the sorghum dataset **[Figure 2-5]**. To evaluate whether these overlaps represented a significant overrepresentation of heritable orders, we performed a permutation test (n=10,000) in which we resampled 100 random OTUs from the 1189 total sorghum OTUs. Notably, we found the overlap between heritable sorghum OTUs and both the individual and combined heritable maize OTUs to be significant, compared with the resampled sorghum OTUs **[Figure 2-5]**. This result suggests that there is conservation between the bacterial orders being recruited by maize and sorghum. An identification of bacterial orders containing the greatest number of heritable OTUs across all three datasets identified Actinomycetales, Burkholderiales and Myxococcales as the three most frequently represented **[Figure 2-6a]**. These results are likely in part driven by the overall frequency of these lineages within the rhizosphere microbiome, with more common lineages resulting in a greater fraction of heritable microbes due to their ubiquity. To help account for this, we normalized the frequency of heritable sorghum OTUs (n=100) by total sorghum OTU counts (n=1189) belonging to each order **[Figure 2-6b]**. These results demonstrate that while the prevalence of Actinomycetales and Myxococcales among heritable microbes is consistent with their general prevalence in the overall dataset, Burkholderiales (p-value = 0.000984) and two other lineages, including the Verrucomicrobia (p-value = 6.688e-06) and Planctomycetes (p-value = 6.382e-06), exhibited a significant enrichment in the heritable fraction not explained by their overall numbers. Collectively, these data demonstrate a significant overlap in the heritable fraction of the rhizosphere microbiome of two different cereal crops.

2.4 Discussion

2.4.1 Sample type selection for studying the host effect of plant microbiome

Previous studies that have employed a plant-associated microbiome GWAS have most often been conducted with leaf samples, and have had mixed success in identifying loci that correlate with microbiome phenotypes (Horton *et al.*, 2014; Wallace *et al.*, 2018; Roman-reyna *et al.*, 2019). In this study, we compared the overall correlation between host genotype and bacterial microbiome distances across leaf, root, and rhizosphere of *Sorghum bicolor*, and demonstrate that of the three, the rhizosphere represents the most promising compartment for conducting experiments to untangle the heritability of the sorghum microbiome. Notably, the degree of correlation between sorghum phylogenetic distance and microbiome distance was highest in the rhizosphere and lowest in the leaves. This could in part be due to the phyllosphere microbiome's relative simplicity. Even Arabidopsis rosette leaves, which are in close proximity to soil, harbor a distinct and relatively simple bacterial community compared to the root (Bergelson, Mittelstrass and Horton, 2019). By contrast, the rhizosphere represents a direct interface with the highly diverse and populated soil microbiome, which offers a greater pool of microbes upon which the host may exert influence (Bodenhausen, Horton and Bergelson, 2013). Alternatively, the greater correlation observed in the rhizosphere could be due to the plant's reduced ability to select epiphytes in its aboveground microbiome; while the arrival of phyllosphere colonists is largely thought to be driven by wind and rainfall dispersal (Copeland *et al.*, 2015), rhizosphere exudation is known to control chemotaxis and other colonization activities of select members of the surrounding soil environment. This provides a direct mechanism for host selection of its microbial inhabitants prior to direct interaction with the plant surface (Badri *et al.*, 2013; Zhang *et al.*, 2014; Zhalnina *et al.*, 2018). Notably, sorghum is known to be an atypically strong producer of root exudates (Baerson *et al.*, 2008). Given this, it is possible that other plant hosts may demonstrate the greatest selective influence within tissues other than the rhizosphere. Future efforts to investigate host control of the microbiome through GWAS or related techniques would benefit from careful selection of sample type following pilot studies designed to explore heritability across different host tissues.

2.4.2 Heritable microbes within the rhizosphere are phylogenetically clustered and may be similar across hosts

Within the rhizosphere, we demonstrate that a fraction of the microbiome can be considered heritable, and that these heritable taxa show a strong degree of overlap with heritable lineages identified in maize, spanning fifteen different bacterial orders (Walters *et al.*, 2018). In particular, three of these orders, Verrucomicrobiales, Burkholderiales, and Planctomycetales were significantly enriched in the heritable fraction ($H2 > 0.15$) of our dataset. As members of Burkholderiales are known to form strong associations with both plant and animal hosts (Angus *et al.*, 2014; Martínez-Hidalgo and Hirsch, 2017),

and some are well known to colonize only specific members of a host genus or species (Shu *et al.*, 2018), it is feasible that the evolution of such relationships has necessitated additional genetic potential that allows for finer scale discrimination between hosts. In Burkholderia, this could be facilitated by their relatively large pan-genome, with diversity driven by large multi-replicon genomes and abundant genomic islands (Mannaa, Park and Seo, 2018). These observations suggest that heritability may be correlated with the degree to which an organism is symbiotic. This suggests that evaluating bacterial heritability may also be useful in identifying new lineages for which close but previously undetected associations with plant hosts exist. For example, we observed several lineages with high heritability that are common in soil, yet prior evidence of plant-microbe interactions in the literature is lacking, including Verrucomicrobiales (Bergmann *et al.*, 2011) and Planctomycetales (Erbilgin, McDonald and Kerfeld, 2014). Heritability in these lineages might be facilitated by the presence of the Planctomycetes and Verrucomicrobia-type BMC gene cluster, which confers the ability to degrade certain plant polysaccharides (Erbilgin, McDonald and Kerfeld, 2014). However, whether these microbes are behaving as saprophytes or contain plant-beneficial characteristics is yet to be established.

Unfortunately, low phylogenetic resolution of illumina-based 16S sequencing and the use of different 16S rRNA variable regions sequenced between studies precludes our ability to associate heritability with individual microbial species. In cases of crop-associated microbiomes, the ability to distinguish between closely related bacteria is especially important, where beneficial species may be closely related to human pathogens, as is the case with *Burkholderia* spp. (Eberl and Vandamme, 2016). Nonetheless, the overlap of heritable lineages at higher taxonomic classification demonstrates that heritability of certain rhizosphere microorganisms may extend outside an individual host species and be a property of the rhizosphere environment more generally.

Distinctions between the heritability of bacterial lineages identified in our data compared with Walters *et al.*, 2018 also suggests the presence of sorghum and maize-specific heritable lineages. These differences may be partially driven by species-specific exudates. Previous estimations by Bulgarelli *et al.*, 2012 suggest that approximately 40% of Arabidopsis root colonists are attracted by lignocellulosic matrices, while the remaining 60% may be influenced by root exudates. Supporting this hypothesis, sorghum is an atypical exuder (Baerson *et al.*, 2008), and a previous study of 18 plant species found sorghum also produces an atypical rhizosphere microbiome relative to other plant species (Naylor *et al.*, 2017). However, disentangling species-specific patterns of heritability can be challenging; field soils are also a major environmental driver of microbial community diversity (Bulgarelli *et al.*, 2012; Lundberg *et al.*, 2012; Edwards *et al.*, 2015; Walters *et al.*, 2018), and rhizosphere diversity shifts across a growing season (Shi *et al.*, 2016; Edwards *et al.*, 2018; Xu *et al.*, 2018) and between years (Shi *et al.*, 2016; Walters *et al.*, 2018). Because many studies of plant microbiomes do not encompass multiple field sites, plant species, or years, our ability to disentangle which bacterial lineages are field or year specific, rather than broadly species-specific are limited. Additional dissection of the heritable fraction of the rhizosphere microbiome across a wider range of plant hosts and environments will help establish to what degree

the heritable lineages detected in this study constitute the heritable members of the plant rhizosphere more generally.

3.5 Methods

Field experimental design and root microbiome samples collection The experimental field used in this study is an agricultural field site located in Albany, California (37.8864°N, 122.2982°W), characterized by a silty loam soil with pH 5.2 (Naylor *et al.*, 2017). Germplasm for the US sorghum association panel (SAP) used in this study (Casa *et al.*, 2008) were obtained from GRIN (www.ars-grin.gov). To ensure a uniform starting inoculum for all sorghum seedlings and to control their planting density, seeds were first sown into a thoroughly homogenized field soil mix in a growth room with controlled environmental factors (25 °C, 16hr photoperiods) followed by transplanting to the agricultural field. To prepare the soil for seed germination, 0.54 cubic meters of soil was collected at depth of 0 to 20 cm from the field site that was subsequently used for planting. This soil was homogenized by separately mixing 4 equally sized batches with irrigation water in a sterilized cement mixer followed by manually homogenization on a sterilized tarp surface. Soil was then transferred to sterilized 72-cell plant trays. To prepare seeds for planting, seeds were surface-sterilized through soaking 10 min in 10% bleach + 0.1% Tween-20, followed by 4 washes in sterile water. Following planting, sorghum seedlings were watered with approximately 5 ml of water using a mist nozzle every 24 hrs for the first three days, and bottom watered every three days until the 12th day, then transplanted to the field.

The field consisted of three replicate blocks, with each block containing 200 plots for each of 200 selected genotypes. Six healthy sorghum seedlings of each genotype were transplanted to their respective plots at three positions separated by 15.2cm, and thinning to three seedlings per plot was performed at two weeks post transplanting. The subplots were organized in an alternating pattern with respect to the irrigation line to maximize the distance between each plant **[Figure 2-7]**. Plants were watered for one hour, three times per week, using drip irrigation with 1.89 L/hour rate flow emitters, resulting in 5.67 L of water per plant per week. To reduce the impact of weeds on plant growth and microbiome selection, manual weeding was performed three times per week throughout the growing season.

To ensure that the genotypes were at a similar stage of development and that the host-associated microbiome had sufficient time to develop, collection of plant-associated samples was performed nine weeks post germination. Only the middle plant within each subplot was harvested to help mitigate potential confounding plant-plant interaction effects resulting from contact with roots from neighboring plants of other genotypes. Rhizosphere, leaf, and root samples were collected as described previously (Simmons *et al.*, 2018).

Germplasm selection In order to ensure that microbiome profiling was performed on a representative subset of the broad genetic diversity present in the 378 member SAP panel, subsets of 200 genotypes were randomly sampled from the 378 member panel 10,000 times and aggregate nucleotide diversity score was calculated for each using the R package “PopGenome”. From these data, the subset of 200 lines with the maximum diversity value was selected **[Figure 2-1; Table 2-1]**. For the pilot

experiment used to determine the appropriate sample type for GWAS, a subset of 24 lines were selected that included genotypes from a wide range of phylogenetic distances **[Figure 2-1]**. The lines used in this study are listed in **[Table 2-1]**. The phylogenetic tree of sorghum accessions was generated using the online tool: Interactive Tree Of Life (iTOL) v5 (Letunic and Bork, 2019).

DNA extraction and PCR amplification DNA extraction for all samples was performed using extraction kits (MoBio PowerSoil DNA Isolation Kit, MoBio Inc., Carlsbad, CA) following the manufacturer's protocol. We amplified the V3-V4 region of 16S rRNA gene using a dual-indexed 16s rRNA Illumina iTags primer (341F (5'-CCTACGGGNBGCASCAG-3') and 785R (5'-GACTACNVGGGTATCTAATCC-3') as described in (Xu *et al.*, 2018) using 5-Prime Hot Master Mix (catalog No. 2200410). After DNA extraction, DNA samples were diluted to 5 ng/ µl and randomized in 96-well plates. Two wells received water rather than template on each 96-well plate as negative controls. PNA clamps described in were used to minimize host-derived amplicons from both chloroplast and mitochondrial 16S rRNA gene sequences (Lundberg *et al.*, 2013). Reactions included 11.12 µL DNase-free sterile H20, 0.4 µg BSA, 10.0 µL 5-Prime Hot Master Mix, and 2 µL template, and 0.75 µM of chloroplast and mitochondria PNAs. PCR reactions were performed in triplicate in three thermocyclers (to account for possible thermocycler bias) with the following conditions: initial 3 min cycle at 94°C, then 30 cycles of 45 seconds at 94°C, 10 sec at 78°C, 1 min at 50°C, and 1.5 min at 72°C, followed by a final cycle of 10 min at 72°C. Triplicates were then pooled (128 samples per library) and DNA concentration for each sample was quantified using a Qubit reader with the Broad Range kit (ThermoFisher Scientific, Waltham, MA). Pools of amplicons were constructed using 100 ng for each PCR product. Before submitting for sequencing, pooled samples were cleaned up with 1.0X volume Agencourt AMPureXP (Beckman-Coulter, West Sacramento, CA) beads according to the manufacturer's directions, except for the modifications of using 1.0X rather than 1.6X volume beads per sample, dispensing 1500 µL 70% EtOH to each well rather than 200 µL, and eluting in 100 µL DNase-free H20 rather than 40 µL. DNA extractions, PCR amplification, and amplicon pooling were performed as described previously (Simmons *et al.*, 2018). An aliquot of the pooled amplicons was diluted to 10 nM in 30µL total volume before submitting to the QB3 Vincent J. Coates Genomics Sequencing Laboratory facility at the University of California, Berkeley for sequencing using Illumina Miseq 300bp pair-end with v3 chemistry. Sequences were returned demultiplexed and with adaptors removed.

Amplicon sequence processing and taxonomic assignment Sequencing data were analyzed using the iTagger pipeline to obtain OTUs (Bolyen *et al.*, 2019). The iTagger pipeline was developed by the U.S. Department of Energy's Joint Genome Institute, which wraps several packages for the filtering, merging, clustering and taxonomy assignment, including CUTADAPT, FLASH, USEARCH, and RDP. In brief, after filtering 80,323,748 16S rRNA raw reads for known contaminants (Illumina adapter sequence and PhiX), primer sequences were trimmed from the 5' ends of both forward and reverse reads. Low-quality bases were trimmed from the 3' ends prior to assembly of forward and reverse reads with FLASH55. The remaining 66,524,452 high-quality merged reads were clustered with simultaneous chimera removal using UPARSE58.

After clustering, 37,819,153 read counts mapped to 10,006 OTUs at 97% identity. The resulting reads produced on average approximately 42974, 41162 and 34445 reads per sample for soils, rhizospheres, and roots respectively. Taxonomies were assigned to each OTU using the RDP Naïve Bayesian Classifier with custom reference databases. For the 16S rRNA V3-V4 data, this database was compiled from the May 2013 version of the GreenGenes 16S database v13, trimmed to the V3-V4 region. After taxonomies were assigned to each OTU, OTUs were discarded if they were not assigned a Kingdom level RDP classification score of at least 0.5, or if they were not assigned to Kingdom Bacteria. To remove low abundance OTUs that are in many cases artifacts generated through the sequencing process, OTUs without at least 3 reads in at least 3 samples were removed. Samples that had less than 10,000 reads were also removed, which yielded 3,818 high-abundance OTUs for downstream analyses. To account for differences in sequencing read depth across samples, all samples were normalized to an even read depth of 15,000 reads per sample random subsampling for specific analyses, or alternatively, by dividing the reads per OTU in a sample by the sum of usable reads in that sample, resulting in a table of relative abundance frequencies. OTUs, which were reduced to less than one read per OTU after rarefaction, were discarded to yield a total of measurable, normalized reads for downstream analysis. The raw sequencing reads for this project will be deposited in the NCBI Short Read Archive.

Estimates of broad sense heritability (H2) of OTU abundance in rhizosphere

To calculate the broad-sense heritability (H2) for individual OTU abundances, we fitted the following linear mixed model to OTU abundances of each individual OTU (n=1189) following a cumulative sum scaling (CSS) (Paulson *et al.*, 2013) normalization procedure that adjusted for differences in sequencing depth and fit a normal distribution:

$$Y_{ijk} = u + Gi + Rj + Bjk + e$$

In this model for a given OTU, Y_{ijk} denotes the OTU abundance of the i^{th} genotype evaluated in the kth block of the jth replicate; u denotes the overall mean; Gi is the random effect of the ith genotype; Rj is the random effect of the jth replicate; Bjk is the random effect of the kth block nested within the jth replicate; e denotes the residual error. To account for the spatial effects in the field, additional spatial variables were fitted as random effects using 2-dimensional splines in the above model using an R add-on package "sommer" (Covarrubias-Pazaran, 2016). Broad-sense heritability was estimated as the amount of variance explained by the genotype term (V_G) relative to the total variance ($V_G + V_E/j$). Here j is the number of replications. To get the null distribution of H2, each OTU was randomly shuffled 1,000 times and then fitted to the same model as described above. Permutation p value was calculated as the probability of the permuted H2 values bigger than the observed H2.

Figure 2-1

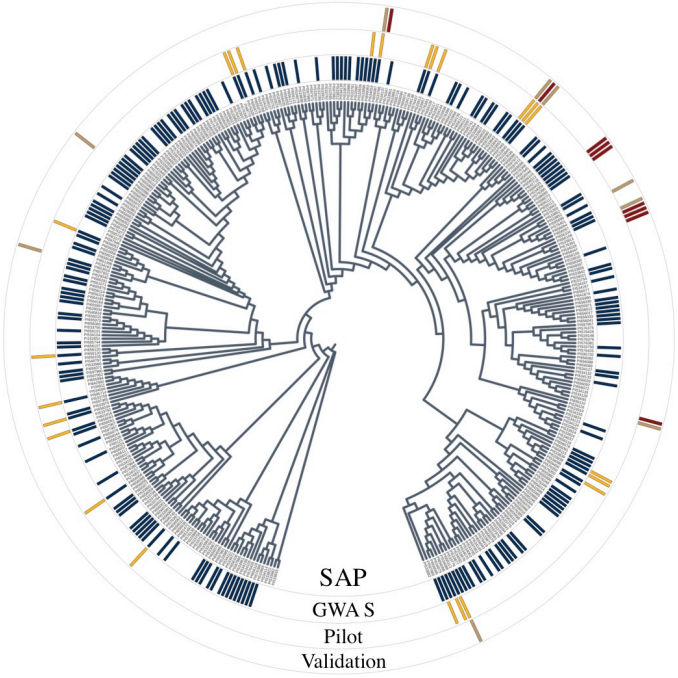

Figure 2- 1. Experimental design and line selection

Phylogenetic tree representing the 378 member sorghum association panel (SAP, inner ring), subset of 200 lines selected for GWAS (2nd ring from the center, in blue), 24 lines used for sample type selection (Pilot, 3rd ring from the center, in yellow), and 18 genotypes used for GWAS validation containing either the Chromosome 4 minor allele (red) or major allele (brown) identified by GWAS (outer ring).

Figure 2-2

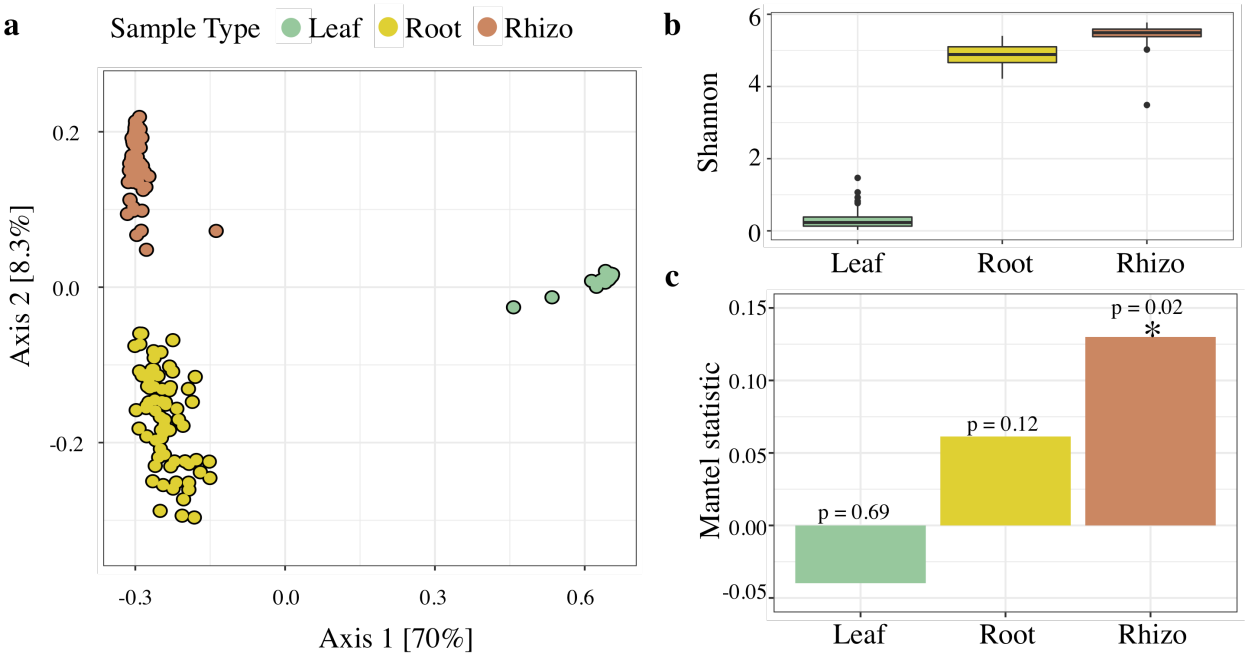

Figure 2- 2. Selection of sample type for quantitative genetic analyses

a. Shannon's Diversity values from 16S rRNA amplicon datasets for the leaf (green), rhizosphere (red), and root (yellow) sample types across all 24 genotypes used in the pilot experiment. **b.** Principal coordinate analysis generated using Bray-Curtis distance for the 24 genotypes across leaf (green), rhizosphere (red), and root (yellow). **c.** Mantel's R statistic plotted for each sample type indicating the degree of correlation between host genotypic distance and microbiome distance.

Figure 2-3

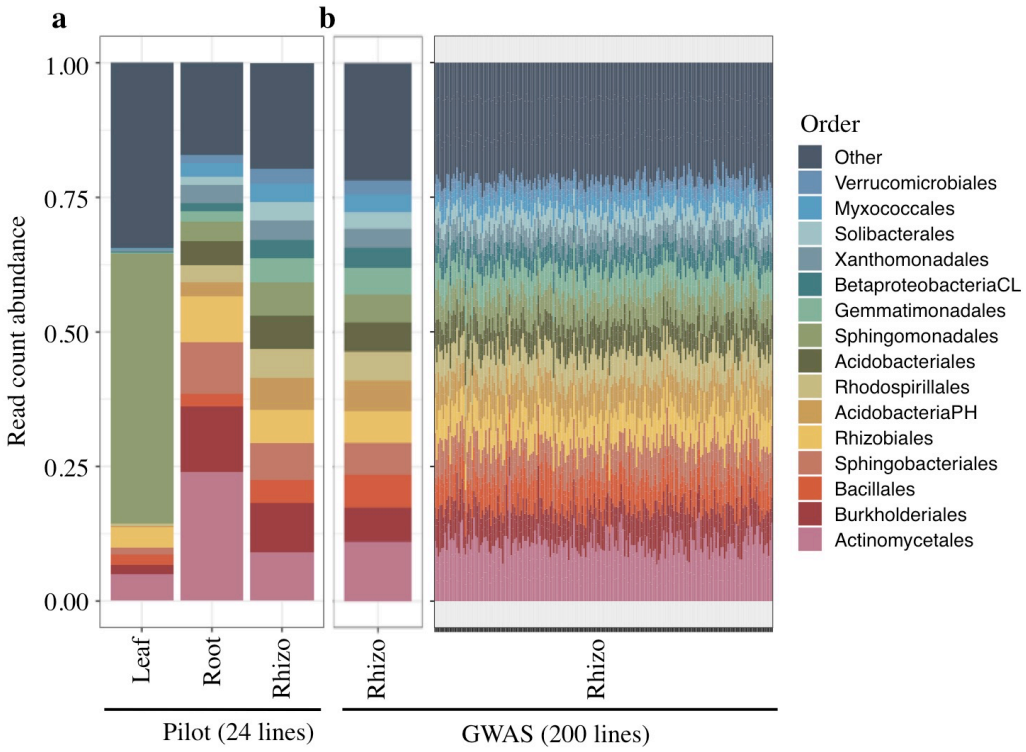

Figure 2- 3. Composition of the sorghum microbiome across sample types

a. Order level relative abundances of sorghum leaf, root, and rhizosphere (rhizo) microbiomes used for GWAS sample type selection (Pilot). **b.** Relative abundance of the top 15 orders in the 200 rhizosphere microbiomes used for GWAS, displayed both as an average relative abundance and separated by individual line.

Figure 2-4

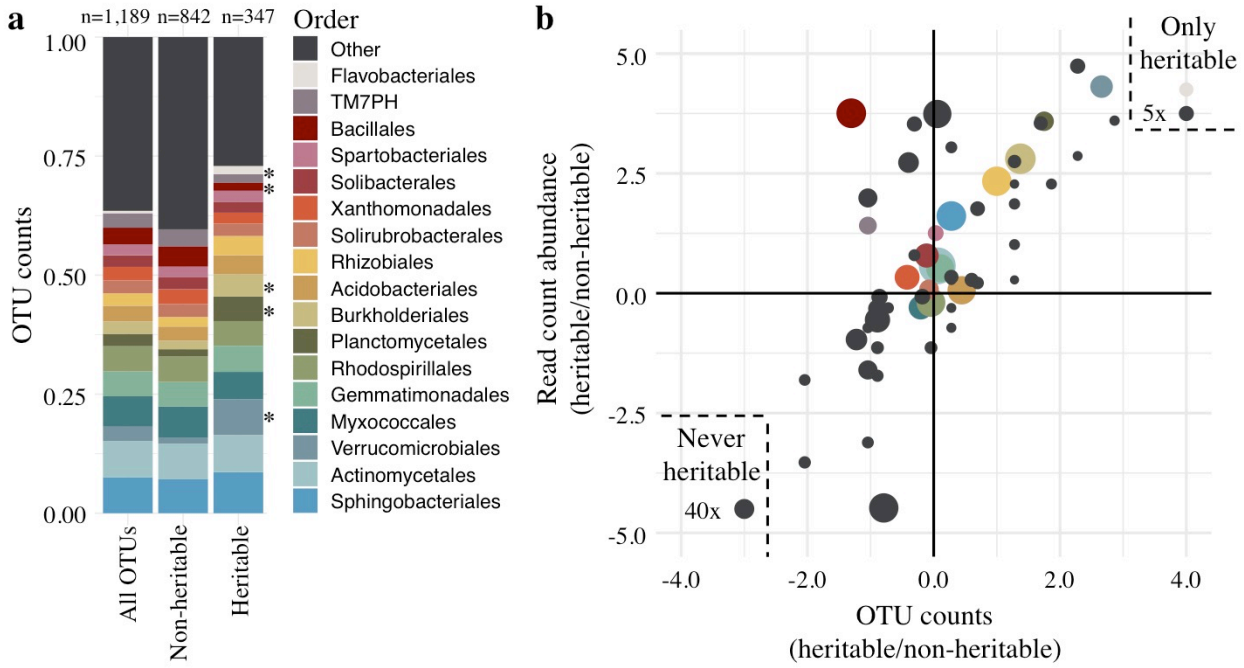

Figure 2- 4. Taxonomic classification of heritable rhizosphere microbes

a. The relative percentage of total OTUs belonging to each of the top 17 bacterial orders for all OTUs (left bar), non-heritable OTUs (middle bar), or heritable OTUs (right bar). Orders with significantly greater number of OTUs in the heritable as compared to the non-heritable fraction as determined by Fisher's exact test ($P<0.05$) are indicated with asterisks. **b.** Order-level scatterplot of the ratio between heritable and non-heritable OTU counts (x-axis) and OTU abundance (y-axis). Size of the points represents the total number of read counts represented by each bacterial order. Points outside the dashed lines indicate bacterial orders that were present only in the heritable (n=6) or non-heritable (n=40) fractions.

Figure 2-5

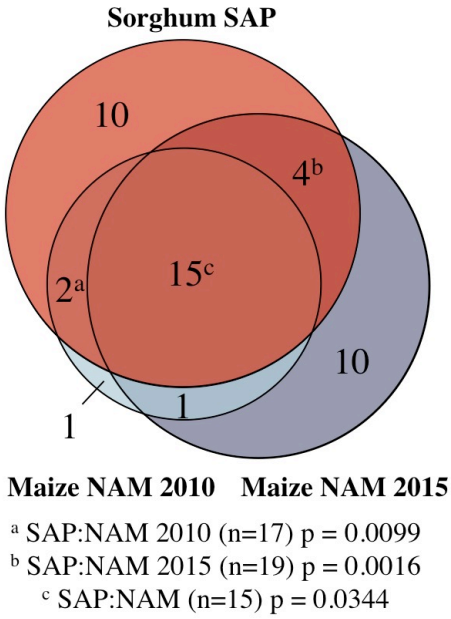

Figure 2- 5. Comparison of the top 100 most heritable OTUs between maize and sorghum dataset

Proportional Venn diagram of bacterial orders containing heritable OTUs identified in this study (Sorghum SAP), compared with those found in a large-scale field study of maize nested association mapping (NAM) parental lines grown over two separate years, published in Walters et al., 2018. The top 100 heritable OTUs (based on H2) from each dataset were classified at the taxonomic rank of order to generate the Venn diagram. NAM heritable orders only present in the SAP non-heritable fraction are represented by the blue sections. Superscript letters indicate the frequency that a random subsampling of 100 sorghum OTUs (10,000 permutations) overlap with maize OTUs from either single year (a/b) or both (c).

Figure 2-6

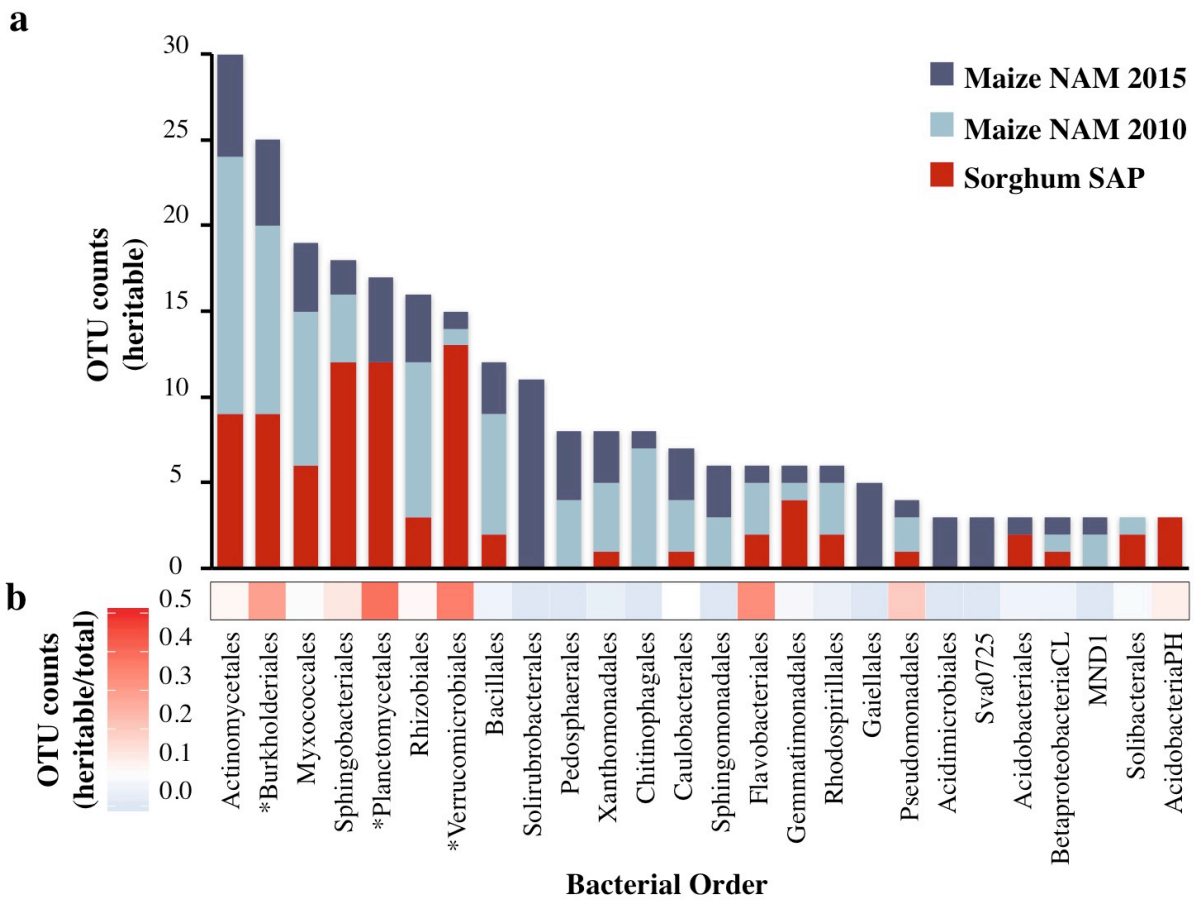

Figure 2- 6. Heritability of rhizosphere microbes across maize and sorghum

a. Stacked barplot displaying cumulative counts (y-axis) of OTUs identified as heritable in any of the three datasets for all bacterial orders (x-axis), which have a total of at least three heritable OTUs. **b.** The fraction of heritable sorghum OTUs relative to all sorghum OTUs within each order are displayed as a heatmap. Asterisks indicate orders enriched in heritable OTUs (Fisher's exact test, $P<0.001$).

Figure 2-7

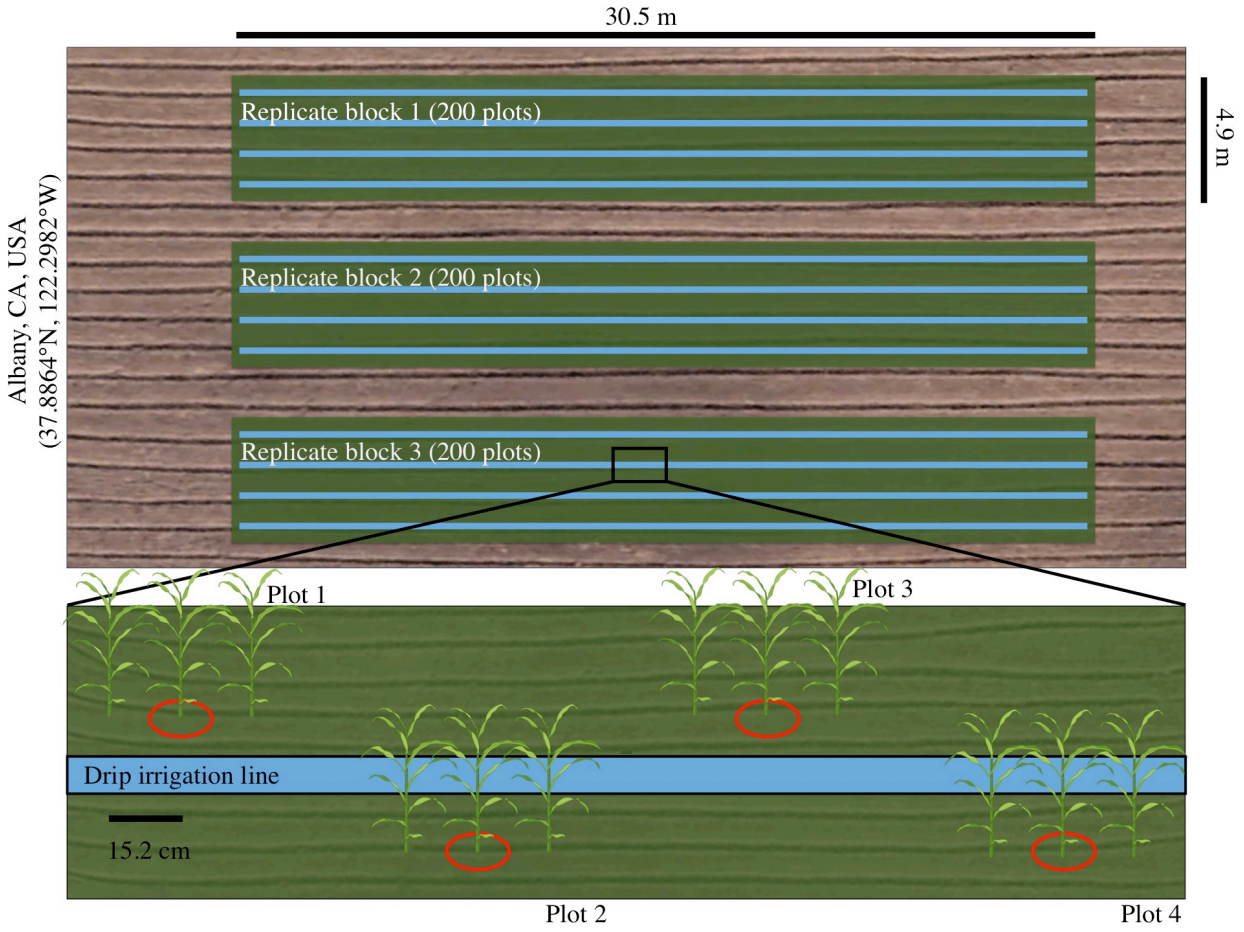

Figure 2- 7. Field experimental design

Three replicate blocks, each containing 200 plots. Each plot contains a different sorghum genotype with three replicate plants. Leaf, root, and rhizosphere from the center plant of each plot were harvested (Red circles). Sorghum plant models were created with Biorender.com.

Table 2-1

Table 2- 1. 200 genotypes selected from the Sorghum Association Panel (SAP)

Subsets of 200 genotypes were randomly sampled from the SAP germplasm collection (Casa *et al.,* 2008).

SAP ID	Line name	Country	Grain type
PI48770	WHITE KAFIR	United States	na
PI533750	K.3 Perimanjial Irungu Cholam	India	bicolor
PI533776	KA 3	na	na
PI533789	Kodilib	Sudan	caudatum-guinea
PI533800	IS 12666C	Ethiopia	caudatum
PI533810	Karad 2-7-11	India	durra
PI533838	BA45 Faria Bonkum	na	na
PI533839	K037 Camjin	Nigeria	guinea
PI533841	ZA41 Danye	na	na
PI533842	Nandyal	India	durra
PI533843	SV 34	na	na
PI533845	EC 18246 (preconverted)	na	na
PI533852	Cholia Talijhari	na	na
PI533855	Chanan Singoo	India	bicolor-guinea
PI533866	255 Tirter	Sudan	bicolor-caudatum
PI533869	Msumbji SB 117	Tanzania	guinea
PI533876	KA 12 Janjari	Nigeria	caudatum
PI533877	KA 15 Yazgar Giwa	Nigeria	caudatum
PI533902	Orange No. l, Baijo	Ethiopia	bicolor-durra
PI533912	Mendo	Sudan	caudatum
PI533913	Nyan Dok	Sudan	caudatum-guinea
PI533927	J.A.T.S. #67	Ethiopia	bicolor
PI533936	1903 AS 4633	Tanzania	guinea
PI533938	AS 5826 Holcus	Zaire	caudatum
PI533939	AS 4055 N Kambwa	Mosambique	caudatum
PI533940	AS 4601 Pawaga	Tanzania	bicolor
PI533956	MN 586 (preconverted)	Congo	bicolor-durra
PI533961	Wit Lichtenburg DL/59/1530	South Africa	caudatum
PI533965	SAP-148	Uganda	bicolor-caudatum
PI533967	SB-283	Uganda	caudatum
PI533970	2033Z-3	Uganda	caudatum
PI533976	Framiola DL/59/1539	South Africa	caudatum-kafir
	Bulfontein White Kafir Corn		
PI533979	DL/60/133	South Africa	kafir
PI533985	Barking 119	Sudan	caudatum
PI533986	Huria White 621	Sudan	caudatum
PI533987	Kireniga 317	Sudan	caudatum
PI533991	Sinidyil 177	Sudan	caudatum-guinea
PI533996	Hamaisi 38	Sudan	caudatum-durra
PI533997	MN 1592 (preconverted)	Arabia	durra-guinea
PI533998	Brawley	United States	bicolor
PI534009	Yerra Jonna Goda	India	durra
PI534021	Jola Nandyal	India	durra
PI534053	T 28	Uganda	caudatum
PI534063	BO 36	na	na
PI534070	BE 25	Nigeria	guinea
PI534075	KA 24	Nigeria	caudatum
PI534088	ZA 6	Nigeria	caudatum
PI534092	ZA 71	Nigeria	caudatum

PI534101	A-106	Japan	caudatum
PI534105	EC 21361 G30	Uganda	caudatum-kafir
PI534108	EC 21428 SB 63	Uganda	caudatum-kafir
PI534112	EC 21463 STR 5/1	Uganda	caudatum
PI534123	No. 25 Gobo, Kaichama Ethiopia	Ethiopia	bicolor-durra
PI534127	No. 37 Ubi, Abelti Ethiopia	Ethiopia	bicolor-kafir
PI534132	No. 59 Bekedjie, Kembolcha Ethiopia	Ethiopia	durra
PI534139	Tuery 11	Sudan	caudatum-guinea
PI534167	F.R. Miller	na	na
PI542718	SAN CHI SAN	na	na
PI548797	Tx2891	na	na
PI561071	RTx436	na	na
PI564164	RTX433	na	na
PI564165	RTX434	na	na
PI576332	SAP-287	China	bicolor
PI576333	SAP-306	United States	guinea-kafir
PI576339	SAP-312	Zimbabwe	caudatum-kafir
PI576345	SAP-294	South Africa	kafir
PI576347	SAP-342	United States	bicolor
PI576348	SAP-341	United States	bicolor
PI576349	SAP-343	United States	bicolor
PI576350	SAP-417	na	na
PI576352	Marupantse	Botswana	caudatum-kafir
PI576359	Butivori	India	durra-guinea
PI576364	Chari Uri	India	caudatum
PI576366	Jowar Red Jankinagar	India	bicolor-durra
PI576373	SAP-325	Japan	bicolor-caudatum
PI576385	SAP-135	Nigeria	na
PI576386	SAP-347	Uganda	na
PI576387	Awanlek	Sudan	na
PI576390	Kharuth Waragel	India	durra
PI576393	MN 708 (preconverted)	Ethiopia	caudatum
PI576394	Lambas	Sudan	caudatum
PI576399	SAP-224	Sudan	caudatum
PI576418	SAP-147	Nigeria	guinea
PI576422	SAP-141	South Africa	kafir
PI576435	Lula	Uganda	bicolor-kafir
PI576437	SAP-155	Brazil	na
PI595699	SAP-340	India	durra
PI595702	SAP-311	Zimbabwe	caudatum-kafir
PI595714	SAP-139	Sudan	caudatum
PI595720	SAP-149	Ethiopia	bicolor-durra
PI595741	SAP-134	Senegal	guinea
PI595743	SAP-154	na	caudatum-guinea
PI595744	SAP-157	Guatemala	caudatum-kafir
PI597949	SAP-146	India	guinea
PI597950	SAP-250	India	durra-guinea
PI597951	SAP-264	West Volta	guinea
PI597952	SAP-323	Sudan	caudatum
PI597957	SAP-151	Ethiopia	bicolor-durra
PI597960	SAP-137	Nigeria	bicolor-caudatum
PI597964	SAP-166	Ethiopia	caudatum
PI597966	SAP-158	Venezuela	caudatum
PI597972	SAP-398	na	na
PI597973	SAP-171	Sudan	bicolor-durra
PI597976	SAP-172	Mali	guinea
PI597980	SAP-173	Mali	caudatum
PI597982	SAP-175	Sudan	caudatum

PI598069	RTx2909	na	na
PI607931	Tx2911	na	na
PI609456	SAP-386	na	guinea
PI629040	Tx2917	na	na
PI629059	BTx2928	na	na
PI641849	WACONIA	na	na
PI641874	DAY MILO	na	na
PI642998	Black Spanish	na	na
PI651492	COWLEY	na	na
PI653616	WRAY	na	na
PI655970	Standard Blackhull Kafir	na	na
PI655971	STANDARD WHITE MILO	na	na
PI655973	SPUR FETERITA	na	na
PI655974	TEXAS BLACKHULL KAFIR	na	na
PI655975	WHEATLAND	na	na
PI655976	RED KAFIR	na	na
PI655977	TAM2566	na	na
PI655979	Tx2741	na	na
PI655980	Tx2785	na	na
PI655981	SAP-50	na	na
PI655982	SAP-380	na	na
PI655983	SUGAR DRIP	na	na
PI655985	PLAINSMAN	na	na
PI655986	CAPROCK	na	na
PI655987	MARTIN	na	na
PI655988	COMBINE KAFIR-60	na	na
PI655989	REDBINE-60	na	na
PI655991	TX 378 (REDLAN) B LINE	na	na
PI655995	DEER	na	na
PI655997	R TX 431	na	na
PI656000	RTx432	na	na
PI656001	TX2783	na	na
PI656010	RTx2536	na	na
PI656012	90M	na	na
PI656013	94Q63	na	na
PI656014	Acme Broomcorn	na	na
PI656015	Ajabsido	na	na
PI656016	B KS66	na	na
PI656017	100M	na	na
PI656018	BTx2752	na	na
PI656019	BTx3042	na	na
PI656022	BTx615	na	na
PI656024	SEPON82	na	na
PI656025	Shan Qui Red	na	na
PI656028	Town	na	na
PI656029	BTx642	na	na
PI656030	58M	na	na
PI656031	CE151-262-A1	na	na
PI656033	60M	na	na
PI656034	Dorado	na	na
PI656035	El Mota	na	na
PI656036	ICSV 1089BF	na	na
PI656041	80M	na	na
PI656043	KAT83369	na	na
PI656046	LianTang Ai	na	na
PI656048	Malisor 84-7	na	na
PI656049	Marupantse	na	na
PI656050	Mota Maradi	na	na

PI656051	MR732	na	na
PI656052	N250B	na	na
PI656053	N290B	na	na
PI656056	P850029	na	na
PI656058	P9517	na	na
PI656059	Pinolero 1	na	na
PI656064	Chiragon-2	na	na
PI656067	Rambharose	na	na
PI656069	SC 326-6	na	na
PI656070	SC 748-5	na	na
PI656071	SC 1019	na	na
PI656075	SC 1251	na	na
PI656076	SC 1271	na	na
PI656078	SC 1424	na	na
PI656079	SC 1426	na	na
PI656081	SC 1439	na	na
PI656083	SC 1451	na	na
PI656085	SC 1465	na	na
PI656087	SC 1476	na	na
PI656093	SC 295	na	na
PI656094	SC 301	na	na
PI656095	SC 373	na	na
PI656096	SC 391	na	na
PI656101	SC 525	na	na
PI656103	SC 610	na	na
PI656104	SC 621	na	na
PI656106	SC 695	na	na
PI656107	SC 702	na	na
PI656113	SC 1277	na	na
PI656114	SC 134	na	na
PI656115	SC 1440	na	na
PI656118	SC 332	na	na
PI656120	SC 790	na	na
PI656121	SC 947	na	na

Table 2-2

Table 2- 2. The broad-sense heritability (H2) of individual OTUs

The IDs are the OTU ID from itagger pipeline. Taxonomies were assigned to each OTU using the RDP Naïve Bayesian Classifier with custom reference databases (See method for detailed information). 347 OTUs with heritability scores > 0.15 are shown here.

ID	H2	Taxonomy Information			
		Phylum	Class	Order	Family
X35	0.66	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae
X59	0.59	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae
X9	0.57	Chloroflexi	Ktedonobacteria	KtedonobacteriaCL	KtedonobacteriaCL
X565	0.54	Bacteroidetes	Sphingobacteria	Sphingobacteriales	SphingobacterialesOR
X7	0.54	unclassified	unclassified	unclassified	unclassified
X61	0.49	Bacteroidetes	Sphingobacteria	Sphingobacteriales	SphingobacterialesOR
X155	0.49	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae
X5	0.47	Actinobacteria	ActinobacteriaPH	Actinomycetales	Catenulisporaceae
X353	0.46	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Sphingobacteriaceae
X4	0.46	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae
X171	0.45	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae
X265	0.41	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Flexibacteraceae
X157	0.41	Actinobacteria	ActinobacteriaPH	Actinomycetales	Streptomycetaceae
X359	0.41	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae
X300	0.41	unclassified	unclassified	unclassified	unclassified
X425	0.4	unclassified	unclassified	unclassified	unclassified
X650	0.39	Acidobacteria	AcidobacteriaPH	Acidobacteriales	AcidobacterialesOR
X159	0.38	unclassified	unclassified	unclassified	unclassified
X250	0.38	Acidobacteria	AcidobacteriaPH	Acidobacteriales	AcidobacterialesOR
X74	0.38	Proteobacteria	Betaproteobacteria	Burkholderiales	BurkholderialesOR
X225	0.38	Proteobacteria	Deltaproteobacteria	Myxococcales	MyxococcalesOR
X892	0.38	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Flexibacteraceae
X2	0.37	Firmicutes	Bacilli	Bacillales	Bacillaceae
X131	0.37	Actinobacteria	ActinobacteriaPH	Actinomycetales	Nocardioidaceae
X51	0.37	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Sinobacteraceae
X484	0.36	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	Gemmatimonadaceae
X494	0.36	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Flexibacteraceae
X132	0.36	Actinobacteria	ActinobacteriaPH	Actinomycetales	Nocardiaceae
X730	0.36	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae
X521	0.35	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae
X30	0.35	Actinobacteria	ActinobacteriaPH	Actinomycetales	Frankiaceae
X343	0.35	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X758	0.35	Actinobacteria	ActinobacteriaPH	Actinomycetales	Nocardioidaceae

X566	0.35	unclassified	unclassified	unclassified	unclassified
X245	0.35	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
X509	0.35	Acidobacteria	Solibacteres	Solibacterales	Solibacteraceae
X212	0.35	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae
X662	0.35	Bacteroidetes	Flavobacteria	Flavobacteriales	Cryomorphaceae
X179	0.34	Proteobacteria	Deltaproteobacteria	Myxococcales	Haliangiaceae
X204	0.34	unclassified	unclassified	unclassified	unclassified
X1215	0.34	Acidobacteria	AcidobacteriaPH	AcidobacteriaPH	Koribacteraceae
X340	0.34	Proteobacteria	Alphaproteobacteria	Sphingomonadales	SphingomonadalesOR
X318	0.33	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae
X85	0.33	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X139	0.33	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X557	0.33	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X384	0.32	unclassified	unclassified	unclassified	unclassified
X551	0.32	TM7	TM7PH	TM7PH	TM7PH
X702	0.32	unclassified	unclassified	unclassified	unclassified
X515	0.32	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X593	0.32	Chloroflexi	Thermomicrobia	ThermomicrobiaCL	ThermomicrobiaCL
X31	0.32	Bacteroidetes	Sphingobacteria	Sphingobacteriales	SphingobacterialesOR
X355	0.32	unclassified	unclassified	unclassified	unclassified
X1101	0.32	Proteobacteria	Deltaproteobacteria	Myxococcales	Polyangiaceae
X473	0.31	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae
X370	0.31	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Flexibacteraceae
X246	0.31	unclassified	unclassified	unclassified	unclassified
X485	0.31	Acidobacteria	Solibacteres	Solibacterales	Solibacteraceae
X947	0.31	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae
X1313	0.31	Proteobacteria	Betaproteobacteria	Rhodocyclales	unclassified
X363	0.31	unclassified	unclassified	unclassified	unclassified
X14	0.31	Bacteroidetes	Sphingobacteria	Sphingobacteriales	SphingobacterialesOR
X129	0.31	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae
X123	0.31	Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae
X71	0.31	Actinobacteria	ActinobacteriaPH	MC47	MC47OR
X46	0.31	unclassified	unclassified	unclassified	unclassified
X369	0.31	Bacteroidetes	Sphingobacteria	Sphingobacteriales	SphingobacterialesOR
X277	0.31	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X404	0.31	unclassified	unclassified	unclassified	unclassified
X475	0.31	Proteobacteria	Deltaproteobacteria	Syntrophobacterales	Syntrophobacteraceae
X744	0.3	Proteobacteria	Gammaproteobacteria	Thiotrichales	ThiotrichalesOR
X128	0.3	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
X683	0.3	Actinobacteria	ActinobacteriaPH	Actinomycetales	Actinospicaceae
X342	0.3	Proteobacteria	Alphaproteobacteria	Rhizobiales	Xanthobacteraceae
X1422	0.3	Proteobacteria	Betaproteobacteria	BetaproteobacteriaCL	BetaproteobacteriaCL
X360	0.3	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR

X1056	0.3	Proteobacteria	Alphaproteobacteria	AlphaproteobacteriaCL	AlphaproteobacteriaCL
X232	0.3	Proteobacteria	Deltaproteobacteria	Myxococcales	MyxococcalesOR
X135	0.3	Proteobacteria	Deltaproteobacteria	Myxococcales	Haliangiaceae
X659	0.3	Chloroflexi	Anaerolineae	A4b	A4bOR
X449	0.29	Proteobacteria	Alphaproteobacteria	Rhizobiales	RhizobialesOR
X181	0.29	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	Gemmatimonadaceae
X577	0.29	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae
X1380	0.29	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X110	0.29	Bacteroidetes	Sphingobacteria	Sphingobacteriales	SphingobacterialesOR
X13	0.29	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X1333	0.29	Proteobacteria	Deltaproteobacteria	Myxococcales	Haliangiaceae
X455	0.29	Firmicutes	Bacilli	Bacillales	Alicyclobacillaceae
X317	0.28	Acidobacteria	AcidobacteriaPH	AcidobacteriaPH	Koribacteraceae
X1304	0.28	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	Gemmatimonadaceae
X253	0.28	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae
X895	0.28	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	Gemmatimonadaceae
X237	0.28	Proteobacteria	Betaproteobacteria	Rhodocyclales	RhodocyclalesOR
X312	0.28	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X572	0.28	unclassified	unclassified	unclassified	unclassified
X117	0.28	Actinobacteria	ActinobacteriaPH	Actinomycetales	Mycobacteriaceae
X830	0.28	Verrucomicrobia	Spartobacteria	Spartobacteriales	Spartobacteriaceae
X210	0.28	Verrucomicrobia	Spartobacteria	Spartobacteriales	Spartobacteriaceae
X813	0.28	Acidobacteria	AcidobacteriaPH	AcidobacteriaPH	AcidobacteriaPH
X535	0.28	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X928	0.28	Bacteroidetes	Sphingobacteria	Sphingobacteriales	SphingobacterialesOR
X329	0.28	Chloroflexi	Thermomicrobia	ThermomicrobiaCL	ThermomicrobiaCL
X249	0.28	Actinobacteria	ActinobacteriaPH	Actinomycetales	Nocardioidaceae
X362	0.28	Verrucomicrobia	Opitutae	Opitutales	Opitutaceae
X178	0.27	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae
X890	0.27	Acidobacteria	AcidobacteriaPH	AcidobacteriaPH	AcidobacteriaPH
X222	0.27	Actinobacteria	ActinobacteriaPH	Acidimicrobiales	AcidimicrobialesOR
X235	0.27	Proteobacteria	Deltaproteobacteria	Myxococcales	Haliangiaceae
X612	0.27	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X109	0.27	Bacteroidetes	Sphingobacteria	Sphingobacteriales	SphingobacterialesOR
X173	0.27	Acidobacteria	Chloracidobacteria	ChloracidobacteriaCL	ChloracidobacteriaCL
X737	0.27	Chloroflexi	Ktedonobacteria	KtedonobacteriaCL	KtedonobacteriaCL
X859	0.27	Acidobacteria	Solibacteres	Solibacterales	Solibacteraceae
X1042	0.27	unclassified	unclassified	unclassified	unclassified
X694	0.26	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae
X795	0.26	Chloroflexi	SOGA31	SOGA31CL	SOGA31CL
X856	0.26	Actinobacteria	ActinobacteriaPH	Actinomycetales	Pseudonocardiaceae
X116	0.26	Verrucomicrobia	Opitutae	Opitutales	Opitutaceae
X18	0.26	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae

X457	0.26	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae
X323	0.26	Chloroflexi	ChloroflexiPH	Herpetosiphonales	Herpetosiphonaceae
X458	0.26	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	Gemmatimonadaceae
X272	0.26	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae
X88	0.26	Bacteroidetes	Sphingobacteria	Sphingobacteriales	SphingobacterialesOR
X411	0.26	Nitrospirae	NitrospiraePH	Nitrospirales	Nitrospiraceae
X162	0.26	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	Gemmatimonadaceae
X113	0.26	Chloroflexi	SOGA31	SOGA31CL	SOGA31CL
X91	0.25	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
X185	0.25	TM7	TM7PH	TM7PH	TM7PH
X442	0.25	Actinobacteria	ActinobacteriaPH	Solirubrobacterales	SolirubrobacteralesOR
X1196	0.25	TM7	TM7PH	TM7PH	TM7PH
X933	0.25	Bacteroidetes	Sphingobacteria	Sphingobacteriales	SphingobacterialesOR
X571	0.25	Chloroflexi	Thermomicrobia	ThermomicrobiaCL	ThermomicrobiaCL
X142	0.25	Actinobacteria	ActinobacteriaPH	Actinomycetales	Micromonosporaceae
X487	0.24	Verrucomicrobia	Opitutae	Opitutales	Opitutaceae
X970	0.24	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X729	0.24	Proteobacteria	Deltaproteobacteria	Desulfuromonadales	Geobacteraceae
X759	0.24	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X1125	0.24	Proteobacteria	Deltaproteobacteria	Myxococcales	Myxococcaceae
X1008	0.24	Actinobacteria	ActinobacteriaPH	Actinomycetales	ActinomycetalesOR
X594	0.24	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Sphingobacteriaceae
X1112	0.24	Verrucomicrobia	Spartobacteria	Spartobacteriales	Spartobacteriaceae
X763	0.24	Bacteroidetes	Sphingobacteria	Sphingobacteriales	SphingobacterialesOR
X393	0.24	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae
X862	0.24	Elusimicrobia	ElusimicrobiaPH	FAC88	FAC88OR
X168	0.24	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae
X675	0.23	Actinobacteria	ActinobacteriaPH	Solirubrobacterales	Patulibacteraceae
X188	0.23	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae
X595	0.23	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
X238	0.23	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
X72	0.23	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	Gemmatimonadaceae
X211	0.23	Chloroflexi	SOGA31	SOGA31CL	SOGA31CL
X197	0.23	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae
X94	0.23	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae
X422	0.23	Proteobacteria	Alphaproteobacteria	Rhizobiales	Phyllobacteriaceae
X489	0.23	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae
X1190	0.23	Acidobacteria	AcidobacteriaPH	Acidobacteriales	AcidobacterialesOR
X153	0.23	Acidobacteria	AcidobacteriaPH	Acidobacteriales	Acidobacteriaceae
X375	0.23	Chloroflexi	ChloroflexiPH	Roseiflexales	Kouleothrixaceae
X434	0.23	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X1171	0.23	Bacteroidetes	Sphingobacteria	Sphingobacteriales	unclassified
X143	0.23	Bacteroidetes	Sphingobacteria	Sphingobacteriales	SphingobacterialesOR

X1065	0.23	Planctomycetes	Planctomycea	Pirellulales	PirellulalesOR
X400	0.23	Proteobacteria	Betaproteobacteria	BetaproteobacteriaCL	BetaproteobacteriaCL
X108	0.23	TM7	TM7PH	TM7PH	TM7PH
X954	0.23	Gemmatimonadetes	GemmatimonadetesPH	GemmatimonadetesPH	GemmatimonadetesPH
X703	0.23	Armatimonadetes	CH21	CH21CL	CH21CL
X83	0.23	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Flexibacteraceae
X997	0.23	Elusimicrobia	ElusimicrobiaPH	ElusimicrobiaPH	ElusimicrobiaPH
X130	0.23	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae
X501	0.23	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
X70	0.23	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
X802	0.22	Proteobacteria	Betaproteobacteria	Rhodocyclales	RhodocyclalesOR
X240	0.22	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae
X998	0.22	Verrucomicrobia	Spartobacteria	Spartobacteriales	Spartobacteriaceae
X746	0.22	Firmicutes	Bacilli	Bacillales	Paenibacillaceae
X807	0.22	Acidobacteria	AcidobacteriaPH	AcidobacteriaPH	AcidobacteriaPH
X874	0.22	Actinobacteria	ActinobacteriaPH	ActinobacteriaPH	ActinobacteriaPH
X407	0.22	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	Gemmatimonadaceae
X960	0.22	Actinobacteria	ActinobacteriaPH	Solirubrobacterales	SolirubrobacteralesOR
X22	0.22	Acidobacteria	AcidobacteriaPH	Acidobacteriales	Acidobacteriaceae
X985	0.22	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
X214	0.22	Chloroflexi	Thermomicrobia	Thermomicrobiales	ThermomicrobialesOR
X288	0.22	TM7	TM7PH	TM7PH	TM7PH
X133	0.22	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	Gemmatimonadaceae
X416	0.22	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Sphingobacteriaceae
X189	0.22	Actinobacteria	ActinobacteriaPH	Actinomycetales	Nocardioidaceae
X392	0.22	Actinobacteria	ActinobacteriaPH	Actinomycetales	Mycobacteriaceae
X366	0.22	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
X198	0.22	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae
X797	0.22	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
X734	0.21	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
X424	0.21	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Sinobacteraceae
X172	0.21	Actinobacteria	ActinobacteriaPH	Actinomycetales	Thermomonosporaceae
X0	0.21	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae
X789	0.21	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae
X65	0.21	Acidobacteria	Solibacteres	Solibacterales	Solibacteraceae
X1210	0.21	BRC1	BRC1PH	BRC1PH	BRC1PH
X881	0.21	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X269	0.21	Actinobacteria	ActinobacteriaPH	Acidimicrobiales	EB1017
X736	0.21	Chloroflexi	SOGA31	SOGA31CL	SOGA31CL
X925	0.21	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X896	0.21	Acidobacteria	AcidobacteriaPH	Acidobacteriales	AcidobacterialesOR
X104	0.21	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Sphingobacteriaceae
X490	0.21	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae

X127	0.21	Verrucomicrobia	Spartobacteria	Spartobacteriales	Spartobacteriaceae
X298	0.21	unclassified	unclassified	unclassified	unclassified
X606	0.21	Actinobacteria	ActinobacteriaPH	Solirubrobacterales	SolirubrobacteralesOR
X976	0.21	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X290	0.21	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
X105	0.21	Actinobacteria	ActinobacteriaPH	Actinomycetales	Intrasporangiaceae
X207	0.21	Actinobacteria	ActinobacteriaPH	Solirubrobacterales	Patulibacteraceae
X934	0.21	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae
X107	0.2	Actinobacteria	ActinobacteriaPH	Solirubrobacterales	SolirubrobacteralesOR
X1145	0.2	Proteobacteria	Deltaproteobacteria	Myxococcales	MyxococcalesOR
X339	0.2	Proteobacteria	Deltaproteobacteria	Myxococcales	MyxococcalesOR
X1079	0.2	Proteobacteria	Deltaproteobacteria	Myxococcales	Polyangiaceae
X1256	0.2	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X766	0.2	Proteobacteria	Deltaproteobacteria	Myxococcales	MyxococcalesOR
X735	0.2	Chloroflexi	ChloroflexiPH	Chloroflexales	ChloroflexalesOR
X29	0.2	Actinobacteria	ActinobacteriaPH	Actinomycetales	Intrasporangiaceae
X832	0.2	Verrucomicrobia	Spartobacteria	Spartobacteriales	Spartobacteriaceae
X64	0.2	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	GemmatimonadalesOR
X576	0.2	Actinobacteria	ActinobacteriaPH	Acidimicrobiales	Iamiaceae
X60	0.2	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae
X161	0.2	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
X413	0.2	Proteobacteria	Deltaproteobacteria	Myxococcales	Haliangiaceae
X260	0.2	Proteobacteria	Deltaproteobacteria	Myxococcales	Haliangiaceae
X1081	0.2	TM6	TM6PH	TM6PH	TM6PH
X285	0.2	TM7	TM7PH	TM7PH	TM7PH
X491	0.19	Chloroflexi	Ktedonobacteria	KtedonobacteriaCL	KtedonobacteriaCL
X1281	0.19	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Flexibacteraceae
X600	0.19	Proteobacteria	Alphaproteobacteria	Rhizobiales	RhizobialesOR
X804	0.19	Actinobacteria	ActinobacteriaPH	Actinomycetales	Nocardiaceae
X24	0.19	Proteobacteria	Betaproteobacteria	BetaproteobacteriaCL	BetaproteobacteriaCL
X809	0.19	Acidobacteria	Solibacteres	Solibacterales	Solibacteraceae
X1012	0.19	unclassified	unclassified	unclassified	unclassified
X95	0.19	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae
X531	0.19	Verrucomicrobia	Spartobacteria	Spartobacteriales	Spartobacteriaceae
X1165	0.19	Proteobacteria	Deltaproteobacteria	DeltaproteobacteriaCL	DeltaproteobacteriaCL
X949	0.19	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	Gemmatimonadaceae
X119	0.19	Acidobacteria	AcidobacteriaPH	Acidobacteriales	Acidobacteriaceae
X454	0.19	Proteobacteria	Betaproteobacteria	BetaproteobacteriaCL	BetaproteobacteriaCL
X93	0.19	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Sphingobacteriaceae
X1051	0.19	Bacteroidetes	Sphingobacteria	Sphingobacteriales	SphingobacterialesOR
X798	0.19	Actinobacteria	ActinobacteriaPH	Actinomycetales	Streptomycetaceae
X767	0.19	Chloroflexi	Anaerolineae	Anaerolineales	Anaerolinaceae
X11	0.19	Firmicutes	Bacilli	Bacillales	Bacillaceae

X647	0.19	Chloroflexi	Anaerolineae	Anaerolineales	Anaerolinaceae
X399	0.19	Actinobacteria	ActinobacteriaPH	ActinobacteriaPH	ActinobacteriaPH
X1205	0.19	Proteobacteria	Deltaproteobacteria	MIZ46	MIZ46OR
X98	0.19	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae
X936	0.19	Acidobacteria	AcidobacteriaPH	Acidobacteriales	AcidobacterialesOR
X352	0.19	Firmicutes	Bacilli	Bacillales	Thermoactinomycetaceae
X81	0.18	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae
X1092	0.18	Proteobacteria	Deltaproteobacteria	Myxococcales	MyxococcalesOR
X456	0.18	Acidobacteria	Chloracidobacteria	ChloracidobacteriaCL	ChloracidobacteriaCL
X894	0.18	Proteobacteria	Gammaproteobacteria	Legionellales	Coxiellaceae
X1200	0.18	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	Gemmatimonadaceae
X451	0.18	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X618	0.18	Proteobacteria	Deltaproteobacteria	Myxococcales	Haliangiaceae
X857	0.18	Chloroflexi	Ktedonobacteria	KtedonobacteriaCL	KtedonobacteriaCL
X1207	0.18	Proteobacteria	Betaproteobacteria	Rhodocyclales	RhodocyclalesOR
X678	0.18	unclassified	unclassified	unclassified	unclassified
X445	0.18	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae
X114	0.18	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae
X727	0.18	Proteobacteria	Deltaproteobacteria	Myxococcales	MyxococcalesOR
X1283	0.18	Proteobacteria	Gammaproteobacteria	unclassified	unclassified
X713	0.18	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	GemmatimonadalesOR
X638	0.18	Elusimicrobia	ElusimicrobiaPH	ElusimicrobiaPH	ElusimicrobiaPH
X821	0.18	Verrucomicrobia	Spartobacteria	Spartobacteriales	Spartobacteriaceae
X295	0.18	Acidobacteria	AcidobacteriaPH	Acidobacteriales	AcidobacterialesOR
X3	0.18	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae
X749	0.18	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X771	0.18	Acidobacteria	Solibacteres	Solibacterales	Solibacteraceae
X598	0.18	Proteobacteria	Betaproteobacteria	Burkholderiales	BurkholderialesOR
X430	0.18	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
X213	0.18	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Flexibacteraceae
X963	0.17	Gemmatimonadetes	GemmatimonadetesPH	GemmatimonadetesPH	GemmatimonadetesPH
X987	0.17	Proteobacteria	Deltaproteobacteria	Myxococcales	MyxococcalesOR
X326	0.17	Actinobacteria	ActinobacteriaPH	MC47	MC47OR
X514	0.17	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae
X36	0.17	Actinobacteria	ActinobacteriaPH	Actinomycetales	Micrococcaceae
X1097	0.17	Gemmatimonadetes	GemmatimonadetesPH	GemmatimonadetesPH	GemmatimonadetesPH
X585	0.17	Acidobacteria	AcidobacteriaPH	Acidobacteriales	AcidobacterialesOR
X273	0.17	Acidobacteria	AcidobacteriaPH	Acidobacteriales	AcidobacterialesOR
X1309	0.17	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	Gemmatimonadaceae
X38	0.17	Acidobacteria	Sva0725	Sva0725CL	Sva0725CL
X918	0.17	BRC1	BRC1PH	BRC1PH	BRC1PH
X96	0.17	Acidobacteria	Solibacteres	Solibacterales	Solibacteraceae
X27	0.17	Acidobacteria	Solibacteres	Solibacterales	Solibacteraceae

X48	0.17	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae
X206	0.17	Actinobacteria	ActinobacteriaPH	ActinobacteriaPH	ActinobacteriaPH
X526	0.17	Actinobacteria	ActinobacteriaPH	Actinomycetales	Actinospicaceae
X1132	0.17	unclassified	unclassified	unclassified	unclassified
X325	0.17	Firmicutes	Bacilli	Bacillales	Bacillaceae
X287	0.17	Actinobacteria	ActinobacteriaPH	Actinomycetales	Pseudonocardiaceae
X279	0.17	Chloroflexi	Thermomicrobia	ThermomicrobiaCL	ThermomicrobiaCL
X1030	0.17	Verrucomicrobia	Opitutae	Opitutales	Opitutaceae
X440	0.17	Bacteroidetes	Sphingobacteria	Sphingobacteriales	SphingobacterialesOR
X337	0.17	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	Gemmatimonadaceae
X414	0.17	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae
X382	0.17	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae
X462	0.16	Acidobacteria	AcidobacteriaPH	Acidobacteriales	Acidobacteriaceae
X124	0.16	Chloroflexi	ChloroflexiPH	Roseiflexales	Kouleothrixaceae
X1293	0.16	Chloroflexi	TK17	TK17CL	TK17CL
X778	0.16	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	Gemmatimonadaceae
X247	0.16	Actinobacteria	ActinobacteriaPH	Actinomycetales	ActinomycetalesOR
X848	0.16	Armatimonadetes	CH21	CH21CL	CH21CL
X258	0.16	Acidobacteria	AcidobacteriaPH	Acidobacteriales	AcidobacterialesOR
X506	0.16	Chloroflexi	Anaerolineae	Caldilineales	Caldilineaceae
X1407	0.16	Bacteroidetes	Sphingobacteria	Sphingobacteriales	SphingobacterialesOR
X917	0.16	Elusimicrobia	ElusimicrobiaPH	ElusimicrobiaPH	ElusimicrobiaPH
X869	0.16	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	Gemmatimonadaceae
X706	0.16	Actinobacteria	ActinobacteriaPH	Solirubrobacterales	Patulibacteraceae
X49	0.16	Nitrospirae	NitrospiraePH	Nitrospirales	Nitrospiraceae
X698	0.16	unclassified	unclassified	unclassified	unclassified
X544	0.16	Actinobacteria	ActinobacteriaPH	Actinomycetales	Micromonosporaceae
X793	0.16	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Sinobacteraceae
X959	0.16	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae
X1155	0.16	Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae
X930	0.16	Actinobacteria	ActinobacteriaPH	Acidimicrobiales	Microthrixaceae
X1186	0.16	unclassified	unclassified	unclassified	unclassified
X201	0.16	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae
X505	0.16	Actinobacteria	ActinobacteriaPH	Actinomycetales	Nocardioidaceae
X40	0.16	Proteobacteria	Alphaproteobacteria	Rhizobiales	RhizobialesOR
X831	0.16	Chloroflexi	Ktedonobacteria	KtedonobacteriaCL	KtedonobacteriaCL
X692	0.16	Bacteroidetes	Sphingobacteria	Sphingobacteriales	SphingobacterialesOR
X532	0.16	Actinobacteria	ActinobacteriaPH	MC47	MC47OR
X754	0.16	Proteobacteria	Betaproteobacteria	Rhodocyclales	RhodocyclalesOR
X528	0.16	Actinobacteria	ActinobacteriaPH	Solirubrobacterales	SolirubrobacteralesOR
X556	0.16	Proteobacteria	Deltaproteobacteria	Myxococcales	MyxococcalesOR
X190	0.15	Actinobacteria	ActinobacteriaPH	Solirubrobacterales	Patulibacteraceae
X632	0.15	unclassified	unclassified	unclassified	unclassified

X76	0.15	Proteobacteria	Alphaproteobacteria	AlphaproteobacteriaCL	AlphaproteobacteriaCL
X938	0.15	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Sphingobacteriaceae
X580	0.15	Proteobacteria	Deltaproteobacteria	Syntrophobacterales	Syntrophobacteraceae
X75	0.15	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
X587	0.15	Proteobacteria	Deltaproteobacteria	Myxococcales	Myxococcaceae
X472	0.15	Chloroflexi	ChloroflexiPH	Chloroflexales	ChloroflexalesOR
X209	0.15	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	Gemmatimonadaceae
X145	0.15	Acidobacteria	AcidobacteriaPH	Acidobacteriales	Acidobacteriaceae
X47	0.15	Actinobacteria	ActinobacteriaPH	Actinomycetales	Geodermatophilaceae
X230	0.15	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae
X92	0.15	Actinobacteria	ActinobacteriaPH	Actinomycetales	Catenulisporaceae
X54	0.15	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	Gemmatimonadaceae
X943	0.15	Actinobacteria	ActinobacteriaPH	Acidimicrobiales	Iamiaceae

Chapter 3

Dissecting Host Genetic Control of the Sorghum Rhizosphere Microbiome Through Genome-Wide Association Studies

3.1 Abstract

Host genetics has recently been shown to be a driver of plant microbiome composition in many plant species. However, identifying the underlying genetic loci controlling microbial selection remains challenging. Although recent efforts have begun to dissect the impact of individual genes on microbiome composition, there is a lack of global approaches to comprehensively examine the host genotype influences on microbial community composition. Genome-Wide Association Study (GWAS) represents a potentially powerful and unbiased method to identify microbes sensitive to host genotype, and to connect them with the genetic loci that control them. To identify genetic loci in sorghum that are associated with root microbiome regulation, we performed a comprehensive GWAS analysis to investigate host genotype-dependent influences on microbial community structure. Here, we utilize this same dataset, a population-level microbiome analysis of the rhizosphere microbiomes of 200 sorghum genotypes, to demonstrate the utility of GWAS in dissecting host genetic influence on the plant microbiome. We show that GWAS can be used to identify host loci that correlate with the abundance of specific subsets of the rhizosphere microbiome. Surprisingly, we further demonstrate that these results can be used to predict the rhizosphere microbiome structure for an independent panel of sorghum genotypes based solely on knowledge of host genotypic information.

3.2 Introduction

Despite consistent evidence of the interaction between host genetics and plant microbiome composition, the identification of specific genetic elements driving host-genotype dependent microbiome acquisition and assembly in plants remains a challenge. Recent efforts guided by *a priori* hypotheses of gene involvement have begun to dissect the impact of individual genes on microbiome composition (Lebeis *et al.*, 2015; Castrillo *et al.*, 2017). However, these studies are inherently limited in scope to the relatively small fraction of plant genes for which we have sufficient knowledge to make predictions about their influence in microbiome-related processes. Additionally, many plant traits predicted to impact microbiome composition and activity, such as root exudation (Zhalnina *et al.*, 2018) and root system architecture (Saleem *et al.*, 2018), are inherently complex and potentially governed by a very large number of genes. Collectively, this suggests the need for alternative large-scale and unbiased methods for identifying the genes that regulate the host-mediated selection of the microbiome.

Genome-wide association studies (GWAS) represent a powerful approach to map loci, which are associated with complex traits in a genetically diverse population. Though pioneered for use in human genetics, to date the majority of GWAS studies have been conducted in plants (Brachi, Morris and Borevitz, 2011), and it has become an increasingly popular tool for studying the genetic basis of natural variation and traits of agricultural importance. When inbred lines are available, GWAS can be particularly useful because once these lines have been genotyped, they can be phenotyped multiple times, making it possible to study many different traits in many different environments (Atwell *et al.*, 2010). While GWAS is typically used in the context of a single quantitative phenotypic trait, analyses of multivariate molecular traits, such as RNA-seq expression data and metabolite profiles, have also been conducted (Schaefer *et al.*, no date; Wu *et al.*, 2018). More recently, several attempts have been made to use host-associated microbiome census data as an input to GWAS, which in theory will allow for the identification of host genetic loci controlling microbiome composition (Davenport *et al.*, 2015; Wang *et al.*, 2016).

In plants, a recent GWAS in *Arabidopsis thaliana* that used leaf microbial community data as the phenotypic trait suggested that plant loci responsible for defense and cell wall integrity affect microbial community variation (Horton *et al.*, 2014), establishing GWAS as a potential tool for studying the role of plant host genetic heterogeneity in shaping plant-associated microbial communities. Several other recent phyllosphere studies that have performed GWAS to identify genetic factors controlling these associations in the leaf microbiome, with mixed degrees of success (Horton *et al.*, 2014; Wallace *et al.*, 2018; Roman-reyna *et al.*, 2019). However, the use of GWAS in the context of root-associated microbiome has yet to be extensively explored; here, selection of sample type (rhizosphere or endosphere) and host system may be critical factors that determine the success of such efforts. Previous work comparing the root microbiomes of a broad range of cereal crops has demonstrated the degree to which microbial communities correlate with host phylogenetic distance is strongest in the root endosphere (Naylor *et al.*, 2017), while another study of different rice cultivars found that host

genotype correlated with microbial association most strongly in the rhizosphere (Edwards *et al.*, 2015). Collectively, these data suggest the sample type exhibiting the strongest correlation may different for each host, and that an initial evaluation of the degree of correlation between genotype and microbiome phenotype across sample types may be helpful for successful implementation of GWAS.

We show that GWAS can be used to identify specific genetic loci within the host genome that are correlated with the abundance of specific heritable lineages, and that differences in microbiome composition can be predicted solely from genotypic information. This work described in this chapter demonstrates the utility of GWAS for the analysis of host-mediated control of rhizosphere microbiome phenotypes.

3.3 Results

3.3.1 Genetic loci correlated with rhizosphere microbial abundance through genome-wide association

Recent work in the leaf microbiome has demonstrated the potential utility of GWAS for uncovering loci correlated with microbiome composition. Here, we sought to use GWAS analyses with rhizosphere microbiome datasets using both global properties of the OTU dataset and the abundances of individual OTUs. We first conducted a principal component analysis on our dataset. Bi-plots of principal components (PCs) 1, and 2 are presented in **[Figure 3-1a]**. For overall community composition, a subset of PCs was selected from an analysis of the abundance patterns of the 1189 OTUs. To prioritize individual PCs for inclusion in our GWAS analysis, the heritability of the top ten PCs were determined **[Table 3-1]**; PC1 through PC10 collectively explain 75% of the total amount of variance in our dataset **[Figure 3-1b, Table 3-1]**. Among the top ten PCs, PC1 and PC5 had the two highest heritability scores (H2 =0.35 and 0.42, respectively). PCs with heritability scores larger than 0.25 were selected for downstream analyses, including PC1, PC3, PC5, PC9, and PC10. GWAS analysis was performed first using eigenvectors from PCA as input, and Manhattan plots were generated to show the association between genotypes and each community-level trait, PCs **[Figure 3-2]**. The Manhattan plot of GWAS analysis performed on PC1 revealed a significant correlation between community composition and a locus of approximately 0.93 Mb on Chromosome 4 with a moderately stringent threshold of $-\log(10^{-4})$**[Figure 3-3a]**. The Manhattan plot of PC5 and PC10 also revealed an identifiable peak on chromosome 6. The peak on PC5 was slightly below the threshold of significance **[Figure 3-3b]**, and PC10 had a moderately stringent threshold of $-\log(10^{-4})$ **[Figure 3-3a].**

As results from PC1, PC5, and PC10 are derived from the properties of overall microbiome composition, it is unclear which specific microbial lineages drive these correlations. To determine if the observed correlations at these two loci (Chromosome 4 and Chromosome 6) were driven by one common or several different sets of OTUs, we performed separate GWAS analyses using the abundances of each single OTU as input **[Figure 3-4]**. From these analyses, two distinct sets of OTUs were found to have significant correlations with the loci on chromosomes 4 (n=40) and 6 (n=11), expect OTU18 from *Burkholderiales* order **[Figure 3-4]**. Taxonomy information of these OTUs are presented in **[Table 3-2]**. Notably, significant correlations between OTU abundance and genotype at each locus appear to be unique to a select group of microbes **[Figure 3-4]**, suggesting that different sorghum genes influence the abundance patterns of different groups of microbes.

3.3.2 Expression pattern of candidate genes within the genetic loci

To determine if distinct sets of OTUs identified by GWAS were associated with host genotype, we evaluated the ratio of relative abundance of each OTU between the chromosome 4 major and minor allele containing sorghum lines. Notably, bacterial clades with a single membrane (monoderm) were more heavily enriched in sorghum containing the major allele, while bacteria with two membranes (diderms) were more heavily enriched in the minor allele group **[Figure 3-5a]**. The chromosome 4 locus contains approximately 27 candidate genes, several of which exhibited strong root specific expression patterns **[Figure 3-5b]**. Two notable candidates include gamma carbonic anhydrase-like 2 and a putative Beta-1,4 endoxylanase, which exhibited the strongest root specific expression pattern in publicly available sorghum RNA-Seq data obtained from phytozome v12.1 (Goodstein *et al.*, 2012) **[Figure 3-5b]**.

3.3.3 Microbiome composition prediction using sorghum genotypic data

To validate that genetic differences across sorghum genotypes at the candidate loci on chromosome 4 identified by GWAS are indeed responsible for observed differences in the rhizosphere microbiome composition a follow up experiment was performed using eighteen additional sorghum lines, including genotypes not present in the original study. To help disentangle phylogenetic-relatedness from loci-specific effects, the new selected sorghum genotypes spanned the diversity panel, and for each minor allele genotype (n=9), we included a phylogenetically related major allele line (n=9) **[Figure 2-1].** Following two weeks of growth in a mixture of calcined clay and field soil in the growth chamber, the rhizosphere microbiomes of two replicates of each genotype were collected and microbiome composition was analyzed using 16S rRNA amplicon sequencing as in the main study. A canonical analysis of principal coordinates (CAP) ordination constrained on genotypic group demonstrates that the rhizospheres of genotypes belonging to major and minor allele groups separate into distinct clusters **[Figure 3-7]**.

Next, we performed indicator species analysis to identify OTUs that distinguished the two chromosome 4 allele groups. We identified 16 indicator OTUs for the major allele, and 4 indicators for the minor allele. Notably, 15 of 16 major allele indicators were monoderm bacterial lineages, while all 4 minor allele indicators were diderm bacteria **[Table 3-3]**. To determine whether patterns of indicator OTUs were consistent between the growth chamber validation and our primary field experiment, we used indicator species analysis to compare the six minor allele group members present in the 200 line GWAS panel against six closely related major allele containing lines. Comparing the top 100 indicators from each allele group for this analysis, we observed similar trends in abundance of indicator OTUs across bacterial orders, including two monoderms, *Actinomycetales* and *Solirubrobacterales*, as abundant indicators of the major allele, and three diderms, *Burkholderia*, *Sphingobacterales*, and *Xanthomonadales*, as indicators of the minor allele. Collectively, these experiments support the findings of our sorghum GWAS, that a locus located on chromosome 4 is capable of selecting for specific bacterial lineages, with signals that persist across two developmental timepoints, and is stable between growth chamber and field grown sorghum plants.

3.4 Discussion

3.4.1 Multiple sorghum loci are responsible for controlling the rhizobiome

In this study, we demonstrate that GWAS can be used to successfully identify correlations between specific host genetic loci and the abundance of specific microbes within the host microbiome, as well as overall rhizosphere community structure. To our knowledge, this is the first example of such work in a crop rhizosphere. Similar to previous plant microbiome GWAS studies (Horton *et al.*, 2014; Wallace *et al.*, 2018; Roman-reyna *et al.*, 2019), no single strong effect loci was observed, providing additional evidence that microbiome association is a polygenic trait. In this study, several loci were identified to have strong associations with the microbiome structure, and the most significant of these maps to a locus on chromosome 4. This locus contains several strong candidate genes exhibiting root specific expression and natural genetic variation across sorghum varieties. One candidate gene located near the center of this locus encodes a beta 1,4 endo xylanase. Xylanases are responsible for the degradation of xylan into xylose, and are one of the primary catabolizers of hemicellulose, a major component of the plant cell wall (Meents, Watanabe and Samuels, 2018). A loss or alteration of function of a beta 1,4 endo xylanase could potentially impact the rhizosphere microbiome through one of several mechanisms. First, such enzymes are responsible for loosening the cell wall architecture, and may control the degree of plasticity in the barrier between the root and surrounding rhizosphere environments. This loosening could potentially allow for greater release of cell wall or apoplast derived metabolites, including carbohydrates, into the rhizosphere environment (Sasse, Martinoia and Northen, 2018). Plant-derived metabolites may contribute to chemotactic response of certain microbes into the rhizosphere and/ or proliferation of fast growing bacterial clades capable growing on diverse carbon substrates (Goldfarb *et al.*, 2011; Chaparro *et al.*, 2013). Alternatively, altered xylanase activity could lead to shifts in carbohydrate profiles within the cell wall, leading to heightened plant immune responses (Claverie *et al.*, 2018; Hou *et al.*, 2019). The catabolic byproducts of microbially-produced xylanase used in pathogen invasion are in part responsible for triggering innate immune responses in plants, and various components of the plant immune signalling network have been shown to influence microbiome structure (Lebeis *et al.*, 2015; Castrillo *et al.*, 2017).

Another candidate gene within the chromosome 4 locus, that also displays root-specific expression, encodes γ carbonic anhydrase-like 2. In plants, carbonic anhydrases (CA) facilitate the reversible hydration of CO_2 to bicarbonate (Parisi *et al.*, 2004; DiMario *et al.*, 2017). Sorghum is predicted to encode 17 CAs, include 3 γCAs (DiMario *et al.*, 2017). γCA and γCA-like isoforms are nuclear encoded, mitochondrial-localized proteins, that make up the mitochondrial Complex I (NADH-ubiquinone oxidoreductase) (Sunderhaus *et al.*, 2006). While studies of plant CAs have mainly focused on photosynthetic tissue, CA activity has also been observed in non-photosynthetic tissues, including roots (DiMario *et al.*, 2017). In rice, CA1 expression is increased in both shoots and roots in response to osmotic and salt stresses, and expression of rice CA1 in

Arabidopsis thaliana conferred enhanced salt tolerance (Yu *et al.*, 2007). Previous studies have also implicated CA activity in plant-microbe interactions (Floryszak-Wieczorek and Arasimowicz-Jelonek, 2017); CA was first observed in root nodules of legumes inoculated with Rhizobium (Atkins, 1974), and has since been implicated in disease resistance, having both antioxidant activity and salicylic acid (SA) binding capability (Slaymaker *et al.*, 2002; Restrepo *et al.*, 2005). In *Arabidopsis thaliana*, a chloroplast-localized CA, salicylic acid-binding protein 3 (AtSABP3), is regulated by nitric oxide via S-nitrosylation. S-nitrosylation of AtSABP3 suppresses both its interaction with SA and CA activity (Wang *et al.*, 2009). Collectively, these studies suggest that a loss or alteration of function of CA could potentially impact the composition of the rhizosphere microbiome. Future validation experiments using genetic mutants within this and other candidate genes can be used to help elucidate the underlying genetic element(s) responsible for modulation of the rhizosphere microbiome.

3.4.2 Host genotypic data can predict microbiome differences

Here we show GWAS can be used as a non-candidate approach to predict microbiome structure based solely on host genetic information, building on previous studies that have observed inter- and intra-species variation capable of generating variation in microbiomes (Bulgarelli *et al.*, 2012; Lundberg *et al.*, 2012; Peiffer *et al.*, 2013; Horton *et al.*, 2014; Haney *et al.*, 2015; Naylor *et al.*, 2017; Fitzpatrick *et al.*, 2018; Walters *et al.*, 2018). Although the underlying cause of these shifts are not well understood, candidate driven approaches have implicated disease resistance (Lebeis *et al.*, 2015; Castrillo *et al.*, 2017), nutrient status (Hiruma *et al.*, 2016; Khan *et al.*, 2016; Castrillo *et al.*, 2017), sugar signaling (Yamada *et al.*, 2016), and plant age (Wagner *et al.*, 2016; Edwards *et al.*, 2018) as major factors. However, these directed approaches limit the potential to capture novel mechanisms compared to non-candidate approaches such as GWAS.

What has become clear from previous studies of plant-associated microbiomes, is the complex nature of this interaction. Many of these traits are likely to be controlled by a large number of genes, with each having only a small effect on the overall phenotype. As such, future studies should take advantage of large panels to draw statistically significant associations. Despite these challenges, a recent GWAS of the rice phyllosphere microbiome identified genomic regions overlapping with quantitative trait loci (QTLs) for resistance and carbohydrate metabolism that regulate microbiome composition. Using near-isogenic lines that vary in cellulose and salicylate production, they predictably modulated the abundance of certain bacterial lineages that associate with the plant (Roman-reyna *et al.*, 2019).

Collectively, our study adds to a growing list of evidence that genetic factors of plant hosts are capable of modulating their associated microbiomes. Moving forward, additional efforts will promote a more mature understanding of the host molecular mechanisms underlying the assembly of microbiomes and identification of the beneficial functions microbes provide to their hosts. These efforts will facilitate future breeding

efforts aimed at promoting beneficial microbiomes capable of promoting plant yield under a number of challenging biotic and abiotic stresses.

3.5 Methods

GWAS For each OTU, GWAS was conducted separately using the best linear unbiased predictors (BLUPs) obtained from the linear mixed model. Population structure was accounted for using statistical methods that allow us to detect both population structure (Q) and relative kinship (K) to control spurious association. The Q model ($y = S\alpha + Q\nu + e$), the K model ($y = S\alpha + Zu + e$), and the Q + K model ($y = X\beta + S\alpha + Q\nu + Zu + e$) described previously in (Yu *et al.*, 2008), used in our study. In the model equations, y is a vector of phenotypic observation; α is a vector of allelic effects; e is a vector of residual effects; ν is a vector of population effects; β is a vector of fixed effects other than allelic or population group effects; u is a vector of polygenic background effects; Q is the matrix relating y to ν; and X, S, and Z are incidence matrices of 1s and 0s relating y to β, α, and u, respectively. To account for the population structure and genetic relatedness, the first three principal components (PCs) and kinship matrix were calculated using the SNPs obtained from (Morris, Ramu, *et al.*, 2013) and fitted into the MLM-based GWAS pipeline for each OTU using GEMMA (Zhou and Stephens, 2012).

GWAS Validation For the GWAS validation experiment, the 378 genotypes of the SAP were first subset into lines containing the major (n=343) and minor (n=14) allele for the two haplotypes found at the peak on chromosome 4 described in the text. From the 200 genotypes not originally selected as part of the GWAS, a total of nine sorghum genotypes belonging to the minor allele were selected, with an effort to include genotypes spanning the phylogenetic tree. For each of these nine minor allele lines, another genotype containing the major allele with close overall genetic relatedness was selected, resulting in nine major and nine minor allele containing linges (n=18). Two replicates of each line were grown in growth chambers (33°C/28°C, 16h light/ 8h dark, 60% humidity) in a 10% vermiculite/ 90% calcined clay mixture rinsed with a soil wash prepared from a 2:1 ratio of field soil to water from the field site used in the main experiment. Plants were watered daily with approximately 5 ml of autoclaved Milli-Q water using a spray bottle for the first three days, followed by top watering with 15 ml of water every three days. An additional misting was performed to the soil surface every 24 hrs to prevent drying. Following two weeks of growth, plants were harvested and rhizosphere microbiomes extracted as described for the field experiment.

Microbiome Statistical Analysis All statistical analyses of the amplicon datasets were performed in R using the normalized reduced dataset with 1189 OTUs and 598 samples, with each sample containing 10,000 reads, unless stated otherwise. For alpha-diversity measurement, Shannon's Diversity was calculated as e^X, where X is Shannon's Entropy as determined with the diversity function in the R package vegan (Oksanen *et al.*, 2016). Principal coordinate analyses were performed with the function pcoa in the R package ape (Paradis, Claude and Strimmer, 2004), using the Bray-Curtis distance obtained from function vegdist in the R package vegan (Oksanen *et al.*, 2016). Mantel's tests were used to determine the correlation between host phylogenetic distances and microbiome distances using the mantel function in the R package vegan with 9 999 permutations, and using Spearman's correlations to reduce the effect of outliers. Indicator species analyses were performed using the function indval in the R package labdsv

(Roberts and Roberts, 2016), with p-values based on permutation tests run with 10,000 permutations. To account for multiple testing performed for all 430 genera in our dataset, multiple testing corrections were performed with an FDR of 0.05 using the p.adjust function in the base R package stats. Canonical Analysis of Principal Coordinates (CAPs) was performed for the final validation experiment to test the amount of variance explained by genotypic group using the capscale function in the R package vegan (Oksanen *et al.*, 2016); an ANOVA like permutation test using the sum of all constrained eigenvalues was performed to determine the percent variance explained by each factor using the function anova.cca in package vegan.

Figure 3-1

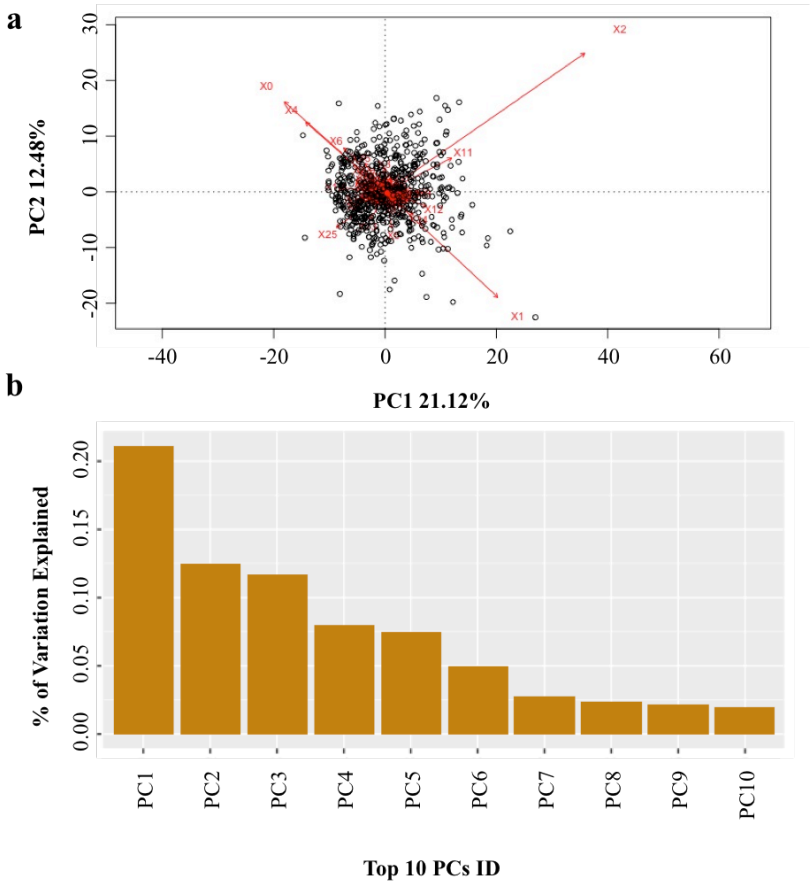

Figure 3- 1. Principal component analysis (PCA) on rhizosphere samples

a. PCA bi-plots of principal components 1 and 2 generated using all 598 rhizosphere samples. Each dot represents a sorghum accession. The red arrows indicate OTUs. **b.** The amount of variance explained by the top 10 PCs.

Figure 3-2

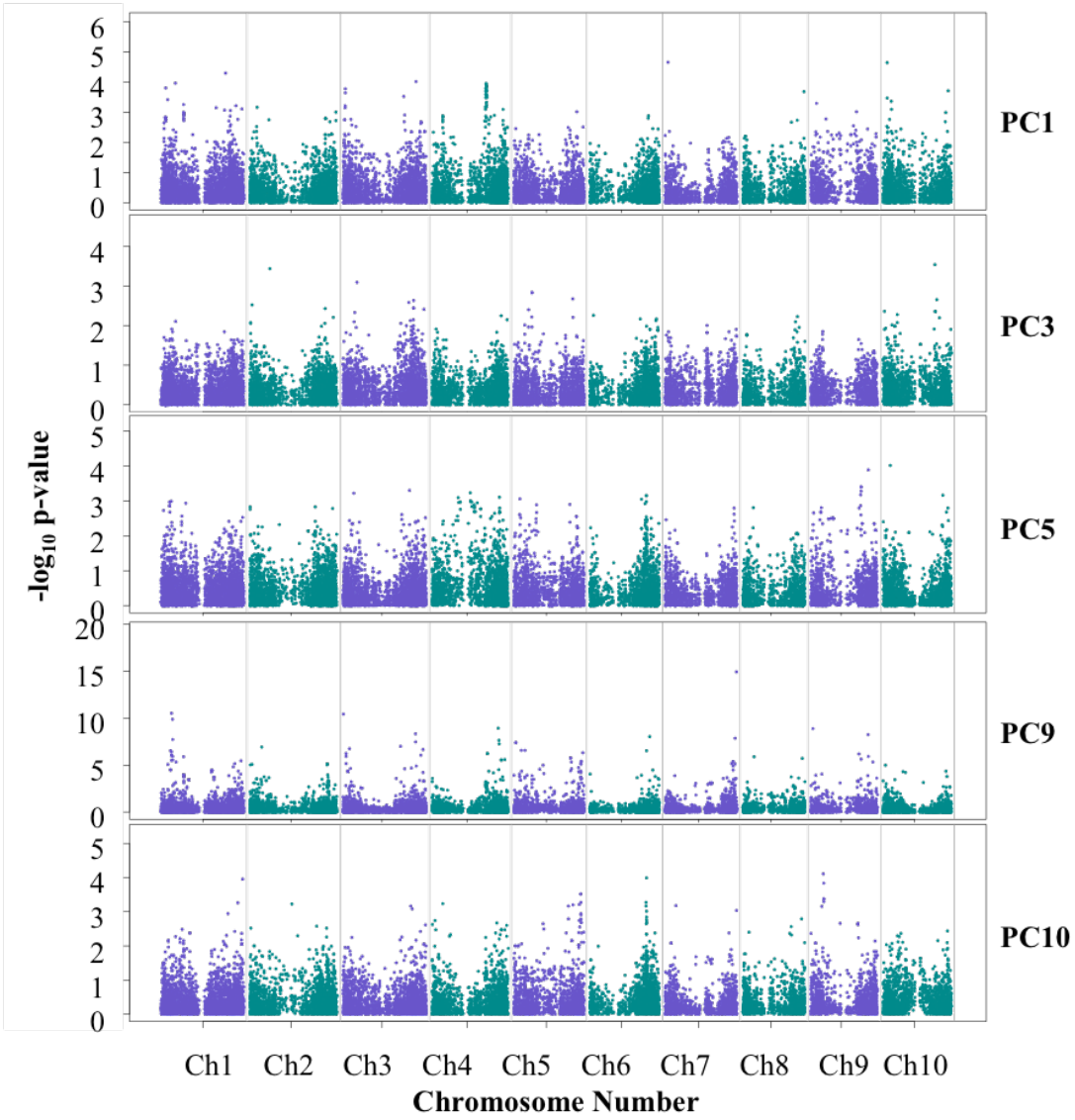

Figure 3- 2. GWAS on heritable PCs

Manhattan plots shows the results of GWAS for heritable PCs with broad sense heritability >= 0.25, including PC1, PC3, PC5, PC9, and PC10.

Figure 3-3

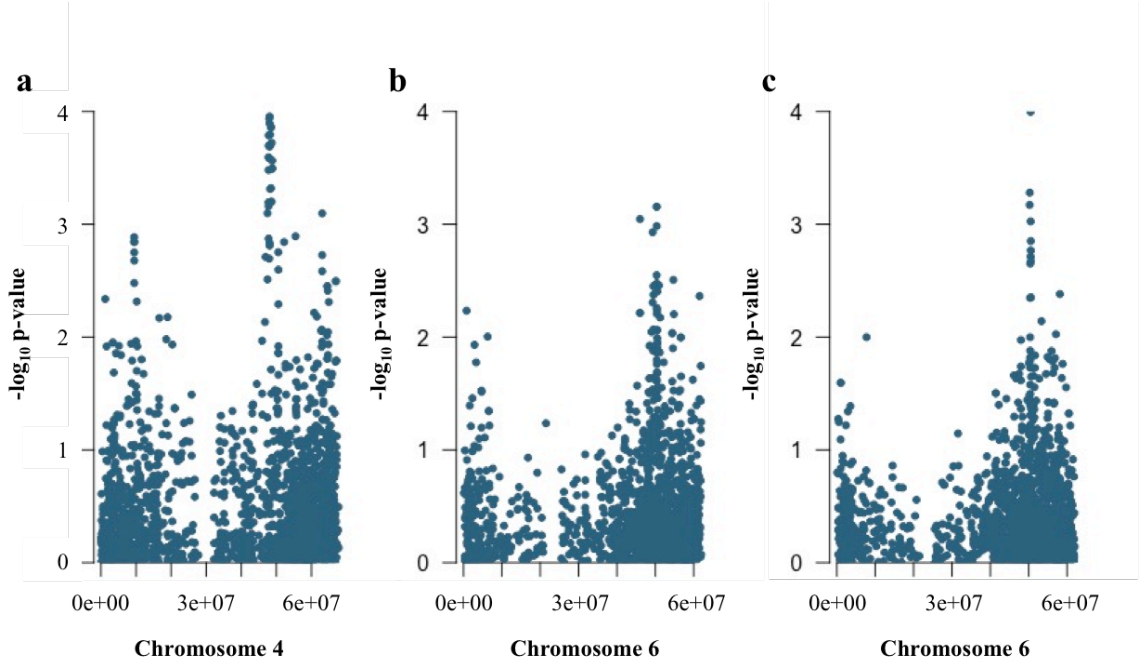

Figure 3- 3. GWAS peaks for PC1, PC5 and PC10

a. Manhattan plot of PC1 community analysis GWAS, displaying a locus on chromosome 4. **b**. Manhattan plot of PC5 community analysis GWAS, displaying a locus on chromosome 6. **c.** Manhattan plot of PC10 community analysis GWAS, displaying a locus on chromosome 6.

Figure 3-4

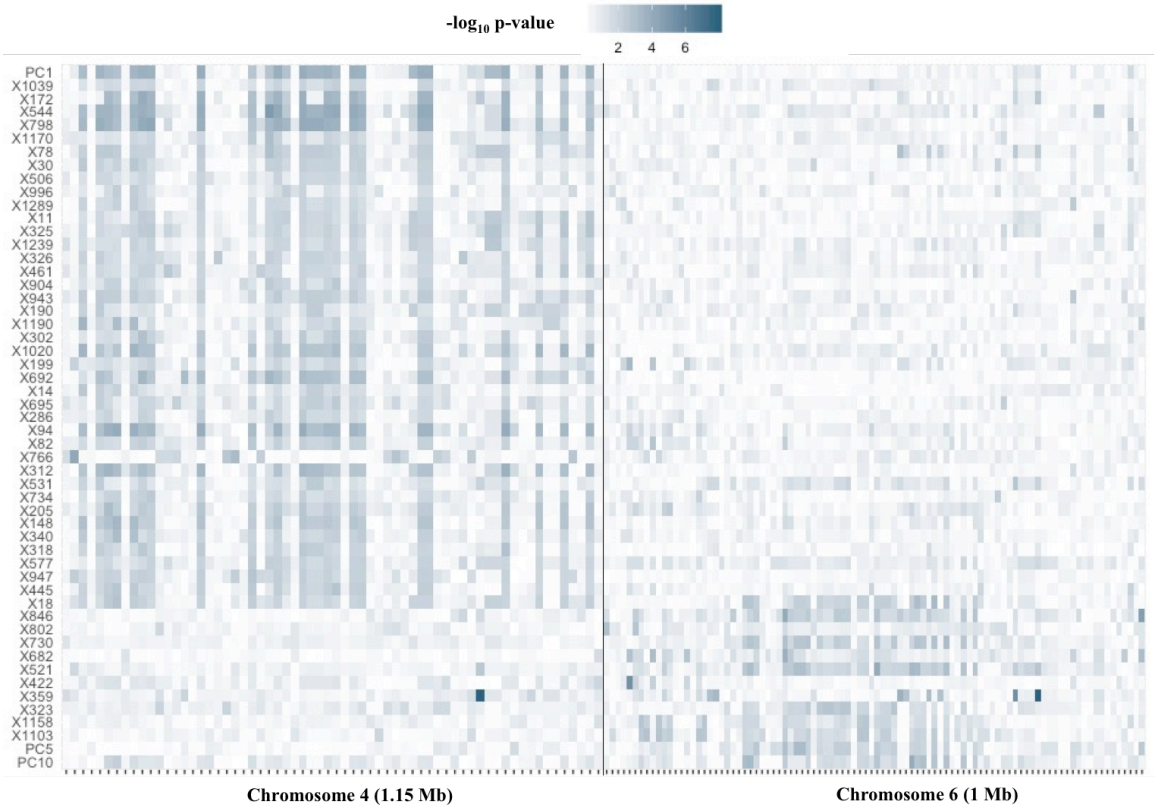

Figure 3- 4. GWAS analyses using the abundances of individual OTUs

Individual OTUs have correlations with the loci on chromosomes 4 and 6. There are 40 OTUs showed significant correlations with the locus on chromosome 4, and 11 OTUs showed significant correlations with the locus on chromosome 6. OTU18 shared between both chromosomes. Taxonomy information of these OTUs is presented in [Table 3-2].

Figure 3-5

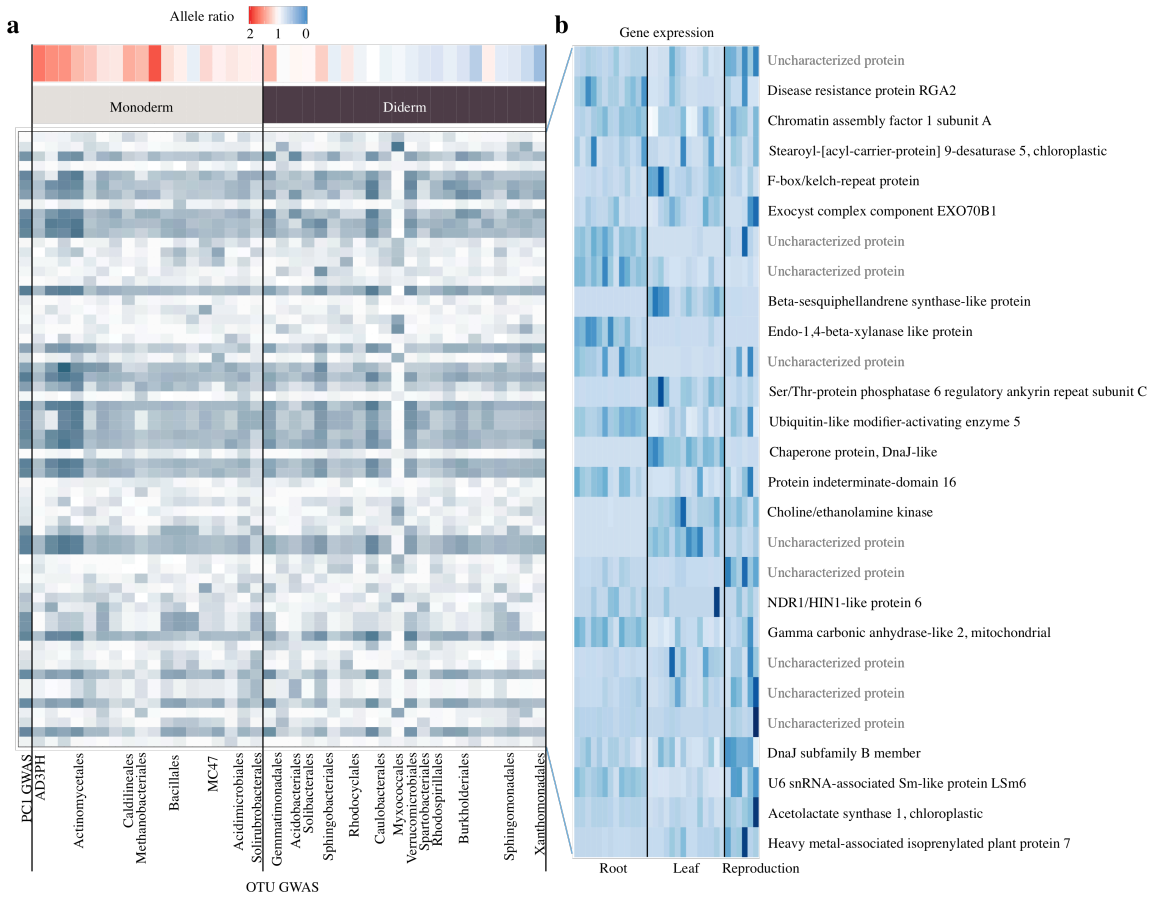

Figure 3- 5. A sorghum genetic locus on chromosome 4

a. GWAS using individual OTU abundance. OTUs with at least 5 SNPs above -log (10^{-3}) in the window (size = 1.15 Mb) identified on the same chromosome 4 locus as PC1 GWAS. Relative abundance of OTUs that associate with the sorghum major (red) or minor (blue) allele groups within this locus (upper heat map). OTUs were grouped based on the predicted presence of one or two membranes (monoderm or diderm) within each bacterial order. **b.** Tissue-specific gene expression data for sorghum genes within the chromosome 4 locus. The publicly available sorghum RNA-Seq data was obtained from phytozome v12.1.

Figure 3-6

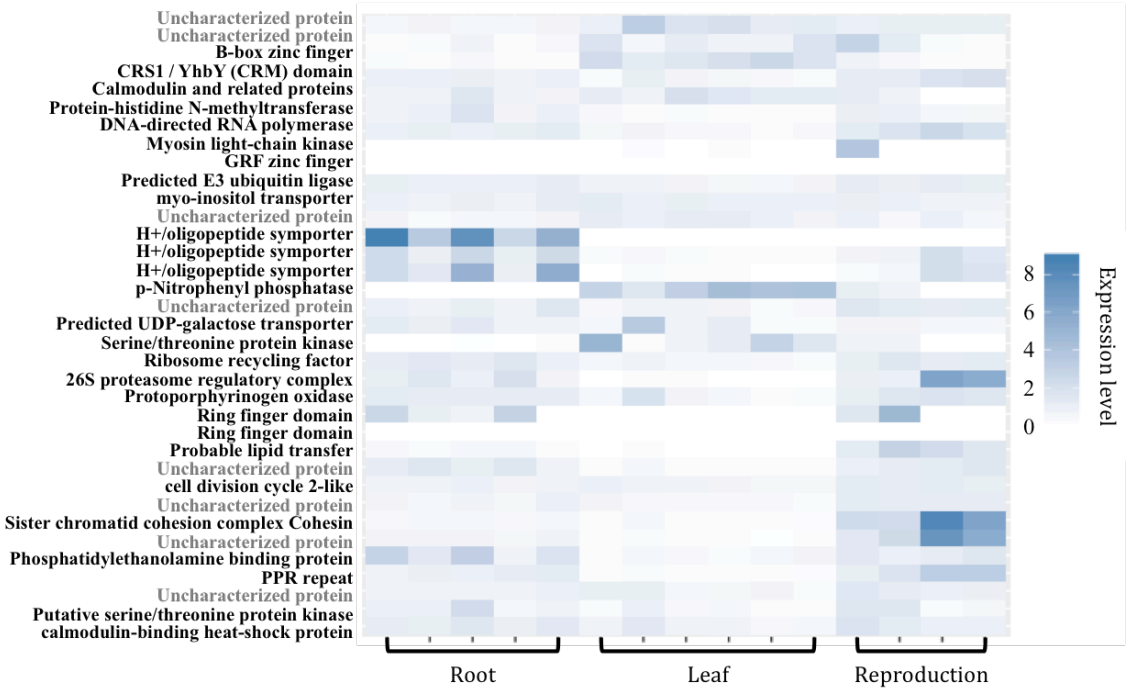

Figure 3- 6. Tissue-specific gene expression data for sorghum genes within the chromosome 6 locus

Tissue-specific gene expression data for sorghum genes within the chromosome 6 locus. The publicly available sorghum RNA-Seq data was obtained from phytozome v12.1

Figure 3-7

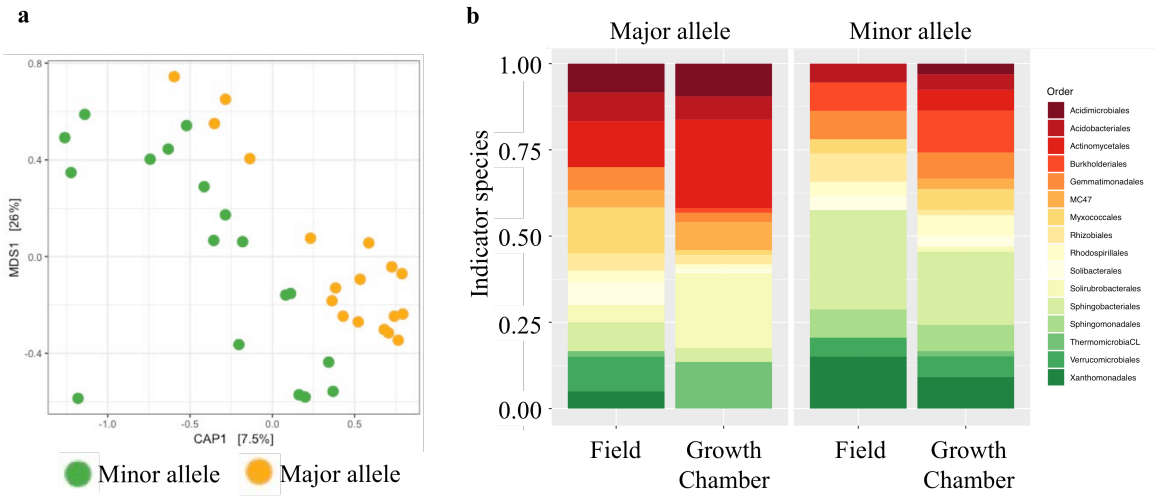

Figure 3- 7. Sorghum genetic information can be used to predict rhizosphere microbiome composition

a. Canonical Analysis of Principle Coordinates of the rhizosphere microbiome for nine major allele genotypes (yellow) and nine minor allele genotypes (green). **b.** Relative abundance of chromosome 4 indicator OTUs. Six minor allele and six phylogenetically related major allele genotypes present in the 200 line GWAS field experiment were compared to the nine minor/major allele indicator OTUs used for the growth chamber validation experiment. Colors represent indicator OTUs grouped at the taxonomic level of order.

Table 3-1

Table 3- 1. Proportion of variance explained by the top ten PCs and the heritability scores

Top 10 PCs	Proportion of Variance	Cumulative Proportion	Broad Sense Heritability
PC1	0.2112	0.2112	0.3501
PC2	0.1248	0.3361	0.0295
PC3	0.1166	0.4526	0.3108
PC4	0.0795	0.5321	0.2102
PC5	0.0750	0.6071	0.4212
PC6	0.0500	0.6570	0.0982
PC7	0.0274	0.6845	0.1973
PC8	0.0241	0.7086	0.1230
PC9	0.0216	0.7301	0.3220
PC10	0.0202	0.7504	0.2484

Table 3- 2. Taxonomy information of OTUs associated with two genetic loci on chromosome 4 and 6

GWAS analyses using the abundances of individual OTUs obtained 50 OTUs correlated with genetic loci on chromosomes 4 and 6. There are 40 OTUs showed significant correlations with the locus on chromosome 4, and 11 OTUs showed significant correlations with the locus on chromosome 6. OTU18 shared between both chromosomes.

OTU ID	Taxonomy Information			
	Phylum	Class	Order	Family
X445	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae
X312	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	VerrucomicrobialesOR
X521	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae
X730	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae
X996	unclassified	unclassified	Methanobacteriales	unclassified
X148	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae
X205	Proteobacteria	Alphaproteobacteria	Sphingomonadales	SphingomonadalesOR
X340	Proteobacteria	Alphaproteobacteria	Sphingomonadales	SphingomonadalesOR
X359	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae
X14	Bacteroidetes	Sphingobacteria	Sphingobacteriales	SphingobacterialesOR
X682	Bacteroidetes	Sphingobacteria	Sphingobacteriales	SphingobacterialesOR
X692	Bacteroidetes	Sphingobacteria	Sphingobacteriales	SphingobacterialesOR
X1158	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Flammeovirgaceae
X531	Verrucomicrobia	Spartobacteria	Spartobacteriales	Spartobacteriaceae
X846	Verrucomicrobia	Spartobacteria	Spartobacteriales	Spartobacteriaceae
X190	Actinobacteria	ActinobacteriaPH	Solirubrobacterales	Patulibacteraceae
X302	Acidobacteria	Solibacteres	Solibacterales	Solibacteraceae
X734	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
X286	Proteobacteria	Betaproteobacteria	Rhodocyclales	RhodocyclalesOR
X695	Proteobacteria	Betaproteobacteria	Rhodocyclales	RhodocyclalesOR
X802	Proteobacteria	Betaproteobacteria	Rhodocyclales	RhodocyclalesOR
X422	Proteobacteria	Alphaproteobacteria	Rhizobiales	Phyllobacteriaceae
X766	Proteobacteria	Deltaproteobacteria	Myxococcales	MyxococcalesOR
X326	Actinobacteria	ActinobacteriaPH	MC47	MC47OR
X461	Actinobacteria	ActinobacteriaPH	MC47	MC47OR
X323	Chloroflexi	ChloroflexiPH	Herpetosiphonales	Herpetosiphonaceae
X199	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	Gemmatimonadaceae
X1020	Gemmatimonadetes	GemmatimonadetesPH	Gemmatimonadales	Gemmatimonadaceae
X1103	Elusimicrobia	ElusimicrobiaPH	FAC88	FAC88OR
X82	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae
X94	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae
X506	Chloroflexi	Anaerolineae	Caldilineales	Caldilineaceae

X18	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae
X318	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae
X577	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae
X947	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae
X11	Firmicutes	Bacilli	Bacillales	Bacillaceae
X325	Firmicutes	Bacilli	Bacillales	Bacillaceae
X1239	Firmicutes	Bacilli	Bacillales	Bacillaceae
X1289	Firmicutes	Bacilli	Bacillales	Planococcaceae
X1039	AD3	AD3PH	AD3PH	AD3PH
X30	Actinobacteria	ActinobacteriaPH	Actinomycetales	Frankiaceae
X78	Actinobacteria	ActinobacteriaPH	Actinomycetales	Thermomonosporaceae
X172	Actinobacteria	ActinobacteriaPH	Actinomycetales	Thermomonosporaceae
X544	Actinobacteria	ActinobacteriaPH	Actinomycetales	Micromonosporaceae
X798	Actinobacteria	ActinobacteriaPH	Actinomycetales	Streptomycetaceae
X1170	Actinobacteria	ActinobacteriaPH	Actinomycetales	unclassified
X1190	Acidobacteria	AcidobacteriaPH	Acidobacteriales	AcidobacterialesOR
X904	Actinobacteria	ActinobacteriaPH	Acidimicrobiales	EB1017
X943	Actinobacteria	ActinobacteriaPH	Acidimicrobiales	Iamiaceae

Table 3- 3. Indicator species analysis on the validation dataset

OTU ID	p-value	Allele	Order	Bacterial Lineages
X211	1.47E-05	Major	SOGA31CL	monoderm
X1448	0.000728477	Major	Roseiflexales	monoderm
X1535	0.001353937	Major	Actinomycetales	monoderm
X68	0.001662987	Major	Solirubrobacterales	monoderm
X1099	0.001743929	Major	SOGA31CL	monoderm
X71	0.002656365	Major	MC47	monoderm
X764	0.003061074	Major	Actinomycetales	monoderm
X3724	0.004518028	Major	ThermomicrobiaCL	monoderm
X113	0.005886681	Major	SOGA31CL	monoderm
X505	0.006460633	Major	Actinomycetales	monoderm
X1576	0.006497425	Major	Actinomycetales	monoderm
X712	0.0065195	Major	MC47	monoderm
X471	0.006556291	Major	Acidimicrobiales	diderm
X2530	0.007520235	Major	Solirubrobacterales	monoderm
X131	0.007689478	Major	Actinomycetales	monoderm
X1318	0.007836645	Major	Solirubrobacterales	monoderm
X18	0.005128771	Minor	Burkholderiales	diderm
X2114	0.005945548	Minor	ChloracidobacteriaCL	diderm
X148	0.006129507	Minor	Sphingomonadales	diderm
X1299	0.007534952	Minor	CH21CL	diderm

Chapter 4

Evaluation of a Plant Growth-Promoting Microbial Soil Amendment Effects on Root Microbiome in Strawberry Plants

A modified version of this chapter is submitted as "Deng, S., Wipf, H.M.L., Pierroz, G., Raab, T.K., Khanna, R. and Coleman-Derr, D., 2019. A Plant Growth-Promoting Microbial Soil Amendment Dynamically Alters the Strawberry Root Bacterial Microbiome. Scientific reports, 9(1), pp.1-15."

4.1 Abstract

Given concerns about food sufficiency in the U.S. and worldwide for the coming decades, every effort must be made to optimize the environmental conditions on crop production. There is an increasing demand for using microbial strategies to improve plant productivity in agricultural production. However, much work still remains to fully elucidate how beneficial plant-microbe associations are established and function, as well as what role soil amendments may play in shaping these interactions. Here, we describe a set of experiments to test the effect of a commercially available soil amendment product on the soil and strawberry (*Fragaria x ananassa* Monterey) root microbiome. We tested the hypothesis that this product's application may correlate with distinct shifts in the root and rhizosphere microbiomes. The bacterial communities of the soil, rhizosphere, and root endosphere from amendment-treated and untreated fields were profiled at four time points across the strawberry growing season using 16S rRNA amplicon sequencing on the Illumina MiSeq platform. In all sample types, bacterial community composition and relative abundance were significantly altered with amendment application. Importantly, time point effects on community composition are more pronounced in the root and rhizosphere, which suggests an interaction between plant development and the effect of amendment treatment. Surprisingly, there was a slight overlap between the taxa within the amendment and those enriched in plant and soil following amendment application, suggesting that VESTA may act to rewire existing networks of organisms through an, as of yet, uncharacterized mechanism. Taken together, these results demonstrate that a commercial microbial soil amendment can impact the community structure of both plant roots and the surrounding environment.

4.2 Introduction

Agriculture faces numerous abiotic and biotic challenges in optimizing crop production. Drought and excess salinity impact 45% and 19.5% of agricultural land, respectively (Flowers and Yeo, 1995; Dos Reis, Lima and de Souza, 2012; Onaga and Wydra, 2016), and plant disease is one of the leading constraints of agricultural productivity (Savary *et al.*, 2012; Carroll *et al.*, 2017) that accounts for total annual yield losses of US$220 billion globally (Chakraborty and Newton, 2011). Achieving food security is further compounded by the effects of climate change and rapid population growth (Alexandratos and Bruinsma, 2012; Myers *et al.*, 2017). While preventative measures and early management of stresses are key to fortifying crop yield, current strategies often have calamitous ramifications and are unsustainable. Widely employed methods include the heavy application of exogenous fertilizers (Havlin *et al.*, 2005), growth enhancers (Nawaz *et al.*, 2016), pesticides (Oerke, 2006), and soil sterilization methods, such as fumigation with methyl bromide (MeBr) (Braun and Supkoff, 1994). While these treatments can be effective for managing stressors, they also can have dire consequences for human health, the environment, and long-term soil quality and health (Yates, Wang and Ernst, 1997). Adverse impacts of these methods include water, air, soil, and food contamination, increased incidences of cancer, the disruption of reproductive, neurological, and respiratory systems of mammals, birds, and fish, reductions in pollination and other important ecosystem services, ozone depletion, and eutrophication (Pimentel, 2005; Savci, 2012; Stehle and Schulz, 2015; Kim, Kabir and Jahan, 2017; Sabarwal, Kumar and Singh, 2018; Siviter *et al.*, 2018). In recognition of these consequences, the regulation of several widely-used pesticides has increased in recent years (Duniway, 2002). Often in conjunction with pesticide use, chemical fertilizers are applied to help stimulate crop fitness and boost yields. Yet, not only is the supply of raw source materials for fertilizer production rapidly dwindling, several considerable issues arise with their intensive use (Gilbert, 2009). Only ~30% to 40% of applied fertilizer is actually taken up and utilized by crops, with the majority being lost to the environment (Yu *et al.*, 2015). Runoff, leaching, and volatilization of fertilizers pollute air and water systems, contributing to ozone depletion, eutrophication and marine 'dead zones', and increased incidence of human disease (Anderson, Glibert and Burkholder, 2002; Savci, 2012). Due to the lack of sustainability and the environmental impacts of pesticide and fertilizer use, new approaches are needed to further support and improve crop performance.

One promising alternative that is being actively pursued is the use of plant growth-promoting microbes (PGPM), where there are a growing number of microbial amendments being commercialized for various crops (Berg, 2009; Coleman-Derr and Tringe, 2014; Vejan *et al.*, 2016; Kumar and Verma, 2018). Along with the enhancement of plant productivity and yield, PGPM can allow for significant reductions in the application of chemical fertilizers and pesticides (Singh, Singh and Prabha, 2016; Timmusk *et al.*, 2017). Recent work has shown that plants recruit specific microbial colonists via root and leaf metabolites, some of which are exuded into the soil and surrounding environment (Singh *et al.*, 2004). In exchange for these compounds, PGPM can promote plant growth through a variety of mechanisms, including increasing soil

nutrient bioavailability, improving water acquisition, decreasing herbivore damage, and suppressing plant disease (Yang, Kloepper and Ryu, 2009; Berendsen, Pieterse and Bakker, 2012; Turner, James and Poole, 2013; Pii *et al.*, 2015).

The agronomic production of strawberry, an important horticultural commodity crop valued at US$2.8 billion (Suh, Guan and Khachatryan, 2017), faces a number of disease-related challenges, including the devastation by *Verticillium dahliae*, *Pythium*, *Rhizoctonia*, and *Cylindrocarpon* spp. (Martin and Bull, 2002). Pathogen attack can occur at many different stages of plant development and remains one of the biggest threats to the strawberry market (Daugaard and Lindhard, 2000). Pre-planting soil fumigation with MeBr is found to effectively suppress a broad range of strawberry diseases and has been the predominant strategy used for commercial strawberry production in California. As the use of this pesticide is being discontinued in many parts of the United States (Duniway, 2002), strawberry serves as a good agronomic system for the study of PGPM use as an alternative to traditional disease control methods (Martin and Bull, 2002). Strawberry further presents itself as a relevant biological system with its use in many recent development and fruit growth studies. Previous studies have utilized strawberry in investigating the effects of animal compost as a non-synthetic alternative to MeBr on root disease suppression (Millner, Ringer and Maas, 2004). Plant-associated *Stenotrophomonas* strains were also shown to stimulate root growth and root hair development by efficiently colonizing the strawberry rhizosphere (Suckstorff and Berg, 2003). An amplicon-based metagenomics approach has also identified microbes that potentially play a role in yield decline in strawberry (Xu *et al.*, 2015), where low abundances of beneficial bacteria and a nematode fungus, in addition to high levels of fungal root rot pathogens and wet soil conditions, have contributed to yield decline. Taken together, these recent findings suggest that manipulating the strawberry microbiome may be an effective way to increase both fitness and yield.

There is ongoing research on how certain PGPM affect crop performance and the resident soil and plant-associated microorganisms (Castro-Sowinski *et al.*, 2007; Trabelsi and Mhamdi, 2013). Field inoculation of rhizobia strains enhanced populations of Alpha- and Gamma-proteobacteria, together with Firmicutes and Actinobacteria, in the bulk soil of common bean plants (Trabelsi *et al.*, 2011). However, inoculation of Azospirillum strains showed no prominent effects on the population structure of rhizobacterial communities in maize (Herschkovitz *et al.*, 2005). Yet, few studies have explored the effect of community-level microbial amendments on the native soil microbiome, the survival of applied microorganisms in the field, and the temporal community dynamics of host-associated epiphytes and endophytes following amendment application. Additional research is needed to understand how microbial-based methods can be employed for improving crop growth.

In this study, we explore the effect of a commercially available microbial soil amendment, VESTA, on the microbiome of strawberry plants. VESTA is a fermented liquid product composed of a broad spectrum of microbes, fermentation by-products and organic acids (VESTA, SOBEC Corporation, Fowler, CA), and the direct effect of VESTA on microbial communities in the soil or plant host is unknown. Here, we

investigate the effect of the amendment on soil physicochemistry, strawberry growth, and bacterial community composition, diversity, and function. We observed phenotypic differences in overall biomass and lateral root growth in amendment-treated strawberry plants. 16S rRNA gene amplicon-based Illumina sequencing revealed substantial changes in the strawberry microbiome following treatment with the amendment, but suggests that this effect is the result of the amendment modulating the compositional profile of existing root and soil microbes, rather than through replacement of the community with the product's microorganisms. Shotgun sequencing of the amendment largely corroborated the community composition profile obtained through amplicon sequencing, and offers insight into its potential function. Together, these findings demonstrate that a commercial microbial soil amendment can alter the community structure of the strawberry root and surrounding soil environment.

4.3 Results

4.3.1 The microbial soil amendment VESTA enhances strawberry yield and alters soil chemistry

Roots from strawberries grown in amendment-treated fields displayed enhanced growth, as evidenced by significantly greater fresh weight across time points in comparison to control roots **[Figure 4-1a]**; treated plants displayed enhanced above-ground growth at each time point as well **[Figure 4-1b]**. This indicates that amendment-treated samples may differ from control in their greater ability to absorb water from the surrounding soil. Visual inspection revealed that this effect was likely from enhanced root-proliferation at the crown and greater aboveground shoot biomass **[Figure 4-1b]**. Additionally, amendment-treated roots only had significantly greater dry weight in the first and second time points out of the three time points that this trait was measured **[Figure 4-2]**. Lastly, the water content of amendment-treated roots was uniformly higher than in the control plants **[Figure 4-1a]**, which can impact photosynthesis, plant performance, and actualized growth potential(Lambers, Stuart Chapin and Pons, 2008).

In comparison to control soils, soils treated with the soil amendment had a higher wet-cohesive strength, greater mass plant-derived materials, and increased average soil particle aggregate-size. Additionally, treated soil had slightly increased soil pH (7.44 versus 7.32 for control) **[Table 4-1a]**. Electrical conductivity (EC_{10}) was also higher in amendment-treated soil compared to control soils (0.242 dS/m vs. 0.184 dS/m) **[Table 4-1a]**. Comparing nitrate levels, amendment-treated soil was composed of distinctly less nitrate than the controls **[Table 4-1a]**. Total nitrogen of treated soils was 0.188% dry weight, while for controls it was 0.177% dry weight; for total carbon content, treated soils contained 0.385% dry weight while control soils contained 0.249% dry weight **[Table 4-1a]**. Based on energy-dispersive X-ray fluorescence (EDXRF), overall micronutrients were lower in the treated soils than controls **[Table 4-1b]**. In comparison to control plants, the roots of plants treated with the amendment product also had significantly higher phosphorus levels (2315 μg/g vs. 1747 μg/g of tissue) and significantly lower levels of aluminum, iron, and molybdenum; foliar tissue collected from treated plants at time point four showed generally greater levels of micronutrients than those collected from control plants **[Table 4-2]**. Taken together, these results demonstrate that amendment application has a significant effect not only on root growth, but on soil physiochemistry as well. However, with the constraints of a two-block design, we acknowledge that there is an unknown degree of spatial influence on soil physiochemistry, as well as on bacterial community assembly, and additional studies utilizing a fully randomized block design are needed to confirm these reported effects of the microbial soil amendment.

4.3.2 Amendment treatment effects on bacterial diversity and composition of the strawberry root microbiome

To investigate the impact of amendment application on bacterial community composition in the root, rhizosphere, and surrounding soil, we performed 16S rRNA amplicon sequencing on the Illumina MiSeq platform using custom V3-V4 dual-indexed primers(Takahashi *et al.*, 2014; Naylor *et al.*, 2017). To allow for a direct comparison of microbes present in the amendment and microbes observed in our treated samples, samples of the product were also processed and sequenced using the same protocol. After demultiplexing, quality control, clustering based on 97% similarity threshold, and assigning taxonomies, 13,545 OTUs were generated. To remove underrepresented sequences which often represent sequencing artifacts(Tremblay *et al.*, 2015), we filtered these OTUs by removing OTUs not seen more than 5 times in at least 3% of the samples (n=5). After filtering, we resampled the OTU table to normalize for differences in read count abundance between samples to a common read depth of 16,542 reads per sample. The resulting dataset of 3,387 OTUs was used in the downstream statistical analyses.

We hypothesized that the addition of a complex microbial amendment would increase overall microbial diversity within the soil and root microbiomes. Rarefaction curves were used to estimate species richness as a function of sequencing depth and suggest that saturation in sequencing was achieved, as curves began to plateau or reached their asymptote for all sample types by treatment **[Figure 4-3]**. Unexpectedly, we found that total community diversity, as measured by Shannon's Diversity, decreased in treated samples as compared to control samples, for both soil and rhizosphere sample types at every time point **[Figure 4-4a; Table 4-3a]**. By contrast, Shannon's Diversity in the root was not significantly different between treated and control samples **[Figure 4-4a; Table 4-4]**. Additionally, there is a significant effect of time point on Shannon's Diversity (F value = 21.555, $p<0.001$) **[Table 4-3b]**. From time point 1 to 4, Shannon's Diversity decreases in treated soil samples **[Figure 4-4a]**. In the rhizosphere, Shannon's Diversity decreases from the 1st to the 3rd time point in treated samples; however, in the 4th, it increases to a level more similar to the Shannon's diversity observed in the control samples **[Figure 4-4a]**. In the root, Shannon's diversity in treated samples is similar to levels in control samples, increasing from time point 1 to 2, and then gradually decreases again for all successive time points **[Figure 4-4a]**. Finally, the mean number of observed species was found to be lower in amendment-treated samples than control samples in all sample types and time points **[Figure 4-5a]**. Taken together, these results suggest that amendment application alters root-associated microbial communities and is correlated with an overall decrease in bacterial diversity in the treated area, but that this effect is mitigated in some way within the plant host.

To test which factors in our experimental design contribute to differences in the beta diversity of bacterial communities associated with the soil, rhizosphere, and root, permutation multivariate analysis of variance (PERMANOVA) analyses were performed independently on each sample type using Bray-Curtis distances. These results indicated that the effect of amendment application on community composition was larger in soil than in the root and rhizosphere samples **[Table 4-5]**. These results also revealed that time point influenced community composition **[Figure 4-5b; Table 4-5]**, and that this effect was least pronounced in soil samples as compared to root and rhizosphere. To further visualize whether amendment treatment influenced bacterial community

composition in the soil, rhizosphere, and root, we conducted principal coordinate analysis (PCoA) for all sequenced samples using both Bray-Curtis and UniFrac distances **[Figure 4-4b; Figure 4-5b]**. PCoA using the Bray-Curtis dissimilarities **[Figure 4-4b]** displayed strong clustering of samples by sample type along the first PCoA axis, which explains 36.8% of the variance. Additionally, we observed that samples cluster by treatment type along the second PCoA axis, which explains 15.7% of the variance. PCoA based on weighted UniFrac distances **[Figure 4-5b]** revealed similar trends. The result that the effect of time point was least pronounced in soil samples is observable in the PCoA plot, where for root and rhizospheres, amendment-treated samples belonging to the fourth time point were found to cluster closer to the untreated samples **[Figure 4-5b]**. Notably, samples taken at time point four were treated more than a month prior to sampling, whereas samples collected at all other time points had been treated one or two weeks prior to sampling. This suggests that in the interim between treatments, the soil and root microbiomes in the treated field may drift back towards a compositional profile more similar to that of control fields. To determine the percent of variance explained by the factors shown to be significant by PERMANOVA (time and treatment, and their interaction) **[Table 4-5]**, we next performed a canonical analysis of principal coordinates (CAPS) using the Bray–Curtis distance independently on each sample type **[Table 4-6a]**. Results were very similar to those determined by PERMANOVA, where treatment explained the largest proportion of variance across all sample types (42%, 32%, and 39% in the soil, rhizosphere, and root, respectively) **[Figure 4-6a]**. The percent variance explained by time point is largest in the rhizosphere (21%), followed by the root (18%) and smallest in the soil (14%) **[Figure 4-6a]**. A CAPS analysis performed separately for each time point and constrained for treatment type showed that time point effects on community composition are more pronounced in the root and rhizosphere **[Figure 4-6b, Table 4-6b]**, which suggests that the extent to which the amendment impacts bacterial communities varies across plant development and may be influenced by shifts in microbial recruitment and community maintenance via exudation rates and profiles. In addition, treatment explains less variation at time point four in root and rhizosphere, but not in soil **[Figure 4-6b, Table 4-6b]**. Taken together, these results suggest that treatment with the amendment has an effect on bacterial community composition not only in the soil, where it is applied, but also within and on the plant root.

4.3.3 Amendment treatment is correlated with increases in *Betaproteobacteria*

Amendment treatment was correlated with changes in relative abundance at different taxonomic levels within all sample types **[Figure 4-7b]**. Changes in the roots were characterized at the class level by a significant increase in *Betaproteobacteria* ($P < 0.001$) and a concomitant decrease in the abundance of *Actinobacteria (*$P < 0.001$). An indicator species analysis was used to identify at higher taxonomic resolution the individual bacterial genera with enrichment or depletion patterns in roots treated with the amendment as compared to controls. These analyses revealed that genera from the families *Alteromonadaceae*, *Comamonadaceae*, *Burkholderiales*, which are known to contain species with beneficial properties to plants(Kyselková *et al.*, 2009), were all

observed to be more abundant in amendment-treated roots. In addition, several families and genera of *Betaproteobacteria* (*Ramlibacter, Rhodocyclales, Methylophilaceae, Methylotenera, Acidovorax, Comamondaceae*) that were significantly enriched in amendment-treated samples versus control samples have the capabilities of nitrogen fixation(Loy *et al.*, 2005; Shang and Yi, 2015; Yan *et al.*, 2017), denitrification(Kalyuhznaya *et al.*, 2009; Mustakhimov *et al.*, 2013), and sulfur cycling(Schmalenberger *et al.*, 2008) **[Table 4-7]**. However, it is important to note that the specific strains observed in our study that are represented by these indicator lineages may or may not possess these properties noted in the literature; additional work in the future to reveal the functional capacities of individual isolates through isolate sequencing and phenotyping will help to resolve this knowledge gap.

Based on our observation that the amendment has an overall larger effect on soil microbiome composition than roots, we hypothesized that a larger number of OTUs would be shared between treated and control samples in roots, as compared to soils. However, we observed that the percentage of differentially present OTUs between amendment-treated and control samples is greatest in roots, where 594 (27.2%) and 396 (18.2%) OTUs are distinct to control and amendment-treated, respectively, with only 54.6 % of the OTUs held in common **[Figure 4-7b]**. By comparison, 2399 (79.0%) and 2378 (77.6%) OTUs are shared across soil and rhizosphere samples, respectively.

4.3.4 Community compositional changes associated with VESTA treatment are not driven by an increase in abundance of VESTA organisms

To investigate whether the changes in community composition in the soil, root and rhizosphere following inoculation with the amendment product are due to increased colonization by microorganisms present within VESTA product, we first sought to characterize the microbial diversity, community composition, and compositional stability within the product. As the product VESTA is created by first mixing two raw pre-products, BHF-10 and SOBEC, we collected initial samples of both pre-products, and weekly samples of VESTA for 14 weeks. Bray-Curtis distance of pairwise treated time point samples also showed that the product changes significantly over time **[Figure 4-8, Figure 4-9]**. As amendment application is typically administered using product prepared within one to two weeks, we considered the composition of VESTA samples after one and two weeks after the mixing of the pre-products as representative of the VESTA community applied in our strawberry field study for downstream analysis.

To help eliminate potential taxonomic bias introduced by our choice of 16S rRNA primers used in our study, a shotgun metagenomic analysis of the VESTA product was also performed. Both shotgun and 16S datasets revealed a community composition of roughly 500 organisms **[Figure 4-10a]** that was largely comprised of the genera *Mycobacterium*, *Caulobacter*, *Novosphingobium*, *Bacillus*, *Flavobacterium*, and *Pseudomonas*, many of which are reported to contain strains with plant growth-promoting characteristics(Marek-Kozaczuk and Skorupska, 2001; Kyselková *et al.*, 2009;

Soltani *et al.*, 2010; Shafi, Tian and Ji, 2017). Due to the unexpectedly complex nature of the VESTA community, an investigation into the functional capacity of the product provided limited information on the product's potential relevance for plant growth-promoting capabilities; however, we did observe an abundance of genes related to microbial stress responses, including dormancy, sporulation, and secondary metabolism, as well as phosphorus, sulfur, nitrogen, iron, and potassium metabolism, which could play a role in improving nutrient acquisition for the host plant **[Figure 4-10b]** (Terrazas *et al.*, 2016).

We next explored the overlap between OTUs present in the VESTA product with all those uniquely enriched in amendment-treated and control samples across the three different sample types. Surprisingly, we did not observe a large enrichment in abundance of the majority of the OTUs belonging to the product within treated versus control samples in any compartment **[Figure 4-11a; Figure 4-12]**. However, a comparison of the lists of OTUs uniquely enriched in control and treated samples across all sample types as determined by indicator species analysis with the list of OTUs present in the product revealed that a larger proportion (30% or 39 OTUs) were members of the product in the treated samples than OTUs in the controls (14% or 19 OTUs) **[Figure 4-11b]**. Additionally, the mean rank abundance within the product for all OTUs uniquely enriched in treated samples was roughly 1.5 fold higher than that for OTUs enriched in the control samples **[Figure 4-11c]**. Of the 639 OTUs present in the product, 436 (68.2%), 426 (66.7%), and 327 (51.2%) were observed to also be present within the amendment-treated soil, rhizosphere, and root compartments, respectively **[Figure 4-12]**. By comparison, 396 (62%), 413 (65%), and 288 (45%) OTUs from the product were observed to also be present in the control soil, rhizosphere, and root samples, respectively.

Taken together, these results demonstrate that amendment treatment leads to a broad restructuring of the root microbiome, and that while treatment results in significant changes in microbial community composition, increases in the abundance of the microbes present in the product itself are not the primary driver of this difference.

4.4 Discussion

Our 16S rRNA gene amplicon-based metagenomic analysis of soil, rhizosphere and root samples treated with the amendment product reveals that bacterial communities can be significantly altered by microbial amendment application. Unexpectedly, application of the amendment was associated with decreases in soil and rhizosphere Shannon Diversity. Similar results have been observed in previous studies; functional and bacterial and fungal diversity was shown to decrease post amendment application (Siles *et al.*, 2014), and the shift was observed to be dependent on specific parameters, including soil type, soil pore size (15-20 μm), and year (White *et al.*, 1994; Koyama *et al.*, 2014). However, other studies reveal that inoculation with certain soil amendments can instead increase or not impact soil and rhizosphere alpha diversity, including application of biochar (De Tender *et al.*, 2016; Jaiswal *et al.*, 2017; Kolton *et al.*, 2017), vermicompost (Strauss, Stover and Kluepfel, 2015), a heavy metal immobilizer amendment(Touceda-González *et al.*, 2015), and bacterial inoculations(Tian and Gao, 2014). One explanation for these discrepancies in effects may be the inherent complexity of each inoculant, and VESTA represents a more complex microbial system as compared with many of these other treatments. More generally, we expect that changes in microbial diversity in response to amendment application are likely dependent on a wide variety of factors, including edaphic and abiotic parameters of the environment being amended, but also the individual genetic and phenotypic characteristics of the microbes present in the amendment (Garbeva, van Veen and van Elsas, 2004; Pii *et al.*, 2016). Additional studies in controlled settings that allow for additional environmental factors to be tested within the experimental design would further our understandings of the role of amendment application on soil and plant-associated microbial diversity.

As a pre-existing community of microorganisms, VESTA differs from many commercially available products in that its constituents already exist within the bounds of an established nutrient exchange system. The absence of these types of supporting relationships is often cited as the reason many single microbe inoculants that show promise in greenhouse studies fail to produce significant effects when tested in more complex field-based trials. While it has been observed that successful establishment of introduced microbes occurs less readily in diverse soils(van Elsas *et al.*, 2012; Vivant *et al.*, 2013), and some research also indicates that rhizosphere microbiomes are highly buffered against microbial invaders (reviewed by (Ambrosini, de Souza and Passaglia, 2016), we anticipated that the product's diversity would allow for increased colonization of its constituents within the root and rhizosphere. Surprisingly, we observed that the changes in community composition following treatment with the amendment were not primarily shifts in the abundance of bacterial organisms present in the product, but instead shifts in bacteria native to the environment. This suggests that the amendment is instead acting through an unknown mechanism to rewire the abundance patterns of existing soil and root microbes. As precedent for this, past research has found that bacterial inoculants on seeds are correlated with improved plant growth due to the stimulation of native microflora (reviewed in (Khare and Arora, 2015)).

One of the primary shifts we observed across all sample types was an increase in the relative abundance of *Betaproteobacteria* with amendment application. *Betaproteobacteria* are Gram-negative aerobic or facultative bacteria that are comprised of chemolithotrophs and phototrophs, and they have a broad range of metabolic capabilities, including the ability to fix nitrogen (Falkow *et al.*, 2006). Increases in this class suggest treatment could hypothetically increase nutrient bioavailability for both the plant and the surrounding microbial communities. With an indicator species analysis, family and genera under the phylum of *Betaproteobacteria* that are highly enriched under amendment treatment have been ascribed the functions of denitrification and sulfur cycling; this may partly explain the reduced nitrate in amendment-treated soils, as compared to control soils. Another primary shift we found with amendment application was a decrease in abundances of *Actinobacteria*, particularly in amendment-treated roots. As two recent studies have demonstrated that water stress in crop roots systems is correlated with an increase in the abundance of many Actinobacterial lineages(Naylor *et al.*, 2017; Santos-Medellín *et al.*, 2017), the observed decrease in Actinobacteria may be a result of the observed increase in water uptake in amendment-treated plants. Additional studies in other field sites will reveal which of these observed changes are typical for the application of the amendment, and which are specific to this experimental environment.

There are several mechanisms by which treatment with the amendment may be enacting community changes. Microbes in the product may outcompete a few, select native microbes and eliminate existing community hubs by means of greater resource use efficiency and metabolic potential. Amendment application may also cause organisms from the amendment and/or native microbes to produce new or higher levels of antimicrobials that affect overall diversity (Hol *et al.*, 2015). Metabolites produced initially from VESTA microbes may also directly impact plant growth, which then may cause indirect stimulation of a different community by influencing subsequent root growth and exudate release. In addition, the amendment product may directly contribute fungi or other microfauna that may then influence bacterial communities. While currently unknown, the specific mechanisms through which the amendment acts to modulate community structure, and whether this mechanism is also responsible for the observed growth promoting traits, will require further experimentation. VESTA has been utilized by commercial farmers in the past for its ability to improve plant disease resistance, growth, and yield. While the observed increase in plant biomass with amendment application may correlate with increased yield (Lima *et al.*, 2017), additional testing is needed to determine how soil treatment directly relates to microbial and genetic signatures of soil fertility, plant growth, and - in particular - yield and fruit quality (such as flavor, resistance to rot, and nutritional profile), as well as to overall shifts in microbial abundances and the fungal communities associated with the plant and soil. Furthermore, although this study was conducted under agriculturally relevant field conditions, and in a pair of plots which had received identical treatment in prior years, the two-block design used in this study limits our ability to draw conclusions regarding the statistical significance of the effect of VESTA in shaping the root microbiome; future work using additional replication and randomization, and in additional soil environments, will be useful for determining how generalizable these findings are.

Finally, we observed that many of the shifts associated with amendment application are dependent on time. In the final time point, the treated samples are more similar to control samples for all sample types. Importantly, in the fourth time point, samples were collected nearly one month after the most recent amendment application, in contrast to the other time points, in which the amendment had been applied two weeks prior to collection. This result is typical for studies conducted on the introduction of single organisms into complex soil systems. Several studies have found the general reduction of levels of individual inoculants introduced to soils(van Veen, van Overbeek and van Elsas, 1997; Matos, Kerkhof and Garland, 2005; Kröber *et al.*, 2014; Schreiter *et al.*, 2014), finding as in little as a week there was more than a 99% reduction in abundance. Another study with pathogenic *Pseudomonas aeruginosa* showed a decline to below detectable levels after 3–5 weeks post introduction under non-sterile microcosms, while the population was maintained at high abundance under sterilized microcosms(Deredjian *et al.*, 2014). Our data suggest that even in cases where amendments represent complex communities of microorganisms, soil and root communities both have a resilience that will lean towards eventual recapitulation of the native state following microbial inoculation. This has important implications for the successful use of such products in commercial agriculture, and suggests that repeated applications may be beneficial for the persistence of the plant growth promoting agents.

Our 16S rRNA gene amplicon-based metagenomic analysis of soil, rhizosphere and root samples treated with the amendment product reveals that bacterial communities can be significantly altered by microbial amendment application. To our knowledge, this is the first reported instance of a microbial and nutrient-based soil amendment causing shifts in the bacterial communities present in the rhizosphere and root endosphere of a crop species. We find that the prescribed concentration of the product is effective in inducing changes in local bacterial communities, which may contribute to the root growth-promotion observed under treatment with the amendment.

4.5 Method

Site description and amendment application To investigate the effect of the microbial soil amendment on field-grown strawberries, soil properties, and bacterial communities associated with the host and soil, we designed an experiment that allowed for collection of amendment–treated and untreated control samples from a commercial strawberry farm in Guadalupe, CA (34.9716° N, 120.5718° W) **[Figure 4-12]**. The sandy loam soil of the strawberry operation is a mixture of sandstone-derived alluvium from the Coast Range and wind-borne materials (Watson, 1919). Plugs of a day-neutral strawberry cultivar (*Fragaria x ananassa* Monterey) were transplanted to beds in 12-acre plots, and limited by the commercial farm site's irrigation system, the microbial soil amendment VESTA was applied to one of two neighboring plots in a two-block design. The field utilized had been uniformly planted with broccoli (*Brassica oleracea* var. *italica*) the previous season and received identical watering and fertilizing treatments and crop rotations before the start of this study. The microbial soil amendment VESTA was fertigated via drip irrigation systems to the treated plot starting at seven days after planting, and the freshly prepared product was applied each month from January through August during the year of 2015. At the time of each application, 75 liters of VESTA was applied per acre. Water in the system was run for approximately 30 minutes in order to transfer the product into the strawberry root zone.

Sample collection and processing Four replicate samples each of soil and whole root systems were collected from both treatments before (time point 1), during (time points 2 and 3) and after (time point 4) the fruit harvesting season, which were approximately two, three, four and seven months after planting. Soil samples were collected from the topsoil approximately 15 centimeters (cm) from each individual plant that was sampled, and whole root systems were collected by shoveling to a depth of approximately 20 cm and sent to the lab overnight with ice packs (Supporting Information Fig. S1). Samples were stored at -80°C until processing. The root endosphere and rhizosphere fractions were separated as detailed by(Simmons *et al.*, 2018). In brief, fractions were placed in epiphyte removal buffer (0.75% KH2PO4, 0.95% K2HPO4, 1% Triton X-100 in ddH2O; filter sterilized at 0.2 μM) and separated using a sonication method (pulses at 160 W for 30 seconds, separated by a 30 second pause for 10 minutes at 4 °C). Root endosphere and rhizosphere samples were then stored in sterile epiphyte removal buffer at -80 °C until DNA extraction.

Soil chemistry analysis For a broad survey of the amendment product's influence on bulk soil chemistry, we utilized EDXRF for non-destructive analysis of soils **[Table 4-1a]**. Samples weighing 1.80 to 1.85 grams of the <2 mm sieved and dried soils were analyzed in a Spectro XEPOS HE spectrometer (AMETEK Inc.; Berwyn, Pennsylvania, USA). Each sample was interrogated at 4 spots on a 35 mm-diameter planchet, and the values averaged after normalizing to a Compton scattering background.

To standardize the chemical and physical tests and permit ease of handling, control and amendment treated soils were gently warmed from -8 °C to 3 °C overnight. Soil physical observations were performed using a <2 mm sieve to determine wet-

cohesive strength, and physical manipulation was used to determine textural characteristics and average aggregate-size using oven-dried (55 °C for 48 to 72 hours). Two grams of soil solubilized in the saturated paste extract method was used in determining soil temperature-corrected pH with an Ag/AgCl electrode (Fisher Scientific), and electrical conductivity (EC_{10}) with a conductivity probe Russell RL060-C (W.W. Grainger, Inc., Lake Forest, IL, USA) **[Table 4-1b]** (Rhoades *et al.*, 1989). The <2 mm sieved fraction of two soil samples per treatment were analyzed first for soil carbon and nitrogen using a Carlo Erba 1500NA elemental analyzer (Carlo Erba; Milan, Italy). Separate subsamples were extracted(Keeney and Nelson, 1982) with 2M KCl to determine nitrate and ammonium levels with a WESCOR SmartChem 200 discrete analyzer (Unity Scientific; Milford, MA), and exchangeable cations from soil estimated using 1 M (pH=7.0) ammonium acetate extractions, followed by inductively coupled plasma-optical emission spectrometry (ICP-OES).

Strawberry plant and product nutrient analysis The edible portions, strawberry aggregates of achenes, rely on the concerted efforts of the root zone and attendant microbiota to mine available nutrients from the soil that are then transported to the fruit. For all below ground tissue harvested at time points 1-4, Ca, Mg, K, Na, P, S, and micronutrients were analyzed with ICP-OES **[Table 4-2]**. Tissue was rinsed with distilled water, dried slowly at 55^{o}C (48-72 h), and ground to flour consistency using a reciprocating ball-mill (SPEX Inc.; Metuchen, NJ). Powder was then digested using an overnight soak in concentrated HNO_3 at room temperature, followed by heating to 180-210 ^{o}C in a programmable heating block in 2:1 (v/v) HNO_3:$HClO_4$, as described in(Miller, 1997). All values are expressed on a mg/kg oven-dry basis, utilizing subsamples dried to constant weight at 105 ^{o}C. Foliar tissues were only compared at time point 4, the final harvest, in order to not interfere with commercial fruit harvest. Composited from 5 plants spatially-dispersed in the control field and the amendment-treated field each, leaves were dried and ball-milled as described above. Energy-dispersive X-ray fluorescence (EDXRF) was then used to analyze a suite of elements similar to that for roots **[Table 4-2]**. The nutrient profile of the liquid product VESTA was assessed by D&D Agricultural Laboratory, Inc. (Fresno, CA) by means of a pH analyzer, electrical conductivity meter for soluble salts, FP-528 (Leco Corporation, St. Joseph, MI) for total nitrogen, and Optima 8000 ICP-OES (PerkinElmer, Inc., Waltham, MA) for all other parameters **[Table 4-8]**.

DNA extraction and PCR amplification Soil and rhizosphere DNA was extracted using extraction kits (MoBio PowerSoil DNA Isolation Kit, MoBio Inc., Carlsbad, CA) following the manufacturer's protocol. For root endosphere samples, roots were homogenized using mortar and pestle in liquid nitrogen and DNA was extracted using a modified CTAB DNA extraction procedure(Naylor *et al.*, 2017). Due to high humic substances in root endosphere DNA, which potentially inhibit PCR reaction, we performed a cleanup step after DNA extraction using a modified MoBio PowerSoil kit protocol supplied by the manufacturer. We previously tested that DNA extraction method had no significant effect on the DNA preparation method(Naylor *et al.*, 2017). We amplified V3-V4 region of 16S ribosomal gene using a dual-indexed 16s rRNA Illumina iTags primer (341F (5'-CCTACGGGNBGCASCAG-3') and 785R (5'-

GACTACNVGGGTATCTAATCC-3') as described in(Takahashi *et al.*, 2014) using 5-Prime Hot Master Mix (catalog No. 2200410). After DNA extraction, DNA samples were diluted to 5 ng/ µl and randomized in 96-well plates. Water blanks were included on each 96-well plate as negative controls. PNA clamps were used to minimize host-derived amplicons from both chloroplast and mitochondrial 16S rRNA gene sequences(Lundberg *et al.*, 2013). Reactions included 11.12 µL DNase-free sterile H20, 0.4 µg BSA, 10.0 µL 5-Prime Hot Master Mix, and 2 µL template, and 0.75 µM of chloroplast and mitochondria PNAs. PCR reactions were performed in triplicate in three thermocyclers (to account for possible thermocycler bias) with the following conditions: initial 3 min cycle at 94°C, then 30 cycles of 45 seconds at 94°C, 10 sec at 78°C, 1 min at 50°C, and 1.5 min at 72°C, followed by a final cycle of 10 min at 72°C. Triplicates were then pooled (128 samples per library) and DNA concentration for each sample was quantified using Qubit reader. Pools of amplicons were constructed using 100 ng for each PCR product. Before submitting for sequencing, pooled samples were cleaned up with 1.0X volume Agencourt AMPureXP (Beckman-Coulter, West Sacramento, CA) beads according to the manufacturer's directions, except for the modifications of using 1.0X rather than 1.6X volume beads per sample, dispensing 1500 µL 70% EtOH to each well rather than 200 µL, and eluting in 100 µL DNase-free H20 rather than 40 µL. An aliquot of the pooled amplicons was diluted to 10 nM in 30µL total volume before submitting to the QB3 facility at UC Berkeley for sequencing using Illumina Miseq 300bp pair-end with v3 chemistry.

Amplicon sequence processing, OTU classification and taxonomic assignment

Sequencing data was analyzed using the iTagger pipeline developed by the U.S. Department of Energy's Joint Genome Institute(Tremblay *et al.*, 2015). This pipeline wraps several packages for the filtering, merging, clustering and taxonomy assignment, including CUTADAPT, FLASH, USEARCH, and RDP(Wang *et al.*, 2007; Edgar, 2010; Magoč and Salzberg, 2011; Martin, 2011). In brief, after filtering 28,581,170 16S rRNA raw reads for known contaminants (Illumina adapter sequence and PhiX), primer sequences were trimmed from the 5' ends of both forward and reverse reads. Low-quality bases were trimmed from the 3' ends prior to assembly of forward and reverse reads with FLASH(Magoč and Salzberg, 2011). The remaining 18,717,158 high-quality merged reads were clustered with simultaneous chimera removal using UPARSE(Edgar, 2013). After clustering, 11,204,438 read counts mapped to 13, 545 operational taxonomic units (OTUs) at 97% identity. The resulting reads produced on average approximately 72,298, 88,493 and 76,070 reads per sample for soils, rhizospheres, and roots respectively. Taxonomies were assigned to each OTU using the RDP Naïve Bayesian Classifier(Wang *et al.*, 2007) with custom reference databases. For the 16S rRNA V3-V4 data, this database was compiled from the May 2013 version of the GreenGenes 16S database(DeSantis *et al.*, 2006), the Silva 16S database(Quast *et al.*, 2013) and additional manually curated 16S rRNA sequences, trimmed to the V3-V4 region. After taxonomies were assigned to each OTU, we discarded 1) all OTUs that were not assigned a Kingdom level RDP classification score of at least 0.5, 2) all OTUs that were not assigned to Kingdom Bacteria. To remove low abundance OTUs that are in many cases artifacts generated through the sequencing process, we removed OTUs without at least 5 reads in at least 5 samples. We also removed samples have less than 10,000 reads, which yielded

3,387 high-abundance OTUs (respectively) for downstream analyses. These thresholds were found to be suitable using technical replicates in a dataset published previously(Coleman-Derr *et al.*, 2016). To account for differences in sequencing read depth across samples, all samples were rarefied to 16,542 reads per sample for specific analyses, or alternatively, by dividing the reads per OTU in a sample by the sum of usable reads in that sample, resulting in a table of relative abundance frequencies; OTUs which were reduced to less than one read per OTU after rarefaction were discarded to yield 2,464,758 measurable, rarefied reads for downstream analysis. The raw sequencing reads for this project will be deposited in the NCBI Short Read Archive SUB3636449.

Shotgun metagenomic sequence processing and analysis of amendment product In order to assess the metagenomic composition of the VESTA product itself, three 50mL aliquots were made from the same batch of VESTA and treated as technical replicates. To extract genomic DNA, the 50 mL aliquots of homogenized product were centrifuged at 10,000 rpm for 20 minutes. The supernatant was discarded, and the pellet was then re-homogenized. Of the resulting sediment, 0.250 grams were processed according to manufacturer instructions with the PowerSoil DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA). DNA was sheared to 300 bp using a Covaris Focused-ultrasonicator (Covaris, Inc., Woburn, MA, USA), and then used to generate libraries with the Kapa LTP Library Preparation Kit (Kapa Biosystems, Inc., Wilmington, MA) and adapters and unique barcodes provided by the QB3-Berkeley facility to allow for multiplexing. Sequencing was performed at QB3-Berkeley on the Illumina HiSeq2500 system with 150 bp paired-end reads. De-multiplexing was performed by the California Institute for Quantitative Biosciences (QB3-Berkeley) Functional Genomics Laboratory. We received sequences for 266 million reads, or an average of 88.5±4.1 million reads per technical replicate, for a total of 45.5 billion bases. Raw reads were uploaded to Metagenomics RAST server (MG-RAST) for paired-end joining, quality control, and taxonomic and functional annotation(Meyer *et al.*, 2008). The raw metagenome sequencing reads for this project will be deposited in the NCBI Short Read Archive SUB3688024. After quality control, one of the three technical replicates was revealed to be of much lower quality and was discarded. The two replicates we chose to retain accounted for a total of 172 million reads, of which 21.1 million (12.3%) failed quality control, 11.4 million (6.63%) could not be identified, and 139 million (81.1%) were assigned functional and/or taxonomic annotations according to default parameters (alignment length cutoff = 15bp, e-value cutoff = e-5, percent identity cutoff = 60%) using RefSeq database(O'Leary *et al.*, 2016). Of those sequences which could be identified, 329,113 sequences (0.24%) contain ribosomal RNA genes, 63.6 million sequences (45.7%) contain predicted proteins with known functions, and 75.2 million sequences (54.0%) contain predicted proteins with unknown function. Over 97% of all reads were determined to belong to bacteria, compared to only 0.12% of reads being assigned fungal identity. For this reason, we decided to focus solely on the bacterial sequences in this study. Functional annotations according to the SEED subsystem database(Overbeek *et al.*, 2014) are presented in this publication.

Statistical analysis RStudio (version 1.0.136; RStudio Team) was utilized for all statistical analyses with the packages phyloseq(McMurdie and Holmes, 2013) and

vegan(Oksanen *et al.*, 2016). For plant phenotype data, scatter plots were generated using ggplot2, and Analysis of Variance (ANOVA) was performed with function aov. For the Alpha diversity measurement, Shannon Index of diversity and observed OTUs were calculated with the estimate_richness function in the R package phyloseq. ANOVAs were performed with function aov for Sample Type, Treatment, and Time Point. A Tukey's Post Hoc test was performed using function TukeyHSD in the stats package and with HSD.test in the package agricolea to test which levels were significantly different from one another. Beta diversity was measured using Bray-Curtis distances and UniFrac distance with function ordinate in the R package phyloseq. For UniFrac distances, trees were built with default parameters using FastTree(Price, Dehal and Arkin, 2010) with an alignment constructed in Muscle(Edgar, 2004). Canonical Analysis of Principal Coordinates (CAPs) was performed for subsets of the data with each sample type and time point to determine the percent variance explained by treatment, time point and replicate, or treatment, sample type, and replicate, respectively, using the capscale function in the R package vegan(Oksanen *et al.*, 2016). The non-parametric Kruskal-Wallis test in R was used to compare Shannon indices and class-level relative abundances between treated and untreated within each time point and sample type. Indicator species analyses run on root samples to determine genera that were enriched for either control or amendment treatments were performed using R package indicspecies(De Caceres, Jansen and De Caceres, 2016), with p-values < 0.01 based on permutation tests run with 999 permutations. All scripts used can be found at a public repository on github (https://github.com/siwendeng/Strawberry-Microbiome).

Figure 4-1

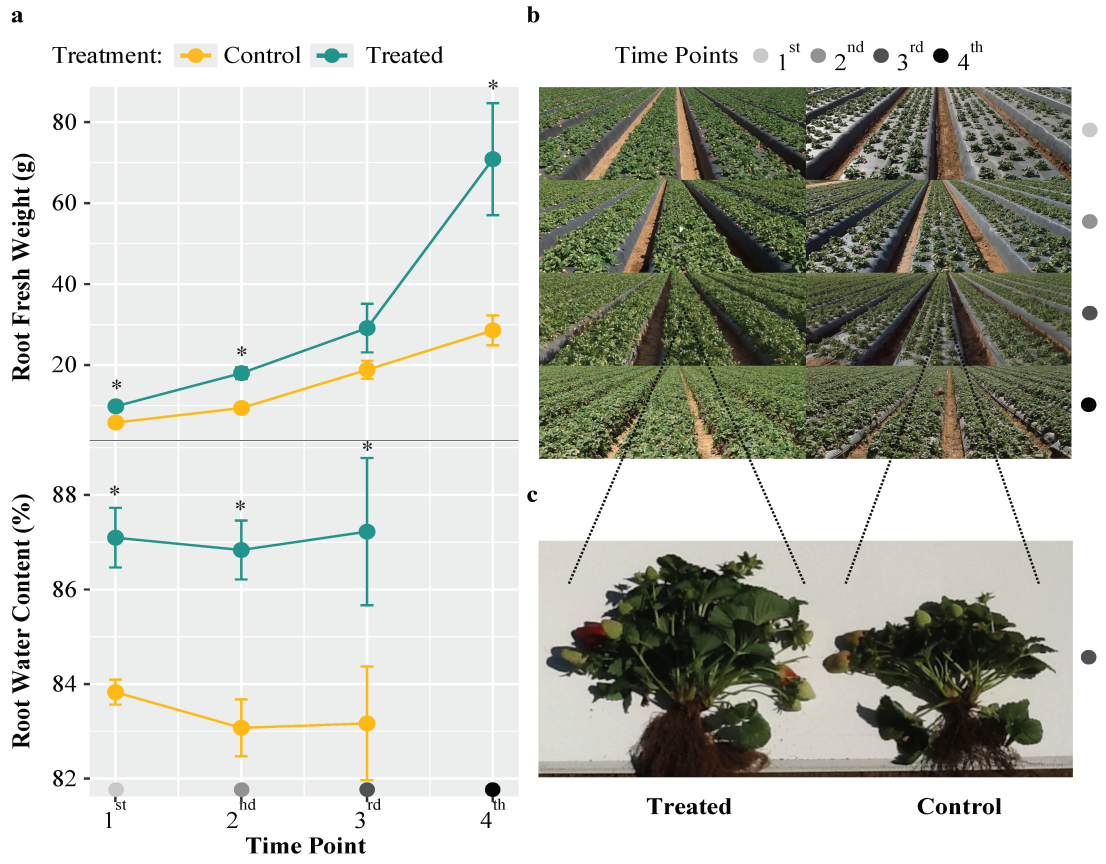

Figure 4- 1. Treatment effect on strawberry root growth

a. Line plot illustrating the average trend of 5 to 6 replicates per treatment per time point for root fresh weight (RFW) in grams and percentage Root Water Content (RWC) in amendment-treated (indicated in blue) and control (yellow) strawberry plants across the four sampling time points. RWC was calculated using (RFW – root dry weight)/RFW. Data shown are mean ± SE. ('*' indicates p-value < 0.01) **b.** Photographs of amendment-treated and control strawberry fields across four time points in upper panel. Four time points correspond to sampling dates of March 18 (light grey circle), April 8 (grey circle), May 13 (dark grey circle), and August 24 (black circle) in 2015. **c.** Two representative plants from treated and control fields from the 3rd sampling time point in lower panel. Amendment-treated strawberry plant displays enhanced root proliferation at the crown and greater aboveground shoot biomass.

Figure 4-2

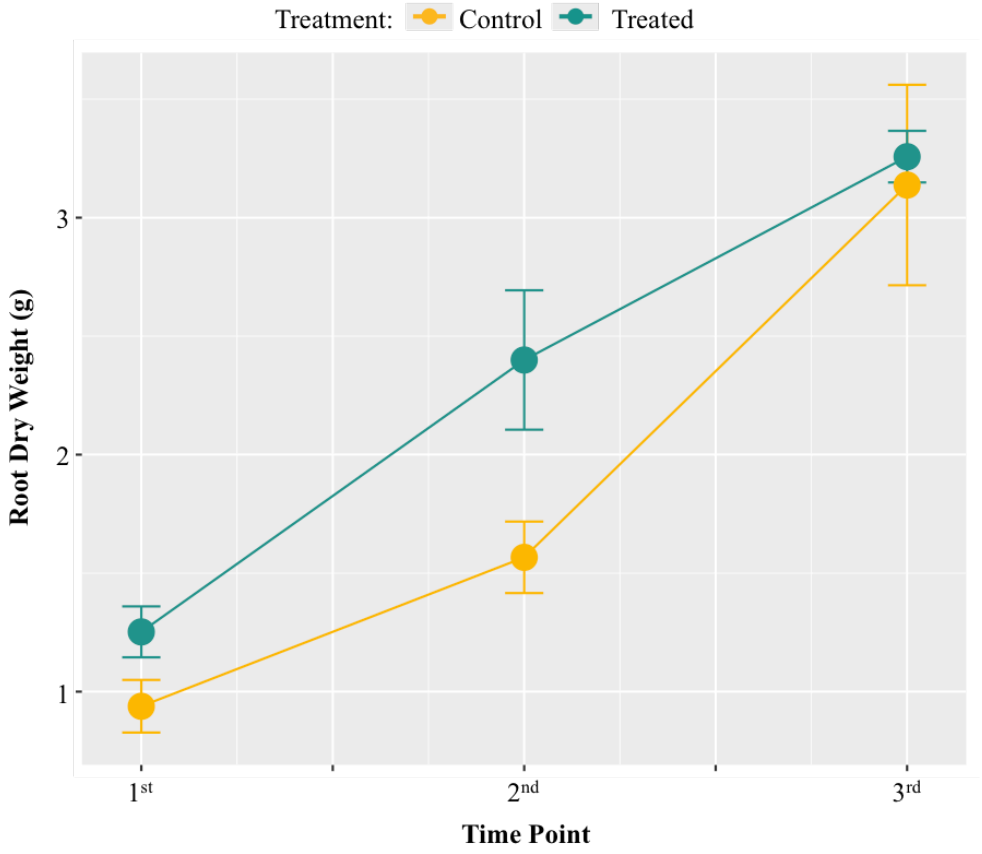

Figure 4- 2. Treatment effects on strawberry plant dry weight

Line plot illustrating the average trend of 5 to 6 replicates per treatment per time point for root dry weight across the three sampling time points.

Figure 4-3

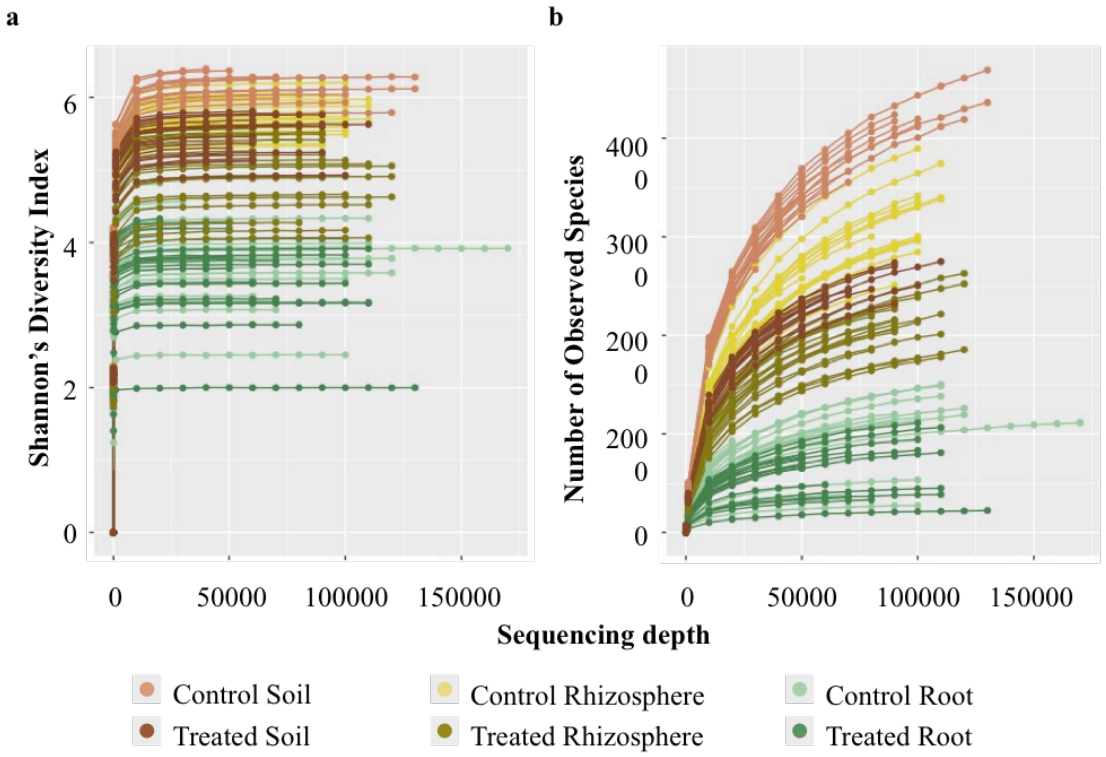

Figure 4- 3. Rarefaction curves to estimate alpha diversity

a. Shannon's diversity indices, and **b.** number of observed species, as a function of sequencing depth.

Figure 4-4

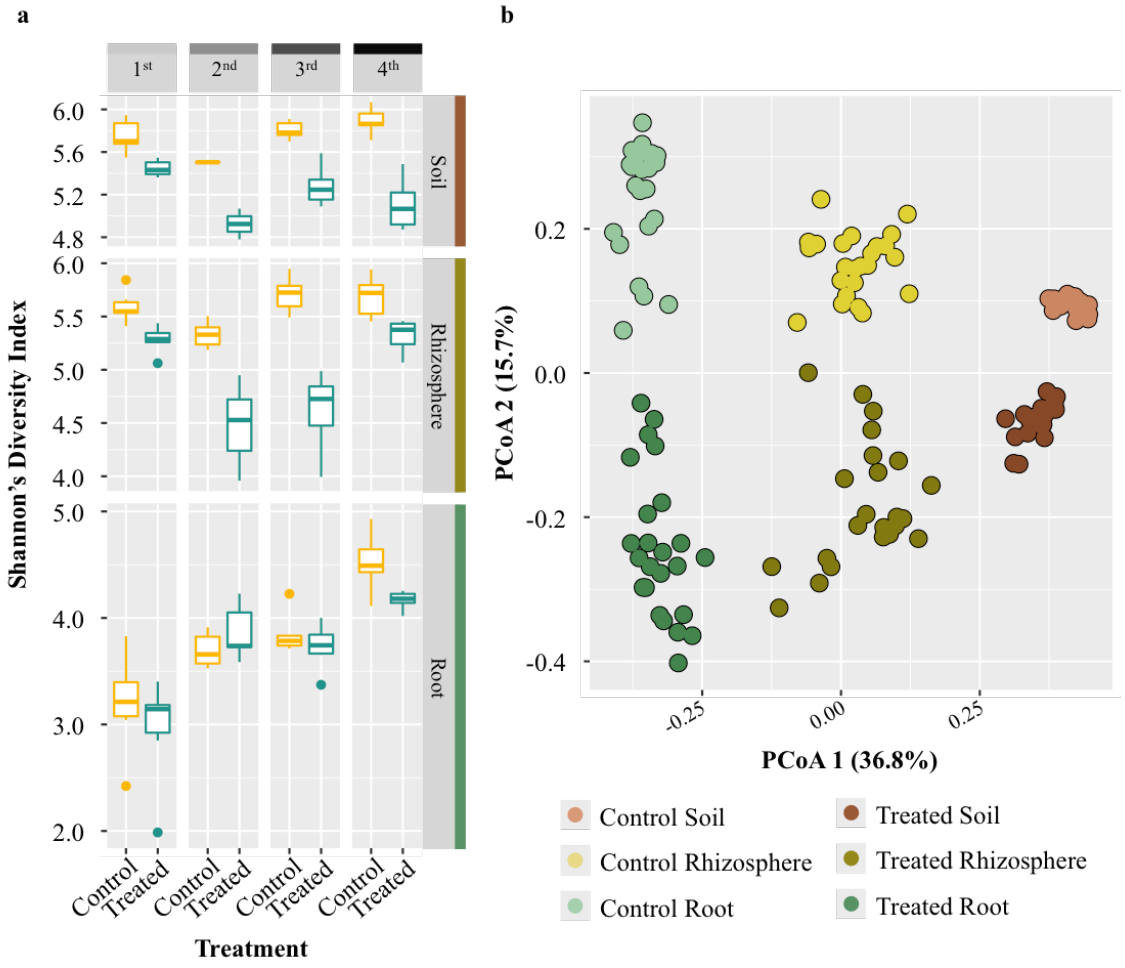

Figure 4- 4. Treatment effect on strawberry root microbiome

a. Box and whisker plots representing Shannon's Diversity indices for all samples based on Bray-Curtis distances in amendment-treated (blue) and control (yellow) samples for each sample type and at each of the four sampling time points; soil (upper panel), rhizosphere (middle panel), and root (lower panel) indicate that amendment treatment correlates with a decrease in the within-sample diversity in soil and rhizosphere samples, but not in roots. The horizontal line within each box represents the median index value, and the bottom and top edges of each box indicate the 25th and 75th percentiles, respectively. Individual points are outliers. **b.** Principal coordinate analysis (PCoA) plot for all samples using Bray-Curtis distances indicates that the largest source of variation between microbial communities is sample type (PCo 1, 36.8%) and the second largest source of variation is treatment (PCo 2, 15.7%). Both control soil (light brown circles), rhizosphere (light yellow), and root (light green) samples and amendment-treated soil (dark brown), rhizosphere (dark yellow), and root (dark green) samples cluster together within their respective treatment and sample types.

Figure 4-5

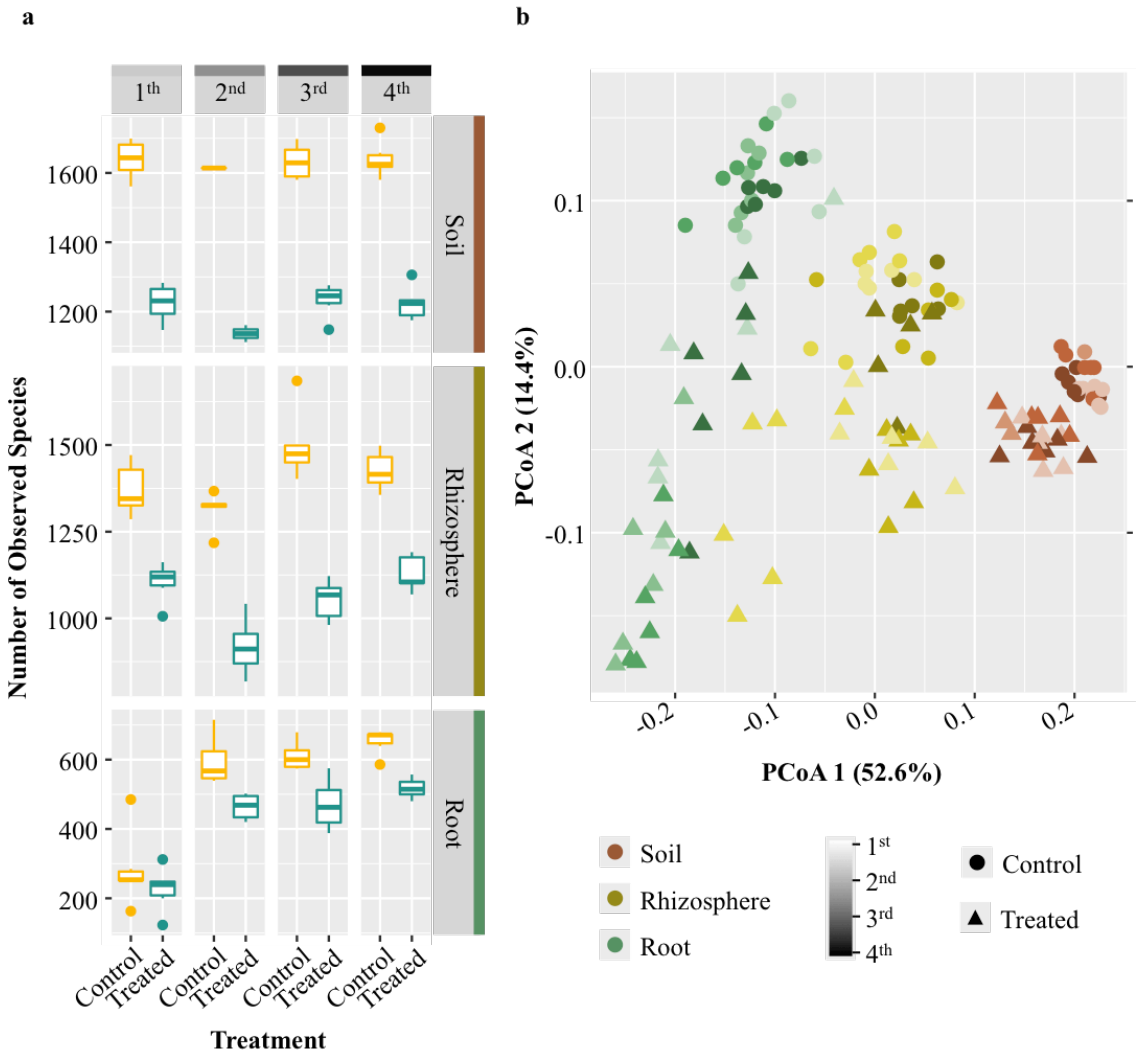

Figure 4- 5. Amendment application correlates to reduced numbers of observed species, and samples cluster by treatment

a. Box-and-whisker plots of the numbers of observed species in amendment-treated (blue) and control (yellow) samples for each sample type (soil, rhizosphere, root) and across the four sampling time points (1st, 2nd, 3rd, 4th). The horizontal line within each box represents the median. The bottom and top edges of each box indicate the 25th and 75th percentiles, respectively. Individual points are outliers. **b.** Principal coordinate analysis (PCoA) plot for all samples generated based on weighted UniFrac distances. The first two axes explain 67.0% of the data, with the primary axis (52.6% of variance) primarily distinguishing samples by sample type and the secondary axis (14.4% of variance) distinguishing samples by treatment type (amendment-treated or control).

Figure 4-6

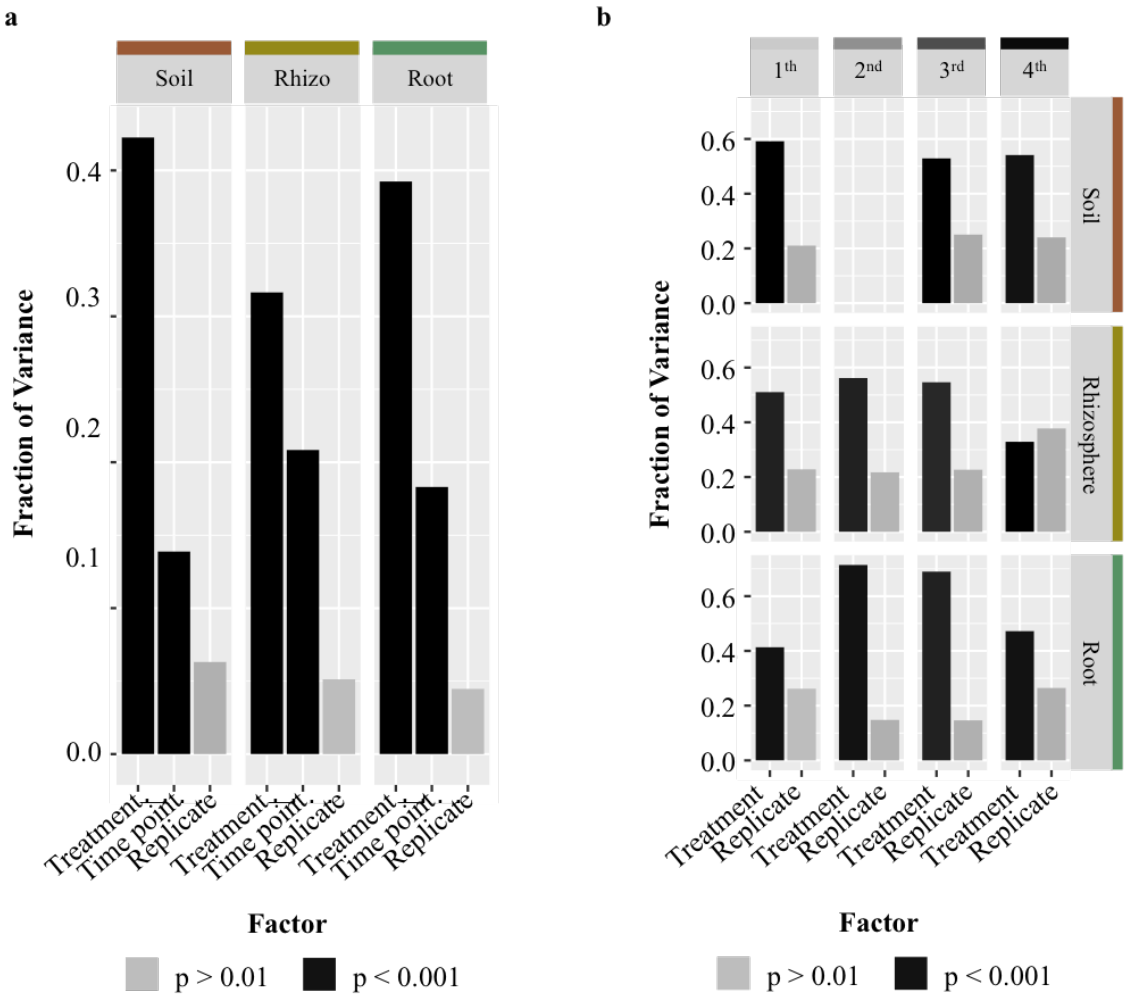

Figure 4- 6. Treatment type explains the greatest bacterial community variance for each sample type and time point

a. Bar plots displaying the fraction of variance explained by treatment, time point and replicate, as determined by canonical analysis of principal coordinates (CAPS) using Bray-Curtis distances. The y-axis indicates the fraction of variance explained by each factor, and the shade of the bar indicates the level of significance ($p>0.01$ is in light grey; $p<0.001$ is in dark grey). **b.** Bar plots displaying the fraction of variance determined by CAPS using Bray-Curtis distances and performed separately for each time point (1st, 2nd, 3rd, and 4th) on soil (shown in the upper panel), rhizosphere (middle panel), and root (lower panel) samples. The fraction of variation is indicated by the y-axis, and the shade of the bar indicates the level of significance.

Figure 4-7

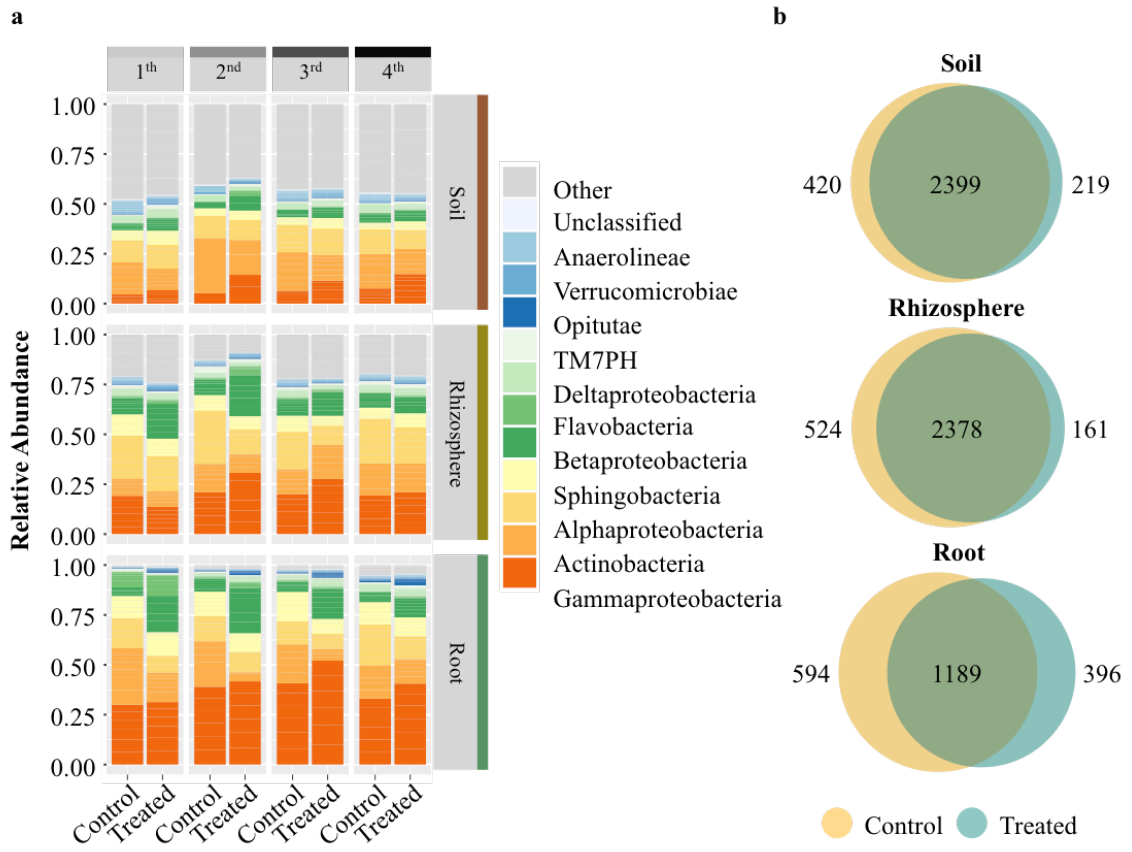

Figure 4- 7. Class-level relative abundance of bacterial communities indicate distinct soil, rhizosphere, and root profiles

a. Relative abundances bar graphs for the top 12 most abundant bacterial classes for amendment-treated and control samples in soil (upper panel), rhizosphere (middle panel), and root (lower panel) across four time points (1st, 2nd, 3rd, 4th). **b.** Venn-diagrams displaying numbers of OTUs shared and distinct between amendment-treated (blue) and control (yellow) samples in soil (upper panel), rhizosphere (middle panel), and root (lower panel) compartments.

Figure 4-8

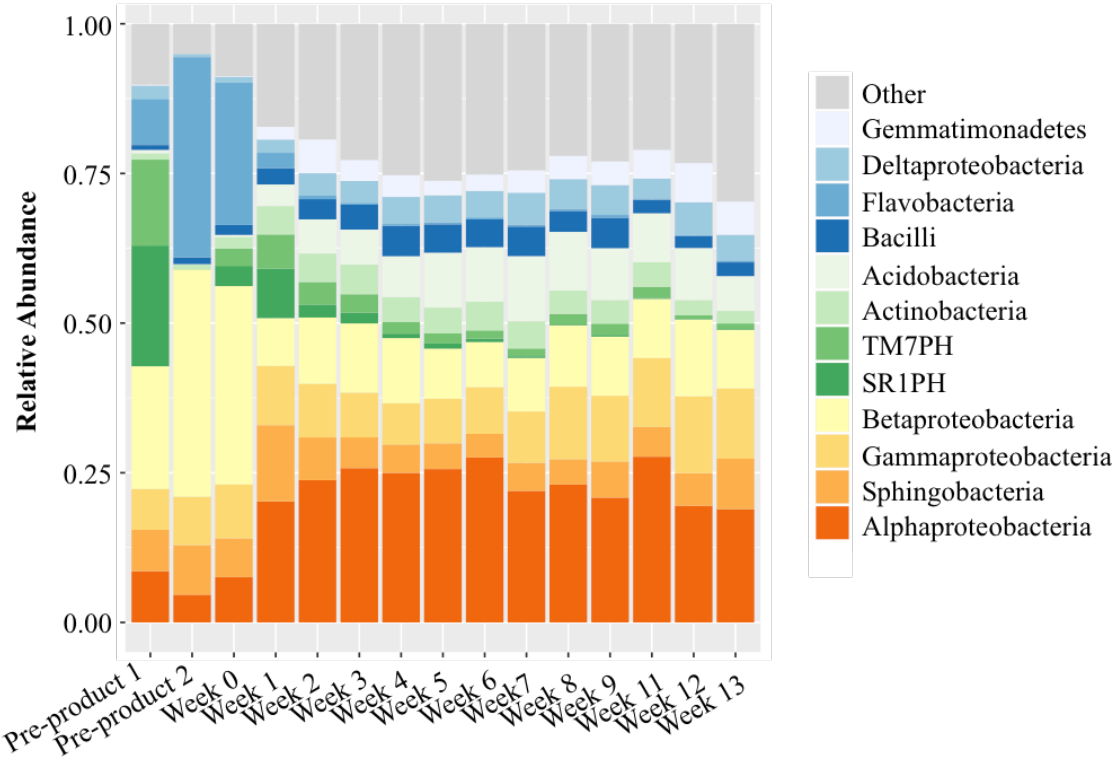

Figure 4- 8. Relative abundance of bacterial communities of the amendment

Relative abundance bar graphs showing bacterial communities shift from the pre-products throughout a 13 weeks time course of the amendment Relative abundance bar graphs of the bacterial classes present in the initial pre-product. 1 BHF-10® and pre-product 2 SOBEC®, as well as samples of the product VESTA® collected over a time course of 13 weeks. Week 0 was collected immediately after mixing the two pre-products, Week 1 was collected after one week, Week 2 was collected after two weeks, etc.

Figure 4-9

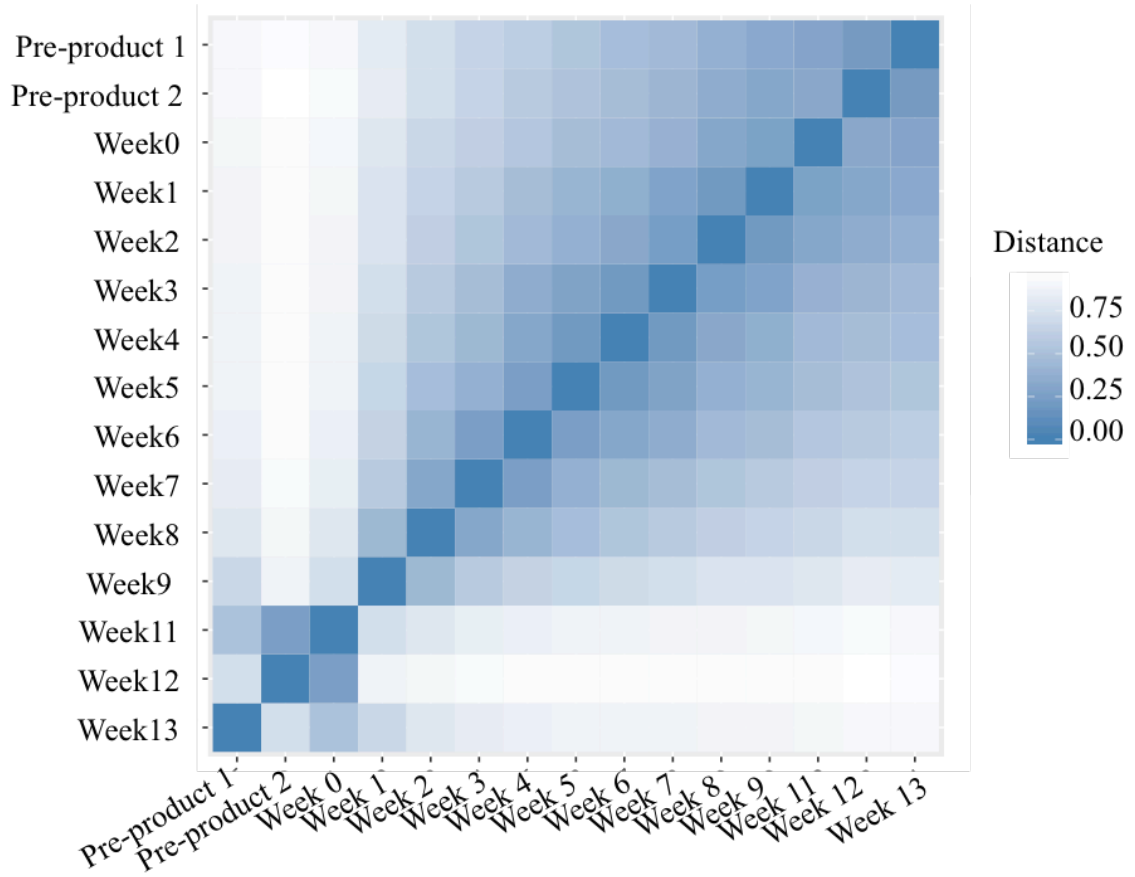

Figure 4- 9. Pairwise distances between the amendment products across time points

Heat map representative of the pairwise Bray-Curtis distances between samples of the two pre-products and the samples of the amendment product across the 13 week time course the time points of which the amendment product was sampled. Degree of blue shading corresponds to Bray-Curtis distance value, where darker shading corresponds to less distance between the bacterial community profiles of samples.

Figure 4-10

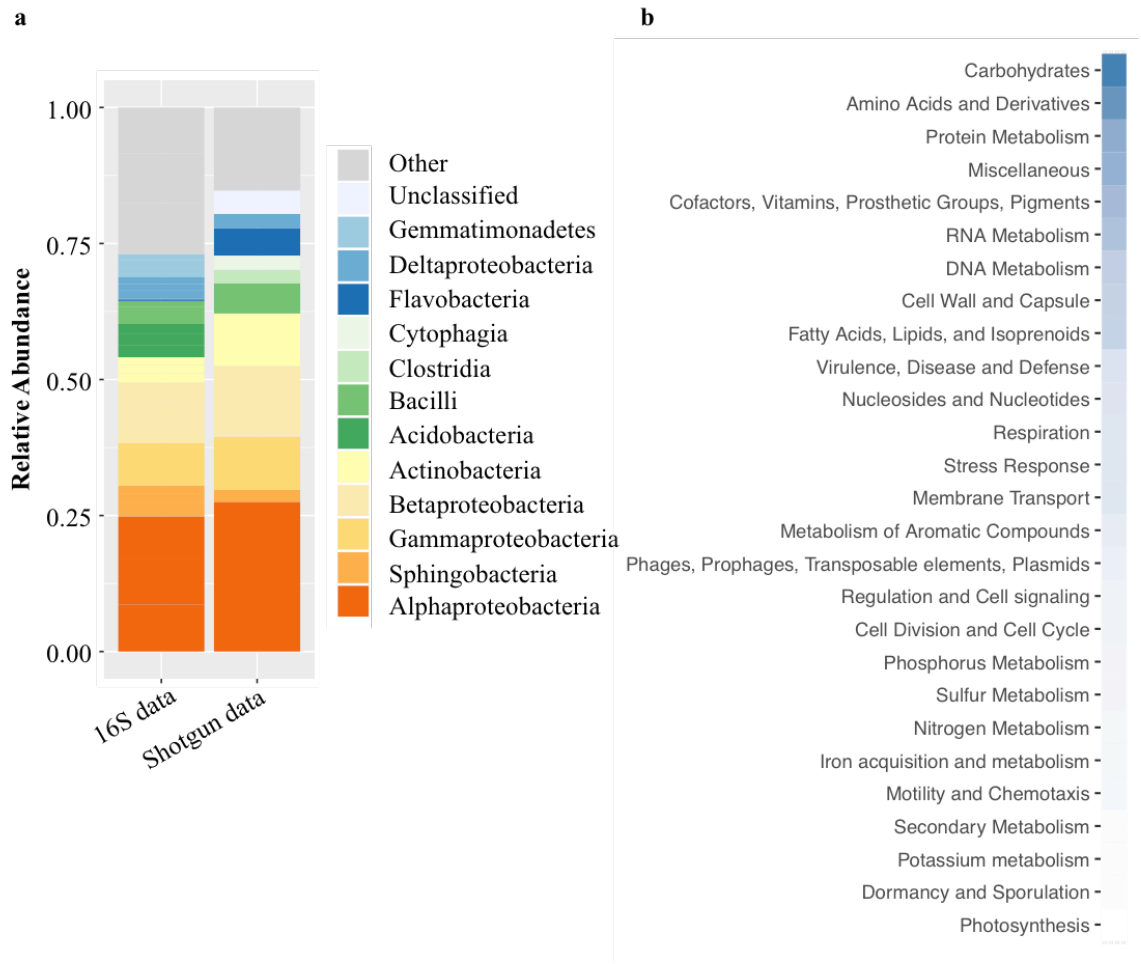

Figure 4- 10. Shotgun metagenomic analysis of the amendment product

a. Comparison of the relative abundances of bacterial classes between 16S rRNA amplicon sequencing and shotgun whole genome sequencing of the amendment product. This slightly differs from 16S rRNA bacterial relative abundances. **b.** Functional profile of the amendment product and the relative abundances of genes ascribed to the listed categories, where relative abundances range from 0.1% - 14.5% and greater enrichment is indicated by corresponding block being more darkly shaded. There is high enrichment in carbohydrate, amino acid, and protein metabolism.

Figure 4-11

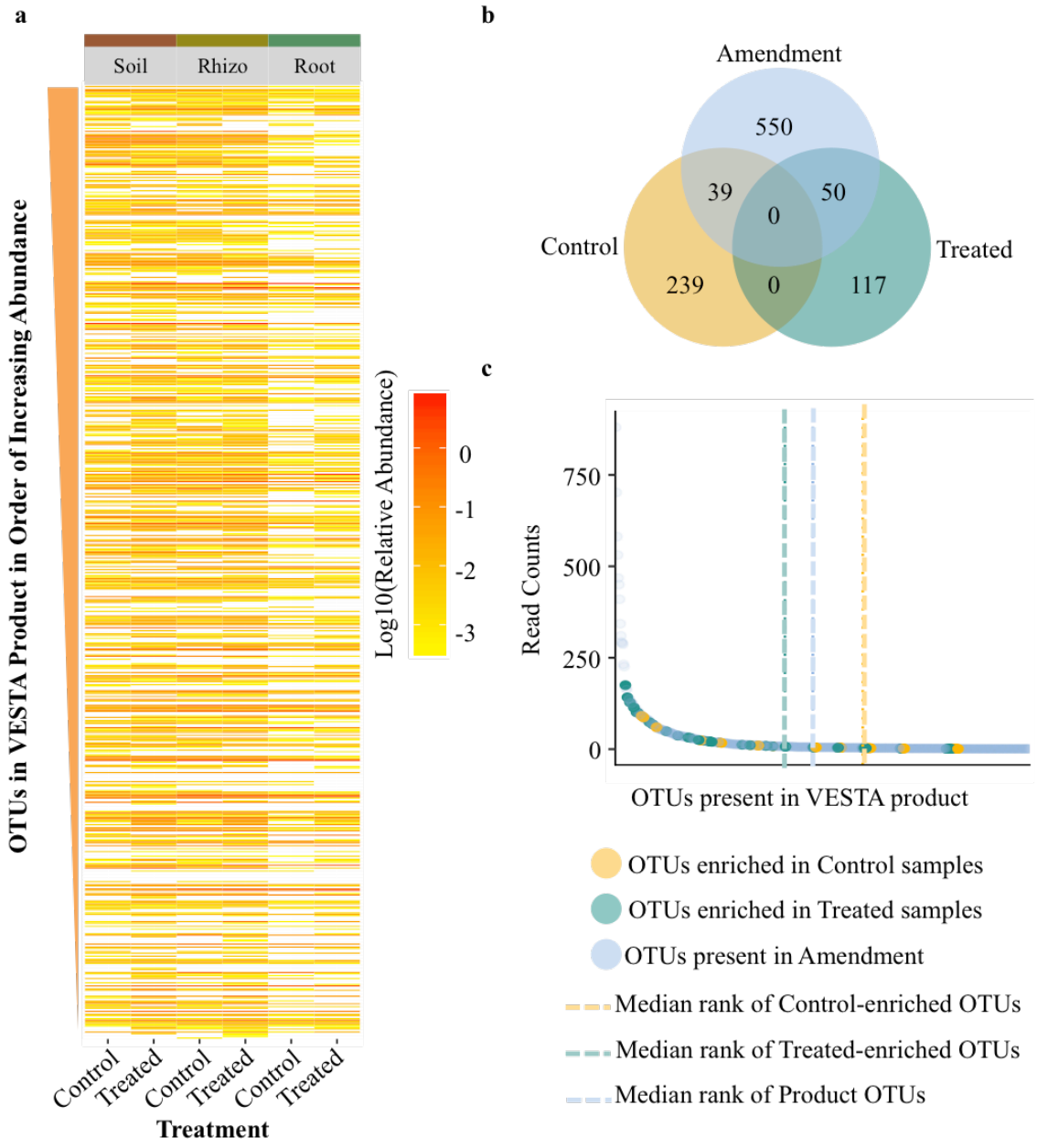

Figure 4- 11. OTUs in the product are not highly enriched in treated or control samples

a. Heatmap displaying degree of enrichment by log10 relative abundance in treated and control soil, rhizosphere, and root samples for the 639 OTUs present in the product. Product OTUs were sorted according to their relative abundance from top (highest) to the bottom (lowest). Darker shades of color indicate higher relative abundance within the samples, whereas lighter shades indicate lower relative abundance. Columns for amendment-treated soils, rhizospheres, and roots generally are darker, indicating greater

enrichment in product OTUs than controls, but there was no significant enrichment. **b.**Venn-Diagram displaying the OTUs shared and distinct to the product and those uniquely enriched in the amendment-treated or control roots, as determined with indicator species analysis. OTUs uniquely enriched in control (yellow) and treated (green) samples were compared with those present in the product (light blue). More OTUs from the product were distinctly shared with treated samples (50 vs. the 39 OTUs shared between the product and controls), and treated samples shared a larger proportion of total OTUs (30% or 50 of 167 OTUs total) with the product than controls (14% or 39 of 278 OTUs total). **c.** All OTUs present in the product were ranked by their relative abundance, with the y-axis representing the average read count for each product OTU. The median rank for OTUs which showed enrichment in the control (yellow circles; n=39) or treated roots (green circles; n=50) was calculated and is displayed with yellow and green dashed lines, respectively; the median rank for all OTUs in the product with no enrichment (light blue circles; n=550) is shown by the blue dashed line. This demonstrates that the median rank abundance in the product was roughly 1.5x higher in treated roots than that for OTUs enriched in controls.

Figure 4-12

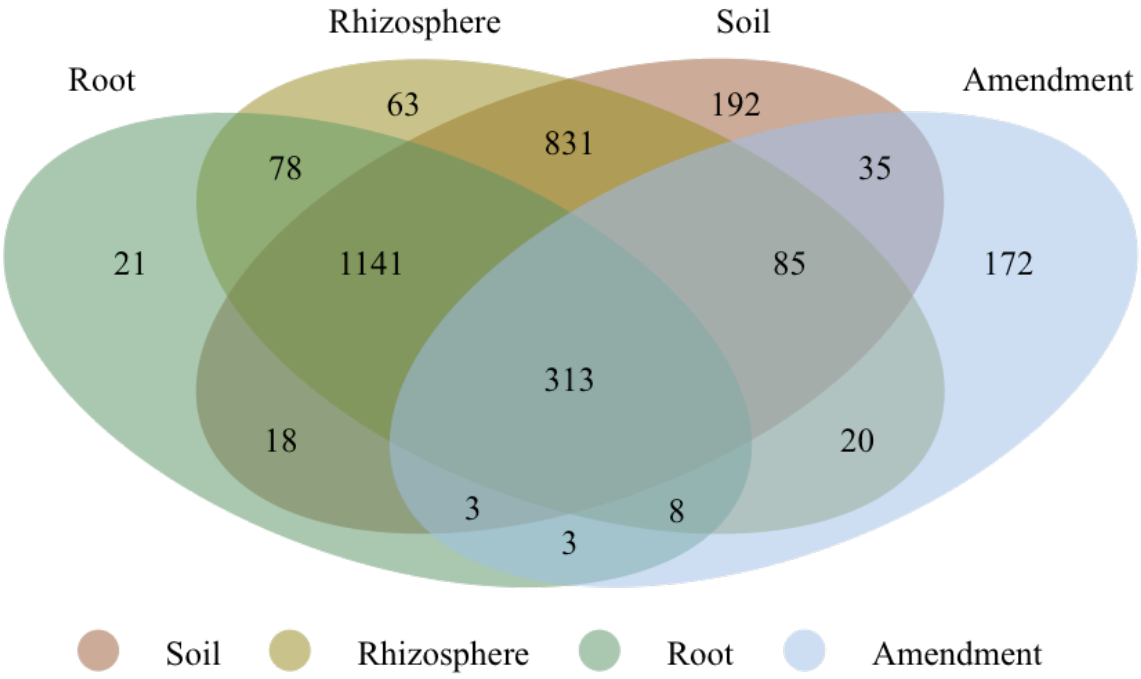

Figure 4- 12. Venn-Diagram of OTU numbers across different sample types

A majority of OTUs in the amendment product (as shown by the blue sphere) are present either within the soil (brown), rhizosphere (yellow), or root (green) compartment of treated samples. Venn-Diagram displaying the numbers of OTUs from amendment-treated soil, rhizosphere, and root samples that are shared or unique to the OTUs found in the amendment. OTUs used as determined by 16S amplicon sequencing.

Table 4-1

Table 4- 1. Field soil chemical analyses in treated and control soils

a. Soil pH was measured with 2.00 grams of soil with a Fischer Tris-electrode after a 3-point calibration (n=13 control; m=9 treated). Electrical conductivity (EC) was also measured from 2 grams of water solubilized soil with a Thermo-Russell RL060-C conductivity meter, and total nitrogen and carbon were measured with a CarloErba1500N elemental analyzer. Total nitrogen and carbon data reported here is the average of two replicates. Nitrate levels were determined by WESCOR analyses on n=11 soil samples from control, and m=10 from the amendment-treated fields. Exchangeable Ca analyzed over 3 time points, mean + standard error of mean for n=10 controls, m=9 treated. **b.** Total elemental composition of soils, as determined by energy-dispersive X-ray fluorescence (EDXRF); values are expressed as mean ± standard error of the mean in dry weight or µg/g dry wt, and n=5 controls, m=4 for amendment-treated plants. Though many other elements were determined, their values were near instrumental detection limits, as EDXRF is only suitable for elements between Z> 11 (Na) and Z<92 (U). MANOVA significance values when tested for differences between values for treated and control samples: * (P<0.05), ** (P<0.01), and *** (P<0.001).

a. Chemical parameters measured in soil from treated and control fields		
Parameter	Amount in Control Soil	Amount in Treated Soil
pH	7.32 ± 0.04	7.44 ± 0.04
EC	184 ± 15 µS/cm	242 ± 22 µS/cm *
Total Nitrogen	0.177% dry weight	0.188% dry weight
Total Carbon	0.249% dry weight	0.385% dry weight
Nitrate	23.2 ± 1.51 ppm	18.6 ± 1.51 ppm *
Ammonium	6.36 ± 0.96 ppm	5.80 ± 0.93 ppm
Exchangeable [Ca^{++}]	1256 ± 34 mg/kg	1427 ± 23 mg/kg ***
b. Total elemental composition of soils from treated and control fields		
Parameter	Amount in Control Soil	Amount in Treated Soil
pH	7.32 ± 0.04	7.44 ± 0.04
EC	184 ± 15 µS/cm	242 ± 22 µS/cm *
Total Nitrogen	0.177% dry weight	0.188% dry weight
Total Carbon	0.249% dry weight	0.385% dry weight
Nitrate	23.2 ± 1.51 ppm	18.6 ± 1.51 ppm *
Ammonium	6.36 ± 0.96 ppm	5.80 ± 0.93 ppm
Exchangeable [Ca^{++}]	1256 ± 34 mg/kg	1427 ± 23 mg/kg ***

Table 4-2

Table 4- 2. Chemical analyses of treated and untreated strawberry plants

a. Acid-digestible nutrient pools from strawberry root tissues collected at time points 1-3, as determined by inductively coupled plasma emission spectroscopy (ICP-EOS); values are expressed as mean ± standard error of the mean in µg/g dry wt, and n=8, m=8 for control and treated plants. **b.** Total elemental composition by EDXRF of strawberry foliage; values are composited from n=5 controls, m=5 for VESTA-treated plants. NBS-1533a Standard pine needles were also run on the same instrument, and elements for which certificate values are registered were satisfactorily quantified/recovered. MANOVA significance values when tested for differences between values for control and treated samples: * (P<0.05), ** (P<0.01), and *** (P<0.001).

a. Acid-digestible nutrient pools from strawberry root tissues of plants grown in treated and control fields, collected at time points 1-3.		
Nutrient	Amount in Control Roots (µg/g)	Amount in Treated Roots (µg/g)
Aluminum	3114 ± 371	2073 ± 172 *
Boron	25.65 ± 8.61	15.41 ± 6.71
Calcium	8760 ± 692	7068 ± 521
Potassium	4006 ± 374	4511 ± 294
Magnesium	3916 ± 229	3684 ± 283
Phosphorus	1747 ± 107	2315 ± 89.6 ***
Silicon	511.4 ± 70.3	447.3 ± 71.8
Sulfur	18496 ± 3029	16192 ± 1282
Iron	2995 ± 340	1951 ± 151 *
Copper	33.9 ± 14.9	19.96 ± 1.81
Manganese	103.6 ± 12.1	105.6 ± 9.90
Molybdenum	3.505 ± 0.288	2.602 ± 0.193 *
Sodium	1096 ± 159	1415 ± 191
Zinc	53.62 ± 9.37	52.44 ± 2.65
b. Total elemental composition of strawberry foliage from plants grown in treated and control fields, collected at time point 4.		
Nutrient	Amount in Control Leaves	Amount in Treated Leaves
Sodium	0.31%	0.39%
Magnesium	0.25%	0.29%
Aluminum	< 0.0020 %	< 0.0020 %
Silicon	0.13%	0.27%
Phosphorous	0.18%	0.25%

Sulfur	0.12%	0.17%
Chlorine	0.07%	0.11%
Potassium	0.74%	0.99%
Calcium	0.47%	0.69%
Titanium	18.5 mg/kg	27.5 mg/kg
Vanadium	0.2 mg/kg	0.7 mg/kg
Chromium	0.6 mg/kg	0.8 mg/kg
Manganese	34.9 mg/kg	52.3 mg/kg
Iron	177.6 mg/kg	283.1 mg/kg
Cobalt	< 3.0 mg/kg	< 2.9 mg/kg
Nickel	6.7 mg/kg	9.4 mg/kg
Molybdenum	6.1 mg/kg	3.4 mg/kg

Table 4-3

Table 4- 3. Factors explaining the variation in Shannon's Diversity of all samples

a. Results from a Tukey's HSD (honest significant difference) test of Shannon's Diversity indices, as based on Bray-Curtis distances. 'SD' = standard deviation. **b.** ANOVA analyses for Shannon's Diversity across all samples based on the mixed linear model with factors sample type, treatment, time point and replicate as explanatory variables. Significant factors indicated by ***, denoting a p value of less than .001.

a. Results from a Tukey's HSD (honest significant difference) test of Shannon's Diversity indices, as based on Bray-Curtis distances.						
SampleType by Treatment	Means	S.D.	R	Min	Max	Groups
Soil.Control	5.7984	0.1464	19	5.5038	6.0671	a
Rhizosphere.Control	5.5786	0.2135	24	5.1865	5.9477	ab
Soil.Treated	5.2411	0.2392	20	4.7793	5.5888	bc
Rhizosphere.Treated	4.9072	0.4673	23	3.9585	5.4575	c
Root.Control	3.8148	0.5585	24	2.4223	4.9298	d
Root.Treated	3.6779	0.5394	24	1.9864	4.2547	d
b. ANOVA analyses for Shannon's Diversity across all samples based on the MLM with factors sample type, treatment, time point and replicate as explanatory variables.						
Factor	DF	Sum of Squares	Mean Squares	F Statistics	p-value	
SampleType	2	82.59	41.3	271.701	<0.001	***
Treatment	1	6.68	6.68	43.968	<0.001	***
Time point	3	4.2	1.4	9.202	<0.001	***
Replicate	5	0.32	0.06	0.417	0.836	
Residuals	122	18.54	0.15			

Table 4-4

Table 4- 4.Kruskal-Wallis test on Shannon indices between treatments

Table of Kruskal-Wallis test results comparing Shannon indices between treated and untreated samples within each sample type and time point. Significant factors indicated by *, **, and ***, denoting p value of less than .05, .01, and .001, respectively.

Sample Type	Time Point	Kruskal-Wallis Test		
		chi-squared	p-value	
Soil	1	8.308	0.0039	**
Soil	2	NA	NA	
Soil	3	8.308	0.0039	**
Soil	4	8.308	0.0039	**
Rhizosphere	1	7.410	0.0065	**
Rhizosphere	2	8.308	0.0039	**
Rhizosphere	3	8.308	0.0039	**
Rhizosphere	4	6.533	0.0106	*
Root	1	1.256	0.2623	
Root	2	1.256	0.2623	
Root	3	0.641	0.4233	
Root	4	5.026	0.0250	*

Table 4-5

Table 4- 5. Factors explaining the variance in bacterial community for all sample types

Treatment and time point are highly significant factors explaining the variance in bacterial community differences between samples within each compartment (soil, rhizosphere, and root). Table of PERMANOVA analysis results using Bray-Curtis distances and performed independently on each sample type using the adonis function within the package vegan in R 70. 'TRT" = treatment factor, 'DF' = degrees of freedom, and significant factors indicated by *, **, and ***, denoting p value of less than .05, .01, and .001, respectively

Sample Type	Factors	DF	Sum of Squares	Mean Squares	F Statistic	R^2	p-value	
Soil	Treatment	1	1.4555	1.4555	35.278	0.4226	<0.001	***
Soil	Time point	3	0.4781	0.1594	3.863	0.1388	<0.001	***
Soil	TRT : Time point	3	0.2314	0.0772	1.87	0.0672	0.0286	*
Soil	Residuals	31	1.279	0.0413	0.371			
Soil	Total	38	3.444			1		
Rhizosphere	Treatment	1	1.8962	1.8962	31.429	0.3164	<0.001	***
Rhizosphere	Time point	3	1.2491	0.4164	6.901	0.2084	<0.001	***
Rhizosphere	TRT : Time point	3	0.4942	0.1647	2.731	0.0825	<0.001	***
Rhizosphere	Residuals	39	2.353	0.0603	0.393			
Rhizosphere	Total	46	5.9926			1		
Root	Treatment	1	3.4715	3.4715	45.377	0.3924	<0.001	***
Root	Time point	3	1.6196	0.5399	7.057	0.1831	<0.001	***
Root	TRT : Time point	3	0.6952	0.2317	3.029	0.0786	0.0014	**
Root	Residuals	40	3.0601	0.0765	0.346			
Root	Total	47	8.8464			1		

Table 4-6

Table 4- 6. CAPS on testing the factors explaining the variation in samples

a. Table displaying results of a canonical analysis of principal coordinates (CAPS) for all samples testing the hypothesis that sample type, treatment, time point, and replicate are significant factors explaining the variation between all samples in either Bray-Curtis or weighted UniFrac distances. **b.** Table displaying results of CAPS done for samples in each sample type (soil, rhizosphere, and root) individually, to test the hypothesis that treatment, time point, and replicate are significant factors in explaining the variation between samples in either Bray-Curtis or weighted UniFrac distances. 'DF' = degrees of freedom, and significant factors indicated by ***, denoting p value of less than .001.

a. Results from CAPS using Bray-Curtis and UniFrac distances to determine significant factors explaining variance across bacterial communities.

Factors	Distance Metric	DF	Sum of Squares	F Statistics	p-value	
Sample Type	UniFrac	2	2.8844	108.4102	0.001	***
Treatment	UniFrac	1	0.5436	40.8648	0.001	***
Time point	UniFrac	3	0.2954	7.4008	0.001	***
Replicate	UniFrac	5	0.0652	0.9804	0.447	
Residual	UniFrac	122	1.623			
Sample Type	Bray-Curtis	2	13.5589	68.5415	0.001	
Treatment	Bray-Curtis	1	4.0932	41.3828	0.001	***
Time point	Bray-Curtis	3	1.6578	5.5868	0.001	***
Replicate	Bray-Curtis	5	0.465	0.9402	0.548	***
Residual	Bray-Curtis	122	12.067			

b. Results from CAPS using Bray-Curtis and UniFrac distances done separately for each sample type.

Sample Type	Distance Metric	Factors	DF	Sum of Squares	F Statistic	p-value	% Variance
Soil	UniFrac	Treatment	1	0.1171	34.8504	0.001	0.43
Soil	UniFrac	Time point	3	0.0443	4.3922	0.001	0.16
Soil	UniFrac	Replicate	5	0.0166	0.9847	0.487	0.06
Soil	UniFrac	Residual	29	0.0975	0.2755	0.354	
Soil	Bray	Treatment	1	1.4555	32.6404	0.001	0.42
Soil	Bray	Time point	3	0.4781	3.5739	0.001	0.14
Soil	Bray	Replicate	5	0.2173	0.9746	0.453	0.06
Soil	Bray	Residual	29	1.2931	3.444	0.375	
Rhizosphere	UniFrac	Treatment	1	0.1971	24.2593	0.001	0.26

Rhizosphere	UniFrac	Time point	3	0.2171	8.9069	0.001	0.29
Rhizosphere	UniFrac	Replicate	5	0.0322	0.793	0.759	0.04
Rhizosphere	UniFrac	Residual	37	0.3007	0.7472	0.402	
Rhizosphere	Bray	Treatment	1	1.8962	27.6167	0.001	0.32
Rhizosphere	Bray	Time point	3	1.2491	6.064	0.001	0.21
Rhizosphere	Bray	Replicate	5	0.3068	0.8936	0.635	0.05
Rhizosphere	Bray	Residual	37	2.5405	5.9926	0.424	
Root	UniFrac	Treatment	1	0.5103	33.942	0.001	0.34
Root	UniFrac	Time point	3	0.3451	7.6526	0.001	0.23
Root	UniFrac	Replicate	5	0.0779	1.0361	0.419	0.05
Root	UniFrac	Residual	38	0.5713	1.5045	0.38	
Root	Bray	Treatment	1	3.4715	39.2592	0.001	0.39
Root	Bray	Time point	3	1.6196	6.1052	0.001	0.18
Root	Bray	Replicate	5	0.3951	0.8937	0.613	0.04
Root	Bray	Residual	38	3.3602	8.8464	0.38	

Table 4-7

Table 4- 7. OTUs enriched in treated root samples

Several members of the kingdom Bacteria, phylum Proteobacteria, and class Betaproteobacteria that are more enriched in treated root samples have potential functions of sulfur cycling, denitrification, and nitrogen fixation. Summary of read counts and potential activities of Betaproteobacteria genera that are enriched in root samples with amendment treatment with potential function predicted based on a literature search. 'RC' = number of read counts.

OTU Rank	Order	Family	Genus	p-value	Control RC	Treated RC	Potential Function
13	*Burkholderiales*	*Comamonadaceae*	*Acidovorax*	0.001	2261	16523	Sulfur cycling
11	*Burkholderiales*	*Burkholderiales*	*unclassified*	0.003	1007	14086	.
15	*Methylophilales*	*Methylophilaceae*	*Methylotenera*	0.008	4930	8628	De- nitrification
20	*Methylophilales*	*Methylophilaceae*	*Methylotenera*	0.001	878	8547	De- nitrification
22	*Burkholderiales*	*Comamonadaceae*	*Comamonadaceae*	0.001	768	3966	Sulfur cycling
55	*Methylophilales*	*Methylophilaceae*	*Methylophilaceae*	0.001	569	2454	De- nitrification
158	*Burkholderiales*	*Comamonadaceae*	*Ramlibacter*	0.001	20	232	Nitrogen fixation
465	*Methylophilales*	*Methylophilaceae*	*Methylophilaceae*	0.001	8	126	De- nitrification
479	*Burkholderiales*	*Comamonadaceae*	*Comamonadaceae*	0.001	1	103	Sulfur cycling
261	*Rhodocyclales*	*Rhodocyclales*	*Rhodocyclales*	0.008	16	79	Nitrogen fixation
96	*Rhodocyclales*	*Rhodocyclales*	*Rhodocyclales*	0.007	18	70	Nitrogen fixation
1054	*Rhodocyclales*	*Rhodocyclaceae*	*Methyloversatilis*	0.001	2	55	.
1703	*Burkholderiales*	*Burkholderiaceae*	*Chitinimonas*	0.003	0	24	.
233	*Neisseriales*	*Neisseriaceae*	*Vogesella*	0.004	0	20	.
412	*Burkholderiales*	*Comamonadaceae*	*Comamonadaceae*	0.007	1	16	Sulfur cycling

Table 4-8

Table 4- 8. Nutrient analysis of the product VESTA

The nutrient profile of the liquid product VESTA assessed by D&D Agricultural Laboratory, Inc. (Fresno, CA) by means of a pH analyzer, electrical conductivity meter for soluble salts, FP-528 (Leco Corporation, St. Joseph, MI) for total nitrogen, and Optima 8000 ICP-OES (PerkinElmer, Inc., Waltham, MA) for all other parameters.

Parameter	Amount
pH	6.89
Soluble Salts (Electrical Conductivity)	0.61 dS/m
Total Nitrogen	0.00%
Phosphorus (P)	0.017
Diphosphorus Pentoxide (P_2O_5)	0.04%
Potassium (K)	0.10%
Potassium Oxide (K_2O)	0.12%
Calcium Carbonate ($CaCO_3$)	0.01%
Magnesium Carbonate ($MgCO_3$)	0.00%
Zinc	0.49 ppm
Manganese	1.05 ppm
Iron	172.0 ppm
Copper	0.84 ppm
Boron	3.17 ppm
Sodium	0.00%
Chloride	0.00%

Reference

Agler, M. T. *et al.* (2016) 'Microbial Hub Taxa Link Host and Abiotic Factors to Plant Microbiome Variation', *PLoS biology*, 14(1), p. e1002352.

Aleklett, K. *et al.* (2015) 'Wild plant species growing closely connected in a subalpine meadow host distinct root-associated bacterial communities', *PeerJ*, 3, p. e804.

Alexandratos, N. and Bruinsma, J. (2012) 'World agriculture towards 2030/2050: the 2012 revision'. large.stanford.edu. Available at: http://large.stanford.edu/courses/2014/ph240/yuan2/docs/ap106e.pdf.

Ambrosini, A., de Souza, R. and Passaglia, L. M. P. (2016) 'Ecological role of bacterial inoculants and their potential impact on soil microbial diversity', *Plant and soil*, 400(1), pp. 193–207.

And, D. H. and Keel, C. (2003) 'REGULATION OF ANTIBIOTIC PRODUCTION IN ROOT-COLONIZING PSEUDOMONAS SPP. AND RELEVANCE FOR BIOLOGICAL CONTROL OF PLANT DISEASE'. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA. doi: 10.1146/annurev.phyto.41.052002.095656.

Anderson, D. M., Glibert, P. M. and Burkholder, J. M. (2002) 'Harmful algal blooms and eutrophication: Nutrient sources, composition, and consequences', *Estuaries*. Springer-Verlag, 25(4), pp. 704–726.

Angus, A. A. *et al.* (2014) 'Plant-associated symbiotic Burkholderia species lack hallmark strategies required in mammalian pathogenesis', *PloS one*, 9(1), p. e83779.

Arzanesh, M. H. *et al.* (2011) 'Wheat (Triticum aestivum L.) growth enhancement by Azospirillum sp. under drought stress', *World journal of microbiology & biotechnology*. Springer Netherlands, 27(2), pp. 197–205.

Ashelford, K. E. *et al.* (2005) 'At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies', *Applied and environmental microbiology*, 71(12), pp. 7724–7736.

Atkins, C. A. (1974) 'Occurrence and some properties of carbonic anhydrases from legume root nodules', *Phytochemistry*. Elsevier, 13(1), pp. 93–98.

Atwell, S. *et al.* (2010) 'Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines', *Nature*, 465(7298), pp. 627–631.

Badri, D. V. *et al.* (2013) 'Application of natural blends of phytochemicals derived from the root exudates of Arabidopsis to the soil reveal that phenolic-related compounds predominantly modulate the soil microbiome', *The Journal of biological chemistry*,

288(7), pp. 4502–4512.

Baerson, S. R. *et al.* (2008) 'A functional genomics investigation of allelochemical biosynthesis in Sorghum bicolor root hairs', *The Journal of biological chemistry*, 283(6), pp. 3231–3247.

Baker, G. C., Smith, J. J. and Cowan, D. A. (2003) 'Review and re-analysis of domain-specific 16S primers', *Journal of microbiological methods*, 55(3), pp. 541–555.

Benson, A. K. *et al.* (2010) 'Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors', *Proceedings of the National Academy of Sciences of the United States of America*, 107(44), pp. 18933–18938.

Berendsen, R. L., Pieterse, C. M. J. and Bakker, P. A. H. M. (2012) 'The rhizosphere microbiome and plant health', *Trends in plant science*, 17(8), pp. 478–486.

Bergelson, J., Mittelstrass, J. and Horton, M. W. (2019) 'Characterizing both bacteria and fungi improves understanding of the Arabidopsis root microbiome', *Scientific reports*, 9(1), p. 24.

Berg, G. (2009) 'Plant-microbe interactions promoting plant growth and health: perspectives for controlled use of microorganisms in agriculture', *Applied microbiology and biotechnology*, 84(1), pp. 11–18.

Bergmann, G. T. *et al.* (2011) 'The under-recognized dominance of Verrucomicrobia in soil bacterial communities', *Soil biology & biochemistry*. Elsevier, 43(7), pp. 1450–1455.

Bodenhausen, N., Horton, M. W. and Bergelson, J. (2013) 'Bacterial communities associated with the leaves and the roots of Arabidopsis thaliana', *PloS one*, 8(2), p. e56329.

Bogas, A. C. *et al.* (2015) 'Endophytic bacterial diversity in the phyllosphere of Amazon Paullinia cupana associated with asymptomatic and symptomatic anthracnose', *SpringerPlus*, 4, p. 258.

Bolyen, E. *et al.* (2019) 'Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2', *Nature biotechnology*, 37(8), pp. 852–857.

Borevitz, J. O. and Nordborg, M. (2003) 'The impact of genomics on the study of natural variation in Arabidopsis', *Plant physiology*, 132(2), pp. 718–725.

Bottini, R., Cassán, F. and Piccoli, P. (2004) 'Gibberellin production by bacteria and its involvement in plant growth promotion and yield increase', *Applied microbiology and biotechnology*, 65(5), pp. 497–503.

Brachi, B., Morris, G. P. and Borevitz, J. O. (2011) 'Genome-wide association studies in plants: the missing heritability is in the field', *Genome biology*, 12(10), p. 232.

Braun, A. L. and Supkoff, D. M. (1994) 'Options to methyl bromide for the control of soil-borne diseases and pests in California with reference to the Netherlands'. cdpr.ca.gov. Available at: http://www.cdpr.ca.gov/docs/emon/methbrom/soilsol.pdf.

Bresson, J. *et al.* (2013) 'The PGPR strain Phyllobacterium brassicacearum STM196 induces a reproductive delay and physiological changes that result in improved drought tolerance in Arabidopsis', *The New phytologist*. Wiley Online Library, 200(2), pp. 558–569.

Bulgarelli, D. *et al.* (2012) 'Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota', *Nature*, 488(7409), pp. 91–95.

Bulgarelli, D. *et al.* (2015) 'Structure and function of the bacterial root microbiota in wild and domesticated barley', *Cell host & microbe*, 17(3), pp. 392–403.

Burton, P., Clayton, D., Cardon, L. et al. (2007) 'Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls', *Nature*, 447(7145), pp. 661–678.

Carroll, C. *et al.* (2017) *Crop Disease and Agricultural Productivity*. Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w23513.

Casa, A. M. *et al.* (2008) 'Community Resources and Strategies for Association Mapping in Sorghum', *Crop science*. Madison, WI: Crop Science Society of America, 48, pp. 30–40.

Castrillo, G. *et al.* (2017) 'Root microbiota drive direct integration of phosphate stress and immunity', *Nature*. Macmillan Publishers Limited, part of Springer Nature. All rights reserved., 543, p. 513.

Castro-Sowinski, S. *et al.* (2007) 'Effects of inoculation with plant growth-promoting rhizobacteria on resident rhizosphere microorganisms', *FEMS microbiology letters*, 276(1), pp. 1–11.

Chakraborty, S. and Newton, A. C. (2011) 'Climate change, plant diseases and food security: an overview', *Plant pathology*. Blackwell Publishing Ltd, 60(1), pp. 2–14.

Chaparro, J. M. *et al.* (2013) 'Root exudation of phytochemicals in Arabidopsis follows specific patterns that are developmentally programmed and correlate with soil microbial functions', *PloS one*. journals.plos.org, 8(2), p. e55731.

Chopra, R. *et al.* (2017) 'Genome-wide association analysis of seedling traits in diverse Sorghum germplasm under thermal stress', *BMC plant biology*, 17(1), p. 12.

Claverie, J. *et al.* (2018) 'The Cell Wall-Derived Xyloglucan Is a New DAMP Triggering Plant Immunity in Vitis vinifera and Arabidopsis thaliana', *Frontiers in plant science*, 9, p. 1725.

Clemente, J. C. *et al.* (2012) 'The impact of the gut microbiota on human health: an

integrative view', *Cell*, 148(6), pp. 1258–1270.

Cohen, A. C., Bottini, R. and Piccoli, P. N. (2008) 'Azospirillum brasilense Sp 245 produces ABA in chemically-defined culture medium and increases ABA content in arabidopsis plants', *Plant growth regulation*, 54(2), pp. 97–103.

Colebrook, E. H. *et al.* (2014) 'The role of gibberellin signalling in plant responses to abiotic stress', *The Journal of experimental biology*, 217(Pt 1), pp. 67–75.

Coleman-Derr, D. *et al.* (2016) 'Plant compartment and biogeography affect microbiome composition in cultivated and native Agave species', *The New phytologist*, 209(2), pp. 798–811.

Coleman-Derr, D. and Tringe, S. G. (2014) 'Building the crops of tomorrow: advantages of symbiont-based approaches to improving abiotic stress tolerance', *Frontiers in microbiology*, 5, p. 283.

Copeland, J. K. *et al.* (2015) 'Seasonal community succession of the phyllosphere microbiome', *Molecular plant-microbe interactions: MPMI*, 28(3), pp. 274–285.

Covarrubias-Pazaran, G. (2016) 'Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer', *PloS one*, 11(6), p. e0156744.

Creus, C. M. *et al.* (2005) 'Nitric oxide is involved in the Azospirillum brasilense-induced lateral root formation in tomato', *Planta*, 221(2), pp. 297–303.

Cruaud, P. *et al.* (2014) 'Influence of DNA extraction method, 16S rRNA targeted hypervariable regions, and sample origin on microbial diversity detected by 454 pyrosequencing in marine chemosynthetic ecosystems', *Applied and environmental microbiology*, 80(15), pp. 4626–4639.

Cuevas, H. E. *et al.* (2019) 'Genome-Wide Association Mapping of Grain Mold Resistance in the US Sorghum Association Panel', *The plant genome*, 12(2). doi: 10.3835/plantgenome2018.09.0070.

Daugaard, H. and Lindhard, H. (2000) 'Strawberry cultivars for organic production', *cabdirect.org*. Available at: https://www.cabdirect.org/cabdirect/abstract/20003033695.

Davenport, E. R. *et al.* (2015) 'Genome-Wide Association Studies of the Human Gut Microbiota', *PloS one*. Public Library of Science, 10(11), p. e0140301.

De Caceres, M., Jansen, F. and De Caceres, M. M. (2016) 'Package "indicspecies"'. Available at: ftp://r-project.org/pub/R/web/packages/indicspecies/indicspecies.pdf.

Deredjian, A. *et al.* (2014) 'Low occurrence of Pseudomonas aeruginosa in agricultural soils with and without organic amendment', *Frontiers in cellular and infection microbiology*, 4, p. 53.

DeSantis, T. Z. *et al.* (2006) 'Greengenes, a chimera-checked 16S rRNA gene database

and workbench compatible with ARB', *Applied and environmental microbiology*, 72(7), pp. 5069–5072.

Desgarennes, D. *et al.* (2014) 'Diazotrophic potential among bacterial communities associated with wild and cultivated Agave species', *FEMS microbiology ecology*, 90(3), pp. 844–857.

De Tender, C. *et al.* (2016) 'Dynamics in the Strawberry Rhizosphere Microbiome in Response to Biochar and Botrytis cinerea Leaf Infection', *Frontiers in microbiology*, 7, p. 2062.

DiMario, R. J. *et al.* (2017) 'Plant Carbonic Anhydrases: Structures, Locations, Evolution, and Physiological Roles', *Molecular plant*, 10(1), pp. 30–46.

Dos Reis, S. P., Lima, A. M. and de Souza, C. R. B. (2012) 'Recent molecular advances on downstream plant responses to abiotic stress', *International journal of molecular sciences*, 13(7), pp. 8628–8647.

Duniway, J. M. (2002) 'Status of chemical alternatives to methyl bromide for pre-plant fumigation of soil', *Am Phytopath Society*. Available at: http://apsjournals.apsnet.org/doi/abs/10.1094/PHYTO.2002.92.12.1337.

Eberl, L. and Vandamme, P. (2016) 'Members of the genus Burkholderia: good and bad guys', *F1000Research*, 5. doi: 10.12688/f1000research.8221.1.

Edgar, R. (2010) 'Usearch'. Available at: https://www.osti.gov/scitech/servlets/purl/1137186 (Accessed: 17 May 2018).

Edgar, R. C. (2004) 'MUSCLE: multiple sequence alignment with high accuracy and high throughput', *Nucleic acids research*, 32(5), pp. 1792–1797.

Edgar, R. C. (2013) 'UPARSE: highly accurate OTU sequences from microbial amplicon reads', *Nature methods*, 10(10), pp. 996–998.

Edwards, J. *et al.* (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(8), pp. E911–20.

Edwards, J. A. *et al.* (2018) 'Compositional shifts in root-associated bacterial and archaeal microbiota track the plant life cycle in field-grown rice', *PLoS biology*, 16(2), p. e2003862.

van Elsas, J. D. *et al.* (2012) 'Microbial diversity determines the invasion of soil by a bacterial pathogen', *Proceedings of the National Academy of Sciences of the United States of America*, 109(4), pp. 1159–1164.

Erbilgin, O., McDonald, K. L. and Kerfeld, C. A. (2014) 'Characterization of a planctomycetal organelle: a novel bacterial microcompartment for the aerobic degradation of plant saccharides', *Applied and environmental microbiology*, 80(7), pp.

2193–2205.

Fadrosh, D. W. *et al.* (2014) ‘An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform’, *Microbiome*, 2(1), p. 6.

Falkow, S. *et al.* (2006) *The Prokaryotes: Vol. 5: Proteobacteria: Alpha and Beta Subclasses*. Springer Science & Business Media.

Fierer, N. (2017) ‘Embracing the unknown: disentangling the complexities of the soil microbiome’, *Nature reviews. Microbiology*, 15(10), pp. 579–590.

Fierer, N. and Lennon, J. T. (2011) ‘The generation and maintenance of diversity in microbial communities’, *American journal of botany*, 98(3), pp. 439–448.

Fitzpatrick, C. R. *et al.* (2018) ‘Assembly and ecological function of the root microbiome across angiosperm plant species’, *Proceedings of the National Academy of Sciences of the United States of America*, 115(6), pp. E1157–E1165.

Floryszak-Wieczorek, J. and Arasimowicz-Jelonek, M. (2017) ‘The multifunctional face of plant carbonic anhydrase’, *Plant physiology and biochemistry: PPB / Societe francaise de physiologie vegetale*, 112, pp. 362–368.

Flowers, T. J. and Yeo, A. R. (1995) ‘Breeding for Salinity Resistance in Crop Plants: Where Next?’, *Functional plant biology: FPB*. CSIRO PUBLISHING, 22(6), pp. 875–884.

Fonseca-García, C. *et al.* (2016) ‘The Cacti Microbiome: Interplay between Habitat-Filtering and Host-Specificity’, *Frontiers in microbiology*, 7, p. 150.

Garbeva, P., van Veen, J. A. and van Elsas, J. D. (2004) ‘MICROBIAL DIVERSITY IN SOIL: Selection of Microbial Populations by Plant and Soil Type and Implications for Disease Suppressiveness’, *Annual review of phytopathology*. Annual Reviews, 42(1), pp. 243–270.

German, M. A. *et al.* (2000) ‘Effects of Azospirillum brasilense on root morphology of common bean (Phaseolus vulgaris L.) under different water regimes’, *Biology and fertility of soils*, 32(3), pp. 259–264.

Ghyselinck, J. *et al.* (2013) ‘The effect of primer choice and short read sequences on the outcome of 16S rRNA gene based diversity studies’, *PloS one*, 8(8), p. e71360.

Gilbert, N. (2009) ‘Environment: the disappearing nutrient’, *nature.com*. Available at: https://www.nature.com/articles/461716a.

Glazier, A. M., Nadeau, J. H. and Aitman, T. J. (2002) ‘Finding genes that underlie complex traits’, *Science*, 298(5602), pp. 2345–2349.

Glick, B. R. (2005) ‘Modulation of plant ethylene levels by the bacterial enzyme ACC deaminase’, *FEMS microbiology letters*, 251(1), pp. 1–7.

Glick, B. R. (2012) ‘Plant growth-promoting bacteria: mechanisms and applications’, *Scientifica*, 2012, p. 963401.

Goldfarb, K. C. *et al.* (2011) ‘Differential growth responses of soil bacterial taxa to carbon substrates of varying chemical recalcitrance’, *Frontiers in microbiology*, 2, p. 94.

Gontia-Mishra, I. *et al.* (2017) ‘Molecular diversity of 1-aminocyclopropane-1-carboxylate (ACC) deaminase producing PGPR from wheat (Triticum aestivum L.) rhizosphere’, *Plant and Soil*, pp. 213–227. doi: 10.1007/s11104-016-3119-3.

Goodstein, D. M. *et al.* (2012) ‘Phytozome: a comparative platform for green plant genomics’, *Nucleic acids research*, 40(Database issue), pp. D1178–86.

Haichar, F. el Z. *et al.* (2008) ‘Plant host habitat and root exudates shape soil bacterial community structure’, *The ISME journal*, 2(12), pp. 1221–1230.

Han, B. and Huang, X. (2013) ‘Sequencing-based genome-wide association study in rice’, *Current opinion in plant biology*, 16(2), pp. 133–138.

Haney, C. H. *et al.* (2015) ‘Associations with rhizosphere bacteria can confer an adaptive advantage to plants’, *Nature plants*, 1(6). doi: 10.1038/nplants.2015.51.

Harman, G. E. *et al.* (2004) ‘Trichoderma species--opportunistic, avirulent plant symbionts’, *Nature reviews. Microbiology*, 2(1), pp. 43–56.

Harman, G. E. (2006) ‘Overview of Mechanisms and Uses of Trichoderma spp’, *Phytopathology*, 96(2), pp. 190–194.

Havlin, J. L. *et al.* (2005) *Soil fertility and fertilizers: An introduction to nutrient management*. Pearson Prentice Hall Upper Saddle River, NJ.

van der Heijden, M. G. A., Bardgett, R. D. and van Straalen, N. M. (2008) ‘The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems’, *Ecology letters*, 11(3), pp. 296–310.

Herschkovitz, Y. *et al.* (2005) ‘Azospirillum brasilense does not affect population structure of specific rhizobacterial communities of inoculated maize (Zea mays)’, *Environmental microbiology*, 7(11), pp. 1847–1852.

Hiruma, K. *et al.* (2016) ‘Root Endophyte Colletotrichum tofieldiae Confers Plant Fitness Benefits that Are Phosphate Status Dependent’, *Cell*. Elsevier, 165(2), pp. 464–474.

Hiscox, J. *et al.* (2015) ‘Priority effects during fungal community establishment in beech wood’, *The ISME journal*, 9(10), pp. 2246–2260.

Hol, W. H. G. *et al.* (2015) ‘Non-random species loss in bacterial communities reduces antifungal volatile production’, *Ecology*, 96(8), pp. 2042–2048.

Horton, M. W. *et al.* (2014) ‘Genome-wide association study of Arabidopsis thaliana leaf microbial community’, *Nature communications*, 5, p. 5320.

Hou, S. *et al.* (2019) 'Damage-Associated Molecular Pattern-Triggered Immunity in Plants', *Frontiers in plant science*, 10, p. 646.

Huang, J. *et al.* (2013) 'Genome-wide Association Analysis of Ten Chilling Tolerance Indices at the Germination and Seedling Stages in Maize', *Journal of Integrative Plant Biology*, pp. 735–744. doi: 10.1111/jipb.12051.

Huang, X. *et al.* (2011) 'Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm', *Nature genetics*, 44(1), pp. 32–39.

Huang, X. and Han, B. (2014) 'Natural variations and genome-wide association studies in crop plants', *Annual review of plant biology*, 65, pp. 531–551.

Hwang, E.-Y. *et al.* (2014) 'A genome-wide association study of seed protein and oil content in soybean', *BMC genomics*, 15, p. 1.

Jaiswal, A. K. *et al.* (2017) 'Linking the Belowground Microbial Composition, Diversity and Activity to Soilborne Disease Suppression and Growth Promotion of Tomato Amended with Biochar', *Scientific reports*, 7, p. 44382.

Jalili, F. *et al.* (2009) 'Isolation and characterization of ACC deaminase-producing fluorescent pseudomonads, to alleviate salinity stress on canola (Brassica napus L.) growth', *Journal of plant physiology*, 166(6), pp. 667–674.

Kalyuhznaya, M. G. *et al.* (2009) 'Methylophilaceae link methanol oxidation to denitrification in freshwater lake sediment as suggested by stable isotope probing and pure culture analysis', *Environmental microbiology reports*, 1(5), pp. 385–392.

Kamada, N. *et al.* (2013) 'Role of the gut microbiota in immunity and inflammatory disease', *Nature reviews. Immunology*, 13(5), pp. 321–335.

Kang, S.-M. *et al.* (2014) 'Gibberellin secreting rhizobacterium, Pseudomonas putida H-2-3 modulates the hormonal and stress physiology of soybean to improve the plant growth under saline and drought conditions', *Plant physiology and biochemistry: PPB / Societe francaise de physiologie vegetale*, 84, pp. 115–124.

Keeney, D. R. and Nelson, D. W. (1982) 'Nitrogen—inorganic forms', *Methods of soil analysis. Part 2*. dl.sciencesocieties.org. Available at: https://dl.sciencesocieties.org/publications/books/abstracts/agronomymonogra/methodsofsoilan2/643.

Khan, G. A. *et al.* (2016) 'Phosphate Deficiency Induces the Jasmonate Pathway and Enhances Resistance to Insect Herbivory', *Plant physiology*, 171(1), pp. 632–644.

Khare, E. and Arora, N. K. (2015) 'Effects of Soil Environment on Field Efficacy of Microbial Inoculants', in Arora, N. K. (ed.) *Plant Microbes Symbiosis: Applied Facets*. New Delhi: Springer India, pp. 353–381.

Kim, K.-H., Kabir, E. and Jahan, S. A. (2017) 'Exposure to pesticides and the associated human health effects', *The Science of the total environment*, 575, pp. 525–535.

Kim, Y.-C. *et al.* (2012) 'Enhancement of Plant Drought Tolerance by Microbes', *Plant Responses to Drought Stress*, pp. 383–413. doi: 10.1007/978-3-642-32653-0_15.

Kolton, M. *et al.* (2017) 'Biochar-stimulated plant performance is strongly linked to microbial diversity and metabolic potential in the rhizosphere', *Wiley Online Library*. Available at: http://onlinelibrary.wiley.com/doi/10.1111/nph.14253/full.

Korte, A. and Farlow, A. (2013) 'The advantages and limitations of trait analysis with GWAS: a review', *Plant methods*, 9(1), p. 29.

Koyama, A. *et al.* (2014) 'Soil bacterial community composition altered by increased nutrient availability in Arctic tundra soils', *Frontiers in microbiology*, 5, p. 516.

Kozich, J. J. *et al.* (2013) 'Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform', *Applied and environmental microbiology*, 79(17), pp. 5112–5120.

Krajmalnik-Brown, R. and Ilhan, Z. E. (2012) 'Effects of gut microbes on nutrient absorption and energy regulation', *Nutrition in Clinical*. Wiley Online Library. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1177/0884533611436116?casa_token=Za5bF8kLJ7AAAAAA:kok0AQj3n1Q8lIKMPgCr9pniyFNdoTB4KE0dvRTKHJ6RzE3EPr6Y1Og93vmdzGfkBZuFn2diMV8XMQ.

Kröber, M. *et al.* (2014) 'Effect of the strain Bacillus amyloliquefaciens FZB42 on the microbial community in the rhizosphere of lettuce under field conditions analyzed by whole metagenome sequencing', *Frontiers in microbiology*, 5, p. 252.

Kumar, A. and Verma, J. P. (2018) 'Does plant—Microbe interaction confer stress tolerance in plants: A review?', *Microbiological research*, 207, pp. 41–52.

Kyselková, M. *et al.* (2009) 'Comparison of rhizobacterial community composition in soil suppressive or conducive to tobacco black root rot disease', *The ISME journal*, 3(10), pp. 1127–1138.

Lambers, H., Stuart Chapin, F. and Pons, T. L. (2008) 'Plant Water Relations', in *Plant Physiological Ecology*. Springer, New York, NY, pp. 163–223.

Lebeis, S. L. *et al.* (2015) 'PLANT MICROBIOME. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa', *Science*, 349(6250), pp. 860–864.

Letunic, I. and Bork, P. (2019) 'Interactive Tree Of Life (iTOL) v4: recent updates and new developments', *Nucleic acids research*, 47(W1), pp. W256–W259.

Li, F. *et al.* (2014) 'Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (Brassica napus L.)', *DNA research: an international journal for rapid publication of reports on genes and genomes*, 21(4), pp. 355–367.

Lima, M. de F. *et al.* (2017) 'Molecular mechanisms of biomass increase in plants', *Biotechnology Research and Innovation*, 1(1), pp. 14–25.

Linkies, A. *et al.* (2010) 'The evolution of seeds', *The New phytologist*, 186(4), pp. 817–831.

Lin, Y. R., Schertz, K. F. and Paterson, A. H. (1995) 'Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population', *Genetics*, 141(1), pp. 391–411.

Liu, F. *et al.* (2013) 'Cytokinin-producing, plant growth-promoting rhizobacteria that confer resistance to drought stress in Platycladus orientalis container seedlings', *Applied microbiology and biotechnology*, 97(20), pp. 9155–9164.

Loy, A. *et al.* (2005) '16S rRNA Gene-Based Oligonucleotide Microarray for Environmental Monitoring of the Betaproteobacterial Order "Rhodocyclales"', *Applied and environmental microbiology*, 71(3), pp. 1373–1386.

Lundberg, D. S. *et al.* (2012) 'Defining the core Arabidopsis thaliana root microbiome', *Nature*, 488(7409), pp. 86–90.

Lundberg, D. S. *et al.* (2013) 'Practical innovations for high-throughput amplicon sequencing', *Nature methods*, 10(10), pp. 999–1002.

Lu, Y. *et al.* (2006) 'Structure and activity of bacterial community inhabiting rice roots and the rhizosphere', *Environmental microbiology*, 8(8), pp. 1351–1360.

Lynch, M., Walsh, B. and Others (1998) *Genetics and analysis of quantitative traits*. Sinauer Sunderland, MA.

Magoč, T. and Salzberg, S. L. (2011) 'FLASH: fast length adjustment of short reads to improve genome assemblies', *Bioinformatics* , 27(21), pp. 2957–2963.

Mannaa, M., Park, I. and Seo, Y.-S. (2018) 'Genomic Features and Insights into the Taxonomy, Virulence, and Benevolence of Plant-Associated Burkholderia Species', *International journal of molecular sciences*, 20(1). doi: 10.3390/ijms20010121.

Marek-Kozaczuk, M. and Skorupska, A. (2001) 'Production of B-group vitamins by plant growth-promoting Pseudomonas fluorescens strain 267 and the importance of vitamins in the colonization and nodulation of red clover', *Biology and fertility of soils*. Springer-Verlag, 33(2), pp. 146–151.

Martínez-Hidalgo, P. and Hirsch, A. M. (2017) 'The Nodule Microbiome: N2-Fixing Rhizobia Do Not Live Alone', *Phytobiomes Journal*. Scientific Societies, 1(2), pp. 70–

82.

Martin, F. N. and Bull, C. T. (2002) 'Biological approaches for control of root pathogens of strawberry', *Phytopathology*, 92(12), pp. 1356–1362.

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal*, 17(1), pp. 10–12.

Masson-Boivin, C. and Sachs, J. L. (2018) 'Symbiotic nitrogen fixation by rhizobia—the roots of a success story', *Current opinion in plant biology*, 44, pp. 7–15.

Matos, A., Kerkhof, L. and Garland, J. L. (2005) 'Effects of microbial community diversity on the survival of Pseudomonas aeruginosa in the wheat rhizosphere', *Microbial ecology*, 49(2), pp. 257–264.

McCarthy, M. I. *et al.* (2008) 'Genome-wide association studies for complex traits: consensus, uncertainty and challenges', *Nature reviews. Genetics*, 9(5), pp. 356–369.

McMurdie, P. J. and Holmes, S. (2013) 'phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data', *PloS one*, 8(4), p. e61217.

Meents, M. J., Watanabe, Y. and Samuels, A. L. (2018) 'The cell biology of secondary cell wall biosynthesis', *Annals of botany*, 121(6), pp. 1107–1125.

Meijón, M. *et al.* (2014) 'Genome-wide association study using cellular traits identifies a new regulator of root development in Arabidopsis', *Nature genetics*, 46(1), pp. 77–81.

Mendes, R. *et al.* (2011) 'Deciphering the rhizosphere microbiome for disease-suppressive bacteria', *Science*, 332(6033), pp. 1097–1100.

Mendes, R., Garbeva, P. and Raaijmakers, J. M. (2013) 'The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms', *FEMS microbiology reviews*, 37(5), pp. 634–663.

Meyer, F. *et al.* (2008) 'The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes', *bmcbioinformatics.biomedcentral*. Available at: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-386.

Miller, R. (1997) 'Microwave Digestion Of Plant Tissue In A Closed Vessel', in *Handbook of Reference Methods for Plant Analysis*.

Millner, P. D., Ringer, C. E. and Maas, J. L. (2004) 'Suppression of Strawberry Root Disease with Animal Manure Composts', *Compost science & utilization*. Taylor & Francis, 12(4), pp. 298–307.

Moe, L. A. (2013) 'Amino acids in the rhizosphere: from plants to microbes', *American journal of botany*, 100(9), pp. 1692–1705.

Molina-Favero, C. *et al.* (2008) 'Aerobic nitric oxide production by Azospirillum

brasilense Sp245 and its influence on root architecture in tomato', *Molecular plant-microbe interactions: MPMI*, 21(7), pp. 1001–1009.

Morrell, P. L., Buckler, E. S. and Ross-Ibarra, J. (2011) 'Crop genomics: advances and applications', *Nature reviews. Genetics*, 13(2), pp. 85–96.

Morris, G. P., Rhodes, D. H., *et al.* (2013) 'Dissecting Genome-Wide Association Signals for Loss-of-Function Phenotypes in Sorghum Flavonoid Pigmentation Traits', *G3: Genes|Genomes|Genetics*, pp. 2085–2094. doi: 10.1534/g3.113.008417.

Morris, G. P., Ramu, P., *et al.* (2013) 'Population genomic and genome-wide association studies of agroclimatic traits in sorghum', *Proceedings of the National Academy of Sciences of the United States of America*, 110(2), pp. 453–458.

Mustakhimov, I. *et al.* (2013) 'Insights into denitrification in Methylotenera mobilis from denitrification pathway and methanol metabolism mutants', *Journal of bacteriology*, 195(10), pp. 2207–2211.

Myers, S. S. *et al.* (2017) 'Climate Change and Global Food Systems: Potential Impacts on Food Security and Undernutrition', *Annual review of public health*, 38, pp. 259–277.

Nasholm, T., Huss-Danell, K. and Hogberg, P. (2001) 'Uptake of glycine by field grown wheat', *The New phytologist*, 150(1), pp. 59–63.

Nawaz, H. *et al.* (2016) 'Exogenous Application of Growth Enhancers Mitigate Water Stress in Wheat by Antioxidant Elevation', *Frontiers in plant science*, 7, p. 597.

Naylor, D. *et al.* (2017) 'Drought and host selection influence bacterial community dynamics in the grass root microbiome', *The ISME journal*, 11(12), pp. 2691–2704.

O'Brien, S. L. *et al.* (2016) 'Spatial scale drives patterns in soil bacterial diversity', *Environmental microbiology*, 18(6), pp. 2039–2051.

Oerke, E.-C. (2006) 'Crop losses to pests', *cambridge.org*. Available at: https://www.cambridge.org/core/journals/journal-of-agricultural-science/article/crop-losses-to-pests/AD61661AD6D503577B3E73F2787FE7B2.

Ogura, T. and Busch, W. (2015) 'From phenotypes to causal sequences: using genome wide association studies to dissect the sequence basis for variation of plant development', *Current opinion in plant biology*, 23, pp. 98–108.

Oksanen, J. *et al.* (2016) 'Vegan: community ecology package. software'.

O'Leary, N. A. *et al.* (2016) 'Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation', *Nucleic acids research*, 44(D1), pp. D733–45.

Onaga, G. and Wydra, K. (2016) 'Advances in Plant Tolerance to Abiotic Stresses', in Abdurakhmonov, I. Y. (ed.) *Plant Genomics*. InTech.

Overbeek, R. *et al.* (2014) 'The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST)', *Nucleic acids research*, 42(Database issue), pp. D206–14.

Owen, A. G. and Jones, D. L. (2001) 'Competition for amino acids between wheat roots and rhizosphere microorganisms and the role of amino acids in plant N acquisition', *Soil biology & biochemistry*, 33(4), pp. 651–657.

Paradis, E., Claude, J. and Strimmer, K. (2004) 'APE: Analyses of Phylogenetics and Evolution in R language', *Bioinformatics* , 20(2), pp. 289–290.

Parisi, G. *et al.* (2004) 'Gamma carbonic anhydrases in plant mitochondria', *Plant molecular biology*, 55(2), pp. 193–207.

Pasam, R. K. *et al.* (2012) 'Genome-wide association studies for agronomical traits in a world wide spring barley collection', *BMC plant biology*, 12, p. 16.

Pasaniuc, B. *et al.* (2012) 'Extremely low-coverage sequencing and imputation increases power for genome-wide association studies', *Nature genetics*, 44(6), pp. 631–635.

Paulson, J. N. *et al.* (2013) 'Differential abundance analysis for microbial marker-gene surveys', *Nature Methods*, pp. 1200–1202. doi: 10.1038/nmeth.2658.

Peiffer, J. A. *et al.* (2013) 'Diversity and heritability of the maize rhizosphere microbiome under field conditions', *Proceedings of the National Academy of Sciences of the United States of America*, 110(16), pp. 6548–6553.

Pieterse, C. M. J. *et al.* (2014) 'Induced systemic resistance by beneficial microbes', *Annual review of phytopathology*, 52, pp. 347–375.

Pii, Y. *et al.* (2015) 'Microbial interactions in the rhizosphere: beneficial influences of plant growth-promoting rhizobacteria on nutrient acquisition process. A review', *Biology and fertility of soils*, 51(4), pp. 403–415.

Pii, Y. *et al.* (2016) 'The interaction between iron nutrition, plant species and soil type shapes the rhizosphere microbiome', *Plant physiology and biochemistry: PPB / Societe francaise de physiologie vegetale*, 99, pp. 39–48.

Pimentel, D. (2005) 'Environmental and economic costs of the application of pesticides primarily in the United States', *Environment, Development and Sustainability*. Springer, 7(2), pp. 229–252.

Price, M. N., Dehal, P. S. and Arkin, A. P. (2010) 'FastTree 2--approximately maximum-likelihood trees for large alignments', *PloS one*, 5(3), p. e9490.

Quast, C. *et al.* (2013) 'The SILVA ribosomal RNA gene database project: improved data processing and web-based tools', *Nucleic acids research*, 41(Database issue), pp. D590–6.

Reinhold-Hurek, B. *et al.* (2015) 'Roots shaping their microbiome: global hotspots for microbial activity', *Annual review of phytopathology*, 53, pp. 403–424.

Restrepo, S. *et al.* (2005) 'Gene profiling of a compatible interaction between Phytophthora infestans and Solanum tuberosum suggests a role for carbonic anhydrase', *Molecular plant-microbe interactions: MPMI*, 18(9), pp. 913–922.

Rhoades, J. D. *et al.* (1989) 'Soil Electrical Conductivity and Soil Salinity: New Formulations and Calibrations', *Soil Science Society of America journal. Soil Science Society of America*. Madison, WI: Soil Science Society of America, 53, pp. 433–439.

Richter-Heitmann, T. *et al.* (2016) 'Evaluation of Strategies to Separate Root-Associated Microbial Communities: A Crucial Choice in Rhizobiome Research', *Frontiers in microbiology*, 7, p. 773.

Roberts, D. W. and Roberts, M. D. W. (2016) 'Package "labdsv"', *Ordination and Multivariate*. Available at: https://www.ohio.edu/plantbio/staff/mccarthy/multivariate/Package-labdsv.pdf.

Roman-reyna, V. *et al.* (2019) 'The Rice Leaf Microbiome Has a Conserved Community Structure Controlled by Complex Host-Microbe Interactions', *-HOST-MICROBE-D …*. doi: 10.2139/ssrn.3382544.

Sabarwal, A., Kumar, K. and Singh, R. P. (2018) 'Hazardous effects of chemical pesticides on human health-Cancer and other associated disorders', *Environmental toxicology and pharmacology*, 63, pp. 103–114.

Saleem, M. *et al.* (2018) 'Impact of root system architecture on rhizosphere and root microbiome', *Rhizosphere*, 6, pp. 47–51.

Saleem, M., Hu, J. and Jousset, A. (2019) 'More Than the Sum of Its Parts: Microbiome Biodiversity as a Driver of Plant Growth and Soil Health', *Annual review of ecology, evolution, and systematics*. Annual Reviews. doi: 10.1146/annurev-ecolsys-110617-062605.

Salter, S. J. *et al.* (2014) 'Reagent and laboratory contamination can critically impact sequence-based microbiome analyses', *BMC biology*, 12, p. 87.

Santos-Medellín, C. *et al.* (2017) 'Drought Stress Results in a Compartment-Specific Restructuring of the Rice Root-Associated Microbiomes', *mBio*, 8(4). doi: 10.1128/mBio.00764-17.

Saravanakumar, D. and Samiyappan, R. (2007) 'ACC deaminase from Pseudomonas fluorescens mediated saline resistance in groundnut (Arachis hypogea) plants', *Journal of applied microbiology*, 102(5), pp. 1283–1292.

Sasse, J., Martinoia, E. and Northen, T. (2018) 'Feed Your Friends: Do Plant Exudates Shape the Root Microbiome?', *Trends in plant science*, 23(1), pp. 25–41.

Sauvage, C. *et al.* (2014) 'Genome-Wide Association in Tomato Reveals 44 Candidate Loci for Fruit Metabolic Traits', *Plant physiology*, 165(3), pp. 1120–1132.

Savary, S. *et al.* (2012) 'Crop losses due to diseases and their implications for global food production losses and food security', *Food Security*, 4(4), pp. 519–537.

Savci, S. (2012) 'An agricultural pollutant: chemical fertilizer', *search.proquest.com*. Available at: http://search.proquest.com/openview/d0a1967effb187e43367e3f46ae630d1/1?pq-origsite=gscholar&cbl=2027401.

Schaefer, R. J. *et al.* (no date) 'Integrating co-expression networks with GWAS to prioritize causal genes in maize'. doi: 10.1101/221655.

Schlaeppi, K. *et al.* (2014) 'Quantitative divergence of the bacterial root microbiota in Arabidopsis thaliana relatives', *Proceedings of the National Academy of Sciences of the United States of America*, 111(2), pp. 585–592.

Schmalenberger, A. *et al.* (2008) 'The role of Variovorax and other Comamonadaceae in sulfur transformations by microbial wheat rhizosphere communities exposed to different sulfur fertilization regimes', *Environmental microbiology*, 10(6), pp. 1486–1500.

Schreiter, S. *et al.* (2014) 'Soil type-dependent effects of a potential biocontrol inoculant on indigenous bacterial communities in the rhizosphere of field-grown lettuce', *FEMS microbiology ecology*, 90(3), pp. 718–730.

Sessitsch, A. *et al.* (2005) 'Burkholderia phytofirmans sp. nov., a novel plant-associated bacterium with plant-beneficial properties', *International journal of systematic and evolutionary microbiology*, 55(Pt 3), pp. 1187–1192.

Shafi, J., Tian, H. and Ji, M. (2017) 'Bacillus species as versatile weapons for plant pathogens: a review', *Biotechnology, biotechnological equipment*. Taylor & Francis, 31(3), pp. 446–459.

Shakir, M. A., Bano, A. and Arshad, M. (2012) 'Short Communication&58; Rhizosphere bacteria containing ACC-deaminase conferred drought tolerance in wheat grown under semi-arid climate', *Soil & Environment*. ingentaconnect.com. Available at: https://www.ingentaconnect.com/content/doaj/20749546/2012/00000031/00000001/art00015.

Shang, S. and Yi, Y. (2015) 'A Greenhouse Assay on the Effect of Applied Urea Amount on the Rhizospheric Soil Bacterial Communities', *Indian journal of microbiology*, 55(4), pp. 406–414.

Shi, S. *et al.* (2016) 'The interconnected rhizosphere: High network complexity dominates rhizosphere assemblages', *Ecology letters*, 19(8), pp. 926–936.

Shu, L. *et al.* (2018) 'Symbiont location, host fitness, and possible coadaptation in a symbiosis between social amoebae and bacteria', *eLife*, 7. doi: 10.7554/eLife.42660.

Siles, J. A. *et al.* (2014) 'Microbial diversity of a Mediterranean soil and its changes after biotransformed dry olive residue amendment', *PloS one*, 9(7), p. e103035.

Simmons, T. *et al.* (2018) 'Exploring the Root Microbiome: Extracting Bacterial Community Data from the Soil, Rhizosphere, and Root Endosphere', *Journal of visualized experiments: JoVE*, (135). doi: 10.3791/57561.

Singh, B. K. *et al.* (2004) 'Unravelling rhizosphere–microbial interactions: opportunities and limitations', *Trends in microbiology*, 12(8), pp. 386–393.

Singh, D. P., Singh, H. B. and Prabha, R. (2016) *Microbial Inoculants in Sustainable Agricultural Productivity: Vol. 1: Research Perspectives*. Springer.

Siviter, H. *et al.* (2018) 'Quantifying the impact of pesticides on learning and memory in bees', *The Journal of applied ecology*, 55(6), pp. 2812–2821.

Slaymaker, D. H. *et al.* (2002) 'The tobacco salicylic acid-binding protein 3 (SABP3) is the chloroplast carbonic anhydrase, which exhibits antioxidant activity and plays a role in the hypersensitive defense response', *Proceedings of the National Academy of Sciences of the United States of America*, 99(18), pp. 11640–11645.

Slovak, R. *et al.* (2014) 'A Scalable Open-Source Pipeline for Large-Scale Root Phenotyping of Arabidopsis', *The Plant Cell*, pp. 2390–2403. doi: 10.1105/tpc.114.124032.

Soltani, A.-A. *et al.* (2010) 'Plant growth promoting characteristics in some Flavobacterium spp. isolated from soils of Iran', *ccsenet.org*. Available at: http://www.ccsenet.org/journal/index.php/jas/article/download/8404/6221.

Stehle, S. and Schulz, R. (2015) 'Agricultural insecticides threaten surface waters at the global scale', *Proceedings of the National Academy of Sciences of the United States of America*, 112(18), pp. 5750–5755.

Stewart, A. and Hill, R. (2014) 'Chapter 31 - Applications of Trichoderma in Plant Growth Promotion', in Gupta, V. K. et al. (eds) *Biotechnology and Biology of Trichoderma*. Amsterdam: Elsevier, pp. 415–428.

Strauss, S. L., Stover, J. K. and Kluepfel, D. A. (2015) 'Impact of biological amendments on Agrobacterium tumefaciens survival in soil', *Applied soil ecology: a section of Agriculture, Ecosystems & Environment*, 87(Supplement C), pp. 39–48.

Suckstorff, I. and Berg, G. (2003) 'Evidence for dose-dependent effects on plant growth by Stenotrophomonas strains from different origins', *Journal of applied microbiology*, 95(4), pp. 656–663.

Suh, D. H., Guan, Z. and Khachatryan, H. (2017) 'The impact of Mexican competition on the US strawberry industry', *wageningenacademic.com*. Available at: http://www.wageningenacademic.com/doi/abs/10.22434/IFAMR2016.0075.

Sunderhaus, S. *et al.* (2006) 'Carbonic anhydrase subunits form a matrix-exposed domain attached to the membrane arm of mitochondrial complex I in plants', *The Journal of biological chemistry*, 281(10), pp. 6482–6488.

Sun, Y., Cheng, Z. and Glick, B. R. (2009) 'The presence of a 1-aminocyclopropane-1-carboxylate (ACC) deaminase deletion mutation alters the physiology of the endophytic plant growth-promoting bacterium Burkholderia phytofirmans PsJN', *FEMS microbiology letters*, 296(1), pp. 131–136.

Sutlović, D. *et al.* (2005) 'Taq polymerase reverses inhibition of quantitative real time polymerase chain reaction by humic acid', *Croatian medical journal*, 46(4), pp. 556–562.

Sutlovic, D. *et al.* (2008) 'Interaction of humic acids with human DNA: proposed mechanisms and kinetics', *Electrophoresis*, 29(7), pp. 1467–1472.

Swigonova, Z. *et al.* (2004) 'On the tetraploid origin of the maize genome', *Comparative and functional genomics*, 5(3), pp. 281–284.

Takahashi, S. *et al.* (2014) 'Development of a prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using next-generation sequencing', *PloS one*, 9(8), p. e105592.

Terrazas, R. A. *et al.* (2016) 'Plant--microbiota interactions as a driver of the mineral turnover in the rhizosphere', in *Advances in applied microbiology*. Elsevier, pp. 1–67.

Thompson, L. R. *et al.* (2017) 'A communal catalogue reveals Earth's multiscale microbial diversity', *Nature*. Nature Research. doi: 10.1038/nature24621.

Tian, F. *et al.* (2011) 'Genome-wide association study of leaf architecture in the maize nested association mapping population', *Nature genetics*, 43(2), pp. 159–162.

Tian, Y. and Gao, L. (2014) 'Bacterial diversity in the rhizosphere of cucumbers grown in soils covering a wide range of cucumber cropping histories and environmental conditions', *Microbial ecology*, 68(4), pp. 794–806.

Timmusk, S. *et al.* (2017) 'Perspectives and Challenges of Microbial Application for Crop Improvement', *Frontiers in plant science*, 8, p. 49.

Touceda-González, M. *et al.* (2015) 'Combined amendment of immobilizers and the plant growth-promoting strain Burkholderia phytofirmans PsJN favours plant growth and reduces heavy metal uptake', *Soil biology & biochemistry*, 91(Supplement C), pp. 140–150.

Trabelsi, D. *et al.* (2011) 'Effect of on-field inoculation of Phaseolus vulgaris with rhizobia on soil bacterial communities', *FEMS microbiology ecology*, 77(1), pp. 211–222.

Trabelsi, D. and Mhamdi, R. (2013) 'Microbial inoculants and their impact on soil microbial communities: a review', *BioMed research international*, 2013, p. 863240.

Tremblay, J. *et al.* (2015) 'Primer and platform effects on 16S rRNA tag sequencing', *Frontiers in microbiology*, 6, p. 771.

Turner, T. R., James, E. K. and Poole, P. S. (2013) 'The plant microbiome', *Genome biology*, 14(6), p. 209.

Van de Peer, Y., Chapelle, S. and De Wachter, R. (1996) 'A quantitative map of nucleotide substitution rates in bacterial rRNA', *Nucleic acids research*, 24(17), pp. 3381–3391.

van Veen, J. A., van Overbeek, L. S. and van Elsas, J. D. (1997) 'Fate and activity of microorganisms introduced into soil', *Microbiology and molecular biology reviews: MMBR*, 61(2), pp. 121–135.

Vejan, P. *et al.* (2016) 'Role of Plant Growth Promoting Rhizobacteria in Agricultural Sustainability-A Review', *Molecules* , 21(5). doi: 10.3390/molecules21050573.

Viterbo, A. *et al.* (2010) 'Characterization of ACC deaminase from the biocontrol and plant growth-promoting agent Trichoderma asperellum T203', *FEMS microbiology letters*, 305(1), pp. 42–48.

Vivant, A.-L. *et al.* (2013) 'Microbial diversity and structure are drivers of the biological barrier effect against Listeria monocytogenes in soil', *PloS one*, 8(10), p. e76991.

van der Voort, M. *et al.* (2016) 'Impact of soil heat on reassembly of bacterial communities in the rhizosphere microbiome and plant disease suppression', *Ecology letters*, 19(4), pp. 375–382.

Wagner, M. R. *et al.* (2016) 'Host genotype and age shape the leaf and root microbiomes of a wild perennial plant', *Nature communications*, 7, p. 12151.

Wallace, J. G. *et al.* (2018) 'Quantitative Genetics of the Maize Leaf Microbiome', *Phytobiomes Journal*. Scientific Societies, 2(4), pp. 208–224.

Walters, W. A. *et al.* (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), pp. 7368–7373.

Wang, J. *et al.* (2016) 'Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota', *Nature genetics*, 48(11), pp. 1396–1406.

Wang, Q. *et al.* (2007) 'Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy', *Applied and environmental microbiology*, 73(16), pp. 5261–5267.

Wang, Y.-Q. *et al.* (2009) 'S-nitrosylation of AtSABP3 antagonizes the expression of plant immunity', *The Journal of biological chemistry*, 284(4), pp. 2131–2137.

Watson, E. B. (1919) *Soil Survey of the Santa Maria Area, California*. U.S. Government Printing Office.

Weiss, S. *et al.* (2014) 'Tracking down the sources of experimental contamination in microbiome studies', *Genome biology*, 15(12), p. 564.

White, D. *et al.* (1994) 'Effect of an introduced inoculum on soil microbial diversity', *FEMS microbiology ecology*. Oxford University Press, 14(2), pp. 169–178.

von Wintzingerode, F. *et al.* (2000) 'Peptide nucleic acid-mediated PCR clamping as a useful supplement in the determination of microbial diversity', *Applied and environmental microbiology*, 66(2), pp. 549–557.

Wu, S. *et al.* (2018) 'Mapping the Arabidopsis Metabolic Landscape by Untargeted Metabolomics at Different Environmental Conditions', *Molecular plant*, 11(1), pp. 118–134.

Xue, Y. *et al.* (2013) 'Genome-wide association analysis for nine agronomic traits in maize under well-watered and water-stressed conditions', *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 126(10), pp. 2587–2596.

Xu, L. *et al.* (2018) 'Drought delays development of the sorghum root microbiome and enriches for monoderm bacteria', *Proceedings of the National Academy of Sciences of the United States of America*, 115(18), pp. E4284–E4293.

Xu, X. *et al.* (2015) 'Amplicon-based metagenomics identified candidate organisms in soils that caused yield decline in strawberry', *Horticulture research*, 2, p. 15022.

Yamada, K. *et al.* (2016) 'Regulation of sugar transporter activity for antibacterial defense in Arabidopsis', *Science*, 354(6318), pp. 1427–1430.

Yan, G. *et al.* (2017) 'Effects of different nitrogen additions on soil microbial communities in different seasons in a boreal forest', *Ecosphere* , 8(7). doi: 10.1002/ecs2.1879.

Yang, B., Wang, Y. and Qian, P.-Y. (2016) 'Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis', *BMC bioinformatics*, 17, p. 135.

Yang, J., Kloepper, J. W. and Ryu, C.-M. (2009) 'Rhizosphere bacteria help plants tolerate abiotic stress', *Trends in plant science*, 14(1), pp. 1–4.

Yang, W. *et al.* (2014) 'Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice', *Nature communications*, 5, p. 5087.

Yates, S. R., Wang, D. and Ernst, F. F. (1997) 'Methyl Bromide Emissions from Agricultural Fields: Bare-Soil, Deep Injection', *Environmental science & technology*. American Chemical Society, 31(4), pp. 1136–1143.

Yedidia, I., Benhamou, N. and Chet, I. (1999) 'Induction of Defense Responses in Cucumber Plants (Cucumis sativus L.) by the Biocontrol AgentTrichoderma harzianum', *Applied and environmental microbiology*. American Society for Microbiology, 65(3), pp. 1061–1070.

Yu, J. *et al.* (2008) 'Genetic design and statistical power of nested association mapping in maize', *Genetics*, 178(1), pp. 539–551.

Yu, Q. *et al.* (2015) 'Effects of combined application of organic and inorganic fertilizers plus nitrification inhibitor DMPP on nitrogen runoff loss in vegetable soils', *Environmental science and pollution research international*, 22(1), pp. 472–481.

Yu, S. *et al.* (2007) 'Expression of a carbonic anhydrase gene is induced by environmental stresses in rice (Oryza sativa L.)', *Biotechnology letters*, 29(1), pp. 89–94.

Zahir, Z. A. *et al.* (2008) 'Effectiveness of rhizobacteria containing ACC deaminase for growth promotion of peas (Pisum sativum) under drought conditions', *Journal of microbiology and biotechnology*, 18(5), pp. 958–963.

Zhalnina, K. *et al.* (2018) 'Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly', *Nature microbiology*, 3(4), pp. 470–480.

Zhang, N. *et al.* (2014) 'Effects of different plant root exudates and their organic acid components on chemotaxis, biofilm formation and colonization by beneficial rhizosphere-associated bacterial strains', *Plant and soil*, 374(1), pp. 689–700.

Zhang, T. and Yao, Y.-F. (2015) 'Endophytic Fungal Communities Associated with Vascular Plants in the High Arctic Zone Are Highly Diverse and Host-Plant Specific', *PloS one*, 10(6), p. e0130051.

Zhang, Z. *et al.* (2019) 'Demystifying the manipulation of host immunity, metabolism, and extraintestinal tumors by the gut microbiome', *Signal transduction and targeted therapy*, 4, p. 41.

Zhou, X. and Stephens, M. (2012) 'Genome-wide efficient mixed-model analysis for association studies', *Nature genetics*, 44(7), pp. 821–824.

Zhou, Y. *et al.* (2019) 'Semiautomated Feature Extraction from RGB Images for Sorghum Panicle Architecture GWAS', *Plant physiology*, 179(1), pp. 24–37.