ABSTRACT

EYRE, ALEXANDER WAYNE. Identification and Characterization of the Rice Seed Microbiome and Other Works. (Under the direction of Dr. Ralph Dean and Dr. David Bird)

Rice (*Oryza Sativa*) is a staple food for over half of the world and its demand is expected to increase as the human population grows. Consequently, there is a need for tools that improve the yield of these crops, such as products that provide plant nutrients or protect against pests. Over the past several decades, synthetic chemical fertilizers and pesticides have made widespread appearance for crops of every kind, however their effectiveness and sustainability are limited. The plant microbiome, or the collection of bacteria, fungi, and other microbes that inhabit the inner and outer compartments of plants, is responsible for performing many of the same tasks the chemical products fulfill for crops and the recent development of high throughput sequencing technologies have enabled their detailed study. The composition and diversity of bacterial and fungal communities were deduced for six different rice seeds from two different rice genotypes, sourced from two different locations in Arkansas, USA, and from two harvest years. The seeds were divided into four compartments (grain, outer grain, husk, outer husk) and the microbes identified with Illumina MiSeq 300bp paired end sequencing. Read processing and OTU picking was performed in the bioinformatic tool QIIME, and additional analyses performed with R and TBAS. Both the bacterial and fungal microbiomes experienced declines in abundance of unique OTUs from the outer to the inner most seed compartments. Principal component analyses revealed that the samples formed distinct groupings only when visualized based on their compartment with exception to the fungal samples factored based on year. A core microbiome, or a set of OTUs shared between all samples, was deduced for each of the seed compartments. Among the bacterial genera present are *Enterobacter*, *Pantoea*, *Sphingomonas*, and *Paenibacillus*, some species of which have been reported to support rice growth. Far fewer fungi were present in the core thought to be due to the large amount of fungal variability between seed types, however some species are *Alternaria longissima* and

Cladosporium delicatulum. This core may represent a conserved set of microbes for these rice plants, and the best candidates for manipulation in the pursuit of improving plant yield.

The filamentous fungus, *Magnaporthe oryzae*, is the causal agent of rice blast disease and most destructive pathogen of rice worldwide. This fungus has been established as a model plant pathogen and has had its genome sequenced in addition to many isolates and members of its family. The abundance of available genetic data has enabled the estimation of rates of nonsynonymous (dN) to synonymous (dS) mutations occurring in gene alignments whose ratio predict whether they experience diversifying or purifying selection. These ratios were deduced for a collection of putative effector genes, proteins that modulate immune responses in plants to increase infection, by employing a maximum likelihood analysis of codon substitution models using CodeML. Regardless of utilizing sequences from 43 independently sequenced isolates, there was insufficient variability to infer selection pressure for these genes. The same algorithm was used also with 6,518 clusters of orthologous genes containing sequences from *M. oryzae* and two related species. 79% of these were estimated to be under neutral selection, 19% under diversifying selection, and 2% under purifying selection. Investigation into the putative functions of genes in each of these categories suggested metabolism and binding-related proteins are more highly conserved, while transcription factor and regulation-related proteins are actively changing. These studies of selection pressure with sets of genes can give insight into whether they are integral and conserved for the organism or are no longer useful and in need of modification or disposal.

Identification and Characterization of the Rice Seed Microbiome and Other Works

by
Alexander Wayne Eyre

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Plant Pathology

Raleigh, North Carolina

2018

APPROVED BY:

Ralph Dean Co-Chair of Advisory Committee	David Bird Co-Chair of Advisory Committee
Ignazio Carbone	

BIOGRAPHY

Alexander Wayne Eyre was born on May 23rd, 1990 and was raised alongside his younger sister in eastern Ohio. In his early years, he avidly tinkered with electronics and explored the outdoors with the Boy Scouts of America. He graduated from Watkins Memorial High School in 2008 and later pursued electrical engineering at University of Cincinnati. The following year, in a quest to gain understanding of the most fascinating machine, the human being, he transferred to The Ohio State University to pursue an undergraduate degree in molecular genetics and biomathematics. Alex graduated magna cum laude in 2013 and accepted a graduate position in the Genomic Sciences department at North Carolina State University the following year. After rotating through several labs, he joined the Fungal Genomics Laboratory supervised by Dr. Ralph Dean with interest of studying the dynamics and genomics of fungi and their plant hosts. During his graduate career, he joined the Department of Entomology and Plant Pathology and got involved in projects seeking to absorb as many skills as possible. Two years later, a change of heart led him to switch to a Master program from the original doctoral path with a desire to explore the world of scientific application and industry. After graduation, Alex intends to travel the United States to visit various labs and companies to explore his mysterious future. He spends much of his free time reading, creating gadgets or artwork, and exploring the outdoors. He hopes to find a career path that dives deeply into the complexity of biological systems in order to develop holistic technologies.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 3

LIST OF FIGURES

Chapter 3

Chapter 4

Appendix A

CHAPTER 1

Engineering of the Plant Seed Microbiome for Agriculture

Alexander W. Eyre

Fungal Genomics Laboratory, Center for Integrated Fungal Research, Department of Plant Pathology, North Carolina State University, Raleigh NC, 27606, USA

Abstract

Microbial products have been used for the promotion of beneficial crop qualities and defense against pests for over a century. Their effectiveness is often limited due to the application methods, their stability, environmental variability, and the immense crop acreages needing treatment. Engineering of the seed endophytic microbiome directly may be an alternative that negates these issues and introduces novel microbes at the earliest stage of the plant life cycle, however, more research needs to be performed to identify the most appropriate microbes and develop methodologies for their introduction. In this review, the current literature on the acquisition of the seed microbiome, its constituents, and functions for the plant during and after germination will be covered. This diverse microbiota has a role in the protection of seedling when it is most vulnerable stage. By understanding the mechanisms that formulate these seed communities and the rules that govern their stability, the manipulation of the seed endophytic microbiome may become an effective strategy for the agricultural industry. Once established in the seedling, microbial additions would be potentially maintained in the mature plant, eliminating the need to apply additional microbes, creating a more sustainable means of promoting beneficial plant traits.

Introduction

By the year 2050, the human population is estimated to surpass 9.2 billion, an increase of nearly 1.6 billion people from today's global population estimate (Bongaarts et al., 2009). As

a result, there will be a continuous growth in demand for agricultural foodstuffs. A naive solution to this problem would be to allocate more land for crop usage, however it is already limited with more than 37% of the world's land area and nearly half of the area of the United States devoted to various agricultural practices alone (Bigelow, 2017; The World Bank, 2018). More importantly, the poor soil quality, lack of water, and need for additional resources makes this strategy unsustainable. Instead, the efficiency of existing farmland will need to be improved to produce either greater quantity and/or higher quality products (Robertson et al., 2005). The usage of chemical sprays to control pests or the application of synthetic fertilizers to increase soil nutrition may have value when and where available. These methods run into significant drawbacks however, such as a limited number of functional pest controlling applications, off target effects, low bioavailability of the nutrients, and runoff of the chemicals to pollute the environment surrounding the crop (Carvalho, 2006; Aktar et al., 2009).

Over the past few decades the topic of the plant microbiome, or the collection of microbes that inhabit in and around the plant tissues, has been of growing interest. The recent development of PCR and high throughput sequencing technologies has enabled the exponential growth of microbiome studies since the 1990's. This research has shown that plant tissues have an abundance of microbes, many of whose role has been discovered to provide benefits for the plant and/or microbial community (Lebeis et al., 2012; Bulgarelli et al., 2013). Moreover, a large amount of research has demonstrated that plant physiology is better understood when microbial communities are included as part of the plant system, suggesting that future studies should focus on the holobiont, or the plant-microbe macroorganism (Vandenkoornhuyse et al., 2014). This idea exemplifies the importance of the microbes and raises the question, how can these be improved to bolster plant qualities? Not surprisingly, agricultural products already exist that utilize one or more microbes to either aid in nutrient uptake or provide defense against a pathogen to improve plant yield (Calvo et al., 2014). In some cases, these are economically feasible alternatives to the use of chemical fertilizers and pesticides (O'Callaghan, 2016),

however they run into similar drawbacks as their chemical counterparts due to their methods of application. In addition, many microbial products are often restrained to the specific plant system in which they were developed, not allowing for their usage with other plant species or in different geographic locations (Timmusk et al., 2017).

An alternative to the external application of microbes is the modification of the community within the plants themselves. In theory, this kind of approach provide a more long-term association of the microbes with the plant without the need of reapplication or consideration of microbial runoff. One recently explored method has been the engineering of the microbiome within the plant seed, which may represent a strategy capable of being scaled to an industrial level. Research as early as the 1940's has demonstrated that all plant seeds have an abundance of microbes whose composition is unique to the species of study (Wallace et al., 1951). More recent works have demonstrated the role these microbes have on the successful germination of the seed, thought to be due to defense against soil borne pathogens (Munkvold, 2009). Most importantly, members of these microbiomes are later found in both the foliar and root structures of the plant (Hardiom et al., 2012), suggesting that any stable additions or modifications to the seed would later be reflected in the mature plant. Not only could such a strategy modify the microbes within the plant, but it could establish beneficial microbes at the earliest possible stage of plant life. To realize the research needed to develop this methodology for agriculture, this review presents current literature on the nature of the seed microbiome, its evolution through the seed life cycle, and how contemporary microbial product difficulties can inform the engineering of this novel strategy.

Seed Microbiome Acquisition

Seed development represents the earliest stage of a seed's lifecycle and the period in which most of its microbiome is acquired. During this time, microbes from the neighboring environment and plant tissues colonize both the seed interior and exterior, or endophytic and epiphytic tissues respectively. Several mechanisms are known to supply the constituents of

these communities, although the relative contribution of each to the final microbiome is not well characterized (Vandenkoornnhuyse et al., 2015). Identification of the major pathways can enable the modification of these communities, as will be explored later in the review. In general, the mechanisms are broken up into two inheritance categories: vertical and horizontal inheritance (Gundel et al., 2011).

Vertically inherited microbes are those that are passed on to the seed from the parent plant. Three pathways have been described, the first of which is through the funiculus and chalaza, vascular tissues joining the seed endophyte and the plant. During seed development, these structures are responsible for the transfer of nutrients into the seed for some time before the chalaza closes off at the antipodal side of the seed (Agarwal et al., 1996). Microbes that occupy the vascular spaces, particularly in the plant xylem, can travel along with the nutrients and colonize the seed interior. However, there is often a time constraint for the availability of this pathway that may exclude microbes without their own means of motility through these tissues. A second pathway is through the shoot apical meristem that grows into the floral structures. Microbes that inhabit this bundle of multipotent cells are later found in many of the tissues that emerge during floral development, including those involved in the growth of the seed. This route appears to be less specialized and more effective than the vascular pathways allowing for a broader range of microbes to colonize the developing seed, but it also has limitations due to the transience of many plant species' flowers (Pirttilia et al., 2000; Darsonval et al., 2009). Finally, microbes are passed to the seed from the plant gametes. While the plant pollen and ovule are a product of floral development, their microbiome is typically different from the other tissues (Malfanova et al., 2013; Junker et al., 2015). This mechanism likely has a greater contribution to community diversity for those plants that are not self-pollinating.

Horizontally inherited microbes are those that are passed to the seed from its surrounding environment. Many of these mechanisms have sporadic and short-lived contributions to the final seed microbiome due to their chance of occurring, but collectively make

a significant contribution. Here, only the major pathways that may take part in agricultural systems will be considered, excluding those that may impact the seed after maturation and harvest. There are several natural means of horizontal inheritance, the most prevalent of which being the transfer of microbes from surrounding plant tissues not involved in the seed development process (Won et al., 2003). This may be a major factor for crops with more exposed seed, such as many of the cereal grains. In this case, microbes on the surface of petals, leaves, and other tissues are transferred to the seed in the event of wind, rainfall, or even air movements during periods of warming weather (Lindemann et al., 1985). For crops that rely heavily on insect pollination such as apple, legumes, and cucurbits, in addition to the contributions from the pollen are those microbes transferred from the insect itself (Ambika Manirajan et al., 2016). Ushio et al. (2016) found sizeable microbiomes on the surfaces of a variety of pollinating insects. These insects were allowed to pollinate the flowers, and before-and-after snapshots of the floral microbiome revealed significant changes in diversity due to this contact. Finally, microbes may colonize the seed through aerial dispersal, however their success is largely dependent on their size and subsequent ability to be airborne (Lindemann et al., 1985; Wilkinson et al., 2012).

There are a few contributions from the seed harvesting process itself, although these microbes should be considered contaminants. Some of these pathways include contact with the harvest machinery, contact with other biological material, and handling during processing or storage (Hill et al., 2009). These may be more important for those crops that account for terroir, or the thorough consideration of farming methods, soil type, climate, and more. These crops include coffee beans, wine grapes, tobacco, and more, and may simply be exposed to additional microbes due to increased handling by their growers (Klaedtke et al., 2016). Overall, horizontal inheritance mechanisms largely impact the seed post-development due to the protection given by the floral structures, suggesting that these pathways primarily impact the microbial composition of the seed epiphyte. Altogether, microbes that make up the seed

endophyte are mainly derived from vertical inheritance mechanisms, while those in the seed epiphyte are a mixture of both vertical and horizontal inheritance. The contribution of each of these pathways to the final seed microbiome remain ambiguous for bacteria, however studies have suggested that up to 90% of the inhabiting fungi are acquired through vertical inheritance means (Ngugi et al., 2006).

Research into acquisition mechanisms is primarily performed by comparing the microbiomes of the seed to their hypothetical source. Rigorous methodology is generally lacking, however, as the association between the seed and its source is typically by implication. For example, it would be impossible to distinguish the origin of a seed microbe if it were found both in the plant vascular tissues and a pollinating insect. To resolve this issue, individual bacteria or fungi need to be tracked using markers to visualize their transfer. This has been previously performed with the GUS labelling of bacterial species, then staining to observe their localization (Lubeck et al., 2002). A more contemporary method is the utilization of fluorescent reporters, such as GFP. Fluorescence in situ hybridization (FISH) can then be employed, which allows for quality visualization and computer aided measurement (Cardinale, 2014). Additionally, microbes identified in plant tissues not directly associated with the inheritance mechanisms should not be considered as a microbial source, as the microbiomes can fluctuate significantly among the tissues (Junker et al., 2001; Nelson, 2017). Altogether, while there is an abundance of information available about the various inheritance mechanisms, there remains a great deal of research is needed to investigate the contribution of each to the final seed microbiome.

Seed Microbiome Composition

Once the plant seed has been harvested and prepared for planting, its microbiome is established. Study of these microbial communities began as early as the 1940's when researchers realized that bacteria could be cultured from different crop seed (Wallace et al., 1951). Several decades later, PCR-based identification methods revealed that not only were

there bacteria present, but they were abundant and were often unique to each plant species (Hallmann et al., 1997). As high throughput sequencing technologies emerged, the number of microbiome studies increased exponentially, and many seed microbiome studies have revealed a highly diverse community of bacteria, fungi, viruses, and oomycetes (Finkel et al., 2017). The study of the microbial composition in plants is particularly important as a high diversity of microbes often corresponds to improved plant traits and adaptability to environmental changes (Crutsinger et al., 2006), and maintenance of this diversity in the seed may correspond to success for the plant at its earliest stage. Due to the relatively low abundance of studies, further discussion of oomycetes and viruses will not be covered here.

Like most plant tissues, studies have shown that the seed exhibits a reduction in the abundance and diversity of microbes as one moves from its outer- to its innermost compartments (Sanchez-Canizares et al., 2017). Most knowledge of the seed microbiome come from bacterial studies, because they are the most abundant microbe in plants and have more often been associated with improved plant qualities. Thorough investigation into these bacterial communities have been annotated for rice (Okunishi et al., 2005; Cottyn et al., 2009), maize (Rijavec et al., 2007), wheat (Robinson et al., 2016), and other plants, with emphasis on those of high agricultural value. These studies revealed hundreds of unique species existing within the seed, with 131 genera from 4 different phyla being reported as naturally occurring among 25 plant species. Among the most commonly shared bacterial genera are: *Agrobacterium, Bacillus, Burkholderia, Enterobacter, Paenibacillus, Pantoea, and Pseudomonas* (Nelson 2004; Truyens et al., 2015). These members are often of little surprise, however, as they also predominate in most soil communities (Fierer et al., 2012; Finch-Boekweg et al, 2013).

The fungal seed community follows similar trends as the bacteria where the diversity decreases from the epiphyte to the endophyte, and highly represented members are commonly found in the soil. However, studies on fungal constituents of plants have primarily focused on

potential pathogens as many have been associated with disease, aside from mycorrhizae and their symbiosis with plant roots. Many of the beneficial fungal members have been recently uncovered due to the improved throughput of the methods employed (Vanderkoornhuyse et al., 2002). As a result, these have been annotated for several crop seeds, including wheat (Links et al., 2014), maize (Abe et al., 2015), and more (Barret et al., 2015). Among these communities, there tends to be a high prevalence of fungi from the class *Dothideomycetes* in the phylum *Ascomycota* with a wide range of possible functions for the plant. There are several fungal genera that tend to dominate the seed compartments, such as: *Alternaria, Cladosporium, Epichloe, Epicoccum, Fusarium, Microdochium, Phaeosphaeria, Stagonospora, and Xylaria* (Barret et al., 2015; Nelson, 2017; Geisen et al., 2017).

When assessing the composition of plant microbiomes, a large amount of variation is often present within a single crop, which can be an issue when attempting to summarize the constituents of a single species' seed. Several factors are known to contribute to this effect, including the plant genotype, soil composition, and seasonal environmental conditions (Buyer et al., 1999; Hacquard et al., 2016). Plant genotype is thought to impact the microbiome primarily through the variable expression of host genes. These may alter plant metabolism causing a change in the root exudate profile, a collection of secreted carbon-rich compounds used to recruit soil borne microbes. Additionally, these genes may alter the plant innate immune system that directly interacts with its microbiome, acting as a kind of selection mechanism (Khalid et al., 2004; Bulgarelli et al., 2013). Microbiome shifts due to seasonal environmental changes are widely acknowledged, as microbes have often been considered to aid short-term environmental adaptation. Fluctuations in water, temperature, light, soil acidity, and more can contribute to these shifts through alterations of the conditions within the plant tissues occupied by microbes (Barret et al., 2015). The soil often makes the greatest contribution to variation in the plant microbiome, as its physical and chemical composition has a direct impact on the microbiome

(Hardiom et al., 2012; Wagner et al., 2016). Nutrient levels, their bioavailability, water retention, organic matter, and more can cause variation in the soil microbiome the roots recruit from.

A recently pursued means for circumventing the effects of these factors has been to develop ‘core microbiomes’ for plants and their tissues. A plant core microbiome is defined the collection of microbes present among all samples taken from various genotypes, derived from variable soil or environmental conditions (Huse et al., 2012; Lundberg et al., 2014). This group represents a conserved subset of the microbiome, most likely to be identified if a plant is randomly sampled within the context of the factors included in the core. For seed, a core microbiome would be constructed by collecting mature seed from multiple plant genotypes, soil plots, and seasons, then their microbiomes compared to identify those that are present in all samples. If comparing operational taxonomic units (OTUs) between these samples, an OTU is typically considered part of the core if it appears in at least 90% of the samples (Huse et al., 2012). This idea runs into theoretical limitations, however, as the size of this group would approach zero as the number of different seed samples increases due to natural variability. One potential solution to this dilemma is to relax the taxonomic constraints, possibly only defining a member as part of the core down to the family level instead of the genera or species level. This may include many more members, but their functional role for the plant or microbiome will become more ambiguous. Additionally, due to the great impact of the soil and environment, the core microbiome could simply be adjustable. For example, a core microbiome from all possible regions and genotypes around the globe might approach zero members, but for some work it might be better to narrow the geographic context of the core allowing for more microbes to be represented. Regardless, as high throughput sequencing technologies and microbiome-associated methodology improves, insights into the sources of the immense community variability may help resolve these issues and provide a clearer picture of these microbiome compositions.

Function for the Seed and Seedling Microbiome

Over the past several decades, many bacteria and fungi have been discovered to provide benefits for plants (Finkel et al., 2017). These microbes have powerful roles in the promotion of plant growth or resistance to biotic and abiotic stressors. The understanding of these microbes' functions has enabled the development of commercial microbial products, as will be later reviewed. These effects have been observed as early as germination, which represents the most dynamic stage of a plant's life cycle. Extending this knowledge to the seed microbiome may enable the stable introduction of highly functional microbes at the earliest stage of a plant's development. While it remains to be studied if the microbiome has a preservation effect for the dormant seed, much has been observed of its impacts during and immediately after germination.

After being planted in sufficient soil conditions, the seed imbibes water and opens. During this process, the water displaces nutrient reserves called the seed exudates into the surrounding soil. These exudates are a concentrated and diverse collection of carbon- and nitrogen-rich compounds that are quickly utilized by the surrounding microbes. This develops what is known as the spermosphere, or a zone of elevated microbial activity, an area no larger than 5-10mm around the seed (Slykhuis, 1947; Nelson, 2004). This space is competed for by microbes from the seed epiphyte, endophyte, and surrounding soil. The chemical constituents of the exudates play an important role in the microbes that dominate this region, as some can utilize compounds better than others. Steinauer et al. (2016) demonstrated this phenomena by adding artificial exudate solutions to mono- and polyculture plant systems. The solutions contained either a low or high diversity of either carbon- or nitrogen-rich compounds. Controlling for plant species richness, which has been observed to also contribute to microbial diversity, the authors observed significant increases in microbial biomass with reductions in Shannon diversity as the diversity of the nutrients in the exudates increased, more so for the C-rich solutions. These results suggested that the diversity of nutrients within the seed may be

able to act as a selective mechanism for the microbes that best thrive in the spermosphere. However, the source of the microbes that dominated this compartment remains unclear. Some studies provide evidence that they are mainly derived from the soil (Klaedtke et al. 2016), while others have shown that they come from the seed itself (Barret et al., 2015; Cope-Selby 2016).

The microbes in the spermosphere are known to aid the germinating seed in several ways. While the emerging plant is at its most vulnerable stage, the seed-borne microbes can protect the seedling by reproducing the fastest in the exudates to either antagonize or act as a physical barrier to soil-borne pathogens. The latter idea was demonstrated by Bacilio-Jimenez et al. (2001) where two bacterial species from the endophyte of rice seed were able to occupy the spermosphere and early roots that grew into this space. These bacteria were observed coating the root and competing with the growth-inhibiting bacteria, *Azospirillum brasilense*. The success of this defense mechanism is additionally dependent on the pathogen load, where if a pathogen is in abundance in the soil or on the seed coat, the germination rate will be drastically reduced regardless if the seed is normally resistant (Darrassee et al., 2007). Seed microbes have also been observed regulating the rate of germination through the release of phytohormones. Goggin et al. (2015) demonstrated the dependence of annual ryegrass on cytokinin-producing bacteria for breaking seed dormancy. The presence of the bacteria enabled even germination of the seed relative to those that were artificially induced (2015). These growth promoting members appear to be vital for the plant at these early stages, as they limit the time spent in this most vulnerable stage (Beckstead et al., 2007).

After germination, the seedling emerges from the soil and matures along with its microbiome. At this stage, the spermosphere represents a fraction of the area taken up by the seedling, although its microbial constituents can be found throughout the plant tissues (Nelson et al. 2004; Darrasse et al. 2010). This suggests this small community can act as an initial inoculum for the plant, however its impact on the seedling roots is ambiguous as they secrete exudates of their own to recruit microbes from the surrounding soil. In rice, this idea was

studied, and the root microbes appeared to be derived from soil itself, although the distinction was difficult to make as the seed and soil communities already resembled each other. This phenomenon was mainly observed as a loss of those microbes unique to the seed (Hardiom et al., 2012). Conversely, while the microbiome of the belowground plant structures resembled that of the soil, some of the unique microbes carried by the seed were found in abundance in the mature plant (Compant et al., 2010; Huang et al., 2016). This suggests that functional microbes carried by the seed are capable of being conserved across generations of the plant. Although the overall contribution of the seed microbes to the mature plant microbiome remains unclear (Johnston-Monje et al., 2014), this idea has been the main driver in the engineering of seed endophytes.

The function of microbes for the seedling can be classified into having plant growth promoting (PGP) or defensive roles. Those with PGP roles produce or make bioavailable nutrients or phytohormones involved in plant growth, such as nitrogen (Desbrosses et al., 2011), phosphorous (de Freitas et al., 1999), auxin (Dobbelaere et al., 1999), and many more. Among the most studied and important PGP microbes for agriculture are the nitrogen fixing rhizobacteria that form symbiotic relationships with the roots of leguminous crops, such as *Azospirillum* spp. These bacteria also provide benefits to non-legumes. Two *Azospirillum* species were shown to contribute 7-12% of the total nitrogen to wheat (Malik et al., 2002), while *Azospirillum diazotrophicus* was shown to contribute 60-80% of the total nitrogen to sugarcane (Boddey et al., 1991), exemplifying these species' importance for crop success. Additionally, several PGP fungi have been annotated. Mycorrhizal fungi, such as *Rhizophagus clarus*, are known to provide benefit to a wide range of plants, promoting growth through the modification of plant hormonal pathways with targeted effector proteins to the roots (Sedzielewska Toro et al., 2016). Some microbes also have defensive roles against the various pests and pathogens of plants. These are often able to protect the plant through direct interaction with the pathogen, as is the case with many fungi in the *Hypocreaceae* family. Genera such as *Trichoderma* can

antagonize many plant pathogenic fungi or oomycetes, and in some cases also secrete plant growth promoting compounds (Contreras-Cornejo et al., 2016). Other defensive microbes function indirectly by bolstering the plant immune system after triggering systemic induced resistance (SAR), causing a plant-wide preparation against future infection (Pieterse et al., 2014). These are of special interest to crops that exhibit pathogen vulnerability at early stages of growth, as they can preemptively bolster the immune system.

Some considerations need made in the study of microbial function, foremost that the research is often performed through plant co-culturing experiments in search for improved plant qualities. This requires that the microbes are first culturable, a challenge made evident by the large divide in the number of identified species between culture-based and sequencing-based experiments. This suggests that either sufficient media has not yet been formulated, or that the microbes are dependent on some component of their community or plant host, which would represent a knowledge gap in the understanding of plant microbiomes. Additionally, the functional aspects of both bacteria and fungi are often transient in culture, disappearing after several hours (Finkel et al., 2017). Therefore, some of what is observed in laboratory experimentation might be a misrepresentation of what is occurring in the field.

More importantly, the function of microbes may be better understood in terms of the genes they express that are utilized by the plant (Bulgarelli et al., 2013). A number of these protein encoding genes have been annotated, carrying basic functions such as cytokinin synthesis, phosphate solubilization, or antifungal compound synthesis. The concept changes the microbiome landscape from a collection of unique species to a collection of microbial 'vehicles' that carry particular genes or operons (Lemanceau et al., 2017). For example, a *nifH* vehicle for the fixation of nitrogen might consist of multiple diazotroph species (Bouffaud et al., 2016). This idea would transform the core microbiome into a core metagenome, where the metagenome is comprised of functional genes shared between plant samples taken from different genotypes, soil types, etc. However, there remains a great number of genes to be

identified as some members of a core microbiome may not necessarily be a vehicle for plant beneficial genes but carry some alternative function, possibly support of the microbial community. This concept may offset the diminishing size of the core as additional samples are considered, however its pursuit may require whole genome sequencing and annotation of the majority of a plants' microbiome (Busby et al., 2017). It also carries intriguing implications for those microbes outside the core, where these members' genes may carry 'niche' functions. These might be important for the community's adaptation to different plant genotypes, soil conditions, or environmental fluctuations (Tucker et al., 2016). While this method may work well for identifying those members important for the improvement of plant qualities post-germination, a separate set of genes need to be identified for microbes that proliferate in the spermosphere. These would likely include metabolism genes corresponding to the primary exudate compounds that emerge from a seed.

Seed Microbiome Engineering for the Agricultural Industry

The first commercial plant microbial product was patented more than 100 years ago under the name of 'Nitragin,' a rhizobacterium for the improvement of leguminous crops through nitrogen fixation. Since then, hundreds of microbial products have been manufactured utilizing a wide range of microbial species and application strategies (Bashan, 1998; Deaker et al., 2004). These are broadly categorized depending on the function they provide to the plant: biocontrol agent (BCA) and plant growth promoting (PGP) products. Over the past five years, the market for PGP products, also known as biostimulants, have had a compound growth rate of over 12% on the global market. These numbers are expected to grow with the increasing interest in plant microbes as alternatives to their chemical counterparts and the development of technologies needed to study the complex interactions (Calvo et al., 2014).

A variety of methods exist for the application of microbes such as soil granules, foliar sprays, and seed coats, each with their own advantages and disadvantages. Soil granules are typically microbial suspensions imbued into a porous pellet to be added to soil or fertilizer

mixtures and may be substituted with liquid forms of inoculants. Many of these products have been successful at increasing yield, such as MycoGold™ for nutrient uptake and Symbion-P® for the solubilization of phosphate, however negative effects on crops are possible if little microbial diversity exists in the soil prior to application. This is primarily an issue for farms that do not implement sustainability practices or till prior to planting, which strongly disrupts the existing soil microbiome (Magarey 1999; Hartmann et al., 2015; Parnell et al., 2016; Zhao et al., 2016). Additionally, the microbes might have difficulty finding a hospitable niche within the soil due the presence of competitive microbes and heterogeneous soil conditions (van Elsas et al., 1993). Foliar sprays have often successfully improved qualities of crops, whether it be from application of microbial byproducts or microbes themselves. Often these have unknown mechanisms due to the complexity of the interactions in consideration (Saa et al., 2015). Like most chemical foliar sprays, there is the possibility that they are washed away with precipitation. However, this may work in a microbe's favor if it is able to reach the local root systems of the maturing plants.

A more recent microbial application has been the addition of microbes to the seed coat. This method uses a microbial slurry or polymeric matrix, such as alginate, to cover the seed prior to sowing (O'Callaghan, 2016). In theory, the large concentration of microbes should participate in the competition over the spermosphere and later find their way into the mature plant. This has proved troublesome, where in some instances more than 95% of the microbial inoculum has expired within 4 hours after seed inoculation. Consequently, the best strategy has been inoculating immediately prior to planting, however most farmers prefer to utilize pre-inoculated seed (Bashan, 1998; Deaker et al., 2004). This problem may simply represent a knowledge gap in the stability of epiphytic microbes and has not stopped biotechnology companies in their search for new seed coat microbes (Broadfoot, 2016).

Modification of the seed endophytic microbiome may be an alternative that avoids many of the drawbacks the external microbial applications experience. In this review, evidence that this community can act as both a protectant for the germinating seed and an initial inoculum for

the emerging seedling has been presented. Therefore, if the microbiome can be manipulated, this method may provide a means of altering the endophytic plant microbiome free from the effects of environmental changes and soil composition. An application might resemble treatment of plants with the microbe of interest sometime around seed development before maturation. These modified seeds would effectively be pre-inoculated and, if they do provide beneficial properties, should be of greater interest to farmers given they are produced in a cost-effective manner (Deaker et al., 2004). However, there is a great deal of research still needed performed to make this application a reality.

The first challenge will be to develop methodology that introduces microbes into the interior of the seed. This technique will need to consistently incorporate relatively even levels of inoculum to the endophyte, and it is not evident that a mechanical process working with matured seed would be effective or efficient at handling this task. Literature on seed microbiome acquisition mechanisms contain clues, however. Since the interior of the seed is the primary target, many of the horizontal inheritance mechanisms will not be useful as these typically affect the seed epiphyte, however those with vertical inheritance mechanisms have promise. During seed development, the shoot apical meristem grows into the floral structures, then internal vascular tissues transport valuable nutrients to the seed site. This represents a period in which a microbial inoculum could spread into the emerging flower and seed. The time at which the inoculum is applied should be the most important factor, which will vary depending on the plant species. Application after the formation of the floral shoot meristems may be an optimal starting point, as it has the least selectivity in transporting the microbes to the site of seed development (Darsonval et al., 2009). This idea may be optimal for those microbes that have the means to travel through the vascular tissues as it could allow for the utilization of two mechanisms for seed colonization. Alternatively, an inoculum could be applied during flowering, however this would require careful attention to timing and the protection the flower offers to the seed.

After an introduction method has been developed, a second challenge will be to determine a microbe(s) to add to the seed community. These can function to either provide the seed protection if germination rate is an issue, or act as a plant growth promoting (PGP) or biocontrol agent (BCA) product. Since the microbes that largely aid the germinating seed seem to be those that thrive in the spermosphere, the best candidates for seed protection will be those that can grow most optimally in the seed exudates and carry antagonistic effects if there is a known soil borne pathogen. Literature has yet to be published in this field, however finding candidates might begin with identifying the main compounds in the seed exudates then screening for microbes that thrive on minimal media containing a similar nutrient profile. Co-culturing of these with a pathogen in search for antagonistic effects, if possible, may pinpoint ideal microbes if the goal is to protect the germinating seed. These experiments would run into the prevalent issue of getting much of the microbiome into culture, where some microbes might be dependent on some plant or microbiome factor. The screening may be remedied through the search of key metabolism genes if the metagenome concept is first pursued.

If the goal is to add microbes with PGP or BCA roles to the plant microbiome, studies investigating these for the mature plant will be sufficient. However, there is the possibility that the seed inoculum goes through a bottleneck during germination due to the small concentration of microbes in competition with those from the soil and seed epiphyte that might cause loss of the additions. Identification of those that can thrive in the seed exudates as previously mentioned might mitigate this issue although this might drastically limit the candidates. A naïve remedy would be to increase the abundance of the microbe in the seed, but this may cause unfavorable changes to the microbiome diversity. Alternatively, reduction of the soil microbes prior to sowing may reduce competition, however this procedure may be unfavorable for the plant root microbiome and further experimentation would need performed to find balance (Zhao et al., 2016). A more suitable solution has been outlined by Busby et al. following the development of the core metagenome for plant microbiomes. The authors suggest that

comparison of the genomes of the plant and the main microbial constituents may yield a set of 'rules' that drive the formation of the microbiome. Such a study may reveal a set of genes or other factors that dictate the vital plant-microbe interactions (2017). These results would inform stability of a microbe addition through the number and quality of interactions with the plant and other members of the microbiome.

Once the microbe has been successfully introduced into the seed and taken residence in the plant, a final consideration is its impact on the mature plant microbiome. If the microbe is introduced at high levels it may reduce the overall diversity. This may result in the plant's decreased ability to adapt to niches, environmental changes, and may affect its overall productivity (Crutsinger et al., 2006; Weidner et al., 2015). This phenomenon exemplifies the importance of refining the methods of microbial introduction. The level of inoculum should be adjustable so that if the effects on diversity are too great, the quantity can be decreased in future by an amount corresponding to a resolution of the problem. Additionally, the method will need to add the microbe at uniform levels among all seed, or else great variability in productivity may be observed in the field.

A recently published study was able to successfully apply many of these principles to add a microbe to seeds' endophytic communities. The authors successfully introduced *Paraburkholderia phytofirmans*, a powerful plant growth promoting bacterium, into various crop seed. Maize, soy, and pepper were inoculated in the laboratory, and wheat in both the laboratory and field, then the presence of the new community additions confirmed with GUS-staining. Each of the plants were sprayed with a bacterial inoculum at growth stage 61-63 on the BBCH scale, a system for uniformly coding phenological growth stages of all mono- and dicotyledonous plant species (Meier, 2001). This corresponded to periods immediately before or during flowering of the included plants. The wheat grown from treated seed exhibited significantly earlier emergence of their grain-bearing ears and increased numbers of ears per plot, although with great variability that might be explained by the consistency of their treatment

method. Conversely, the microbial diversity shifted significantly not only due to the introduction of the bacterium at high levels, but an increased abundance of others denoting an unprecedented microbial interaction. The *P. phytofirmans* did not last more than a single generation, suggesting instability of the microbe within the wheat plant or its microbiome. Regardless, the study demonstrated successful methodology for the introduction of microbes into the seed endophyte and a future for engineering pursuits (Mitter et al., 2017).

In this review, current literature assessing the nature of plant seed and its microbiome was covered to unveil details that may aid in its engineering. Acquisition mechanisms indicated pathways to introduce microbes to the seed endophyte, and the variety of functions for both the germinating seed and plant offered means to improve plant traits. A recent study demonstrated successful addition of a novel bacterium to several crops' seed, however microbial diversity and the stability of the bacterium were lacking. Research progress in this area has broad implications for agricultural production and sustainability practices. For example, highly resistant crop cultivars that exhibit a reduction in yield due to the energy expended on their heightened defense might be able to restore lost yield through the manipulation of the seed microbial communities. This would develop robust high-yielding cultivars, however there remains a great deal of work to be performed in the stability of microbial communities to make such an endeavor possible. Additionally, engineered seed would have an established microbiome that would not obviate issues for farmers with inoculum applications, and with the understanding of the rules that govern stable communities might allow for a one-time acquisition by the seed. Overall, this novel strategy offers opportunities for creative and sustainable solutions to the human populations increased demand for food production.

References

Abe CAL, Faria CB, de Castro FF, de Souza SR, dos Santos FC, da Silva CN, Tessmann DJ, and Barbosa-Tessmann IP (2015). Fungi Isolated from Maize (*Zea mays* L.) Grains and Production of Associated Enzyme Activities. International Journal of Molecular Sciences 15, 15328-15346.

Agarwal PD and Sinclair JB (1996). Principles of Seed Pathology. Boca Raton, FL, USA: Lewis Publication.

The World Bank: Agricultural land (% of land area) (2018). Retrieved from https://data.worldbank.org/indicator/AG.LND.AGRIZS

Aktar MW, Sengupta D, and Chowdhury A (2009). Impact of pesticides use in agriculture: their benefits and hazards. Interdisciplinary Toxicology 2:1, 1-12.

Ambika Manirajan B, Ratering S, Rusch V, Schwiertz A, Geissler-Plaum R, Cardinale M, and Schnell S (2016). Bacterial microbiota associated with flower pollen is influenced by pollination type, and shows a high degree of diversity and species-specificity. Environmental Microbiology 18:12, 5161-5174.

Barret M, Briand M, Bonneau S, Preveaux A, Valiere S, Bouchez O, Hunault g, Simoneau P, and Jacques MA (2015). Emergence Shapes the Structure of the Seed Microbiota. Applied Environmental Microbiology. 81:4, 1257-1266.

Bacilio-Jimenez M, Aguilar-Flores S, del Valle MV, Perez A, Zepeda A, and Zenteno E (2001). Endophytic bacteria in rice seeds inhibit early colonization of roots by *Azospirillum brasilense*. Soil Biology and Biochemistry 33, 167-172.

Barampuram S, Allen G, and Krasnyanski S (2014). Effect of various sterilization procedures on the in vitro germination of cotton seeds. Plant Cell 188, 179-185.

Bashan Y (1998). Inoculants of plant growth-promoting bacteria for use in agriculture. Biotechnology Advances 16:4, 729-770.

Beckstead J, Meyer SE, Molder CJ, and Smith C (2007). A race for survival: can *Bromus tectorum* seeds escape *Pyrenophora semeiperda*-caused mortality by germinating quickly? Annual Botany 99, 907-914.

Berg G (2009). Plant-microbe interactions promoting plant growth and health: perspectives for controlled use of microorganisms in agriculture. Applied Microbiology and Biotechnology 84, 11–18.

Bigelow D (2017). USDA ERS - Major Land Uses. https://www.ers.usda.gov/data-products/major-land-uses/

Bongaarts J (2009). Human population growth and the demographic transition. Phil. Trans. R. Rosc. B 364, 2985-2990.

Bouffaud ML, Renoud S, Moenne-Loccoz Y, and Muller D (2016). Is plant evolutionary history impacting recruitment of dizotrophs and nifH expression in the rhizosphere? Scientific Reports 6:21690, 1-9.

Broadfoot M (2016). Microbes Added to Seeds Could Boost Crop Production. Scientific American.

Bulgarelli D, Schlaeppi K, Spaepen S, van Themaat EVL, and Schulze-Lefert P (2013). Structure and Functions of the Bacterial Microbiota of Plants. Annual Review of Plant Biology 64, 807-838.

Buyer JS, Roberts DP, and Russek-Cohen E (1999). Microbial community structure and function in the spermosphere as affected by soil and seed type. Canadian Journal of Microbiology 45, 138-144.

Calvo P, Nelson L, and Kloepper JW (2014). Agricultural uses of plant biostimulants. Plant Soil 383, 3-41.

Cardinale M (2014). Scanning a microhabitat: plant-microbe interactions revealed by confocal laser microscopy. Frontiers in Microbiology 5:94, 1-10.

Carvalho FP (2006). Agriculture, pesticides, food security, and food safety. Environmental Science and Policy 9, 685-692.

Compant S, Clement C, and Sessitsch A (2010). Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization. Soil Biology and Biochemistry 42:5, 669-678.

Cope-Selby N, Cookson A, Squance M, Donnison I, Flavell R, and Farrar K (2016). Endophytic bacteria in Miscanthus seed: implications for germination, vertical inheritance of endophytes, plant evolution and breeding. GCB Bioenergy 9, 57-77.

Cottyn B, Debode J, Regalado E, Mew TW, and Swings J (2009). Phenotypic and genetic diversity of rice seed-associated bacteria and their role in pathogenicity and biological control. Journal of Applied Microbiology 107, 885-897.

Crutsinger GM, Collins MD, Fordyce JA, Gompert Z, Nice CC, and Sanders NJ (2006). Plant genotypic diversity predicts community structure and governs an ecosystem process. Science 313, 966-968.

Darrasse A, Darsonval A, Boureau T, Brisset MN, Durand K, Jacques MA (2010). Transmission of plant-pathogenic bacteria by nonhost seed without induction of an associated defense reaction at emergence. Applied Environmental Microbiology 76, 6787-6796.

Darsonval A, Darrassee A, Durand K, Bureau C, Cesbron S, and Jacques MA (2009). Adhesion and fitness in the bean phyllosphere and transmission to seed of *Xanthomonas fuscans* subsp. Fuscans. Molecular Plant Microbe Interactions 22, 747-757.

Deaker R, Roughley RJ, and Kennedy IR (2004). Legume seed inoculation technology – a review. Soil Biology and Biochemistry 36, 1275–1288.

Desbrosses GJ and Stougaard J (2011). Root nodulation: a paradigm for how plant–microbe symbiosis influences host developmental pathways. Cell Host and Microbe 10, 348-358.

Dobbelaere S, Croonenborghs A, Thys A, Vande Broek A, and Vanderleyden J (1999). Phytostimulatory effect of *Azospirillum brasilense* wild type and mutant strains altered in IAA production on wheat. Plant Soil 212, 153-162.

de Fritas JR, Banerjee MR, and Germida JJ (1997). Phosphate-solubilizing rhizobacteria enhance the growth and yield but not phosphorus uptake of canola (*Brassica napus* L.). Biology and Fertility of Soils 24:4, 358-364.

Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, Owens S, Gilbert JA, Wall DH, and Caporaso JG (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. PNAS 109:52, 21390-21395.

Finch-Boekweg H, Allen P, and Meyer S (2013). Exposure to Low Water Potentials and Seed Dormancy Favor the Fungus in the *Pyrenophora semeniperda-Bromus tectorum* Pathosystem. Plant Protection Science 49, S15-S20.

Finkel OM, Castrillo G, Paredes SH, Gonzalez IS, and Dangl JL (2017). Understanding and exploiting plant beneficial microbes. Current Opinion in Plant Biology 38, 155-163.

Geisen S, Kostenko O, Cnossen MC, ten Hooven FC, Vres B, and van der Putten WM (2017). Seed and Root Endophytic Fungi in a Range Expanding and a Related Plant Species. Frontiers in Microbiology 8:1645, 1-11.

Gemell LG, Hartley EJ, and Herridge DF (2005). Point-of-sale evaluation of preinoculated and custom-inoculated pasture legume seed. Australian Journal of Experimental Agriculture 45, 161–169.

Glick BR (2012). Plant Growth-Promoting Bacteria: Mechanisms and Applications. Scientifica, 1-15.

Goggin DE, Emery RJ, Kurepin LV, and Powles SB (2015). A potential role for endogenous microflora in dormancy release, cytokinin metabolism and the response to fluridone in *lolium rigidum* seeds. Annals of Botany 115, 293-301.

Gundel PE, Rudgers JA, and Ghersa CM (2011). Incorporating the process of vertical transmission into understanding of host-symbiont dynamics. Oikos 120, 1121-1128

Hacquard S (2016). Disentangling the factors shaping microbiota composition across the plant holobiont. New Phytologist 209, 454-457.

Hallmann J, Quadt-Hallmann A, Mahaffee WF, and Kloepper JW (1997). Bacterial endophytes in agricultural crops. Canadian Journal of Microbiology 43:10, 895-914.

Hardiom PR, Hardiom CCP, van Overbeek LS, and van Elsas JD (2012). Dynamics of Seed-Borne Rice Endophytes on Early Plant Growth Stages. PLoS ONE 7:2, 1-13.

Hartley EJ, Gemell LG, and Deaker R (2012). Some factors that contribute to poor survival of rhizobia on preinoculated legume seed. Crop Pasture Science 63, 858–865.

Hartmann M, Frey B, Mayer J, Mader P, and Widmer F (2015). Distinct soil microbial diversity under long-term organic and conventional farming. ISME Journal 9:5, 1177-1194.

Hill NS and Roach PK (2009). Endophyte Survival during Seed Storage: Endophyte-Host Interactions and Heritability. Crop Science 49:3, 1425-1430.

Hong M, Peng G, Keyhani NO, and Xia Y (2017). Application of the entomogenous fungus, *Metarhizium anisopliae* for leafroller (*Cnaphalocrocis medinalis*) control and its effect on rice phyllosphere microbial diversity. Applied Microbiology Biotechnology, 1-15.

Huang Y, Kuang Z, Wang W, and Cao L (2016). Exploring potential bacterial and fungal biocontrol agents transmitted from seeds to sprouts of wheat. Biological Control 98, 27-33.

Huse SM, Ye Y, Zhou Y, and Fodor AA (2012). A Core Human Microbiome as Viewed through 16S rRNA Sequence Clusters. PLoS One 7:6, 1-12.

Johnston-Monje D and Raizada MN (2014). Impact of swapping soils on the endophytic bacterial communities of pre-domesticated, ancient and modern maize. BMC Plant Biology 14:233, 1-19.

Junker RR, Loewel C, Gross R, Dotterl S, Keller A, and Bluthgen N (2011). Composition of epiphytic bacterial communities differs on petals and leaves. Plant Biology 13, 918-924.

Junker RR and Keller A (2015). Microhabitat heterogeneity across leaves and flower organs promotes bacterial diversity. FEMS Microbiology Ecology 91:9, 1-9.

Kepler RM, Maul JE, and Rehner SA (2017). Managing the plant microbiome for biocontrol fungi: examples from Hypocreales. Current Opinion in Microbiology 37, 48-53.

Khalid A, Arshad M, and Zahir ZA (2004). Screening plant growth-promoting rhizobacteria for improving growth and yield of wheat. Journal of Applied Microbiology 96:3, 473-480.

Klaedtke S, Jacques MA, Raggi L, Preveaux A, Bonneau S, Negri V, Chable V, and Barret M (2016). Terroir is a key driver of seed-associated microbial assemblages. Environmental Microbiology 18:6, 1792-1804.

Lebeis SL, Rott M, Dangl JL, Schulze-Lefert P (2012). Culturing a plant microbiome community at the cross-Rhodes. New Phytologist 196, 341-344.

Lehr P (2010). Biopesticides: The Global Market. Report code CHM029B, BCC Research, Wellesley, Massachusetts.

Lemanceau P, Blouin M, Muller D, and Moenne-Loccoz Y (2017). Let the Core Microbiota Be Functional. Trends in Plant Science 22:7, 583-595.

Lindemann J and Upper CD (1985). Aerial Dispersal of Epiphytic Bacteria over Bean Plants. Environmental Microbiology 50:5, 1229-1232.

Links Mg, Demeke T, Grafenhan T, Hill JE, Hemmingsen SM, and Dumonceaux TJ (2014). (2014). Simultaneous profiling of seed-associated bacteria and fungi reveals antagonistic interactions between microorganisms within a shared epiphytic microbiome on *Triticum* and *Brassica* seeds. New Phytologist 202, 542-553.

Lubeck M, Knudsen IMB, Jensen B, Thrane U, Janvier C, and Funck Jensen D (2002). GUS and GFP transformation of the biocontrol strain *Clonostachys rosea* IK726 and the use of these marker genes in ecological studies. Mycology Research 106:7, 815-826.

Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, Tremblay J, Engelbrektson A, Kunin V, del Rio TG, Edgar RC, Eickhorst T, Ley RE, Hugenholtz P, Tring SG, and Dangl JL (2014). Defining the core *Arabidopsis thaliana* root microbiome. Nature 488:7409, 86-90.

Magarey RC (1999). Reduced productivity in long term monoculture: where are we placed? Australasian Plant Pathology 28, 11-20.

Malfanova N, Lugtenberg BJJ, and Berg G (2013). Bacterial Endophytes: Who and Where, and What are they doing there? Hoboken, NJ, USA: Wiley-Blackwell, 391-403.

Malik KA, Mirza MS, Hassan U, Mehnaz S, Rasul G, Haurat J, Bally R, and Normand P (2002). The role of plant-associated beneficial bacteria in rice-wheat cropping system. 73-83.

Meier U (2001). Growth Stages of Mono- and Dicotyledonous plants. 2nd Edn. Braunschweig: Federal Biological Research Center for Agriculture and Forestry.

Mitter B, Pfaffenbichler N, Flavell R, Compant S, Antonielli L, Petric A, Berninger T, Naveed M, Sheibani-Tezerji R, von Maltzahn G, and Sessitsch A (2017). A New Approach to Modify Plant Microbiomes and Traits by Introducing Beneficial Bacteria at Flowering into Progeny Seeds. Frontiers in Microbiology 8:11, 1-10.

Montesinos E (2003). Development, registration and commercialization of microbial pesticides for plant protection. International Microbiology 6:4, 245-252.

Munkvold GP (2009). Seed Pathology Progress in Academia and Industry. Annual Review of Phytopathology 47, 285-311.

Naeem S, Loreau M, and Inchausti P (2002). Biodiversity and Ecosystem Functioning: the emergence of a synthetic ecological framework. In: Loreau M, Naeem S, Inchausti P, eds.Biodiversity and ecosystem functioning: synthesis and perspectives, 1st edn. Oxford, UK: Oxford University Press, 3–11.

Nelson EB (2004). Microbial Dynamics and Interactions in the Spermosphere. Annual Review of Phytopathology 42, 271-309.

Nelson EB (2017). The seed microbiome: Origins, interactions, and impacts. Plant Soil, 1-28.

Ngugi HK and Scherm H (2006). Biology of flower-infecting fungi. Annual Review of Phytopathology 44, 261-282.

O'Callaghan M (2016). Microbial inoculation of seed for improved crop performance: issues and opportunities. Applied Microbial Biotechnology, 1-18.

Okunishi S, Sako K, Mano H, Imamura A, and Morisaki H (2005). Bacterial flora of endophytes in the maturing seed of cultivated rice (Oryza sativa). Microbes Environment 20, 168-177.

Parnell JJ, Berka R, Young HA, Sturino JM, Kang Y, Barnhart DM, and DiLeo MV (2016). From the Lab to the Farm: An Industrial Perspective of Plant Beneficial Microorganisms. Frontiers in Plant Science 7:1110, 1-12.

Pieterse CMJ, Zamioudis C, Berendsen RL, Weller DM, Van Wees SCM, Baaker PAHM (2014). Induced Systemic Resistance by Beneficial Microbes. Annual Review of Phytopathology 52, 347-375.

Pikaar I, Matassa S, Rabaey K, Bodirsky BL, Popp A, Herrero M, and Verstraete W (2017). Microbes and the Next Nitrogen Revolution. Environmental Science & Technology 51:13, 7297-7303.

Pirttila AM, Laukkanen H, Pospiech H, Myllya R, and Hohtola A (2000). Detection of Intracellular Bacteria in the Buds of Scotch Pine (*Pinus sylvestris* L.) by In Situ Hybridization. Applied Environmental Microbiology 66, 3073-3077.

Rijavec T, Lapanje A, Dermastia M, and Rupnik M (2007). Isolation of bacterial endophytes from germinated maize kernels. Canadian Journal of Microbiology 3, 802-808.

Robertson GP, and Swinton SM (2005). Reconciling agricultural productivity and environmental integrity: a grand challenge for agriculture. Frontiers in Ecology and the Environment 3:1, 38-46.

Robinson RJ, Fraaije BA, Clark IM, Jackson RW, Hirsch PR, and Mauchline TH (2016). Wheat seed embryo excision enables the creation of axenic seedlings and Koch's postulates testing of putative bacterial endophytes. Nature Scientific Reports 6, 1-9.

Rosas SB, Andres JA, Rovera M, and Correa NS (2006). Phosphate-solubilizing *Pseudomonas putida* can influence the rhizobia-legume symbiosis. Soil Biology and Biochemistry 38:12, 3502-3505.

Saa S, Rio AO, Castro S, and Brown PH (2015). Foliar application of microbial and plant based biostimulants increases growth and potassium uptake in almond (*Prunus dulcis* [Mill.] D. A. Webb). Frontiers of Plant Science 6:87, 1-9.

Sanchez-Canizares C, Jorrin B, Poole PS, and Tkacz A (2017). Understanding the holobiont: the interdependence of plants and their microbiome. Current Opinion in Microbiology 38, 188-196.

Schuster A, and Schmoll M (2010). Biology and biotechnology of Trichoderma. Applied Microbiology and Biotechnology 87, 787–799.

Sedzielewska Toro K and Brachmann A (2016). The effector candidate repertoire of the arbuscular mycorrhizal fungus *Rhizophagus clarus*. BMC Genomics 17:101, 1-13.

Shade A, Jacques M, and Barret M (2017). Ecological patterns of seed microbiome diversity, transmission, and assembly. Current Opinion in Microbiology 35, 15-22.

Sharma KK, Singh US, Sharma P, Kumar A, and Sharma L (2015). Seed treatments for sustainable agriculture-a review. Journal of Applied Natural Science 7, 521-539.

Steinauer K, Chatzinotas A, and Eisenhauer N (2016). Root exudate cocktails: the link between plant diversity and soil microorganisms? Ecology and Evolution 6, 7387-7396.

Slykhuis JT (1947). Studies on *Fusarium culmorum* blight of crested wheat and brome grass seedlings. Canadian Journal of Forest Research 25, 155-180.

Timmusk S, Behers L, Muthoni J, Muraya A, and Aronsson A (2017). Perspectives and Challenges of Microbial Application for Crop Improvement. Frontiers in Plant Science 8, 1-49.

Truyens S, Weyens N, Cuypers A, and Vangronsveld J (2015). Bacterial seed endophytes: genera, vertical transmission and interaction with plants. Environmental Microbiology Reports 7:1, 40-50.

Tucker CM, Shoemaker LG, Davies KF, Nemergut DR, and Melbourne BA (2016). Differentiating between niche and neutral assembly in metacommunities using null models of B-diversity. Oikos 125, 778-789.

Ushio M, Yamasaki E, Takasu H, Nagano AJ, Fujinaga S, Honjo MN, Ikemoto M, Sakai S, and Kudoh H (2016). Microbial communities on flower surfaces act as signatures of pollinator visitation. Scientific Reports 5:8695, 1-7.

Van Elsas JD and van Overbeek LS (1993). Bacterial responses to soil stimuli. Starvation in Bacteria, S. Kjelleberg (ed.), 55-79, Plenum Press, New York.

Vandenkoornhuyse P, Quaiser A, Duhamel M, Le Van A, and Dufresne A (2015). The importance of the microbiome of the plant holobiont. New Phytologist 206, 1196-1206.

Vogel C, Bodenhausen N, Gruissem W, and Vorholt JA (2016). The Arabidopsis leaf transcriptome reveals distinct but also overlapping responses to colonization by phyllosphere commensals and pathogen infection with impact on plant health. New Phytologist 212, 192-207.

Wagner MR, Lundberg DS, del Rio TG, Tringe SG, Dangl JL, and Mitchell-Olds T (2016). Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. Nature Communications 7, 1-15.

Wallace RH and Lochhead AG (1951). Bacteria Associated With Seeds Of Various Crop Plants. Soil Science 71:3, 159-166.

Weidner S, Koller R, Latz E, Kowalchuk G, Bonkowski M, Scheu S, and Jousset A (2015). Bacterial diversity amplifies nutrient-based plant-soil feedbacks. Functional Ecology 29, 1341-1349.

Wilkinson DM, Koumoutsaris S, Mitchell EAD, and Bey I (2012). Modelling the effect of size on the aerial dispersal of microorganisms. Journal of Biogeography 39, 89-97.

Won H and Renner SS (2003). Horizontal gene transfer from flowering plants to Gnetum. PNAS 100:19, 10824-10829.

Yeoh YK, Dennis PG, Paungfoo-Lonhienne C, Weber L, Brackin R, Ragan MA, Schmidt S, and Hugenholtz P (2017). Evolutionary conservation of a core root microbiome across plant phyla along a tropical soil chronosequence. Nature Communications 8:215, 1-9.

Zhao YP, Lin S, Chu L, Gao J, Azeem S, and Lin W (2016). Insight into structure dynamics of soil microbiota mediated by the richness of replanted *Pseudostellaria heterophylla*. Scientific Reports 6, 1-9.

Zhou J, Deng B, Zhang Y, Cobb AB, and Zhang Z (2017). Molybdate in Rhizobial Seed-Coat Formulations Improves the Production and Nodulation of Alfalfa. PLoS ONE 12:1, 1-14.

CHAPTER 2

Identification and Characterization of the Core Rice Seed Microbiome

Alexander W. Eyre

Fungal Genomics Laboratory, Center for Integrated Fungal Research, Department of Entomology and Plant Pathology, North Carolina State University, Raleigh NC, 27606, USA

Abstract

The utilization of microbes in agriculture is an emerging alternative to the usage of chemical fertilizers and pesticides to increase plant yield and other beneficial characteristics. In this study, the microbiomes of six different rice (*Oryza sativa*) seeds were characterized. Rice seed was sourced from two locations in Arkansas, USA using two different genotypes and two harvest years for microbiome comparison. The bacterial and fungal communities were identified in each of four seed compartments (grain, outer grain, husk, and outer husk) using high throughput Illumina MiSeq 300bp paired end sequencing then analyzed using QIIME. More unique OTUs with greater diversity were identified in the outer seed husk and least in the grain compartment for both the fungal and bacterial microbiomes. Principal component analysis indicated that each tissue compartment harbored relatively distinct bacterial and fungal communities, more so for the bacteria. A bacterial and fungal core microbiome shared among the six seed types for each compartment was identified. Key bacterial genera in the core were *Enterobacter*, *Pantoa*, *Sphingomonas*, and *Paenibacillus*, members of which have been reported to support rice growth. Far fewer fungi were part of the core thought to be due to the large amount of variability between seed types, however key species found were *Alternaria longissima* and *Cladosporium delicatulum*. These core members represent valuable candidates for manipulating the rice microbiome, decreasing the use of chemicals while increasing plant performance.

Introduction

In 1909, the Haber-Bosch process was introduced to the world, enabling the production of ammonia from atmospheric nitrogen on an industrial scale. This technology allowed for the creation of some of the first synthetic fertilizers and the eventual development of agriculture into the industry it is today (Pikaar et al., 2017). Since then, these chemical fertilizers have been added to and modified for just about every crop application. With the boom of large monoculture cropping systems also came the need for pesticides to control the various pathogens (Oerke, 2006). Today, chemical products continue to increase in demand as the global population grows, but many suffer from problems such as limited effectiveness, low plant bioavailability, and widespread environmental pollution (Carvalho, 2006; Aktar et al., 2009).

One emerging alternative of interest has been the exploitation of the plant microbiome, or collection of bacteria, fungi, oomycetes, and viruses that inhabit plant tissues (Finkel et al. 2017). Recent research, particularly into the bacterial and fungal members of these communities, has led to the development of microbial applications that fulfill the same role of the chemicals employed, although through different mechanisms (Busby et al., 2017). This has been realized by the agricultural industry, as the market for plant growth promoting (PGP) products has been growing by roughly 12% annually (Calvo et al., 2014). However, microbials often run into issues similar to their chemical counterparts, such as limited effectiveness due to environment fluctuations and runoff outside the crop system where they do not function. The impact of this microbial runoff into the environment surrounding a crop remains largely unexplored, except in a few instances, and it remains unclear whether these may act as pollutants (Joung et al., 2000). Regardless, microbial products have often been sufficient to increase valuable crop characteristics making them an economically feasible method of increasing yield (Esiken et al., 2006; Saa et al., 2015)

A recently realized alternative that may negate issues associated with microbial applications has been the manipulation of the plant seed microbiome. Studies have shown that

microbes contained in the endophyte, or interior, of the seed are later found in the endophyte of the mature plant (Nelson et al., 2004; Darrasse et al., 2010). Therefore, direct manipulation of the seed microbiome may offer a unique avenue for shaping of the plant microbiome to provide a lasting, chemical-free means of plant growth promotion or pest resistance (Nelson et al., 2017). Several factors affect the landscape of the plant microbial community such as the soil composition, plant genotype, and environmental fluctuations, which often severely limit the geographical regions in which microbial applications are effective (Peiffer et al., 2013; Barret et al., 2015). Manipulation of microorganisms that are associated with the seed may circumvent some of these issues, however there is a need to identify optimal candidates particularly those that are part of the core seed microbiome. An ideal plant core microbiome is defined as the microbial community shared between several plant genotypes grown in various geographical areas that contain different soil compositions and environmental conditions (Lundberg et al., 2012; Vandenkoornhuyse et al., 2015). It is likely that a number of these core microbes are maintained for their beneficial properties, and as such it is not unreasonable to presume that their manipulation may lead to enhanced improvement in plant performance.

To begin such an endeavor, we deduced a preliminary core seed microbiome for rice, or *Oryza sativa*. Rice is among the most important staple crops for much of the world and for which substantial microbiome and genomic data already exists (Kawahara et al., 2013; Edwards et al., 2015; Breidenbach et al., 2016). Therefore, introduction of seed microbiome data and the development of a core microbiome will expand upon this resource and may provide valuable insight to future microbe-related endeavors. In this study, we sourced six related rice seeds of two different genotypes from two separate years and locations in Arkansas, USA. Each seed was separated into four distinct tissue compartments (grain, outer grain, husk, inner husk), the bacterial and fungal members identified for each, and additional analyses of population richness and structure were performed from which a core microbiome was deduced.

Results

Defining the rice seed microbiome

In this study, rice seed was obtained from two rice research plots in Arkansas, USA. Seed from two different genotypes were selected with variable resistance to the fungal rice pathogen, *Magnaporthe oryzae*: Katy (generally resistant) and M202 (highly susceptible). Additionally, seeds from two different research field locations and two separate harvest years were used for a total of six different seed lots (**Table 1**). The seeds themselves were divided into four distinct compartments to examine microbial trends within and between various seed tissues (**Figure 1**). The outer husk compartment was collected as a thorough wash of the raw seed. The seed was subsequently husked, both the rice husk and grain were thoroughly washed, and samples pooled to form the outer grain compartment. The washed husk and the grain were then collected and pulverized to constitute the husk and grain compartments. The grain itself contained the bran, endosperm, and the germ tissues of the rice seed, negating any effects on community structure that these may provide. In total, 24 samples for microbiome analysis were produced among the six seed types.

From these samples, the v3-v4 region of the bacterial 16S ribosomal gene and ITS1 region of the fungal genome were PCR amplified using primers designed from current literature (**Table 2**) including Illumina sequencing adapters. Bacterial and fungal amplifications were performed separately, quantified, pooled for each sample, then sequenced. Illumina MiSeq 300bp paired end sequencing produced 18.70M raw reads, which yielded 4.46M joined and quality filtered reads (**Table 3**). Of the quality reads, 1.11M belonged to the bacterial and 3.35M belonged to the fungal community. The disparity in the number of reads may reflect the longer length of the bacterial v3-v4 region hence lower quality of the overlapping sequences, forcing the joining process to discard more of the read pairs.

The most well-manicured databases were sourced for OTU picking: the SILVA bacterial 16S rRNA v128 database and the UNITE fungal ITS v7.1 database. The reads and databases

were imported for use in QIIME (Caporaso et al., 2010), open reference OTU picking performed, and any sequences corresponding to contaminants were removed. One sample from each of the bacterial (seed C: outer husk) and fungal (seed F: grain) microbiomes did not have enough reads to meet the cutoff for the analyses due to the overabundance of plant contaminant sequences in the bacterial sample and lack of reads in the fungal sample. Analysis of the OTUs was performed in QIIME and α-diversity rarefaction plots indicated the sampling depth reached saturation, suggesting each sample is evenly represented. The databases utilizing a 97% similarity cutoff were opted for over the 99% databases, as the later produced a large amount of variability in taxonomic classification. Overall, most bacterial v3-v4 sequences were able to be resolved to the genera level, while the fungal ITS1 sequences were a mix of various taxonomic levels.

Rice Seed Microbe Members and Trends

To deduce the microbial members of the seeds, QIIME core diversity analyses were employed on the OTU datasets for each of the 24 seed samples. Overall, a total of 2,718 unique bacterial OTUs (**Figure 2**) and 828 unique fungal OTUs (**Figure 3**) were identified as having at least one read between all samples. For both bacteria and fungi, several similar trends were observed. In both instances, the outer husk contained a majority of the unique OTUs, 81.6% for the bacteria and 76.0% for the fungi. In contrast, the grain contained the fewest, 15.9% for the bacterial and 21.5% for the fungal OTUs. Of the bacterial members, approximately 27.9% were found in a single seed compartment with 72.1% being shared with at least one other, and 7.4% shared between all compartments. Of the fungal OTUs, approximately 31.5% were found in a single compartment with 68.5% being shared with at least one other, and 13.8% were shared between all compartments. In addition, for the grain compartment OTUs, approximately half (46.6% and 64.6% for the bacteria and fungi respectively) were shared with all other compartments and contained the fewest number of unique OTUs.

Analysis of the bacterial v3v4 reads yielded 185 taxonomic classifications predominately at the genus level. To summarize the distribution of bacterial taxa across the four seed compartments, taxa ranks with at least 1% of the total reads within all samples were examined of which there were 10 that contained 89.59% of the total reads. Figure 4 shows the distribution of reads and the proportion of distinct OTUs in each of the taxonomic groups identified (**Figure 4**). Based on read abundance, the *Gammaproteobacteria* class was found to be most abundant between all seed compartments ranging from 62.8% to 99.8%, while the *Alphaproteobacteria* class were most abundant in the outer seed tissues (36.8%) and were nearly absent from the grain (0.2%). The most apparent trend between the compartments was the reduction of diversity from the outer husk (1357 OTUs) to the interior of the grain (328 OTUs), particularly a near loss of members of the *Alphaproteobacteria* class, *Chryseobacterium*, and *Xanthomonas*. Interestingly, *Chryseobacterium* was primarily found only on surfaces of the husk and grain, suggesting it might not be able to infiltrate the tissues. This is consistent with studies demonstrating its presence and travel through the plant xylem and might indicate a mechanism for finding its way into the seed during development (Achari et al., 2014).

To get a more in-depth view of the bacterial members, those taxa that were present with greater than 0.1% of the total reads were examined (**Table 4**). To identify microbial distribution patterns across seed compartments, the extracted data was normalized, subjected to k-means clustering, and graphically visualized (**Figure 5**). OTUs assigned to clusters A and B were found predominately in the grain of the seed and contained members of the *Enterobacteriaceae* family that typically occupy inner plant tissues (Hardiom et al., 2013). Conversely, OTUs found in clusters D and G appeared mainly in or on the husk of the seed and contained members of the *Alphaproteobacteria* class and *Actinobacteria* phylum. Like *Chryseobacterium* previously described, a few other species follow the same surface patterning in cluster E, species of the respective genera have been described as being xylem-bound (Achari et al., 2014). Cluster C was nearly exclusively found on the epiphyte of the seed and contained the *Mucilaginibacter*

and *Siphonobacter* genera. Such OTUs may not be closely associated with the seed and may represent contaminants.

Analysis of the fungal ITS1 reads yielded 136 taxonomic classifications of fungi ranging from the phylum to the genus level. Fungal taxa that contained at least 1% of the total fungal reads were extracted of which thirteen were identified as containing 94.60% of the total reads (**Figure 6**). Like the bacterial taxa distribution, the grain compartment was largely comprised of a single taxa class, the unidentified *Dothideomycetes*. Largely absent from the grain were five *Basidiomycete* genera, particularly found in the seed surface compartments where they represented >50% of the total read abundance. This observation may be due to these genera mainly being isolated in unicellular yeast forms in nature (Nakase et al., 1985; Garcia et al., 2010; Nutaratat et al., 2014), forcing colonization of the interstitial spaces. Finally, there were unidentified fungal members found in greater abundance in the surface tissue compartments, although limited information can be deduced from these without the use of additional loci.

The fungi representing at least 0.1% of the total reads within respective taxa were extracted (**Table 5**) and clusters constructed (**Figure 7**) as above. OTUs assigned to fungal clusters A and D were among the most extreme, primarily confined to the outer grain and husk compartments respectively, although they contained the single rather low abundance members *Ustilaginaceae* and *Naganishia*. Cluster B exhibited a similar extreme for the outer husk tissue but contained five members with *Cladosporium* and *Occultifur* in highest abundance. Like bacterial cluster C, the fungal OTUs represented within this cluster may not be closely associated with rice seed and may instead be contaminants (Shade et al., 2017). Cluster C contained the largest number of fungal members whose abundance increased from the interior to the exterior of the seed, as is most common when exploring trends among plant tissues (Edwards et al., 2014; Sanchez-Canizares et al., 2017). In clusters E and F, no obvious trends in the represented fungi such as *Culvaria* and *Microdochium* were noticeable across compartments that might have explained the abundance in the respective tissues. More

interesting was the large differences in representation of the fungal members when analyzed based on the other seed factors. For example, *Pleosporales* was found throughout the seed, yet had a 22.95% difference in abundance between seed sourced from 2013 and 2014. Additional disparities are a 13.78% difference in abundance between locations for *Hanella*, and 12.46% difference in abundance between genotypes for *Papilotrema* (**Table 5**). Such differences in representation occurred quite frequently, especially when considering the seed year, although contributions to this effect were unknown.

Microbial Diversity

To evaluate if the microbiomes formed distinct communities when pooled depending on the four factors, a principal component analysis (PCA) was first generated for the 24 samples for both the bacterial and fungal members as part of the core QIIME analyses. The resulting analyses (**Figure 8** & **9**) revealed that the samples when grouped by seed compartment were relatively distinct, although this was much more apparent for the bacterial rather than fungal community. For the bacterial PCA, the grain and husk tissues formed the most distinct clusters, while the outer husk and outer grain had some overlap. For the fungal PCA, the three innermost tissues exhibited some overlap, while the outer grain was most distinct. When the results from the PCA were compared based on genotype, location, or seed year, separation was only observed in the fungal year (**Appendix A**).

Two additional approaches were used to evaluate the seed microbiome diversity across seed compartment, year, genotype, and location utilizing packages in R. Shannon diversity, a measure of species diversity and evenness (Lemos et al., 2011), revealed an expected reduction in the diversity index from the outer to the innermost tissue compartments for both the bacteria and the fungi (**Table 6**). This result is consistent with the observation that microbial abundance and diversity tends to decrease from the epiphyte to the endophyte of most plant tissues, especially within the root (Edwards et al., 2014; Sanchez-Canizares et al., 2017). In relation to the seed, this might suggest the presence of some type of selective mechanism that

limits the occupancy by certain microbes. Whether this mechanism is imposed by the biology of the microbes or selection of the plant remains unclear. For the bacterial members, there did not seem to be a large impact of factors other than tissue, with the location having a modest impact on diversity. On the other hand, for the fungal members, genotype, location, and year all affected diversity to some noticeable extent.

To determine if any of these factors had a statistically significant impact on diversity, an analysis of similarities (AnoSim) analysis was employed. This analysis compares the variation in species abundance and composition between sampling units pooled by their respective factors (Xia et al., 2017). The results indicated that tissue compartment was the primary driver of both bacterial and fungal community composition, explaining 51.3% and 30.4% of the dissimilarity between the samples, respectively (**Table 7**). Interestingly, the seed year factor also showed significance for the fungal community explaining 35.4% of the dissimilarity between samples, more than compartment alone. This result is consistent with the PCA and might largely be due to the lack of seed from the UA research station for the year 2013, although this effect is only apparent for the fungi. When the samples were pooled by the combination of compartment with the other factors, they were all found to be significant suggesting that their role on microbial composition is a function of the seed compartment. The combination of compartment and genotype explained most of the bacterial dissimilarity between seeds at 54.7%. Expectedly, the combination of tissue and year explained the most fungal dissimilarity at 54.5%. The seed compartments were then separated and tested for each of the three remaining factors finding no significance, suggesting that the tissue compartment is the most important factor for microbial community in the seed.

Rice Seed Core Microbiome and Hypothetical Functions

A key motivator for this research was to elucidate a core microbiome for rice seed, a set of microbes that one could expect to find within seed independent of its genotype, harvest year, and location. To define the per-compartment seed core microbiome, the OTUs that were found

in each seed sample for a compartment were identified. Often in the deduction of a core microbiome, it is suggested to use a 90% threshold value for species presence among samples (Huse et al., 2012), however due to the limited number of samples and potential for expansion with new data, a threshold was not used. Both the bacterial core (**Table 8**) and the fungal core (**Table 9**) contained primarily members that were found in high abundance in previous analyses, but a few additional genera were identified. The unique OTUs identified in the bacterial core represented 5.0% of their total unique OTUs, while the fungal core represented 5.8% of their total unique OTUs. The microbes found in each of the seed types, occurring in less than 0.1% of the entire community, could be considered rare yet conserved members of the rice core microbiome.

Within the bacterial grain core, all taxa except for the *Enterobacter* genera were shared amongst the tissue compartments, which was present only in outer grain. This genus is of interest as it is typically confined to the endophyte of the rice plant and some species, such as *Enterobacter oryziphilus* and *oryzendophyticus*, have been noted to have rice growth promoting effects through the supply of nitrogen and phosphorous (Hardiom et al., 2013). At high abundance among all tissue compartments were a variety of *Pantoea*, *Sphingomonas*, *Pseudomonas*, and *Paenibacillus*. These genera are commonly found associated with rice plants and seed (Midha et al., 2016). Certain species of both *Sphingomonas* and *Pantoea* have valuable nitrogen fixing mechanisms, and the latter can further promote growth through the production of a variety of phytohormones (Xie et al., 2006; Videria et al. 2008; Feng et al., 2008). *Pseudomonas* is noted to both promote plant growth through the secretion of phytohormones and induce systemic resistance to a variety of other bacterial pathogens (Vleesschauwer et al., 2008). *Methylobacterium* is of interest to rice and the environment, as it able to both fix nitrogen and oxidize methane, which wetland rice is responsible for releasing ~20% of the world's content annually (Tani et al., 2015). Some genera contain species that are more associated with pathogenesis of rice such as *Xanthomonas*, *Pseudomonas*, and *Pantoea*,

although whether this is the role of the unidentified species identified here remains unknown. Many more of the genera do not yet have known functions to the plant or microbial community, although many have been recently cultured from various rice crops (Liu et al. 2010; Elbeltagy et al. 2001).

The fungal core rice microbiome contained a slightly greater fraction of the overall OTU members compared to the bacterial core, although many of its members were confined to single compartments. We expect that this number might decline had greater resolution of the taxonomic identifications been possible, as it is likely that many of the unidentified OTUs at the phylum or class level are comprised of multiple taxa. Regardless, several fungal genera and species were identified among the tissue compartments, most of whose functions are yet to be deduced. *Alternaria longissima* was one of the few species identified among the tissue compartments and has been previously reported in rice (Agarwal et al., 1975), and reported to produce possibly active metabolites (Wang et al., 2014). A closely related species and potential member of the genera classifications, *A. padwickii*, is a common leaf spot pathogen of rice (Gutierrez et al., 2009). *Cladosporium delicatulum* was also identified among the central compartments of the seed, which has been noted to have suppressing effects on rice blast disease when applied before or during infection by *Magnaporthe oryzae*, which causes losses of ~20 million tons of rice yield annually (Chaibub et al., 2016). For most of the remaining core fungal species, no function has been reported, with the exception that many Basidiomycetes around the grain and within the husk largely exist in the form of yeasts (Robert et al., 2005).

Discussion

In this study, six related rice seed types were used, divided into four different tissue compartments, their bacterial/fungal microbiomes sequenced, and analyzed via QIIME. Overall, 2,718 unique bacterial and 828 unique fungal OTUs were identified from the seed, which formed distinct communities when sorted by seed tissue compartment. Ten bacterial and thirteen fungal members constituted the taxa with greater than 1% of the total reads in the samples,

which experienced variable abundance trends among the seed tissues. For the bacterial communities, the seed compartments explained most of the dissimilarity in the samples, which improved with the inclusion of the rice genotypes. For the fungal communities, the seed year explained most of the dissimilarity in the samples, which improved with the addition of seed compartment. This effect may reflect the limited availability of seeds from the UA research station from the year 2013, but other soil, environmental, or fungal longevity factors could play a significant role. Finally, the core microbiome was deduced, which contained the microbial members shared by all samples in a tissue compartment. The bacterial core contained many genera, many with known plant growth promoting functions and others whose species have been cultured. The fungal core was largely comprised of genera documented with species that are yeasts or linked to pathogens with little indication as to what roles these may play for the plant or microbial community.

The abundance trends of many of these members have interesting implications for inheritance mechanisms. For example, Chryseobacterium, Paenibacillus, and others are primarily found in the interstitial compartments containing surfaces and spaces suggesting these might be transferred to the seed through the xylem or other plant interspaces. Other organisms with the ability to transverse through the plant tissue such as bacteria and filamentous fungi through the action of cell wall degrading enzymes (Aparna et al., 2009; Kubicek et al., 2014), might be able to penetrate the seed tissue to occupy the grain or husk compartments. Those without these abilities may find their way into the seed endophyte due to abundance in flowering structures during seed development. Further interrogation of inheritance mechanisms, however, is beyond the scope of this study and would require detailed cytological analyses using strains tagged with reporter genes (Lubeck et al., 2002). The core microbiome in this study provides several ideal candidates for such research.

Current literature was searched against the members of the core microbiome in search for documented cases of their association with rice plant or seed. While there were numerous

accounts of bacteria, the study of fungi as a member of plant microbiome is still in its infancy for many plant species and estimates could not be made. More importantly, the limited resolution for the unidentified taxonomic classifications of fungi makes this troublesome. This was due to the ITS1 region being assigned to two or more species' OTUs that belong to different families, orders, etc., which could be resolved further with the inclusion of the ITS2, and/or other regions in the future (Tedersoo et al., 2016). Regardless, review of literature for the bacterial core showed that approximately 91.7%, 78.0%, 71.4%, and 67.8% of the core microbiome from the inner to the outer tissue compartments have been reported as being associated with rice seed (Bertani et al., 2016; Midha et al., 2016). Additionally, many species of these reported genera have been cultured from various rice plants around the globe enabling their future investigation or modification.

The primary focus of this research was to identify the core microbiome members for rice seed, a group of microbes that appear to be ubiquitously present in seed regardless of the impact of seed genotype, year, and location. These microbes are of interest for the development of agricultural products, as they likely represent a conserved grouping that might thrive for multiple rice genotypes and over a large geographic area. This idea is important to the industry, as some microbial applications that have flourished during research do not function outside of conditions in which it was performed (Barret et al., 2015, Parnell et al., 2016). Additionally, members of the core might be more tightly associated with the plant and/or microbial community that, if modified, would increase the addition's stability or long-term association with the plant. This issue was demonstrated by a group who successfully introduced a novel plant growth promoting bacterium into a variety of crop seed and achieve increased yield in wheat, however the phenomena lasted a single generation due to the bacterium's insufficient stability as a member of the microbiome (Mitter et al., 2017).

The rice seed core microbiome presented in this study represents only a small subset of the world rice population. Additional cultivars would provide the depth of investigation needed

to create a useful resource for future research and product development. The primary theoretical issue with this pursuit is that as more plant genotypes or conditions are included in the study, the size of the core might approach zero. There are some workarounds such as relaxing the taxonomic classifications allowing for more individuals to be propagated into the core at a cost of resolution, or simply to construct the core and redefine it depending on the particular genotypic or geographic context of interest. Regardless, our work provides a preliminary evaluation into what such a community looks like for rice seed, and future work will help develop the core into a powerful tool for research.

Materials and Methods

Rice Seed and Experimental Design

Rice seed used in the study was sourced from Dr. Yulin Jia at the USDA Dale Bumpers National Rice Research Center. Six different seed types were utilized from two different rice genotypes (Katy and M202), two different years (2013 and 2014), and two different locations (Dale Bumpers and University of Arkansas). Unfortunately, 2013 seed from the University of Arkansas was not available. The samples were sent in enclosed envelopes containing 50g of seed through standard mail and stored dry at 4°C while unused.

Tissue Sample Collection and Preliminary Analysis

For sample collection, 80 rice seeds were selected for each type, placed into sterile 50 mL Falcon tubes with 15 mL PBS buffer, and vortexed for 2 minutes. The supernatant was collected, an additional 15 mL PBS buffer was added to the seed, sonication performed for 1 minute to remove tightly adhering microbes, and the resulting supernatant pooled with the vortexed aliquot to form the outer husk sample. Two additional 1-minute sonications were performed and supernatants discarded to rinse the husk. The husk was then removed from the grain with sterile tweezers in a flow hood and each collected in their own sterile 50 mL Falcon tubes to which 15 mL PBS buffer was added. Each tube was vortexed for 2 minutes, supernatant collected, additional 15 mL PBS buffer added, sonicated for 1 minute, and the

supernatant collected. All of the supernatant from the husk/grain washes were collected into a single Falcon tube. The husks and grains were stored and the samples in PBS buffer were centrifuged at 12,000 rpm for 15 minutes. The remaining tissues and pellets were stored at 4°C until DNA extraction (Bulgarelli et al., 2012; Bulgarelli et al., 2013; Lundberg et al., 2012).

Preliminary analyses were performed by performing a serial dilution of 1 mL of the supernatant that was collected from the outer husk and outer grain compartments. 10 husks and grains were then ground via sterile mortar and pestle, sterile water added, and 1 mL of these solutions serially diluted. Bacteria presence was confirmed on LB plates incubated at 37°C and fungi presence was confirmed on PDA plates incubated at 25°C.

DNA Extraction

The husk, grain, and pellets were placed into their own sterile mortar and pestle, liquid nitrogen was added, and the samples were thoroughly grinded. DNA was extracted from each sample using the Wizard® Genomic DNA Purification Kit by Promega following the provided instructions (Fadrosh et al., 2014). DNA quality and concentration were checked on a NanoDrop spectrophotometer.

16S V3-V4 and ITS1 PCR Amplification and Sequencing

Primers were developed for the bacterial 16S rRNA gene V3-V4 region and the fungal ITS1 region based off the most currently adopted primer sequences for both (**Table 2**) (Fadrosh et al., 2014; Walters et al., 2015). Overhang adapters were added to the primers for compatibility with the Nextera Illumina Index Kits. The functionality of the primers was first assessed using PCR whose products were run on 1% agarose gels in search for bands corresponding to the size of the microbial regions. Following confirmation, samples were PCR amplified, concentrations of DNA checked, and equal proportions of bacterial and fungal sequence were appropriately pooled for each of the samples. These were sent to the Genomic Sciences Laboratory at North Carolina State University for Illumina MiSeq 300bp paired end sequencing.

Data Processing & OTU Picking

Illumina raw paired end reads arrived from the sequencing service already demultiplexed, and their quality was assessed using FastQC v0.11.5 in search for passing sequencing and tile quality. The reads for each sample were joined then quality filtered in QIIME v1.9.0 using the provided Python scripts and the resulting sequences merged into one large fasta file (**Table 3**). The OTUs were selected using the pick_open_reference_otus.py script in QIIME with reverse strand matching enabled. The SILVA 16S rRNA version 128 database was used to pick the bacterial OTUs, utilizing the 97% database with 7 level taxonomy (Quast et al., 2013; Yilmaz et al. 2013). The UNITE ITS version 7.1 database was used to pick the fungal OTUs, also utilizing the 97% database (Abarenkov et al., 2010). The resulting OTUs from each organism type were filtered of contaminant sequences, rarefaction plots generated, and samples without enough reads for analysis were discarded. The sampling depth for the analyses were set to the highest number of reads without discarding any of the good samples. The resulting OTU and composition data was exported from QIIME for further analysis.

Data Visualization and Statistical Analyses

Venn diagrams were created using custom Python scripts that extracted OTU data, compared it between samples, and generated the figures utilizing the matplotlib_venn package. Bar charts were generated in R using the ggplot2 package and legends were customized in standard image processing utilities. Ecological analyses were performed in R utilizing the vegan package v 2.4-3 with default parameters (Oksanen et al., 2017). In order to assess the diversity within each compartment, the Shannon (H') diversity index was estimated using *diversity*. To identify the principal factors that influence microbial community composition, analysis of similarity (AnoSim) was used with 999 permutations unless otherwise adjusted. A distance matrix was generated from the data using *vegdist*, followed by computation of the statistical test with *anosim*. Clustering was performed in R using *vegdist* matrices and standard hierarchical clustering methods.

Figures

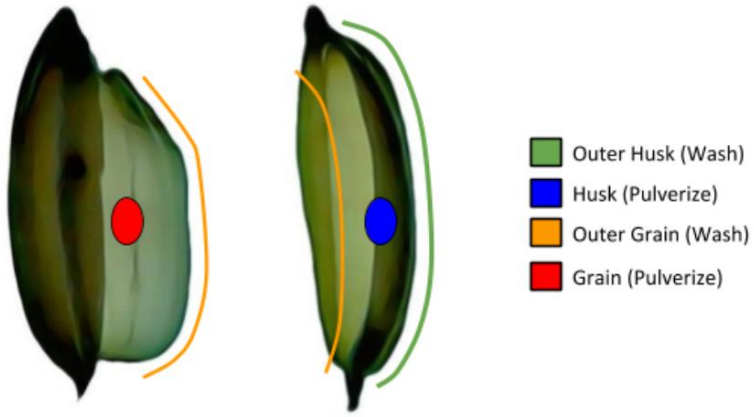

Figure 3.1 Seed Compartment Diagram: Illustration of the rice seed and its four tissue compartments for sample collection. Wash indicates the tissue sample was vortexed and sonicated to collect the microbes in solution. Pulverize indicates the tissue sample was frozen and crushed.

Compartment	Number Unique OTUs
Outer Husk	2219
Husk	1398
Outer Grain	1694
Grain	431
Total	2718

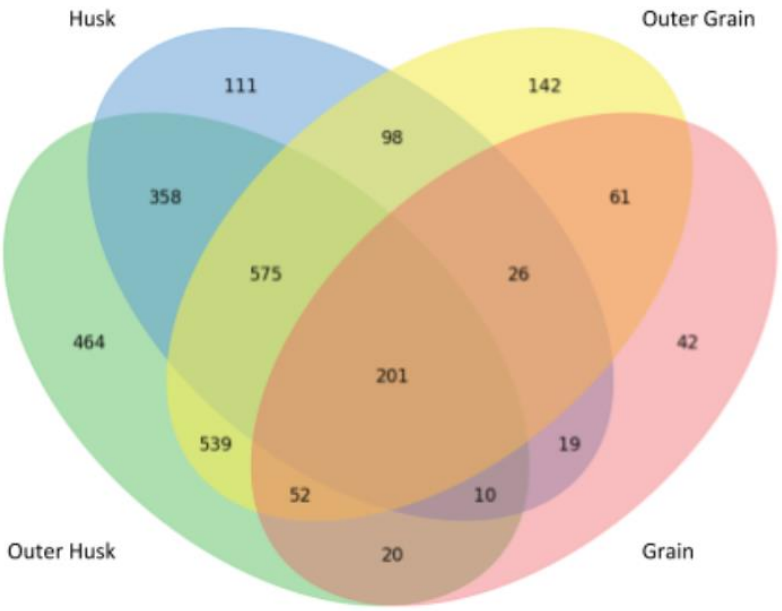

Figure 3.2 Bacterial Unique OTUs Venn Diagram: Distribution of unique bacterial OTUs separated by seed compartment. A value represents that at least one of the six samples belonging to a tissue compartment contain a unique OTU.

Compartment	Number Unique OTUs
Outer Husk	629
Husk	421
Outer Grain	592
Grain	178
Total	828

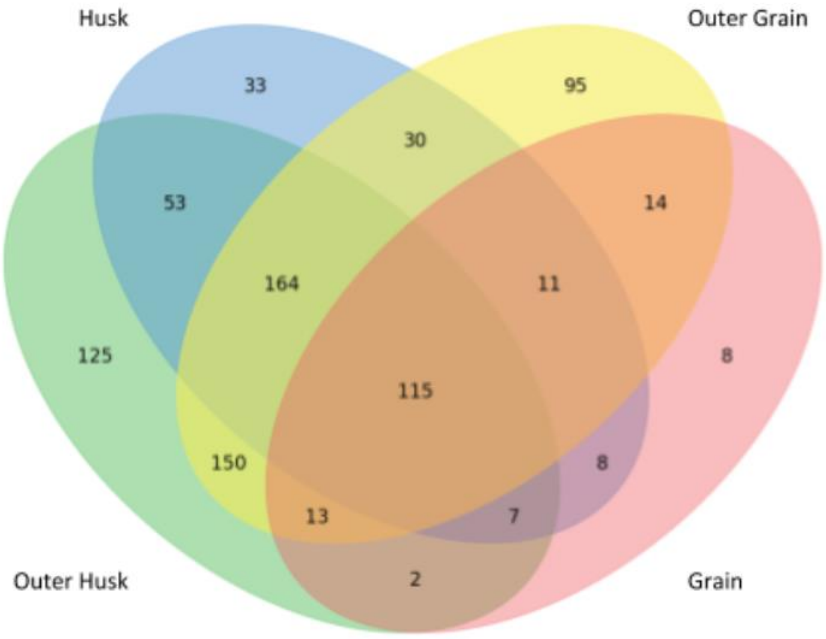

Figure 3.3 Fungal Venn Diagram: Distribution of unique fungal OTUs separated by seed compartment. A value represents that at least one of the six samples belonging to a tissue compartment contain a unique OTU.

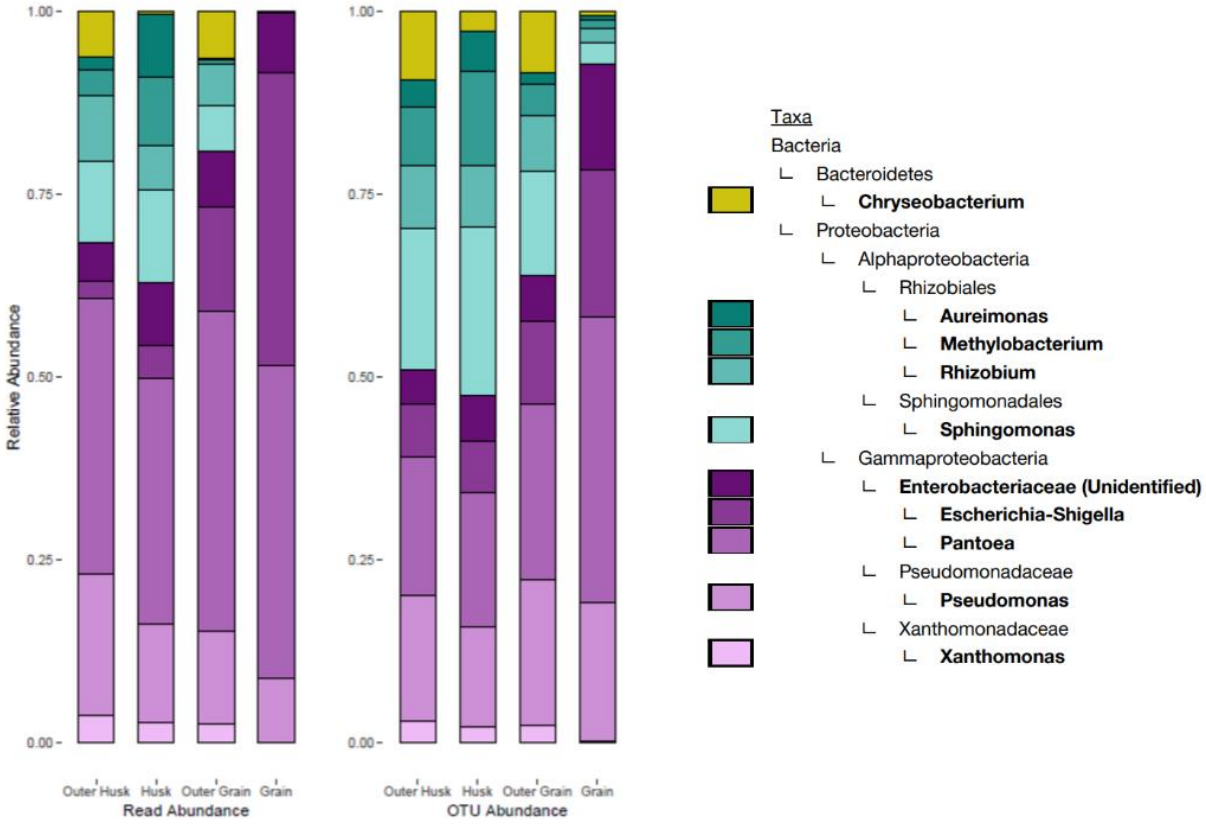

Figure 3.4 Bacterial Tissue Diversity: Relative abundance plots of the bacterial taxa with greater than 1.0% of total read abundance after being pooled based on seed compartment. Left displays the total read abundance belonging to the respective taxonomic classification. Right displays the abundance of unique OTUs belonging to the taxonomic classifications. The four tissue compartments contain 1357, 937, 1142, and 328 unique OTUs belonging to all taxa represented.

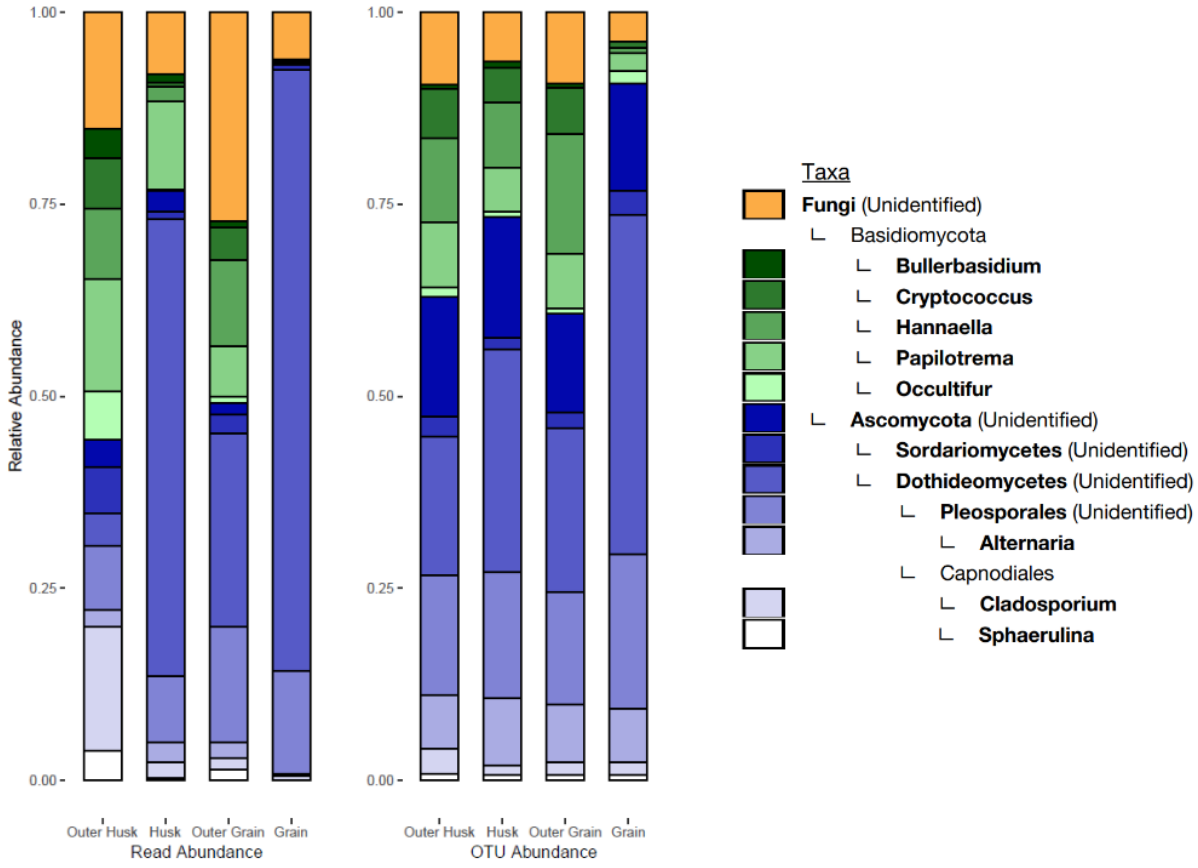

Figure 3.5 Fungal Tissue Diversity: Relative abundance plots of the fungial taxa with greater than 1.0% of total read abundance after being pooled based on seed compartment. Left displays the total read abundance belonging to the respective taxonomic classification. Right displays the abundance of unique OTUs belonging to the taxonomic classifications. The four tissue compartments contain 349, 262, 384, and 129 unique OTUs belonging to all taxa represented.

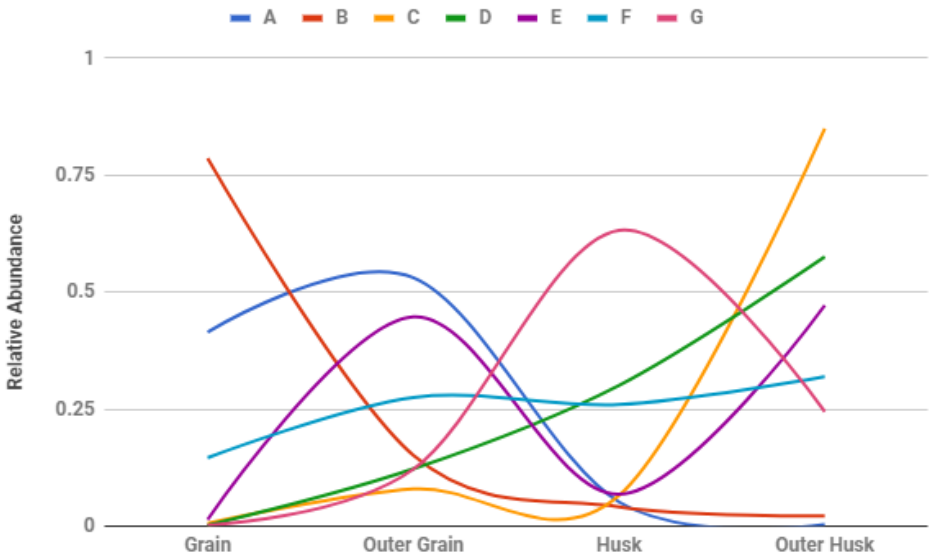

Figure 3.6 Bacterial Cluster Chart: Clustering of the normalized relative abundance values for all bacterial species with greater than 0.1% of the total reads. A line's values represent the average of the normalized abundance values within a cluster for each of the seed compartments.

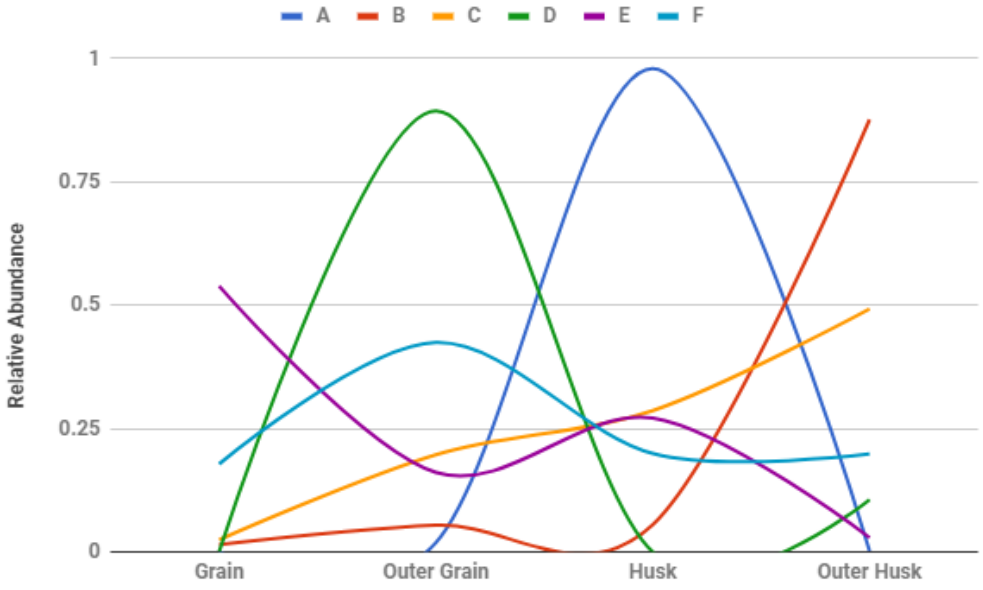

Figure 3.7 Fungal Cluster Chart: Clustering of the normalized relative abundance values for all fungal species with greater than 0.1% of the total reads. A line's values represent the average of the normalized abundance values within a cluster for each of the seed compartments.

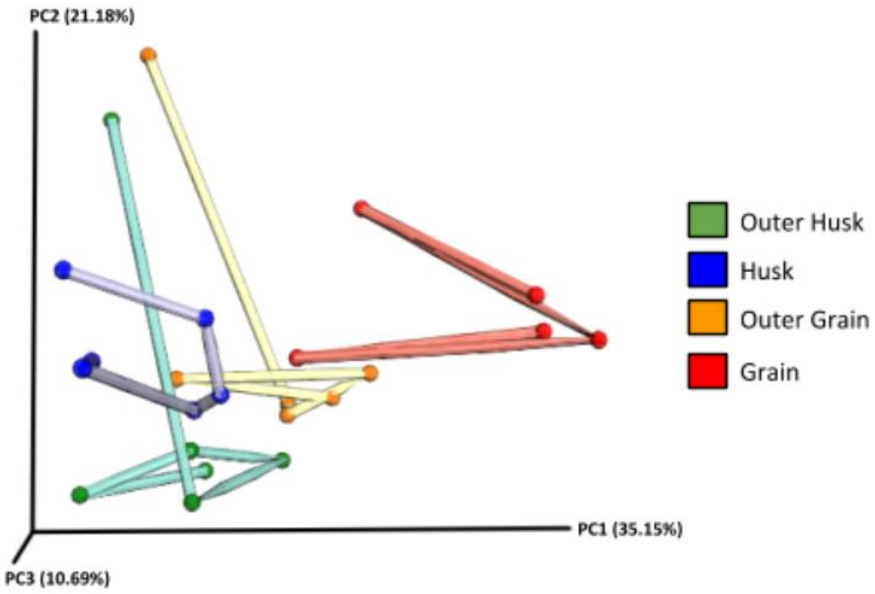

Figure 3.8 Bacterial PCA Analysis: Principal component analysis (PCA) performed on the 24 samples and colorized for seed compartment.

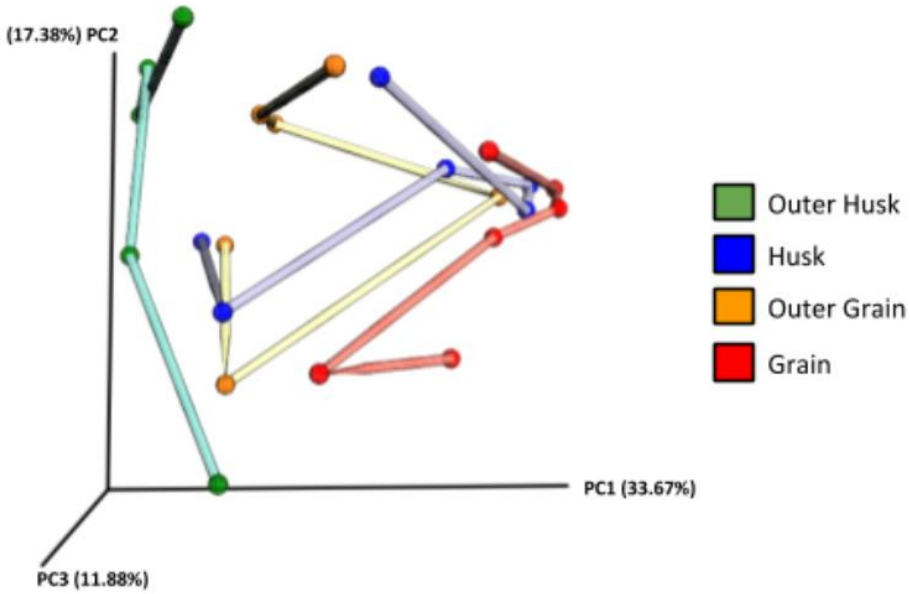

Figure 3.9 Fungal PCA Analysis: Principal component analysis (PCA) performed on the 24 samples and colorized for seed compartment.

Tables

Table 3.1 Rice Seed Information: The six rice seeds used in the study and their attributes.

Sample #	Genotype	Year	Location
A	Katy	2013	DB
B	Katy	2014	DB
C	M202	2014	UA
D	M202	2013	DB
E	Katy	2014	UA
F	M202	2014	DB

Table 3.2 Bacterial and Fungal PCR Primer Sequences: Primer sequences used for amplification of the bacterial and fungal communities.

Target	Forward Primer Sequence (5'-3')	Reverse Primer Sequence (5'-3')
Bacterial 16S v3-v4 Region	CCTACGGGNGGCWGCAG	GACTACHVGGGTATCTAATCC
Fungal ITS1 Region	CTTGGTCATTTAGAGGAAGTAA	GCTGCGTTCTTCATCGATGC

Table 3.3 Read Quality Information: Number of raw and post-processing reads separated by tissue compartment source. Each compartment contains reads from six samples. Quality checked read values are those before contaminant filtering.

Tissue Compartment	RAW	Joined, Quality Checked	
		Bacteria	Fungi
Grain	4,563,000	902,000	
Outer Grain	5,677,000	1,638,000	
Husk	3,757,000	533,000	
Outer Husk	4,699,000	1,386,000	
Total	18,697,000	1,110,000	3,349,000
Average per Sample	779,000	46,250	139,541

Table 3.4 Identified Bacteria: All bacterial OTUs with greater than 0.1% of the total reads sorted by their k-means clustering assignment. The headers represent the four seed factors. The values are the average relative abundance for the samples pooled based on the respective factor.

Taxonomic Classification	Cluster	Grain	Outer Grain	Husk	Outer Husk	2013	2014	DB	VA	Katy	M202
Klebsiella	A	1.02	1.31	0.13	0	0.3	0.65	0.15	1.03	1.07	0.15
Enterobacter	B	8.29	0.58	0.12	0.06	2.58	0.26	1.33	0.39	0.38	1.31
Escherichia-Shigella	B	33.12	12.09	3.51	1.84	11.43	6.8	11.87	3.54	14.16	3.32
Siphonobacter	C	0	0.12	0.02	0.63	0	0.44	0	0.71	0.35	0.3
Mucilaginibacter	C	0	0.34	0.45	4	0.88	2.33	0.68	3.43	0.77	2.84
Pedobacter	C	0	0.02	0	0.37	0	0.23	0	0.37	0	0.3
Brevundimonas	C	0	0.02	0.04	0.28	0.06	0.16	0.04	0.24	0.05	0.19
Microbacterium	D	0	0.04	0.13	0.26	0.15	0.14	0.11	0.19	0.05	0.22
Novosphingobium	D	0	0.06	0.15	0.27	0.2	0.15	0.16	0.18	0.12	0.21
Chryseobacterium	E	0	5.56	0.3	5.74	3.28	4.91	4.45	4.47	5.74	3.46
Sphingobacterium	E	0	1.67	0.04	1.67	0	1.81	0	2.88	0.89	1.64
Paenibacillus	E	0.33	2.45	0.64	2.73	2.08	2.18	2.19	2.11	1.98	2.29
Luteibacter	E	0	0.87	0.27	0.95	0.28	0.94	0.3	1.29	0.96	0.59
Rhizobium	F	0.03	4.82	4.76	7.43	6.47	5.23	4.75	6.56	6.02	5.23
Enterobacteriaceae	F	8.55	8.2	8.28	5.23	5.9	7.33	7.67	6.07	6.97	6.91
Pantoea	F	41.01	41.18	30.26	32.15	26.25	39.32	35.91	35.47	32.56	38.17
Pseudomonas	F	7.35	11.12	11.04	15.8	15.75	11.75	11.69	14.25	12.17	13.4
Xanthomonas	F	0	2.14	2.15	3.06	2.1	2.48	1.91	2.93	2.98	1.9
Kineococcus	G	0	0.02	0.05	0.14	0.26	0.1	0.17	0.12	0.04	0.23
Methylobacterium	G	0.05	0.76	13.97	3.93	5.76	3.22	4.13	3.68	3.04	4.62
Acidovorax	G	0	0.09	0.18	0.13	0	0.15	0.02	0.22	0.2	0.04
Curtobacterium	G	0.01	0.61	1.8	0.59	0.72	0.72	0.71	0.72	0.6	0.82
Aureimonas	G	0.01	0.27	7.58	1.6	2.91	1.44	1.95	1.73	1.77	1.91
Sphingomonas	G	0.12	5.66	13.65	11.14	12.64	7.23	9.82	7.42	7.16	9.95

Table 3.5 Identified Fungi: All fungal taxa with greater than 0.1% of the total reads sorted by their k-means clustering assignment. The headers represent the four factors considered. The values are the average relative abundance for the samples pooled based on the respective factor.

Taxonomic Classification	Cluster	Grain	Outer Grain	Husk	Outer Husk	2013	2014	DB	VA	Katy	M202
Ustilaginaceae	A	0	0.09	4.78	0	1.82	0.05	1	0.03	0.09	1.16
Cladosporium	B	0.49	1.27	1.82	14.1	9.24	1.32	5.69	1.09	3.57	4.5
Russuales	B	0	0.02	0.02	1.56	0.63	0.22	0.39	0.32	0.39	0.35
Occultifur	B	0.04	0.65	0.23	5.5	4.12	0.2	2.32	0.15	0.1	2.81
Sporobolomyces	B	0.05	0.09	0.16	1.07	0.79	0.06	0.44	0.06	0.11	0.49
Sporidiobolales	B	0	0.02	0	0.94	0.56	0.04	0.33	0.02	0.02	0.4
Phaesphaeriaceae	C	0.09	0.38	0.53	0.74	0.25	0.5	0.5	0.25	0.26	0.55
Alternaria	C	0.36	2.29	3.12	2.24	0.74	2.58	1.73	2.33	2.6	1.38
Ascomycota	C	0.21	2.06	4	4.04	2.42	2.28	3.16	0.81	1.26	3.25
Papilotrema	C	0.26	5.79	9.84	13.3	9.02	5.52	7.21	5.86	13.43	0.97
Basidiomycota	C	0.02	0.09	0.32	0.36	0.32	0.08	0.21	0.08	0.07	0.25
Nectriaceae	C	0.09	0.72	0.23	0.75	0.9	0.32	0.61	0.34	0.45	0.84
Hannaella	C	0.11	15.25	2.17	9.56	2.82	11.84	3.86	17.64	5.28	11.67
Tremellales	C	0	0.68	0.09	0.56	0.38	0.44	0.28	0.68	0.05	0.73
Sphaerulina	C	0.07	1.28	0.22	3.33	0.41	1.81	1.03	1.86	1.82	0.9
Sordariomycetes	C	0.86	2.62	0.86	5.55	1.3	3.36	1.98	3.87	3.75	1.7
Cyphellaceae	C	0	0.25	0.05	0.9	0.61	0.16	0.48	0	0.43	0.23
Bulleribasidium	C	0	0.81	0.94	3.44	1.34	1.18	1.09	1.49	1.31	1.16
Cryptococcus	C	0.13	4.41	1.36	7.13	5.12	2.84	3.44	3.97	1.7	5.3
Naganishia	D	0	0.52	0	0.06	0.5	0.08	0.32	0.06	0.46	0.03
Curvularia	E	1.08	0.3	0.2	0.08	1.1	0.06	0.6	0.084	0.72	0.15
Exserohilum	E	0.26	0.08	0.19	0	0.04	0.17	0.18	0.01	0.24	0.02
Dothideomycetes	E	76.67	22.99	53.52	4.03	18.01	44.38	27.27	49.83	48	24.23
Pleosporales	F	13.21	13.4	8.21	7.47	26.32	3.37	16.14	2.5	9.02	13.32
Microdochium	F	0.04	0.17	0.09	0.05	0.04	0.13	0.15	0.01	0.02	0.18
Fungi	F	5.94	23.76	6.97	13.21	11.21	17.02	19.56	6.64	5.15	23.48

Table 3.6 Shannon Diversity: Diversity of taxa within the samples pooled based on the four seed factors. A higher value can indicate a greater number of species and/or more even relative abundance between the species.

Factor	Bacteria	Fungi
Outer Husk	2.5762	2.7299
Husk	2.3873	1.8335
Outer Grain	2.2452	2.2236
Grain	1.4930	0.8882
Katy	2.5063	2.0825
M202	2.5749	2.4083
DB	2.3908	2.3615
UA	2.6175	1.8350
2013	2.5088	2.4415
2014	2.4966	1.9720

Table 3.7 AnoSim Diversity: Analysis of similarity using Bray-Curtis dissimilarity matrices across all rice samples. A P-value <0.05 indicates significance and the AnoSim R value indicates whether the samples pooled based on their factor have more (>0) or less (<0) diversity relative to the sample considered. Subscript numbers for each factor indicate the number of levels, number of replicates in each level, and number of permutations performed by the AnoSim test.

		Bacterial 16S v3-v4		Fungal ITS1	
Sample	**Factor**	**AnoSim R**	**P**	**AnoSim R**	**P**
Global (all)	Tissue $_{4,6,999}$	0.5133	0.001	0.3039	0.001
	Genotype $_{2,12,999}$	-	0.148	-	0.083
	Year $_{2,12,999}$	-	0.223	0.3539	0.001
	Location $_{2,12,999}$	-	0.872	-	0.192
	Tissue by Genotype $_{8,3,999}$	0.547	0.001	0.266	0.014
	Tissue by Year $_{8,3,999}$	0.4575	0.001	0.5449	0.001
	Tissue by Location $_{8,3,999}$	0.3747	0.002	0.2245	0.043
Outer Husk	Genotype $_{2,6,719}$	-	0.600	-	0.200
	Year $_{2,3,719}$	-	0.467	-	0.100
	Location $_{2,3,719}$	-	0.667	-	0.800
Husk	Genotype $_{2,3,719}$	-	0.200	-	0.700
	Year $_{2,3,719}$	-	0.067	-	0.067
	Location $_{2,3,719}$	-	0.467	-	0.200
Outer Grain	Genotype $_{2,3,719}$	-	1.000	-	0.200
	Year $_{2,3,719}$	-	0.667	-	0.533
	Location $_{2,3,719}$	-	0.400	-	0.733
Grain	Genotype $_{2,3,119}$	-	0.100	-	0.200
	Year $_{2,3,119}$	-	0.900	-	0.133
	Location $_{2,3,119}$	-	1.000	-	0.600

Table 3.8 Rice Seed Bacterial Core Microbiome: The bacterial OTUs and their representative taxa shared between all samples of a particular seed compartment. The numbers in parentheses represent the number of OTUs belonging to the bacteria taxa that are shared with other compartments in order according to the header. The bold number represents the number of OTUs belonging to the compartment of interest.

Grain	Outer Grain	Husk	Outer Husk
Pantoea (**11**,11,11,11)	Pantoea (11,**19**,12,14)	Sphingomonas (-,12,**17**,14)	Sphingomonas (1,13,14,**26**)
Enterobacteriaceae (**6**,6,6,6)	Sphingomonas (-,**13**,12,13)	Pantoea (11,12,**13**,13)	Pantoea (11,14,13,**18**)
Enterobacter (**2**,2,-,-)	Pseudomonas (1,**11**,4,9)	Enterobacteriaceae (6,10,**10**,10)	Pseudomonas (1,9,4,**14**)
Escherichia-Shigella (**2**,2,2,2)	Enterobacteriaceae (6,**10**,10,10)	Aureimonas (-,3,**9**,5)	Enterobacteriaceae (6,10,10,**10**)
Paenibacillus (**2**,2,2,2)	Paenibacillus (2,**7**,5,7)	Methylobacterium (-,5,**8**,5)	Paenibacillus (2,7,5,**9**)
Pseudomonas (**1**,1,1,1)	Escherichia-Shigella (2,**6**,4,4)	Curtobacterium (-,5,**5**,4)	Chryseobacterium (-,1,-,**8**)
	Rhizobium (-,**6**,4,5)	Escherichia-Shigella (2,4,**5**,5)	Rhizobium (-,5,4,**8**)
	Curtobacterium (-,**5**,5,4)	Paenibacillus (2,5,**5**,5)	Escherichia-Shigella (2,4,5,**6**)
	Methylobacterium (-,**5**,5,5)	Rhizobium (-,4,**5**,4)	Mucilaginibacter (-,3,3,**6**)
	Aureimonas (-,**3**,3,3)	Pseudomonas (1,4,**4**,4)	Aureimonas (-,3,5,**5**)
	Mucilaginibacter (-,**3**,2,3)	Mucilaginibacter (-,2,**3**,3)	Methylobacterium (-,5,5,**5**)
	Enterobacter (2,**3**,-,-)	Microbacterium (-,-,**2**,2)	Curtobacterium (-,4,4,**4**)
	Xanthomonas (-,**3**,2,3)	Roseomonas (-,-,**2**,2)	Xanthomonas (-,3,2,**4**)
	Chryseobacterium (-,**1**,-,1)	Xanthomonas (-,2,**2**,2)	Luteibacter (-,1,1,**3**)
	Eucidaris (-,**1**,-,1)	Kineococcus (-,1,**1**,-)	Microbacterium (-,-,2,**3**)
	Herbaspirillum (-,**1**,-,1)	Luteibacter (-,-,**1**,1)	Herbaspirillum (-,1,-,**2**)
	Kineococcus (-,**1**,1,-)	Novosphingobium (-,-,**1**,1)	Ochrobacterium (-,1,1,**2**)
	Luteibacter (-,**1**,-,1)	Ochrobacterum (-,1,**1**,1)	Roseomonas (-,-,2,**2**)
	Ochrobacterum (-,**1**,1,1)		Stenotrophomonas (-,-,-,**2**)
			Brevundimonas (-,-,-,**1**)
			Eucidaris (-,1,-,**1**)
			Flavobacterium (-,-,-,**1**)
			Herbiconiux (-,-,-,**1**)
			Novosphingobium (-,-,1,**1**)
			Variovorax (-,-,-,**1**)

Table 3.9 Rice Seed Fungal Core Microbiome: The fungal OTUs and their representative taxa shared between all samples of a particular seed compartment. The numbers in parentheses represent the number of OTUs belonging to the bacteria taxa that are shared with other compartments in order according to the header. The bold number represents the number of OTUs belonging to the compartment of interest.

Grain	Outer Grain	Husk	Outer Husk
Dothideomycetes (**3**,3,3,-) Pleosporales (**2**,2,2,-) Ascomycota (**1**,-,1,-) Sordariomycetes (**1**,1,1,-)	Dothideomycetes (3,**9**,4,-) Pleosporales (2,**7**,3,-) Alternaria (-,**4**,2,-) Ascomycota (-,**2**,1,-) Cladosporium (-,**1**,1,-) Culvaria (-,**1**,1,-) Exserohilum (-,**1**,-,-) Periconia (-,**1**,-,-) Pyricularia (-,**1**,-,1) Naganishia (-,**1**,-,-) Sordariomycetes (1,**1**,1,-)	Dothideomycetes (3,4,**6**,-) Pleosporales (2,3,**4**,-) Ascomycota (1,1,**3**,-) Alternaria (-,2,**2**,-) Papilotrema (-,-,**2**,-) Basidiomycota (-,-,**1**,-) Bullera (-,-,**1**,-) Cladosporium (-,1,**1**,-) Culvaria (-,1,**1**,-) Cyphellaceae (-,-,**1**,1) Dioszegia (-,-,**1**,-) Kondoa (-,-,**1**,-) Neoascochyta (-,-,**1**,-) Occultifur (-,-,**1**,-) Sphaerulina (-,-,**1**,-) Sporobolomyces (-,-,**1**,-)	Ascomycota (-,-,-,**2**) Cyphellaceae (-,-,1,**1**) Pyricularia (-,1,-,**1**)

References

Abarenkov K, Nilsson RH, Larsson K, Alexander IJ, Eberhardt U, Erland S, Hoiland K, Kjoller R, Larsson E, Pennanen T, Sen R, Taylor AFS, Tedersoo L, Ursing BM, Vralstad T, Liimatainen, Peintner U, and Koljaljg U (2010). The UNITE database for molecular identification of fungi - recent updates and future perspectives. New Phytologist 186, 263-264.

Achari GA, and Ramesh R (2014). Diversity, Biocontrol, and Plant Growth Promoting Abilities of Xylem Residing Bacteria from Solanaceous Crops. International Journal of Microbiology, 1-14.

Agarwal VK, Singh OV, and Modgal SC (1975). Influence of different doses of nitrogen and spacing on the seed-borne infections of rice. Indian Phytopathology 28:1, 38-40.

Aktar MW, Sengupta D, and Chowdhury A (2009). Impact of pesticides use in agriculture: their benefits and hazards. Interdisciplinary Toxicology 2:1, 1-12.

Aparna G, Chatterjee A, Sonti RV, and Sankaranarayanan R (2009). A Cell Wall-Degrading Esterase of *Xanthomonas oryzae Requires a Unique Substrate Recognition Module for Pathogenesis on Rice*. Plant Cell 21:6, 1860-1873.

Barret M, Briand M, Bonneau S, Preveaux A, Valiere S, Bouchez O, Hunault G, Simoneau and Jacquesa MA (2015). Emergence shapes the structure of the seed microbiota. Applied Environmental Microbiology 81:4, 1257-1266.

Berg G (2009). Plant-microbe interactions promoting plant growth and health: perspectives for controlled use of microorganisms in agriculture. Applied Microbiology and Biotechnology 84, 11–18.

Bertani I, Abbruscato P, Piffanelli P, Subramoni S, and Venturi V (2016). Rice bacterial endophytes: isolation of a collection, identification of beneficial strains and microbiome analysis. Environmental Microbiology Rep.

Breidenbach B, PUmp J, and Dumont MG (2016). Microbial Community Structure in the Rhizosphere of Rice Plants. Frontiers of Microbiology 6, 1-12.

Bulgarelli D, Rott M, Schlaeppi K, Ver Loren van Themaat E, Ahmadinejad N, Assenza F, Rauf P, Huettel B, Reinhardt R, Schmelzer E, Peplies J, Gloeckner FO, Amann R, Eickhorst T, Schulze-Lefert P (2012). Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. Nature 2:488, 91-95.

Bulgarelli D, Schlaeppi K, Spaepen S, Ver Loren van Themaat E, and Schulze-Lefert P (2013). Structure and functions of the bacterial microbiota of plants. Annual Review of Plant Biology 64, 807-838.

Busby PE, Soman C, Wagner MR, Friesen ML, Kremer J, Bennett A, Morsy M, Eisen JA, Leach JE, and Dangl JL (2017). Research priorities for harnessing plant microbiomes in sustainable agriculture. PLOS Biology 15:3, 1-14.

Calvo P, Nelson L, and Kloepper JW (2014). Agricultural uses of plant biostimulants. Plant Soil 383, 3-41.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, and Knight R (2010). QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7:5, 335-336.

Carbone I, White JB, Miadlikowska J, Arnold AE, Miller MA, Kauff F, U'Ren JM, May G, and Iutzoni F (2017). T-BAS: Tree-Based Alignment Selector toolkit for phylogenetic-based placement, alignment downloads, and metadata visualization: an example with the Pezizomycotina tree of life. Bioinformatics 33:8, 1160-1168.

Carvalho FP (2006). Agriculture, pesticides, food security, and food safety. Environmental Science and Policy 9, 685-692.

Chaibub AA, de Carvalho JC, de Sousa Silva C, Collevatti RG, Goncalves FJ, de Carvalho Varros Cortes MV, de Filippi MC, de Faria FP, Lopes DC, and de Araujo LG (2016). Defense responses in rice plants in prior and simultaneous applications of *Cladosporium* sp. during leaf blast suppression. Environmental Science Pollution Research 23, 215554-21564.

Darrasse A, Darsonval A, Boureau T, Brisset MN, Durand K, Jacques MA (2010). Transmission of plant-pathogenic bacteria by nonhost seed without induction of an associated defense reaction at emergence. Applied Environmental Microbiology 76, 6787-6796.

Edwards J, Johnson C, Santos-Medellin C, Lurie E, Podishetty NK, Bhatnagar S, Eisen JA, Sundaresan V (2015). Structure, variation, and assembly of the root-associated microbes of rice. PNAS 112:8, 911-920.

Elbeltagy A, Nishioka K, Sato T, Suzuki H, Ye B, Hamada T, Isawa T, Mitsui H, and Minamisawa K (2001). Endophytic Colonization and In Planta Nitrogen Fixation by a *Herbaspirillum* sp. Isolated from Wild Rice Species. Applied and Environmental Microbiology, 67:11, 5285-5293.

Esitken A, Pirlak L, Turan M, and Sahin F (2006). Effects of floral and foliar application of plant growth promoting rhizobacteria (PGPR) on yield, growth, and nutrition of sweet cherry. Scientia Horticulturae 110:4, 324-327.

Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, and Ravel J (2014). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome 2:6, 1-7.

Feng Y, Shen D, and Song W (2006). Rice endophyte *Pantoea agglomerans* YS19 promotes host plant growth and affects allocations of host photosynthates. Journal of Applied Microbiology 100, 938-945.

Finkel OM, Castrillo G, Paredes SH, Gonzalez, and Dangl JL (2017). Understanding and exploiting plant beneficial microbes. Current Opinion in Plant Biology 38, 155-163.

Fonseca-Garcia C, Coleman-Derr D, Garrido E, Visel A, Tringe SG, and Partida-Martinez LP (2016). The Cacti Microbiome: Interplay between Habitat-Filtering and Host-Specificity. Frontiers in Microbiology 5:150, 1-16.

Garcia V, Brizzio S, Russo G, Rosa CA, Boekhout T, Theelen B, Libkind D, and van Broock (2010). *Cryptococcus spencermartinsiae* sp. Nov., a basidiomycetous yeast isolated from glacial waters and apple fruits. International Journal of Systematic and Evolutionary Microbiology 60, 707-711.

Gutierrez SA, Carmona MA, and Reis EM (2010). Methods for Detection of *Alternaria padwickii* in Rice Seeds. Journal of Phytopathology 158:8, 523-526.

Hardiom PR, Nazir R, Sessitsch A, Elhottova D, Korenblum E, van Overbeek LS, and van Elsas JD (2013). The new species *Enterobacter oryziphilus* sp. Nov. and *Enterobacter oryzendophyticus* sp. Nov. are key inhabitants of the endosphere of rice. BMC Microbiol. 13:164.

Huse SM, Ye Y, Zhou Y, and Fodor AA (2012). A Core Human Microbiome as Viewed through 16S rRNA Sequence Clusters. PLoS One 7:6, 1-12.

Joung K, and Cote J (2000). A review of the environmental impacts of the microbial insecticide *Bacillus thuringiensis*. Government of Canada Publications, 1-16.

Kawahara Y, de la Bastide M, Hamilton JP, Kanamori HK, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, Childs KL, Davidson RM, Lin H, Quesada-Ocampo L, Vaillancourt B, Sakai H, Lee SS, Kim J, Numa H, Itoh T, Buell CR, and Matsumoto T (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. Rice 6:4, 1-10.

Kubicek CP, Starr TL, and Glass NL (2014). Plant Cell Wall-Degrading Enzymes and Their Secretion in Plant-Pathogenic Fungi. Annual Review of Phytopathology 52, 427-451.

Lemos LN, Fulthorpe RR, Triplett EW, Roesch LFW (2011). Rethinking microbial diversity analysis in the high throughput sequencing era. Journal of Microbiological Methods 86:1, 42-51.

Liu Y, Liu L, Qiu F, Schumann P, Shi Y, Zou Y, Zhang X, and Song W (2010). *Paenibacillus hunanensis* sp. nov., isolated from rice seeds. International Journal of Systematic and Evolutionary Microbiology 60, 1266-1270.

Lubeck M, Knudsen IMB, Jensen B, Thrane U, Janvier C, and Funck Jensen D (2002). GUS and GFP transformation of the biocontrol strain *Clonostachys rosea* IK726 and the use of these marker genes in ecological studies. Mycology Research 106:7, 815-826.

Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, Tremblay J, Engelbrektson A, Kunin V, del rio TG,, Edgar RC, Eickhorst T, Ley RE, Hugenholtz P, Tringe SG, and Dangl JL (2012). Defining the core *Arabidopsis thaliana* root microbiome. Nature 488, 86-90.

Mano, H. and Morisaki, H (2008). Endophytic bacteria in the rice plant. Microbes and Environments 23, 109-117.

Midha S, Bansal K, Sharma S, Kumar N, Patil PP, Chaudhry V, and Patil PP, Chaudhry V, and Patil PB (2016). Genomic Resource of Rice Seed Associated Bacteria. Frontiers in Microbiology 6:1551.

Miller MA, Schwartz T, Pickett BE, He S, Klem EB, Scheuermann RH, Passarotti M, Kaufman S, and O'Leary MA (2015). A RESTful API for Access to Phylogenetic Tools via the CIPRES Science Gateway. Evolutionary Bioinformatics Online 11, 43-48.

Mitter B, Pfaffenbichler N, Flavell R, Compant S, Antonielli L, Petric A, Berninger T, Naveed M, Sheibani-Tezerji R, von Maltzahn G, and Sessitsch A (2017). A New Approach to Modify Plant Microbiomes and Traits by Introducing Beneficial Bacteria at Flowering into Progeny Seeds. Frontiers in Microbiology 8:11, 1-10.

Nakase T and Suzuki M (1985). Ballistospore-forming yeasts found on the Surface of the Japanese Rice Plant, *Oryza sativa*. Journal of General and Applied Microbiology 31, 457-474.

Nelson EB (2004). Microbial Dynamics and Interactions in the Spermosphere. Annual Review of Phytopathology 42, 271-309.

Nelson EB (2017). The seed microbiome: Origins, interactions, and impacts. Plant Soil, 1-28.

Nutaratat P, Srisuk N, Arunrattiyakorn P, and Limtong S (2014). Plant growth-promoting traits of epiphytic and endophytic yeasts isolated from rice and sugar cane leaves in Thailand. Fungal Biology 118:8, 683-694.

O'Callaghan M (2016). Microbial inoculation of seed for improved crop performance: issues and opportunities. Applied Microbiology Biotechnology 100, 5729-5746.

Oerke EC (2006). Crop losses to pests. Journal of Agricultural Science 144, 31–43.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, and Wagner H (2017)**.** vegan: Community Ecology Package. https://CRAN.R-project.org/package=vegan

Parnell JJ, Berka R, Young HA, Sturino JM, Kang Y, Barnhart DM, and DiLeo MV (2016). From the Lab to the Farm: An Industrial Perspective of Plant Beneficial Microorganisms. Frontiers in Plant Science 7:1110, 1-12.

Pikaar I, Matassa S, Rabaey K, Bodirsky BL, Popp A, Herrero M, and Verstraete W (2017). Microbes and the Next Nitrogen Revolution. Environmental Science & Technology 51:13, 7297-7303.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, and Glockner FO (2014). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research 41, 590-596.

Robert V, Stegehuis G, and Stalpers J (2005). The MycoBank engine and related databases. http://www.mycobank.org

Robertson GP, and Swinton SM (2005). Reconciling agricultural productivity and environmental integrity: a grand challenge for agriculture. Frontiers in Ecology and the Environment 3:1, 38-46.

Saa S, Rio AO, Castro S, and Brown PH (2015). Foliar application of microbial and plant based biostimulants increases growth and potassium uptake in almond (*Prunus dulcis* [Mill.] D. A. Webb). Frontiers in Plant Science 6:87, 1-9.

Sanchez-Canizares C, Jorrin B, Poole PS, and Tkacz A (2017). Understanding the holobiont: the interdependence of plants and their microbiome. Current Opinion in Microbiology 38, 188-196.

Shade A, Jacques M, and Barret M (2017). Ecological patterns of seed microbiome diversity, transmission, and assembly. Current Opinion in Microbiology 37, 15-22.

Tani A, Sahin N, Fujitani Y, Kato A, Sato K, and Kimbara K (2015). *Methylobacterium* Species Promoting Rice and Barley Growth and Interaction Specificity Revealed with Whole-Cell Matrix-Assisted Laser Desorption/Ionization-Time-of-Flight Mass Spectrometry (MALDI-TOF/MS) Analysis. PLoS ONE 10:6, 1-15.

Tedersoo L, and Lindahl B (2016). Fungal identification biases in microbiome projects. Environmental Microbiology Reports 8:5, 774-779.

Vandenkoornhuyse P, Quaiser A, Duhamel M, Le Van A, and Dufresne A (2015). The importance of the microbiome of the plant holobiont. New Phytologist 206, 1196-1206.

Videira SS, de Araujo JL, Rodrigues Lda S, Baldani VLD, and Baldani JI (2008). Occurrence and diversity of nitrogen-fixing *Sphingomonas* bacteria associated with rice plants grown in Brazil. FEMS Microbiol. Lett. 293, 11-19.

Vleesschauwer DD, Djavaheri M, Bakker P, and Hofte M (2008). *Pseudomonas fluorescens* WCS374r-Induced Systemic Resistance in Rice against *Magnaporthe oryzae* Is Based on Pseudobactin-Mediated Priming for a Salicylic Acid-Repressible Multifaceted Defense Response. Plant Physiology 148:4, 1996-2012.

Xia Y and Sun J (2017). *Hypothesis testing and statistical analysis of the microbiome*. Genes and Diseases 4:3, 138-148.

Xie C, and Yokota A (2006). *Sphingomonas azotifigens* sp. nov., a nitrogen-fixing bacterium isolated from the roots of *Oryza sativa*. International Journal of Systematic and Evolutionary Microbiology 56, 889-893.

Walters W, Hyde ER, Berg-Lyons D, Ackermann G, Humphrey G, Parada A, Gilbert JA, Jansson JK, Caporaso JG, Fuhrman JA, Apprill A, and Knight R (2015). Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. mSystems 1:1, 1-10.

Wang Y, Yang MH, Wang XB, Li TX, and Kong LY (2014). Bioactive metabolites from the endophytic fungus *Alternaria alternata.* Fitoterapia 99, 153-158.

Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, and Glockner FO (2013). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks 42, 643-648.

Zhu B, Liu H, Tian W, Fan X, Li B, Zhou X, Jin G, and Xie G (2012). Genome Sequence of *Stenotrophomonas maltophilia* RR-10, Isolated as an Endophyte from Rice Root. Journal of Bacteriology 194:5, 1280-1281.

CHAPTER 3

Adaptive Evolution Estimates for *M. oryzae* Putative Effectors and Orthologous Clusters

Alexander Eyre

Preface

This chapter contains work done in collaboration with Laura H. Okagaki, Joshua K. Sailsbery and William Sharpee while members of the Dean laboratory. For two projects, I conducted analyses of selection pressures on alignments of genes using CodeML, a package part of the Phylogenetic Analysis by Maximum Likelihood (PAML) program that uses maximum likelihood estimation with models of codon evolution to predict whether the alignment is undergoing neutral, diversifying, or purifying selection. In the first project, I estimated selection pressures for eleven putative *M. oryzae* effectors using sequences from 43 sequences isolates. In the second project, I aligned and predicted selection pressures for 6518 orthologous gene clusters formed between *M. oryzae* and two sequenced members of its family: *M. poae* and *G. graminis* var. *tritici*.

Abstract

Comparative genomic studies are becoming more prevalent and powerful as the amount of sequence data for related organisms increases. These evolutionary analyses are useful for identifying genes with similar functions between species, especially fungal pathogenic genes involved in infection or modulation of the host immune system. *Magnaporthe oryzae* is a fungal plant pathogen of rice responsible for rice blast disease whose genome has recently been sequenced in addition to some isolates and members of its family. Hypothesized effector genes were identified in an effectoromics screen, transiently expressed and tested in *N. benthamiana*, and eleven were classified as suppressors of plant death. An analysis of selection pressures was conducted using CodeML for these putative effector genes using sequences extracted from 43 *M. oryzae* isolates. Only one of the eleven genes contained enough sequence variability to

inform the selection pressures, *SPD4*, which experienced purifying selection. Additionally, 6518 orthologous clusters were identified between *M. oryzae* and two of its related family members, *Magnaporthe poae* and *Gaeumannomyces graminis* var. *tritici*. Selection pressures were identified for each cluster, which revealed 79% undergoing neutral, 19% diversifying, and 2% purifying selection. Analysis of the mode of selection in relation to distance to repetitive elements or whether the genes are small and/or secreted revealed no statistical correlation. Overall, these results indicate the need for sufficient genomic data to assess mode of selection and no evidence that the proximity to repetitive elements effects the diversification of genes.

Introduction

As the cost of DNA sequencing has been falling and its quality improving with the introduction of high throughput sequencing technologies, there has been a massive increase in the abundance of genetic data. Many new organisms' genomes and their isolates are now available for research, which has made comparative genomic analyses more prevalent and powerful, allowing for increased insight into evolutionary relationships within and between species. These studies are useful for investigation into the genetics of fungal plant-pathogens where the comparison of related fungi may reveal core genes related to fungal growth and maintenance. Usage of more closely related pathogenic species may reveal key genes involved in fungal pathogenicity and those regulating their lifestyles (O'Connell et al., 2012). Unique genes without orthologs in related species may unveil proteins unique to a pathogen used to promote virulence on their host(s), such as effectors (Creuzburg et al., 2011; Shang et al., 2016).

Magnaporthe oryzae is a filamentous ascomycete fungus and the causal agent of rice blast disease responsible for the destruction of millions of hectares of rice annually (Wilson *et al*. 2009). The fungus beings its life cycle as an aerial asexual spore that, upon landing on a rice plant, germinates and develops an appressorium. This dome-shaped structure produces high internal turgor pressure that pushes a penetration peg through the plant cell wall to start

invasion. Once inside the host, it grows and feeds from living rice tissue for about 72 hours, at which point it kills the occupied cells and feeds from the dead tissue (Talbot, 2003). The fungus regulates infection during this time by secreting effector proteins, small proteins that block host recognition, defense response pathways, and alter host metabolism (Hogenhout et al., 2009; Djamei et al., 2011). Once environmental conditions are optimal, the fungus sporulates and produces several thousand spores from each infection site that repeat the cycle on the host or neighboring rice plants (Howard *et al*. 1996).

Over the years, *M. oryzae* has established itself as a model fungal plant pathogen for research due to its economic significance, simple culturing requirements, and possibility of maintaining a haploid stage allowing for ease of genetic manipulation. This classification has driven the sequencing of not only its genome, but many isolates from around the globe (Talbot, 2003). Within the Magnaporthaceae family are a few additional fungal pathogens also with their genomes sequenced, namely *Mangaporthe poae* and *Gaeumannomyces graminis* var. *tritici* causing summer patch on turfgrass and take-all disease on wheat respectively. Like *M. oryzae*, these pathogens also have great economic impact and are important subjects of research (Besi *et al*. 2009, Okagaki *et al*. 2015).

When comparing two organisms' genomes, the primary focus has been the search for homologs, genes related by descent from a common ancestral DNA sequence. This classification usually entails that the two genes share a minimum of 50% sequence similarity. There are two main classes of homologs: orthologs where the genes in each organism share the same function, and paralogs where the function of one or more of the genes have changed over time. This distinction has interesting implications for which functions are important to the respective organisms, where orthologous genes may be considered more highly conserved to preserve function. Functional differentiation is thought to be largely driven by the selection pressures on the genes, where a drive to conserve function over time might change the DNA sequence in such a way that the amino acids of the proteins are not altered (Rech et al., 2014).

In these two studies, selection pressures for two *M. oryzae* datasets were analyzed using CodeML, a package part of Phylogenetic Analysis by Maximum Likelihood (PAML). CodeML uses maximum likelihood estimation and several models of codon evolution to predict whether a gene alignment is experiencing neutrality or positive selection. If a set of genes are predicted to experience positive selection, the algorithm calculates of the rate of nonsynonymous (dN) and synonymous (dS) substitutions in the alignment. The ratio of these values (dN/dS) for the set of genes can inform the direction of selection, where a ratio greater than one suggests diversifying and less than one suggests purifying selection (Yang, 2007). In the study of fungal pathogens, one might expect the vital growth and lifestyle-related genes to undergo purifying selection in order to maintain their function over time. Conversely, effector and related genes that interact with the plant immune system can be recognized by host defense proteins, triggering the plant hypersensitive response (HR) resulting in immunity to the pathogen. Therefore, one might expect these genes to experience diversifying selection in order to alter their amino acid sequence in such a way that they avoid detection by the host (Huang *et al*. 2014).

In the first study, a set of 247 hypothesized effector genes were identified from the *M. oryzae* genome using a bioinformatic pipeline, eleven of which were classified as putative effector genes after expression experiments in *N. benthamiana*. Access to the GenBank database provided additional sequences for these genes from 43 isolates around the globe. The selection pressures for each of these gene sets were estimated, however only one of the eleven genes contained enough sequence variability to inform the selection pressures, *SPD4*, which experienced purifying selection (Sharpee *et al.* 2016). In the second study, orthologs of *M. oryzae* genes were identified in two related species from the *Magnaporthaceae* family, *M. poae*, and *G. graminis* var. *tritici*, to form 6518 orthologous clusters. Each cluster was aligned and selection pressures estimated, about 21% of which were identified as undergoing positive selection. Major functions clusters in each of the diversifying and purifying categories were

identified and cellular functions were found that dominated each class. Finally, correlations of the diversifying clusters to protein size and distance to repetitive elements was assessed to identify no significant correlation (Okagaki *et al*. 2015). Altogether, the work provides an abundance of evolutionary insight into the *M. oryzae* genome and genes involved in pathogenicity.

Results and Discussion

CodeML Analysis of Putative M. oryzae Effector Genes

Effector proteins are vital for the pathogenicity of a wide variety of plant pathogens, where they block host recognition or response pathways, and alter the host metabolism to promote disease. Their genes often do not have homologs in related species making their identification difficult, however they have a set of unique characteristics that make their screening possible. Preliminary work into the selective pressures these genes experience involved a bioinformatic pipeline to identify candidate effector genes from the whole *M. oryzae* genome. This screening required the hypothetical effectors to be shorter than 250 amino acids, have equal to or greater than 3% cysteine content, contain an N terminal signal peptide, and lack a transmembrane domain. Additional candidates were added if they did not meet some of the above criteria but contained an N terminal signal peptide and were highly upregulated during infection. This yielded 247 candidates from the 12,991 sequenced *M. oryzae* genes of which 73 were successfully cloned into *Agrobacterium tumefaciens* vectors for functional verification. These effectors were then screened for plant cell death suppressing activity in *N. benthamiana* by agroinfiltration of leaf discs with the gene vector followed 24 hours later with agroinfiltration with Nep1, a known inducer of plant cell death by *M. oryzae* (Zhang *et al*. 2012). 11 of the 73 cloned genes significantly reduced plant cell death in the leaf discs, were confirmed in a similar manner with BAX, also an inducer of cell death, and classified as necrosis suppressing effectors (Sharpee *et al*. 2016).

The general evolutionary hypothesis for effectors is that they actively experience diversifying selection to avoid detection by their hosts, since detection elicits the HR response and immunity to the pathogen. However, this might not always be the case if an effector is not being employed during pathogenesis. In order to investigate whether the eleven putative and two additional documented effectors were subject to such selection, their gene sequences were extracted from the Broad Institute database (Broad Institute, 2015) and run through BLASTn with the whole genome shotgun database containing 43 additional *M. oryzae* isolates' sequences from around the world. After reconstruction, these were aligned then analyzed using CodeML M1/M2 and M7/M8 models. For each effector, there were two likelihood ratio tests performed to identify whether the data was informative enough to deduce the type of selection occurring for the group of genes. All but one of these tests failed, indicating the datasets did not have enough sequence diversity and subsequent information. Review of the alignments revealed eight of the eleven effectors only contained a single allele among the isolates, suggesting there was insufficient information to calculate selection pressure.

One putative effector, *SPD4*, contained enough information between its 11 alleles present to produce positive selection on one of the MLE tests. The results indicated the gene was undergoing purifying selection with an estimated dN/dS ratio of 0.7517, although it was only able to pass the simpler of the two selection models (M1/M2). The second set of selection models were developed to be more realistic, relying on a beta distribution and additional parameters (Yang, 1997). This presumably required more information in the datasets, which was not available between the eleven alleles present. Alignment of the SPD4 alleles (**Figure 1**) revealed all the differentiation between the alleles were single nucleotide polymorphisms with no indels present to vary sequence lengths. The results together reveal the difficulties in the evolutionary study of effectors, as their specificity to a pathogen limits additional sequences from related species to increase the amount of sequence data available. One could work around this issue through the period sequencing of isolates from around the globe over many

years. This idea might capture more alterations to these genes that would better inform the type of selection on these genes over time, however this approach would take an unspecified amount of time to reach a conclusion. Targeted laboratory experiments, such as evolutionary studies involving an array of rice plants with variable effector-related resistance to *M. oryzae*, may be able to better explore the diversification of effector genes hypothesis.

CodeML Analysis of M. Oryzae Orthologous Clusters

Computation of selection pressures for orthologous clusters across related organisms can shed light on important patterns in gene conservation. Clusters experiencing purifying selection might represent a need to maintain the functionality of the genes, while those experiencing diversifying selection might indicate a need for modified functions or removal of deleterious genes. To identify genes orthologous to the *M. oryzae* genome within the *Magnaporthaceae* family, OrthoMCL was employed with the genomes from two other members of the family, *M. poae* and *G. graminis* var. *Tritici*. These datasets included 12991, 14650, and 12329 sequences for the three fungi, respectively. Clusters were constructed that contain at least one gene from each of the three species, resulting in a total of 6518 clusters, which were considered the *Magnaporthaceae* "core proteome." Within these clusters, 87% contained a single gene from each of the three species, while 13% contained at least one additional paralog.

These clusters were aligned then subject to CodeML analysis to identify selection and estimate their dN/dS ratio. Unlike the *M. oryzae* putative effector analysis, all the clusters contained enough sequence variability to pass the likelihood ratio statistical analysis allowing the selection results to be interpreted. Of the core proteome clusters, 79% were found to be under neutrality, 19% to be under diversifying selection, and 2% to be under purifying selection. These clusters were then divided up based on whether they contained additional paralogs, and the proportions of those undergoing neutral vs. positive selection identified. Those with no paralogs had a profile similar to all clusters, while those with paralogs had a great increase in the proportion of those undergoing positive selection (**Figure 3.2**), although they showed no

bias toward purifying or diversifying selection. To shed light on the functions of the genes in the selection categories, Blast2GO was employed on clusters undergoing their respective selection type. About 55% of the diversifying and 38% of the purifying categorized clusters had no annotation, and 14 of the top 20 classifications for those with annotations were shared between the two. Only present in the purifying category were binding, nucleotide binding, nucleoside metabolic, and lipid metabolic processes classifications. Alternatively, only present in the diversifying category were regulation of transcription, nucleus, and zinc binding classifications. Together, this suggests metabolism and binding related proteins are more highly conserved, while transcription factor and regulation related proteins are actively being changed.

Previous studies have demonstrated long-terminal repeat (LTR) transposon regions carry a tendency for repeat induced point (RIP) mutations, and that genes in proximity to these regions or other repetitive elements have increased rates of diversification due to the increased rates of mutation (Ikeda *et al.,* 2002). Therefore, it was hypothesized that genes in proximity to these repetitive elements may undergo a higher rate of selection. To test this hypothesis, repetitive element libraries were built for each *Magnaporthaceae* species and only those with lengths greater than 200bp were considered for the analysis. The distances to these elements for purifying or diversifying clusters were first deduced and little difference between the two groups was found. The distance to the closest repetitive element for each gene was then compared to the dN/dS score as reported by PAML, graphed, and a Wann-Whitney Rank Sum test was performed on these data (**Figure 3.3**). Overall, there was no significant correlation between the distance to the nearest repetitive element and rates of positive selection on the genes with coefficients of determination (R^2) values between 0.0003 and 0.0007 for the three species.

Finally, there remained a possibility that effector proteins, which are hypothesized to be undergoing higher rates of diversifying selection to avoid detection by their respective

pathogens, might be in proximity of these repetitive elements and undergoing diversifying selection. To test this hypothesis, the unique proteins, or those genes that do not cluster with either of the other two species, were run through SignalP and TargetP to identify those with a signal peptide that targets the respective proteins out of the cell. Interestingly, there was an enrichment in the quantity of secreted proteins among the unique proteins relative to the whole genome for the three species. These proteins were then classified into either secreted or small secreted (<250 amino acids), where the small secreted proteins are more likely to be the effectors of interest. To test if the secreted proteins are more likely to be closer to the repetitive elements, the proteins in each category were compared with the genome average (**Figure 3.4**). For *M. oryzae*, both secreted and small secreted were found to be closer to the repetitive elements than the genome average, although there was no significant difference between the two. In *M. poae* and *G. graminis* var. *Tritici*., only the secreted proteins were found closer to the repetitive elements.

Overall, the results demonstrated no significant correlation between the mode of selection pressure of a gene and its distance to repetitive elements or its putative functionality as an effector. These do not support the hypotheses that closeness to a repetitive element that experiences higher rates of mutation, or that effectors expected to undergo diversifying selection to continuously challenge the plant innate immune system. This suggests that there may be more at play than what can be interpreted from genomic data, such as an epigenetic mechanism that may alter the rate of selection on a gene. Conversely, the functional annotation of clusters experiencing variable selection pressure offered insight into the conservation of genes in fungal genomes. Prevalent in the purifying category were nucleotide binding, nucleoside metabolic, and lipid metabolic processes, suggesting the importance of maintaining the functionality of genes related to DNA maintenance or expression and other key metabolic processes. In the diversifying category were regulation of transcription and zinc binding classifications, both of which have to do with DNA binding functions. This might suggest the

need to vary the amino acids related to the targeting of proteins to different sections of DNA, possibly allowing for cross-generational adjustment of gene expression levels.

Materials and Methods

Effector Data Acquisition

Gene sequences from the candidate suppressing genes were extracted from the *M. oryzae* 70-15 isolate, downloaded from the Fungal Genome Initiative at Broad Institute of Harvard and Massachusetts Institute of Technology (Broad Institute, 2015). These were run through BLASTn against the shotgun sequences of 43 *M. oryzae* strains deposited in the Whole Genome Shotgun (WGS) database located in GenBank, and manually reconstructed into whole gene sequences.

CodeML of Putative Effectors

Putative effector sequences were aligned using MUSCLE v3.8.31 (Edgar, 2004), reiterating the alignments until reaching convergence and generating phylogenetic trees from the second alignment iteration. The nonsynonymous to synonymous (dN/dS) substitution rates of the sequences were estimated using the CodeML algorithm as part of PAML v4.8 (Yang, 2007). Likelihood ratio tests of site-specific selection were used, comparing M1 (neutral) to M2 (selection) and M7 (beta) to M8 (beta & w) using the test statistic 2*(lnL1-lnL2) = 2ΔL. The gene was considered undergoing positive selection if both the M1/M2 and M7/M8 likelihood ratio tests were significant under a chi-square test with $p < 0.05$.

Data Acquisition and Ortholog Identification

Whole genome, transcriptome, and proteome sequences for 74 fungal species, including *M. oryzae*, *M. poae*, and *G. Graminis*, were downloaded from the Fungal Genome Initiative at Broad Institute of Harvard and Massachusetts Institute of Technology (Broad Institute). The protein sequences were compared using BLASTp (all-vs-all) with a maximum E-value of 1e-5. From the resulting BLASTp hits, OrthoMCL was used to identify homologous and paralogous relationships at 50% similarity. Markov clustering was used to further refine orthologous

clusters (Li *et al.*, 2003). OrthoMCL clusters that contained at least one gene from each Magnaporthaceae species were parsed and transcripts for genes within each cluster were retrieved from the Broad Institute transcript files using custom Python scripts.

CodeML on M. oryzae Orthologous Clusters

Magnaporthaceae species were parsed and transcripts for genes within each cluster were retrieved from the corresponding Broad Institute transcript files using custom Python scripts. The paired sequence files were aligned using command line MUSCLE v3.8.31 (Edgar, 2004), reiterating the alignments until reaching convergence and generating phylogenetic trees from the second alignment iteration. Custom Python scripts removed alignment columns containing at least 65% gap characters. The nonsynonymous to synonymous (dN/dS) substitution rates were estimated using the CodeML algorithm as part of PAML v4.8 (Yang, 2007) and scripted using BioPython v1.65 (Chapman *et al*., 2000). Likelihood ratio tests of site-specific selection were used, comparing M1 (neutral) to M2 (selection) and M7 (beta) to M8 (beta & w) using the test statistic $2*(\ln L1-\ln L2) = 2\Delta L$. The cluster was considered undergoing positive selection if both the M1/M2 and M7/M8 likelihood ratio tests were significant under a chi-square test with $p < 0.05$.

Cluster Function, Repetitive Element Locality, and Secreted Protein Analysis

The putative function of genes or clusters were identified using the Blast2GO suite (Conesa *et al*. 2005) utilizing BLASTn, InterPro protein domain identification, and Gene Ontology annotation using the *Aspergillus* slim. InterProScan v5.14 was used to predict the function of unique genes not shared with the Gene Ontology annotation (Jones *et al*. 2014). Repetitive element analysis was performed using RepeatModeler and RepeatMasker programs. The RMBlast NCBI search engine within RepeatModeler created *de novo* repetitive element libraries, and custom perl scripts were used to determine the repetitive element flanking distance to the right or left for each gene in the three Magnaporthaceae genomes. To determine if there was correlation between repetitive element distance and dN/dS ratio, boxplots

were generated, and Mann-Whitney Rank Sum statistical tests were performed using SigmaPlot v12.5.

Secreted proteins were identified by first searching whole proteomes for a signal peptide within the protein sequence using SignalP v4.1 (Peterson *et al*. 2011). Those with a signal peptide were then searched for subcellular localization to a secretory pathway using TargetP v1.1 (Emanuelsson *et al.* 2000). To determine if the putative secreted proteins are enriched near repeat regions, proportions were set and tested with VasserStats. Additionally, those proteins undergoing purifying or diversifying selection were checked with the secretory protein pipeline to check for enrichment.

Figures

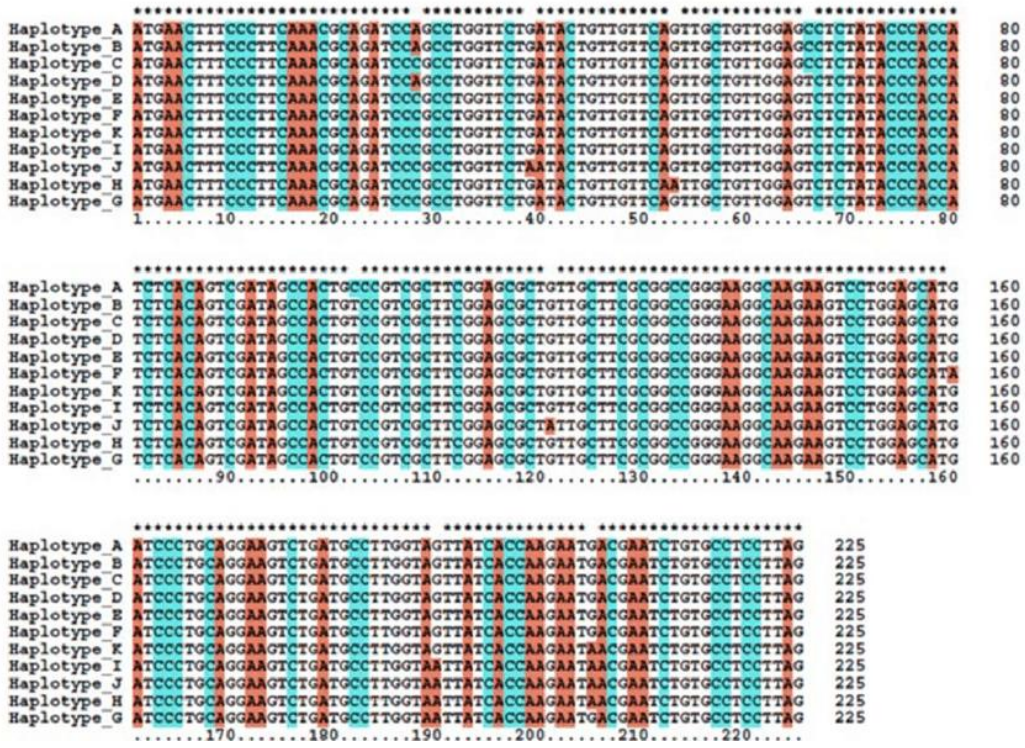

Figure 4.1 Alignment of the eleven SPD4 alleles: The major differences between the eleven alleles are single nucleotide polymorphisms.

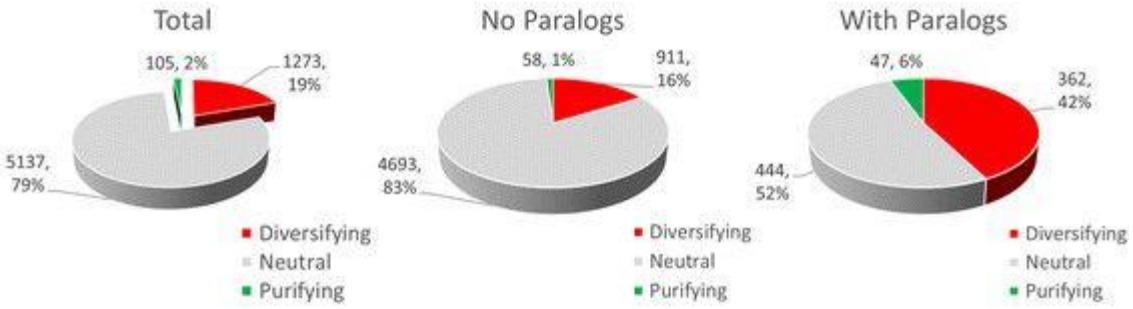

Figure 4.2 Distribution of Selection Pressures Among Clusters: Proportion of neutral, purifying, and diversifying clusters among all the clusters (left), all of the clusters containing paralogs (right), and those without paralogs (middle).

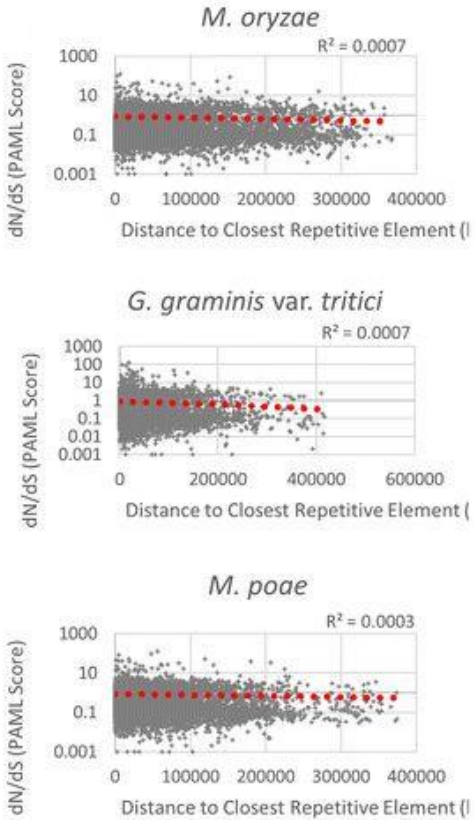

Figure 4.3 dN/dS Repetitive Element Distance: distance to the closest repetitive element and dN/dS ratio for the three fungal pathogens. Statistical analysis showed no correlation for all three.

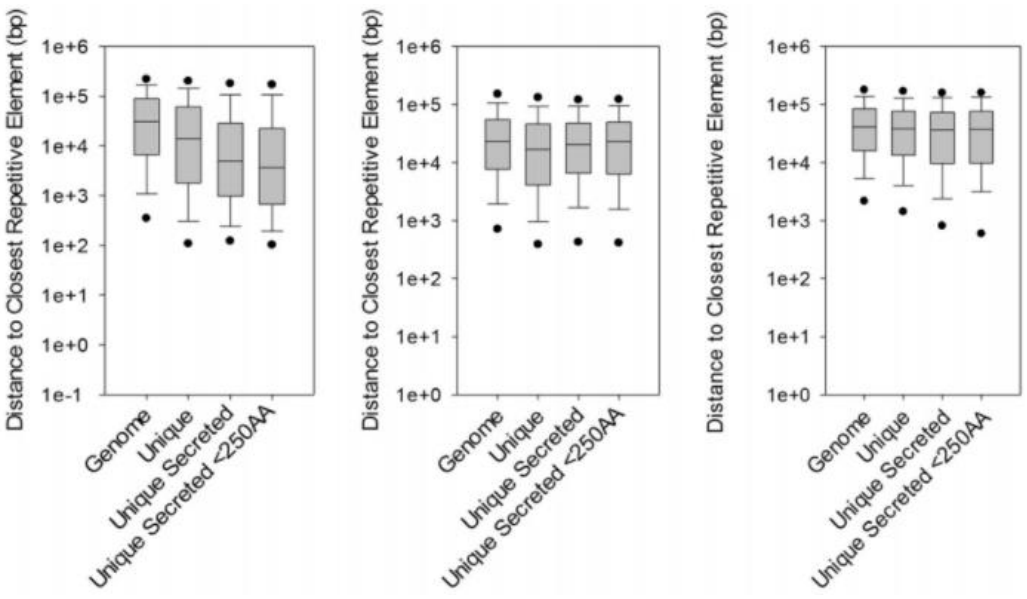

Figure 4.4 Unique Repetitive Element Distance: distance to the closest repetitive elements for the four cluster categories sorted by *M. oryzae* (left), *M. poae* (middle), and *G. graminis* var *tritici* (right).

References

Besi MI, Tucker SL, Sesma A (2009). Magnaporthe and its Relatives. Published Online.

Broad Institute of Harvard and MIT. https://www.broadinstitute.org/scientific-community/science/projects/fungal-genome-initiative/fungal-genomics

Chapman BA, Chang JT (2000). Biopython: Python tools for computational biology. ACM SIGBIO Newsletter 20, 15-19.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:18, 3674-3676.

Creuzburg K, Middendorf B, Mellmann A, Martaler T, Holz C, Fruth A, Karch H, and Schmidt H (2011). Evolutionary analysis and distribution of type III effector genes in pathogenic Escherichia coli from human, animal and food sources. Environmental Microbiology 13:2, 439-452.

Djamei, A, Schipper K, Rabe F, Ghosh A, Vincon V, Kahnt J, Osorio S, Tohge T, Fernie AR, Feussner I, Feussner K, Meinicke P, Stierhof YD, Schwarz H, Macek B, Mann M and Kahmann R (2011). Metabolic priming by a secreted fungal effector. Nature 478, 395-398.

Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:5, 1792-1797.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Journal of Molecular Biology 2000:300, 1005-1016.

Hogenhout SA, Van der Hoorn RAL, Terauchi R, Kamoun S (2009). Emerging Concepts in Effector Biology of Plant-Associated Organisms. Molecular Plant-Microbe Interactions 22:2, 115-122.

Howard RJ and Valent B (1996). Breaking and entering: Host penetration by the fungal rice blast pathogen *Magnaporthe grisea*. Annual Review of Microbiology 50, 491–512.

Huang, J., Si, W., Deng, Q., Li, P. and Yang, S. Rapid evolution of avirulence genes in rice blast fungus Magnaporthe oryzae. BMC Genetics 15:45, 1-10 (2014).

Ikeda KI, Nakayashiki H, Kataoka T, Tamba H, Hashimoto Y, Tosa Y, and Mayama S (2002). Repeat-induced point mutation (RIP) in Magnaporthe grisea: implications for its sexual cycle in the natural field context. Molecular Microbiology 45, 1355–1364.

Jones P, Binns D, Chang H, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S, Lopez R, Hunter S. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:9, 1236-1240 (2014).

Li L, Stoeckert CJ, and Roos DS (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Research 13, 2178-2189

O'Connell RJ, Thon MR, Hacquard S, Amyotte SG, Kleemann J, Torres MF, Damm U, Buiate EA, Epstein L, Alkan N, Altmuller J, Alvarado-Balderrama L, Bauser CA, Becker C,

Birren BW, Chen Z, Choi J, Crouch JA, Duvick JP, Farman MA, Gan P, Heiman D, Henrissat B, Howard RJ, Kabbage M, Koch C, Kracher B, Kubo Y, Law AD, Lebrun M, Lee Y, Miyara I, Moore N, Neumann U, Nordstrom K, Panaccione DG, Panstruga R, Place M, Proctor RH, Prusky D, Rech G, Reinhardt R, Rollins JA, Rounsley S, Schardl CL, Schwartz DC, Shenoy N, Shirasu K, Sikhakolli UR, Stuber K, Sukno SA, Sweigard JA, Takano Y, Takahara H, Trail F, van der Does HC, Voll LM, Will I, Young S, Zeng Q, Zhang J, Zhou S, Dickman MB, Schulze-Lefert P, van Themaat EVL, Ma L, and Vaillancourt LJ (2012). Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. Nature Genetics 44, 1060-1065.

Okagaki LH, Nunes CC, Sailsbery J, Clay B, Brown D, John T, Oh Y, Young N, Fitzgerald M, Haas BJ, Zeng Q, Young S, Adiconis X, Fan L, Levin JZ, Mitchell TK, Okubara PA, Farman ML, Kohn LM, Birren B, Ma L, and Dean RA (2015). Genome Sequences of Three Phytopathogenic Species of the Magnaporthaceae Family of Fungi. G3 (Bethesda) 5:12, 2539-2545.

Petersen TN, Brunak S, von Heijne G, and Nielsen H (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods 8, 785-786.

Rech GE, Sanz-Martin JM, Anisimova M, Sukno SA, and Thon MR (2014). Natural Selection on Coding and Noncoding DNA Sequences Is Associated with Virulence Genes in a Plant Pathogenic Fungus. Genome Biology and Evolution 6:9, 2368-2379.

Shang Y, Xiao G, Zheng P, Cen K, Zhan S, and Wang C (2016). Divergent and Convergent Evolution of Fungal Pathogenicity. Genome Biology and Evolution 8:5, 1374-1387.

Talbot NJ (2003). On the Trail of a Cereal Killer: Exploring the Biology of *Magnaporthe grisea*. Annual Review of Microbiology 57, 177-202.

Wilson RA and Talbot NJ (2009). Under pressure: investigating the biology of plant infection by *Magnaporthe oryzae*. Nature Reviews Microbiology 7, 185-195.

Yang, Z (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. Computer Applications in BioSciences 13, 555-556.

Yang Z (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution 24:8, 1586-1591.

Zhang H, Li D, Wang M, Liu J, Teng W, Cheng B, Dong S, Zheng X, and Zhang Z (2012). Nicotiana benthamiana MAPK cascade and WRKY transcription factor participate in Nep1Mo-triggered plant responses. Molecular Plant–Microbe Interactions 25, 1639-1653.

APPENDIX

Appendix A: Supplemental PCA Figures

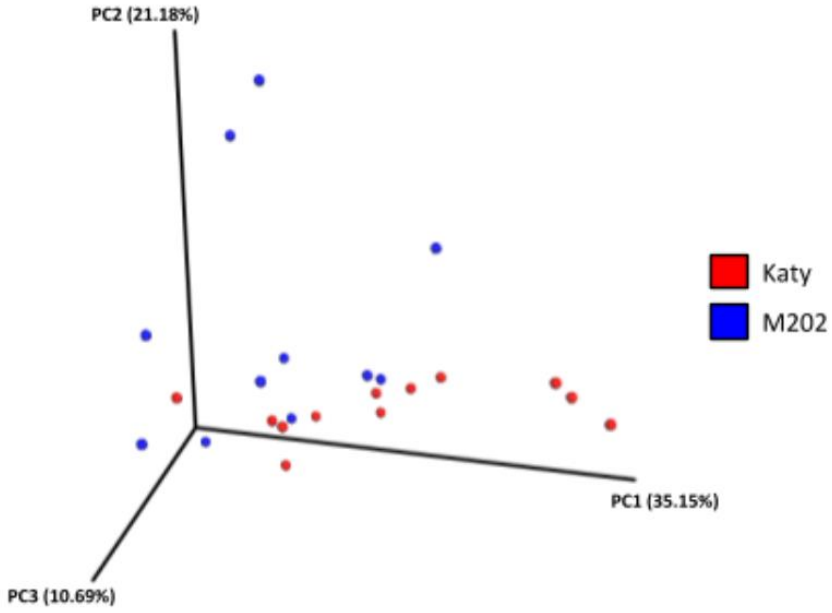

Figure A.1 Bacteria Genotype PCA

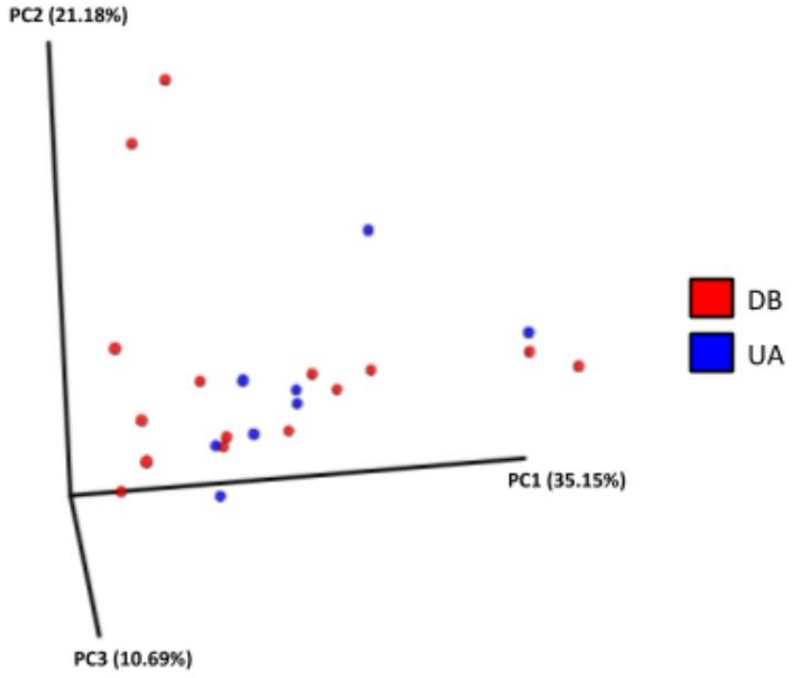

Figure A.2 Bacterial Location PCA

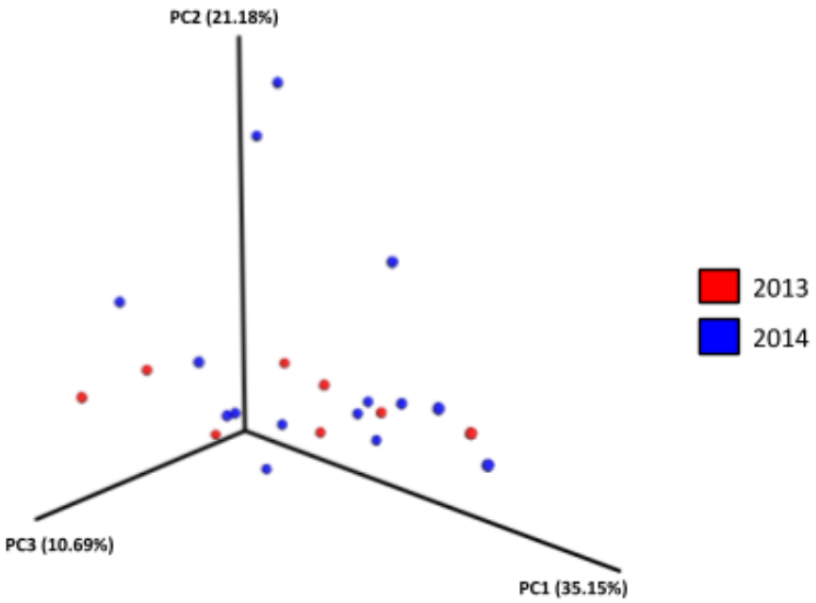

Figure A.3 Bacterial Year PCA

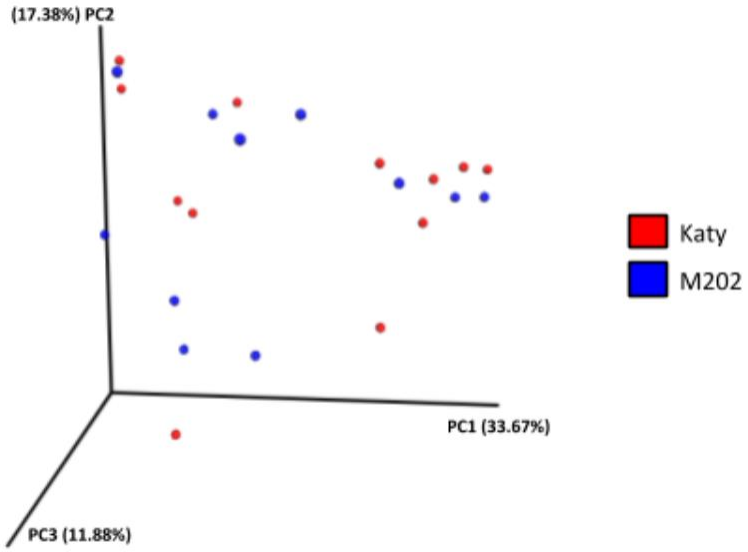

Figure A.4 Fungal Genotype PCA

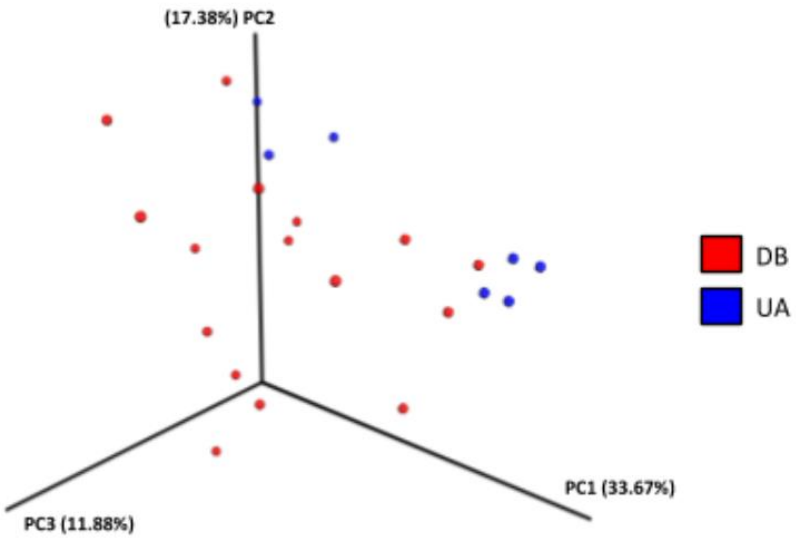

Figure A.5 Fungal Location PCA

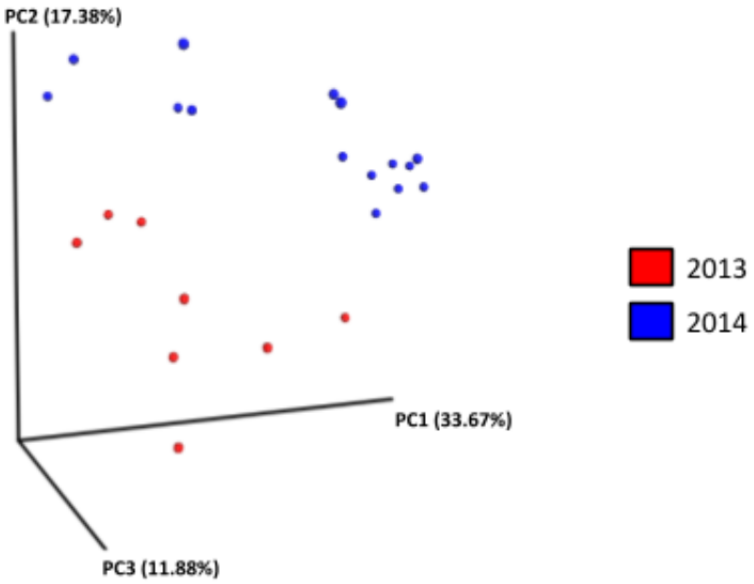

Figure A.6 Fungal Year PCA