INVESTIGATION OF MISCELLANEOUS ASSOCIATED MICROBIOME: EFFECTS OF BIOTIC AND ABIOTIC FACTORS

BY

DONGFANG LI

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Natural Resources and Environmental Sciences in the Graduate College of the University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Assistant Professor Katy D. Heath, Chair
Associate Professor Angela D. Kent, Director of Research
Associate Professor Adam S. Davis
Assistant Professor Erik J. Sacks
Associate Professor Thomas B. Voigt
Assistant Professor Anthony C. Yannarell
ABSTRACT

Long-term sustainable production of biofuel feedstock crops like Miscanthus species requires minimum anthropogenic energy inputs as well as the maintenance of soil fertility. Beneficial microbes, such as nitrogen-fixing bacteria, contribute to plant yield and fitness, representing under-explored plant nutrient sources. Optimizing the positive interactions between Miscanthus species and associated bacteria has high potential to support sustainable crop production. Bacterial community functions are directly related to community structures, which rely on surrounding biotic and abiotic factors. Therefore, before Miscanthus-bacteria interactions can be optimized for better plant productivity, it is necessary to investigate the factors that regulate community assemblage of these microbes.

The goal of this dissertation is to identify factors that significantly explain the Miscanthus-associated bacterial community variation. To achieve this goal, a combination of observational and experimental approaches was used. Miscanthus plants were collected from native, naturalized and cultivated habitats. These habitats represent a broad range of biotic and abiotic factors. Both Miscanthus endophytic compartment (rhizomes) and their surrounding rhizosphere soil were collected from each plant because both endophytes and rhizosphere bacteria are known to contribute to mutualistic plant-microbe interactions. Here, I accessed and compared both endophytic and rhizosphere bacterial communities, and evaluated whether plant genotypes and soil edaphic factors may influence them differently. As one of the most important beneficial bacteria groups, Miscanthus-associated diazotrophs assemblages were also examined in this work.
I found endophytic and rhizosphere communities correlated differently with their surrounding environment. Overall, I found that endophytic communities were more likely to be genotype-specific than the rhizosphere bacteria while the rhizosphere communities tended to change when soil conditions changed. The effect of agricultural practices could be hard to maintain and not significantly change microbial communities. Therefore, the goal of optimizing rhizosphere communities by changing soil conditions cannot be achieved easily. Instead, Miscanthus-associated endophytes are relatively stable to environmental change, and I identified some key endophytic-enriched taxa were shared among Miscanthus sites at a global scale. Hence, facilitate beneficial taxa to colonize Miscanthus endophytic compartment during rhizome propagation or pre-treatment of Miscanthus seeds with beneficial strains may be effective strategies to enhance the mutualism between Miscanthus and bacteria.

Even closely related Miscanthus genotypes tended to harbor very different endophytic N-fixing bacteria. On one hand, this result indicates the possibility of identifying plant genetic markers that control endophytic bacteria recruitment. On the other hand, the contribution of beneficial bacteria to Miscanthus must be evaluated on a genotype-by-genotype basis.

I observed many highly similar bacteria taxa from sites located on different sides of the world, indicating plant-associated taxa might not be dispersal limited. However, the interconnections between bacteria species are context-depend. Miscanthus-associated bacterial taxa were much more tightly connected with each other in the native sites than in the naturalized sites. This result indicates that the presence of similar bacteria taxa
does not necessarily lead to similar functions. Understanding the plant microbiome assemblage is simply the first step exploring the plant-bacteria interactions.

This work contributes to our knowledge of the plant-associated microbiome, which provides guidance for optimizing *Miscanthus*-microbe interactions for the long-term sustainable cultivation of *Miscanthus* species.
ACKNOWLEDGEMENTS

I would like to express my greatest gratitude and appreciation to those helped me accomplish this study. First and foremost, I would like to thank my advisor, Dr. Angela Kent, for her continued advice, encouragement and support during graduate school. I am deeply grateful for your enthusiasm, openness to ideas, as well as the faith and confidence you placed on me. I would like to sincerely thank all other committee members - Dr. Anthony Yannarell, Dr. Katy Heath, Dr. Adam Davis, Dr. Thomas Voigt, and Dr. Erik Sack for supporting my work, providing inspirational ideas and asking insightful questions.

This work would not have been possible without those who provided technical guidance and assistance. I would like to thank Dr. Jeff Dangl from The University of North Carolina at Chapel Hill for kindly providing support for sequencing Miscanthus-associated microbiome and welcoming me to use computational resources in his lab. I owe special thanks to Dr. Dangl’s graduate students Scott Yourstone and Sur Herrera Paredes for patiently helping me get started with high-throughput DNA sequencing data analysis. I also wish to acknowledge Dr. Lindsay Clark and Melina Salgado from Dr. Erik Sacks’s lab for conducting the microsatellite analysis on the native Miscanthus plants. Thanks also extended to Dr. Lauren Quinn, Dr. Jacob Barney, Ryan Dougherty, Dr. Eoghan Smyth, Corey Mitchell, L. Hou, I. Tseng, and C. Kuoh for field and lab assistance.

I take this opportunity to thank my past and present lab-mates. A big thanks goes to Yi Lou for always willing to offer help, and always willing to listen and discuss research problems with me. I extend particular thanks to Daniel, Sara, Ariane, Yu-rui,
Chinmay, Dora, Natalie Stevenson, Alyssa, Sam and Maya. Thank you all for your generous help and constructive feedback.

Last but not least, I would like to express thanks to my family members - my parents Xiaoqiu Feng and Sulun Li, and my husband Jing Jin. Thank you for always being there for me, and supporting my life decisions.
# TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION........................................................................................................... 1

CHAPTER 2: PLANT AND SOIL EFFECTS ON BACTERIAL COMMUNITIES ASSOCIATED WITH *MISCANTHUS ×GIGANTEUS* RHIZOSPHERE AND RHIZOMES........................................................................................................ 14

CHAPTER 3: PLANT MICROBIOME ASSOCIATED WITH NATIVE *MISCANTHUS* PLANTS.......................................................................................................................... 45

CHAPTER 4: COMPARISON OF BACTERIAL COMMUNITIES ASSOCIATED WITH NATIVE AND NATURALIZED *MISCANTHUS SINENSIS* .................................................................................. 82

CHAPTER 5: THE EFFECTS OF PLANT GENOTYPES AND LOCAL SOIL ON DIAZOTROPHS ASSOCIATED WITH *MISCANTHUS* ............................................................................... 107

CHAPTER 6: SUMMARY AND CONCLUSION................................................................................ 138

APPENDIX A: SUPPLEMENTARY INFORMATION FOR CHAPTER 2 ......................... 144

APPENDIX B: SUPPLEMENTARY INFORMATION FOR CHAPTER 3 ......................... 145

APPENDIX C: SUPPLEMENTARY INFORMATION FOR CHAPTER 4 ......................... 146

APPENDIX D: SUPPLEMENTARY INFORMATION FOR CHAPTER 5 ......................... 149
CHAPTER 1: INTRODUCTION

Increasing concerns about energy security, degradation of the environment have highlighted the importance of developing, implementing, and intensifying high-yielding bioenergy crops production. *Miscanthus*, particularly one high-yield genotype *Miscanthus × giganteus* is one of the most promising bioenergy crops (Heaton et al., 2010). *Miscanthus* is a genus of perennial rhizomatous C$_4$ grass (Lewandowski et al., 2000) originating from East Asia. It shows great capacity for adapting to various tropical and subtropical climate and soil conditions. Long-term trials cultivating *M. × giganteus* as a biofuel candidate have been successfully established in many locations throughout Europe and the United States (Christian et al., 2008; Heaton et al., 2010; Lewandowski et al., 2003).

Sustainable production is always an important issue in agricultural systems, and it is particularly important for biofuel feedstocks, given that sustainable energy supply is the strongest motivation for developing bioenergy crop. Minimizing anthropogenic energy inputs, including fertilizer inputs is desirable for economically and energetically sustainable bioenergy production. Merely relying on soil nutrient will deplete soil fertility and harm natural biogeochemical cycles, but improving the mutualistic interactions between bioenergy crops and their associated microbes may help to meet these goals.

Plant-associated beneficial microbes contribute significantly to plant growth and health (Berendsen et al., 2012; Compant et al., 2010; Hallmann et al., 1997; Reinhold-Hurek & Hurek, 2011). Terrestrial plants, especially their belowground structures, harbor and are surrounded by a large number of bacteria. There are about $10^4$ - $10^8$ bacteria per gram of plant root, and $10^6$ - $10^9$ per gram of rhizosphere soil (Bulgarelli et al., 2013). Admittedly, some of these bacteria are pathogens that cause plant diseases. However, such pathogenic interaction is simply an extreme
case of all the diverse plant bacteria interactions (Bulgarelli et al., 2013). Plant growth-promoting bacteria conduct various functions to bring multiple benefits to their hosts. These functions include nitrogen fixation (Reinhold-Hurek & Hurek, 1998), mineral phosphates solubilization (Rodríguez & Fraga, 1999), plant hormone production (Bulgarelli et al., 2013), and pathogen suppression (Kloepper et al., 1980; Whipps, 2001). Previous work has shown that Miscanthus species actively interact with beneficial bacteria (Straub et al., 2013) and benefit from their bacterial associates (Ye et al., 2005), especially from nitrogen fixers (Davis et al., 2010; Keymer & Kent, 2014).

Biologically available nitrogen (N) is typically the most limiting nutrient in terrestrial ecosystems, and the inputs of N fertilizer account for a significant portion of fossil fuel use in the agricultural systems (Vitousek et al., 1997). Miscanthus species generally have a high N-use efficiency. After senescence and before harvesting, N is translocated from aboveground Miscanthus tissues to their rhizomes for future uses (Scally et al., 2001). Nevertheless, M. × giganteus still prefers external N inputs. A review study conducted by Heaton implies that M. × giganteus yields showed a significant positive response to nitrogen fertilization inputs (Heaton, Voigt, & Long, 2004). The result is representative because it is based on 21 peer-review articles and 174 representative observations, which were almost all the available records at the time of the study. Additionally, the yield differences due to nitrogen input increased with years of cultivation (Miguez et al., 2008). Successive Miscanthus biomass harvests gradually remove N from soil. Cadoux and associates (2012) summarized studies that investigated Miscanthus and its nutrient usage. They found on average 4.9 gram of N per kilogram of dry biomass were removed by harvests. Similarly, Dohleman et al. (2012) reported that Miscanthus shoots contained 193 kg N/ ha during December harvesting. This number is even much higher than the amount of N input
even for those Miscanthus trials with high N treatment (Lewandowski et al., 2003; Maughan et al., 2012). These observations call into question the prospects for long-term sustainable production of Miscanthus - the highly nitrogen using efficiency of Miscanthus could not prevent the eventual depletion of soil nitrogen in long-term cultivation.

Nitrogen budget modeling results and $^{15}$N isotope experiment revealed that besides fertilizer and aerial deposition, diazotrophs potentially contribute to the nitrogen content of Miscanthus. Davis et al. used DAYCENT model simulations to evaluate biogeochemical cycles associated with the cultivation of $M. \times$giganteus (Davis et al., 2010). Model results demonstrated a deficit in the nitrogen budget – that is, output N exceeds the input N. After testing all other potential nitrogen sources of nitrogen, the authors hypothesized that nitrogen fixation may contribute up to 25 g N m$^{-2}$y$^{-1}$ of the nitrogen requirement of $M. \times$giganteus. In support of the hypothesis, seven $nifH^+$ strains of Enterobacteriaceae were isolated from Miscanthus rhizome and stem tissues, and these strains showed positive in vitro nitrogenase activities. In vitro nitrogenase activities assays also carried out on Miscanthus rhizomes providing additional evidence in support of nitrogen fixation (Davis et al., 2010). Following the work of Davis et al., Keymer and Kent conducted a $^{15}$N isotope dilution experiment to quantify the amount of N contributed by diazotrophs to the first-year field $M. \times$giganteus plants. Their results indicated that 16% of plant N came from N fixation process, and they also detected that the diazotroph assemblage changed according to plant N demand dynamics, which indicated that plant associated diazotrophs contributed to plant N (Keymer & Kent, 2014). To conclude, the quantity of Miscanthus N that contributed by biological nitrogen fixation (BNF) process was measurable but small.
More promising results have been reported for plant species that closely related to *Miscanthus*. *Setaria viridis*, a species from family Poaceae, is closely related to bioenergy grasses such as *Miscanthus*, switchgrass, maize, sorghum and sugarcane, and services as a model for the C₄ photosynthesis for these bioenergy crops (Brutnell et al., 2010). ¹³N tracer studies provide clear evidence that *Setaria viridis* takes up biologically fixed nitrogen and incorporates it into plant protein (Pankievicz et al., 2015). Inoculating *Setaria viridis* with diazotrophic strain *Azospirillum brasilense* significantly increased plant biomass under a nitrogen-deficient condition (Pankievicz et al., 2015). Sugarcane (*Saccharum officinarum*), a very close relative of *Miscanthus*, has been long known to benefit from their association with nitrogen fixers (Baptista et al., 2014; Boddey, 1995; Taulé et al., 2012; Thaweenut et al., 2011; Urquiaga et al., 2012). BNF could help to meet 77% of nitrogen demand for some sugarcane genotypes (Baptista et al., 2014). Elephant grass (*Pennisetum purpureum*), another C₄ biofuel candidate in the family Poaceae, has also been reported to gain 18% - 70% plant nitrogen through BNF, with variations due to genotypes and soil conditions (de Morais et al., 2012; Reis et al., 2001). Biologically fixed nitrogen represents a promising N source for *Miscanthus* species and related C₄ species, but the contribution of diazotrophs could be context-dependent.

Both biotic and abiotic factors determine community structure and function of plant-associated bacteria (Bulgarelli et al., 2013; Hallmann et al., 1997; Rosenblueth & Martinez-Romero, 2006). The plant-associated bacteria assemblages can be considered as results of successive sets of ecological filters. Soil edaphic factors such as pH, nutrients represent an ecological filter that generates local species pools. Along with plant rhizodeposits, these factors limit the microbes that have access to the plant rhizosphere (Bulgarelli et al., 2013; Fierer et al., 2007; Lauber et al., 2009; Marschner et al., 2001). Plant tissue structures, defense systems, and
plant - microbe signaling pathways represent another layer of ecological filters and enable plants to recruit and select endophytic microbes from rhizosphere microbes (Bulgarelli et al., 2013; Hardoim et al., 2008). As a result, the relative influence of biotic factors (plant genotypes) and abiotic factors (soil edaphic factors) differentially influence endophytic and rhizosphere bacteria. Endophytic populations may be better protected from environmental stresses than rhizosphere populations, and rhizosphere population may be less specific to plant genotypes (Bulgarelli et al., 2013).

Currently, all the Miscanthus species used in the fields are genetically unmodified. M. ×giganteus is a triploid grass that can only reproduce by rhizome transplanting and in vitro culture, which certainly reduces its invasiveness (Hodkinson et al., 2002), but increases the field establishment costs (Lewandowski et al., 2003). Due to its sterility, M. ×giganteus has a narrow genetic variation (Heaton et al., 2010), and, therefore, large-scale cultivation of this uniform species will be vulnerable to ecological threats, such as pest and pathogen attacks (Heaton et al., 2010). Several other Miscanthus species have shown potential bioenergy production and desirable traits such as cold-tolerance, late flowering and low lignin content (Kim et al., 2012; Shumny et al., 2010; Zub & Brancourt-Hulmel, 2010). In conclusion, improved Miscanthus species are possible and are needed for long-term sustainable production as well as high profit. In fact, breeding new Miscanthus varieties is ongoing (Heaton et al., 2010; Sacks et al., 2013).

Basic questions about the ecology of plant-associated bacteria need to be answered if plant-microbe associations are to be managed and maintain for large-scale sustainable biofuel production. As mentioned earlier, both soil conditions and plant genotype highly influence plant-associated microbial community structure and function. Investigating ecological drivers that structure Miscanthus-associated microbial populations will provide guidance in breeding and
selecting new *Miscanthus* varieties that leverage mutualistic plant-microbe interactions, as well as identify key environmental elements that need to be monitored.

The central hypothesis of this dissertation is that biotic and abiotic factors form strong and predictable ecological filters that determine *Miscanthus*-associated plant microbiomes.

To address different aspects of this hypothesis, four studies were conducted using a combination of observational and experimental approaches. To be more specific, the following hypotheses were addressed in these studies:

1) Endophytic bacteria living inside of the plant are directly influenced by plant environment and receive indirect influences from the surrounding soil, whereas rhizosphere communities receive direct influences from the bulk soil. The key ecological drivers and their influence on endophytic and rhizosphere communities are different. Here I hypothesize that endophytic communities are more sensitive to plant genotype differences, whereas rhizosphere communities are more sensitive to bulk soil differences.

2) The majority of endophytic taxa are originated from rhizosphere soil, but they are not a random subset of the rhizosphere communities. To access and successfully colonize the inner plant tissues, bacteria taxa need to have a specific combination of characteristics. Therefore, here I hypothesize that only specific bacterial taxa are able to pass the successive ecological filter and colonize inside *Miscanthus* plant tissue.

3) Plant traits controlling microbe colonization are often plant genotype-specific. Thus, I hypothesize that closely related *Miscanthus* genotypes harbor different microbes. That is, breeding new *Miscanthus* varieties will alter the *Miscanthus*-associated microbiome.
To address these hypotheses, the Miscanthus-associated total and diazotroph bacterial communities were surveyed by investigating the microbiome of Miscanthus plants collected from natural environments (native and introduced), and cultivated fields. I also experimentally cultivated several Miscanthus genotypes in two different soil sources to evaluate how plant genotypes and soil sources determine the total bacterial and diazotroph assemblages. By addressing these hypotheses, I identified generalizable microbial responses to Miscanthus plants and soil.

Distinct total bacterial and diazotroph communities were found in the endophytic compartment and the rhizosphere soil of M. ×giganteus, and they responded differently to the surrounding environmental conditions. In Chapter 2 - Plant and soil effects on bacterial communities associated with Miscanthus ×giganteus rhizosphere and rhizomes, to exclude the variation introduced by plant genotypes, I investigated microbial communities associated with M. ×giganteus rhizosphere and rhizomes from four different agricultural fields. The rhizosphere communities were clearly distinct among sampling locations, while the endophytic communities from the different locations were less distinct. Following this observation, I found that the rhizosphere communities correlated strongly with environmental factors, and the correlation between endophytic communities and surrounding soil was much weaker. Additionally, total bacteria and diazotrophs also showed correlations with different sets of environmental factors. Results from this chapter suggest that endophytic microbial communities are robust to environmental changes and suggest the possibility to introduce stable beneficial microbial via propagation.

In Chapter 2, I also observed that the same bacterial taxa were found to associate with M. ×giganteus endophytic and rhizosphere compartment in all sampling sites. I investigated if a
similar pattern occurred in Miscanthus species living in their native habitats. In Chapter 3 - Plant microbiome associated with native Miscanthus plants, two native Miscanthus species (M. sinensis and M. floridulus) living in a wide range of soil nutrient conditions were collected throughout Taiwan. I identified groups of bacterial taxa that were consistently enriched in the endophytic compartments (EC-enriched) or in the rhizosphere soil (R-enriched). The number of such bacterial taxa was small, but they comprised a considerable portion (30% - 60%) of the total bacterial communities. Moreover, the relative abundance of these compartment-enriched taxa captured the patterns of total bacterial community distance amount sites. Therefore, these compartment-enriched taxa represent the core Miscanthus-associated members in their corresponding compartment. I further investigated if highly similar (if not identical) compartment-enriched taxa can be found at the global scale in Chapter 4. In the study describe in Chapter 4, additional M. sinensis plants were collected in their naturalized habitat - Eastern United States. The majority Miscanthus-associated taxa can be found in both native and naturalized habitats. 16 EC-enriched and 55 R-enriched bacterial taxa were shared between native and naturalized habitats. These shared compartment-enriched taxa comprise 20 - 30% of total bacterial communities. The majority of EC-enriched bacterial taxa were from Burkholderiaceae and Enterobacteriaceae, and the majority of R-enriched taxa belonged to Bradyrhizobiaceae and Hyphomicrobiaceae. Surprisingly, even though these taxa were present in both habitats, patterns of co-occurrence were completely different. The community assembly patterns in native site indicate these communities followed a niche-driven assembly role while the communities in naturalized site fit an outcome described by neutral theory (Faust & Raes, 2012).
Plant microbiome from both endophytic compartment and rhizosphere rely on plant species and the local environment. In Chapter 3, I compared the relative importance of plant genetic distance and the variation of the environment in shaping plant-associated bacteria communities. In general, the results indicated these two aspects correlated with bacterial assembly pattern to a similar extent. In Chapter 5, I experimentally introduced four closely related Miscanthus genotypes - triploid M. ×giganteus, diploid M. sinensis, diploid M. sacchariflorus and potential parental genotype tetraploid M. sacchariflorus, and examined how these treatments influence the diazotroph taxa that are recruited by Miscanthus. My results show that diploid M. sinensis tended to recruit many more endophytic diazotroph taxa than the other three genotypes. Additionally, quite a number of endophytic diazotroph taxa are more enriched in diploid M. sinensis than in other genotypes. In contrast, diploid M. sacchariflorus limited the number of diazotroph taxa living in their rhizosphere soil, and some rhizosphere taxa were specifically enriched when associated with M. sacchariflorus. As a result, I propose that diazotroph-genotype interactions need to be taken into account when optimizing the positive association between Miscanthus and diazotrophs. There is a possibility to modify and improve Miscanthus associated microbiome through breeding.

The mutualistic interactions between non-legume crops and endophytic diazotrophs represent under-investigated nutrient resources that could enhance sustainable crop production (Boddey et al., 1995; Reinhold-Hurek & Hurek, 1998). There is much evidence suggesting that microbial community functions are directly related to community composition (Zak et al., 2003). Hence, the first steps for enhancing the benefits derived from Miscanthus-microbe association are to explore the endophytic community composition and to investigate the biotic and abiotic factors that influence this composition. In this dissertation, I have contributed to the knowledge
of the Miscanthus-associated microbiome under different environmental conditions, identified a list of compartment-enriched bacterial taxa warranting further study, and also provided guidance to optimize Miscanthus-microbe interactions for the long-term cultivation of Miscanthus as a biofuel feedstock.

References


CHAPTER 2

PLANT AND SOIL EFFECTS ON BACTERIAL COMMUNITIES ASSOCIATED WITH *MISCANTHUS × GIGANTEUS* RHIZOSPHERE AND RHIZOMES

Abstract

Bacterial assemblages, especially diazotroph assemblages residing in the rhizomes and the rhizosphere soil of *Miscanthus × giganteus* contribute to plant growth and nitrogen use efficiency. However, the composition of these microbial communities has not been adequately explored, nor have the potential ecological drivers for these communities been sufficiently studied. This knowledge is needed for understanding and potentially improving *M. × giganteus* -microbe interactions, and further enhancing sustainability of *M. × giganteus* production. In this study, cultivated *M. × giganteus* from four sites in Illinois, Kentucky, Nebraska, and New Jersey were collected to examine the relative influences of soil conditions and plant compartments on assembly of the *M. × giganteus*-associated microbiome. Automated ribosomal intergenic spacer (ARISA) and terminal restriction fragment length polymorphism (T-RFLP) targeting the *nifH* gene were applied to examine the total bacterial communities and diazotroph assemblages that reside in the rhizomes and the rhizosphere. Distinct microbial assemblages were detected in the endophytic and rhizosphere compartments. Site soil conditions had strong correlation with both total bacterial and diazotroph assemblages, but in different ways. Nitrogen treatments showed no significant effect on the composition of diazotroph assemblages in most sites. Endophytic compartments of different *M. × giganteus* plants tended to harbor similar microbial communities across all sites, whereas, the rhizosphere soil of different plant tended to harbor diverse microbial

---

1 This is the peer reviewed version of the following article: Li D, Voigt TB, Kent AD (2015) Plant and soil effects on bacterial communities associated with *Miscanthus × giganteus* rhizosphere and rhizomes. GCB Bioenergy. DOI: 10.1111/gcbb.12252. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.
assemblages that were distinct among sites. These observations offer insight into better understanding of the associative interactions between *M. ×giganteus* and diazotrophs, and how this relationship is influenced by agronomic and edaphic factors.

**Keywords:** bioenergy feedstock, *Miscanthus ×giganteus*, bacterial community composition, diazotroph, endophytic compartment, rhizosphere
Introduction

Biologically fixed nitrogen provided by diazotrophs represents a promising nitrogen source for sustainable production of non-leguminous crops (Bhattacharjee et al., 2008; James, 2000), which is an especially desired trait for a biofuel feedstock such as Miscanthus × giganteus. As a perennial C4 grass, M. × giganteus has high nitrogen use efficiency (Long et al., 2001), however, successive biomass harvests of M. × giganteus still leads to substantial loss of soil nitrogen (Dohleman et al., 2012). Yet surprisingly, the biomass yield of M. × giganteus does not appear to be limited by low nitrogen input. Multiple field experiments and meta-analysis studies have indicated that the biomass yield of M. × giganteus shows little or no response to nitrogen additions (Cadoux et al., 2012; Heaton et al., 2004; Maughan et al., 2012; Miguez et al., 2008). Plants obtaining fixed nitrogen from association with diazotrophs have been observed in multiple non-legume grasses from family Poaceae, which includes Miscanthus. Multiple studies have demonstrated that sugarcane (Saccharum officinarum), a very close relative of Miscanthus, benefits from biological nitrogen fixation (BNF) (Baptista et al., 2014; Boddey, 1995; Taulè et al., 2012; Thaweenut et al., 2011; Urquiaga et al., 2012). Depending on local soil conditions and plant genotypes, sugarcane may obtain up to 77% of their nitrogen content through BNF (Baptista et al., 2014). Elephant grass (Pennisetum purpureum), another C4 biofuel candidate in the Poaceae family, also has been shown to obtain plant nitrogen (18% - 70%) from BNF (de Morais et al., 2012; Reis et al., 2001). Similar to these species, nitrogen-balance experiments and modeling studies suggest that the unaccounted for nitrogen source supporting M. × giganteus in soils having limited nitrogen is likely provided by diazotrophs through biological nitrogen fixation (BNF) (Christian et al., 2008; Davis et al., 2010). Indeed, our initial field trials of M. × giganteus have demonstrated that 16% or more of plant N acquired over a single growing
season could be attributed to N fixation (Keymer & Kent, 2014). Even though BNF seems to contribute less N to Miscanthus than to other Poaceae, the DAYCENT model used to estimate biogeochemical cycles associated with cultivation of M. × giganteus indicated that BNF contributed up to 25 g ha\(^{-1}\) m\(^{-2}\) to the growth of M. × giganteus (Davis et al., 2010). Thus far, most studies investigating the association between diazotrophs and M. ×giganteus plants have focused on isolates cultured from M. ×giganteus tissues (Davis et al., 2010; Eckert et al., 2001; Kirchhof et al., 2001; Miyamoto et al., 2004; Straub et al., 2013). While these previous studies have demonstrated the presence of associative diazotrophs in M. ×giganteus rhizosphere and tissues, the knowledge of the composition of diazotroph communities associated with M. ×giganteus is limited, and the potential ecological drivers of these communities have not been sufficiently explored. Lack of this information limits our capability further investigate the contribution of uncultured diazotrophs to plant nitrogen and sustainability.

Both the plant and the surrounding soil environment exert influences on diazotrophs and other microorganisms residing in the rhizosphere and endophytic compartment (Marschner et al., 2004; Reinhold-Hurek & Hurek, 1998). Soil chemical factors such as pH, nutrients, and soil organic matter act as ecological filters that limit the microbes that have access to the plant rhizosphere and create local species pools (Fierer et al., 2007; Lauber et al., 2009; Marschner et al., 2001). Plant tissue structures, defense systems, and the interactions between plant and microbes enable plants to recruit and select endophytic microbes from rhizosphere microbes, representing a second layer of ecological filters (Bulgarelli et al., 2013; Hardoim et al., 2008). These successive sets of ecological filters ultimately determine the microbial assemblages associated with M. ×giganteus.
Comparing the microbial communities residing in endophytic compartment and rhizosphere of \textit{M. \times giganteus} plants cultivated in different geographic regions provides opportunities for exploring selective ecological factors, and insights into the interactions between plant and microbes. In this study, the putative endophytic compartments (for the purpose of brevity, we will refer this term as endophytic compartments) - represented by surface-sterilized rhizomes, and rhizosphere soil from genetically identical \textit{M. \times giganteus} were collected from cultivated sites in Illinois, Kentucky, Nebraska, and New Jersey to examine the relative influence of geography and plant compartments on assembly of the \textit{M. \times giganteus}-associated microbiome. We hypothesize that the distinct soil edaphic factors in these locations will lead to different total bacterial and diazotroph assemblages in these sampling locations. We also hypothesize that the plant’s effect on recruiting and retaining endophytes will yield distinct microbial communities in the endophytic compartment and the rhizosphere.

\textbf{Materials and methods}

\textit{Sampling sites}

Field sites were located in Urbana, IL, Lexington, KY, Mead, NE, and Adelphia, NJ. Soil pH, organic matter, total organic carbon, total nitrogen, bulk density, phosphorus, potassium, and iron (Table 2.1) for soil collect at depth of 0 – 30 cm (Table 2.1). Nitrogen, as urea, was applied at three rates (0, 60, 120 kg N ha\(^{-1}\)) to 4 replicated plots at each site. Site conditions and soil chemical analyses were previously described by Maughan \textit{et al.} (2012), previously published yield data for this study is presented in Table A.1. Here, all measures were standardized using the z-score transformation (Legendre & Legendre, 1998) before further data analysis.
Sample processing

Two rhizomes and the surrounding rhizosphere soil were collected from each of the 12 plots (3 N treatments \(\times\) 4 replicates) at each location in October 2011. Rhizome processing was conducted using the methods described by Keymer and Kent (2014). Rhizosphere soil was washed and collected from rhizomes using 40 mL sterile deionized water. Rhizosphere soil was then frozen at \(-80\) °C in sterile containers, and lyophilized prior to DNA extraction. Total genomic DNA was extracted from surface-sterilized rhizomes and lyophilized rhizosphere soil using the FastDNA SPIN Kit and FastDNA SPIN Kit for Soil (MP Biomedicals, Aurora, OH), respectively. For rhizosphere samples, genomic DNA was further purified using the cetyl trimethyl ammonium bromide (CTAB) extraction procedure to remove contaminating humic acids (Sambrook & Russell, 2001). DNA concentration for all samples was adjusted to a standard concentration of 10 ng/µL prior to DNA analyses.

Analysis of microbial community composition

In this study, automated ribosomal intergenic spacer analysis (ARISA) and terminal restriction enzyme fragment length polymorphism (T-RFLP) were used to access the microbial community compositions. Such DNA fingerprinting methods have been demonstrated to reveal patterns in beta diversity similarly to high throughput sequencing methods, and have the advantage of being relatively inexpensive and high throughput, allowing analysis of sufficient replicates to provide high statistical power for distinguishing treatment effects (Jami et al., 2014; Spencer et al., 2011; van Dorst et al., 2014).

Bacterial community composition and diversity were assessed using ARISA (Fisher & Triplett, 1999) as described previously (Kent et al., 2004; Yannarell & Triplett, 2005). Primers 1406f, 6-FAM-5’-TGYACACACCGCCCGT-3’ (universal, 16S rRNA gene), and 23Sr, 5’-
GGTTBCCCCATTTCR5’ (bacteria-specific, 23S rRNA gene) were used for ARISA polymerase chain reaction (PCR). The PCR contained 1X Tris buffer (50 mM Tris-HCl, pH 8.3 at 25 °C), 0.25 mg of bovine serum albumin per mL (NEB, Beverly, MA), 2.5 mM MgCl₂ (Promega, Madison, WI), 0.20 mM of dNTPs (NEB, Beverly, MA), 0.5 µM of each primer, 0.05 U/µL of Taq polymerase (Promega, Madison, WI) and 20 ng of DNA template in a final volume of 25 µL. The PCR cycling conditions include an initial denaturation at 94 °C for 2 min, followed by 30 cycles of 94 °C for 35 s, 55 °C for 45 s, and 72 °C for 2 min, with a final extension carried out at 72 °C for 2 min.

To assess the community composition of the diazotrophs, T-RFLP targeting the nifH gene was conducted for each sample. The nifH PCR was carried out using primers PolF (6-FAM-5’-TGCGAYCCSAARGCGBACT-3’) and PolR (HEX-5’-ATSGGCATCATYTCRCCGGA-3’) (Poly, Monrozier, et al., 2001). PCRs were conducted as described by Keymer and Kent (2014). The PCR buffer contained 1X GoTaq buffer (Promega, Madison, WI), 0.5 mg of bovine serum albumin per mL (NEB, Beverly, MA), and 2.0 mM MgCl₂ (Promega, Madison, WI), 0.20 mM of dNTPs (NEB, Beverly, MA), 0.5 µM of each primer, 0.05 U/µL of Taq polymerase (Promega, Madison, WI) and 20 ng of DNA template in a final volume of 25 µL. A touchdown PCR program was performed as follows: an initial denaturation step at 94 °C for 5 min, followed by 35 cycles of denaturation at 94 °C for 45 s, annealing at variable temperatures for 45 s, and extension at 72 °C for 45 s. In the first cycle, the annealing temperature was 64 °C. The 2nd and 3rd cycles had an annealing temperature of 62 °C. The 4th to 6th cycles had an annealing temperature of 60 °C, and the 7th to 10th cycles used an annealing temperature of 58 °C. The last 25 cycles used an annealing temperature of 56 °C. The program ended with a 10 min extension at 72 °C. After purifying the nifH amplicons with MinElute 96 UF PCR Purification Kit (Qiagen,
Valencia, CA), the amplicons were digested with restriction enzymes MboII and MnlI (NEB, Beverly, MA) for 16 hours at 37 °C following the protocol recommended by the enzyme manufacturer.

Denaturing capillary electrophoresis was performed by the Keck Center for Functional Genomics at University of Illinois (Urbana, IL) for ARISA PCR products and digested nifH PCR products using an ABI 3730xl Genetic Analyzer (PE Biosystems, Foster City, CA). Size-calling was carried out using GeneMarker 2.4 (SoftGenetics, State College, PA). Capillary electrophoresis results in minor run-to-run variations in observed vs. actual fragment length were resolved using the allele-calling features in GeneMarker before analysis. To include the maximum number of peaks while excluding background fluorescence, a threshold of 200 fluorescence units was used. The signal strength (i.e. peak area) of each peak was normalized to account for run-to-run variations in signal detection by dividing the area of individual peaks by the total fluorescence (area) detected in each profile, expressing each peak as a proportion of the observed community (Rees et al., 2004; Yannarell & Triplett, 2005).

Statistical analysis

Correlations between environmental factors and microbial communities were evaluated with canonical correspondence analysis (CCA) and linear regression. CCA was conducted on both the total bacterial community data generated by ARISA and diazotroph assemblage data generated by nifH T-RFLP. Normalized soil chemical and physical variables were included to calculate the explained variance. Longitude and latitude for each site were also included to capture unmeasured environmental variables. Correspondence analysis (CA) was used to visualize the community similarities among samples. The dispersion of samples within a group (within group similarity) was estimated using the betadisper function in the vegan package of R.
In order to test the correlation between microbial assemblages and environmental factors, the community similarity (1 – Bray-Curtis dissimilarity) was linearly regressed against the environmental similarity, represented by the Euclidean distance generated using all available normalized soil edaphic variables as well as longitude and latitude data.

The effect of N rates on the diazotroph communities was tested using permutational multivariate analysis of variance (PerMANOVA) (Anderson, 2001; McArdle & Anderson, 2001), and Bray-Curtis dissimilarity was used to measure the distance between all pairs of diazotroph communities. The significance of the N treatment was determined through 999 random permutations.

Redundancy analysis (RDA) of the nifH T-RFLP data was performed to select the terminal restriction fragments (TRFs) that were consistently enriched in either the rhizosphere or the endophytic compartment. We used the Bray-Curtis dissimilarity for diazotroph community as the response variable, and the predictor variable was the compartment containing the diazotroph community (rhizosphere or endophytic compartment). Given that we only had one predictor variable, the first RDA axis represented the variance explained by the plant compartment. Each TRF had one orthonormal species score on the first RDA axis. The more positive the orthonormal species score is, the more likely this TRF would be a rhizosphere specialist. Similarly, the more negative the score is, the more likely this TRF would be a putative endophytic specialist. Fifty TRFs with the highest likelihood of detection in the endophytic or rhizosphere compartment (100 TRFs total) were chosen for each site. These enriched TRFs were compared among sites, and the results were represented using a Venn diagram.

All CA and CCA were conducted using Canoco 5.6 (Biometris, Wageningen, The Netherlands). The calculation of community diversity, PerMANOVA, betadisper, linear
regression, RDA, and Venn diagrams were carried out with the packages vegan (Oksanen et al., 2011), stats, and VennDiagram (Chen & Boutros, 2011) for the R statistical environment (R Development Core Team, 2010).

**Results**

*Edaphic factors*

The results demonstrated that edaphic factors explained a larger amount of variance for the rhizosphere microbial community composition than for the putative endophytic communities (Table 2.2). For the rhizosphere bacterial community (assessed by ARISA), all the listed factors except for total soil N explained a significant portion of variance. All factors except for soil pH were significant for explaining variance in the rhizosphere diazotroph community structure (assessed by *nifH* T-RFLP). For microbes residing in the endophytic compartment, a significant portion of variance in the composition of total bacterial communities was explained by total organic carbon, bulk soil density, soil phosphorus and potassium, while variance in endophytic diazotrophs communities was explained by total soil nitrogen, total organic carbon, soil iron and potassium. Soil pH was a significant explanatory variable for the composition of the total bacterial community residing in the rhizosphere, but not for the endosphere bacterial assemblages or diazotroph communities. Interestingly, soil N content did not explain a significant amount of variance in the total bacterial assemblage, but was a significant edaphic factor for diazotroph communities. Diazotroph communities in both the endosphere and rhizosphere showed no response to N treatments except in the NE rhizosphere.

As shown in Figures 2.1 and 2.2, location was the main factor in shaping total bacterial community composition (PerMANOVA: endophytic total bacteria, $R^2 = 0.2568$; rhizosphere total bacteria, $R^2 = 0.2677$). Similarly, location also explained a statistically significant portion of
variance in the diazotroph communities. However, the relationship between location and endophytic diazotroph communities (PerMANOVA: endophytic diazotrophs, $R^2 = 0.1304$) was less pronounced than that for rhizosphere diazotroph communities (PerMANOVA: rhizosphere diazotrophs, $R^2 = 0.3449$).

Significant trends of decay in community similarity with environmental distance (including the effects of both geographic distance and dissimilarity in soil factors) were observed when linear regression was applied; the trends were relatively stronger for rhizosphere communities (regression $R^2$: total bacteria $R^2 = 0.3633$, diazotrophs $R^2 = 0.1681$). Meanwhile, the distance-decay trends were weak for the endophytic communities (regression $R^2$: total bacteria $R^2 = 0.08139$, diazotrophs $R^2 = 0.03151$). Therefore, bacterial communities were more similar to each other when environmental conditions were similar, and the composition of bacteria inhabiting the rhizosphere was more sensitive to external environmental changes. Comparable results were obtained using an eigenanalysis-based method: a considerable amount of variance among bacterial communities can be explained by the soil factors (Table 2.2).

**Plant compartment**

Across all sites, both total and diazotroph bacteria had distinct communities residing in the endophytic and rhizosphere compartments (Figure 2.3; PerMANOVA: total bacteria $R^2 = 0.2089$; diazotrophs $R^2 = 0.2893$). Additionally, microbial assemblages residing in the endophytic compartments were more similar to each other than assemblages in the rhizosphere soil (Betadisper: total bacteria $p < 0.001$; diazotrophs $p < 0.001$). That is, endophytic compartments of different *M. × giganteus* plants tended to harbor similar microbial communities across all sites, whereas, the rhizosphere soil of different plants tended to harbor diverse microbial communities.
When examining the differences in detected TRFs between the endophytic compartment and the rhizosphere, the results showed that the majority of diazotroph TRFs were capable of colonizing both endosphere and rhizosphere. There were more diazotroph TRFs found exclusively in the endosphere than in the rhizosphere except for the IL sites (Table 2.3). The correlations between the loadings of each diazotroph taxon on the endosphere/rhizosphere RDA axis across different states are high in general, ranging from 0.56 to 0.92 (Table 2.4), meaning that diazotrophs that tend to be found in the endophytic compartments in one location tend to be found in endophytic habitats elsewhere.

From a total of 798 diazotroph TRFs recovered, the most enriched TRFs (50 TRFs per site per habitat) in the endophytic compartment or rhizosphere at each site were chosen by redundancy analysis. As shown in Figure 2.4, half of these enriched endophytic diazotrophs were found in all four sites, while about 25% of the top rhizosphere diazotrophs were found in all sites. While the NJ site had more unique rhizosphere diazotrophs, it still shared the majority of endophytic diazotrophs with other sites. Overall, across sites, the top diazotrophs in the endosphere were more similar than those found in rhizosphere, which agrees with the correspondence analysis results. These results indicate that the endophytic compartment differs from the rhizosphere in terms of recruiting or retaining diazotrophs.

**Discussion**

The overall bacterial community and *Miscanthus*-associated diazotrophs responded differently to their surrounding environment, thus, to improve the interaction between *M. × giganteus* and its associative nitrogen fixers, soil edaphic factors need to be managed targeting this specific functional group. Geographic location is the main determinant of *nifH* diversity (Collavino *et al.*, 2014), especially for rhizosphere communities. Both total bacteria and
diazotroph assemblages were strongly correlated with soil edaphic factors (Table 2.2). However, the main contributors for these correlations differed between these two microbial assemblages. Mainly, we observed three interesting differences. First, total soil nitrogen was strongly linked to the assemblage of diazotrophs in both endophytic compartment and rhizosphere compartment, but failed to explain much variance for the overall bacterial community composition. Previous studies that investigated correlations between total soil N and diazotroph communities had complex results. The \textit{nifH} gene diversity showed no connection to N availability in some cropping systems (Collavino \textit{et al.}, 2014; Silva \textit{et al.}, 2011; Wakelin \textit{et al.}, 2010) and upland soils on a regional scale (Poly, Ranjard, \textit{et al.}, 2001), but did correspond to the micro-scale differences in N availability in upland soil (Poly, Ranjard, \textit{et al.}, 2001), and alpine soil (Zhang \textit{et al.}, 2006). Thus far, the connection between total soil N and diazotrophs associated with grasses have been rarely reported. The correlation observed here between total soil N and diazotroph community structure may be caused by the plant host regulating the colonization of diazotrophs according to the amount of available nitrogen. Previous work studying legume-diazotroph interactions indicated that with low nitrogen availability, plants recruit nitrogen fixers with both high and low nitrogen-fixing capacities. With an increase in available nitrate, an agricultural soil showed decrease in \textit{nifH} abundance (Collavino \textit{et al.}, 2014), and the plants preferentially associated with diazotrophs that have high nitrogen fixing performance (Kiers \textit{et al.}, 2006). Our second interesting finding is that soil iron content had a significant connection to the putative endophytic diazotroph community assemblages, but not to the putative endophytic total bacteria. This is likely due to the requirement of iron as a cofactor for nitrogenase (Burgess, 1990), and may be a limiting factor for the development of diazotrophic populations. Iron as a limiting factor for nitrogen fixation has been widely observed in marine diazotrophs (Mills \textit{et al.}, 2004;
Moore et al., 2009). Previous work (A. Kent, unpublished data) also suggested that iron content influences the abundance and composition of diazotrophs associated with *M. ×giganteus* at field locations throughout IL. Finally, soil pH only showed a significant correlation with the total bacterial communities residing in the rhizosphere. Soil pH has been suggested as the most important edaphic factor shaping the soil bacterial assemblage patterns (Lauber et al., 2009). Our results indicate that soil pH did not have the same impact on either endophytic communities or diazotroph communities. A possible explanation is that when compared to rhizosphere communities, the endophytic assemblages experience a more uniform environment within plant tissue, and thus, respond less to edaphic factors such as soil pH. However, this result is not consistent across plant species. For example, endophytic communities that reside in sweet sorghum, another biofuel candidate that is closely related to *M. ×giganteus*, were significantly influenced by soil pH (Ramond et al., 2013). Regarding the effect of pH on diazotroph communities, other studies have also found that pH may not significantly affect the diazotrophic community composition. Similarly, previous diazotroph communities studies conducted in agricultural sites (Hsu & Buckley, 2009; Silva et al., 2011) with pH ranging from acidic to slightly alkaline found no strong connection between *nifH* gene diversity and soil pH.

Although we observed a significant effect of total soil nitrogen across all sites, within each site, we did not see significant differences in diazotroph community composition (Table A.2) or abundance (data not shown) among N application rates. This observation differs from some previous results reported for diazotrophs associated with other grasses (Coelho et al., 2008; Prakamhang et al., 2009). In these studies, fertilization has a long-term effect on diazotrophs, and N is usually the overriding influential element. Coelho and associates (2008) found that sorghum cultivars receiving different levels of nitrogen fertilizer showed significantly different
rhizosphere diazotroph community patterns. Prakamhang et al. (2009) investigated how diazotroph communities associated with rice reacted to N fertility, and they observed a more diverse diazotroph community in unfertilized fields compared to fertilized fields.

In our study sites, however, it is worth noting that the total soil N data did not reflect differences in nitrogen treatments (Maughan et al., 2012), implying that the applied nitrogen may be absorbed quickly by the plant, or lost through leaching or volatilization. Meanwhile, as previously noted, total soil N still had a significant correlation with the community composition of diazotrophs living in both endophytic compartment and rhizosphere of M. ×giganteus. Therefore, it is possible that N application only boosted soil N content, and influenced the diazotroph community composition for a short time after fertilization. These results highlight the importance of soil monitoring and management, especially for field sites. In order to maintain the optimized plant growth conditions, in field monitoring appears to be essential for ensuring soil management practices such as fertilization and irrigation can be applied on time.

Besides the soil edaphic factors, the plant compartment (endophytic or rhizosphere) where the bacterial communities reside had a determinative effect on both total bacterial and diazotroph community composition. An interesting observation is that despite the rhizosphere communities showing significantly different site-specific assemblages, the endophytic diazotroph communities from different sampling sites were very similar (Figure 2.3). Moreover, half of the most enriched endophytic diazotroph TRFs from one site could be found in all other locations. The difference between endophytic and rhizosphere microbial compositions may due to their origins. Rhizosphere microbes most likely originate from bulk soil, and are influenced by plant exudates (Bais et al., 2006) and soil nutrient status (Kent & Triplett, 2002). Endophytes, however, can be either rhizome-borne or originate from rhizosphere soil (Rosenblueth &
Martinez-Romero, 2006). In Table 2.3, the percentage of diazotroph TRFs that only occurred in the endosphere in Urbana is much lower than those in other locations. This may due to some location-specific traits such as low soil N content in Urbana (Table 2.1), and low winter survival rate of 2008 (Maughan et al., 2012). Another explanation is that the cross-site similarity in endophyte assemblages resulted from rhizome propagation procedures. Bacteria can be transmitted from generation to generation through vegetative plant propagation, and the transmission of diazotrophs through vegetative settes has been previously observed in sugarcane (Dong et al., 1994; Reis et al., 1994). In this study, the rhizomes from all four field sites were harvested and propagated in Urbana, IL. The harvested rhizomes may harbor endophytes that originated from Urbana rhizosphere. The single origin for M. ×giganteus rhizomes may have allowed colonization of these diazotrophs in rhizomes before transplanting, and these diazotroph taxa could then benefit from priority effects, where pre-existing species may have a competitive advantage over later immigrants (Young et al., 2001). In our study, when rhizomes were transplanted in the field sites, rhizome-borne diazotrophs that benefit from priority effects could potentially outcompete diazotrophs originating from soil for the plant resources. The rhizomes sampled for this study were new rhizomes generated after 3 years of propagation at each field site rather than the original transplanted rhizomes, however, transmission of the endophytic microbiome to daughter rhizomes (Rosenberg & Zilber-Rosenberg, 2013) could allow the propagation of the original priority effect of beneficial microbes to progeny rhizomes. This has potential implications for engineering the rhizome microbiome through careful selection of nursery propagation conditions. In fact, a previous study has demonstrated that vegetative sugarcane settes receiving selected diazotroph inoculants were able to transmit these inoculants effectively into sugarcane seedlings (da Silva et al., 2012).
The plant and rhizosphere as ecological filters, can also explain the distinct endophytic and rhizosphere community composition we observed. For rhizosphere diazotrophs, access to the endophytic compartment is limited by the plant defense systems, cell walls and plant tissue structure (Rosenblueth & Martinez-Romero, 2006). Only taxa with a specified combination of characteristics may colonize within the plant tissue (Keddy, 1992). Once they have successfully colonized inside of the plant, endophytes in general are better protected against environmental stresses (Hallmann et al., 1997). Thus, they may evolve and adapt to the reduced stress of the microenvironment inside plant tissues, leading to more dissimilarity between the endophytic and rhizosphere microbial communities (Krause et al., 2006). In our study, opportunistic endophytes (taxa occurring in both the rhizosphere and endophytic compartment) were major components of the diazotroph community in both endosphere and rhizosphere. Despite the fact that diazotrophs had quite distinct community composition in the rhizosphere, the endophytic diazotroph communities were very similar across all locations. This suggests that the relative abundance of opportunistic endophytes changed after entering the plant tissue, which is likely the results of stresses imposed by plant during colonization (Compant et al., 2010), as well as the different microenvironment inside and outside the plant rhizome (Rosenblueth & Martinez-Romero, 2006).

Sustainability is a critical attribute for the development of biofuel feedstock candidates. Improving the N-fixing capacity of diazotroph communities associated with *M. × giganteus* is a promising, yet little explored strategy for achieving sustainable production of this feedstock. Our results provide basic knowledge towards better understanding of the associated interactions between *Miscanthus × giganteus* and diazotrophs. In this study, we showed various soil factors, and plant compartments explained a significant portion of variance for *M. × giganteus*-
associated diazotrophs, and these factors had different impacts on total bacteria and diazotrophs. The results of this study suggest that soil management and rhizome propagation processes targeted at enhancing specific diazotroph groups are promising strategies for enhancing this associative interaction and improving biofuel feedstock sustainability.

Acknowledgements

The authors thank SunGrant/U.S. Department of Energy Regional Biomass Feedstock Partnership for funding this study, Dr. M. Maughan and Dr. R. Darmody for kindly providing soil chemical data, as well as the University of Illinois Energy Farm, S. Bonos, M. Sousek, D. Williams and L. Williams for land and labor support. This work was also supported by the Cooperative State Research, Education, and Extension Service, US Department of Agriculture, under project number ILLU 875-374.

References


Tables

Table 2.1. Soil characteristics for the 0 – 30 cm soil in the sampling sites. Data are presented as mean (standard deviation) here.

<table>
<thead>
<tr>
<th>Site</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Soil pH</th>
<th>Soil organic matter (%)</th>
<th>Soil total organic carbon (%)</th>
<th>Total Soil nitrogen (%)</th>
<th>Soil Bulk Density (g/cm³)</th>
<th>Soil P (mg/kg)</th>
<th>Soil K (mg/kg)</th>
<th>Soil Fe (mg/kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL</td>
<td>40°06'20&quot; N</td>
<td>88°19'18&quot; W</td>
<td>5.85</td>
<td>1.9</td>
<td>1.11</td>
<td>0.11</td>
<td>1.63</td>
<td>41.9</td>
<td>127.85</td>
<td>167.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.15)</td>
<td>(0.16)</td>
<td>(0.09)</td>
<td>(0.06)</td>
<td></td>
<td>(16.69)</td>
<td>(15.41)</td>
</tr>
<tr>
<td>KY</td>
<td>38°07'45&quot; N</td>
<td>84°30'08&quot; W</td>
<td>5.53</td>
<td>3.46</td>
<td>1.26</td>
<td>2.74</td>
<td>1.51</td>
<td>318.63</td>
<td>149.68</td>
<td>357.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.21)</td>
<td>(0.21)</td>
<td>(0.78)</td>
<td>(3.71)</td>
<td></td>
<td>(29.28)</td>
<td>(35.11)</td>
</tr>
<tr>
<td>NE</td>
<td>41°10'07&quot; N</td>
<td>96°28'10&quot; W</td>
<td>6.5</td>
<td>3.98</td>
<td>0.26</td>
<td>8.55</td>
<td>1.34</td>
<td>81.22</td>
<td>648.95</td>
<td>174.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.10)</td>
<td>(0.23)</td>
<td>(0.03)</td>
<td>(0.52)</td>
<td></td>
<td>(25.11)</td>
<td>(19.79)</td>
</tr>
<tr>
<td>NJ</td>
<td>40°13'31&quot; N</td>
<td>74°14'54&quot; W</td>
<td>5.43</td>
<td>2.07</td>
<td>0.11</td>
<td>10.92</td>
<td>1.49</td>
<td>214</td>
<td>101.94</td>
<td>328.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.17)</td>
<td>(0.39)</td>
<td>(0.03)</td>
<td>(1.00)</td>
<td></td>
<td>(34.88)</td>
<td>(13.59)</td>
</tr>
</tbody>
</table>

§ IL (Urbana, Illinois); KY (Lexington, Kentucky); NE (Mead, Nebraska); NJ (Adelphia, New Jersey)
Table 2.2. Percentage of microbial community variance explained by edaphic factors determined by canonical correspondence analysis.

<table>
<thead>
<tr>
<th></th>
<th>Total bacteria</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Endophytic compartment</td>
<td>Rhizosphere</td>
<td>Endophytic compartment</td>
<td>Rhizosphere</td>
<td></td>
</tr>
<tr>
<td>Percentage of variance explained by edaphic factors</td>
<td>15.76</td>
<td>20.15</td>
<td>12.5</td>
<td>24.83</td>
<td></td>
</tr>
<tr>
<td>Percentage of variance explained solely by Total nitrogen</td>
<td>1.41</td>
<td>0.95</td>
<td>1.42*</td>
<td>6.70**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.82**</td>
<td>3.8**</td>
<td>1.42*</td>
<td>5.58**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.94</td>
<td>4.99**</td>
<td>1.42</td>
<td>1.67*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.11*</td>
<td>1.90**</td>
<td>0.71</td>
<td>2.79**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>1.66**</td>
<td>0.71</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.17</td>
<td>1.19*</td>
<td>2.13**</td>
<td>3.91**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.17**</td>
<td>1.19*</td>
<td>3.55**</td>
<td>1.67*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.41*</td>
<td>4.51**</td>
<td>1.42</td>
<td>1.12*</td>
<td></td>
</tr>
</tbody>
</table>

Significance codes: ** p < 0.01, * 0.01 ≤ p < 0.05
Table 2.3. The percentage of diazotrophic TRFs found in habitats (only in endosphere, only in rhizosphere or both) within each field site.

<table>
<thead>
<tr>
<th>Habitat</th>
<th>IL</th>
<th>KY</th>
<th>NE</th>
<th>NJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRFs occurring only in endosphere</td>
<td>15.55%</td>
<td>30.80%</td>
<td>20.90%</td>
<td>35.26%</td>
</tr>
<tr>
<td>TRFs occurring only in rhizosphere</td>
<td>19.35%</td>
<td>7.26%</td>
<td>7.28%</td>
<td>5.93%</td>
</tr>
<tr>
<td>TRFs occurring in both compartments</td>
<td>65.10%</td>
<td>61.95%</td>
<td>71.83%</td>
<td>58.81%</td>
</tr>
</tbody>
</table>

Table 2.4. Correlation between the species score of each TRF among sampling sites. High correlation indicates species enriched in one compartment (endophytic compartment or rhizosphere compartment) on one site are likely to be found enriched in the same compartment on another site.

<table>
<thead>
<tr>
<th></th>
<th>IL</th>
<th>KY</th>
<th>NE</th>
<th>NJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KY</td>
<td>0.92</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NE</td>
<td>0.78</td>
<td>0.86</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>NJ</td>
<td>0.67</td>
<td>0.58</td>
<td>0.64</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Figure 2.1. Correspondence analysis ordination of total bacterial community residing in endophytic compartment (a) and rhizosphere compartment (b) at each location. Each point represents the total bacterial community from one sample, with the shading indicate the site sampled. Arrows and their lengths indicate direction and strength of the environmental variables: Soil pH, organic matter (SOM), total organic carbon (TOC), total nitrogen (N), bulk density, phosphorus (P), potassium (K), and iron (Fe). Percentage of community variance explained by each axis is indicated in parentheses.
Correspondence analysis ordination of diazotroph communities residing in endophytic compartment (a) and rhizosphere compartment (b) at each location. Each point represents the diazotroph assemblage from one sample, with the shading indicating the site sampled. Arrows and their lengths indicate direction and strength of the environmental variables: Soil pH, organic matter (SOM), total organic carbon (TOC), total nitrogen (N), bulk density, phosphorus (P), potassium (K), and iron (Fe). Percentage of community variance explained by each axis is indicated in parentheses.
Figure 2.3. Correspondence analysis ordination of total bacterial community composition (a) and diazotroph community composition (b). Each point represents the microbial community from one sample, with the shading indicate the compartment sampled. Percentage of community variance explained by each axis is indicated in parentheses.
Figure 2.4. Venn diagram showing the percentage of shared enriched diazotroph TRFs in the endophytic compartment (a) and rhizosphere compartment (b).
CHAPTER 3: PLANT MICROBIOME ASSOCIATED WITH NATIVE MISCANTHUS PLANTS

Abstract

Microbes in rhizosphere soil and plant tissue impact nutrient cycling and plant growth. Understanding the drivers of plant-associated microbial communities and their assemblage process will enhance the understanding towards positive associative interactions between plant and microbes. The objectives of this study were to characterize bacterial communities associated with the biofuel feedstock candidate Miscanthus in its native habitats, as well as identify the ecological drivers of these bacterial communities. Microbial communities residing in the rhizosphere and endophytic compartment of wild Miscanthus in Taiwan were assessed using both automated ribosomal intergenic spacer analysis and Illumina MiSeq sequencing. Plant compartment is the primary factor that shapes the associated bacterial communities. The majority of endophytic bacterial taxa appeared to originate from the rhizosphere soil. However, distinct bacterial assemblages were observed in the endophytic compartments and rhizosphere soil of Miscanthus, suggesting that plants regulate their endophytic community composition. A number of bacteria taxa specifically enriched in either the rhizosphere or the plant endophytic compartment have been observed and identified in this study. Local environmental factors and plant phylogenetic distances showed strong correlations with microbial community composition in both compartments. Additionally, I found that Miscanthus plants under stressful environmental condition have different endophytic recruiting strategies. To conclude, endophytic communities associated with Miscanthus are distinct yet derived from the rhizosphere community. Bacterial community compositions from both compartments rely on plant species and the local environment.
Introduction

Plant roots or rhizomes and the rhizosphere soil surrounding them are habitats with intense plant-microbe interactions. Rhizodeposition releases about 11% of net fixed plant carbon and about 10% of non-legume plant nitrogen into the rhizosphere soil (Jones et al., 2009), creating nutrient-rich niches that attract rhizosphere communities with high richness and broad diversity (Berendsen et al., 2012). Subsequently, a small fraction of the rhizosphere bacterial communities is filtered by the plant root structure and defense systems to colonize the plant endophytic compartment (EC) through host-dependent selection (Bulgarelli et al., 2013). Plant hosts often rely on the benefits provided by bacterial species inhabiting the plant endophytic compartment and rhizosphere. These benefits include suppressing pathogens, enhancing environmental stress tolerance, and improving nutrient acquisition (Compant et al., 2005; Malinowski & Belesky, 2000; Rosenblueth & Martinez-Romero, 2006).

Both biotic and abiotic factors influence the community structure of plant-associated bacteria (Berg & Smalla, 2009; Lundberg et al., 2012; Rosenblueth & Martinez-Romero, 2006). These factors can be translated into functional changes of the microbial community that then affect plant fitness and productivity (Ehrenfeld et al., 2005). Plant genotype-specific microbiomes have been found to associate with model plant Arabidopsis thaliana, as well as cultivated crops such as maize (Peiffer et al., 2013) and potato (van Overbeek & van Elsas, 2008). Plants actively control the diversity of microbial populations and recruit beneficial bacteria by releasing genotype-specific root exudates (Berendsen et al., 2012) and regulating plant-microbe signaling pathways (Bulgarelli et al., 2013). Soil type and edaphic factors are also important drivers of bacterial communities (Berg & Smalla, 2009; Garbeva et al., 2008). They can alter the rhizosphere community by influencing plant root exudates (Jones et al., 2009) and
by limiting the local soil microbiome pool (Bulgarelli et al., 2013). Thus far, repeated observations indicate that plant-associated microbiomes rely on both plant genotype and soil factors. However, the relative importance of these two aspects has not been fully explored.

Although evidence indicates that the structure of plant-associated microbiome depends on the local environment, recently, the existence of a core plant-associated microbiome has been introduced by recent studies (Lundberg et al., 2012; Podolich et al., 2014). The core plant-associated microbiome is a group of microbes that consistently associate with a particular plant species and their structures are relatively stable over different growth stages (Lundberg et al., 2012; Podolich et al., 2014). Investigation of the community structure and function of these core plant microbiome, and comparing the core microbiome between plant species will enhance the understanding of plant needs at the species level.

*Miscanthus* species are promising bioenergy feedstock candidates as they are capable of sustainable and economically favorable biomass production (Heaton et al., 2004). An under-explored characteristic of perennial C₄ grasses is their potential to harbor associative microbes capable of promoting plant growth fitness. Basic questions about the ecology of plant-associated microbes need to be answered if these associations are to be managed for sustaining and economically viable production of biofuel feedstocks. The first step in such an investigation is to examine the array of biotic and abiotic factors with the potential to influence the community composition of endophytic and rhizosphere microbes associated with *Miscanthus*. Evaluation of key ecological drivers that shape *Miscanthus*-associated microbial populations will identify environmental factors that are critical to monitor, in order to optimize beneficial plant-microbe interactions.
Native populations of Miscanthus in Asia offer an opportunity to evaluate microbial communities associated with long-term sustainable plots of this potential bioenergy crop. Miscanthus plots with plant genetic differences in Taiwan can be found in a variety of soil fertility, chemistry, moisture and other environmental conditions.

From the last Chapter, I have learned that the plant compartment is the key factor that drives microbial community differences. My central hypothesis is that the compartment differences is due to the fact that only specific bacterial taxa are able to access and successfully colonize inside Miscanthus plant tissue. In this Chapter, with the help of high-throughput sequencing, I explored which bacterial taxa were specifically enriched in Miscanthus endophytic compartment (EC) or rhizosphere. The genetic variation among native plants and local edaphic variation allows me to determine the relative importance of plant genetic distance and edaphic factors for shaping the assemblages of endophytic and rhizosphere microbial populations associated with Miscanthus, and identify the key suite of ecological drivers that influence composition of Miscanthus-associated bacterial communities.

Materials and Methods

Sample collection

Sample collection sites were located throughout Taiwan (Figure 3.1 and Table 3.1). Sites were selected to represent gradients in soil pH, fertility, soil organic matter, elevation, soil moisture and plant species. Detailed information on each sampling location is included in Table 3.1. At each sampling site, four quadrats (1 m²) were randomly established, and three Miscanthus plants were harvested within each quadrat. Rhizomes were removed with a shovel, shaken to remove loosely adhered soil. Rhizomes were then separated from aboveground tissues and placed into a plastic bag. Samples were transported on ice and stored at 4°C prior to processing.
Within each quadrat, six soil cores were collected (0-12 cm depth using a 1.9-cm diameter soil core) and used for soil chemical and physical analyses. Soil samples were stored on ice for transport back to the lab.

**Sample processing**

**Soil samples**

Gravimetric soil moisture was determined for each soil sample. Air-dried soil samples were submitted to the Iowa State University Soil Test Lab (Ames, IA) for chemical analyses. Total N, total C, C/N and percent organic matter were determined, along with P, K, Ca, Fe, NH$_4$+, and NO$_3$-

**Plant rhizome and rhizosphere soil**

Rhizosphere soil was washed off the rhizomes using 40 ml sterile deionized water and collected in a sterile container for characterization of rhizosphere bacterial populations. This soil was frozen at -80°C and lyophilized prior to DNA extraction. Rhizomes were surface-sterilized following the methods of Chelius and Triplett (Chelius & Triplett, 2001) with modifications. Each rhizome was placed in a 1L container containing 100 ml 95% ethanol and shaken for 30 seconds. The ethanol was then replaced by 100 ml 5.25% sodium hypochlorite and shaken for 15 minutes at 150 rpm. Sterilized rhizomes were rinsed three times with 300 ml sterile distilled water to remove all traces of sodium hypochlorite. Using ethanol-sterilized pruners, rhizomes were chopped into small pieces (3 - 5 cm in length), and placed into a sterilized Waring blender cup with 50 ml of PBS + 0.1% Tween 80. Rhizomes were ground in the blender and placed into a sterile centrifuge tube containing five sterile glass beads. Pulverized rhizomes were washed gently to release endophytic bacteria following the methods of Brulc et al. with modification
(Brulc et al., 2009). Grounded rhizomes were shaken gently (100 rpm) on ice for 1 hour, and plant material was removed by filtration through a sterile 3-inch No. 25 US Standard Test Sieve. Endophytic bacteria contained in the filtrate were concentrated by centrifugation prior to DNA extraction.

**DNA extraction and purification**

Total genomic DNA was extracted from freeze-dried rhizosphere soil samples collected from all rhizomes using the FastDNA SPIN Kit for Soil (MP Biomedicals, Solon OH). Genomic DNA was further purified using cetyl trimethyl ammonium bromide (CTAB) extraction to remove contaminating humic acids (Sambrook & Russell, 2001). Genomic DNA was extracted from concentrated endophytic bacterial samples using the FastDNA Spin Kit (MP Biomedicals, Solon OH) following the manufacturer’s protocol. DNA concentration for all samples was adjusted to a standard concentration of 10 ng/µl prior to DNA analyses.

**Bacterial analyses using community fingerprinting approach**

Bacterial community composition and diversity were assessed using automated ribosomal intergenic spacer analysis (ARISA) (Fisher & Triplett, 1999). The polymerase chain reaction (PCR) conditions were conducted as described previously (Li et al., 2015). Denaturing capillary electrophoresis was performed by the Keck Center for Functional Genomics at University of Illinois (Urbana, IL) for each PCR using an ABI 3730xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) with ROX1000 size standards (Applied Biosystems, Foster City, CA, USA). Capillary electrophoresis results in minor run-to-run variations in observed against actual fragment length. This problem is solved using the allele-calling features in GeneMarker 2.4 (SoftGenetics, State College, PA) before analysis. To include the maximum number of peaks while excluding background fluorescence, a threshold of 100 fluorescence units was used. The
signal strength (i.e. peak area) of each peak was normalized to account for run-to-run variations in signal detection.

**Bacterial community analyses using Illumina sequencing approach**

Extracted genomic DNA of the endophytic and rhizosphere samples were sent to DOE Joint Genome Institute (JGI, Walnut Creek, CA), where the V4 region of bacterial 16S rRNA genes were amplified using primers 515F (5’-GTGCCAGCMGCCGCGGTAA-3’) and 806R (5’-GGACTACVSGGGGTATCTAAT-3’) with bar (Caporaso et al., 2011). The amplicons were then purified and sequenced on Illumina MiSeq instrument with a 2 × 250 bp reads configuration.

**Plant genetic variation analysis**

The genetic similarity of the collected plants was assessed using simple sequence repeat (SSR) markers. The plant tissues were screened with 16 SSRs, including 14 nuclear markers ESSR024, ESSR026, GSSR04, GSSR041, GSSR046, GSSR08, GSSR11, GSSR12, GSSR13, GSSR15, GSSR19, GSSR25, GSSR35, and GSSR48 (Swaminathan et al., 2012) and two chloroplast markers sac-17 and sac-26 (de Cesare et al., 2010). PCR conditions were the same as described by Głowacka et al. (2015). Denaturing capillary electrophoresis separated the PCR fragments. It was performed using ABI 3730xl Genetic Analyzer (Applied Biosystems) by the Keck Center for Functional Genomics at University of Illinois (Urbana, IL). The STRand software v. 2.4.59 (http://www.vgl.ucdavis.edu/informatics/strand.php) was used to conduct marker scoring.
Data analysis

Plant genetic data

To compare plant genetic variation among samples, pairwise distance matrix between all samples were calculated with R package polysat v1.3-3 (Clark & Jasieniuk, 2011) using Lynch distance (Lynch, 1990).

Community fingerprinting data

To compare microbial community composition among samples, the Bray-Curtis dissimilarity was calculated among all samples pairs based on the Hellinger-transformed ARISA data (Legendre & Legendre, 1998).

Permutational Multivariate Analysis of Variance (PerMANOVA) (Anderson, 2001) was used to estimate the marginal variances explained by the potential ecological drivers such as plant compartments (endophytic compartment and rhizosphere). The significance of potential explanatory variables was estimated using a permutational test. Correspondence analysis (CA) was performed to visualize the patterns of bacterial communities from different plant compartments.

The contributions of host plant genetic distance, geographic distance and local environmental factors were estimated by variance partition analysis. Variance partitioning was conducted based on redundancy analysis (RDA) and used adjusted $R^2$ to estimate the variance partition that explained by explanatory variables. Variation of the Hellinger-transformed bacterial community matrix was partitioned by three explanatory matrices – environmental, geographic distance and plant genetic matrices. The environmental matrix had all soil edaphic measures and the elevation of the sampling site (Table 3.1), and these numbers were normalized.
using z-score transformation before fitting into the model. The geographic distance matrix consisted of the latitude and longitude coordinates for each sampling site. To avoid collinearity, the plant genetic matrix was generated based on the Lynch distance matrix for all plant pairs instead of using the plant marker scoring results directly. The Lynch distance matrix for all plant pairs was converted to coordinates of each sample on the first k principal-component (PC) axes. These coordinates were used to fit in the variation partition model as the plant genetic matrix. Here, k is equal to the number of PCs that had positive eigenvalues.

Canonical correspondence analysis (CCA) was performed on ARISA data to identify the main ecological drivers (Table 3.1). The type III effect (marginal effect) of each term was calculated with all the other terms in the model. Separate significance test were performed for each marginal term.

Functions from Package vegan version 2.2-1 (Oksanen et al., 2015) for R Language (R_Development_Core_Team) were used to conduct the Hellinger transformation, Bray-Curtis dissimilarity calculation, PerMANOVA, CA, variation partitioning, CCA and test of marginal effects.

Community sequencing data

Sequences reported in this paper have been deposited in MG-RAST (Project 13491) (Meyer et al., 2008).

Each sequence read was assigned to its original sample according to the oligo index. The forward and reverse read of each paired-end sequence were merged as one using software FLASH (Fast Length Adjustment of SHort reads) (Magoč & Salzberg, 2011). After sorting and merging, both forward and reverse primers were trimmed and any sequence that contained
ambiguous base “N” was removed from downstream processing. All above steps were conducted using customized scripts. Sequences that have more than 10% bases with the quality score lower than 30 were also removed from downstream processing.

The filtered sequences were then clustered into operational taxonomic units (OTUs) using USEARCH6 (Edgar, 2010). USEARCH performed the following steps: (1) de-replicate sequences and remove singletons to reduce the data size and calculating time, (2) Remove any chimeras contained in the sequences using gold database (version microbiomeutil-r20110519) as a reference (http://drive5.com/uchime/uchime_download.html), (3) form clusters of 97% identity sequences and represent each OTU by consensus sequences (representative sequences).

The cluster files were converted into an OTU table using a customized script derived from QIIME (Caporaso et al., 2010). These consensus sequences were used as representative sequences in each OTU, and the taxonomic attribution of filtered sequences were conducted using the RDP classifier (Wang et al., 2007) trained on the 4 February 2011 Greengenes database with 50% confidence interval. The taxonomic information of representative sequences assigned as Cyanobacteria or “unidentified bacteria” were reconfirmed with nucleotide blast (http://blast.ncbi.nlm.nih.gov/Blast.cgi) in GenBank. Any potential mitochondrial or chloroplast OTUs were removed from the OTU table. After removing all the plant sequences, any samples with less than 200 reads were dropped from the analysis.

The OTUs that differentially represented between compartments (compartment-enriched OTUs) were identified by comparing their abundances between compartments using ANOVA. Script group_significance.py in QIIME was applied to find these OTUs. OTUs were considered to have significantly different relative abundance between compartments if the FDR corrected P value was less than 0.05. The abundances of compartment-enriched OTUs per sampling site
were visualized using a heat map. Only OTUs that reached 0.5% averaged relative abundance in the enriched compartment were included in this heat map. Sample sites on the heat map were arranged according to their unweighted UniFrac distance, and OTUs were arranged according to their phylogenetic relationships generated using make_phylogeny.py function (the default method) in QIIME. The heat map was generated using heatmap.2 function in Package gplots for R.

To compare bacterial community composition among samples, phylogenetic beta diversity metrics were calculated using UniFrac, which takes the phylogenetic relationship among OTUs into consideration. Script beta_diversity.py in QIIME were used to calculate the matrix. Similar to the community fingerprinting data analysis, PerMANOVA was carried out to evaluate how much beta-diversity could be explained by plant compartment. Principal coordinate analysis (PCoA) ordinations based on unweighted UniFrac distances were used to visualize the community composition of bacteria from different compartments.

The correlation between bacterial community distances obtained by ARISA and Illumina sequencing for shared samples was compared using Mantel’s test (function mantel of the vegan package in R). Only samples that present in both ARISA and OTU table were selected for this comparison. After conducting the Hellinger transformation for both ARISA data and the OTU table, Bray-Curtis dissimilarity was calculated for both matrices separately. Mantel’s test was then performed on these two distance matrices with 999 permutations.
Results

Comparing the community results generated by ARISA and Illumina sequencing

In this study, ARISA identified a total of 564 OTUs for bacteria associated with Miscanthus from 200 endophytic samples and 200 rhizosphere samples. Illumina sequencing generated 9,270,935 16S rRNA gene reads after removing low-quality reads. However, a significant amount of reads from the endophytic compartment were from the plant chloroplast and mitochondria. After removing all reads assigned to plant DNA, there are 1,259,262 reads from 67 endophytic samples and 175 rhizosphere samples. These reads were clustered into 2570 OTUs based on 97% similarity.

The Mantel test indicated that bacterial community dissimilarity matrices obtained using ARISA and Illumina sequencing had a moderate correlation with each other (Mantel statistic r: 0.5216, p < 0.001).

Plant compartment determines bacterial communities

Despite the differences in pH, soil fertility, geographic location, and plant species, the strongest clustering in similarity among all microbial community samples was observed between rhizosphere and endophyte communities. Both ARISA and sequencing results showed that distinct microbial assemblages were observed in the endosphere and rhizosphere of native Miscanthus (Figure 3.2). The compartment differences explained 5.37% and 12.87% of bacterial community variance based on the ARISA and Illumina sequencing data respectively.

The taxonomic groups associated with Miscanthus varied between compartments and species (Figure 3.3). Both endophytic and rhizosphere communities were predominated by Proteobacteria. In general more Acidobacteria were found in the rhizosphere soil than in the
endophytic compartment. Slightly more *Bacteroidetes* and *Firmicutes* were found to associate with *M. floridulus* than *M. sinensis*. At the class level of *Proteobacteria*, relatively more *Gamma-proteobacteria* were found in the endophytic compartment than in the rhizosphere soil, especially for the endophytic compartment of *M. floridulus*. Meanwhile, more *Alpha-proteobacteria* were found in the rhizosphere than in the endophytic compartment. More than half *Proteobacteria* found in *M. sinensis’* rhizosphere soil were *Alpha-proteobacteria*. Little *Delta-proteobacteria* was found in the detected bacterial communities.

At the family level, the composition of *Alpha-proteobacteria* detected from endophytic samples was distinct between plant species. For *Alpha-proteobacteria* associated with the rhizosphere, however, the detected families and their relative distribution was quite similar between plant species. Within *Beta-proteobacteria* species, more *Burkholderiaceae* and fewer *Comamonadaceae* and *Oxalobacteraceae* were associated with the *M. sinensis* samples than the *M. floridulus* samples. *Gamma-proteobacteria* dominated the endophytic compartment of *M. floridulus*, and most of these *Gamma-proteobacteria* were found to be species from *Enterobacteriaceae*. *Enterobacteriaceae* species were also found to be abundant in other species and plant compartment. Comparing between *Miscanthus* species, relatively more *Sinobacteraceae* and *Xanthomonadaceae* were found in *M. sinensis* samples than in *M. floridulus* samples.

*Endophytic compartment enriched OTUs and rhizosphere enriched OTUs*

To further explore the OTUs that lead to distinct endophytic and rhizosphere community composition, OTUs that were significantly enriched in either endophytic compartment (EC-enriched) or rhizosphere compartment (R-enriched) were identified (Figure 3.4, Table B.1), and results were compared between plant genotype. In total, 53 EC-enriched OTUs and 488 R-
enriched OTUs were identified, and these compartment-enriched OTUs comprised 40% - 70% relative abundance in the compartment where they were enriched in. Six EC-enriched OTUs were enriched in the endophytic compartment of both species (Figure 3.4a), and all these six OTUs were all from Enterobacteriaceae family (Table B.1). The relative abundance of these six OTUs represented 67.18% of the endophytic communities in M. floridulus, and 13.14% endophytic communities in M. sinensis. Species-specific EC-enriched OTUs had relative abundances of 6.27% and 44.66% for the endophytic communities of M. floridulus and M. sinensis correspondingly. 129 R-enriched OTUs were found to be R-enriched in both plant species (Figure 3.4b), and they came from diverse taxonomic groups (Table B.1). These 129 general R-enriched OTUs built 24.29% and 22% of bacterial communities in the rhizosphere soil of M. floridulus and M. sinensis. Species-specific R-enriched OTUs composed 21.18% and 27.29% of bacterial communities in the rhizosphere soil of M. floridulus and M. sinensis respectively. These findings indicate that specific bacterial taxa are favored in the plant endophytic compartment and rhizosphere soil.

The relative abundance of compartment-enriched OTUs captured the patterns of community distance among sites. Figure 3.4c visualized the UniFrac distance among communities detected from each sampling site. As shown on the left side of Figure 3.4c, the samples can be classified into three clusters. Cluster 1 contains rhizosphere samples from both species. In Cluster 1, the main compartment-enriched OTUs are general R-enriched OTUs found in both species, and R-enriched OTUs found in M. floridulus. Cluster 2 contained rhizosphere soil from M. sinensis and two EC samples from M. sinensis. These two EC samples are both from sites where Miscanthus plants grew near hot springs. In Cluster 2, the main compartment-enriched OTUs are R-enriched OTUs found in both species, R-enriched OTUs found in M.
sinensis, and EC-enriched OTUs that were specific to hot spring EC samples. The EC-enriched OTUs for the two hot spring site were found in both EC and rhizosphere of these hot spring samples, but were rarely detected in EC from other locations. The last cluster contains only endophytic samples, and not surprisingly, the main compartment-enriched OTUs are EC-enriched OTUs.

Local environment and plant genetic variation overwhelm the effect of geographic distance

Variance partition analysis was employed to evaluate the relative importance of local environmental factors, plant genetic variation and geographic distance to the bacterial communities (Figure 3.5). The results indicated that about 25% variation of bacterial community compositions could be explained by these variables. In both compartments, the percentage of conditional variation that could be explained by local environmental factors and plant genetic variation are similar (around 10%). In the endophytic compartment, local environmental factors explained a little more bacterial community variation than plant genetics. Opposite results were observed for the rhizosphere samples, where plant genetic variation explained slightly more variation than the environmental factors. Surprisingly, geographic distances only independently explained 2% of bacterial community variation for both compartments, which was only one fifth to one-third of the variation explained by surrounding environments or plant genotyping data. Variation explained by interactions between any two of the three explanatory matrices was small, so was the interaction among all three matrices.

Local environmental variables influence microbial communities

Table 3.2 shows the marginal community variation explained by each environmental variable. In both endophytic and rhizosphere communities, elevation, soil pH and C/N ratio were the top three variables that explained most bacterial community variation. Besides these three,
soil P, Ca, Fe, Total N, NH$_4^+$, and NO$_3^-$ also had a significant correlation with endophytic bacterial communities. All the measured environmental variables had a significant correlation with rhizosphere bacterial communities. Rhizosphere communities had significant correlations with more environmental variables than the endophytic communities, likely due to the fact that endophytes reside inside plants and, therefore, are less sensitive to the surrounding soil conditions.

Low pH along with poor soil nutrient condition may lead to the observation (Figure 3.4) that endophytic communities from the two hot spring samples were more similar to their corresponding rhizosphere communities rather than the rest of the endophytic samples. These hot spring samples site had the lowest soil pH among all sites (Table 3.1, Figure 3.6a). Miscanthus rhizosphere community richness is negatively correlated with pH under low pH condition (Figure 3.6b). However, the endophytic community richness was stably maintained under a range of pH conditions (Figure 3.6c).

**Discussion**

*The determinative role of compartment*

In this study, the bacterial communities from endophytic compartments significantly differed from those of rhizosphere samples; this is reflected in both alpha-diversity and beta-diversity. An order-of-magnitude fewer OTUs were found in the endophytic compartment than in the rhizosphere (Figure 3.6b, Figure 3.6c). Distinct beta-diversity patterns were detected between communities from these two compartments using both sequencing and community fingerprinting methods (Figure 3.2). Similar results have been found in bacterial communities associated with endophytic and rhizosphere compartment of other plant species using both conventional cultural methods and molecular approaches (Bulgarelli et al., 2013; Lundberg et al.,
2012; Mahaffee et al., 1992; Roesch et al., 2008). The differences between endophytic microbial communities and rhizosphere ones are likely due to their origins and the distinct biotic and abiotic conditions in rhizosphere and endosphere. Rhizosphere bacteria are most likely to originate from bulk soil (Bulgarelli et al., 2013) and are influenced by the plant exudates (Bais et al., 2006) and soil nutrient status (Kent & Triplett, 2002). On the other hand, some endophytes are vertically transmitted from the parental plants while others originate from rhizosphere soil and phylloplane (Rosenblueth & Martinez-Romero, 2006). The dynamics of endophytic microbes largely depend on plant nutrient supply. Thus, endophytes can be indirectly influenced by factors that affect plant growth (Hallmann et al., 1997). Additionally, compared to rhizosphere microbes, endophytic microorganisms tend to have a relatively lower richness and diversity (Hallmann et al., 1997; Rosenblueth & Martinez-Romero, 2006). These results together with this study indicated that soil edaphic factors and plant species shape the endophytic and rhizosphere communities in different ways. Investigating these differences will enhance the understanding of the plant needs and how they recruit the endophytic communities.

**Proteobacteria predominated the community**

Comparing with findings from other plant species, the taxonomic diversity of bacteria that associated with Miscanthus is narrower. OTUs from phylum other than Proteobacteria only took a very small part of the entire community. Proteobacteria predominated both Miscanthus endophytic compartment and rhizosphere. Depending on Miscanthus species, Proteobacteria made up to 80 - 95% of endophytic community and 56% - 78% rhizosphere communities that associated with Miscanthus. Studies investigating Arabidopsis (Bulgarelli et al., 2012; Lundberg et al., 2012), rice (Edwards et al., 2015), maize (Peiffer et al., 2013) and Populus (Gottel et al., 2011) using high-throughput sequencing approach also revealed Proteobacteria as important
members associated with plants, and *Proteobacteria* comprise a greater portion of microbial assemblages in the endophytic compartments than the rhizosphere, except for those associated with *Populus* (reported by Gottel et al. (2011)). *Proteobacteria* formed 20 - 30% of the *Arabidopsis* and rice rhizosphere community and 45% of maize rhizosphere communities. Bacteria from this phylum comprise 25% or 50-60% of the *Arabidopsis* endophytic community, and 50% of the rice endophytic community (Edwards et al., 2015; Lundberg et al., 2012). Greater representation of bacteria from *Chloroflexi, Acidobacteria, Actinobacteria*, and *Bacteriodetes* were reported by these studies. The endophytic communities in those studies refer to root microbiome rather than rhizome microbiome, however, it is still clear that *Miscanthus* species tend to harbor a higher percentage of *Proteobacteria* than the other plant species mentioned. Thus, here I conclude that the taxonomic spectrum of plant-associated bacteria is species-specific.

**The effect of plant species**

The endophytic and rhizosphere bacteria population can be variable depending on the *Miscanthus* species. I discovered that some compartment-enriched OTUs were species-specific, whereas, others were generally enriched in the same compartment for both species. Previous studies found some differences in physiological aspects and ecological habitats of the two species included in this study, which may lead to distinct interactions between the microbes and their plant host. *M. sinensis* tends to be found in soil with pH of 4 to 6, can occupy habitat with a wild range of elevation and latitude, and have a short inflorescence axis. Meanwhile, *M. floridulus* is more likely to be found in sea level tropical areas and have a long inflorescence axis (Clifton-Brown et al., 2008; Stewart et al., 2009). Besides plant physiology and natural habitat aspects, plant phylogenetic history could also impact associated bacterial communities. Johnston-
Monje and Raizada (2011) found the compositions of seed-associated communities correlated with maize phylogeny (Johnston-Monje & Raizada, 2011).

Podolich (2014) reviewed genotype-specific microbiome and suggested the existence of a core endophytic community at genotype level, and the core endophytic communities were stable across different plant growth stages and extrinsic factors (Bulgarelli et al., 2012; Lundberg et al., 2012). In this study, these compartment-enriched OTUs comprised half of the plant-associated bacterial communities, and the presence of these OTUs was reasonably robust across different sampling sites. These OTUs could represent the core microbiome associated with Miscanthus species. The 6 EC-enriched OTUs found in both plant species were all from family Enterobacteriaceae. Previous work conducted in Dr. Kent’s lab cultured nitrogen-fixing bacteria from native Miscanthus endophytic compartment and the majority of the nitrogen-fixing isolates were from the Enterobacteriaceae family. Thus, lack of nitrogen could be a general challenge for these native Miscanthus plants.

**Bacterial communities in a stressful condition**

Plant-associated bacteria are not independent of each other or their surrounding environment. In this study, I found that rhizosphere communities respond to stressful external conditions. Miscanthus plants were collected near two hot springs, where soil pH fell below the optimum range for M. sinensis growth. The low pH, along with low soil organic matter content in these sites reflected stressful plant growth environment. Rhizosphere communities showed reduced significant richness in these sites. This is likely due to the low pH soil condition representing a strict environmental filter, only allowing those adapted to this acidic condition to thrive. However, endophytic community richness did not show reduced richness under the same condition. Podolich et al. (2014) consider plants and their associated endophytes as a super
organism. They proposed that under stressful conditions, plants would revive some beneficial species that are in a latent state under normal conditions, resulting in an endophyte-mediated plant resistance to environmental stress. Their theory provides one putative mechanism explaining the observation from this study.

Another explanation proposed here is that plants fail to select or relax the selection criteria under stressful conditions. Even though endophytic communities and rhizosphere assemblages showed consistently distinct beta-diversity across many plant species (Bulgarelli et al., 2013), the rhizosphere bacteria community is still the most important source pool for endophytes. The endophytic communities and rhizosphere ones are distinct and yet dependent. So now the question is since the rhizosphere bacterial pool is much smaller under low pH conditions, would it still possible for Miscanthus to recruit beneficial endophytic bacteria from their surrounding rhizosphere soil? It is possible that the number of microbial niches in the plant endophytic compartment does not change much in a stressful environment, and rhizosphere bacteria occupied these niches. As a result, the endophytic richness did not decrease much in stressful conditions. Future work investigating how plants and their endophytes respond to external stress will clarify potential mechanisms for this observation and provide guidance to manipulate recruitment of the most beneficial endophytic communities.

Conclusion

Results from this study show that the Miscanthus endophytic compartment and rhizosphere harbor different bacterial taxa. The Miscanthus genotype and soil attributes also contributed to the differences in microbial community structures. Improving the understanding of the influence of these factors and plant compartment-enriched microbiomes will allow to better
select plant genotypes and soil conditions that enhance mutualistic plant-microbe interactions, which will ultimately lead to sustainable biofuel crop production.

Acknowledgements

This chapter was completed in collaboration with Dr. A.D Kent, Dr. L.V. Clark, Dr. E.J. Sacks, and Dr. J.L. Dangl. DOE Joint Genome Institute conducted the Illumina sequencing as part of a Community Sequencing Project awarded to J.L. Dangl. We would like to thank P.-CL. Hou, I.-C. Tseng, and C.-S. Kuoh from National Cheng Kung University, Taiwan for field work support and guidance. Thanks are also extended to N. Gottel, M. Salgado, and Y. Chang for field and lab assistance. I thank S. Yourstone and S.H. Paredes developed sequencing analysis pipeline and customized scripts. I thank Dr. C.P. Soman and Dr. A.C. Yannarell for contributing constructive feedbacks on earlier versions of this manuscript.

References


Tables
Table 3.1A. Site name, code, locations, plant species of the *Miscanthus* collecting sites in 2008.

<table>
<thead>
<tr>
<th>Site Code</th>
<th>Site Name</th>
<th>Site Location</th>
<th>Plant Species</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Elevation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Siaoyoukeng Hot Springs</td>
<td>North Taiwan</td>
<td><em>M. sinensis</em></td>
<td>25.17558</td>
<td>121.54726</td>
<td>813</td>
</tr>
<tr>
<td>B</td>
<td>The Gate</td>
<td>North Taiwan</td>
<td><em>M. sinensis</em></td>
<td>25.16615</td>
<td>121.57548</td>
<td>753</td>
</tr>
<tr>
<td>C</td>
<td>Cingtiangang</td>
<td>North Taiwan</td>
<td><em>M. sinensis</em></td>
<td>25.16531</td>
<td>121.57629</td>
<td>749</td>
</tr>
<tr>
<td>D</td>
<td>Menghuanhu Ecological Conservation Zone</td>
<td>North Taiwan</td>
<td><em>M. sinensis</em></td>
<td>25.16566</td>
<td>121.55976</td>
<td>888</td>
</tr>
<tr>
<td>E</td>
<td>Mt. Cising (base)</td>
<td>North Taiwan</td>
<td><em>M. sinensis</em></td>
<td>25.17559</td>
<td>121.5582</td>
<td>761</td>
</tr>
<tr>
<td>F</td>
<td>Mazao Bridge Hot Springs</td>
<td>North Taiwan</td>
<td><em>M. sinensis</em></td>
<td>25.17747</td>
<td>121.56193</td>
<td>614</td>
</tr>
<tr>
<td>G</td>
<td>Mazao Bridge</td>
<td>North Taiwan</td>
<td><em>M. sinensis</em></td>
<td>25.17793</td>
<td>121.56252</td>
<td>600</td>
</tr>
<tr>
<td>H</td>
<td>Datunshan peak</td>
<td>North Taiwan</td>
<td><em>M. sinensis</em></td>
<td>25.17943</td>
<td>121.52219</td>
<td>1124</td>
</tr>
<tr>
<td>I</td>
<td>Datunshan mid-elevation</td>
<td>North Taiwan</td>
<td><em>M. sinensis</em></td>
<td>25.18251</td>
<td>121.52402</td>
<td>952</td>
</tr>
<tr>
<td>J</td>
<td>Datunshan Nature Center (base)</td>
<td>North Taiwan</td>
<td><em>M. sinensis</em></td>
<td>25.18743</td>
<td>121.52163</td>
<td>800</td>
</tr>
<tr>
<td>K</td>
<td>Houlong/Miaoli (West Coast)</td>
<td>Central Taiwan</td>
<td><em>M. floridulus</em></td>
<td>24.60799</td>
<td>120.75047</td>
<td>21</td>
</tr>
<tr>
<td>L</td>
<td>Yushan NP</td>
<td>Central Taiwan</td>
<td><em>M. sinensis</em></td>
<td>23.48448</td>
<td>120.8964</td>
<td>2520</td>
</tr>
<tr>
<td>M</td>
<td>Lulin Mountain (peak)</td>
<td>Central Taiwan</td>
<td><em>M. floridulus</em></td>
<td>23.46774</td>
<td>120.88077</td>
<td>2881</td>
</tr>
<tr>
<td>N</td>
<td>Alishan Area</td>
<td>Central Taiwan</td>
<td><em>M. floridulus</em></td>
<td>23.49433</td>
<td>120.7938</td>
<td>1965</td>
</tr>
<tr>
<td>O</td>
<td>Between Alishan and Chiayi</td>
<td>Central Taiwan</td>
<td><em>M. floridulus</em></td>
<td>23.41466</td>
<td>120.64457</td>
<td>975</td>
</tr>
<tr>
<td>P</td>
<td>Near Jhongpu</td>
<td>Central Taiwan</td>
<td><em>M. floridulus</em></td>
<td>23.4473</td>
<td>120.58086</td>
<td>260</td>
</tr>
</tbody>
</table>
Table 3.1B. The edaphic values of the sampling sites.

<table>
<thead>
<tr>
<th>Site Code</th>
<th>P (ppm)</th>
<th>K (ppm)</th>
<th>Ca (ppm)</th>
<th>Fe (ppm)</th>
<th>Total C (%)</th>
<th>Total N (%)</th>
<th>CN ratio</th>
<th>pH</th>
<th>NH4-N (ppm)</th>
<th>NO3-N (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>19 (5.27)</td>
<td>49.2 (44.7, 171.5)</td>
<td>84.1 (26.1, 109.3)</td>
<td>397.9 (183.9, 562.5)</td>
<td>0.6 (0.2, 1.7)</td>
<td>0.1 (0.1, 0.1)</td>
<td>7 (4.3, 12.3)</td>
<td>3 (2.9, 3.2)</td>
<td>1.0 (0.7, 1.6)</td>
<td>0.3 (0.2, 0.3)</td>
</tr>
<tr>
<td>B</td>
<td>2.4 (1.4)</td>
<td>76.5 (57.2, 98.2)</td>
<td>62.7 (29.6, 89.2)</td>
<td>212.7 (106.8, 339.4)</td>
<td>3.7 (2.3, 4.9)</td>
<td>0.3 (0.2, 0.4)</td>
<td>13.3 (12.8, 13.5)</td>
<td>4.5 (4.3, 4.6)</td>
<td>1.6 (1.1, 1.9)</td>
<td>0.6 (0.4, 0.8)</td>
</tr>
<tr>
<td>C</td>
<td>3.5 (1.75)</td>
<td>91.8 (65.2, 106.6)</td>
<td>47.4 (29.8, 76.4)</td>
<td>379.7 (230.9, 533.7)</td>
<td>5.6 (2.3, 7.9)</td>
<td>0.4 (0.2, 0.5)</td>
<td>13.5 (12.2, 15.4)</td>
<td>4.1 (3.8, 4.2)</td>
<td>1.7 (1.2, 2)</td>
<td>0.8 (0.3, 1.5)</td>
</tr>
<tr>
<td>D</td>
<td>7.8 (4.5, 13)</td>
<td>163.6 (126.1, 197.8)</td>
<td>104.6 (91.4, 117.9)</td>
<td>380 (238.6, 499.8)</td>
<td>13 (10.6, 15.3)</td>
<td>0.8 (0.7, 1)</td>
<td>15.4 (15.2, 15.7)</td>
<td>3.7 (3.5, 4)</td>
<td>3.0 (2.5, 3.6)</td>
<td>2.6 (2.3, 3)</td>
</tr>
<tr>
<td>E</td>
<td>56.3 (7.5, 110.5)</td>
<td>166.8 (43.6, 299.2)</td>
<td>75.1 (57.8, 108.6)</td>
<td>596.7 (345.1, 927.5)</td>
<td>17.8 (3.1, 39.9)</td>
<td>0.8 (0.2, 1.7)</td>
<td>20.1 (16.2, 23.9)</td>
<td>3.6 (3.3, 4.1)</td>
<td>2.7 (1.5, 3.6)</td>
<td>0.6 (0.4, 0.8)</td>
</tr>
<tr>
<td>F</td>
<td>38.6 (6.5, 85.5)</td>
<td>40.3 (7.8, 88.6)</td>
<td>763.4 (189.7, 1512)</td>
<td>555.4 (254.7, 893.5)</td>
<td>1.4 (0.6, 3.1)</td>
<td>0.1 (0.1, 0.2)</td>
<td>10.8 (8.2, 16.4)</td>
<td>3.1 (2.9, 3.6)</td>
<td>1.5 (0.8, 1.8)</td>
<td>0.4 (0.3, 0.5)</td>
</tr>
<tr>
<td>G</td>
<td>47.7 (6.5, 110.5)</td>
<td>112.5 (7.8, 2992)</td>
<td>422.7 (57.8, 1512)</td>
<td>644.7 (254.7, 1476.5)</td>
<td>9.8 (0.6, 39.9)</td>
<td>0.5 (0.1, 1.7)</td>
<td>15.3 (8.2, 23.9)</td>
<td>3.3 (2.9, 4.1)</td>
<td>2.7 (1.6, 4.5)</td>
<td>0.3 (0.3, 0.5)</td>
</tr>
<tr>
<td>H</td>
<td>36.6 (19.5, 65.5)</td>
<td>286.5 (245.8, 323.8)</td>
<td>200.5 (155.2, 268.7)</td>
<td>137.4 (117.3, 181.5)</td>
<td>19.8 (16.8, 25.4)</td>
<td>1.3 (1, 1.7)</td>
<td>14.6 (14.3, 15.3)</td>
<td>4.1 (4.1, 4.2)</td>
<td>6.3 (4.7, 7.9)</td>
<td>3.8 (2.8, 5.8)</td>
</tr>
<tr>
<td>I</td>
<td>41.5 (4.5, 138.3)</td>
<td>131.6 (117.6, 164.3)</td>
<td>1298.8 (205.8, 3696)</td>
<td>96.4 (80.9, 121.4)</td>
<td>63 (4.8, 82)</td>
<td>0.5 (0.4, 0.7)</td>
<td>14.1 (12.6, 17.2)</td>
<td>5.3 (4.3, 7.5)</td>
<td>3.0 (1.7, 4.9)</td>
<td>2.5 (0.8, 4.8)</td>
</tr>
<tr>
<td>J</td>
<td>4.9 (3.6, 5)</td>
<td>171.1 (156, 185)</td>
<td>1059.6 (433.6, 2120.5)</td>
<td>77.4 (61.3, 97.2)</td>
<td>3.5 (2.9, 4.3)</td>
<td>0.3 (0.3, 0.4)</td>
<td>11.5 (9.7, 12.1)</td>
<td>5.6 (5.2, 6.2)</td>
<td>2.8 (2.4, 3)</td>
<td>1.5 (0.9, 2.2)</td>
</tr>
<tr>
<td>K</td>
<td>8 (7.9)</td>
<td>914.9 (57.9, 103.9)</td>
<td>91.4 (800, 1001)</td>
<td>116.6 (100.3, 130.5)</td>
<td>0.6 (0.4, 1)</td>
<td>0.1 (0.1, 0.1)</td>
<td>5.6 (4.4, 7.5)</td>
<td>6.1 (5.7, 6.4)</td>
<td>1.2 (0.7, 2.3)</td>
<td>0.6 (0.4, 1.3)</td>
</tr>
<tr>
<td>L</td>
<td>5.4 (3.9)</td>
<td>56.7 (25, 134.4)</td>
<td>1657 (1308.5, 2178.5)</td>
<td>183.6 (1308.5, 2569)</td>
<td>2.7 (1.6, 9)</td>
<td>0.2 (0.1, 0.6)</td>
<td>9.5 (7.9, 12.5)</td>
<td>6.9 (5.1, 7.9)</td>
<td>1.1 (0.5, 2.4)</td>
<td>0.9 (0.3, 2.2)</td>
</tr>
<tr>
<td>M</td>
<td>71.8 (19.5, 220)</td>
<td>549.1 (2186.1, 1465.5)</td>
<td>868 (543.5, 1793)</td>
<td>272.9 (155, 334.6)</td>
<td>20.5 (14.3, 34)</td>
<td>1.2 (1.1, 1.4)</td>
<td>16.4 (13.6, 243)</td>
<td>4.1 (4.1, 4.1)</td>
<td>2.5 (2.4, 2.5)</td>
<td>0.4 (0.3, 0.4)</td>
</tr>
<tr>
<td>N</td>
<td>17.4 (10.3, 26)</td>
<td>133.6 (115.1, 153.7)</td>
<td>905.3 (451.2, 1671.5)</td>
<td>135.4 (92.6, 171.6)</td>
<td>2.4 (1.6, 3.6)</td>
<td>0.2 (0.2, 0.3)</td>
<td>12.6 (10.2, 142)</td>
<td>5.2 (4.6, 6.1)</td>
<td>2.6 (1.3, 4.9)</td>
<td>0.3 (0.1, 0.3)</td>
</tr>
<tr>
<td>O</td>
<td>7.8 (6.5, 9)</td>
<td>102.5 (91.9, 119.5)</td>
<td>972.1 (651.5, 1190)</td>
<td>190.4 (165.1, 210.3)</td>
<td>1.6 (1.1, 1.9)</td>
<td>0.2 (0.1, 0.2)</td>
<td>8.4 (8.9)</td>
<td>5.3 (4.9, 5.6)</td>
<td>1.9 (1.5, 2.3)</td>
<td>0.4 (0.3, 0.5)</td>
</tr>
<tr>
<td>P</td>
<td>76.4 (45, 108.5)</td>
<td>137.5 (52, 282)</td>
<td>1051.9 (960.5, 1265.5)</td>
<td>235.4 (213.8, 260.6)</td>
<td>1.2 (0.7, 1.9)</td>
<td>0.1 (0.1, 0.2)</td>
<td>8.5 (7.2, 10.7)</td>
<td>5.8 (5.7, 5.9)</td>
<td>1.9 (1.4, 2.8)</td>
<td>1.1 (0.8, 1.3)</td>
</tr>
</tbody>
</table>
Table 3.2. Percentage of marginal variation explained by each soil edaphic factors.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sequencing EC&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Sequencing R&lt;sup&gt;a&lt;/sup&gt;</th>
<th>ARISA EC</th>
<th>ARISA R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevation</td>
<td>3.93%*</td>
<td>0.94%*</td>
<td>2.27%***</td>
<td>1.93%***</td>
</tr>
<tr>
<td>P</td>
<td>1.10%</td>
<td>4.13%***</td>
<td>1.35%**</td>
<td>1.36%***</td>
</tr>
<tr>
<td>K</td>
<td>0.50%</td>
<td>0.73%</td>
<td>0.71%</td>
<td>1.27%***</td>
</tr>
<tr>
<td>Ca</td>
<td>11.50%***</td>
<td>0.83%</td>
<td>1.27%**</td>
<td>1.59%***</td>
</tr>
<tr>
<td>Fe</td>
<td>5.14%**</td>
<td>0.89%*</td>
<td>1.09%*</td>
<td>1.48%***</td>
</tr>
<tr>
<td>Total C</td>
<td>3.74%*</td>
<td>1.65%**</td>
<td>0.67%</td>
<td>1.00%**</td>
</tr>
<tr>
<td>Total N</td>
<td>6.97%**</td>
<td>0.82%</td>
<td>0.87%*</td>
<td>1.57%**</td>
</tr>
<tr>
<td>C/N Ratio</td>
<td>2.91%*</td>
<td>1.10%*</td>
<td>1.98%***</td>
<td>2.31%***</td>
</tr>
<tr>
<td>OM</td>
<td>6.29%***</td>
<td>1.39%*</td>
<td>0.75%</td>
<td>1.09%**</td>
</tr>
<tr>
<td>pH</td>
<td>26.50%***</td>
<td>3.15%***</td>
<td>1.88%***</td>
<td>4.74%***</td>
</tr>
<tr>
<td>NH4-N</td>
<td>2.39%*</td>
<td>0.15%</td>
<td>1.22%*</td>
<td>1.52%***</td>
</tr>
<tr>
<td>NO3-N</td>
<td>11.79%***</td>
<td>1.64%**</td>
<td>1.09%*</td>
<td>1.16%***</td>
</tr>
</tbody>
</table>

<sup>a</sup> EC represents endosphere sampels; R respresents rhizosphere samples.

<sup>b</sup> *0.01<p<0.05, **p<0.01
Figure 3.1. Map of sample collection sites for native *Miscanthus* in Taiwan. Each point represents a sampling site. Detailed information about each location is contained in Table 1.
Figure 3.2. Distinct bacterial assemblages found in Miscanthus endophytic compartment and rhizosphere. Correspondence analysis ordination of total bacterial community composition associated with native Miscanthus in Taiwan using ARISA relative florescence (a) and principal coordinate analysis ordination of total bacterial community composition accessed with Illumina sequencing (b). Each point represents the bacterial assemblage from one sample, with the color and shape indicate the plant compartment sampled. EC is short for endophytic compartment, and R represents rhizosphere.
Figure 3.3. The main taxonomic groups associated with native Miscanthus species and compartments. (a) Histogram showing the relative distribution of detected phyla. (b) Histogram showing the relative distribution of detected Proteobacteria classes. (c) Histograms displaying the relative distribution of families detected in Alphaproteobacteria (α), Betaproteobacteria (β) and Gammaproteobacteria (γ).
Figure 3.4. Compartment specific OTUs found in both endophytic compartment and rhizosphere soils. (a) Number of OTUs that are EC-enriched only in *M. floridulus*, in both species, and only in *M. sinensis*. (b) Number of OTUs that are R-enriched only in *M. floridulus*, in both species, and only in *M. sinensis*. (c) A heat map showing the number of reads for compartment-enriched OTUs across sampling sites. Each row of the heat map represents a sampling site, and each column of the heat map represent compartment-enriched OTU. The four samples in the yellow square are samples collect from hot spring sites. Color and shape at the left and right side indicate the plant compartment and species found in this site. Color and shape at the upper and lower side indicate the plant compartment and species this particular OTU was enriched in. EC is short for endophytic compartment, and R represents rhizosphere.
Figure 3.5. Venn Diagram showing how the bacterial community variation based on ARISA data can be partitioned among local environmental factors, geographic distance and plant genetic distance. (a) The variation partitioning results for endophytic communities. (b) The variation partitioning results for rhizosphere communities.
Figure 3.6. Community richness changed along the pH gradient. (a) pH variation among sampling sites. Detailed information about each site code is contained in Table 1. (b) The rhizosphere community richness along pH gradient. (c) The endophytic community richness along pH gradient. Here richness was represented by number of OTUs for every 200 sequence reads.
CHAPTER 4: COMPARISON OF BACTERIAL COMMUNITIES ASSOCIATED WITH
NATIVE AND NATURALIZED MISCANTHUS SINENSIS

Abstract

Land plants intensively interact with their associated microbes, and together they act as super organisms. Recent studies of the plant microbiome using high-throughput sequencing approaches have suggested the existence of core plant microbiome. However, these studies were typically conducted either in a greenhouse setting or within a relatively small geographic region. Here, I investigated and compared the core microbiome associated with Miscanthus sinensis - a C\textsubscript{4} perennial grass, in the plant’s native (Eastern Asia) and naturalized habitats (Eastern United States). Using Illumina sequencing targeting the bacterial 16S rRNA gene, I observed a large number of bacterial taxa shared between the native and the naturalized habitats. 20\% - 30\% of plant-associated bacteria were specifically enriched in the plant endophytic compartment or rhizosphere soil in both habitats. These microbes were identified as the core microbiome associated with M. sinensis. I also found that although similar bacterial members were found in both habitats, the patterns were distinct between habitats. The results of this study should assist future studies to identify the function of plant core microbiome and ultimately shed light on plant microbiome assembly rules.
Introduction

Various microbes colonize land plant rhizomes and the surrounding rhizosphere soil. Many of them form intimate mutualistic interactions with their plant hosts by suppressing pathogens, improving nutrient acquisition, and releasing plant hormone (Compant et al., 2005; Malinowski & Belesky, 2000; Rosenblueth & Martinez-Romero, 2006). Such plant-microbe interactions have considerable influence on both associated microbial communities (Bulgarelli et al., 2013) and the plant hosts (Friesen et al., 2011). Plants actively recruit associated microbes from the surrounding soil. To encourage the colonization of plant growth promoting bacteria, plant hosts differentiate rhizosphere microbial members from the surrounding pool of soil microorganisms through selective root exudate production (Berendsen et al., 2012). Plant genotype specific characteristics, such as plant-microbe signaling pathway, further filter subsets of rhizosphere microbial communities to colonized the endophytic compartment (EC) of the hosts (Bulgarelli et al., 2013).

High-throughput sequencing techniques bring new tools as well as new insight into studies of the complex microbial communities associated with plants. Recent studies of the plant-microbiome topic have suggested the existence of core endophytic bacteria communities (Edwards et al., 2015; Lundberg et al., 2012; Podolich et al., 2014). The core endophytic bacterial community is a group of microbes that commonly found to be enriched in plant endophytic compartments, and it is reasonably stable in different soil types (Lundberg et al., 2012). This concept has not yet been extended to the rhizosphere communities. However, giving the important plant growth promoting effects provided by rhizosphere communities (Berendsen et al., 2012; Singh et al., 2004; van Loon et al., 1998; Yang et al., 2009), I consider evaluating the composition of both core endophytic and core rhizosphere communities is essential. Identify
core taxa that consistently associate with specific plant species will potentially reveal which plant needs, such as nutrient demands or hormone regulation, are fulfilled by their microbial associates. Up to the present, studies that investigate core endophytic bacterial communities were conducted either in a greenhouse setting or within a relatively small geographic region. The existence of a core plant microbiome has not yet been reported on a global scale.

Sequencing data can also be used to explore the direct and indirect connection between co-existing microbial taxa (Faust & Raes, 2012). Co-occurrence patterns and network inferences have been used to study microbial communities in multiple environments, including human-microbiomes (Arumugam et al., 2011; Qin et al., 2010), marine microbes (Fuhrman et al., 2015), soil microbes (Barberán et al., 2014) and plant microbiomes associated with rice (Edwards et al., 2015). These techniques help to elucidate microbial assembly rules (Faust & Raes, 2012). Strong interspecies connections between co-existing microbial taxa usually suggest a niche-based communities (Cody & Diamond, 1975). Otherwise, weak interspecies connections are more likely to support the neutral theory (Rosindell et al., 2011). With sufficient number of time-series samples and extremely deep sequencing data, microbial taxa co-existent patterns can facilitate predicting marine microbial community composition at a global scale (Larsen et al., 2012). Although modeling plant microbiome remains challenging, comparing the network influences of land plant microbial communities from different locations will be a vital first step to see if it is possible to predict land microbial communities in a similar way.

*Miscanthus sinensis* is a C₄ perennial grass that originated from eastern Asia (Lewandowski et al., 2000). *M. sinensis* cultivars were initially introduced to Unite States as ornamental grasses during the late 19th century (Dougherty et al., 2014). Later, *M. sinensis* plants were documented to establish natural populations (naturalized) in the eastern US (Dougherty et
al., 2014). Recently, *M. sinensis* has also been considered as an important bioenergy feedstock candidate (Shumny et al., 2010). Cultivars of *M. sinensis* have shown potential for high quality cellulose production (Shumny et al., 2010) and the capacity for adapting to a wide range of climatic and soil conditions (Dougherty et al., 2014; Lewandowski et al., 2000; Stewart et al., 2009; Shumny et al., 2010). Microbial communities associated with *M. sinensis* could facilitate their colonization in various conditions. It has been suggested by multiple studies that further plant biotechnology should take optimizing plant-microbe interactions as a crucial practice to improve plant yield and fitness (Peiffer et al., 2013; Podolich et al., 2014). So far, the microbiome associated with bioenergy candidates, their correlations with plant fitness, and the possibility to model these microbial communities have not been fully explored.

In this study, I investigated the microbiome residing in the rhizome and rhizosphere soil of *M. sinensis* in both the native and naturalized ranges for this plant species. The hypotheses of this study include 1) similar to other non-legume plants mentioned above, the core microbiome can be identified in the *M. sinensis* endophytic compartment for plants that are living in relatively close sites. 2) On the global scale, however, this observation may not hold true due to bacterial dispersal limitation and the distinct surrounding environment. 3) The bacteria taxa co-occurrence pattern in the native sites and naturalized sites are different due to their different plant core microbiomes. To address these hypotheses, *Miscanthus sinensis* rhizome and rhizosphere samples were collected from its native site in Taiwan as well as its naturalized site in Eastern United States. Additionally, based on the plant size information collected for naturalized *M. sinensis*. I identified the potential bacteria taxa that most likely contribute to the fitness (represented by size) of naturalized *M. sinensis*. 

85
Materials and methods

Sampling sites

Native Miscanthus sinensis rhizomes and rhizosphere soil from 140 plants were collected from Taiwan as described in Chapter 3. Native M. sinensis sites represent a wide range of soil edaphic factors, elevation.

Naturalized Miscanthus sinensis samples were collected from multiple sites across nine states in eastern United States (Table C.1) (Dougherty et al., 2014). These sites represented various soil chemical conditions and covered nearly the entire latitude of naturalized M. sinensis habitat in eastern US (Dougherty et al., 2014). The habitats of these naturalized M. sinensis fell into several categories - forest edge, forest understory, open field and roadside. Previous work conducted by Dougherty and co-workers (2014) found that M. sinensis favored sites with more disturbances, such as roadside and forest edge. To characterize plant morphology and fitness, tiller height and diameter, the total number of tillers, and basal diameter were recorded for 20 plants for each site. If less than 20 plants were found in one site, the plant morphology data for all plants were collected (Dougherty et al., 2014).

Characterizing plant-microbiome

Endophytic and Rhizosphere communities from total 148 naturalized M. sinensis distributed in 15 sites were characterized. Half of plants from each sampling sites (except for NJ-02) were randomly selected to have their rhizome endophytic and rhizosphere bacterial community characterized. Only two M. sinensis plants were found in site NJ-02, and the bacterial communities associated with both plants were characterized. Plant rhizomes and rhizosphere processing, as well as DNA extraction, were conducted following the protocol
described in Li et al. (2015). The extracted DNA was then diluted to 20 ng/ul and sent to Joint Genome Institute (Walnut Creek, CA, USA), where the V4 region of 16S rRNA genes was sequenced using primers 515F and 806R (Caporaso et al., 2011) (with adaptors and barcodes for Illumina sequencing) to address the microbiome in each sample. Details were described in Chapter 3. Sequences reported in this paper have been deposited in MG-RAST (Native Miscanthus: Project 13491; naturalized Miscanthus: Project 13560) (Meyer et al., 2008).

**Statistical analysis**

A sample by bacterial species table (OTU table) was generated from the sequencing results of both native and naturalized samples using the protocol described in Chapter 3. Using this protocol, poor quality reads, chimeric reads, and plant chloroplast and mitochondrial reads were all removed prior to downstream analyses. Reads were then clustered into operational taxonomic units (OTUs) at 97% similarity. Taxonomic information was then assigned to each OTU using Greengenes 12_10 as a reference database. Species accumulation curves were generated to check if enough samples were collected for each compartment by habitat combination.

**Comparing OTUs from native and naturalized habitat**

A Venn diagram was generated to present numbers of endophytic OTUs detected from native, naturalized *M. sinensis* and shared between these two habitats. To remove the effect due to different sample sizes in these two habitats, a random subset of samples from the one with more samples (in this case, naturalized sites) were rarefied. The sample size of the randomly chosen subset equaled the sample size of the habitat with fewer samples, and the number of detected endophytic OTUs and the shared endophytic OTUs were calculated based on this random subset. The subsampling process was permutated 100 times, and the mean number of
detected OTUs and shared OTUs were used on the Venn diagram. The same analyses were applied to the rhizosphere samples. Function sample() and draw.pairwise.venn() from R packages base and VennDiagram were used for these analyses.

**Compartment-enriched OTUs**

The OTUs that specifically enriched in the endophytic compartment (EC-enriched) or in the rhizosphere (R-enriched) were identified for the native and naturalized samples separately. To prevent any rare OTUs identified as compartment-enriched OTUs by chance, a frequency filter was applied to the OTU table first, so that only OTUs present in at least 25% of all samples were included in the calculation. The relative abundances of each OTU in endophytic and rhizosphere samples were then compared using ANOVA. An OTU was considered a compartment-enriched OTU if it had significantly (False discovery rate adjusted p-value ≤ 0.05) higher relative abundance in one compartment than the other. All the calculations were carried out using script group_significance.py in QIIME V1.7 (Caporaso, Kuczynski, et al., 2010).

**Co-occurrence and network**

Non-random OTU co-occurrence patterns were tested for native EC and rhizosphere soil, and naturalized EC and rhizosphere soil respectively. Only OTUs present in at least 25% of the selected habitat × compartment samples were included in the tests. Checkerboard score (C-score) (Stone & Roberts, 1990) was calculated based on OTUs present/absent in samples from each habitat × compartment combination. A community that has a higher C-score is less likely to be randomly assembled (Stone & Roberts, 1990). The significance of the C-score was determined by comparing the C-score from actual OTU-table with C-scores obtained from a series of simulated null models. If the actual C-score is higher than 95% of the simulated ones, then I
consider the observed OTU occurrence pattern was non-random. Function oecosimu() and C.score() from R packages vegan and bipartite were used for these analyses.

Network inferences were generated to illustrate the OTU occurrence patterns. Spearman’s rank correlations between OTU pairs were calculated. Pairs that had the absolute value of \( \rho \) larger than 0.6, and statistically significant (P-value < 0.01) were considered as valid correlated OTU pairs. Basic network characters (average node connectivity, geodesic distance, clustering coefficient, modularity) were calculated to describe the network inferences (Faust & Raes, 2012). Nodes that did not link to any other nodes were removed from the visualized network topology. Whether the distribution of node degree fits power law distribution was tested. All the network analyses were conducted in the R statistical environment. Main functions include rcorr() and graph.adjacency(), power.law.fit() from Hmisc and igraph.

**Potential plant growth-promoting OTUs**

A plant fitness index was generated to represent the fitness of naturalized *M. sinensis* plants collected in this study. The partial principal component analysis was conducted on a normalized plant-by-plant morphology (tiller height, tiller diameter, the total number of tillers, and basal diameter) matrix. Dummy variables representing sampling site were used as the conditional matrix, so that potential site-to-site genetic variations would not influence the results. *M. sinensis* plants from site MA-01 were trimmed one month before the sampling, resulting in plant heights in that site that were not representative of their true height. Therefore, plants from site MA-01 were excluded in this normalized matrix. All four plant morphological traits increased along the first principle component (PC). That is, the higher the score plant sample had on the first PC, the larger the plant was. As a result, each plant sample’s score on the first principle component was extracted and used as the plant fitness index. The plant fitness index
was then used as a response variable, and regressed against the Hellinger transformed bacterial OTU table using partial least squares regression (PLSR). PLSR is a modeling technique that finds a set of orthogonal multi-dimensional directions (latent variables) in the X space (in this case, Hellinger transformed OTU table) to maximize the explained variance in Y space (in this case, plant fitness index). PLSR is especially useful when a number of variables in X space are correlated (Wold et al., 2001). The first latent variable in the X space represents the best linear combination of X variables (in this case, Hellinger transformed OTU abundances) in explaining the Y space. The top five percent of OTUs that had the highest loading on the first latent variable were considered to have the highest correlation with plant fitness. Therefore, these OTUs were considered as potential growth promoting bacteria in this system. The potential growth-promoting OTUs were identified separately for both the endophytic compartment and rhizosphere of naturalized Miscanthus. Hellinger transformation was conducted using function decostand() in package vegan for R, and PLSR was conducted using function plsr() in package pls for R.

**Results**

After removing all the low-quality reads, singleton reads, and plant DNA, I recovered 830,336 reads from native M. sinensis sites (51 endophytic samples and 119 rhizosphere samples), and 691,338 from the naturalized sites (111 endophytic samples and 105 rhizosphere samples). A total of 806 and 2884 OTUs were detected in the native M. sinensis endophytic compartment and the rhizosphere soil respectively. A total of 1261 and 2055 M. sinensis-associated OTUs were detected from the naturalized sites. Species accumulation curves for the endophytic samples did not reach a plateau, indicating that larger endophytic sample sizes or
more reads per sample would increase the number of OTUs detected both native and naturalized sites (Figure C.1a)

**M. sinensis from native and naturalized sites shared associated bacteria**

A set of bacterial OTUs was found to be associated with native and naturalized *M. sinensis* plants, especially in the rhizosphere soil, even though their habitats were on different continents. In this study, I found that almost 90% of the OTUs detected in naturalized *M. sinensis* rhizosphere were also detected in the native *M. sinensis* rhizosphere as well (Figure 4.1a). Compared to the rhizosphere community, *M. sinensis* from different habitats shared fewer endophytic OTUs. About 40% - 50% OTUs were detected in both habitats (Figure 4.1b).

***EC-enriched OTUs and R-enriched OTUs came from different families***

Here I found 45 EC-enriched OTUs and 326 R-enriched OTUs associated with native *M. sinensis*. Whereas, there were 54 EC-enriched OTUs and 133 R-enriched OTUs associated with naturalized *M. sinensis*. Although the number of compartment-enriched OTUs was only a small fraction of the total number of detected OTUs, they were important members of the bacterial communities - these compartment-enriched OTUs comprise half of the *M. sinensis*-associated bacterial communities, based on the relative abundance of sequencing reads (Table C.2). There were 16 EC-enriched OTUs and 55 R-enriched OTUs shared between the native and naturalized *M. sinensis* habitats. These OTUs comprised about 20% of the bacterial communities associated with natural *M. sinensis* and about 30% for the naturalized *M. sinensis* (Figure 4.2). These shared EC-enriched and R-enriched OTUs came from distinct taxonomic groups (Figure 4.2). Acetobacteraceae was the only family found in both shared EC-enriched and R-enriched OTUs. Most shared EC-enriched OTUs were assigned to Burkholderiaceae and Enterobacteriaceae.
while the majority of the shared R-enriched OTUs belonged to Bradyrhizobiaceae and Hyphomicrobiaceae.

**Co-occurrence pattern and network**

I observed a non-random (p < 0.001) co-occurrence pattern for the endophytic and rhizosphere bacterial communities associated with native *M. sinensis*, but not for the communities associated with naturalized ones (Table 4.1). I further visualized the correlation network of co-occurring bacterial OTUs (Figure C.2, Table 4.1). The network inferences of native *M. sinensis* microbiome had larger network sizes, higher network densities (nodes were better connected), larger clustering coefficients than the ones of naturalized *M. sinensis* microbiome. These network characteristics indicated that bacterial members in the native *M. sinensis* microbiome formed close connections with each other, while close interactions were less common in the naturalized *M. sinensis* microbiome. Native *M. sinensis* co-occurrence networks have smaller modularity than the naturalized ones. This suggests that the network inferences of naturalized *M. sinensis* microbiome are likely to form sub-clustering within the network. All nodes degree distributions were found to follow power law distribution, indicating that the OTUs within each network inferences showed a scale-free network property (Faust & Raes, 2012; Zhou et al., 2010). It means the OTUs that within the network are not randomly connected. A typical characteristic of scale-free network is that a few nodes (such nodes are known as hubs) are joined with numerous others, while most nodes are only connected to a few others (Barabási, 2009; Zhou et al., 2010).

**Potential plant growth promoting OTUs**

Naturalized *Miscanthus sinensis* morphology data allowed us to identify the OTUs that had the highest correlations with plant size. These OTUs are likely to carry traits or produce
growth-promoting hormones to enhance the fitness of *M. sinensis*. 1256 OTUs were detected from naturalized endophytic compartment, and 2049 OTUs from naturalized rhizosphere compartment. I considered the top 5% OTUs (63 OTUs of the endophytic compartment and 103 OTUs of the rhizosphere) that had the correlation with plant fitness index as potential *M. sinensis* growth promoting species. The relative abundance of these 5% of OTUs reached 40% in the endophytic compartment and the rhizosphere soil. Members of Burkholderiaceae and Enterobacteriaceae families were found to be abundant among these potential plant growth-promoting OTUs (Figure 4.3).

**Discussion**

Baas Becking once said of microbial ecology “everything is everywhere, but the environment selects” (Baas-Becking, 1934). Up to this date, I still cannot claim that every bacterium is everywhere. However, I did observe highly similar bacterial taxa occupying the *M. sinensis*-associated niches in distant geographic regions. In this study, a large portion of the bacterial OTUs was shared among native and naturalized habitats, especially for these residing in the rhizosphere soil. Key endophytic and rhizosphere taxa also were found shared among habitats. These findings support the microbial seed bank theory to some extent. This theory considers the microbial community turnover to be an outcome of the relative abundance fluctuation of persistent microbial taxa (Caporaso, Bittinger, et al., 2010; Lennon & Jones, 2011). It has been shown that marine microbial assemblages follow the seed bank theory (Caporaso et al., 2012; Gibbons et al., 2013). With great sequence depth, sequences from one location can be well overlapped with a global database, and a sequence depth of $1.93 \times 10^{11}$ is required to achieve 100% genetic overlap (Gibbons et al., 2013). To examine if the same theory applied to a terrestrial ecosystem, even deeper sequence depth would be required for the same coverage level.
due to the complexity of the terrestrial ecosystem. Future research with more endophytic samples (Figure C.1) and greater sequence depth will be required to explore if the seed bank theory also can be applied to the plant microbiome ecosystems.

This study provides evidence for the presence of core microbiome residing in the endophytic compartment and rhizosphere of \textit{M. sinensis}. Here I identified a number of compartment-enriched OTUs shared between both habitats. These OTUs are commonly found in the samples collected from disparate locations, indicating they are consistently attracted by \textit{M. sinensis} plants, and their relative abundances reached 20 - 30\% in the corresponding compartment (Figure 4.2). Therefore, even though the number of shared compartment-enriched OTUs is small (16 EC-enriched OTUs, 55 R-enriched OTUs), the dominance of these OTUs indicates they are a good representation of the core microbiome in \textit{M. sinensis}.

I also observed that distinct bacterial species were enriched in the endophytic compartment and rhizosphere. Species from family Burkholderiaceae and Enterobacteriaceae are the dominant members of the EC-enriched OTUs (Figure 4.2). Members of these two families were also found to have strong correlations with plant fitness in naturalized \textit{M. sinensis} (Figure 4.3). Burkholderiaceae and Enterobacteriaceae were commonly found to associate with non-legume plants (Estrada-De los Santos et al., 2001; Rosenblueth & Martinez-Romero, 2006; Ryan et al., 2008). As EC-enriched OTUs, species from these families possess certain traits that facilitate entry to the endophytic compartment of \textit{M. sinensis}, such as 1-aminocyclopropane-1-carboxylate (ACC) deaminase production. Burkholderiaceae and Enterobacteriaceae species can promote plant growth by releasing ACC deaminase to lower plant ethylene levels (Hayat et al., 2010; Onofre-Lemus et al., 2009). Ethylene-related pathways are often involved with plant-microbe signaling and plant defense systems, and can be used as markers for showing how the
non-legume plants respond to the inoculation of diazotrophic strains (Nogueira et al., 2001). Therefore, ACC deaminase activity also enables bacteria to adjust the plant ethylene-related gene expression levels to achieve successful colonization. Other benefits brought by plant growth promoting species from Burkholderiaceae and Enterobacteriaceae families includes production of indole-3-acetic acid (IAA) (Halda-Alija, 2003), releasing plant growth promoting volatile organic compounds (VOCs) (Blom et al., 2011), fixing nitrogen (Bhattacharjee et al., 2008; Estrada-De los Santos et al., 2001), and enhancing plant cold tolerance (Fernandez et al., 2012). Interactions between plants and endophytic strains of family Burkholderiaceae and Enterobacteriaceae need to be investigated in future studies. Understanding such interactions will reveal how *M. sinensis* may benefit from species from these families and why *Miscanthus* consistently associates with them.

Species from family Bradyrhizobiaceae and Hyphomicrobiaceae are the dominant members of R-enriched OTUs. Species from Bradyrhizobiaceae are well known for their important roles in nitrogen cycling. Genus *Bradyrhizobium* includes many strains with high nitrogen-fixing capacity (Kaneko et al., 2002). Many *Bradyrhizobium* species form symbiotic relationship with legumes (Brewin, 1991; Young & Johnston, 1989) and also show great potential to be used as a plant growth promoting rhizobacteria (PGPR) for non-legumes (Antoun et al., 1998; Rashad et al., 2001). In our earlier studies, sequencing results showed *Bradyrhizobium* spp. were the most abundant nitrogen fixers associated with *M. ×giganteus* (Li, 2011). Therefore, those *Bradyrhizobium* are likely to form positive associative interactions with *M. sinensis* in natural sites. *Nitrobacter* is also a genus within the Bradyrhizobiaceae, and species from genus *Nitrobacter* are important nitrifying bacteria that oxidize nitrite to nitrate in soil (Prosser, 1989). Compared to Bradyrhizobiaceae, studies investigating plant growth promoting
effect of Hyphomicrobiaceae members are much less abundant. In terms of plant-microbe interactions, genus *Devosia* from this family includes species that form nodules on land and aquatic legumes (Bautista et al., 2010; Rivas et al., 2002). So far, no pathogenic activities have been reported from species of this family.

The importance of plant species overwhelmed geographic distance in determining the core endophytic microbiome. In the previous chapter, I found only six bacterial OTUs were shared between native *M. sinensis* and *M. floridulus* (Chapter 3, Figure 3.4a), but 16 bacterial OTUs were shared between native and naturalized *M. sinensis*. This result indicates endophytic community response to fine plant genetic variation. Studies of endophytic bacteria associated with other non-legume plants have shown different communities structures among genotypes (Johnston-Monje & Raizada, 2011; Sessitsch et al., 2004; van Overbeek & van Elsas, 2008), and the difference reflects the plant phylogenetic variation (Johnston-Monje & Raizada, 2011). On the other hand, the effect of geographic distance was more important than plant species for shaping the rhizosphere core microbiome. Many more R-enriched OTUs were shared between native *M. sinensis* and *M. floridulus* plants (Chapter 3, Figure 3.4b) than between native and naturalized *M. sinensis* plants. Results from Chapter 3 and previous studies investigating the rhizosphere community associated with maize suggest rhizosphere community structure correlate with plant genotype (Bouffaud et al., 2012; Peiffer et al., 2013). However, the majority of rhizosphere communities originate from surrounding soil bacterial pool. Results from Chapter 2 and Chapter 3 both indicate that rhizosphere bacterial communities are more sensitive to external environment than the endophytic ones, and compared with the endophytic communities, plant genotype has less control over the rhizosphere bacteria assemblages. Therefore, I conclude that
core endophytic microbes are more resistant to external environmental changes than core rhizosphere microbes, and more likely to form robust heritable interaction with the plant hosts.

Distinct bacterial co-occurrence patterns were observed in native and naturalized sites, but this is not due to different core microbiomes as I hypothesized. Although a large portion of the detected OTUs was similar in native and naturalized *M. sinensis* habitats, but these taxa did not function the same way. Bacterial communities from native and naturalized *M. sinensis* habitats showed very different co-occurrence patterns (Table 4.1, Figure C.2). Bacterial taxa from the native habitat were tightly interconnected, and members of communities that show this pattern usually have overlapped niches, and form strong ecological interaction. The ones from the naturalized habitat were loosely interconnected, and the assembly of such communities typically follows the neutral theory (Faust & Raes, 2012; Rosindell et al., 2011). It is possible that the after long-term interactions with the plant and other taxa, *M. sinensis*-associated bacteria have become stable and form close inter-species connections. The naturalized *M. sinensis* plants bring a series of relatively new bacterial niches to their introduced habitat, and the associated bacterial communities have not yet reached the equilibrium status. The naturalized *M. sinensis* co-occurrence networks have lower modularity.

**Conclusion**

In this study, I found a large portion of the detected OTUs were associated with *M. sinensis* plants in both their native and naturalized habitats, and 20 - 30% of the detected bacteria formed core microbiomes associated with *M. sinensis* at a global level. I was able to narrow down the dominant core plant-microbiome members to several families - Burkholderiaceae, Enterobacteriaceae, Bradyrhizobiaceae, and Hyphomicrobiaceae. Members of Burkholderiaceae, Enterobacteriaceae were also found to have the highest correlation with plant size. Using strains
of these families as model bacteria will reveal the plant microbiome recruitment process in future studies. Lastly, even though the OTUs were detected in both native and naturalized habitats of *M. sinensis*, their co-occurrence patterns were distinct. Understanding the plant microbiome assemblage is simply a first step exploring the plant-bacteria interaction. Both the present of the beneficial taxa and their interactions with surrounding bacteria could be important to the final outcome.

**Acknowledgements**

This chapter was completed in collaboration with Dr. Angela Kent, Dr. Jeffery Dangl, Dr. Lauren D. Quinn, Dr. Jacob N. Barney and Ryan F. Dougherty. DOE Joint Genome Institute conducted the Illumina sequencing. Scott Yourstone and Sur Herrera Paredes provided sequencing analysis assistance. I thank Yi Lou for statistical assistance. Dr. Anthony Yannarell contributed constructive comments to earlier versions of this manuscript.

**References**


Baas-Becking LGM (1934) *Geobiologie; of inleiding tot de milieukunde*. WP Van Stockum & Zoon NV.


Tables

Table 4.1. Network statistics for microbial association networks in native and naturalized *M. sinensis* sites. EC stands for endophytic compartment and R stands the rhizosphere soil.

<table>
<thead>
<tr>
<th></th>
<th>Co-occurrence (mean c-score)</th>
<th>Network size</th>
<th>Network density</th>
<th>Number of connected component</th>
<th>Average geodesic distance</th>
<th>Clustering coefficient</th>
<th>Modularity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native EC</td>
<td>0.35*</td>
<td>81</td>
<td>0.13</td>
<td>2</td>
<td>2.99</td>
<td>0.59</td>
<td>0.54</td>
</tr>
<tr>
<td>Naturalized EC</td>
<td>0.25*</td>
<td>67</td>
<td>0.03</td>
<td>15</td>
<td>2.08</td>
<td>0.47</td>
<td>0.83</td>
</tr>
<tr>
<td>Native R</td>
<td>0.3</td>
<td>575</td>
<td>0.10</td>
<td>1</td>
<td>2.60</td>
<td>0.57</td>
<td>0.01</td>
</tr>
<tr>
<td>Naturalized R</td>
<td>0.27</td>
<td>174</td>
<td>0.03</td>
<td>11</td>
<td>4.03</td>
<td>0.47</td>
<td>0.68</td>
</tr>
</tbody>
</table>

* The co-occurrence pattern is statistically non-random

Figures

a. Rhizosphere

b. Endophytic compartment

Figure 4.1. Venn diagram of the numbers of OTUs detected in the endophytic compartments (a) and the rhizosphere soil (b) in native and naturalized *M. sinensis* plants. Circle area represents number of OTUs detected. The overlapped areas are number of OTUs that found in both native and naturalized *M. sinensis* plants.
Figure 4.2. The relative abundance and family-level taxonomic information of the 16 EC-enriched OTUs and 55 R-enriched OTUs that found in both native and naturalized *M. sinensis* plants.
Figure 4.3. The relative abundance and family-level taxonomic information of the top 5% OTUs in endophytic compartment (a) and rhizosphere soil (b) that most correlated with plant fitness index.
CHAPTER 5: THE EFFECTS OF PLANT GENOTYPES AND LOCAL SOIL ON DIAZOTROPHS ASSOCIATED WITH MISCANTHUS

Abstract

*Miscanthus × giganteus* is a promising biofuel feedstock candidate that can be cultivated under various soil and climate conditions. Currently, *M. × giganteus* in the field of United States is one single genotype that is sterile and genetically unmodified. The narrow genetic variation within this genotype associates with potential ecological risks such as vulnerability to pests and crop diseases. Selecting and breeding new *Miscanthus* genotypes is imperative for the sustainable production of this plant. Previous work have suggested that diazotrophs associated with *M. × giganteus* contributes the plant nitrogen. However, the community composition of bacteria associated with the plant is under the influence of biotic factors such as plant genotypes and local soil bacteria. Therefore, the positive interaction between *Miscanthus* and diazotrophs may not remain effective across different locations and genotypes.

To better understand this association between *Miscanthus* and diazotrophs, the effects of *Miscanthus* genotypes and soil on the associative diazotrophs were examined. A common garden experiment using four *Miscanthus* genotypes (diploid *M. sinensis*, diploid *M. sacchariflorus*, triploid *M. × giganteus* and tetraploid *M. sacchariflorus*) and two soil sources was conducted. Community fingerprinting method - terminal restriction fragment length polymorphism, as well as 454-sequencing approach targeting *nifH* gene were used to profile the *Miscanthus*-associated diazotroph community.

Both approaches indicated that *Miscanthus* genotypes had a significant effect on the diazotroph communities residing in both endophytic compartment and rhizosphere of *Miscanthus*. The impact of soil source was strong on the rhizosphere communities, but not on the
endophytic ones. I also found that *Miscanthus* genotypes varied in their ability to enrich specific bacterial taxa. *M. sinensis* had the highest capability in enriching specific endophytic taxa, while diploid *M. sacchariflorus* recruit genotype-specific rhizosphere diazotrophs.

These results demonstrated that plant genotype is a key biotic factor shaping the diazotrophic assemblages. The plant-soil-microbe interactions need to be considered when evaluating beneficial plant-associated microbes.
Introduction

The microbiome residing in plant rhizosphere and endophytic compartment provide important ecological functions for their plant hosts, such as generating plant hormones, nutrient solubilization, and biological nitrogen fixation (BNF) (Hurek et al., 2002; Rashad et al., 2001; Rodríguez & Fraga, 1999). Among these functions, BNF conducted by diazotrophs provides an important but under-explored nitrogen source for non-legume plant species such as Miscanthus (Davis et al., 2010; Keymer & Kent, 2014). Nitrogen (N) is typically the most limiting nutrient in the agricultural system, and commercial N fertilizers represent a major fossil fuel use in this system (Vitousek et al., 1997). Strategies that decrease the need for N fertilizers while maintaining crop yield and soil fertility will bring large economic and ecological benefits. Developing such strategies is particularly important for biofuel crops because these strategies will increase the net bioenergy produced and ensure the environment-friendly bioenergy supply. Both outcomes help to achieve sustainable renewable energy production. Diazotrophs have been detected in the rhizomes and rhizosphere of Miscanthus plants (Davis et al., 2010; Keymer & Kent, 2014; Kirchhof et al., 1997; Li et al., 2015) and a previous study indicated that these diazotrophs contribute up to 16% percent of plant nitrogen demand to first year M. ×giganteus plants (Keymer & Kent, 2014).

The structure and functional diversity of non-legume associated microbes depend on by edaphic factors, plant genotypes, growth stage and other environmental factors (Berg & Smalla, 2009; Lauber et al., 2009; Lundberg et al., 2012; Tan et al., 2003). Biological N fixation has previously been detected in sugarcane – a close relative of Miscanthus. These previous studies showed that the benefit sugarcane plants obtained through BNF varied widely, depending on cultivar and cultivation locations. In Brazil, the percentage of sugarcane nitrogen provided by the
associative diazotrophs ranged from 34.8 – 77% for different cultivars (Baptista et al., 2014; Taulé et al., 2012). Additionally, the results also varied in different locations. $^{15}$N dilution and $^{15}$N natural abundance studies conducted in Brazil indicated that BNF was a significant nitrogen source for sugarcane and contributed to the sustainability of this gramineous bioenergy crop. However, the proportions of plant nitrogen gained through BNF were reported to be lower or negligible in Japan (Thaweenut et al., 2011) and Africa (Hoefsloot et al., 2005). Based on these observations in sugarcane, I predict that the positive interaction between Miscanthus and diazotrophs may not remain effective across locations and different genotypes, even for the clone “Illinois” planted in different locations.

Currently in the US, all Miscanthus biomass production is based on a single sterile clone - “Illinois” (Matlaga & Davis, 2013). This leads to concerns towards sustainability and productivity. This narrow genetic basis exposes Miscanthus to ecological risks such as pests and crop diseases (Stefanovska et al., 2011; Stewart & Cromey, 2011). Due to its sterility, M. ×giganteus “Illinois” is reproduced through vegetation propagations, which brings high establishment costs (Zub & Brancourt-Hulmel, 2010). Genotypes that reproduce by seeds are needed to reduce energy input and make more profit. Additionally, Miscanthus breeding projects have shown that late flowering, or no flowering genotypes have higher yields (Heaton et al., 2008; Sacks et al., 2013). To sum up, there is still much space to improve Miscanthus characteristics, and breeding new genotypes is imperative.

Positive interactions between Miscanthus ×giganteus “Illinois” clone and their diazotroph associates have been observed (Keymer & Kent, 2014). However, there is no guarantee that this mutual interaction can be maintained under other conditions. Understanding the ecological drivers of bacterial communities associated with Miscanthus genotypes is the first step to
estimate how the associative interaction between Miscanthus species and diazotrophs may vary across different field sites and different Miscanthus genotypes. Thus far, most studies reporting on diazotrophs associated with Miscanthus species have employed culture-based approaches (Davis et al., 2010; Miyamoto et al., 2004; Rothballer et al., 2008; Ye et al., 2005), and the community level work has been conducted only on M. ×giganteus (Keymer & Kent, 2014; Li et al., 2015). In this study, a greenhouse experiment was conducted to investigate how Miscanthus genotype and local soil sources may influence the Miscanthus-associated diazotrophs. A high-throughput sequencing approach was used to access the diazotroph assemblages. I hypothesize that since the majority of rhizosphere bacteria originate from local soil, the rhizosphere diazotroph assemblage will be similar to local soil diazotroph pool. Another hypothesis of this study is that plant genotype-specific factors control the plant-diazotroph interactions, and, therefore, diazotroph assemblages associated with Miscanthus differ among genotypes.

Materials and methods

Planting and initiation of experiment

Four Miscanthus genotypes were included in this experiment - triploid M. ×giganteus, diploid M. sinensis, tetraploid M. sacchariflorus and diploid M. sacchariflorus (Table 5.1). The genetic origin of triploid M. ×giganteus is not fully understood. One hypothesis is that the tetraploid M. sacchariflorus crossed with diploid M. sinensis, resulting in triploid M. ×giganteus (Clark et al., 2015; Hodkinson et al., 2002; Rayburn et al., 2009). Currently, there is no direct evidence suggesting the parental species of tetraploid M. sacchariflorus, but it could be a result of hybridization between diploid M. sinensis, and diploid M. sacchariflorus. No genetically identical Miscanthus seeds are available, so to avoid plant-to-plant genetic variation within a
genotype, young clonal ramets (~ 10 cm tall) genetically identical within each genotype were used. Sources of these ramets were described in Table 5.1.

To evaluate the effect of the local soil community, I used field soil from Miscanthus plots known to have distinct diazotroph communities, in terms of both composition and abundance (Figure D.1). The bulk soil was collected from Miscanthus fields located in Dixon Springs, IL (37.45, -88.67) and Urbana, IL (40.08, -88.22). These two sites were chosen because earlier experiments conducted in Dr. Kent’s lab indicated that soil communities in these two farms contained diazotrophs that differ in both composition and abundance (Figure D.1). The soil was passed sequentially through sieves to a 4-mm size, and the bulk soil from each location was mixed and homogenized at a 1:2 ratio with fine sands. Subsets of the soil samples from both locations were collected to accesses the initial soil microbial community composition. The homogenized soil-sand mixture was distributed in 1-gallon pots. Young Miscanthus plants were transplanted into these 1-gallon pots so that each pot contains one Miscanthus plant. Subsets of the plant with eight replicates for each genotype were collected to access the initial microbial communities residing in the rhizosphere and endophytic compartment. Pots were arranged in a completely randomized manner and watered generously with tap water.

Plant harvest and processing

Harvest occurred at 10 and 20 weeks after transplantation. During each harvest, eight plants from each genotype × soil source combination were collected. Rhizome processing was conducted using the methods described by Keymer and Kent (2014). Bacterial communities were extracted from each rhizome (endophytic compartment) with FastDNA Spin Kit (MP Biomedicals, Irvine, CA). Rhizosphere compartment is defined as the soil washed from rhizomes after shaking off all the loose soil. Rhizosphere soil was lyophilized prior to DNA
extraction. Bacterial communities residing in the rhizosphere compartment were extracted using the FastDNA SPIN Kit for Soil (MP Biomedicals, Irvine, CA), and further purified with a cetyl trimethyl ammonium bromide (CTAB) extraction procedure to remove contaminating humic acids (Sambrook & Russell, 2001).

**Molecular fingerprinting approaches**

Diazotroph communities were accessed using terminal restriction enzyme fragment length polymorphism (T-RFLP) targeting the *nifH* gene. The *nifH* gene was amplified using primers PolF (6-FAM-5’-TGCGAYCCSAARGCBGACTC-3’) and PolR (HEX-5’-ATSGCCATCATYTCRCCGGA-3’) (Poly et al., 2001). PCR reactions were conducted as described by Keymer and Kent (2014). The *nifH* amplicons were purified with MinElute 96 UF PCR Purification Kit (QIAGEN, Valencia, CA). Next, the amplicons were digested using restriction enzyme *Mbo*II and *Mnl*II for 16 hours at 37°C following the procedures recommended by New England Biolabs (Beverly, MA). The lengths of the fluorescently labeled terminal restriction fragments were determined by denaturing capillary electrophoresis performed by Roy J. Carver Biotechnology Center (Urbana, IL). Size-calling was carried out using GeneMarker 2.4 (SoftGenetics, State College, PA). Capillary electrophoresis results in minor run-to-run variations in observed against actual fragment length were resolved using the allele-calling features in GeneMarker before analysis. To include the maximum number of peaks while excluding background fluorescence, a threshold of 200 fluorescence units was used. The signal strength (i.e. peak area) of each peak was normalized to account for run-to-run variations in signal detection by dividing the area of individual peaks by the total fluorescence (area) per restriction enzyme per fluorescent tag, expressing each peak as a proportion of the observed community (Rees et al., 2004; Yannarell & Triplett, 2005). The normalized T-RFLP data were
then concatenated to include information from both restriction enzymes and both fluorescent tag in one sample by operational taxonomic unit (OTU) table prior to analysis.

Based on the T-RFLP data, the dissimilarity between diazotroph communities from different samples was calculated using Bray-Curtis dissimilarity matrix. Based on the dissimilarity matrix, the diazotroph community assemblage patterns were visualized using correspondence analysis (CA) results generated by CANOCO 5.6 (Biometrics, Wageningen, The Netherlands), and the effect of plant genotype, soil source and harvesting time were determined using Permutational Multivariate Analysis of Variance (PerMANOVA) (Anderson, 2001) in R (R_Development_Core_Team, 2010).

*High-throughput sequencing*

To access the diazotroph species enriched by different plant genotypes, tag-encoded 454 pyrosequencing of the *nifH* gene was performed on representative samples chosen based on the *nifH* T-RFLP results. These samples were selected from 1) the initial plants and 2) plants receiving Urbana farm soil harvested at the 20th week. PCR for *nifH* pyrosequencing contained 1X Phusion® HF buffer (Thermo Scientific, Hudson, NH), 0.25 mg of bovine serum albumin per mL (New England Biolabs), 0.20 mM of dNTPs (New England Biolabs), 0.5 µM of each primer, 0.05 U/µL of Phusion® high fidelity polymerase (Thermo Scientific, Hudson, NH) and 30 ng of DNA template in a final volume of 25 µL. A touchdown program was employed for a more precise amplification. The program started with an initial step at 98 °C for 2 min, followed by 30 cycles of denaturation at 98 °C for 10 s, annealing at variable temperatures for 30 s, and extension at 72 °C for 20 s. Among the 30 cycles, the annealing temperature for the 1st cycle was 64 °C, the 2nd and 3rd cycles were 62 °C, the 4th to 6th cycles were 60 °C, the 7th to 10th cycle
were 58 °C, and the last 20 cycle were 56 °C. Finally, the program ends with extension at 72 °C for 10 min. The PCR amplicons were purified using QIAquick Gel Extraction Kit (QIAGEN) following the manufacturer's protocol. The \textit{nifH} amplicons were sequenced by Roy J. Carver Biotechnology Center using Roche/454 Genome Sequencer FLX+ (Roche/454 Life Sciences, Branford, CT).

Sequences reported in this work have been deposited in MG-RAST (Project 13328) (Meyer et al., 2008). Each raw sequence read was assigned to its original sample according to the oligo index, and both forward and reverse primers were trimmed. Sequences with an average quality score lower than 25 or containing any ambiguous “N” bases were removed. The remaining sequences were further processed using FunFrame pipeline (Weisman et al., 2013). Through FunFrame, I (1) corrected sequencing errors by comparing sequencing reads to \textit{nifH} hmm frame provided by Fungene Functional Gene Repository (http://fungene.cme.msu.edu/hmm_download.spr?hmm_id=328); (2) removed chimeric sequences by calling UCHIME (Edgar et al., 2011); (3) clustered the sequences to 97% similarity using ESPRIT-Tree (Cai & Sun, 2011); (4) selected representative OTUs and generated a sample by OTU table. The taxonomic attributes of the representative OTUs were determined using uclust (Edgar, 2010), and the references data sets are derived from the Fungene \textit{nifH} database (http://fungene.cme.msu.edu/hmm_download.spr?hmm_id=328). All singleton reads were removed before conducting any community diversity or similarity analysis.

Phylogenetic dissimilarities of diazotroph communities among samples were estimated using UniFrac distance (Lozupone et al., 2011), and visualized based on the results of principal coordinates analysis (PCoA) performed with QIIME (Caporaso et al., 2010). To show how detected OTUs were partitioned among samples, OTU network analyses were conducted, and the
results were visualized using Cytoscape 2.8.3 (Shannon et al., 2003). The OTUs enriched in specific plant genotypes were identified using the R function indval from package labdsv (Roberts, 2013). This method takes into account both the frequency of each OTU present in a plant genotype as well as the relative abundance of different genotypes. The effect of plant genotype on the \textit{nifH} gene diversity was determined using PerMANOVA (Anderson, 2001) in R.

**Results**

\textit{Diazotroph community composition pattern based on T-RFLP results}

A total of 887 diazotroph terminal restriction fragments (TRFs) were detected in all samples using \textit{nifH} T-RFLP. Community composition of diazotrophs differed significantly among plant genotypes (Figure 5.1, Table 5.2). A strong correlation between plant genotype and associated diazotroph community composition was observed for all harvesting times, in both soil types and also in both compartments. Comparisons of diazotroph community composition between different bulk soil sources were also conducted (Table 5.3). Before plant cultivation started, the different soil types explained 38.38\% (p = 0.093) of observed variance among diazotroph community. After 10 weeks of cultivation, endophytic diazotroph communities growing in different soil showed no significant differences in all plant genotypes except for tetraploid \textit{M. sacchariflorus}, and after 20 weeks the soil source became non-significant in all genotypes. For the rhizosphere communities after 10 weeks cultivation, the percentage of diazotroph community variance that can be explained by soil type was lower than that of the original bulk soil, but still significant in all genotypes. After 20 weeks, the rhizosphere communities of plant growing in different soil types became less different, and the difference among soil types became non-significant for triploid \textit{M. ×giganteus}. However, I did observe that diazotrophs residing in the rhizosphere soil became more and more similar to the communities
from original bulk soil (Figure D.2 and Figure D.3). The results indicated that, in general, the plant genotype effects were very strong throughout the experiment. The effect of bulk was mainly on the rhizosphere communities.

_Diazotroph community composition pattern based on sequencing results_

To further investigate which taxa were retained or recruited leading to the different diazotroph communities observed in different plant genotypes, 454 pyrosequencing on *nifH* gene was conducted for selected representative samples from different genotypes that were harvested at time 0 and 20 weeks cultivation. Only samples growing in Urbana soil were included in the sequence analysis.

A total of 277,350 reads with good quality were detected from the 67 selected samples using the 454 sequencing method, ranging from 1270 to 8133 reads per sample. After clustering at 97% similarity and removing all the singletons, the reads were classified into 5043 OTUs. There were clearly more detected OTUs (Figure 5.2) and higher alpha-diversity (Figure D.4) in the rhizosphere than those in the endophytic compartment. After 20 weeks cultivation, the number of estimated diazotroph species (Chao1 index) in rhizosphere samples increased at least 50% to 100%. The rhizosphere Chao1 rarefaction curve did not reach a plateau at the sequencing depth used. On the other hand, the number of estimated endophytic diazotroph species was maintained at a similar level after 20 weeks cultivation, except for *M. sinensis* (Figure D.5).

Similar to the results from T-RFLP approach, plant genotype had a strong effect on the community assemblages of endophytic diazotrophs (Table 5.4). As the plant-OTU network shows in Figure 5.2, I found that different *Miscanthus* genotypes tended to harbor distinct OTUs in their endophytic compartment, and also there were some plant-to-plant variations within the same genotype. Rhizosphere communities from different plant genotypes started with dissimilar
diazotroph communities. However, after 20 weeks cultivation in the same soil source, they became less distinguishable except for diploid *M. sacchariflorus* (Figure 5.2). Soil factors had a stronger effect on the rhizosphere diazotroph assemblages than on the endophytic ones.

*The relative abundance of taxonomic groups varied among Miscanthus genotypes*

Taxonomic groups detected differed across plant genotypes, between compartments, and changed over time (Figure 5.3, Table D.1). Most taxa showed greater variance in the endophytic compartment than in the rhizosphere compartments for the same genotype, indicating a plant-to-plant variation within the same genotype.

*Proteobacteria* is the most dominant *Miscanthus*-associated diazotrophs. The detected *Proteobacteria* mainly belong to *Alpha*, *Beta*, and *Deltaproteobacteria* classes. Interestingly, *Miscanthus*-associated *Deltaproteobacteria* was almost undetectable at the beginning of the cultivation. The relative abundance of *Deltaproteobacteria*, especially species belonging genera *Geobacter* and *Anaeromyxobacter*, increased during cultivation, which probably resulted from the high abundance of these genera in bulk soil from Urbana fields (Table D.1). I also observed that there were more species from genera *Desulfovibrio* following cultivation even though they were not detected in the bulk soil or the rhizosphere soil at time 0 (Table D.1). These results suggest some soil bacteria migrate to and enriched in the plant rhizosphere during cultivation.

*Bradyrhizobium* from *Alphaproteobacteria* and *Burkholderia* from *Betaproteobacteria* were the most abundant *Miscanthus*-associated diazotroph genera that were identified. Their relative abundance varied a lot among different plant genotypes, and generally decreased over time. Diploid *M. sinensis* had a consistent association with *Bradyrhizobium* while diploid *M. sacchariflorus* formed a consistent association with *Burkholderia*. 
Among OTUs with relatively high abundance, the relative abundance of OTU12752 showed interesting genotype-to-genotype variation. OTU12752 is an uncultured *Proteobacteria* (80% confidence) most closely related to *Pseudomonas azotifigens* (42% confidence) (Figure D.6). When compared with the NCBI blast nucleotide database, the *nifH* gene sequence of OTU12752 was found to match no known species at 97% or higher similarity. However, the sequence does match 100% or 99% to uncultured diazotrophs from two aquatic microbial studies (Jayakumar et al., 2012; Turk et al., 2011). OTU12752 was found in high abundance in almost all samples from the endophytic compartment, but was rarely found in the rhizosphere or bulk samples. Additionally, it was enriched significantly in the endophytic compartment of the two hybrid species triploid *M. ×giganteus*, and tetraploid *M. sacchariflorus* after 20 weeks cultivation (Figure 5.3, Table D.1).

**OTUs enriched in specific genotype**

At the OTU level, Miscanthus genotypes varied in their abilities of recruiting and retaining diazotrophs. Figure 5.4 shows the number of OTUs whose relative abundance differed significantly among different genotypes for each harvesting time × compartment combination. For the endophytic compartments harvested at the twentieth week, the majority of the significantly enriched OTUs were observed in *M. sinensis*.

At the 80% confidence level, these 31 OTUs enriched in *M. sinensis* were classified as members of genus *Bradyrhizobium*, order *Rhizobiales*, or uncultured *Proteobacteria* (Table D.2). The endophytic compartments of *M. sinensis* also had the highest estimated diazotroph species among all genotypes (Figure D.5). This evidence indicates that the *M. sinensis* endophytic compartment tends to have a distinct diazotroph assemblage and greater diazotroph richness compared with other Miscanthus genotypes.
Eight out of these 42 OTUs were enriched in diploid *M. sacchariflorus*, including OTUs from order *Burkholderiales* and *Rhizobiales*, as well as two OTUs that were uncultured *Proteobacteria* (Table D.2). Meanwhile, even though the total overall estimated diazotroph species in the diploid *M. sacchariflorus* rhizosphere was lower than the other genotypes (Figure D.5d), more than 100 OTUs were specifically enriched in the rhizosphere compartments of this genotype (Figure 5.4), much more than the other genotypes. This agrees with the results shown in Figure 5.2, and that diploid *M. sacchariflorus* tended to have distinct rhizosphere diazotroph communities after 20 weeks of cultivation.

Three endophytic diazotroph OTUs were specifically enriched in triploid *M. ×giganteus*, including OTU12752 mentioned above. The average relative abundance of this OTU in the endophytic compartment of triploid *M. ×giganteus* reached 80%.

**Discussion**

In this Chapter, I have explored the rhizosphere and endophytic diazotroph communities that are associated with four *Miscanthus* genotypes grown in two different soils. The objectives were to investigate the effects of soil and plant genotype on the *Miscanthus*-associated diazotrophs, as well as to examine how genotypes differed in recruited and retained endophytic diazotroph taxa. I found the plant genotype had a strong influence on the diazotrophs living in endophytic compartment and rhizosphere compartment, and this influence was consistently significant over time. Different bulk soil sources leaded to distinct rhizosphere diazotrophs over time, but the effect of soil source on endophytic diazotrophs was weaker than for the rhizosphere diazotrophs. Among all the genotypes used in this study, *M. sinensis* endophytic compartments had more genotype specialists, and the richness of endophytic diazotroph communities living in *M. sinensis* is much higher than the other genotypes. Tetraploid *M. sacchariflorus* and triploid *M.
×giganteus, the genotype currently being used for bioenergy production, tended to enrich an uncharacterized diazotroph from *Proteobacteria* phylum.

Our results support the two-step selection model that has been used to explain the microbiome formed in plant root or rhizomes (Bulgarelli et al., 2013). The first selection step aims at selecting rhizosphere communities. Rhizosphere communities likely originated from bulk soil (Bais et al., 2006) and rhizosphere community assemblage is regulated by the plant cell wall features and rhizodeposition. Correspondingly, I found that rhizosphere diazotroph communities were strongly influenced by the local soil communities (Lundberg et al., 2012; Peiffer et al., 2013). For plants cultivated in Dixon Springs farm soil (Figure D.2), rhizosphere communities became similar to the Dixon Springs bulk soil over time. Moreover, diazotrophs such as those from *deltaproteobacteria* initially were present in the bulk soil but not associated with plants, and were later found to be quite abundant after 20 weeks’ cultivation. The different rhizosphere diazotroph assemblages across genotypes may due to distinct rhizome structures (Kaack et al., 2003) and different plant exudate compositions (Kaňová et al., 2010; Kochian et al., 2005). But differing from the model’s prediction, the diazotroph community richness in the rhizosphere is higher than that in bulk soil. One potential explanation is that enriched and diverse carbon sources in the rhizosphere (Alexander, 1977) boosted the abundance of diazotrophs whose abundance was below the detection threshold in the bulk soil. In the second step of plant microbiome assembly, genotype-dependent factors determine the subset of microbes that can access the internal plant tissue. Diazotroph richness inside the plant significantly lower than in the rhizosphere and bulk soil, indicating that the access to plant endophytic compartment was limited. In this study, both *nifH* T-RFLP (Figure 5.1, Table 5.2) and 454-sequencing (Figure 5.2, Figure 5.4, Table 5.4) approaches showed that plant genotype plays a key role in determining the
endophytic diazotrophs (Lundberg et al., 2012; Tan et al., 2003). These results implied that the nutritional requirements and genetic markers that interact with diazotrophs differently among genotypes. By comparing the plant microbiomes of closely related plant genotype and their plant genetic marker, future work can identify genetic markers that regular plant-microbe interaction. This result suggested the possibility of intentionally modulate the plant microbiome through plant breeding (Peiffer et al., 2013).

The results generated in this study have shown that diazotroph communities associated with Miscanthus received a combined effect from both plant genotype and local soil effects. The relative strength of plant or soil influence on the associated diazotrophs varied between compartments and across genotypes. In general, inside the plant rhizome, the plant genotype effect is the primary driver of diazotroph communities the diazotroph communities (Table 5.2, Figure 5.2), which may be determined by some plant genetic markers, whereas local soil communities are the main driver for rhizosphere diazotrophs. I observed that diploid M. sacchariflorus had a strong genotype effect on both endophytic and rhizosphere communities while the influence of plant genotype was observed mostly for communities in the endophytic compartment for the other three genotypes (Figure 5.2). It has been shown that the plant can actively attract rhizosphere bacteria by releasing root exudate, and bacterial taxa that assimilate root exudates could form assemblages distinct from the surrounding soil bacteria pool (Berendsen et al., 2012). Comparing the root exudates among these genotypes would be an interesting next step for a better understanding of substance that leads to these different diazotroph assemblages and potential interaction signals released by the plant.

As reported by other studies that have investigated endophytes associated with non-legumes (Bahulikar et al.; Gottel et al.; Lundberg et al.; Thaweenut et al., 2011), endophytic
diazotroph assemblages are dominated by *Proteobacteria* species. Within the known *Proteobacteria* detected, *Alphaproteobacteria*, especially *Bradyrhizobium* species, had high abundance in almost all samples. A handful of *Bradyrhizobium* species were also found to be specifically enriched in the endophytic compartment of diploid *M. sinensis* and diploid *M. sacchariflorus*. *Bradyrhizobium* species are common known symbiotic bacteria that forming nitrogen-fixing nodules for legumes (Brewin, 1991; Young & Johnston, 1989). *Bradyrhizobium* species have been widely used in agricultural systems due to their high nitrogen fixing capabilities (Kaneko et al., 2002). When inoculating *Bradyrhizobium* strains on non-legumes, growth increases have been shown as well (Chaintreuil et al., 2000). Weak nitrogenase activity can be detected (Chaintreuil et al., 2000; Rouws et al., 2013), but not enough to support the growth increase observed. In fact, the growth-promoting effect on non-legumes may results from the phytohormone production by *Bradyrhizobium*. *Bradyrhizobium* had a promising growth stimulatory effect by producing indole acetic acid (IAA), gibberellic acid (GA₃) and abscisic acid (ABA) (Antoun et al., 1998; Boiero et al., 2007; Rashad et al., 2001). It is likely that the high abundance of *Miscanthus*-associated *Bradyrhizobium* brings benefits that other than BNF to their hosts.

The fact that OTU12752 occurs almost exclusively in the endophytic compartment and its high abundance in the endophytic component of hybrid *Miscanthus* species is noteworthy. Unfortunately, the taxonomic information of this sequence is very limited. Based on the known species pool, it likely belongs to the *Pseudomonas* genus. *Pseudomonas* species are known as one of the best root colonizers (Lugtenberg et al., 2001; Lugtenberg & Dekkers, 1999), and have been reported to colonize rhizosphere soil (Thomashow et al., 1990; Vázquez et al., 2000) and roots of variety non-legume crops (Mendes et al., 2007; Ramos-González et al., 2005). Different
species can act as either plant pathogens (Guttman et al., 2002; Sarkar & Guttman, 2004) or plant growth promoting microbes (Vessey, 2003; Xie et al., 1996). The nifH amplicon sequence of OTU12752 is identical to the uncultured diazotrophs in an oxygen-deficient marine system and tropical aquatic system. It is possible that OTU12752 is an OTU that is adapted to and thrives under the low oxygen condition in plant endophytic compartment. The nitrogen fixation process is sensitive to oxygen (Gallon, 1981), and therefore the low oxygen microenvironment is essential for the growth of and best nitrogen fixation performance of endophytic diazotrophs (Reinhold-Hurek & Hurek, 1998). Given its high abundance in the endophytic compartment, OTU12752 is potentially the main contributor of BNF detected in Miscanthus. Future studies that explore the genomic information of this OTU are needed in order to understand its function and roles in the endophytic compartment of Miscanthus.

Conclusion

In summary, this study shows evidence that both Miscanthus genotypes and local soil sources influence plant-associated diazotroph assemblages. As a result, to maintain the sustainable production of selected bioenergy crops across different location, we need to take the interaction of diazotrophs-genotype and diazotroph-soil interactions into consideration. Plant alleles responsible for diazotroph variation are not clear yet. Larger bioenergy crop population surveys are necessary to characterize plant genetic markers and plant-microbe signal exchanges behind the observations.

Acknowledgements

This chapter was completed in collaboration with Dr. Angela Kent. We would like to thank Dr. Erik Sacks for kindly provided the hybrid plants used in this work. Dr. Daniel P. Keymer contributed to the experimental design and provided technical assistance in the greenhouse. Dr.
Anthony C. Yannarell and Dr. Chinmay Prakash Soman contributed valuable comments to an earlier version of this chapter. This work was supported by the Energy Biosciences Institute and Campus Research Board Programs - University of Illinois at Urbana-Champaign.

References


Tables

Table 5.1. The genetic information and sources of the four genotypes used in this study

<table>
<thead>
<tr>
<th>Species</th>
<th>Variety</th>
<th>Haploid number of chromosomes</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. sinensis</td>
<td>'Gracilimus'</td>
<td>diploid</td>
<td>Emerald Coast Growers, Pensacola, FL</td>
</tr>
<tr>
<td>M. sacchariflorus</td>
<td>'Robustus'</td>
<td>diploid</td>
<td>Kurt Bluemel, Inc., Baldwin, MD</td>
</tr>
<tr>
<td>M. sacchariflorus</td>
<td>'Ogi'</td>
<td>tetraploid</td>
<td>Provided by Dr. Erik Sacks</td>
</tr>
<tr>
<td>M. × giganteus</td>
<td>'Illinois'</td>
<td>triploid</td>
<td>Provided by Dr. Erik Sacks</td>
</tr>
</tbody>
</table>

Table 5.2. Variances explained by Miscanthus genotype in each compartment and soil source over time. Results are calculated based on the nifH T-RFLP community fingerprinting data using PerMANOVA

<table>
<thead>
<tr>
<th>Compartment/Soil source</th>
<th>Endophytic</th>
<th>Rhizosphere</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 week</td>
<td>10 week</td>
<td>20 week</td>
<td>0 week</td>
<td>10 week</td>
<td>20 week</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urbana</td>
<td>0.41341 ***</td>
<td>0.1903 ***</td>
<td>0.28174 ***</td>
<td>0.37166 ***</td>
<td>0.26012 ***</td>
<td>0.22091 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dixon Spring</td>
<td>0.20757 ***</td>
<td>0.16797 ***</td>
<td></td>
<td>0.29763 ***</td>
<td>0.15945 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significance codes: *** p < 0.001, ** 0.001 < p < 0.01, * 0.01 < p < 0.05
Table 5.3. Variances explained by soil source in each compartment and Miscanthus genotypes over time. Results are calculated based on the nifH T-RFLP community fingerprinting data using PerMANOVA.

<table>
<thead>
<tr>
<th>Compartment/Genotype</th>
<th>Bulk Soil</th>
<th>Endophytic</th>
<th>Rhizosphere</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 week</td>
<td>10 week</td>
<td>20 week</td>
</tr>
<tr>
<td>Triploid M. × giganteus</td>
<td>0.38384</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diploid M. sacchariflorus</td>
<td>0.05098</td>
<td>0.07833</td>
<td>0.20358***</td>
</tr>
<tr>
<td>Diploid M. sinensis</td>
<td></td>
<td>0.05986</td>
<td>0.04809</td>
</tr>
<tr>
<td>Tetraploid M. sacchariflorus</td>
<td>0.12057*</td>
<td>0.05672</td>
<td>0.21164***</td>
</tr>
</tbody>
</table>

Significance codes: '***' p < 0.001, '**' 0.001 < p < 0.01, '*' 0.01 < p < 0.05

Table 5.4. Variances explained by Miscanthus genotypes in each compartment at Week 0 and Week 20. Results are calculated based on the nifH 454-sequencing data using PerMANOVA. Only Week 0 bulk soil and plants cultivated in Urbana soil at Week 0 and Week 20 were included in 454-sequencing analysis.

<table>
<thead>
<tr>
<th>Compartment</th>
<th>Endophytic</th>
<th>Rhizosphere</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 week</td>
<td>20 week</td>
</tr>
<tr>
<td></td>
<td>0.31603*</td>
<td>0.43787***</td>
</tr>
</tbody>
</table>

Significance codes: '***' p < 0.001, '**' 0.001 < p < 0.01, '*' 0.01 < p < 0.05
Figure 5.1. Correspondence analysis biplots of diazotroph assemblages residing in the endophytic (A) and rhizosphere (B) of Miscanthus after 20 week’s cultivation in Urbana farm soil. These biplots are based on the results of nifH T-RFLP based community fingerprinting data. Each point represents a diazotroph assemblage from one sample and points are color-coded by Miscanthus genotype.
Figure 5.2. 454-sequencing results based network analysis plots display how detected OTUs are partitioned among week 0 endophytic samples (A), week 0 rhizosphere samples (B), week 20 endophytic samples (C), and week 20 rhizosphere samples (D). Each small dot represents an OTU that detected in this study, and each large circle represents a plant sample. The samples are color-coded by Miscanthus genotype. Samples with more shared OTUs are clustered closer together.
Figure 5.3. Bar plots of the relative abundance of main taxonomic groups, detected in week 0 endophytic samples (A), week 0 rhizosphere samples (B), week 20 endophytic samples (C), week 20 rhizosphere samples (D), and week 0 bulk soil (E). Different taxonomic levels (class, genus, and OTU) are separated by dash lines. Bars are color-coded by Miscanthus genotype. Error bars represent the standard error.
Figure 5.4. Logarithm with base 10 of the relative abundances of OTUs that significantly enriched in one genotype. These OTUs were selected using indval function in the R package labdev with familywise p value < 0.05. A small value ($10^{-5}$) was added to the relative abundance of OTUs before taking logarithm to avoid the not defined value of $\log_{10} 0$. Each point represent the relative abundance of the OTU labeled on Y-axes in one sample, can they are color-coded based on the Miscanthus genotype of that sample.
CHAPTER 6: SUMMARY AND CONCLUSION

This work examined the bacterial communities that associated with Miscanthus species and brought new angles for future investigation into plant-bacteria interaction and plant breeding.

Admittedly, this work has some limitations.

1) I examined the total bacterial community compositions and their taxonomic information, but the bacterial functions were not directly measured. Certain taxa were considered to have potential to promote plant growth because i) phylogenetically close related species have been reported to be plant growth promoting taxa; or ii) they showed strong positive correlation with Miscanthus plant fitness that measure by plant size. Diazotroph taxa detected in this work carry the key gene encoding nitrogenase reductase, and previous studies detected positive nitrogen-fixing activity from Miscanthus plant tissue (Davis et al., 2010; Keymer & Kent, 2014). However, I did not measure the nitrogenase activity in this work.

2) For the sequence data, the variation of the number of reads per sample is very large. In Chapter 3, more than half endophytic samples generated too few reads to be included in the statistical analysis. This mainly due to the primers (515F and 806R) amplified both plant and bacterial DNA. As a result, the raw 16S rRNA sequence reads generated from the endophytic samples contained a large proportion (in some cases 90% or higher) of plant DNA. To generate more endophytic bacterial sequences, each endophytic sample was sequenced twice on MiSeq, but this approach still failed to eliminate the problem. However, the sequence data were still adequate for me to examine the taxa found and enriched in the endophytic compartment of native Miscanthus plants, as well as how the endophytic communities correlated with the local environment.
3) The sequence depth in this work is relatively shallow, meaning I may fail to detect locally rare bacterial taxa. On average I obtained approximately 3,000 reads per sample for the 16S rRNA gene (Chapter 3 & 4), and I obtained about 4000 reads per sample targeting the nifH gene (Chapter 5). Caporaso and co-workers conducted deep-sequencing on a marine sample, from which they generated 10 million sequence reads (Caporaso et al., 2012). A comparison between reads generated from this work to a global marine bacterial database suggested that a sequencing depth of $1.93 \times 10^{11}$ is needed to capture all the phylogenetic diversity reported in the global database (Gibbons et al., 2013), and shallow sequencing approaches fail to detect members of core microbiota. The terrestrial ecosystem is usually more complicated and contains more niches than the aquatic system. Therefore, it is possible that I need to have an even deeper sequencing depth to capture all the phylogenetic information. Hence, the species forming the core endophytic or rhizosphere microbiome reported in this work may not be a complete list.

Even with these limitations being considered, this work still generated some solid conclusions as follows.

1) Miscanthus plant genotype actively shapes the associated total bacterial and diazotroph community compositions. As described in Chapter 3, plant-to-plant genetic variation explained a significant amount of Miscanthus-associated bacterial community variation. In Chapter 5, I observed that these closely related Miscanthus genotypes tended to recruit very different N-fixing bacteria. It also has been reported that when different diazotroph strains attempted to colonize the rhizome of Miscanthus sinensis, the plant responded differently (Straub et al., 2013). Finding the plant genetic markers that regulate plant-bacteria pathway and control endophytic colonization is a key next step towards better understanding of plant-microbe interactions.
2) Plant endophytes and rhizosphere taxa responded differently to the surrounding edaphic factor. The majority of endophytic taxa came from the rhizosphere soil, and the endophytic communities have much lower richness than the rhizosphere communities, but they are not random subsets of rhizosphere communities. I consistently observe distinct endophytic and rhizosphere bacterial communities in this work. Endophytic communities were more sensitive to plant genetic differences and less sensitive to soil edaphic factors than the rhizosphere communities.

Compartment-enriched taxa were detected in both endophytic compartment and rhizosphere compartment. Here the compartment-enriched taxa were defined as common species that were present in at least 25% samples and had significantly higher relative abundance in one of the two plant compartments (endophytic and rhizosphere). Different responses to abiotic and biotic factors were observed between the EC-enriched and R-enriched taxa. From Chapter 4, I found that a plant is more likely to share EC-enriched bacteria with phylogenetically close plants and are more likely to share R-enriched bacteria with geographically close plants.

To conclude, comparing with the rhizosphere bacteria taxa, endophytic taxa are more likely to be plant species-specific and stable in various soil conditions. While it is still under debate whether endophytes or rhizosphere communities contribute more to plant fitness, my results suggest that to establish a stable mutualistic plant-bacteria partnership for commercial usage, plant endophytic bacteria are better candidates than the rhizosphere bacteria.

3) The presence of similar bacterial taxa does not naturally lead to similar inter-species interactions. Understanding the plant microbiome assemblage is simply a first step exploring the plant-bacteria interaction. In Chapter 4, highly similar (if not identical taxa) were observed in both the native and naturalized habitats of *M. sinensis*, however, the co-occurrence pattern of
these endophytes were completely different in these two habitats, implying that they may functioning in a distinct manner in these habitats. For future studies optimizing plant-microbe interactions, both the present of the beneficial taxa and their interactions with surrounding bacteria are important to the final outcome.

To further investigate plant microbiome and plant-microbe interactions, future studies are needed.

1) To reveal of the mechanism of plant controlling bacteria recruitment, plant genotyping is needed to provide candidate genetic markers that regulate plant-microbe signal exchanges. Genome-wide genotyping for model C₄ plant Setaria italica using a microsatellite technique has been conducted (Pandey et al., 2013). Future work that links bacterial community composition with plant genetic markers is still needed.

2) The microbial seed bank hypothesis considers the spatial and temporal microbial community turnover as an outcome of relative abundance fluctuation of microbial taxa that are always present (Caporaso et al., 2012). This concept facilitates evaluating microbial community resilient to external environmental change. Deep sequencing approached confirmed this hypothesis in the marine system (Caporaso et al., 2012). Plant microbial ecosystems are less dynamic than marine microbial systems system. However, I observed highly similar bacteria taxa from geographically distant plants, which supported the microbial seed bank hypothesis to some extent. It would be interesting to use deep sequencing approaches to test 1) if the seed-bank hypothesis is valid in the plant-microbiome system; 2) whether the plant microbiome is resilient to environment changes that are critical to agricultural systems, so that once established, beneficial microbial communities are able to maintain their stability.
3) Little is known about the criteria non-legume plants use to recruit their endophytic associates. Bacteria taxa vary greatly in their plant growth promoting capability (Pankiewicz et al., 2015). By inhibiting bacterial genes that involved in plant growth promoting and plant-microbial signaling pathway, the establishment of bacterial endophytic colonization will be better understood.

**Final remarks**

My work suggests that local soil and plant genotype are critical and differentially determine the Miscanthus-associated bacterial communities. Local soil condition is more important in determine Miscanthus rhizosphere communities and less important in determining Miscanthus endophytic communities than plant genetic distance. For future work aimed at maximizing contribution of bacteria to plant nutrients, endophytic bacteria are better candidates than the rhizosphere colonists as Miscanthus endophytic members are more stable than the rhizosphere ones across habitats. Challenges remain, as little is known about the mechanism of endophytic bacteria recruitments. Closely-related Miscanthus genotypes showed distinct preferences in bacterial recruitment, suggesting the possibility of identifying key plant genetic markers that regulate beneficial bacteria association. Finally, I found that the presence of similar bacteria taxa does not equal similar bacterial interspecies interactions, making it difficult to predict microbial community assemblages and functions.

**References**


APPENDIX A: SUPPLEMENTARY INFORMATION FOR CHAPTER 2

Table A.1. (Modified based on Table 7, Maughan et al. (2012)): The average dry biomass yield (Mg ha\(^{-1}\)) of *M. × giganteus* across different nitrogen fertilization rate at each field site. Yields from IL are not presented here due to the large percentage of replanting at 2009.

<table>
<thead>
<tr>
<th>Average dry biomass (Mg ha(^{-1}))</th>
<th>2009</th>
<th></th>
<th></th>
<th>2010</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N rate (Kg N ha(^{-1}))</td>
<td></td>
<td>KY</td>
<td>NE</td>
<td>NJ</td>
<td>KY</td>
<td>NE</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>16.5</td>
<td>15.7</td>
<td>15.2</td>
<td>18.2</td>
<td>26.8</td>
</tr>
<tr>
<td>60</td>
<td></td>
<td>17.6</td>
<td>15.9</td>
<td>17.9</td>
<td>19.4</td>
<td>28</td>
</tr>
<tr>
<td>120</td>
<td></td>
<td>17.1</td>
<td>15.2</td>
<td>17.6</td>
<td>19.5</td>
<td>27.7</td>
</tr>
</tbody>
</table>

Table A.2. Percentage of diazotroph community variance explained by nitrogen fertilization treatment determined by canonical correspondence analysis.

<table>
<thead>
<tr>
<th>Percentage of variance explained by nitrogen treatment</th>
<th>Compartment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>Endosphere</td>
</tr>
<tr>
<td>IL</td>
<td>0.08842</td>
</tr>
<tr>
<td>KY</td>
<td>0.10314</td>
</tr>
<tr>
<td>NE</td>
<td>0.09496</td>
</tr>
<tr>
<td>NJ</td>
<td>0.11828</td>
</tr>
</tbody>
</table>

Significance code: ** p < 0.01, * 0.01 ≤ p < 0.05

Reference

Table B.1. Compartment-enriched OTUs, the compartment and plant species they enriched in and their taxonomic information

(Attached file: APPENDIX B-Table B.1.xlsx)
APPENDIX C: SUPPLEMENTARY INFORMATION FOR CHAPTER 4

Table C.1. Locations of the naturalized *Miscanthus sinensis* sampling sites (modified based on Table 1, Dougherty et al. (2014)).

<table>
<thead>
<tr>
<th>Sampling site</th>
<th>State</th>
<th>Site ID</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quaker Hill</td>
<td>CT</td>
<td>CT-01</td>
<td>41.397477</td>
<td>-72.114556</td>
</tr>
<tr>
<td>Cape Cod</td>
<td>MA</td>
<td>MA-01</td>
<td>41.581563</td>
<td>-70.525619</td>
</tr>
<tr>
<td>Monkton</td>
<td>MD</td>
<td>MD-02</td>
<td>39.59974</td>
<td>-76.604694</td>
</tr>
<tr>
<td>Fallston</td>
<td>MD</td>
<td>MD-03</td>
<td>39.508233</td>
<td>-76.382042</td>
</tr>
<tr>
<td>Henderson</td>
<td>NC</td>
<td>NC-01</td>
<td>35.270553</td>
<td>-82.412192</td>
</tr>
<tr>
<td>Sam's Gap</td>
<td>NC</td>
<td>NC-03</td>
<td>35.954031</td>
<td>-82.564233</td>
</tr>
<tr>
<td>Baptist Church</td>
<td>NC</td>
<td>NC-04</td>
<td>35.940247</td>
<td>-82.558189</td>
</tr>
<tr>
<td>Bernardsville</td>
<td>NJ</td>
<td>NJ-01</td>
<td>40.732999</td>
<td>-74.576325</td>
</tr>
<tr>
<td>Heckscher State Park</td>
<td>NY</td>
<td>NY-01</td>
<td>40.709308</td>
<td>-73.148865</td>
</tr>
<tr>
<td>Seatuck NWR</td>
<td>NY</td>
<td>NY-02</td>
<td>40.716871</td>
<td>-73.208331</td>
</tr>
<tr>
<td>Fort Washington</td>
<td>PA</td>
<td>PA-01</td>
<td>40.119787</td>
<td>-75.223539</td>
</tr>
<tr>
<td>Quakertown</td>
<td>PA</td>
<td>PA-02</td>
<td>40.415249</td>
<td>-75.313303</td>
</tr>
<tr>
<td>Green Hill</td>
<td>RI</td>
<td>RI-01</td>
<td>41.373201</td>
<td>-71.593845</td>
</tr>
<tr>
<td>Amherst</td>
<td>VA</td>
<td>VA-01</td>
<td>36.745903</td>
<td>-79.740539</td>
</tr>
<tr>
<td>Martinsville</td>
<td>VA</td>
<td>VA-02</td>
<td>36.748914</td>
<td>-79.733953</td>
</tr>
</tbody>
</table>

Reference

Table C.2. Number of detected compartment-enriched OTUs and their relative abundance in their enriched compartment.

<table>
<thead>
<tr>
<th>Habitat type</th>
<th>EC-enriched Counts</th>
<th>Relative abundance</th>
<th>R-enriched Counts</th>
<th>Relative abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native</td>
<td>45</td>
<td>53.8%</td>
<td>326</td>
<td>49.3%</td>
</tr>
<tr>
<td>Naturalized</td>
<td>54</td>
<td>52.7%</td>
<td>133</td>
<td>46.7%</td>
</tr>
</tbody>
</table>

Figure C.1. Species accumulation curves of calculated for the endophytic samples (a) and rhizosphere samples (b) in native (blue) and naturalized (green) *M. sinensis* habitats.
Figure C.2. OTU co-occurrence patterns of bacterial communities resided in the native endophytic (a), native rhizosphere (b), naturalized endophytic(c) and naturalized rhizosphere (d) compartments of *M. sinensis*. Each node represents a bacterial OTU and each edge represents a moderately high pairwise correlation (r > 0.6) between OTUs. The nodes in red represent the shared compartment-enriched OTUs.
APPENDIX D: SUPPLEMENTARY INFORMATION FOR CHAPTER 5

Table D.1. The mean of relative abundances of detected genera at each compartment, genotype and time harvested combination

(Attached file: APPENDIX D-Table D.1.xlsx)

Table D.2. Taxonomic information of OTUs that significantly enriched in the endophytic compartment of a certain genotype after 20 weeks cultivation

(Attached file: APPENDIX D-Table D.2.xlsx)

Table D.3. Number of OTUs that significantly enriched in certain genotype at each harvest time and compartment

<table>
<thead>
<tr>
<th>Compartment</th>
<th>Time Harvested</th>
<th>Diploid M. sacchariflorus</th>
<th>Diploid M. sinensis</th>
<th>Triploid M. × giganteus</th>
<th>Tetraploid M. sacchariflorus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endophytic</td>
<td>Week 0</td>
<td>35</td>
<td>12</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Rhizosphere</td>
<td>Week 0</td>
<td>75</td>
<td>96</td>
<td>87</td>
<td>72</td>
</tr>
<tr>
<td>Endophytic</td>
<td>Week 20</td>
<td>8</td>
<td>31</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Rhizosphere</td>
<td>Week 20</td>
<td>103</td>
<td>28</td>
<td>63</td>
<td>10</td>
</tr>
</tbody>
</table>
Figure D.1. Correspondence analysis biplots based on the \textit{nifH} T-RFLP community fingerprinting from a previous study. Triploid \textit{Miscanthus \times giganteus} rhizosphere soil samples were collected from \textit{Miscanthus} farms located in Dixon Springs (37.45, -88.67) and Urbana (40.04, -88.24) in 2009. The result showed rhizosphere diazotrophs that associated \textit{M. \times giganteus} differed significantly in Urbana and Dixon Springs (PerMANOVA $R^2 = 0.406$, $p = 0.001$). Each point represents one rhizosphere diazotrophs assemblage and the color of points represent sampling location.
Figure D.2. Correspondence analysis biplots based on the *nifH* T-RFLP community fingerprinting data showing rhizosphere diazotrophs that associated with diploid *M. sinensis* (A), diploid *M. sacchariflorus* (B), triploid *M. ×giganteus* (C) and tetraploid *M. sacchariflorus* (D) changed over time. Each point represents one rhizosphere diazotrophs assemblage cultivated in Dixon Spring farm soil, and the symbols of points represent the time harvest and whether this sample is from rhizosphere or bulk soil.
Figure D.3. Correspondence analysis biplots based on the *nifH* T-RFLP community fingerprinting data showing rhizosphere diazotrophs that associated with diploid *M. sinensis* (A), diploid *M. sacchariflorus* (B), triploid *M. ×giganteus* (C) and tetraploid *M. sacchariflorus* (D) changed over time. Each point represents one rhizosphere diazotrophs assemblage cultivated in Urbana farm soil, and the symbols of points represent the time harvest and whether this sample is from rhizosphere or bulk soil.
Figure D.4. Compartment based rarefaction curves. Points and curves are color-coded by compartment.
Figure D.5. Genotype based rarefaction curves of Week 0 endophytic communities (A) and rhizosphere communities (B), and Week 20 endophytic (C) and rhizosphere (D) communities. Points and curves are color-coded by Miscanthus genotype.
Figure D.6. Phylogenetic tree of the diazotrophic OTU12752 recovered from Miscanthus rhizome tissues. The tree is built based on the MUSCLE alignment (Edgar, 2004) results of nifH gene sequences that can be amplified using PoLF and PoLR primers (Poly et al.). DNA, and the tree is generated by fasttree algorithm (Price et al., 2009) in QIIME (Caporaso et al., 2010). The numbers represent the percent of confidence in generating that branch split.

References


