PROFILING EUKARYOTIC AND BACTERIAL COMMUNITIES OF HIGH- AND LOW-PRODUCTIVITY BLUEBERRY FARM AND FOREST SOILS

By

SEDA MIRZOYAN

A thesis submitted to the

School of Graduate Studies
Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science
Graduate Program in Microbial Biology

Written under the direction of

Lee J. Kerkhof

And approved by

_________________________________

_________________________________

_________________________________

_________________________________

New Brunswick, New Jersey

May 2020
ABSTRACT OF THE THESIS

Profiling Bacterial and Eukaryotic Communities in High- and Low-Productivity Blueberry Farm and Forest Soils

by SEDA MIRZOYAN

Thesis Director:
Lee J. Kerkhof

Highbush blueberry (*Vaccinium corymbosum*) is native to North America, commercially cultivated, and grows in acidic soils with high organic content. We attempted to characterize the core microbiome of two different highbush blueberry farm soils compared to forest soils. Stable isotope probing (SIP) was employed on two low-productivity, two high-productivity blueberry farm, and two nearby forest soils using $^{13}$C-$^{15}$N-Bioexpress (amino acids mixture) to detect the differences between the resident and amino acid-active soil microbial communities. The SIP experiments were coupled with bacterial and eukaryotic near full-length ribosomal profiling using Oxford Nanopore MinION sequencing technology. Bacterial rRNA operons were screened against the EzBioCloud database to characterize the resident and active rhizosphere bacterial communities, while eukaryotic rRNA operon sequences were screened against the UNITE/all eukaryotes database. Over 13,000 distinct taxonomic units were detected including a number of very abundant ones and thousands of rare members of the bacterial and eukaryotic communities inhabiting these soil systems. The results demonstrated that multiple *Bacillus* species were abundant in the amino acid-active communities of low-productivity soils, while high-productivity soils were dominated by amino acid-active
Paraburkholderia species. Profiling the eukaryotic members of these communities indicated that mycorrhizal fungi belonging to the Glomeromycota phylum were abundant in high-productivity soils, while low-productivity soils were enriched with fungi-like organisms of the phyla Rozellomycota and Oomycota (found in the amino acid-active community fraction only). Secondly, we tried to understand whether there was hidden diversity within a subset of the most abundant fungal and bacterial taxonomic units detected in this study. High-quality and unambiguous ribosomal RNA operons of several abundant fungal and bacteria were reconstructed from the best BLAST hits by generating long-read consensus sequences (LRC). The phylogenetic analysis of the fungal LRCs was performed based on the 18S and 28S rRNA genes, while bacterial LRCs were compared based on the 16S rRNA gene. The data demonstrate that most taxonomic assignments by BLAST were accurate, but new bacterial species and/or strains were shown to potentially exist in these soil systems. Novel clades of fungi (Glomeraceae- and Rozellomycota-related) were detected as well. In conclusion, SIP and MinION sequencing elucidated the resident and active microbial communities, as well as demonstrated the differences between high- and low-productivity farm soils. In the future, this approach can be applied using other heavy isotope-containing substrates for SIP to better characterize the active members of these communities that may be crucial in shaping blueberry plant health and productivity. In addition, this approach coupled with sequencing whole genomes should provide more useful information about the individual phylogenies of the detected organisms as well as their metabolic and functional capabilities that are responsible for unique plant-microbe interactions.
ACKNOWLEDGMENTS

I would like to thank my thesis advisor, Dr. Lee J. Kerkhof, for his constant support and guidance.
# TABLE OF CONTENTS

ABSTRACT OF THE THESIS .................................................................................................................. ii
ACKNOWLEDGMENTS ...................................................................................................................... iv
LIST OF TABLES .................................................................................................................................... vii
LIST OF ILLUSTRATIONS .................................................................................................................. viii
CHAPTER ONE: THE RHIZOSPHERE MICROBIOME ................................................................. 1
   Stable isotope probing ...................................................................................................................... 5
   Profiling ribosomal RNA operons for microbiome studies ...................................................... 10
REFERENCES ...................................................................................................................................... 14
CHAPTER TWO: DETECTION OF BOTH RESIDENT AND ACTIVE BACTERIA
   AND EUKARYOTES IN AGRICULTURAL AND FOREST SOILS VIA SIP AND
   MINION SEQUENCING ........................................................................................................... 19
   Introduction .................................................................................................................................. 19
   Methods and materials ................................................................................................................... 21
      Sample collection and stable isotope probing ........................................................................ 21
      DNA extraction .......................................................................................................................... 22
      PCR amplification of bacterial and eukaryotic ribosomal RNA operons .......................... 24
      Library preparation and MinION sequencing ...................................................................... 26
      Demultiplexing and BLAST analysis .................................................................................... 26
      Statistical analysis and data visualization ............................................................................. 27
   Results .......................................................................................................................................... 28
      Bacterial taxa ............................................................................................................................. 30
      Eukaryotic taxa ......................................................................................................................... 37
   Discussion ..................................................................................................................................... 38
REFERENCES ...................................................................................................................................... 44
CHAPTER THREE: DETECTING THE UNSEEN DIVERSITY WITHIN ABUNDANT
   TAXONOMIC UNITS ............................................................................................................. 49
   Introduction .................................................................................................................................. 49
   Methods and materials ................................................................................................................... 51
      Selection of highly similar ribosomal RNA operon sequences ....................................... 51
      Definition of sequence clusters ............................................................................................ 52
LIST OF TABLES

TABLE 1. Different $^{13}$C-, $^{15}$N-, and $^{18}$O- containing substrates used for stable isotope probing in microbial ecology experiments .................................................................9
FIGURE 1. Figure A provides the overview of the experimental design of the study as well as the schematic structure of bacterial and eukaryotic ribosomal RNA operons; Figure B describes stable isotope probing using heavy isotope-labeled substrates........23

FIGURE 2. NMDS (k=3) plots of A) bacterial and B) eukaryotic communities based on the species-abundance data showing community composition patterns in soil samples collected from two different blueberry farms (Farm 1 and Farm 2) and color-coded by soil type: Forest (adjacent forest soils); LP Soil (low-productivity soils); HP Soil (high-productivity soils); –resident indicates resident microbial communities derived from light ($^{12}$C/$^{14}$N) DNA bands of CsCl gradients, -AA active indicates amino acid-active (amino acid assimilating) microbial communities derived from heavy ($^{13}$C/$^{15}$N) DNA bands of CsCl gradients; light blue ovals incorporate all amino acid-active communities, orange ovals incorporate all resident communities.................................................................29

FIGURE 3. Bubble plots showing the relative abundances of ten most abundant bacterial taxonomic units in each of the samples (80 taxonomic units in total) color-coded by bacterial families in A) resident communities (amplicons from the light ($^{12}$C/$^{14}$N) DNA bands) marked with “-Resident”; B) amino acid-active (assimilating amino acids) communities (amplicons from the heavy ($^{13}$C/$^{15}$N) DNA bands) marked with “-AA-Active”. Three soil types are presented from two different farms in triplicates (1-6): Forest: adjacent forest soils, HP Soil: high-productivity soils, LP Soil: low-productivity soils ......................................................................................................................................32

FIGURE 4. Bubble plots showing the relative abundances of ten most abundant eukaryotic taxonomic units in each of the samples (58 taxonomic units in total) color-coded by eukaryotic phyla in A) resident communities (amplicons from the light ($^{12}$C/$^{14}$N) DNA bands) marked with “-Resident”; B) amino acid-active (assimilating amino acids) communities (amplicons from the heavy ($^{13}$C/$^{15}$N) DNA bands) marked with “-AA-Active”. Three soil types are presented from two different farms in triplicates (1-6): Forest: adjacent forest soils, HP Soil: high-productivity soils, LP Soil: low-productivity soils. The names of the eukaryotic taxonomic units from the UNITE database that were not classified at the species level were changed to include “_unclassified” to avoid confusions in common taxonomic nomenclature..................35

FIGURE 5. Schematic overview of the main steps of the procedure of the reconstruction of high-quality ribosomal RNA operons for different fungal and bacterial taxonomic units ..........................................................52

FIGURE 6. MUSCLE alignment heatmaps of unique sequences resulted from the first (A) and the ninth (B) internal BLAST searches for unclassified Mortierella-related sequences, and the total alignment heatmap (C) comprising sequences from all the observed high-similarity clusters (shown in black boxes) from the nine original alignments........................................................................................................................................55
FIGURE 7. A) The total MUSCLE alignment heatmap for *Paraburkholderia oxyphila*-related sequences after nine iterative BLAST searches against a database of 5,000 sequences, B) the total alignment heatmap after BLAST searches of nine groups of high-similarity sequences (shown in black boxes) detected in the first total alignment against a database of 105,000 sequences .................................................................56

FIGURE 8. The maximum-likelihood FastTree 18S rRNA partial gene phylogenetic tree (1351 positions) including fungal consensus sequences of ribosomal RNA operon fragments ........................................................................................................59

FIGURE 9. The maximum-likelihood FastTree 28S rRNA partial gene phylogenetic tree (2361 positions) including fungal consensus sequences of ribosomal RNA operon fragments ........................................................................................................60

FIGURE 10. The maximum-likelihood FastTree 16S rRNA partial gene (1297 positions) phylogenetic tree including *Bacillus niacini*-associated LRC sequences of ribosomal RNA operons ........................................................................................................61

FIGURE 11. The maximum-likelihood FastTree 16S rRNA partial gene (1297 positions) phylogenetic tree including *Paraburkholderia oxyphila*-associated LRC sequences of ribosomal RNA operons ........................................................................................................62
CHAPTER ONE: THE RHIZOSPHERE MICROBIOME

The rhizosphere microbiome is a community of diverse microorganisms physically associated with plant roots that directly or indirectly interact with the host plant and contribute to its health. The sum of microbial genomes in the rhizosphere substantially outnumbers that of the host plant and is considered to be its second genome (Berendsen et al., 2012). The rhizosphere microbiome can contain both beneficial and pathogenic microorganisms for the plant, and some microbes may be commensal. Moreover, the microbial composition of the rhizosphere, containing a larger number of organisms than the bulk soil, is unique to each plant species and genotype (Bakker et al., 2013). Organic carbon that is largely available in close proximity to plant roots enables the formation of distinct communities of microorganisms with a variety of metabolic capabilities (Berendsen et al., 2012; Bakker et al., 2013). Plant roots usually secrete various phytochemicals, which select for particular microorganisms depending on the plant genotype, bulk soil properties, and the availability of specific nutrients such as iron, nitrogen, and phosphorus. This phenomenon is described as rhizosphere priming (Pii et al., 2016; Dijkstra et al., 2013). Common root exudates include different organic acids such as citric acid and malic acid, sugars, phenolic compounds, and even benzoxazinoids that are known to inhibit the growth of some rhizosphere bacteria (Berendsen et al., 2012). Plants may also selectively produce chemicals that stimulate or suppress the production of signal molecules, mediating quorum sensing processes among rhizosphere microorganisms (Berendsen et al., 2012). Metatranscriptomic approaches by identifying the 16S and 18S rRNA fragments in the total microbial RNA pool demonstrate that the rhizosphere microbiomes of distinct plant species can also be very different at the kingdom level, which
suggests the existence of complex interactions between various organisms including bacteria, fungi, archaea, nematodes, ciliates and others inhabiting similar ecological habitats (Turner et al, 2013). Although the chemical and physical characteristics of the soil can influence the composition of the rhizosphere microbiome, the plant (both genotype and particular species) plays a more important role in defining the diversity and richness of root-associated microorganisms. For instance, different breeds of maize have been shown to host different rhizosphere microbiomes when grown in the same agricultural field. In contrast, different fields (with different soils) did not significantly contribute to the microbial community composition of the plant rhizosphere (Peiffer et al. 2013). Furthermore, plant developmental stages have been shown to affect the rhizosphere microbiome of cotton (Qiao et al., 2017). This is also the case for Arabidopsis thaliana; bacterial phyla including Acidobacteria, Actinobacteria, and Bacteroides had different abundances in the rhizosphere at the seedling versus other developmental stages of the plant (Chaparro et al., 2014). It has also been suggested that plant domestication has affected the richness of the plant rhizosphere microbiome since wild plant species are more successful in generating associations with potentially beneficial microorganisms than their cultivated counterparts (Pérez-Jaramillo et al. 2016).

In addition, the transition from living in the wild to domestication has changed the root architecture of the common bean, which has resulted in lower abundances of Bacteroidetes species and increased abundances of Actinobacteria and Proteobacteria in cultivated forms of the plant (Pérez-Jaramillo et al. 2017). Likewise, the rhizosphere of the wild beet, Beta vulgaris, has lost almost 40% of the abundant OTUs upon domestication, including bacteria resistant to abiotic stress (e.g. osmolarity and reactive oxygen species
stress) and has become more enriched with bacteria exhibiting phytopathogenic activities (Zachow et al., 2014). Although microbial communities inhabiting the plant rhizosphere can affect many aspects of the plant life (i.e. nutrition, productivity, and immunity against diseases), relatively few members of these communities are well-characterized. Thus, the mechanisms by which these organisms influence a given host plant remain unclear (Mendes et al., 2013). An interesting example of bacterial-fungal interactions that affects plant immunity is the increased abundances of the bacterial families Oxalobacteraceae, Burkholderiaceae, Sphingobacteriaceae, and Sphingomonadaceae, observed in the sugar beet rhizosphere upon the invasion of a potentially pathogenic fungus *Rhizoctonia solani*. Moreover, stress related genes were upregulated in these bacteria (Chapelle et al., 2016).

Knowing how rhizosphere microbiome perturbations affect the plant health and recognizing their causative mechanisms are vital for understanding the plant disease state and how it can be restored.

Another example of potential benefits by rhizosphere microbes is observed in “disease suppressive soils”, a term used to describe soils able to improve plant immunity against an infectious disease by controlling the growth and/or negative effects of plant-pathogenic microorganism(s) (Bakker et. al, 2013). This phenomenon is thought to result from microbial community members competing with a pathogen or suppressing its growth by other mechanisms. For instance, some pseudomonads are thought to contribute to fungal disease suppression by producing a putative chlorinated lipopeptide (DeSantis et al., 2011). In addition to competing for common sources of food, these “beneficial” bacteria may also produce various antibiotics, exotoxins, and bacteriocins. Some of these substances have been shown to promote plant health/productivity and are of great interest
as biocontrol agents for agriculture. Specifically, the bacteriocin Bac IH7 was found to be beneficial for tomato and musk melon plants (Subramanian and Smith, 2015). Moreover, other bacterial species may act as protective community members against bacterial pathogens; a flavobacterium strain was isolated from the roots of a disease resistant tomato variety and was shown to protect the plant against a soil-borne pathogenic bacterium, *Ralstonia solanacearum* (Kwak et. al, 2018). From this perspective, microbiome engineering aims to promote beneficial plant-microbe interactions with the goal of improving different aspects of plant health through manipulations in the rhizosphere microbiome (Quiza et al., 2015). *Pseudomonas* species, for example, have been observed to reduce the influence of plant pathogens when introduced to the rhizosphere of the tomato plant (Hu et al., 2016). The banana rhizosphere, when manipulated with a potentially protective *Bacillus* species, has substantially lowered the incidence of Panama disease by suppressing an infectious agent, *Fusarium oxysporum* (Xue et al., 2015). Interestingly, the abundance of other bacterial genera such as *Sphingobium*, *Dyadobacter*, and *Cryptococcus* was also found to be positively correlated with the low incidence of Panama disease in bananas (Fu et al., 2017). In addition, another study compared the community composition of the rhizosphere microbiomes of several common bean varieties and found that Pseudomonadaceae, Bacillaceae, Solibacteraceae, and Cytophagaceae were enriched in the roots of the plant varieties resistant to *Fusarium oxysporum*. The study also demonstrated that genes involved in protein secretion and the production of antifungal substances (including phenazines and rhamnolipids) exhibited increased abundances in root microbiomes of the pathogen-resistant plants (Mendes et al., 2018). Abiotic factors such as heat have also been shown to cause shifts in the rhizosphere microbiome structure, which
can affect the plant’s resistance against the fungal pathogen, *Rhizoctonia solani* (Voort et al., 2016).

Given the complex interactions between various microorganisms within the plant rhizosphere, it has been suggested that communities of beneficial bacteria and fungi might be increased by taking under consideration various growth promoting factors (Yergeau et al., 2015). Thus, a systematic approach needs to be employed to manipulate root-associated communities in agricultural fields (Wallenstein, 2017). For example, finding potential carbon and nitrogen-rich food sources that beneficial microorganisms may utilize in the plant rhizosphere could be one avenue in promoting their growth in agricultural settings to foster plant health and productivity.

**Stable isotope probing**

Stable isotope probing (SIP) is widely used in microbial ecology as a method to detect active microorganisms in complex environmental samples based on their ability to metabolize specific substrates (Dumont and Morrell, 2005). In the original SIP study (Radajewski et al., 2000), heavy $^{13}$C-labeled methanol was metabolized by different microorganisms and became incorporated into cellular biomolecules, such as nucleic acids. After labeling, cesium chloride (CsCl) centrifugation allowed the separation of $^{12}$C- and $^{13}$C-containing DNA molecules, which differentiates the microbial community members that are present in the sample from the actively metabolizing and growing ones. Additionally, this approach enabled the identification of methylotrophic microorganisms in forest soils and the rice rhizosphere (Dumont and Morrell, 2005; Lueders et. al., 2004). Often, cultivation-based methods have been insufficient for studying complex microbial communities. Other methods, such as stable isotope probing, can be a good alternative.
since it is a culture-independent technique and can be used along with other nucleic acid-based approaches such as amplicon analysis and novel metagenomic techniques to discover active microbes utilizing a particular substrate (Lu and Conrad, 2005; Dumont and Morrell, 2005; Prosser et al., 2006). The rhizosphere microbiome of rice has for instance been investigated via SIP by using $^{13}$C-labeled CO$_2$ to track methanogenic microorganisms; as a result, an uncultured group of archaea was detected (Lu and Conrad, 2005). In addition, Lu et. al (2007) showed that eukaryotes and gram-negative bacteria were mainly involved in the biochemical transformation of root-derived carbon, while gram-positive bacteria metabolized different carbon sources in the bulk soil instead. Methanotrophic microorganisms were detected by SIP as well from samples of a methane-rich cave and the genetic capacity for methanotrophy was revealed in those organisms including some *Methyldomonas* and *Methylococcus* species (Hutchens et al., 2004). SIP was also successfully used to identify bacteria able to metabolize biphenyl compounds in the horseradish rhizosphere growing in PCB-contaminated soils (Uhlik et al., 2009). Moreover, several methylotrophic bacterial genera including *Methylocella* and *Methylocystis* were detected via SIP in acidic environments such as forest soil systems (Radajewski et al., 2002). In a different study, the use of $^{13}$C-labeled acetate enabled the identification of active bacteria in permafrost soils and showed that nearly 80% of the members of resident bacterial communities were in fact actively growing (Tuorto et al., 2014). Likewise, Gadkari et al. (2020) was successful in demonstrating temperature optima range differences in the active bacteria found in Arctic soils by using $^{13}$C-cellobiose as a substrate for SIP. Furthermore, $^{13}$C-labeled inorganic and organic carbon sources allowed determining the heterotrophic mode of carbon acquisition by Crenarchaeota in salt marshes
(Seyler et al., 2014). In addition to heterotrophy, autotrophy and mixotrophy were also observed in marine archaea in deep sea samples by using a combination of inorganic and organic SIP substrates, in this case $^{13}$C-labeled acetate, urea, and bicarbonate (Seyler et al., 2018).

Furthermore, SIP studies were helpful in discerning unique interactions in rhizosphere microbiomes in terms of carbon flow and indicate that some rhizosphere bacteria can obtain carbon from mycorrhizal fungi attached to plants roots in addition to the carbon received from the roots. Mycorrhizal fungi, in contrast, usually obtain their carbon from the host plant only (Ballhausen and de Boer, 2016). Carbon flow in rhizosphere systems indeed plays an important role in plant health, but many questions remain unanswered about the intricate structure of complex food webs existing near plant roots. Techniques using stable isotopes have been extensively used to reveal these unique relationships between various microorganisms and plant roots and identify carbon transformation and stabilization pathways. The heavy $^{13}$C isotope allows tracking carbon flow in the plant rhizosphere, while $^{15}$N may also be used to reveal nutrient cycles. The use of $^{18}$O-labeled water is also very effective in detecting actively growing microorganisms. The presence of plant-derived $^{13}$C in proteins and enzymes related to the assimilation of carbohydrates and several amino acids suggests that root exudates contain biomolecules, which are being utilized by rhizosphere microorganisms (Pett-Ridge and Firestone, 2017). Moreover, a DNA-SIP study with labeled CO$_2$ elucidated that microbial communities usually actively respond to plant-derived carbon availability upon land abandonment from agriculture and bacteria are the main organisms metabolizing carbon in recently abandoned lands. Fungi, on the other hand, play a larger role in rhizosphere carbon flow at later stages.
of land abandonment. In addition, there is a community composition shift from beneficial, mycorrhiza-forming fungi to potentially pathogenic saprotrophic fungal species as land abandonment progresses in time (Hannula et al., 2017). SIP-based techniques can also detect food webs existing among other organisms such as flagellates and ciliates in addition to bacteria and fungi; it was shown that beneficial root-associated fungi act as a carbon reservoir for many bacteria inhabiting the plant rhizosphere (Nanninghaus et al., 2019). Distinct bacterial species have been detected in the rhizosphere of oilseed rape via SIP; it was shown that the plant roots and rhizosphere soil environments harbored different communities of abundant bacterial groups (Gkarmiri et al., 2017). In addition, stable isotope probing can be helpful in studying the expression of active genes when used in combination with transcriptomics (Achouak and Haichar, 2018). For instance, SIP detected important enzymes and other macromolecules that are being actively produced during various metabolic processes taking place in the plant rhizosphere when proteomics and metabolomics techniques are employed as well (Haichar et al., 2016). Stable isotope probing of DNA and mRNA molecules in the rhizosphere of Arabidopsis thaliana found highly expressed non-coding RNAs and genes responsible for the production of the enzymes 1-aminocyclopropane-1-carboxylate deaminase and nitrous oxide reductase suggesting the role of the corresponding microbiome-associated processes in the plant roots (Haichar et al., 2012). Overall, SIP is a useful tool for studying complex microbial communities like the plant rhizosphere especially when employed in combination with novel -omics approaches. The technique is able to provide profound insights into the structure of actively growing microorganisms that play crucial roles in nutrient cycles associated with plant roots and contribute to plant health and productivity.
Table 1 provides a summary of heavy isotope containing substrates used for various SIP studies.

Table 1. Different $^{13}$C-, $^{15}$N-, and $^{18}$O- containing substrates used for stable isotope probing in microbial ecology experiments.

<table>
<thead>
<tr>
<th>Examples of heavy isotope containing substrates used for SIP</th>
<th>System</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>$^{13}$C substrates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$^{13}$CH$_4$</td>
<td>Cave groundwater, mats</td>
<td>Hutchens, et al., 2004</td>
</tr>
<tr>
<td>$^{13}$C-CO$_2$</td>
<td>Soil, rice rhizosphere</td>
<td>Lu and Conrad, 2005; Lu et. al, 2007;</td>
</tr>
<tr>
<td>$^{13}$C-bicarbonate</td>
<td>Salt marsh sediments; deep sea samples</td>
<td>Seyler et. al, 2014; Seyler et al., 2018</td>
</tr>
<tr>
<td>$^{13}$C-urea</td>
<td>Salt marsh sediments; deep sea samples</td>
<td>Seyler et. al, 2014; Seyler et al., 2018</td>
</tr>
<tr>
<td>$^{13}$C-hexadecane</td>
<td>Oilfield methanogenic consortium</td>
<td>Cheng et al., 2013</td>
</tr>
<tr>
<td>$^{13}$C-biphenyl</td>
<td>Soil, horseradish rhizosphere,</td>
<td>Uhlik et al., 2009</td>
</tr>
<tr>
<td>$^{13}$C-toluene</td>
<td>Pond sludge; contaminated aquifer water</td>
<td>Jehmlich et al., 2008; Geyer et. al, 2005</td>
</tr>
<tr>
<td>$^{13}$C-benzene</td>
<td>Pond sludge; contaminated aquifer water</td>
<td>Jehmlich et al., 2008; Geyer et. al, 2005</td>
</tr>
<tr>
<td>$^{13}$C-labeled amino acids</td>
<td>Coastal marine bacterioplankton; marine sediments; salt marsh sediments</td>
<td>Bryson et al., 2016; MacGregor et al. et al, 2006; Seyler et al. 2014</td>
</tr>
<tr>
<td>$^{13}$C-xylose</td>
<td>Agricultural soil</td>
<td>Pepe-Ranney et al., 2016</td>
</tr>
<tr>
<td>$^{13}$C-cellobiose</td>
<td>Tundra soil</td>
<td>Gadkari et. al., 2020</td>
</tr>
<tr>
<td>$^{13}$C-cellulose</td>
<td>Agricultural soil; forest soil; insect gut microbiome</td>
<td>Pepe-Ranney et al., 2016; Bastias et. al, 2009; Alonso-Pernas et. al, 2017</td>
</tr>
<tr>
<td>$^{13}$C-glucose</td>
<td>Bacterial cultures; marine sediments</td>
<td>Lerch et al., 2011; MacGregor et al. et al, 2006</td>
</tr>
<tr>
<td>$^{13}$C-acetate</td>
<td>Marine sediments; bacterial cultures; activated sludge; permafrost soils; deep sea samples</td>
<td>MacGregor et al. et al, 2006; Auclair et al. et al, 2012; Osaka et al., 2006; Tuerto et al., 2014; Seyler et al., 2018</td>
</tr>
<tr>
<td>$^{13}$C-propionate</td>
<td>marine sediments</td>
<td>MacGregor et al. et al, 2006</td>
</tr>
<tr>
<td>$^{13}$C-methanol</td>
<td>Activated sludge; forest soil; agricultural soil</td>
<td>Osaka et al., 2006; Radajewski et al., 2000; Lueders et. al., 2004</td>
</tr>
<tr>
<td><strong>$^{15}$N-substrates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$^{15}$N$_2$</td>
<td>Pulp and paper mill effluent</td>
<td>Addison et. al. 2010</td>
</tr>
<tr>
<td>$^{15}$N-urea</td>
<td>Insect gut microbiome</td>
<td>Alonso-Pernas et. al. 2017</td>
</tr>
<tr>
<td>$^{15}$N-nitrate</td>
<td>Soil</td>
<td>Reay et. al. 2019</td>
</tr>
<tr>
<td>$^{15}$N-ammonium</td>
<td>Pond sludge; soil</td>
<td>Jehnlich et al., 2008; Reay et. al, 2019</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>$^{18}$O substrates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$^{18}$O-H$_2$O</td>
<td>Soil</td>
<td>Blazewicz and Schwartz, 2011</td>
</tr>
</tbody>
</table>

**Profiling ribosomal RNA operons for microbiome studies**

16S rRNA gene sequencing is widely used to characterize bacterial members of environmental microbial communities, while fungal members of these communities are usually detected via the sequencing of internal transcribed spacer (ITS) regions of fungal ribosomal RNA operons. ITS regions are the most accurate biomarkers for the identification of fungal species belonging to district taxonomic groups and can detect variance with a single species. In contrast, large subunit (LSU) genes have been shown to be helpful in detecting early fungal evolutionary linages, while the small subunit (SSU) is less efficient in species-level fungal taxonomic assignments (Schoch et al., 2012). Therefore, ITS region sequencing has been useful in elucidating fungal communities in various samples, such as dairy products where nineteen distinct fungal genera were detected (Buehler et al., 2017). However, *in silico* studies have shown that ITS sequencing alone can introduce biases in fungal species identification, via mechanisms such as ITS1F and ITS2 primer mismatches. It has been recommended that sequencing of eukaryotic ITS regions containing parts of the 18S and 28S rRNA genes lessens bias and can be more accurate (Tedersoo, 2016). Fungal ITS2 regions were also found to form chimeric structures when amplified, which is also a major drawback for proper species-level assignment of fungi (Aas and Kauserud, 2016). Thus, amplifying and profiling full-length ribosomal RNA operons alleviates many of these issues and has a huge potential for
microbiome studies, even though most of the reference databases include partial ribosomal genes or single ITS regions generated by previous research only targeting those fragments.

The sequencing of PCR-amplified almost full-length ribosomal RNA operons including both small and large subunits and ITS regions has been shown to be more effective in discerning phylogenies of prokaryote organisms in various microbiomes than traditional short-read partial 16S rRNA gene sequencing techniques and provides species-level resolution for bacteria (Kerkhof et al., 2017; Cusco et al., 2018; Martijn et al., 2019). Quality control protocols of PacBio long-read ribosomal RNA operons usually result in up to 40,000 high-quality reads that may elucidate only some of the most abundant bacterial species in a given sample (Martijn et al., 2019). The MinION long-read sequencing technology, on the other hand, was shown to generate at least as many long reads of ribosomal RNA operons (~4200 bp) as the PacBio platform does and provides strain-level resolution for bacterial members of complex microbiomes (Kerkhof et al., 2017). The taxonomic assignment of high-quality MinION reads was proven to be as accurate as 99.5% despite of the relatively high sequencing error rate (Brown et al., 2018). A different study has conducted MinION sequencing of full-length 16S rRNA genes (~1900 bp) to investigate the dog skin microbiome, but since the ITS region was omitted, strain-level resolution was not achieved (Cusco et al., 2017). Although some bacterial species have several copies of ribosomal RNA operons in their genomes and these copies sometimes have variability between them affecting the proper investigation of prokaryote phylogenies, profiling full ribosomal RNA operons is still a more robust approach than partial 16S rRNA gene sequencing since full operons involve more variable regions
resulting in more accurate taxonomic assignments of prokaryotes (de Oliveira Martins et al., 2020).

**The aim of this thesis**

The main purpose of this thesis was to elucidate the differences in the microbial communities between high- and low-productivity agricultural farm soils, where the commercial highbush blueberry plant is grown. Our goal was to uncover both resident and active microbial communities metabolizing various amino acids in these systems in addition to the ones present in nearby forest soils as control sites. We attempted to profile both bacterial and eukaryotic members of blueberry farm soils with the ultimate goal of developing a reliable tool for discerning high- and low-productivity farms in the future. To provide high taxonomic resolution, we employed stable isotope probing and the MinION DNA sequencing technology to generate long reads of PCR-amplified almost full-length bacterial and eukaryotic ribosomal RNA operons. These specific soil systems have not been previously characterized in terms of their microbial diversity via long-read sequencing of ribosomal RNA operons, thus we aimed to establish their core microbial communities and potentially gain insight into unique interactions between the identified organisms. This information can help us better understand the diverse processes including nutrient cycling taking places in the blueberry rhizosphere microbiome.

Another question we tried to answer was whether there were undetected groups of microorganisms related to some of the abundant bacterial and fungal taxonomic units identified by this approach. Since we were able to generate almost full-length ribosomal RNA operons covering both highly conserved and variable genetic regions, it was possible to use multiple alignment techniques to reveal cryptic groups of microorganisms based on
the pairwise identities of the reads matching these groups. Phylogenetic analysis of the consensus sequences generated from highly similar ribosomal RNA operons enabled us to gain insight into the evolutionary relationships of these cryptic clusters of fungi and bacteria with their previously identified close relatives.
REFERENCES


CHAPTER TWO: DETECTION OF BOTH RESIDENT AND ACTIVE BACTERIA AND EUKARYOTES IN AGRICULTURAL AND FOREST SOILS VIA SIP AND MINION SEQUENCING

Introduction

Highbush blueberry (*Vaccinium corymbosum* L.) is a long-lived woody perennial native to North America and an important economic crop. The U.S. is the world’s largest producer of blueberries with over 89,000 acres harvested in 2018 and a production value of $797 million (NASS, 2018; Brannen et al., 2014). Additionally, there is renewed interest in blueberry agriculture due to their richness in natural antioxidants and other bioactive compounds (Prior et al., 1998). Highbush blueberry fields can remain productive for decades, but older fields tend to decline in plant health and productivity. Replanting often does not solve the problem, suggesting the issue is soil-associated (Brannen et al., 2014; Jagdale et al., 2013; Noe et al., 2014). This condition, termed replant disease, is common in woody perennial crops and is well studied in tree fruits (Waschkies et al., 1994; Mazolla et al., 2012).

The soil-associated issues may be related to such things as depletion of organic matter and/or changes in the soil microbial community. Plant roots harbor diverse microbial communities that are critical in plant health and productivity. Interestingly, plants not only shape their own rhizosphere microbiomes but can become enriched with protective bacteria and fungi when challenged by pathogenic fungi and insects (Berendsen et al., 2012). Recently, efforts have been made to promote beneficial microorganisms in plant root/soil systems. For example, agricultural soil management types (conventional vs. organic farming) and tillage intensities have been shown to influence bacterial and fungal communities associated with plant roots differently, suggesting that cropping practices can
be helpful in altering the root microbiomes in favor of beneficial microbes (Hartman et al., 2018). Furthermore, each plant species appears to have a unique composition of root-associated microbes (Turner et al., 2013). Unfortunately, many of these important indigenous microorganisms have yet-to-be cultivated in the laboratory and the unique mechanisms by which these organisms contribute to plant health remain largely unknown.

Despite advances in sequencing technology, the taxonomic identification of rhizosphere microorganisms and their placement into ecological guilds (i.e. determining their functional role in the microbial community) is challenging. The short-read sequencing approach usually targets a small portion of the 16S rRNA gene for bacteria or the intergenic transcribed spacer (ITS) region in the ribosomal RNA operon for eukaryotes. Often, only short amplified fragments (100-250 bp) are sequenced. These short sequences can be insufficient for accurate taxonomic placement to the species level. Additionally, this method is PCR-based and can lead to bias towards inaccurate relative quantitation due to differences in ribosomal RNA operon copy number. Metagenomic shotgun sequencing can be performed to sample all of the genes in all organisms in a complex sample, such as rhizosphere soil. However, this approach requires high read depth, is relatively costly, and can be low in sensitivity.

Regardless of the chosen sequencing approach, the rhizosphere soil can contain billions of cells from many different microorganisms. Some of these may affect plant health directly (e.g. pathogens and mycorrhizal fungi) or indirectly (e.g. denitrifying bacteria). Finally, there are those that fill an important ecological niche, such as saprophyles, but may be neutral with regard to their effect on plant growth. For example, the common soil fungus, *Fusarium oxysporum*, exists in the soil as a saprophyte, but some
strains can cause *Fusarium* wilt or root rot in many plant species while other non-pathogenic strains can be used as biocontrol agents (Fravel et al., 2003).

To begin addressing some of these issues, we targeted ribosomal RNA operon amplification coupled with long-read sequencing. Oxford Nanopore MinION enables sequencing of long DNA fragments (Jain et al., 2015; Jain et al., 2018) and has been shown to provide species-level identification for bacteria by sequencing rRNA operons (Cusco et al., 2018; Benitez-Paez and Sanz, 2017; Kerkhof et al., 2017). Here, we attempt to reveal the active bacterial and/or fungal community members assimilating amino acids that may play critical roles in shaping overall plant health. We utilized stable isotope probing (SIP), coupled with rRNA operon profiling, to characterize both fungal and bacterial members of farm and forest soils. SIP is a technique used to characterize the active members of the microbial communities in various systems based on the uptake of $^{13}$C and/or $^{15}$N into newly synthesized DNA (Radajewski et al., 2000). The results demonstrate it is possible to identify the actively growing microorganisms by sequencing near full-length rRNA operons to better characterize important members of soil microbial communities. This approach may lay a foundation for developing a molecular tool for differentiating low- and high-productivity blueberry fields.

**Methods and materials**

**Sample collection and stable isotope probing**

Two types of field soils (exhibiting low- and high-productivity) were sampled from two commercial blueberry farms, along with two adjacent forest soils. All six soils were sampled in triplicate. The soils were collected from the rhizosphere of ‘Bluecrop’ blueberry plants in the cultivated fields. The forest soils were collected from the rhizosphere of wild
blueberries adjacent to the cultivated fields. Five hundred milligrams of the original soil samples were inoculated with either \( ^{12}\text{C}/^{14}\text{N} \) or \( ^{13}\text{C}/^{15}\text{N} \) 10X BioExpress Cell Growth Media (Cambridge Isotope Laboratories Inc., Tewksbury, MA; 100 \( \mu \text{L/sample} + 900 \mu \text{L} \) of sterile water to fully wet the soil). The samples were incubated for two days in lighted growth chamber (Percival Scientific, Perry, IA) with a 12-hour day/night cycle and fluorescent illumination \( (80 \mu \text{mol m}^{-2}\text{s}^{-1}) \) at 25 °C, after which the soil was collected and processed as detailed below.

**DNA extraction**

The total genomic DNA was extracted from the soil near the center of the tube using a modified CTAB/phenol-chloroform procedure. (Porebski et al, 1997). Specifically, 50 mg of soil was wetted with 50 \( \mu \text{L} \) of Solution 1 \( (50 \text{ mM glucose}, 10 \text{ mM EDTA}, 25\text{mM Tris-Cl (pH 8.0)}) \) and subjected to five freeze-thaw cycles by incubated in a -80°C freezer and in a 55°C hot water bath. Each sample was then incubated with 50 \( \mu \text{L} \) of lysozyme solution \( (4.0 \text{ mg/mL in Solution 1}) \) for 10 minutes at room temperature and 450 \( \mu \text{L} \) CTAB solution \( (100 \text{ mM Tris-Cl, 20 mM EDTA, 1.4 M NaCl, 4\% (w/v) CTAB, and 1\% }\beta\text{-mercaptoethanol}) \) was added. This aqueous mixture was extracted twice with 800 \( \mu \text{L} \) phenol:chloroform:isoamyl alcohol \( (24:25:1) \) to remove proteins/lipids. The aqueous phase was then ethanol precipitated with 20 \( \mu \text{g} \) of glycogen. DNA pellets were resuspended in 50 \( \mu \text{L} \) of 10 mM Tris-HCl and stored at -80°C.

The \( ^{12}\text{C}/^{14}\text{N} \)- and \( ^{13}\text{C}/^{15}\text{N} \)-labeled DNA extracts, purified as described above, were physically separated via cesium chloride (CsCl) gradient ultracentrifugation \( (225,000 \times g \text{ for 36 hours}) \) using an Optima MAX-TL ultracentrifuge with a TLA-110 rotor (Beckman Coulter Inc., Indianapolis, IN, U.S.A.). Prior to centrifugation, \( ^{12}\text{C} \)- and \( ^{13}\text{C} \)-labeled
archaeal carrier DNA from *Halobacterium salinarum* was added to the gradients to enhance the visualization of the $^{12}$C/$^{14}$N- and $^{13}$C/$^{15}$N-labeled genomic DNA (Tuorto et al., 2014). The light ($^{12}$C/$^{14}$N) and heavy ($^{13}$C/$^{15}$N) DNA bands were collected by pipette and CsCl was removed by micro-dialysis against 10 mM Tris-HCl. The workflow of laboratory experiments is summarized in Figure 1.
**Figure 1.** Figure A provides the overview of the experimental design of the study as well as the schematic structure of bacterial and eukaryotic ribosomal RNA operons; Figure B describes stable isotope probing using heavy isotope-labeled substrates.

**PCR amplification of bacterial and eukaryotic ribosomal RNA operons**

Near-full length ribosomal RNA operons of bacteria were amplified from $^{12}$C/$^{14}$N- and $^{13}$C/$^{15}$N-DNA bands collected from both experimental ($^{13}$C/$^{15}$N-treated) and control ($^{12}$C/$^{14}$N-treated) samples. The universal bacterial 16S forward 27F (5' AGAGTTTGATCCTGGCTCAG 3') and 23S reverse 2241R (5' ACCGCCCAAGTHAAACT 3') primers were used to amplify bacterial ribosomal RNA operon fragments. DNA extracts (1.5 µM) from light and heavy DNA bands were subjected to PCR amplification with bacterial and eukaryotic primers separately and a Hi-Fidelity Taq polymerase (Bimake LLC, Houston, TX, USA) as previously described in Kerkhof et al. (2017). Touchdown PCR involved initial denaturation at 95 °C for 4 min; 2 cycles of 95 °C for 20 s, 68 °C for 15 s, 72 °C for 75 s; 2 cycles of 66 °C for primer annealing; 2 cycles of 64 °C for primer annealing; 2 cycles of 62 °C for primer annealing - all with denaturation/extension; followed by 22 cycles of denaturation, 60 °C for 15 s, extension; and a final extension at 72 °C for 7 min.

At the end of the 16th cycle (8 touchdown + 8 standard cycles) of the touchdown PCR, 10 µl of the PCR amplification mixture was removed and stored at -80 °C. The bacterial PCR amplifications were continued until the end of 30 cycles. Agarose gel electrophoresis was used to visualize the PCR products and ensure that the heavy DNA bands of $^{12}$C/$^{14}$N-treated control samples did not show any PCR amplification (Suppl. Figure 1). The 16th cycle PCR products from $^{12}$C/$^{14}$N or $^{13}$C/$^{15}$N bands were then purified with AMPure magnetic beads (Beckman Coulter, Brea, CA, USA) as follows: 10 µL DNA
extract was added to 30 µL of 10 mM Tris-HCl, 15 µL of AMPure beads, 20 µL 5 M NaCl, and 20 µL 30% PEG/1.5 M NaCl and bound to beads for 15 min, later washed twice with 70% ethanol and resuspended in 1 µL of 10 mM Tris-HCl. The bead-purified PCR products were stored at -80°C.

The PCR products of bacterial rRNA operons were then barcoded via a second round of PCR with universal primers attached to the ONT (Oxford Nanopore Technologies Ltd.) barcode sequences for multiplex sequencing. The barcode primers for both bacterial and eukaryotic rRNA operons are provided in Suppl. Table 1. The barcode PCR program for bacterial operons included 5 minutes denaturation at 94 °C, followed by 2 cycles of 94°C for 20 s, 56°C for 15 s and 72°C for 75 s; 2 cycles of 94°C for 20 s, 60°C for 15 s and 72°C for 75 s; 2 cycles of 94°C for 20 s, 64°C for 15 s and 72°C for 75 s; 2 cycles of 94°C for 20 s, 66°C for 15 s and 72°C for 75 s; 2 cycles of 94°C for 20 s, 68°C for 15 s and 72°C for 75 s; 20 cycles of 94°C for 20 s and 72°C for 75 s, followed by a final extension cycle at 72°C for 7 min.

For eukaryotic rRNA operon amplifications, the primers Euk 18S F 5’ TACCTGGTTGATQCTGCCAGT 3’ and Euk 28S R 5’ TCTGACTTAGAGGCGTTTCAGT CATAAT 3’ with a similar touchdown PCR program was employed, which included initial denaturation at 95 °C for 3 min; 2 cycles of 95 °C for 20 s, 66 °C for 15 s, 72 °C / 75 seconds for extension; 2 cycles of 95 °C /20 s for denaturation, 64 °C /15 s for primer annealing, 72 °C / 75 s for extension, 2 cycles of 95 °C /20 s for denaturation, 62 °C /135 s for primer annealing, 72 °C / 75 s for extension followed by 22 cycles of denaturation, 60 °C /15 s for primer annealing, 72 °C / 135 s extension; and a final extension at 72 °C for 7 min.
A portion of the amplification was also collected at 16 cycles, bead cleaned, and re-amplified with barcoded primers as described above. The barcode PCR program for eukaryotic operons included 5 minutes denaturation at 94°C, followed by 2 cycles of 94°C for 30 s, 60°C for 15 s and 72°C for 75 s, 2 cycles of 94°C for 30 s, 64°C for 15 s and 72°C for 75 s, 2 cycles of 94°C for 30 s, 68°C for 15 s and 72°C for 75 s, 24 cycles of 94°C for 30 s and 72°C for 75 s, followed by a final extension cycle at 72°C for 7 min. Barcoded rRNA operon amplicons were analyzed by agarose gel electrophoresis so that their concentration could be determined.

**Library preparation and MinION sequencing**

MinION libraries were prepared using the Ligation Sequencing Kit (SQK-LSK109). Twelve samples (each sample having 150 ng of DNA (1800 ng in total)) were combined, end-repaired, dA-tailed as per manufacturer’s instructions (Oxford Nanopore Technologies Ltd.) instructions using NEB kits (New England Biolabs, Ipswich, MA, USA; supplemented with 1 µl of 100 mM dNTP’s) and ligated to ONT adaptors with the Blunt/TA ligase master mix (NEB) amended with 1 µL of freshly made ATP solution (~4 mg/mL) and 1 µL of 100 mM dNTP mix (Bimake LLC, Houston, TX, USA) to facilitate ligation. All libraries were analyzed on FLO-MIN106 R9.4.1 flow cells on run on MinION Mk1B devices.

**Demultiplexing and BLAST analysis**

All raw sequence reads were base-called with the Guppy 3.2.2 software (Oxford Nanopore Technologies Ltd.) via the High-Accuracy algorithm. Each library resulted in roughly 2-5 million raw reads. Demultiplexing was performed via a custom workflow using the Geneious R11 software (Biomatters Ltd.), which included: filtering the DNA
sequence reads by size (3700-4700 bp for bacterial operons; 4000-6300 bp for eukaryotic operons), truncating the assigned ONT name, annotating the reads with barcode primers (90% pairwise identity), extracting reads containing specific barcodes and grouping all reads for each barcode. An assessment of whether annotated sequences could be assigned to multiple barcode folders indicated a 0.375% “contamination” of mis-assigned reads per sample. This low percentage of mis-assigned reads were deemed too low to warrant extensive detection and clean-up. The barcode-categorized reads were exported as fasta files and screened against the EzBioCloud 16S rRNA gene database for bacteria and the UNITE/all eukaryotes ITS database (UNITE Community, 2019) for eukaryotes via the BLAST algorithm (version 2.9.0). MegaBLAST settings included a word size of 28 bp, a match/mismatch of 1/-2, and a linear gap penalty.

**Statistical analysis and data visualization**

The BLAST results were imported as hit tables to Microsoft Excel in CSV format. Pivot tables were generated based on count of species for all the soil samples originating from both $^{12}\text{C}/^{14}\text{N}$ (light) and $^{13}\text{C}/^{15}\text{N}$ (heavy) DNA bands and imported to R studio (Version 1.2.5033). Non-metric multidimensional scaling (NMDS) was performed on the species abundance data using the ‘vegan’ package (Oksanen et al., 2018). $k = 3$ number of dimensions were used for the analysis based on the NMDS stress plots and $R^2$ values for model fit. The $R^2$ value for non-metric fit of observed dissimilarities was 0.99 for the eukaryotic data and 0.997 for bacterial species-abundance data when $k = 3$. The NMDS stress plots and the ordination distance graphs are provided in Suppl. Figures 2 and 3.
Results

Farm and forest soils were amended with $^{12}\text{C}/^{14}\text{N}$ or $^{13}\text{C}/^{15}\text{N}$ growth media and the bacterial and eukaryotic communities were profiled by rRNA operon sequencing. Heavy DNA bands originating from $^{12}\text{C}/^{14}\text{N}$-treated control samples did not show any PCR amplification with neither bacterial nor eukaryotic rRNA primers indicating that there was no heavy-isotope incorporation into nucleic acids by the microbial communities in the control samples. In contrast, both light and heavy DNA bands isolated from $^{13}\text{C}/^{15}\text{N}$-treated experimental samples produced PCR amplicons of both bacterial and eukaryotic rRNA operons (Suppl. Fig. 1), which were then successfully sequenced by MinION sequencing. This approach yielded information on the overall microbial community ($^{12}\text{C}/^{14}\text{N}$-resident) and the growing component ($^{13}\text{C}/^{15}\text{N}$-active) which utilized $^{13}\text{C}/^{15}\text{N}$-amino acids for new DNA biosynthesis. Overall, the study produced 2,024,407 rRNA operon reads for bacteria, while the eukaryotic profiling produced 3,447,857 reads. The alignment length and % pairwise identity for bacterial and eukaryotic BLAST hits are presented in Suppl. Fig. 4. The bacterial rRNA operon pairwise identity ranged from 64-100% with an average of 89 ± 4% over a BLAST alignment length of 1452 ± 133 bp. The pairwise identity for eukaryotic rRNA operon BLAST alignments ranged from 70-100% and averaged at 88 ± 4% over an alignment length of 2008 ± 618 bp. Only 1.6% of the alignments with bacterial 16S rRNA genes were smaller than 1000 bp, possibly because the EzBioCloud database contains near-full length 16S rRNA sequences (~1500 bp). In contrast, the smallest reference sequence in the UNITE/all eukaryotes database is 93 bp; the percentage of the sequence alignments with query eukaryotic reads shorter than 93 bp was 0.13%.
In order to determine if there were overall differences between study sites or between the resident and amino acid-active members of the soil microbial communities, a non-metric multidimensional scaling (NMDS) analysis was performed. Figure 2 shows the NMDS plots for bacterial and eukaryotic communities. It can be observed that amino acid assimilating microbial communities are distinct from resident communities both for bacteria and eukaryotes. Moreover, the amino acid-active eukaryotic communities of all soil types form a more compact cluster, while the resident communities exhibit a larger spread among soil types. Interestingly, the resident bacterial communities overlap in high- and low-productivity farm soils but are different from the forest communities. In contrast, the two types of farm soils are more distinct in the amino-acid active fraction of the bacterial communities.
Figure 2. NMDS (k=3) plots of A) bacterial and B) eukaryotic communities based on the species-abundance data showing community composition patterns in soil samples collected from two different blueberry farms (Farm 1 and Farm 2) and color-coded by soil type: Forest (adjacent forest soils); LP Soil (low-productivity soils); HP Soil (high-productivity soils); –resident indicates resident microbial communities derived from light (\(^{13}C/^{14}N\)) DNA bands of CsCl gradients, -AA active indicates amino acid-active (amino acid assimilating) microbial communities derived from heavy (\(^{13}C/^{15}N\)) DNA bands of CsCl gradients; light blue ovals incorporate all amino acid-active communities, orange ovals incorporate all resident communities.

**Bacterial taxa**

Overall, the rRNA operon profiling detected 10,897 bacterial taxonomic units, belonging to 44 different bacterial phyla. Bacterial communities were abundant in
Proteobacteria and Actinobacteria with more Proteobacteriota found in the amino acid-active communities than in the resident ones except one forest soil sample. Acidobacteria were also identified in all resident communities with the high-productivity soils relatively more enriched with the AD3 clade (Suppl. F. 5). However, there was an increase in the number of actively growing Firmicutes species observed in low-productivity farm soils. These differences in community composition patterns were more easily observed when the relative abundances of the top ten eukaryotic and bacterial taxonomic units, grouped by dominant eukaryotic phyla and bacterial families, in each sample were visualized as bubble plots (58 eukaryotic and 80 bacteria taxonomic in total) (Figures 3 and 4). The resident bacterial communities were different in terms of the relative abundances of several taxonomic units from the FJ625337_f group and the family Gaiellaceae between the high-productivity soils from two agricultural sites. Low-productivity soils, on the other hand, were very similar to each other and exhibited higher abundances of DQ480487_s and AY913269_s (Rhodospirillaceae family). Amino acid-active bacteria in high-productivity soils belonged to the family Burkholderaceae; *Paraburkholderia oxyphyla* constituted more than 75% of the total reads. In contrast, low-productivity soils were enriched with amino acid-active Bacilli: *Bacillus cucumis, Bacillus fumaroli, Bacillus niacini, Bacillus soli, Bacillus novalis*, and DQ129292_s. *Bacillus* species did not show amino acid assimilation in forest soils, which exhibited abundant Streptomycetaceae species such as *Streptacidiphilus neutriniunicus, Streptomyces herbaricolor*, and *Streptacidiphilus melanogenes* instead. Moreover, the two forest soils were different from each other in terms of the abundances of active Streptomycetaceae and Burkholderaceae species.
Figure 3. Bubble plots showing the relative abundances of ten most abundant bacterial taxonomic units in each of the samples (80 taxonomic units in total) color-coded by bacterial families in A) resident communities (amplicons from the light ($^{12}\text{C}/^{14}\text{N}$) DNA bands) marked with “-Resident”; B) amino acid-active (assimilating amino acids) communities (amplicons from the heavy ($^{13}\text{C}/^{15}\text{N}$) DNA bands) marked with “-AA-Active”. Three soil types are presented from two different farms in triplicates (1-6): Forest: adjacent forest soils, HP Soil: high-productivity soils, LP Soil: low-productivity soils.
Figure 4. Bubble plots showing the relative abundances of ten most abundant eukaryotic taxonomic units in each of the samples (58 taxonomic units in total) color-coded by eukaryotic phyla in A)
resident communities (amplicons from the light \( {^{12}\text{C}/^{14}\text{N}} \) DNA bands) marked with “-Resident”; B) amino acid-active (assimilating amino acids) communities (amplicons from the heavy \( {^{13}\text{C}/^{15}\text{N}} \) DNA bands) marked with “-AA-Active”. Three soil types are presented from two different farms in triplicates (1-6): Forest: adjacent forest soils, HP Soil: high-productivity soils, LP Soil: low-productivity soils. The names of the eukaryotic taxonomic units from the UNITE database that were not classified at the species level were changed to include “_unclassified” to avoid confusions in common taxonomic nomenclature.

**Eukaryotic taxa**

Nineteen major clades of eukaryotic organisms were identified constituting 51 eukaryotic phyla and 2501 distinct taxonomic units. Fungi constituted an average of 76% of the total eukaryotic reads in the resident communities of all soil types, while they were the majority of microorganisms (98%) detected in the amino acid-active communities (Suppl. F. 6). Soil resident communities were also rich in various protist species belonging to Alveolata and contained, for instance, plant and animal reads as well (Suppl. F. 6). Suppl. Figure 7 shows the top eukaryotic phyla and their abundances in forest and blueberry farm soils. Mortierellomycota, Basidiomycota and Ascomycota species were abundant in resident soil communities, while soil communities metabolizing amino acids were more enriched with Mortierellomycota species. Farm soils, especially from high-productivity sites, were enriched with Mucoromycota species as well (more in the amino acid-active fraction of the microbial communities), compared to forest soils. The high-productivity soils were both enriched with Glomeromycota species in their resident communities, while low-productivity soils showed high abundances of Rozellomycota species instead. Glomeromycota species were almost absent in low-productivity soils and very few reads for Rozellomycota were found in high-productivity soils.
As seen from the community composition bubble plot (Figure 4), the two different forest soils also show clear differences in their resident eukaryotic communities; the most abundant taxonomic unit in one of the forests is an unclassified Tremellodendropsidales member (phylum Basidiomycota), whereas the samples of the other forest are mostly abundant in unclassified Mortierella. An unclassified fungus belonging to the family Glomeraceae dominates in high-productivity soils from two different farms, whereas the two low-productivity farms both exhibit higher abundances in unclassified GS08, GS11, Rozellomycota members (10-25 times higher), and ciliates from the genera Meseres and Gastrostyla. Additionally, high-productivity farm soils were also populated by more species from the genera Mucor and Rhizopus. Reads for earthworms (Elsenia fetida, Tubifex tubifex) were found in some of the samples, which mildly altered the relative abundances of other eukaryotic taxonomic units in those samples. The amino acid-active eukaryotic communities were dominated with an unclassified Mortierella representative for all soil types. However, high-productivity farm soils were more enriched with species for the genera Mucor, Rhizopus, and Umbelopsis actively metabolizing amino acids, whereas low-productivity soils exhibited an increased abundance in unclassified Pythiaceae members (phylum Oomycota), which were almost absent in high-productivity soils.

Discussion

The highbush blueberry plant grows in organic, well-aerated acidic sandy soils that have a pH of 4-5 (Haynes, 1988). Little is known about the highbush blueberry rhizosphere, but a recent study investigating the rhizosphere of the wild blueberry plant as well as the bulk soil has discovered plant-influenced species sorting in this system affecting soil
eukaryotes more dramatically than bacterial species in the soil. In addition, the fruit productivity of the plant has shown a negative correlation with several fungal and protist taxa (Yurgel et al., 2018). Additionally, treating the soil with different nutrients, such as phosphorus, was found to alter the rhizosphere microbiome of the blueberry plant (Pantigoso et al., 2018).

This study was initiated to ascertain if high resolution profiling of ribosomal RNA operons or detection of amino acid-active microbiota could be used to elucidate differences between high productivity and low productivity soils on blueberry farms. Our results demonstrate that rRNA operon profiling can provide nearly all of the SSU-ITS-LSU fragments from bacteria and eukaryotes in soils. This is important since in silico experiments have suggested that fungal identification biases arise when only one of the ITS regions is sequenced and it is recommended to generate reads covering the whole ITS region as well as parts of the 18S and 28S rRNA genes (Tedersoo and Lindahl, 2016). Furthermore, our SIP analysis identifies those bacteria and fungi from the sites which can actively metabolize $^{13}$C/$^{15}$N-amino acids. For example, *Paraburkholderia* species were found in all types of soil. However, these microorganisms were more active on $^{13}$C/$^{15}$N-amino acids in forest and high-productivity farm soils than in low-productivity farm soils. Some *Paraburkholderia* representatives are considered to be beneficial plant symbionts able to fix nitrogen and promote plant health (Kaur et al., 2017). In our study, *Paraburkholderia oxyphyla* was found to be the most actively growing bacterial species in high-productivity soils. This bacterium, first isolated from acidic forest soils, is not capable of N$_2$-fixation, unlike most of its phylogenetically closely related *Paraburkholderia* species. However, a novel species from the same genus, a symbiont capable of fixing
nitrogen, has recently been isolated from the nodules of *Mimosa gymnas* and shown to share almost 99% similarity with *Paraburkholderia oxyphyla* at the 16S rRNA gene level (Otsuka et al., 2011; Paulitsch et al., 2019). This suggests that there might be cryptic diversity within this specific taxonomic unit, which is not observed because of the lack of more closely related reference sequences to this bacterial species in the database.

Our study also found *Bacillus niacini* to be very active in low-productivity farm soils. *B. niacini* is a soil bacterium capable of metabolizing nicotinate (Nagel and Andreesen, 1991). Although the genome of the bacterium has been sequenced, it is unclear whether the organism can have any potential influence on plant health (Harvey and Snider, 2014). *Bacillus cucumis*, on the other hand, is a rhizosphere bacterium isolated from cucumber roots (Kampfer et al., 2016). *Bacillus soli* and *Bacillus novalis* are common soil bacteria isolated from hay fields (Heyrman et al., 2004). *Bacillus fumaroli* is another interesting bacterium inhabiting geothermal soils (De Clerck et al., 2004). Although we are not aware of any potential direct or indirect interactions of these *Bacillus* species with the blueberry plant inhabiting low-productivity farm soils, it is known that many bacilli can synthesize antibacterial and antifungal substances, and these natural products have been used in agriculture as microbial pesticides and fungicides (Pérez-García and de Vicente, 2011). Lastly, the existence of amino acid metabolizing *Streptacidiphilus* and *Streptomyces* species in forest soils is consistent with previous research that has identified *Streptomyces* species in acidic forest soils with capabilities of fighting soil-borne plant pathogens and *Streptacidiphilus* species in spruce forest soils (Hagedorn, 1976; Chen et al., 2012; Golinska et al., 2013).
With respect to eukaryotes, high-productivity blueberry farm soils were enriched with fungi from the Glomeromycota phylum. These fungi are an ancient group of eukaryotes and thought to be obligatory plant symbionts, forming arbuscular mycorrhizas with the roots of many vascular plants (Oehl et al. 2011; Redecker et al., 2013; Redecker and Raab, 2006). These fungi are considered beneficial symbionts for the plants and thought to contribute to the plant health by fighting fungal pathogens, harmful nematodes, aid in the uptake of phosphate/other inorganic ions, and to increase the fitness of plants in polluted environments (Schüßler et al., 2001; Redecker et al., 2013). It had been previously shown that multiple Glomeromycota species can enhance the height and dry weight of blueberry seedlings and increase the levels of N and P in the plant roots (Farias et al., 2014). Glomeromycota species were also found in the rhizosphere of the wild blueberry plant and were relatively more abundant in the rhizosphere itself than in the bulk soil (Yurgel et al., 2018). Another study investigating mycorrhizal fungi in the roots of *Vaccinium oldhamii* Miq. found a unique OTU very similar to *Rhizophagus diaphanus* (Baba et al., 2016) which suggests that Glomeromycota species may, in fact, colonize the roots of ericaceous plants though previously reported to be symbionts of other vascular plants. However, more research is required to understand whether these beneficial fungi are critical in the health of the highbush blueberry plant.

In contrast, low-productivity farm soils were enriched with Rozellomycota species, which are mainly unicellular endoparasites of algae, amoebae, oomycetes, and other fungi. These fungi-like organisms form zoospores and infect the host cell via a penetration tube. Most of the members of this phylum are uncultured environmental clones and not much information is known about their pathogenic lifestyle and potential hosts (Corsaro et al.
Since Rozellomycota species are pathogens of other fungi, they may also possess mechanisms for infecting their plant hosts. Furthermore, a recent study found that the abundance of Rozellomycota species was substantially higher in disease conducive soil (DCS) for tobacco cultivation (Gao et al., 2020). Our results also indicate that in high-productivity soils, where the members of the phylum Rozellomycota have very low abundance, Glomeromycota members thrive instead and vice versa. In addition, members of the Oomycota phylum are known pathogens of agricultural crops, nematodes, algae, zooplankton and fish. They are fungi-like organisms and can either form hyphae or live as very simple holocarpic thalli. Some species, though, are not predominantly pathogens; they are saprophytes but can act like opportunistic pathogens when the immune system of the host is compromised (Hassett et al., 2019; Klinter et al., 2019). Interestingly, amnio acid-active Oomycota species were abundant in low-productivity soils but were almost absent in high-productivity soils. Low-productivity farms were also relatively enriched with ciliates. Although, it is not clear how ciliates interact with plant rhizospheres, it has been shown that their diversity does not depend on the plant species unlike fungi and bacteria. Rather, the richness of ciliates in rhizosphere microbiomes may depend on the availability of inorganic N or the composition of rhizosphere bacterial communities (Acosta-Mercado and Lynn, 2004), suggesting predation may play a major role in removing beneficial microbes from the rhizosphere in low productivity soils.

Although we do not know whether these unique groups of organisms affect the blueberry plant in situ, these fungi may interfere with the health of the blueberry plant directly or indirectly and their presence and/or absence in some soils and not in others is a very valuable piece of information for further, more-targeted research. One future direction
for this research project would be sequencing whole genomes from the soil samples to reveal the functional capabilities of these unique bacterial and eukaryotic communities as well to gain insight into the potential mechanisms they can use to thrive in their environment. Additionally, other substrates for SIP can be utilized instead of, or in combination with, amino acids to preferentially select for specific microorganisms that are particularly interesting to study in these systems, for example $^{13}$C-lignin for saprophytic fungi.
REFERENCES


32. Noncitrus Fruits and Nuts 2018 Summary (June 2019) USDA, National Agricultural Statistics Service


CHAPTER THREE: DETECTING THE UNSEEN DIVERSITY WITHIN ABUNDANT TAXONOMIC UNITS

Introduction

Next-generation sequencing technologies reveal the microbial composition of complex environmental niches and it has become relatively simple to obtain information about unculturable organisms (Reza et al., 2019). Ribosomal RNA gene profiling (mostly 16S rRNA gene analysis) not only provides an estimate of the taxonomic and functional diversity of microbiomes but also enables researchers to track community dynamics over time, space etc. (Simon and Daniel, 2013). By using these 16S rRNA gene approaches, previously inaccessible and extreme environments for instance hydrogen sulfide-rich caves are elucidated in terms of microbial diversity and bacteria responsible for the formation of unique biofilms are identified (Jones et al., 2012). However, different steps involved in the most commonly performed small subunit (SSU) 16S rRNA gene investigations can introduce bias into the analysis. Those steps include DNA extraction, PCR methods, and data analysis (Brooks et al., 2015; Pollock et al., 2018). For example, all 16S rRNA gene studies rely on the comprehensiveness and quality of a given reference database, and most databases are comprised of reference sequences from microorganisms which have been cultured in the laboratory and do not include many reference sequences from the vast majority of microbial species inhabiting environmental samples. Although this bias exists, it is not clear how low database representation affects the results of the 16S rRNA gene analysis (Pignatelli et al., 2008; Breitwieser and Salzberg, 2019). Indeed, different classification methods used to assess prokaryotic diversity in complex communities employ different algorithms and results are not often identical (Illeghems et al., 2012).
Although scientists have been extensively sequencing total DNA extracted from various environments for the last few decades, many microbial taxonomic groups remain unclassified and better approaches are required to fill the corresponding gaps in the tree of life (Steen et al., 2019). What is more striking is that some habitats may harbor over 60% uncultured phyla (Lloyd et al., 2018). Long-read sequencing approaches provide an opportunity to analyze near full-length ribosomal genes and generate high-quality assemblies of newly discovered operons to include into reference databases of both prokaryotes and eukaryotes in the future (Karst et al., 2018). Long-read sequencing in this regard can be used to estimate the representation of a given reference database since it provides species-level identification of prokaryotic organisms based on complete ribosomal RNA operons (Martijn et al, 2019; Cuscó et al, 2018; Kerkhof et al., 2017; Benítez-Páez et al., 2017).

In Chapter two, I summarized how several million reads of near full-length ribosomal RNA operon sequences of both bacteria and eukaryotes were used for the analysis of the microbial community composition of blueberry farm and forest soil systems. We used some of the most comprehensive publicly available reference databases and more than 13,000 distinct taxonomic units were identified. A number of taxonomic were highly abundant in some samples, while a few exhibited >40% abundances (Figures 3 and 4 from Chapter two). However, we were curious to assess whether the taxonomic assignment of the several hundreds of thousands of sequences representing a single taxonomic unit by the BLAST algorithm contained cryptic groups of microorganisms. For example, some abundant taxonomic units were classified only at high taxonomic levels such as family, order and phylum in the reference databases (e.g. unclassified Ceratobasidiaceae,
Glomeraceae, Rozellomycota reference sequences), thus it was important to come up with a pipeline enabling us to estimate the true diversity of these microbial communities phylogenetically. We were able to detect not only abundant bacterial and eukaryotic species, but also thousands of very rare ones. This suggests that our approach is effective in discerning subtle differences between different complex soil samples, but limitations arise from the reference databases that we have used. Hence, it is crucial to reconstruct ribosomal RNA operons for several taxonomic units based on the available high sequencing coverage (also given the relatively high sequencing error rate of 5-15%) and analyze the phylogenies of those reconstructed operons to be able to identify the actual microorganisms inhabiting farm and forest samples more accurately and improve existing databases for future studies.

**Methods and materials**

*Selection of highly similar ribosomal RNA operon sequences*

In order to begin the re-construction of rRNA operons to test for cryptic variation from the Chapter two studies, it was necessary to identify 60-100 raw sequence reads that were highly similar. Figure 5 gives an overview of the data analysis pipeline. For this study, we performed a series of internal BLAST searches of sequences yielding the same best BLAST hit from the UNITE or EZ BioCloud databases from our blueberry farm/forest SIP study. For example, 5000 sequences related to unclassified *Rozellomycota* were randomly selected from a pool of 27,000 rRNA operon sequences originating from two different soil samples. A custom BLAST database was made from these particular 5000 sequences. Then, iterative BLAST searches were conducted with 50 randomly selected query sequences taken from the initial 5000. This was done using MegaBLAST with the
following default parameters: Index length: 28bp; Target sequences: 10, reward/penalty values: 0/none, match/mismatch scoring: 1/-2. The resulting hit table contained 500 sequences of the most similar rRNA operon sequences to the original search set. Duplicate hits were then removed to generate a pool of unique, highly similar sequences and the process was repeated on another 50 randomly selected sequences after removal of these sequences from the initial 5000 custom BLAST database. The previous steps of blasting 50 randomly selected sequences and making sequence alignments were repeated eight more times. In this manner, more than the 50% of initial 5000 sequences were quickly sampled and associated with their most similar sequences.

Figure 5. Schematic overview of the main steps of the procedure of the reconstruction of high-quality ribosomal RNA operons for different fungal and bacterial taxonomic units.

**Definition of sequence clusters**

In order to verify the MegaBLAST selection of rRNA operon sequences, a distance matrix/heatmap was created to visualize the sequence similarity. That is, multiple MUSCLE alignments (version 3.8.425) were created using the unique rRNA operon sequences (in the proper orientation verified by 16S or 18S rRNA universal primer
annotation), the alignment was grouped by similarity (rather than file order), and the distance matrix was color-coded based on similarity percentages. All clusters were identified visually. Following the initial clustering, highly similar sequences (from 8-20) within each of the clusters from the nine heatmaps were combined to make a total global alignment and reveal if the observed clusters were unique or were highly similar to each other.

Consensus reconstruction using LASTZ alignments

Once unique clusters of similar ribosomal RNA operons were identified, >60 sequences were chosen to make a high-quality long read consensus using an iterative approach (Kerkhof et al., 2017). Initially, a MUSCLE alignment was performed on the 10 operons having the highest pairwise identity from the >60 read alignment. This “mock” alignment was then used to map the next 20 “best” reads to create a “new” consensus sequence using LASTZ (Large-Scale Genome Alignment Tool; version 1.02.00) with a step length of 7 bp and a seed pattern of 12 of 19 (Harris, 2007). The “new” LASTZ consensus was then mapped with the “best” 40 sequences to generate a second LASTZ consensus. The second LASTZ consensus was then mapped against all 60 sequences and the final LASTZ consensus sequence was termed the “Long-read consensus (LRC)”.

If the total global alignment of ribosomal RNA operons of a given taxonomic unit resulted in multiple distinct similarity clusters containing less than 60 sequences, a larger custom BLAST database was created for that particular taxonomic unit containing >100k operons. BLAST searches were conducted using sequences (from 5-15) comprising similarity clusters in the total alignment as queries against the new large custom database. The number of target sequences was set to 50. This procedure was done, for instance, for
nine groups of bacterial sequences with the highest similarity to *Paraburkholderia oxyphila* and nine sequence alignments were made after removing duplicate hits. After selecting high-similarity regions from each of the alignment heatmaps, a new total global alignment was made. Unique clusters of >60 similar sequences could then be identified in this final total alignment and LRC sequences were created as described above.

**Phylogenetic analysis of LRC ribosomal RNA operons**

LRC sequences from eukaryotic and bacterial rRNA operons were compared with closely related rRNA operons from the NCBI NR gene database. Two sets of alignments were created for eukaryotic ribosomal genes (18S and 23S) using MUSCLE as described above and one set for bacterial ribosomal genes (16S). Maximum-likelihood phylogenetic trees were created on unambiguously aligned sequences with the FastTree algorithm (version 2.1.11) (Price et al., 2009).

**Results**

Three eukaryotic taxonomic units were selected for testing of cryptic variation from the Blueberry/Forest Soil dataset: unclassified *Mortierella* (the most abundant eukaryotic taxonomic unit in the amino acid-active fraction of the communities of all the soil samples), unclassified *Glomeraceae* (the most abundant taxonomic unit in high-productivity farm soil resident communities) and unclassified *Rozellomycota* (highly enriched in low-productivity farm soils). In addition, two bacterial taxonomic units classified at the species level: *Bacillus niacini* (actively abundant in low-productivity soils), and *Paraburkholderia oxyphila* (more active in high- than low-productivity soils) were also tested for cryptic variation. An example of the first and the ninth and the total MUSCLE alignment heatmaps
for one of the detected abundant eukaryotic taxonomic units (unclassified Mortierella) made during the iterative blasting procedure described above are shown Figure 6.

**Figure 6.** MUSCLE alignment heatmaps of unique sequences resulted from the first (A) and the ninth (B) internal BLAST searches for unclassified Mortierella-related sequences, and the total alignment heatmap (C) comprising sequences from all the observed high-similarity clusters (shown in black boxes) from the nine original alignments.

The ribosomal RNA operons originating from bacteria exhibited a slightly lower range of within taxonomic unit similarity percentages than those of eukaryotic taxonomic units. Figure 7 A shows the total MUSCLE alignment heatmap for Paraburkholderia
As it is seen from the heatmap, high similarity regions are very small and there is huge variance among those clusters. The new total alignment heatmap for *P. oxyphila* (Figure 7 B) after resampling a 20 times bigger database resulted in additional sequences in each of the high-similarity regions, thus it was possible to build high-quality consensus sequences. Additionally, *Paraburkholderia* tRNAs were used to annotate the operons to make sure each cluster represents a single operon (Suppl. Figure 8). *Bacillus niacini* operons were aligned and consensus sequences were made as explained above. Although *B. niacini* operons did not get annotated with any tRNAs from a pool of 40 various *Bacillus* tRNAs, it was possible to manually remove operons containing groups of extra nucleotides in their ITS region to create valid consensus sequences. All MUSCLE alignments including total alignments for all fungal and bacterial taxonomic units are provided in Supplemental Figures 9 – 14.
Figure 7. A) The total MUSCLE alignment heatmap for *Paraburkholderia oxyphila*-related sequences after nine iterative BLAST searches against a database of 5,000 sequences, B) the total alignment heatmap after BLAST searches of nine groups of high-similarity sequences (shown in black boxes) detected in the first total alignment against a database of 105,000 sequences.

LASTZ alignments resulted in high-quality consensus sequences with as few as 0 - 3 nucleotide ambiguities across 5.0-5.7 kbp for fungal ribosomal RNA operons and 4.0-4.4 kbp for bacterial operons. A picture of the final LASTZ alignment for unclassified *Mortierella* sp. Cluster2 including the unambiguous consensus sequence is shown in Suppl. Fig. 15. In total, four long-read consensus (LRC) sequences were generated for unclassified *Mortierella*, two for unclassified Rozellomycota and one for unclassified Glomeraceae. Bacterial ribosomal RNA operon LRCs were nine for *Paraburkholderia oxyphila*-associated reads and four for *Bacillus niacini*-associated reads. Eight of the *P. oxyphila* LRCs were annotated similarly with the same *Paraburkholderia* tRNAs, whereas the ninth LRC did not follow the annotation pattern of the others.

Maximum-likelihood FastTree phylogenetic trees of 18S (based on an alignment of 1351 bp) and 28S (2361 bp) partial rRNA gene sequences are shown in Figures 8 and 9. Both trees exhibited the same topology; Glomeraceae representatives, *Glomus* and *Funneliformis* species, form an outgroup in both trees. Zygomycota, Ascomycota, and Basidiomycota members form a distinct cluster together although these species dominated in the BLAST hits of both unclassified Rozellomycota and Glomeraceae LRCs. The two Rozellomycota-related LRCs form a monophyletic group with known Rozellomycota members such as *Rozella, Paramicrosporidium*, and *Nucleophaga* species, and some Glomeromycota members (*Rhizophagus* species). The Rozellomycota-associated LRCs are closely related to each other, which suggests that these may represent two different
species of the Rozellomycota, but both do not seem to belong to any of the above-mentioned known genera of the Rozellomycota. These sequences seem to be more closely related to *Paramicrosporidium* species.

The 28S rRNA gene tree validates that the Rozellomycota LRCs are most related to *Rozella* species since this pattern is also observed on the 18S rRNA gene tree. This suggests that the NCBI ribosomal partial gene reference databases are not rich with these lineages, so common BLAST hits are not actually related to our query sequences and it was necessary to manually add additional sequences to the trees from a number of taxa from closely related taxonomic groups to complete the analysis. The only Glomeraeaceae-associated LRC from our study did not cluster with any of the known Glomeraeaceae genera, *Glomus* and *Funneliformis*. In fact, it seems to be very distinct from all the other sequences on both trees as well, and the only observation that can be made is that it belongs to the same lineage of most Rozellomycota members. The *Mortierella*-related LRCs cluster with other known *Mortierella* species and each of them is more related to another *Mortierella* species rather than to each other as seen on the 18S rRNA gene. Overall, the similar topology between the 18S and 28S rRNA gene trees suggests that the rRNA operon profiling method does not create chimeric artifacts during amplification.
Figure 8. The maximum-likelihood FastTree 18S rRNA partial gene phylogenetic tree (1351 positions) including fungal consensus sequences of ribosomal RNA operon fragments.
Figure 9. The maximum-likelihood FastTree 28S rRNA partial gene phylogenetic tree (2361 positions) including fungal consensus sequences of ribosomal RNA operon fragments.

The 16S rRNA trees for *Bacillus niacini* (Figure 10) and *Paraburkholderia oxyphila*-associated LRCs (Figure 11) exhibit similar topology; *Staphylococcus* species represent an outgroup on both trees. Is can also be seen that most known *B. niacini* strains cluster together, but our *B. niacini* LCRs are more related to the *B. niacini* strain Et9/1, which does not cluster tightly with the other *B. niacini* strains. *P. oxyphila* LRC sequences, on the other hand, exhibit a greater diversity among each other and show clear separation between some of the sequences e.g. Cluster 1, 5, and 9 LRCs.
Figure 10. The maximum-likelihood FastTree 16S rRNA partial gene (1297 positions) phylogenetic tree including *Bacillus niacini*-associated LRC sequences of ribosomal RNA operons.
Figure 11. The maximum-likelihood FastTree 16S rRNA partial gene (1297 positions) phylogenetic tree including *Paraburkholderia oxyphila*-associated LRC sequences of ribosomal RNA operons.

Discussion

It was previously demonstrated that databases harboring microbiological data introduce bias in favor of microorganisms isolated from the human body, thus it becomes difficult to interpret metagenomics analysis results for microbial communities associated with other host organisms as well as those originating from environmental niches (Hildebrand et al., 2012). Ribosomal RNA based-metagenomic approaches introduce another bias for prokaryotic members of microbial communities because of operon copy number heterogeneity. For example, *in silico* experiments have shown that the abundance
of bacterial species with higher operon copy numbers is usually overestimated (Crosby and Criddle, 2003). Algorithms have been developed to try to correct for ribosomal RNA operon copy number related errors, but some biases towards specific organisms still exist (Guo et al., 2016; Vetrovsky and Baldrian, 2013). Recombination and horizontal gene transfer events have also been suggested to have occurred in the 16S rRNA gene of some bacteria. This is potentially another source of error when conducting 16S rRNA gene-based metagenomic analysis (Rajendhran and Gunasekaran, 2011).

*In silico* experiments also suggest biases can be introduced by different data processing approaches using distinct software, algorithms, and databases which may be partially resolved by manual inspection of taxonomic assignments and well-curated databases (Escobar-Zepeda et al., 2018). It is worth mentioning that effort was put forth to evaluate a few of widely used 16S rRNA gene databases by using a mock community and the best performance was designated to EzBioCloud, which was also used in our SIP study as a bacterial reference database (Park and Won, 2018). To account for some of these biases we attempted to assemble high-quality bacterial and fungal ribosomal RNA operons and were able to successfully reconstruct almost unambiguous consensus sequences for those operons, which were then used to investigate the phylogenies of some of the most abundant microorganisms in our samples. The phylogenetics analysis also demonstrated that known *B. niacini* strains exhibit a lot of diversity in their 16S rRNA gene and some *B. niacini* strains such as NBK36, Marseille-P3556, and NBR5A form a distinct group of bacilli within that genus. In addition, *B. niacini* strain Et9/1 was the closest relative of the reconstructed *B. niacini*-related LCRs. In fact, it is known for *Bacillus* species to originate from multiple evolutionary clades regardless of their common habitant (Hernández-
B. niacini was shown to be related to many Bacillus soil isolates than any other Bacillus species based on the 16S rRNA gene similarity (Felske et. al., 2004). The ribosomal RNA operon copy number of B. niacini strains is unknown, but the draft genome of the Strain JAM F8 indicated the presence of 13 rRNA genes (Kurata et. al, 2014). Some Bacillus spp. may contain up to 15 rRNA operons in their genomes and this was thought to be an evolutionary adaptation to nutrient availability in some strains (Valdivia-Anistro et al., 2016; Wu et. al, 2017). Thus, it is difficult to assess whether our LRCs for B. niacini represent multiple rRNA operons from a single strain or they come from different strains of the species. However, it is clear that the taxonomic assignment of this abundant taxonomic unit is correct and sequencing whole genomes could potentially shed light on understanding the copy number variation of rRNA operons in our B. niacini communities.

P. oxyphila, originally isolated from acidic soils, was shown to be more closely related to B. sacchari (Otsuka et. al, 2011). Some Paraburkholderia members isolated for a plant rhizosphere shared highest 16S rRNA gene similarity with P. oxyphila among other Paraburkholderia species, while housekeeping genes indicated that the isolate was more closely related to a different species, P. nodosa (Paulitsch et al., 2019). Several forest soil isolates, on the other hand, formed a distinct clade with P. oxyphila based on maximum-likelihood FastTree phylogenetics of 92 essential genes, but not based on the 16S rRNA gene (Xiao et. al., 2019). This suggests that proper taxonomic assignment for Paraburkholderia species relies on multi-faceted analysis of not only rRNA genes, but also core housekeeping genes, and culture-dependent characterization is also very important in discerning these bacterial species. Our results demonstrate that eight out of nine LRCs of
P. oxyphila-related sequences form a monophyletic group with the known P. oxyphila strain OX-01, although the clade clearly possesses internal diversity as well (Figure 11). The other P. oxyphila-associated LRC seems not to be related to this clade and it is most closely related to P. nodosa. This was also observed when the sequences were annotated with Paraburkholderia tRNAs; P. oxyphila LRC Cluster9 did not follow the tRNA annotation patterns of the other eight P. oxyphila LRCs. Hence, we have identified at least one new species very similar to P. oxyphila or a number of P. oxyphila strains. In addition, an uncultured Paraburkholderia species was detected (P. oxyphila LRC Cluster9), which was recognized as P. oxyphila based on BLAST assignments.

Fungi belonging to the Mortierella genus, mostly found in soil, are thought to be very diverse in the environment, but only some species have been isolated and cultured. It is important to establish a methodology of discerning Mortierella species based on both phylogenetic and morphological characteristics (Nam et. al, 2018; Nguyen et. al, 2019). Mortierella phylogenies are usually being investigated based on the 28S rRNA gene partial alignments of ~630 bp (Takashima et. al, 2018; Takashima et. al, 2019). Our approach is more robust since we used an almost 4 times longer alignment (2361 bp), but very few reference partial genes could be found to able to make an alignment of this length. This limits our ability to compare the detected four Mortierella-related LRCs with other Mortierella species, but the 18S rRNA gene tree clearly demonstrates that all LRCs are different from each other.

Glomeromycota species are common arbuscular mycorrhizal fungi and their species diversity depends on soil pH. Interestingly, Glomeromycota dominate the communities of mycorrhizal fungi in soils with pH > 6.5 (Stürmer et al, 2018). Mycorrhizal
fungi are known to undertake co-evolution with their plant host, which may take millions of years. Rhizophagus species are examples of such mycorrhizal fungi and are considered to be distinct members of the Glomeromycota, forming a separate group (Merckx et al., 2017). In fact, our trees demonstrate that Rhizophagus species are strongly divergent from other Glomeromycota such as Glomus and Funneliformis species. Hence, it is possible to expect that the Glomeraceae-related LRC also undertook unique evolutionary transitions with its host, showing clear separation from other Glomeromycota. Within the Glomeraceae family itself, several known species possess diverse strategies for the establishment of symbiotic relationships with plant roots, suggesting unique adaptations due to their individual evolutionary histories (Pepe et al., 2016). Others also point out that the phylogenies of Glomeromycota species remain largely unknown and morphological characteristics such as the formation of glomoid spores (the Glomeraceae family) plays a role in their taxonomy and classification as well (Jobim et al, 2019). Sequencing complete rRNA operons indeed enables deciphering Glomeraceae phylogenies and a new genus (Halonatospora) has recently been identified and classified with this approach (Blaszkowski et. al, 2018).

Rozellomycota members are also very diverse organisms representing mostly obligate pathogens of animals, amoebae, and humans. Their evolutionary history is unknown as well but are thought to have originated from early linages of fungi and their relatives (Corsaro et al., 2016). Morphological features such as the formation of flagellated zoospores versus immobile spores played a role in the classification of these organisms. New genera have been proposed in the phylum Rozellomycota based on 18S rRNA gene and ITS phylogenies as well (Corsaro et. al, 2020). Interestingly, some microsporidia-like
Rozellomycota members possess ITS-2 regions subjected to great reduction during their evolution from common ancestors, while others have lost their ITS-2 region completely (Corsaro et. al, 2019). These findings suggest the existence of great evolutionary divergence between many Rozellomycota species. The 18S and 28S rRNA gene trees demonstrate the distinctness of the two detected Rozellomycota-related LRCs from other Rozellomycota species and suggest that our samples may harbor a novel clade of these remarkable fungi-like organisms.

Overall, our approach of reconstructing almost full-length rRNA operons for some of the abundant taxonomic units detected in our study was successful in investigating their phylogenies via long alignment rRNA gene trees. Our results were consistent with previous research in terms of estimating the diversity of these organism, and we were able to identify distinct linages of uncultured fungi, as well as new species and/or strains of some bacteria. Our approach seems to be robust, however, these bacterial and fungal phylogenies would be revealed more accurately if complete genomes of the organisms in question were studied as well.
REFERENCES


The biogeographical history of the interaction between mycoheterotrophic *Thismia* (Thismiaceae) plants and mycorrhizal *Rhizophagus* (Glomeraceae) fungi. *Journal of Biogeography, 44*(8), 1869–1879.


CONCLUSIONS

Stable isotope probing and MinION sequencing were effective in profiling resident and amino acid assimilating members of bacterial and eukaryotic communities inhabiting commercial blueberry farm soils as well as forest soils. This approach revealed some differences between high- and low-productivity farm soils at the microbial level. Most of the differences between the two types of soils were detected in the amino acid-active fraction of bacterial communities and the resident portion of fungal community members. This implies that the use of specific SIP substrates is important and suggests that other $^{13}\text{C}$ and/or $^{15}\text{N}$ substrates could be utilized as well to target particular members of soil and blueberry rhizosphere microbiomes. Additionally, assembling high-quality rRNA operons was possible for a number of abundant fungal and bacterial taxonomic units. The phylogenetic analysis of the multiple long-read consensus sequences uncovered cryptic diversity within these taxonomic units and detected the presence of distinct bacterial and fungal species/strains as well as novel fungal clades.

Moreover, it could potentially be possible to reconstruct at least one (and up to 4-5) ribosomal RNA operons for about 400 more different eukaryotic taxonomic with the available data given the existing sequencing coverage. It was estimated that if three times more sequencing data were generated, then rRNA operons could be assembled for at least another 300 eukaryotic taxonomic units from these soil systems; most of these taxonomic units belong to fungi. It would be very beneficial for similar metagenomic research to employ a database of complete, high-quality, unambiguously assembled rRNA operons since most taxonomic assignments for fungi rely of the ITS region only with existing reference databases. Hence, these reconstructed rRNA operons could be a great addition to
the available databases. This applies to bacterial taxonomic units as well, since rRNA operons belonging to different strains/or new species can potentially be identified.
### APPENDICES

**Table S1.** Bacterial (A) and eukaryotic (B) universal rRNA primers attached to ONT barcode sequences.

#### A.

<table>
<thead>
<tr>
<th>Bacterial rRNA Operon Barcode (BC) Primers</th>
<th>Nucleotide Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Forward primers</strong></td>
<td></td>
</tr>
<tr>
<td>BC1 Forward</td>
<td>GGTGCTGAAGAAAGTTGTCGGTGTTTTGTGAGAGTTTGATCCTGGCTCAG</td>
</tr>
<tr>
<td>BC2 Forward</td>
<td>GGTGCTGATCGATTCCGTGTAGTGCTGCTGAGTTTGATCCTGGCTCAG</td>
</tr>
<tr>
<td>BC3 Forward</td>
<td>GGTGCTGGAGTCTTGTGTCCCAGTAGACAGAGTTTGATCCTGGCTCAG</td>
</tr>
<tr>
<td>BC4 Forward</td>
<td>GGTGCTTGGATTCTAATCCTGATTTCCCTAAGAGTTTGATCCTGGCTCAG</td>
</tr>
<tr>
<td>BC5 Forward</td>
<td>GGTGCTGCTTTGTCCAGGGTTTGATACCTTAGAGTTTGATCCTGGCTCAG</td>
</tr>
<tr>
<td>BC6 Forward</td>
<td>GGTGCTTTCTCGAAAGGCAGAAGTTAGTCAGAGTTTGATCCTGGCTCAG</td>
</tr>
<tr>
<td>BC7 Forward</td>
<td>GGTGCTGTTACCGTGAGAATCCCTAGAGTTTGATCCTGGCTCAG</td>
</tr>
<tr>
<td>BC8 Forward</td>
<td>GGTGCTGTTACGGGAACAAGCAAGTTACGTAGAGTTTGATCCTGGCTCAG</td>
</tr>
<tr>
<td>BC9 Forward</td>
<td>GGTGCTGAACCTAGGCAGCGAGTGCTTGTAGTTAGTTTGATCCTGGCTCAG</td>
</tr>
<tr>
<td>BC10 Forward</td>
<td>GGTGCTGAAGCGTTGAAACCTTTGTCCCTCAGAGTTTGATCCTGGCTCAG</td>
</tr>
<tr>
<td>BC11 Forward</td>
<td>GGTGCTGTTCTCCAGGGTAAGTACGTAGAGTTTGATCCTGGCTCAG</td>
</tr>
<tr>
<td>BC12 Forward</td>
<td>GGTGCTGCAAGGTAGAAAGCAAAGGAAATCGGAAGAGTTTGATCCTGGCTCAG</td>
</tr>
<tr>
<td><strong>Reverse primers</strong></td>
<td></td>
</tr>
<tr>
<td>BC1 Reverse</td>
<td>GGTGCTGAAGAAAGTTGTCGGTGTTTTGTGACCAGCCAGGAATACCTGGCTCAG</td>
</tr>
<tr>
<td>BC2 Reverse</td>
<td>GGTGCTGATCGATTCCGTGTAGTGCTGCTGAGTTTGATCCTGGCTCAG</td>
</tr>
<tr>
<td>BC3 Reverse</td>
<td>GGTGCTGGAGTCTTGTGTCCCAGTAGACAGAGTTTGATCCTGGCTCAG</td>
</tr>
<tr>
<td>BC4 Reverse</td>
<td>GGTGCTTGGATTCTAATCCTGATTTCCCTAAGAGTTTGATCCTGGCTCAG</td>
</tr>
<tr>
<td>BC5 Reverse</td>
<td>GGTGCTGCTTTGTCCAGGGTTTGATACCTTAGAGTTTGATCCTGGCTCAG</td>
</tr>
<tr>
<td>BC6 Reverse</td>
<td>GGTGCTTTCTCGAAAGGCAGAAGTTAGTCAGAGTTTGATCCTGGCTCAG</td>
</tr>
<tr>
<td>BC7 Reverse</td>
<td>GGTGCTGTTACCGTGAGAATCCCTAGAGTTTGATCCTGGCTCAG</td>
</tr>
<tr>
<td>BC8 Reverse</td>
<td>GGTGCTGCAAGGTAGAAAGCAAAGGAAATCGGAAGAGTTTGATCCTGGCTCAG</td>
</tr>
<tr>
<td>Eukaryotic rRNA Operon Barcode (BC) Primers</td>
<td>Nucleotide Sequence (5’ to 3’)</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td><strong>Forward primers</strong></td>
<td></td>
</tr>
<tr>
<td>BC1 Forward GGTGCTGAAGAAAGTTGTCGGTGTTTGTGTAACCTGGTTGATQCTGCC AGT</td>
<td></td>
</tr>
<tr>
<td>BC2 Forward GGTGCTGTCGATTCCGTTTGTAGTCGTCTGTTACCTGGTTGATQCTGCCA GT</td>
<td></td>
</tr>
<tr>
<td>BC3 Forward GGTGCTGGAGTCTTGTGCCCTACACCAGGTTACCTGGTTGATQCTGCC AGT</td>
<td></td>
</tr>
<tr>
<td>BC4 Forward GGTGCTGTTCGGATTCTATCGTGTTTCCCTATACCTGGTTGATQCTGCCA GT</td>
<td></td>
</tr>
<tr>
<td>BC5 Forward GGTGCTGCTTGTCCAGGGTTTGTGTAACCTTTACCTGGTTGATQCTGCC AGT</td>
<td></td>
</tr>
<tr>
<td>BC6 Forward GGTGCTGGTGTTACCGTGGAATGAATCCTTTACCTGGTTGATQCTGCC AGT</td>
<td></td>
</tr>
<tr>
<td>BC7 Forward GGTGCTGGTGTTACCGTGGAATGAATCCTTTACCTGGTTGATQCTGCC AGT</td>
<td></td>
</tr>
<tr>
<td>BC8 Forward GGTGCTTTCAGGGAACAAACCAAAGTTACCTGGTTGATQCTGCC AGT</td>
<td></td>
</tr>
<tr>
<td>BC9 Forward GGTGCTGAAGAAAGTTGTCGGTGTTTGTGTAACCTGGTTGATQCTGCC AGT</td>
<td></td>
</tr>
<tr>
<td>BC10 Forward GGTGCTGAAGAAAGTTGTCGGTGTTTGTGTAACCTGGTTGATQCTGCC AGT</td>
<td></td>
</tr>
<tr>
<td>BC11 Forward GGTGCTTAATTTCTCATCTCAGGGAATGGATACCTTTACCTGGTTGATQCTGCC AGT</td>
<td></td>
</tr>
<tr>
<td>BC12 Forward GGTGCTGCAAGTAGAAAGAAGCAGAATCCGATACCTGGTTGATQCTGCC AGT</td>
<td></td>
</tr>
<tr>
<td><strong>Reverse primers</strong></td>
<td></td>
</tr>
<tr>
<td>BC1 Reverse GGTGCTGAAGAAAGTTGTCGGTGTTTGTGTAACCTGGTTGATQCTGCC AGT</td>
<td></td>
</tr>
<tr>
<td>BC2 Reverse GGTGCTGTCGATTCCGTTTGTAGTCGTCTGTTACCTGGTTGATQCTGCCA GT</td>
<td></td>
</tr>
<tr>
<td>BC3 Reverse GGTGCTGGAGTCTTGTGCCCTACACCAGGTTACCTGGTTGATQCTGCC AGT</td>
<td></td>
</tr>
<tr>
<td>BC4 Reverse GGTGCTGTTCGGATTCTATCGTGTTTCCCTATACCTGGTTGATQCTGCCA GT</td>
<td></td>
</tr>
<tr>
<td>BC5 Reverse GGTGCTGCTTTCAGGGAACAAAGTTAATGCTCAGGTTGATQCTGCC AGT</td>
<td></td>
</tr>
<tr>
<td>BC6 Reverse GGTGCTGTTCAGGGAACAAAGTTAATGCTCAGGTTGATQCTGCC AGT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>BC7 Reverse</td>
<td>GGTGC\textit{c}GTG\textit{i}GTTC\textit{c}CGTGGGAA\textit{t}TAGAAT\textit{c}CTTTCTGACTTAGAGGCG\textit{t}TCAGTC\textit{a}ATAAT</td>
</tr>
<tr>
<td>BC8 Reverse</td>
<td>GGTGC\textit{c}GTG\textit{c}TTCA\textit{g}GGGAA\textit{a}CACAACCAAC\textit{g}AG\textit{t}TTAG\textit{c}TGACTTAGAGGCCG\textit{t}TCAGTC\textit{a}ATAAT</td>
</tr>
<tr>
<td>BC9 Reverse</td>
<td>GGTGC\textit{c}GTG\textit{a}ACTAGGC\textit{a}CAGCGAGTCTTGGTTGTTCTGACTTAGAGGCCG\textit{t}TCAGTC\textit{a}ATAAT</td>
</tr>
<tr>
<td>BC10 Reverse</td>
<td>GGTGC\textit{c}GTG\textit{a}AGCGTTGAAACCTTCTTGTCTCTTCTCTGACTTAGAGGCCG\textit{t}TCAGTC\textit{a}ATAAT</td>
</tr>
<tr>
<td>BC11 Reverse</td>
<td>GGTGC\textit{c}GTG\textit{c}TTTATCTATGC\textit{a}GGAAT\textit{g}ATAG\textit{t}CTGACTTAGAGGCCG\textit{t}TCAGTC\textit{a}ATAAT</td>
</tr>
<tr>
<td>BC12 Reverse</td>
<td>GGTGC\textit{c}GTG\textit{c}GAGTA\textit{a}AGA\textit{g}G\textit{a}AGCGAG\textit{a}TCTG\textit{a}G\textit{a}G\textit{a}TCTGACTTAGAGGCCG\textit{t}TCAGTC\textit{a}ATAAT</td>
</tr>
</tbody>
</table>

**Figure S1.** Pictures of agarose gel electrophoresis of the PCR products of ribosomal RNA operons amplified from total genomic DNA derived from light ($^{12}\text{C}/^{14}\text{N}$) and heavy ($^{13}\text{C}/^{15}\text{N}$) and isotope containing DNA bands separated in CsCl gradients.
Figure S2. The NMDS stress plots based A) eukaryotic and B) bacterial species-abundance data indicating that $k = 3$ (number of dimensions) has stress values lower than 0.01 for both analyses, which is within the optimal range of stress values that provide a good model fit for observed dissimilarities.
Figure S3. The ordination distance plots for the NMDS analysis performed with $k = 3$ (number of dimensions) based on the A) bacterial and B) eukaryotic species-abundance data showing $R^2$ values for non-metric and linear fit of the models.
Figure S4. The correlation between the sequence length of bacterial and eukaryotic BLAST alignments of query and references sequences and the percent pairwise identity for those BLAST hits based on a randomly selected subset of 10,000 BLAST alignments.
**Figure S5.** Bar charts showing the average relative abundances of seven most abundant bacterial phyla in resident and amino acid-active (AA-Active) microbial communities of three types of soils (Forest, high-productivity, and low-productivity) from two blueberry farms.

**Figure S6.** Six most abundant eukaryotic clades detected in resident and amino acid-active microbial communities based on their average relative abundances in all soil samples.
Figure S7. Bar charts showing the average relative abundances of seven most abundant fungal phyla in resident and amino acid-active (AA-Active) microbial communities of three types of soils (Forest, high-productivity, and low-productivity) from two blueberry farms.

Figure S8. Putative *Paraburkholderia oxyphila*-related ribosomal RNA operons annotated with 16S rRNA gene primers and *Paraburkholderia* tRNAs.
Figure S9. MUSCLE alignment heatmaps of Rozellomycota-associated sequences, first through ninth (Groups A - F) and the total alignment made from high-similarity clusters detected in all initial alignments.
Figure S10. MUSCLE alignment heatmaps of Glomeraceae-associated sequences, first through ninth (Groups A - F) and the total alignment made from high-similarity clusters detected in all initial alignments.
Figure S11. MUSCLE alignment heatmaps of *Mortierella*-associated sequences, first through ninth (Groups A - F) and the total alignment made from high-similarity clusters detected in all initial alignments.
Figure S12. MUSCLE alignment heatmaps of *Bacillus niacini*-associated sequences, first through ninth (Groups A - F) and the total alignment made from high-similarity clusters detected in all initial alignments.
Figure S13. MUSCLE alignment heatmaps of *Paraburkholderia oxyphila*-associated sequences, first through ninth (Groups A - F) and the total alignment made from high-similarity clusters detected in all initial alignments.
Figure S14. MUSCLE alignment heatmaps of *Paraburkholderia oxyphila*-associated sequences after BLAST searches of nine groups of sequences (Groups A - F) originating from nine high-

...
similarity clusters detected in the total alignment made from all initial alignments (Suppl. Fig. 12) against a 105,000 sequence containing database.

**Figure S15.** The second LASTZ alignment and the unambiguously assembled consensus sequence generated from 60 “best” *Mortierella*-associated ribosomal RNA operons comprising the second high-similarity cluster on the total alignment of *Mortierella*-associated sequences.