Predictive Functional Profiling of Soil Microbes under Different Tillages and Crop Rotations in Ohio

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in the Graduate School of The Ohio State University

By

Janani Hariharan

Graduate Program in Environmental Science

The Ohio State University

2015

Master's Examination Committee:

Dr. Parwinder S. Grewal (Co-advisor)
Dr. Warren A. Dick (Co-advisor)
Dr. Margaret E. Staton
Abstract

Food production and security is dependent on maintaining soil health and quality. Thus, the emphasis on sustainable and healthy soil function is a top priority for scientists and land managers. One of the most important factors that influences soil function is the microbial community. Recent advances have allowed us to quantify more accurately the composition of such communities, but there is still a knowledge gap with regard to the contribution of microorganisms to various processes occurring in the soil. Understanding this will facilitate the development of healthier agroecosystems.

In this thesis, a predictive functional approach is used to elucidate bacterial species–function relationships. Bacterial community profiles were compared across two tillage systems and two crop rotations in Northern Ohio (Wooster and Hoytville). 16S rRNA gene-targeted sequencing was performed and the raw data obtained were filtered, denoised and processed using QIIME. Open-reference OTU picking and taxonomic assignment was performed using the Greengenes database. I then used a computational approach called PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) to predict metagenomes and the most likely functions performed by individual species of bacteria.

Sequence analysis reveals a large number of unidentified OTUs, which is consistent with our expectations of the soil ecosystem. Comparison of sequencing
data from different platforms indicates that the dataset generated using Illumina sequencing provided better hits with the reference database than pyrosequencing, and was associated with a greater number of putative soil bacterial functions.

PICRUSt allows an estimation of the level of involvement each OTU has with a specific gene function, which enables comparisons to be made across bacterial species and treatment conditions. Predicted functions of the bacterial community revealed a large number of proteins connected with metabolism and maintenance of natural organic molecules in soil as well as enzymes related to degradation of xenobiotics. Using this approach, I was also able to map specific OTUs to their functional potential. Bacterial enzymes implicated in the cycling of nitrogen, sulfur, carbon and methane through the soil were examined, as were enzymes that catalyzed the oxidative degradation of hydrocarbon compounds that are considered soil pollutants.

Specialized groups of bacteria were linked to functions like nitrogen fixation and degradation of compounds like atrazine and chlorohydrocarbons. A broader range of OTUs was found to contain genes for carbon utilization and sulfur metabolism. These predictions are supported by previous ecological studies. There were other OTU-function relationships predicted in these studies that are novel and could be valuable in identifying commercially important microorganisms. These leads will require experimental validation.

A clear difference was seen between the no-till and plow-till treatments, with no-till being functionally enriched for most major nutrient cycles. No such differences were observed between the different crop rotations. *Proteobacteria*, *Actinobacteria* and *Acidobacteria* were some of the most abundant phyla found in these soil samples,
along with *Nitrospirae*, and *Bacteroidetes*. I concluded that long-term and continuous application of different tillage systems, and to a lesser extent crop rotation, result in unique bacterial communities that affect the overall functioning of the soil.
Acknowledgements

My heartfelt gratitude goes out to my co-advisors, Drs. Warren Dick and Parwinder Grewal. Dr. Dick has been an incredible support through this challenging journey, and made me feel welcome in a new country and working environment when I was adjusting to life as a graduate student. Dr. Grewal has been an enthusiastic and warm presence, all through the admission process, failed experiments and thesis writing. For this and more, I will be indebted to you both. I immensely appreciate Dr. Staton’s efforts to guide my bioinformatics work despite being in a different location. Her willingness to discuss ideas over Skype calls and prompt responses to any queries I had were vital in the completion of this project.

I would certainly be remiss if I did not thank Aditi Sengupta, a colleague and PhD candidate. Thank you for all the advice when things weren’t going right and for being someone to celebrate with when they did. The same can be said about other friends in Wooster like Ashley Markazi and Debora Liabeuf, who have always been supportive and positive. Thanks are due to all the members of both Dick and Grewal labs, for their unwavering assistance with things big and small. I have learnt valuable things from many of you, especially Drs. Harit Bal and Ruisheng An.

It goes without saying that I would not have been able to complete this thesis without love and unconditional support from my parents. Thank you for standing by me through all the tough times. These two years have been wonderfully rewarding and a
great learning experience, and I hope that this will continue to hold true for a long time.
Vita

2009 ..........................................................S.B.O.A School and Junior College, Chennai, India

2013 ..........................................................B.Tech. Biotechnology, Anna University

2013 – 2014 .................................................University Graduate Fellow, The Ohio State University

2014 - 2015 ..................................................CFAES Environmental Graduate Research Fellow, The Ohio State University

Fields of Study

Major Field: Environmental Science
Table of Contents

Abstract...........................................................................................................................................ii

Acknowledgements.........................................................................................................................iv

Vita..................................................................................................................................................vii

List of Tables...................................................................................................................................x

List of Figures....................................................................................................................................xii

Chapter 1: Introduction and Rationale............................................................................................1

1.1 Overview.....................................................................................................................................1

1.2 The Concept of Tillage...............................................................................................................2

1.3 Crop Rotations...........................................................................................................................4

1.4 Biogeochemical Cycles and Soil Enzyme Activity.................................................................7

1.5 Measuring Microbial Diversity in Soils.....................................................................................9

1.6 16S rRNA Gene Sequencing....................................................................................................13

1.7 Rationale for Studying Community Structure and Function..............................................15

1.8 Hypotheses...............................................................................................................................17

1.9 Objectives.................................................................................................................................18

References.........................................................................................................................................19

Chapter 2: Pyrosequencing Pilot Study............................................................................................31

2.1 Introduction...............................................................................................................................31

2.2 Materials and Methods............................................................................................................33

2.3 Results and Discussion.............................................................................................................40
Chapter 3: Functional Profiling of Tillage and Crop Rotation Effects on Hoytville Soils

3.1 Introduction

3.2 Materials and Methods

3.3 Results and Discussion

3.4 Summary

References

Chapter 4: Summary and Future Directions

4.1 Summary

4.2 Future Directions

Bibliography
List of Tables

Table 2.1 Alpha diversity indices for soil samples under different tillage treatments.45

Table 2.2 KEGG functions taken into consideration for assessment of soil biodiversity, as modified from Jones & Bradford, 2001.................................46

Table 2.3 Summary of KEGG genes and Greengenes OTUs involved in nitrogen metabolism.................................................................51

Table 2.4 Summary of KEGG genes and Greengenes OTUs involved in sulfur metabolism.................................................................59

Table 2.5 Summary of KEGG genes and Greengenes OTUs involved in hydrocarbon degradation.............................................................66

Table 2.6 Summary of KEGG genes and Greengenes OTUs involved in carbon fixation.................................................................71

Table 2.7 Summary of KEGG genes and Greengenes OTUs involved in methane metabolism.............................................................81

Table 3.1 Sample treatments in Hoytville plots.................................................95

Table 3.2 Alpha diversity indices for soil samples under different treatments........106
Table 3.3 Summary of most abundant KEGG genes and Greengenes OTUs involved in nitrogen metabolism ........................................................................................................112

Table 3.4 Summary of most abundant KEGG genes and Greengenes OTUs involved in sulfur metabolism ........................................................................................................121

Table 3.5 Summary of most abundant KEGG genes and Greengenes OTUs involved in hydrocarbon degradation ........................................................................................................128

Table 3.6 Summary of most abundant KEGG genes and Greengenes OTUs involved in carbon fixation ........................................................................................................140

Table 3.7 Summary of most abundant KEGG genes and Greengenes OTUs involved in methane metabolism ........................................................................................................150
List of Figures

Figure 2.1 Workflow for library preparation ...........................................39

Figure 2.2 Bioinformatics workflow .........................................................40

Figure 2.3 Total number of OTUs picked at different sequence similarity levels .....41

Figure 2.4 Total number of reference OTUs picked at different sequence similarity levels .................................................................41

Figure 2.5 Increase in number of predicted functional relationships with sequence similarity ..........................................................42

Figure 2.6 Rarefaction curve of OTUs generated with pyrosequencing depth ............................................................44

Figure 2.7 Top ten phyla in No-Till and Plow-Till soil samples, ranked in order of abundance ..........................................................44

Figure 2.8 Comparison of soil functional enrichment across No-Till and Plow-Till ecosystems ..........................................................48

Figure 2.9 Genes involved in nitrogen metabolism through the nitrogen cycle ....49

Figure 2.10 Nitrogen metabolism enzymes with highest no-till:plow-till gene abundance ratios .............................................................50
Figure 2.11 Sulfur cycle.................................................................56

Figure 2.12 Sulfur metabolism enzymes with highest no-till:plow-till gene abundance ratios.................................................................57

Figure 2.13 Hydrocarbon degradation enzymes with highest no-till:plow-till gene abundance ratios..........................................................63

Figure 2.14 Carbon fixation enzymes with highest no-till:plow-till gene abundance ratios.................................................................70

Figure 2.15. Methane metabolism with highest no-till:plow-till gene abundance ratios.................................................................80

Figure 3.1 Workflow for processing of soil samples in the laboratory..............98

Figure 3.2 Workflow for bioinformatics analysis of 16S rRNA amplicon data using PICRUSt.................................................................102

Figure 3.3 Rarefaction curve of OTUs generated with sequencing depth.........104

Figure 3.4 Visualization of top ten phyla in Hoytville samples..................105

Figure 3.5 No-till/Plow-till (NT/PT) gene abundance ratios in Hoytville samples.................................................................108

Figure 3.6 Corn-soybean/Continuous corn (CS/CC) gene abundance ratios in Hoytville samples.................................................................109
Figure 3.7 Nitrogen metabolism enzymes with highest no-till:plow-till gene abundance ratios ................................................................. 111

Figure 3.8 Nitrogen metabolism enzymes with highest corn-soybean:continuous corn gene abundance ratios ................................................................. 112

Figure 3.9 Sulfur metabolism enzymes with highest no-till:plow-till gene abundance ratios .................................................................................... 119

Figure 3.10 Sulfur metabolism enzymes with highest corn-soybean:continuous corn gene abundance ratios ............................................................................. 120

Figure 3.11 Hydrocarbon degradation enzymes with highest no-till:plow-till gene abundance ratios .................................................................................... 126

Figure 3.12 Hydrocarbon degradation enzymes with highest corn-soybean:continuous corn gene abundance ratios ............................................................................. 127

Figure 3.13 Carbon fixation enzymes with highest no-till:plow-till gene abundance ratios .................................................................................... 138

Figure 3.14 Carbon fixation enzymes with highest corn-soybean:continuous corn gene abundance ratios .................................................................................... 139

Figure 3.15 Methane metabolism enzymes with highest no-till:plow-till gene abundance ratios .................................................................................... 148
Figure 3.16 Methane metabolism enzymes with highest corn-soybean:continuous corn
gene abundance ratios………………………………………………………………………………149
CHAPTER 1: INTRODUCTION AND RATIONALE

1.1 OVERVIEW

Agriculture, a landmark in the evolution of mankind, has now become a powerful force in our society. One of the most essential requirements for food production is the substrate, soil. Soil has various physical and chemical factors that contribute to its fertility and health, but perhaps the most important one is biological. The richness and diversity of organisms that inhabit soil are unique. Different types of soils around the world support very different types of organisms, both micro and macro (Bates et al., 2011; Sait et al., 2002).

While extensive studies have been conducted on the importance of insects, nematodes, earthworms and other organisms to crop productivity (Lattanzio et al., 2006; Thompson & Goggin, 2006; Scheu, 2003; Williamson & Gleason, 2003; Brown et al., 1994), the precise functions fulfilled by microorganisms in an agroecosystem still remain relatively unknown. Research in the past few decades has indicated that a vast number of microbes thrive and reproduce in the soil, with some estimates ranging up to billions of microorganisms in each gram of soil (Torsvik & Øvreås, 2002).

Soil microorganisms are integral components of a wide range of ecosystem processes, including cycling of nutrients, soil carbon storage, production of greenhouse gases, degradation of xenobiotics and pollutants, and maintenance of soil structure (Groffman & Bohlen, 1999). Breakdown of soil community structure has
been predicted to have dramatic and frequently negative effects on the soil food web, ecosystem processes, ecosystem stability and community composition (Moore & de Ruiter, 1997). In fact, it has been argued that a world without microbes would be unable to sustain itself for longer than a few months, during which there would be widespread chaos and anarchy (Gilbert & Neufeld, 2014).

For all their importance, microbial interactions are rarely one-dimensional or easy to classify. A large number of microbes show some level of adaptability and can co-operate with other microbes to form mutually beneficial or commensalistic relationships. This could be with other microbes or with plants (one example of this would be nitrogen fixation by *Rhizobia*) and in some cases, also with nematodes and small animals (Tan & Shapira, 2011; Dillon & Dillon, 2004). No matter the type of interaction, it is clear that microorganisms are invisible but ubiquitous parts of agriculture.

1.2 THE CONCEPT OF TILLAGE

Historically, farming has been practiced by tilling the soil during the growing season. Plowing was the preferred method of tillage until a few decades ago, when Edward Faulkner suggested that the plow itself could be a major cause of soil erosion. This prompted investigation into alternative, less damaging methods of tillage such as the chisel till. Eventually, the concept of ‘no-till’ or leaving the soil undisturbed from season to season was formed (Faulkner, 1943).

The no-till practice relies on previous years’ crop residue to enhance the availability of nutrients and increase microbial activity in the soil. Soil erosion and runoff losses are minimized due to minimal disturbance of the topsoil layer and the
protective residue layer over the soil, and it has been seen to promote microbial enzyme activity of the ecosystem as well (Dick, 1984). This is because absence of tillage allows the build-up of bacterial communities and fungal hyphae networks in the soil, which in turn can lead to increased concentration of amino sugars in the soil (Zhang et al., 2012). Differences in tillage have an effect on macroaggregates, microbial biomass (Hernandez-Hernandez & Lopez-Hernandez, 2002) and enzymatic activity (Dick, 1984) in the soil. Soil carbon content and nutrients are distributed differently in the soil profile under no till compared to plow till (Mestelan et al., 2008). They are concentrated more at the soil surface with levels generally significantly higher under no-till than plow till. These changes in nutrient dynamics are accompanied by concomitant shifts in microbial community structure, suggesting a possible relationship between microbial community composition and ecosystem function (Calderon et al., 2001).

More recently, the no-tillage system has captured the attention of environmentalists as it acts as a strong sink for carbon sequestration (Lal, 2004). The no-tillage method is also thought to reduce the emission of greenhouse gases like methane and nitrous oxide compared to more intensive tillage, thus lessening the impact of global warming (Omonode et al., 2011). However, the tillage effect on nitrous oxide can also be the opposite (Rochette et al., 2007). Agriculture has been identified as one of the top contributors to climate change (Vermeulen et al., 2012) and no-till is being viewed as one of the possible remedies to lessen this impact. The ability of microorganisms to lock carbon in the soil is vital to this, and the absence of tillage has been shown to improve this ability. Long-term no-tillage is known to increase the sizes of the carbon pools (active, slow and passive) in soil, especially in
the case of continuous corn (CC) cultivation (Sherrod et al., 2005). However, it is not yet clear whether this phenomenon (no-tillage to enhance soil carbon sequestration) remains constant over long periods of time. In a tillage treatment experiment on continuous corn cultivation, microbial biomass, respiration and earthworm population were shown to be enhanced by the no-till treatment, as were phosphorus and potassium nutrient stratification (Karlen et al. 1994). All these factors are thought to enhance overall soil quality and health as well.

Although there are still some unanswered questions pertaining to the no-till system, it is considered to be a beneficial land management practice overall. The popularity and public interest towards this method of farming is increasing, largely due to numerous grants and incentives provided by the government to farmers who choose to participate in no-tillage (USDA, 2014).

With increasing awareness of sustainable land management practices and the crucial role of microorganisms in agricultural productivity and security, farmers and land managers have become more open to the idea of using microbial augmentation (addition of microorganisms to the soil to improve upon some existing functions) to increase the fertility and productivity of their soils. There is a real need to understand the functions performed by each of these species, and how they will respond to the presence of introduced microbes in their habitat. One of the important things this study focuses on is a comparative analysis of microbial functions in soils under different tillage practices.
1.3 CROP ROTATIONS

Approximately 80% of the U.S. corn crop is now grown in two-year rotations with soybeans or three-year rotations with soybeans and wheat (Baldwin, 2004). Similar crop rotations have been used in the plots studied in these experiments as well. Such crop rotations are generally preferred over continuous monocropping cultivations for a variety of reasons, related to soil fertility and function, as well as to reduce pest and disease buildup (Karlen et al., 2006; Murphy et al., 2006; Peters et al., 2003; McLaughlin & Mineau, 1995).

Soil structure, as measured by soil aggregate stability, bulk density, water infiltration rate and soil erosion, has generally been found to be improved under different crop rotations than under a continuous monocropping system (Bullock, 1992). In addition, various studies conducted in the past 30 years have suggested that crop rotations that include legumes are beneficial to aggregate stability and produce favorable soil structure (Robinson et al. 1996; Raimbault & Vyn 1991; Perfect et al. 1990; Power 1990; Tisdall & Oades 1982). Soil erosion is reduced by the introduction of crop rotation systems, specifically those including forages and other surface covers (Johnston et al., 1942). However, this is not seen for corn-soybean rotations, where rates of soil loss through erosion have been higher under the rotations than under continuous corn (CC) (Van Doren et al., 1984). This has been attributed to the fact that the amount of surface residue left behind following soybean is very low (Papendick & Elliott, 1984). Corn-soybean (CS) rotations also decrease soil organic matter (SOM) content in comparison to continuous corn cultivation, simply because soybean does not provide as much biomass as corn (Eckert et al., 1986). Rotation
length, tillage, and fertilizer and manure interactions appear to be the most important factors in estimating the effect of crop rotation on SOM (Karlen et al. 1994).

Crop rotation increases plant yields in U.S. agriculture with yields reported 5-20% above those obtained with continuous corn cultivation (Crookston et al., 1991). Aside from increasing yields, crop rotations have also been proven to help complementary root systems fully exploit water and nutrients (Karlen and Sharpley, 1994). Crop rotations, especially with soybean, also improve nitrogen use and efficiency in agroecosystems, with some studies finding that soybean can both fix atmospheric nitrogen and scavenge residual soil nitrogen, especially on medium and heavy textured soils (Havlin et al., 1990; Olsen et al., 1970). Potassium and total micronutrient concentrations (except calcium and magnesium) were also reported to increase for the 2-year corn-soybean rotation (Copeland & Crookston, 1992).

Aside from influencing soil physicochemical properties and quantity and uptake of nutrients by soil microbiota and plants, crop rotations also affect ecosystem health by reducing pest pressure from certain kinds of pests and insects (Flint & Roberts 1988; Ware, 1980). For example, crop rotations help control weed growth in combination with other treatments like smother crops and mechanical cultivation (Regnier and Janke, 1990), control the growth of most plant parasitic nematodes (Dabney et al. 1988; Ferris, 1967) and have also been suggested as a means of preventing plant diseases (Roder et al., 1988).

Crop rotations (in particular, the corn-soybean rotation that was used in our study), thus have an overall beneficial impact on soil health, although some physical properties are negatively impacted by the soybean crop. Maintenance of soil physical structure provides microhabitats for microbes to colonize and contribute to ecosystem
functions, whereas nutrient uptake is shared between plants and the microbes that thrive in a soil ecosystem. The beneficial as well as detrimental effects of crop rotation are thus important factors influencing the structure and function of the soil microbial community.

1.4 BIOGEOCHEMICAL CYCLES AND SOIL ENZYME ACTIVITY

Biogeochemical cycles are the pathways through which nutrients and minerals get used, recycled or stored in the soil. Various macro- and micronutrients have been identified as crucial to soil and plant health; namely nitrogen, phosphorus, potassium, calcium, magnesium and sulfur (macronutrients), boron, copper, iron, chloride, manganese, molybdenum and zinc (micronutrients). These nutrients are classified as either ‘micro’ or ‘macro’ depending on the concentrations that are considered necessary to support plant growth (Epstein, 1965).

Microorganisms are key players in soil nutrient cycling, performing various functions like SOM decomposition, nitrogen fixation in plant roots, catabolism and degradation of hydrocarbons and phytotoxic compounds, mineralization of nutrients like sulfur and phosphorus, etc. Of all the soil microorganisms, bacteria are 20-30% efficient at recycling carbon, have a high nitrogen content, and a lower carbon content (Hoorman & Islam, 2010). Gram negative bacteria are important for different aspects of nitrogen cycling, and are usually dominant in agroecosystems and prairies (Lloyd et al., 2009). Forest ecosystems, however, are typically dominated by fungi, which are important for the breakdown of recalcitrant organic material and forming symbiotic relationships with plants (Buée et al., 2009). Fungal functions in no-tillage systems
also seem to be more important compared to plow-tillage systems (Hendrix et al., 1986).

Enzymes catalyzing reactions in the various biogeochemical cycles have widely been used to study microbial activity and diversity, in addition to being used as indicators of soil quality (Bowles et al. 2014; Burns et al. 2013; Keeler et al. 2009; Wallenstein et al. 2009; Allison & Vitousek 2005; Huang et al. 2005; Schloter et al. 2003; Kennedy & Papendick 1995; Kanazawa & Filip 1986; Dick 1984; Burns 1982; Nannipieri et al. 1981). Such studies are usually conducted using biochemical assays and measure the production or disappearance of a reagent during an enzymatic reaction. It is misleading to assume that all of the enzyme activity measured in soils comes from living, active microbes. It is thought that 40-60% of enzyme activity can come from stabilized enzymes in the soil, and so enzyme activity is typically a cumulative effect of long-term microbial activity as well as activity of the viable microbial population at the time of sampling (Dick, 1994). Thus, enzyme assays reflect both past and present potential microbial activity and must be viewed as an index, rather than an absolute value (Tabatabai, 1994). Some of the most commonly used enzymes for measurement of soil microbial activity include beta glucosidase, amidase, urease, phosphatase, invertase and sulfatase (Acosta-Martinez & Tabatabai 2000; Bandick & Dick 1999; Miller & Dick 1995; Dick et al. 1988; Eivazi & Tabatabai 1988; Bolton et al. 1985; Frankenberger & Dick 1983; Eivazi & Tabatabai 1977; Pancholy & Rice 1973; Tabatabai & Bremner 1972; Tabatabai & Bremner 1970).

Soil enzyme activities can provide a snapshot of the functional processes occurring in the soil, but it is not always evident which microbes are responsible for
these processes. Microbiological and phylogenetic methods for identification of microbial species, thus, go hand in hand with enzymatic activity measurements, and provide a more detailed view of the soil microbiome’s functioning.

1.5 MEASURING MICROBIAL DIVERSITY IN SOILS

Soils have been shown to have very rich microbial communities, but they also display the least variability across different habitats for species-time relationships. Whether this is a result of true resistance to temporal change or simply an effect of insufficiently large time-scales in studies is open for discussion (Shade et al., 2013).

Traditional methods of measuring microbial activity in the soil involve biomass estimations using techniques such as fumigation-incubation, fumigation-extraction and substrate-induced respiration or SIR and ATP extraction (Sparling et al. 1981; Oades & Jenkinson 1979; Anderson & Domsch 1978). While these methods could provide insight into the abundance of microbial communities at a particular habitat, enzyme assays are used to measure functional properties of the microorganisms in the soil (Taylor et al. 2002; Dick et al. 1996). The major drawback of these approaches is that they are not able to provide stepwise details of all the biochemical processes occurring in the soil’s microbiome, and they are not sensitive enough to taxonomically identify the microbial species performing said functions with a high degree of accuracy.

Microbial diversity can be assessed using plate counts, but this method is restricted by difficulty in dislodging bacteria or spores from soil particles or biofilms, growth medium selection, providing appropriate growth conditions and the inability to culture a large number of bacterial and fungal species with current techniques. In
addition to this, culture plating tends to favor microorganisms with higher growth rates and fungi that produce large number of spores, thus influencing the apparent diversity of the microbial community (Kirk et al., 2004). Thus, culturing as a means of microbial identification has been replaced or substituted by other methods for characterizing microorganisms in environmental samples. This is particularly due to the fact that only about 1-10% of the microorganisms in the soil can be cultured by current methods (Torsvik et al., 1998).

Other biochemical methods to characterize soil microorganisms, including microbial diversity, are fatty acid methyl ester analysis (FAME), community level physiological profiling (CLPP) or sole source carbon utilization patterns (SSCU). These methods use natural variations in different biomolecules as a basis for species identity; FAME differentiates communities based on groupings of signature fatty acids, while CLPP identifies groups of organisms based on the carbon source preferentially utilized (Ibekwe & Kennedy, 1998; Garland & Mills, 1991).

Today, most biochemical methods of microbial biodiversity estimation have been replaced with molecular ones. These techniques focus on classifying microorganisms based on their genome content, and not physiological or biochemical properties. DNA reassociation and hybridization was one of the earliest molecular methods used for species identification and served as the gold standard for later definitions of sequence identity and species delineation (Torsvik et al. 1990a,b). This method is based on the principle that the rate of reassociation of denatured DNA will depend on the similarity of sequences present. Under specific conditions, the half-association value of the DNA can be used as a diversity index.
Examining differences in the guanine and cytosine content of microbial biomass (G+C content) was another method for molecule-based identification of diversity (Nüsslein & Tiedje, 1999). This method provides a coarse level of taxonomic resolution, given that similar taxonomic groups only differ between 3-5% in their G+C content. More recently, DNA-DNA hybridization has served as the basis for the development of DNA microarrays. Microarrays are valuable tools for the detection and identification of bacterial species since a single array chip could contain up to thousands of DNA sequences (Cho & Tiedje, 2001).

With the advent of the polymerase chain reaction or PCR, various techniques to assess microbial diversity and function were developed that involved the cloning or amplification of specific regions of prokaryotic DNA. One of the most popular techniques in this category is denaturing gradient gel electrophoresis or DGGE. The DNA extracted from soil samples is attached to a 35-40 bp GC clamp and separated on a polyacrylamide gel with an increasing concentration of denaturants (Muyzer et al. 1993). This allows the DNA to denature and melt in sequence-specific domains that then migrate differentially through the gel. TGGE uses the same principle as DGGE, except that temperature is used to denature the DNA. These techniques were originally developed to detect point mutations in DNA sequences.

Another tool that utilizes prokaryotic DNA polymorphisms is restriction fragment length polymorphism (RFLP). PCR-amplified ribosomal DNA is digested with a 4-base pair cutting restriction enzyme. Different fragment lengths are detected using agarose or non-denaturing polyacrylamide gel electrophoresis, and these banding patterns can then be used to identify shifts in bacterial community structure (Liu et al., 1997). A variation of RFLP is terminal restriction fragment length
polymorphism (T-RFLP). In T-RFLP, one of the PCR primers is labeled with a fluorescent dye, which then allows detection of only labeled fragments. This allows for a simpler banding pattern as well as better taxonomic resolution, as each labeled band can be thought of as a single operational taxonomic unit or OTU (Tiedje et al., 1999). Similar in principle to RFLP and T-RFLP, ribosomal intergenic spacer analysis (RISA) provides ribosomal-based fingerprinting of the microbial community. Here, PCR is used to amplify the intergenic spacer region between the 16S and 23S regions of the 16S ribosomal subunits. This amplified DNA is denatured and separated using DGGE, and can be used to differentiate even closely related strains of bacterial species (Fisher & Triplett, 1999).

The 16S rRNA gene is perhaps the most popular tool used for taxonomic identification of microorganisms in environmental studies, due to a variety of reasons. The combination of cost-effective sequencing technology and the taxonomic importance of the 16S rRNA gene are responsible for unprecedented progress in the field of soil microbial ecology. 16S rRNA gene-based identification of microbial species is dealt with in further detail in the next section.

Sanger sequencing provides a way to elucidate the base composition of chromosomes and indeed, entire genomes. Two corporations, Roche and Illumina, built upon this premise and produced technologies that offered faster and more reliable sequencing alternatives, using sequencing-by-synthesis. Roche’s 454 pyrosequencing technology uses emulsion-based clonal amplification, whereas Illumina uses fluorescently labelled probes. Both platforms have their respective pros and cons, but have been extensively used in microbial ecology studies. With the introduction of these ‘next-generation’ sequencing technologies, the scientific
community now has access to completely sequenced genomes. Studies can now be conducted based on gene content and not just phenotype (Snel et al., 1999).

1.6 16S rRNA GENE SEQUENCING

Whole genome sequencing can reveal important information about the genome content of an organism, as well as its metabolic functions, but targeted gene sequencing is more often used for analysis of community profiles. One of the most popular target genes for such investigations is the rRNA gene, due to it being highly conserved and functionally homogeneous across species (Woese & Fox, 1977).

The bacterial rRNA genetic locus (rrn) consists of three regions: 16S (1541 nt), 23S (2904 nt) and 5S (120 nt). Of these, the rrs or the 16S region is the most popular one for sequencing analysis. Some studies have suggested that an intermediate region (outside the 16S rRNA region) called the internal transcribed spacer region or ITS may be better for obtaining higher species resolution than 16S rRNA gene sequencing, as it has more variability than the 16S region (White et al., 1990).

One of the biggest questions in post-sequencing data analysis was the definition of a species. Earlier, species were classified according to morphological characteristics observed on growth media, but a multitude of unclassified, uncultivable microorganisms have been found. What basis, then, should be used to define a group of organisms as a species? Initially, DNA-DNA reassociation studies were identified as the best way to reconcile the molecular and phylogenetic properties of an organism. For unique species delineation, at least 70% DNA reassociation (which correlates to around 97% sequence similarity) was considered mandatory.
Later, DNA reassociation studies were replaced with the MLSA (Multi Locus Sequence Analysis) technique, but the sequence similarity standard for species delineation prevailed, with most studies opting for 97% or 98.8% as the standard value (Stackebrandt et al. 2002; Stackebrandt & Goebel 1994). Recently, it has been found that resolving species identities at the 97% level produces around 70% ANI (average nucleotide identity), which is not stringent enough for resolution at very minute scales (Konstantinidis & Tiedje, 2007).

Another interesting feature of this kind of species delineation is that the cut-offs do not account for the ages or evolutionary histories of species populations, they usually match the variations found among strains of traditionally named and characterized organisms (Cohan, 2002). This is a potential pitfall of this technique, especially for environmental populations that contain diverse organisms including Archaea. Criticisms aside, the ANI-method of species assignment is used predominantly in microbial ecology today. Using sequence similarity parameters to define a species leaves us with a new unit of classification, termed by as ‘sequence-discrete clusters’ but better known to us as OTUs or Operational Taxonomic Units (Blaxter et al., 2005).

Different pipelines are currently used for analysis of microbial metagenomic data, the most common of which are MG-RAST or MetaGenomic Rapid Annotation using Subsistent Technology (Meyer et al., 2008), MEGAN or MEtaGenome ANalyzer (Huson et al., 2007), QIIME or Quantitative Insights Into Microbial Ecology (Caporaso et al., 2011) and mothur (Schloss et al., 2009).
1.7 RATIONALE FOR STUDYING COMMUNITY STRUCTURE AND FUNCTION

Species-rich communities render abundant and diverse ecosystem services. Addition of microbial species to species-poor communities increases the probability that high-functioning species will be selected out at the evolutionary scale, where high-functioning species refers to a species that contributes significantly to functioning of the ecosystem (Bell et al., 2005).

As with many other microbial communities, soils are dominated by a small number of highly abundant organisms, but most of the species of interest are not easily observed (Bent & Forney, 2008). One of the biggest problems in characterizing members of the soil microbial community is the presence of these rare taxa; species that are low in abundance but are presumed to fulfill an important ecological niche. These rare taxa could act as reservoirs of activity to rapidly respond to environmental changes (Shade et al., 2014).

One of the suggested workarounds to dealing with this problem of low abundance but functionally important microorganisms is a genomic trait-based evaluation of the community, as opposed to phenotypic evaluation. This is based on the observation that genomic functional potential (as seen in genome size and protein family domains) correlates better with habitat occupancy of the microorganism than phylogeny (Barberán et al., 2014). As with some eukaryotes, prokaryotes also exhibit taxa-area relationships. Similar members of a microbial community tend to be located geographically close to each other (Horner-devine et al., 2004). An interesting point here is that variation in soil metabolic function has also been found to correspond to geography (climate) (Doni et al., 2014; Sun et al., 2013), prompting us to examine the link between microbial community diversity and associated ecosystem functions.
Other studies have established that functional process rates can change with no accompanying detectable change in community composition. As a corollary, differences in microbial community composition may also lead to no change in function, a phenomenon known as functional redundancy (Yin et al., 2000). The explanation postulated for this observation is that measures of whole microbial community composition do not identify groups of organisms of critical importance to the process being measured. It is also possible that shifts in composition may occur that are unrelated to the process being measured (Waldrop & Firestone, 2004). Thus, a more comprehensive approach to measuring environmental processes is needed, one that takes into account the fact that each organism may perform more than one function in the ecosystem, and that changes in community composition need not be correlated with changes in functional processes. It is also important to monitor simultaneously the biodiversity of all groups of organisms, because events that change species diversity in one group of organisms will also affect functions in other groups (Jones & Bradford, 2001).

To provide a specific example for soil microbial communities, a study in 1998 hypothesized that changes in functional groups that degrade macromolecular carbon are more critical to ecosystem processes than those degrading simple carbon because of lower species richness of microbes with enzymatic capacity for macromolecular C degradation (Schimel & Gulledge, 1998). This then begets the question – would species richness or diversity contribute more to increased ecosystem functioning? Different schools of thought exist on this subject, but a recent study in Nature suggests that biodiversity and composition of the soil microbial community change
according to its functioning, at least with respect to soil carbon stability (Mau et al., 2014).

One of the simplest ways to investigate this connection between phylogeny and function, as well as the variation of soil function between land management practices like tillage and crop rotation is to evaluate the variation of soil enzyme levels across these variables. Given that enzymes are the products of gene expression, getting an estimate of gene abundances for various important genes associated with nutrient cycling in the soil would be a feasible way to track the contributions of microorganisms to nutrient processing in the soil. Various databases are available today with information on genetic markers, copy number variation and phylogenetic similarity between different microbes, which provides a good foundation for this task. Here, we use a predictive tool called PICRUST (Phylogenetic Inference of Communities through Reconstruction of Unobserved States) (Langille et al., 2013) to execute this idea. Based on the previous background information about tillage, crop rotation and soil microbe functions, the following hypotheses were formed.

1.8 HYPOTHESES

1. OTU and predicted gene abundances will be higher under the no-till treatment, as compared to plow-till.

2. OTU and predicted gene abundances will be higher under the corn-soybean rotation, as compared to continuous corn.

To test out these hypotheses, the following objectives were formulated.
1.9 OBJECTIVES

a. To analyze the phylogenetic composition of the soil microbiome to identify microbial communities present in the samples

b. To use the functional prediction tool PICRUSt to estimate putative gene abundances from the phylogenetic composition seen in (a)

c. To identify and describe possible OTU-function relationships related to ecosystem services based on PICRUSt predictions

Two studies were carried out in order to fulfill these objectives, and are described in detail in the subsequent chapters. Chapter 2 contains the first study, a pilot-scale experiment designed to use PICRUSt to evaluate functional differences between the two tillage treatments using pyrosequencing. Chapter 3 describes the second study, which included the additional variable of crop rotation, and substantiates upon the results obtained in Chapter 2.
REFERENCES


Stackebrandt, E., & Goebel, B. M. (1994). Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species


CHAPTER 2: PYROSEQUENCING PILOT STUDY

2.1. INTRODUCTION

Sequencing of DNA and RNA has heralded a new era in the identification of microbial communities across different environments. Initial sequencing was done using the technique of Sanger. Sanger sequencing was developed by Frederick Sanger and his team in 1977, and is based on the selective incorporation of dideoxynucleotides during \textit{in vitro} DNA replication (Sanger & Coulson, 1975). The initial drawbacks of Sanger sequencing were improved upon by using radioactive labels or fluorescent labels on the dideoxynucleotides used in the reaction as well as the revolutionary dye-terminator sequencing method, which enabled sequencing of all the bases in a single reaction (as compared to four separate reactions in the earlier version). The greatest challenges with the Sanger technique included poor base quality for the initial part of the read (15-40 bases), sequencing parts of the cloning vector if the samples had been cloned prior to sequencing and insufficient power to resolve large DNA fragments (above 1000 nt).

In contrast to this, next-generation sequencing technologies like pyrosequencing do not rely on the use of a cloning vector for sample preparation. Pyrosequencing, in particular, is based on the ‘sequencing-by-synthesis’ principle, wherein a single DNA strand is used as the template and its complementary strand is synthesized using a DNA polymerase enzyme. The sequencer is able to detect the
complementary base that was added, and thus identify the original base on the template DNA strand. Addition of the correct base triggers a chemical reaction, which is catalyzed at the final step by the chemoluminescent enzyme luciferase to emit a flash of light. This technology was developed in 1996 and was first commercialized by Pyrosequencing AB in Sweden (Ronaghi et al., 1996). Currently, 454 Life Sciences (owned by Roche) is the licensed provider for pyrosequencing services, but support for this technology was discontinued as of 2013.

Pyrosequencing technology was a pioneer in making genome sequencing an integral part of soil microbial community studies. 16S amplicon pyrosequencing has been a very popular technique used in the study of soil microbial communities (Hu et al., 2015; Bates et al., 2011; Nacke et al., 2011; Roesch et al., 2010; Buée et al., 2009; Jones et al., 2009). This has been, in part, due to the popularity of the 16S gene marker for metagenomic studies. Another reason for the popularity of 16S amplicon pyrosequencing studies was that pyrosequencing also offered a substantial improvement upon Sanger sequencing technology, in terms of processing time, number of reads generated, elimination of the need for a cloning vector, amount of sample used for sequencing, sensitivity and ease of automation (Shendure & Ji, 2008).

The pyrosequencing technique was chosen for this pilot study to provide preliminary insight into the microbial diversity in the agricultural soil at the long-term no-till plots in Wooster, Ohio. Another important objective of this study was to evaluate the usefulness of PICRUST, a predictive functional genomics tool (Langille et al., 2013), to investigate the functional properties of these soil microbial communities.
2.2. MATERIALS AND METHODS

2.2.1. Experimental design

This study was conducted on the Van Doren-Triplett long-term no-tillage plots in Wooster, Ohio. These plots are renowned for being the longest continuously maintained no-till plots (51 years at the time of sampling) in the world. Thus, these plots are especially suited to our objective of comparing microbial diversity and functional properties across tillage treatments. These sites have extensive collections of soil microbiological and biochemical data, which makes them extremely useful for functional assessments (Jacinthe et al., 2014; Martens & Dick, 2003; Dick et al., 1991; Dick & Van Doren, 1985; Dick, 1984; Dick, 1983).

These plots are under three different tillage systems (no-tillage (NT), moldboard plow-tillage (PT) and chisel tillage or minimum-till (MT)) and three types of crop rotations (continuous corn, corn-soybean, corn-oats-alfalfa). Of these, two agricultural treatments were selected for sampling, one of which was under plow-till (PT) and the other under no-tillage (NT). Both tillage treatments were under continuous corn cultivation (CC), which enabled us to make valid functional comparisons between tillage treatments alone. The soil at this location was classified as silt loam.

2.2.2. Sample collection and processing

Soil cores were collected in triplicate from each plot and pooled together to make a composite sample. There were three randomized plots for each treatment, and so six composite samples in total for the two treatments. Sampling was done at a
depth of 10 cm, as this is where most of the annual flux of carbon and nutrients occurs. This zone is also thought to contain maximum bacterial diversity (Eilers et al., 2012). DNA was extracted from this soil using the UltraClean® Soil DNA Isolation Kit by Mo Bio Laboratories, Carlsbad, CA. The manufacturer’s instructions were followed step-by-step to obtain genomic DNA. DNA extracted from each technical replicate was then pooled to form a composite DNA sample for each tillage treatment. These DNA samples were then sent to ChunLab, Seoul National University, South Korea for library preparation and sequencing using the 454 Titanium FLX platform (Roche). The V1 – V3 regions of the 16S rRNA gene were targeted during library preparation and the raw data generated by sequencing was supplied as a demultiplexed SFF (Standard Flowgram Format) file for further analysis.

The inclusion of biological replicates in sequencing studies is a current topic of interest, especially in the context of these long-term plots. It has been suggested that dense sampling studies involving a larger number of samples and treatments may be more beneficial than sequencing biological replicates on different runs (Caporaso et al., 2012). In our case, the long-term nature of the NT and PT treatments decrease the odds that any effects observed are due to chance variation, and substantiate the nature of our results. Simply put, the long-term treatments could be regarded as temporal replications of each treatment, as opposed to spatial replications commonly seen in other studies. Phylogenetic analysis of the microbial community at this site reveals a stable community, with many known members of soil microbiota being identified (Sengupta & Dick, 2015). However, increased statistical power and
detection of rare taxa are certainly benefits of replication and these should not be disregarded lightly.

2.2.3. Analysis of 454 Titanium FLX Data

The raw reads from the sequencer were denoised, filtered and passed for downstream analysis using QIIME version 1.8 (Caporaso et al., 2011). QIIME, which stands for Quantitative Insights Into Microbial Ecology, is a widely used bioinformatics tool that combines various open-source programs and scripts to form a robust pipeline for analysis of metagenomic data. The steps used for downstream analysis of sequencing data can be divided into three main categories – denoising of reads, OTU picking and table making, and OTU analysis using PICRUSt. The steps and parameters conducted during each of these stages are described in detail below.

2.2.4. Denoising 454 reads with Denoiser

In preparation for denoising, the raw SFF files supplied by ChunLab were separated into 3 files – a flowgram, a FASTA file and a file containing the quality scores assigned to all the sequences in the FASTA file. QIIME’s “split_libraries.py” script was then used to filter out low-quality reads (defined as reads with Phred score < 25) from this raw file. All sequences with lengths less than 200 bp were also removed. As the files had already been demultiplexed by the sequencing facility, primer and barcode removal were not necessary.

For denoising, the abovementioned files were passed into Denoiser, which was built for rapid denoising of pyrosequencing amplicon data (Reeder & Knight, 2010). This algorithm acts by devising an initial sequence distribution based on read prefixes. It then sorts the prefix clusters in descending order of abundance. Each run
is then denoised using its quality-filtered output from “split_libraries.py” and the flowgram file. All flowgrams without a match in the quality-filtered output file are discarded from the final output. QIIME’s built-in functionality “denoise_wrapper.py” was used to complete this workflow. Once the denoised sequences were obtained, the results were inflated using “inflate_denoiser_output.py” and looped back into QIIME.

2.2.5. OTU (Operational Taxonomic Unit) picking

Open-reference OTU picking was done on the sequences so as to obtain both closed-reference OTUs, which are necessary for PICRUSt analysis, as well as unclassified de novo OTUs (in order to get an estimate of total abundance in the sample). The reference database chosen was the Greengenes database (May 2013 version), which is a curated collection of 16S rRNA sequences (DeSantis et al., 2006). In brief, the OTU picking process occurs through seven steps:

1. Using UCLUST (Edgar, 2010) to cluster sequences against the Greengenes database. This generates clusters of sequences or OTUs using UCLUST’s heuristic approach. These clusters are then compared against the reference sequences in the Greengenes collection at the specified similarity level. If a sequence produces a hit, the cluster is assigned to that GG identifier. The remaining, unclassified sequences are then clustered against each other to produce de novo OTUs. This takes place with prefiltering and iterations to improve the accuracy of the process.

2. Picking a representative set of sequences from each cluster. The most abundant seed from each cluster was chosen for this analysis.

3. Alignment of the representative sequences using PyNAST (Caporaso et al., 2010)
4. Checking for chimeric sequences using ChimeraSlayer (Haas et al., 2011)

5. Assigning taxonomy to the UCLUST-generated OTUs using the Greengenes taxonomic annotations (QIIME uses UCLUST for taxonomy assignment as a default)

6. Construction of the OTU table – this step involved removal of failures from the sequence alignment step (#3) and chimeric sequences identified by ChimeraSlayer (#4) from the final OTU table

7. In addition to these steps, the PyNAST-aligned sequences were filtered and used to construct a phylogenetic tree using the FastTree algorithm (Price et al., 2009).

The above steps were repeated for 4 different sequence similarity thresholds – 91%, 94%, 97% and 99%. This was done in order to assess the effect that sequence similarity levels have on the OTU picking process, especially for a complex environment such as soil. The different similarity thresholds were also used to examine variations in PICRUST’s metagenome and functional predictions.

2.2.6. Predicting community function with PICRUSt

PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) is a bioinformatics software package designed to predict metagenome functional content from 16S rRNA gene surveys (Langille et al., 2013). This algorithm uses a fast Ancestral State Reconstruction method from the APE (Analysis of Phylogenetics and Evolution) R package to predict the gene content of the closest known relative in the reference dataset (Paradis et al., 2004).
PICRUSt utilizes the reference OTUs picked in the previous step for functional analysis. The process required for metagenome and functional predictions is described below in some detail. 16S rRNA copy numbers are predicted using the IMG database (Markowitz et al., 2012) and copy number normalization is done to achieve unbiased estimates of overall diversity. The final predicted metagenome is then generated by multiplying the gene counts associated with each OTU with the abundance of that OTU in that sample. These OTUs are then matched against the KEGG (Kyoto Encyclopedia of Genes and Genomes) database (Kanehisa & Goto, 2000) to predict functional traits associated with those gene families.

From this OTU table, the taxonomic classification can be collapsed down to any hierarchical level and then analyzed at that particular level. Another useful PICRUSt function utilized in this study is its ability to identify the OTUs contributing to specific functions, as implemented by the “metagenome_contributions.py” script. A workflow depicting this analysis is shown in Figs. 2.1 and 2.2 for easier comprehension. PICRUSt enables the user to use either the COG (Clusters of Orthologous Groups of proteins) (Tatusov et al., 2000) or RFAM (RNA family) (Griffiths-Jones et al., 2003) databases as references for functional annotations as well. For this analysis, the KEGG ortholog and pathway predictions were found to give the best predictions.
Figure 2.1 Workflow for library preparation; NT-CC and PT-CC stand for no-till continuous corn and plow-till continuous corn treatments.
2.3. RESULTS AND DISCUSSION

2.3.1. Effect of sequence similarity on OTU picking

This dataset had 7,040 sequences in total (after denoising), with 4,093 sequences derived from the NT sample and 2,947 from the PT sample. As the sequence similarity threshold was increased, it was seen that more sequences were assigned to the final OTU table, and the number of OTUs per sample also showed an upward trend (Fig. 2.3). Of special interest were the OTUs mapped to the Greengenes database (henceforth referred to as reference OTUs), as PICRUSt utilizes only these
hits to perform metagenome prediction. The upward trend was even more pronounced in the case of these reference OTUs, as can be visualized from Fig. 2.4.

![Figure 2.3 Total number of OTUs picked at different sequence similarity levels.](image1)

![Figure 2.4 Total number of reference OTUs picked at different sequence similarity levels.](image2)

These results suggest that an increase in the stringency of the sequence similarity threshold results in finer resolution of OTUs, with overall seed cluster size being smaller but average similarity of each cluster being higher at higher sequence
similarities. This trend has been seen in other studies as well, with higher sequence similarity levels producing OTUs that contain lesser number of sequences overall, or sometimes even singleton OTUs (Poretsky et al., 2014). This trend is not seen in Fig. 2.3, which depicts both de novo and reference OTUs. It is possible that the number of de novo OTUs clustered at different sequence similarities does not vary linearly, and this affects the difference in distribution between Fig.s 2.3 and 2.4.

![Graph showing number of predicted functional relationships with sequence similarity.](image)

Figure 2.5 Increase in number of predicted functional relationships with sequence similarity.

Given these observations, the OTU table resulting from the 99% similarity OTU picking was chosen for further analysis using PICRUSt. The underlying assumption beneath this choice was that an increase in the number of OTUs (coupled with better taxonomic resolution at the more stringent OTU picking threshold) would yield better PICRUSt predictions, especially for a small dataset such as this one.
2.3.2. OTU table summary

At a sequence similarity threshold of 99%, a total of 4,014 sequences were clustered (~57% of the denoised sequence collection). Of these, 2,486 sequences were assigned to the NT and 1,528 to PT. Around 33% of these sequences were identified as belonging to the Greengenes database. This resulted in a total of 506 OTUs (324 from NT, 259 from PT) being identified, with 87 of these belonging to the reference collection.

2.3.3. Sequencing depth and sample diversity

Multiple rarefactions were performed on the OTU table to generate a rarefaction curve, which is depicted in Fig. 2.6. The rarefaction curve is close to reaching saturation, which implies that deeper sequencing would not have yielded a proportional increase in microbial diversity. This might seem confusing, given the number of OTUs, but soil rarefaction curves have sometimes been known to display this behavior (Sengupta & Dick, 2015).

A visualization of the top ten phyla in these samples (Fig. 2.7) show that Acidobacteria and Proteobacteria are most abundant, followed by Bacteroidetes, Actinobacteria and Nitrospirae. These results are consistent with those seen in soil samples sequenced in other recent studies (Navarro-Noya et al., 2013; Peralta et al., 2013; Li et al., 2014).
Figure 2.6 Rarefaction curve of OTUs generated with pyrosequencing depth; the top (i.e blue dashed) line depicts the No-Till treatment and the bottom (i.e red dotted) line stands for the Plow-Till treatment.

Figure 2.7 Top ten phyla in no-till and plow-till soil samples, ranked in order of abundance.
In terms of phylum-level OTU diversity, the NT sample is almost always higher than PT (except in the case of some *Proteobacteria*, *Gemmatimonadetes* and *Chlorobi*). This is corroborated by the alpha diversity indices shown in Table 2.1, which indicate that the species richness is skewed in favor of the no-till treatment. Bacterial diversity is generally thought to be higher under no-tillage than plow tillage (Lupwayi et al., 2012; van Capelle et al., 2012; Lupwayi et al., 2007; Adl et al., 2006), so these results are consistent with some earlier reports. However, there are studies that have found no changes between the tillage treatments (Hartmann et al., 2014; Jiang et al., 2011). Thus, there is still no clear consensus on the overall effect of tillage on bacterial diversity, but these results correspond with the studies showing an increase in diversity caused by no-tillage. The rationale behind these observations is that no-tillage prevents the breakdown of soil physical structure and properties (as would be the case under regular tillage), as well as protects the unique microhabitats necessary for microorganisms to thrive (Hill, 1990). Increased concentrations of nutrients like nitrogen, phosphorus and sulfur under NT are also related to higher microbial survival under this treatment (Tracy et al., 1990b; Dick, 1983). In this particular case, the higher amount of crop residue on the surface of the NT plots could also provide a substrate for the microbes to grow, especially just below the soil surface (Doran, 1979).

<table>
<thead>
<tr>
<th>Diversity index</th>
<th>No-Till</th>
<th>Plow-Till</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chao1</td>
<td>333.58</td>
<td>274.55</td>
</tr>
<tr>
<td>ACE</td>
<td>353.18</td>
<td>296.92</td>
</tr>
<tr>
<td>Shannon</td>
<td>5.09</td>
<td>4.88</td>
</tr>
<tr>
<td>Simpson</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>Fisher</td>
<td>99.45</td>
<td>89.47</td>
</tr>
</tbody>
</table>

Table 2.1 Alpha diversity indices for soil samples under different tillage treatments.
2.3.4. **PICRUSt predictions**

The OTU table generated at 99% sequence similarity was processed using PICRUSt to yield information on putative functions performed by microorganisms across the two different treatments. Using the KEGG database, these functions were classified into various categories such as energy metabolism, amino acid metabolism, hydrocarbon degradation, transcription and translation, replication and repair, membrane transport, signal transduction, cell motility, etc.

The most abundant of these predicted functions were mapped to purine and pyrimidine metabolism, metabolism of cofactors and vitamins, lipid and glycan metabolism, biosynthesis and degradation of secondary metabolites, amino acid metabolism, and carbohydrate metabolism. All the abovementioned functions are related to maintenance of cell function and structure and so, would be performed by every species in the community. Of special interest, however, are certain highly specific functions related to soil health and crop productivity. These functions were identified mainly from Jones & Bradford (2001) and are listed in Table 2.2.

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycling of nutrients</td>
<td>Nitrogen metabolism</td>
</tr>
<tr>
<td>Cycling of nutrients</td>
<td>Sulfur metabolism</td>
</tr>
<tr>
<td>Sequestration of carbon</td>
<td>Carbon fixation by prokaryotic pathways</td>
</tr>
<tr>
<td>Production and consumption of trace gases</td>
<td>Methane metabolism (also implicated in sequestration of carbon)</td>
</tr>
<tr>
<td>Degradation of soil, water and air pollutants</td>
<td>Hydrocarbon degradation</td>
</tr>
</tbody>
</table>

Table 2.2 KEGG functions taken into consideration for assessment of soil biodiversity, as modified from Jones & Bradford, 2001.

PICRUSt’s NSTI (Nearest Sequenced Taxon Index) provides a reliable way to estimate the confidence of these predictions. NSTI is the sum of phylogenetic
distances for each organism in the OTU table to its nearest relatives with a sequenced reference genome, measured in terms of substitutions per site in the 16S rRNA gene and weighted by the frequency of that organism in the OTU table (Langille et al., 2013). For these soil samples, PICRUSt predicted an NSTI value of 0.17 for NT and 0.15 PT, which is similar to the value obtained for the soil test datasets in Langille et al. While these values are not ideal, they are to be expected given the yet undiscovered portion of the soil microbiome. A closer look at the OTU table generated by QIIME indicates that a substantial portion of the OTUs was unclassified, and this is reflected in the PICRUSt functional predictions as well. Around 15% of the reference OTUs were not mapped to any protein/function. For now, the NSTI value serves as a benchmark for the confidence and accuracy of PICRUSt predictions. For all predicted functions in this dataset, the no-till system was found to be at least twice as functionally enriched as the plow-till system (Fig. 2.8). Although reports of bacterial diversity across different tillage systems can be ambiguous, the functional potential of NT soil has always been estimated to be higher, as quantified by enzyme activity and soil organic carbon measurements (Mathew et al., 2012; Green et al., 2007; Doran et al., 1998; Dick, 1984; Dick, 1983). KEGG function predictions in our samples only validate these previously established theories.
The different functional categories listed in Table 2.2 are explained in more detail below. It is worth noting that significant differentiation was observed between the no-till and plow-till treatment for four of these functional categories (methane metabolism was not significant), as indicated by Welch's t-test (p<0.05). As PICRUSt results are predictive, they could be biased due to the sequencing depth as well as current taxonomic classification limitations. This point will be further explored in the next chapter, which deals with Illumina MiSeq data that possesses better read quality and number.

2.3.5. Nitrogen metabolism

This is an area of special interest, given that one of the established benefits of no-tillage is increased nitrogen uptake by plants and more nitrogen fixation by the belowground microbiota (Soon & Arshad, 2005; Thiagalingam et al., 1991). Nitrogen fixation and metabolism is also important in the context of greenhouse gas emissions,
given that nitrous oxide (N$_2$O) is one of the most influential greenhouse gases, accounting for 5% of all GHG emissions from human activities (EPA, 2010).

The nitrogen cycle is the process by which uptake and conversion of atmospheric and soil nitrogen by microorganisms and plants takes place. This is depicted in Fig 2.9. Identification of the genes involved in this cycle and their abundances in any soil sample is key to estimating the nitrogen fixing potential of the soil microbiome. Using PICRUSt, it was possible to get a glimpse of the presence or absence of genes that could potentially be involved in various steps of the nitrogen cycle, as well as their absolute and relative abundances. Linking this information to the Greengenes database also provides a KEGG function-OTU relationship, wherein the OTUs contributing to that particular function can be identified.

![Figure 2.9 Genes involved in nitrogen metabolism through the nitrogen cycle (Credit: Jennifer Bowen, University of Massachusetts, Boston).]
Forty-two enzymes from the KEGG database were identified as involved in nitrogen metabolism. A summary of these enzymes is presented in Table 2.3, with information on the genes associated with the enzyme, overall gene abundance, rank based on differentiation between treatments and OTUs that were mapped to this function.

Overall, the catalytic subunit of assimilatory nitrate reductase, periplasmic nitrate reductase (NapA protein), cytochrome c-type protein (NapB), formamidase, and carbamate kinase were the rare few enzymes that were enriched by tillage. Most enzymes involved in nitrogen metabolism appear to be preferentially enriched under no-till, and the most differentiated enzymes (as inferred from gene abundance ratios between treatments) are depicted in Fig. 2.10.

![Figure 2.10 Nitrogen metabolism enzymes with highest no-till:plow-till gene abundance ratios. Percentages indicate contribution to the pie chart, not contribution to the overall category.](image-url)
<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Function</th>
<th>Gene</th>
<th>Total gene count</th>
<th>Rank</th>
<th>Number of OTUs</th>
<th>Most abundant OTU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbonic anhydrase</td>
<td>Cyanate degradation</td>
<td><em>cynT, can</em></td>
<td>1384</td>
<td>8</td>
<td>54</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>544</td>
<td></td>
<td>41</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td>Glutamine synthetase</td>
<td>Ammonia degradation</td>
<td><em>glnA</em></td>
<td>1282</td>
<td>15</td>
<td>61</td>
<td>Ellin329</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>578</td>
<td></td>
<td>44</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td>Glutamate synthase small chain</td>
<td>Glutamate formation</td>
<td><em>gltD</em></td>
<td>1219</td>
<td>16</td>
<td>59</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>558</td>
<td></td>
<td>43</td>
<td>Candidatus Solibacter</td>
</tr>
<tr>
<td>Glutamate synthase large chain</td>
<td>Glutamate formation</td>
<td><em>gltB</em></td>
<td>1123</td>
<td>13</td>
<td>59</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>462</td>
<td></td>
<td>47</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td>Nitronate monooxygenase</td>
<td>Nitrite synthesis</td>
<td><em>ncd2, npd</em></td>
<td>845</td>
<td>14</td>
<td>44</td>
<td>Koribacteraceae</td>
</tr>
</tbody>
</table>

Continued

Table 2.3 Summary of KEGG genes and Greengenes OTUs involved in nitrogen metabolism; rank is based on order of No-Till:Plow-Till ratios. Upper values represent the NT sample and lower values represent the PT sample.
Table 2.3 continued

<table>
<thead>
<tr>
<th>Nitronate monooxygenase</th>
<th>Nitrite synthesis</th>
<th>ncd2, npd</th>
<th>369</th>
<th>31</th>
<th>Bradyrhizobium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nitrilase</td>
<td>Ammonia formation</td>
<td></td>
<td>946</td>
<td>1</td>
<td>11 Rhodoplanes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>49</td>
<td>10</td>
<td>Sinobacteraceae</td>
</tr>
<tr>
<td>Nitrite reductase small subunit</td>
<td>Nitrate to ammonia conversion</td>
<td>nirD</td>
<td>416</td>
<td>12</td>
<td>34 Koribacteraceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>169</td>
<td>26</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td>Glutamate dehydrogenase (NAD(P)+)</td>
<td>Glutamate and ammonia formation</td>
<td>gdhA</td>
<td>412</td>
<td>6</td>
<td>40 Koribacteraceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>152</td>
<td>27</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td>Ferredoxin-nitrite reductase</td>
<td>Nitrate to ammonia conversion</td>
<td>nirA</td>
<td>389</td>
<td>4</td>
<td>22 Koribacteraceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>124</td>
<td>10</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td>Nitrogenase molybdenum-iron protein alpha chain</td>
<td>Nitrate to ammonia conversion</td>
<td>nifD</td>
<td>157</td>
<td>21</td>
<td>17 Ellin329</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>96</td>
<td>15</td>
<td>Nitrospira</td>
</tr>
</tbody>
</table>
Interestingly, the most differentially abundant enzymes in this functional category are all related to the process of denitrification and nitrate reduction (both assimilatory and dissimilatory). Thus, the processes of nitrogen and ammonia formation are obviously abundant and dynamic processes that are enriched by no-tillage. It is important, however, to account for the roles of individual species in ammonia metabolism, given that the activity of this process is apt to vary even between classes of the same group of bacteria (Nyerges & Stein, 2009).

Nitrogen fixation (conversion of nitrogen to ammonia) activity was inferred from the presence of the nitrogenase iron protein (\textit{nifH}), and the nitrogenase molybdenum-iron protein (alpha and beta chains; \textit{nifD} and \textit{nifK}). These genes were differentially enriched to almost the same level and had very similar gene abundances, possibly influenced by the fact that they belong to the same gene family.

The process of nitrification (ammonia to nitrite) is catalysed by the \textit{amo} genes, all of which were present in these samples in very low abundance. They also did not show any variation in activity between the treatments and were almost exclusively mapped to the \textit{Nitrosomonadaceae} family. The PT sample of the \textit{amoC} gene showed some variation, with \textit{Bacillus sp.} and \textit{Bradyrhizobium sp.} also contributing to its nitrifying activity. This correspondence of the \textit{amo} genes to ammonia-oxidizing species like \textit{Nitrosomonadaceae} and \textit{Bradyrhizobium} is corroborated by previous findings (Francis et al., 2003; Dionisi et al., 2002; Purkhold et al., 2000; Rotthauwe et al., 1997).

The process of nitrate reduction can be assimilatory or dissimilatory, depending on whether the environment is aerobic or anaerobic and the organisms carrying out the process. Bacteria are capable of both assimilatory and dissimilatory
nitrate reduction, which can be seen from these results as well. Assimilatory or aerobic nitrate reduction activity, was observed from the assimilatory nitrate reductase catalytic subunit (nasA gene, interestingly enriched under the PT treatment) and ferredoxin-nitrite reductase (nirD). The dissimilatory or anaerobic process seemed to be much more common in these samples, and closely coupled to the overall denitrification process (nitrate to nitrogen conversion). It is worth noting that the nitrilase enzyme (nit), which is involved in the production of ammonia, had the highest gene abundance in nitrogen metabolism.

The NapA and NapB proteins (nitrate reductases), the delta and gamma subunits of the nitrate reductase enzymes (narI and narJ), nitrite reductase (nirB, nirD, nrfA), and nitrate reductase (narG, narH) were all implicated in the denitrification process through the dissimilatory mechanism. With the exception of the nasA and amoA genes, all these genes were preferentially enriched under no-tillage.

It seems that nitrogen fixation and denitrification are performed by a mix of OTUs, the most prominent of which belong to class Alphaproteobacteria. Some species of Nitrospirae and Verrucomicrobia also appear to assist in these processes, with the nitrite reductases being linked to a rare OTU, classified as family Chthoniobacteraceae, genus DA101. All of these classes have been linked to different aspects of nitrogen fixation or nitrite reduction in soil and other ecosystems (Wertz et al., 2012; Swanner & Templeton, 2011; Chen et al., 2003).

To complete the nitrogen cycle, subunit C of nitric oxide reductase (norC), nitrous-oxide reductase (nosZ), and nitrite reductase (nirK) are highly abundant enzymes that perform denitrification and cause nitrogen to be emitted from the soil.
Each of these enzymes relates to a single OTU that seems to perform almost all of the predicted enzymatic activity across both treatments; for instance, $\text{norC}$ is linked almost exclusively with classes $\text{Beta}$- and $\text{Gamma}$-proteobacteria, $\text{nosZ}$ with $\text{Betaproteobacteria}$ and $\text{Bacteroidetes}$ (the $\text{Chitinophagaceae}$ family) and $\text{nirK}$ with $\text{Deltaproteobacteria}$ (the $\text{Myxococcales}$ order).

An interesting class of proteins, the nodulation proteins, was also predicted from this OTU collection. The $\text{nodA}$ gene, which codes for an acyltransferase, and $\text{nodB}$, which produces a chitoligosaccharide deacetylase were both present in these samples, and showed no differences between the NT and PT treatments. The $\text{nod}$ gene, which translates to a transcriptional regulator, showed slightly higher occurrence under the PT sample. It is not possible to make any conclusions from this, however, given that this is a small sample size and the predicted gene abundances are smaller than average. As expected from previous studies, only rhizosphere-associated bacteria ($\text{Rhodoplanes sp.}$ and $\text{Bradyrhizobium sp.}$ from $\text{Alphaproteobacteria}$) were associated with nodulation activity (Lafay & Burdon, 2006; Young & Haukka, 1996).

Most nitrogen metabolism enzymes, thus, may not show high gene abundances overall but are expressed by specific members of the soil microbial community. Disturbing this balance of species could prove adversarial to the nitrogen fixing potential of soil.

2.3.6. Sulfur metabolism

Most of the sulfur in soils is present in the form of sulfate esters or sulphonates rather than in the elemental form. The rate of sulfur cycling is related to the microbial community present in the soil and its metabolic activity. Microbial
processes that control the movement of sulfur and other related compounds through the soil are also key influencers of plant sulfur intake (Kertesz & Mirleau, 2004). A simplified representation of the sulfur cycle in soil is shown in Fig. 2.11.

Figure 2.11 Sulfur cycle (sourced from http://textbookofbacteriology.net).
Most of the reactions involving organic sulfur molecules are preferentially enriched under the no-till treatment (Fig. 2.12). Studies have found sulfur mineralization to be higher in no-till soils than plow-tilled ones (Tracy et al., 1990a; Tracy et al., 1990b). This could partly be due to the higher dry deposition of soil sulfur under no-tillage conditions, which is an indirect effect of tillage (Morra & Dick, 1988). This could, hence, translate to a higher functional activity for the sulfur cycle genes under no-till treatment.

A total of forty-six enzymes in this functional category were predicted from our samples, and the most abundant of these are shown in Table 2.4. Unlike nitrogen metabolism, certain enzymes involved in sulfur metabolism were found in only NT or PT soils. For instance, ThiS adenyltransferase (*thiF*) was only found in the plow-tilled
soil sample (WS3, *Gemmatimonadetes*), whereas the PAPSS enzyme (PAPSS1, PAPSS2) (*Myxococcales*), tetrathionate reductase (*trrB*) and S-ribosylhomocysteine lyase (*luxS*) were only found in the no-till system. The PAPSS1, PAPSS2 and *trrB* genes were mapped to OTUs containing order *Myxococcales*, while predicted *luxS* activity was performed by the relatively low-functioning *Bacillus* sp. The NT-only *thiF* activity was restricted to phyla WS2 and *Gemmatimonadetes*.

Preferential enrichment for plow-tillage was found in cystathione beta-lyase (*patB*) and D-cysteine desulfhydrase (*dcyD*). Both these KEGG enzymes were associated with a mix of OTUs across both treatments. This, and the fact that PICRUST predictions are performed using normalized OTU tables, makes it safe to rule out this out as a singularity.
<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Function</th>
<th>Gene</th>
<th>Total gene count</th>
<th>Rank</th>
<th>Number of OTUs</th>
<th>Most abundant OTU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thioredoxin reductase (NADPH)</td>
<td>Formation of reduced disulfide bonds</td>
<td>trxB</td>
<td>1878</td>
<td>14</td>
<td>60</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>751</td>
<td></td>
<td>43</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td>Cysteine synthase A</td>
<td>Sulfide degradation</td>
<td>cysK</td>
<td>1101</td>
<td>15</td>
<td>56</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>451</td>
<td></td>
<td>40</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td>Cysteine desulfurase</td>
<td>Sulfur transfer</td>
<td>iscS</td>
<td>1044</td>
<td>18</td>
<td>57</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>432</td>
<td></td>
<td>42</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td>5-methyltetrahydrofolate-homocysteine methyltransferase</td>
<td>Cysteine degradation</td>
<td>metH</td>
<td>973</td>
<td>13</td>
<td>56</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>384</td>
<td></td>
<td>42</td>
<td>Candidatus Solibacter</td>
</tr>
<tr>
<td>Thiosulfate/3-mercaptopuruvate sulfurtransferase</td>
<td>Sulfite formation</td>
<td>sseA</td>
<td>866</td>
<td>31</td>
<td>52</td>
<td>Ellin329</td>
</tr>
</tbody>
</table>

Table 2.4 Summary of KEGG genes and Greengenes OTUs involved in sulfur metabolism; rank is based on order of No-Till:Plow-Till ratios. Upper values represent the NT sample and lower values represent the PT sample.
<table>
<thead>
<tr>
<th>Table 2.4 continued</th>
<th>Thiosulfate/3-mercaptopryruvate sulfurtransferase</th>
<th>Sulfite formation</th>
<th>sseA</th>
<th>481</th>
<th>36</th>
<th><em>Bradyrhizobium</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Serine O-acetyltransferase</td>
<td>Cysteine formation</td>
<td>cysE</td>
<td>627</td>
<td>29</td>
<td>58</td>
<td><em>Koribacteraceae</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>326</td>
<td></td>
<td>42</td>
<td><em>Bradyrhizobium</em></td>
</tr>
<tr>
<td>Sulfate transport system permease protein</td>
<td>Extracellular sulfate transport</td>
<td>cysU</td>
<td>662</td>
<td>17</td>
<td>52</td>
<td><em>Koribacteraceae</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>272</td>
<td></td>
<td>35</td>
<td><em>Nitrospira</em></td>
</tr>
<tr>
<td>5-methyltetrahydropteroyl triglutamate-homocysteine methyltransferase</td>
<td>Cysteine degradation</td>
<td>metE</td>
<td>574</td>
<td>23</td>
<td>45</td>
<td><em>Koribacteraceae</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>256</td>
<td></td>
<td>27</td>
<td><em>Koribacteraceae</em></td>
</tr>
<tr>
<td>Sulfate transport system substrate-binding protein</td>
<td>Extracellular sulfate transport</td>
<td>cysP, sbp</td>
<td>577</td>
<td>22</td>
<td>51</td>
<td><em>Koribacteraceae</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>241</td>
<td></td>
<td>35</td>
<td><em>Koribacteraceae</em></td>
</tr>
<tr>
<td>Sulfate adenylyltransferase subunit 2</td>
<td>Sulfate formation</td>
<td>cysD</td>
<td>527</td>
<td>24</td>
<td>53</td>
<td><em>Koribacteraceae</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>241</td>
<td></td>
<td>42</td>
<td><em>Koribacteraceae</em></td>
</tr>
</tbody>
</table>
There are two major groups of reactions in sulfur metabolism, which involve the formation and reduction of sulfates (Fig. 2.11). In these samples, both assimilatory and dissimilatory sulfate reduction are seen to occur equally. Most of the KEGG enzymes in this category are involved in the conversion of sulfate to H₂S or sulfides that may eventually be used to build microbial biomass. Sulfate adenyltransferase (sat/met3, cysN, cysD), phosphoadenosine phosphosulfate reductase (cysH), adenylsulfate kinase (cysC), the sulfite reductase hemoprotein (cysI), and the CysC/CysN bifunctional enzyme (cysCN) are all mapped to assimilatory processes, where the sulfate molecule is taken inside the microorganism.

The different genes associated with sulfate adenyltransferase enzyme are all associated with a range of OTUs, with a preference for the *Proteobacteria* phylum, although this could also be related to the fact that the *Proteobacteria* are some of the most abundant OTUs in this dataset. The same can be said about cysH, cysC, and cysCN, all of which appear to be high-abundance enzymes that can be coded for by a major portion of the microbial community observed in these samples. cysI, however, is linked specifically to *Proteobacteria* (all classes) and the *Chitinophagaceae* family from phylum *Bacteroidetes*, with minor contribution from *Bacillus sp.* and *Opitutus sp.* as well. As far as the assimilatory mechanism goes, most OTUs in this sample set seem to have high sulfate-reducing capacity. Also, depending on the accuracy of these predictions, it is possible that even if certain microbial species in the soil are affected, sulfate-reducing activity could still persist with minor changes.

From the sulfate reduction genes involved in the assimilatory mechanism, the sulfate adenyltransferase genes alone seem to participate in sulfate reducing reactions outside the microbe through the dissimilatory mechanism, along with adenylsulfate
reductase (aprA, aprB) and sulfite reductase (dsrA, dsrB). Nitrospira sp. and family 0319-6A21 from the Nitrospirae phylum seem primarily responsible for the expression of the aprA and aprB genes, and the same applies for the dsrA and dsrB genes. This is a remarkable occurrence, where most of the contribution for a single function comes from a specific OTU, which is by itself not typically associated with that class of enzymes. However, further validation (by enzyme measurements or RT-PCR quantification targeting these genes) is required before this link can be established.

The enzymes catalyzing the various processes in sulfur metabolism are among the most abundant ones predicted by PICRUSt. It is also evident from this brief description that these activities are not easily generalized, and need to be carefully considered on a case-by-case basis.

2.3.7. Hydrocarbon degradation

Hydrocarbons are organic compounds that mimic natural compounds found in the soil, but which do not belong in the soil. These compounds can often be toxic to plants and small animals and insects living in the soil. Degradation of hydrocarbon compounds and naturally occurring large-chain aromatic compounds is an important function performed by microorganisms that keeps the soil free of toxic compounds and contaminants. Categories of hydrocarbons taken under consideration in this category included benzoates (including amino- and fluorozenoates), chloroalkanes and chloroalkenes, chlorocyclohexanes and chlorobenzenes, toluene, nitrotoluene, ethylbenzene, styrenes, atrazine, caprolactam, DDT, polycyclic aromatic hydrocarbons (PAHs), and steroids.
Some of the most abundant enzymes predicted in this functional category are atrazine chlorohydrolase, phenylacetaldehyde dehydrogenase, benzoate-CoA ligase and pentachlorophenol monooxygenase, all of which are involved in the processing and degradation of aromatic compounds and herbicides like atrazine. This category, though preferentially enriched for NT soil as well, showed a higher incidence of functional genes that are more abundant under the PT treatment (Fig. 2.13 and Table 2.5).

Figure 2.13 Hydrocarbon degradation enzymes with highest no-till:plow-till gene abundance ratios. Percentages indicate contribution to the pie chart, not contribution to the overall category.

Most genes associated with higher functional potential in the PT plots are involved in degradation of aromatic compounds, or more specifically,
aminobenzoates. Mandelate racemase (*mdlA*), vanillate O-demethylase (*vanA, vanB*), benzyolformate decarboxylase (*tpmT*) and phenol 2-monoxygenase (PHH) are all examples of such enzymes.

A total of 141 enzymes related to hydrocarbon degradation were predicted by PICRUSt. Analysis of the predictions in this category was done based on the class of compounds they degraded. Thus, these hydrocarbon-degrading enzymes were grouped into 5 categories: benzoate-degrading, aminobenzoate-degrading, chlorohydrocarbon-degrading, atrazine-degrading, and PAH-degrading. Each of these compounds is toxic to the environment in a unique way. Benzoates and aminobenzoates are developmental and reproductive toxins, while also being hazardous air pollutants. Chloroalkanes and chloroalkenes are carcinogenic, water pollutants and bioaccumulative, while PAHs are considered carcinogenic, mutagenic and teratogenic. Atrazine occupies a special category of its own, given that Ohio is one of the top states contributing to atrazine use in the United States (per unit area). Atrazine is most commonly used on corn crops, and is the most widely used herbicide currently in use in the U.S. It has, however, shown signs of potential hormone effects in laboratory animals with short term exposure. Atrazine levels in water are currently regulated at 3.0 ppb under the Safe Drinking Water Act (EPA, 2007).

Benzoate-degrading enzymes were the most numerous in this category. These enzymes are also involved in the degradation of fluorobenzoate molecules. 4-hydroxybenzoyl-CoA thioesterase (*ybgC*), protocatechuate 3,4-dioxygenase (*pcaG, pcaH*), p-hydroxybenzoate 3-monoxygenase (*pobA*), benzoate-CoA ligase (*badA*), and 2,3-dihydroxybenzoate decarboxylase (DHBD) are a few of the important enzymes catalyzing the oxidative degradation of benzoate molecules.
Aminobenzoate-degrading enzymes were sometimes distinct from those degrading benzoates/flurobenzoates, and thus, were placed in a category of their own. Many of these enzymes were differentially enriched under the PT treatment compared to the NT treatment, like vanillate O-demethylase monooxygenase (vanA, vanB), benzoyl formate decarboxylase (mdlC), and mandelate racemase (mdlA). Other, NT-enriched aminobenzoate degraders are 2-aminobenzoate-CoA ligase (abmG) and anthraniloyl-CoA monooxygenase (namA).

The trend of strong functional linkage to OTUs under Proteobacteria continues in these enzymes too - vanA and vanB, mdlC (Rhodoplanes sp. and Myxococcales), mdlC (Alphaproteobacteria – Rhodoplanes sp., Bradyrhizobium sp. and Rhodospirillaceae), abmG (Alphaproteobacteria, Janthinobacterium sp.) and namA (Alpha and Deltaproteobacteria, Gaiellaceae from Actinobacteria). Thus, it can be seen that benzoate degradation is probably mainly performed by the Proteobacteria members. Previous literature has linked benzoate degradation with members of the Alpha- and Betaproteobacteria, specifically Thauera aromatic and Rhodopseudomonas palustris (Carmona et al., 2009; Schüle et al., 2003; Egland et al., 1995). Thus, it is reasonable to investigate the closely related genotypes of these members for benzoate degradation activity, keeping in mind the restriction imposed by the range of OTUs identified in our particular samples.
<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Function</th>
<th>Gene</th>
<th>Total gene count</th>
<th>Rank</th>
<th>Number of OTUs</th>
<th>Most abundant OTU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enoyl-CoA hydratase</td>
<td>Benzoate and caprolactam degradation</td>
<td>paaF, echA</td>
<td>2146</td>
<td>76</td>
<td>52</td>
<td>Rhodoplanes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1337</td>
<td>36</td>
<td></td>
<td>Bradyrhizobium</td>
</tr>
<tr>
<td>Acetyl-CoA C-acetyltransferase</td>
<td>Benzoate degradation</td>
<td>atoB</td>
<td>1118</td>
<td>74</td>
<td>59</td>
<td>Ellin329</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>671</td>
<td>42</td>
<td></td>
<td>Bradyrhizobium</td>
</tr>
<tr>
<td>Amidase</td>
<td>Aminobenzoate and styrene degradation</td>
<td>amiE</td>
<td>1009</td>
<td>66</td>
<td>49</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>515</td>
<td>34</td>
<td></td>
<td>Bradyrhizobium</td>
</tr>
<tr>
<td>Carboxymethyle-nebutenolidase</td>
<td>Chlorobenzene, toluene and fluorobenzoate degradation</td>
<td></td>
<td>1829</td>
<td>49</td>
<td>51</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>469</td>
<td>38</td>
<td></td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td>Aldehyde dehydrogenase (NAD+)</td>
<td>Limonene and pinene degradation</td>
<td></td>
<td>912</td>
<td>69</td>
<td>57</td>
<td>Ellin329</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>473</td>
<td>42</td>
<td></td>
<td>Rhodoplanes</td>
</tr>
</tbody>
</table>

Table 2.5 Summary of KEGG genes and Greengenes OTUs involved in hydrocarbon degradation; rank is based on No-Till:Plow-Till ratios. Upper values represent the NT sample and lower values represent the PT sample.
<table>
<thead>
<tr>
<th>3-hydroxybutyryl-CoA dehydrogenase</th>
<th>Benzoate degradation</th>
<th>paaH, hbd, fadB, mmgB</th>
<th>949</th>
<th>40</th>
<th>47</th>
<th>Koribacteraceae</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acid phosphatase</td>
<td>Aminobenzoate degradation</td>
<td>pho</td>
<td>810</td>
<td>29</td>
<td>34</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td>Alcohol dehydrogenase</td>
<td>Caprolactam degradation</td>
<td>adh</td>
<td>570</td>
<td>60</td>
<td>48</td>
<td>Koribacteraceae</td>
</tr>
<tr>
<td>Gluconolactonase</td>
<td>Caprolactam degradation</td>
<td>gn1</td>
<td>549</td>
<td>54</td>
<td>39</td>
<td>Rhodoplanes</td>
</tr>
<tr>
<td>Alcohol dehydrogenase (cytochrome c)</td>
<td>Chloroalkane and chloroalkene degradation</td>
<td>exaA</td>
<td>559</td>
<td>44</td>
<td>33</td>
<td>Ellin6513</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>236</td>
<td>24</td>
<td></td>
<td>Bradyrhizobium</td>
</tr>
</tbody>
</table>
Chloroalkanes and chloroalkenes are among the most prevalent hydrocarbon hydrocarbons in the environment, and thus, they have a higher number of enzymes associated with their degradation. Carboxymethylenebutenolidase (CMBL), aldehyde dehydrogenase (ALDH), alcohol dehydrogenase (adhP) (these enzymes are also involved in naphthalene degradation), acetaldehyde dehydrogenase (adhE), catechol 2,3-dioxygenase (dmpB), catechol 1,2-dioxygenase (catA), haloalkane dehalogenase (dhaA), haloacetate dehalogenase (dehH), 2,4-dichlorophenoxyacetate dioxygenase (tfda), and biphenyl-2,3-diol 1,2-dioxygenase (bphC) were some of the more prominent enzymes involved in chloro-hydrocarbon degradation.

CMBL, adhP, and dhaA did not show particular affinity for any microbial community, while ALDH was primarily predicted as being linked to Rhodoplanes sp., and the Chitinophagaceae and Rhodospirillaceae families, which appear together in a number of hydrocarbon-degrading functions. adhE was linked specifically to three members of the Acidobacteria family, the Koribacteraceae, Acidobacteraceae and the Solibacteraceae families. Of the catechol degrading genes, catA was linked exclusively to Alphaproteobacteria (as were dehH and tfda), while dmpB was putatively expressed by Alpha- (Rhizobiales) and Betaproteobacteria. Lastly, bphC was a low-abundance, PT-enriched gene that was expressed by Bacillus sp. in NT and Bradyrhizobium sp. under PT.

Atrazine-degrading or mineralizing soil bacteria were first reported in 1995 (Radosevich et al., 1995). Since then, microbes from phyla Actinobacteria (Arthrobacter sp., Nocardioides sp.) and Proteobacteria (Rhizobium leguminosarum) have been reported to contain atrazine-degrading genes (Vibber et
al., 2007; Smith & Crowley, 2006; Piutti et al., 2003). Allophanate hydrolase (DUR1 enzyme), hydroxyatrazine ethylaminohydrolase (K03382, atzB), atrazine chlorohydrolase (atzA), and urease (ureB, ureA) were the atrazine-degradation related enzymes found in our samples. All of these genes were linked to OTUs classified as Alphal/Betaproteobacteria (Rhodoplanes sp., Hyphomicrobium sp., Limnohabitans sp.), with some Actinobacteria (families Solirubrobacterales and Gaiellales) in the case of atzB and Verrucomicrobia (Opitutus sp.) with the urease genes.

Finally, salicylate hydroxylase (NHG) (also implicated in dioxin and naphthalene degradation), 4,5-dihydroxyphthalate decarboxylase (pht5) and protocatechuate 3,4-dioxygenase (pcaH, pcaG) were identified as key enzymes in the degradation of PAHs. All these genes mapped primarily to the Alphaproteobacteria, with the exception of the Chitinophagaceae family linked to pcaH alone.

There are traces of other enzymes predicted in the degradation of hydrocarbons such as styrene, bisphenol, xylene, dioxin, nitrotoluene, etc. but these are not discussed in detail here, as the OTU-function link was either too ambiguous or the predicted gene abundance was on the lower end of the spectrum.

2.3.8. Carbon fixation

This is one of the most active areas of research in land management, given its enormous significance in strategies that address climate change. Soil microbes are currently viewed as one of the most promising strategies for improving the availability and storage of carbon in the soil, and so, it is important to narrow down the species that could contribute to this soil in agricultural systems.
Around 71 enzymes from the KEGG database were predicted as belonging to the pathways of carbon fixation in prokaryotes. Most of these enzymes are involved in the breakdown of malate, glutamate or pyruvate compounds (Table 2.6 and Fig. 2.14).

Figure 2.14 Carbon fixation enzymes with highest no-till:plow-till gene abundance ratios. Percentages indicate contribution to the pie chart, not contribution to the overall category.
<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Function</th>
<th>Gene</th>
<th>Total gene count</th>
<th>Rank</th>
<th>Number of OTUs</th>
<th>Most abundant OTU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isocitrate dehydrogenase</td>
<td>Isocitrate formation</td>
<td>icd</td>
<td>2146</td>
<td>47</td>
<td>58</td>
<td>Koribacteriaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1384</td>
<td>41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acetyl-CoA C-acetyltransferase</td>
<td>Acetyl CoA formation</td>
<td>atoB</td>
<td>1118</td>
<td>44</td>
<td>59</td>
<td>Rhodoplanes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>671</td>
<td>42</td>
<td></td>
<td>Ellin329</td>
</tr>
<tr>
<td>Methylenetetrahydrofolate dehydrogenase</td>
<td>Formate degradation</td>
<td>folD</td>
<td>1178</td>
<td>20</td>
<td>61</td>
<td>Koribacteriaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>526</td>
<td>44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transketolase</td>
<td>Ribose 5-phosphate formation</td>
<td>tktA, tktB</td>
<td>977</td>
<td>29</td>
<td>61</td>
<td>Nitrospira</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>473</td>
<td>44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acetyl-CoA synthetase</td>
<td>Acetate degradation</td>
<td>acs</td>
<td>863</td>
<td>35</td>
<td>60</td>
<td>Rhodoplanes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>435</td>
<td>44</td>
<td></td>
<td>Bradyrhizobium</td>
</tr>
<tr>
<td>Glyceraldehyde 3-phosphate dehydrogenase</td>
<td>Glyceraldehyde 3-phosphate formation</td>
<td>gapA</td>
<td>668</td>
<td>19</td>
<td>61</td>
<td>Koribacteriaceae</td>
</tr>
</tbody>
</table>

Table 2.6 Summary of KEGG genes and Greengenes OTUs involved in carbon fixation; rank is based on order of No-Till:Plow-Till ratios. Upper values represent the NT sample and lower values represent the PT sample.
Table 2.6 continued

<table>
<thead>
<tr>
<th>Glyceraldehyde 3-phosphate dehydrogenase</th>
<th>Glyceraldehyde 3-phosphate formation</th>
<th>gapA</th>
<th>294</th>
<th>44</th>
<th>Nitrospira</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malate dehydrogenase</td>
<td>Malate formation</td>
<td>mdh</td>
<td>645</td>
<td>22</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>293</td>
<td>44</td>
<td>Nitrospira</td>
</tr>
<tr>
<td>Fumarate reductase, iron-sulfur subunit</td>
<td>Fumarate formation</td>
<td>sdhB, frdB</td>
<td>589</td>
<td>24</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>270</td>
<td>44</td>
<td>Nitrospira</td>
</tr>
<tr>
<td>Fumarate reductase, flavoprotein subunit</td>
<td>Fumarate formation</td>
<td>sdhA, frdA</td>
<td>593</td>
<td>17</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>260</td>
<td>44</td>
<td>Cand. Solibacter</td>
</tr>
<tr>
<td>Aconitate hydratase</td>
<td>Citrate formation</td>
<td>acnA</td>
<td>593</td>
<td>14</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>257</td>
<td>44</td>
<td>Koribacteriaceae</td>
</tr>
</tbody>
</table>
Relatively few enzymes were enriched in the PT treatment compared to NT, as was the consistent trend across all previous categories (aside from hydrocarbon degradation). 3-hydroxypropionate dehydrogenase (mcl), 3-methylfumaryl-CoA hydratase (meh), pyruvate carboxylase (pyc), and the flavoprotein subunit of fumarate reductase (frdA) all belong to this category. meh expression is dominated by the Alphaproteobacteria, with a rare appearance by Limnohabitans sp. from Betaproteobacteria. The pyc gene, on the other hand, shows almost no Proteobacteria involvement, being expressed mainly by the Acidobacteria. A similar trend can be seen with the predictions for the frdA gene, while the mcl gene showed an almost equal mix of classes Alphaproteobacteria and Chloracidobacteria coding for it.

Four enzymes in this category were found under NT treatment alone. These included two subunits of fumarate reductase (frdB and frdC), 2-methylfumaryl isomerase (mct) and a putative phosphotransacetylase (pduL). All these enzymes were associated with only one OTU each, varying between Bacillus sp. (pduL), and family Rhodospirillaceae (mct, frdB and frdC).

Surprisingly, groups of enzymes with similar functions tended to have almost the same NT/PT ratio in the prokaryotic carbon fixation category. For instance, all three subunits of carbon monoxide dehydrogenase (large, small and medium) have ratios between 1.1-1.2, pyruvate dikinases (water and orthophosphate) have NT:PT ratios of 1.7 and 1.9 respectively, and the alpha and beta subunits of fumarate hydratase have ratios of 1.5 and 1.47 respectively. This suggests that the various sub-processes underlying carbon fixation functions, at least in soil prokaryotes, might be more closely coupled than other functions investigated using PICRUSt.
All of the genes associated with carbon monoxide dehydrogenase (CODH enzyme) showed a slight preference for the NT treatment (relative abundance ratios between 1.1 and 1.2). Despite belonging to the same gene family, different OTUs are associated with the expression of the different subunits of CODH, as described below:

a. **coxS** – Mix of all four classes of Proteobacteria, isolated OTUs belonging to phyla AD3, Nitrospirae, and Planctomycetes

b. **coxM** – Predominantly Alpha- and Deltaproteobacteria, Saprospiraceae family from Bacteroidetes, AD3 and Actinobacteria (Gaiellaceae). Two rare OTUs from Betaproteobacteria were displayed specific to each treatment – *Janthinobacterium sp.* from the family Oxalobacteraceae in NT, and *Limnohabitans sp.* from family Comamonadaceae.

c. **coxL** – This is related almost exclusively to Alphaproteobacteria (*Rhodoplanes sp.*, order Ellin329 and the *Rhodospirillaceae* family) and Deltaproteobacteria (order Myxococcales). The family Gaiellaceae from Actinobacteria contributes 2 units of gene abundance under the PT treatment alone.

This is an exciting possibility for exploration, as it indicates that different species of microorganisms could be involved at different levels in the expression of proteins that are closely regulated to perform a single ecosystem function, in this case carbon sequestration as affected by tillage.

The **ppdK** and **ppsA** genes, corresponding to the pyruvate dikinase enzymes mentioned previously, do not show such clear correlations with any particular phylum, but have microbial composition distributed almost evenly across the most
abundant OTUs. Both these enzymes catalyze the conversion of pyruvate to PEP, with the difference that the water dikinase causes an irreversible reaction, in contrast to the reversible reaction catalyzed by PPDK. Pyruvate dikinase (water) and pyruvate kinase (orthophosphate or PPDK) are important enzymes in the reductive carboxylate cycle in CO₂ fixation. They regulate the efficiency of CO₂ uptake in environments with a lot of light, thus helping to make the C4 pathway of photosynthesis more efficient (Pocalyko et al., 1990).

Examination of various carbon fixation pathways in prokaryotic systems helped to group the enzymes in this category effectively. The enzymes and their corresponding genes, absolute gene abundances and microbial communities associated with these genes are described in detail below.

Most carbon fixation in these particular soil samples appears to be taking place through the reductive citrate cycle or the Arnon-Buchanan cycle (or the reductive/reverse citrate cycle), which also shares enzymes with other citrate-processing cycles such as the dicarboxylate-hydroxybutyrate cycle. As the name suggests, the reverse citrate cycle takes up molecules of CO₂ and converts it to the form of carbon compounds like acetate, citrate, malate, fumarate and succinate. It is mostly found in anaerobic and microaerobic bacteria (Buchanan & Evans, 1966).

Specifically, the enzymes identified as belonging to the Arnon-Buchanan cycle are 2-oxoglutarate ferredoxin oxidoreductase (korB/oorB, korA/oorA, korC/oorC), pyruvate-ferredoxin/flavodoxin oxidoreductase (por, porA, gamma, porB, porD), succinyl-CoA synthetase (sucC, sucD), aconitate hydratase (acnA), fumarate hydratase (fumA, fumC, frdC, frdA, fumB), fumarate reductase (frdA, frdB, frdD, frdB, frdC), pyruvate carboxylase (pyc, pycA, pycB), pyruvate dikinase (ppsA), isocitrate
dehydrogenase (icd), and 2-methylisocitrate dehydratase (acnB).

All subunits of pyruvate ferredoxin oxidoreductase are linked to different OTUs belonging to the *Nitrospirae* family. Even though some OTUs belonging to *Planctomycetes* and *Deltaproteobacteria* (family *Syntrophobacteraceae*) are also observed, *Nitrospirae* is responsible for most of this function. 2-oxoglutarate ferredoxin oxidoreductase does not show any such OTU preferences. Various highly abundant phyla contribute to the formation of different subunits of this enzyme in near-equal measure, as also in the case of aconitate hydratase, pyruvate kinase and isocitrate dehydrogenase. The two subunits of succinyl CoA-synthetase have high absolute gene abundances and show differential enrichment under the NT treatment (NT/PT ratios of 2.4) but the OTU contributions to this enzyme are evenly distributed among the *Acidobacteria* and *Proteobacteria*. Similarly, pyruvate carboxylase is associated mainly with *Nitrospira sp.* and *Syntrophobacteraceae*.

Fumarate reductase, like most of the other enzymes in this cycle, is also produced mainly by members of *Proteobacteria* and *Acidobacteria*, aside from the flavoprotein subunit which also has a number of *Nitrospirae* OTUs associated with it. This phenomenon is echoed by fumarate hydratase as well. 2-methylisocitrate dehydrogenase shows a marked difference from these enzymes as it is linked to some rare OTUs. *Gammaproteobacteria* is the dominant class mapping to this enzyme (families *Sinobacteraceae*, *Piscirickettsiaceae*, and *Xanthomonodaceae*), followed by *Limnnohabitans sp.* and *Janthinobacterium sp.* from *Betaproteobacteria.*
Another cycle of interest is the Calvin cycle or the ribulose bisphosphate pathway, which can convert atmospheric CO$_2$ into sugar. This is probably the most important carbon fixation pathway known thus far (Bassham et al., 1950). Since prokaryotes do not have chloroplasts like plants, it has been suggested that this series of enzymatic reactions occurs in a thylakoid-like structure in their cytoplasm (Suss et al., 1995). Ribulose-bisphosphate carboxylase (rbcS), phosphoribulokinase (prkB), ribose 5-phosphate isomerase A (rpiA), fructose-bisphosphate aldolase (fbaA, ALDO), fructose-1,6-bisphosphatase (fhp, glpX), transketolase (tktA, tktB), methylenetetrahydrofolate dehydrogenase (metF), and phosphoglycerate kinase (pgk) are the enzymes from this cycle that were identified by PICRUSt in these samples.

The rbcS gene is predicted to be expressed by Alpha-, Beta-, and Gammaproteobacteria exclusively, as are prkB and fbp. rpiA is also related to mostly these three classes of Proteobacteria but some Bacteroidetes, WS3 and Gemmatimonadetes can also be seen contributing to this enzyme. In the case of the aldolase genes, ALDO and fbaA, a mix of the most abundant OTUs is seen, as also with glpX, tktA/B, metF and pgk.

Another carbon sink in bacterial metabolism is the 3-hydroxypropionate pathway, wherein CO$_2$ is utilized to generate pyruvate (Strauss & Fuchs, 1993). PICRUSt predicted the presence of the enzymes catalyzing these reactions, namely acetyl-CoA carboxylase transferase (accA, accB) and 3-hydroxypropionate dehydrogenase (mcr). The accA and accB genes were not exclusive to any OTUs, and most OTUs in these samples were able to contribute to carbon fixation in this pathway, but mcr was found only under the NT treatment and characterized by the
exclusive contributions of the Chloracidobacteria. This is another illustration of the peculiar linkage between OTU and function for some enzymes.

Another pathway used for the conversion of CO₂ to acetyl CoA in some bacteria and archaea is the reductive acetyl-CoA pathway or the Wood-Ljungdahl pathway. In this pathway, carbon dioxide is reduced to carbon monoxide and then acetyl CoA. This is typically found in methanogens and acetate-producing bacteria (Ragsdale & Pierce, 2008).

Acetyl CoA synthase (acsB), CO dehydrogenase (cooS), and acetyl-CoA decarbonylase (cdhB, cdhC) are the most important enzymes in this pathway, and their presence has been predicted by PICRUSSt. Three OTUs are involved in the synthesis of acetyl-CoA decarbonylase and the cooS gene; these are Nitrospira sp., family 0319-A621 from Nitrospirae and Syntrophobacteraceae from Deltaproteobacteria. Interestingly, the same trend was seen in the prediction of acetyl-CoA synthase, except that the PT sample showed traces of Acidobacteria involvement (all from the Acidobacteria-6 family).

Most of the important microorganisms that have been implicated in the functioning of the major carbon fixation pathways described above were not a part of the microbial diversity in our soil samples. For instance, the Wood-Ljungdahl pathway is carried out in bacteria like Clostridium aceticum and Moorella thermoacetica (Ragsdale & Pierce, 2008), while the Arnon-Buchanan cycle has been found to occur mainly in phylum Aquificae, and possibly in some Archae (Berg et al., 2007). While the Calvin cycle is indeed associated with Proteobacteria, the purple bacteria (Rhodospirillaceae) as well as Cyanobacteria are the groups proven to carry out this process in prokaryotes (Swan et al., 2011). It is possible that these bacteria
were not present in these soil samples, or the absence of these species could be indicative of insufficient information for taxonomic or functional identification.

It must be noted that the PICRUSt predictions suggest the closest related genetic neighbors of these prokaryotic groups as containing these genes, which corresponds with the mechanism of the PICRUSt algorithm. However, it is important that we seek to improve the accuracy of such predictions, both by improving the accuracy of sequence-based taxonomic identification as well as generating better databases for diverse but poorly characterized environments like soil.

2.3.9. Methane metabolism

Methane is an important greenhouse gas that accounted for 10% of total GHG emissions in the USA in 2013. Natural, microbe-mediated processes in the soil help remove methane from the atmosphere, which is crucial given that the effect of methane on GHG emissions is thought to be 25% higher than that of CO$_2$, pound for pound (EPA, 2010). Around thirty enzymes were implicated in methane metabolism of soil microbes in these samples, most of which dealt with the processing of formate molecules. The top ten most abundant enzymes (depicted in Fig. 2.15 and Table 2.7) are the only ones displaying above average gene abundances compared to other enzymes in this functional category.

Formate dehydrogenase genes (FDH, $f$ds$D$, and $f$dh$A$) and the gene encoding for S-(hydroxymethyl) glutathione synthase ($gfa$) are the only enzymes enriched in the plow-till sample, but it is important to note that neither of these are enzymes that are specific to methane metabolism. They are also implicated in carbohydrate and energy metabolism as well as general carbon metabolism (Jormakka et al., 2003).
Figure 2.15. Methane metabolism enzymes with highest no-till:plow-till gene abundance ratios. Percentages indicate contribution to the pie chart, not contribution to the overall category.
<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Function</th>
<th>Gene</th>
<th>Total gene count</th>
<th>Rank</th>
<th>Number of OTUs</th>
<th>Most abundant OTU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formate dehydrogenase major subunit</td>
<td>CO₂ formation</td>
<td>fdoG, fdfH</td>
<td>1104</td>
<td>8</td>
<td>60</td>
<td>Koribacteriaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>457</td>
<td>39</td>
<td>Koribacteriaceae</td>
<td></td>
</tr>
<tr>
<td>Formate dehydrogenase iron-sulfur subunit</td>
<td>CO₂ formation</td>
<td>fdoH</td>
<td>934</td>
<td>7</td>
<td>43</td>
<td>Koribacteriaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>353</td>
<td>29</td>
<td>Koribacteriaceae</td>
<td></td>
</tr>
<tr>
<td>Formate dehydrogenase alpha subunit</td>
<td>CO₂ formation</td>
<td>fdhA1</td>
<td>584</td>
<td>4</td>
<td>5</td>
<td>Nitrospira</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>195</td>
<td>2</td>
<td>Nitrospira</td>
<td></td>
</tr>
<tr>
<td>Formylmethanofuran dehydrogenase subunit E</td>
<td>Formate synthesis</td>
<td>fwdE, fmdE</td>
<td>444</td>
<td>5</td>
<td>27</td>
<td>Koribacteriaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>153</td>
<td>17</td>
<td>Nitrospira</td>
<td></td>
</tr>
<tr>
<td>Formate dehydrogenase subunit gamma</td>
<td>CO₂ formation</td>
<td>fdoI</td>
<td>272</td>
<td>13</td>
<td>22</td>
<td>Ellin6513</td>
</tr>
</tbody>
</table>

Table 2.7 Summary of KEGG genes and Greengenes OTUs involved in methane metabolism; rank is based on order of No-Till:Plow-Till ratios. Upper values represent the NT sample and lower values represent the PT sample.
Table 2.7 continued

<table>
<thead>
<tr>
<th>Enzyme Complex</th>
<th>Subunit Type</th>
<th>Gene Symbol(s)</th>
<th>Accessory</th>
<th>Accessory</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formate dehydrogenase subunit gamma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carbon monoxide dehydrogenase small subunit</td>
<td>CO₂ formation</td>
<td>coxS</td>
<td>198</td>
<td>20</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>167</td>
</tr>
<tr>
<td>Carbon monoxide dehydrogenase medium subunit</td>
<td>CO₂ formation</td>
<td>coxM, cutM</td>
<td>175</td>
<td>21</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>159</td>
</tr>
<tr>
<td>Carbon monoxide dehydrogenase large subunit</td>
<td>CO₂ formation</td>
<td>coxL, cutL</td>
<td>148</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>122</td>
</tr>
<tr>
<td>S-formylglutathione hydrolase</td>
<td>Formate synthesis</td>
<td>frmB, fghA</td>
<td>87</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>48</td>
</tr>
<tr>
<td>Formate dehydrogenase subunit delta</td>
<td>CO₂ formation</td>
<td>fdsD</td>
<td>37</td>
<td>25</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>43</td>
</tr>
</tbody>
</table>
Two major processes were observed from the KEGG ontologies related to methane processing – methanogenesis, or the conversion of carbon dioxide to methane and methane oxidation (typically performed by methanotrophs), where methane is converted to formaldehyde. Two molecules (the trimethylamine-corrinoid protein Co-methyltransferase, \textit{mttB} gene and the trimethylamine corrinoid protein, \textit{mttC}) were shown to catalyze the formation of methane from methylamine compounds.

Eight genes can be seen contributing to the process of methanogenesis. These are formylmethanofuran-tetrahydromethanopterin N-formyltransferase (\textit{ftr}), methenyltetrahydromethanopterin cyclohydrolase (\textit{mch}), formylmethanofuran dehydrogenase (\textit{fmd/fwd A,B,C,E}), the trimethylamine corrinoid protein (\textit{mttC}), and the trimethylamine-corrinoid protein Co-methyltransferase (\textit{mttB}).

The \textit{fmdA}, \textit{fmdB} and \textit{fmdC} genes all have similar OTUs linked to them, classified as phyla \textit{Proteobacteria} or \textit{Planctomycetes} (the \textit{Pirellulaceae} family). The \textit{fmdE} gene, however, displays much higher abundance and involvement of a much more diverse set of OTUs. Formylmethanofuran dehydrogenase subunit E is the fifth most abundant enzyme in methane metabolism, and the most prominent OTUs involved in its expression belong to the phyla \textit{Acidobacteria} and \textit{Nitrospirae}. Orders \textit{Acidobacteriia}, \textit{Acidobacteriia-6}, \textit{Sva0725}, \textit{Solibacteres}, \textit{Chloracidobacteria} and \textit{DA052} from \textit{Acidobacteria} and families \textit{Nitrospiraceae} and 0319-6A21 from \textit{Nitrospirae} were the Greengenes OTUs corresponding to these KEGG IDs, with the family \textit{Syntrophobacteraceae} being the only OTU represented from \textit{Proteobacteria}. 

For the reverse reaction, breakdown and oxidation of methane, the \textit{pmo} set of genes (closely related to and having the same abundance and functional microbial composition as the \textit{amo} genes) is the major contributor, followed by cytochrome c-associated methanol dehydrogenase enzymes (\textit{mdh1, mdh2}). A key feature of the methanol dehydrogenase enzymes is that they were observed in the PT sample alone, and at very low gene abundances (1 each). \textit{Hyphomicrobium sp.} from the \textit{Alphaproteobacteria} class was the only organism predicted in the expression of these enzymes. The \textit{Hyphomicrobium} genus has earlier been identified as a potential commercial source for the enzyme methanol dehydrogenase (Duine & Frank, 1980).

\textit{CO}_2 is an important end product of methane catabolism and so, the reversible conversion between \textit{CO}_2 and CO, which is catalyzed by CODH, is of special importance. Three genes involved in expression of this enzyme (\textit{coxS, coxM, coxL}) were present in both the NT and PT samples. All of these genes showed a slight preference for the NT treatment (relative abundance ratios between 1.1 and 1.2). The microbial composition associated with these enzyme subunits has been described in the previous section. To recap briefly, \textit{Proteobacteria} was found to be strongly associated with this function but different classes of this phylum were mapped to different subunits, suggesting that closely related but distinct genotypes could be involved in different sub-functions of the same enzyme.

Evidently, only a small portion of the enzymes involved in the utilization and cycling of methane in the soil have been predicted by PICRUSt from these soil samples. The issue of soil serving as a sink for methane is a subject of intense research and debate, and utilizing the potential of soil microbes to do so is certainly one of our best chances of achieving this. Perhaps deeper sequencing and
improvement of functional genomic tools like PICRUSt (and the associated databases) would enable us to make strides in this direction.

2.4. SUMMARY

This pilot study of two tillage samples has been conducted from sites where NT and PT have been continuously maintained for almost 50 years, and can thus be considered to have reached some sort of equilibrium. This has, thus, given us an idea of the capabilities as well as drawbacks of the PICRUSt algorithm. Perhaps the most important limiting factor in utilizing PICRUSt to predict microbial community function in soil is the lack of comprehensive identification and/or annotation for the microbial species in soil. A much larger and well-characterized reference dataset for terrestrial microorganisms would certainly have improved the specificity of PICRUSt predictions. It is certainly useful as a tool for preliminary assessment of the microbial population in any given sample, and the functional potential therein, but further validation would be needed before it could qualify as a standard for functional prediction.

There is also the question of how better sequence quality and more reads per sample (thereby resulting in better OTU identification, both qualitatively as well as quantitatively) would affect these PICRUSt predictions. With this idea in mind, the 16S rDNA from six soil samples, from the Northwest Agricultural Research Station in Hoytville, Ohio, was sequenced using the Illumina platform. The findings from this experiment are described in the following chapter.
REFERENCES


Piutti, S., Semon, E., Landry, D., Hartmann, a., Dousset, S., Lichtfouse, E., & Martin-Laurent, F. (2003). Isolation and characterisation of Nocardioides sp. SP12, an


CHAPTER 3: FUNCTIONAL PROFILING OF TILLAGE AND CROP ROTATION EFFECTS ON HOYTVILLE SOILS

3.1 INTRODUCTION

The initial promising results from the pyrosequencing analyses conducted in the previous chapter led to the design of a larger study at a new site. In this study, we decided to study the effects of an additional variable, crop rotation, on microbial community composition and function predictions. Two positive control sites, a forest and grassland, were also included as references for comparison.

One important modification in this experiment is the usage of Illumina MiSeq sequencing in the place of 454 pyrosequencing. It has been demonstrated that Illumina sequencing reads provide equal, if not better, read coverage and quality than pyrosequencing, in addition to producing a much greater number of reads per sample (Claesson et al., 2010). Many soil metagenomic studies have adopted Illumina sequencing technologies for this reason. This study also utilized the Illumina MiSeq platform, which resulted in increased OTU identification and classification as well as more PICRUSt-predicted functions.

3.2 MATERIALS AND METHODS

3.2.1 Experimental design

This study was conducted on agricultural plots in the Northwest Agricultural Research Station, Hoytville, Ohio. The soil at this location was identified as silty clay
loam. The details of the treatments studied at these plots are shown in Table 3.1. In addition to these treatments, soil samples were collected in triplicate from an adjoining forest and grassland site, to be used as positive controls. There was no elevational difference between the forest or grassland sites and the agricultural plots, and similar soil type was observed.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Tillage</th>
<th>Crop rotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>NT-CC</td>
<td>No Tillage (NT)</td>
<td>Continuous Corn (CC)</td>
</tr>
<tr>
<td>NT-CS</td>
<td>Plow Tillage (PT)</td>
<td>Corn-Soybean (CS)</td>
</tr>
<tr>
<td>PT-CC</td>
<td>No Tillage (NT)</td>
<td>Continuous Corn (CC)</td>
</tr>
<tr>
<td>PT-CS</td>
<td>Plow Tillage (PT)</td>
<td>Corn-Soybean (CS)</td>
</tr>
<tr>
<td>G</td>
<td>Grassland (G)</td>
<td>Control</td>
</tr>
<tr>
<td>F</td>
<td>Forest (F)</td>
<td>Control</td>
</tr>
</tbody>
</table>

Table 3.1 Sample treatments in Hoytville plots

The moldboard plow was used to till and prepare the plots under the plow-till treatment. The no-till plots at Hoytville have not had any tillage other than seed placement and anhydrous ammonia application since 1964 (50 years at the time of sampling). As with the Wooster site, a combination of continuous corn (CC), corn-soybean (CS) and corn-oats-alfalfa (COA) crop rotations are applied to these plots. In this study, only the effects of the CC and CS rotations were considered (in addition to the comparison between NT and PT treatments).

The inclusion of biological replicates in sequencing studies is a current topic of interest, especially in the context of these long-term plots. Studies have indicated
that sample beta diversity does not vary between sequencing runs or sequencing depths (Caporaso et al., 2012). Thus, it has been suggested that dense sampling studies involving a larger number of samples and treatments may be more beneficial than sequencing biological replicates on different runs. In our case, the long-term nature of the NT and PT treatments decrease the odds that any effects observed are due to chance variation, and substantiate the nature of our results. Simply put, the long-term treatments could be regarded as temporal replications of each treatment, as opposed to spatial replications commonly seen in other studies. However, increased statistical power and detection of rare taxa are certainly benefits of replication and these should not be disregarded lightly.

3.2.2 Sample collection and processing

Soil core samples were collected in triplicate from each plot, with three randomized plots for the six treatments described in Table 3.1. The cores were then pooled together to form a composite sample for each replicate of each treatment. DNA was extracted from these soil samples using the UltraClean® Soil DNA Isolation Kit manufactured by Mo Bio Laboratories, Carlsbad, CA. The manufacturer’s instructions were followed to obtain genomic DNA. Following this, the samples underwent one cycle of PCR to attach partial Illumina adapters (Nextera™ XT DNA Sample Preparation Kit, Illumina, San Diego, CA) and to amplify the region of interest, i.e. the V3 region of the 16S rRNA gene (spanning 170 bp).

Agarose gel electrophoresis was used to separate this amplified DNA into fragments, and the desired fragment was cut out from the gel and amplified again.
using PCR (initial denaturation at 94 °C for 1 min; 20 cycles of denaturation at 94 °C for 30 s, annealing at 49 °C for 60 s, and extension at 72 °C for 60 s; and a final extension at 72 °C for 5 min). The resulting amplicon products from each replicate were then pooled together to make a composite sample for each treatment. Finally, indices required for sequencing library preparation were attached, along with sequencing primers, and these libraries were sent for sequencing on the MiSeq platform (Illumina) at the Molecular Cellular Imaging Center (MCIC), OARDC, Wooster, Ohio. Details of the protocol for locus-specific primer and partial adapter attachment, as well as Illumina index and Nextera primer attachment can be found in Aditi Sengupta’s doctoral dissertation (Sengupta, 2015).

The raw data were supplied as demultiplexed FASTA files for analysis. A workflow depicting the sequence of steps used for library preparation is shown in Fig 3.1.
Figure 3.1 Workflow for processing of soil samples in the laboratory. Here, NT-CC and PT-CC refer to the sample treatments (no till-continuous corn and plow till-continuous corn).

3.2.3 Analysis of Illumina MiSeq Data

The raw reads from the sequencer were denoised, filtered and passed for downstream analysis using QIIME version 1.8 (Caporaso et al., 2011). QIIME, which stands for Quantitative Insights Into Microbial Ecology, is a widely used bioinformatics tool that combines various open-source programs and scripts to form a robust pipeline for analysis of metagenomic data. The steps used for downstream analysis of sequencing data can be divided into three main categories – filtering of reads, Operational Taxonomic Unit (OTU) picking and tabulation, and OTU analysis using PICRUSt. The steps and parameters conducted during each of these stages are described in detail below.
3.2.4 Quality filtering and trimming of Illumina reads

The raw reads were examined using parameters including quality scores, overrepresented sequences, k-mers and sequence lengths using FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). Based on the FastQC-generated reports, the parameters for sequence trimming were set. TrimGalore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was used to trim the adapter sequences from the 3' ends (the Illumina adapter AGATCGGAAGAGC was used during library preparation). In addition to this, sequences with lengths less than 160 bp were removed, as were sequences with Phred quality score less than 25. As the files had already been demultiplexed by the sequencing facility, primer and barcode removal were unnecessary. Finally, the trimmed and filtered reads were assembled using the fastq-join program in the QIIME workflow (ea-utils; Aronesty, 2011).

3.2.5 OTU picking

Open-reference OTU picking was done on the sequences so as to obtain both closed-reference OTUs, which are necessary for PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) analysis, as well as unclassified de novo OTUs (in order to get an estimate of total diversity in the sample). The reference database chosen was the Greengenes database (May 2013 version), which is a curated collection of 16S rRNA sequences (DeSantis et al., 2006). In brief, the OTU picking process occurs through seven steps:
1. Using UCLUST (Edgar, 2010) to cluster sequences against the Greengenes database. This generates clusters of sequences or OTUs using UCLUST’s heuristic approach. These clusters are compared against the reference sequences in the Greengenes collection at the specified similarity level. If sequence produces a hit, the cluster is assigned to that Greengenes identifier. The remaining, unclassified sequences are then clustered against each other to produce *de novo* OTUs. This takes place with prefiltering and iterations to improve the accuracy of the process.

2. Picking a representative set of sequences from each cluster; the most abundant seed from each cluster was chosen for this analysis.

3. Alignment of the representative sequences using PyNAST (Caporaso et al., 2010)

4. Checking for chimeric sequences using ChimeraSlayer (Haas et al., 2011)

5. Assigning taxonomy to the UCLUST-generated OTUs using the Greengenes taxonomic annotations (QIIME uses “UCLUST” for taxonomy assignment as a default)

6. Construction of the OTU table. This step involved removal of failures from the sequence alignment step (#3) and chimeric sequences identified by ChimeraSlayer (#4) from the final OTU table.

7. In addition to these steps, the PyNAST-aligned sequences were filtered and used to construct a phylogenetic tree using the FastTree algorithm (Price et al., 2009).
3.2.6 Predicting community function with PICRUSt

PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) is a bioinformatics software package designed to predict metagenome functional content from 16S rRNA gene surveys (Langille et al., 2013). This algorithm uses a fast Ancestral State Reconstruction method from the APE (Analysis of Phylogenetics and Evolution) R package to predict the gene content of the closest known relatives of the reference dataset (Paradis et al., 2004). The key point of interest here is that Ancestral State Reconstruction is used for reconstruction of the most likely metagenome associated with a particular amplicon sample, and not ancestral relatives of the same.

PICRUSt utilizes the reference OTUs picked in the previous step for functional analysis. The process required for metagenome and functional predictions is described below in some detail. 16S rRNA copy numbers are predicted using the IMG database (Markowitz et al., 2012) and copy number normalization is done to achieve unbiased estimates of overall diversity. The final predicted metagenome is then generated by multiplying the gene counts associated with each OTU with the abundance of that OTU in that sample. These OTUs are then matched against the KEGG (Kyoto Encyclopedia of Genes and Genomes) database (Kanehisa & Goto, 2000) to predict functional traits associated with those gene families. From this OTU table, the taxonomic classification can be collapsed down to any hierarchical level and then analyzed at that particular level. Another useful PICRUSt function utilized in this study is its ability to identify the OTUs contributing to specific functions, as implemented by the “metagenome_contributions.py” script. A workflow depicting this analysis is shown in Fig. 3.2.
PICRUST enables the user to use either the COG (Clusters of Orthologous Groups of proteins) (Tatusov et al., 2000) or RFAM (RNA family) (Griffiths-Jones et al., 2003) databases as references for functional annotations as well. For this analysis, the KEGG ortholog and pathway predictions were found to give the best predictions.

3.3 RESULTS AND DISCUSSION

3.3.1 OTU table summary

At a sequence similarity threshold of 99%, a total of 1,663,232 sequences were clustered (~97% of the filtered sequence collection). Of these, 44,110 sequences were identified as chimeras by “ChimeraSlayer”. Around 92% of the total non-chimeric clustered sequences were identified as belonging to the Greengenes
database. This resulted in a total of 28,892 OTUs being identified, with 13,214 of these belonging to the reference collection. Thus, clustering sequences generated by Illumina sequencing produced more reads and OTUs than pyrosequencing and hence, better mapping to the Greengenes database (than 454 pyrosequencing), which has been verified by other studies as well (Liu et al., 2012; Luo et al., 2012).

3.3.2 Sequencing depth and sample diversity

Multiple rarefactions were performed on the OTU table to generate a rarefaction curve, which is depicted in Fig. 3.3. The rarefaction curve has not reached a saturation plateau yet but is close, which implies that deeper sequencing might have yielded slight increase in diversity. The effect of increasing sequencing depth is most apparent with the NT-CS treatment, whereas the grassland and forest samples (positive controls) are the least affected by sequencing depth. The PT-CS treatment had the least number of sequences, and thus, does not extend all the way through the X-axis.

A visualization of the top ten phyla in these samples (Fig. 3.4) show that Proteobacteria and Actinobacteria are most abundant, followed by Acidobacteria, Chloroflexi and Gemmatimonadetes. Unlike the previous study, Illumina sequencing reveals the presence of other, rare phyla such as Cyanobacteria, Spirochaetes, Elusimicrobia, Lentisphaerae and Tenericutes. These results are consistent with those seen in soil samples sequenced in other recent studies (Li et al., 2014; Navarro-Noya et al., 2013; Peralta et al., 2013; Quadros et al., 2012; Hill et al., 2006; Pankratova et al., 2006).
Figure 3.3 Rarefaction curve of OTUs generated with sequencing depth; the lines depict (from the top) the following treatments: NT-CS (plain blue), PT-CC (dotted orange line with squares), NT-CC (dotted red line with triangles), PT-CS (green line with crosses), Forest (dotted purple line with circles) and Grassland (dotted blue line with pluses) (NT-No Till, PT-Plow Till, CC-Continuous Corn, CS-Corn Soybean).
In terms of phylum-level OTU diversity, the NT sample is generally higher than PT. This is corroborated by the alpha diversity indices shown in Table 3.2, which indicate that the species richness is in favor of the no-till treatment. In the case of crop rotation, no clear correlation can be seen between any of the treatments and species abundance.

The diversity indices appear to be higher than those observed in the Wooster soil (Chapter 2 of this thesis) in all cases except for the Simpson indices. In these soil samples, the Simpson index indicates low diversity, which is consistent across both locations. Also, because crop rotation does not show correlation with diversity indices, this may suggest that the effects of tillage and location overshadow those of crop rotations, as seen in a different phylogenetic analysis of soil samples from this location (Aditi Sengupta, personal communication). The Shannon and Simpson
indices indicate that species richness and evenness is almost equal across all the treatments, but the NT-CS treatment contains a greater number of rare OTUs (typically represented by singletons), as shown by the increased Chao1 diversity estimate. This is corroborated by the ACE index as well. All of the diversity indices calculated here indicate that the NT-CS treatment shows greatest microbial richness and evenness of distribution, closely followed by the Grass plot and the PT-CC and NT-CC treatments.

<table>
<thead>
<tr>
<th>Diversity Index</th>
<th>NT-CC</th>
<th>NT-CS</th>
<th>PT-CC</th>
<th>PT-CS</th>
<th>G</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chao1</td>
<td>19180</td>
<td>23850</td>
<td>20350</td>
<td>18060</td>
<td>20120</td>
<td>17640</td>
</tr>
<tr>
<td>ACE</td>
<td>20160</td>
<td>25240</td>
<td>21330</td>
<td>18480</td>
<td>21300</td>
<td>18610</td>
</tr>
<tr>
<td>Shannon</td>
<td>7.59</td>
<td>7.67</td>
<td>7.44</td>
<td>7.32</td>
<td>7.08</td>
<td>7.44</td>
</tr>
<tr>
<td>Simpson</td>
<td>0.998</td>
<td>0.998</td>
<td>0.997</td>
<td>0.997</td>
<td>0.995</td>
<td>0.997</td>
</tr>
<tr>
<td>Fisher</td>
<td>3203</td>
<td>3920</td>
<td>3270</td>
<td>2903</td>
<td>3111</td>
<td>3025</td>
</tr>
</tbody>
</table>

Table 3.2 Alpha diversity indices for soil samples under different treatments. NT-CC stands for n till-continuous corn, NT-CS for no till-corn-soybean, PT-CC for plow till-continuous corn, PT-CS for plow till-corn-soybean, G for Grass and F for Forest.

Between the NT treatments, the CS rotation shows higher abundance and diversity across almost all of the diversity indices measured. This would fit in with the hypothesized benefits of crop rotations over monocrop cultivation, but the same trend is not observed in the PT treatments.

### 3.3.3 PICRUSt predictions

The OTUs generated at 99% sequence similarity were processed using PICRUSt to yield information on putative functions performed by microorganisms.
across the two different treatments. Using the KEGG database, these functions were
classified into various categories such as energy metabolism, amino acid metabolism,
hydrocarbon degradation, transcription and translation, replication and repair,
membrane transport, signal transduction, cell motility, etc.

The most abundant of these predicted functions were mapped to membrane
transport, DNA repair and recombination, signal transduction, purine metabolism,
translation-related protein processing, oxidative phosphorylation and bacterial
motility protein secretion. Most of these functions are related to maintenance of cell
function and structure and so, would be performed by every species in the
community. However, for our study, the functions mentioned in Table 2.1 only were
considered.

PICRUSt’s NSTI (Nearest Sequenced Taxon Index) provides a reliable way to
estimate the confidence of these predictions. NSTI is the sum of phylogenetic
distances for each organism in the OTU table to its nearest relatives with a sequenced
reference genome, measured in terms of substitutions per site in the 16S rRNA gene
and weighted by the frequency of that organism in the OTU table (Langille et al.,
2013). For these soil samples, PICRUSt predicted NSTI values between 0.17 – 0.20,
which are similar to the values obtained for the soil test datasets in Langille et al. as
well as the NSTI values obtained in the pyrosequencing study in Chapter 2 (0.17-
0.19). In these samples, the grassland had the lowest NSTI value, while NT-CS, PT-
CS and Forest had the highest values. In simple terms, this would indicate that the
grassland samples had the closest phylogenetically related reference genomes. As
seen in the pilot study, a large portion of the PICRUSt functional predictions were
found to be unclassified (around 13% of the reference OTUs were not mapped to any protein/function).

In this dataset as well, the no-till system was found to be more enriched functionally as compared to plow-till (Fig. 3.5). These results agree with those seen in the previous study (Chapter 2), as well as previous research conducted on the Hoytville site (Puget & Lal, 2005; West & Post, 2002; Dick et al., 1991; Dick, 1984).

In the case of crop rotations however, such a clear difference was not seen. The link between crop rotation and functional enrichment is tenuous at best (Fig. 3.6).

![Figure 3.5 No-till/plow-till (NT/PT) gene abundance ratios in Hoytville samples.](image)
Significant difference was observed between the no-till and plow-till treatment for carbon fixation, as indicated by Welch’s t-test (p<0.05). No significant differences could be observed between any of the other enzyme categories under tillage, or for any of these enzymes under the crop rotation variable. In the case of the pyrosequencing results (Chapter 2), it was argued that sequencing depth could be a limiting factor for PICRUSt predictions. Here, however, Illumina sequencing is known to provide superior sequencing depth and quality and thus, should compensate for this shortcoming. The PICRUSt predictions for the functional categories previously described are detailed in the following sections.

3.3.4 Nitrogen metabolism

Forty-three nitrogen metabolism-related enzymes were identified from the KEGG database by PICRUSt. The most differentially enriched enzymes (by gene abundance ratios) are represented in Figs. 3.7 and 3.8. To provide a brief overview of the enzymes in this category, the ten most abundant enzymes involved in nitrogen
metabolism are shown in Table 3.4. Most of the enzymes enriched by no-tillage and the corn-soybean crop rotation are the same, which is attributed to the higher gene abundances for those enzymes (hence, resulting in greater contrasting ratios under different treatments) or because some crucial enzymes are more influenced by the land management. For better understanding, they were grouped into four distinct processes – nitrogen fixation, denitrification, nitrification, and assimilatory and dissimilatory nitrate reduction.

Nitrogen fixation is the process by which atmospheric nitrogen is converted to ammonia, and this is an essential function of soil microbiota that contributes to soil health and productivity. Three \textit{nif} genes (\textit{nifK}, \textit{nifH} and \textit{nifD}), coding for different subunits of the nitrogenase enzyme, were found as potential end products of various OTUs present in these soil samples. These genes did not show much enrichment under either tillage treatment or cropping system, and were above average in terms of relative gene abundance, compared to the other enzymes in the nitrogen metabolism category. All three \textit{nif} genes showed similar levels of gene abundance across all the samples, with the exception of the forest and grassland control samples, which had elevated gene abundance. The \textit{nifH} gene encodes the nitrogenase iron protein, and is putatively expressed by \textit{Rhodoplanes sp.}, \textit{Hyphomicrobium sp.}, \textit{Geobacter sp.}, \textit{Bradyrhizobium sp.}, \textit{Balneimonas sp.} and \textit{Bosea sp.} (\textit{Bradyrhizobiaceae} family), the \textit{Syntrophobacteraceae} family, the \textit{Sphingomonadaceae} family, the \textit{Entotheonellaceae} family (\textit{Proteobacteria}), the Ellin5301 and Ellin5290 families (\textit{Gemmatacomonadetes}), the \textit{Sporichthyaceae} family, \textit{Frankia sp.} (\textit{Actinobacteria}), the \textit{Nitrospiraceae} family, \textit{Opitutus sp.} (\textit{Verrucomicrobia}) and the \textit{Kouleothrixaceae} family from \textit{Chloroflexi}. 

110
Figure 3.7 Nitrogen metabolism enzymes with highest no-till:plow-till gene abundance ratios. Percentages indicate contribution to the pie chart, not contribution to the overall category.
### Table 3.3 Summary of most abundant KEGG genes and Greengenes OTUs involved in nitrogen metabolism; rank is based on order of No-Till:Plow-Till gene abundance ratios.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Gene</th>
<th>Function</th>
<th>Total gene count</th>
<th>Rank</th>
<th>Most abundant OTU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glutamine synthetase</td>
<td>glnA</td>
<td>Ammonia utilization</td>
<td>2,199,257</td>
<td>22</td>
<td><em>Kribella sp.</em></td>
</tr>
<tr>
<td>Glutamate synthase (NADPH/NADH)</td>
<td>gltD</td>
<td>Glutamate formation</td>
<td>1,120,980</td>
<td>29</td>
<td><em>Kribella sp.</em></td>
</tr>
<tr>
<td>Glutamate synthase (NADPH/NADH)</td>
<td>gltB</td>
<td>Glutamate formation</td>
<td>845,149</td>
<td>30</td>
<td><em>Kribella sp.</em></td>
</tr>
</tbody>
</table>

Figure 3.8 Nitrogen metabolism enzymes with highest corn-soybean:continuous corn gene abundance ratios. Percentages indicate contribution to the pie chart, not contribution to the overall category.
Table 3.4 continued

<table>
<thead>
<tr>
<th>Protein/Enzyme</th>
<th>Gene</th>
<th>Activity</th>
<th>Abundance</th>
<th>OTU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nitrate monooxygenase</td>
<td>ncd2</td>
<td>Nitrite formation</td>
<td>481,455</td>
<td><em>Bradyrhizobium elkanii</em></td>
</tr>
<tr>
<td>Glutamate dehydrogenase (NAD(P)+)</td>
<td>gdhA</td>
<td>Ammonia formation</td>
<td>372,811</td>
<td><em>Rhodococcus sp.</em></td>
</tr>
<tr>
<td>Nitrite reductase (NADH) small subunit</td>
<td>nirD</td>
<td>Dissimilatory nitrate reduction</td>
<td>305,128</td>
<td><em>Kribella sp.</em></td>
</tr>
<tr>
<td>Glutamate dehydrogenase (NADP+)</td>
<td>gdhA</td>
<td>Ammonia formation</td>
<td>284,593</td>
<td><em>Propionibacteria-ceae</em></td>
</tr>
<tr>
<td>MFS transporter, NNP family, nitrate/nitrite transporter</td>
<td>narK</td>
<td>Nitrate assimilation</td>
<td>267,374</td>
<td><em>Kribella sp.</em></td>
</tr>
<tr>
<td>Nitrite reductase (NADH) large subunit</td>
<td>nirB</td>
<td>Dissimilatory nitrate reduction</td>
<td>230,442</td>
<td><em>Kribella sp.</em></td>
</tr>
<tr>
<td>Ferredoxin-nitrite reductase</td>
<td>nirA</td>
<td>Assimilatory nitrate reduction</td>
<td>183,874</td>
<td><em>Kribella sp.</em></td>
</tr>
</tbody>
</table>

These OTUs were identified to be contributing the most to gene abundance, and there are other rare OTUs which have not been mentioned here, but which could also have made noticeable contributions to the expression of the nitrogen fixation enzymes. A similar community of microorganisms can be seen contributing to the expression of *nifD* (nitrogenase molybdenum-iron protein alpha chain) and *nifK* (nitrogenase molybdenum-iron beta chain) genes, and this does not vary greatly across samples. Thus, the process of nitrogen fixation is performed mainly by the aforementioned genera of *Proteobacteria, Bradyrhizobia* and *Nitrospirae* in these soil samples, which is in agreement with our findings in the previous chapter.
For the process of denitrification, conversion of nitrate into nitrogen gas, three genes were identified in our samples – *norB*, *norC* and *nirK*. *norB* or nitric oxide reductase subunit B, is mainly predicted to be expressed by the *Propionibacteraceae* family, *Mycobacterium sp.*, as well as the *Phyllobacteraceae, Solibacteraceae, Xanthomonadaceae* and *Syntrophobacteraceae* families. Nitric oxide reductase subunit C or *norC*, on the other hand, is predominantly linked to *Proteobacteria* – mainly the Alpha- and Betaproteobacteria (*Methylococcaceae, Phyllobacteraceae, Nitrospiraceae* and *Comamonadaceae* families).

In contrast to the previous two denitrification enzymes (nitrogenase and nitric oxide reductase), the most abundant members producing *nirK*, or NO-forming nitrite reductase, belong to *Verrucomicrobia, Nitrospira* or *Chloroflexi* (*Nitrospirae, Chthoniobacteraceae, Propionibacteraceae, Bradyrhizobaceae* and *Phyllobacteraceae* families, genus *Nitrospira*). For both nitrogen fixing and denitrifying enzymes, the OTUs performing similar functions are predicted to be the same across all samples, even though their predicted gene abundance varies by treatment (the positive controls having highest abundance, followed by the NT treatments and lastly the PT treatments).

The major role of *Proteobacteria* and *Nitrospirae* in these processes was predicted by PICRUST for the Wooster soil samples, and links between these microbial species and nitrogen fixation/denitrification have been established in other experiments (Wertz et al., 2012; Swanner & Templeton, 2011; Seneviratne & Jayasinghearachchi, 2005; Chen et al., 2003). The predicted involvement of *Methylococcaceae* in denitrification could be due to the close coupling of anerobic
methane oxidation and ammonium reduction processes (Ettwig et al., 2009), thereby resulting in some phylogenetic similarities.

Nitrate reduction could be assimilatory or dissimilatory, and we identified four enzymes contributing to assimilatory and nine enzymes contributing to dissimilatory nitrate reduction, which contribute to the formation of ammonia in soils. narB, or ferredoxin-nitrate reductase showed uncharacteristically low abundance in the forest sample, but appears to be almost exclusively linked to *Cyanobacteria*. PICRUSt analysis indicates that nitrate reductase production/activity is concentrated mostly in the chloroplast, especially in orders *Streptophyta* and *Stramenopiles*, family *Phormidiaceae* and *Leptolyngbya* sp. *Cylindrospermopsis* sp. and *Nodularia* sp. are also implicated in *narB* expression. Although these species are known to perform nitrogen fixation in other environments, this relationship has not been established thus far in agricultural soils (Lyra et al., 2005). Thus, this pathway is definitely a potential avenue for exploration. However, the presence of the *narB* gene in *Cyanobacteria* and the mechanism of their nitrate reduction are well understood (Flores et al., 2005; Rubio et al., 1996).

*nasB*, the gene that codes for assimilatory nitrate reductase electron transfer subunit, is mostly linked to the *Actinobacteria*. The *Nocardioidaeae* family is the single most abundant producer of this subunit, and its expression is predicted in *Aeromicrobium* sp. and *Nocardioides* sp., along with some contribution from *Catellatospora* sp. and *Pimelobacter* sp. *nasA*, which codes for the catalytic subunit of the same enzyme, does not display the same affinity for *Actinobacteria*. Some *Verrucomicrobia* and *Alphaproteobacteria* OTUs were also mapped to *nasA* (genera DA101, *Candidatus Xiphinematobacter* and *Kribella*, *Bradyrhizobium elkanii*,

115
Sphingomonas azotifigens, Rhodoplanes sp.), and the Ktedonobacteria and Ellin6529 orders from phylum Chloroflexi. Interestingly, ferredoxin nitrite reductase (nirA) shows a similar community profile as nasA, with the notable exceptions being Rhodococcus sp., Mycobacterium sp., Couchioplanes sp. and Edaphobacter sp. from the Acido- and Actinobacteria.

Dissimilatory nitrate reduction is primarily governed by the nitrate and nitrite reductases (as with the assimilatory process), with the main difference being that these products are not incorporated into the metabolism of the microorganism. The large subunit of nitrite reductase (nirB) was one of the most abundant proteins involved in nitrogen metabolism, and families Chthoniobacteraceae and Propionibacteraceae were among the highest producers of this enzyme, along with Nitrospiraceae and order Ellin6529 from Chloroflexi. The small subunit (nirD) also maps to the OTUs mentioned for nirB, in addition to Bradyrhizobium elkanii, Sphingomonas sp., Pseudonocardia sp., Mycobacterium sp., and the Acidobacteriaceae family. Cytochrome c-552-associated nitrite reductase (nrfA), on the other hand, shows high levels of Chthoniobacteraceae but not Propionibacteraceae. Nitrospira sp., the Syntrophobacteraceae family and order Anaerolineae from Chloroflexi, all of which are typical for this category of enzymes, were also linked to nrfA.

Clone and RT-PCR analyses of the nir genes in soil show similar levels of high diversity associated with these genes. Bradyrhizobium sp., Hyphomicrobium zavarzinii, Rhodobacter sp., and Azotobacter sp. are some of the organisms that were found to be involved in this process (Priemé et al., 2002), which corresponds with the community profiles predicted by PICRUSt for these samples.
The NT-CC sample shows noticeably low levels of the nitrate reductase delta subunit (narJ), which is mainly linked to Actinobacteria. Similar to nasB, genera *Kribella*, *Couchioplanes*, *Cellulomonas*, *Nonomuraea*, *Mycobacterium* and *Actinoallomurus iriomotensis* were linked to narJ production. The alpha subunit of the same enzyme, narG, showed links to the same genera of *Actinobacteria* (along with *Actinomadura vinacea*), but also some *Betaproteobacteria* and *Nitrospira sp*. The narH gene (nitrate reductase beta subunit) shows similar linkages to all the aforementioned genera, along with order *Anaerolineae* and family *Rubrobacteraceae*. The gamma subunit, narI, on the other hand, is exclusively linked to the *Actinobacteria* phylum, including all the genera mentioned above. It is important to remember that these enzymes are not exclusive to any one cycle or process, and that some of these enzymes could be involved in multiple processes. The process of dissimilatory nitrate reduction, for example, is closely related to denitrification and some genes (narJ, narG, narH, narI, napA, napB) are common to both. Previous literature indicates the involvement of not just *Actinobacteria*, but also *Proteobacteria* like *Geobacter sp.*, *Nitrobacter sp.*, *Pseudomonas aeruginosa*, *Bacillus subtilis*, *E. coli*, and rare species such as *Psychrobacter arcticus*, *Halomonas halodenitrificans*, and *Marinobacter aquaeolei* in the nitrate and nitrite reduction process (Smith et al., 2007) It is possible that the overwhelming prediction of *Actinobacteria* for these genes in our samples is due to their abundance in the phylogenetic composition in the samples.

Lastly, the nodulation proteins seen in the Wooster soil were also predicted for this dataset. All the nodulation proteins (except *nodC*) were enriched under the NT treatment, and marginally higher under the corn-soybean rotation. Nodulation protein
A (nodA) was associated with *Bradyrhizobium elkanii*, *Rhodoplanes sp.*, *Pseudonocardia sp.*, *Balneimonas sp.*, *Mesorhizobium sp.*, and *Rhizobium leguminosarum*, all of which are classic examples of nodulating bacterial species. *nodB*, which is a chitooligosaccharide deacetylase, showed the same OTU-enzyme associations, as did *nodD* (a nod box-dependent transcriptional activator) and *nodC*. An interesting deviation in the case of *nodC* (N-acetylglucosaminyltransferase) was the presence of *Burkholderia glathei* in the PT-CC treatment and *Azorhizobium sp.* in the forest sample. This association of rhizobial bacteria with nodulation was consistent with our expectations, as well as numerous previous studies on leguminous plants and their associated bacteria (Zhang et al., 2014; Graham, 1992).

Thus, with nitrogen metabolism, as in the pyrosequencing studies (Chapter 2), there appears to be a well-defined association between OTUs and functions, indicating that some groups of bacteria could have evolved (perhaps symbiotically) to perform the specialized functions of nitrogen fixation and denitrification.

### 3.3.5 Sulfur metabolism

Inorganic sulfur can be metabolized by the soil microbiome either via an assimilatory or dissimilatory process, similar to nitrates. Out of the 33 enzymes related to sulfur metabolism that were predicted by PICRUST, 15 were connected to the process of sulfate reduction. As with nitrogen metabolism, tillage produced a more visible differential effect than crop rotation (see Fig.s 3.9 and 3.10). Also, the maximum differentiation was seen in the same enzymes for both tillage and crop rotation. The ten most abundant enzymes in this category are reported in Table 3.5.
Figure 3.9 Sulfur metabolism enzymes with highest no-till:plow-till gene abundance ratios. Percentages indicate contribution to the pie chart, not contribution to the overall category.
Figure 3.10 Sulfur metabolism enzymes with highest corn-soybean:continuous corn gene abundance ratios. Percentages indicate contribution to the pie chart, not contribution to the overall category.
<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Gene</th>
<th>Function</th>
<th>Total gene counts</th>
<th>Rank</th>
<th>Most abundant OTU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thiosulfate/3-mercaptopropruvat-sulfurtransferase</td>
<td>sseA</td>
<td>Sulfite formation</td>
<td>1,727,290</td>
<td>20</td>
<td><em>Kribella sp.</em></td>
</tr>
<tr>
<td>Phosphoadenosine-phosphosulfate reductase</td>
<td>cysH</td>
<td>Assimilatory sulfate reduction</td>
<td>663,717</td>
<td>25</td>
<td><em>Kribella sp.</em></td>
</tr>
<tr>
<td>Sulfate adenylyltransfer-ase subunit 2</td>
<td>cysD</td>
<td>Sulfate reduction</td>
<td>633,834</td>
<td>22</td>
<td><em>Kribella sp.</em></td>
</tr>
<tr>
<td>Bifunctional enzyme CysN/CysC</td>
<td>cysNC</td>
<td>Assimilatory sulfate reduction</td>
<td>516,080</td>
<td>19</td>
<td><em>Kribella sp.</em></td>
</tr>
<tr>
<td>Alkanesulfonate monoxygenase</td>
<td>ssuD</td>
<td>Sulfite formation</td>
<td>509,910</td>
<td>11</td>
<td><em>Bradyrhizob-ium elkanii</em></td>
</tr>
<tr>
<td>Sulfate adenylyltransfer-ase subunit 1</td>
<td>cysN</td>
<td>Sulfate reduction</td>
<td>403,814</td>
<td>30</td>
<td><em>Kribella sp.</em></td>
</tr>
<tr>
<td>FMN reductase</td>
<td>ssuE</td>
<td>Sulfite formation</td>
<td>343,984</td>
<td>31</td>
<td><em>Kribella sp.</em></td>
</tr>
<tr>
<td>Sulfate transport system substrate-binding protein</td>
<td>cysP</td>
<td>Sulfate transport</td>
<td>288,320</td>
<td>29</td>
<td><em>Bradyrhizob-ium elkanii</em></td>
</tr>
<tr>
<td>Adenylylsulfate kinase</td>
<td>cysC</td>
<td>Assimilatory sulfate reduction</td>
<td>282,341</td>
<td>14</td>
<td><em>Kribella sp.</em></td>
</tr>
<tr>
<td>Sulfate transport system permease protein</td>
<td>cysW</td>
<td>Sulfate transport</td>
<td>243,858</td>
<td>27</td>
<td><em>Bradyrhizob-ium elkanii</em></td>
</tr>
</tbody>
</table>

Table 3.4 Summary of most abundant KEGG genes and Greengenes OTUs involved in sulfur metabolism; rank is based on order of No-Till:Plow-Till gene abundance ratios

Sulfite reductase is an enzyme that regulates dissimilatory sulfate reduction, the process that converts sulfate to hydrogen sulfide. The alpha subunit (*dsrA*) is noticeably enriched by NT and the CS rotation, and is predicted to be produced by *Bradyrhizob-ium elkanii*.
mainly *Nitrospira sp.*, across all treatments. Higher gene abundance is observed under the NT-CC and NT-CS treatments, further illustrating the effect of NT.

*Pedomicrobium sp.* and *Hyphomicrobium sp.* from Alpha- and Betaproteobacteria are also involved in *dsrA* production, mainly in the positive controls and NT-CS samples. The beta subunit, *dsrB*, shows the same community distribution. *Nitrospira sp.* is not usually linked to sulfate reduction, but the *Proteobacteria* have recently been reported to perform this process (Baker et al., 2015).

Adenylsulfate reductase is an oxidoreductase that utilizes sulfite molecules, and PICRUSt predicted the presence of the alpha and beta subunits of this enzyme in our samples. While the *Nitrospira* (0319-6A21 and *Nitrospiraceae* families) are linked with the production of these genes as well, greater OTU diversity is seen in the case of this enzyme. Various members of phylum *Gemmatimonadetes* were observed in the predictions for *aprA*, mainly the N1423WL and *Gemmatimonadetes*. *aprB* is exclusively related to the *Nitrospirae* phylum, mostly to *Nitrospira sp.* High *aprB* abundances were also seen in Ellin6067, *Entotheonellaceae* and *Azohydromonas sp.* from the *Proteobacteria* also mapped to *aprB*.

The final gene found to be involved in dissimilatory sulfate reduction is *cysN*, which codes for sulfate adenyltransferase subunit 1. *cysN* is linked to *Actinobacteria* (*Kribella sp.*, *Propionibacteraceae*, , *Acidobacteria* (classes Acidobacteria-6, *Chloracidobacteria*, *Acidobacteriia*, *Koribacteraceae* and *Solibacteres*), *Gemmatimonadetes* (classes Gemm-1 and *Gemmatimonadetes*), *Verrucomicrobia* (*Chthoniobacteraceae* and *Opitutus sp.*) and *Betaproteobacteria*.

Results indicating linkage of these *apr* and *dsr* genes to the *Nitrospirae* and *Proteobacteria* phyla seem consistent with those observed with the Wooster soils, but
experimental validation will be necessary to distinguish whether these relationships arose due to phylogenetic relatedness, or a possible involvement in their real-time function.

The assimilatory sulfate reduction mechanism also converts sulfate to hydrogen sulfide, except that the sulfur gets used in the organism’s metabolism. The bifunctional enzyme CysN/CysC (cysNC) used in this pathway was predictively linked to Alphaproteobacteria and Actinobacteria (Bradyrhizobium elkanii, Sphingomonas sp., Rhodoplanes sp., Mycobacterium sp., Pedomicrobium australicum, Kribella sp., Geobacter sp., Skermanella sp., the Gaiellaceae family and Nitrospira sp. from Nitrospirae). A related gene, cysC, codes for adenylsulfate kinase, but is mostly expressed by Actinobacteria, Nitrospirae, Chloroflexi and some Acidobacteria (Kribella sp., Chthoniobacteraceae, Nitrospira sp., Gaiellaceae, Propionibacteraceae, Pseudonocardia sp., Syntrophobacteraceae, Steroidobacter sp., Candidatus Xiphimenatobacter, Skermanella sp., Acidobacteriaceae and Candidatus Solibacter).

The cysD gene, which produces subunit 2 of the sulfate adenylationase enzyme, is associated with Kribella sp., Bradyrhizobium elkanii, Rhodococcus sp., Sphingomonas sp., Mycobacterium sp., Pseudonocardia sp., Opitutus sp. and the Gaiellaceae family, and cysH (phosphoadenosine phosphosulfate reductase) showed a similar trend. Involvement of Geobacter sp., Dactylosporangium sp., Methylibium sp., Balneimonas sp., Catellatospora sp. and the Sinobacteraceae family was also seen in these two cases. The sulfite reductase hemoprotein component (cysI) was mapped to the Cytophagaceae and Chitinophagaceae families, Azoarcus sp., Cohnella sp., Brevbacillus thermoruber, Limnohabitans sp., Flavobacterium sp.,
*Kaistobacter sp.*, *Bosea sp.*, but mainly *Bradyrhizobium elkanii*, *Sphingomonas sp.*, *Rhodoplanes sp.*, *Skermanella sp.*, *Methylococcaceae*, *Pedomicrobium sp.*, *Hyphomicrobium sp.* and *Steroidobacter sp.*, in a trend that can be visualized in almost all of the *cys* genes. Consistently, *cysJ*, the sulfite reductase alpha flavoprotein component, is predicted as expressed by *Singulisphaera sp.*, *Janthinobacterium sp.*, *Ramlibacter sp.* and *Dokdonella sp.* in addition to all the OTUs mentioned previously for the *cysI* gene. This diversity of OTU-function relationships associated with the *cys* gene family was seen in the earlier study (Chapter 2) as well, but the greater genetic diversity observed in these samples has allowed the prediction of a greater diversity of OTUs involved in sulfate reduction.

*met3* or sulfate adenyltransferase is most likely coded for by the *Chloroflexi* and *Nitrospirae* (*Anaerolinae*, *Nitrospira sp.*, *Kouleothrixaceae*, *Nitrospiraceae*, Ellin6529) and shows preference for the NT-CS and NT-CC treatments over the PT ones. Similarly, the *sir* gene (related to the sulfite reductase ferredoxin enzyme) is related almost exclusively to the *Anaerolinae* family from *Chloroflexi*, along with some *Planctomycetes* (*Pirellula sp.*) and *Bacteroidetes* (*Flavobacterium sp.*). Finally, the PAPSS enzyme is exclusively to OTUs belonging to class *Deltaproteobacteria* (*Sorangium sp.*, *Chondromyces sp.* and other members of the *Polyangiaceae* family).

Thus, sulfur metabolism in these soils is governed by a relatively lower number of enzymes than nitrogen metabolism, but possibly with more diversity of bacteria influencing the overall process.
3.3.6 Hydrocarbon degradation

Most enzymes involved in the degradation of hydrocarbons appear to be more abundant under the NT treatment, and unlike other categories of enzymes, show a definite increase under the CS rotation. The ten most enriched enzymes (or genes) under the NT and CS treatments are reported in Fig.s 3.11 and 3.12 respectively. A total of 170 enzymes were predicted as potentially degrading hydrocarbon compounds in the soil, and the most important ones for each class of compounds will be covered in this section. Of these, the most abundant enzymes are listed out in Table 3.6. It is important to remember that this does not indicate presence/absence of hydrocarbon compounds or pollutants in the soil. It is more of an indicator of the hydrocarbon degrading potential of the microbial community in the soil.
Figure 3.11 Hydrocarbon degradation enzymes with highest no-till:plow-till gene abundance ratios. Percentages indicate contribution to the pie chart, not contribution to the overall category.
Figure 3.12 Hydrocarbon degradation enzymes with highest corn-soybean:continuous corn gene abundance ratios. Percentages indicate contribution to the pie chart, not contribution to the overall category.
<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Gene</th>
<th>Function</th>
<th>Total gene abundance</th>
<th>Rank</th>
<th>Most abundant OTU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enoyl-CoA hydratase</td>
<td>paaF</td>
<td>Aminobenzoate degradation</td>
<td>4,508,397</td>
<td>64</td>
<td>Bradyrhizobium elkanii</td>
</tr>
<tr>
<td>Amidase</td>
<td>amiE</td>
<td>Aminobenzoate, styrene degradation</td>
<td>1,222,030</td>
<td>80</td>
<td>Bradyrhizobium elkanii</td>
</tr>
<tr>
<td>Carboxymethylenebutenolidease</td>
<td>Nil</td>
<td>Fluorobenzoate, chlorohydrocarbon, degradation</td>
<td>1,132,346</td>
<td>105</td>
<td>Kribella sp.</td>
</tr>
<tr>
<td>3-hydroxyacyl-CoA dehydrogenase</td>
<td>fadJ</td>
<td>Limonene and pinene degradation</td>
<td>898,244</td>
<td>88</td>
<td>Kribella sp.</td>
</tr>
<tr>
<td>Alcohol dehydrogenase</td>
<td>adh</td>
<td>Chlorohydrocarbon, naphthalene degradation</td>
<td>786,076</td>
<td>102</td>
<td>Kribella sp.</td>
</tr>
<tr>
<td>S(hydroxymethyl) glutathione dehydrogenase</td>
<td>frmA</td>
<td>Chlorohydrocarbon, naphthalene degradation</td>
<td>730,316</td>
<td>60</td>
<td>Bradyrhizobium elkanii</td>
</tr>
<tr>
<td>Alkaline phosphatase D</td>
<td>phoD</td>
<td>Aminobenzoate degradation</td>
<td>706,488</td>
<td>113</td>
<td>Kribella sp.</td>
</tr>
<tr>
<td>Alcohol dehydrogenase, propanol-preferring</td>
<td>adhP</td>
<td>Chlorohydrocarbon, naphthalene degradation</td>
<td>619,144</td>
<td>94</td>
<td>Bradyrhizobium elkanii</td>
</tr>
<tr>
<td>Gluconolactonase</td>
<td>gnl</td>
<td>Caprolactam degradation</td>
<td>448,082</td>
<td>101</td>
<td>Bradyrhizobium elkanii</td>
</tr>
<tr>
<td>4-oxalocrotonate tautomerase</td>
<td>prac</td>
<td>Benzoate, dioxin, xylene degradation</td>
<td>410,810</td>
<td>112</td>
<td>Bradyrhizobium elkanii</td>
</tr>
</tbody>
</table>

Table 3.5 Summary of most abundant KEGG genes and Greengenes OTUs involved in hydrocarbon degradation; rank is based on order of No-Till:Plow-Till gene abundance ratios
**Benzoate degradation**

The majority of hydrocarbon-degrading enzymes was implicated in benzoate degradation, and the most important ones are described as follows. Benzoate-4-monooxygenase is associated with communities of *Actinobacteria* – *Nonomurea sp.*, *Kutzneria sp.*, *Actinomycetospora sp.*, *Streptomyces radiopugnans* and *Pseudonocardia halophobica*. 4-hydroxybenzoyl-CoA reductase (*hbaB, hbaD, hbaA*) appears to be linked to mostly *Proteobacteria* communities, including *Syntrophobacteraceae*, *Azoarcus sp.*, *Bosea sp.*, and *Bradyrhizobium sp.* Both these enzymes perform similar functions and show marked increase in gene abundance under the NT treatment. *hbaC* shows the same trend, with the additional involvement of *Magnetospirillum sp.*, and *Thauera sp.*, both from *Proteobacteria* as well.

The DHBD or 2,3-dihydroxybenzoate decarboxylase enzyme is also linked to a mix of *Actinobacteria* and *Proteobacteria* communities, following a prevalent trend with benzoate-degrading microorganisms. *Novosphingobium sp.*, *Phyllobacterium sp.*, *Beijerinckia sp.*, *Mycobacterium sp.*, and *Amycolatopsis sp.* are all the most abundant OTUs associated with this enzyme. Benzoylformate decarboxylase (*mdlC*) shows the same collection of associated microbiota, including some novel ones such as *Haliangium sp.*, *Plesiostis sp.*, *Rubrobacter sp.*, *Solwaraspora sp.*, *Catellatospora sp.*, *Verrucosispora sp.* and *Nannocystis sp.*

4-hydroxybenzoate decarboxylase is a low abundance enzyme predicted for these soil samples, and is linked exclusively to *Burkholderia sp.* and *Ochrobactrum sp.* from the *Alpha- and Betaproteobacteria*. p-hydroxybenzoate 3-monooxygenase (*pobA*) performs oxidative cleavage of benzoate compounds and shows high gene abundance in these samples. *Proteobacteria* and *Actinobacteria* communities are
associated with this gene, such as *Bradyrhizobium elkanii, Rhodococcus sp.*, *Propionibacteriaceae, Pedomicrobium australicum, Streptomyces sp., Skermanella sp., Rhodoplanes sp.* and *Pseudonocardia sp.* Enoyl-CoA hydratase (*paaF*) shows one of the highest predicted gene abundances in the hydrocarbon degradation category, and is also linked mostly to *Proteobacteria* and *Actinobacteria* (*Mycobacterium sp.*, *Rhodococcus sp.*, *Kribella sp.*, *Bradyrhizobium sp.* and *Pseudonocardia sp.*).

The *abmG* gene codes for 2-aminobenzoate-CoA ligase, which specifically cleaves aminobenzoate compounds. *abmG* shows a similar community profile as the previously mentioned genes, but with some rare OTUs like *Limnohabitans sp.*, *Devosia sp.*, *Euzebya sp.*, *Polaromonas sp.* and *Amycolatopsis thermoflava*, *Rubrobacter sp.* and the *Oxalobacteraceae*.

The *bad* gene family is important in benzoate degradation, and five genes from this family were observed in these PICRUSt output for these samples. *badD*, *badE*, *badF* and *badG* are all genes coding for different subunits of the benzoyl-CoA reductase enzyme. *badD* is also associated with *Proteobacteria* communities like *Pedomicrobium australicum, Bosea sp., Hyphomicrobium sp.* and *Bradyrhizobium sp.*, which is also the case with *badE, badF* and *badG*. The *badA* gene (benzoate-CoA ligase), however, is predicted for more novel OTUs such as *Geobacter sp.*, *Amycolatopsis thermoflava, Saccharomonospora sp.* and *Nannocystis sp.*, but mostly from the *Proteobacteria* phylum.

The *benA, benC* and *benB* genes code for benzoate/toluate 1,2-dioxygenase subunits. Both of these genes are associated with a mix of *Actinobacteria* and *Proteobacteria*, including but not limited to *Rhodococcus sp.*, *Amycolatopsis thermoflava, Aeromicrobium sp.*, *Nocardioides sp.*, *Kaistobacter sp.*, *Sphingomonas*
wittichii (benB) and A. thermaflava, Mycobacterium sp., Rhodococcus fascians, Geodermatophilus obscurus, Cupriavidus sp., and Gordonia sp. (benC). A related enzyme, 2-chlorobenzoate 1,2-dioxygenase, is associated with the more dominant OTUs in Proteobacteria – Novosphingobium sp., Bradyrhizobium sp., Rhizobium leguminosarum, Mycobacterium sp., and Gordonia sp.

Recently, 16S rRNA sequencing of a mock soil community to look for bacteria involved in biphenyl, naphthalene and benzoate degradation revealed a metagenome that is very similar to that predicted for benzoate (and naphthalene) degradation in our soil samples (Uhlik et al., 2012). Our previous study (Chapter 2) suggested that Proteobacteria were closely involved in benzoate degradation processes. Illumina sequencing reveals a greater diversity of OTUs, and also the strong possibility of Actinobacterial involvement in this process. This is a concept that has been suggested by some previous findings as well (Weidow, 2013).

**Chlorohydrocarbon (chloroalkanes, chloroalkenes, cyclohexane and cyclobenzene)**

degradation

Haloacetate dehalogenase (dehH) degrades 1,2-dichloroethane in particular, and is linked to Proteobacteria (Bradyrhizobium elkanii, Pseudonocardia sp., Rhodoplanes sp.), Actinobacteria (Kribella sp., Actinomycetales, Skermanella sp.) and Chloroflexi (Kouleothrixaceae). dhaA codes for haloalkane dehalogenase and is associated with an even more diverse range of OTUs – Proteobacteria (Rhodoplanes sp., Bradyrhizobium elkanii), Actinobacteria (Pseudonocardia sp., Gaiellaceae, Kribella sp., Mycobacterium sp., Nonomuraea sp., Actinoallomurus sp.), Chloroflexi (P2-11E, Ellin6529, Kiedonobacteria), and Verrucomicrobia (Chthoniobacteraceae). 2-
haloacid dehalogenase, a related enzyme, is predicted with almost every phylum in
the sample – *Firmicutes* (*Bacillus* sp.), *Proteobacteria* (*Myxococcales, Rhodocyclaceae, Herminiimonas* sp., *Rhodoplanes* sp., *Azoarcus* sp., *Methylibium* sp., *Rickettsiales*), *Actinobacteria* (*Geodermatophilaceae, Kutzneria* sp., *Nocardioidaceae, Iamia* sp., *Kribella* sp.), *Acidobacteria* (*Candidatus Solibacter*), and *Bacteroidetes* (*Chitinophagaceae*), but the most abundant OTUs are all consistent
with the trend in this category.

Allophonate hydrolase, like most enzymes in this category, links to
*Proteobacteria* (*Rhodoplanes* sp., *Bradyrhizobium elkanii*, and *Skermanella* sp.) and
*Actinobacteria* (*Mycobacterium* sp., *Pseudonocardia* sp.) as well as *Edaphobacter* from *Acidobacteria*. Finally, soluble epoxide hydrolase is also involved in the
degradation of chloroalkanes, and links to the *Acidobacteria* (*Terriglobus* sp.,
*Acidobacteriaceae*), *Actinobacteria* (*Gordonia* sp., *Streptomyces radiopugnans*) and
*Proteobacteria* (*Myxococcales*).

Thus, the pathways for chloroalkane and chloroalkene degradation are
mediated by a similar mix of *Proteobacteria* and *Actinobacteria* communities, many
of which have been similarly implicated in bioremediation-targeted experiments
(Jiménez et al., 2014; Allpress & Gowland, 1999).

**Toluene**

Catechol 1,2 dioxygenase (*catA*) carries out oxidative cleavage of catechol
compounds. The *catA* gene is closely associated with *Proteobacteria*, especially
*Herminiimonas* sp., *Bradyrhizobium elkanii*, *Rhodoplanes* sp., *Mycobacterium* sp.,
*Gaiellaceae, Rhodococcus* sp. and *Rhodospirillaceae* (all Alphaproteobacteria,
except *Gaiellaceae* which is from *Actinobacteria*). *catB* codes for muconate cycloisomerase, and degrades benzoates, fluorobenzoates and chlorobenzene as well. This gene, interestingly, is linked most often to families from phylum *Verrucomicrobia* (class *Opitutae*, families R4-41B, OPB35, and Ellin517). In this context, it may be of interest to note that a recent study described a microaerophilic genotype and phenotype in an isolated strain of *Verrucomicrobia* (Wertz et al., 2012) and that microaerophilic bacteria are usually associated with hydrocarbon degradation in different environments (Ivanov, 2011).

Almost all of the subunits of benzylsuccinyl CoA transferase were predicted for this dataset. *bbsA* does not appear to be present in the forest sample, and is linked mostly with *Azoarcus* sp. and *Geobacter* sp., which is the predominant trend. The same is seen in the case of genes *bbsB, bbsC, bbsD, bbsG,* and *bbsH.* *bbsE* alone shows slightly higher gene abundance (but still low abundance overall), and maps to *Deltaproteobacteria* (*Geobacter* sp.), *Betaproteobacteria* (*Thaurea* sp., *Azoarcus* sp. and *Burkholderia* sp.), *Actinobacteria* (*Streptosporangium pseudovulgare,* *Nonomurea* sp., *Nocardiosis* sp., *Mycobacterium vaccae,* *Gordonia* sp. and *Frankia* sp.). *bbsF* also shows low gene abundance and similar community distribution as the first few genes (*Burkholderia* sp., *Azoarcus* sp. and *Geobacter* sp.), but with 1 OTU mapping to the forest sample.

Phenol-2-monooxygenase is also involved in toluene degradation, and is linked with a wide diversity of OTUs – *Actinobacteria* (*Pseudonocardia* sp., *Salinibacterium* sp., *Mycobacterium* sp. and *Rhodococcus* sp.), *Acidobacteria* (*Acidobacteriaceae,* *Solibacterales,* and *Koribactericeae*) and *Proteobacteria* (*Rhodoplanes* sp., *Methylbium* sp., and SC-I-84 from *Betaproteobacteria*). Aryl-
alcohol dehydrogenase is another enzyme involved in the degradation of different classes of hydrocarbon compounds, including toluene and xylene. This enzyme is mainly linked to the *Actinobacteria*, especially the *Gaiellaceae, Conexibacteraceae* and *Nocardioidaceae* families and *Rhodococcus sp.*, *Actinoallomorus iriomotensis* and *Salinibacterium sp.* (*Microbacteriaceae*).

**Polycyclic Aromatic Hydrocarbon (PAH) degradation**

The enzyme cis-1,2-dihydro-1,2-dihydroxynaphthalene/dibenzothiophene dihydrodiol catalyzes the degradation of PAH compounds as well as naphthalene. This is coded for by the *nahB* gene, which is a low abundance gene putatively expressed by *Acidovorax caeni* and *Polaromonas sp.* (rare OTUs from the *Betaproteobacteria*). Another enzyme that acts on PAH and naphthalene compounds is naphthalene 1,2-dioxygenase (beta subunit, *nahAd* was predicted), which showed the same OTU association. The ferredoxin subunit of the naphthalene 1,2-dioxygenase system (*nahAb*) shows higher gene abundance than the previous two genes, and more diversity in terms of associated OTUs – *Bradyrhizobium sp.*, *Limnohabitans sp.*, *Rubrivivax gelatinosus*, *Azospirillum sp.*, *Leptothrix sp.*, *Ramlibacter sp.*, *Variovorax paradoxus*, *Hylemonella sp.*, *Pelomonas sp.*, *Achromobacter sp.*, and *Polaromonas sp.*, all from *Proteobacteria*. *nahAc*, the alpha subunit, maps to *Novosphingobium sp.*, *Sphingopyxis alaskensis*, *Acidovorax caeni*, and *Lutibacterium sp.* *nahAa* (ferredoxin-NAD(P)+ reductase) also shows the same distribution (as *nahAb*). Some of these species are known to be involved in the degradation of polychlorinated biphenyl (PCB) compounds (Leigh et al., 2006), but of all the species described above, the link with the *nah* gene family has been well
explored for *Polaromonas sp.*, and *Acidovorax sp.* (Pesce & Zylstra, 2008; Pumphrey & Madsen, 2008; Jeon et al., 2006).

The *pca* family of genes acts on the protocatechuate compounds, and various subunits of these enzymes were identified for this dataset. *pcaH* (beta subunit) is associated with a mix of *Proteobacteria* and *Actinobacteria* (*Herminiimonas sp.*, *Bradyrhizobium elkanii*, *Rhodoplanes sp.*, *Actinoallomurus iriomotensis*, *Solwaraspora sp.*, *Pseudonocardia sp.*, *Salinibacterium sp.*, *Skermanella sp.* and the *Phyllobacteriaceae*, as is *pcaG* (alpha subunit).

4-(2-carboxyphenyl)-2-oxobut-3-enoate aldolase (*phdJ*) is exclusively linked to the *Actinobacteria* (*Mycobacterium vaccae*, *Citrococcus alkalitolerans*, *Arthrobacter psychrolactophilus*, and *Micrococcus luteus*). Another member of the *phd* gene family, extradiol oxygenase (*phdF*) also shows the same close association with *Actinobacteria*, including *Amycolatopsis thermoflava*, *Mycobacterium llatzerense*, *Burkholderia sp.* and *Streptomyces sp.* The same is seen for *phdI*, 1-hydroxy-2-naphthoate dioxygenase, which is seen exclusively in *Mycobacterium vaccae*.

The *nid* genes (*nidA* and *nidB*) code for PAH dioxygenase, which catalyzes the oxidative cleavage of PAH compounds and is one of the most important enzymes in this category. *nidA* is predicted as being produced by *Mycobacterium vaccae* and *M. arupense*, as is *nidB*. 4,5-dihydroxyphthalate decarboxylase (*ph5*) shows a mix of common and novel OTUs – *Rhodoplanes sp.*, *Nonomurea sp.*, *Limnohabitans sp.*, *Kutzneria sp.*, *Edaphobacter*, *Sterptomyces ahygroscopicus*, *Actinomycetospora sp.*, *Balneimonas sp.*, *Amycolatopsis thermoflava*, *Devosia sp.*, *Polaromonas sp.*, but...
predominantly by the *Proteobacteria*. *ligB*, or protocatechuate 4,5-dioxygenase (beta chain) degrades benzoate and PAH compounds, and shows a similar distribution as *pht5*. A few OTUs are distinct though, namely *Novosphingobium sp.* and *Sphingobium sp.*, *Cryptosporangium japonicum*, *Alicyclobacillus*, and *Actinoplanes sp.* Along with *Proteobacteria*, *Actinobacteria* can also be observed with this particular gene. The *ligA* gene shows similar abundance and OTU distribution as *ligB*. Thus, the predominant OTUs involved in breakdown of toluene, chlorohydrocarbon compounds and benzoates are also predicted for the degradation of PAHs, with the exception of some specialized microorganisms. One of the probable reasons for this observation could be the structural similarities between these compounds, which could result in similar oxidative/cleavage reactions being required for their degradation. A related cause could be that all of these compounds are preferentially degraded under similar conditions (like sulfate reducing conditions), which was suggested by a study conducted on marine harbor sediments (Coates et al., 1997).

**Atrazine degradation**

As a test of the ability of PICRUSt to look at a more narrow family of genes, we looked at the atrazine degrading set of genes. The *atz* gene family is crucial for the degradation of atrazine, and different units of the *atz* genes were predicted in these soil samples. Atrazine chlorohydrolase is coded for by *atzA* and linked to dominant *Proteobacteria* (*Skermanella sp.*, *Rhodospirillaceae*, *Phyllobacterium sp.*) and *Actinobacteria* (*Streptomyces sp.*, and *Nocardioidaceae*). Hydroxyatrazine ethylaminohydrolase is encoded by the *atzB* gene, which is linked to a number of
different phyla, i.e. *Proteobacteria* (*Rhodoplanes* sp.), *Actinobacteria* (*Gaiellaceae*, *Solirubrobacterales*, *Rubrobacteraceae*, *Salinibacterium* sp., *Pseudonocardia* sp., *Sporichthya* and *Rhodococcus* sp.) and *Chloroflexi* (*P2-11E, Anaerolinae*). The *atzC* gene, associated with N-isopropylammelide isopropylaminohydrolase, maps mainly to *Actinobacteria* OTUs (*Aeromicrobium* sp., *Nocardioides* sp. and *Streptomyces* sp.). The *atzD* gene codes for cyanuric acid amidohydrolase and is associated with *Actinobacteria* OTUs – *Knoellia subterranea, Rhodococcus* sp., *Actinoallomurus* sp., *Nonomurea* sp., *Aeromicrobium* sp., *Nocardioides* sp., *Geodermatophilaceae*, and *Nakamurellaceae*. As seen in the previous chapter, these predictions are consistent with earlier reports, and this study provides an additional level of classification to the OTUs mentioned earlier.

### 3.3.7 Carbon fixation in prokaryotes

A total of sixty-two enzymes were predicted by PICRUSt for the carbon fixation category, and the enzymes with highest NT:PT and CS:CC ratios are depicted in Figs 3.13 and 3.14. As with previous categories, no pronounced differences were observed between the CS and CC rotations. In contrast, all but one enzyme (acetyl CoA carboxylase) showed high NT:PT ratios. The most abundant carbon fixation enzymes are shown in Table 3.7.

Carbon fixation by microorganisms also involves the processes that govern internal carbon metabolism in the microbial system, including the citrate cycle and related pathways. Keeping this in context, we identified four carbon metabolism pathways that play major roles in carbon sequestration by soil microorganisms – the Calvin cycle, the CAM cycle (light and dark), the reductive citrate cycle or the
Arnon-Buchanan cycle and the reductive acetyl-CoA pathway or the Wood-Ljungdahl pathway.

Figure 3.13 Carbon fixation enzymes with highest no-till:plow-till gene abundance ratios. Percentages indicate contribution to the pie chart, not contribution to the overall category.
Figure 3.14 Carbon fixation enzymes with highest corn-soybean:continuous corn gene abundance ratios. Percentages indicate contribution to the pie chart, not contribution to the overall category.
<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Gene</th>
<th>Function</th>
<th>Total gene counts</th>
<th>Rank</th>
<th>Most abundant OTU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetyl-CoA C-acyltransferase</td>
<td>atoB</td>
<td>Acetyl-CoA synthesis</td>
<td>2,291,404</td>
<td>28</td>
<td>Bradyrhizobium elkanii</td>
</tr>
<tr>
<td>Methylenetetrahydrofolate dehydrogenase (NADP+)</td>
<td>folD</td>
<td>Formate utilization</td>
<td>1,498,628</td>
<td>45</td>
<td>Kribella sp.</td>
</tr>
<tr>
<td>Acetyl-CoA synthetase</td>
<td>acs</td>
<td>Acetate breakdown</td>
<td>1,192,157</td>
<td>31</td>
<td>Bradyrhizobium elkanii</td>
</tr>
<tr>
<td>Succinate dehydrogenase, iron-sulfur subunit</td>
<td>sdhB</td>
<td>Fumarate synthesis</td>
<td>825,357</td>
<td>53</td>
<td>Kribella sp.</td>
</tr>
<tr>
<td>Succinate dehydrogenase, flavoprotein subunit</td>
<td>sdhA</td>
<td>Fumarate synthesis</td>
<td>815,811</td>
<td>40</td>
<td>Kribella sp.</td>
</tr>
<tr>
<td>Aconitate hydratase</td>
<td>acnA</td>
<td>Citrate synthesis</td>
<td>727,633</td>
<td>44</td>
<td>Kribella sp.</td>
</tr>
<tr>
<td>Succinate dehydrogenase, cytochrome b subunit</td>
<td>sdhC</td>
<td>Fumarate synthesis</td>
<td>720,188</td>
<td>48</td>
<td>Kribella sp.</td>
</tr>
<tr>
<td>Malate dehydrogenase -ase</td>
<td>mdh</td>
<td>Oxaloacetate formation</td>
<td>693,524</td>
<td>46</td>
<td>Kribella sp.</td>
</tr>
<tr>
<td>Carbon-monoxide dehydrogenase -ase small subunit</td>
<td>coxS</td>
<td>CO formation</td>
<td>678,401</td>
<td>13</td>
<td>Bradyrhizobium elkanii</td>
</tr>
<tr>
<td>Isocitrate dehydrogenase -ase</td>
<td>icd</td>
<td>Isocitrate synthesis</td>
<td>668,211</td>
<td>52</td>
<td>Kribella sp.</td>
</tr>
</tbody>
</table>

Table 3.6 Summary of most abundant KEGG genes and Greengenes OTUs involved in carbon fixation; rank is based on order of No-Till:Plow-Till gene abundance ratios
Phosphoribulokinase (prkB) catalyzed the transformation of ribulose-5-phosphate to ribulose-5-bisphosphate and maps to the Proteobacteria and Actinobacteria. OTUs in our samples associated with prkB included Rhodoplanes sp., Friedmaniella sp., Methylibium sp., Steroidobacter sp., Hyphomicrobium sp., Balneimonas sp., Bosea sp., Devosia sp., Actinomycetospora sp., Pseudonocardia sp., Pedomicrobium australicum, Methylococcaceae, Bradyrhizobium elkanii and interestingly, Nostoc from Cyanobacteria. Another major enzyme implicated in ribose-5-phosphate processing is ribose-5-phosphate isomerase, of which genes rpiA and rpiB were predicted for these samples. rpiA is associated with Proteobacteria (Rhodoplanes sp., Hyphomicrobiaceae, Oxalobacteraceae, Afifella sp., Bradyrhizobiaceae, Sphingomonas sp.) as well, but also Firmicutes (Paenibacillus sp., and other Bacillaceae), Gemmatimonadetes (class Gemm-1, Ellin5301), Planctomycetes (Singulisphaera), Bacteroidetes (Chitinophagaceae) and Chloroflexi (Anaerolinae) but the Chlorophyta order from the Cyanobacteria. rpiB is linked to the same phyla, but to slightly different OTUs (Geobacter sp., Candidatus Solibacter, Kribella sp., Knoellia subterranea, Agromyces sp., Conexibacteraceae, Gemmata sp.). However, the most abundant OTUs associated with both rbi genes are still the same. Finally, the rbcS and rbcL genes (ribulose-bisphosphate carboxylase small and large subunits) were also associated with these samples. rbcL shows a similar microbial community distribution to the rbiA and rbiB genes, as does rbcS. It is worth noting that the carbon fixation enzymes show the greatest abundances and diversity among all the functional categories described in this study.

The dark reactions of the CAM (Crassulacean acid metabolism) cycle include phosphoenolpyruvate carboxylase and malate dehydrogenase. Two genes associated
with the production of these enzymes were predicted by PICRUSt – \textit{ppc} and \textit{mdh}. The \textit{ppc} gene shows links to \textit{Herminiimonas sp.}, \textit{Rhodocyclaceae}, \textit{Azoarcus sp.}, \textit{Rhodoplanes sp.}, \textit{Nonomurea sp.}, \textit{Paenibacillus chondroitinus}, \textit{Steroidobacter sp.}, \textit{Iamia sp.}, \textit{Cytophagaceae}, \textit{Limnohabitans sp.}, as well as the dominant members involved with the Calvin cycle genes. The same was observed in the case of the \textit{mdh} gene.

The light-dependent reactions of the CAM pathway involve pyruvate dikinase (orthophosphate) and malate dehydrogenase (NADP+ enzyme, oxaloacetate-decarboxylating). \textit{ppdK} or the pyruvate dikinase also shows similar microbial community distribution, along with the more rare OTUs corresponding to Ellin506, \textit{Bdellovibirio sp.}, \textit{Kaistobacter sp.}, \textit{Rickettsiales}, \textit{Acidimicrobiales}, \textit{Erthyrobacteraceae} and \textit{Clostridium sp.}. This was also the only carbon fixation gene associated with \textit{Nitrospira sp.} and other families from the \textit{Nitrospirae} phylum. The \textit{maeB} gene shows the same trend, except for the \textit{Actinobacteria} involvement.

The reductive citrate cycle is important in the utilization of CO\textsubscript{2} and showed the highest number of carbon fixation genes associated with it. The \textit{pyc} family of genes codes for pyruvate carboxylase. The \textit{pycA} and \textit{pycB} genes, related to subunits A and B of pyruvate carboxylase, were predicted by PICRUSt. They were found associated mainly with the \textit{Proteobacteria} (\textit{Syntrophobacteraceae}, \textit{Bdellovibrio sp.}, \textit{Spirobacillales}), \textit{Planctomycetes} (\textit{Phycisphaerae}), \textit{Nitrospira} (\textit{Leptospirillaceae}, \textit{Nitrospira sp.}) and \textit{Gemmatimonadetes} (Gemm-1, Ellin5290, Ellin5301).

The \textit{por} family of genes is also involved in this cycle, and the alpha, beta, delta and gamma subunits (\textit{porA}, \textit{porB}, \textit{porD} and \textit{porG}) were predicted in these samples. \textit{porA} is linked to \textit{Geobacter sp.}, \textit{Geobacillus thermodenitrificans}, \textit{Nitrospira}
sp., Bdellovibrio sp., Azoarcus sp., Fimbriomonas sp., Nocardioidaceae, Paenibacillus sp., Nonomurea sp., as well as involvement of FAC88 from Elusibacteria. The most abundant OTU associated with porA (and porG) is, however, Nitrospira sp. porB shows links to the same OTUs as porA, except for the major involvement of Syntrophobacteraceae in addition to Nitrospira sp. In the case of porD, Nitrospira sp., Bradyrhizobium sp. and family Entotheonellaceae are the major OTUs.

Genes coding for three subunits of fumarate hydratase were also predicted for these samples – fumA, fumB and fumC. Several rare OTUs were found associated with fumA, including phyla Elusimicrobia (order FAC88), Armatimonadetes (Fimbriomonas sp.), OD1, WS3 and TM7. OTUs from phyla Gemmatimonadetes and Nitrospirae were predicted to be the most abundant expressors of fumA. The same trend was observed for fumB as well, but fumC showed lower gene abundances than the other two. The microbial communities associated with fumC were more similar to the rpi genes, however.

Four subunits of the fumarate reductase enzyme were predicted in these samples (frdA, frdB, frdC and frdD – corresponding to the flavoprotein subunit, iron-sulfur subunit, subunits C and D). frdA links to Nonomurea sp., Kutzneria sp., Pedomicrobium sp., Actinomycetospora sp., Clostridium sp., Rhodospirillaceae, but mainly with Acidobacteria (Acidobacteria-6) and Proteobacteria (Alcaligenaceae, SC-I-84) and Nocardioidaceae (Actinobacteria). frdB, frdC and frdD show relatively low gene abundances and all of them map mostly to Pedomicrobium sp., Hyphomicrobium sp., Rhodococcus fascians, Rubrivivax gelatins, Comamonadaceae, Enterobacteriaceae, Paucibacter sp., Sporomusa sp., but mostly to the Alphaproteobacteria.
Aconitate hydratase (*acnA* and *acnB*) catalyzes the conversion of citrate to isocitrate. *acnA* shows high gene abundance and high diversity of associated OTUs, similar to *rpiA*, *rpiB* and *fumC*. The only differences seen between these distributions was that *acnA* showed linkage to *Brevibacillus thermoruber* and *Brevibacillus reuszeri*. *acnB*, on the other hand, seems exclusively connected to *Proteobacteria*, with the involvement of *Dokdonella sp.*, *Chitinophaga arvensicola*, *Anaeromyxobacter sp.*, *Luteimonas sp.*, *Phormidium sp.* and *Sinobacteraceae*.

Succinate dehydrogenase is another important enzyme involved in the citrate (and reductive citrate) cycle, coded for by the *sdh* gene family. *sdhA* is related to *Herminiimonas sp.*, *Candidatus Solibacter*, *Geobacter sp.*, *Friedmanniella sp.*, *Azoarcus sp.*, *Singulisphaera sp.*, but mostly with *Actinobacteria* (*Kribella sp.*, *Rhodococcus sp.*, *Gaiellaceae, Couchioplanes sp.*), *Chloroflexi* (*Anaerolinae*) and *Gemmatimonadetes* (Gemm-1), *Bradyrhizobium elkanii*, *Nitrospirales* and *Sphingomonas sp.*. are *sdhB*, *sdhC* and *sdhD*. It is interesting to note that even though the microbial profiles associated with these genes are similar, their gene abundances vary in the order subunits B, C, A and D.

A related enzyme, succinyl-CoA synthetase, was also predicted for this dataset. The alpha and beta subunits of this enzyme are coded for by *sucD* and *sucC*. *sucD* has the same OTUs associated with it as the *sdh* genes, but *sucC* differs in having OTUs such as *Geobacter sp.*, *Gemmata sp.*, *Cytophagaceae* and *Chitinophagaceae*, *Knoellia subterranea*, *Myxococcales*, *Bradyrhizobiaceae*, *Flavobacterium sp.* and *Acidobacteria* (*Acidobacteria-6*) mapped to it.

The *kor* gene family produces 2-oxoglutarate ferredoxin oxidoreductase, which is another essential enzyme in the TCA cycle. *korA* and *korB* have the same
diversity of microorganisms associated with them, aside from the Acidobacteria.

korD, however, shows association of Acidobacteria as well, with orders Acidobacteria-6, Solibacterales, RB41, Sva0725, as well as the phylum Elusimicrobia. korC shows additional involvement of Ruminococcaceae (Firmicutes), Cyanobacteria (ML635J-21 class) and Streptomyces sp., as well as higher abundances of Anaerolinae and Nitrospiraceae.

Another TCA cycle enzyme, isocitrate dehydrogenase, is expressed by the icd gene. icd shows the same microbial community profile as the succinate-metabolizing class of genes. However, ppsA (pyruvate, water dikinase) is associated with Herminiimonas sp., Geobacter sp., Knoellia subterranea, Friedmanniella sp., Azoarcus sp., Kribella sp., Iamia sp., Streptomyces sp., Nonomuraea sp., Paenibacillus chondroitinus, Steroidobacter sp., Limnohabitans sp., Actinoallomurus sp., Sinobacteraceae, but the most abundant OTUs remain constant across most carbon fixation genes.

Another important mechanism of utilizing carbon dioxide (in the soil as well as the atmosphere) involves the carbon monoxide dehydrogenase enzyme. This enzyme is made up of three subunits – small, medium and large (coxS, coxM and coxL). coxS is linked to a diversity of phyla (Actinobacteria, Proteobacteria, Chloroflexi, Cyanobacteria), including Syntrophobacteraceae, Geobacter sp., Rhodoplanes sp., Knoellia subterranea, Conexibacteraceae, Azoarcus sp., Geodermatophilaceae, Solirubrobacteraceae, Actinoallomurus sp., Kaistobacter sp., Bdellovibrio sp., Kutzneria sp., but mainly to Bradyrhizobium elkanii, Gaiellaceae, and Mycobacterium sp., as are coxM and coxL. This is a marked difference from the limited range of Proteobacteria OTUs that we observed in the pyrosequencing study
The most important reason could be simply the increased beta diversity offered by Illumina sequencing, which could in turn lead to better OTU-function relationships by PICRUSt. Thus, cox gene abundance in the soil is influenced by a diverse mix of phyla, and may be more complex than previously predicted.

CO dehydrogenase also has two other subunits associated with it in our dataset, cooS and cooF. cooS codes for the catalytic subunit of the enzyme, and is linked to almost the same OTUs as the cox family, but with additional involvement of *Nitrospira sp.* cooS is exclusively linked to *Plantomycetes* (classes *Phycisphaerae*, OM190, vadinHA49, WD2101, Pla3, BD7-11, and C6).

The final pathway to be considered in prokaryotic carbon fixation was the Wood-Ljungdahl or the reductive acetyl-CoA pathway. The *acs* family of genes is related to metabolism of acetyl CoA, and is hence important to this pathway. *acsB*, or acetyl CoA synthetase, is linked to *Proteobacteria* (*Syntrophobacteraceae*, *Bdellovibrio*), *Acidobacteria* (*Acidobacteria*–6) and *Planctomycetes* (*Phycisphaerae*), but the most abundant OTUs are *Nitrospira sp.* and classes TK10 and *Thermomicrobia* from *Chloroflexi*. *acsC*, related to acetyl-CoA decarbonylase complex subunit gamma, shows additional involvement of *Geobacter sp.*, but the main OTUs remain the same. *acsD* is the delta subunit of the same complex, and has the same community distribution as *acsB*. Lastly, *acsE*, 5-methyltetrahydrofolate corrinoid protein methyltransferase (which is also related to methane metabolism), is linked only to *Nitrospira sp.* and *Phycisphaerae* from *Planctomycetes*.

Two other enzymes related to methane metabolism as well as carbon fixation were identified. The first of these is *fdhB1* or formate dehydrogenase (beta subunit), associated with sulfur-metabolizing bacteria (*Geobacter sp.* from
Desulfuromonadales, Bosea sp., Phycisphaerae from Planctomycetes. metF, coding for methylenetetrahydrofolate reductase (NADPH), has Kribella sp., Bradyrhizobium elkanii, Propionibacteriaceae, Sphingomonas sp., Pseudonocardia sp., Rhodoplanes sp., Skermanella sp., Couchioplanes sp., Ellin5301 (Gemmatimonadetes) and Syntrophobacteraceae as most abundant OTUs.

As seen with the pyrosequencing study (Chapter 2), carbon fixation enzymes show high abundance and diversity in terms of associated microbial communities. The different subunits of the same enzyme, however, show definite similarities to each other in terms of associated OTUs as well as gene abundances. This further substantiates the hypothesis (described in chapter 2) that enzymes involved in carbon fixation are both more abundant and more closely linked in function, as evidenced by similar gene abundance ratios of different enzyme subunits, than those in other categories.

3.3.8 Methane metabolism

The process of methane metabolism by methanogenic and methanotrophic bacteria is closely linked to greenhouse gases and climate change. Methane-related enzymes show the same differentiation between different types of tillages, but do not show much differentiation in gene abundance between the CS and CC rotations. In fact, methane metabolism enzymes seem to show a greater number of enzymes being enriched under the CC rotation than most of the other categories studied in this chapter. Fig.s 3.15 and 3.16 depict the enzymes with highest NT:PT and CS:CC ratios respectively. The enzymes with highest predicted gene abundances are shown in Table 3.8.
Figure 3.15 Methane metabolism enzymes with highest no-till:plow-till gene abundance ratios. Percentages indicate contribution to the pie chart, not contribution to the overall category.
Figure 3.16 Methane metabolism enzymes with highest corn-soybean:continuous corn gene abundance ratios. Percentages indicate contribution to the pie chart, not contribution to the overall category.
<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Gene</th>
<th>Function</th>
<th>Total gene counts</th>
<th>Rank</th>
<th>Most abundant OTU</th>
</tr>
</thead>
<tbody>
<tr>
<td>5,10-methylenetetrahydro-methanopterin reductase</td>
<td>mer</td>
<td>Methanogenesis</td>
<td>309,913</td>
<td>3</td>
<td>Kribella sp.</td>
</tr>
<tr>
<td>Formylmethanofuran dehydrogenase subunit E</td>
<td>fwdE</td>
<td>Methanogenesis</td>
<td>246,352</td>
<td>13</td>
<td>envOPS12 (c. Anaerolinae)</td>
</tr>
<tr>
<td>Heterodisulfide reductase subunit A</td>
<td>hdrA</td>
<td>Methanogenesis</td>
<td>119,088</td>
<td>4</td>
<td>0319-6A21 (c. Nitrospirales)</td>
</tr>
<tr>
<td>Trimethylamine-corrinoid protein Co-methyltransferase</td>
<td>mttB</td>
<td>Methanogenesis</td>
<td>84,355</td>
<td>12</td>
<td>envOPS12 (c. Anaerolinae)</td>
</tr>
<tr>
<td>Heterodisulfide reductase subunit B</td>
<td>hdrB</td>
<td>Methanogenesis</td>
<td>43,035</td>
<td>7</td>
<td>0319-6A21 (c. Nitrospirales)</td>
</tr>
<tr>
<td>Methenyltetrahydromethanopterin cyclohydrolase</td>
<td>mch</td>
<td>Methanogenesis</td>
<td>42,801</td>
<td>11</td>
<td>envOPS12 (c. Anaerolinae)</td>
</tr>
<tr>
<td>Heterodisulfide reductase subunit C</td>
<td>hdrC</td>
<td>Methanogenesis</td>
<td>39,401</td>
<td>6</td>
<td>0319-6A21 (c. Nitrospirales)</td>
</tr>
<tr>
<td>Formylmethanofuran dehydrogenase subunit A</td>
<td>fwdA</td>
<td>Methanogenesis</td>
<td>23,986</td>
<td>5</td>
<td>Methyloccaceae</td>
</tr>
<tr>
<td>Formylmethanofuran dehydrogenase subunit B</td>
<td>fwdB</td>
<td>Methanogenesis</td>
<td>22,455</td>
<td>8</td>
<td>Ellin6067</td>
</tr>
<tr>
<td>Formylmethanofuran dehydrogenase subunit C</td>
<td>fwdC</td>
<td>Methanogenesis</td>
<td>22,455</td>
<td>9</td>
<td>Ellin6067</td>
</tr>
</tbody>
</table>

Table 3.7 Summary of most abundant KEGG genes and Greengenes OTUs involved in methane metabolism; rank is based on order of No-Till:Plow-Till gene abundance ratios

A total of twenty-one enzymes were identified as specific to methane metabolism, out of which methanogenesis and methane oxidation are the two major processes. The methane oxidation process converts methane to formaldehyde and is carried out by methanotrophs. Methanotrophs are often classified as either Type I (usually Gammaproteobacteria, utilize the RuMP pathway for carbon assimilation) or
Type II (Alphaproteobacteria, use the serine pathway for the same function) (Hanson & Hanson, 1996).

Methanol dehydrogenase is an important enzyme involved in methane oxidation, and subunits 1 and 2 of cytochrome c-associated methanol dehydrogenase were predicted by PICRUSSt. mdh1 and mdh2 are connected to methanotrophic families such as Methylobacteriaceae, Methylococcaceae and Methylocystaceae from the Alphaproteobacteria. Another related family from Alphaproteobacteria (belonging to the same Rhizobiales order) is Bradyrhizobiaceae, which also links to mdh1 and mdh2 synthesis. It is possible that Bradyrhizobiaceae showed up on a methanotrophic functional prediction mainly because of higher OTU abundance and close phylogenetic relationship with some methanotrophs. Other similar microbial members include Beijerinckia sp., Balneimonas sp., Labrys sp. and Hyphomicrobium sp. However, a Ce(3+)-induced gene analysis study of Bradyrhizobium sp. showed that Ce(3+)-induced methanol dehydrogenase may be involved in the methanol metabolism of this species (Fitriyanto et al., 2011).

Another important set of genes involved in methane oxidation is the pmo genes, namely pmoA, pmoB and pmoC. These genes code for methane monooxygenase subunits A, B and C. As with chapter 2, the pmoC subunit shows slightly higher gene abundances in all of the samples. Nitrosomonas tenuis and other members of the Nitrosomonadaceae family were mapped to pmoA function (possibly due to the high phylogenetic similarity between the pmo and amo genes), and pmoB reflected this trend. The only exception in the case of pmoB function was the presence of Crenothrix sp. from the Methylococcaceae family (also seen in previous experiments) ((Stoecker et al., 2006). pmoC is linked to all the microbial species
predicted for *pmoA* and *pmoB*, as well as *Mycobacterium vaccae* (from *Actinomycetales*).

Methanogenesis is the process of methane formation, and this can involve different substrates. The *mtrH* gene (which putatively expresses the tetrahydromethanopterin S-methyltransferase subunit H) is optimally expressed under the PT-CS treatment, which is unusual compared to the data we have seen thus far. This enzyme aids in the production of methane from acetate, and is exclusively linked to *Clostridia* from the *Firmicutes* phylum (*Desulfosporosinus meridiei*, *Pelosinus sp.*, and the *Veillonellaceae* family). It is unclear whether this association is a characteristic of this enzyme, or just a limitation caused by the microorganisms present in these soil samples. It should be noted that this enzyme has usually been reported in methanogenic archaea, with sequence similarity to *E. coli* and *Clostridium thermoaceticum* (Hippler & Thauer, 1999). Another enzyme uses methanol as a substrate to generate methane, and it is coded for by the *mtaB* gene (methanol-5-hydroxybenzimidazolylcobamide Co-methyltransferase). *mtaB* is predictively associated with the same OTUs, but is absent in the NT-CC sample. This gene also shows the lowest gene abundance among all the enzymes involved in methane metabolism. It is interesting to note the prediction of *Desulfosporosinus meridiei* for methane metabolism, given that this species was recently discovered as an active member in sulfur metabolism (Pester et al., 2010). It would be worth investigating whether this organism is actually involved in methane metabolism, given that sulfur-reducing and methane-metabolizing bacteria can sometimes thrive in similar anaerobic environments.
The trimethylamine-corrinoid proteins help in the conversion of the methylamine group of compounds to methane. The *mttB* and *mttC* genes code for the protein-associated Co-methyltransferase and the corrinoid protein itself, respectively.

The *Chloroflexi* and *Alphaproteobacteria* were the most abundant microbes associated with *mttB* (class *Anaerolinae*, order envOPS12; *Skermanella* sp., *Mesorhizobium* sp., *Phyllobacteriaceae* family, *Afifella* sp.), but in the case of *mttC*, *Opitutus* sp. from *Verrucomicrobia* was the unanimous dominant OTU. *Pelosinus* sp. and *Desulfosporosinus meridiei* were also linked to *mttC*. Most of the enzymes involved in methane formation, however, use CO₂ as a substrate. *mtrA* (a relative of *mtrH*), encodes tetrahydromethanopterin S-methyltransferase subunit A and is only present in the CC (NT and PT) and Forest samples. *Magnetospirillum* sp. from *Alphaproteobacteria* is responsible for *mtrA* expression under all three conditions.

The *fwd* family of genes codes for the enzyme formylmethanofuran dehydrogenase, which is important for the conversion of CO₂ to methane. The *fwdA* gene is associated with *Planctomyces* (*Isopharaceae* and *Pirellelulaceae* family, *Planctomyces* sp.), and *Proteobacteria* (*Methylococcaceae*, *Rhodocyclaceae*, *Methylibium* sp., *Hyphomicrobium* sp., *Methyllosinus* sp. and *Rubrivivax gelatinosus*). *fwdB* shows a similar microbial community distribution associated with it, as does *fwdC*. The *fmdE/fwdE* gene is associated with a variety of OTUs, including *Chloroflexi*, *Proteobacteria*, *Acidobacteria* and *Nitrospirae* (*Candidatus Solibacter*, *Nitrospirales*, *Syntrophobacteraceae*, *Acidobacteriaceae*, *Kouleothrixaceae* and *Koribacteraceae* are the most abundant).

Heterodisulfide reductase is typically found in methanogenic archaea (Hedderich et al., 2005), and our samples were predicted to contain HDR activity as
well. Three genes associated with heterodisulfide reductase (*hdrA*, *hdrB* and *hdrC*) were predicted for these samples. *hdrA* also shows affiliation with a large diversity of OTUs, including *Proteobacteria*, *Cyanobacteria*, *Firmicutes*, *Nitrospirae*, and *Chloroflexi*, but most abundantly with the *Nitrospirales* order and the *Syntrophobacteraceae* family. *hrdB* is linked to *Bdellovibrio sp.*, *Geobacter sp.*, *Alcicyclobacillus sp.*, *Saccharomonaspora sp.* and *Cylindrospermopsis sp.*, but has the same dominant OTUs as *hrdA*, as does *hrdC*.

*mer* or the 5,10-methylenetetrahydromethanopterin reductase is mainly associated with the *Chloroflexi*, *Proteobacteria* and *Actinobacteria* (*Kribella sp.*, *Rhodococcus sp.*, *Mycobacterium sp.*, *Gaiellaceae* family, *Sporichthyaceae* family and *Couchioplanes sp.*). A closely related enzyme, methylenetetrahydromethanopterin cyclohydrolase (*mch*) shows involvement of *Chloroflexi* and *Proteobacteria* as well but *Planctomycetes* instead of *Actinobacteria*, mainly the *Anaerolinae* family, Ellin6067, *Methylococcaceae*, *Balneimonas sp.* and *Methylibium sp.* Another related enzyme, formylmethanofuran-tetrahydromethanopterin N-formyltransferase (*ftr*) shows a similar distribution to *mch* overall, but is dominantly linked with *Methylibium sp.*, the *Methylococcaceae* family, *Balneimonas sp.*, *Pirellula sp.* and *Gemmata sp.*

The OTUs described as being involved in methane metabolism here were not the only methanotrophic/methanogenic OTUs identified in our samples. However, PICRUSt predicted greater methane metabolizing activity for some other OTUs (that are not typically associated with this process). This is very likely due to the low abundance (but high specificity) of methane metabolizing microorganisms.
3.4 SUMMARY

Illumina sequencing has provided us with a more accurate and deeper insight into the soil microbiome than pyrosequencing, as seen from the community composition and diversity estimates. The pyrosequencing study (Chapter 2) conducted on the Wooster soil samples served as a pilot study for the evaluation of PICRUSt as a prediction tool, but the information gleaned from this study adds substantially to our results. Aside from the relationships established among a number of novel OTUs with rare or low abundance enzymes, this dataset shows a more coherent picture of the abundance of these genes in the soil. Peculiarities related to the metabolism of certain groups of compounds could be seen more clearly, which could help in formulating future hypotheses to investigate the regulation of these metabolic pathways.

High gene abundances were seen specifically in enzymes related to sulfur metabolism and carbon fixation. Processes like nitrogen fixation and methane metabolism, however, appear to be more specialized across OTUs. Few microorganisms appear to have the genes necessary to perform these functions, suggesting certain the presence of certain functional niches in the soil. Between the two variables, only tillage produced a significant effect on the processes studied with no-till showing higher carbon fixing potential than plow-till.
REFERENCES


160


CHAPTER 4: SUMMARY AND FUTURE DIRECTIONS

4.1 SUMMARY

The studies described in this thesis used a predictive algorithm to glean information on the specific processes being performed by key players of the soil bacterial community. Given the importance of soil tillage and crop rotation variables in agricultural and environmental issues, I decided to use a predictive functional approach to elucidate bacterial species–function relationships as affected by these variables. Bacterial community profiles were compared across two tillage systems (i.e. no-tillage and plow tillage) and two crop rotations (i.e. continuous corn and corn rotation every other year with soybean) in plots located near Wooster and Hoytville in Northern Ohio. 16S rRNA gene-targeted sequencing was performed and the raw data obtained were filtered, denoised and processed using QIIME. Open-reference OTU picking and taxonomic assignment was performed using the Greengenes database. I then used a computational approach called PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) to predict metagenomes and the most likely functions performed by individual species of bacteria. Five important biochemical processes (nitrogen metabolism, sulfur metabolism, hydrocarbon degradation, carbon fixation and methane metabolism) occurring in the soil were examined, and valuable information on the roles played by various enzymes and microbes was obtained.
Differences in reference genome mapping were apparent between the Roche 454 and Illumina sequencing platforms, which were also evident in the number of OTUs identified per sample. The no-till treatment showed higher phylogenetic and functional diversity in most cases, compared to the plow-till treatment. Differences were not significant between the two crop rotations studied.

Nitrogen cycling enzymes showed selective association by bacteria, usually among the *Nitrospirae* or *Bradyrhizobia* phyla. This is consistent with earlier reports, and reinforces the idea that certain functions like nitrogen fixation could be evolutionarily conserved in certain species. The same was seen in the case of hydrocarbon degrading genes, with specific microorganisms (e.g. high numbers of Actinobacteria) being linked with the breakdown of specific classes of compounds.

Sulfur and carbon cycling enzymes were more diverse and abundant, as seen from both predicted gene abundances as well as number of OTUs linked to each enzyme/protein. Carbon fixation, in particular, was significantly increased under the no-till system in both studies. Methane metabolism, a specific carbon function in soil, was neither abundant nor specialized and there were very few methanogenic or methanotrophic bacteria identified in these samples. Thus, studying the mechanisms of methane generation and oxidation in soils might require sampling the soil at deeper depths and usage of targeted primers for those microorganisms.

Overall, this *in silico* analysis provides a snapshot of the processes occurring in the soil of an agroecosystem, including who the key microbial players are, which enzymes they produce and how much they contribute to each process. Some details are definitely needed to provide better resolution, such as factoring in interactions between microorganisms, tracking the real-time conversion of gene expression to

163
protein synthesis, accounting for intracellular and extracellular enzyme levels and greater replication.

4.2 FUTURE DIRECTIONS

This study is one of the first attempts at linking bacterial phylogeny to their functional attributes in the soil. Whether this information is quantitative or qualitative in nature, it sets the foundation for the formation of more specific hypotheses in soil microbiology (particularly for agroecosystems). The approach used here to study the five biochemical processes (i.e. nitrogen metabolism, sulfur metabolism, hydrocarbon degradation, carbon fixation and methane metabolism) could be extended to many other processes.

One of the most challenging aspects of such studies is determining which portion of the microbiome is actively producing enzymes (and other proteins), and involved in ecosystem processes at any given point of time. Bacterial communities are apt to vary spatially due to many other physicochemical factors. Another challenge is that many bacterial species have not yet been isolated for study. However, the recent increase in the use of molecular biology and bioinformatics techniques in soil microbiology and biochemistry has brought us closer to solving problems related to the non-culturability of microbes. Indeed, most recent studies on soil microbiomes have a molecular component. This has ushered in an era of phylogenetic analyses and discovery of many novel organisms and biomarker genes, which still have to be cultured or validated by established and traditional methods.

It is important to exercise caution while using such bioinformatics tools. Soil microbiota are not well characterized in comparison to the microbiomes from other
systems such as the human body, and so, bioinformatics analyses on such systems are conducted with limited references. This affects the accuracy and confidence intervals of these predictions. This skewing of known reference microbial genomes (with functional annotations) in favor of human tissues and organs also affects predictional capabilities. Also, the mere presence of a gene does not immediately indicate its expression. For example, the presence of hydrocarbon-degrading genes should not be taken as signs of soil contamination. However, the PICRUSt predictions do predict or provide an estimate of the functional potential of the soil microbiome. Phylogenetic and evolutionary relationships between microorganisms should be investigated thoroughly before a specific hypothesis is constructed using this data.

In a similar vein, functional gene validation (using qPCR or similar methods) and experimental validation of enzyme activity would provide validation for these results. RNA-Seq experiments could also be useful in narrowing down microbial species that are functionally active at the time of sampling, and this could help investigate differential gene expression under pre-planned treatment conditions as well.

As with all tools, PICRUSt has certain areas that could be improved upon. One such prominent need is to integrate the workflow with other databases such as the SILVA database. Intensive pre-computational steps are required in order to use databases other than Greengenes currently, especially for marker genes other than the 16S rRNA gene. Inclusion of information on the microorganisms’s habitat could also help with the interpretation of results, especially in the case of some geographically exclusive species. Lastly, minor improvements in terms of annotations for the KEGG
IDs as well as Greengenes taxonomies in the “metagenome_contributions.py” output would greatly improve the user friendliness of the program.

One possible area where predictional data could be useful is in the identification of candidate microorganisms for novel or commercially important genes. We can build upon our existing knowledge of microbial phylogeny and functional annotations to identify and test specific functional genes in microorganisms of interest (for instance, nitrogen fixation genes or genes related to the process of carbon sequestration). If found to be of significant importance, processes can be developed to optimize the expression of these genes to produce protein products that will be of commercial importance.

Conducting such in silico analyses of 16S amplicon sequencing data prior to designing functional gene analyses and/or whole genome shotgun sequencing experiments could potentially save thousands of dollars by more precisely focusing research questions, in addition to saving time spent on sample collection, processing and analysis. While such studies are not meant to be a substitute for whole genome sequencing studies, they could provide critical information on where to look during metagenomic and metatranscriptomic data analyses.

The results in this thesis provide a number of species-function relationships based on the phylogeny in soil microbiota. These are of practical value in understanding functional processes in a terrestrial ecosystem and could provide useful starting points for (1) specific investigations into nutrient cycling and other such ecosystem functions and (2) isolation of novel genes and microorganisms of scientific and commercial value.
BIBLIOGRAPHY


169


174
175


under conventional and zero tillage. Canadian Journal of Soil Science, 84, 403–410. doi:10.1016/j.still.2007.01.007


180


188