Winter 2016

METAGENOMIC ANALYSIS OF THE MICROBIAL COMMUNITY ASSOSCIATED WITH THE ROOT NODULES OF THE ACTINORHIZAL PLANT CASUARINA GLAUCA

Timothy Stephen D'Angelo
University of New Hampshire, Durham

METAGENOMIC ANALYSIS OF THE MICROBIAL COMMUNITY ASSOSCIATED WITH THE ROOT NODULES OF THE ACTINORHIZAL PLANT *CASUARINA GLAUCA*

BY

TIMOTHY S. D'ANGELO

B.S Biology, Rivier University, 2012

THESIS

Submitted to the University of New Hampshire

In Partial Fulfillment of

the Requirements for the Degree of

Master of Science

In

Microbiology

December, 2016

This thesis has been examined and approved in partial fulfillment of the requirements for the degree of Master in Microbiology by:

Louis S. Tisa, Professor of Microbiology and Genetics

Kelley Thomas, Hubbard Professor of Genomics

Serita Frey, Professor of Natural Resources

On November 21st 2016

Original approval signatures are on file with the University of New Hampshire Graduate School.

ACKNOWLEDGEMENTS

I have many people to thank for the acheivements I have accomplished so far. For this achievement, I would like to thank Dr. Tisa for allowing me to explore the actinorhizal symbiosis with the techniques that I desired to learn during my time at the University of New Hampshire. I must acknownledge all of the wisdom I gleened off of Rediet Oshone while working on various side-projects in the lab, as well as the interesting and often fun conversations about politics and culture we had that broke up research time. I would also like to thank Jackie LeMaire, Katy Parise and Mike D'Angelo for being the people that I commiserated with about the stresses and pressure of being a graduate student.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE PAGE

ABSTRACT

METAGENOMIC ANALYSIS OF THE MICROBIAL COMMUNITY ASSOSCIATED WITH THE ROOT NODULES OF THE ACTINORHIZAL PLANT *CASUARINA GLAUCA*

BY

TIMOTHY S. D'ANGELO

University of New Hampshire, December 2016

The symbiosis between *Frankia* and actinorhizal plants in the genus *Casuarina* are intentionally exploited in efforts to reclaim degraded lands and improve agricultural production through the practice of agroforestry, particularly in countries on the continent of Africa. Laboratory investigations have shown that some non-*Frankia* bacteria isolated from the root nodules of *Casuarina* plants can play active and positive roles in this symbiosis, but a culture-independent survey of the microbial community in system had not yet been conducted. To fill this knowledge-gap, a metagenomic survey of the endophytic microbial community of *Casuarina glauca* root nodules was performed at three sites in the country of Tunisia. The sites were chosen to evaluate the difference in community structure across a steep gradient of aridity. To complement this survey, a shotgun metagenomic analysis of the endophytic and epiphytic microbial community of *C. glauca* root nodules at these sites was also performed. Results of both analyses found a sharp decrease in the abundance of the symbiont *Frankia* in the semi-arid and arid samples. Other microbial taxa that were known to be active in the system as well as previously unobserved taxa were found in increased abundance in samples with decreased *Frankia* abundance. These results show a strong effect of climate on the endophytic community associated with *C. glauca* and provide insights for continued laboratory and field testing of this system.

CHAPTER 1

INTRODUCTION

Frankia and the Actinorhizal Symbiosis

Actinorhizal plants are a group of woody dicotyledons that are comprised of 200 species and are members of 8 families from 3 angiosperm orders (Benson et. al, 2004). These plants are of particular research interest due to their ability to host nitrogen-fixing microsymbionts within structures on their roots termed nodules. In contrast to legumes, which host several genera of nitrogen-fixing bacteria belonging to the proteobacteria phylum, actinorhizal plants host a mutualistic symbiotic relationship with members of the filamentous bacterial genus *Frankia,* which are members of the actinobacteria phylum (Normand et al., 2007). Depending on the taxonomy of the actinorhizal host, members of the genus *Frankia* infect the plant through a root hair infection pathway, or by a direct penetration of root cells (Berry et al., 1986; Valverde and Wall, 1998). When residing within the nodule, *Frankia* expresses genes in the *nif* operon, which produce the enzymatic machinery able to reduce atmospheric dinitrogen to usable forms such as ammonia (NH_3). Nitrogen can be a limiting nutrient in many environments, and therefore the ability of *Frankia* to convert inert atmospheric N_2 to a usable form confers a fitness advantage to the plant. *Frankia* benefits from this symbiosis by being supplied photosynthesized carbon by the plant that allows for growth and enables the energy-intensive process of nitrogen fixation.

The eight families of actinorhizal plants have unique distributions that collectively achieve a global distribution, except for the continent of Antarctica (Benson and Dawson, 2007). These plants contribute to the primary productivity of a diverse set of ecosystem and habitat types across their respective ranges. Nitrogen has been found to be a limiting nutrient controlling net primary production in most terrestrial ecosystems (LeBauer and Treseder, 2008). Due to the input of nitrogen from *Frankia*,

actinorhizal plants are an important part of the nitrogen budget in many ecosystems. Estimated nitrogen contributions vary by the taxonomy of the actinorhizal plant and location. For example, *Alnus rubra* grown in Scotland have been estimated to contribute 10 – 12 kg N Ha^{-1} in a three-year period (Wheeler et al., 1986). In certain temperate ecosystems, *Mycria gale* contributes 3.0 – 3.7 g N m^{-2} $year^{-1}$ (Schwintzer, 1984) and in tropical environments, actinorhizals in the genus *Casuarina,* are estimated to accrue 15- 94 kg N Ha^{-1} $year^{-1}$ (Dommergues, 1997). Nitrogen fixed by actinorhizal plants can be transferred to other plants either via litter degradation or direct exudation and absorption of nitrogenous compounds from roots (Roggy et al., 2004). Thus, nitrogen fixed by actinorhizal plants becomes available for other members of the local ecosystem, and plays an important role in large-scale ecosystem processes. For example, using stable isotope analysis, it has been observed that during primary succession processes on recently deglaciated areas, nitrogen fixed by actinorhizal plants *Alnus* and *Dryas* can be traced to other non-nitrogen-fixing primary successional plant species such as *Salix* (Kohls et al., 2003).

The ability of actinorhizal plants to contribute to the primary productivity of a given ecosystem has been utilized by humans in several different venues. For example, land reclamation projects in the hydropower industry have long used *Alnus spp* to reclaim degraded lands after large construction projects (Perinet et al., 1985). In these scenarios actinorhizal plants are playing analogous roles to the ecological ones previously mentioned (Kohls et al., 2003). Besides the input of nitrogen into ecosystems, actinorhizal plants and *Frankia* possess other advantageous traits used in land reclamation, such as tolerance to high levels of metals, organic pollutants and salinity. *Alnus* – infective Frankia strains have been demonstrated to be able to degrade a variety of organic pollutants (Baker et al., 2015, Rehan et al., 2014). These traits of *Frankia* have been shown to improve the health and biomass of *Alnus* used in land reclamation of hydrocarbon-polluted soils in the Canadian tar sands industry (Bissonnette et al., 2014). *Casuarina*, a native tropical actinorhizal plant, have become important in soil rehabilitation and

agroforestry operations in developing areas of the globe that have an amenable tropical environment (Dommergues, 1997). *Casuarina* are nodulated by several salt-tolerant strains of *Frankia* that have been studied for their use in reclaiming soils that have become over salinated due to extensive agriculture (Ng, 1987).

Actinorhizal plants play very important ecological and utilitarian roles. Although actinorhizal plants are not a direct agricultural crops like many legumes, with the aid of the microsymbiont *Frankia* they contribute to sustainable resource development and land reclamation, particularly in developing parts of the globe. Thus, much of the research into this system has been related to studying the unique abilities of the plants and *Frankia* to tolerate environmental stresses that can be utilized for practical purposes.

Biogeography of Actinorhizal Plants and *Frankia*

Phylogenetic analysis shows that all nodulating plants that host nitrogen-fixing bacterial endosymbionts, including actinorhizal plants and legumes, belong to one clade (Soltis et al., 1995; Swensen and Mullin, 1997). The individual families that make up actinorhizal plants are distantly related and there is phylogenetic evidence that the actinorhizal symbiosis has evolved several times (Swensen, 1996; Jeong et al., 1999). In support of this theory, different families of actinorhizal plants display specificity as to which members of the genus *Frankia* they will host as symbionts. These different families of actinorhizal plants have distinct geographical ranges that appear to coincide with the patterns of host-symbiont specificity (Normand et al., 2007). The actinorhizal plant family *Casuarinaceae* contains the genera *Casuarina*, *Allocasuarina* and *Gymnostoma*. *Casuarinaceae* has the smallest geographic range of any actinorhizal plant family and has only recently been actively introduced outside of its native range of Australia and the south pacific (Simonet et al., 1999). The other prominent families of actinorhizal plants, *Betulaceae, Myricaceae, Elaeagnaceae, Rhamnaceae* and *Rosaceae* have

much wider global distributions (Normand et al., 2007). The actinorhizal families of *Coriariaceae* and *Datiscaceae* have distjunct distributions spread across the globe (Benson and Dawson, 2007).

Members of the genus *Frankia* cluster into distinct clades that appear to coincide with the geographic range and evolutionary history of their host plants. As previously mentioned, Casuarinaceae have the smallest geographic range of any actinorhizal plant, and are considered evolutionarily basal (Simonet et al., 1999). *Frankia* Cluster 1 has two subgroups, one being the narrow host range *Casuarina*-infective group. Members of this subgroup have the narrowest host specificity and also have the smallest genomes of all sequenced *Frankia* strains. The other sub-group are considered to have intermediate host-specificity and forms symbiosis with the genus *Alder* and *Myrica*. Cluster 2 *Frankia* have only recently been isolated in pure culture and form symbiosis with plants in Rosaceae, Coriariaceae, Datiscaceae and *Ceanothus* (Gtari et al., 2015). Cluster 3 *Frankia* have the most promiscuous symbiotic range and form symbiosis with the families Rhamnaceae, Elaeagnaceae, Myricaceae, and *Gymnostoma* of the Casuarinaceae (Benson et al., 2004). *Frankia* members of cluster three, along with a fourth Cluster of non-nitrogen-fixing and non-infective strains, have the larger genomes of the sequences *Frankia* isolates (Normand et al., 2007).

The large differences in genome size and host-specificity of *Frankia* strains are also related to the global geographic distribution of *Frankia*. Analysis of *nif*H sequences from soils from five continents, not including Australia revealed a diversity of *Alnus* and *Elaeagnus* infective strains with a cosmopolitan distribution, while very few *Casuarina* infective sequences were retrieved (Mirza et al., 2009). Conversely, *Elaeagnus* compatible *Frankia* strains have been isolated from Tunisian soils, although there are no native *Elaeagnus* plants on the continent of Africa (Gtari et al., 2004; Gtari et al., 2007). Comparative genomic analysis has shown that *Casuarina*-infective strains, which have the smallest *Frankia* genomes, lack extracellular cellulases that the more cosmopolitan *Frankia* strains possess (Normand et al., 2007). This suggests that broad host-specificity *Frankia* strains are equipped to, and

persist as saprophytic soil organisms, while *Casuarina* infective strains are constrained to the geographic distribution of their hosts and may lack free-living capabilities. Molecular fingerprinting techniques have identified that *Casuarina* nodules harvested from the native geographic range of Australia host *Frankia* populations that are genetically diverse (Rouvier et al., 1996). On the contrary, it has been observed that *Casuarina* planted outside of their native range are nodulated by a single type, or very closely related types of *Frankia* (Simonet et al., 2009). It is possible that *Casuarina* infective *Frankia* have been introduced to these regions during the importation of the host plant, and that the natural diversity of *Casuarina*- infective *Frankia* is lower in introduced environments. This seems plausible, as it has been noted that nodulation of *Casuarina* in Africa is not always achieved without intentional inoculation of the plant with a compatible *Frankia* strain (Gtari et al., 2007).

Casuarina **Habitat and Uses in Africa**

The plant family *Casuarinaceae* including the genera *Casuarina*, *Allocasuarina,* and *Gymnostoma,* have a native range of encompassing Australia and the Pacific Rim. Due to their natural preference for tropical habitats, members of this family, specifically in the genus *Casuarina* have been transplanted elsewhere on the globe. Beginning in the 1800's, the genus *Casuarina* was planted extensively outside of their native range, particularly in Africa (Gtari and Dawson, 2011). Prior to this event, the continent of Africa was not devoid of actinorhizal plants, as members from *Betulaceae*, *Myricaceae*, *Coriariaceae* and *Rhamnaceae* have native ranges on the continent (Benson and Dawson, 2007; Gtari and Dawson, 2011). In their native range, *Casuarina,* especially *C. cunninghamiana* and *C. glauca,* tend to occupy riparian ecosystems in arid environments (Woolfrey and Ladd, 2001). This natural habitat preference, coupled with input of fixed nitrogen from *Frankia* has made *Casuarina* an attractive crop for many purposes on the continent of Africa.

Some of the earliest uses of *Casuarina* in Africa was for sheltering agricultural crops from wind and other elements of the environment (El Lakany, 1983). Those initial operations were performed in Egypt, where *Casuarina* is the most prevalent genera used in forestry operations (Zhong et al., 2013). These techniques were adapted to other areas, for example, in Senegal *Casuarina* are used as windbreaks for stabilization of agricultural soils, and their litter, containing fixed nitrogen, is used as fertilizer by farmers (Mailly and Margolis, 1992). As previously discussed, the input of nitrogen from actinorhizal plants can stimulate primary productivity in a given ecosystem. Nitrogen fixation has been exploited in agroforestry, where actinorhizal plants are grown alongside agricultural food crops. This enhances food production through the input of nitrogen from the actinorhizal plant, and also provides timber for fuel or structural use. This has been successfully implemented with *Casuarina* in countries such as Papua New Guinea (Bourke, 1985). The rehabilitation of lands that have been degraded by human over-use is an issue that is very important to peoples living in the arid regions of Africa (Houerou, 2000). Consequently, *Casuarina* has been transplanted to many countries in these regions for the purpose of land rehabilitation (Diagne et al., 2013).

Since the diversity of *Frankia* with the ability to nodulate *Casuarina* is thought to be lower in Africa than in the plants native range, there have been complications with achieving nodulation on field-grown *Casuarina* in Africa (Gtari et al., 2007). Field studies in Senegal have shown that pre-inoculation of *Casuarina* with compatible Frankia strains increases frequency of nodulation and rate of nitrogen fixation (Gauthier et al., 1985). Further studies have shown that nitrogen fixation is dependent on the *Frankia* strain chosen to use as an inoculant (Reddell and Bowen, 1985). These factors are important to take into consideration, as much of the benefits of using *Casuarina* for the aforementioned reasons are dependent on nitrogen input from *Frankia.* Besides the apparent lack of diversity of the *Casuarina* – infective *Frankia* in Africa, much of the microbial interactions of this system have not been investigated. *Casuarina* growing wild in Africa or those being used for particular purposes are likely influencing by a

myriad of interactions between the plant and the microbial communities in the soil they inhabit. Other interactions between *Frankia*, *Casuarina*, soil microbes and the environment might have an effect on the performance of *Casuarina* in a given application. Investigating the other microbial aspects to this system besides just *Frankia* could likely shed new light on the microbial ecology of introduced *Casuarina* growing in Africa.

Rhizosphere Microbial Ecology and Plant-Growth Promoting Bacteria

The rhizosphere is defined as the area of soil that is in contact with and under direct influence of the roots of plants (Hartmann et al., 2008). The environment of the rhizosphere is unique from soil that is unaffected by plant roots and thus comprises a niche for soil microorganisms (Walker et al., 2003). Plants exude low molecular weight (LMW) carbon compounds including amino acids, small organic acids and sugars in different proportions depending on plant type (Vancura and Hovadik, 1965). Along with LMW carbon compounds, plant roots also exude mucilage comprised of dead cells and complex carbohydrates derived from cell wall material (Knee et al., 2001). Included in this mixture of exudates are signaling molecules that mediate specific interactions between plants and soil microorganisms (Somers et al., 2004). These compounds are collectively termed root exudates and have strong influence on the structuring of the plant rhizosphere.

Both the LMW and mucilage fractions of plant root exudates have been demonstrated to act as carbon sources for microbial growth (Baudoin et al., 2003; Knee et al., 2001). Introduction of a mixture of LMW carbon compounds structurally similar to the root exudates of *Zea Mays* resulted in significantly more microbial abundance in soil mesococsms (Baudoin et al., 2003). Knee et. al (2001) fractioned pea root exudates using gel-filtration HPLC to yield only compounds with weights greater than 20,000 Daltons. Several strains of pea rhizosphere bacteria were able to sustain growth using this mixture as a sole carbon source (Knee et al., 2001). Stable-isotope labelled CO_2 experiments have shown that a

significant portion of photosynthesized carbon is respired by soil microorganisms (Weixen et al., 1993; Kuzyakov and Cheng, 2001). The input of photosynthesized carbon in the rhizosphere contrasts the oligotrophic nature of many soil environments and is thus a niche of increased microbial activity (Dennis et al., 2009).

The environment of the rhizosphere has also been shown to select for microbial communities that are distinct from bulk- soil microbial communities. Mesocosm experiments have shown that bacterial communities within the rhizosphere of soybean are more closely related than bulk soil across multiple soils that were spatial and temporally distinct (Mendes et al., 2014). Similarly, different plant species select for unique microbial communities from the bulk soil (Grayston et al., 1998; Haichar et al., 2008). By growing Wheat, Maize, Rape, and Barrel Clover under C^{13} conditions and employing Stable Isotope Probing of 16s rDNA differences in the bacterial community that are able to assimilate plant derived carbon from different plant species (Haichar et al., 2008). These studies taken together show that plant root exudates play an active role in selecting a unique rhizosphere microbial community from the soil.

Microcosm experiments have shown that *Frankia* populations in soils exhibit increased growth in the rhizosphere of host plants and non-actinorhizal plants. The abundance of *Frankia* increased 10 – 100 fold in the rhizosphere of host plants *C. equisetifolia* and *A. glutinosa* when compared to populations in bulk soils (Samant et al., 2015). In that study the most notable increase was *Casuarina*-infective *Frankia* Clade I in the rhizosphere of *C. equisetifolia*. A similar study comparing *Frankia* growth in the rhizosphere and leaf litter of the non-actinorhizal plant *Betula pendula* found that increased growth of *Alnus* and *Eleagnus* infective *Frankia* lineages, but did not detect *Casuarina* infective lineages (Samant et al., 2015b). Pure culture studies have shown that aqueous root exudates of *Casuarina cunninghamiana* elicit increased growth rates of *Frankia* CcI3. The root exudates did not support growth of *Frankia* CcI3 without propionate as an exogenous carbons source, indicating that the growth

response is caused by recognition of signaling molecule (Beauchemin et al., 2011). These findings highlight the differences in the *Frankia* clades, with *Casuarina* infective groups having the narrowest symbiotic specificity and are also the most metabolically limited. These studies show that root exudates of actinorhizal plants elicit increased growth in members of the genus *Frankia* by different mechanisms depending on the metabolic diversity of the bacteria and co-evolutionary relationship between the host and symbiont. Exposure of common soil microbes, *Streptomyces albus*, *Bradyrhizobium japonicum* and *Bacillus subtillis* to tissue extracts of *C. cunninghamiana* caused increased growth of *S. albus* while an inhibition of *B. subtilis* and *B. japonicum* (Zimpfer et al., 2004). These results demonstrate that chemical compounds produced by *Casuarina* have differential effects on common soil microbes, and thus are likely a selective pressure in the rhizosphere. The selective pressure that root exudates may be exhibiting on the soil microbial community likely results in an altered community structure within the rhizosphere of *Casuarina.* The studies outlined above on *Casuarina* and other actinorhizal plants primarily used targeted qPCR assays and microcosm studies looking at particular isolates. The complex nature of interactions between microorganisms is not fully captured in these assays. Thus culture independent techniques would provide a fuller picture of the interactions occurring in these systems.

Plant Growth Promoting Bacteria and Endophytic symbionts

Bacteria that naturally inhabit soil have been shown to act beneficially to promote the growth of plants are termed Plant Growth Promoting Bacteria(PGPB) (Compant et al., 2010). They occupy the rhizosphere and some are able to colonize an area known as the rhizoplane (the surfaced of plant roots and the superficial primary layers of the root surface) or they can be endophytic (inhabiting the interior of plant roots) (Rosenblueth and Martinez-Romero, 2006). There are several traits of PGPB that benefit plants that have been studied in detail and exploited in agriculture such as phytohormone production, nutrient acquisition and pathogen defense.

Native soil bacteria are known to produce plant-like hormones and regulate the levels of plant hormones that play a role in plant development and growth regulation. Some PGPB produce the enzyme 1-aminocyclopropane-1-carboxylate (ACC) deaminase. ACC is the precursor to the major plant hormone, ethylene. The action of this enzyme regulates ethylene levels and promotes plant growth (Glick, 2005). Plant auxins, with Indole Acetic Acid (IAA) being the primary pathway in plants, influence plant cell growth by governing regulatory cascades that export enzymes that degrade the cell wall and allow for cell expansion (Teale et al., 2006). The production of this compound by a diversity of soil bacteria in several phyla has been observed (Patten and Glick, 1996; Manulis et al., 1994). Bacterial derived auxin enhances the growth of plant roots and enhances root system development in plants (Patten and Glick, 2002; Loper and Shroth, 1986). For non-pathogenic bacteria these traits have shown to be beneficial to the plant, but auxin production has been implicated in disease formation by plant pathogens such as *Xanthomonas, Pseudomonas* and *Erwinia* (Fett et al., 1987; Manulis et al., 1998).

The production of IAA by *Frankia* residing in nodules of *Casuarina glauca in planta* has been documented and is implicated in formation and establishment of root nodules and promotion of the symbiosis (Perrin-Walker et al., 2010). Non-*Frankia* isolates from genera including *Streptomyces, Micromonospora, Actinoplanes,* and *Nocardia* have been isolated from actinorhizal nodules from *Casuarina, Alnus, Eleagnus* and *Ochetophila*. These isolates have been demonstrated to produce IAA (Ghodhbane-Gtari et al., 2009; Solans et al., 2011). Other plant hormones such as gibberellic acid and zeatin were also produced by some of actinorhizal endophytes (Solans et al., 2007). These isolates increase the nodulation and health of the actinorhizal plant *Discaria trinervis* when co-inoculated with *Frankia* (Solans et al., 2007). Soil bacteria capable of producing IAA benefit plant health through their synergistic effects on the formation of these mutualistic forms of symbiosis. It could be presumed that the promotion of stable nitrogen-fixing symbiotic interactions by PGPB therefore has benefits for ecosystems in the form of increased input of nitrogen.

Plant growth promoting bacteria can have positive effects on plant immunity through multiple mechanisms. PGPB can aid in the plant immune system by triggering induced systemic resistance, which heightens the plants immune response to potential pathogens. Inhibition of one phyla of bacteria by another through natural products or competition for resources can have implications on plant health if the member out-competed is a plant pathogen. Inoculation of cucumber with *Bacillus pumilis*, *Bacillus subtilis* or *Curtobacterium flaccumfaciens* has been shown to significantly reduce disease symptoms caused by the plant pathogen *Pseudomonas syringae* (Raupach and Kloepper, 1998). Mechanisms for pathogen suppression has been linked to production of secondary metabolites. It has been shown that volatile secondary metabolites produced by a broad range of PGPB can inhibit the growth of the plant pathogenic fungus *Rhizoctonia solani* (Kai et al., 2007).

Rhizosphere dwelling PGPB can promote plant health through increased mineral nutrient acquisition. Cationic nutrients are often a part of insoluble mineral complexes in soils, and thus challenging for plants to acquire. Bacteria produce extracellular molecules called siderophores that bind and remove iron from mineral complexes which allows it to become biologically available (Neilands, 1995). Since siderophores are exported out of the bacterial cell, they can act as a public good in the soil solution. Graminaceous plants produce phytosiderophores that act analogously to microbial siderophores (Romheld and Marschner, 1990). Dicotyledons, such as actinorhizal plants, rely on net efflux of protons in rhizosphere to solubilize iron as well as microbial siderophores for iron uptake (Crowley et al., 1991). The plant *Vigna unguiculata* can accumulate more iron when grown with a filtered supernatant containing siderophores produced by *Streptomyces acidiscabies* (Dimkpa et al., 2008). In a mesocosm study, *Vigna radiata* supplied with Hoaglans solution or a filtered supernatant of *Pseudomonas sp.* strain GRP_3 as a sole mineral nutrient source, showed significant increases in chlorophyll content and less chlorotic symptoms with the culture supernatant (Sharma and Johria, 2003).

Phosphorous, another important plant nutrient, is often bound in in soluble complexes in soil and unavailable for plants. In desert environments where calcareous soils dominate, phosphorous is often bound in insoluble calcium complexes (Lajtha and Schlesinger, 1988). Rhizosphere PGPB can increase the availability of phosphorous for plants through secretion of phosphatase enzymes that cleave C-P bonds and organic acids that dissolve mineral phosphate complexes (Rodriguez and Fraga, 1999). Inoculation of Mangrove with phosphate solubilizing *Bacillus licheniformis* increased plant growth and also increase nitrogen fixation when co-inoculated with *Phyllobacterium* (Rojas et al., 2001). Similarly, phosphorous deficiency limits nodule biomass on actinorhizal plants in the genus *Alnus*, and thus phosphorous solubilization by rhizosphere PGPB may play an important role in the actinorhizal symbiosis (Tobita et al., 2009).

Casuarina that have been transplanted to regions in Africa often inhabit poorly-evolved, low-nutrient calcareous soils (Mtimet, 2001). In such soils, cationic nutrients specifically iron and phosphorous can exist as insoluble mineral complexes that are inaccessible to plants. Members of the plant family *Casuarinaceae* form structures called cluster roots to increase root surface area under nutrient deficient conditions in order to increase nutrient acquisition (Arahou and Diem, 1997). The iron and phosphorous acquisition enhancing properties of rhizosphere PGPB are important to establishment of *Casuarina* in harsh conditions. The use of beneficial microorganisms with these properties to promote the health of *Casuarina* growing in Africa has been demonstrated (Diagne et al., 2013). Experimental inoculation of *C. equestifolia* with *Paenibacillus polymxa*, a known siderophore producer, *Frankia*, and a fungus *G. geosporum,* a known phosphate solubilizer resulted in the most nodule biomass and best plant health (Muthukumar and Udaiyan, 2010). This is attributed to the iron, nitrogen and phosphorous aquistion capabilities of the organisms, respectively. Phosphorous deficiency limits nodule biomass on actinorhizal plants in the genus *Alnus* (Tobita et al., 2009). These results taken together show that the productivity of *Frankia* and *Actinorhizal* plants is aided by PGPB that inhabit soil and the

rhizosphere. The studies mentioned early demonstrate that the root exudates of *Casuarina* display differential effects on different soil bacteria and *Frankia.* This is evidence that *Casuarina* may be selecting for beneficial bacteria that possess plant-growth-promoting traits that aid the plant in occupying harsh environments.

Symbiotic Helper Bacteria

There is a subset of PGPB that have been shown to have a more direct effect on the symbiosis between *Frankia* and actinorhizal plants. It has been noted that there are beneficial bacteria that cause an increase in nodulation of the host plant by *Frankia* and an increase in plant health when co-inoculated with *Frankia* and the host plant. Knowlton and Dawson (1983) first observed this when the symbiosis between *Frankia* ArI3 and *Alnus rubra* was enhanced when co-cultured with *Pseudomonas cepacia* (now recognized *as Burkholderia cepacia*). Bacteria with these properties have been termed "Helper" bacteria. *B. cepacia* was shown to cause root hair deformation on the host plant, which is a step in in the formation of root nodules. This phenomenon has been observed in other clades of the actinorhizal symbiosis. A screening of 50 *Casuarina cunninghamiana* rhizosphere isolates identified two members identified as belonging to the genus *Bacillus* and *Ochrobactrum* that caused a significant increase in growth of *C. cunninghamiana* in the presence of *Frankia* (Echbab et al., 2004). There results show that bacteria naturally inhabiting the rhizosphere of actinorhizal plants can play an active role in promoting the symbiosis.

Several isolates of actinobacteria from the nodules, rhizoplane and rhizosphere of actinorhizal plants have proven to be "helper" bacteria. These isolates have come from genera including *Streptomyces, Actinoplanes,* and *Micromonospora* (Solans et al., 2011; Ghodhbane-Gtari et al., 2009). Many of these isolates have been demonstrated to possess plant-growth-promoting traits previously discussed like the production of IAA and antagonism against plant pathogens (Ghodhbane-Gtari et al.,

2009). Recently, two members of the genus *Nocardia* were isolates from the root nodules of *C. gluaca* that act as helper bacteria. They are also able to form pseudo-nodules on *Casuarina* independent of the presence of *Frankia* (Ghodhbane-Gtari, unpublished). The experiments discussed demonstrate the ability and possible mechanisms that helper bacteria have on the actinorhizal symbiosis. There have been many studies published on actinobacteria and non-actinobacterial helper bacteria, showing that the traits needed to be a symbiosis-promoting species are possessed by a diverse group of bacteria. The experiments demonstrating these abilities have all been conducted primarily in axenic microcosms. There is currently a knowledge gap regarding the stoichiometry and population dynamics of these bacteria in natural environments.

The functions played by helper bacteria are nonspecific to actinorhizal plants, as similar effects have been observed in legumes. Similar to what has been seen with co-inoculation of *C. cunninghamiana* with *Frankia* and *Bacillus* (Echbab et al., 2004), the co-inoculation of the legume *Phaseolus vulgaris* and it's microsymbiont *Bradyrhizobium japonicum* with *Bacillus s*p. CECT 450 causes a significant increase in nodulation (Camacho et al., 2001). Members of the genus *Burkholderia*, which promote the actinorhizal symbiosis, also nodulate members of the leguminous shrub genus *Mimosa* (Lammel et al., 2015). The same group of benefical actinobacteria isolated from the actinorhizal plant *Ochetophila trinervis* act as helper bacteria in the symbiosis between the legume *Lotus tenuis* and *Mezorhizobium loti* (Solans et al., 2007; Solans et al., 2015). Other examples include the genus *Micromonospora,* which has been recognized as a broad-ranging nodule occupier (Trujillo et al., 2015). Representatives of the genus have been isolates from the nodules of the legume *Lupinus angustifolius* (Truijollo et al., 2007) and *Medicago sativa* (Martinez-Hidalgo et al., 2014) as well as the actinorhizal plants *Coriara myrtifolia* (Trujillo et al., 2006) and *C. equestifolia* (Niner et al., 1996). Members of the genus *Micromonospora* have been categorized as *Rhizobium* Helper Bacteria (RHB) due to their ability to increase nodulation of *Medicago sativa* by *Rhizobium* (Martinez-Hidalgo et al., 2014). Members of

Micromonospora have previously be identified as helper strains in the actinorhizal symbiosis (Solans et al., 2007; Ghodhbane-Gtari et al., 2009). These results taken together show that there is a community of soil dwelling microorganisms that promote the health of nitrogen-fixing symbiosis and thus indirectly contribute to ecosystem functions.

Due to the benefits of the symbiosis, the pre-inoculation of actinorhizal plants with *Frankia* has been used for large-scale land reclamation projects by industries such as hydro-electricity (Perinet et al., 1985) and oil extraction (Mallet and Roy, 2014). *Casuarina* grown in Africa are often used to attempt to revegetate land that has lost the natural ability to do so. The promotion of symbiosis with *Frankia* allows for the fixation of nitrogen and increased nutrients input into depleted environments (Sayed, 2011). The utility of symbiosis-promoting helper bacteria as well as fungi to improve *Casuarina* health has been actively researched and demonstrated in field in these climates (Muthukumar and Udaiyan, 2010; Diagne et al., 2013). If co-inoculation of *Casuarina* with *Frankia* and with helper bacteria can increase nodulation and rates of nitrogen fixation, these strategies can be used to improve land-reclamation efforts. Research in other systems has found that inoculation of plants with bacteria that possess advantageous traits for a particular environment can improve plant growth. For example, inoculation of wheat seedling with halotolerant species from the genera *Bacillus* and *Halobacillus* allowed for 90% more root growth than an un-inoculated control when grown under salt stress (Ramadoss et al., 2013). Therefore, it can be thought that screening for helper bacteria that also possess traits that can help alleviate abiotic stresses to *Casuarina* would be an advantageous strategy for land-reclamation efforts using this system.

Research Goals

Previous research shows that the root nodules of actinorhizal plants, including *Casuarina* growing in Africa are not monocultures of *Frankia*, but include PGPB and helper bacteria. Much of the

functional information known about how PGPB and helper bacteria has been determined through laboratory experimentation. Co-inoculation studies have shown that the interactions between *Frankia* and actinorhizal plants are influenced by members of the soil microbial community. These studies have been useful in demonstrating the importance of community processes in the formation of the actinorhizal symbiosis, and they have helped discover promising bio-inoculants for use in field projects utilizing actinorhizal plants.

The drawback from these studies is that they are culture-dependent and therefore do not include the entire community that is present in the natural environment. Culture-based techniques bias studies because they depend on easily cultured organisms and they tell us no information about natural community structure of these systems. It has long been known that vast majority of prokaryotic life has not been cultured (Amann et al., 1995). Thus it can be assumed that there are many genera of bacteria that could be important to *Casuarina* inhabiting harsh environments that have not been isolated or characterized. The goal of the work described here is to complement the current culture-based knowledge of the microbial ecology of *Casuarina* in Africa with a culture-independent study using metagenomic techniques. Using these techniques will allow us to gain new insight into actinorhizal microbial ecology and start foundations for new lines of laboratory-based inquiry.

The first major goal of this work it to get an unbiased view of what other genera besides *Frankia* are residing within *Casuarina glauca* root nodules that are growing in the field in Tunisia. Many other genera of bacteria have been isolated from root nodules, but there has not been a culture-independent assessment of the community structure. The second goal is to determine if *C. glauca* root exudates are structuring the bacterial community within the rhizosphere. In laboratory experiments, *Casuarina* root exudates have physiological effects on *Frankia* and other soil organisms (Beauchemin et al., 2011; Zimpfer et al., 2004). It is hypothesized that the root exudates of *C. glauca* shape the microbial community within the rhizosphere and that the stimulated genera may be performing some of the

nutrient-acquisition roles that PGPB play in the rhizosphere. Lastly, these questions are being compared across an environmental gradient of aridity. The genus *Casuarina* tend to grow in riparian areas within their native range, but are tolerant of arid environments (Woolfry and Ladd, 2001). *C. glauca* are used in the Humid-Mediterranean areas of Tunisia, but are also used in more arid environments. It is hypothesized that such a steep environmental gradient will have an effect on the microbial community associated with *C. glauca.* Understanding any changes in community structure may allow for better development of bio-inoculation strategies or more targeted isolation attempts for specific genera that may be important PGPB for *C. glauca* growing in harsh environments.

To achieve these goals, two different metagenomics strategies were used. Shotgun metagenomic sequencing of DNA extracted from *C. glauca* root nodules was performed. This strategy allows for reconstruction and functional annotation of genomes from organisms present in the samples (Tyson el al., 2004). Shotgun sequencing was performed on non-surface sterilized nodules to capture the endophytic community as well as the rhizoplane community. The second strategy was to use a PCR amplification based survey using the 16s ribosomal RNA (rRNA) gene as a taxonomic marker. Amplification of 16s rRNA from environmental DNA samples has greatly improved our view of the diversity of uncultured microbes (Pace and Norman, 1997). Sequencing of 16s rRNA genes was analyzed using the Quantitative Insights into Microbial Ecology pipeline (QIIME), which facilitates taxonomic identification and statistical analysis of microbial communities (Caporaso et al., 2010). Amplicon-based samples are easily replicated and were used to investigate the differences between soil and rhizosphere at the sample sites. Replicated amplicon samples were also used to validate the observations from the shotgun metagenomic dataset, which was not replicated.

CHAPTER 2

METHODS

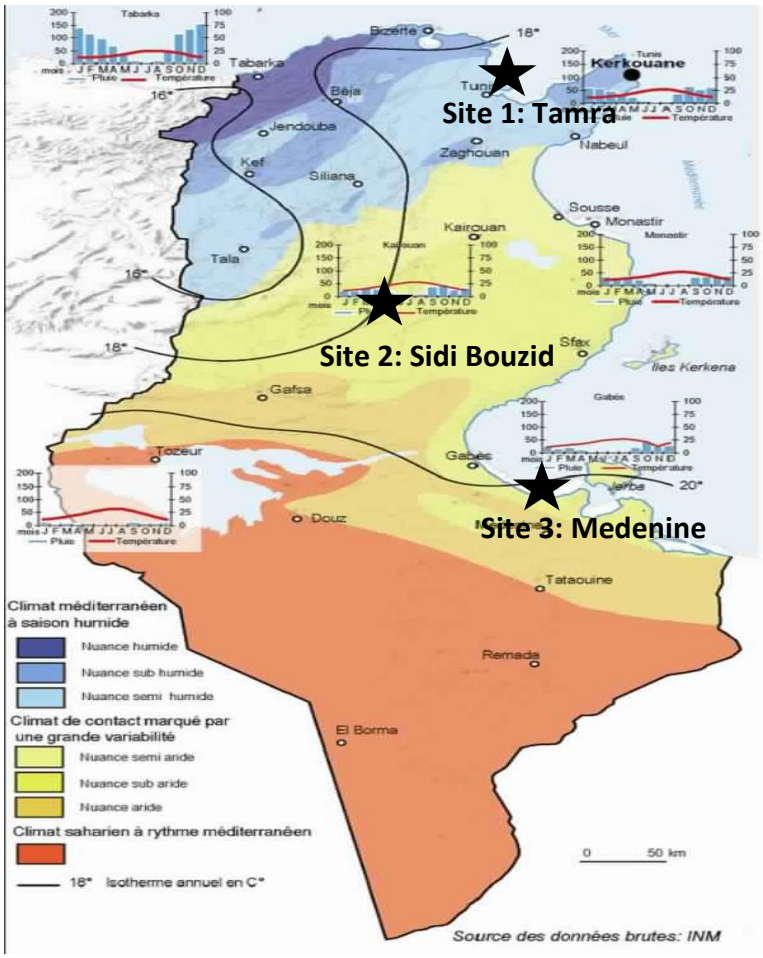
DNA Extraction for 16s rRNA Survey (Performed by collaborators in the Ghodhbane Laboratory, University of Tunis)

Three root nodule samples from *Casuarina glauca* were collected per bioclimatic region (Figure 1). Three to four root nodules were used per extraction. Nodules were surface sterilized as previously described (Ghodhbane-Gtari et al., 2009). Nodules were thoroughly rinsed with sterile H_20. Surface sterilization was conducted by shaking root nodules in solution of 30% (v/v) H_20_2 for 30 min. Sterilized root nodules were rinsed thoroughly with sterile water. Water used for rinsing was checking for sterility by plating onto nutrient agar before the DNA extraction procedure continued. Using sterile mortar and pestle the nodule tissue was homogenized using liquid nitrogen. DNA was extracted from nodule tissue using the Qiagen DNA Plant Mini Kit™, following the manufacturers protocol.

For soil and rhizosphere sample extractions, three replicates of 0.5g of soil were collected from each sampling region. DNA was extracted using the Qiagen DNA kit for soil, following the manufacturer's instructions. Bulk soil that was not in contact with the root system of any plant was collected for extraction. Rhizosphere soil was collected by removing organic litter and 15 cm of surrounding soil from around *C. glauca* roots. Soil that was attached to roots after removal of surrounding soil was collected. Nodule, soil and rhizosphere samples were collected in the month of February, 2016.

DNA sequences corresponding to the V4 hypervariable region of the 16s subunit rRNA gene were amplified using the Earth Microbiome Project 515f/806R universal primers with PCR conditions as previously described (Caporaso et al., 2012). For each sample, triplicate amplifications were performed.

The triplicates were pooled and all samples were normalized to equal DNA concentration. Paired-end sequencing of the amplification product was performed using the Illumina HiSeq2000 platform.



Site 1: Humid Nodules ⟷ Rhizosphere Soil ⟷ Bulk Soil

Site 2: Semi - Arid Nodules ⟷ Rhizosphere Soil ⟷ Bulk Soil

Site 3: Arid Nodules ⟷ Rhizosphere Soil ⟷ Bulk Soil

N = 3

Figure 1: Map and Schematic of Amplicon Survey Sampling Design: For each sample type from each site, 3 replicates were sampled.

Sequence Quality Filtering, Clustering and Taxonomic Assignment

The QIIME (Quantitative Insights into Microbial Ecology) pipeline (Caporaso et al., 2010) was initially used to process the sequence reads. Paired-end reads were joined using Fastqjoin (Erik Aronesty, 2011). Paired reads were truncated from positions where nucleotides received a Phred score of 20 or less. Reads with more than 3 consecutive low quality nucleotides were discarded. Reads with less than 75% high-quality base-call scores were also discarded. This processing resulted in a total of 176,319 reads with an average length of 253 base pairs.

The reads were clustered as described in the open reference based protocol provided in QIIME (version 1.9.0-dev) (Rideout et al., 2014). OTUs were clustered by UCLUST, using a 97% percent identity threshold to define an OTU (Edgar RC.,2010). Taxonomy of each OTU cluster was assigned using the UCLUST consensus taxonomy assigner by aligning the centroid sequence of each cluster to the Green Genes 13_8 rRNA database using PyNAST (Caporaso et al., 2010; DeSantis et al., 2006; Edgar RC. 2010). OTUs that were not able to aligned to the Green Genes 13_8 database by PyNAST were excluded from downstream analysis. OTUs that only contained one sequence (singletons) were also excluded from downstream analysis. The resulting BIOM format OTU table files were filtered to exclude all unassigned sequences, and sequences that assigned to mitochondrial or chloroplast taxonomy.

Data Normalization

The OTU table resulting from the process above was rarified to an equal sampling depth based on the smallest library size (1,762 reads) by random sub-sampling of all samples. Alternatively, Cumulative Sum Scaling (CSS), available through the metagenomeSeq R package and facilitated through QIIME, was used to normalize the OTU table (Paulson et al., 2013). Because choice of normalization method has been shown to effect downstream analysis (McMurdie and Holmes, 2014), both methods were used to compare results of normalization techniques. For Alpha diversity analysis, the original OTU

table was converted to closed-reference format and corrected for variation in 16s rRNA copy number using the online Galaxy server version of PiCRUST (Langille et al., 2013). This correction was also performed on CSS normalized OTU tables and raw OTU tables in order to assess effects on ordination results.

Diversity Analysis

Alpha diversity for each sample was calculated using the Shannon Diversity Index using scripts provided in the QIIME pipeline. Significant differences in alpha diversity between sample type (Nodule, Rhizosphere, Soil) were determined using ANOVA and the T-Test in Microsoft Excel. Alpha diversity was also calculated in the same manner for nodule samples from individual sampling locations.

Beta Diversity between samples was measured by the UniFrac distance (Lozupone and Knight, 2005). Both the weighted (quantitative) and unweighted (qualitative) versions of the distance metric were used. Both of these metrics can yield different but complementary results from the same dataset (Lazupone et al., 2007). Distance matrices produced using these metrics were ordinated using Principle Coordinates Analysis (PcoA). Ordinations were created for distance matrices produced by CSS (for both 16s copy number corrected and not corrected), for raw OTU tables with 16s operon copy number correction and for evenly rarefied OTU tables (Figures 8 – 9).

Jackknifing, which is the repeated subsampling of a dataset, was used to determine the confidence level in Beta diversity analysis. This technique was performed by randomly subsampling 100 sequences per sample from the rarefied OTU table. For each subsampling, an Unweighted Pair Group Method with Arithmetic Mean (UPGMA) tree was created using the unweighted UniFrac distance matrix between samples, based on those 100 randomly subsampled reads. This process was repeated 100 times. A consensus tree was built from the 100 trees with jackknife values for each node representing

the number of times, out of the 100 jackknifed trees, that the consensus configuration was observed (Figure 7).

Two-Way Cluster Analysis of Nodule Samples

Nodule samples were analyzed by two-way cluster analysis to investigate the difference in nodule community structure across the environmental gradient in the sampling design. An OTU table was constructed that was corrected for 16s operon copy number abundance by PiCRUST and summarized to relative abundance at the genus level. The samples were grouped by hierarchical clustering using the average linkage method. The Bray-Curtis distance between each sample based on taxonomic composition was used for sample clustering. The taxa were clustered by their abundance patterns across samples using the Bray-Curtis distance between metric. Clustering was performed using the data for all genera present. However, for visualization purposes, only taxa that made up greater than 1% of the community were displayed (Figure 10).

Significance of Ordination Groupings

Multi-Response Permutation Procedures (MRPP) is a nonparametric procedure that tests for differences between groups of samples in a distance matrix by comparing the differences between the within-group distances to a permuted randomization of the data (Mielke and Berry, 1984). This analysis was performed in R (Version 3.2.1) using the Vegan Library (Oksanen et al., 2016). MRPP was performed to determine if there were significant differences in phylogenetic distances based on the sample type (Nodule, Soil, or Rhizosphere). To determine if nodules formed a statistically significant group based on the weighted and unweighted UniFrac distance matrices, the samples were partitioned into two categories, nodule (n = 9) and non-nodule (n = 18). Significance was calculated by permutation of the distance matrices 999 times to determine if the groups had within-group distances that were

different to what would be expected by chance. All MRPP analysis was performed on the weighted and unweighted UniFrac distance matrices produced by the different normalization methods described previously (Lozupone et al., 2007, McMurdie and Holmes, 2014).

Soil and Rhizosphere samples were analyzed to test the hypothesis that *C. glauca* maintains a unique microbial community within the rhizosphere at the different sampling locations. Soil (n = 9) and Rhizosphere (n = 9) samples were from all sites were pooled together and analyzed independently of nodule samples in the same fashion as described above. To determine if there were significant groupings between soil and rhizosphere by climate type, the soil and rhizosphere samples were grouped into a humid group (n = 6) and non-humid group (n = 18). MRPP was performed as above with this grouping using the unweighted UniFrac distance matrix. This metric calculates phylogenetic distance between samples based off of the presence or absence of taxa, and is thus useful for identifying differences between samples that come from environments that host distinct communities.

Indicator Species and Differentially Abundant OTUs of Identified Groups

Indicator Species Analysis (ISA) was performed in R (Version 3.2.1) using the Indicspecies Library (De Caceres and Legendre, 2009). All ISA analysis was performed on OTU tables that were normalized by 16s rRNA copy number and summarized to relative abundance at the Genus level. The groupings chosen for ISA were determined based on the results of the MRPP tests described above. ISA was performed to determine indicator species for nodule and rhizosphere samples. ISA was also performed to determine indicator species for the pooled soil and rhizosphere samples from different climactic regions.

In conjunction with ISA on all analysis, the DESeq2 algorithm was used to identify differentially abundant OTUs between sample types (Love et al., 2014). ISA identifies species that have high fidelity and high abundance within the group being analyzed. DESeq2 identifies OTUs that are

differentially abundant in different sample types even if they are present in all types. The differences between these analyses yield different kinds of information that can be complementary. This analysis was used to determine OTUs that were differentially abundant in nodule samples when compared soil samples. To determine OTUs that were significantly different in nodule samples from the different sampling locations, this analysis was performed in a pairwise fashion on the nodule samples from the different sampling locations.

Soil and Rhizosphere Analysis

Indicator species analysis and DESeq2 differential abundance analysis was performed to determine indicator species and differentially abundant OTUs in the pooled rhizosphere samples. Alternatively, rhizosphere and soil samples from individual sites were compared to determine what OTUs were differentially abundant in rhizosphere samples from differing environments. Lastly, rhizosphere and soil samples from arid environments were grouped together and compared to rhizosphere and soil samples from the humid sampling site to determine differentially abundant OTUs between humid and non-humid samples. Indicator Species analysis was performed on pooled soil and rhizosphere samples from each sampling site.

Nodule Collection DNA Extraction for Shotgun Metagenomic Sequencing

Nodules were collected from *C. glauca* growing in three different bioclimatic regions of Tunisia in the month of March, 2015 (Figure 1). Samples sites were in Tamra, Sidi Bouzid and Gabes. These three locations represent Mediterranean-humid, semi-arid, and arid bioclimatic regions, respectively. Nodules were rinsed with sterile water prior to DNA extraction. Three to four nodules were used per DNA extraction. One set of DNA extractions for the three sampling sites were performed using the FastDNA™ Spin Kit for Soil (MP Biomedicals, Santa Ana, California, USA) per the manufacturers

instruction. DNA was resuspended in 20 µl of TE buffer. A second set of DNA extractions using three to four nodules per site were performed using the CTAB/Chloroform extraction technique (Murray and Thompson, 1980).

A phenol chloroform extraction was performed on each CTAB DNA sample to remove contaminants. Briefly, DNA samples were suspended in 100 µl TE buffered at a pH 8.0. An equal volume of Phenol: Chloroform: Isoamyl alcohol pH 8.0 was added and the tube was inverted several times. The tubes were centrifuged at 13,000 g for 2 min. The aqueous layer was removed and extracted a second time as described above. To each tube, 20 µl of 3M Sodium Acetate and 440 µl of 100% ethanol was added and the DNA was precipitated by incubation overnight at -80˚ C. After DNA precipitation, the tubes were centrifuged at 13,000 g for 30 min at 4˚ C. The supernatant was decanted and the DNA pellet was washed with 1 ml of 70% ethanol The pellet was centrifuged at 13,000 g for 5 min at C. After the ethanol was decanted, the pellet was air-dried and resuspended in 50 µl of TE buffer. Samples of 20 µl of resuspended DNA from both DNA extraction methods (FastDNA™ Spin Kit and CTAB-procedure) were combined together. The pooled DNA samples were quantified using the Qubit Fluorometric Quantitation system (ThermoFischer Scientific).

Illumina Sequencing Library Preparation and Sequencing

Sequencing libraries for the three pooled samples were prepared using the Nextera Library protocol per the manufacturer's instructions (Illumina Inc., San Diego, CA). The libraries were sequenced using paired-end reads with a length of 150 base-pairs using the Illumina HiSeq2000 by the Hubbard Center for Genome Studies (UNH, Durham, NH).

Quality Filtering of Reads

Trimmomatic was used to remove nextera adapters and filter the raw reads based on quality (Bolger et al., 2014). The following parameters were used. Three base-pairs were trimmed off of the tailing and leading ends of each read. The reads were scanned with a sliding window of four base-pairs, and sections of reads were discarded when the average quality score of the base-pairs was below 15. After this process, only reads that retained a minimum length of 36 base pairs were used for downstream processes. Of these resulting reads, only reads that retained both forward and reverse pairs were used for assembly.

Co-Assembly of Metagenomic Reads

In order to perform differential-coverage binning of draft genomes, the read sets from each of the three sampling locations were combined *in silico* and a composite co-assembly was created. SPAdes Version 3.6.0 (Bankevich et al., 2012) was used to assemble the combined reads using the metagenomic setting, without error correction. The assembly process used Khmer lengths of 21, 33, and 55 to produce the final assembled contigs.

Differential-Coverage Binning of potential Draft Genomes

Draft genomes of individual members of the metagenomic community were binned using the mmgenome package (Albertsen et al., 2013) for R (r-project.org) using the Rstudio IDE (Rstudio.com). This package facilitates distinguishing individual genomes from metagenomes primarily by the unique coverage values of assembled sequences belonging to organisms present in multiple samples, but at different abundances. Coverage values for each contig per-site was performed by individually aligning each read-set to the co-assembled contigs using the Bowtie2 alignment tool (Version 2.2.5) (Langmead et. al, 2009). The resulting Sequence Alignment Map (SAM) files were

converted to binary format and sorted to the left-most coordinates on the assembled sequences and depth of coverage per-base was calculated using Samtools (Version: 0.1.19-96b5f2294a) (Li et. al, 2009). The resulting files containing coverage per-base per-contig were converted to average coverage per-contig scripts in the Mmgenome package.

Additional information for assembled contigs was generated following the workflow as described by the mmgenome package developers (Albertsen et al., 2013). Open Reading Frames (ORFs) within the assembled metagenomic contigs were predicted and translated to amino acid sequence using the metagenomic setting of Prodigal (Version 2.6.2) (Hyatt et al., 2010). The translated amino acid fasta file of predicted ORFs was searched against Hidden Markov Models (HMMs) of 111 essential prokaryotic single copy genes provided in the mmgenome package, using Hmmsearch (S.R Eddy, 1998). ORFs identified as essential genes were extracted from the co-assembly using Perl scripts provided by the mmgenome package. Taxonomic information was assigned to each ORF identified as essential gene by searching against the refseq protein database using BlastP with an E value cutoff of $1e^{-5}$, with five maximum target sequences. The resulting Blastp output in XML format was uploaded into MEGAN (Version 5) (Huson et al., 2007). The Lowest Common Ancestor (LCA) algorithm was used to make a consensus taxonomic assignment for each ORF based off of the top five BlastP hits using the following settings; Top Percent = 5, Minimum Support = 1, Rank = Species. The resulting file was parsed using the hmm.majority.vote.pl script to derive a consensus taxonomic assignment for each assembled contig based off of the taxonomy of the ORFs it contained.

The coverage and taxonomic information produced above was loaded into the Rstudio environment along with the composite metagenomic co-assembly using BioStrings (Pages et al., 2016). The GC content and tetranucleotide frequency distribution of each assembled contig was calculated using the mmgenome R package. A principle component analysis (PCA) of normalized tetranucleotide frequencies of all contigs was performed in R using the mmgenome package. The composite co-

assembly was visualized in Rstudio and binning was performed using the tools available in the mmgenome R package. The data that was produced as described above and accompanying Rmarkdown files allow the reproduction of the binned draft genomes.

Removal of Contaminant Contigs from Draft Genome, Reassembly, Annotation

The contigs for reach resulting genome bin were converted to amino acid sequence using Prodigal (Version 2.6.2) (Hyatt et al., 2015). The resulting amino acid fasta files were searched against the refseq protein database using BlastP with an E value cutoff of $1e^{-5}$. The resulting XML files were uploaded into MEGAN Version 5 (Huson et al., 2007). The consensus taxonomic assignment of the ORFs was identified using the LCA algorithm as described above. Contigs that did not annotate with the vast majority of the contigs in a given bin were removed from the fasta files that were exported from mmgenome.

Calculation of Coverage and Relative Abundance for Comparisons

Coverage profiles of individual genomes with PCR duplicates removed were calculated in order to compare genome coverages at different sites. Reads were aligned to the metagenomic assembly using the Bowtie2 alignment tool (Version 2.2.5) (Langmead et. al, 2009). The SAM file resulting from alignment with Bowtie2 was converted to BAM format. The aligned reads were sorted to the left-most coordinate on the indexed metagenomic assembly using Samtools (Version: 0.1.19-96b5f2294a) (Li et. al, 2009). Duplicate reads, those having the exact same coordinates on the reference fasta file as another read, were removed using Picard Tools MarkDuplicates (http://broadinstitute.github.io/picard/). Coverage for the contigs was calculated from the resulting SAM file using Bedtools (v2.17.0) (Quinlan and Hall, 2010). The coverage calculations per-contig were averaged for all of the contigs within a given genome bin. The average coverage for the entire genome

was normalized by the amount of reads per data set corresponding to the different sampling sites. The percentage of the given read sets that aligned to plant-derived DNA was used to normalize each genome coverage value based on the differing amounts of plant-derived DNA in the different read sets (Figure 17).

The read sets, with plant-derived reads removed, were aligned to the *Frankia* genome using Bowtie2. The percentage of the reads that aligned to the *Frankia* genome bin were recorded from the standard output statistics in order to compare the relative abundance of *Frankia* in the shotgun data in comparison to the replicated amplicon data (Figure 5).

Mining of Metagenome for Plant-Growth-Promoting Genes and Secondary Metabolites

Reads from each site were assembled individually using SPAdes-3.6.0 using the metagenomic setting (Bankevich et al., 2012). The resulting assembled contigs were annotated using the Prokka annotation pipeline using the metagenomic settings (Seamann et al., 2014). The Genome Feature File (GFF) produced by Prokka was parsed using a custom python script to evaluate the presence and taxonomy of plant-growth-promoting genes and genes involved in cellulose, chitin and pectin degradation within the nodule community. Genes known for plant auxin production (Indole acetic acid and Phenyl acetic acid), phosphate solubilization (alkaline phosphatases and phytases), ACC Deaminase activity and degradation of chitin, cellulose and pectin were identified for each of the three data sets. The presence of these enzymes were identified by parsing the GFF file for the Enzyme Commission (EC) number assigned to a given ORF by the Prokka pipeline. Enzymes that were identified by their EC numbers were annotated by aligning the amino acid sequence of the entry against the refseq protein database using BlastP with an E value cutoff of $1e^{-5}$ and a maximum output of five entries. Cladograms of the BlastP output were constructed using the Lowest Common Ancestor (LCA) algorithm in MEGAN (Version 5.0) using the following settings; Top Percent = 5, Minimum Support = 1, Rank = Genus (Figures

18 – 20). A summary of the EC numbers used in this analysis is in Tables 1 -3. The antiSmash pipeline (version 3.0) was used to identify secondary metabolites in the binned genomes (Weber et al., 2015). AntiSmash was run using the metagenomic setting to improve gene prediction for highly fragmented metagenomic assemblies.

Table 1: Enzyme Commission (EC) Numbers used for mining metagenome for potential Indole Acetic Acid Biosynthetic genes.

Pathway	Enzyme	Enzyme Commission Number(s)
IPyA	Tryptophan-pyruvate aminotransferase	2.6.1.99, 2.6.1.27
IPyA	Indole-3-Pyruvate decarboxylase	4.1.1.74
IPyA	Indole-3-acetaldyde dehydrogrenase	1.2.1.3
IAM	Tryptophan monooxegenase	1.13.12.3
IAM	Indole-3-acetamide hydroloase	3.5.1
TAM	Tryptophan decarboxylase	4.1.1.28
TAM	Amine oxidase	1.4.3.21
IAN	Arylacetonitrilase	3.5.5.5
PAA	Phenylalinine aminotransferase	2.6.1.58
PAA	Phenylpyruvate decarboxylase	4.1.1.43
PAA	Phenylacetaldehyde dehydrogenase	1.2.1.39

Table 2: Enzyme Commission (EC) numbers used for mining Chitin, Pectin and Cellulose degrading genes

Enzyme	Action	Enzyme Commission Number(s)
Chitinase	Cleavage of Glycosidic Bonds	3.2.1.14
Pectinase	Cleaveage of polysaccharide bonds in pectin	3.3.1.15, 4.2.2.2
Exocellulase	Cleavage of small polysaccharides off the ends polymer chains	3.2.1.91
Endocellulase	Cleavage of internal bonds in cellulose polymers	3.2.1.4
Beta-glucosidase	Hydrolyze product of exocellulases into monosaccharides	3.2.1.21
1,4-beta-glucosidase	Remove of glucose units from glucans and other oligosaccharides	3.2.1.74
Cellobiohydrolase	Hydrolysis of (1-4)-beta-D-glucosidic linkages in cellulose	3.2.1.176

Table 3: Enzyme Commission (EC) numbers used for mining Alkaline Phosphatase, Phytase and ACC Deaminase genes

Enzyme	Action	Enzyme Commission Number
Alkaline Phosphatase	Dephosphorylation, Phosphate solubilization	3.1.3.1
3-Phytase	Cleavage of phytic acid	3.1.3.8
4-Phytase	Cleavage of phytic acid	3.1.3.26
5-Phytase	Cleavage of phytic acid	3.1.3.72
1-Aminocyclopropane-1-Carboxylate Deaminase	Regulation of plant ethylene levels	3.5.99.7

CHAPTER 3

RESULTS

16s Amplification

Preprocessing of the 16s amplicon sequences resulting in a total of 7,855 OTUs comprised of 149,879 paired reads. Individual library sizes ranged from 1,762 to 11,542 reads per sample, with a mean of 5,551 reads per sample. Rarefaction plots were performed in order to estimate the completeness of community sampling at the sequencing depth performed. The rarefaction plots for average observed OTUs per sample type is shown in Figure 2. These plots show that the nodule samples were sequenced close to saturation, while increased sequencing depth could improve the documentation of rare OTUs for rhizosphere and soil samples.

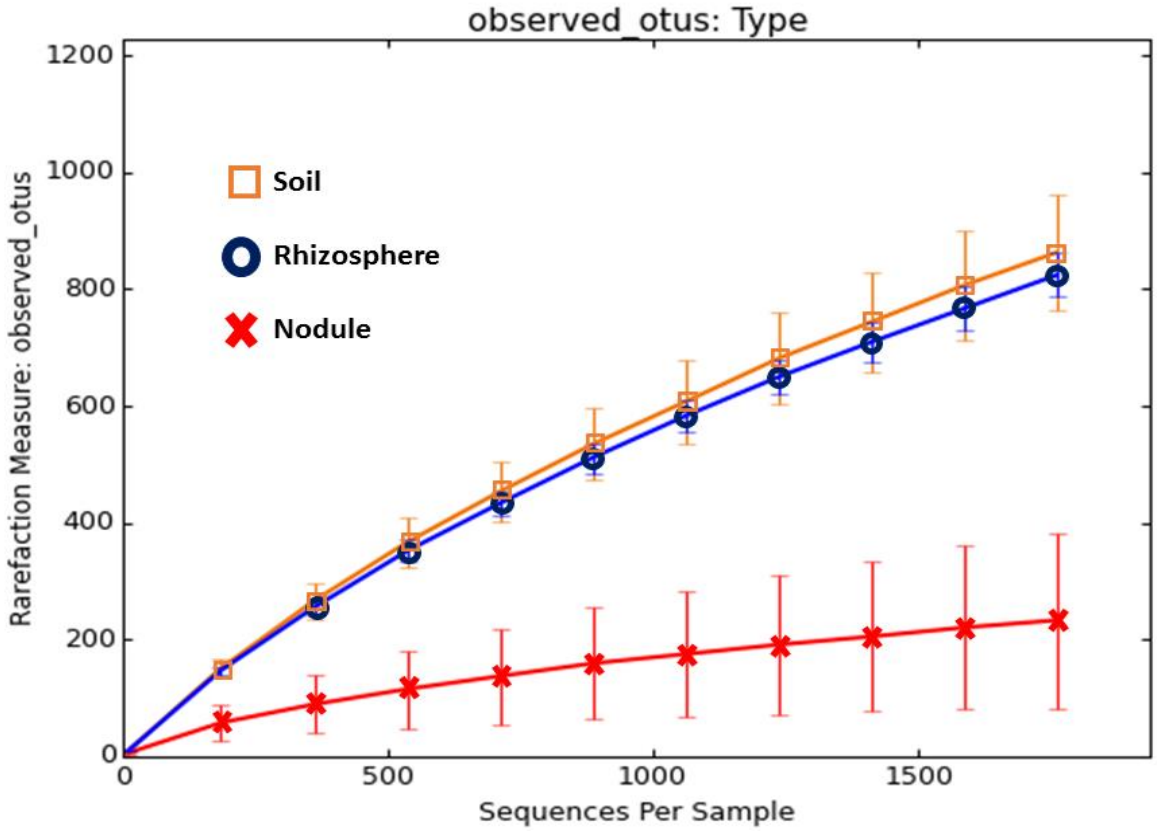


Figure 2: Average Rarefaction Plots Summarized for Sample Type. Rarefaction plots were constructed by subsampling 175 sequences from the rarefied OTU table ten times and plotting the number of observed OTUs after each subsampling.

Alpha Diversity

The Shannon Diversity Index (SDI), was used to measure alpha diversity. This metric was measured on the relative abundance after correction for 16s operon copy number. Diversity measures were grouped by sample type and are as follows; Nodules: 4.47 ± 2.29 (sd), Rhizosphere: 9.33 ± 0.233 (sd) and Soil: 9.52 ± 0.69 (sd), displayed in Figure 3. The alpha diversity was significantly less for nodules than for rhizosphere and soil samples (Students T-Test, $p = 0.00001$). Alpha diversity was calculated for the nodule samples at the three different samples locations. Nodules from Tamra had an average SDI of 2.28 (s.d = 0.584), nodules from Sidi Bouzid had an average SDI of 7.32 (s.d = 0.389) and nodules from Mednine had an SDI of 3.82 (s.d = 0.394) (Figure 4).

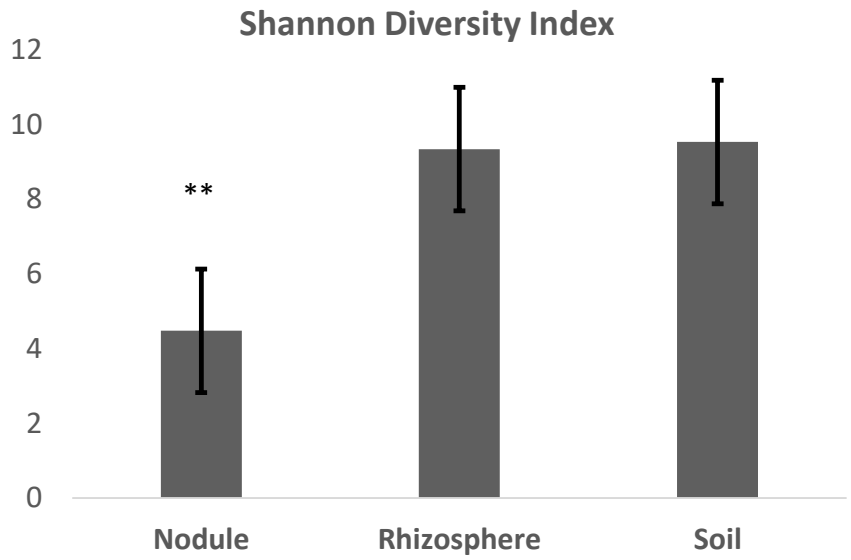


Figure 3: Alpha diversity of sample types measured by the Shannon Diversity Index. Asterisks demark that the nodule samples were significantly less diverse than both the rhizosphere and soil samples. (One-tailed T-Test $P < 0.5$)

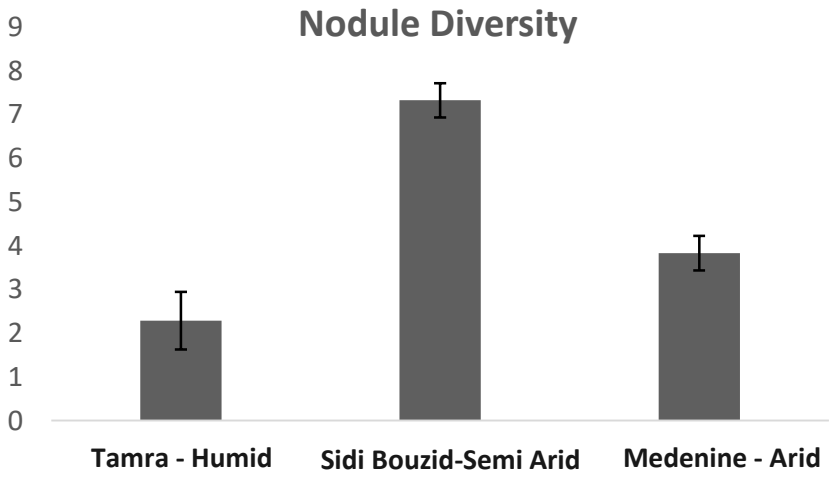


Figure 4: Alpha Diversity of Nodules by Site using the Shannon Diversity Index. All nodule sample sites differed significantly from another using the one-tailed Students T-Test ($P < 0.05$)

After correction for 16s operon copy number, the relative abundance of taxa present in the root nodules was analyzed. At the humid sampling site, the relative abundance of *Frankia* was an average of 80% of the nodule community. This prevalence dropped to 1.5% and 0.5% and the semi-arid and arid sampling sites, respectively. These data are similar to the shotgun sequencing data (Figure 17), where reads mapping to the binned *Frankia* genome comprised 60% of the humid site read set and that level dropped to 1.71% and 1.63% for the semi-arid and arid sampling sites, respectively. Figure 5 shows the relative abundance of *Frankia* at the three sampling sites using the shotgun and amplicon data. Figure 6 gives a taxonomic summary of the nodule samples at the class level.

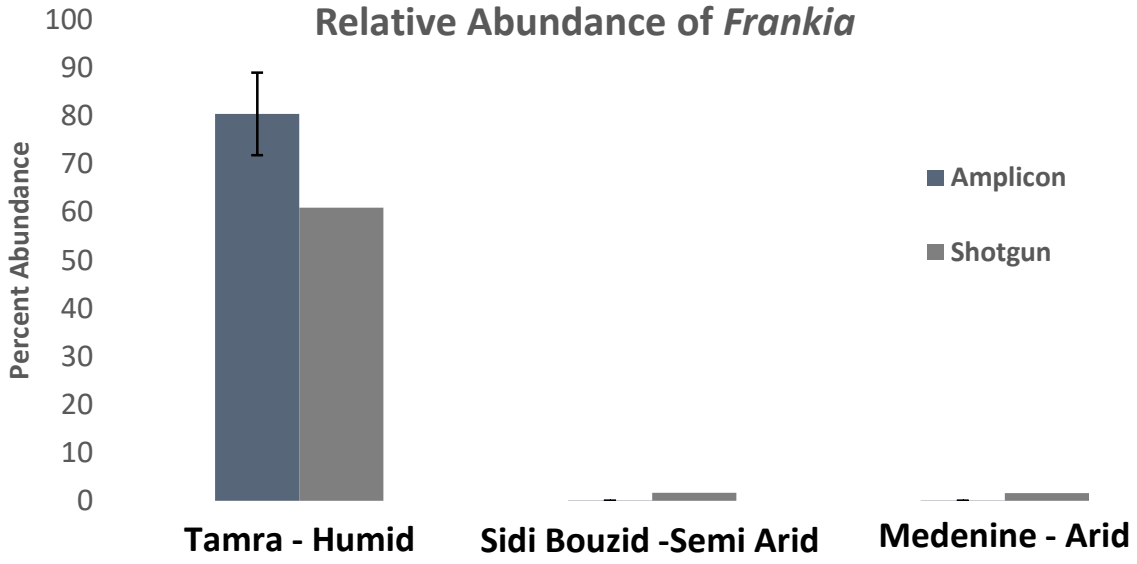


Figure 5: Relative Abundance of *Frankia* in Root Nodule Samples. Amplicon data refers to average relative abundance of *Frankia* after 16s copy number correction. Shotgun data refers to the percentage of read sets that aligned to the binned *Frankia* genome

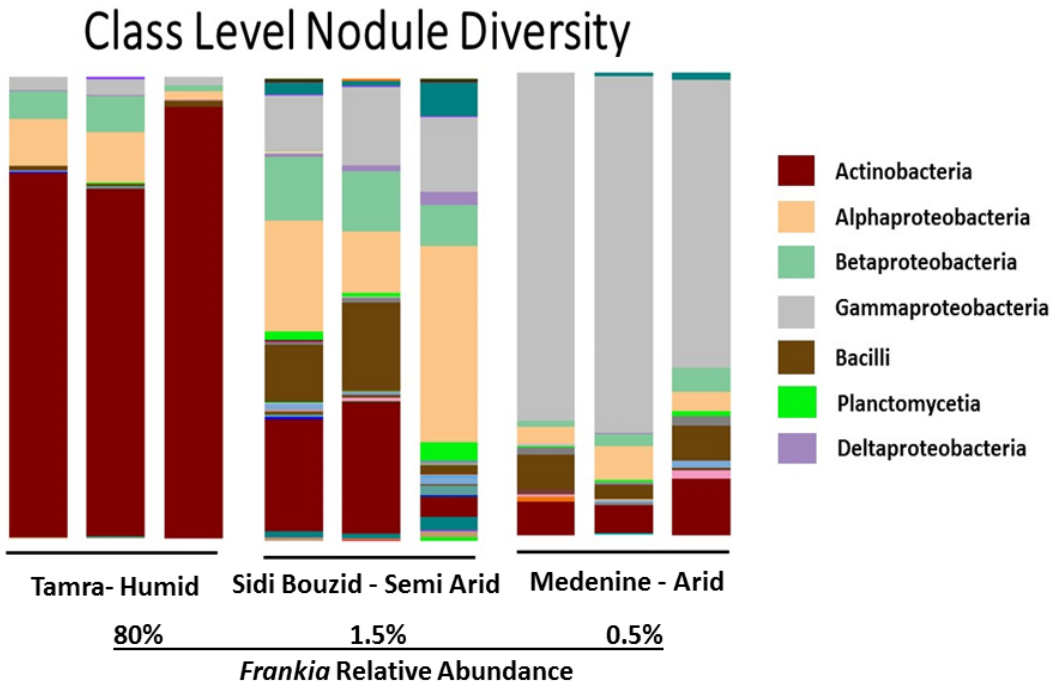


Figure 6: Class Level Taxonomic Summary of Nodule Samples. The relative abundance of the nodule samples was summarized at the class level after corrected for 16s operon copy number variation. The humid and arid samples are significantly less diverse than the semi-arid samples, as can be seen in Figure 4. This can be seen here, as the humid site is dominated by Actinobacteria (the genus *Frankia*), and the arid site is dominated by *Acinetobacter* and *Pseudomonas*

Beta Diversity

Beta diversity was analyzed by measuring the weighted and unweighted UniFrac distances (Lozupone and Knight, 2007). An Unweighted Pair Ground Method with Arithmetic Mean (UPGMA) consensus tree was built from the unweighted UniFrac distance matrix (Figure 7). Numbers on the nodes of the tree represent the percentage of trees out of 100 bootstrapped trees that had the configuration of the final consensus tree. The grouping on the tree shows that nodule samples from all sites form one clade when analyzed by the presence or absence of taxa in the samples, after removing one outlier (Sample ID: N3SIDI). This shows that although the relative abundance of taxa within the nodule differ from sampling location (Figure 5, Figure 6), the nodules are distinct based on what taxa they host when compared to the soil and rhizosphere samples. The rhizosphere and soil primarily cluster based off of the sampling location, this trend is also evident in ordinations of the data using the unweighted UniFrac distance metric (Figure 9).

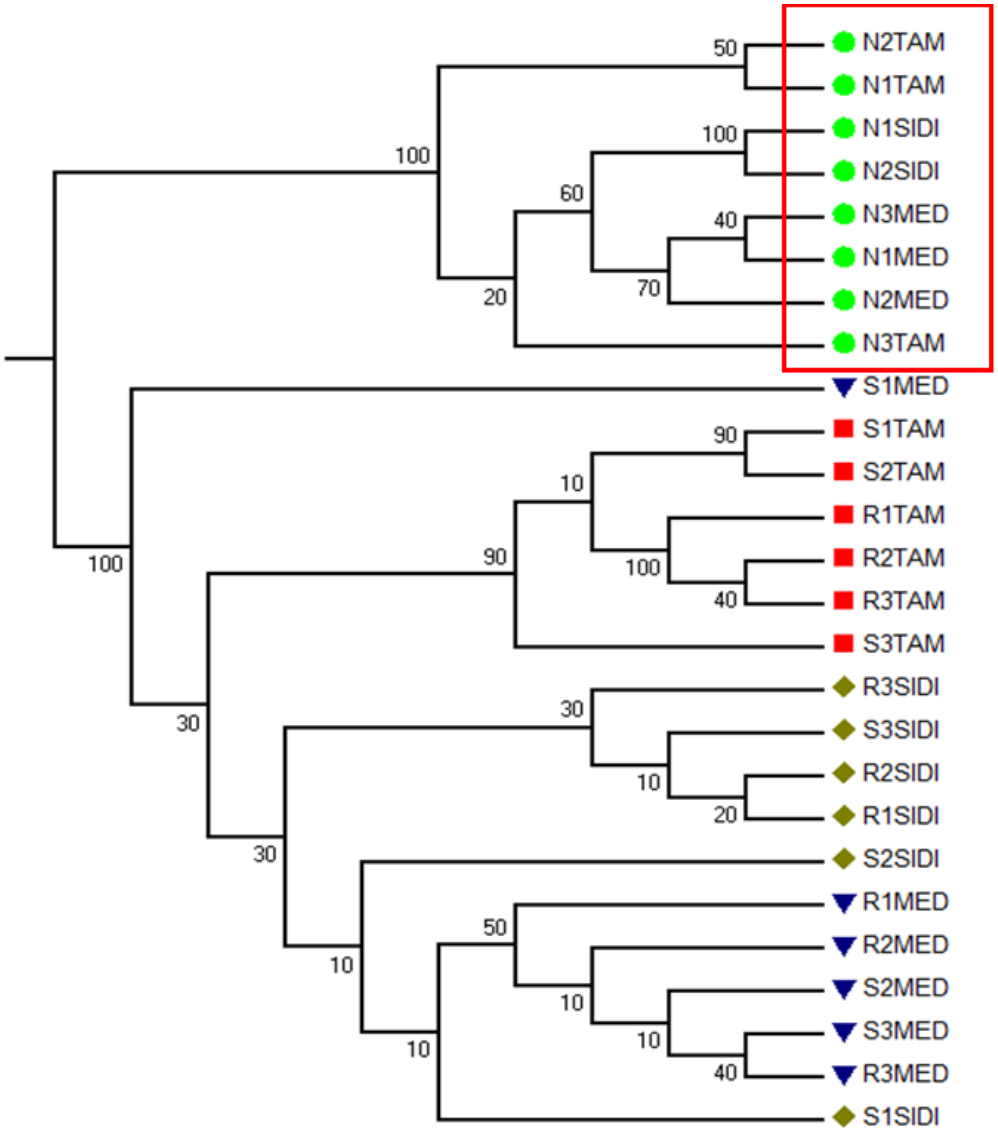


Figure 7: Unweighted Pair Group Method with Arithmetic Mean (UPGMA) Consensus Tree. The tree was constructed from the rarefied unweighted UniFrac distance Matrix. The consensus tree was built by creating 100 trees from 100 randomly subsampled reads from each sample. The numbers on the nodes represent the percentage of the 100 trees that conformed to this configuration. The red box outlines that all nodule samples form one cluster on the tree. The other clades on the tree show that the soil and rhizosphere communities cluster mostly by sampling site.

The beta diversity distance matrices were ordinated using PcoA. Distance matrices were constructed using the weighted and unweighted UniFrac distances on OTU tables that had been normalized by rarefaction, a raw OTU table with ribosomal operon copy number correction and by cumulative sum-scaling (with and without ribosomal operon copy number correction). PcoA ordinations of these distance matrixes can be seen in Figures 8-9. In all ordinations, nodule samples are separated from the soil and rhizosphere samples along the first Principal Coordinate, which explains the most phylogenetic variation in the data. Multi-response Permutation Procedures (MRPP) agreed with the results seen in the UPGMA tree, with significant grouping of nodules for every normalization type using the both forms of the UniFrac distance metric. Table 4 summarizes the chance-corrected results of the MRPP analysis. The A Statistics, the chance-corrected within-group agreement, is a value between 1 and 0 that describes the within-group heterogeneity of a given group. If all samples in a group are identical A = 1, if the heterogeneity in a groups is equal to what would be expected by chance A = 0 (Mielke and Berry, 1984).

Table 4: Significance of Nodule Grouping by MRPP. Statistics for Cumulative Sum Scaling (CSS) normalized, Rarefied and Relative Abundance with 16s copy number correction, using the Weighted Unifrac (WU) and Unweighted UniFrac (UW).

Normalization Type	CSS - WU	CSS -UW	CSS-16s-WU	CSS-16s-UW	Rarefield-WU	Rarefield-UW	16s-Rel-WU	16s-Rel-UW
A Statistic	0.2529	0.1138	0.2926	0.1451	0.177	0.1079	0.1832	0.1437
P Value	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

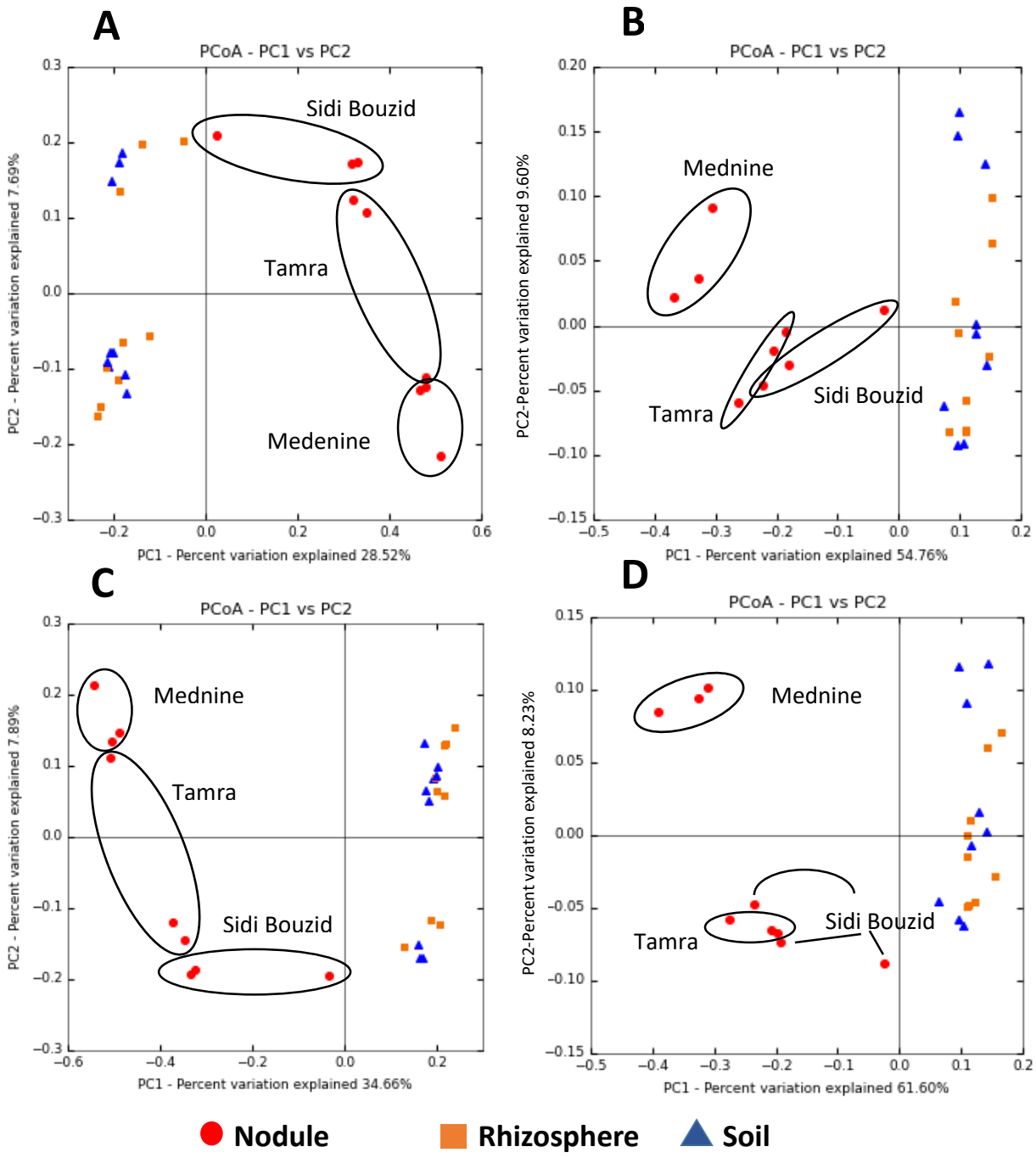


Figure 8: Principal Coordinates Ordination of CSS Normalized UniFrac Distance Matrices: Ordinations **A** and **B** where normalized using Cumulative Sum Scaling. Ordinations **C** and **D** were first adjusted for 16s copy number variation and then normalized by Cumulative Sum Scaling. Ordinations **A** and **C** were made using the unweighted UniFrac distance and ordinations **B** and **D** were made using the weighted UniFrac distance. For all ordinations MRPP showed a statistically significant grouping of the nodule samples, as outlined in Table 7.

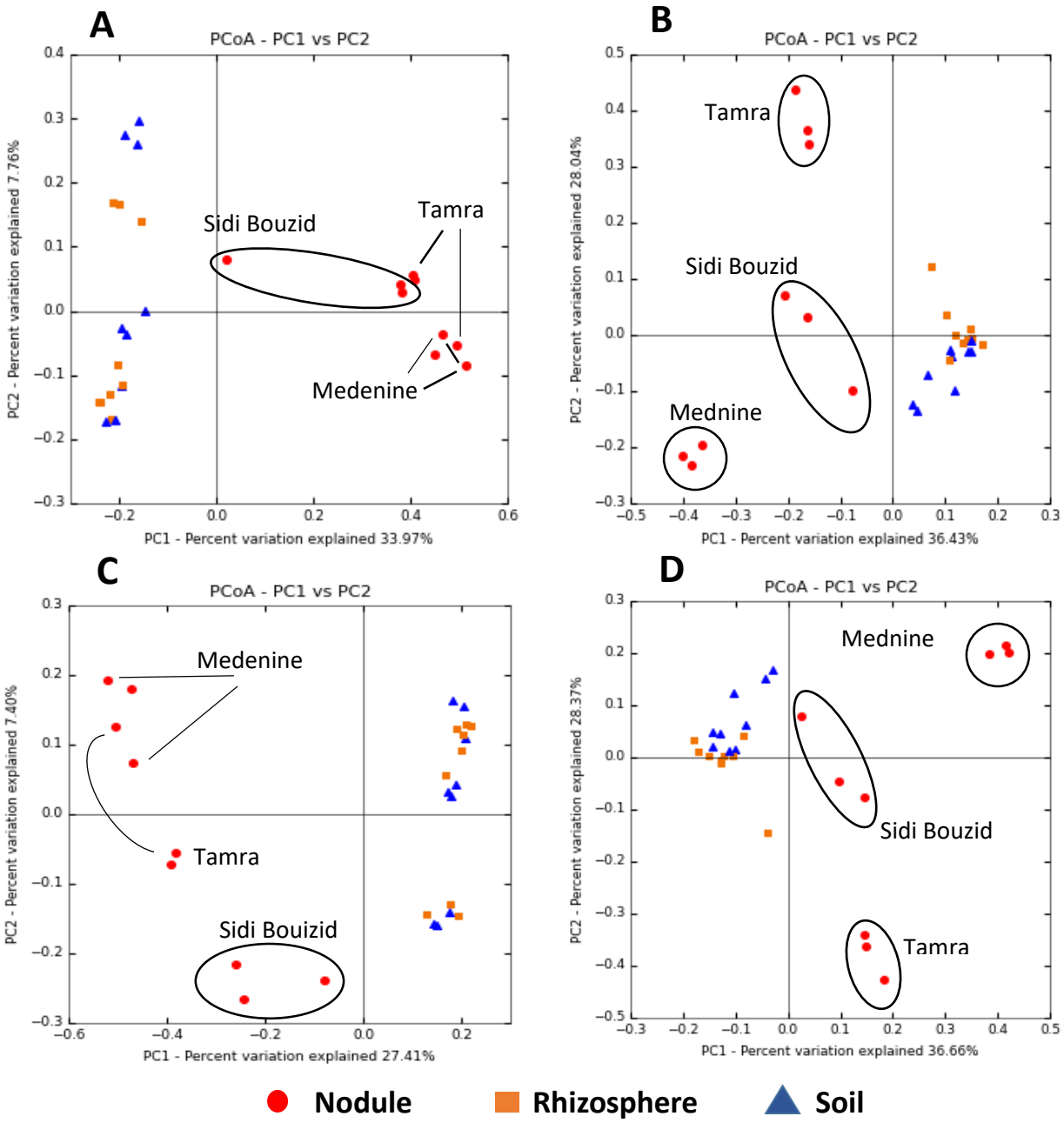


Figure 9: Principal Coordinates Ordination of UniFrac Distance Matrices: Ordinations **A** and **B** were produced from raw OTU tables that were 16s copy number corrected. Ordinations **C** and **D** were produced from evenly rarefied OTU tables. Ordinations **A** and **C** were made using the unweighted UniFrac distance and ordinations **B** and **D** were made using the weighted UniFrac distance. For all ordinations MRPP showed a statistically significant grouping of the nodule samples, as outlined in Table 7. Geographical sampling site labels on B and D are there to show that the nodules to separate along the second PcoA axis, which explains the second most amount of taxonomic variation in the data set.

Nodule Analysis

The results of the UPGMA Tree (Figure 7) and the PcoA ordinations (Figures 8 and 9) show that nodules host a distinct prokaryotic communitycompared to the soil and rhizosphere samples. Ordination and MRPP results show that this grouping is statistically significant whether the unweighted UniFrac or weighted UniFrac is used. Indicator Species Analysis was used to determine what species are significant indicators of nodule samples grouped from all locations. This analysis takes into account relative abundance of a given taxa and the relative frequency of that taxa in the group being analyzed when determining and indicator species value. The results for this analysis are outlined in Table 5. The DESeq2 algorithm was also used to determine differentially abundant OTUs in the pooled nodule samples. This method uses a generalized linear module to determine OTUs that are differentially abundant in certain sample types over others. The results to this analysis, which were very similar to Indicator Species Analysis, can be seen in Table 6.

For the 22 statistically significant indicator species identified, 12 shared the taxonomy of OTUs that were determined to be differentially abundant by DESeq2, these include; *Acinetobacter*, *Corynebacterium*, *Rheinheimera*, *Steptococcus*, *Peptonophilus*, *Enhydrobacter*, *Finegoldia*, *Enterococcus*, *Anaerococcus*, *Propionibacterium*, *Brachybacterium* and *Citrobacter*. Indicator Species Analysis takes into account fidelity within the group (nodules), while DESeq2 measures log fold change within the group compared to abundance across all samples. OTUs that are present in all sample types, but are differentially more abundant only in nodules are not identified as indicator species, but are identified by DESeq2 (Table 6). Inversely, taxa that have low fidelity to nodule samples are not identified as an indicator species. *Frankia* was only present in five of the nine nodule samples, which explains why it was not identified as an indicator species. DESeq2 did identify four *Frankia* OTUs that were differentially abundant in nodules. All of these OTUs annotated with a 99% identity to *Frankia* CcI3 using BlastN.

Table 5: Indicator Species of Nodules as a group. Data shown here was produced by ISA analysis on all nodule samples as a pooled group. ISA was performed on OTU tables that were corrected for variation in 16s operon copy number and summarized to relative abundance at the genus level.

Taxonomy (Genus)	Indicator Value	P value
Acinetobacter	0.998	0.002
Streptococcus	0.989	0.001
Corynebacterium	0.981	0.001
Rheinheimera	0.933	0.002
Exiguobacterium	0.846	0.002
Brevibacterium	0.791	0.004
Neisseria	0.791	0.012
Shewanella	0.791	0.005
Citrobacter	0.787	0.011
Anaerococcus	0.785	0.008
Peptoniphilus	0.779	0.01
Enterococcus	0.771	0.007
Alloiococcus	0.765	0.021
Propionibacterium	0.759	0.024
Actinomyces	0.707	0.006
Lactococcus	0.707	0.007
Enhydrobacter	0.707	0.005
Brachybacterium	0.701	0.037
Finegoldia	0.693	0.011
Delftia	0.689	0.023
Sediminibacterium	0.686	0.038
Serratia	0.612	0.049

Table 6: OTUs that are Differentially Abundant in Nodule Samples compared to Soil samples: These OTUs were determined to be differentially abundant in nodule samples by DESeq2 when all the nodule samples were analyzed as one group. The level of taxonomy being Genus, species, family, class and order is specified by the letters g, s, f, c, and o prior to the taxonomic name, respectively. The top 40 significant OTUs are shown here, the four *Frankia* OTUs are bolded

OTU ID	Log_2 Fold Change	Adjusted P Value	Taxonomy
988314	6.064795	1.52E-22	*g__Acinetobacter*
569335	4.325368	1.77E-11	*g__Rheinheimera*
40609	**4.220041**	**9.14E-10**	***g__Frankia***
1097359	3.819698	2.15E-09	*g__Acinetobacter*
New.CleanUp.ReferenceOTU396	3.772423	5.40E-09	*g__Shewanella*
557368	3.762968	7.18E-09	*g__Rheinheimera*
836783	3.73751	7.37E-09	*g__Shewanella*
4434190	3.668226	8.56E-09	*g__Pseudomonas*
932696	3.27909	1.34E-07	*g__Serratia*
922761	3.109828	1.34E-07	*f__Enterobacteriaceae*
1009894	3.152304	7.81E-07	*g__Acinetobacter; s__johnsonii*
1565317	3.162687	1.30E-06	*g__Geobacillus*
1083194	3.018852	2.32E-06	*g__Streptococcus*
4470206	2.941488	2.82E-06	*g__Pseudomonas*
667570	2.816011	3.64E-06	*g__Serratia; s__marcescens*
441265	2.882592	3.67E-06	*g__Corynebacterium*
646549	2.776396	4.16E-06	*g__Pseudomonas*
1003206	2.817493	5.11E-06	*g__Sphingomonas*
984831	2.803555	5.78E-06	*g__Corynebacterium*
80221	2.851964	7.85E-06	*g__Geobacillus*
1084865	2.932581	7.85E-06	*g__Staphylococcus*
570385	2.719618	1.57E-05	*g__Agrobacterium*
990864	2.55939	5.26E-05	*g__Enhydrobacter*
4422519	2.34888	6.67E-05	*g__Corynebacterium*
1047041	2.49602	0.000155	*g__Corynebacterium*
446403	2.413816	0.000158	*g__Corynebacterium; s__kroppenstedtii*
New.CleanUp.ReferenceOTU14579	**2.387998**	**0.000169**	***g__Frankia***
1083508	2.29939	0.000187	*f__Xanthomonadaceae*
938948	2.363384	0.000226	*g__Fusobacterium*
388951	2.270781	0.00027	*g__Acinetobacter*

136379	**2.160607**	**0.00063**	***g__Frankia***
579608	2.038119	0.000765	g__Streptococcus
1075600	2.044416	0.001289	f__Rhizobiaceae
580077	2.049696	0.001332	g__Rheinheimera
564411	1.769023	0.003357	g__Polynucleobacter
4328471	**1.83804**	**0.003367**	**g__Frankia**
450209	1.805817	0.003452	g__Corynebacterium
1099914	1.79344	0.004222	g__Peptoniphilus
1101451	1.754427	0.004556	g__Micrococcus

Differing Nodule Community By Location

Nodules form a distinct group based on the taxa they host, but the relative composition of this group of taxa changes drastically along the environmental gradient sampled (Figure 6). In Figure 9 (Ordinations B and D – Weighted UniFrac Distance) the ordinations show the grouping of nodule samples by location along the second principal coordinate, which explains the second largest amount of taxonomic variation in the dataset. Two-way cluster analysis of the nodule samples was used to investigate which taxa are responsible for this clustering. The DESEq2 algorithm was used to make pairwise comparisons between the nodule samples to determine OTUs that were differentially abundant in nodules from the different sampling locations (Figures 11- 13).

The heatmap of this analysis shows that sampling sites form distinct clusters based on their taxonomic composition (Figure 10). Clustering was performed using the Bray-Curtis distance metric using an OTU table that was summarized to the genus level and corrected for 16s operon copy number. The humid sampling site, Tamra, formed one cluster, with *Frankia* having an average relative abundance of 80%. The semi-arid site, Sidi Bouzid, did not have single dominant taxa, but the cluster analysis showed that the genera *Corynebacterium*, *Staphylococcus*, *Streptococcus, Rheinheimera* and *Pseudomonas* were important in determining the clustering of these samples. These results agree with the alpha diversity analysis (Figure 9) showing that Sidi Bouzid had significantly higher alpha diversity

than the other two sampling sites. The arid sampling site, Medenine, was largely dominated by the genus *Acinetobacter*, which comprises an average of 52.95% of the nodule community. The cluster of taxa on the far right of the heatmap including *Corynebacterium*, *Staphylococcus*, *Streptococcus, Rheinheimera* and *Pseudomonas,* which were predominant in the Sidi Bouzid samples, were also more prevalent in the Medenine samples suggesting that these genera are common nodule occupiers in arid environments. These genera also have high indicator values, showing a high fidelity and abundance within the group of nodule samples (Table 5) and are significantly more prevalent in arid nodules compared to humid nodules (Figure 11).

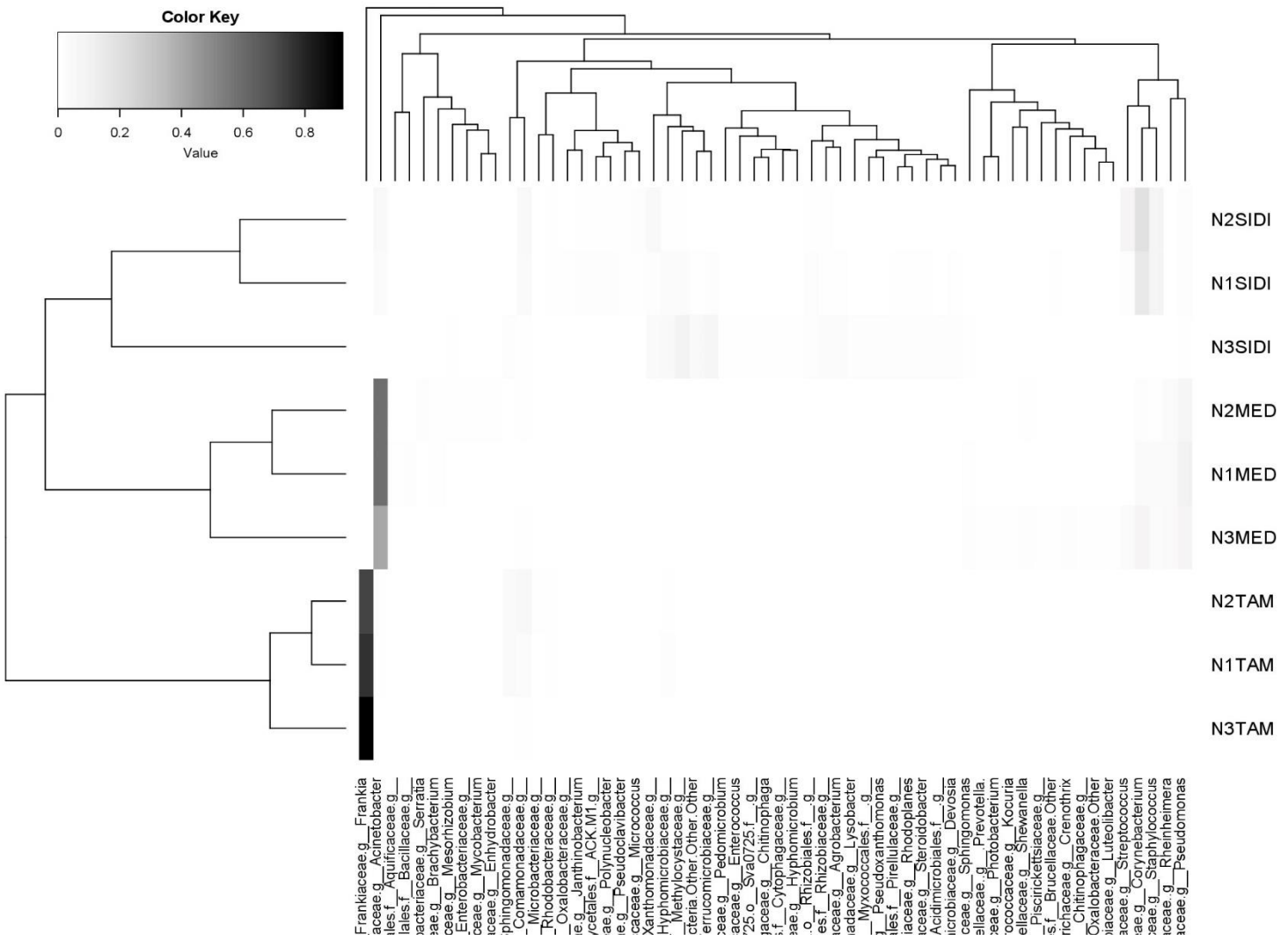


Figure 10: Two-way Cluster Analysis of Nodule Samples: Nodule samples were summed to Genus level and corrected for 16s operon copy number. Hierarchical clustering using the average linkage method was performed using the Bray-Curtis distance metric. The vertical clustering dendrogram represents how sample are related to each other based their taxonomical composition. The horizontal dendrogram represents how taxa are related to each other based on what samples they appear in. Clustering was performed based on all genera identified at the genus level (n=669), but only genera that represented more than 1% of the total community are displayed here (N=59).

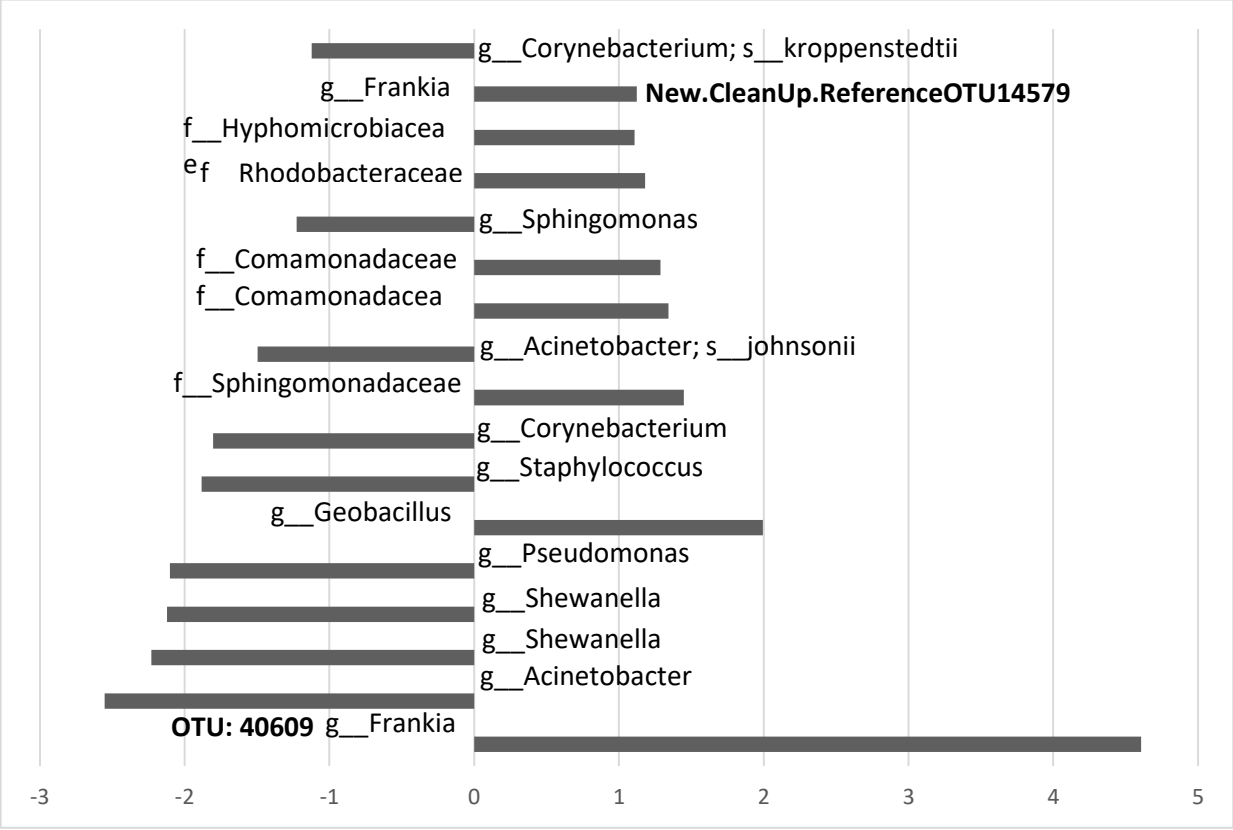


Figure 11: Pairwise Differential Abundance of OTUs between Tamra(Humid) and Medenine (Arid) samples. Pairwise comparisons were made using the DESeq2 algorithm. The X axis represents a base-two log fold change. Negative values represent differential abundance in the Medenine samples, while positive numbers represent differential abundance values in the Tamra samples. All differentially abundant OTUs reported had a Bonferonni adjusted P Value less than 0.05.

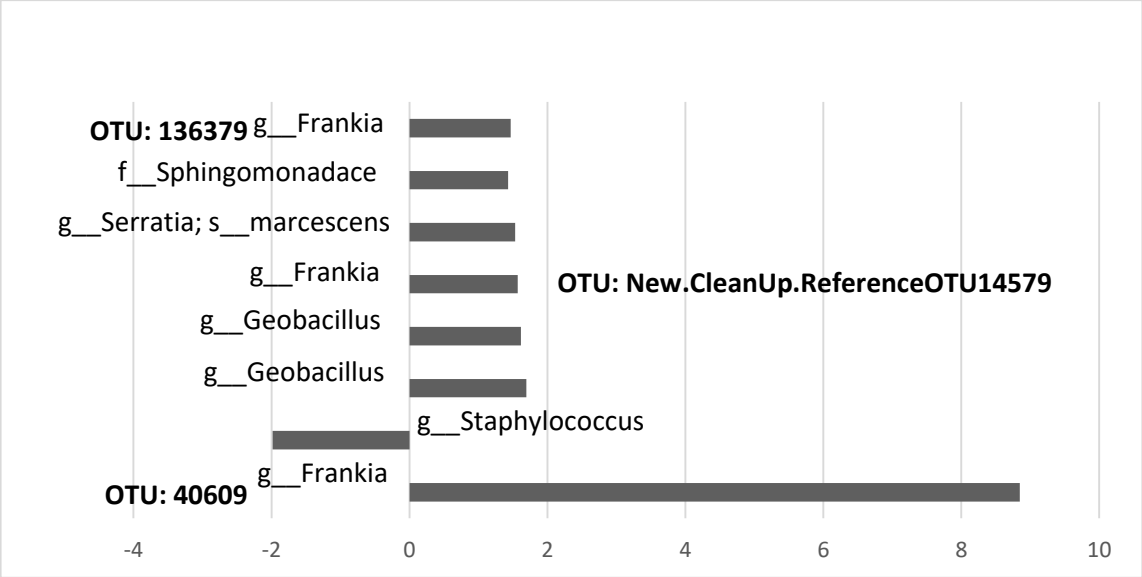


Figure 12: Pairwise Differential Abundance of OTUs between Tamra(Humid) and Sidi Bouzid (Semi - Arid) samples. Pairwise comparisons were made using the DESeq2 algorithm. The X axis represents a base-two log fold change. Negative values represent differential abundance in the Sidi Bouzid samples, while positive numbers represent differential abundance values in the Tamra samples. All differentially abundant OTUs reported had a Bonferonni adjusted P Value less than 0.05.

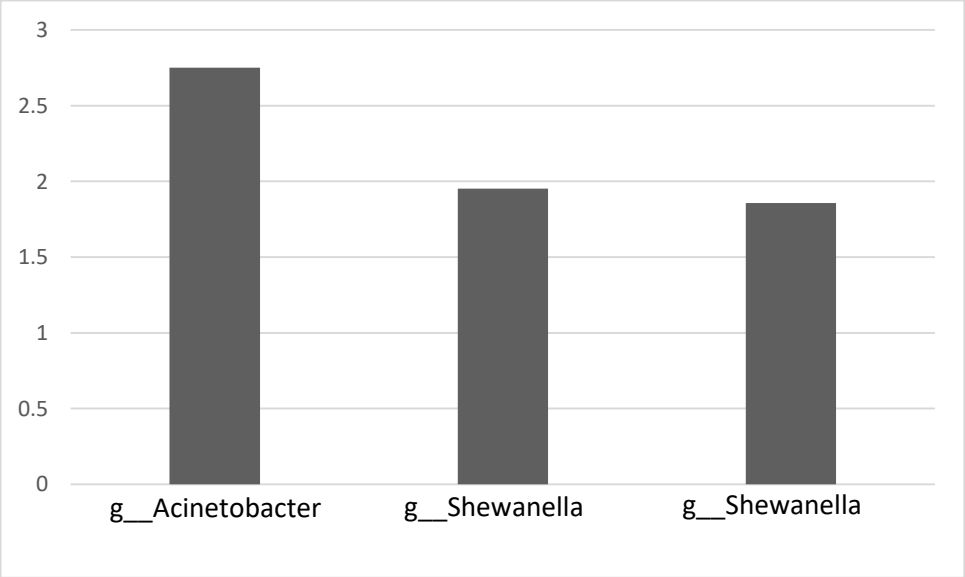


Figure 13: Pairwise Differential Abundance of OTUs between Sidi Bouzid (Semi-Arid) and Medenine (Arid) samples. Pairwise comparisons were made using the DESeq2 algorithm. The DESeq2 algorithm only identified statistically significant OTUs that were differentially abundant in the Medenine samples, which are reported here.

Rhizosphere and Soil Analysis

To understand differences in the rhizosphere and soil communities across the sampling sites, these samples were analyzed separately from the nodules samples. Rhizosphere soils from all sites were analyzed as one sample category to test the hypothesis that *C. glauca* hosts a distinct microbial community within the rhizosphere across environmental gradients. When soil and rhizosphere samples from all sites were grouped together by type the MRPP results were inconsistent in their significance, depending on the OTU table normalization strategy. A summary of these results are found in Table 7. Statistical significance ($P < 0.05$) was only observed with the evenly rarefied OTU table and the raw OTU table that corrected for 16s operon copy number Variation. Ordinations of these data (Figures 8 - 9) show that the rhizosphere and soil samples form two clusters only on the raw OTU table and the evenly rarefied OTU table (Figure 9B – 9D). CSS normalization produces different results with the soil and rhizosphere samples not forming distinct groups.

The DESeq2 Algorithm used to identify OTUs that are differentially abundant in the grouped rhizosphere samples. Indicator Species Analysis was also used to identify rhizosphere indicator species. Table 8 shows the OTUs that are upregulated in the rhizosphere when the three sampling sites are grouped together. Indicator species and the accompanying Indicator Values are found in Table 9. Three of the indicator species overlap with those identified as differentially abundant using the DESeq2 algorithm including an OTU in the class Betaproteobacteria, an OTU in Betaproteobacteria order MDN1 and an OTU in the family *Piscirickettsiaceae*.

Table 7: MRPP Results of pooled Soil and Rhizosphere samples

Normalization Type	Raw-Relative	16s Copy Number Relative	Rarefied	CSS	16s Copy Number CSS
A statistic	0.03552	0.03858	0.0323	0.01597	0.0233
P Value	0.063	0.035	0.049	0.167	0.104

Table 8: Differentially Abundant OTUs in pooled rhizosphere samples: The level of taxonomy being family, class and order is specified by the letters f, c, and o prior to the taxonomic name, respectively.

Log_2 Fold Change	P Value	Taxonomy
2.266173001	0.004712214	f__Piscirickettsiaceae
2.099161072	0.010132725	c__Betaproteobacteria; o__MND1
1.95473451	0.018675705	c__Betaproteobacteria; o__MND1
1.928568486	0.018825131	c__Betaproteobacteria; o__MND1
1.81172522	0.022660199	c__Betaproteobacteria

Table 9: Indicator Species Analysis for pooled rhizosphere samples: The level of taxonomy being family, class and order is specified by the letters f, c, and o prior to the taxonomic name, respectively.

IV	P Value	Taxonomy
83.7	0.001	c__Betaproteobacteria;o__MND1;f__
71.3	0.0008	c__Deltaproteobacteria;o__NB1-j;f__MND4
70.9	0.0012	c__Deltaproteobacteria;o__Entotheonellales;f__Entotheonellaceae
66.8	0.0038	c__Deltaproteobacteria;o__NB1-j;f__NB1-i
66.5	0.0082	c__Betaproteobacteria
66.1	0.001	c__Deltaproteobacteria;o__NB1-j;f__
64.8	0.0014	c__Deltaproteobacteria;o__Syntrophobacterales;f__Syntrophobacteraceae
64.3	0.0014	c__Solibacteres;o__Solibacterales;f__PAUC26f
63	0.0008	c__Gammaproteobacteria;o__Thiotrichales;f__Piscirickettsiaceae
62.9	0.0124	c__Acidimicrobia;o__Acidimicrobiales;f__koll13
61.7	0.0044	c__Pedosphaerae;o__Pedosphaerales;f__auto67_4W
60.9	0.0076	c__Gammaproteobacteria;o__[Marinicellales];f__[Marinicellaceae]
60.4	0.0002	c__Alphaproteobacteria;o__Rhodospirillales;f__Rhodospirillaceae
60.3	0.0056	c__Acidobacteria-5
60.2	0.0022	c__Gemmatimonadetes

The ordinations using the unweighted UniFrac distance, which takes into account the presence and absence of taxa, show that soil and rhizosphere samples clustered based on climate. When the data is ordinated using this metric, two distinct clusters of rhizosphere and soil samples formed (Figure 14). These clusters correspond to the humid sampling site (Tamra) and the two arid sampling sites (Sidi Bouzid and Mednine). For all normalization types, a significant grouping was identified by MRPP (Table 10),showing that the environmental conditions significantly change the taxa present between the humid and arid soils that were sampled. A taxonomic summary of these samples at the class level are presented in Figure 15. DESeq2 was used to find differentially abundant OTUs between the humid group and the arid group (Tables 11 and 12). Indicator Species Analysis was performed to identify indicator species of each sample site. Both of these analysis show distinct abundant OTUs and Indicator Species for each group of samples that explain the differences in class-level taxonomic structure (Figure 15).

Table 10: MRPP results for Humid-Arid grouping in ordinations seen in Figure 15.

Normalization Type	Raw-Relative	16s Copy Number Relative	Rarefied	CSS	16s Copy Number CSS
A Statistic	0.06809	0.07799	0.05062	0.06019	0.07526
P Value	0.001	0.001	0.001	0.001	0.001

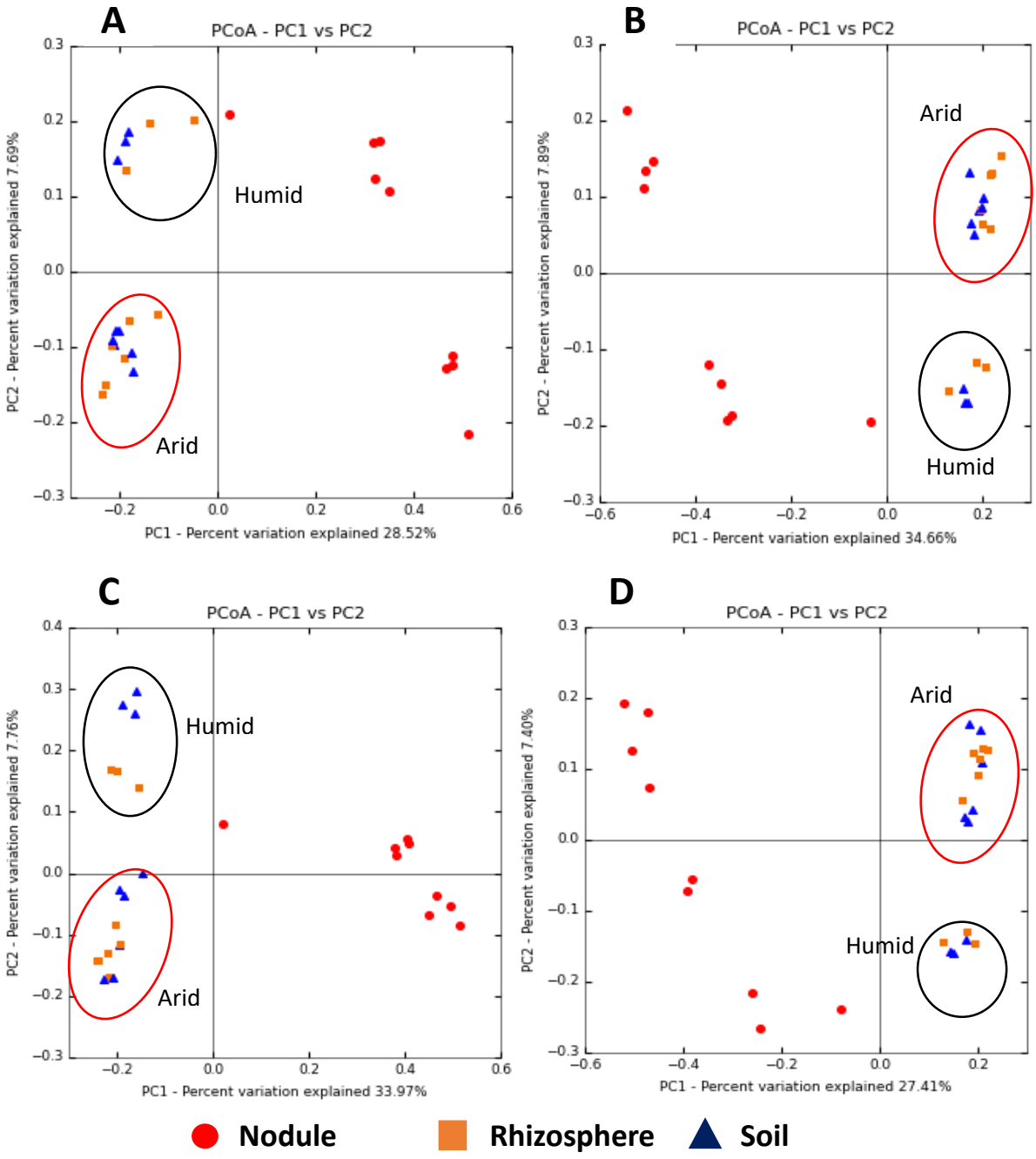


Figure 14: Principal Coordinate Ordinations of Unweighted UniFrac Distance Matrices. For all ordinations, samples grouped by black circles are soil and rhizosphere samples from the humid sampling site and samples grouped by red circles are samples from the semi-arid and arid sampling sites. **A:** Ordination from distance matrix produced from OTU table that was normalized by Cumulative Sum Scaling. **B:** Ordination from distance matrix that was produced from OTU table that was corrected for 16s operon copy number before Cumulative Sum Scaling. **C:** Ordination from raw OTU table that was corrected for 16s operon copy number. **D.** Ordination from distance matrix that was produced from evenly rarefied OTU table. MRPP statistics for these ordinations can be seen in Table 12.

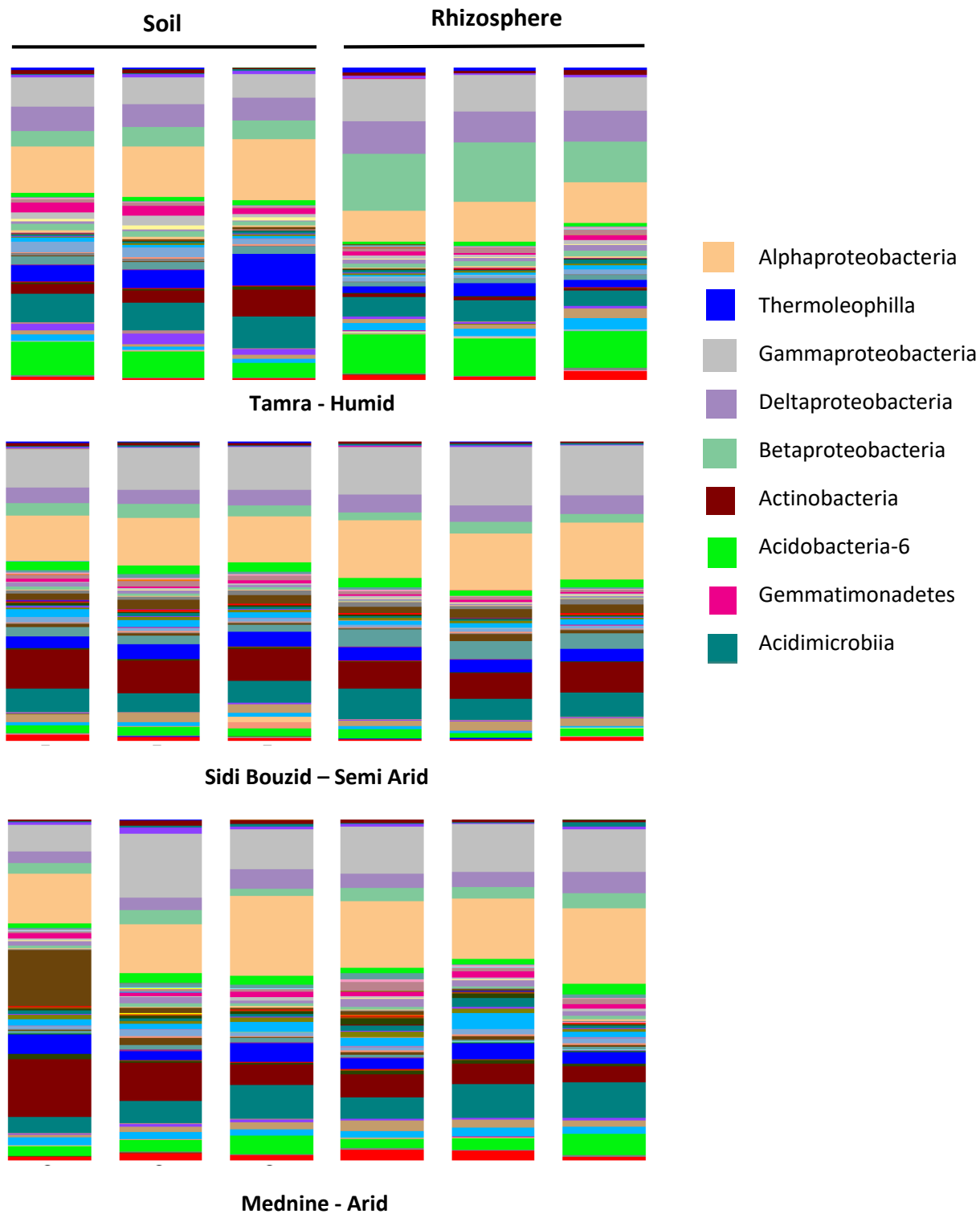


Figure 15: Class level Diversity of Rhizosphere and Soil samples grouped by sampling site. These charts show relative abundance of taxa at the class level after correct for 16s operon copy number.

Table 11: OTUs determined to be differentially abundant in humid samples by DESeq2

Log_2 Fold Change	P Value	Taxonomy
4.959574336	5.77E-23	c__Chloracidobacteria; o__RB41; f__Ellin6075
4.598712383	9.90E-17	c__Saprospirae; o__[Saprospirales]; f__Chitinophagaceae
3.531274288	9.90E-17	c__Acidimicrobiia; o__Acidimicrobiales
3.332793773	9.90E-17	c__Acidobacteria-6; o__iii1-15
3.604908707	2.32E-14	c__Deltaproteobacteria; o__Syntrophobacterales; f__Syntrophobacteraceae
3.011458273	1.81E-11	c__Deltaproteobacteria; o__Myxococcales
5.745838009	2.78E-11	c__Betaproteobacteria
3.962822517	1.00E-10	c__Thermoleophilia; o__Gaiellales; f__Gaiellaceae
5.054478211	2.52E-10	c__Gammaproteobacteria; o__Thiotrichales; f__Piscirickettsiaceae
3.331682099	4.11E-10	k__Bacteria; p__WS3; c__PRR-12; o__Sediment-1; f__; g__; s__
4.399049076	3.60E-09	c__Betaproteobacteria; o__MND1
2.303728882	2.18E-08	c__Alphaproteobacteria; o__Rhizobiales; f__Hyphomicrobiaceae;
4.312490696	4.03E-08	c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae
2.65576562	1.80E-07	c__Alphaproteobacteria; o__Rhizobiales; f__Hyphomicrobiaceae; g__Pedomicrobium
2.183215189	9.80E-07	c__Deltaproteobacteria; o__Myxococcales
3.011514151	2.99E-05	Gemmatimonadetes; c__Gemm-1
2.086280704	0.002241175	c__Gammaproteobacteria; o__Xanthomonadales; f__Sinobacteraceae; g__Steroidobacter; s__
2.012966006	0.003428656	c__Saprospirae; o__[Saprospirales]; f__Saprospiraceae

Table 12: OTUs determined to be differentially abundant in arid samples by DESeq2: The level of taxonomy being Genus, species, family, class and order is specified by the letters g, s, f, c, and o prior to the taxonomic name, respectively.

Log_2 Fold Change	P value	Taxonomy
3.701818252	1.90E-09	c__Gammaproteobacteria; o__Methylococcales; f__Methylococcaceae; g__Methylocaldum; s__
3.993209332	3.10E-09	c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__; s__
3.438966269	4.19E-09	c__Cytophagia; o__Cytophagales; f__Cytophagaceae; g__; s__
3.474614071	9.48E-09	c__Actinobacteria; o__Actinomycetales; f__Micrococcaceae; g__; s__
3.409066406	1.06E-06	c__Gammaproteobacteria; o__Xanthomonadales; f__Sinobacteraceae; g__; s__
2.937409327	2.49E-06	c__Bacilli; o__Bacillales; f__Planococcaceae
2.531167224	2.63E-06	c__Actinobacteria; o__Actinomycetales; f__Nocardioidaceae; g__Aeromicrobium
2.911540897	3.08E-06	c__Anaerolineae; o__GCA004; f__; g__; s__
2.872951457	7.87E-06	c__Actinobacteria; o__Actinomycetales; f__Micrococcaceae; g__; s__
2.661715854	7.87E-06	c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae; g__; s__
2.799189289	1.27E-05	c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__; s__
2.682036047	1.27E-05	c__Bacilli; o__Bacillales; f__Bacillaceae; g__Bacillus; s__foraminis
2.894568485	1.44E-05	c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__; s__
2.301437427	2.51E-05	c__Actinobacteria; o__Actinomycetales; f__Nocardioidaceae; g__; s__
2.402010725	3.19E-05	p__Chloroflexi; c__Ellin6529; o__; f__; g__; s__
2.753316249	3.75E-05	c__Actinobacteria; o__Actinomycetales; f__Micrococcaceae; g__Microbispora; s__rosea
2.3579123	4.74E-05	c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__Pseudomonas
2.361744126	9.42E-05	c__Acidimicrobiia; o__Acidimicrobiales; f__; g__; s__
2.454038606	9.42E-05	p__Gemmatimonadetes; c__Gemm-1; o__; f__; g__; s__
2.192287053	0.000148148	c__Acidimicrobiia; o__Acidimicrobiales; f__; g__; s__
2.206600522	0.000179435	c__Deltaproteobacteria; o__Syntrophobacterales; f__Syntrophobacteraceae; g__; s__
2.310342028	0.00031282	c__Acidimicrobiia; o__Acidimicrobiales; f__; g__; s__
2.411303582	0.000376269	c__Gammaproteobacteria; o__Xanthomonadales; f__Sinobacteraceae; g__; s__
2.491759743	0.000381755	c__Alphaproteobacteria; o__Rhizobiales; f__Phyllobacteriaceae; g__Mesorhizobium; s__
2.75590494	0.000381755	c__Actinobacteria; o__Actinomycetales; f__Nocardiopsaceae; g__Prauseria; s__
2.009172175	0.000611399	c__Cytophagia; o__Cytophagales; f__Cytophagaceae; g__; s__
2.049792422	0.000729016	c__Acidimicrobiia; o__Acidimicrobiales; f__koll13; g__; s__
2.507438697	0.000729016	c__Anaerolineae; o__GCA004; f__; g__; s__
2.408631518	0.00077915	c__Alphaproteobacteria; o__Sphingomonadales; f__Sphingomonadaceae; g__Sphingomonas; s__suberifaciens
2.034842908	0.00077915	c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae; g__Agrobacterium; s__
2.395749992	0.000782068	c__Actinobacteria; o__Actinomycetales; f__Thermomonosporaceae; g__Actinomadura; s__vinacea
2.389055895	0.00080139	c__Acidimicrobiia; o__Acidimicrobiales; f__; g__; s__

2.5964642	0.001023856	c__Gammaproteobacteria; o__Aeromonadales; f__Aeromonadaceae; g__Zobellella
2.071039545	0.001341302	c__Bacilli; o__Bacillales; f__; g__; s__
2.315881175	0.001487573	c__Alphaproteobacteria; o__Rhodospirillales; f__Rhodospirillaceae; g__; s__
2.063821574	0.00153166	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Xanthomonadales; f__Sinobacteraceae; g__Steroidobacter; s__
2.364802139	0.001966049	c__Actinobacteria; o__Actinomycetales; f__Nocardiaceae; g__Nocardia; s__
2.178421837	0.001972025	c__Gammaproteobacteria; o__Xanthomonadales; f__Sinobacteraceae; g__; s__
2.379334712	0.002103051	c__Acidimicrobiia; o__Acidimicrobiales; f__koll13; g__; s__
2.355674322	0.00222185	c__Gammaproteobacteria; o__Alteromonadales; f__Idiomarinaceae; g__Idiomarina
2.01936029	0.002238486	c__Alphaproteobacteria; o__Rhodospirillales; f__Rhodospirillaceae; g__; s__
2.287330276	0.002281869	c__Gammaproteobacteria; o__Xanthomonadales; f__Sinobacteraceae; g__; s__
2.040018242	0.00235143	c__Bacilli; o__Bacillales; f__Bacillaceae; g__Bacillus; s__
2.274906518	0.003006528	c__Flavobacteriia; o__Flavobacteriales; f__Flavobacteriaceae; g__; s__
2.106005408	0.003207193	c__Actinobacteria; o__Actinomycetales; f__Streptomycetaceae; g__Streptomyces
2.050509525	0.005212536	c__Clostridia; o__Clostridiales; f__Clostridiaceae; g__Clostridium; s__
2.005806798	0.005348471	c__Alphaproteobacteria; o__Sphingomonadales; f__Erythrobacteraceae; g__; s__
2.165891788	0.005951688	p__Acidobacteria; c__Sva0725; o__Sva0725; f__; g__; s__
2.129264302	0.006756092	c__Flavobacteriia; o__Flavobacteriales; f__Flavobacteriaceae; g__; s__
2.049605895	0.007418112	c__Gammaproteobacteria; o__Legionellales; f__; g__; s__
2.031922694	0.007907731	k__Bacteria; p__Chloroflexi; c__Anaerolineae; o__H39; f__; g__; s__

Indicator Species Analysis was performed for each site by pooling the rhizosphere and soil samples together to determine if there are taxa that are indicative of the individual sites. The results of this analysis are shown in Tables 13-15. These analyses show distinct groups of taxa that are associated with each site. Some of the taxa identified in this analysis agree with the differential abundance analysis performed on the humid and arid grouping that was seen in the PcoA ordinations.

Table 13: Indicator Species Analysis of Tamra soil and rhizosphere samples: The level of taxonomy being Genus, species, family, class and order is specified by the letters g, s, f, c, and o prior to the taxonomic name, respectively.

Indicator Value	P Value	Taxonomy
0.977	0.001	p__Acidobacteria;c__RB25;o__;f__;g__
0.955	0.001	c__Betaproteobacteria;o__SC-I-84;f__;g__
0.816	0.016	c__Anaerolineae;o__WCHB1-50;f__;g__
0.816	0.01	p__Proteobacteria;c__TA18;o__PHOS-HD29;f__;g__
0.816	0.021	c__Spartobacteria;o__[Chthoniobacterales];f__[Chthoniobacteraceae];g__DA101
0.795	0.009	c__Gemmatimonadetes;o__Gemmatimonadales;f__Gemmatimonadaceae;g__Gemmatimonas
0.776	0.026	c__Gemmatimonadetes;o__Ellin5290;f__;g__
0.774	0.017	c__Deltaproteobacteria;o__[Entotheonellales];f__[Entotheonellaceae];g__
0.736	0.027	k__Bacteria;p__Elusimicrobia;c__Elusimicrobia;o__FAC88;f__;g__

Table 14: Indicator Species Analysis of Sidi Bouzid soil and rhizosphere samples: The level of taxonomy being Genus, species, family, class and order is specified by the letters g, s, f, c, and o prior to the taxonomic name, respectively.

IV	P Value	Taxonomy
1	0.001	c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__
1	0.001	c__Deltaproteobacteria;o__NB1-j;f__JTB38;g__
0.993	0.001	c__Actinobacteria;o__Actinomycetales;f__Pseudonocardiaceae;g__Saccharomonospora
0.991	0.001	c__Clostridia;o__MBA08;f__;g__
0.989	0.001	c__Bacilli;o__Bacillales;f__Bacillaceae;Other
0.987	0.001	c__Gammaproteobacteria;o__Methylococcales;f__Methylococcaceae;g__Methylobacter
0.985	0.001	c__Bacilli;o__Bacillales;f__Bacillaceae;g__Lentibacillus
0.985	0.001	c__Gammaproteobacteria;o__Methylococcales;f__Methylococcaceae;Other
0.984	0.001	c__Bacilli;o__Bacillales;f__Planococcaceae;g__Planomicrobium
0.982	0.001	c__Gammaproteobacteria;o__Alteromonadales;f__OM60;g__
0.98	0.001	c__Bacilli;o__Bacillales;f__Bacillaceae;g__Virgibacillus
0.975	0.001	c__Actinobacteria;o__Actinomycetales;f__Streptosporangiaceae;Other
0.974	0.001	c__Bacilli;o__Bacillales;f__Thermoactinomycetaceae;g__
0.971	0.001	c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__Alkaliphilus
0.967	0.015	c__Bacilli;o__Bacillales;f__Bacillaceae;g__
0.963	0.001	c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Coprococcus
0.957	0.001	c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__
0.957	0.001	c__Gammaproteobacteria;o__HTCC2188;f__HTCC2089;g__
0.955	0.003	c__Actinobacteria;o__Actinomycetales;f__Microbacteriaceae;g__Salinibacterium
0.947	0.001	c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;g__Tepidibacter
0.937	0.002	c__Bacilli;o__Bacillales;f__Paenibacillaceae;g__

Table 15: Indicator Species Analysis of Medenine soil and rhizosphere samples: The level of taxonomy being Genus, species, family, class and order is specified by the letters g, s, f, c, and o prior to the taxonomic name, respectively.

Indicator Value	P Value	Taxonomy
1	0.002	c__Actinobacteria;o__Actinomycetales;f__Nocardiaceae;g__Nocardia
1	0.002	c__Alphaproteobacteria;o__Rhizobiales;f__Xanthobacteraceae;Other
1	0.002	c__Gammaproteobacteria;o__Aeromonadales;f__Aeromonadaceae;g__Zobellella
0.998	0.002	c__Actinobacteria;o__Actinomycetales;f__Nocardiopsaceae;g__Prauseria
0.943	0.003	c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__
0.916	0.005	c__Actinobacteria;o__Actinomycetales;f__Nocardioidaceae;g__Kribbella
0.913	0.006	c__Gammaproteobacteria;o__Alteromonadales;f__Alteromonadaceae;g__Microbulbifer
0.909	0.005	c__Thaumarchaeota;o__Cenarchaeales;f__Cenarchaeaceae;g__Nitrosopumilus
0.903	0.032	c__Gammaproteobacteria;o__Legionellales;f__;g__
0.899	0.003	c__Alphaproteobacteria;o__Rhizobiales;f__Rhizobiaceae;g__Sinorhizobium
0.886	0.048	c__Gammaproteobacteria;o__Alteromonadales;f__Alteromonadaceae;g__Cellvibrio
0.871	0.007	p__Armatimonadetes;c__0319-6E2;o__;f__;g__
0.847	0.005	p__Planctomycetes;c__Pla4;o__;f__;g__
0.815	0.03	c__Bacilli;o__Lactobacillales;f__Leuconostocaceae;g__Weissella
0.81	0.017	p__Planctomycetes;c__ODP123;o__;f__;g__
0.79	0.011	c__Thaumarchaeota;o__Cenarchaeales;f__Cenarchaeaceae;g__
0.783	0.033	c__Gammaproteobacteria;o__34P16;f__;g__
0.778	0.015	c__[Pedosphaerae];o__[Pedosphaerales];f__OPB35;g__
0.732	0.034	c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingopyxis

The rhizosphere and soil samples from each site were analyzed separately to determine differentially abundant OTUs in the rhizosphere and soil at each sampling location. Figures 16 - 18 show the difference in increased OTUs at the three different sampling sites: Tamra, Sidi Bouzid, and Medenine. Although both Medenine and Tamra have increased OTUs of the Betaproteobacteria Order MDN1, the groups of increaded OTUs in the rhizosphere at the different sites were distinct. The log-fold changes for OTUs in the rhizosphere for the humid sampling site, Tamra, were higher than those found at the semi-arid and arid sampling sites, suggesting a stronger influence of rhizosphere affects in this climate.

At the humid sampling site, Tamra, OTUs of the families *Syntrophobacteracea* and *Piscirickettsiaceae* were significantly more abundant in the rhizosphere. *Syntrophobacteracea* increased from an average abundance level of 1.63% in the soil to an average of 5.48% in the rhizosphere, while *Piscirickettsiaceae* increased from 1.9% to 5.99% in the soil and the rhizosphere, respectively. These OTUs were responsible for the increased relative abundance of Gammaproteobacteria in the rhizosphere (Figure 16). Two OTUs in Betaproteobacteria order MDN1 were significantly more abundant in the rhizosphere, increasing from 0.36% in soil to 10.1% in the rhizosphere. An OTU classifiedas the class Betaproteobacteria was also identified and increased from 1% to 4.3%. These OTUs values help explain the increase in Betaproteobacteria in the rhizosphere samples at this site (Figure 16). An OTU in the family *Syntrophobacteracea* was also significantly increased from 1.5% in soil to 5.5% in the rhizosphere. This increase helps explain the higher abundance of Deltaproteobacteria in Figure 16.

The rhizosphere samples at the semi-arid and arid sampling sites contained different increased OTUs compared to the humid sampling site, except for the Betaproteobacteria order MND1 at the Mednine sampling site. The lower log-fold changes and different increased OTUs show that the rhizosphere influence was distinctive and less powerful at the semi-arid and arid sampling sites.

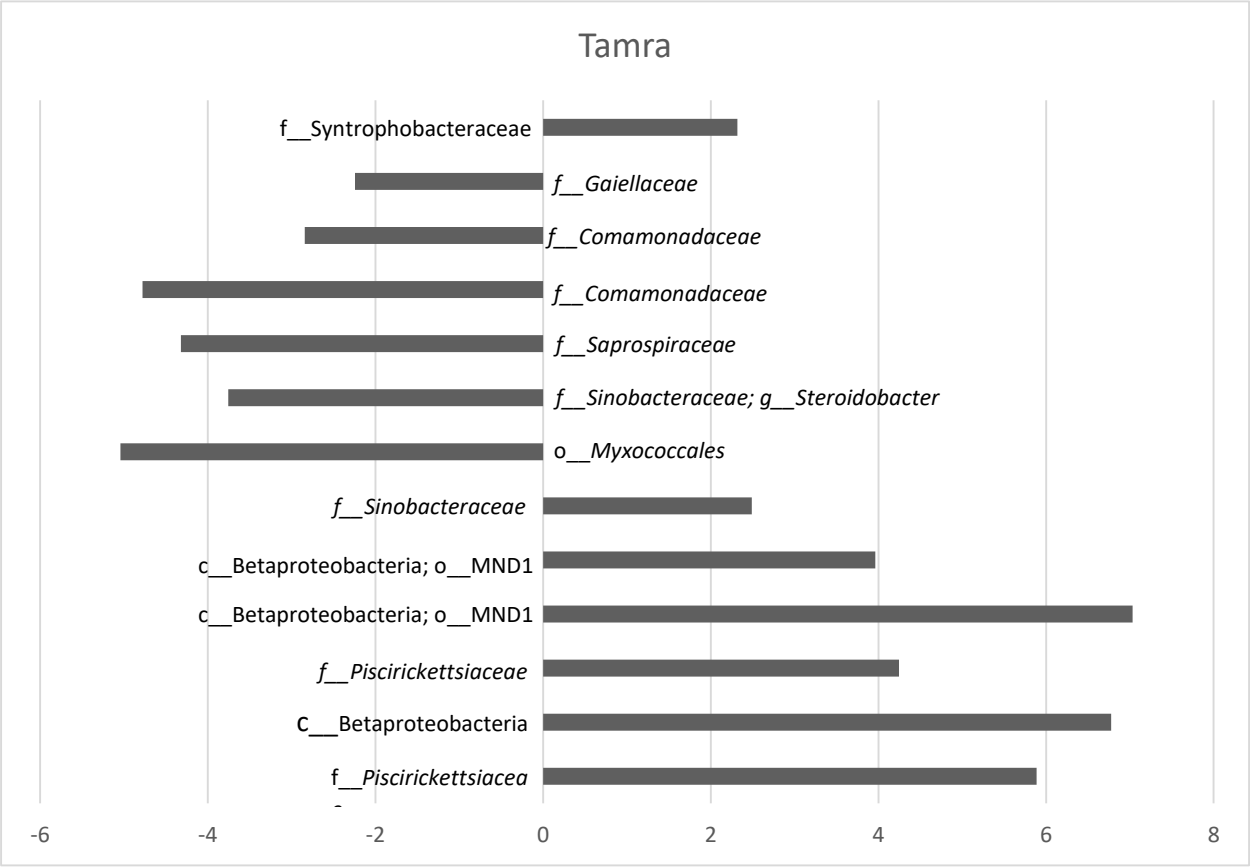


Figure 16: Differentially abundant OTUs between the rhizosphere and soil at the Tamra sampling site. Low abundance OTUs that were represented by less than 100 sequences were discarded before analysis. The numbers on the X axis represent log fold change (base 2). Negative numbers indicate log fold change in the soil, while positive numbers are log fold change in the rhizosphere.

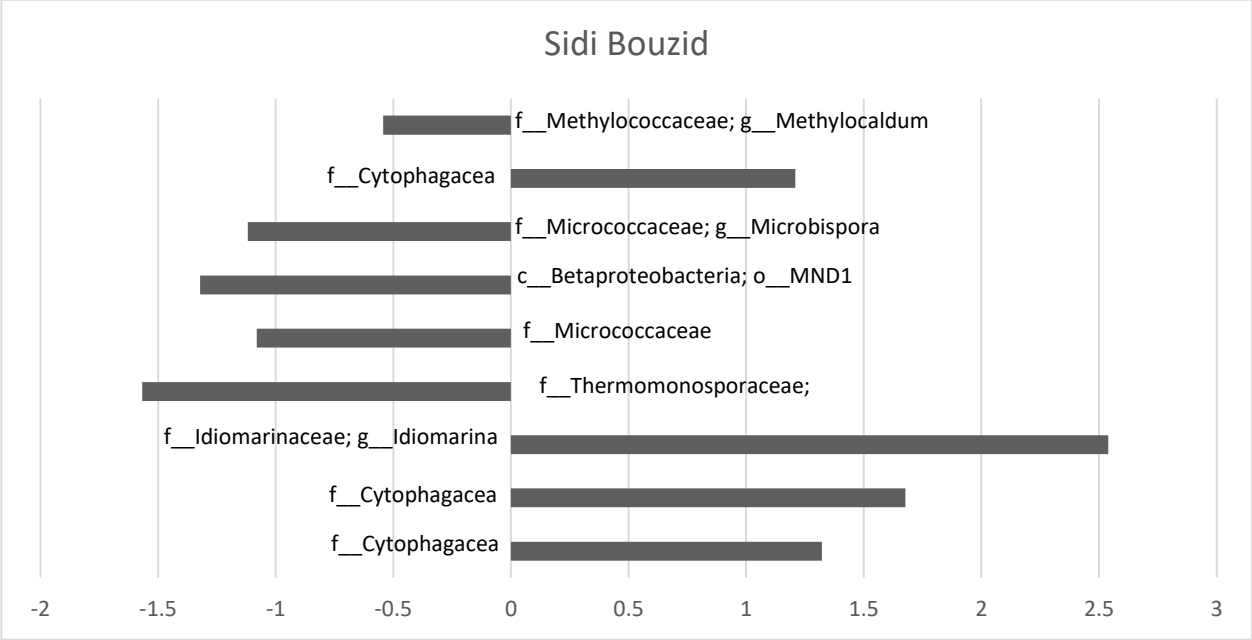


Figure 17: Differentially abundant OTUs between the rhizosphere and soil at the Sidi Bouzid sampling site. Low abundance OTUs that were represented by less than 100 sequences were discarded before analysis. The numbers on the X axis represent log fold change (base 2). Negative numbers indicate log fold change in the soil, while positive numbers are log fold change in the

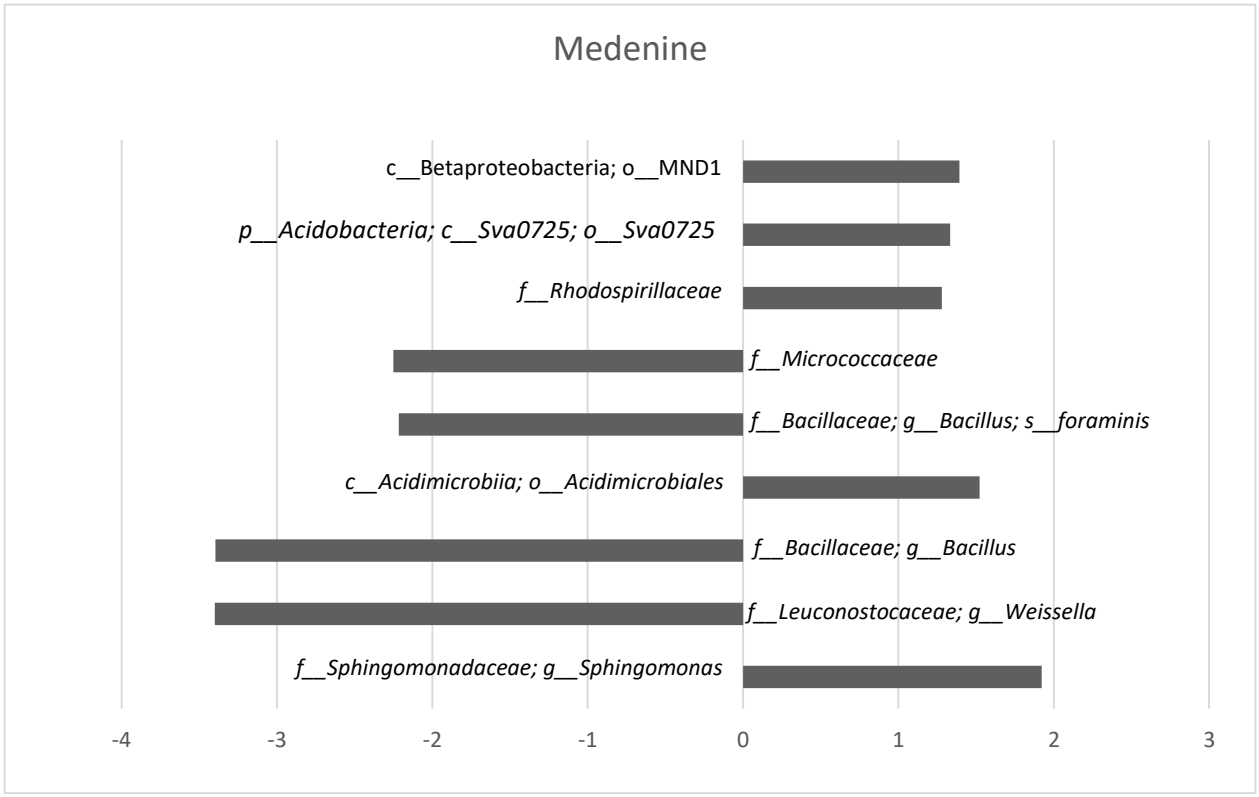


Figure 18: Differentially abundant OTUs between the rhizosphere and soil at the Medenine sampling site. Low abundance OTUs that were represented by less than 100 sequences were discarded before analysis. The numbers on the X axis represent log fold change (base 2). Negative numbers indicate log fold change in the soil, while positive numbers are log fold change in the rhizosphere.

Shotgun Metagenomic Sequencing

Quality filtering of shotgun metagenomic reads resulted in a total of 83,027,118 paired end reads. The library for Sites 1, 2 and 3 were comprised of 33,051,352, 34,239,886 and 15,735,880 reads, respectively. *In* silico co-assembly of the reads resulted in an assembly contained 3,153,711 contigs totaling 56,964,841 base pairs in length. The statistics for the assembly produced by Quast (Gurevich et al., 2013) are outlined in Table 16.

Table 16: Assembly Statistics from Metagenome Assembly

Number of contigs (>= 0 bp)	3,153,711
Number of contigs (>= 1000 bp)	122,849
Total length (contigs >= 0 bp)	1,320,376,721
Total length (contigs >= 1000 bp)	365,121,218
Number of Contigs >= 500 bp	432,339
Largest Contig	1,267,956
Total Length (contigs >= 500 bp)	569,464,841
GC(%)	47.20
N50	1683

Draft Genomes Reconstructed using Differential Coverage Binning

Seven predominant genomes were binned using the mmgenome R package. A plot of the metagenomic assembly can be seen in Figure 19. The binned genomes were in the genera *Frankia*, *Micromonospora*, *Bacillus* (two isolates), *Phyllobacterium*, *Paenibacillus* and *Afipia.* The DNA isolated for shotgun sequencing was obtained from nodules that were superficially washed, but not surface sterilized. The nodules from the amplicon dataset were surface sterilized, therefore the community profiled by 16s amplification can be considered true endophytes. Only the abundance of *Frankia* was consistent between the two datasets, leading the deduction that the other binned genomes were from organisms that were surface-colonizing epiphytes.

Two of the binned genomes annotated to the phylum actinobacteria. The first genome bin contained 4,111 ORFs that annotated to the genus *Frankia*. The annotations for the ORFs in this bin had

top hits on many of the *Frankia* clade 1c *Casuarina* infective strains. The most top annotations that hit on a single sequenced strain was 775 of the 4,111 ORFs annotated to *Frankia* sp. BMG5.23, a salt tolerant Tunisian isolate (Ghodhbane-Gtari et al., 2014). Reads mapping to the binned *Frankia* draft genome comprised 60% of the reads in from the humid sampling site, Tamra. Of the reads from semi-arid Sidi Bouzid and arid Gabes, 1.7% and 1.6% of the reads mapped to the *Frankia* genome bin, respectively. Genome bin two resulted in an assembly that contained 6,522 ORFs with 6,053 annotating to the genus *Micromonospora.* The ORFs that annotated to Micromonospora were mainly comprised of annotations to *Micromonosopora purpeochromogens,* with 4,565 ORFs.

Three genomes that were binned annotated to the phylum Firmicutes. The first contained 6,268 ORFs with 5,866 of them annotating to the genus *Paenibacillus.* The second Firmicutes genome bin contained 4,275 ORFs, 4,060 of which annotated to the genus *Bacillus.* Of those ORFs annotating to the genus *Bacillus*, 3,416 of them annotated to *B. simplex.* The last Firmicutes genome had 4,215 ORFs that also annotated to the genus *Bacillus* with 3,313 of them annotating to *B. aquimaris* and 569 annotating to *B. vietnamensis*. Two of the binned genomes annotated to the phylum proteobacteria. The first binned genome contained 4,370 ORFs that annotated to the genus *Afipia.* The second proteobacteria genome bin contained 3,068 ORFs that annotated to the genus *Phyllobacterium*. General genome characteristics of the binned genomes are summarized in Table 17.

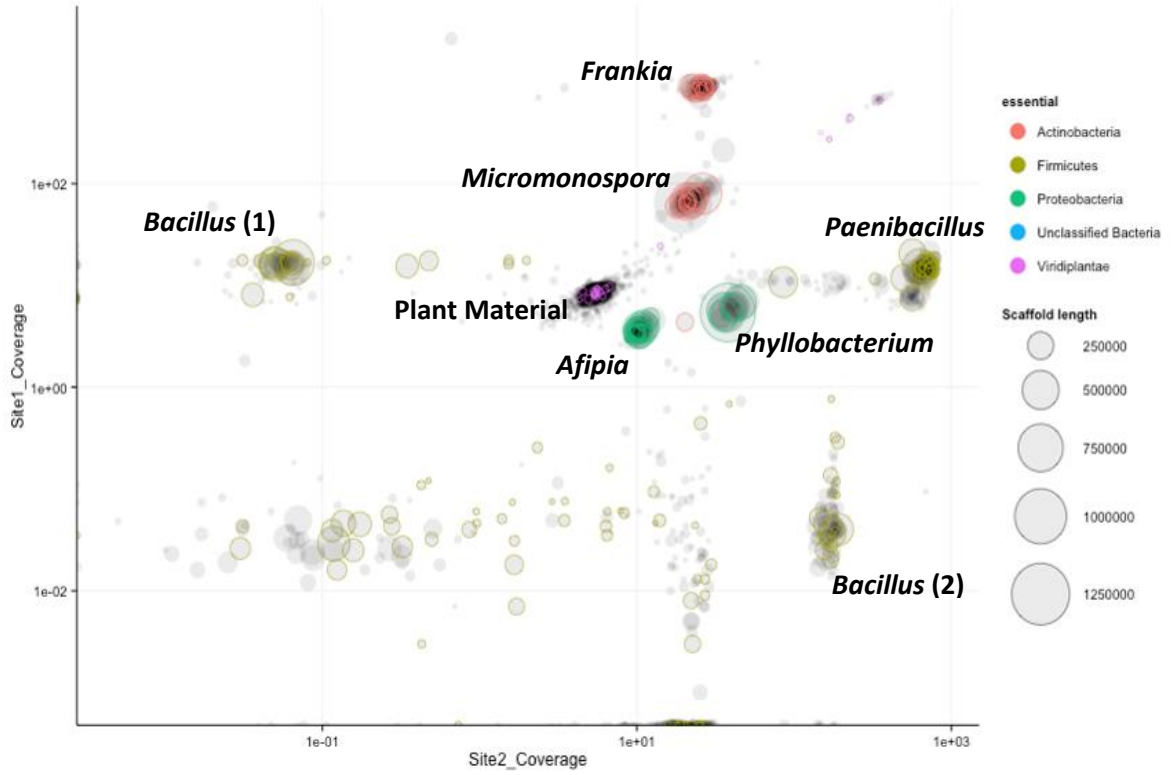


Figure 19: Plot of Metagenomic Contigs larger than 5kbp produced by the Mmgenome R Package. The seven dominant genomes that were binned using the mmgenome R package are labelled with their annotated genus. The purple and black cluster in the center of the image are contigs derived from *C. glauca* DNA.

Table 17: General Assembly Characteristics of each Binned Genome. The assembly length and GC content were calculated by Quast. Open reading frames were predicted using Prodigal. Closest annotations were determined after aligning all ORFs to the refseq protein database using BlastP.

Isolate Genus	Closest Annotation	Assembly Length	Open Reading Frames	GC %
Frankia	*Frankia* BMG5.23	4.92 Mbp	4,111	70.25%
Micromonospora	*M. purpeochromogenes*	7.28 Mbp	6,522	72.26%
Bacillus (1)	*B. Simplex*	4.64 Mbp	4,060	40.0%
Bacillus (2)	*B. aquimaris/vietnamensis*	4.7 Mbp	4,215	41.53%
Phyllobacterium	*P. sp UNC302MFCol5.2*	4.8 Mbp	3,068	59.42%
Afipia	*A. broomeae*	5.3 Mbp	4,370	61.64%
Paenibacillus	*P sp*. HGF5	7.0 Mbp	5,866	49.56%

Coverage Values for Genome Bins

Genome coverage for binned genomes was calculated in order to compare coverage patterns across sites to the replicated 16s amplicon survey. The only binned genome which followed the same patterns as the replicated analysis was the genome bin that annotated as a member of *Frankia* (Figure 9) compared to the relative abundance of *Frankia* between the two datasets. For those calculations, reads that aligned to plant-derived DNA was first removed from the readsets by removing reads that aligned to the genome bin that contained plant material (Figure 19). The resulting read sets for each site were aligned to the reassembled *Frankia* genome bin using Bowtie2. The percentage of those reads that aligned to the genome bin was recorded. The *Frankia* genome bin was the dominant member of the humid site, but coverage values were much lower for the semi-arid and arid sampling sites (Figures 5 and 20). Coverage values for all genome bins are found in Figure 20.

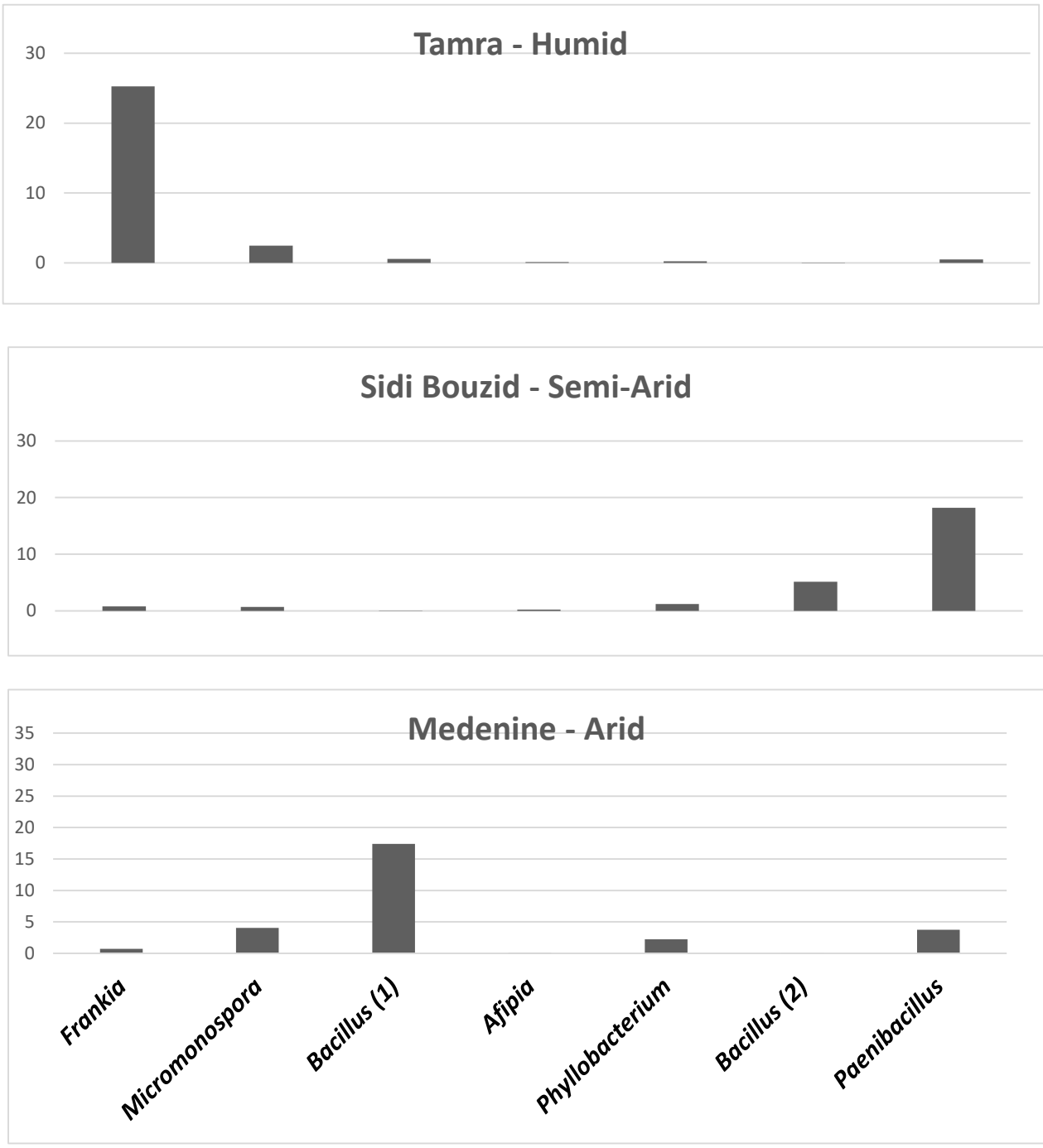


Figure 20: Normalized Genome Coverage of Genome Bins. Reads from each of the three sampling sites were aligned to the coassembled contigs. Samtools was used to sort and index the coordinates of the aligned reads and Picard MarkDuplicates was used to remove duplicate reads with identical coordinates. Bedtools was then used to calculate coverage per contig per site. Coverage values were normalized by millions of reads per data set and the relative amount of plant DNA that comprised each data set.

Secondary Metabolites Identified in Binned Genomes

The draft genomes reconstructed from the metagenomic dataset were analyzed for the presence of secondary metabolite biosynthetic pathways using AnitSMASH (Weber et al., 2015). The secondary metabolite biosynthetic clusters that were identified in the binned genomes are outlined in Tables 18-20. Predicted biosynthetic clusters from the binned genomes produced siderophores, terpenes, non-ribosomal peptide synthases and other secondary metabolites. The genome *Frankia* and *Micromonospora* genome bins contained the most secondary metabolite clusters, 25 for each bin although only clusters which have a predicted product are listed here.

Table 18: Secondary metabolites in *Frankia* genome bin identified by AntiSmash.

Type	Most Similar Known Cluster	Percent Identity
Terpene	Sioxanthin	60%
Type-Two PKS	Frankiamicin	100%
Lassopeptide-Type-One PKS	Maklamicin	13%
Terpene	Hopene	38%
Other	Streptolydigin	7%
NRPS	Triostin	16%
Type-One PKS	Divergolide	13%
Terpene	Collismycin A	7%
Type-Two PKS	Medermycin	30%
Ketide Synthase- Butyrolactone	Abyssomicin	10%
Other	A47934	10%
Type-Three Polyketide Synthase	Feglymycin	52%

Table 19: Secondary metabolites in *Micromonospora* genome bin identified by AntiSmash

Type	Most Similar Known Cluster	Percent Identity
T1PKS	Malkamacin	19%
NRPS-T1pks	Nostopeptolide	37%
Other	Diazepinomicin	75%
Oligosaccharide-PKS-Terpene	Brasilicardin A	54%
Siderophore	Desferrioxamine B	80%
T2PKS	Xantholipin	16%
Terpene	Nocathiacin	4%
Lantipeptide	Pentalenolactone	15%
NRPS	Friulimicin	12%
T3PKS	Alkyl-O-Dihydrogeranyl-Methoxyhydroquinone	71%
Nrps-T1pks-Lantipeptide	Bleomycin	12%
Terpene	Sioxanthin	80%
Nrps-Lantipeptide-T1pks-Otherks	Naphthyridinomycin	14%
T1PKS	Calicheamicin	13%
Terpene-Bacteriocin	Lymphostin	30%
T1pks-Nrps	Rifamycin	35%
T1pks	Leucanicidin	100%

Table 20: Secondary Metabolites in genomes bins identified by AntiSmash

Type	Most Similar Known Cluster	Percent Identity
Paenibacillus		
Ectoine	Ectoine	75%
Nrps	Bacillibatin	53%
Trans-At Pks	Bacillaene	21%
Bacillus (1)		
Siderophore	Desotamide	9%
Nrps	Koranimine	87%
Bacillus (2)		
Terpene	Carotenoid	33%
Siderophore		16%
Afipia		
Terpene	Malleobactin	11%
Phyllobacterium		
Nrps	Vicibactin	77%

Functional Mining of Metagenomes

The shotgun metagenomic data was mined for genes of interest that could be of functional importance to the actinorhizal symbiosis and *C. glauca* health. Genes potentially responsible for plant – growth- promoting traits were mined from GFF files produced by the Prokka Annotation pipeline (Seamann et al., 2014). The list of these genes are presented in the methods section (Table 1-3). Figures 21 – 23 shows the results of this analysis as outlined by site and are cladograms for the taxonomy of the genes identified. The number of genes for each function is reported in each figure. The sampling site 1, site 2 and site 3 had 159, 183 and 202 genes of interest, respectively. Since shotgun sequencing was not performed with replicates, these results are presented qualitatively as presence or absence data.

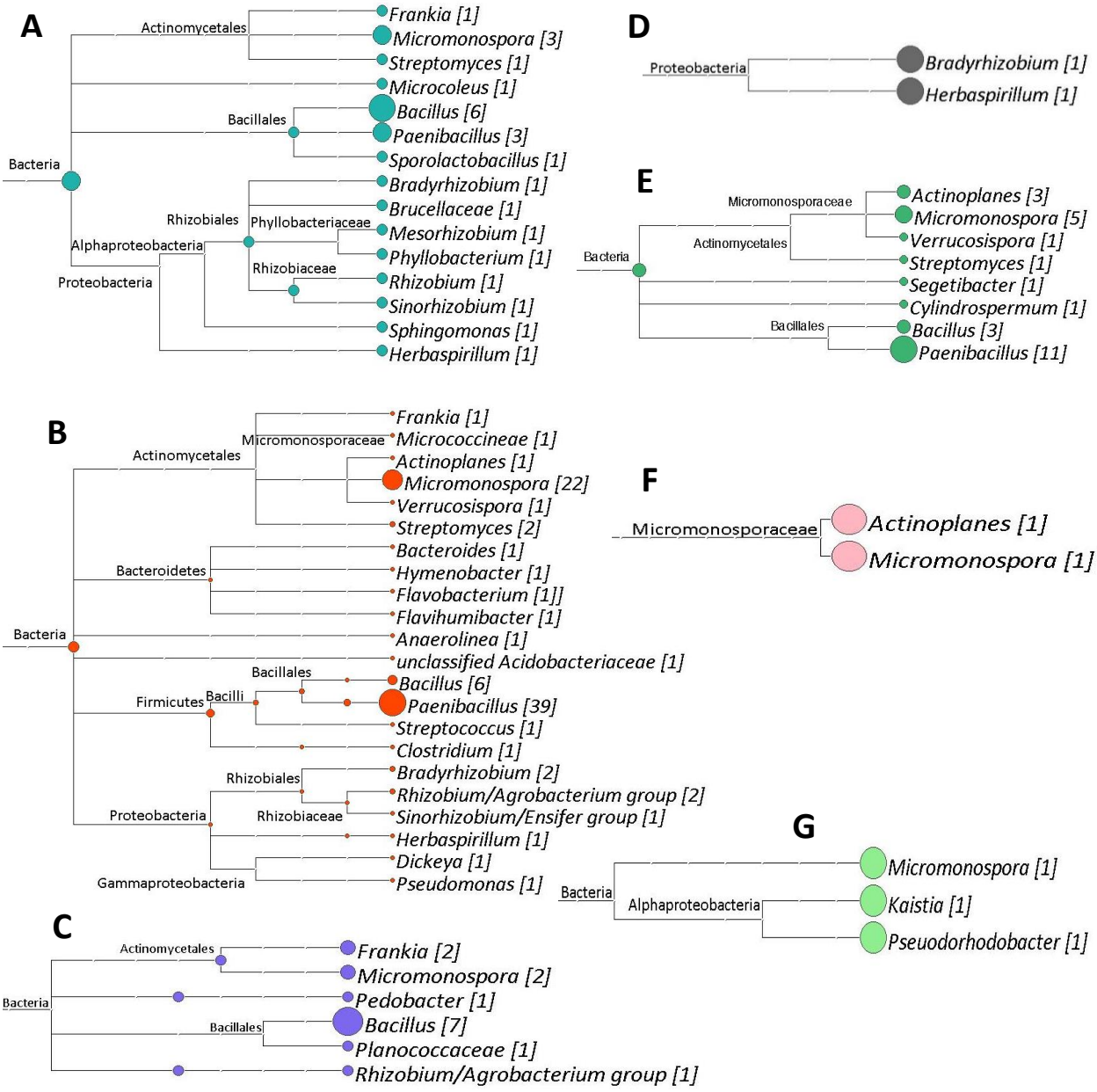


Figure 21: Cladograms of functional genes from Site 1 (Tamra), that were produced by the LCA Algorithm in MEGAN (Version 5). The sizes of colored circles represent that relative contribution of a taxa to the functional genes analyzed in a given cladogram. Numbers in Brackets represent that actual amount of genes that annotated to a given taxon for that set of genes. **A: Potential Indole Acetic Acid** biosynthetic genes. Total genes, 24. **B: Cellulose** degrading genes. Total genes identified, 89 **C. Alkaline Phosphatase** genes. Total genes identified, 26. **D: ACC Deaminase** genes. Total genes Identified, 2. **E: Chitin** degrading genes. Total genes identified, 14. **F: Pectinase** genes. Total genes identified, 2. **G: Phytase** genes. Total genes identified, 2.

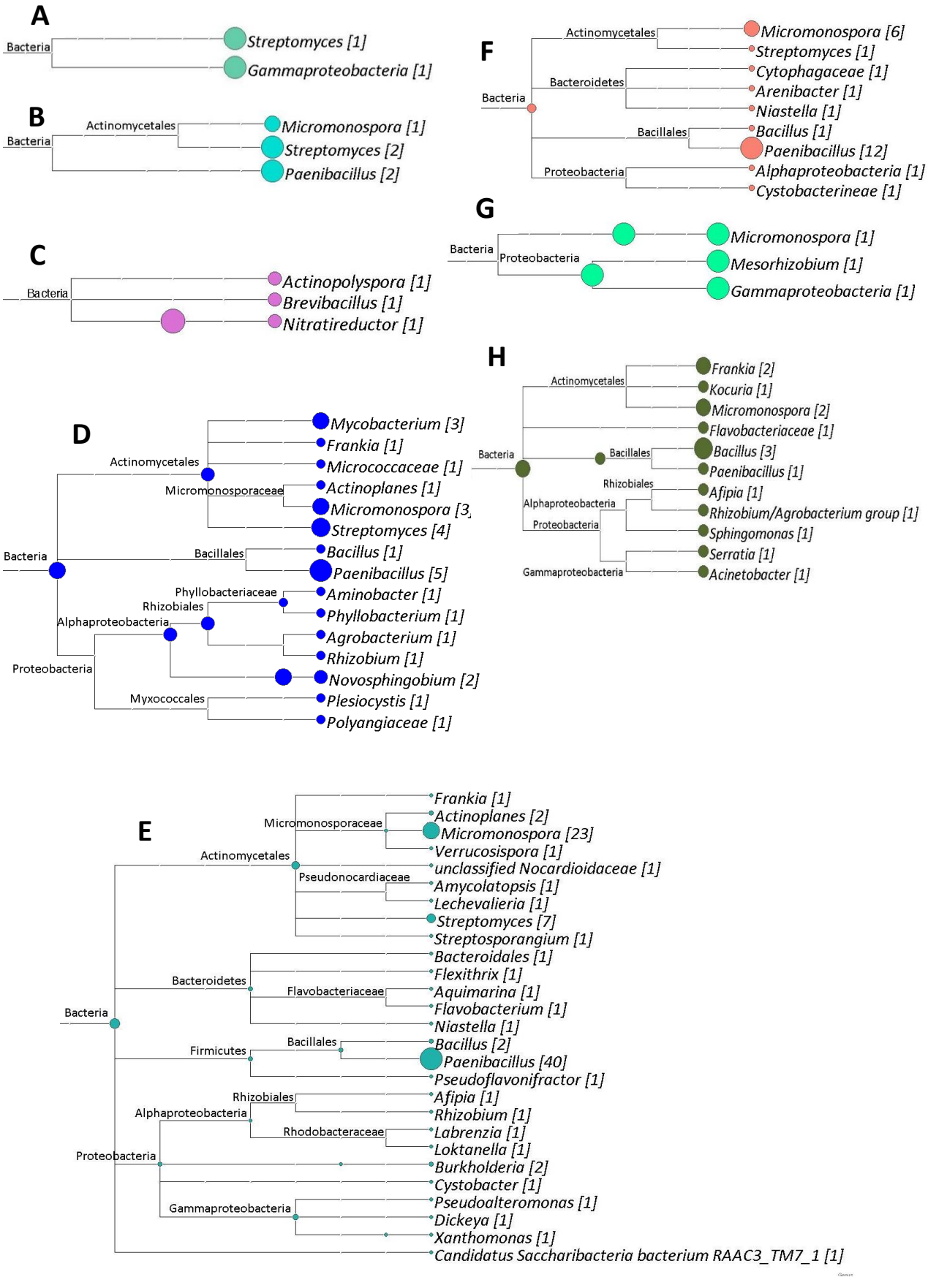

A
Bacteria
Streptomyces [1]
Gammaproteobacteria [1]
B
Actinomycetales
Bacteria
Micromonospora [1]
Streptomyces [2]
Paenibacillus [2]
C
Bacteria
Actinopolyspora [1]
Brevibacillus [1]
Nitratireductor [1]
D
Mycobacterium [3]
Frankia [1]
Actinomycetales
Micrococcaceae [1]
Micromonosporaceae
Actinoplanes [1]
Micromonospora [3]
Streptomyces [4]
Bacteria
Bacillales
Bacillus [1]
Paenibacillus [5]
Phyllobacteriaceae
Aminobacter [1]
Rhizobiales
Phyllobacterium [1]
Alphaproteobacteria
Agrobacterium [1]
Rhizobium [1]
Proteobacteria
Novosphingobium [2]
Myxococcales
Plesiocystis [1]
Polyangiaceae [1]
E
Frankia [1]
Micromonosporaceae
Actinoplanes [2]
Micromonospora [23]
Verrucosispora [1]
Actinomycetales
unclassified Nocardioidaceae [1]
Pseudonocardiaceae
Amycolatopsis [1]
Lechevalieria [1]
Streptomyces [7]
Streptosporangium [1]
Bacteroidales [1]
Bacteroidetes
Flexithrix [1]
Flavobacteriaceae
Aquimarina [1]
Bacteria
Flavobacterium [1]
Niastella [1]
Bacillales
Bacillus [2]
Firmicutes
Paenibacillus [40]
Pseudoflavonifractor [1]
Rhizobiales
Afipia [1]
Alphaproteobacteria
Rhizobium [1]
Rhodobacteraceae
Labrenzia [1]
Loktanella [1]
Proteobacteria
Burkholderia [2]
Cystobacter [1]
Pseudoalteromonas [1]
Gammaproteobacteria
Dickeya [1]
Xanthomonas [1]
Candidatus Saccharibacteria bacterium RAAC3_TM7_1 [1]
F
Actinomycetales
Micromonospora [6]
Streptomyces [1]
Cytophagaceae [1]
Bacteroidetes
Arenibacter [1]
Bacteria
Niastella [1]
Bacillales
Bacillus [1]
Paenibacillus [12]
Proteobacteria
Alphaproteobacteria [1]
Cystobacterineae [1]
G
Bacteria
Micromonospora [1]
Proteobacteria
Mesorhizobium [1]
Gammaproteobacteria [1]
H
Frankia [2]
Actinomycetales
Kocuria [1]
Micromonospora [2]
Flavobacteriaceae [1]
Bacteria
Bacillales
Bacillus [3]
Paenibacillus [1]
Rhizobiales
Afipia [1]
Alphaproteobacteria
Rhizobium/Agrobacterium group [1]
Proteobacteria
Sphingomonas [1]
Serratia [1]
Gammaproteobacteria
Acinetobacter [1]

Figure 22: Cladograms of functional genes from Site 2 (Sidi Bouzid), that were produced by the LCA Algorithm in MEGAN (Version 5). The sizes of colored circles represent that relative contribution of a taxa to the functional genes analyzed in a given cladogram. Numbers in Brackets represent that actual amount of genes that annotated to a given taxon for that set of genes. **A: Phytase** genes. Total genes identified, 8. **B: Pectinase** enzymes. Total genes identified, 5. **C: ACC Deaminase** genes. Total genes identified, 3. **D: Potential Indole Acetic Acid** biosynthesis genes. Total genes identified, 27. **E: Cellulose degrading** enzymes. Total genes identified, 97. **F: Chitin** degrading enzymes. Total genes identified, 25. **G: Potential Phenyl Acetic Acid** biosynthesis genes. Total genes identified, 3. **H: Alkaline Phosphatase** genes. Total genes identified, 15.

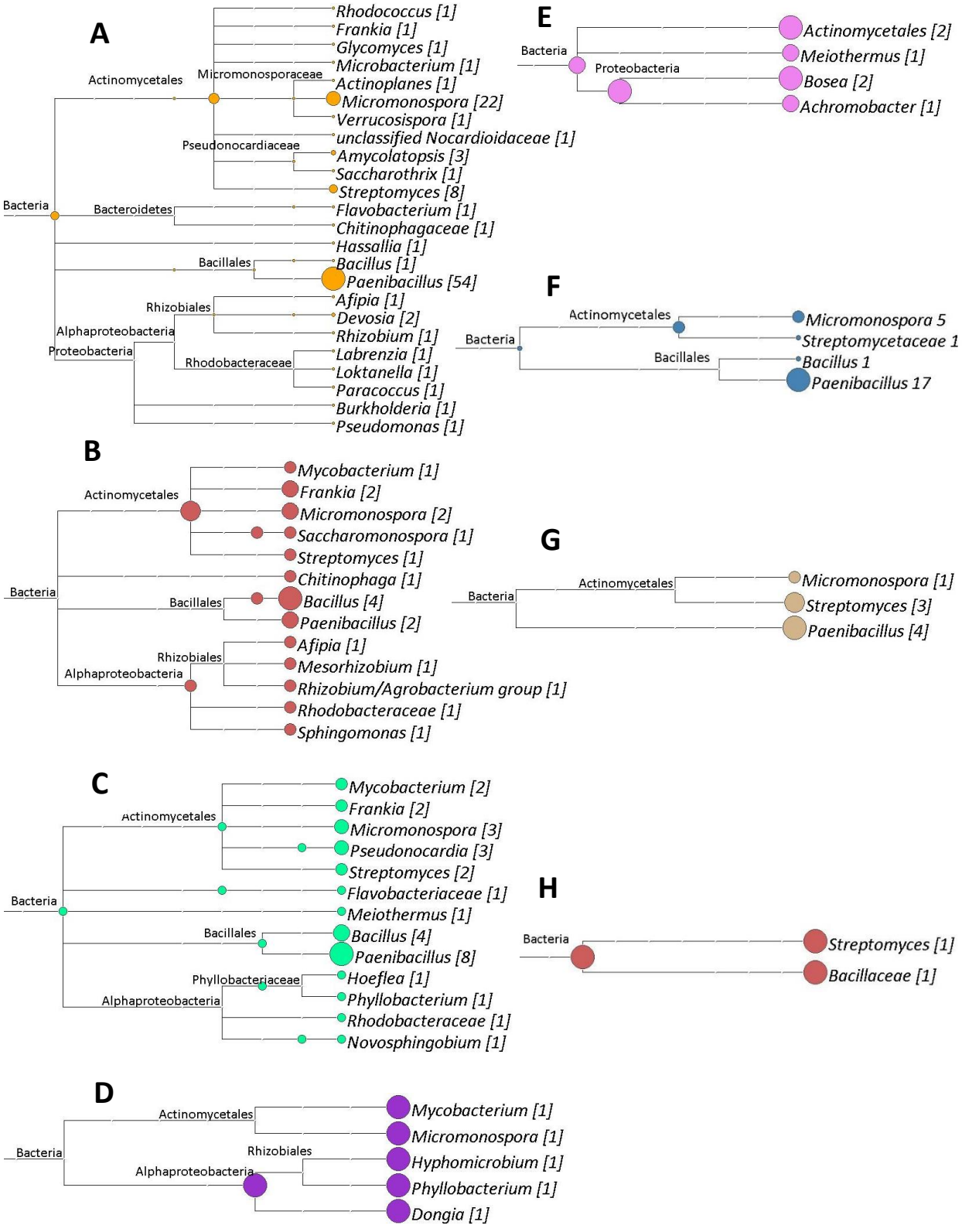

A
Rhodococcus [1]
Frankia [1]
Glycomyces [1]
Microbacterium [1]
Actinoplanes [1]
Micromonospora [22]
Verrucosispora [1]
unclassified Nocardioidaceae [1]
Amycolatopsis [3]
Saccharothrix [1]
Streptomyces [8]
Flavobacterium [1]
Chitinophagaceae [1]
Hassallia [1]
Bacillus [1]
Paenibacillus [54]
Afipia [1]
Devosia [2]
Rhizobium [1]
Labrenzia [1]
Loktanella [1]
Paracoccus [1]
Burkholderia [1]
Pseudomonas [1]
Actinomycetales
Micromonosporaceae
Pseudonocardiaceae
Bacteria
Bacteroidetes
Bacillales
Rhizobiales
Alphaproteobacteria
Proteobacteria
Rhodobacteraceae
E
Actinomycetales [2]
Meiothermus [1]
Bosea [2]
Achromobacter [1]
Bacteria
Proteobacteria
F
Micromonospora 5
Streptomycetaceae 1
Bacillus 1
Paenibacillus 17
Actinomycetales
Bacteria
Bacillales
B
Mycobacterium [1]
Frankia [2]
Micromonospora [2]
Saccharomonospora [1]
Streptomyces [1]
Chitinophaga [1]
Bacillus [4]
Paenibacillus [2]
Afipia [1]
Mesorhizobium [1]
Rhizobium/Agrobacterium group [1]
Rhodobacteraceae [1]
Sphingomonas [1]
Actinomycetales
Bacteria
Bacillales
Rhizobiales
Alphaproteobacteria
G
Micromonospora [1]
Streptomyces [3]
Paenibacillus [4]
Actinomycetales
Bacteria
C
Mycobacterium [2]
Frankia [2]
Micromonospora [3]
Pseudonocardia [3]
Streptomyces [2]
Flavobacteriaceae [1]
Meiothermus [1]
Bacillus [4]
Paenibacillus [8]
Hoeflea [1]
Phyllobacterium [1]
Rhodobacteraceae [1]
Novosphingobium [1]
actinomycetales
Bacteria
Bacillales
Phyllobacteriaceae
Alphaproteobacteria
H
Streptomyces [1]
Bacillaceae [1]
Bacteria
D
Mycobacterium [1]
Micromonospora [1]
Hyphomicrobium [1]
Phyllobacterium [1]
Dongia [1]
Actinomycetales
Bacteria
Rhizobiales
Alphaproteobacteria

Figure 23: Cladograms of functional genes from Site 3 (Gabes), that were produced by the LCA Algorithm in MEGAN (Version 5). The sizes of colored circles represent that relative contribution of a taxa to the functional genes analyzed in a given cladogram. Numbers in Brackets represent that actual amount of genes that annotated to a given taxon for that set of genes. **A: Cellulose** degrading genes. Total genes identified, 108. **B: Alkaline Phosphatase** genes. Total genes identified, 19. **C. Potential Indole Acetic Acid** biosynthetic genes. Total genes identified, 30. **D: Potential Phenyl Acetic Acid** biosynthesis genes. Total genes identified, 5. **E: ACC Deaminase** genes. Total genes Identified, 6. **F: Chitin** degrading genes. Total genes identified, 24. **G: Pectinase** genes. Total genes identified, 8. **H: Phytase** genes. Total genes identified, 2.

CHAPTER 4

DISCUSSION

Significance

The data presented here represents the first time that Next Generation Sequencing (NGS) technologies has been used to capture the microbial community associated with the actinorhizal plant *C glauca*. Since *C. glauca* is a model actinorhizal plant and is prevalent in field projects (Mailly and Margolis, 1992; Muthukumar and Udaiyan 2010; Zhong et al., 2013), its use in in this study was appropriate. The advantageous physiological and ecological traits of *C. glauca* have been used extensively in Africa, and thus understanding the taxonomic composition and genetic potential of the microbiome of this plant outside of its native range can allow us to gain more insight into this system.

The culture-based nature of most of the studies of this system have provided insight into strategies to improve this system for the practical uses it is employed for. This study has used culture-independent techniques to add to this knowledge-base by investigating the community structure *in planta* without the bias introduced by culturing organisms. These two contrasting approaches can be used together to investigate the basic ecology of this system and for the potential development of applications of bioinoculants.

Abundance of *Frankia* in nodule samples

The shotgun and amplicon sequencing results agree with previous work showing that *Casuarina* outside of their native range only host one genotype of *Frankia*, or a few very closely related strains (Simonet et al., 1999). The shotgun sequencing data contained only one predominant genome

bin annotating to *Frankia.* This genome bin most closely annotated to *Frankia* sp. BMG5.23, which is a Tunisian isolate from *C. equestifolia* (Ghodhbane-Gtari et al., 2014). There were four OTUs annotating to the genus *Frankia* that were differentially abundant in the nodule samples (Table 6, Figures 11 – 12). All of these OTUs were very closely related, sharing 99% identity with *Frankia* CcI3. These results also agree with the previous observation that *Casuarina* nodules in Africa contain *Frankia* populations with low genetic diversity (Simonet et. al, 1999).

Reads mapping to the *Frankia* genome bin consisted of 60% of the read-set for the humid sampling site. This number dropped to 1.7% and 1.63% in the semi-arid and arid sampling sites, respectively. This decrease in *Frankia* abundance in nodules in arid environments was a major finding of this work that was not initially expected. In the amplicon survey data, *Frankia* comprised 80% of the nodule community and 1.5% and 0.5% in the semi-arid and arid sampling sites, respectively. These numbers do not match exactly with the shotgun data, but that is likely because of the increased DNA associated with epiphytes in the shotgun dataset. Regardless, the overall trend that *Frankia* abundance is drastically decreased in arid environments is apparent in both datasets. These results suggest that *C. glauca* growing in the harsh environments in semi-arid and arid regions of Africa may be contributing less nitrogen to these plants than those in humid conditions. Annotation of the functional metagenome by Prokka resulted in the detection of only *Frankia nif* genes, indicating the absence of other nitrogen-fixing organisms. The addition of nitrogen into nutrient-limited environments is an important property of the actinorhizal symbiosis for land reclamation, therefore this observation has strong implications for this field of study, if confirmed. The nitrogen fixation capacity of the nodules was not tested. However, the plants were observed to be green and healthy at all of the sampling sies (Ghodhbane-Gtari, Unpublished).

Our results show lower *Frankia* nodule occupancy of *C. glauca* nodules, which was surprising. Detecting *Casuarina* infective clades of *Frankia* through plant-trapping assays have proven difficult (Gtari

et al., 2002, Gtari, et al., 2007) and could be explained by the scarcity of detection of *Casuarinaceae* infective *Frankia* outside of the plants native range (Mirza et al., 2009). It is possible that *Casuarina* infective *Frankia* are more persistent in humid environments that are more similar to the riparian habitat that is preferred by *Casuarina* within their host range. These results are congruent with the need for intentional inoculation with a compatible *Frankia* strain necessary for nodulation in some land reclamation projects in Africa (Gauthier et al., 1985). With increased sequencing depth, it may be possible to detect *Frankia* in these samples, but it would still be in very low relative abundance. In this analysis, sequences annotating to *Frankia*, at the genus level, were identified in five of the eighteen rhizosphere and soil samples (Data not shown). Four of those samples contained one or two annotated reads, while the fifth contained 24 annotated reads in a rhizosphere sample from Medenine, the arid sampling site. This could suggest that the *Casuarina* rhizosphere is an environment that hosts can active *Frankia* population, as been previously observed (Beauchemin et al., 2011, Samant et al., 2015). This observation could benefit from a more thorough investigation of rhizosphere dynamics of *Frankia* populations in the *Casuarina* rhizosphere.

One possible explaination of the phenomenon of low *Frankia* abundance in these nodules is that the physical and chemical characteristics of semi-arid and arid environments are not benefical for the maintenance of the actinorhizal symbiosis or *Frankia* persistence. Soils in the semi-arid and arid regions of Tunisia are poorly evolved calcareous soils (Mtimet, A, 2001). In these soils, phosphorous is known to be immobilized by formation of complexes with calcium cations (Tunesi et al., 1999). Low phosphorus levels have been shown to decrease the frequency of nodulation of *C. cunninghamiana* by *Frankia* (Yang, 1995) and decreases nodule weight of *Frankia-C. equestifolia* (Sanginga et al., 1989). Similarly, drying and heat decrease the ability of some *Frankia* strains to infect *C. equestifolia* (Sayed et al., 1997). High heat and low levels of available phosphorous and water are characteristics of soils in arid

environments and these factors could be influencing the low abundance and absence of *Frankia* in nodules at these environments.

However, this reasoning does not explain the absence of *Frankia* within mature nodules. The absence of detected *Frankia* in four of the nine nodule samples leads to the thought that another member of the community may be responsible for the formation of nodules on *C. glauca*. *Agrobacterium rhizogenes* will induce pseudonodules on *Elaeagnus angustifolia* that are indistinguishable from *Frankia* – induced nodules (Berg et al., 1991). Our results show an OTU of genus *Agrobacterium* that was as differentially abundant in nodule samples (Table 6). The genus *Agrobacterium* was present in eight of the nine nodule samples ranging from 0.01% to 2.8% relative abundance. The previous work with *Agrobacterium* and its differential abundance in nodules in this study suggests that a member of this genus could be play a role in nodule formation when *Frankia* is in low abundance.

Another hypothesis is that this absence is a seasonal variation and part of the life cycle of F*rankia* under semi-arid and arid conditions. A seasonal study of the changes in community structure could address this hypothesis.

Distinct Nodule Community Across Sampling Sites

MRPP analysis found that nodule samples formed a statistically significant group compared to the rhizosphere and soil samples (Figures 8 - 9). Nodule samples are separated from the rhizosphere and soil samples by the first principal coordinate axis, which explains the most taxonomic information in the samples. This result is also confirmed with the jackknifed UPGMA tree built from the unweighted UniFrac distance matrix (Figure 7), which shows a distinct clade where all nodule samples clustered, despite location. Alpha diversity of the nodule samples was also significantly lower than for the rhizosphere or soil samples (Figure 3). These observations are similar to comparative analysis of the

root microbiome of the model plant *Arabidopsis thaliana,* which has shown that endophytic communities from plants growing in chemically distinct soil types are highly overlapping and less diverse than surrounding soils (Lundberg et al., 2012). The community selected by *C. glauca* was distinct from the soil and rhizosphere communities across a steep environmental gradient and was significantly less diverse, similar to the pattern of the microbiome of *A. thaliana*. The similar patterns of microbiome recruitment by *C. glauca* and *A. thaliana* displays that phylogenetically distant plant groups tend to select for a particular endophytic community across environmental gradients.

Nodule Composition in Arid Environments

Although the nodule samples do host a unique community, the relative composition of these taxa change drastically across the environmental gradient (Figure 10). *Frankia* was the predominant community member in the humid environment. In the semi-arid environment, there was no dominant member and the alpha diversity was the highest among the three sampling locations (Figure 4). In the arid environment, Medenine, 59% of the community was comprised of the genus *Acinetobacter*. This is the first finding of this genus as an endophyte of actinorhizal plants. Members of *Acinetobacter* have been isolated from the leguminous shrub *Prosopis farcta* (Fterich et al., 2011) and *Acacia tortilis* subsp. *Raddiana* (Fterich et al., 2012) in the same regions of Tunisia that the samples were collected for this study. These studies were isolation-based and did not collect any information about the composition of community structure. *Acinetobacter* isolates from arid environments in Mexico have been demonstrated to produced plant auxins and solubilize phosphate (Moreno-Ramirez et al., 2014). In the metagenomic data from this study, phosphate solubilizing genes were found that annotated to *Acinetobacter*. *Acacia* is also an introduced species in Africa that has been imported for much of the same reasons that *Casuarina* has. These results suggest that *Acinetobacter* could be important as

endophytes for nodulating woody plants in arid environments. The specific interaction between *Acinetobacte*r and the actinorhizal symbiosis could benefit from laboratory investigation.

Functions of Taxa Differentially Abundant Nodules

Analysis of the amplicon data identified a group of taxa that are significantly more abundant in the nodule samples (Table 6). Functional mining of the shotgun data set found potential plant growth promoting genes that annotated to taxa that were identified as differentially abundant (Figures 18 - 20) suggesting a recruitment of potential PGPB by *C. glauca*. Since these two datasets are independent of each other, the differential abundance of a particular function sharing the taxonomy of an OTU can only be deduced. OTUs that were differentially abundant in the nodule samples share taxonomy with potential plant growth promoting genes that annotated to *Sphingomonas*, *Serratia*, *Acinetobacter*, *Agrobacterium* and the family *Micrococceae*. Genes with potential phosphate solubilization and indole production annotated to *Sphingomonas*, *Agrobacterium* and *Acinetobacter*, while there were phosphate solubilizing genes that annotated to *Serratia* and potential indole producing genes that annotated to the family *Micrococceae*.

These genera have been previously identified as endophytes of many kinds of plants. *Sphingomonas*, a PGPB, is an endophyte of tomato (Khan et al., 2014). *Agrobacterium*, an endophyte of the legume shrub *Retama raetam* in Tunisia (Mahdhi et al., 2008), forms pseudonodules on *E. angustifolia* (Berg et al., 1991). *Acinetobacter* is a legume-nodule occupier in similar arid environments of Tunisia as the samples for our study. An OTU of *Serratia marcescens* was more abundant in nodule samples and the genus *Serratia* was identified as a weak indicator species of nodule samples. *S. marcescens* have both plant-pathogenic and nonpathogenic varieties (Zhang et. al, 2005) and endophytic subgroups have been identified (Tan et al., 2001). The amplicon based results, in combination with the shotgun metagenomic data show that Serratia are are more abundant in *C. glauca* nodules having the genetic capability for plant growth promotion.

Actinobacteria with Plant Growth Promoting Genes

There have been many studies showing that non-*Frankia* actinobacteria act as helper-bacteria in the formation of the actinorhizal symbiosis (Ghodhbane-Gtari et al., 2009; Solans et al., 2009, 2011). Our results show that some of the genera and their relative abundance were previously present in wild nodules and. Members of the genera *Streptomyces* and *Actinoplanes* isolated from surface-sterilized nodules from actinorhizal plants (Ghodhbane-Gtari et al., 2009) have been identified as helper-bacteria in the symbiosis between *Discaria trinervis* and *Frankia* (Solans, 2007). Isolates from these genera from root nodules of *Ochetophila trinervis* have been demonstrated to produce IAA (Solans et al., 2011). *Streptomyces* was found in six of the nine root nodules samples, with an abundances ranging from 0.022 % to 0.22%. Mining the metagenomes for PGPB associated traits identified phytase, alkaline phosphatase, potential IAA production, cellulose degradation and pectinase genes that annotated to *Streptomyces* (Figures 18 - 20). The genus *Actinoplanes* was also identified in six of the nine nodule samples, with abundances ranging from 0.03% to 0.3%. Genes involved in cellulose, chitin and pectin degradation and IAA production annotated to *Actinoplanes*. It is possible that actinobacteria with saprophytic capabilities and helper-bacteria are regular nodule inhabitants of *C. glauca*.
In contrast to the differential abundant OTUs and indicator species discussed earlier, the genera discussed here were not identified as differentially abundant in nodules or indicator species of nodule samples. This result suggests that there is no preference for endophytic or free-living niches for these actinobacteria

Epiphyte and Rhizoplane Occupying Taxonomy and Function

The sampling design for the shotgun DNA sequencing captured the epiphyte and endophyte community associated with *C. glauca* nodules because the root nodules were not surface sterilized prior to DNA extraction and sequencing. The replicated nodule samples for the amplicon analysis were

surface sterilized. This sampling design allows the epiphyte and endophyte community in the shotgun data set to be captured. Surface colonizing taxa play active roles in nutrient acquisition and can influencing the health of *C. glauca* and the actinorhizal symbiosis, therefore this fraction of the community is of interest. The most notable differences in these data are the absence of the taxa of the binned genomes in the same proportions in the replicated amplicon data. These abundance patterns suggest that *Paenibacillus*, *Afipia*, *Phyllobacterium* and the *Bacillus* genome bins are surface colonizing epiphytes. The genus *Micromonospora* has been identified as an endophyte of actinorhizal and leguminous plants (Solans et al., 2011, Trujillo et al., 2006, 2007), but in this analysis it was only found as an epiphyte. The coverage values for these genome bins was variable. With the absence of replication for this sequencing set, it can only be concluded that these organisms are likely epiphytes that colonize the surface of the plant.

The taxonomy of these epiphytes agrees with previous studies (Echbab et al., 2004; Solans et al., 2011, Zahkia et al., 2006) and the potential genetic functions identified suggest roles for these genera. Two different genome bins in the genus *Bacillus* were identified as epiphytes, one being closely related to *B. simplex* the other being closely related to *B. aquimaris* and *B. vietnamisis*. *B. simplex* has been studied for its adaptations to arid environments (Sikorski and Nevo, 2005). Co-inoculation studies have shown that *B. simplex*is a helper-bacteria for the symbiosis between *Pisum sativum* and *Rhizobium legiminosarum* bv *viciae* (Schwartz et al., 2013). *Bacillus* isolates are helper-bacteria for the *Frankia-Casuarina* symbiosis (Echbab et al., 2004). The current analysis identified biosynthetic clusters for siderophore production in the two *Bacillus* genomes and also included PGP genes for potential auxin product and phosphate solubilization. The localization of *Bacillus* on the root-nodule surface could be an advantageous to the plant. *Paenibacillus* genes similar to those of *Bacillus* were found for secondary metabolic clusters for siderophore production, ectoine, an osmoprotectant, and bacillaene, an antibiotic. Inoculation of *C. equestifolia* with this *Paenibacillus polymyxa* and a fungus, *Glomus*

geosporum resulting in the highest seedling quality in a nursery trial (Muthukumar and Udaiyan 2010). Our survey shows that these PGPB occupy the root nodule surface of *C. glauca* growing wild under natural conditions.

Members of the genus *Phyllobacterium* have also been identified as PGPB, including as an opportunistic endophyte of legumes in arid parts of Tunisia (Zakhia et al., 2006). Growth of the root structure of *Brassica napus* is enhanced by inoculation with a *Phyllobacterium* sp. isolate (Larcher et al., 2003). Potential *Phyllobacterium* IAA and PAA producing genes were identified (Figures 18,20) suggesting that members of *may* actias a PGPB for *C. glauca.*

The genus *Micromonospora* has received much attention as a nodule occupier in actinorhizal systems (Carro et al., 2013, Ghodhbane-Gtari et al., 2009, Solans et al., 2007, Solans et al., 2011, Trujillo et al., 2006) and legume systems (Martinez-Hidalgo et al., 2014, Solans et al., 2009, Truijillo et al., 2010, Truijillo et al., 2015). Members of this genus are recognized as an important source for biocontrol agents (Hirsch and Valdes, 2010). *Micromonospora* is an antagonist of fungal plant pathogens through the production of extracellular glucanases (El-Tarabily et al., 2009) and antifungal secondary metabolites (Ismet et al., 2004). The *Micromonospora* genome found in the nodule contained 25 secondary metabolite clusters including potential auxin biosynthesis genes, pectinase genes, cellulose degrading genes, alkaline phosphatase and phytase genes, and chitinase genes. In our study, *Micromonospora* was not identified as an endophyte of the surface sterilized nodules, but it was detected as a surface-colonizer in all three sampling sites. This suggests that *Micromonospora* likely acts as a PGPB and helper bacteria in the actinorhizal symbiosis in a diversity of environments.

Rhizosphere and Soil Analysis

One hypothess is that root exudates affect the soil community by creating a distinct rhizosphere community, which that is maintained despite sampling location. Our data showed a weak grouping of

samples into distinct soil or rhizosphere clusters when all sites were combined for analysis. Grouping was statistically significant by MRPP for the OTU table that was summarized at relative abundance with a 16s operon copy number normalization and the OTU table that was evenly rarefied, but was not significant with the data that was normalized using CSS. Differential abundance analysis revealed OTUs that were differentially upregulated in the rhizosphere when all sampling sites were analyzed as one group. Three of these OTUs were Betaproteobacteria that annotated only to the order MND1, and one OTU annotated only to the class Betaproteobacteria. The Betaproteobacteria order MND1 has been identified as an abundant unclassified lineage in surveys of soil microbial communities, but there are no isolates of this order and no physiological data to aid in interpreting the abundance of this OTU (Spain et al., 2009). This order was also the highest indicator species of the pooled rhizosphere sample. It can be seen in Figures 13 and 15 that OTUs from this order are upregulated in rhizosphere samples from Tamra (humid) and Medenine (arid). This suggests that members of this order occupy the rhizosphere of C. glauca in varying environments. There are a number of uncultured representatives in the Deltaproteobacteria (orders NB1-j, NB-i) that are indicator species of the pooled rhizosphere samples that have no literature to aid in the interpretation of their indicator values in the rhizosphere.

The clustering pattern of soil and rhizosphere samples were markedly different when ordinated using the unweighted UniFrac. The ordination shows samples clustering by environmental type, with samples from the Tamara, the humid sample site, clustering together and samples from the semi-arid and arid sample sites forming another cluster. The unweighted UniFrac distance metric measures the phylogenetic distance between samples based on the presence or absence of taxa between samples. Significant differences between communities seen using the unweighted UniFrac metric suggests that the two environments have strong differences based what taxa are able to persist between the samples. Clustering of microbial communities using the unweighted UniFrac due to chemical and physical properties of environments has been previously documented (Lozupone et al., 2007). It is likely that the

soil chemistry and environmental differences between the humid site and the arid sites select for organisms with different physiological requirements that explains this pattern. These data show that the differences between a humid and arid soils structure the soil community, while effects of the *C. glauca* rhizosphere do have a slightly significant influence of the relative abundance of select taxa.

Although climate shapes bacterial community structure in a stronger fashion than the effects of the *C. glauca* rhizosphere, there are differentially abundant OTUs within each individual rhizosphere (Figures 13 – 15). This suggests that different taxa are able to utilize the carbon sources supplied by *C. glauca* in different environments. As discussed above, the differences in environment could be excluding taxa in arid environments that are stimulated in the rhizosphere in humid environments. This could explain the difference in taxonomy of the differentially abundant OTUs identified in the rhizosphere samples at the different locations.

References

Albertsen M, **Hugenholtz P**, **Skarshewski A**, **Nielsen KL**, **Tyson GW**, **Nielsen PH**. 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat Biotechnol **31**:533–8.

Amann, R.I., Ludwig, W. and Schleifer, K.H., 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. rev **59:**143-169.

Arahou M, **Diem HG**. 1997. Iron deficiency induces cluster (proteoid) root formation in Casuarina glauca. Plant Soil **196**:71–79.

Baker, E., Tang, Y., Chu, F. and Tisa, L.S. 2015. Molecular responses of *Frankia* sp. strain QA3 to naphthalene. Canadian Jour Micrbiol. **61:**281-292.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D. and Pyshkin, A.V. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journ Comp Biol. **19:**455-477.

Baudoin E, **Benizri E**, **Guckert A**. 2003. Impact of artificial root exudates on the bacterial community structure in bulk soil and maize rhizosphere. Soil Biol Biochem **35**:1183–1192.

Benson DR, **Heuvel BD Vanden**, **Potter D**. 2004. Actinorhizal symbioses : diversity and biogeography. Plant microbiology. BIOS Scientific Publishers Ltd, Oxford. 99–130.

Berg RH, **Liu LX**, **Dawson JO**, **Savka MA**, **Farrand SK**. 1992. Induction of Pseudoactinorhizae by the Plant Pathogen Agrobacterium-Rhizogenes. Plant Physiol. **98**:777–779.

Beauchemin, N.J., Furnholm, T., Lavenus, J., Svistoonoff, S., Doumas, P., Bogusz, D., Laplaze, L. and Tisa, L.S. 2012. Casuarina root exudates alter the physiology, surface properties, and plant infectivity of Frankia sp. strain CcI3. App Env Microbiol. **78:**575-580.

Benson DR, **Dawson JO**. 2007. Recent advances in the biogeography and genecology of symbiotic Frankia and its host plants. Physiol Plant. **130**:318–330.

Berry, K.J. and Mielke, P.W. 1984. Computation of exact probability values for multi-response permutation procedures (MRPP). Comm in Statistics-Simulation and Comp. **13**:417-432.

Berry, A.M., McIntyre, L. and McCully, M.E. 1986. Fine structure of root hair infection leading to nodulation in the Frankia-Alnus symbiosis. Canadian Journ. of Bot. **64**:292-305.

Bissonnette C, **Fahlman B**, **Peru KM**, **Khasa DP**, **Greer CW**, **Headley JV**, **Roy S**. 2014. Symbiosis with Frankia sp. benefits the establishment of Alnus viridis ssp. crispa and Alnus incana ssp. rugosa in tailings sand from the Canadian oil sands industry. Ecol Eng. **68**:167–175.

Bolger, A.M., Lohse, M. and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 1-7

Bourke, R.M. 1985. Food, coffee and casuarina: an agroforestry system from the Papua New Guinea highlands. Agroforestry Systems. **2**:273-279.

Cáceres, M.D. and Legendre, P. 2009. Associations between species and groups of sites: indices and statistical inference. Ecology. **90**:3566-3574.

Caporaso JG, **Kuczynski J**, **Stombaugh J**, **Bittinger K**, **Bushman FD**, **Costello EK**, **Fierer N**, **Peña AG**, **Goodrich JK**, **Gordon JI**, **Huttley G a**, **Kelley ST**, **Knights D**, **Koenig JE**, **Ley RE**, **Lozupone C a**, **McDonald D**, **Muegge BD**, **Pirrung M**, **Reeder J**, **Sevinsky JR**, **Turnbaugh PJ**, **Walters W a**, **Widmann J**, **Yatsunenko T**, **Zaneveld J**, **Knight R**. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods **7**:335–6.

Caporaso JG, **Lauber CL**, **Walters W a**, **Berg-Lyons D**, **Huntley J**, **Fierer N**, **Owens SM**, **Betley J**, **Fraser L**, **Bauer M**, **Gormley N**, **Gilbert J a**, **Smith G**, **Knight R**. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J **6**:1621–1624.

Carro L, **Pujic P**, **Trujillo ME**, **Normand P**. 2013. Micromonospora is a normal occupant of actinorhizal nodules. J Biosci **38**:685–693.

Camacho, M., Santamaria, C., Temprano, F., Rodriguez-Navarro, D.N. and Daza, A. 2001. Co-inoculation with Bacillus sp. CECT 450 improves nodulation in Phaseolus vulgaris L. Canad Journ Microbiol. **47:**1058-1062.

Compant S, **Clement C**, **Sessitsch A**. 2010. Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization. Soil Biol Biochem **42**:669–678.

Crowley, D.E., Wang, Y.C., Reid, C.P.P. and Szaniszlo, P.J. 1991. Mechanisms of iron acquisition from siderophores by microorganisms and plants. Plant and Soil. **130**:179-198.

Dennis PG, **Miller AJ**, **Hirsch PR**. 2010. Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities? FEMS Microbiol Ecol. **72**:313–327.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P. and Andersen, G.L. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. App. And Env. Micro. **72:**5069-5072.

Diagne N, **Diouf D**, **Svistoonoff S**, **Kane A**, **Noba K**, **Franche C**, **Bogusz D**, **Duponnois R**. 2013. Casuarina in Africa : Distribution , role and importance of arbuscular mycorrhizal , ectomycorrhizal fungi and Frankia on plant development. J Environ Manage. **128**:204–209.

Dimkpa, C., Svatoš, A., Merten, D., Büchel, G. and Kothe, E. 2008. Hydroxamate siderophores produced by Streptomyces acidiscabies E13 bind nickel and promote growth in cowpea (Vigna unguiculata L.) under nickel stress. Canad journ of Micro. **54**:163-172.

Dommergues YR. 1997. Contribution of actinorhizal plants to tropical soil productivity and rehabilitation. Soil Biol Biochem. **29**:931–941.

Echbab H, **Prin Y**, **Ducousso M**. 2004. Helper bacteria associated with Casuarina cunninghamiana-Frankia symbiosis : selection of isolates for their effects on plant growth in axenic conditions. Acta Botanica Gallica : Bot Lett. **151**:429–440.

Eddy, S.R., 1998. Profile hidden Markov models. Bioinformatics. **14**:755-763.

Edgar, R.C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. **26:**2460-2461.

El-Lakany, M.H. 1983. A review of breeding drought resistant *Casuarina* for shelterbelt establishment in arid regions with special reference to Egypt. For. Eco. and Manage. **6:**129-137.

El-Tarabily KA, **Nassar AH**, **Hardy GESJ**, **Sivasithamparam K**. 2009. Plant growth promotion and biological control of Pythium aphanidermatum, a pathogen of cucumber, by endophytic actinomycetes. J Appl. Microbiol. **106**:13–26.

Fett, W.F., Osman, S.F. and Dunn, M.F.1987. Auxin production by plant-pathogenic pseudomonads and xanthomonads. App and Env Microbio. **53:**1839-1845.

Fterich A, **Mahdhi M**, **Caviedes MA**, **Pajuelo E**, **Rivas R**, **Rodriguez-Llorente ID**, **Mars M**. 2011. Characterization of root-nodulating bacteria associated to Prosopis farcta growing in the arid regions of Tunisia. Arch Microbiol. **193**:385–397.

Fterich A, **Mahdhi M**, **Lafuente A**, **Pajuelo E**, **Caviedes MA**, **Rodriguez-Llorente ID**, **Mars M**. 2012. Taxonomic and symbiotic diversity of bacteria isolated from nodules of Acacia tortilis subsp. raddiana in arid soils of Tunisia. Can J Microbiol. **58**:738–751.

Gauthier, D., Diem, H.G., Dommergues, Y.R. and Ganry, F. 1985. Assessment of N2 fixation by *Casuarina equisetifolia* inoculated with Frankia ORS021001 using 15 N methods. Soil Biol and Biochem. **17**:375-379.

Ghodhbane-Gtari F, **Essoussi I**, **Chattaoui M**, **Chouaia B**, **Jaouani A**, **Daffonchio D**, **Boudabous A**, **Gtari M**. 2009. Isolation and characterization of non-Frankia actinobacteria from root nodules of Alnus glutinosa, Casuarina glauca and Elaeagnus angustifolia. Symbiosis **50**:51–57.

Ghodhbane-Gtari, F., Hurst, S.G., Oshone, R., Morris, K., Abebe-Akele, F., Thomas, W.K., Ktari, A., Salem, K., Gtari, M. and Tisa, L.S. 2014. Draft genome sequence of Frankia sp. strain BMG5. 23, a salt-tolerant nitrogen-fixing actinobacterium isolated from the root nodules of Casuarina glauca grown in Tunisia. Genome announcements **2:** pp.e00520-14.

Glick BR. 2005. Modulation of plant ethylene levels by the bacterial enzyme ACC deaminase. FEMS Microbiol Lett. **251**:1–7.

Grayston SJ, **Wang S**, **Campbell CD**, **Edwards AC**. 1998. Selective Influence of Plant Species on Microbial Diversity in the Rhizosphere. Soil Biol. And Biochem. **30**:369–378.

Gtari, M., Brusetti, L., Aouani, M.E., Daffonchio, D. and Boudabous, A., 2002. Frankia nodulating Alnus glutinosa and Casuarinaceae in Tunisia. Ann. of microbiol. **52**:145-154.

Gtari M, **Brusetti L**, **Hassen A**, **Mora D**, **Daffonchio D**, **Boudabous A**. 2007. Genetic diversity among Elaeagnus compatible Frankia strains and sympatric-related nitrogen-fixing actinobacteria revealed by nifH sequence analysis. Soil Biol Biochem **39**:372–377.

Gtari M, **Brusetti L**, **Skander G**, **Mora D**, **Boudabous A**, **Daffonchio D**. 2004. Isolation of Elaeagnus-compatible Frankia from soils collected in Tunisia. FEMS Microbiol Lett. **234**:349–355.

Gtari, M., Ghodhbane-Gtari, F., Nouioui, I., Ktari, A., Hezbri, K., Mimouni, W., Sbissi, I., Ayari, A., Yamanaka, T., Normand, P. and Tisa, L.S. 2015. Cultivating the uncultured: growing the recalcitrant cluster-2 Frankia strains. Sci. Rep. **5**

Gtari M, **Dawson JO**. 2011. An overview of actinorhizal plants in Africa. Funct Plant Biol. **38**:653–661.

Gurevich, A., Saveliev, V., Vyahhi, N. and Tesler, G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics **29**:1072-1075.

Knee EM, **Gong FC**, **Gao M**, **Teplitski M**, **Jones a R**, **Foxworthy a**, **Mort a J**, **Bauer WD**. 2001. Root mucilage from pea and its utilization by rhizosphere bacteria as a sole carbon source. Mol Plant Microbe Interact **14**:775–784.

Knowlton, S. and Dawson, J.O. 1983. Effects of Pseudomonas cepacia and cultural factors on the nodulation of Alnus rubra roots by Frankia. Canad Journ Bot. **61**:2877-2882.

Le Houerou, H.N. 2000. Restoration and rehabilitation of arid and semiarid Mediterranean ecosystems in North Africa and West Asia: a review. Arid Soil Res and Rehab. **14:**3-14.

Hartmann, A., Rothballer, M. and Schmid, M. 2008. Lorenz Hiltner, a pioneer in rhizosphere microbial ecology and soil bacteriology research. Plant and Soil. **332**:7-14

Haichar FEZ, **Marol C**, **Berge O**, **Rangel-Castro JI**, **Prosser JI**, **Balesdent J**, **Heulin T**, **Achouak W**. 2008. Plant host habitat and root exudates shape soil bacterial community structure. ISME J **2**:1221–30.

Hirsch AM, **Valdés M**. 2010. Micromonospora: An important microbe for biomedicine and potentially for biocontrol and biofuels. Soil Biol Biochem **42**:536–542.

Huson, D.H., Auch, A.F., Qi, J. and Schuster, S.C. 2007. MEGAN analysis of metagenomic data. Genome Res. **17:** 377-386.

Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W. and Hauser, L.J. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC bioinformatics. **11:**119-130.

Ismet A, **Vikineswary S**, **Paramaswari S**, **Wong WH**, **Ward A**, **Seki T**, **Fiedler HP**, **Goodfellow M**. 2004. Production and chemical characterization of antifungal metabolites from micromonospora sp. M39 isolated from mangrove rhizosphere soil. World J Microbiol Biotechnol **20**:523–528.

Jeong SC, **Ritchie NJ**, **Myrold DD**. 1999. Molecular Phylogenies of Plants and Frankia Support Multiple Origins of Actinorhizal Symbioses **13**:493–503.

Kai M, **Effmert U**, **Berg G**, **Piechulla B**. 2007. Volatiles of bacterial antagonists inhibit mycelial growth of the plant pathogen Rhizoctonia solani. Arch Microbiol **187**:351–360.

Khan AL, **Waqas M**, **Kang SM**, **Al-Harrasi A**, **Hussain J**, **Al-Rawahi A**, **Al-Khiziri S**, **Ullah I**, **Ali L**, **Jung HY**, **Lee IJ**. 2014. Bacterial endophyte Sphingomonas sp. LK11 produces gibberellins and IAA and promotes tomato plant growth. J Microbiol. **52**:689–695.

Kohls SJ, **Baker DD**, **Van Kessel C**, **Dawson JO**. 2003. An assessment of soil enrichment by actinorhizal N2 fixation using 15N values in a chronosequence of deglaciation at Glacier Bay, Alaska. Plant Soil **254**:11–17.

Kuzyakov Y and Weng, W. 2001. Photosynthesis controls of rizosphere respiration and organic mater decomposition . Soil Biol Biochem. **33**:1915–1925.

Lajtha, K. and Schlesinger, W.H.1988. The biogeochemistry of phosphorus cycling and phosphorus availability along a desert soil chronosequence. Ecology. 24-39.

Lammel DR, **Cruz LM**, **Mescolotti D**, **Stürmer SL**, **Cardoso EJBN**. 2015. Woody Mimosa species are nodulated by Burkholderia in ombrophylous forest soils and their symbioses are enhanced by arbuscular mycorrhizal fungi (AMF). Plant Soil. 123–135.

Langille, M.G., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Thurber, R.L.V., Knight, R. and Beiko, R.G. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat. Biotech. **31:**814-821

Langmead, B. and Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Meth. **9:**357-359.

Larcher M, **Muller B**, **Mantelin S**, **Rapior S**, **Cleyet-Marel JC**. 2003. Early modifications of Brassica napus root system architecture induced by a plant growth-promoting Phyllobacterium strain. New Phytol. **160**:119–125.

LeBauer, D.S. and Treseder, K.K. 2008. Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. Ecology **89:**371-379

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics. **25:**2078-2079.

Loper J.E, and Schroth M.N. 1986. Influence of Bacterial Sources of Indole-3-acetic Acid on Root Elongation of Sugar Beet. Phytopathology **74**:386-389

Love, M.I., Huber, W. and Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. **15:**1-21

Lozupone C, **Knight R**. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol **71**:8228–35.

Lozupone, C.A., Hamady, M., Kelley, S.T. and Knight, R. 2007. Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. App Environ microbial. **73:**1576-1585.

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

Mahdhi M, **Nzoué A**, **de Lajudie P**, **Mars M**. 2008. Characterization of root-nodulating bacteria on Retama raetam in arid Tunisian soils. Prog Nat Sci. **18**:43–49.

Mailly, D. and Margolis, H.A. 1992. Forest floor and mineral soil development in *Casuarina equisetifolia* plantations on the coastal sand dunes of Senegal. Forest ecol. and manage. **55**:259-278.

Mallet, P, Roy S. 2014. The Symbiosis between Frankia and Alder Shrubs Results in a Tolerance of the Environmental Stress Associated With Tailings from the Canadian Oil Sands Industry. J Pet Environ Biotechnol **5.**

Manulis, S., Shafrir, H., Epstein, E., Lichter, A. and Barash, I. 1994. Biosynthesis of indole-3-acetic acid via the indole-3-acetamide pathway in Streptomyces spp. Micro. **140:**1045-1050.

Manulis, S., Haviv-Chesner, A., Brandl, M.T., Lindow, S.E. and Barash, I. 1998. Differential involvement of indole-3-acetic acid biosynthetic pathways in pathogenicity and epiphytic fitness of *Erwinia herbicola* pv. gypsophilae. Molec Plant-Microbe Interact. **11:**634-642.

Martínez-Hidalgo P, **Galindo-Villardón P**, **Igual JM**, **Martínez-Molina E**. 2014. *Micromonospora* from nitrogen fixing nodules of alfalfa (Medicago sativa L.). A new promising Plant Probiotic Bacteria. Sci Rep. **4**.

McMurdie, P.J. and Holmes, S. 2014. Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comput Biol. **10:**1003531.

Mendes LW, **Kuramae EE**, **Navarrete A a**, **van Veen J a**, **Tsai SM**. 2014. Taxonomical and functional microbial community selection in soybean rhizosphere. ISME J **8:**1577–1587.

Mirza BS, **Welsh A**, **Rieder JP**, **Paschke MW**, **Hahn D**. 2009. Diversity of *Frankia* in soils from five continents. Syst Appl Microbiol. **32**:558–570.

Moreno-ramirez L, **Gonzalez-mendoza D**, **Gutierrez-miceli F**, **Aviles-marin M**, **Duran CC-**, **Grimaldo-juarez O**. 2014. Phytohormones Production and Phosphate Solubilization Capacities of Acinetobacter sp . Strains Isolation from Mexicali Valley. Jour Pure Appl Microbiol. **8**.

Mtimet A. 2001. Soils of Tunisia **262**:243–268.

Murray, M.G. and Thompson, W.F. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic acids res. **8:**4321-4326

Muthukumar T, **Udaiyan K**. 2010. Growth response and nutrient utilization of Casuarina equisetifolia seedlings inoculated with bioinoculants under tropical nursery conditions. New For **40**:101–118.

Neilands, J.B. 1995. Siderophores: structure and function of microbial iron transport compounds. Journ of Biol Chem. **27:**26723-26726.

Ng, B.H. 1987. The effects of salinity on growth, nodulation and nitrogen fixation of *Casuarina equisetifolia*. Plant and Soil. **103**:123-125.

Normand P, **Lapierre P**, **Tisa LS**, **Gogarten JP**, **Alloisio N**, **Bagnarol E**, **Bassi C a.**, **Berry AM**, **Bickhart DM**, **Choisne N**, **Couloux A**, **Cournoyer B**, **Cruveiller S**, **Daubin V**, **Demange N**, **Francino MP**, **Goltsman E**, **Huang Y**, **Kopp OR**, **Labarre L**, **Lapidus A**, **Lavire C**, **Marechal J**, **Martinez M**, **Mastronunzio JE**, **Mullin BC**, **Niemann J**, **Pujic P**, **Rawnsley T**, **Rouy Z**, **Schenowitz C**, **Sellstedt A**, **Tavares F**, **Tomkins JP**, **Vallenet D**, **Valverde C**, **Wall LG**, **Wang Y**, **Medigue C**, **Benson DR**. 2007. Genome characteristics of facultatively symbiotic Frankia sp. strains reflect host range and host plant biogeography. Genome Res **17**:7–15.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M.H.H., Oksanen, M.J. and Suggests, M.A.S.S., 2007. The vegan package. Community ecology package, 10.

Pace, N.R. 1997. A molecular view of microbial diversity and the biosphere. Science **276:**734-740.

Patten, C.L. and Glick, B.R. 1996. Bacterial biosynthesis of indole-3-acetic acid. Canadian Journal of Microbiology. **42**:207-220.

Patten CL, **Glick BR**. 2002. Role of *Pseudomonas putida* indole acetic acid in development of the host plant root system. Appl Environ Microbiol **68**:3795–801.

Paulson, J.N., Stine, O.C., Bravo, H.C. and Pop, M. 2013. Differential abundance analysis for microbial marker-gene surveys. Nat. Meth. **10:**1200-1202.

Perrine-Walker F, **Doumas P**, **Lucas M**, **Vaissayre V**, **Beauchemin NJ**, **Band LR**, **Chopard J**, **Crabos A**, **Conejero G**, **Péret B**, **King JR**, **Verdeil J-L**, **Hocher V**, **Franche C**, **Bennett MJ**, **Tisa LS**, **Laplaze L**. 2010. Auxin carriers localization drives auxin accumulation in plant cells infected by Frankia in Casuarina glauca actinorhizal nodules. Plant Physiol. **154**:1372–1380.

Perinet, P., Brouillette, J.G., Fortin, J.A. and Lalonde, M. 1985. Large scale inoculation of actinorhizal plants with *Frankia*. Plant and Soil. **87**:175-183.

Quinlan, A.R. and Hall, I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. **26:**841-842.

Ramadoss D, **Lakkineni VK**, **Bose P**, **Ali S**, **Annapurna K**. 2013. Mitigation of salt stress in wheat seedlings by halotolerant bacteria isolated from saline habitats. Springerplus **2**:1-7.

Raupach GS, **Kloepper JW**. 1998. Mixtures of plant growth-promoting rhizobacteria enhance biological control of multiple cucumber pathogens. Phytopathology **88**:1158–1164.

Reddell P, **Bowen G**, **Robson A**. 1986. Nodulation of Casuarinaceae in Relation to Host Species and Soil Properties. Aust J Bot. **34**:435.

Rehan M, **Kluge M**, **Fränzle S**, **Kellner H**, **Ullrich R**, **Hofrichter M**. 2014. Degradation of atrazine by Frankia alni ACN14a: gene regulation, dealkylation, and dechlorination. Appl Microbiol Biotechnol. **98**:6125–35.

Rideout, J.R., He, Y., Navas-Molina, J.A., Walters, W.A., Ursell, L.K., Gibbons, S.M., Chase, J., McDonald, D., Gonzalez, A., Robbins-Pianka, A. and Clemente, J.C. 2014. Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. PeerJ 24:e454

Rodríguez, H. and Fraga, R.1999. Phosphate solubilizing bacteria and their role in plant growth promotion. Biotech Advances **17**:319-339.

Rojas A, **Holguin G**, **Glick BR**, **Bashan Y**. 2001. Synergism between Phyllobacterium sp. (N2-fixer) and Bacillus licheniformis (P-solubilizer), both from a semiarid mangrove rhizosphere. FEMS Microbiol Ecol. **35**:181–187.

Roggy JC, **Moiroud A**, **Lensi R**, **Domenach AM**. 2004. Estimating N transfers between N2-fixing actinorhizal species and the non-N2-fixing *Prunus avium* under partially controlled conditions. Biol Fertil Soils **39**:312–319.

Römheld, V. and Marschner, H., 1990. Genotypical differences among graminaceous species in release of phytosiderophores and uptake of iron phytosiderophores. Plant and Soil. **123:**147-153.

Rosenblueth M, **Martínez-Romero E**. 2006. Bacterial endophytes and their interactions with hosts. Mol Plant Microbe Interact **19**:827–837.

Rouvier C, **Prin Y**, **Reddell P**, **Normand P**, **Simonet P**. 1996. Genetic Diversity among Frankia Strains Nodulating Members of the Family Casuarinaceae in Australia Revealed by PCR and Restriction Fragment Length Polymorphism Analysis with Crushed Root Nodules. App and Env Micrbiol. **62**:979–985.

Samant S, **Dawson JO**, **Hahn D**. 2015. Growth responses of indigenous Frankia populations to edaphic factors in actinorhizal rhizospheres. Syst Appl Microbiol 3–7.

Samant S, **Huo T**, **Dawson JO**, **Hahn D**. 2015. Abundance and Relative Distribution of Frankia Host Infection Groups Under Actinorhizal Alnus glutinosa and Non-actinorhizal Betula nigra Trees. Microb Ecol.

Sanginga N, **Danso SKA**, **Bowen GD**. 1989. Nodulation and growth response of Allocasuarina and Casuarina species to phosphorus fertilization. Plant Soil **118**: 125-132

Sayed WF. 2011. Improving Casuarina growth and symbiosis with Frankia under different soil and environmental conditions. Foili Micobiol. **56:**1–9.

Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics btu153.

Schwartz A, **Ortiz I**, **Maymon M**, **Herbold C**, **Fujishige N**, **Vijanderan J**, **Villella W**, **Hanamoto K**, **Diener A**, **Sanders E**, **DeMason D**, **Hirsch A**. 2013. Bacillus simplex—A Little Known PGPB with Anti-Fungal Activity—Alters Pea Legume Root Architecture and Nodule Morphology When Coinoculated with Rhizobium leguminosarum bv. viciae. Agronomy **3**:595–620.

Schwintzer, C.R. Production, decomposition and nitrogen dynamics of Myrica gale litter. 1984. Plant and Soil. **78:**245-258

Sharma A, **Johri BN**. 2003. Growth promoting influence of siderophore-producing Pseudomonas strains GRP3A and PRS9 in maize (Zea mays L.) under iron limiting conditions. Microbiol Res. **158**:243–248.

Simonet P, **Navarro E**, **Reddell P**, **Zimpfer J**, **Bardin Â**, **Combarro P**, **Cedex V**. 1999. Co-evolution between Frankia populations and host plants in the family Casuarinaceae and consequent patterns of global dispersal. Environ Microbiol. **1**:525–533.

Sikorski J, **Nevo E**. 2005. Adaptation and incipient sympatric speciation of Bacillus simplex under microclimatic contrast at "Evolution Canyons" I and II, Israel. Proc Natl Acad Sci U S A **102**:15924–9.

Solans M. 2007. Discaria trinervis - Frankia symbiosis promotion by saprophytic actinomycetes. J Basic Microbiol **47**:243–250.

Solans M, **Vobis G**, **Wall LG**. 2009. Saprophytic actinomycetes promote nodulation in Medicago sativa-Sinorhizobium meliloti symbiosis in the presence of high N. J Plant Growth Regul **28**:106–114.

Solans, M., Vobis, G., Cassán, F., Luna, V. and Wall, L.G. 2011. Production of phytohormones by root-associated saprophytic actinomycetes isolated from the actinorhizal plant *Ochetophila trinervis*. World Journ of Microbiol and Biotech. **27:**2195-2202.

Solans M, **Ruiz OA**, **Wall LG**. 2015. Effect of actinobacteria on Lotus tenuis – Mesorhizobium loti symbiosis: preliminary study. Symbiosis **65**:33–37.

Soltis DE, **Soltis PS**, **Morgan DR**, **Swensen SM**, **Mullin BC**, **Dowd JM**, **Martin PG**. 1995. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proc Natl Acad Sci. **92**:2647–2651.

Somers, E., Vanderleyden, J. and Srinivasan, M. 2008. Rhizosphere bacterial signalling: a love parade beneath our feet. Critical reviews in microbiology. **30**

Spain AM, **Krumholz LR**, **Elshahed MS**. 2009. Abundance, composition, diversity and novelty of soil Proteobacteria. ISME J **3**:992–1000.

Swensen, S.M. 1996. The evolution of actinorhizal symbioses: evidence for multiple origins of the symbiotic association. American Journal of Botany. **83**:1503-1512.

Swensen SM and Mullin BC. 1997. Phylogenetic relationships among actinorhizal plants. The impact of molecular systematics and implications for the evolution of actinorhizal symbioses. Physiol Plant **99**:565–573.

Tan Z, **Hurek T**, **Gyaneshwar P**, **Ladha JK**, **Reinhold-Hurek B**. 2001. Novel endophytes of rice form a taxonomically distinct subgroup of Serratia marcescens. Syst Appl Microbiol **24**:245–251.

Teale WD, **Paponov I a.**, **Palme K**. 2006. Auxin in action: signalling, transport and the control of plant growth and development. Nat Rev Mol Cell Biol **7**:847–859.

Tobita H, **Uemura A**, **Kitao M**, **Kitaoka S**, **Utsugi H**. 2010. Interactive effects of elevated CO2, phosphorus deficiency, and soil drought on nodulation and nitrogenase activity in alnus hirsuta and alnus maximowiczii. Symbiosis **50**:59–69.

Trujillo ME, **Kroppenstedt RM**, **Schumann P**, **Carro L**, **Martínez-Molina E**. 2006. Micromonospora coriariae sp. nov., isolated from root nodules of Coriaria myrtifolia. Int J Syst Evol Microbiol **56**:2381–2385.

Trujillo ME, **Kroppenstedt RM**, **Fernández-Molinero C**, **Schumann P**, **Martínez-Molina E**. 2007. Micromonospora lupini sp. nov. and Micromonospora saelicesensis sp. nov., isolated from root nodules of Lupinus angustifolius. Int J Syst Evol Microbiol **57**:2799–2804.

Trujillo ME, **Riesco R**, **Benito P**, **Carro L**. 2015. Endophytic Actinobacteria and the Interaction of Micromonospora and Nitrogen Fixing Plants. Front Microbiol **6**:1–15.

Tunesi S, **Poggi V**, **Gessa C**. 1999. Phosphate adsorption and precipitation in calcareous soils: the role of calcium ions in solution and carbonate minerals. Nutr Cycl Agroecosystems **53**:219–227.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S. and Banfield, J.F. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature **428**:37-43.

Valverde, C. and Wall, L.G. 1999. Time course of nodule development in the *Discaria trinervis* Rhamnaceae–Frankia symbiosis. New Phytologist. **141**:345-354.

Vančura, V. and Hovadik, A. 1965. Root exudates of plants: II. Composition of root exudates of some vegetables. Plant and Soil. 21-32.

Walker, T.S., Bais, H.P., Grotewold, E. and Vivanco, J.M. 2003. Root exudation and rhizosphere biology. Plant physiology. **132:**44-51.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R., Lee, S.Y., Fischbach, M.A., Müller, R., Wohlleben, W. and Breitling, R. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic acids res. 43:237-243

Weixin, C., Coleman, D.C., Carroll, C.R. and Hoffman, C.A. 1993. In situ measurement of root respiration and soluble C concentrations in the rhizosphere. Soil Biol and Biomchem. **25:**1189-1196.

Wheeler, C.T., Hooker, J.E., Crowe, A. and Berrie, A.M.M. 1986. The improvement and utilization in forestry of nitrogen fixation by actinorhizal plants with special reference to Alnus in Scotland. Plant and Soil. 90:393-406.

Woolfrey AR, **Ladd PG**. 2001. Habitat preference and reproductive traits of a major Australian riparian tree species *Casuarina cunninghamiana*. Aust J Bot **49**:705–715.

Yang, Y. 1995. The effect of phosphorus on nodule formation and function in the *Casuarina-Frankia* symbiosis. Plant Soil **176**:161–169.

Zakhia F, **Jeder H**, **Willems A**, **Gillis M**, **Dreyfus B**, **De Lajudie P**. 2006. Diverse bacteria associated with root nodules of spontaneous legumes in Tunisia and first report for nifH-like gene within the genera Microbacterium and Starkeya. Microb Ecol **51**:375–393.

Zimpfer, J.F., Igual, J.M., McCarty, B., Smyth, C. and Dawson, J.O. 2004. Casuarina cunninghamiana tissue extracts stimulate the growth of Frankia and differentially alter the growth of other soil microorganisms. Journ Chem Ecol **30**:439-452.

Zhang Q, **Zhou L**, **Najar FZ**, **Roe B a**, **Fletcher J**. 2005. Genomic Comparison of Plant Pathogenic and Nonpathogenic Serratia marcescens Strains by Suppressive Subtractive Hybridization. Microbiology **71**:7716–7723.

Zhong, C., Mansour, S., Nambiar-Veetil, M., Bogusz, D. and Franche, C. 2013. *Casuarina glauca*: a model tree for basic research in actinorhizal symbiosis. Journal of bioscie. **38**:815-823.