Identification and Characterization of Bacterial Genes Involved in the *Medicago-Sinorhizobium* symbiosis

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Matthew S Nelson

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Michael J Sadowsky
Advisor

September 2017

Acknowledgements

First and foremost, I am extremely grateful for all of the help and support my advisor, Mike Sadowsky, has provided me over the years. Mike has been an amazing mentor and has always been there to provide advice and guidance.

I would like to thank my committee members, Debby Samac, Nevin Young and Christine Salomon. Thank you for taking the time to help guide me and my research.

Also, I would like to highlight Chanlan Chun. During my first few years, Chanlan was critical to helping me develop as a scientist. Chanlan was extremely giving of her time, and always willing to provide support in whatever way I needed it.

Numerous people have helped me during my graduate career and I would like to highlight their contributions. I would like to acknowledge the help of all the Sadowsky Lab members. Chris Staley helped edit the Introduction and Chapter 1. Chanlan Chun, helped in experimental design in Chapter 1. Mindy Dornbusch helped with qRT-PCR in Chapter 1 and 2, and performed plant transformation of *Medicago sativa*. Sue Miller helped in qRT-PCR in Chapter 1 and 2. Colby Starker provided advice and the estrogen inducible vector in Chapter 2. Shaun Curtin provide advice on *Medicago truncatula* transformation in Chapter 2. Roxanne Denny provided support in transformation, *Medicago* growth, and general assistance in Chapter 1 and 2. Diana Trujillo provided help in *M. truncatula* transformation and growth in Chapter 2. Steve Gantt provide the yeast two-hybrid libraries and vectors used in Chapter 2. Erik Solhaug provided the BiFC vectors in Chapter 2. Gerit Bethke provided vectors and *Nicotiana benthamiana* seeds used in Chapter 2. Peter Tiffin helped provide insight, guidance, advice and critically

editing in Chapter 3. I would like to specifically highlight Joseph Guhlin's contribution to Chapter 3. He aligned the Illumina reads to the PacBio sequences, generated the Circos figure and did all the work pertaining to repeats and transposable elements. Brendan Epstein generated the phylogenetic tree and did the Mantel test in Chapter 3.

From a personal perspective I would like to thank Mohamed Yakub and Matthew Hamilton for their support and advice.

Dedication

This work is dedicated to my family, for all their love, support and sacrifice.

Abstract

Biological nitrogen fixation is both ecologically and agriculturally important. Rhizobia are soil bacteria which form a symbiosis with legumes, fixing nitrogen in exchange for carbon. The goal of this research was to identify and determine the function of genes which are not required for the establishment of symbiosis, but have an effect on the efficacy of the symbiotic interaction. In Chapter 1, I demonstrate the type IV secretion system has an effect on symbiosis. Bacterial proteins were tested for their ability to translocate through the type IV secretion system into the host cell. One protein, TfeA, was transported into the host cell. Deleting *tfeA* resulted in decrease nodule number and strain competitiveness for nodule occupancy. TfeA-like proteins were also identified in 12 different rhizobia species. In Chapter 2, I showed TfeA binds to *Medicago truncatula* ARF2 *in vivo* and *Glycine max* ARF27. Constitutively expressed *tfeA* was unable to be recovered in *Medicago sativa* (alfalfa) or *M. truncatula.* Inducible *tfeA* was successfully transformed in *M. truncatula*, but expression is inconsistent. TfeA binding to ARF2 *in planta* offers additional insights into the role of auxin during nodule formation. In Chapter 3, I use PacBio sequencing technology to assemble 10 strains of *Sinorhizobium*. Analysis of 16 *Sinorhizobium* complete genomes showed that most strains have small accessory plasmids. These plasmids carry genes which are not required for symbiosis, but have an effect on the symbiotic process. Accessory plasmids combine with pSymA to allow for the gain and loss of symbiosis genes. The type IV secretion system and *tfeA* was found on pSymA or an accessory plasmid depending on the strain. Overall, this research identified important bacterial genes which modulate symbiosis.

Table of Contents

List of Tables

List of Figures

Introduction

Chapter 1

Chapter 2

Chapter 3

Introduction

Secretion systems and signal exchange between nitrogen-fixing rhizobia and legumes

A version of this introduction was published in Frontiers in Plant Science

Nelson, M. S., and Sadowsky, M. J. 2015. Secretion systems and signal exchange between nitrogen-fixing rhizobia and legumes. Front. Plant Sci. 6:491

Summary

The formation of symbiotic nitrogen-fixing nodules on the roots and/or stem of leguminous plants involves a complex signal exchange between both partners. Since many microorganisms are present in the soil, legumes and rhizobia must recognize and initiate communication with each other to establish symbioses. This results in the formation of nodules, where rhizobia exchange fixed nitrogen for carbon from the legume. Symbiotic relationships can become non-beneficial if one partner ceases to provide support to the other. As a result, complex signal exchange mechanisms have evolved to ensure continued, beneficial symbioses. Proper recognition and signal exchange is also the basis for host specificity. Additional nodule formation seems to always provide a fitness benefit to rhizobia, but does not always provide a fitness benefit to legumes. Therefore, legumes have evolved a mechanism to regulate the number of nodules that are formed, called autoregulation of nodulation. Sequencing of many different rhizobia have revealed the presence of several secretion systems - and the type III, type IV and type VI secretion systems are known to be used by pathogens to transport effector proteins. These secretion systems are also known to have an effect on host specificity and are a determinant of overall nodule number on legumes. This review focuses on signal exchange between rhizobia and legumes, particularly focusing on the role of secretion systems involved in nodule formation and host specificity.

Introduction

Plants interact with many different types of microbes, and these associations can be pathogenic, mutualistic or commensal in nature. The type of relationship between a specific microbe and plant can vary based on external factors, such as changes in environment, or due to intrinsic factors of both organisms. Both pathogenic and mutualistic interactions are dependent on communication between host and microbe and are primarily based on signal exchange (Tseng et al. 2009). The symbiotic relationship between rhizobia and legumes has long been a focus of study because of the nitrogen fixation that occurs during the symbiosis. This symbiosis requires the rhizobia to be in close physical proximity to the legume to allow for exchange of nutrients. Nitrogen is essential for all agricultural crops and legumes can access nitrogen from the atmosphere through symbiosis with rhizobia. Signal exchange between rhizobia and legumes has been studied as a potential process for regulating symbiosis in non-legume plants and as a mechanism by which to increase nitrogen fixation in legumes.

The symbiosis between legumes and rhizobia has evolved to incorporate many different levels of signal exchange, from initial contact to senescence. Two primary reasons for this signal exchange are to distinguish between symbionts and pathogens and to ensure mutualism through the exchange of carbon and fixed nitrogen. The line between symbiont and pathogen is not always clear, as both partners can have a fitness benefit to alter the relationship to their advantage. Symbiotic associations may shift from mutually beneficial to pathogenic or vice versa, such as in the case of the plant pathogen *Agrobacterium*, having a common ancestral history with rhizobia. It has been suggested that rhizobia can be viewed as refined pathogens (Deakin and Broughton 2009). The

symbiotic relationship between rhizobia and legumes can easily turn pathogenic if the plants loses the ability to regulate the total number of nodules formed or the rhizobia form nodules that do not fix nitrogen - with the plant experiencing decreased fitness by providing too much carbon to the rhizobia (Kiers et al. 2003; Herridge and Rose 2000). Co-evolution between rhizobia and legumes is more complex because rhizobia selection can oscillate between pathogen and symbiont mode.

The evolutionary arms race between pathogens and plants has long been understood (Jones and Dangl 2006). Pathogens develop new strategies for creating infections, such as evolving secretion systems to alter the host cell. In response, plants develop new strategies for detecting pathogens, such as micro-associated molecular patterns (MAMPs), and R (resistance) genes (Dodds and Rathjen 2010). Sequencing of various rhizobia strains has shown the presence of secretion systems similar to those used by pathogens to transfer proteins into the hosts' cytosol. These secretion systems include the Type III (T3SS), Type IV (T4SS), and Type VI Secretion Systems (T6SS) (Fauvart and Michiels 2008). The evolutionary presence of these secretion systems suggests that while rhizobia and legumes co-evolved a system allowing establishment and maintenance of a symbiosis, a relationship similar to a pathogen/plant interaction also co-evolved. This review focuses on legume-rhizobia signal exchange that occurs during nodule formation, plant mechanisms for limiting nodule number, and potential strategies used by rhizobia to overcome the plants ability to limit nodule number using the T3SS, T4SS or T6SS.

Signal exchange during nodule formation

Rhizobia are free-living, soil saprophytes, prior to symbiosis with plants in the family Leguminosae. Rhizobia, once inoculated into soil, can persist at low levels in the absence of a suitable host (Howieson 1995). The plant initiates symbiosis by secreting flavonoids, which are detected by the rhizobia. Flavonoids vary by plant species and are only recognized by certain, yet specific, rhizobia species, offering the first level of symbiotic specificity (Hassan and Mathesius 2012). The flavonoids diffuse across the membrane of the rhizobia and induce synthesis of the NodD protein to activate transcription of other genes involved in nodulation including nod factor (NF) production (Wang et al. 2012). NFs are a primary signal molecule produced by bacteria and detected by the plant to induce nodule organogenesis. Structurally NFs are lipochitooligosaccharides (LCOs) with a chitin oligomer backbone (Oldroyd and Downie 2008). The *nodABC* genes encode for the proteins required to make the core NF structure and are conserved across all rhizobia species, except two *Aeschynomene*-infective species (Perret et al. 2000; Giraud et al. 2007). The NF core is then modified by species-specific proteins resulting in various substitutions on both the reducing and non-reducing end, including glycosylation and sulfation (Long 1996). These substitutions are specific for each host legume and offer another level of symbiotic specificity (Dénarié et al. 1996; Long 1996).

Many surface polysaccharides are also involved in symbiosis specificity including lipopolysaccharides (LPSs), extracellular polysaccharides (EPSs), and capsular polysaccharides (KPSs) (Deakin and Broughton 2009). The specific structure of LCOs is known to be important for recognition by host nod factor receptors (NFRs), which are

receptor kinases containing lysin motifs (LysM) (Radutoiu et al. 2007). Leucine rich repeat receptor-like kinases (LRR-RLKs) are then involved in signaling pathway in nod factor detection and signal transduction (Endre et al. 2002).

Root hair curling and crack entry are the two infection mechanisms used by rhizobia. Crack entry involves rhizobia entering through cracks at the lateral root bases or stems (Goormachtig et al. 2004). Root hair curling involves recognition of NFs and results in both calcium spiking and the curling of the root hair (Esseling et al. 2003). This is thought to involve a change in the plant cells' polarity, resulting in a new growing direction of the root hair tip (Gage 2004). This results in the root tip growing inward, forming an infection thread (Fournier et al. 2008). The continued growth of the infection thread is dependent on NF specificity as well as extracellular rhizobia polysaccharides (EPS) (Jones et al. 2007a). Both the epidermis and the cortex recognize NFs, the epidermis regulates rhizobia infection, whereas the root cortex is responsible for nodule formation. (Oldroyd and Downie 2008). Cortical cells develop into a nodule primordium. When the infection thread reaches the nodule primordium, the rhizobia enter into the inner cells and become encapsulated within the plant membrane (Oldroyd and Downie 2008).

There are two main types of nodules, indeterminate and determinate, and this is determined by the legumes. For indeterminate nodules, cell division begins in the inner cortex (Ferguson et al. 2010). Indeterminate nodules maintain a persistent meristem and form distinct zones, from rhizobia invasion, active nitrogen fixation and senescence (Udvardi and Poole 2013). These zones contain rhizobia in various developmental states with the proximal zones having a low generation rate (Mergaert et al. 2006). Legumes

belonging to the inverted repeat-lacking clade induce indeterminate nodule formation through secretion of cysteine-rich peptides, which induce membrane permeabilization, endoreduplication, and loss of independent viability (Van de Velde et al. 2010; Oldroyd et al. 2011; Mergaert et al. 2006). In contrast, cell division begins in the outer cortex for determinate nodules (Ferguson et al. 2010). Determinate nodules do not have a persistent meristem and form a homogenous group of rhizobia with high viability (Saeki 2011). In mature nodules, plants exchange small carbon molecules for ammonia with the rhizobia. Another important aspect of symbiosis regulation is amino acid exchange and cycling between the plant and the rhizobia. During symbiosis some plants secrete branched chained amino acids and in return the rhizobia secrete aspartate and, in some cases, alanine. This exchange allows rhizobia to shut down ammonium assimilation and allows the plant to incorporate ammonium into aspartate to produce asparagine (Prell et al. 2009; Lodwig et al. 2006).

After many weeks of plant growth, nodules begin to senescence, with a maximum lifespan well short of that of the host plant (Puppo et al. 2005). Light and water stress, defoliation or addition of nitrate can initiate premature nodule senescence (González et al. 1998; Matamoros et al. 1999; Hernández-Jiménez et al. 2002). This suggest that the plant controls the duration of the symbiosis by being able to induce nodule senescence. These external factors are thought to lead to an increase in *Medicago truncatula* oxygen species, which initiates senescence (Puppo et al. 2005). After nodule senescence, the host plant initiates plant cell death and some rhizobia survive this process and return to a saprophytic state in the soil (Hernández-Jiménez et al. 2002).

Plant signaling limits nodule number

The symbiotic relationship between rhizobia and legumes has the potential to become pathogenic if the plant loses the ability to regulate the total number of nodules or perceives the rhizobia as a pathogen. Rhizobia will generally initiate nodule formation because a symbiotic relationship always has a fitness benefit. However, if the plant forms too many nodules then there is a negative effect on vegetative growth and yield (Takahashi et al. 2003; Matsunami et al. 2004; Herridge and Rose 2000). Legumes use a process called autoregulation of nodulation (AON) to control nodule number by preventing new nodule formation (Mortier et al. 2012). The AON is thought to involve a root-derived signal being transported to the shoot, which induces a shoot-derived signal to be transported to the root - this inhibits nodule formation (Suzaki et al. 2015).

After nodule formation, the plant cell begins to produce CLV3/ESR-related (CLE) peptides. These peptides are thought to be the signal molecule transported from the roots to the shoot (Reid et al. 2011a, 2013). The CLE-RS2 is a post-translationally arabinosylated glycopeptide derived from the CLE domain, and is sufficient to inhibit nodule formation (Okamoto et al. 2013). The CLE-peptides are then recognized by a leucine-rich-repeat receptor-like kinases (LRR-RLKs)(Sasaki et al. 2014; Krusell et al. 2002; Nishimura et al. 2002). The CLE-peptides most likely act as the root-derived signal involved in AON. These receptors then cause a signal cascade which results in cytokinins being transported from the shoot to the root, which could act as the shoot-derived signal to suppress nodule formation (Sasaki et al. 2014). In the *Lotus japonicas tml* mutant, shoot-applied cytokinin does not suppress nodule formation (Sasaki et al. 2014). This

implies that TML acts downstream of cytokinins, and may act directly in the root cells to suppress nodulation. TML encodes a Klech repeat-containing F-box protein and has been hypothesized to target a protein for degradation which has a positive role in nodule formation (Takahara et al. 2013; Suzaki et al. 2015).

AON signaling is a complex process involving numerous steps, some of which are still unknown. Disruption of AON at many different steps has been shown to results in a hyper-nodulation phenotype. This suggests that the AON signaling process could be potential targets for rhizobia to disrupt, in order to increase nodule formation. Inhibition of AON, could result in the symbiotic relationship between rhizobia and legumes becoming a pathogenic(Herridge and Rose 2000).

Bacteria secretion systems

Bacteria use a wide variety of secretion systems to export proteins and other compounds across their membranes and cell walls. Interaction with the external environment is vital to bacterial survival, and many different transmembrane channels have evolved independently to fulfill this need (Wooldridge 2009). There have been reports of up to many different secretion systems, but only the first seven have been significantly investigated (Tseng et al. 2009). These secretion systems have evolved independently, each containing a different set of core proteins. Each secretion system itself diverged into unique subfamilies based on different functions. The T1SS, T2SS and T5SSs are thought to simply transport proteins and compounds outside of the cell. The T3SS, T4SS and T6SSs contain subfamilies with the ability to transport effector proteins into the cytosol of eukaryotic cells (Wooldridge 2009). This is important because it

allows for the direct communication with, and modification of, the eukaryotic cytosol. These three secretion systems are well understood for their role in pathogenesis as key factors in virulence and, in some cases, symbiosis.

Rhizobia secretion systems

As discussed above, rhizobia enter into unique symbioses with eukaryotic cells, through the formation of relationship with legumes. Sequencing of rhizobia strains has shown they typically contain multiple secretion systems. However, the presence of these systems in the bacterial genome does not mean they have a role in symbiosis. Rhizobia surface polysaccharides (LPS) have been known to suppress plant immune responses, but the T3SS and T4SS have also been speculated to have a role in suppressing the plant immune system (Masson-Boivin et al. 2009).

The T3SS and T4SS are both sub-divided into seven families based on function and protein homology (Wooldridge 2009; Sugawara et al. 2013). The T3SS, T4SS and T6SSs have been identified throughout various rhizobia genera and sequence homology shows similarity between known secretion systems used by bacterial pathogens. Specifically, sequence analysis of *Sinorhizobium* has shown they can contain either the T3SS, T4SS or the T6SS, but typically only have one involved in symbiosis per strain (Sugawara et al. 2013). The T3SS, T4SS and T6SS have all been shown to be involved in symbiosis and translocate effector proteins during symbiosis. These effector proteins could potentially have a function either promoting nodule formation, disrupting AON, or suppressing the plant's immune response during invasion. In plant pathogens, the T3SS effectors have been shown to target and suppress the plant immune response (Macho and Zipfel 2015). Deletion of a specific sub-family of the T3SS or the T4SS has been shown

to reduce nodule number and affect host range specificity (Sugawara et al. 2013; Tampakaki 2014). However, their role in symbiosis is still not very well understood.

Type III secretion system (T3SS)

The T3SS is a structure composed of 20-27 different proteins, and this transporter is responsible for secretion of type III effector proteins (T3Es) (Tampakaki 2014; Ghosh 2004). Approximately 50% of these proteins are conserved in most T3SSs (Ghosh 2004). These proteins are generally found clustered in a 22 – 50 kb pathogenicity island (Tampakaki 2014). The T3SS complex spans the bacterial inner and outer membrane as well as the hosts' membranes and allows protein transport into the host. Regions flanking the pathogenicity island can contain genes that encode for effector proteins, but most effector genes are scattered throughout the genome (Lindeberg et al. 2008).

Many different variations of T3SS, with varying functions, are found throughout the kingdom of bacteria. In the literature, the T3SS is first grouped by species, and then grouped by homology. The genes encoding the rhizobia T3SSs are called rhc (rhizobium conserved). The rhc are further subdivided into four families based on phylogenetic analyses, Rhc-1 to Rhc-4 (Gazi et al. 2012). Of these four families, only Rhc-I has been showed to be involved in symbiosis (Tampakaki 2014). The functions of the other families are still unknown. The T3SS is among the best studied secretion systems in rhizobia due to the wide species distribution of Rhc-1 and its role in symbiosis.

T3SS – Rhc-I effect on symbiosis

Early studies of the T3SS – Rhc-1 focused on knocking out the entire system through deletions or disruption of core genes. A diverse range of rhizobia species are known to contain a functional T3SS – Rhc-1 and are listed in Table 0.1. The influence of

T3SSs on nodulation can vary from positive, in which nodulation is increased, to negative, in which nodulation is reduced. In *S. fredii* strain NGR234, the T3SS has both a positive and negative affect on multiple different legume species, but may also have a neutral phenotype, where nodulation is not affected, for example on *Vigna unguiculatal* (Viprey et al. 1998; Skorpil et al. 2005; Kambara et al. 2009). Similarly, rhizobia with the T3SS – Rch-1 show host-dependent phenotypes in regard to nodulation efficiency. This could explain why the T3SS – Rch-1 is found in many genera of rhizobia, but is not ubiquitous at the strain level.

The horizontal transfer of the T3SS could be an important evolutionary driver towards symbiosis or pathogenesis between bacteria and plants. The pathogen *Ralstonia solanacearum* was shown to be able to nodulate *Mimosa pudica* when the T3SS was deleted and the symbiotic plasmid of *Cupriavidus taiwanensis* was added (Marchetti et al. 2010). However, deleting the T3SS effector protein GALA7 prevented pathogenic infection of *Medicago truncatula* (Angot et al. 2006). In addition, *C. taiwanensis* was able to nodulate *Leucaena leucocephala* when the T3SS in *C. taiwanensis* was deleted (Saad et al. 2012). These examples show the impact the T3SS can have on the ability of bacteria to form nodules on legumes.

Regulation of the T3SS – Rhc-1

Expression of the T3SS is induced by plant flavonoid recognition through production of the transcriptional activator TtsI (Viprey et al. 1998; Kobayashi et al. 2004; Krause et al. 2002). TtsI initiates transcription of the T3SS genes and effector proteins by binding to specific *cis*-elements, known as *tts* boxes (Wassem et al. 2008). The number and location of *tts* boxes varies between species and *Bradyrhizobium japonicum*

USDA110 is known to have 52 different *tts* boxes. Proteins secreted by the T3SS are found downstream of *tts* boxes.

There is not a consensus motif for proteins secreted through the T3SS. However, the signal sequence is typically found in the first ~15 amino acids, on the N-terminus, of translocated proteins (Ghosh 2004). In addition not all gene transcription activated by *tts* boxes, are effector proteins translocated through the T3SS; some can have other roles in symbiosis such as the production of rhamnose-rich polysaccharides (Marie et al. 2004). These rhamnose-rich polysaccharides were shown to be surface LPSs, important in nodule formation, independent of the T3SS (Broughton et al. 2006). This suggests an interesting link between secretion systems and surface polysaccharides involved in nodule formation specificity.

Proteins secreted by the T3SS – Rhc-1

Early studies to identify proteins secreted through the T3SS focused on using flavonoids to induce expression in culture and compared the external proteins to those found in a T3SS mutant. However, these experiments did not show translocation into the host cytosol. This led to uncertainty as to whether an identified protein was an effector protein, acting inside the plant cell. A new, high-throughput technique was used to properly identify proteins that translocate through the T3SS as well as to identify effector proteins (Kimbrel et al. 2013). However, this technique did not test for rhizobia translocation into legumes, but rather the proxy of translocation through *P. syringae* pv. *tomato* DC3000 into *Arabidopsis* Col-O. The T3E candidates are fused to Δ79AvrRpt2, which induces a hypersensitive response in *Arabidopsis*. Using this technique on three different strains of *S. fredii* and *B. japonicum*, between 13 – 36 T3Es per strain were

identified (Kimbrel et al. 2013). The T3Es can vary between species and strains, but members of the same species tend to use very similar effector proteins.

Proteins secreted by the T3SS can be separated into two categories - pilus forming and effectors. Proteins involved in pilus formation are secreted through the channel to assist in forming a channel through the plants cell wall or plasma membrane. NopA, NopB and NopX are thought to be involved in the terminal formation of the T3SS, forming a pilus that penetrates the plant's cell wall and plasma membrane (Lorio et al. 2004; Saad et al. 2005; Deakin et al. 2005; Saad et al. 2008). The other secreted proteins are thought to be effector proteins, but few of these proteins have a predicted function *in planta* (Table 0.2).

Despite having a known effect on symbiosis, none of these effector proteins has been expressed in legumes. Only the effectors NopL, NopT, and NopM have all been expressed in eukaryotic cells. NopL was first shown to be phosphorylated by plant kinases (Bartsev et al. 2003). Next, NopL was shown to interfere with mitogen-activated protein kinase (MAPK) signaling in *Nicotiana tabacum*. MAPK signaling is involved pathogen recognition in both basal plant defense and R-mediated resistance (Pedley and Martin 2005). Part of the plant defensive response is the induction of hypersensitive response (HR). The plant pathogen *Pseudomonas syringae* uses effector proteins AvrPto and AvrPtoB to interrupt MAPK signaling by degrading the plant protein FLS2 (Göhre et al. 2008; Shan et al. 2008). Overexpression of MAPK signaling in plants induces HR to prevent pathogen infections. NopL was shown to suppress cell death induced by the overexpression of MAPK signaling (Zhang et al. 2011b). NopT when expressed in *N. tabacum* or *Arabidopsis thaliana* elicited a strong HR response and necrotic symptoms.

The authors did suggest that it could function as a protease and had similarity to the effector family YopT – AvrPphB (Dai et al. 2008). AvrPphB is an effector in *P. syringae* and functions as an autoprotease, cleaving itself to expose a myristolation site (Puri et al. 1997; Shao et al. 2002). The addition of myristoyl groups after cleavage, target AvrPphB to the cell membrane (Nimchuk et al. 2000). NopT has been shown to have cysteine protease activity and may use autoproteolysis for target to cell membranes, but its role is still uncertain (Fotiadis et al. 2012). NopM was shown to possess E3 ubiquitin ligase activity. Furthermore, when this ability was lost through a point mutation, the positive effects on nodule formation were also lost (Xin et al. 2012).

Even though the function of many specific proteins has not been determined, the accumulated effect of the T3SS effector proteins can be determined through deletion of the entire secretion system. *Bradyrhizobium elkanii,* containing the T3SS, but not the T3SS mutant, was shown to increase the transcription of two genes involved in early nodulation regulation, *ENOD40* and *NIN,* on nod factor recognition deficient line of soybean (Okazaki et al. 2013). This shows that the T3SS effector proteins may be involved in up-regulating host genes involved in nodule formation, independent of nod factor recognition. Further research is needed to more completely understand how these individual effectors are functioning *in planta*.

Type IV secretion system (T4SS)

The T4SS-b is functionally similar to the T3SS-Rch-1 and is also involved in protein translocation, but has a separate evolutionary origin. The T4SS is generally sub-divided into three families based on function, including conjugation, DNA uptake and release, and protein translocation (Cascales and Christie 2003). These three families can

use similar core proteins to form the main channel and may share sequence similarity. Properly identifying which sub-family is present in a specific strain is key. In rhizobia, the T4SS-b shares strong homology to the VirB/VirD4 subunits found in *Agrobacterium*. The core structure consists of 12 proteins, VirB1-B11 and VirD4. The T4SS-b, in *Agrobacterium tumefaciens*, is used for translocation of both T-DNA and effector proteins (Kuldau et al. 1990; Zupan and Zambryski 1995). The function of the T4SS-b is well understood because of its role in plant transformation. *Agrobacterium* and rhizobia are closely related, and understanding of the T4SS-b in *Agrobacterium* has been leveraged to better understand the T4SS-b in rhizobia.

T4SS-b effect on symbiosis

Unlike the T3SS, there is a paucity of information regarding the role of the T4SS in symbiosis. A functional T4SS-b has only been identified in three different species (Table 0.3). Similar to the T3SS, it can have both a positive or negative effect on symbiosis. In *Mesorhizobium loti* R7A nodulation was reduced, but not completely lost, when the T4SS-b was partially deleted and the mutant was inoculated on *Lotus corniculatus*. *M. loti* R7A gained the ability, by losing the T4SS-b, to form nodules on *Leucaena leucocephala* (Hubber et al. 2004). Deleting the T4SS-b in *Sinorhizobium meliloti* Kh46c resulted in approximately a 50% decrease in nodule number on *Medicago truncatula* A17, but did not have a significant effect on *M. truncatula* F83005-5 (Sugawara et al. 2013). This dual positive and negative selection could explain why only 9 of 33 *S. meliloti* and 11 of 13 *S. medicae* were found to contain the T4SS-b (Sugawara et al. 2013).

Regulation of the T4SS-b

Transcription of the T4SS is controlled by a two-component response regulator VirA/VirG (Stachel and Zambryski 1986). VirA is a membrane bound kinase that phosphorylates VirG in response to external factors (Hansen et al. 1994). In contrast, VirG is a transcriptional activator that binds to *vir* boxes. In *R* these regulators are induced by flavonoids that activate VirG (Hubber et al. 2007). Unlike the T3SS effectors, which can be present throughout the genome, T4SS tend to be near VirG (Tampakaki 2014; Vergunst et al. 2000). Research in *A. tumefaciens* has identified a sequence motif, a positive charged C-terminus, present on effector proteins needed for translocation (Vergunst et al. 2005). This same sequence motif is also present on the only two effector proteins identified, Msi059 and Msi061, both in *M. loti* R7A (Hubber et al. 2004). VirD4 interacts with the positive charge signal sequence to transport the protein through the channel (Vergunst et al. 2005). VirD4, and the requirement of a more specific signal sequence, could result in more specificity in protein transport.

Proteins secreted by the T4SS-b

Thus far, only two proteins have been shown to transport through the T4SS-b, Msi059 and Msi061 in *M. loti* R7A. The Msi059 showed partial protein sequence similarity to a C48 cysteine peptidase. Interestingly, the NopD T3E in *S. fredii* HH103 also was a predicted C48 cysteine peptidase (Rodrigues et al. 2007). The C48 cysteine peptidase family contain the protein XopD, a T3E in the plant pathogen *Xanthomonas campestris* (Hotson et al. 2003). XopD encodes an active cysteine protease, and functions *in planta* to target SUMO-conjugated proteins (Hotson et al. 2003). This interferes with the plant's ability to regulate the expression of specific proteins. Msi061

has shared protein similarity with *A. tumefaciens* effector VirF. The VirF interacts with the host Skp1 to facilitate protein degradation of effector proteins VirE2 and Vip1 to unbind the T-DNA after into the host cell (Schrammeijer et al. 2001; Tzfira et al. 2004). Skp1 is a core component of the E3 ubiquitin ligase, which mediates protein degradation (Schrammeijer et al. 2001). The precise activity of Msi059 and Msi061 are still unknown, but current evidence suggests a role in changing protein expression levels *in planta*.

Type VI secretion system (T6SS)

The T6SS is the most recent section system discovered capable of protein transport. The T6SS is known to contain different subfamilies, but the sub-families and their functions have yet to be clearly defined. The number of proteins involved in forming the core structure seem to vary and there is no known secretion signal for protein transport (Bingle et al. 2008). Additionally, how T6SS expression is regulated is unknown. Still, the T6SS is thought to play an important role in the virulence of multiple pathogens, like *Burkholderia mallei* (Schell et al. 2007).

T6SS effect on symbiosis

The sequence for the T6SS has been found in five different species of rhizobia, *R. leguminosarum*, *B. japonicum*, *M. loti*, *S. saheli,* and *S. fredii* (Bladergroen et al. 2003; Bingle et al. 2008; Sugawara et al. 2013). However, a functional T6SS, with an effect on symbiosis, has only been shown in *R. leguminosarum.* In this bacterium a negative effect on symbiosis was observed, where the T6SS prevented nodulation on *Pisum sativum* cv Rondo. A single protein was identified that is secreted through the T6SS. Sequencing of the first 50 amino acids suggested a role in ribose transport. More information is needed to better define the effector and possible role in symbiosis. More strains containing the

T6SS have been identified, but not experimentally tested for function (Bingle et al. 2008; Sugawara et al. 2013).

Key examples of effector involvement in symbiosis

Most studies have focused on deleting specific genes in the core structure, instead of the effector proteins, and observing the overall phenotypic change. This is likely due to the fact that the core genes, unlike effectors, do not vary between species. Additionally, the phenotypic effect(s) of a single effector knockout might be small, again with some strains containing 36 different T3Es. One of the most well characterized and complete examples of the how the T3SS functions is in *S. fredii* strain USDA257. In this case *S. fredii* USDA257 is both a pathogen and a symbiont.

Legumes limit nodule number, and one mechanism used is to abort nodule formation, through a process similar to HR (Vasse et al. 1993). The *S. fredii* USDA257 strain contains NopL, which suppresses cell death through preventing MAPK signaling from inducing HR and cell death (Bartsev et al. 2004; Zhang et al. 2011b). This would, in theory, increase the total number of nodules formed. Soybean have evolved an R gene, Rfg1, capable of detecting T3Es from *S. fredii* USDA257 (Yang et al. 2010). Rfg1 encodes a TIR-NBS-LRR disease resistance protein, which are known to recognize pathogen effectors to induce disease resistant (Belkhadir et al. 2004). In soybean lines expressing Rfg1, nodulation is prevented by *S. fredii* USDA257 , but not in the T3SS knockout mutant (Yang et al. 2010). In addition, *S. fredii* USDA257 formed almost twice as many nodules on the soybean lines without the Rfg1 gene than the T3SS knockout mutant. Taken together, the T3SS, including NopL, can increase nodulation in soybean. Recognition of the T3Es, by Rfg1, results in complete prevention of nodulation. NopL

restricts the plant's ability to prevent infection and nodule formation, and rhizobia become partially pathogenic through using this strategy. The specific protein which is recognized by Rfg1, either directly or indirectly, is still not known. Though this is just one example, it is consistent with observations from other studies showing both the positive and negative effects of the T3SS as listed in Table 0.1. This dual selection also explains why the T3SS is not found in all strains of Rhizobia.

Proposed model

Most of these studies were done by deleting the entire secretion system, versus knocking out only specific effector proteins. Secretion systems are not found in all strains for any species of rhizobia. Typically, if the T3SS or T4SS has a positive effect on nodulation, then deletion of the T3SS results in ~40-60% reduction in nodule number. This shows that secretion systems are not required for effective nodulation. If the T3SS has a negative effect on nodulation, then knocking out the T3SS or T4SS results in a gain of function phenotype, where the strain is now able for form nodules on a host genotype that it was previously unable to nodulate effectively. This shows that secretion systems have been shown to restrict host range. Taken together, the evidence suggests that effector proteins may act in a pathogenic manner. The function of most effector proteins are not known. Many are predicted to modify *in planta* protein levels, and NopL was shown to suppress defense responses. This suggest that rhizobia effector proteins act in a pathogenic manner, similar to the function of other known bacterial effector proteins (Shames and Finlay 2012).

The model we propose here (Figure 0.1), is to demonstrate three points regarding effector proteins: 1) The role of effector proteins is strictly pathogenic, and not involved in symbiosis communication between the rhizobia and host; 2) The role of effector proteins is to increase nodule number. AON is the plants system for regulating nodule number. The mechanism of action for individual effector proteins will differ, but the unifying aspect is the increase in nodule number. This increase could be achieved through forming additional nodules or the prevention of nodule senescence; and 3) Plants use R genes to recognize effector proteins. This recognition results in a host defense response, which prevents nodulation. This serves to establish a host range for rhizobia strains possessing effector proteins which are recognized by the host.

Conclusion

The T3SS, T4SS, and T6SS all play an important role in nodule formation in the symbiosis between rhizobia and legumes, Host specificity has been known to be affected by Nod factors and surface polysaccharides. These factors are important for host recognition of a symbiont versus a pathogen and facilitate infection for nodule formation. Pathogens have long been known to use effector proteins during invasion to promote virulence. Now all three secretion systems that are known to transport effector proteins have been shown to be present in numerous rhizobia and have a strong impact on nodule formation.

These secretion systems function to transport proteins from rhizobia in the plant cytosol. Once in the cytosol, they act to either increase nodulation or result in decreased nodulation through plant defense recognition. How effectors alter the host *in planta* is

still unknown. Identifying how rhizobia use effector protein could have an important agricultural application. Rhizobia may be using these proteins to suppress or prevent AON, and manipulation of this regulation may lead to the development of new strategies for increasing nodule formation. These effector proteins still have not been expressed *in planta*, in legumes, and thus their functions remain unclear. Although several hypotheses have been postulated, the role of T3SS and T4SS are still not fully understood and warrant further research.

Tables and Figures

Table 0.1: Symbiotic effect of the T3SS – Rch-1 in rhizobia

Strain of rhizobia with T3SS – Rch-1	Secreted proteins	Positive effect on symbiosis	Negative effect on symbiosis	References
Rhizobium etli CNPAF512	2	*Phaseolus vulgaris*	Unknown	(Michiels et al. 1995; Fauvart and Michiels 2008)
Bradyrhizobium elkanii USDA61	8	*Macroptilium atropurpureum*, *Glycine max* ev. Clark, *G. max* cv Enrei	*Vigna radiata* cv. KPS1, *G. max* cv Hill	(Okazaki et al. 2009, 2013)
Mesorhizobium loti MAFF303099	8	*Lotus glaber*, *Lotus japonicus*, *Lotus corniculatus* subsp. *frondsus*, *Lotus filicaulis*	*Leucaena leucocephala*, *Lotus halophilus*, *Lotus peregrinus* var. *carmeli*, *Lotus subbiflorus*	(Hubber et al. 2004; Sánchez et al. 2009; Okazaki et al. 2010; Sánchez et al. 2012)
Sinorhizobium fredii NGR234	15	*Tephrosia vogelii*, *Flemingia congesta*, *Lablab purpureus*	*L. leucocoephala*, *Pachyrhizus tuberosus*, *Crotalaria juncea*	(Viprey et al. 1998; Skorpil et al. 2005; Kambara et al. 2009; Kimbrel et al. 2013)
S. fredii HH103	8	*G. max* cv. Peking, Heinong 33, Kochi, and Williams, *Glycyrrhiza uralensis*	*Erythrina variegata*	(Rodrigues et al. 2007; López-Baena et al. 2008)
S. fredii USDA207	13	Unknown	Unknown	(Kimbrel et al. 2013)

S. fredii USDA257	13	*G. max* cv. Peking and Williams, *M. atropurpeum*	*G. max* cv. McCall, *E. variegata*	(Krishnan et al. 2003; De Lyra et al. 2006; Kimbrel et al. 2013)
Bradyrhizobium japonicum USDA6	33	Unknown	Unknown	(Kimbrel et al. 2013)
B. japonicum USDA110	36	*M. atropurpureum* *Glycine max* cv Williams	*V. radiata* cv. KPS2	(Krause et al. 2002; Wenzel et al. 2010; Kimbrel et al. 2013)
B. japonicum USDA122	31	Unknown	Unknown	(Kimbrel et al. 2013)
B. japonicum USDA123	32	Unknown	Unknown	(Kimbrel et al. 2013)
B. japonicum USDA124	33	Unknown	Unknown	(Kimbrel et al. 2013)
Cupriavidus taiwanensis LMG19424	Unknown	Unknown	*L. leucocephala*	(Saad et al. 2012)

*Only strains with functional T3SS – Rch-1 with a known effect on symbiosis are listed. More strains have been sequenced that contain the T3SS – Rch-1, but these have not been experimentally tested for function (de Souza et al. 2012). The number of secreted proteins includes proteins identified through analysis of proteins found externally after induction of the T3SS, and proteins shown to be transported into the cytosol of *Arabidopsis*.

Table 0.2: Predicted functions of T3SS secreted proteins

T3SS – Rch-1 secreted proteins	Strains containing homolog	Predicted function	References
NopA	*Bradyrhizobium japonicum* USDA110, *Mesorhizobium loti* MAFF303099, *Sinorhizobium fredii* NGR234, *S. fredii* HH103, *S. fredii* USDA257	Part of the T3SS extracellular pilus which spans the plants cell wall	(Deakin et al. 2005; Saad et al. 2008)
NopB	*B. japonicum* USDA110, *M. loti* MAFF303099, *S. fredii* NGR234, *S. fredii* HH103, *S. fredii* USDA257	Part of the T3SS extracellular pilus which spans the plants cell wall	(Saad et al. 2005, 2008)
NopD	*S. fredii* HH103	Homology to a predicted C48 cysteine peptidase	(Hubber et al. 2004; Rodrigues et al. 2007)

NopL	*B. japonicum* USDA110, *S. fredii* NGR234, *S. fredii* HH103, *S. fredii* USDA257	Suppresses cell death induced by mitogen-activated protein kinase (MAPK)	(Zhang et al. 2011b)
NopM	*B. japonicum* USDA110, *S. fredii* NGR234, *S. fredii* HH103	E3 ubiquitin ligase, thought to be involved in protein-protein interactions	(Rodrigues et al. 2007; Xin et al. 2012)
NopP	*S. fredii* NGR234, *Rhizobium etli* CNPAF512, *S. fredii* HH103, *S. fredii* USDA257	Phosphorylated by plant kinases	(Skorpil et al. 2005)
NopT	*S. fredii* NGR234	Cysteine protease	(Fotiadis et al. 2012)
NopX	*M. loti* MAFF303099, *Sinorhizobium fredii* NGR234, *S. fredii* HH103, *S. fredii* USDA257	Terminal part of the T3SS extracellular pilus which spans the plants cell wall	(Saad et al. 2008)
Mlr6361	*M. loti* MAFF303099	Shikimate kinase	(Sánchez et al. 2009)

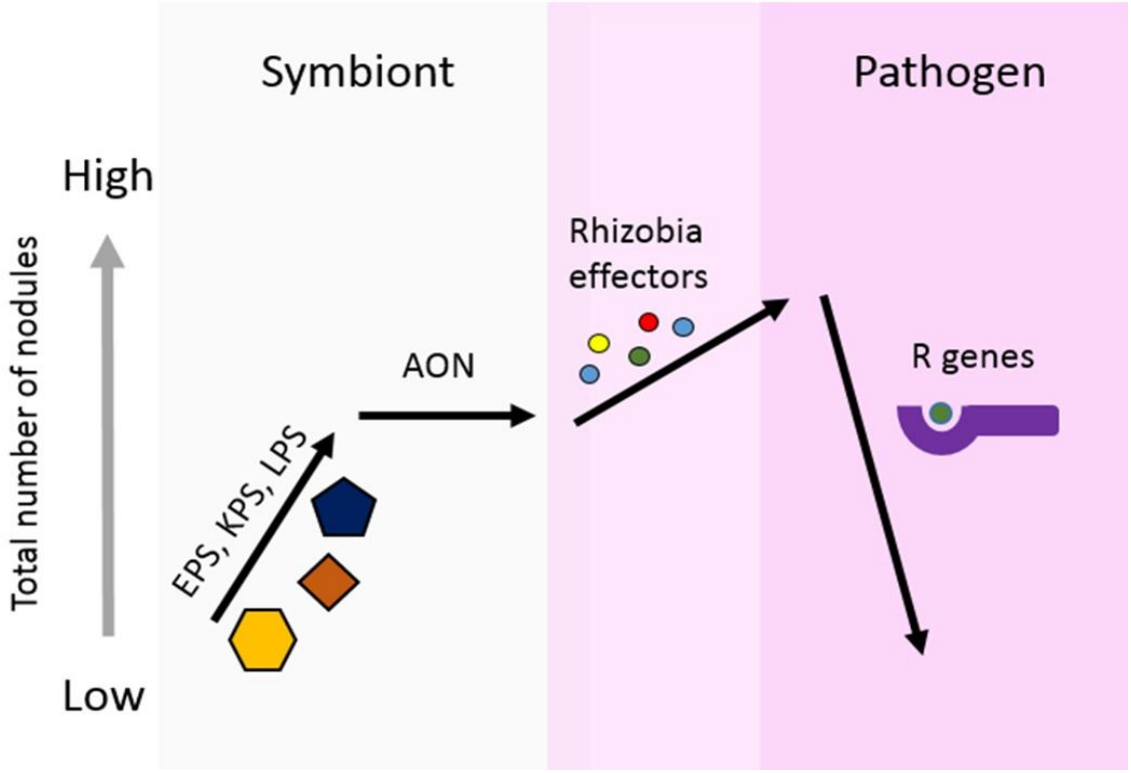
*Subset of known proteins secreted by the T3SS – Rch-1. Only proteins with a predicted function or that have been experimentally tested are listed.

Table 0.3: Symbiotic effect of the T4SS-b

Strain of Rhizobia with T4SS – B	Secreted Proteins	Positive Effect on Symbiosis	Negative Effect on Symbiosis	References
Mesorhizobium loti R7A	2	*Lotus corniculatus*	*Leucaena leucocephala*	(Hubber et al. 2007)
Sinorhizobium meliloti Kh35c	Unknown	*Medicago truncatula* A17, *Medicago tricycla*	Unknown	(Sugawara et al. 2013)
S. medicae M2	Unknown	*M. truncatula* A17	Unknown	(Sugawara et al. 2013)

*Only strains with a functional T4SS-b are listed. More strains containing a T4SS-b have been sequenced, but not experimentally tested for function (Sugawara et al. 2013). Both secreted proteins in *M. loti* R7A, have been shown to be translocated into *Arabidopsis* (Hubber et al. 2004)

Figure 0.1: Proposed model for the role of effector proteins in symbiosis. Surface polysaccharides are known to be involved in determining specificity for nodule formation. These include, extracellular polysaccharides (EPS), capsular polysaccharides (KPS), and lipopolysaccharides (LPS). Legumes limit the total number of nodules formed using autoregulation of nodulation (AON). Rhizobia use effector proteins, similar to pathogens, to alter plant cells to facilitate increased nodule formation. Effectors alter the symbiotic state towards pathogenesis. In response, plants can develop R genes capable of recognizing the presence of these effector protein, either directly or indirectly. Effector recognition results in the plant initiating a defense response and preventing nodule formation.



Chapter 1

Type IV Effector Protein Involved in the *Sinorhizobium*-*Medicago* Symbiosis

A version of this chapter was published in Molecular Plant-Microbe Interactions

Nelson, M. S., Chun, C. L., and Sadowsky, M. J. 2017. Type IV Effector Proteins Involved in the Medicago - Sinorhizobium Symbiosis. Mol. Plant-Microbe Interact. 30:28–34

Summary

In this study, we investigate the genetic elements of the T4SS and the role these elements play in symbiosis. *Sinorhizobium meliloti* and *S. medicae* each contain a type IV secretion system (T4SS), similar to the T4SS used by *Agrobacterium tumefaciens* during pathogenesis. The Cre Reporter Assay for Translocation (CRAfT) was used to validate potential effector proteins. Both *S. meliloti and S. medicae* contain the effector protein TfeA, which is translocated into the host plant. Our sequence analysis revealed the presence of a *nod* box, a nucleotide binding site that initiates transcription of genes involved in symbiosis, upstream of the transcriptional regulator (*virG*) of the T4SS in *Sinorhizobium*. Luteolin, a flavonoid expressed by *Medicago truncatula*, upregulates levels of *tfeA* and *virG*, based on our qRT-PCR analysis. Mutations in the T4SS apparatus or *tfeA* resulted in reduced nodule numbers formed on *M. truncatula* genotypes. In addition, *S. meliloti* strain KH46c, which contains a deletion in the T4SS, was less competitive for nodule formation when co-inoculated with equal numbers of the wild-type strain. Taken together, these results indicate that *Sinorhizobium* uses a T4SS during initiation of symbiosis with *Medicago* and that this transport system plays an important role in symbiosis. To our knowledge, TfeA is the first T4SS effector identified in *Sinorhizobium*. This study suggests that *Sinorhizobium* alters *Medicago* cells *in planta* during symbiosis. This study offers additional bioinformatics evidence that many different rhizobia species might use the T4SS in symbiosis with other legumes.

Introduction

Rhizobia form a well-studied symbiotic relationship with legume plants. In the root or shoot nodule, rhizobia fix nitrogen in exchange for carbon and protection from the plant host. *Sinorhizobium meliloti* (also known as *Ensifer meliloti* and *Rhizobium meliloti*) and *Medicago truncatula* are often used as model organisms for this symbiosis. Nodulation and nitrogen fixation require the concerted efforts of both the plant and the bacterium, using genes involved in recognition, signal exchanges, nodulation, and nitrogen fixation (Saeki 2011). Plant-microbe interactions are complex and fluid, and carry the potential to become commensal or pathogenic (Agrawal 2001). Though rhizobia are often described as symbionts to legumes, they can be potential pathogens.

Whole genome sequencing of symbiotic species of the Rhizobiaceae family has shown the presence of effector secretion systems, including the type III, IV and VI (Nelson and Sadowsky 2015). Although several of these secretion systems are not required for symbiosis, they have been shown to have both positive effects (increasing nodule number) and negative effects (reducing host range) on nodule formation (Deakin and Broughton 2009). Many bacterial species, including pathogens, use secretion systems to transport proteins into eukaryotic host (Cascales and Christie 2003). Specifically, *Agrobacterium tumefaciens* and *A. rhizogenes* use the type IV secretion system (T4SS) to transport effector proteins required for host infection (Berger and Christie 1994; Moriguchi et al. 2001). These proteins alter the host cell to improve virulence. Plants contain resistance (R) genes that recognize effector proteins and activate plant immune responses, which lead to pathogen resistance (Anderson and Frank 2012; Deslandes and Rivas 2012).

Bradyrhizobium elkanii contains the T3SS, which transports effector proteins during soybean symbiosis to increase total nodule number (Okazaki et al. 2009). Some soybean lines contain an R gene that induces effector-triggered immunity, which prevents *Bradyrhizobium elkanii* with the T3SS from forming nodules (Yasuda et al. 2016). *S. fredii* HH103 was also shown to use the T3SS to suppress early defense response by soybean (Jiménez-Guerrero et al. 2015). These interactions are consistent with the plant pathogen zigzag model (Jones and Dangl 2006).

Mesorhizobium loti R7A contains a T4SS similar to that of *Agrobacterium* and transports two effector proteins into the host cell (Hubber et al. 2004). Mutations in the *M. loti* T4SS or effectors resulted in a decrease in the number of nodules *M. loti* formed with *Lotus corniculatus*. However, *M. loti* T4SS mutants could form functioning nodules with *Leucaena leucocephala*, unlike the wild-type strain. These results suggest that *L. leucocephala* contains an R (resistance) gene for these effectors (Hubber et al. 2004). Previous research on the T4SS in *S. meliloti* strain 1021 did not find any relationship between the T4SS and symbiotic effectiveness (Jones et al. 2007b). Illumina sequencing of 48 *Sinorhizobium* strains, revealed 7 different sub-families of the T4SS (Sugawara et al. 2013). T4SSa having a function in conjugation and T4SSb involved in effector protein transport. Strain 1021 contain only T4SSa not T4SSb. Other strains such as *S. meliloti* KH46c and *S. medicae* M2 did contain a T4SS that was phylogenetically similar to the T4SS found in *M. loti* R7A and *Agrobacterium*.

Though many effector proteins have been identified, especially in the T3SS, few have a known function *in planta*. Effector proteins identified in *Sinorhizobium* are of particular benefit because of the genetic modification tools available in *Medicago*. An

understanding of the role and function of effector proteins that are transported through the T4SS could provide new insight into how rhizobia alter the plant cell to increase nodule formation. In this study we report on the identification of effector proteins in *Sinorhizobium* and the phenotypic effects that these proteins have on symbiosis with *Medicago* genotypes.

Methods

Identification of candidate proteins for the CRAfT assay

Candidate proteins that might be translocated through the T4SS system were identified using *Sinorhizobium* sequence data in Mage (https://www.genoscope.cns.fr/agc/microscope/home/). The following criteria were used to identify DNA sequences that code for candidate proteins: 1) sequences are located within 20 kb of *virG*; 2) sequences are present in multiple strains of the same species; 3) strains with candidate sequences have a T4SS, and 4) sequences code for a positively charged C-terminus with at least two arginine residues (Vergunst et al. 2005). Based on these criteria, we identified 5 candidate effector proteins from *S. meliloti* and 2 candidate effector proteins from *S. medicae*.

Construction of plasmids for CRAfT assay

Plasmids containing the nucleotide sequence for the 27 amino-acid C-terminus from each candidate gene were constructed as follows (Hodges et al. 2006). The sense and anti-sense DNA sequence for each C-terminus was purchased from Integrated DNA Technologies, (Coralville, IA). Oligos were designed to contain TCGA on the 5’ end of

the sense strand and CTAG on the 3' end of the anti-sense strand. These strands were annealed so that the nucleotide overhangs matched the overhangs created by restriction enzymes *SalI* and *XbaI*. The dsDNAs were ligated into *SalI/XbaI* digested pSDM3197 (Schrammeijer et al. 2003) and the resulting plasmid constructs were verified by sequencing at the University of Minnesota Genomics Center (St. Paul, MN).

Protein translocation using the CRAfT

Plasmid pSDM3197, *Agrobacterium* strains and *Arabidopsis* seeds were provided by Drs. P. Hooykaas and W. Ream (Vergunst et al. 2000; Hodges et al. 2006). Plasmids were inserted into *A. tumefaciens* strains LBA1100 and LBA 2587 using electroporation as previously described (Hubber et al. 2004). Co-cultivation and Cre recombinase-mediated restoration of kanamycin resistance were performed as previously described (Vergunst et al. 2000).

Construction of candidate gene knockouts in *Sinorhizobium*

Targeted deletions of *Sinorhizobium* genes were obtained by homologous recombination as previously described (Sugawara et al. 2013). The 600 – 800 nt regions flanking the genes to be deleted were amplified by PCR. The region upstream of each gene was digested using *EcoRI* and *BamHI*. The region downstream of each gene was digested using *BamHI* and *HindIII*. The resulting fragment was inserted into the EcoRI/HindIII sites of pk18mob. The resulting plasmid was digested with *BamHI* and ligated with the *BamHI* fragment from the omega Strep/Spec resistance cassette.

The mutants *S. medicae* M2Δ*tfeA* and *S. meliloti* KH46cΔ*tfeA* were created by using homologous recombination that replaced the target gene with a Spec/Strep cassette (Prentki and Krisch 1984). Recombinant strains were selected on TY agar plates

containing 20 µg chloramphenicol, 100 µg spectinomycin, and 100 µg streptomycin per ml.

Nodulation Phenotypes of *Sinorhizobium* strains with *Medicago*

Seeds of *Medicago truncatula* genotypes A17 and R108 were prepared as previously described (Bucciarelli et al. 2006). Plant assays were done in sterile Leonard jars containing a 1:1 mixture of Sunshine #5 (SunGro Horticulture Inc., Vancouver, BC, Canada) and Turface MVP (Profile Product LLC, Buffalo Grove, IL, USA). Plants were inoculated as previously described (Sugawara and Sadowsky 2014) and grown for 6 weeks in a growth chamber. Total plant nitrogen was determined by the Research Analytical Laboratory of the University of Minnesota (St. Paul, MN).

Construction of GFP or RFP tagged *Sinorhizobium*

DNA sequence containing RFP was amplified from vector pPCRRFP6 by PCR and digested with *EcoR1*. It was then ligated into *EcoRI*-digested plasmid pmp6 (Pistorio et al. 2002; Palani 2011). The resulting plasmid, pmp7, was confirmed by sequence analysis. Plasmids pmp6 and pmp7 were transferred to *E. coli* WSM3064 by heat-shock. The transformed *E. coli* strains were co-cultivated with *Sinorhizobium* on TY agar plates. *Sinorhizobium* transconjugants were selected on TY plates containing 9 µg/ml tetracycline and lacking DAP. This resulted in GFP or RFP being incorporated into the chromosome of *Sinorhizobium* strains, as confirmed by PCR, between the *recA* and *alaS* genes (Pistorio et al. 2002).

Competition of nodulation assay

Medicago truncatula genotypes A17 and R108 were grown as described above. The two *Sinorhizobium* strains KH46c and KH46cΔ*B6-B9* were grown in TY medium to an OD_{600}

= 0.7. Strains were combined at a ratio of 1:1 and inoculated on surface sterilized *Medicago* seeds with about 3 x 10^8 cells per seed. After four weeks of growth, nodules were harvested, sterilized by immersion in sodium hypochlorite, and streaked onto TY plates as described by (Sugawara and Sadowsky 2014). Fluorescent microscopy and multiplexed colony PCR was used to examine nodule isolates for the presence of for GFP and/or RFP. Primers were designed using A Plasmid Editor (biologylabs.utah.edu/jorgensen/wayned/ape). PCR products were visualized on 1% agarose gels.

qRT-PCR Assay

Sinorhizobium KH46c was cultured in Vincent minimal medium at 28°. Overnight culture was diluted to 0.1 optical density 600nm. Cultures were grown for 24 hours, luteolin was added 0, 4 or 24 hours prior to RNA collection. Qiagen RNeasy kit was used to collect RNA from all the samples (Qiagen, Chatsworth, CA). All RNA levels were normalized prior to conversion to cDNA. SuperScript II Reverse Transcriptase was used to generate cDNA (ThermoFischer Scientific). qRT-PCR was performed using the iTaq Universal SYBR Green Supermix (BioRad, Hercules, CA). Sample were run in triplicate on a StepOnePlus Real-Time PCR System (Applied Biosystems, Grand Island, NY).

Results

Validation of potential type IV effector proteins

The function and composition of effector proteins vary between rhizobia species (Deakin and Broughton 2009). T3SS effectors translocation is shown by inducing the T3SS and isolating secreted proteins in culture (Kambara et al. 2009). However, this does

not show translocation into the host cell. Previous researchers have validated the location of T4SS effector proteins in rhizobia using the Cre Recombinase Assay for Translocation (CRAfT) (Hubber et al. 2004). This method requires candidate proteins to be tested individually, through the construction of a fused protein. T4SS effectors genes are located near the T4SS, and the proteins have a positive charged C-terminus (Vergunst et al. 2005). Therefore, we chose candidate proteins using the above criteria. In addition, we assumed potential effector proteins would be found in multiple strains containing the T4SS and absence in strains not containing the T4SS. Five candidate effectors were chosen for *S. meliloti* and two for *S. medicae*.

Genetic sequence that codes for the 27 amino acid C-terminus of each candidate protein was fused to Cre in a plasmid (Supplementary Table 1.1). The plasmids were inserted into *A. tumefaciens* using electoporation. The transformed *A. tumefaciens* strains were co-cultivated with *Arabidopsis thaliana* line 3043 (Vergunst et al. 2000). Results of three independent co-cultivations are shown in Table 1.1. Cre:GALLS-27 was used as a positive control. Co-cultivation with Cre:GALLS-27 produced 0.36 kanamycin-resistant calli per root explant on average. Co-cultivation with constructs that code for Candidate 1, from *S. medicae*, and Candidate 4, from *S. meliloti*, both produced 0.29 and 0.32 kanamycin-resistant calli per root explant, respectively. No kanamycin-resistant calli were recovered from co-cultivation with constructs that code for any of the other candidate proteins. Co-cultivation with Cre alone as a negative control did not produce any kanamycin-resistant calli. *A. tumefaciens* LBA2587 (Δ*virD4*), was used as an additional negative control. VirD4 is the protein required for translocating the effector protein (Beijersbergen et al. 1992; Vergunst 2003). Effector proteins translocating

through the T4SS will be unable to restore kanamycin resistance in *A. thaliana* 3043. As expected, constructs coding for Candidate 1 and Candidate 4 did not produce any kanamycin-resistant calli when using this strain.

Genetic layout of the T4SS

Candidate 1 and Candidate 4, share 97% protein identity. These homologous proteins were named TfeA (type IV effector A). The T4SSb is found in 9 *S. meliloti* and 11 *S. medicae* strains (Sugawara et al. 2013). All *Sinorhizobium* strains known to contain the T4SSb also contain *tfeA*. The protein sequence of tfeA is 100% conserved between the strains of each species. *TfeA* is located 1,669 and 1,533 base pairs upstream of *virD4* in *S. meliloti* KH46c and *S. medicae* M2 respectively. Both genes are 888 base pairs long (Figure 1.1).

Nod boxes are conserved sequence motifs upstream of genes involved in symbiosis and nodule formation (Goethals et al. 1992; Suominen 1999). The nod box motif was found upstream of *virG* in *S. meliloti* strain KH46c and *S. medicae* strain M2, as well as in other *Sinorhizobium* strains with the T4SS. This nod box is 84% identical (62 of 74 bp) to the promoter region between *nodD1* and *nodA* in *S. meliloti* and *S. medicae*. VirG is the transcriptional activator responsible for inducing the T4SS apparatus and effector proteins by binding to *vir* boxes (Stachel and Zambryski 1986). There were no identified *vir* boxes upstream of any T4SS proteins based on similarity to *vir* boxes in *Agrobacterium* (Pazour and Das 1990).

***virG* and *tfeA* are upregulated by luteolin**

The plant flavonoid luteolin is the first signal in the *Sinorhizobium/Medicago* nodule formation signaling cascade. It induces expression of genes involved in

nodulation (Peters et al. 1986). *Sinorhizobium* genes involved early in symbiosis are upregulated by exposure to luteolin, which can be quantified using qRT-PCR (Barnett et al. 2004). To validate that the T4SS expression occurs early during symbiosis, we performed a qRT-PCR after exposure to luteolin. Gene expression fold changes were tested at 0, 4 and 24 hours post induction (hpi) (Table 1.2). *UppS* expression levels were previously shown not to by effected by luteolin (Barnett et al. 2004). RNA expression levels were normalized to *uppS*. *NodA* was expression was used as a positive control. *NodA*, *virG* and *tfeA* all had a slight increase in expression at 4 hpi based on an average of two independent experiments. *NodA* and *tfeA* has increased expression at 24 hpi, but *virG* expression was not increased. These results show that *tfeA* is induced by luteolin, which indicates a potential role in symbiosis.

Phenotypic effect of targeted deletions in T4SS or candidate effector proteins

Targeted deletion mutants were used to determine the influence of effector proteins and the entire T4SS on *Sinorhizobium* symbiotic phenotypes. A Spec/Strep cassette was used to replace *tfeA* in *S. medicae* strain M2 and *S. meliloti* strain KH46c. The same technique was used to create targeted knockouts of *virB6-B9,* four proteins part of the T4SS channel (Sugawara et al. 2013).

Medicago genotypes A17 and R108 were inoculated with wild-type or mutant strains. After 6 weeks, the symbiosis phenotypes were measured. *S. meliloti* KH46cΔ*B6-B9* formed a significantly reduced number of nodules relative to the wild-type strain on *M. truncatula* A17 ($p=0.006$) but not on R108 ($p=0.057$). KH46cΔ*tfeA* formed an even smaller number of nodules relative to the wild-type on A17 ($p=0.003$) and R108 ($p=0.004$). *S. medicae* M2Δ*B6-B9* formed marginally fewer nodules than the wild-type on

A17 (p=0.056). *S. medicae* M2Δ*tfeA* formed significantly fewer nodules ($p < 0.001$) than the wild-type on *M. truncatula* A17. Weight per nodule was significantly increased in *S. meliloti* KH46cΔ*B6-B9* relative to the wild-type strain on *M. truncatula* R108 (p=0.068) but not on A17 (p=0.020). KH46cΔ*tfeA* also formed larger nodules relative to the wild-type on A17 ($p < 0.001$) and R108 (p=0.005). *S. medicae* M2Δ*B6-B9* formed larger nodules than the wild-type on A17 (p=0.039). *S. medicae* M2Δ*tfeA* formed larger nodules (p=0.045) than the wild-type on *M. truncatula* A17. Both mutants and wild-type strains produced nodules near the top of plant roots, which were pink in color, which suggests that the mutations did not affect the timing of the start of nodulation. The wild-type strains, however, continued to form nodules farther down the plant root system than did the mutants, and these nodules were smaller and white in color (data not shown).

Plant dry mass of the *Medicago* plants inoculated with the mutant *Sinorhizobium* strains was similar to the plants inoculated with the wild-type strains ($p > 0.05$). None of the plants showed any type of stress response. The chlorophyll content of the *Medicago* plants inoculated with the mutant strains showed no significant difference ($p > 0.05$) from that of the *Medicago* plants inoculated with the wild-type strains (Table 1.3).

Competition for nodulation between *Sinorhizobium meliloti* KH46cΔ*B6-B9* mutant and the wild-type strain

While the T4SS is not present in all *S. meliloti* strains, its presence is associated with an increase in the on the total number of nodules formed on 27 different *Medicago* genotypes (Sugawara et al. 2013). Co-inoculation experiments were performed to determine if the T4SS offered a competitive advantage relative to strains that do not contain the T4SS. To aid in these studies, GFP or RFP were incorporated into the

bacterial chromosome in a region known to not effect nodulation (Pistorio et al. 2002). Wild-type KH46c (GFP) was co-inoculated with KH46cΔ*B6-B9* (RFP) and wild-type KH46c (RFP) was co-inoculated with KH46cΔ*B6-B9* (GFP). Adequate fluorescence could not be detected from strains transformed with RFP. Therefore, PCR was subsequently used for strain identification.

Medicago genotype A17 and R108 were inoculated with the fluorescently-labeled rhizobia. Nodules were harvested four weeks post inoculation and nodule occupancy rates explain what nodule occupancy rates mean for each strain are presented in Table 1.4. Mutant strains KH46cΔ*B6-B9* labeled with GFP and RFP were isolated from 34% of nodules of *M. truncatula* A17 and 36% of nodules of *M. truncatula* R108. The Δ*B6-B9* mutant strains were less competitive for nodulation than the wild-type strains after four weeks of co-cultivation.

TfeA-like proteins found in other Rhizobiaceae strains

The T4SS apparatus with the *virD4* gene that is required for protein translocation is found in multiple members of the Rhizobiaceae family(Nelson and Sadowsky 2015). A BLAST search was used to compare the TfeA amino acid sequence to amino acid sequences in the IMG, NCBI and Microscope databases (Table 1.5). Protein sequences with 22 – 97% identity to TfeA across the entire length of the protein were found in 9 different rhizobia species in the IMG, NCBI, and Microscope databases (Table 1.5). All of these species also contained the T4SSb effector transporting subfamily. This was determined by BLAST comparison of the T4SS to the T4SSb in *S. meliloti*. VirD4 protein percentage identify is shown because *virD4* varies more between the T4SS sub-families. Homologues of VirD4 from the databases shared an amino acid sequence

identity from 80 – 99%. The TfeA-like proteins from the databases contain the T4SS secretion signal, which is required for protein translocation (Vergunst et al. 2005). This conserved region consisted primarily of positively charged arginine residues. Additionally, the *tfeA*-like genes were located in close proximity to *virG* and the T4SS apparatus, all gene lengths and gaps are to scale (Figure 1.1).

Discussion

Determining the molecular interactions that occur between symbionts is important to understanding the symbiotic relationship. Here we identify a new effector protein, TfeA, that is involved in the symbiosis of *Sinorhizobium* with *Medicago*. We show that TfeA is transported through the T4SS into the plant cell. TfeA is upregulated in response to luteolin, indicating it is involved in the early stages of nodulation. Deleting this effector reduces the number of nodules formed on *M. truncatula*.

Despite the large effect that *tfeA* has on symbiosis, it is not found in all *Sinorhizobium* strains. We propose that the *tfeA* may be selected against by specific plant species or specific plant genotypes. The T4SS in *M. loti* R7A is known to restrict host range (Hubber et al. 2004). The T3SS and T4SS effector proteins have been show to repress or enhance nodule formation depending on the host (Deakin and Broughton 2009). The T3SS in *S. fredii* USDA257 is recognized by a *Glycine max* R gene rj2, which prevents nodulation (Yang et al. 2010). It has further been suggested that although these effector proteins are involved in the symbiotic process, they may also be used in pathogen-related responses as well (Marie et al. 2001; Nelson and Sadowsky 2015). If TfeA is acting in a pathogenic manner, instead of a symbiotic manner, then its effects on

altering the host cell will offer important insight into how *Sinorhizobium* modifies nodule formation process.

Interestingly, TfeA has 25% aa identity to VirF in *A. tumefaciens* C58. This *A. tumefaciens* protein is involved in targeted proteolysis by binding to VIP1, an *A. tumefaciens* effector, and Skp1p, a subunit of E3 ubiquitin ligase complex (Schrammeijer et al. 2001; Tzfira et al. 2004). Proteins in *M. truncatula* and other legume species are known to be involved in autoregulation of nodulation (AON) (Kassaw et al. 2015). Effector proteins could potentially target these proteins to limit the plant's ability to restrict nodule formation.

BLAST analysis showed that nine different rhizobia species contain a TfeA-like protein. These proteins were always found associated with the T4SS, often found to be in close genetic proximity (Figure 1.1). Though the TfeA-like proteins have not been experimentally proven to translocate into plant cells, they all contain the signal motif for translocation (Vergunst et al. 2005). While these TfeA-like proteins vary widely in amino acid percent identity to TfeA, VirD4 has > 80% amino acid identity. VirD4 was analyzed for sequence similarity to distinguish between the T4SS sub-families. Unlike the proteins comprising the T4SS channel, effector proteins function inside the host. Effector proteins are likely to experience diversifying selective pressure because they function in many different environments, that is, many different host species. The function of T3SS or T4SS effector proteins *in legume* is still unknown. The continued identification of effector proteins demonstrates their important role in rhizobia/legume symbiosis. Determining the in legume function of these proteins will offer beneficial insight into the rhizobium/legume relationship.

We have shown that deleting *tfeA* (which codes for candidate proteins 1 and 4), significantly reduced the total number of nodules formed on *Medicago*. The nodules that did form were significantly larger than those formed by the wild-type strain, suggesting that the plant was rejecting additional nodules and increasing the amount of carbon provided to each nodule. *M. truncatula* genotypes A17 and R108 infected with the *tfeA* deletion mutants did not show a significantly different plant dry mass, chlorophyll content, or total nitrogen content ($p > 0.05$) from those infected with the wild-type strain. *M. truncatula* A17 is the model genotype, but R108 is more commonly genetically modified. Suggesting, that TfeA could be expressed *in planta*. Plant health, as defined by nitrogen content, plant dry mass, and chlorophyll content, were not altered by the presence of TfeA in the rhizobia. Suggesting, *Medicago* plants used fewer nodules to obtain the same amount of nitrogen, which is consistent with the compensatory nodulation hypothesis (Singleton and Stockinger 1983).

Although TfeA does not appear to affect plant fitness, the T4SS affects nodulation competitiveness. *S. meliloti* KH46c wild-type strain was more competitive at forming nodules than the Δ*virB6-B9* deletion mutant strain. The T4SS may confer a selective advantage to *Sinorhizobium* strains that possess it. However, the T4SS is not found in all *Sinorhizobium* strains and its distribution does not cluster on a phylogenetic tree (Sugawara et al. 2013). The T4SS could be a recent addition to the *Sinorhizobium* population or the T4SS may have been gained and lost at various points in the evolution of this bacterium. Research concerning T3SS in rhizobia suggest that the T4SS could be negatively selected for by a host and subsequently lost from some *Sinorhizobium* strains (Nelson and Sadowsky 2015).

Luteolin, an early symbiosis signaling molecule, acts on the nod-box. A *nod*-box is found upstream of *virG* in *S. meliloti and S. medicae*, so the T4SS is likely upregulated early in symbiosis. Previous research found that expression of the T4SS in *M. loti* R7A is regulated by a *nod*-box found 851 bp upstream of *virA* (Hubber et al. 2007). We found that *tfeA* and *virG* were both upregulated by exposure to luteolin. However, *virG* expression, though upregulated after 4 hpi, was not after 24 hpi. These results are consistent with those found in *M. loti* R7A, where *virG* expression was not significantly increased (Hubber et al. 2007).

This research clarifies the role of the T4SS in *Sinorhizobium*, a model organism. Effector proteins are increasingly being shown to have a large impact on symbiosis, especially nodule number. Increasing nodule number has long been an important goal of research. Currently, only plants with mutations in the AON mechanism result in the hypernodulation phenotype (Oka-Kira and Kawaguchi 2006; Reid et al. 2011b). In soybean, these hypernodulating mutants lack the ability to control nodule number and have lower yields (James 2013). Understanding the molecular action of effector proteins could make it possible to increase nodule number in crop plants.

Conclusion

Sinorhizobium protein TfeA is an effector which translocates through the T4SS. TfeA is induced during symbiosis, increases nodule number and strain competitiveness. TfeA-like proteins are found in 12 different Rhizobiaceae species, and has been shown to be function in three different species. This suggests a conserved mechanism across rhizobia to alter host cells, uses the pathogen associated T4SS.

Tables and Figures

Table 1.1: The number of *A. thaliana* 3043 root explants transformed to express kanamycin resistance after co-cultivation with *A. tumefaciens* containing different constructs.

Strain	Plasmid	Number of kanamycin resistant calli / Total number of root explants	Average number of calli / explant
LBA 1100	Cre	0 / 308	0
	Cre::GALLS[a]	210 / 583	0.36
	Cre::Candidate 1[b]	72 / 251	0.29
	Cre::Candidate 2[b]	0 / 219	0
	Cre::Candidate 3[c]	0 / 208	0
	Cre::Candidate 4[c]	83 / 253	0.32
	Cre::Candidate 5[c]	0 / 185	0
	Cre::Candidate 6[c]	0 / 169	0
	Cre::Candidate 7[c]	0 / 176	0
LBA 2587	Cre::GALLS[a]	0 / 323	0
	Cre::Candidate 1[b]	0 / 214	0
	Cre::Candidate 2[b]	0 / 157	0
	Cre::Candidate 3[c]	0 / 124	0
	Cre::Candidate 4[c]	0 / 237	0

The C-terminus was found in the following microorganisms: [a] *Agrobacterium rhizogenes*, [b] *Sinorhizobium medicae*, and [c] *Sinorhizobium meliloti*

Table 1.2: *S. meliloti* qRT-PCR induction, expression normalized to *uppS*.

S. meliloti KH46c Fold induction			
Hours post luteolin	*nodA*	*virG*	*tfeA*
0	0.71	0.62	0.76
4	2.45 ± 0.50	2.5 ± 0.02	2.41 ± 0.05
24	5.79 ± 1.41	1.09 ± 0.31	5.99 ± 0.28

Table 1.3: Symbiotic phenotype of *Sinorhizobium* mutant and wild-type strains.

A. *Medicago truncatula* A17 inoculated with *Sinorhizobium meliloti* KH46c			
Phenotype	WT	Δ*B6 - B9*	Δ*tfeA*
Nodule number	34.25	22.5*	22.25*
Plant dry mass (g)	0.131	0.128	0.159
Chorophyll content (SPAD)	43.9	43.5	44.1
Per nodule dry mass (mg)	0.141	0.211	0.222**
Total plant nitrogen (%)	4.106	4.012	4.08

B. *Medicago truncatula* R108 inoculated with *Sinorhizobium meliloti* KH46c			
Phenotypes	WT	Δ*B6 - B9*	Δ*tfeA*
Nodule number	28.88	22.38	17*
Plant dry mass (g)	0.17	0.164	0.18
Chorophyll content (SPAD)	36.4	35.9	34.6
Per nodule dry mass (mg)	0.233	0.323*	0.355*
Total plant nitrogen (%)	4.467	4.47	4.479

C. *Medicago truncatula* A17 inoculated with *Sinorhizobium medicae* M2			
Phenotype	WT	Δ*B6 - B9*	Δ*tfeA*
Nodule number	27.13	22.38	14.63**
Plant dry mass (g)	0.113	0.137	0.0887
Chorophyll content (SPAD)	42	41.6	41.6
Per nodule dry mass (mg)	0.139	0.192*	0.202*
Total plant nitrogen (%)	3.824	3.972	4.139

* $p<0.05$, ** $p<0.001$, p-values calculated using t-test (unequal variance)

Table 1.4: Percent nodule occupancy by *Sinorhizobium* KH46cΔ*B6-B9* and the wild-type strain on two *Medicago* genotypes.

	% Nodule occupancy by strain			Nodules tested
Host plant	KH46c	KH46cΔ*B6-B9*	Both	
HM101	64.7	33.3**	2	150
HM340	60.2	35.4**	4.4	113

* $p<0.05$, ** $p<0.001$ p-values calculated using t-test (unequal variance)

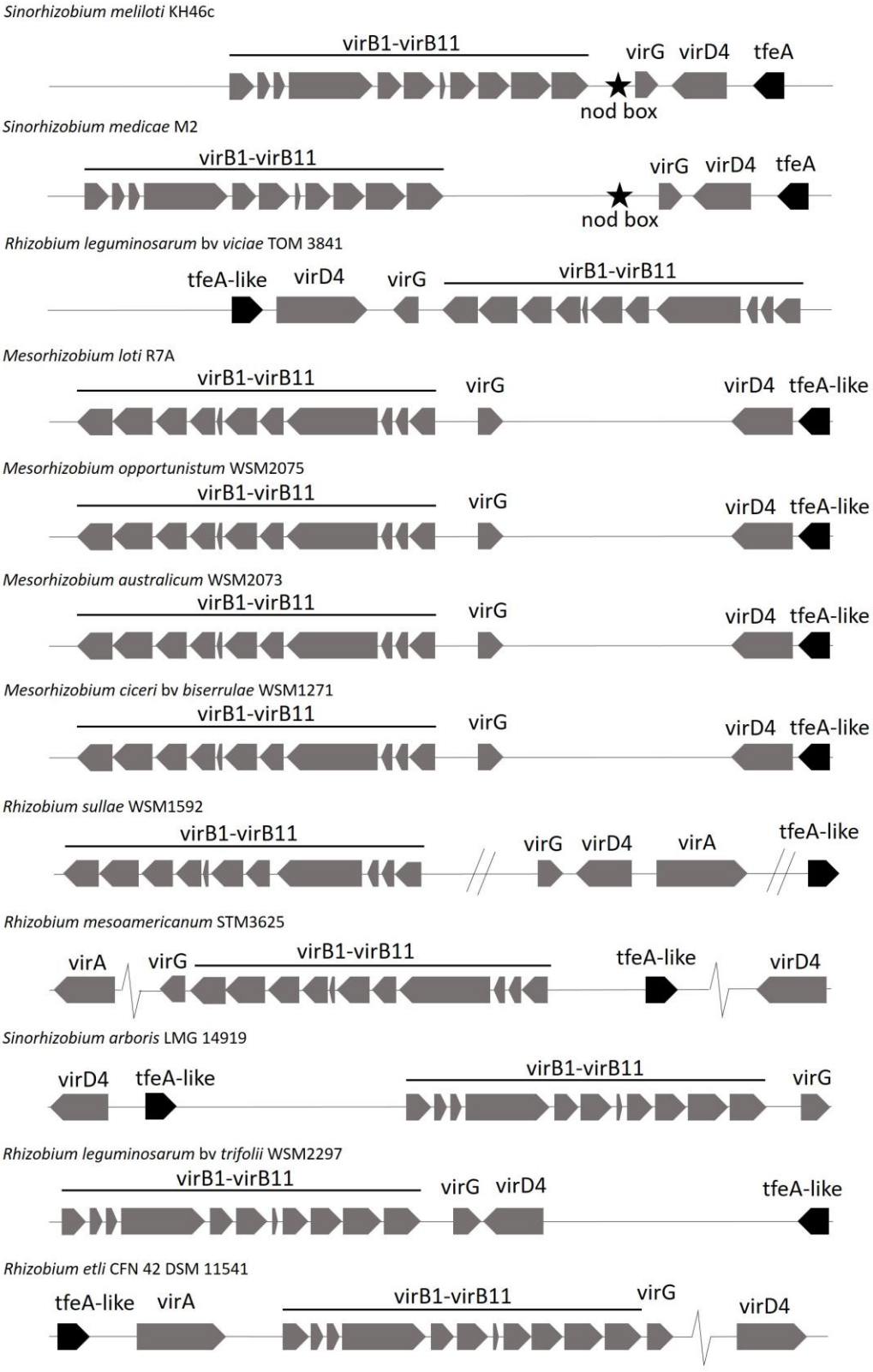
Table 1.5: TfeA-like proteins with the required conserved C-terminus amino acid motif are present in multiple rhizobial species.

Species	TfeA identity (%)	VirD4 identity (%)	27 Amino acid C-terminus [a]
Sinorhizobium meliloti KH46c [b]			ISHAYNHAREDLIASSRSR DRADGTGR
S. medicae M2 [b]	97	99	ISHAYNHAREDLIASSRSR DRADGSGR
Rhizobium leguminosarum bv. *viciae* TOM	70	85	AYNHARDDLRPSSRSRDA AGSRDRTGR
Mesorhizobium. loti R7A [b]	59	81	GRDISHSYNHAREDLMEA RRSRDRTGR
M. opportunistum WSM2075	55	81	GQAINRSYSQARADLQE STRNRDRGGR
M. australicum WSM2073	55	81	GQAINRSYSQARADLQEST RNRDRGGR
M. ciceri bv. *biserrulae* WSM1271	55	81	GQAINRSYSQARADLQEST RNRDRGGR
R. sullae WSM1592	42	90	SISQVSDQARTDLVASFRS RERSDAGR
R. mesoamericanum STM3625	30	80	ESLNRADSDARSELASSSR SRER-NWGR
S. arboris LMG 14919	29	85	RLNRADNDARADLASSSR SRERSNWGR
R. leguminosarum bv. *trifolii* WSM2297	25	84	EHLNRAENEARADLAS-FRSRERSDRGR
R. etli CFN 42, DSM 11541	22	82	ATTEAGEMSRAELMSATR PRRQYDEGR

[a] Amino acid sequencing aligned using Clustal Omega.
[b] Proteins shown to translocate through the T4SS.

Figure Legends

Figure 1.1: Genetic organization of the T4SS and surrounding genes, with the nod box promoter sequence (blue star) in both *S. meliloti* KH46c and *S. medicae* M2 strains.


Sinorhizobium meliloti KH46c
virB1-virB11
virG
virD4
tfeA
nod box
Sinorhizobium medicae M2
virB1-virB11
virG
virD4
tfeA
nod box
Rhizobium leguminosarum bv *viciae* TOM 3841
tfeA-like
virD4
virG
virB1-virB11
Mesorhizobium loti R7A
virB1-virB11
virG
virD4
tfeA-like
Mesorhizobium opportunistum WSM2075
virB1-virB11
virG
virD4
tfeA-like
Mesorhizobium australicum WSM2073
virB1-virB11
virG
virD4
tfeA-like
Mesorhizobium ciceri bv *biserrulae* WSM1271
virB1-virB11
virG
virD4
tfeA-like
Rhizobium sullae WSM1592
virB1-virB11
virG
virD4
virA
tfeA-like
Rhizobium mesoamericanum STM3625
virA
virG
virB1-virB11
tfeA-like
virD4
Sinorhizobium arboris LMG 14919
virD4
tfeA-like
virB1-virB11
virG
Rhizobium leguminosarum bv *trifolii* WSM2297
virB1-virB11
virG
virD4
tfeA-like
Rhizobium etli CFN 42 DSM 11541
tfeA-like
virA
virB1-virB11
virG
virD4

Supplemental Tables

Supplemental Table 1.1: Oligos used to generate candidate fusion proteins.

Species	Candidate	Gene	Oligo	
S. medicae	1	WSM419 Smed_6142	Forward	TCGAATAAGCCACGCTTACAC CACGCCCGAGAGGACCTGATCG CTTCCTCCAGAAGCAGAGATCG CGCTGATGGCTCGGGCCGTTGA
			Reverse	CTAGTCAACGGCCCGAGCCATC AGCGCGATCTCTGCTTCTGGAGG AAGCGATCAGGTCCTCTCGGGCG TGGTTGTAAGCGTGGCTTAT
	2	WSM419 Smed_6146	Forward	TCGACATGGCCCTAAGGGCCTAA ACGAAAGGCGTTTGCCGGCAGA TCCGGAATGCGGCCGGTGTCTCG TCCTACGGAGTGCCCGTTAA
			Reverse	CTAGTTAACGGGCACTCCGTAGG ACGAGACACCGGCCGCATTCCG GATCTGCCGGCAAACGCCTTTCGT TTAGGCCCTTAGGGCCATG
S. meliloti	3	HM007-10 pld620017	Forward	TCGAGGCAGATCCGGAATGCGGC CGGTGTCTCGTCCTACGAAGTG CCCGTTCGGAGCTACACCGGGCA ATCAGAAACGATCGATTTAG
			Reverse	CTAGCTAAATCGATCGTTTCTGAT TGCCCGGTGTAGCTCCGAACGG GCACTTCGTAGGACGAGACACCG GCCGCATTCCGGATCTGCC
	4	HM007-10 pld620021	Forward	TCGAATAAGCCACGCTTACAATC ACGCCCGAGAGGACCTGATCGC TTCCTCCAGAAGCAGAGATCGCG CTGATGGCAGGGGCCGTTGA
			Reverse	CTAGTCAACGGCCCCTGCCATCA GCGCGATCTCTGCTTCTGGAGGA AGCGATCAGGTCCTCTCGGGCGT GATTGTAAGCGTGGCTTAT
	5	HM007-10 pld620023	Forward	TCGACCTATATCTATGCCGAGGA CATCCCCATCGTGCCACTCTATT TCGCCCGCCCGCGCCCGGTGGT CGAGCGCTCGGGCACGCTGA
			Reverse	CTAGTCAGCGTGCCCGAGCGCT CGACCACCGGGCGCGGGCGGGCG AAATAGAGTGGCACGATGGGGA TGTCCTCGGCATAGATATAGG

	6	HM007-10 pld620024	Forward	TCGACGGACCTCCGAGCGCGAAG GGCGTCTGATCGACGGCGATCA GGCCGGCTCCATGGAAAAGAAGA AGCAGGAGGGGTATTTCTGA
			Reverse	CTAGTCAGAAATACCCCTCCTGCT TCTTCTTTTCCATGGAGCCGGC CTGATCGCCGTCGATCAGACGCCC TTCGCGCTCGGAGGTCCG
	7	HM007-10 pld620025	Forward	TCGAGGCACGGGCTCAACAAGGA TCTCGGCTTCACCGACGCGGCA CGTGTGGAGAACATCCGCCGTGTG GCCGAGGTGGCGCGGCTGA
			Reverse	CTAGTCAGCCGCGCCACCTCGGCC ACACGGCGGATGTTCTCCACA CGTGCCGCGTCGGTGAAGCCGAG ATCCTTGTTGAGCCCGTGCC

Chapter 2

Sinorhizobium effector protein TfeA binds to *Medicago truncatula* ARF2

Summary

Effector proteins are used by bacteria to modify the eukaryotic cell. In pathogens, effector proteins promote virulence. Recent studies have shown effector proteins to have a major impact on rhizobia/legume symbiosis. TfeA is a *Sinorhizobium* type IV secretion system effector protein which increases the number of nodules formed. TfeA-like proteins are found in numerous Rhizobiaceae species and has been shown to effect both indeterminate and determinate nodules. In this study, we show that the TfeA binds to *Medicago truncatula* Auxin Response Factor 2 through a yeast two-hybrid screen and bimolecular fluorescence complementation. Similarly, we show that TfeA can also bind to *Glycine max* Auxin Response Factor 27. Auxin response is known to be involved in nodule formation. Transgenic *M. truncatula* and *M. sativa* constitutively expressing *tfeA* were not able to be recovered. Estrogen inducible *tfeA* transgenic *M. truncatula* recovered. Expression of *tfeA* was confirmed uses qRT-PCR. Inheritance of *tfeA* was confirmed by PCR and qRT-PCR. Together this shows that TfeA is used by *Sinorhizobium* to modify host auxin response.

Introduction

Rhizobia/legume symbiosis is a mutualistic interaction where the bacteria supply the plant with nitrogen in exchange for carbon, in specialized organs called nodules (Oldroyd et al. 2011). *Sinorhizobium/Medicago* is a well-studied model for understanding rhizobia/legume symbiosis(Jones et al. 2007a). Nodule organogenesis is formed and maintained through complex signal exchange between the bacteria and the plant(Oldroyd et al. 2011; Jones et al. 2007a). Plants, not bacteria, limit the number of nodules form by a process called auto-regulation of nodulation (AON)(Mortier et al. 2012). Plant cytokinin signaling and auxin response are both thought to be involved in AON(Suzaki et al. 2013).

Some strains of rhizobia, but not all, have been shown to use effector proteins during symbiosis. Effector proteins are used by pathogens to increase virulence(Mattoo et al. 2007; Hansen et al. 1994). In rhizobia/legume symbiosis they have been proposed to suppress autoregulation of nodulation, suppress defense response, or bypass the requirement for nod factors(Nelson and Sadowsky 2015; Okazaki et al. 2016; Miwa and Okazaki 2017). Bacteria use secretion systems to transport effector proteins outside of the cell; the three main families are the type III (T3SS), type IV (T4SS) and type IV (T6SS) secretion systems(Wooldridge 2009). All three systems are found in various rhizobia species, and have a phenotypic effect on nodulation(Deakin and Broughton 2009; Nelson and Sadowsky 2015). Effector proteins are known to inhibit, decrease or increase nodulation depending on host(Deakin and Broughton 2009).

Even though numerous effector proteins have been identified, the *in planta* function of very few effector proteins is known. T3SS effectors NopL, NopM, and NopT are thought to be involved in suppressing plant defense response by blocking the MAPK signal cascade(Bartsev et al. 2004; Fotiadis et al. 2012; Xin et al. 2012; Zhang et al. 2011b). *Bradyrhizobium* effectors Bel2-5 and Mlr6361 are recognized by R (resistance) genes in *G. max* which results in effector triggered immunity (ETI)(Miwa and Okazaki 2017; Tang et al. 2016). Bel2-5 does have protein similarity with XopD in *Xanthomonas campestris*, which is suppresses defense response in *Arabidopsis*(Canonne et al. 2011). Bel2-5 is thought to promote nodulation by suppressing host defense response, but results in rhizobia/host incompatibility when recognized by a host R gene.

Alternatively, it has been shown that effector proteins can promote nodule formation by directly activating Nod factor signaling, bypassing Nod factor perception(Okazaki et al. 2013, 2016). Although the specific effector responsible for initiating Nod factor signaling has yet to be identified(Miwa and Okazaki 2017).

Recently, we reported a new effector protein TfeA, which is transported through the T4SS in *Sinorhizobium*(Nelson et al. 2017). T4SS effector proteins are transported from the bacteria into the host cytosol(Wooldridge 2009). TfeA, along with a TfeA-like effector found in *Mesorhizobium loti* R7A, increased nodulation, strain competitiveness, and was upregulated during nodulation(Hubber et al. 2004, 2007; Nelson et al. 2017). Here we shown how TfeA binds to *Medicago truncatula* ARF2, both in yeast and *in planta*. Constitutively expressed ARF2 in *M. truncatula* and *M. sativa* was unable to be regenerated. However, estrogen inducible ARF2 was recovered in *M. truncatula*, though

there was not a symbiosis phenotype after two weeks of induction. These results show how *Sinorhizobium* modifies *Medicago* to increase nodule formation.

Methods

Yeast Two-hybrid

The yeast two-hybrid was performed using the ProQuest system from Thermo Fisher Scientific. The prey libraries, provided by Dr. Steve Gantt, were created by cloning in nodule mRNA, from *Medicago truncatula* A17, into the pDest22 prey vector. *TfeA* (Smed_v1_mpb1043) was PCR amplified and ligated into, the Gateway compatible entry vector, pCR8/GW/TOPO vector and sequence confirmed. *TfeA* was transferred, by Gateway cloning, into the destination vectors pDest22 and pDest32. Those vectors were then transformed into *Saccharomyces cerevisiae* strain MaV203 using drop-out media lacking either leucine or tryptophan. *S. cerevisiae* cells containing *tfeA*-pDest32 were transformed with the prey library and plated on SC-LTH + 25mM 3AT plates. Each prey library was used for two large scale transformation. Some transformants were diluted and plated on SC-LT to estimate the total number of transformants generated. This ranged between 3.5 and 6.0 million transformants per large scale transformation. After 3 days, growing at 30 degrees Celsius, the colonies that grew were numbered and re-streaked on SC-LTH + 25mM 3AT and SC-LTU plates. After two days only 3 colonies had grown. These colonies were cultured and DNA extracted use the MoBio Ultraclean DNA extraction kit. Purified DNA was transformed into One Shot Top10 Chemically Competent *E. coli*, by Invitrogen. *E. coli* were selected using 100 ug/ml of ampicillin.

The pDest22 plasmid was purified using the Qiagen Miniprep kit. The plasmid was sequenced and re-transformed back into *S. cerevisiae* MaV203 containing *tfeA*-pDest32 to confirm interaction. Many combinations of pDest22 and pDest32 plasmids were transformed to serve as positive and negative controls.

Glycine max arf27 (Glyma12g28550) was synthesize using GenScript. *GmARF27* was then PCR amplified into, the Gateway compatible entry vector, pCR8/GW/TOPO vector and sequence confirmed. *Arf27* was Gateway cloned into the destination vectors pDest22 and transform into *S. cerevisiae* MaV203 containing *tfeA*-pDest32 to confirm protein-protein interaction.

Bimolecular fluorescence complementation

Bimolecular fluorescenece complementation construct were generated using PCR, the entry vector pCR8/GW/TOPO, and the Gateway compatible destination vectors pSPYNE, and pSPYCE(Walter et al. 2004). *Nicotiana* plants were grown for 6 weeks, infected with *Agrobacterium* GV3101 carrying both YFP halves and p19, and analyzed for YFP fluorescence 5 days later(Waadt and Kudla 2008).

Plant transformation

The *S. medicae* type IV effector (*tfeA*, Smed_v1_mpb1043) and a previous candidate effector protein (Smed_v1_mpb1050) were cloned into pILTAB381 under constitutive expression(Samac et al. 2004). *Medicago sativa* Regen-SY(Bingham 1991) was transformed using *Agrobacterium tumefaciens* LBA4404 as described by Austin et al.(Austin et al. 1995). The *tfeA* entry vector from the Y2H was Gateway cloned into the destination vector pCGS661_XVE_LexA_C1300, provided by Dr. Colby Starker and Dr.

Dan Voytas. *M. truncatula* R108 was transformed using pCGS661_XVE_LexA_C1300(*tfeA*) and pILTAB381(*tfeA*), as described in Cosson et al.(Cosson et al. 2015). All transformations were confirmed by PCR.

qRT-PCR Assay

T0 plant leaves were grown on Sh9 media, containing 0 or 25 uM estradiol dissolved in DMSO, for 24 hours in the dark(Zuo et al. 2000; Cosson et al. 2015). Qiagen Plant RNA purification kit was used to collect RNA from all the samples (Qiagen, Chatsworth, CA). All RNA levels were normalized prior to conversion to cDNA. SuperScript II Reverse Transcriptase was used to generate cDNA (ThermoFischer Scientific). qRT-PCR was performed using the iTaq Universal SYBR Green Supermix (BioRad, Hercules, CA). Sample were run in triplicate on a StepOnePlus Real-Time PCR System (Applied Biosystems, Grand Island, NY). Gene expression was normalized to *secret agent*(Kuppusamy 2004; Schnabel et al. 2005).

Nodulation Assay

15 T1 plants from lines 5 and 6, *containing tfeA*, were grown for two weeks in growth pouches in nitrogen free Fahraeus media, according to the protocol(Nelson et al. 2015). Plants were inoculated with *S. meliloti* Kh46cΔ*tfeA*. Every other day 4 ml of Fahraeus media with or without 200 uM estradiol was added. Nodules were counted after two weeks and the plants leaves were screened for *tfeA*. Between 8 and 10 plants for each treatment contained *tfeA*.

Results

Yeast Two-hybrid screening with TfeA

TfeA is known to translocate into *Medicago* from *Sinorhizobium*, which increases nodule formation(Nelson et al. 2017). TfeA has 25% protein identity with VirF, which is known to be involved in protein-protein interactions *in planta*(Schrammeijer et al. 2001). A yeast two-hybrid (Y2H) screen was performed to identify potential *M. truncatula* proteins which bind to TfeA. *S. medicae tfeA* was cloned into bait expression vector in the ProQuest Y2H. Two independent prey libraries were obtained from *Medicago truncatula* A17 roots. Four (A, B, C and D) large scale transformations, using each library twice, generated an estimated total of 20 million individual transformations. Bait plasmids grow in the absence of leucine (-L), prey plasmids grow in the absence of tryptophan (-T), and if the proteins interact the cell can grow in the absence of histidine (-H, weak selection) or uracil (-U, strong selection). Yeast with the *tfeA* bait vector were transformed with the prey libraries and directly plated onto SC –LTH media. The colonies that grew were numbered and re-streaked onto plates lacking SC -U or SC -H. Three colonies grew on SC -U or SC -H after being re-streaked (A1, B3, and C3), no colonies were obtained in the D transformation. All three prey plasmids were recovered in *E. coli* and sequenced. All three contained *M. truncatula* Auxin Response Factor 2 (ARF2, Mtr_2g005240). Although, the prey plasmids all lacked the first 30 nucleotides as compare to the mRNA XM_003592826.2. The prey plasmids were recovered and individually re-transformed into yeast to confirm interaction with TfeA (Figure 2.1). This shows strong interaction between *S. medicae* TfeA and *M. truncatula* ARF2.

Bimolecular Fluorescence Complementation of *Sinorhizobium* TfeA and *Medicago* ARF2

To determine if TfeA and ARF2 would interact *in planta*, we used bimolecular fluorescence complementation (BiFC). BiFC works by fusing two proteins to the split halves of YFP , where if the proteins interact the two halves of YFP are brought close enough together to fluoresce(Walter et al. 2004). *TfeA* and *arf2* were fused to either the N-terminus or the C-terminus of *YFP*. *Nicotiana benthamiana* was transformed with *Agrobacterium* strains carrying the fused constructs and p19 (Figure 2.2).

TfeA interaction with *Glycine max* Auxin Response Factor 1

TfeA-like proteins have been found 12 different rhizobia species, indicating this a common mechanism to increase nodulation across many legume species(Nelson et al. 2017). *M. truncatula* ARF2 has a 91% percent identity with *Glycine max* ARF27 across 98% of the 671 amino acids. To determine if TfeA could interact with G. max ARF27 we performed a Y2H and BiFC (Figure 2.3). TfeA and *G. max* ARF27 did enable yeast growth on both weak and strong selective media. Additionally, BiFC showed that TfeA and ARF27 interacted *in planta*.

TfeA expression in *Medicago*

TfeA, under a constitutive promoter, was transformed into *Medicago sativa*, an agriculturally important crop, which is also nodulated by *Sinorhizobium*. Another *Sinorhizobium* effector candidate protein (*nodQ-like*) was also transformed into *M. sativa*. After four independent experiments, over a hundred transformants contained

nodQ-like gene and 0 contained *tfeA* were recovered. Constitutive expression of *tfeA* was also unable to be recovered in *M. truncatula*. *TfeA* was then put into an estrogen inducible expression vector and transformants were successfully recovered.

TfeA T0 leaves from lines 5 and 6 were assayed for *tfeA* expression when induced with estradiol (Table 2.1). *TfeA* expression was successfully induced but varied considerably. Silencing of inducible promoters has been observed when pattern-triggered immunity is initiated(Igarashi et al. 2013). *CDC16* expression was also determined because *CDC16* expression is known to be regulated by auxin and effect nodule number(Kuppusamy et al. 2009). However, *CDC16* expression did not change when *tfeA* expression was induced.

M. truncatula R108 – *tfeA* T1 plants from lines 5 and 6 were inoculated with *S. meliloti* Kh46cΔ*tfeA* induced with estradiol to determine if *tfeA* increased nodule number. Plants were grown hydroponically, for two weeks, and estradiol added every other day. After two weeks of induction there was no statistical difference in the number of nodules between induced and un-induced plants (Table 2.2). Plants were confirmed to contain *tfeA* by PCR.

Discussion

In this study we identified *Sinorhizobium* effector protein TfeA as interacting with *M. truncatula* ARF2, using both a Y2H screen and *in planta* BiFC. Effector proteins are used by pathogens to increase virulence, and have been proposed to be used by rhizobia to suppress autoregulation of nodulation (AON)(Mattoo et al. 2007; Hansen et al. 1994;

Nelson and Sadowsky 2015). Auxin response factors (ARFs) regulate gene expression through binding auxin and subsequently bind specific DNA sequences to initiate or repress gene transcription(Ulmasov 1995; Ulmasov et al. 1997; Shen et al. 2015). At low auxin levels, ARFs can repress gene transcription by TOPLESS binding to Aux/IAA proteins(Long et al. 2002; Szemenyei et al. 2008; Guilfoyle and Hagen 2012). At higher auxin levels, ARFs activate gene transcription by Aux/IAA proteins being degraded by SCF complex(Calderón Villalobos et al. 2012). As a result, the function of ARFs is different in every cell depending on the current auxin concentration(Korasick et al. 2014). *M. truncatula arf2* is closely related to *A. thaliana arf1*, which is a transcriptional repressor(Ellis et al. 2005). Which means, that auxin would initiate transcription of genes, but that process is blocked by *arf.* Auxin is a primary plant hormone and required for many cellular processes including nodule formation, root initiation, organogenesis, and cell division(Pii et al. 2007; Guilfoyle and Hagen 2007; Mockaitis and Estelle 2008; Su et al. 2015; Li et al. 2016).

Many recent studies have suggested that auxin response and transport to be involved immediately preceding nodule formation(Suzaki et al. 2013). Shoot-derived inhibitors (SDI) or nodule formation are thought to interfere with the action of auxin. The supernodulating *M. truncatula sunn* mutant, has increased auxin transport to the roots (van Noorden 2006). Furthermore, reducing auxin levels correlates with the onset of AON(van Noorden 2006). Flavonoids are known to act as auxin transport inhibitors (Wasson 2006). Exogenously applying auxin transport inhibitors resulted in the formation of pseudonodules in the absence of rhizobia(Rightmyer and Long 2011).

Preventing auxin transport could result in local accumulation of auxin, and initiate an auxin response.

The auxin response nodulation pathway was shown to be nod-factor independent(Rightmyer and Long 2011). Previous work showed t3ss effector proteins bypass Nod factor perception and directly activate Nod factor signaling(Okazaki et al. 2013, 2016). Though, the specific protein responsible has yet to be identified. Together these studies, suggest a mechanism for how TfeA increases nodule formation by binding to ARF2.

ARFs are a diverse family of genes with 23 in *A. thaliana*, 51 in *G. max*, and 24 in *M. truncatula*(Zhang et al. 2011a; Van Ha et al. 2013; Shen et al. 2015). *MtARF2* is closely related to *AtARF1* and *GmARF27*(Shen et al. 2015). MtARF2, AtARF1, and GmARF27 all contain a B3 DNA binding domain, an auxin response domain, and a dimerization motif(Ulmasov et al. 1999; Van Ha et al. 2013; Shen et al. 2015). AtARF1 is a transcriptional repressor, suppressing *AtARF1* expression increases auxin genes in flowers, by binding the sequence TGTCTC (Ulmasov et al. 1997; Ellis et al. 2005). *GmARF27* is constitutively expressed throughout plant tissue, include nodules and roots(Van Ha et al. 2013). *MtARF2* is highly expressed in roots, shoots, leaves, cotyledons, but absent in flowers(Shen et al. 2015). During the early stages of nodule formation *MtARF2* is decreased in the roots and increased in the shoots(Shen et al. 2015). Treatment with exogenous 1-NOA, which blocks auxin polar transport, also decreased *MtARF2* expression in the roots(Shen et al. 2015). Blocking auxin polar transport

increases nodulation, suggesting that decreasing *MtARF2* might result in increased nodule formation.

Though ARFs effect the transcription of numerous genes, *CDC16* was a candidate for regulation by *MtARF2*. *CDC16* expression reduces nodule formation, contains two ARF binding domains in the promoter region, and is induced by auxin. Conversely, decreasing CDC16 expression results in a 4-fold increase in nodulation (Kuppusamy et al. 2009). However, qRT-PCR analysis did not show a change in *CDC16* expression when *tfeA* was highly expressed. This indicates that *CDC16* expression is not effected by *tfeA.*

Here we show that TfeA not only binds to *M. truncatula* ARF2, but also to *G. max* ARF27. The effector TfeA increases nodulation for *Sinorhizobium* and *M. loti*, and a TfeA-like protein is found in at least 12 different Rhizobiaceae species(Hubber et al. 2004; Nelson et al. 2017). This shows how TfeA is found across a diverse set of rhizobia, including both determinate (*Lotus corniculatus*/*M.loti*) and indeterminate nodules (*M. truncatula*/*Sinorhizobium*) (Mergaert et al. 2006; Handberg and Stougaard 1992). Yet, none of the rhizobia species are known to form symbiosis with *G. max*, which forms determinant nodules(Sprent 2007). Similar to *Medicago*, increasing auxin sensitivity in *G. max* decreases nodule formation(Turner et al. 2013; Mao et al. 2013). TfeA binding to *G. max* ARF27 suggests it could have a nodulation phenotype in *G. max*, an agriculturally important crop.

Transformation of *tfeA* under constitutive expression was not successful, though transformation of a control vector or *tfeA* under an inducible promoter was successful.

Additionally, inducing *tfeA*, did not result in plant cell death. This suggests that *tfeA* may interfere with regenerating plantlets from tissue culture. Key component in shoot formation from callus tissue, is applying auxin in the media. TfeA was shown to bind to *M. truncatula* ARF2, potentially interfering with the plants ability to respond to auxin during shoot formation.

Induction of *tfeA* by estradiol was confirmed by qRT-PCR, but was inconsistent. *CDC16*, did not show any change in expression even when *tfeA* was expressed at high levels (Table 2.1). Though *CDC16* is regulated by auxin, there are 24 ARFs in *M. truncatula*. *CDC16* regulation could be controlled by 1 or more of those ARFs and not by ARF2. Another possibility is that CDC16 expression would only be effected by TfeA under high auxin concentrations.

Growing T1-*tfeA* plants for two weeks, did not have a significant impact on nodule number between induced and un-induced plants. This was consistent with nodulation experiments by the Samac Lab, which showed no nodulation phenotype after two weeks between *S. meliloti* Kh46c and *S. meliloti* Kh46cΔ*virB6-B9* on *M. sativa* (Personal communication). This is also consistent with the concept of TfeA involved in suppressing AON, where a phenotype would not be observed till later in plant development. Both TfeA studies that showed a phenotype were with 6 week old plants(Nelson et al. 2017; Hubber et al. 2004).

Conclusion

Our results show *Sinorhizobium* TfeA binds to *M. truncatula* ARF2 both in yeast and *in planta*. TfeA-like proteins are found in numerous Rhizobiaceae species and TfeA binds to *G. max* ARF27. TfeA was transformed and expressed in *M. truncatula* under an inducible promoter. No nodulation phenotype was observed after two weeks of plant growth. TfeA is the first type IV effector with a known function *in planta*.

Tables and Figures

Table 2.1: *M. truncatula* qRT-PCR induction of T0 leaves

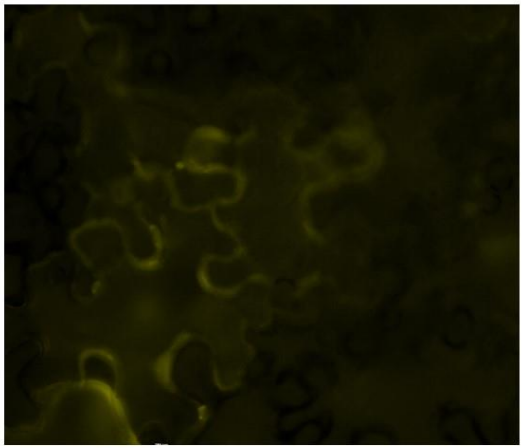
	Leaf *tfeA* Average Fold induction (range)	
Plant Line	*tfeA*	*cdc16*
5	1.49 (1.43 – 1.58)	0.99 (0.88 – 1.1)
6	55.58 (1.13 – 158.98)	1.64 (0.89 – 2.39)

Table 2.2: *M. truncatula*:*tfeA* T1 plants induced with estradiol

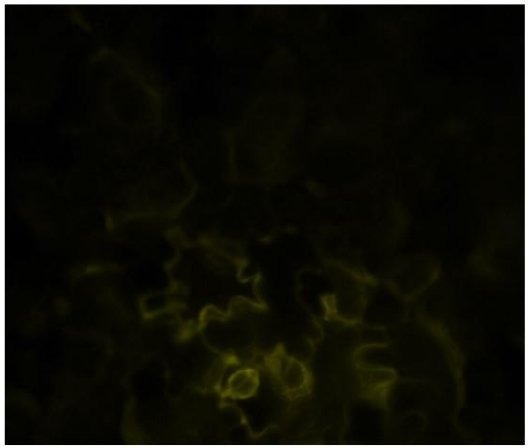
	Total nodules after 2 weeks (standard dev.)	
Plant Line	Control	Induced
5	4.86 (2.23)	4.25 (1.67)
6	4.57 (1.50)	5.00 (1.29)

Figure 2.1: Bi-molecular fluorescence complementation in *Nicotiana* using split YFP.

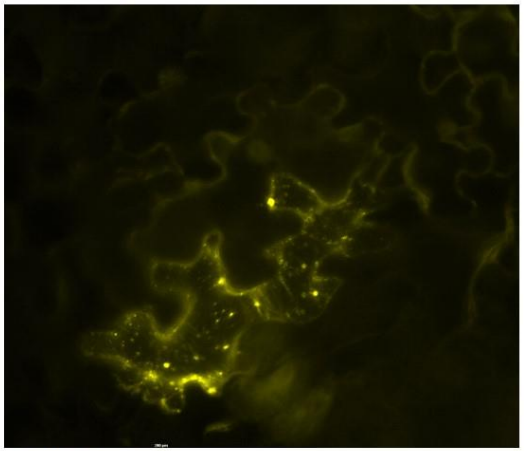
Medicago ARF2:YFP-C
Sinorhizobium TfeA:YFP-N



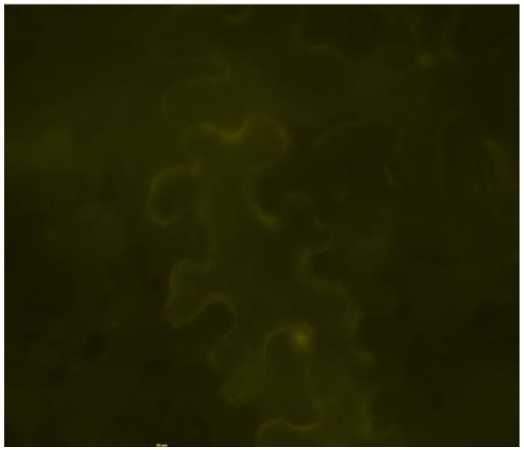
Sinorhizobium TfeA:YFP-C
Medicago ARF2:YFP-N



Medicago ARF2:YFP-C
Medicago ARF2:YFP-N



Saur19:YFP-C
PP2C-D:YFP-N



Sinorhizobium TfeA:YFP-C
Sinorhizobium TfeA:YFP-N

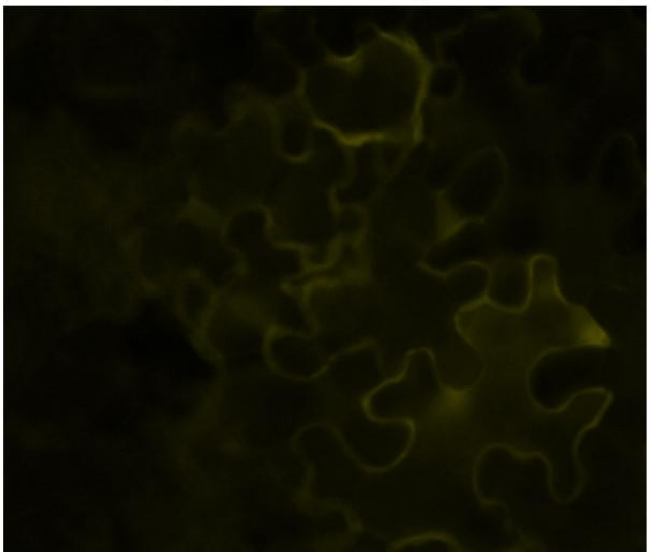
pSPYCE-35S
pSPYNE-35S

Figure 2.2: Yeast two-hybrid showing *M. trunctuala* ARF2 interacts with *Sinorhizobium* TfeA allowing growth on both weak and strong selection

Bait	Prey	SC-LT	SC-LTH+ 3AT 50mM	SC-LTU
Krev1	RalGDS-wt			
Krev1	RalGDS-m1			
Krev1	RalGDS-m2			
TfeA	ARF2			
TfeA	Empty			
Empty	ARF2			
Empty	Empty			

Figure 2.3: Yeast two-hybrid and bi-molecular fluorescence complementation showing interaction between *Glycine max* ARF27 and *Sinorhizobium* TfeA

Bait	Prey	SC-LT	SC-LTH+ 3AT 50mM	SC-LTU
TfeA	*Glycine* ARF1			



Sinorhizobium TfeA:YFP-C
Glycine ARF1:YFP-N

Chapter 3

The Complete Replicons of Sixteen *Sinorhizobium meliloti* Strains Offer Insights into Intra- and Inter-replicon Gene Transfer

Summary

Sinorhizobium meliloti is a model bacterium for understanding legume-rhizobia symbioses. The genome of *S. meliloti* is comprised of three primary replicons (chromosome, pSymB, and pSymA), each with distinct functions and evolutionary histories. Because the majority of previous work has relied on reference genome assemblies using short read lengths, the extent of genome rearrangements within and between replicons has not been previously characterized. Full genome, long-read-based assemblies of 16 strains, 10 of which were sequenced in this study, revealed that movement of genes between replicons is very rare and may act via gene duplications. The main three replicons had distinct gene content, GC percentage, repeat elements and transposable elements. Accessory plasmids, 14 of which were found in 10 strains, were far more similar to pSymA than to the other replicons, in terms of gene-, transposon-, and GC-content. Our results reveal that structural rearrangement within replicons is common but that genes only rarely move between replicons. The three main replicons evolved independently. Accessory plasmids primarily harbored genes from pSymA, which contains most of the nodulation and nitrogen fixation genes. Thus, gene transfer mediated by accessory plasmids may be an important mechanism by which symbiosis genes are transferred between strains and also can explain previous observations that HGT has a greater effect on pSymA than pSymB or chromosomal gene contents.

Introduction

Rhizobia are important bacteria because of the symbioses they form with legume plants. These bacteria provide the plant with fixed nitrogen (N) by converting atmospheric N_2 into a plant-usable form. In exchange, the plant provides carbon to the rhizobia located within root or stem nodules, thereby supporting greater bacterial growth and reproduction than is possible in the soil. The *Medicago truncatula-Sinorhizobium meliloti* symbiosis has been used as a model system to better understand the genetic basis of the rhizobia-legume symbioses and the N-fixation process.

Similar to roughly 10% of bacterial species, *S. meliloti* has more than one replicon. In *S. meliloti*, the non-chromosomal replicons are referred to as symplasmids although non-chromosomal replicons also are referred to as chromids or secondary chromosomes (Harrison et al. 2010). The *S. meliloti* reference genome is ~ 6.5 Mb and is composed of three primary replicons: a chromosome (approximately 3.5 Mb) and two symplasmids (pSymA and pSymB, each of which is 1.5 – 1.7 Mb) (Galibert et al. 2001). Some other strains also contain smaller accessory plasmids (Kuhn et al. 2008). Previous work has shown that the genes involved in certain functions tend to be concentrated on particular replicons: the megaplasmid pSymA has genes with particularly important roles in symbiosis including nodule formation and symbiotic nitrogen fixation (Galibert et al. 2001), pSymB contains a large portion of genes involved in import/export(Finan et al. 2001), and the chromosome contains most of the housekeeping genes. These replicon-specific gene functions have been hypothesized to be the result of an initial acquisition of a plasmid that improves fitness in a particular environment, followed by horizontal gene transfer (HGT)(diCenzo et al. 2016). Relatively little is known about either the origin or

evolution of the accessory plasmids, although they are presumably transient components of the species pan genome.

Several lines of evidence suggest that the three main replicons of *S. meliloti* have distinct evolutionary histories. Among-replicon differences are seen in levels of standing nucleotide variation(Epstein et al. 2012), the effects of purifying and positive selection(Epstein et al. 2014; Galardini et al. 2013), the proportion of duplicated and horizontally transferred genes(Epstein et al. 2014), structural rearrangements (Galardini et al. 2013), and core gene content (Galardini et al. 2013). By all of these measures, the chromosome is the most evolutionary conserved and pSymA is the fastest evolving. However, these and other previous studies have relied primarily on short-read data mapped back to a reference genome or a limited number of complete genomes. Mapping Illumina short-read sequences back to a single reference genome biases the data towards the reference genome, and this method is incapable of revealing large structural variants, characterizing movement of genes between replicons, or accurately assembling repeated sequence(Herring and Palsson 2007; Galardini et al. 2011).

These potential shortcomings can be overcome by using fully assembled, reference quality genomes, which can now be readily obtained through the assembly of long-read, i.e. Pacific Biosciences (PacBio) sequence data(Rhoads and Au 2015). Moreover, reference quality assemblies of multiple genomes allow for identifying gene movement between replicons as well as characterization of transposable elements (TEs) and repeat elements (REs) that are difficult or impossible to robustly align using short-read data. TEs and REs have been inferred to play important roles in gene movement

among bacterial lineages(Frost et al. 2005) and presumably facilitate movement of genic regions between replicons(Guo et al. 2003). TEs are segments of DNA that encode proteins which mediate the movement of DNA within the genome(Frost et al. 2005). In this work, we refer to REs as sequences of DNA repeated one or more times in the genome, which expanded to include more diverged repeats.

Reference-quality genomes can also provide insight into the evolutionary changes of small accessory plasmids, which are not found in the reference. Accessory plasmids are small plasmids that have been found in some *S. meliloti* strains(Kuhn et al. 2008; Crook et al. 2012; Galardini et al. 2011) and can have important effects on host incompatibility and nodule competitiveness(Crook et al. 2012). The limited data on *S. meliloti* accessory plasmids indicate that the genes found on these plasmids are similar to genes found on pSymA or pSymB, but no comprehensive genomic analysis has been done on them(Stiens et al. 2007; Kuhn et al. 2008; Galardini et al. 2011).

Horizontal gene transfer is an important and widespread evolutionary force in bacteria(Boto 2010) and often results in the incorporation of foreign DNA into the heritable genome of an organism. This can be facilitated by inter-replicon gene transfer from a foreign plasmid to the chromosome within a cellular space(Davison 1999). The three independent replicons of *S. meliloti* provide insight into the driving factors of inter-replicon gene transfer.

Here we describe the complete genome sequencing and comparative analysis of 16 *S. meliloti* reference-quality genomes. Ten of these assemblies are new to this study and were completed using high-coverage PacBio data. We used these fully assembled

genome sequences to characterize the: 1) diversity of gene content, TEs and REs found in *S. meliloti*, 2) stability of each of the three main replicons, and 3) genomic composition and relationships among *S. meliloti* accessory plasmids. The analyses revealed that the three main replicons have very different characteristics with respect to gene content, REs, TEs, and the percentage of guanine and cytosine (GC) content. Ten of the strains harbored accessory plasmids, three of which contained more than one accessory plasmid. The gene content of the accessory plasmids was more similar to the gene content of pSymA than the other replicons, but many accessory plasmid genes were not present on the primary replicons of any of the strains. Taken together, our results reveal that the major replicons found within *S. meliloti* have distinct genomic properties and evolutionary histories.

Methods

Genome sequencing, assembly and annotation

We isolated genomic DNA from 10 *Sinorhizobium meliloti* strains using an UltraClean Microbial DNA Isolation Kit from MoBio (MoBio laboratories, Carlsbad, CA, USA). Strains were previously obtained(Sugawara et al. 2013) or isolated from the USDA culture collection. Cultures were grown at 30 °C in TY media(Beringer 1974). Genomic sequence data were generated using a Pacific Biosciences RS II sequencer at the Mayo Clinic (Rochester Minnesota) with 1 SMRT cell per strain. Genomes were assembled using HGAP3(Chin et al. 2013): coverage range (33.3-153.6), read range (20,333 – 100,711), N50 range (17,362 – 20,321) and average read length range (10,826

– 20341). Genomes were sequenced using Gepard(Krumsiek et al. 2007). The assembled genomes were individually polished with Pilon(Walker et al. 2014) using Illumina reads from previous experiments(Sugawara et al. 2013). Base pair changes during the polishing stage for each assembly ranged from 4 to 624 (<0.01% of the genome)(Walker et al. 2014). The protein-coding gene of each assembled genome was annotated using MicroScope(Vallenet et al. 2017). Replicon names were assigned based on gene similarity to the reference strain *S. meliloti* Rm2011(Sallet et al. 2013). The 6 previously sequenced strains were imported from NCBI and were also annotated using MicroScope.

Identification of syntenic regions and core genomes

The "Gene Phyloprofile" tool on MicroScope was used to identify similar genes between bacterial strains or individual replicons, with thresholds set at 80% amino acid identity and 80% length(Vallenet et al. 2017). The MicroScope protein-coding gene annotations were used to generate core-genome and core-replicon gene content. Inter-replicon gene movements were defined at genes present in the core-genome but absent in the core-replicon for the chromosome, pSymA and pSymB.

Identification of transposable and repetitive elements

To predict transposable elements (TEs) *de novo* gene prediction was performed on each genome using Prodigal(Hyatt et al. 2010). Functional predictions were generated using orthologous functional prediction via eggNOG(Huerta-Cepas et al. 2016). TEs were identified based on matching one or more eggNOG-based annotations listed in Supplemental Table 3.1. TEs were then grouped in gene families based on sequence similarity.

Repeats were identified using NUCMER, from the MUMMER package(Kurtz et al. 2004): first, total genomic content of each strain was compared against itself, then the sequences of the matches were extracted and combined with the matches from all strains. NUCMER was then used to compare all repeats to each other and MCL was used to cluster repeats based on sequence matching coverage pairwise for each repeat(Enright et al. 2002). Repeats were then aligned using MAFFT, and poor alignments were trimmed or removed using TrimAL, then re-aligned with MAFFT(Katoh and Standley 2013; Capella-Gutiérrez et al. 2009). HMM profiles were built from these alignments using HMMER(Finn et al. 2011), and the genome of each strain was analyzed for repeat content using nhmmscan with an e-value cutoff of 0.0001(Finn et al. 2011). Regions matching multiple HMM profiles were identified as belonging to the best match, by either length, identity percentage, or score. Principle component analyses (PCA) were performed on the both repetitive element and transposable element contents of each replicon in each strain(Wold et al. 1987).

Mantel Test

We tested for correlations among pairwise genetic distance matrices for each replicon. We first constructed a genetic distance matrix for each replicon using the concatenated alignments of single copy core genes using the dist.dna function from the R(R Core Team 2016) package ape (Paradis et al. 2004) using the TN93 model of evolution(Tamura and Nei 1993). Then, we used a Mantel test (Mantel 1967) implemented in the ade4 package(Dray, S. and Dufour, A.B. and Chessel 2007) with

10,000 permutations to calculate the correlation between distance matrices from different replicons and test significance.

Results

Replicons have distinct gene sets and genomic characteristics

10 new *Sinorhizobium* genomes were *de novo* assembled using PacBio and Illumina sequence, resulting in complete circularized contigs. Each of the 16 completely sequenced *S. meliloti* strains, 10 from this study and 6 previously sequenced(Galibert et al. 2001; Martinez-Abarca et al. 2013; Schneiker-Bekel et al. 2011; Sallet et al. 2013; Galardini et al. 2011), contained three main replicons: a chromosome and two megaplasmids, pSymA and pSymB (Supplemental Table 3.2). Ten strains contained at least one accessory plasmid, two strains contained two accessory plasmids, and one strain contained three accessory plasmids. The total genome sizes of the 16 strains varied from 6.68 to 7.27 Mb with the chromosomes, pSymA, pSymB and the accessory plasmids all having distinct sizes (Table 3.1, Supplemental Table 3.2). Interestingly, the GC content of the chromosome, pSymA, and pSymB were statistically different (Supplemental Table 3.2, all pairwise comparisons $p_{\text{t-test df = 15}} < 0.001$). The accessory plasmids had a GC content similar to pSymA. All genes were annotated by MicroScope(Vallenet et al. 2017). The core genome consisted of 4591 genes with each replicon also having a core gene content: chromosome (2515), psymA (280), psymB (1284).

Transposable elements and repeat elements by replicon

We identified a total of 12,139 TEs, with 683 – 981 TEs per genome. TE sequences comprised 655kb – 856kb (9.3% - 11.7%) of the total genome length. Of the main replicons, pSymA had the highest average density of TEs with a mean 134.1 TEs/Mb, pSymB had 114.4 TEs/Mb, and the chromosome was 87.0 TEs/Mb. Accessory plasmids had an average of 542.6 TEs/Mb, although the accessory plasmids as a group ranged from 78.2 TEs/Mb to 2,360 TEs/Mb. Based on sequence identity, the 12,139 TEs were clustered into 474 TE families, 173 of which were found in every strain, and 42 of which appeared in only a single strain. Of the 173 core TEs, 68 were exclusive to a single replicon and 29 were found on all three of the main replicons. The chromosomes contained 288 TE families, while pSymA, pSymB, and the accessory plasmids contained 191, 161, 209 TE families, respectively. The density of TEs as well as the number of TE families was much higher on the accessory plasmids than the other replicons (Table 3.1) and despite being much smaller than the other replicons, the accessory plasmids contained almost half of all the TE families. TEs on accessory plasmids were also much more likely to be found in only a single strain; more than half of the accessory plasmid TEs were found in only a single strain whereas less than 25% of the chromosomal TEs were found in only a single strain. By contrast, greater than 50% of the chromosomal TEs and greater than 40% of pSymB TEs were found in nearly all of the strains (Supplemental Figure 3.1). Of the TE families, 327 were found only on a single main replicon across all strains, 83 were found on 2 of the main replicons, and 49 were found on all 3 main replicons. Including accessory plasmids as a single replicon class, we find that 232 are found only on a single replicon, 156 found on 2 replicons, 39 found on 3, and

47 found on all 3 main replicons and one or more accessory plasmids. The distribution of TE quantities were structured more by replicon than by the strain from which they were sampled (Figure 3.1).

We identified repeat sequences by first searching for regions of high similarity within each genome using NUCMER(Kurtz et al. 2004), and then clustering these sequences on sequence identity across all genomes to form hidden Markov models (HMMs) that were used to search each genome for more divergent repeat sequences. This process identified 48,133 repeated elements ranging in size from 66 to 12,593 bp that clustered into 688 repeat families. Each strain contained between 2,702 and 3,132 REs for a total of 1.46Mb – 2.04Mb, comprising ~ 21.8% - 28.1% of the total sequence content of each strain. Unlike TEs, which showed a much higher density on pSymA and the accessory plasmids, REs had approximately equal densities on the three main replicons, based on both number and length of repeat: chromosome (417 to 446 repeats,), pSymA (380 to 561), and pSymB (390 to 408), as well as the accessory plasmid (Table 3.1). Approximately 50% of the repeat families were found in all of the strains and less than 5% were found in only a single strain. Whereas TEs tended to be found in very few of nearly all strains (Supplemental Figure 3.1), most REs were found in all strains, with the exception of REs found on the accessory plasmids for which more than 50% were found in only one or two strains (Supplemental Figure 3.1).

Although the majority of REs were not replicon specific, the distribution of REs among all replicons and strains was more strongly structured by replicon than strain (Figure 3.1). Among all strains there were 78 RE families found only on pSymA, 75 only

on the chromosome, 22 on pSymB, and 9 unique to the accessory plasmids. Accessory plasmids and pSymA shared 68 REs that were not found on the chromosome or on pSymB, whereas there were 14 RE families shared by the accessory plasmids and the chromosome but not with pSymA, and 5 RE families were shared by the accessory plasmids and pSymB but not with pSymA.

Inter-replicon gene transfer

To identify inter-replicon gene movement, the core gene content for each replicon was compared to the core genome of all the strains. Two single event gene movements, one in USDA1021 and M162, were responsible for the majority of the inter-replicon gene movement. Removing these two strains changed the core genome (4,687), and the core replicon: chromosome (2,779), psymA (548), psymB (1,297) (Supplemental Table 3.3).

Interestingly, the replication proteins in M162 pSymA are identical to the replication proteins found in accessory plasmids in *E. medicae* strains M2 and WSM419. As per convention, the replicon with the most gene similarity to Rm1021 pSymA was designated pSymA in strain M162. Additionally, the replication proteins on M162 accessory plasmid are identical to the pSymA replication protein found on pSymA replicons from the other *S. meliloti* strains. This suggests that in strain M162, pSymA and an accessory plasmid combined and split unevenly, with the accessory plasmid gaining 800 kb and pSymA losing 800 kb. The other example of large gene transfer was in strain USDA1021, where a 325 kb region (337 genes), moved from the chromosome to pSymB (mpb0525-mpb0862). Consistent with the recent movement of this region, the pSymB in

USDA1021 is ~ 300 kb larger, and the chromosome is ~300 kb smaller than the other strains (Supplemental Table 3.2).

Outside of the two large gene movements (M162 and USDA1021), only 63 core genes experienced inter-replicon translocation (Supplemental Table 3.4). Twenty-three of these translocations were annotated as transposases, nine of the translocations involved multiple genes (from 2 to 5 per translocation), and 17 translocations involved only a single gene. TEs have been associated with gene movement(Frost et al. 2005). The 5 kb flanking sequence of these gene movements were analyzed for TEs, to assess if gene movement was facilitated by TEs. Of the 26 inter-replicon (non-TE) gene movements, only 7 had a TE nearby when comparing the gene location to strain Rm2011 and the alternative gene location in a different strain. All 63 genes were found on multiple replicons within at least one strain. In contrast, the large gene transfer events in USDA1021 and M162 were found in only single copies across all 16 strains, suggesting gene duplication is an important factor for inter-replicon gene transfer.

A core replicon instead of a core genome

Here we define the core genome as the set of genes found in all strains of the same bacterial species (described above), relative to those genes found in the genome Rm2011. For individual strains, the core genome accounted for 58 – 68% of all genes (Supplemental Table 3.5). The core replicon of the chromosome (2,779 genes) represented 66 to 78% of the chromosomal gene content, whereas the core replicon of pSymB (1,297 genes) contained 74 to 80% of the pSymB gene content. In contrast, for pSymA the core genes (548 genes) comprised only 26 to 38% of the gene content

(Supplemental Table 3.5). The percent of genes that are part of the core replicon were significantly different between replicons: chromosome and pSymA ($p = 1.1 \times 10^{-21}$), the chromosome and pSymB ($p = 0.01$), and pSymA and pSymB ($p = 1.8 \times 10^{-22}$).

Accessory plasmids are part of the pSymA pan-replicon

The 14 accessory plasmids identified in the 16 strains contained a total of 4,167 annotated genes, with accessory plasmids having a greater density of annotated genes (1.3 genes / kb) than either the chromosome (1.03 genes / kb), pSymB (1.01 genes / kb) or pSymA (1.17 genes / kb). Unlike the main replicons, there were no genes in common among all the accessory plasmids, and several of these plasmids shared little sequence identity with any of the other accessory plasmids (Figure 3.2) (Krzywinski et al. 2009). Even Replication protein A (RepA), which is required for replication of the plasmid, was not identical across all accessory plasmids(Manen et al. 1992). RepA on pSymA and pSymB proteins are 99% identical across all strains. However, there are many distinct RepA proteins among the 14 accessory plasmids (Supplemental Figure 3.2). Identical RepA proteins found on these accessory plasmids are also found in many other species, including *E. medicae*, *E. fredii*, *A. tumefaciens*, *Rhizobium mesoamericanum*. Furthermore, RepA on USDA1021 accessory plasmid has 52% protein similarity with the RepA on Kh35 accessory plasmid, but 86% with RepA found in *Shinella* DD12. *Shinella* DD12 is found in the gut of zooplankton *Daphnia magna*(Poehlein et al. 2016). This suggests that *Sinorhizobium* is incorporating many different small accessory plasmids from the environment.

Gene content of the accessory plasmids was compared to the pan-replicon of each of the three main replicons for all 16 strains (Table 3.2). The one common element among all of the accessory plasmids was that more genes were exclusively found on pSymA than pSymB or the chromosome. The accessory plasmids contained numerous transposable elements that were often found on multiple replicons and accounted for 12% of the total gene content, on average. Less than 2% of genes were exclusive to the chromosome or pSymB, whereas 32% of the accessory plasmid genes were found exclusively on pSymA, on average. Specifically, the effector transporting Type IV secretion system (T4SS) was found on an accessory plasmid or pSymA. The T4SS is known to be involved, but not required, for nodulation(Nelson et al. 2017). This shows that the accessory plasmids contain genes involved in nodulation and are strongly associated with the pan-replicon of pSymA.

Accessory plasmid B in strain M270 contained 64 genes (>20% of the total) that are were also found on the Ti-plasmid from *A. tumefaciens* strain C58 (Supplemental Table 3.6). Accessory plasmid B did not contain genes required for T-DNA transfer(Hiei et al. 1994),but did contain genes involved in agrocinopine synthases, transport, and catabolism. Agrocinopines are phosphorylated sugars found in tumors induced by *A. tumefaciens*(Ellis and Murphy 1981) Previously, agrocinopines were thought to only be synthesized in crown gall tumors, although a wide variety of bacteria are capable of utilizing them(Moore et al. 1997).

Intra-replicon gene transfer

Synteny statistics from PkGDB (www.genoscope.cns.fr/) were used to gain insight into the extent of structural rearrangements that are segregating within *S. meliloti*(Vallenet 2006; Vallenet et al. 2013, 2017). PkGDB identifies syntenic regions, "syntons", as contiguous linear regions of homologous genes (defined as ≥ 35% identical across 80% of length), in which no two homologous genes are interrupted by more than 5 consecutive non-homologous genes. Thus, the total number of syntons provides a rough estimate for the number of large genomic rearrangements. Because PkGDB calculates syntony statistics between pairs of strains, the genome of strain Rm2011 was used as a reference against which all other strains were compared (Supplemental Table 3.7). Strains Rm1021 and USDA1106 were 100% syntenic (i.e. contained a single synton) with Rm2011. Rm1021 is a lab-derived spontaneous streptomycin-resistant mutant of strain Rm2011(Meade et al. 1982). Each of the other strains showed far less synteny with Rm2011. The number of syntons varied between replicons, with a higher density of rearrangements on the megaplasmid, pSymA (103 – 152 syntons Mb^{-1}), than either pSymB (50 – 85 syntons Mb^{-1}) or the chromosome (34 – 45 syntons Mb^{-1}) (Supplemental Table 3.7). The number of structural rearrangements, as estimated by the number of syntons, was not strongly correlated with percent gene similarity for the chromosome (r^2 =0.03, p=0.92), pSymA (r^2 =0.08, p=0.79), and pSymB (r^2 =0.19, p=0.52).

Replicons evolved independently

In order to evaluate the relative rate of gene content evolution among the replicons of the 16 strains, we examined the percent of genes found in the reference strain that were also found in each of the other strains (Supplemental Table 3.7). Based on

shared gene content, pSymA evolves nearly 3 times faster than pSymB, which evolves approximately 1.5 times faster than the chromosome (on average, the shared gene content of the chromosome, pSymB, and pSymA are 92.3, 87.6, and 66.4%, respectively). Similar results were obtained when other strains were used as the reference (data not shown). A Mantel test revealed no statistically significant correlations in pairwise divergence among strains for any pair of replicons (Mantel test $p > 0.3$, Supplemental Table 3.8 and Supplemental Figure 3.3). The lack of correlation in the rate of divergences can be seen by comparing the order of strains in Supplemental Table 3.7. For example, strain T073 has the fourth most similar chromosome but was only the fourteenth closest to pSymA when compared to Rm2011. In contrast, Rm41 had the twelfth most similar chromosome and was the fourth closest to pSymA.

The faster divergence of pSymA was not due to Rm2011 having gained genes that were not found in the other strains. Approximately 91% of all genes on pSymA of strain Rm2011 were found in at least one pSymA from the other strains (excluding USDA1106 and Rm1021), even though only 75% of the genes were shared by the most similar strain. A similar pattern was found even for the most diverged pSymA in strain T073, for which 80% of the genes were found on pSymA of at least one other strain, even though pSymA shares only 68% of its genes with the next most closely related pSymA (BL225C).

Discussion

Genomic diversity is responsible for phenotypic variation and provides the raw material for evolution. Characterizing genomic diversity is therefore an important step

for both identifying genes responsible for naturally occurring variation as well as gaining insight into past adaptation. Previous work characterizing the genomic diversity of *Sinorhizobium meliloti*, has focused on either using short-read sequence data that have been mapped back to the *S. meliloti* reference genome(Epstein et al. 2012) or assembling a draft genome(Sugawara et al. 2013). Here, we have used high-coverage PacBio sequencing followed by reference-quality genome assembly to examine structural variation and diversity in transposable elements, repetitive elements, and accessory plasmids that are segregating within the *S. meliloti* lineage.

The *Sinorhizobium* genome consists of three replicons found in every strain as well as smaller accessory plasmids that are found in some strains. Previous analyses have found that the replicons differ in their evolutionary histories, the strength of purifying selection to which they are subjected, the extent of horizontal gene transfer, and the proportion of core versus accessory genes(Galardini et al. 2013; Epstein et al. 2012). Here, we show that the three primary replicons also have distinct genomic characteristics, with little evidence for core genes having moved between replicons. The replicons have distinct and largely non-overlapping % GC content and distinct profiles of both TEs and REs.

Perhaps most strikingly, with the exception of a single large translocation event involving the movement of a 300 kb region from the chromosome to pSymB, we detected evidence for only 40 non-transposable-element core genes having moved between replicons – fewer than 1% of the 4,600 core genes. The nearly complete lack of core gene movement between replicons is puzzling given that there is experimental evidence

for frequent genome rearrangements in *Sinorhizobium* under laboratory conditions(Mavingui et al. 2002). Moreover, TEs and REs, both of which are able to contribute to gene movement through the translocation of genes between cells and mediate horizontal gene transfer, are abundant on each replicon and often found on multiple replicons(Wei et al. 2009; Mijnendonckx et al. 2011; Frost et al. 2005; Thomas and Nielsen 2005; Darmon and Leach 2014). The presence of TEs and REs on multiple replicons indicates that movement between replicons is possible, as do the two large translocation events we detected – an 800 kb region that moved from pSymA to an accessory plasmid, and a 325 kb region that moved from the chromosome to pSymB. Although the %GC content consistently differed among replicons, something also found for the multiple chromosomes in *BuKholderia cenocepacia*(Dillon et al. 2015), the magnitude of the differences was small (< 3%) and thus not expected to act as an appreciable barrier to gene exchange. TEs, unlike the Res, did cluster more closely between pSymA and the accessory plasmid (Figure 3.1), suggesting that TEs maybe involved in inter-replicon gene transfer between pSymA and accessory plasmids, as suggested by Table 3.2.

Understanding inter-replicon gene movement is important to understanding evolution because once foreign DNA is inside the cell, it will be destroyed, persist autonomously, or transfer into an autonomously replicating plasmid through inter-replicon gene transfer(Daubin and Szöllősi 2016). Although TEs can clearly have important roles in gene movement, only 7 of the 26 inter-replicon translocation events we detected had TEs in the regions flanking the translocated regions. By contrast, all of the

genes showing evidence of inter-replicon movement were duplicated, with copies on multiple replicons, in at least some strains. These duplicates may suggest that gene duplication is a necessary step in gene transfer between replicons in *Sinorhizobium*, although it is also possible that the duplicates are due to extensive transfer of genes among strains.

Despite finding that inter-replicon gene transfer is rare, we detected evidence for extensive gene movement within replicons, with the chromosome and two megaplasmids each showing evidence for an average of more than 30 "syntons" (Supplemental Table 3.7) per Mb, and pSymA having more than 100 "syntons" per Mb. We calculated the number of syntons using the *S. meliloti* reference genome as a reference and the number would be expected to differ slightly using a different genome as the reference. Moreover, while syntons clearly provide a proxy for a rearrangement event, there is not expected to be a one-to-one relationship between syntons and rearrangements. The much higher number of syntons on pSymA is consistent with previous findings of higher densities of duplicated and horizontally transferred genes on pSymA(Epstein et al. 2014).

In contrast to the relatively rare movement of genes between the primary replicons, we found that movement of genes between pSymA and the accessory plasmids was very high, with more than 40% of accessory plasmids also found on pSymA. The extensive shared gene content between pSymA and the accessory plasmids, together with the very low rates of shared gene content between accessory plasmids and either the chromosome or pSymB, suggests that accessory plasmids might be considered part of the pSymA pan-genome. Moreover, given that small plasmids can play central roles in inter-

strain gene transfer through conjugation(Darmon and Leach 2014), the high rate of gene sharing between pSymA and accessory plasmids suggests that these accessory plasmids might be the primary mechanism by which genes are moved between strains. This is particularly important from a symbiotic and host-range perspective, given that many of the genes that are essential for establishing a functional symbiosis are found on pSymA(Barnett and Kahn 2005).

Accessory plasmids in *S. meliloti* have been shown to cause host incompatibility or increase nodule competitiveness(Crook et al. 2012). Indeed, the T4SS, which increases nodule competitiveness, was found on an accessory plasmid or pSymA(Nelson et al. 2017). The T4SS has also been shown to prevent effective nodule formation between *Mesorhizobium loti* R7A and *Leucaena leucocephala*(Hubber et al. 2004). Demonstrating that the T4SS could have both a positive or negative impact on reproduction depending on the environment. The accessory plasmids exchange gene content with pSymA, allowing for a rapid means for the gain or loss of symbiosis-related genes. This phenomenon may lead to some of the symbiotic instability noted for these and other fast-growing rhizobia, where symbiosis genes are plasmid borne.

While many of the accessory plasmid genes were also found as part of the pSymA pan-genome, this was not the case for all genes and some were clearly more closely related to genes found in other bacterial lineages than to rhizobia, suggesting horizontal gene transfer between bacterial lineages. For example, the repA protein found on one accessory plasmid was more similar to the repA from a *Shinella* strain isolated from the gut of *Daphnia* than it was to the repA from any rhizobia. More striking was the finding

that one of the *S. meliloti* accessory plasmids had 64 *A. tumefaciens*-like genes. *Agrobacterium* and *Sinorhizobium* are closely related and represent different genera within the family Rhizobiaceae(Farrand 2003). *Agrobacterium* Ti-plasmids can be maintained and expressed by *S. meliloti*, though *S. meliloti* is still unable to form tumors on plants(Van Veen et al. 1989). *Rhizobium trifolii* has been shown to induce tumors with the addition of a Ti-plasmid(Hooykaas et al. 1977). Although the *Sinorhizobium* M270 accessory plasmid B did not contain the tumor inducing genes, the presence of the agrocinopine metabolism genes that were previously only identified in *A. tumefaciens* Ti-plasmids(Paulus and Otten n.d.) indicates that rhizobia can obtain genes from *Agrobacterium*. The role of this plasmid remains unknown; however, further analysis could yield insight into how symbiotic rhizobia could use pathogenic mechanisms to facilitate nodule formation.

Conclusion

The complete sequence of 16 *S. meliloti* genomes offers important insight into the evolution of symbiosis in this bacterium. The chromosome, pSymA and pSymB appear to have evolved and function independently. Intra-replicon gene transfer is associated with REs not TEs. Inter-replicon gene transfer is associated with gene duplication rather than TEs. Gene transfer events between accessory plasmids and pSymA demonstrates the mechanism(s) by which the *Sinorhizobium* symbiosis is constantly evolving.

Tables and Figures

Table 3.1: Genomic properties of the chromsomes, megaplasmids, and accessory plasmids from the 16 characterized strains. For each entry the range of values presented in parentheses below the mean. Detailed data are available in Supplemental Table 3.2

	chromsome	pSymA	pSymB	accessory
Size (Mb)	3.69 (3.43 – 3.91)	1.41 (0.89 – 1.63)	1.69 (1.62 – 2.01)	0.23 (0.07 – 0.42)
Number of genes	3807 (3545 - 4203)	1699 (1439 – 2100)	1700 (1620 – 1747)	298 (107 – 497)
% genes that are core	0.73 (0.66 – 0.78)	0.33 (0.26 – 0.38)	0.75 (0.73 – 0.79)	0
% GC content	62.72 (62.57 – 62.83)	60.31 (59.97 – 60.54)	62.40 (62.18 – 62.6)	59.18 (57.45 – 60.66)
RE # / Mb	431.4 (417.0 – 446.5)	437.1 (380.0 – 493.8)	399.2 (390.0 – 408.3)	571.1 (268.0 – 853.3)
RE families	464	360	453	289
TE # / Mb	49.3 (47.0 – 53.35)	59.7 (49.0 – 74.1)	51.4 (48.0 – 53.9)	147.3 (76.0 – 214.3)
TE families	281	184	140	205

Table 3.2: Accessory plasmid replicon genes found in other *Sinorhizobium* strains.

Accessory plasmid strain and name	Total genes	Chromosome Exclusive	pSymA Exclusive	pSymB Exclusive	Multiple replicons
AK83 - plasmid NC_015592.1	107	4	16	1	2
AK83 - plasmid NC_015597.1	391	10	128	4	62
GR4 - plasmid NC_019846.1	213	2	90	1	17

GR4 - plasmid NC_019847.1	374	0	132	2	82
HM006 - accessory plasmid A	248	5	63	2	28
Kh35c - accessory plasmid A	215	3	88	1	24
M162 - accessory plasmid A	497	1	359	1	37
M270 - accessory plasmid A	431	9	83	3	88
M270 - accessory plasmid B	304	11	17	1	46
M270 - accessory plasmid C	181	4	46	2	22
Rm41 - accessory plasmid A	277	3	10	2	12
T073 - accessory plasmid A	197	6	50	0	10
USDA1157 - accessory plasmid A	353	1	216	2	38
USDA1021 - accessory plasmid A	379	13	152	1	87

Figure Legends

Figure 3.1: Principle component analysis plot created from a matric of RE (A) or TE (B) family quantities on each replicon for each strain. Replicons primarily cluster together as opposed to strains clustering together.

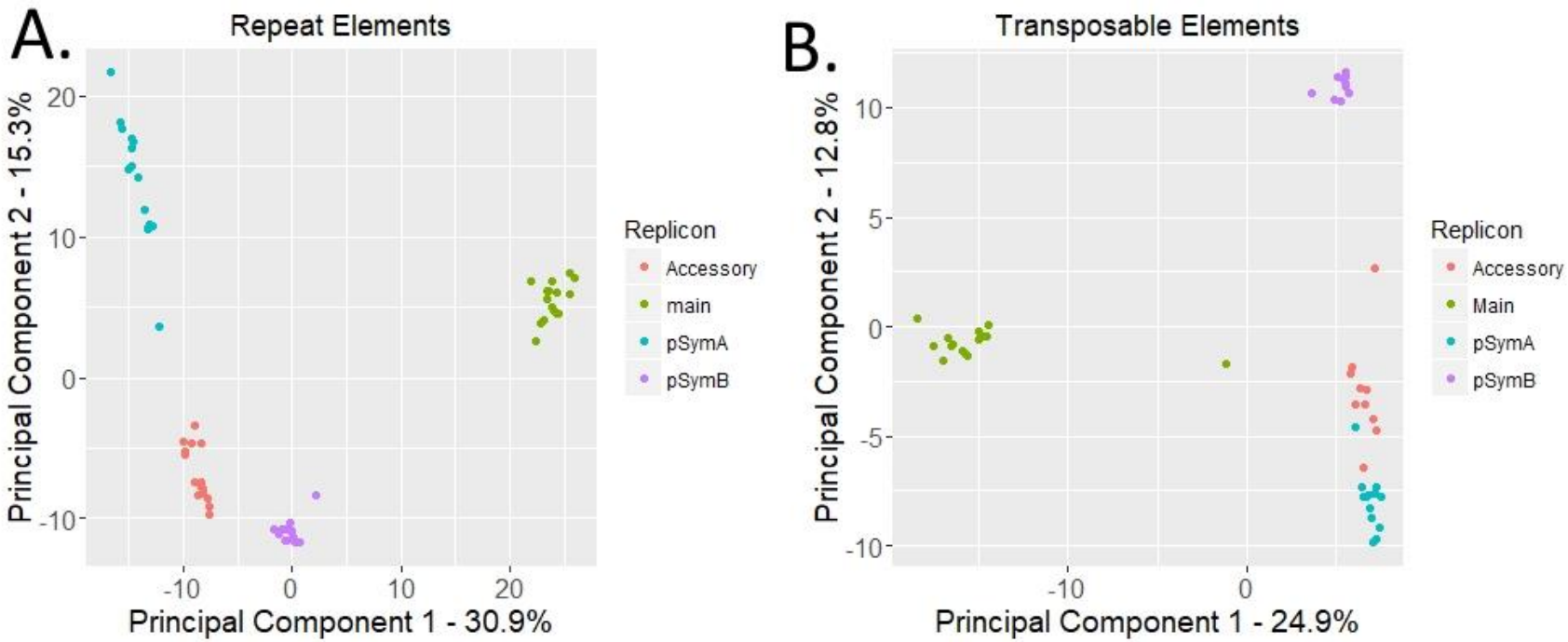
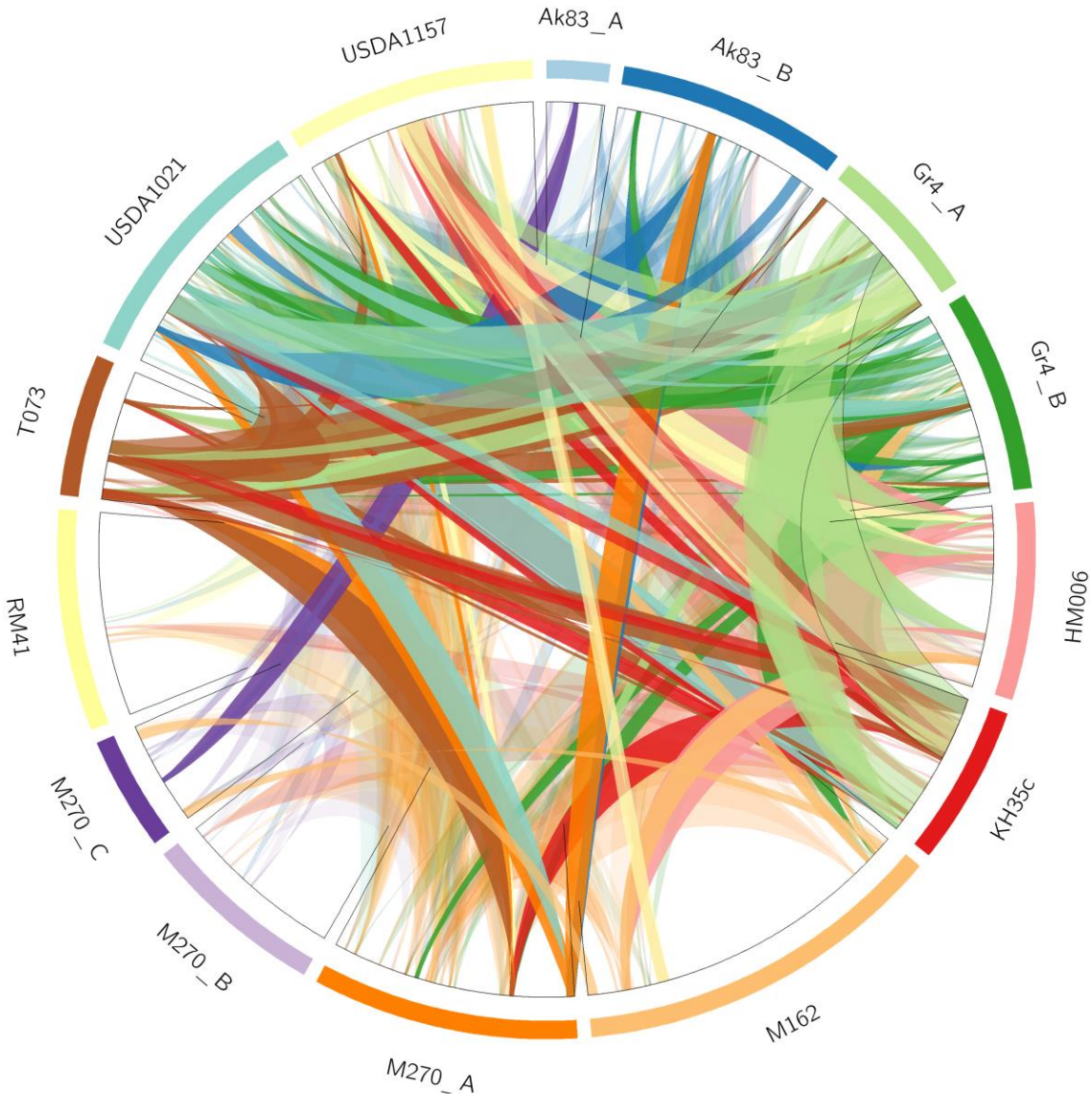

A.
Repeat Elements
Principal Component 2 - 15.3%
Principal Component 1 - 30.9%
20
10
0
-10
-10
0
10
20
Replicon
Accessory
main
pSymA
pSymB
B.
Transposable Elements
Principal Component 2 - 12.8%
Principal Component 1 - 24.9%
10
5
0
-5
-10
-10
0
Replicon
Accessory
Main
pSymA
pSymB

Figure 3.2: Circos plot showing regions (> 2kb) that have ≥80% nucleotide similarity between accessory plasmids in all strains.



Supplemental Tables and Figures

Supplemental Table 3.1: List of the eggNOG-based annotations which were used to identify transposable elements located online at (http://eggnogdb.embl.de/).

Supplemental Table 3.2: List of *Sinorhizobium meliloti* strains with genomic metrics by replicon

Chromosome									
Strain	Size (Mb)	GC %	Genes	RE KB per MB	RE Count per Mb	RE Families	TE KB per MB	TE per Mb	TE Familes
AK83	3.82	62.67	4067	188	417	326	79	48	183
Sm11	3.91	62.65	4212	205	428	336	89	49	192
BL225C	3.67	62.79	3841	186	434	309	78	47	172
GR4	3.62	62.79	3603	175	426	305	83	48	174
Rm2011	3.66	62.72	3623	185	442	332	88	50	182
Rm1021	3.66	62.73	3569	184	441	333	88	50	182
Hm006	3.66	62.77	3778	175	427	304	82	50	182
Kh35c	3.67	62.83	3722	169	419	303	85	47	174
Kh46C	3.70	62.72	3816	172	425	310	81	48	178
M162	3.82	62.60	3990	184	431	349	86	50	191
M270	3.50	62.57	3774	197	429	342	81	49	173
Rm41	3.68	62.75	3789	176	428	325	82	50	185
T073	3.81	62.64	4160	192	446	347	85	49	187
USDA1021	3.43	62.76	3552	194	440	313	82	53	171
USDA1106	3.67	62.72	3744	186	430	332	89	50	183
USDA1157	3.79	62.74	3978	206	437	331	83	50	191
Average	3.69	62.72	3826.13	185.80	431.38	324.81	83.89	49.31	181.25

pSymA									
Strain	Size (Mb)	GC %	Genes	RE KB per MB	RE Count per Mb	RE Families	TE KB per MB	TE per Mb	TE Familes
AK83	1.31	59.97	1703	443	486	269	143	74	97
Sm11	1.63	60.16	2101	413	490	305	133	64	104
BL225C	1.62	60.14	1966	370	444	291	110	52	84
GR4	1.42	60.43	1691	294	380	244	92	51	73
Rm2011	1.35	60.37	1654	322	403	249	110	64	86
Rm1021	1.35	60.40	1658	324	401	249	110	64	86

Hm006	1.48	60.23	1750	384	482	297	118	60	89
Kh35c	1.28	60.26	1469	363	409	254	108	56	72
Kh46C	1.60	60.36	1895	358	450	289	108	56	90
M162	0.89	60.36	1079	338	410	192	117	71	63
M270	1.45	60.31	1745	394	494	287	136	55	80
Rm41	1.55	60.54	1804	332	412	265	106	54	83
T073	1.42	60.35	1589	323	401	252	92	49	70
USDA1021	1.41	60.48	1673	350	432	269	118	60	84
USDA1106	1.36	60.37	1611	335	418	261	115	64	87
USDA1157	1.46	60.20	1703	401	482	300	127	61	89
Average	1.41	60.31	1693.19	359.06	437.08	267.06	115.12	59.65	83.56

pSymB

Strain	Size (Mb)	GC %	Genes	RE KB per MB	RE Count per Mb	RE Families	TE KB per MB	TE per Mb	TE Familes
AK83	1.68	62.37	1828	211	395	239	125	52	87
Sm11	1.66	62.43	1693	188	398	240	116	52	87
BL225C	1.69	62.24	1693	188	395	228	112	51	86
GR4	1.70	62.42	1747	212	404	243	122	52	88
Rm2011	1.68	62.42	1722	210	405	230	119	52	88
Rm1021	1.68	62.40	1743	210	404	229	118	52	87
Hm006	1.63	62.60	1624	193	400	238	117	48	78
Kh35c	1.62	62.60	1624	205	396	225	121	51	83
Kh46C	1.68	62.43	1692	217	408	243	126	53	89
M162	1.68	62.36	1733	205	399	241	121	53	89
M270	1.69	62.18	1732	192	396	232	118	54	91
Rm41	1.67	62.33	1706	200	395	233	116	50	84
T073	1.65	62.46	1718	192	392	233	113	48	79
USDA1021	2.01	62.45	2089	202	390	262	121	50	101

USDA1106	1.68	62.42	1705	211	405	229	117	52	87
USDA1157	1.70	62.31	1747	209	406	239	118	52	89
Average	1.69	62.40	1737.25	202.82	399.20	236.50	118.77	51.40	87.06

Accessory Plasmids

Plasmid	Size (Mb)	GC %	Genes	RE KB per MB	RE Count per Mb	RE Families	TE KB per MB	TE per Mb	TE Familes
AK83 - plasmid NC_015592	0.07	57.91	107	400	457	19	186	214	15
AK83 - plasmid NC_015597	0.26	58.91	391	619	773	117	219	150	39
GR4 - plasmid NC_019846	0.18	59.98	213	378	422	58	122	122	22
GR4 - plasmid NC_019847	0.23	58.58	374	748	665	105	309	204	47
Hm006 - accesory A	0.22	60.00	248	419	377	56	109	105	23
Kh35c - Accesory A	0.18	59.74	215	452	467	61	178	150	27
M162 - accesory A	0.42	60.66	497	425	579	137	117	76	32
M270 - accesory A	0.30	58.38	431	777	853	121	283	150	45
M270 - accesory B	0.20	59.06	304	675	825	76	295	210	42
M270 - accesory C	0.13	59.86	181	500	585	46	208	177	23
Rm41 - accesory A	0.25	59.35	277	245	268	43	124	84	21
T073 - accesory A	0.16	57.45	197	401	519	63	100	113	18
USDA1021 - accessory A	0.30	58.49	379	720	727	120	277	150	45
USDA1157 - accesory A	0.28	60.21	353	447	479	89	167	157	44
Average	0.23	59.18	297.64	514.66	571.08	79.36	192.31	147.30	31.64

Supplemental Table 3.3: The core genes for all strains excluding M162 and USDA1021 by genome and replicon using Rm2011 as the reference, located online at (http://www.genoscope.cns.fr/agc/microscope/home/index.php).

Supplemental Table 3.4: Genes that are part of the core genome but are not exclusive to a single replicon.

Gene	Product	Rm2011
_	Calcium-binding protein	SM2011_c00286
gltB	putative glutamate synthase NADPH large chain	SM2011_c04028
etfB1	Putative electron transfer flavoprotein beta-subunit beta-ETF flavoprotein small subunit	SM2011_c00729
etfA1	Putative electron transfer flavoprotein alpha-subunit alpha-ETF flavoprotein	SM2011_c00728
_	conserved protein of unknown function	SM2011AM_3067
kpsF2	Arabinose-5-phosphate isomerase	SM2011_b20830
_	Calcium binding protein	SM2011_b20838
algI	Putative membrane protein involved in D- alanine export	SM2011_b20843
_	putative transmembrane protein	SM2011_b20842
_	Acyl carrier protein	SM2011_b22007
_	hypothetical protein	SM2011_b20841
_	Hypothetical protein	SM2011_b21695
_	Hypothetical protein	SM2011_b21013
_	Hypothetical protein	SM2011_b20910
_	Putative glycoside hydrolase	SM2011_b20518
_	protein of unknown function	SM2011AM_1303
_	outer membrane protein	SM2011_a1037
_	Multicopper oxidase	SM2011_a1038
_	Copper binding protein	SM2011_a1041
adhC2	AdhC2 glutathione-dependent dehydrogenase	SM2011_a2113
etfB2	EtfB2 electron transport flavoprotein,beta subunit	SM2011_a1391
etfA2	EtfA2 electron-transport flavoprotein,alpha-subunit	SM2011_a1389
aqpZ2	AqpZ2 aquaporin	SM2011_a0627
_	conserved protein of unknown function	SM2011AM_1215
actP	ActP Copper translocating P-type ATPase	SM2011_a1013
hmrR1	Cu(I)-responsive transcriptional regulator	SM2011_a1014
_	Hypothetical protein	SM2011_a2299
cysP1	Bacterial extracellular solute-binding protein,family 1	SM2011_a2069
cysA1	Sulfate/thiosulfate import ATP-binding protein	SM2011_a2067
_	Hypothetical protein	SM2011_a1951
_	hypothetical protein	SM2011_a2065
_	Sensor histidine kinase	SM2011_a2063
_	hypothetical protein	SM2011_a2061
_	Transcriptional regulator,MucR family	SM2011_a0748
_	putative DNA-binding protein	SM2011_a2151
_	hypothetical protein	SM2011_a6501
traC	TraC conjugal transfer protein	SM2011_a0933
_	Hypothetical protein	SM2011_a2071
vapB	Antitoxin VapB32	SM2011AM_1478
_	hypothetical protein	SM2011_a2231
_	Putative transposase for ISRm17 protein	SM2011_c02430

_	Putative transposase for insertion sequence ISRm17	SM2011_c03260
_	TRm17C; corresponds to SMc04039; Putative partial transposase for ISRm17 protein	SM2011_c04039
_	putative TRm24 transposase	SM2011_c01195
_	Putative transposase of insertion sequence ISRm22 protein	SM2011_b20766
_	putative ISRm5 transposase	SM2011_b20060
_	Putative transposase of insertion sequence ISRm2011-2	SM2011_b20778
_	putative ISRm2011-2 transposase protein	SM2011_b20304
_	putative ISRm2011-2 transposase protein	SM2011_b20305
_	Putative transposase of insertion sequence ISRm2011-2,orfA protein	SM2011_b20783
_	TRm5 transposase	SM2011_a0995
_	TRm2011-2b transposase	SM2011_a0020
_	TRm2011-2b transposase	SM2011_a1799
_	TRm2011-2b transposase	SM2011_a2045
_	TRm2011-2b transposase	SM2011_a2375
_	TRm2011-2a transposase	SM2011_a0018
_	TRm2011-2a transposase	SM2011_a1803
_	TRm2011-2a transposase	SM2011_a2373
_	TRm2011-2a transposase	SM2011_a2043
_	TRm17a transposase	SM2011_a0475
_	corresponds to SMa1636; transposase fragment	SM2011_a1636
_	corresponds to SMa0591; transposase fragment	SM2011_a0591
_	corresponds to SMa0787; transposase fragment	SM2011_a0787

Supplemental Table 3.5: Number of genes on each replicon. The percent of core genes in each replicon.

	AK83	SM11	BL225C	GR4	RM2011	RM1021	HM006	KH35c	KH46c	M270	RM41	T073	USDA1106	USDA1157	Core
Chromosome genes	3990	4203	3791	3578	3616	3545	3757	3700	3795	3756	3766	4132	3721	3954	2779
pSymA genes	1439	2100	1768	1642	1640	1656	1745	1469	1895	1739	1800	1585	1607	1701	548
pSymB genes	1721	1692	1668	1730	1720	1743	1620	1624	1690	1732	1704	1712	1703	1747	1297
Accessory A	107			213			248	215		431	277	197		353	
Accessory B	391			374						304					
Accessory C										181					
Total Genes	7648	7995	7227	7537	6976	6944	7370	7008	7380	8143	7547	7626	7031	7755	4687
Core percentage of genes	0.61	0.59	0.65	0.62	0.67	0.67	0.64	0.67	0.64	0.58	0.62	0.61	0.67	0.60	
Chromosome gene percentage	0.70	0.66	0.73	0.78	0.77	0.78	0.74	0.75	0.73	0.74	0.74	0.67	0.75	0.70	
pSymA gene percentage	0.38	0.26	0.31	0.33	0.33	0.33	0.31	0.37	0.29	0.32	0.30	0.35	0.34	0.32	
pSymB gene percentage	0.75	0.77	0.78	0.75	0.75	0.74	0.80	0.80	0.77	0.75	0.76	0.76	0.76	0.74	

Supplemental Table 3.6: Genes with similarity between *Sinorhizobium meliloti* M270 accessory plasmid B and *Agrobacterium tumefaciens* C58 Ti-plasmid

M270 Accessory plasmid B	Begin	End	Evidence	Gene	Product	Agrobacterium tumefaciens C58 plasmid Atu NC_003065
SMEL_v1_pb0017	8645	9763	automatic/finished	noxB	D-nopaline dehydrogenase	Atu6021
SMEL_v1_pb0018	9760	10053	automatic/finished	_	conserved protein of unknown function	Atu6020
SMEL_v1_pb0019	10050	11468	automatic/finished	_	conserved protein of unknown function	Atu6019
SMEL_v1_pb0020	11629	12204	automatic/finished	_	Arginase (fragment)	Atu6018
SMEL_v1_pb0021	12209	12490	automatic/finished	_	Arginase (fragment)	Atu6018
SMEL_v1_pb0022	12566	13648	automatic/finished	_	NAD/NADP octopine/nopaline dehydrogenase	Atu6017
SMEL_v1_pb0023	13696	14757	automatic/finished	arcB	Ornithine cyclodeaminase	Atu6016
SMEL_v1_pb0026	16378	17082	automatic/finished	traR	Transcriptional activator protein TraR	Atu6134
SMEL_v1_pb0027	17395	18048	automatic/finished	yjbB	fragment of putative transporter (part 2)	Atu6136
SMEL_v1_pb0029	18244	18744	automatic/finished	yjbB	fragment of putative transporter (part 1)	Atu6136
SMEL_v1_pb0033	22429	23433	automatic/finished	oppF	oligopeptide transporter subunit ; ATP-binding component of ABC superfamily	Atu6141
SMEL_v1_pb0041	29090	29653	automatic/finished	_	Transcriptional activator protein TraR (fragment)	Atu6134
SMEL_v1_pb0042	29701	30009	automatic/finished	traM	Transcriptional repressor TraM	Atu6131
SMEL_v1_pb0047	32785	36138	automatic/finished	traA	Conjugal transfer protein TraA	Atu6127
SMEL_v1_pb0048	36335	36631	automatic/finished	traC	Conjugal transfer protein TraC	Atu6126
SMEL_v1_pb0049	36606	36851	automatic/finished	traD	Conjugal transfer protein TraD	Atu6125
SMEL_v1_pb0050	36838	38790	automatic/finished	traG	Conjugal transfer protein TraG	Atu6124
SMEL_v1_pb0051	38811	39209	automatic/finished	_	conserved protein of unknown function	Atu6123
SMEL_v1_pb0052	39209	39676	automatic/finished	_	Nuclease	Atu6122
SMEL_v1_pb0054	39933	40169	automatic/finished	_	conserved protein of unknown function	Atu6121
SMEL_v1_pb0055	40307	40585	automatic/finished	hupB	DNA-binding protein HU-beta	Atu6120

SMEL_v1_pb0056	40627	40797	automatic/finished	_	conserved protein of unknown function	AERS4p2_0127
SMEL_v1_pb0058	42951	43268	automatic/finished	_	conserved protein of unknown function	Atu6119, Atu6112
SMEL_v1_pb0060	43939	44775	automatic/finished	_	conserved protein of unknown function	Atu6111
SMEL_v1_pb0071	48631	49230	automatic/finished	_	conserved protein of unknown function	Atu6109
SMEL_v1_pb0072	49747	50052	automatic/finished	phnA	Protein PhnA	Atu6108
SMEL_v1_pb0073	50053	50274	automatic/finished	_	conserved protein of unknown function	Atu6107
SMEL_v1_pb0074	50271	50579	automatic/finished	_	conserved protein of unknown function	Atu6106
SMEL_v1_pb0075	50576	51463	automatic/finished	_	conserved protein of unknown function	Atu6105
SMEL_v1_pb0076	51460	51933	automatic/finished	_	conserved protein of unknown function	Atu6104
SMEL_v1_pb0078	52030	53466	automatic/finished	_	Plasmid partitioning protein (fragment)	Atu6103
SMEL_v1_pb0080	54059	55744	automatic/finished	_	SNF-2-family methyltransferase (fragment)	Atu6101
SMEL_v1_pb0081	55705	57654	automatic/finished	_	SNF-2-family methyltransferase (fragment)	Atu6101
SMEL_v1_pb0083	57762	58856	automatic/finished	_	protein of unknown function	Atu6101
SMEL_v1_pb0085	58986	59291	automatic/finished	_	conserved protein of unknown function	Atu6100
SMEL_v1_pb0086	59398	60000	automatic/finished	_	conserved protein of unknown function	Atu6099
SMEL_v1_pb0087	60380	60841	automatic/finished	_	Transcriptional regulator, HTH family	Atu6098
SMEL_v1_pb0088	60838	61434	automatic/finished	_	conserved protein of unknown function	Atu6097
SMEL_v1_pb0089	61395	62555	automatic/finished	_	conserved protein of unknown function	Atu6096
SMEL_v1_pb0090	62527	62703	automatic/finished	_	conserved protein of unknown function	AERS4p2_0105
SMEL_v1_pb0093	67107	67766	automatic/finished	_	DNA polymerase IV (fragment)	No Hit
SMEL_v1_pb0109	75316	76635	automatic/finished	repC	putative replication protein C	Atu6045
SMEL_v1_pb0110	76873	77883	automatic/finished	_	putative replication protein B	Atu6044
SMEL_v1_pb0111	78107	79327	automatic/finished	repA	putative replication protein A	Atu6043
SMEL_v1_pb0112	79690	80319	automatic/finished	traI	Acyl-homoserine-lactone synthase	Atu6042
SMEL_v1_pb0113	80316	81287	automatic/finished	trbB	Conjugal transfer protein TrbB	Atu6041
SMEL_v1_pb0115	81677	81976	automatic/finished	trbD	Conjugal transfer protein TrbD	Atu6039
SMEL_v1_pb0116	81987	84455	automatic/finished	trbE	Conjugal transfer protein TrbE	Atu6038
SMEL_v1_pb0117	84427	85236	automatic/finished	trbJ	Conjugal transfer protein TrbJ	Atu6037

SMEL_v1_pb0118	85233	85451	automatic/finished	trbK	Conjugal transfer protein TrbK	Atu6036
SMEL_v1_pb0119	85445	86641	automatic/finished	trbL	Conjugal transfer protein TrbL	Atu6035
SMEL_v1_pb0120	86656	87318	automatic/finished	trbF	Conjugal transfer protein TrbF	Atu6034
SMEL_v1_pb0122	87336	88190	automatic/finished	trbG	Conjugal transfer protein TrbG	Atu6033
SMEL_v1_pb0123	88190	88666	automatic/finished	trbH	Conjugal transfer protein TrbH	Atu6032
SMEL_v1_pb0124	88679	89986	automatic/finished	trbI	Conjugal transfer protein TrbI	Atu6031
SMEL_v1_pb0125	90015	90194	automatic/finished	_	conserved protein of unknown function	Atu6030
SMEL_v1_pb0126	90345	91247	automatic/finished	nocR	Regulatory protein NocR	Atu6029
SMEL_v1_pb0127	91500	92273	automatic/finished	hisP	histidine/lysine/arginine/ornithine transporter subunit ; ATP-binding component of ABC superfamily	Atu6028
SMEL_v1_pb0128	92339	93190	automatic/finished	nocT	Nopaline-binding periplasmic protein	Atu6027
SMEL_v1_pb0129	93218	93973	automatic/finished	nocQ	Nopaline transport system permease protein NocQ	Atu6026
SMEL_v1_pb0130	93975	94700	automatic/finished	nocM	Nopaline transport system permease protein NocM	Atu6025
SMEL_v1_pb0131	94814	95734	automatic/finished	_	conserved protein of unknown function	Atu6024
SMEL_v1_pb0132	95828	97807	automatic/finished	_	Hydantoin utilization protein B	Atu6023
SMEL_v1_pb0133	97811	99847	automatic/finished	_	Hydantoin utilization protein A	Atu6022

Supplemental Table 3.7 Similarity of replicons (chromosome, pSymA and pSymB) in each strain relative to the reference strain Rm2011.

Chromosome – RM2011

Strain	Gene Number in Common with Rm2011	Percent Similarity	Number syntons	Relative coding sequences		
				Minimun	Average	Maximum
USDA1106	3522	97.21	1	3522	3522	3522
Rm1021	3458	95.45	1	3458	3458	3458
BL225C	3450	95.22	134	1	27.6	1263
T073	3441	94.98	164	1	22.71	1249
GR4	3422	94.45	132	1	27.92	1895
AK83	3422	94.45	127	1	28.78	822
Sm11	3420	94.4	152	1	24.34	734
Kh46c	3418	94.34	160	1	23.32	1525
USDA1157	3412	94.18	164	1	22.73	1099
HM006	3395	93.71	166	1	22.42	808
M162	3393	93.65	166	1	22.2	890
Rm41	3360	92.74	157	1	23.18	554
Kh35C	3304	91.2	155	1	23.16	880
USDA1021	3142	86.72	148	1	22.94	1050
M270	3016	83.25	161	1	20.44	499

pSymA –RM2011

Strain	Gene Number in Common with Rm2011	Percent Similarity	Number syntons	Relative coding sequences		
				Minimun	Average	Maximum
Rm1021	1493	90.27	1	1493	1493	1493
BL225C	1242	75.09	183	1	8.51	252
Rm41	1220	73.76	164	1	9.36	318
Sm11	1217	73.58	203	1	7.77	183
Kh46c	1216	73.52	183	1	8.41	249
USDA1021	1203	72.73	176	1	8.66	359
USDA1157	1177	71.16	163	1	8.91	216
GR4	1135	68.62	160	1	8.63	270
HM006	1081	65.36	193	1	7.15	175
M270	1038	62.76	215	1	6.2	179
AK83	1017	61.49	196	1	6.64	250
Kh35c	1007	60.88	166	1	7.61	141
T073	985	59.55	182	1	6.78	179
M162	739	44.68	143	1	6.23	183

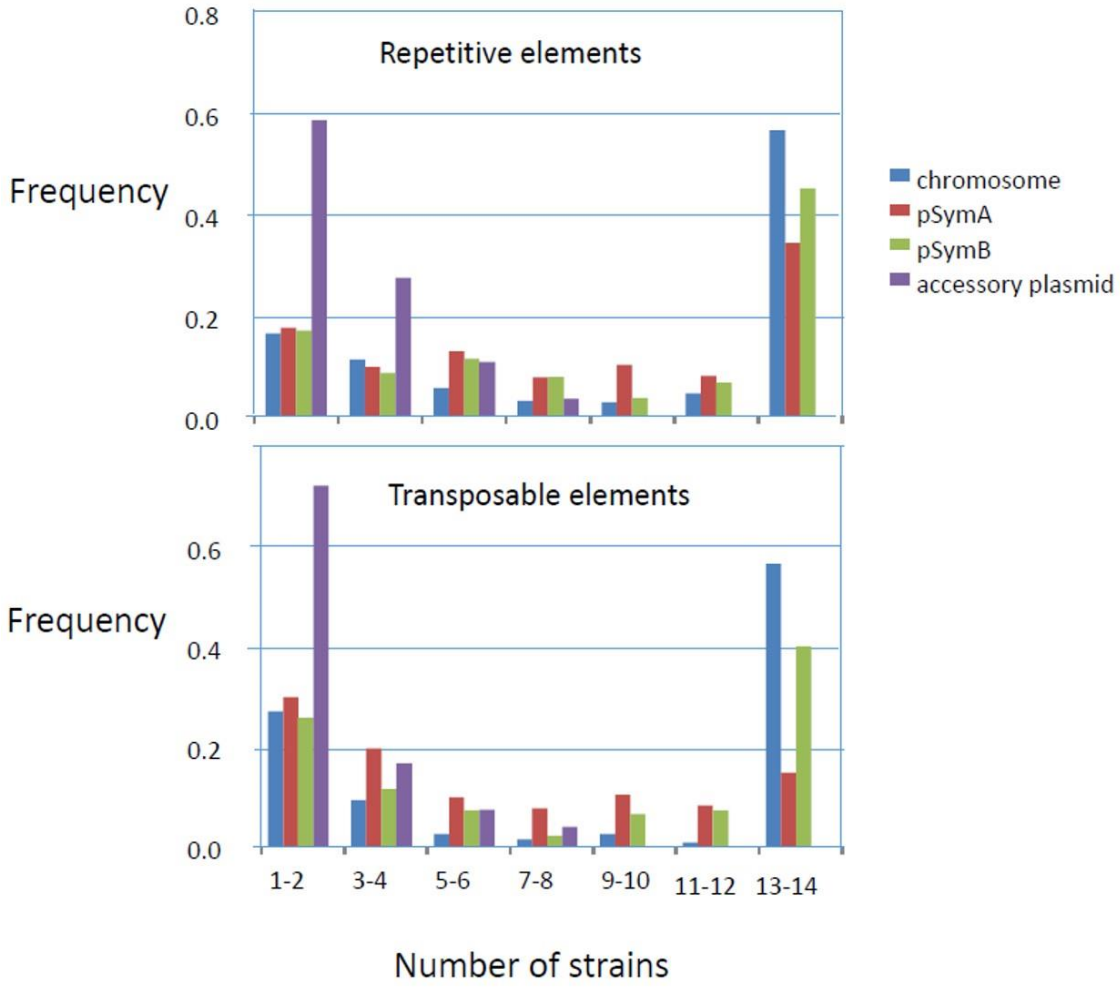
pSymB –RM2011

Strain	Gene Number in Common with Rm2011	Percent Similarity	Number syntons	Relative coding sequences		
				Minimun	Average	Maximum
USDA1106	1643	95.36	1	1643	1643	1643
Rm1021	1642	95.3	1	1642	1642	1642
GR4	1565	90.83	104	1	16.81	578
BL225C	1550	89.96	85	1	20.02	627
Kh46c	1516	87.99	97	1	17.3	530
AK83	1510	87.64	122	1	14.05	331
Sm11	1510	87.64	97	1	17.18	540
T073	1505	87.35	94	1	17.64	536
Kh35c	1504	87.29	96	1	17.44	530
USDA1157	1501	87.12	104	1	16.03	466
M162	1500	87.06	114	1	14.74	532
USDA1021	1494	86.71	143	1	12.16	527
M270	1487	86.3	117	1	14.44	417
HM006	1487	86.3	97	1	17.01	533
Rm41	1484	86.13	121	1	13.93	320

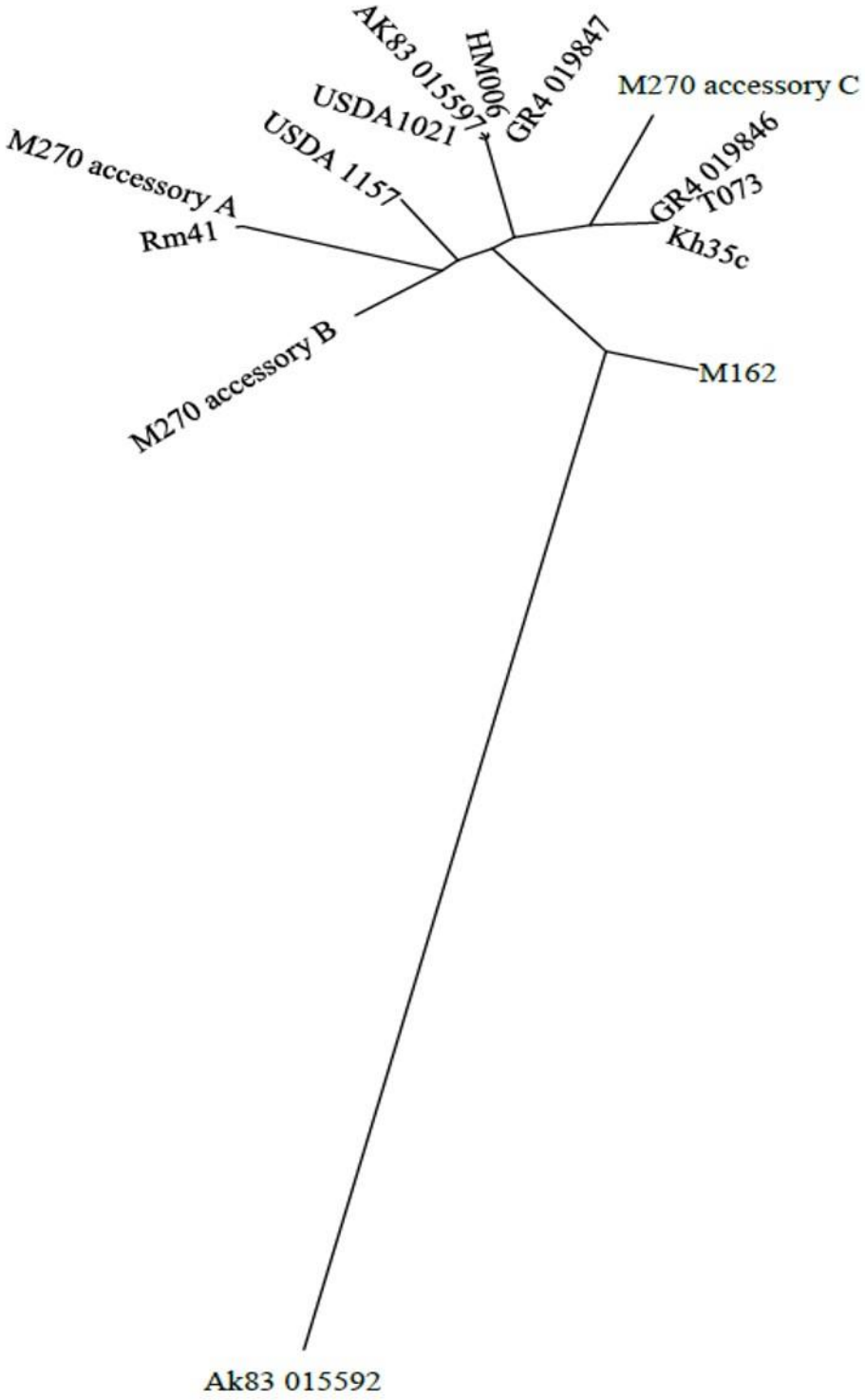
Supplemental Table 3.8: Results of the Mantel test to examine correlation between genetic distance matrices calculated from different replicons.

Replicons	r	p-value
Chr / pSymB	-0.18	0.85
Chr / pSymA	-0.09	0.5
pSymB / pSymA	0.06	0.34

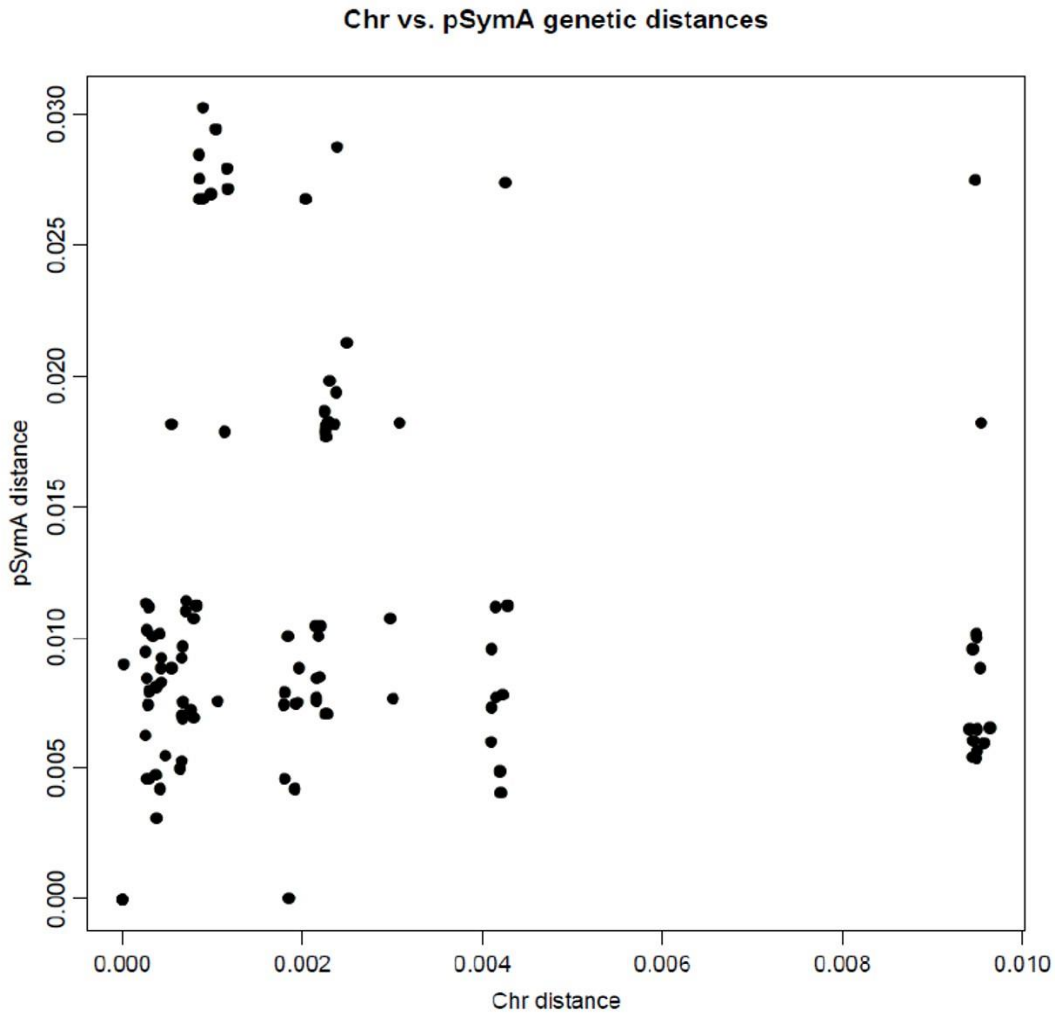
Supplemental Figure 3.1: The frequency of occurrence (number of strains) in which repetitive and transposable elements are found in different numbers of strains. The majority of REs and TEs found on accessory plasmids are found on only a single strain whereas a majority or plurality of REs and TEs found on the chromosome or pSymB are found in all strains. Because of large translocations between plasmids from two strains (USDA1021 and M162) were excluded from analyses leaving a total of 14 primary replicons and 11 accessory plasmids.

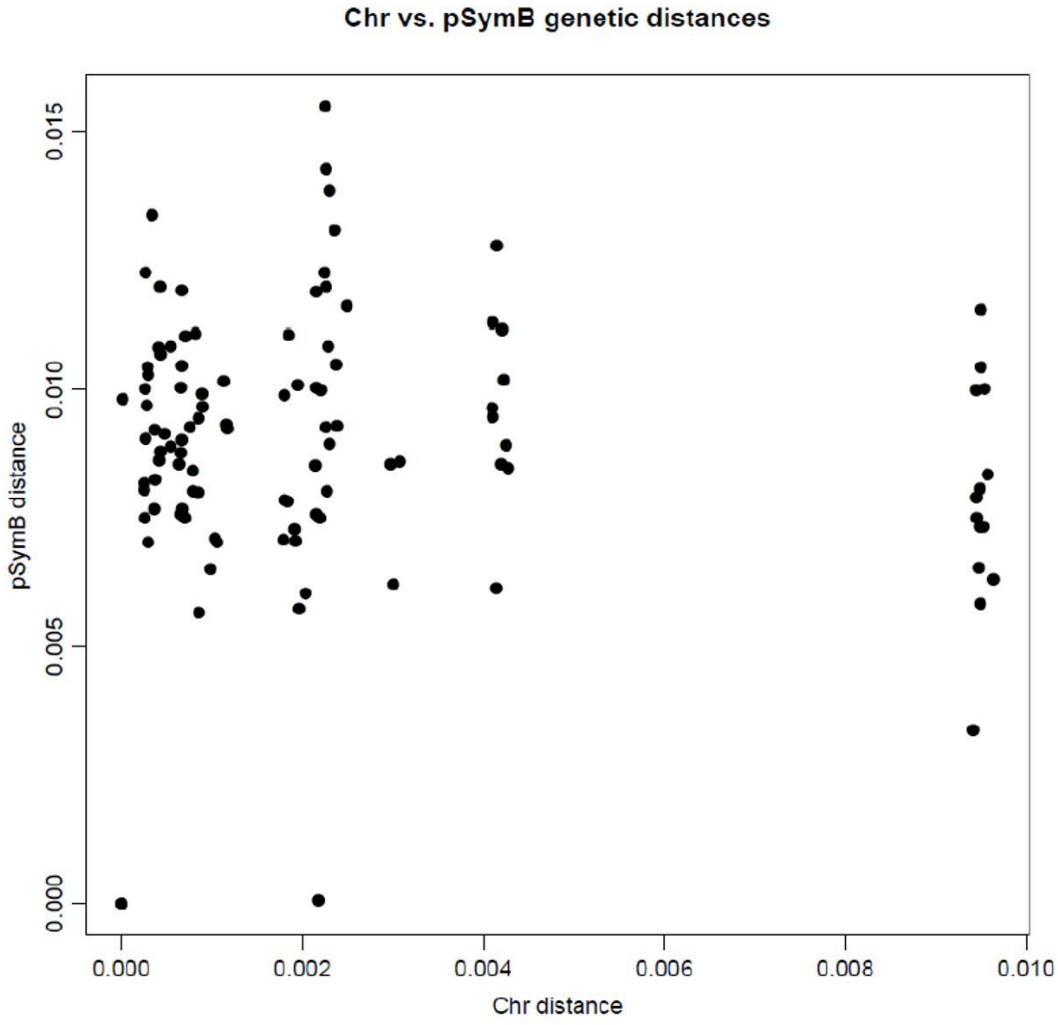


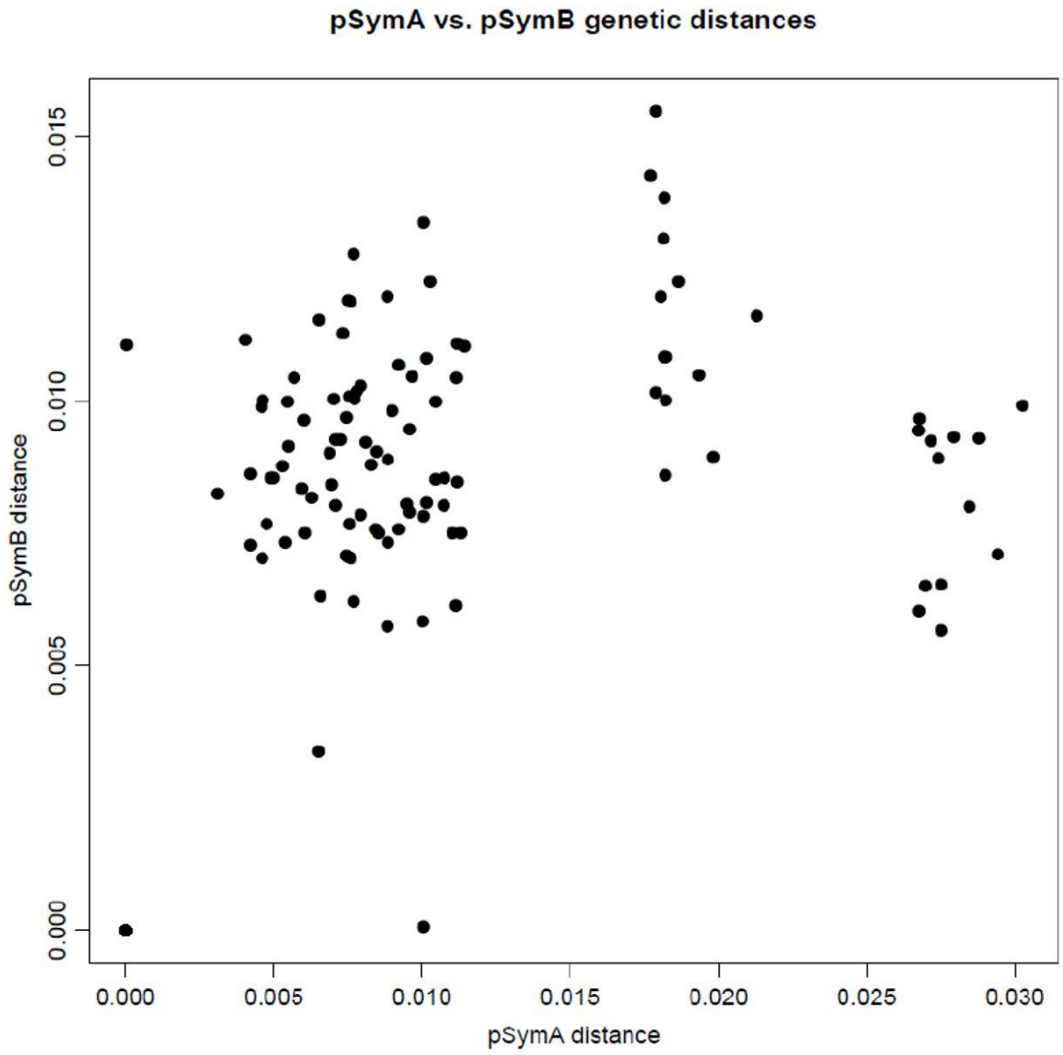
Supplemental Figure 3.2: Neighbor joining unrooted phylogenetic tree of all RepA proteins found on the accessory plasmids.



Supplemental Figure 3.3: Scatterplot showing pairwise comparisons between individual replicons based on genetic distance.




Chr vs. pSymB genetic distances
pSymB distance
0.000
0.005
0.010
0.015
0.000
0.002
0.004
0.006
0.008
0.010
Chr distance


pSymA vs. pSymB genetic distances
pSymB distance
0.000
0.005
0.010
0.015
pSymA distance
0.000
0.005
0.010
0.015
0.020
0.025
0.030

Concluding Remarks

Symbiosis research between rhizobia and legumes has been an important area of study for many decades. This had led to many insights about the complex interaction between rhizobia and legumes, including signal and nutrient exchange. Recent research has shown the T3SS, T4SS, and T6SS all transport effector proteins which alters the symbiosis between rhizobia and legumes. Pathogens have long been known to use effector proteins during invasion to promote virulence. Understanding effector protein effect on symbiosis and *in planta* function can offer critical insight into rhizobia/legume symbiosis. Here I review my key findings and suggest future work.

In Chapter 1, I demonstrate the T4SS increases nodule number and strain competitiveness. I tested candidate proteins to identify effector proteins in *Sinorhizobium*. In *S. meliloti* and *S. medicae* have only one effector protein, TfeA, which translocates through the T4SS. TfeA is induced during symbiosis, increases nodule number and strain competitiveness. TfeA-like proteins are found in 12 different Rhizobiaceae species, suggesting a conserved mechanism across rhizobia to alter plant host cells during nodule formation.

In Chapter 2, I demonstrate the *Sinorhizobium* TfeA binds to *M. truncatula* ARF2 both in yeast and *in planta*. TfeA also binds to *G. max* ARF27. TfeA, through interaction with ARF2, likely prevents constitutive expression in transgenic *Medicago*. *TfeA* was successfully transformed and expressed in *M. truncatula* under an inducible promoter. This offers critical insight into the role of effector proteins, by suggesting TfeA could overcome AON through directly initiating an auxin response.

In Chapter 3, I sequenced 10 *S. meliloti* genomes, and analyze 16 genomes for comparison. The chromosome, pSymA and pSymB appear to have evolved and function independently. Gene transfer events between accessory plasmids and pSymA demonstrates the mechanism(s) by which the *Sinorhizobium* symbiosis is constantly evolving. The core-replicon of pSymA is significantly smaller than pSymB or the chromosome. This suggests the majority of pSymA genes are not required for symbiosis, and instead likely modulate symbiosis. This includes genes involved in the T4SS and *tfeA*. Inter-replicon gene movement is rare, but intra-replicon gene movement is common. Inter-replicon gene transfer is associated with gene duplication rather than TEs. Intra-replicon gene transfer is associated with REs not TEs.

Future research can focus on optimizing *tfeA* induction by estrogen and further characterizing the phenotype of *tfeA* expression, both with low and high auxin conditions. Research is currently being done to identify a minimum gene set for symbiosis based limited inter-replicon gene transfer. Genetic analysis and gene knockouts of accessory plasmids could offer further insight into *Sinorhizobium/Medicago* symbiosis. Accessory plasmids along with the majority of genes on pSymA are not required for symbiosis formation. However, many of these genes could be involved in modulating the symbiosis, like the T4SS.

Bibliography

Agrawal, A. A. 2001. Phenotypic Plasticity in the Interactions and Evolution of Species. Science (80-.). 294:321–326

Anderson, D. M., and Frank, D. W. 2012. Five Mechanisms of Manipulation by Bacterial Effectors: A Ubiquitous Theme. PLoS Pathog. 8:1–4

Angot, A., Peeters, N., Lechner, E., Vailleau, F., Baud, C., Gentzbittel, L., Sartorel, E., Genschik, P., Boucher, C., and Genin, S. 2006. Ralstonia solanacearum requires F-box-like domain-containing type III effectors to promote disease on several host plants. Proc. Natl. Acad. Sci. U. S. A. 103:14620–14625

Austin, S., Bingham, E. T., Mathews, D. E., Shahan, M. N., Will, J., and Burgess, R. R. 1995. Production and field performance of transgenic alfalfa (Medicago sativa L.) expressing alpha-amylase and manganese-dependent lignin peroxidase. Euphytica. 85:381–393

Barnett, M. J., and Kahn, M. L. 2005. Psyma of Sinorhizobium Meliloti: Nitrogen Fixation and More. Pages 113–132in:Genomes and Genomics of Nitrogen-fixing Organisms,Springer-Verlag, Berlin/Heidelberg.

Barnett, M. J., Toman, C. J., Fisher, R. F., and Long, S. R. 2004. A dual-genome

Symbiosis Chip for coordinate study of signal exchange and development in a prokaryote-host interaction. Proc. Natl. Acad. Sci. 101:16636–16641

Bartsev, A. V., Boukli, N. M., Deakin, W. J., Staehelin, C., and Broughton, W. J. 2003. Purification and phosphorylation of the effector protein NopL from Rhizobium sp. NGR234. FEBS Lett. 554:271–274

Bartsev, A. V., Deakin, W. J., Boukli, N. M., McAlvin, C. B., Stacey, G., Malnoë, P., Broughton, W. J., and Staehelin, C. 2004. NopL, an Effector Protein of Rhizobium sp. NGR234, Thwarts Activation of Plant Defense Reactions. PLANT Physiol. 134:871–879

Beijersbergen, A., Dulk-Ras, A. D., Schilperoort, R. A., and Hooykaas, P. J. J. 1992. Conjugative Transfer by the Virulence System of Agrobacterium tumefaciens. Science (80-.). 256:1324–1327

Belkhadir, Y., Subramaniam, R., and Dangl, J. L. 2004. Plant disease resistance protein signaling: NBS–LRR proteins and their partners. Curr. Opin. Plant Biol. 7:391–399

Berger, B. R., and Christie, P. J. 1994. Genetic complementation analysis of the Agrobacterium tumefaciens virB operon: virB2 through virB11 are essential virulence genes. J. Bacteriol. 176:3646–3660

Beringer, J. E. 1974. R Factor Transfer in Rhizobium leguminosarum. Microbiology. 84:188–198

Bingham, G. J. 1991. Registration of Alfalfa Hybrid Regen-Sy Germplasm for Tissue Culture and Transformation Research. Crop Sci. 31:1098

Bingle, L. E., Bailey, C. M., and Pallen, M. J. 2008. Type VI secretion: a beginner's

guide. Curr. Opin. Microbiol. 11:3–8

Bladergroen, M. R., Badelt, K., and Spaink, H. P. 2003. Infection-blocking genes of a symbiotic Rhizobium leguminosarum strain that are involved in temperature-dependent protein secretion. Mol. Plant. Microbe. Interact. 16:53–64

Boto, L. 2010. Horizontal gene transfer in evolution: facts and challenges. Proc. Biol. Sci. 277:819–27

Broughton, W. J., Hanin, M., Relić, B., Kopciñska, J., Golinowski, W., Şimşek, Ş., Ojanen-Reuhs, T., Reuhs, B., Marie, C., Kobayashi, H., Bordogna, B., Le Quéré, a., Jabbouri, S., Fellay, R., Perret, X., and Deakin, W. J. 2006. Flavonoid-inducible modifications to rhamnan O antigens are necessary for Rhizobium sp. strain NGR234-legume symbioses. J. Bacteriol. 188:3654–3663

Bucciarelli, B., Hanan, J., Palmquist, D., and Vance, C. P. 2006. A Standardized Method for Analysis of Medicago truncatula Phenotypic Development. Plant Physiol. 142:207–219

Calderón Villalobos, L. I. A., Lee, S., De Oliveira, C., Ivetac, A., Brandt, W., Armitage, L., Sheard, L. B., Tan, X., Parry, G., Mao, H., Zheng, N., Napier, R., Kepinski, S., and Estelle, M. 2012. A combinatorial TIR1/AFB–Aux/IAA co-receptor system for differential sensing of auxin. Nat. Chem. Biol. 8:477–485

Canonne, J., Marino, D., Jauneau, A., Pouzet, C., Briere, C., Roby, D., and Rivas, S. 2011. The Xanthomonas Type III Effector XopD Targets the Arabidopsis Transcription Factor MYB30 to Suppress Plant Defense. PLANT CELL ONLINE. 23:3498–3511

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25:1972–1973

Cascales, E., and Christie, P. J. 2003. The versatile bacterial type IV secretion systems. Nat. Rev. Microbiol. 1:137–49

Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E. E., Turner, S. W., and Korlach, J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods. 10:563–9

Cosson, V., Eschstruth, A., and Ratet, P. 2015. Medicago truncatula Transformation Using Leaf Explants BT - Agrobacterium Protocols: Volume 1. Pages 43–56in:K. Wang, ed.Springer New York, New York, NY.

Crook, M. B., Lindsay, D. P., Biggs, M. B., Bentley, J. S., Price, J. C., Clement, S. C., Clement, M. J., Long, S. R., and Griffitts, J. S. 2012. Rhizobial plasmids that cause impaired symbiotic nitrogen fixation and enhanced host invasion. Mol. Plant. Microbe. Interact. 25:1026–33

Dai, W. J., Zeng, Y., Xie, Z. P., and Staehelin, C. 2008. Symbiosis-promoting and deleterious effects of NopT, a novel type 3 effector of Rhizobium sp. strain NGR234. J. Bacteriol. 190:5101–5110

Darmon, E., and Leach, D. R. F. 2014. Bacterial Genome Instability. Microbiol. Mol. Biol. Rev. 78:1–39

Daubin, V., and Szöllősi, G. J. 2016. Horizontal Gene Transfer and the History of Life.

Cold Spring Harb. Perspect. Biol. 8:a018036

Davison, J. 1999. Genetic Exchange between Bacteria in the Environment. Plasmid. 42:73–91

Deakin, W. J., and Broughton, W. J. 2009. Symbiotic use of pathogenic strategies: rhizobial protein secretion systems. Nat. Rev. Microbiol. 7:312–20

Deakin, W. J., Marie, C., Saad, M. M., Krishnan, H. B., and Broughton, W. J. 2005. NopA is associated with cell surface appendages produced by the type III secretion system of Rhizobium sp. strain NGR234. Mol. Plant. Microbe. Interact. 18:499–507

Dénarié, J., Debellé, F., and Promé, J. C. 1996. Rhizobium lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. Annu. Rev. Biochem. 65:503–535

Deslandes, L., and Rivas, S. 2012. Catch me if you can: Bacterial effectors and plant targets. Trends Plant Sci. 17:644–655

diCenzo, G. C., Checcucci, A., Bazzicalupo, M., Mengoni, A., Viti, C., Dziewit, L., Finan, T. M., Galardini, M., and Fondi, M. 2016. Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in Sinorhizobium meliloti. Nat. Commun. 7:12219

Dillon, M. M., Sung, W., Lynch, M., and Cooper, V. S. 2015. The Rate and Molecular Spectrum of Spontaneous Mutations in the GC-Rich Multichromosome Genome of Burkholderia cenocepacia. Genetics. 200:935–946

Dodds, P. N., and Rathjen, J. P. 2010. Plant immunity: towards an integrated view of plant-pathogen interactions. Nat. Rev. Genet. 11:539–548

Dray, S. and Dufour, A.B. and Chessel, D. 2007. The ade4 Package--II: Two-table and K-table Methods. R News. 7:47–52

Ellis, C. M., Nagpal, P., Young, J. C., Hagen, G., Guilfoyle, T. J., and Reed, J. W. 2005. AUXIN RESPONSE FACTOR1 and AUXIN RESPONSE FACTOR2 regulate senescence and floral organ abscission in Arabidopsis thaliana. Development. 132:4563–4574

Ellis, J. G., and Murphy, P. J. 1981. Four new opines from crown gall tumours—their detection and properties. Mol. Gen. Genet. MGG. 181:36–43

Endre, G., Kereszt, A., Kevei, Z., Mihacea, S., Kalo, P., and Kiss, G. B. 2002. A receptor kinase gene regulating symbiotic nodule development. Nature. 417:962–966

Enright, A. J., Van Dongen, S., and Ouzounis, C. A. 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30:1575–84

Epstein, B., Branca, A., Mudge, J., Bharti, A. K., Briskine, R., Farmer, A. D., Sugawara, M., Young, N. D., Sadowsky, M. J., and Tiffin, P. 2012. Population Genomics of the Facultatively Mutualistic Bacteria Sinorhizobium meliloti and S. medicae N.A. Moran, ed. PLoS Genet. 8:e1002868

Epstein, B., Sadowsky, M. J., and Tiffin, P. 2014. Selection on Horizontally Transferred and Duplicated Genes in Sinorhizobium (Ensifer), the Root-Nodule Symbionts of Medicago. Genome Biol. Evol. 6:1199–1209

Esseling, J. J., Lhuissier, F. G. P., and Emons, A. M. C. 2003. Nod factor-induced root hair curling: continuous polar growth towards the point of nod factor application. Plant Physiol. 132:1982–1988

Farrand, S. K. 2003. Agrobacterium is a definable genus of the family Rhizobiaceae. Int. J. Syst. Evol. Microbiol. 53:1681–1687

Fauvart, M., and Michiels, J. 2008. Rhizobial secreted proteins as determinants of host specificity in the rhizobium-legume symbiosis. FEMS Microbiol. Lett. 285:1–9

Ferguson, B. J., Indrasumunar, A., Hayashi, S., Lin, M. H., Lin, Y. H., Reid, D. E., and Gresshoff, P. M. 2010. Molecular analysis of legume nodule development and autoregulation. J. Integr. Plant Biol. 52:61–76

Finan, T. M., Weidner, S., Wong, K., Buhrmester, J., Chain, P., Vorholter, F. J., Hernandez-Lucas, I., Becker, A., Cowie, A., Gouzy, J., Golding, B., and Puhler, A. 2001. The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endosymbiont Sinorhizobium meliloti. Proc. Natl. Acad. Sci. 98:9889–9894

Finn, R. D., Clements, J., and Eddy, S. R. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. :gkr367

Fotiadis, C. T., Dimou, M., Georgakopoulos, D. G., Katinakis, P., and Tampakaki, A. P. 2012. Functional characterization of NopT1 and NopT2, two type III effectors of Bradyrhizobium japonicum. FEMS Microbiol. Lett. 327:66–77

Fournier, J., Timmers, A. C. J., Sieberer, B. J., Jauneau, A., Chabaud, M., and Barker, D. G. 2008. Mechanism of infection thread elongation in root hairs of Medicago truncatula and dynamic interplay with associated rhizobial colonization. Plant Physiol. 148:1985–1995

Frost, L. S., Leplae, R., Summers, A. O., and Toussaint, A. 2005. Mobile genetic elements: the agents of open source evolution. Nat. Rev. Microbiol. 3:722–732

Gage, D. J. 2004. Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. Microbiol. Mol. Biol. Rev. 68:280–300

Galardini, M., Mengoni, A., Brilli, M., Pini, F., Fioravanti, A., Lucas, S., Lapidus, A., Cheng, J.-F., Goodwin, L., Pitluck, S., Land, M., Hauser, L., Woike, T., Mikhailova, N., Ivanova, N., Daligault, H., Bruce, D., Detter, C., Tapia, R., Han, C., Teshima, H., Mocali, S., Bazzicalupo, M., Biondi, E. G., Woyke, T., Mikhailova, N., Ivanova, N., Daligault, H., Bruce, D., Detter, C., Tapia, R., Han, C., Teshima, H., Mocali, S., Bazzicalupo, M., and Biondi, E. G. 2011. Exploring the symbiotic pangenome of the nitrogen-fixing bacterium Sinorhizobium meliloti. BMC Genomics. 12:235

Galardini, M., Pini, F., Bazzicalupo, M., Biondi, E. G., and Mengoni, A. 2013. Replicon-Dependent Bacterial Genome Evolution: The Case of Sinorhizobium meliloti. Genome Biol. Evol. 5:542–558

Galibert, F., Finan, T. M., Long, S. R., Pühler, A., Abola, P., Ampe, F., Barloy-Hubler, F., Barnett, M. J., Becker, A., Boistard, P., Bothe, G., Boutry, M., Bowser, L., Buhrmester, J., Cadieu, E., Capela, D., Chain, P., Cowie, A., Davis, R. W., Dréano, S., Federspiel, N. A., Fisher, R. F., Gloux, S., Godrie, T., Goffeau, A., Golding, B., Gouzy, J., Gurjal, M., Hernandez-Lucas, I., Hong, A., Huizar, L., Hyman, R. W., Jones, T., Kahn, D., Kahn, M. L., Kalman, S., Keating, D. H., Kiss, E., Komp, C., Lelaure, V., Masuy, D., Palm, C., Peck, M. C., Pohl, T. M., Portetelle, D., Purnelle, B., Ramsperger, U., Surzycki, R., Thébault, P., Vandenbol, M., Vorhölter, F.-J., Weidner, S., Wells, D. H., Wong, K., Yeh, K.-C., and Batut, J. 2001. The Composite Genome of the Legume Symbiont Sinorhizobium meliloti. Science (80-.). 293:668–

672

Gazi, A. D., Sarris, P. F., Fadouloglou, V. E., Charova, S. N., Mathioudakis, N., Panopoulos, N. J., and Kokkinidis, M. 2012. Phylogenetic analysis of a gene cluster encoding an additional, rhizobial-like type III secretion system that is narrowly distributed among Pseudomonas syringae strains. BMC Microbiol. 12:188

Ghosh, P. 2004. Process of Protein Transport by the Type III Secretion System. Society. 68:771–795

Giraud, E., Moulin, L., Vallenet, D., Barbe, V., Cytryn, E., Avarre, J.-C., Jaubert, M., Simon, D., Cartieaux, F., Prin, Y., Bena, G., Hannibal, L., Fardoux, J., Kojadinovic, M., Vuillet, L., Lajus, A., Cruveiller, S., Rouy, Z., Mangenot, S., Segurens, B., Dossat, C., Franck, W. L., Chang, W.-S., Saunders, E., Bruce, D., Richardson, P., Normand, P., Dreyfus, B., Pignol, D., Stacey, G., Emerich, D., Verméglio, A., Médigue, C., and Sadowsky, M. 2007. Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia. Science. 316:1307–1312

Goethals, K., Van Montagu, M., and Holsters, M. 1992. Conserved motifs in a divergent nod box of Azorhizobium caulinodans ORS571 reveal a common structure in promoters regulated by LysR-type proteins. Proc. Natl. Acad. Sci. U. S. A. 89:1646–1650

Göhre, V., Spallek, T., Häweker, H., Mersmann, S., Mentzel, T., Boller, T., de Torres, M., Mansfield, J. W., and Robatzek, S. 2008. Plant Pattern-Recognition Receptor FLS2 Is Directed for Degradation by the Bacterial Ubiquitin Ligase AvrPtoB. Curr. Biol. 18:1824–1832

González, E. M., Aparicio-tejo, P. M., Gordon, A. J., Minchin, F. R., Royuela, M., and Arrese-igor, C. 1998. Water-deficit effects on carbon and nitrogen metabolism of pea nodules. J. Exp. Bot. 49:1705–1714

Goormachtig, S., Capoen, W., and Holsters, M. 2004. Rhizobium infection: Lessons from the versatile nodulation behaviour of water-tolerant legumes. Trends Plant Sci. 9:518–522

Guilfoyle, T. J., and Hagen, G. 2007. Auxin response factors. Curr. Opin. Plant Biol. 10:453–460

Guilfoyle, T. J., and Hagen, G. 2012. Getting a grasp on domain III/IV responsible for Auxin Response Factor–IAA protein interactions. Plant Sci. 190:82–88

Guo, X., Flores, M., Mavingui, P., Fuentes, S. I., Hernández, G., Dávila, G., and Palacios, R. 2003. Natural genomic design in Sinorhizobium meliloti: novel genomic architectures. Genome Res. 13:1810–7

Van Ha, C., Le, D. T., Nishiyama, R., Watanabe, Y., Sulieman, S., Tran, U. T., Mochida, K., Van Dong, N., Yamaguchi-Shinozaki, K., Shinozaki, K., and Tran, L.-S. P. 2013. The Auxin Response Factor Transcription Factor Family in Soybean: Genome-Wide Identification and Expression Analyses During Development and Water Stress. DNA Res. 20:511–524

Handberg, K., and Stougaard, J. 1992. Lotus japonicus, an autogamous, diploid legume species for classical and molecular genetics. Plant J. 2:487–496

Hansen, G., Das, A., and Chilton, M. D. 1994. Constitutive expression of the virulence genes improves the efficiency of plant transformation by Agrobacterium. Proc. Natl.

Acad. Sci. . 91:7603–7607

Harrison, P. W., Lower, R. P. J., Kim, N. K. D., and Young, J. P. W. 2010. Introducing the bacterial "chromid": not a chromosome, not a plasmid. Trends Microbiol. 18:141–148

Hassan, S., and Mathesius, U. 2012. The role of flavonoids in root-rhizosphere signalling: Opportunities and challenges for improving plant-microbe interactions. J. Exp. Bot. 63:3429–44

Hernández-Jiménez, M. J., Mercedes Lucas, M., and de Felipe, M. R. 2002. Antioxidant defence and damage in senescing lupin nodules. Plant Physiol. Biochem. 40:645–657

Herridge, D., and Rose, I. 2000. Breeding for enhanced nitrogen fixation in crop legumes. F. Crop. Res. 65:229–248

Herring, C. D., and Palsson, B. Ø. 2007. An evaluation of Comparative Genome Sequencing (CGS) by comparing two previously-sequenced bacterial genomes. BMC Genomics. 8:274

Hiei, Y., Ohta, S., Komari, T., and Kumashiro, T. 1994. Efficient transformation of rice (Oryza sativa L.) mediated by Agrobacterium and sequence analysis of the boundaries of the T-DNA. Plant J. 6:271–282

Hodges, L. D., Vergunst, A. C., Neal-McKinney, J., Den Dulk-Ras, A., Moyer, D. M., Hooykaas, P. J. J., and Ream, W. 2006. Agrobacterium rhizogenes GALLS protein contains domains for ATP binding, nuclear localization, and type IV secretion. J. Bacteriol. 188:8222–30

Hooykaas, P. J. J., Klapwijk, P. M., Nuti, M. P., Schilperoort, R. A., and Rörsch, A. 1977. Transfer of the Agrobacterium tumefaciens Ti plasmid to avirulent agrobacteria and to Rhizobium ex planta. Microbiology. 98:477–484

Hotson, A., Chosed, R., Shu, H., Orth, K., and Mudgett, M. B. 2003. Xanthomonas type III effector XopD targets SUMO-conjugated proteins in planta. Mol. Microbiol. 50:377–389

Howieson, J. G. 1995. Rhizobial persistence and its role in the development of sustainable agricultural systems in Mediterranean environments. Soil Biol. Biochem. 27:603–610

Hubber, A. M., Sullivan, J. T., and Ronson, C. W. 2007. Symbiosis-induced cascade regulation of the Mesorhizobium loti R7A VirB/D4 type IV secretion system. Mol. Plant. Microbe. Interact. 20:255–61

Hubber, A., Vergunst, A. C., Sullivan, J. T., Hooykaas, P. J. J., and Ronson, C. W. 2004. Symbiotic phenotypes and translocated effector proteins of the Mesorhizobium loti strain R7A VirB/D4 type IV secretion system. Mol. Microbiol. 54:561–574

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., Rattei, T., Mende, D. R., Sunagawa, S., Kuhn, M., Jensen, L. J., von Mering, C., and Bork, P. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 44:D286–D293

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

Igarashi, D., Bethke, G., Xu, Y., Tsuda, K., Glazebrook, J., and Katagiri, F. 2013. Pattern-Triggered Immunity Suppresses Programmed Cell Death Triggered by Fumonisin B1 E.J.M. van Damme, ed. PLoS One. 8:e60769

James, B. 2013. The Development and Regulation of Soybean Nodules. in: A Comprehensive Survey of International Soybean Research - Genetics, Physiology, Agronomy and Nitrogen Relationships, InTech.

Jiménez-Guerrero, I., Pérez-Montaño, F., Monreal, J. A., Preston, G. M., Fones, H., Vioque, B., Ollero, F. J., and López-Baena, F. J. 2015. The Sinorhizobium (Ensifer) fredii HH103 Type 3 Secretion System Suppresses Early Defense Responses to Effectively Nodulate Soybean. Mol. Plant-Microbe Interact. 28:790–799

Jones, J. D. G., and Dangl, J. L. 2006. The plant immune system. Nature. 444:323–329

Jones, K. M., Kobayashi, H., Davies, B. W., Taga, M. E., and Walker, G. C. 2007a. How rhizobial symbionts invade plants: the Sinorhizobium-Medicago model. Nat. Rev. Microbiol. 5:619–33

Jones, K. M., Lloret, J., Daniele, J. R., and Walker, G. C. 2007b. The Type IV Secretion System of Sinorhizobium meliloti Strain 1021 Is Required for Conjugation but Not for Intracellular Symbiosis. J. Bacteriol. 189:2133–2138

Kambara, K., Ardissone, S., Kobayashi, H., Saad, M. M., Schumpp, O., Broughton, W. J., and Deakin, W. J. 2009. Rhizobia utilize pathogen-like effector proteins during symbiosis. Mol. Microbiol. 71:92–106

Kassaw, T., Jr, W. B., and Frugoli, J. 2015. Multiple autoregulation of nodulation (AON)

signals identified through split root analysis of Medicago truncatula sunn and rdn1 mutants. Plants. 4:209–224

Katoh, K., and Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780

Kiers, E. T., Rousseau, R. a, West, S. a, and Denison, R. F. 2003. Host sanctions and the legume-rhizobium mutualism. Nature. 425:78–81

Kimbrel, J. a, Thomas, W. J., Jiang, Y., Creason, A. L., Thireault, C. a, Sachs, J. L., and Chang, J. H. 2013. Mutualistic co-evolution of type III effector genes in Sinorhizobium fredii and Bradyrhizobium japonicum. PLoS Pathog. 9:e1003204

Kobayashi, H., Naciri-Graven, Y., Broughton, W. J., and Perret, X. 2004. Flavonoids induce temporal shifts in gene-expression of nod-box controlled loci in Rhizobium sp. NGR234. Mol. Microbiol. 51:335–347

Korasick, D. A., Westfall, C. S., Lee, S. G., Nanao, M. H., Dumas, R., Hagen, G., Guilfoyle, T. J., Jez, J. M., and Strader, L. C. 2014. Molecular basis for AUXIN RESPONSE FACTOR protein interaction and the control of auxin response repression. Proc. Natl. Acad. Sci. 111:5427–5432

Krause, A., Doerfel, A., and Göttfert, M. 2002. Mutational and transcriptional analysis of the type III secretion system of Bradyrhizobium japonicum. Mol. Plant. Microbe. Interact. 15:1228–35

Krishnan, H. B., Lorio, J., Kim, W. S., Jiang, G., Kim, K. Y., DeBoer, M., and Pueppke, S. G. 2003. Extracellular proteins involved in soybean cultivar-specific nodulation are associated with pilus-like surface appendages and exported by a type III protein

secretion system in Sinorhizobium fredii USDA257. Mol. Plant. Microbe. Interact. 16:617–625

Krumsiek, J., Arnold, R., and Rattei, T. 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics. 23:1026–1028

Krusell, L., Madsen, L. H., Sato, S., Aubert, G., Genua, A., Szczyglowski, K., Duc, G., Kaneko, T., Tabata, S., de Bruijn, F., Pajuelo, E., Sandal, N., and Stougaard, J. 2002. Shoot control of root development and nodulation is mediated by a receptor-like kinase. Nature. 420:422–426

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., and Marra, M. A. 2009. Circos: An information aesthetic for comparative genomics. Genome Res. 19:1639–1645

Kuhn, S., Stiens, M., PÃ¼hler, A., and SchlÃ¼ter, A. 2008. Prevalence of pSmeSM11a-like plasmids in indigenous Sinorhizobium meliloti strains isolated in the course of a field release experiment with genetically modified S. meliloti strains. FEMS Microbiol. Ecol. 63:118–131

Kuldau, G. a, De Vos, G., Owen, J., McCaffrey, G., and Zambryski, P. 1990. The virB operon of Agrobacterium tumefaciens pTiC58 encodes 11 open reading frames. Mol. Gen. Genet. 221:256–66

Kuppusamy, K. T. 2004. LIN, a Medicago truncatula Gene Required for Nodule Differentiation and Persistence of Rhizobial Infections. PLANT Physiol. 136:3682–3691

Kuppusamy, K. T., Ivashuta, S., Bucciarelli, B., Vance, C. P., Gantt, J. S., and

Vandenbosch, K. A. 2009. Knockdown of CELL DIVISION CYCLE16 reveals an inverse relationship between lateral root and nodule numbers and a link to auxin in Medicago truncatula. Plant Physiol. 151:1155–1166

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S. L. 2004. Versatile and open software for comparing large genomes. Genome Biol. 5:R12

Li, S.-B., Xie, Z.-Z., Hu, C.-G., and Zhang, J.-Z. 2016. A Review of Auxin Response Factors (ARFs) in Plants. Front. Plant Sci. 7:47

Lindeberg, M., Myers, C. R., Collmer, A., and Schneider, D. J. 2008. Roadmap to new virulence determinants in Pseudomonas syringae: insights from comparative genomics and genome organization. Mol. Plant. Microbe. Interact. 21:685–700

Lodwig, E. M., Hosie, A. H. F., Bourdes, A., Findlay, K., Allaway, D., Karunakaran, R., Downie, J. A., Poole, P. S., Table, S., Fig, S., Pcr, S. G., Prism, A. B. I., Information, S., We, A., Izawa, T., Molecular, P., and Areas, P. 2006. Amino-acid cycling drives nitrogen fixation in the legume – Rhizobium symbiosis. Nature. 422:722–726

Long, J. A., Woody, S., Poethig, S., Meyerowitz, E. M., and Barton, M. K. 2002. Transformation of shoots into roots in Arabidopsis embryos mutant at the TOPLESS locus. Development. 129:2797–806

Long, S. R. 1996. Rhizobium symbiosis: nod factors in perspective. Plant Cell. 8:1885–1898

López-Baena, F. J., Vinardell, J. M., Pérez-Montaño, F., Crespo-Rivas, J. C., Bellogín, R.

a., Espuny, M. D. R., and Ollero, F. J. 2008. Regulation and symbiotic significance of nodulation outer proteins secretion in Sinorhizobium fredii HH103. Microbiology. 154:1825–1836

Lorio, J. C., Kim, W. S., and Krishnan, H. B. 2004. NopB, a soybean cultivar-specificity protein from Sinorhizobium fredii USDA257, is a type III secreted protein. Mol. Plant. Microbe. Interact. 17:1259–1268

De Lyra, M. D. C. C. P., López-Baena, F. J., Madinabeitia, N., Vinardell, J. M., Espuny, M. D. R., Cubo, M. T., Bellogín, R. A., Ruiz-Sainz, J. E., and Ollero, F. J. 2006. Inactivation of the Sinorhizobium fredii HH103 rhcJ gene abolishes nodulation outer proteins (Nops) secretion and decreases the symbiotic capacity with soybean. Int. Microbiol. 9:125–133

Macho, A. P., and Zipfel, C. 2015. ScienceDirect Targeting of plant pattern recognition receptor- triggered immunity by bacterial type-III secretion system effectors. Curr. Opin. Microbiol. 23:14–22

Manen, D., Upegui-Gonzalez, L. C., and Caro, L. 1992. Monomers and dimers of the RepA protein in plasmid pSC101 replication: domains in RepA. Proc. Natl. Acad. Sci. U. S. A. 89:8923–7

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27:209–20

Mao, G., Turner, M., Yu, O., and Subramanian, S. 2013. miR393 and miR164 influence indeterminate but not determinate nodule development. Plant Signal. Behav. 8:e26753

Marchetti, M., Capela, D., Glew, M., Cruveiller, S., Chane-Woon-Ming, B., Gris, C., Timmers, T., Poinsot, V., Gilbert, L. B., Heeb, P., Médigue, C., Batut, J., and Masson-Boivin, C. 2010. Experimental evolution of a plant pathogen into a legume symbiont. PLoS Biol. 8

Marie, C., Broughton, W. J., and Deakin, W. J. 2001. Rhizobium type III secretion systems: legume charmers or alarmers? Curr. Opin. Plant Biol. 4:336–342

Marie, C., Deakin, W. J., Ojanen-Reuhs, T., Diallo, E., Reuhs, B., Broughton, W. J., and Perret, X. 2004. TtsI, a key regulator of Rhizobium species NGR234 is required for type III-dependent protein secretion and synthesis of rhamnose-rich polysaccharides. Mol. Plant. Microbe. Interact. 17:958–966

Martinez-Abarca, F., Martinez-Rodriguez, L., Lopez-Contreras, J. A., Jimenez-Zurdo, J. I., and Toro, N. 2013. Complete Genome Sequence of the Alfalfa Symbiont Sinorhizobium/Ensifer meliloti Strain GR4. Genome Announc. 1:e00174-12-e00174-12

Masson-Boivin, C., Giraud, E., Perret, X., and Batut, J. 2009. Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? Trends Microbiol. 17:458–466

Matamoros, M. a, Baird, L. M., Escuredo, P. R., Dalton, D. a, Minchin, F. R., Iturbe-Ormaetxe, I., Rubio, M. C., Moran, J. F., Gordon, a J., and Becana, M. 1999. Stress-induced legume root nodule senescence. Physiological, biochemical, and structural alterations. Plant Physiol. 121:97–112

Matsunami, T., Kaihatsu, A., Maekawa, T., Takahashi, M., and Kokubun, M. 2004.

Characterization of Vegetative Growth of a Supernodulating Soybean Genotype, Sakukei 4. Plant Prod. Sci. 7:165–171

Mattoo, S., Lee, Y. M., and Dixon, J. E. 2007. Interactions of bacterial effector proteins with host proteins. Curr. Opin. Immunol. 19:392–401

Mavingui, P., Flores, M., Guo, X., Davila, G., Perret, X., Broughton, W. J., and Palacios, R. 2002. Dynamics of Genome Architecture in Rhizobium sp. Strain NGR234. J. Bacteriol. 184:171–176

Meade, H. M., Long, S. R., Ruvkun, G. B., Brown, S. E., and Ausubel, F. M. 1982. Physical and genetic characterization of symbiotic and auxotrophic mutants of Rhizobium meliloti induced by transposon Tn5 mutagenesis. J. Bacteriol. 149:114–22

Mergaert, P., Uchiumi, T., Alunni, B., Evanno, G., Cheron, A., Catrice, O., Mausset, A.-E. A.-E., Barloy-Hubler, F., Galibert, F., Kondorosi, A., and Kondorosi, E. 2006. Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium-legume symbiosis. Proc. Natl. Acad. Sci. 103:5230–5235

Michiels, J., Pelemans, H., Vlassak, K., Verreth, C., and Vanderleyden, J. 1995. Identification and characterization of a Rhizobium leguminosarum bv. phaseoli gene that is important for nodulation competitiveness and shows structural homology to a Rhizobium fredii host-inducible gene. Mol. Plant. Microbe. Interact. 8:468–472

Mijnendonckx, K., Provoost, A., Monsieurs, P., Leys, N., Mergeay, M., Mahillon, J., and Van Houdt, R. 2011. Insertion sequence elements in Cupriavidus metallidurans CH34: Distribution and role in adaptation. Plasmid. 65:193–203

Miwa, H., and Okazaki, S. 2017. How effectors promote beneficial interactions. Curr. Opin. Plant Biol. 38:148–154

Mockaitis, K., and Estelle, M. 2008. Auxin Receptors and Plant Development: A New Signaling Paradigm. Annu. Rev. Cell Dev. Biol. 24:55–80

Moore, L. W., Chilton, W. S., and Canfield, M. L. 1997. Diversity of opines and opine-catabolizing bacteria isolated from naturally occurring crown gall tumors. Appl. Environ. Microbiol. 63:201–207

Moriguchi, K., Maeda, Y., Satou, M., Hardayani, N. S., Kataoka, M., Tanaka, N., and Yoshida, K. 2001. The complete nucleotide sequence of a plant root-inducing (Ri) plasmid indicates its chimeric structure and evolutionary relationship between tumor-inducing (Ti) and symbiotic (Sym) plasmids in Rhizobiaceae. J. Mol. Biol. 307:771–784

Mortier, V., Holsters, M., and Goormachtig, S. 2012. Never too many? How legumes control nodule numbers. Plant, Cell Environ. 35:245–258

Nelson, M. S., Chun, C. L., and Sadowksy, M. J. 2015. Evaluation of Nodulation Speed by Sinorhizobium Strains. BIO-PROTOCOL. 5:e1554

Nelson, M. S., Chun, C. L., and Sadowsky, M. J. 2017. Type IV Effector Proteins Involved in the Medicago - Sinorhizobium Symbiosis. Mol. Plant-Microbe Interact. 30:28–34

Nelson, M. S., and Sadowsky, M. J. 2015. Secretion systems and signal exchange between nitrogen-fixing rhizobia and legumes. Front. Plant Sci. 6:491

Nimchuk, Z., Marois, E., Kjemtrup, S., Leister, R. T., Katagiri, F., and Dangl, J. L. 2000.

Eukaryotic fatty acylation drives plasma membrane targeting and enhances function of several type III effector proteins from Pseudomonas syringae. Cell. 101:353–363

Nishimura, R., Hayashi, M., Wu, G.-J., Kouchi, H., Imaizumi-Anraku, H., Murakami, Y., Kawasaki, S., Akao, S., Ohmori, M., Nagasawa, M., Harada, K., and Kawaguchi, M. 2002. HAR1 mediates systemic regulation of symbiotic organ development. Nature. 420:426–429

van Noorden, G. E. 2006. Defective Long-Distance Auxin Transport Regulation in the Medicago truncatula super numeric nodules Mutant. PLANT Physiol. 140:1494–1506

Oka-Kira, E., and Kawaguchi, M. 2006. Long-distance signaling to control root nodule number. Curr. Opin. Plant Biol. 9:496–502

Okamoto, S., Shinohara, H., Mori, T., Matsubayashi, Y., and Kawaguchi, M. 2013. Root-derived CLE glycopeptides control nodulation by direct binding to HAR1 receptor kinase. Nat Commun. 4

Okazaki, S., Kaneko, T., Sato, S., and Saeki, K. 2013. Hijacking of leguminous nodulation signaling by the rhizobial type III secretion system. Proc. Natl. Acad. Sci. U. S. A. 110:17131–6

Okazaki, S., Okabe, S., Higashi, M., Shimoda, Y., Sato, S., Tabata, S., Hashiguchi, M., Akashi, R., Göttfert, M., and Saeki, K. 2010. Identification and functional analysis of type III effector proteins in Mesorhizobium loti. Mol. Plant. Microbe. Interact. 23:223–34

Okazaki, S., Tittabutr, P., Teulet, A., Thouin, J., Fardoux, J., Chaintreuil, C., Gully, D.,

Arrighi, J.-F., Furuta, N., Miwa, H., Yasuda, M., Nouwen, N., Teaumroong, N., and Giraud, E. 2016. Rhizobium–legume symbiosis in the absence of Nod factors: two possible scenarios with or without the T3SS. ISME J. 10:64–74

Okazaki, S., Zehner, S., Hempel, J., Lang, K., and Göttfert, M. 2009. Genetic organization and functional analysis of the type III secretion system of Bradyrhizobium elkanii. FEMS Microbiol. Lett. 295:88–95

Oldroyd, G. E. D., and Downie, J. A. 2008. Coordinating nodule morphogenesis with rhizobial infection in legumes. Annu. Rev. Plant Biol. 59:519–546

Oldroyd, G. E. D., Murray, J. D., Poole, P. S., and Downie, J. A. 2011. The Rules of Engagement in the Legume-Rhizobial Symbiosis. Annu. Rev. Genet. 45:119–144

Palani, N. P. 2011. Metabolic regulation and genetic tools for bacterial neutral lipid production.

Paradis, E., Claude, J., and Strimmer, K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 20:289–90

Paulus, F., and Otten, L. Functional and mutated agrocinopine synthase genes on octopine T-DNAs. Mol. Plant. Microbe. Interact. 6:393–402

Pazour, G. J., and Das, A. 1990. Characterization of the VirG binding site of Agrobacterium tumefaciens. Nucleic Acids Res. 18:6909–13

Pedley, K. F., and Martin, G. B. 2005. Role of mitogen-activated protein kinases in plant immunity. Curr. Opin. Plant Biol. 8:541–547

Perret, X., Staehelin, C., and Broughton, W. J. 2000. Molecular Basis of Symbiotic Promiscuity. Microbiol. Mol. Biol. Rev. 64:180–201

Peters, N., Frost, J., and Long, S. 1986. A plant flavone, luteolin, induces expression of Rhizobium meliloti nodulation genes. Science (80-.). 233:977–980

Pii, Y., Crimi, M., Cremonese, G., Spena, A., and Pandolfini, T. 2007. Auxin and nitric oxide control indeterminate nodule formation. BMC Plant Biol. 7:21

Pistorio, M., Balagué, L. J., Del Papa, M. F., Pich-Otero, a, Lodeiro, a, Hozbor, D. F., and Lagares, a. 2002. Construction of a Sinorhizobium meliloti strain carrying a stable and non-transmissible chromosomal single copy of the green fluorescent protein GFP-P64L/S65T. FEMS Microbiol. Lett. 214:165–70

Poehlein, A., Freese, H., Daniel, R., and Simeonova, D. D. 2016. Genome sequence of Shinella sp. strain DD12, isolated from homogenized guts of starved Daphnia magna. Stand. Genomic Sci. 11:14

Prell, J., White, J. P., Bourdes, a, Bunnewell, S., Bongaerts, R. J., and Poole, P. S. 2009. Legumes regulate Rhizobium bacteroid development and persistence by the supply of branched-chain amino acids. Proc. Natl. Acad. Sci. U. S. A. 106:12477–12482

Prentki, P., and Krisch, H. M. 1984. In vitro insertional mutagenesis with a selectable DNA fragment. Gene. 29:303–313

Puppo, A., Groten, K., Bastian, F., Carzaniga, R., Soussi, M., Lucas, M. M., De Felipe, M. R., Harrison, J., Vanacker, H., and Foyer, C. H. 2005. Legume nodule senescence: Roles for redox and hormone signalling in the orchestration of the natural aging process. New Phytol. 165:683–701

Puri, N., Jenner, C., Bennett, M., Stewart, R., Mansfield, J., Lyons, N., and Taylor, J. 1997. Expression of avrPphB, an avirulence gene from Pseudomonas syringae pv.

phaseolicola, and the delivery of signals causing the hypersensitive reaction in bean. Mol. Plant. Microbe. Interact. 10:247–256

R Core Team. 2016. R: A Language and Environment for Statistical Computing.

Radutoiu, S., Madsen, L. H., Madsen, E. B., Jurkiewicz, A., Fukai, E., Quistgaard, E. M. H., Albrektsen, A. S., James, E. K., Thirup, S., and Stougaard, J. 2007. LysM domains mediate lipochitin-oligosaccharide recognition and Nfr genes extend the symbiotic host range. EMBO J. 26:3923–3935

Reid, D. E., Ferguson, B. J., and Gresshoff, P. M. 2011a. Inoculation- and nitrate-induced CLE peptides of soybean control NARK-dependent nodule formation. Mol. Plant. Microbe. Interact. 24:606–618

Reid, D. E., Ferguson, B. J., Hayashi, S., Lin, Y.-H., and Gresshoff, P. M. 2011b. Molecular mechanisms controlling legume autoregulation of nodulation. Ann. Bot. 108:789–95

Reid, D. E., Li, D., Ferguson, B. J., and Gresshoff, P. M. 2013. Structure-function analysis of the GmRIC1 signal peptide and CLE domain required for nodulation control in soybean. J. Exp. Bot. 64:1575–1585

Rhoads, A., and Au, K. F. 2015. PacBio Sequencing and Its Applications. Genomics. Proteomics Bioinformatics. 13:278–289

Rightmyer, A. P., and Long, S. R. 2011. Pseudonodule Formation by Wild-Type and Symbiotic Mutant Medicago truncatula in Response to Auxin Transport Inhibitors. Mol. Plant-Microbe Interact. 24:1372–1384

Rodrigues, J. a., López-Baena, F. J., Ollero, F. J., Vinardell, J. M., Espuny, M. D. R.,

Bellogín, R. a., Ruiz-Sainz, J. E., Thomas, J. R., Sumpton, D., Ault, J., and Thomas-Oates, J. 2007. NopM and NopD are rhizobial nodulation outer proteins: Identification using LC-MALDI and LC-ESI with a monolithic capillary column. J. Proteome Res. 6:1029–1037

Saad, M. M., Crèvecoeur, M., Masson-Boivin, C., and Perret, X. 2012. The type 3 protein secretion system of Cupriavidus taiwanensis strain LMG19424 compromises symbiosis with Leucaena leucocephala. Appl. Environ. Microbiol. 78:7476–7479

Saad, M. M., Kobayashi, H., Marie, C., Brown, I. R., Mansfield, J. W., Broughton, W. J., and Deakin, W. J. 2005. NopB, a type III secreted protein of Rhizobium sp. strain NGR234, is associated with pilus-like surface appendages. Society. 187:1173–1181

Saad, M. M., Staehelin, C., Broughton, W. J., and Deakin, W. J. 2008. Protein-protein interactions within type III secretion system-dependent pili of Rhizobium sp. strain NGR234. J. Bacteriol. 190:750–754

Saeki, K. 2011. Rhizobial measures to evade host defense strategies and endogenous threats to persistent symbiotic nitrogen fixation: a focus on two legume-rhizobium model systems. Cell. Mol. Life Sci. 68:1327–39

Sallet, E., Roux, B., Sauviac, L., Jardinaud, M.-F., Carrere, S., Faraut, T., de Carvalho-Niebel, F., Gouzy, J., Gamas, P., Capela, D., Bruand, C., and Schiex, T. 2013. Next-Generation Annotation of Prokaryotic Genomes with EuGene-P: Application to Sinorhizobium meliloti 2011. DNA Res. 20:339–354

Samac, D. A., Tesfaye, M., Dornbusch, M., Saruul, P., and Temple, S. J. 2004. A comparison of constitutive promoters for expression of transgenes in alfalfa

(Medicago sativa). Transgenic Res. 13:349–61

Sánchez, C., Iannino, F., Deakin, W. J., Ugalde, R. a, and Lepek, V. C. 2009. Characterization of the Mesorhizobium loti MAFF303099 Type-Three Protein Secretion System. Mol. Plant-Microbe Interact. 22:519–528

Sánchez, C., Mercante, V., Babuin, M. F., and Lepek, V. C. 2012. Dual effect of Mesorhizobium loti T3SS functionality on the symbiotic process. FEMS Microbiol. Lett. 330:148–156

Sasaki, T., Suzaki, T., Soyano, T., Kojima, M., Sakakibara, H., and Kawaguchi, M. 2014. Shoot-derived cytokinins systemically regulate root nodulation. Nat. Commun. 5:1–9

Schell, M. a., Ulrich, R. L., Ribot, W. J., Brueggemann, E. E., Hines, H. B., Chen, D., Lipscomb, L., Kim, H. S., Mrázek, J., Nierman, W. C., and DeShazer, D. 2007. Type VI secretion is a major virulence determinant in Burkholderia mallei. Mol. Microbiol. 64:1466–1485

Schnabel, E., Journet, E.-P., de Carvalho-Niebel, F., Duc, G., and Frugoli, J. 2005. The Medicago truncatula SUNN gene encodes a CLV1-like leucine-rich repeat receptor kinase that regulates nodule number and root length. Plant Mol. Biol. 58:809–22

Schneiker-Bekel, S., Wibberg, D., Bekel, T., Blom, J., Linke, B., Neuweger, H., Stiens, M., Vorhölter, F.-J., Weidner, S., Goesmann, A., Pühler, A., and Schlüter, A. 2011. The complete genome sequence of the dominant Sinorhizobium meliloti field isolate SM11 extends the S. meliloti pan-genome. J. Biotechnol. 155:20–33

Schrammeijer, B., Dulk-ras, A. Den, Vergunst, A. C., and Hooykaas, P. J. J. 2003.

Analysis of Vir protein translocation from Agrobacterium tumefaciens using Saccharomyces cerevisiae as a model: evidence for transport of a novel effector protein VirE3. Nucleic Acids Res. 31:860–868

Schrammeijer, B., Risseeuw, E., Pansegrau, W., Regensburg-Tuïnk, T. J., Crosby, W. L., and Hooykaas, P. J. 2001. Interaction of the virulence protein VirF of Agrobacterium tumefaciens with plant homologs of the yeast Skp1 protein. Curr. Biol. 11:258–62

Shames, S. R., and Finlay, B. B. 2012. Bacterial effector interplay: a new way to view effector function. Trends Microbiol. 20:214–9

Shan, L., He, P., Li, J., Heese, A., Peck, S. C., Nürnberger, T., Martin, G. B., and Sheen, J. 2008. Bacterial Effectors Target the Common Signaling Partner BAK1 to Disrupt Multiple MAMP Receptor-Signaling Complexes and Impede Plant Immunity. Cell Host Microbe. 4:17–27

Shao, F., Merritt, P. M., Bao, Z., Innes, R. W., and Dixon, J. E. 2002. A Yersinia effector and a Pseudomonas avirulence protein define a family of cysteine proteases functioning in bacterial pathogenesis. Cell. 109:575–588

Shen, C., Yue, R., Sun, T., Zhang, L., Xu, L., Tie, S., Wang, H., and Yang, Y. 2015. Genome-wide identification and expression analysis of auxin response factor gene family in Medicago truncatula. Front. Plant Sci. 6:1–13

Singleton, P. W., and Stockinger, K. R. 1983. Compensation against Ineffective Nodulation in Soybean. Crop Sci. 23:69

Skorpil, P., Saad, M. M., Boukli, N. M., Kobayashi, H., Ares-Orpel, F., Broughton, W. J.,

and Deakin, W. J. 2005. NopP, a phosphorylated effector of Rhizobium sp. strain NGR234, is a major determinant of nodulation of the tropical legumes Flemingia congesta and Tephrosia vogelii. Mol. Microbiol. 57:1304–1317

de Souza, J. A. M., Tieppo, E., de Souza Magnani, G., Alves, L. M. C., Cardoso, R. L., Cruz, L. M., de Oliveira, L. F., Raittz, R. T., de Souza, E. M., de Oliveira Pedrosa, F., and de Macedo Lemos, E. G. 2012. Draft genome sequence of the nitrogen-fixing symbiotic bacterium Bradyrhizobium elkanii 587. J. Bacteriol. 194:3547–3548

Sprent, J. I. 2007. Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation. New Phytol. 174:11–25

Stachel, S. E., and Zambryski, P. C. 1986. virA and virG control the plant-induced activation of the T-DNA transfer process of A. tumefaciens. Cell. 46:325–333

Stiens, M., Schneiker, S., PÃ¼hler, A., and SchlÃ¼ter, A. 2007. Sequence analysis of the 181-kb accessory plasmid pSmeSM11b, isolated from a dominant Sinorhizobium meliloti strain identified during a long-term field release experiment. FEMS Microbiol. Lett. 271:297–309

Su, Y. H., Liu, Y. B., Bai, B., and Zhang, X. S. 2015. Establishment of embryonic shoot-root axis is involved in auxin and cytokinin response during Arabidopsis somatic embryogenesis. Front. Plant Sci. 5

Sugawara, M., Epstein, B., Badgley, B. D., Unno, T., Xu, L., Reese, J., Gyaneshwar, P., Denny, R., Mudge, J., Bharti, A. K., Farmer, A. D., May, G. D., Woodward, J. E., Medigue, C., Vallenet, D., Lajus, A. A., Rouy, Z. Z. Z., Martinez-Vaz, B., Tiffin, P.,

Young, N. D., Sadowsky, M. J., Médigue, C., Vallenet, D., Lajus, A. A., Rouy, Z. Z. Z., Martinez-Vaz, B., Tiffin, P., Young, N. D., Sadowsky, M. J., Medigue, C., Vallenet, D., Lajus, A. A., Rouy, Z. Z. Z., Martinez-Vaz, B., Tiffin, P., Young, N. D., and Sadowsky, M. J. 2013. Comparative genomics of the core and accessory genomes of 48 Sinorhizobium strains comprising five genospecies. Genome Biol. 14:R17

Sugawara, M., and Sadowsky, M. J. 2014. Enhanced Nodulation and Nodule Development by nolR Mutants of Sinorhizobium medicae on Specific Medicago Host Genotypes. Mol. Plant-Microbe Interact. 27:328–335

Suominen, L. 1999. Identification of nodulation promoter (nod-box) regions of Rhizobium galegae. FEMS Microbiol. Lett. 177:217–223

Suzaki, T., Ito, M., and Kawaguchi, M. 2013. Genetic basis of cytokinin and auxin functions during root nodule development. Front. Plant Sci. 4:1–6

Suzaki, T., Yoro, E., and Kawaguchi, M. 2015. *Leguminous Plants: Inventors of Root Nodules to Accommodate Symbiotic Bacteria*. Elsevier Ltd.

Szemenyei, H., Hannon, M., and Long, J. A. 2008. TOPLESS Mediates Auxin-Dependent Transcriptional Repression During Arabidopsis Embryogenesis. Science (80-.). 319:1384–1386

Takahara, M., Magori, S., Soyano, T., Okamoto, S., Yoshida, C., Yano, K., Sato, S., Tabata, S., Yamaguchi, K., Shigenobu, S., Takeda, N., Suzaki, T., and Kawaguchi, M. 2013. TOO MUCH LOVE, a novel kelch repeat-containing F-box protein, functions in the long-distance regulation of the LEGUME-rhizobium symbiosis.

Plant Cell Physiol. 54:433–447

Takahashi, M., Arihara, J., Nakayama, N., and Kokubun, M. 2003. Characteristics of Growth and Yield Formation in the Improved Genotype of Supernodulating Soybean (*Glycine max* L. Merr.). Plant Prod. Sci. 6:112–118

Tampakaki, A. P. 2014. Commonalities and differences of T3SSs in rhizobia and plant pathogenic bacteria. Front. Plant Sci. 5:114

Tamura, K., and Nei, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10:512–26

Tang, F., Yang, S., Liu, J., and Zhu, H. 2016. Rj4 , a Gene Controlling Nodulation Specificity in Soybeans, Encodes a Thaumatin-Like Protein But Not the One Previously Reported. Plant Physiol. 170:26–32

Thomas, C. M., and Nielsen, K. M. 2005. Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. Nat. Rev. Microbiol. 3:711–721

Tseng, T.-T., Tyler, B. M., and Setubal, J. C. 2009. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. BMC Microbiol. 9 Suppl 1:S2

Turner, M., Nizampatnam, N. R., Baron, M., Coppin, S., Damodaran, S., Adhikari, S., Arunachalam, S. P., Yu, O., and Subramanian, S. 2013. Ectopic Expression of miR160 Results in Auxin Hypersensitivity, Cytokinin Hyposensitivity, and Inhibition of Symbiotic Nodule Development in Soybean. PLANT Physiol. 162:2042–2055

Tzfira, T., Vaidya, M., and Citovsky, V. 2004. Involvement of targeted proteolysis in plant genetic transformation by Agrobacterium. Nature. 431:6–11

Udvardi, M., and Poole, P. S. 2013. Transport and metabolism in legume-rhizobia symbioses. Annu. Rev. Plant Biol. 64:781–805

Ulmasov, T. 1995. Composite Structure of Auxin Response Elements. PLANT CELL ONLINE. 7:1611–1623

Ulmasov, T., Hagen, G., and Guilfoyle, T. J. 1997. ARF1, a transcription factor that binds to auxin response elements. Science. 276:1865–8

Ulmasov, T., Hagen, G., and Guilfoyle, T. J. 1999. Dimerization and DNA binding of auxin response factors. Plant J. 19:309–319

Vallenet, D. 2006. MaGe: a microbial genome annotation system supported by synteny results. Nucleic Acids Res. 34:53–65

Vallenet, D., Belda, E., Calteau, A., Cruveiller, S., Engelen, S., Lajus, A., Le Fevre, F., Longin, C., Mornico, D., Roche, D., Rouy, Z., Salvignol, G., Scarpelli, C., Thil Smith, A. A., Weiman, M., and Medigue, C. 2013. MicroScope--an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. Nucleic Acids Res. 41:D636–D647

Vallenet, D., Calteau, A., Cruveiller, S., Gachet, M., Lajus, A., Josso, A., Mercier, J., Renaux, A., Rollin, J., Rouy, Z., Roche, D., Scarpelli, C., and Médigue, C. 2017. MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. Nucleic Acids Res. 45:D517–D528

Vasse, J., Billy, F., and Truchet, G. 1993. Abortion of infection during the Rhizobium

meliloti—alfalfa symbiotic interaction is accompanied by a hypersensitive reaction. Plant J. 4:555–566

Van Veen, R. J. M., den Dulk-Ras, H., Schilperoort, R. A., and Hooykaas, P. J. J. 1989. Ti plasmid containing Rhizobium meliloti are non-tumorigenic on plants, despite proper virulence gene induction and T-strand formation. Arch. Microbiol. 153:85–89

Van de Velde, W., Zehirov, G., Szatmari, A., Debreczeny, M., Ishihara, H., Kevei, Z., Farkas, A., Mikulass, K., Nagy, A., and Tiricz, H. 2010. Plant peptides govern terminal differentiation of bacteria in symbiosis. Science (80-.). 327:1122–1126

Vergunst, A. C. 2003. Recognition of the Agrobacterium tumefaciens VirE2 Translocation Signal by the VirB/D4 Transport System Does Not Require VirE1. PLANT Physiol. 133:978–988

Vergunst, A. C., van Lier, M. C. M., den Dulk-Ras, A., Stüve, T. a G., Ouwehand, A., and Hooykaas, P. J. J. 2005. Positive charge is an important feature of the C-terminal transport signal of the VirB/D4-translocated proteins of Agrobacterium. Proc. Natl. Acad. Sci. U. S. A. 102:832–7

Vergunst, A. C., Schrammeijer, B., den Dulk-Ras, A., de Vlaam, C. M. T., Regensburg-Tuïnk, T. J. G., and Hooykaas, P. J. J. 2000. VirB/D4-Dependent Protein Translocation from Agrobacterium into Plant Cells. Sci. . 290:979–982

Viprey, V., Del Greco, A., Golinowski, W., Broughton, W. J., and Perret, X. 1998. Symbiotic implications of type III protein secretion machinery in Rhizobium. Mol. Microbiol. 28:1381–1389

Waadt, R., and Kudla, J. 2008. In Planta Visualization of Protein Interactions Using Bimolecular Fluorescence Complementation (BiFC). Cold Spring Harb. Protoc. 2008:pdb.prot4995-prot4995

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. M. 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement J. Wang, ed. PLoS One. 9:e112963

Walter, M., Chaban, C., Schütze, K., Batistic, O., Weckermann, K., Näke, C., Blazevic, D., Grafen, C., Schumacher, K., Oecking, C., Harter, K., Kudla, J., Grefen, C., Schumacher, K., Oecking, C., Harter, K., and Kudla, J. 2004. Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. Plant J. 40:428–38

Wang, D., Yang, S., Tang, F., and Zhu, H. 2012. Symbiosis specificity in the legume - rhizobial mutualism. Cell. Microbiol. 14:334–342

Wassem, R., Kobayashi, H., Kambara, K., Le Quéré, A., Walker, G. C., Broughton, W. J., and Deakin, W. J. 2008. TtsI regulates symbiotic genes in Rhizobium species NGR234 by binding to tts boxes. Mol. Microbiol. 68:736–748

Wasson, A. P. 2006. Silencing the Flavonoid Pathway in Medicago truncatula Inhibits Root Nodule Formation and Prevents Auxin Transport Regulation by Rhizobia. PLANT CELL ONLINE. 18:1617–1629

Wei, M., Zhang, J.-J., Liu, H., Wang, S.-J., Fu, H., and Zhou, N.-Y. 2009. A transposable class I composite transposon carrying mph (methyl parathion hydrolase) from

Pseudomonas sp. strain WBC-3. FEMS Microbiol. Lett. 292:85–91

Wenzel, M., Friedrich, L., Göttfert, M., and Zehner, S. 2010. The type III-secreted protein NopE1 affects symbiosis and exhibits a calcium-dependent autocleavage activity. Mol. Plant. Microbe. Interact. 23:124–129

Wold, S., Esbensen, K., and Geladi, P. 1987. Principal component analysis. Chemom. Intell. Lab. Syst. 2:37–52

Wooldridge, K. 2009. *Bacterial secreted proteins: secretory mechanisms and role in pathogenesis*. Horizon Scientific Press.

Xin, D. W., Liao, S., Xie, Z. P., Hann, D. R., Steinle, L., Boller, T., and Staehelin, C. 2012. Functional analysis of NopM, a novel E3 ubiquitin ligase (NEL) domain effector of Rhizobium sp. strain NGR234. PLoS Pathog. 8

Yang, S., Tang, F., Gao, M., Krishnan, H. B., and Zhu, H. 2010. R gene-controlled host specificity in the legume-rhizobia symbiosis. Proc. Natl. Acad. Sci. U. S. A. 107:18735–18740

Yasuda, M., Miwa, H., Masuda, S., Takebayashi, Y., Sakakibara, H., and Okazaki, S. 2016. Effector-Triggered Immunity Determines Host Genotype-Specific Incompatibility in Legume– Rhizobium Symbiosis. Plant Cell Physiol. 57:1791–1800

Zhang, H., Jin, J., Tang, L., Zhao, Y., Gu, X., Gao, G., and Luo, J. 2011a. PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. Nucleic Acids Res. 39:D1114–D1117

Zhang, L., Chen, X. J., Lu, H. Bin, Xie, Z. P., and Staehelin, C. 2011b. Functional

analysis of the type 3 effector nodulation outer protein L (NopL) from Rhizobium sp. NGR234: Symbiotic effects, phosphorylation, and interference with mitogen-activated protein kinase signaling. J. Biol. Chem. 286:32178–32187

Zuo, J., Niu, Q. W., and Chua, N. H. 2000. An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. Plant J. 24:265–273

Zupan, J. R., and Zambryski, P. 1995. Transfer of T-DNA from Agrobacterium to the plant cell. Plant Physiol. 107:1041–1047