TOWARDS THE MINIMAL SYMBIOTIC GENOME OF *SINORHIZOBIUM MELILOTI*

TOWARDS THE MINIMAL SYMBIOTIC GENOME OF *SINORHIZOBIUM MELILOTI*

By JIARUI HUANG, B.Sc. MOLECULAR BIOLOGY AND GENETICS

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements for the

Degree of Master of Science

Master of Science (2019)

McMaster University (Biology), Hamilton, Ontario, Canada

Title: Towards the minimal symbiotic genome of *Sinorhizobium meliloti*

Author: Jiarui Huang, B.Sc. Biology (McMaster University)

Supervisor: Dr. Turlough M. Finan

Number of Pages: i-xvi, 1-150

Lay Abstract

Nitrogen is one of the critical elements for life. Biological nitrogen fixation plays a crucial role in providing fixed nitrogen for the ecosystem on Earth. Our Laboratory has endeavored to establish a minimal symbiotic genome in *Sinorhizobium meliloti*, a model nitrogen fixing bacterium which forms symbiosis with certain kinds of legumes. Building this minimal symbiotic genome will improve our understanding of the symbiotic nitrogen fixation process in *S. meliloti* at gene level. It may also help in eventually introducing a nitrogen fixation system into other organisms. In this study, the minimal symbiotic genome of the pSymA replicon in *S. meliloti* was constructed. In addition, a staining method to detect specific *S. meliloti* strains in nodules was established. This method is potentially useful in finding genes related to nodule competitiveness, and these are potentially important for augmenting the genes that constitute the minimal symbiotic genome.

Abstract

Sinorhizobium meliloti is a model bacterium for the study of symbiotic nitrogen fixation (SNF). It infects the roots of alfalfa as well as some other legumes and differentiates into N_2-fixing bacteroids within the plant cells of specialized nodule organs. To understand genes essential for SNF and, in the longer term, to facilitate the manipulation of this SNF process for agricultural purposes, it is highly desirable to construct the minimal genome for SNF in this organism. *S. meliloti* harbors two replicons required for SNF, a 1.7-Mb chromid (pSymB) and a 1.4-Mb megaplasmid (pSymA). A previous deletion analysis revealed that only four gene regions, accounting for <12% of the total sequences of pSymA and pSymB that, were essential for SNF. In the first part of the thesis, I report the cloning of these two pSymA SNF-essential regions on a plasmid (pTH3255) in *Escherichia coli*, and the integration of this plasmid into the genome of a ΔpSymA *S. meliloti* derivative strain (the strain was named as RmP4291 after integration). Plant root dry weight and nitrogenase-catalyzed acetylene reduction assays were carried out on RmP4291 with four host plants, including *Medicago sativa*, *Medicago truncatula*, *Melilotus alba* and *Melilotus officinalis*. Nodule kinetic assays were also performed on RmP4291 and RmP110(wt). The results showed that the SNF-essential regions from pSymA were sufficient to restore the symbiotic capabilities to the ΔpSymA derivative strain with all the host plants tested, except a significant reduction (~40%) in SNF by RmP4291 was

noticed on *M. officinalis* compared to that by wildtype *S. meliloti.* A higher alfalfa nodulation efficiency of RmP4291 compared to that of wildtype RmP110 was also discovered.

In the second part of the thesis, a histochemical staining method for *S. meliloti* nodules was developed by integrating the marker genes *gusA* (β-glucuronidase) and *celB* (β-glucosidase) into the *S. meliloti* genome. This staining method was found to be useful in the study of nodule competitiveness. A nodule competition assay was carried out between RmP4291 and RmP110 using the new staining method. RmP4291 was found to be significantly reduced in nodulation competitiveness compared to wildtype *S. meliloti*. The development of the histochemical staining method for *S. meliloti* nodules will accelerate the identification of genes required for nodule competitiveness in the organism, which will be of crucial importance to the construction of the minimal genome strains with high SNF efficiency.

Acknowledgements

First, I would like to thank Dr. Turlough Finan for giving me the opportunity to study in his lab and do such an interesting project at McMaster University, Canada. From him I have learned not only the work I did itself, more importantly the beauty of microbiology and genetics, which would always keep me interested in the area. I also thank Dr. Marie Elliot for being part of my committee and giving me advice.

Secondly, I'm also thankful to all the members in the Finan lab who have been very helpful to me not only on scientific research, but also in my daily life in general. I thank Dr. Barney Geddes for his guidance on my research and answering my questions, Dr. Zahed Muhammed for all his help when I just got into the lab, Lauren Tiller for her help on reminding me of basically everything, Jason Kearsley for explaining interesting things to me and teaching me English terms, Leah Sather for her help on many little things, Michael MacLeod for teaching me things outside the lab. I would also like to thank all the people who have contributed to the minimal genome project, without their efforts I wouldn't have had the chance to do this very interesting project. I thank Dr. George DiCenzo, even though he was not in the lab during the time, he still helped me in many ways. And I'm very grateful to all the undergraduates who did all the cleaning and lab chores for us, which ensured a comfortable environment for doing research. Lastly, a thank you for my family and friends who

always support me and keep me in company. All in all, thank you thank you thank you.

Table of Contents

List of Figures and Tables

List of Abbreviations and Symbols

SNF	Symbiotic nitrogen fixation
Δ	Genetic deletion
Φ	Bacteriophage transduction
ORF	Open reading frames
wt	Wild type

Declaration of Academic Achievement

I declare that the research contribution which follows is original work, completed and written by me, with editorial assistance from my supervisor Dr. Turlough M. Finan and Dr. Li Huang.

The following individuals contributed to experimentation:

- In Chapter 3.1.2 – The deletion of the “scar” region between A118 and A121 was carried out by Dr. Zahed Muhammed.

Chapter 1 - Literature review and Introduction

1.1 Nitrogen and nitrogen fixation

Nitrogen is one of the critical elements for life. It is an essential component in most of the key biological molecules, such as proteins and nucleic acids (i.e., DNA and RNA). 78% of the total Earth's atmosphere is accounted for by about 10^{15} tons of N_2 gas, which makes nitrogen one of the most ubiquitous element on Earth. However, most living organisms cannot directly utilize this form of nitrogen because of the trivalent bond in the molecule. Fortunately, nature has developed several ways in which N_2 is fixed to form nitrogen-containing compounds; this process is also referred to as nitrogen fixation. There are three major processes of nitrogen fixation: a) conversion of atmospheric nitrogen into nitrogen compounds during abiotic natural processes, such as lightning, b) industrial nitrogen fixation, and c) biological nitrogen fixation. The industrial nitrogen fixation usually entails the use of Haber-Bosch process, which involves a reaction between nitrogen gas and hydrogen gas under high pressures and high temperatures with ammonia as product. Biological nitrogen fixation uses nitrogenase enzymes to convert N_2 into NH_3, contributing more than 60% of the total atmospheric nitrogen fixed in the biogeochemical nitrogen cycle on Earth (Canfield et al., 2010).

1.2 Biological nitrogen fixation

Although some plants and animals are able to establish a symbiotic relationship with nitrogen-fixing microorganisms no eukaryotes are known to possess the ability to fix nitrogen by themselves. However, for prokaryotes, nitrogen fixation is increasingly recognized as a widespread biological function existing in diverse bacteria and archaea, which are collectively referred to as diazotrophs (Vicente and Dean, 2017). For bacteria, nitrogen fixation has been found in most bacterial phylogenetic groups, such as green sulfur bacteria, firmibacteria, actinomycetes, cyanobacteria and all subdivisions of the Proteobacteria (Dixon and Kahn, 2004). On the other hand, methanogens are the main archaea currently known to possess nitrogen fixing ability (Dixon and Kahn, 2004). Diazotrophs are divided into the following three categories.

a) Free-living nitrogen-fixing bacteria are diazotrophs that thrive freely in the environment and fix nitrogen independently. Free-living nitrogen-fixing bacteria are known for their high diversity and include obligate aerobes (e.g., *Azotobacter vinelandii*), obligate anaerobes (e.g., *Clostridium pasteurianum*), and facultative anaerobes (e.g., *Klebsiella oxytoca*).
b) Symbiotic N_2-fixing (SNF) bacteria include rhizobia, Frankia and cyanobacteria and these form association with legumes, actinorhizal plants (non-legume plants), and ferns, respectively. These bacteria usually form a special symbiotic structure with their host plants. For instance, rhizobia are able to infect the roots of a legume plant and induce the formation of specialized nodule organs within which

the bacteria differentiate into N_2-fixing bacteroids. While the N_2-fixing bacteria provide plant hosts with fixed nitrogen, plants supply nutrients to the bacteria in return. Symbiotic N_2-fixing bacteria usually exhibit a strict host specificity (i.e., a narrow host range) and display high efficiency in N_2 fixation.

c) Associative N_2-fixing bacteria, such as *Azospirillum*, fix N_2 in association with plants. However, they don't get into plant cytoplasm or form special structures or organs as symbiotic N_2-fixing bacteria do. Associative N_2-fixing bacteria are often found on the surface of a plant (roots or leaves) or in between plant cortical cells. They usually can associate with a wide range of plant hosts, including important cereal crops, such as rice, wheat, corn, etc.

1.3 Legume-rhizobia symbiosis

Benefiting from symbiosis, symbiotic N_2-fixing bacteria, especially rhizobia, are extraordinarily efficient in N_2-fixation, accounting for about 80% of the total biologically fixed nitrogen in nature. Therefore, rhizobia are the most important contributor of biological N_2-fixation on Earth (Peoples et al., 1995). The general process of legume-rhizobia symbiosis is outlined below.

1.3.1 Rhizobia infection

The legume-rhizobia symbiosis starts with a signal produced by the plant host. This signal usually is a flavonoid such as luteolin, methoxychalcone, naringenin, and

daidzein, or sometimes nonflavonoid such as trigonelline and stachydrine (Long, 1996). The signal serves as an inducer of *nod* gene expression and is sensed by the *nodD* gene product. NodD protein is a transcriptional activator, which promotes the transcription of genes for nodulation (i.e., *nod*, *nol*, *noe*) in rhizobia. Most of the nodulation genes are responsible for the synthesis and secretion of lipo-chitooligosaccharides, also known as the Nod factors. Most of the *nod* genes can be divided into two classes, i.e., common nodulation genes and host-specific nodulation genes. The common nodulation genes, referred to as *nodABC* genes, are conserved among all rhizobia, and encode the pathway for the synthesis of the Nod factor core: *N*-acylated chitooligosaccharide (Carlson, 1994). Host-specific nodulation genes, such as *nodH*, *nodP*, *nodQ* and *nodL*, are involved in the recognition between symbionts and specific plant hosts. The products of these genes structurally modify the Nod factor, usually by making substitutions at the terminal residues or changes on the structure of the acyl chain. For example, NodP, NodQ and NodH are responsible for the addition of a sulphate group on the Nod factor (Lerouge et al., 1990), and this modification is a main determinant of host specificity for *Sinorhizobium meliloti*. NodL (Geurts and Bisseling, 2002) places an acetyl group onto the Nod factor. Inactivation of *nodL* results in a delay in nodule formation and a decrease in nodule number on *Medicago sativa*, and in nodulation efficiency on *Medicago lupulina* and *Medicago littoralis*. Besides, not all of the nodulation genes are directly involved in

the synthesis of Nod factor. For instance, NodIJ are known to play a role in exporting the Nod factor (Carlson, 1994).

Nod factors are detected by the receptors of the plant host and root morphological changes are induced (Geurts et al., 2005). To begin with, root hair deformation and branching are activated, causing isotropic growth of the hairs. While most of the root hairs deformed, only a few of them will continue to the next step: curling. The root hairs attached by rhizobia sense the Nod factor and start to grow toward the side where rhizobia(nod factor) are attached (Esseling et al., 2003). As a result, the root hair bends over and forms a pocket in which the symbionts are encircled. Inside the pocket, the surrounding plant cell walls partially degrade, as induced by the rhizobia. And this allows the bacteria to get into the plant cell and the plant host forms a special structure inside the cytoplasm, that is, an infection thread, which is a tube-like structure made of cellulose. Symbionts can travel inside the tube-like structure to reach the cortex of the roots and differentiate into specialized N_2-fixing bacteroids, thereby forming a nodule primordium (Geurts et al., 2005).

1.3.2 Regulation of N_2-fixation during symbiosis

After the formation of the nodule, genes involved in symbiosis regulation will be activated. In *S. meliloti*, the *fixLJ* genes, which code for an oxygen-sensing two-component system, are the first symbiotic nitrogen fixation regulator genes

activated in response to microoxic conditions inside the nodules (David et al., 1988). The regulatory process starts with the activation of the oxygen-responsive gene *fixL*, which encodes a haemoprotein kinase, by the low oxygen environment. The FixL protein phosphorylates FixJ, which in turn switches on the transcription of two intermediate regulatory genes, *nifA* and *fixK*. NifA is a member of the enhancer binding protein family. As a transcriptional activator, it turns on the transcription of the *nif* operons by interacting with the RNA polymerase sigma factor, σ^{54} (Dixon and Kahn, 2004). NifA is also sensitive to oxygen, and its activity is supressed by the increasing fixed nitrogen sources. On the other hand, FixK is a transcriptional regulator of the Crp/Fnr family (Batut et al., 1989). FixK mainly activates genes involved in microaerobic growth. Among these genes are those in the *fixNOQP* operon, which encodes a high-affinity respiratory oxidase complex required for cell survival under microoxic conditions (Preisig et al., 1993). FixK also activates the expression of the gene coding for FixT (Foussard et al., 1997), a negative regulator of nitrogen fixation. FixT inhibits phosphorylation of FixJ by FixL.

1.3.3 Nitrogen fixation

All known nitrogen-fixing microorganisms produce nitrogenases. Nitrogenases are a family of metalloenzymes consisting of two components, nitrogenase reductase and nitrogenase (Dixon and Kahn, 2004). Nitrogenase reductase, encoded by *nifH* is

known as the Fe protein or component 2. It is a ~65 kDa dimer with each subunit containing a Mg-ATP binding sites and a single [4Fe-4S] cluster at the interface of the two subunits. Nitrogenase is a heterotetrameric protein containing two α subunits (encoded by *nifD*) and two β units (encoded by *nifK*) (Dixon and Kahn, 2004). Three types of nitrogenases have been discovered so far. They differ in metal cofactor at the active site, which could be molybdenum-iron, vanadium-iron or iron-iron (Schindelin et al., 1997; Tezcan et al., 2005). Some organisms contain more than one nitrogenase system, and the choice among the different systems depends on the availability of the metal in the environment (Jacobitz and Bishop, 1992). MoFe protein is the highest while FeFe protein is the lowest among the three nitrogenases in enzyme activity or nitrogen fixation efficiency (Crans et al., 2004; Schlessman et al., 1998). All three nitrogenases catalyze the biological reduction of N_2 according to the following equation: $N_2 + 8\ e^- + 8\ H^+ + 16\ MgATP \rightarrow 2\ NH_3 + H_2 + 16\ MgADP + 16Pi$ (Dixon and Kahn, 2004). As shown in the equation, eight electrons and sixteen ATPs are required for the reduction of each N_2 in the reaction, suggesting that a substantial amount of reducing power and energy is needed for biological nitrogen fixation (Seefeldt et al., 2009). Among the three nitrogenase systems, the molybdenum-iron nitrogenase system (or the MoFe protein) is most widely present in diazotrophs and is the only system in rhizobia. Two types of metal centers are included in the MoFe protein: The P cluster (i.e., an [8Fe–7S] cluster) and the FeMo cofactor (i.e., $MoFe_7$

S_9 ·homocitrate). The electron transport starts with the reduction of Fe protein (nitrogenase reductase) by accepting one electron from donors which usually are ferredoxins or flavodoxin. This is followed by the formation of a complex by the Fe and MoFe proteins. Single electrons are then transferred from the [4Fe-4S] cluster of Fe protein to the P cluster of MoFe protein at the expense of ATP hydrolysis, and eventually moved to the FeMo cofactor ($MoFe_7$ S_9 ·homocitrate). Besides, a number of additional proteins are involved in the assembly of nitrogenases and metal clusters required for the activation of the catalysis (Hu and Ribbe, 2013).

1.4 Nitrogen crisis

1.4.1 Nitrogen in agriculture

The lack of fixed nitrogen available to plants has always been a major concern in the global efforts to improve the productivity of crops. For decades, people have used industrial nitrogen fertilizers to alleviate this problem. However, the overuse of nitrogen fertilizers has caused serious environmental pollution and inflicted high economic costs, especially in developing countries (Galloway et al., 2008). The problem with nitrogen fertilizers is as follows. First, as the Haber-Bosch process used in the industrial fixation of N_2, reaction requires high pressure and high temperature (15~25 MPa, 400~500 °C) and thus the production of ammonium often comes with a large consumption of non-renewable energy (fossil fuels such as oil and coal) as well

as the emissions of the greenhouse gases nitrogen oxide and carbon oxide. Second, chemical fertilizers are used in large quantity because of their low utilization efficiency for crops. Un-utilized fixed nitrogen is left in the environment, causing secondary pollutions such as soil or water acidification and water eutrophication.

It has been shown that human interference has already doubled the rate of nitrogen input into the terrestrial nitrogen cycle (Vitousek et al., 1997). As the world's population continue to increase in the 21st century, the greater food demand will conceivably lead to an overwhelming consumption of nitrogen fertilizers if the current agricultural practice remains unchanged. Therefore, a cleaner and more efficiently usable nitrogen source is becoming one of the most urgent needs of the world. With the advent of synthetic biology, scientists have started to look into the possibility of manipulating biological nitrogen fixation for agricultural and ecological purposes with an ultimate goal of introducing this process into cereals and other crops, in which symbiotic nitrogen fixation does not naturally occur. Three major strategies have been employed to achieve this goal.

a) To engineer bacteria that are naturally associated with non-legume plants, such as cereals, which are of vital importance in agriculture, into nitrogen-fixing bacteria.

b) To introduce the ability to fix nitrogen into eukaryotes, especially plants, allowing them to fix nitrogen by themselves without the help of microorganisms.

c) To engineer non-legume plants, such as cereals, into ones capable of symbiosis with nitrogen fixing bacteria.

1.4.2 Turning a plant-associated bacterium into a nitrogen fixer

The first strategy entails the identification of non-nitrogen fixing bacteria capable of attaching to a target non-legume plant and the introduction of the nitrogen-fixing ability into the bacteria. In this strategy, the bacteria should be able to attach to the roots or invade into the tissue of a target plant. Once a bacterial strain of desired characteristics is identified, it is engineered into a nitrogen-fixing one. Efforts should then be made to initiate intracellular accommodation for the engineered bacterium, control the plant microbiome, and keep cheaters (microorganisms that extract benefits from the symbiosis without paying costs) under control (Mus et al., 2016).

As an example, (Fox et al., 2016) successfully transferred a genomic island with nitrogen fixing activity from a *Pseudomonas* strain into another *Pseudomonas* strain which was inactive in nitrogen fixation. For *Pseudomonas*, the ability to fix nitrogen is a rare feature (Anzai et al., 2000). The authors took advantage of the presence of a nitrogen fixation genomic island on the chromosome of *Pseudomonas stutzeri* A1501. The fact that some *Pseudomonas* strains possess a nitrogen fixation island suggests that: (a) *Pseudomonas* may have a genetic background suitable for the synthesis of a heterologous nitrogenase, (b) the horizontal transfer of a nitrogen fixation island

among *Pseudomonas* strains may occur, and (c) all genes required for the expression of nitrogenase may be efficiently packaged within the nitrogen fixation island of *P. stutzeri* A1501 (Setten et al., 2013). On the other hand, the recipient strain used in the study is a biological control agent that lives in the rhizosphere of a wide variety of plant species, and is able to persist and compete with native soil microorganisms (Paulsen et al., 2005). The authors transferred the nitrogen fixation island (X940) from *P. stutzeri* A15101 into *Pseudomonas protegens* Pf-5 by using a cosmid as a cloning vector. The resulting nitrogen-fixing strain, denoted *P. protegens* Pf-5 X940, displayed high nitrogenase activity, releasing significant quantities of ammonium into the medium in an nitrogen-deficient environment (Setten et al., 2013). Moreover, they also successfully transferred the nitrogen fixation island into other *Pseudomonas* strains, such as *Pseudomonas putida* KT2440, *Pseudomonas veronii* DSM11331, *Pseudomonas taetrolens* IAM1653, *Pseudomona*s balearica SP1402 and *Pseudomonas stutzeri* CCUG11256, using cosmid X940 (Setten et al., 2013). These results demonstrate that the genetic engineering strategy used to obtain a nitrogen-fixing strain is not restricted to *P. protegens* Pf-5 but applicable to other members of the genus *Pseudomonas*.

The authors then introduced the novel nitrogen-fixing strain *P. protegens* Pf-5 X940 into maize and wheat, and analyzed the impact of the strain on the productivity and nitrogen contents of the cereals in an experimental system involving the use of

non-gnotobiotic agronomic soil under greenhouse conditions (Fox et al., 2016). In their study, the plants were either supplemented with fixed nitrogen or left untreated (control), and subsequently inoculated with either the non-nitrogen-fixing wild-type strain Pf-5 or its isogenic nitrogen-fixing strain Pf-5 X940. Compared to the nitrogen-supplemented plants, the untreated plants and the plants inoculated with wild-type Pf-5 strain were low in biomass and nitrogen content in both vegetative and reproductive tissues, whereas the plants inoculated with Pf-5 X940 showed significantly improved phenotypes. Also, the rates of nitrogen fixation in the rhizosphere of the Pf-5 X940-treated maize and wheat plants were significantly increased only in isolated roots and not in soil without roots, indicating that nitrogen fixation indeed occurred in the roots.

1.4.3 Introducing nitrogen fixation ability into plants

The second strategy aims to endow a plant with the ability to fix nitrogen by itself without the need for bacterial interactions. Biological nitrogen fixation is known to require a large quantity of reducing power and energy. Besides, an array of proteins are involved in the assembly of nitrogenases and metal clusters required to activate the catalysis (Hu and Ribbe, 2013). Therefore, Vicente and Dean (2017) believed the following issues need to be addressed in the construction of nitrogen-fixing cereals. (a) How can a minimal set of genes required for nitrogen fixation be transferred into a

plant and expressed successfully? (b) How can the requirement for large amounts of reducing power and energy be satisfied? (c) How can an anoxic or low-oxygen environment be established for oxygen-sensitive nitrogenases? (d) How can enough metal clusters be provided to activate the catalysis?

It is currently believed that a chloroplast or a mitochondrion might be an ideal organelle for housing a nitrogen fixation system(Cheng et al., 2005; Ivleva et al., 2016; López-Torrejón et al., 2016). As a semi-autonomous organelle derived from a prokaryotic cell, a chloroplast or mitochondrion possesses a transcription/translation system similar to that in prokaryotes, allowing genes to be expressed in the form of operons (Scharff and Bock, 2014). In addition, both chloroplasts and mitochondria can potentially provide abundant ATP and reducing power to nitrogenase since they are organelles responsible for energy conversion.

Although these issues remain to be fully addressed, promising progress has been made in recent years. For instance, Yang et al. (2017) came up with an interesting idea to replace the bacterial electron-transport components (ETCs) used in nitrogen fixation with plant ETCs. The authors tested the compatibility of the molybdenum and the iron-only nitrogenases with ETC modules from plant organelles (i.e., chloroplast, mitochondria and root plastid from plants such as *Oryza sativa*) using *Escherichia coli* as a chassis. They found that the ferredoxin–NADPH oxidoreductases (FNRs) and

their cognate ferredoxin counterparts from plant chloroplasts and root plastids were able to support the activities of both types of nitrogenase, but an analogous ETC module from mitochondria failed to function properly in electron transfer to nitrogenase. The incompatibility of the mitochondrial ETC with nitrogenase could be resolved with a hybrid electron-transfer module consisting of a mitochondrial adrenodoxin oxidoreductase and plant-like ferredoxins. This study demonstrates that ETCs from plants are capable of supporting nitrogen fixation in the absence of the bacterial ETCs.

1.4.4 Creating nitrogen-fixing symbiosis in non-legume plants with rhizobia

The third strategy envisions that plants, especially cereals, can be made to communicate and associate with the nitrogen-fixing bacteria. Conceivably, there are two steps in this process of genetic engineering. In the first step, a bacterium needs to contact and "talk" to a target plant, and then recognize the plant as a suitable host for it to inhabit. In the second step, the plant needs to form a nodule or a nodule-like structure to provide a suitable environment for the bacterium to fix nitrogen (Burén and Rubio, 2018). This strategy requires modification mostly in the desired plant hosts. This would apparently lead to the question regarding the plant host mechanisms of the legume-rhizobia symbiosis. As indicated by Rogers and Oldroyd (2014), at least four co-ordinated genetic programs need to be introduced into the plants if one wants to

establish the legume-rhizobia-like symbiosis in them: recognition of Nod factors, organogenesis of the root nodule, bacterial infection, and establishment of a suitable environment for the activity of nitrogenase inside the nodule. Little has been done in this area. However, some relevant research deserves attention. It was found that Nod factors are structurally similar to another type of symbiosis signals, called Myc factors, which activate the common symbiosis signalling (SYM) pathway during the mycorrhizal symbiosis (Maillet et al., 2011). The SYM pathway exists in both legumes and monocotyledons. It may therefore be possible to engineer the SYM pathway in plants such that it would respond to Nod factors and subsequently activate nodulation-specific gene expression.

1.5 Minimal symbiotic genome of *Sinorhizobium meliloti*

Apparently, the long-term goal of manipulating biological N_2-fixation for human goods can be achieved in several manners as discussed above. However, no matter which approach one chooses to use, a thorough understanding of biological nitrogen fixation, or symbiotic nitrogen fixation in many cases, is very much needed. From our point of view, a promising first step could be establishing a minimized genome which only contains genes essential for symbiotic nitrogen fixation. To simplify the problem, the construction of the minimal symbiotic rhizobia genome should be based on a well-studied model symbiosis system, which is *Sinorhizobium meliloti-Medicago sativa* symbiosis in our case. During the process of building the minimal genome,

novel genes involved in symbiosis might be discovered, and a greater understanding of the mechanism of symbiotic nitrogen fixation and interactions between genes would be facilitated. Taking advantage of the minimized and defined symbiotic nitrogen fixation genes in this genome, further genetic manipulations can be carried out much more efficiently.

1.5.1 *S. meliloti* and its megaplasmid pSymA

Our minimal symbiotic genome project is focused on *Sinorhizobium meliloti*, a Gram-negative bacterium of the class alpha-proteobacteria. The organism is known to be able to establish a symbiotic relationship with legumes of the genera *Medicago, Melilotus* and *Trigonella*.

S. meliloti harbors a 3.7-Mb chromosome as well as two additional replicons, i.e., pSymA, a ~1.4-Mb megaplasmid, and pSymB, a ~1.7-Mb chromid (Galibert et al., 2001). Rhizobia are known to have acquired their symbiotic genes via horizontal gene transfer (HGT), and these genes usually exist either on large plasmids or as symbiotic gene islands (MacLean *et al*., 2007). It has been previously shown by many laboratories, including ours, that both pSymA and pSymB are essential for SNF (diCenzo et al., 2016) of *S. meliloti*. While pSymB serves to facilitate the ability of the microbe to synthesize polysaccharides and metabolize various compounds ((Finan et

al., 2001), most of the genes directly involved in symbiosis and nitrogen fixation are on pSymA (Galibert et al., 2001).

The complete DNA sequence of pSymA from *S. meliloti* Rm1021 (wildtype) is 1,354,226 nt. Although pSymA accounts for ~20% of the total *S. meliloti* genome, the lack of pSymA does not affect the growth of free-living *S. meliloti* cells on nutrient-rich media (Oresnik et al., 2000). However, a strain lacking pSymA is defective in the utilization of certain carbon sources such as trigonelline, inosine, 4-aminobutyrate (Barnett et al., 2001). pSymA contains most of the genes required for nodulation and nitrogen fixation of *S. meliloti*, including the *nif* genes, which are responsible for the synthesis of nitrogenase, nodulation-related genes *nod*, *noe*, *nol* and *fix* genes, most of which are required for microaerobic viability and energy metabolism, and ferredoxin genes, such as *fdxN*, *fdxB*, *fixX*, etc. In addition, pSymA is also known to contain genes responsible for versatile nitrogen metabolism, carbon metabolism, nutrient transport, stress response, and toxin resistances (Barnett et al., 2001).

1.5.2 Previous work in our Laboratory

Our minimal symbiotic genome project began several years ago with an attempt to cure both pSymA and pSymB from *S. meliloti* wildtype strain (diCenzo et al., 2013, 2014, 2016; Milunovic et al., 2014). As mentioned above, symbiotic genes are

generally located on either pSymA or pSymB in *S. meliloti*. Complete removal of the two replicons would provide us with a *S. meliloti* platform where we could build up a minimized and well characterized symbiotic nitrogen fixation system. Megaplasmid pSymA was removed as described previously (Oresnik et al., 2000). The resulting strain remains viable. Unlike pSymA, pSymB is a chromid that displays some characteristics of a chromosome (Harrison et al., 2010). The *engA-tRNA-rmlC* (ETR) region of pSymB arose through translocation from the chromosome, as revealed by phylogenetic analysis, and was found to be indispensable for cell viability (diCenzo et al., 2013). The essentiality of the ETR region would apparently limit the potential of the strain lacking both the pSymB and pSymA for use as the genomic platform for the analysis of symbiotic nitrogen fixation. To solve this problem, the ETR region of *Sinorhizobium fredii* NGR234, an extant ETR region most closely related to the predicted ancestral ETR region, was introduced into the chromosome of *S. meliloti*, rendering the ETR region on pSymB redundant. Subsequently, a derivative of this strain lacking both of the symbiotic replicons (pSymA and pSymB, denoted as ΔpSymAB) was constructed. Transfer of pSymA and pSymB back into the derivative strain restored symbiotic capabilities with alfalfa. However, it was later discovered that the ETR of *S. fredii* NGR234 was unable to replace functionally the *S. meliloti* ETR in symbiosis because the *bacA* gene within the ETR region, which contributes to intracellular viability during symbiosis interactions, could not be altered (diCenzo et

al., 2017). Therefore, the *S. fredii bacA* on the chromosome in the ΔpSymAB strain was replaced by *S. meliloti bacA* (diCenzo et al., 2017).

The strain constructed as described above represents a suitable background platform for the gain-of-function analysis, a step in defining the minimal symbiotic genome. As a start, our Laboratory first constructed a *S. meliloti* pSymA/pSymB deletion library representing > 95% of the 2,900 genes of the symbiotic replicons (Milunovic et al., 2014) and recently screened the library on alfalfa (*Medicago sativa*) for the symbiotic phenotype on nitrogen deficient medium (diCenzo et al., 2016). It turned out that only four loci, which accounted for less than 12% of the pSymA and pSymB sequences, were necessary for SNF. To be more specific, six out of 38 regions, i.e., A117 region (402137 ~ 458915 nt), A118 region (459669 ~ 505334 nt), A121 region (624864 ~ 677156 nt), B108 region (1131169 ~ 1169072 nt), B109 region (1170467 ~ 1204769 nt), and B123 region (1529712 ~ 165255 nt), were found to be essential for the process (Fig. 1). Alfalfa plants inoculated with the strains carrying a deletion of any of these regions were unable to grow normally in the absence of a nitrogen fertilizer. It is worth noting that A116 deletion mutant also showed deficiency in symbiosis probably because it contained essential region A117. The remainder of A116 is likely unnecessary since it is part of A303, whose deletion showed no growth defects.

The data obtained in the above studies have paved the way for the construction of the minimum genome for SNF in *S. meliloti*. It is expected that our knowledge about the molecular basis of the legume-rhizobia symbiosis will be expanded considerably during the process of the construction of the minimum genome.

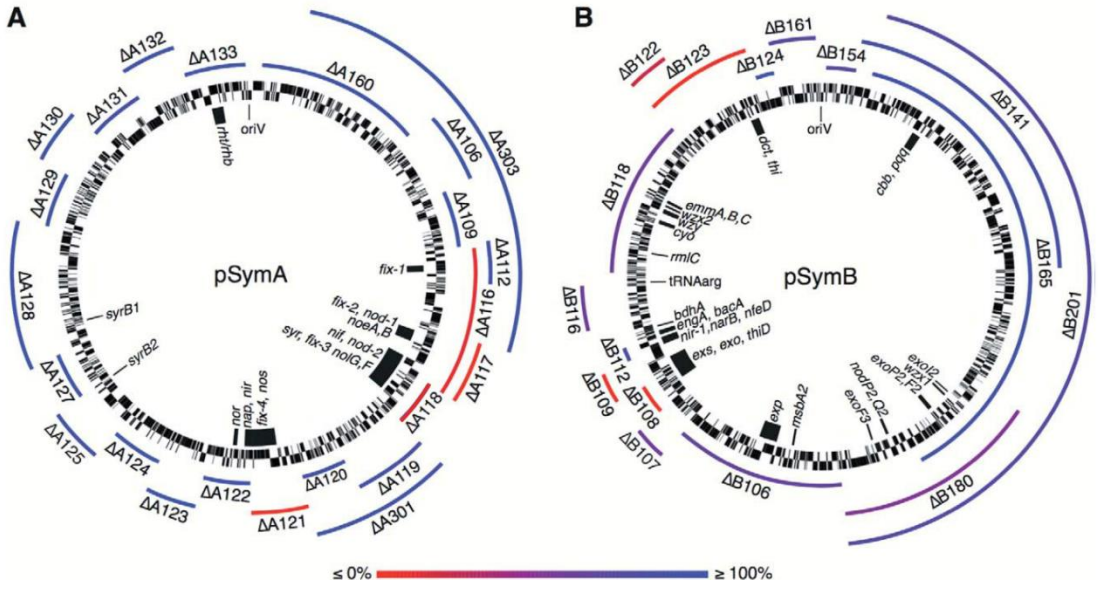


Figure 1. Map of the pSymA and pSymB deletion library. Circular maps of the pSymA (A) and pSymB (B) replicons of *S. meliloti*. The inner circle represents pSymA(B) with individual lines corresponding to annotated genes. The solid outer lines highlight the position of the deletions. The deletions are color coded based on their symbiotic phenotypes, as indicated at the bottom of the figure. Red outer lines indicate that alfalfa plants inoculated with the deletion mutant had a shoot dry weight less than or equal to that of plants inoculated with the control strain which has lost the symbiosis ability with alfalfa. On the other hand, blue outer lines indicate that the shoot dry weight of inoculated alfalfa was ≈100% of that inoculated with the wildtype *S. meliloti*. Several relevant loci are indicated along the inner circle, with shorthand notation referring to gene clusters as follows. **fix-1***: fixI*$_2$*,N*$_3$*,O*$_3$*,P*$_3$*,Q*$_3$*,S*$_2$. **fix-2***: fixK*$_2$*,N*$_2$*,O*$_2$*,P*$_2$*,Q*$_2$*,T*$_2$. **fix-3***: fixA-C,U,X*. **fix-4***: fixG,H,I*$_1$*,J,K*$_1$*,L,M,N*$_1$*,O*$_1$*,P*$_1$*,Q*$_1$*,S*$_1$*,T*. **nif***: nifA,B,D,E,H,K,N,X*. **nod-1***: nodD*$_2$*,L*. **nod-2***: nodA-C,D*$_1$*,D*$_3$*,E-J,M,NP*$_1$*,Q*$_1$. **syr***: syrA,M,B3*. **nos***: nosD,F,L,R,X-Z*. **nap***: napA-F, nnrR*. **nir-1***: nirV,K*. **nir-2***: nirB,D*.

nor*: norB-E,Q.* **rht/rhb***: rhtX,A, rhbA,C-F.* **cbb***: cbbA,F,L,P,R,S,T,X, pqqA-E.* **exp***: wgeA-H, wgdA,B, wggR, wgcA, wgaA,B,D-J.* **exs***: exsA-I.* **exo***: exoA,B,F1,H,I,K-Q,T-Z.* **cyo***: cyoA-D.* **dct***: dctA,B,D.* **thi***: thiC,O,G,E.* This figure is taken from(diCenzo et al., 2016).

1.5.3 Nodule occupancy competitiveness

Nodule occupancy competitiveness is a critical trait for rhizobia to survive in their natural environments. In the rhizosphere, competition exists not only between rhizobia and non-rhizobia, but also among different rhizobia species and other strains. The more competitive a strain is, the larger portion of the total rhizosphere the strain will occupy. This is a problem of practical importance in agriculture. To increase the N_2-fixation efficiency of legume symbioses, the use of high efficiency strains selected in the laboratory is a commonly proposed solution (Triplett and Sadowsky, 1992). However, the high efficiency strains oftentimes fail to increase the legume production as expected. This has been attributed primarily to the failure of the constructed strains to compete with indigenous rhizobia in nodulation occupancy in the field (diCenzo et al., 2018). In other words, a strain with better efficiency in symbiotic nitrogen fixation does not necessarily have a better chance to nodulate. Unlike genes involved in SNF (e.g., *nif, fix, nod*), those responsible for nodule competitiveness are not well characterized either in *S. meliloti* or in other rhizobia. Although some genes, such as *luxR* (Patankar and González, 2009), *putA* (van Dillewijn et al., 2001), *iolA,* and *iolRCDEB* (Kohler et al., 2010), have been claimed to affect nodule

competitiveness, a systematic study on competitiveness-related genes remains to be carried out. It is now time to learn more about the genes contributing to the competitiveness of rhizobia in nodulation and to understand the molecular mechanism of nodule competitiveness on a genome scale as we carry out the minimal symbiotic genome project. Taking advantage of the availability of the pSymA/pSymB deletion library, a rational first step would be screening the deletion mutants for the loss of nodule competitiveness in competition with wildtype *S. meliloti*. However, the screening experiment may not be performed efficiently and on a large scale due to technical difficulties. Traditionally, nodule competition assays are carried out first by picking off nodules from plants inoculated with genetically-marked rhizobia. Then the bacterial cells inside of each nodule need to be isolated and screened on media containing relevant antibiotic or other markers. Therefore, the assays not only require an input of huge amounts of work and time but are also subject to human errors. Moreover, because of the inevitable and high variability of the competition assays, a large number of nodules need to be screened to obtain statistically significant results, further increasing the difficulties of the assays.

1.6 Research objectives

Research described in this thesis aims to achieve the following two main objectives.

1.6.1 Further defining the minimal pSymA region for symbiotic nitrogen fixation by *Sinorhizobium meliloti*

One of the objectives in this study is to further define the minimal symbiotic genome of pSymA. In the previous study of (diCenzo et al., 2016), the SNF essential regions of pSymA were clearly identified and defined, as plants inoculated with each of the three deletion mutants (ΔA117, ΔA118, ΔA121) had less than 20% of the shoot dry weight of those inoculated with the wildtype strain, while plants inoculated with each of the remaining deletions were similar to those inoculated with the wildtype strain in shoot dry weight. These results are not surprising since most of the known symbiotic genes and clusters are located in the three regions. pSymB appears to be more complicated on the other hand. Each of the deletion mutants ΔB108, ΔB109 and ΔB123 showed a highly deficient symbiotic phenotype (less than 20% of the shoot dry weight of the wildtype strain). These regions are thought to be most critically required for SNF. However, mutants with a deletion in regions B106, B107, B116, B118, B154, B161 and B180 produced 60~70% of the wildtype symbiotic shoot dry weight. The importance of these regions to SNF is hard to determine and probably even harder to understand mechanistically. This notion was further strengthened by a recent study in our Laboratory which showed that integration of B108-B109-B123 into the chromosome of ΔpSymB *S. meliloti* was unable to fully restore the symbiotic nitrogen fixation ability of the strain (~16% of the wildtype shoot dry weight) (Kearsley and

Finan, unpublished results). Therefore, the minimal symbiotic pSymB regions required for SNF awaits further clarification. As a first step towards building up the minimal symbiotic genome in *S. meliloti*, we set out to clone A117, A118 and A121, the only regions required for symbiosis in pSymA. A question embedded in this plan is whether the SNF-essential regions from pSymA are sufficient to restore the symbiotic capabilities of the ΔpSymA derivative strain of *S. meliloti* with alfalfa.

In this study, construction of a ΔpSymA *S. meliloti* strain with the three essential regions integrated (A117, A118, A121) was carried out. The project started with the deletion of the region between A118 and A121 to bring together the three regions. Phage lysates were then employed to transduce two FRT sites flanking the SNF region, and then the SNF region was excised as a plasmid using the Flp/FRT site-specific recombination system. The plasmid was later rescued by conjugation to *E. coli* in which it could replicate. The Flp/FRT site-specific recombination system was then utilized again to integrate the plasmid containing the SNF region into a ΔpSymA *S. meliloti* strain on a FRT landing pad located on pSymB(Δ*hypRE::*FRT). Thus, a strain lacking pSymA but containing the three putative pSymA SNF essential regions was obtained. To study the symbiotic phenotype of this strain, plant root dry weight, nodule kinetics, nodule competition and acetylene reduction assays were performed. The host range of this strain was also tested on *M. sativa*, *Medicago truncatula*, *Melilotus officinalis* and *Melilotus alba*.

1.6.2 Establishing a method for determining the competitiveness of *S. meliloti* strains in nodulation based on histochemical staining

Presently used nodule competition assays for *S. meliloti* strains are labour intensive and there is considerable interest in the development of alternate methods. Sessitsch et al. (1996) first designed a nodule competition assay for *Rhizobium tropici* based on histochemical staining. That method employed *gusA*, which encodes the enzyme β-glucuronidase from *E. coli*, and *celB*, which encodes a thermostable and thermoactive β-glucosidase from the hyperthermophilic euryarchaeon *Pyrococcus furiosus*. Because of the lack of β-glucuronidase activity in plant tissues and bacteria (including rhizobia) that interact with plants, nodules formed by *gusA*-marked strains can be easily stained and detected by using specific histochemical substrates, such as Magenta-glcA (5-bromo-6-chloro-3-indoxyl-β-D-glucuronic acid) in their case. The nodules stained red without background interference. On the other hand, β-glucosidase from *P. furiosus* remains highly active following thermal treatment with a half-life of 85 hours at 100°C (Bauer et al., 1996). Since there is no thermostable β-galactosidase in rhizobia or plant hosts, heat-treated rhizobium strains marked with *celB* are readily detected by staining with 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) yielding green/blue color. Thus, rhizobium strains marked separately with the two markers can be counted as

they produce nodules of distinct colors following a ‘gusA staining - thermal treatment - celB staining’ procedure.

As compared to conventional approaches, the histochemical staining assay offers a number of advantages. The procedure is effort- and time-saving and neither nodule picking nor bacterial isolation is required and all nodules can be screened simultaneously. The substrates used in staining procedure are commercially available and generally cheap. No large amount of work and time involved. There are fewer chances for human errors.

Therefore, it was of interest to introduce the histochemical staining method into the *S. meliloti* system. In this study, I employed a cloning procedure in which *gusA* and *celB* genes were either conjugated into *S. meliloti* on a plasmid or integrated into the *S. meliloti* chromosome. The competition assays between wildtype RmP110 and the minimal pSymA strain RmP4291 carrying either of the marker genes were performed to validate the method.

Chapter 2 – Materials and Methods

2.1 Bacteria, phages and plasmids

Bacterial strains and plasmids used in this study are listed in Table 1 with descriptions and references provided.

Table 1. Bacterial strains and plasmids used in this study.

Strain/ plasmid name	Description	Resistance (see table 2 for the full name of abbreviations)	Reference
Sinorhizobium meliloti			
Rm2011	*S. meliloti* with *pstC* frameshift mutation originally isolated in 1939 in New South Wales, Australia, also named as SU47.		
RmG430	Rm1021 *phe232*::Tn5	Sp, Gm	(Oresnik et al., 1994)
RmP110	Rm1021 (streptomycin-resistant derivative of Rm2011), with wild-type *pstC* allele	Sm	(Yuan et al., 2006)
RmP938	RmP110 with pTH1522 (nt 400267–402136), pTH1937 (nt 458916-459668) integrated in pSymA	Sm, Gm, Nm	(Milunovic et al., 2014)
RmP946	RmP110 with pTH1522 (nt 623673–624863), pTH1937 (nt 677157-678150) integrated in pSymA	Sm, Gm, Nm	(Milunovic et al., 2014)
RmP4218	RmP110 with the deletion of region A301 (nt 507338-623673).	Sm, Gm	(diCenzo et al., 2016)
RmP4219	RmP4218 with the scar region (region left between A118 and A121 after the deletion of A301) deleted	Sm	Muhammed and Finan, unpublished
RmP4253	RmP110, ΔpSymA, Δ*hypRE:: FRT*, carrying pTH2505	Sm, Tc	Situ and Finan,

			unpublished
RmP4291	RmP4253, *hypRE::* FRT-A117-A118-A121-FRT	Sm, Gm, Nm, Tc	This work
RmP4306	ΔpSymA, *hypRE::* FRT-A117-A118-A301-A121-FRT	Rif, Gm, Nm, Tc	Situ and Finan, unpublished
RmP4307	RmP4219 with pTH1522 (nt 400267–402136) integrated in pSymA	Sm, Gm	This work
RmP4308	RmP4219, with pTH1937 (nt 677157-678150) integrated in pSymA	Sm, Nm	This work
RmP4309	RmP4219 with pTH1522 (nt 400267–402136) and pTH1937 (nt 677157-678150) integrated in pSymA	Sm, Gm, Nm	This work
RmP4310	RmP4309 carrying pTH2505	Sm, Gm, Nm, Tc	This work
RmP4501	RmP110 (pTH3362)	Sm, Gm	This work
RmP4502	RmP110 (pTH3363)	Sm, Gm	This work
RmP4503	pTH3364 integrated in RmP110	Sm, Gm	This work
RmP4504	pTH3365 integrated in RmP110	Sm, Gm	This work
RmP4505	RmP110 containing chromosomal *gusA*	Sm	This work
RmP4506	RmP110 containing chromosomal *celB*	Sm	This work
RmP4507	phenylalanine auxotroph RmP4291 (φRmG430 *phe232::*Gm^RSp^R)	Sm, Gm, Nm	This work
RmP4508	RmP4291 containing chromosomal *gusA*, made by RmP4507 (φRmP4505)	Sm, Gm, Nm	This work
RmP4509	RmP4291 containing chromosomal *celB*, made by RmP4507 (φRmP4506)	Sm, Gm, Nm	This work
Escherichia coli			

DH5α	*endA1 hsdR17 supE44 thi-1 recA1 gyrA96 relA1* Δ(*argF-lacZYA*) U169 φ80d*lacZ*ΔM15		
J297	DH5α carrying pJQ200sk	Gm	(Quandt and Hynes, 1993)
MT616	DH5α with pRK600 (mobilizing plasmid)	Cm	Strain Collection
M1449	DH5α with pTH2505 (flp plasmid)	Tc	Strain Collection
M928	Rifampicin resistant DH5α	Rif	Strain Collection
M2228	DH5α with pJG592 (I-SceI plasmid)	Tc	Strain Collection
M2608	M928 with pTH3255 (A-SNF plasmid)	Rif, Gm, Km	This work
M2733	DH5α with pTH3362	Gm	This work
M2734	DH5α with pTH3363	Gm	This work
M2735	DH5α with pTH3364	Gm	This work
M2736	DH5α with pTH3365	Gm	This work
Plasmid			
pJG592	pRK7813 with I-SceI meganuclease gene under the control of an IPTG inducible promoter	Tc	Strain Collection
pJQ200sk	p15A origin vector derived from pACYCY184-Gm^R with *Bacillus subtilis sacB*, pUC18 polylinker and *lacZα* fragment	Gm	(Quandt and Hynes, 1993)
pTH1522	Integration and dual reporter vector containing FRT site – ColE1 *oriV*	Gm	(Cowie et al., 2006)
pTH1937	ΔTn903 inverted repeats, pRK2 *oriT*, *nptII* from Tn5	Km	(Milunovic et al., 2014)

pTH2505	pRK7813 with *flp* recombinase gene driven from *pcaD* promoter.	Tc	(White et al., 2012)
pTH3255	A-SNF plasmid, FRT-A117-A118-A121-FRT excised from pSymA of RmP4310	Gm, Nm	This work
pTH3362	nt256639~257195 of Rm1021 S. *meliloti* chromosome, *gusA* with *nifH* promoter of pOPSO253, nt257274~257640 of Rm1021 *S. meliloti* chromosome were golden gate assembled into pOGG024 using BsaI.	Gm	This work
pTH3363	nt256639~257195 of Rm1021 S. *meliloti* chromosome, *celB* with *nifH* promoter of pOPSO254, nt257274~257640 of Rm1021 S. *meliloti* chromosome were golden gate assembled into pOGG024 using BsaI.	Gm	This work
pTH3364	nt256639~257195 of Rm1021 S. *meliloti* chromosome, *gusA* with *nifH* promoter of pOPSO253, nt257274~257640 of Rm1021 S. *meliloti* chromosome were inserted into pJQ200SK with XbaI and PstI.	Gm	This work
pTH3365	nt256639~257195 of Rm1021 S. *meliloti* chromosome, *celB* with *nifH* promoter of pOPSO253, nt257274~257640 of Rm1021 *S. meliloti* chromosome were inserted into pJQ200SK with XbaI and PstI.	Gm	This work

2.2 Bacterial growth and storage

S. meliloti and *E. coli* were cultured in liquid media in test tubes or flasks, or on solid media containing 1.5% (w/v) Difco agar. All *S. meliloti* strains were incubated at 30°C, whereas *E. coli* strains were grown at 30°C or 37°C.

For long term storage, bacterial strains were kept at -80°C in glass vials in LB (for *E. coli*) or LBmc (for *S. meliloti*) containing 7% dimethylsulfoxide (DMSO) and relevant antibiotics. Cells were revived by picking up a small piece of a frozen culture from the storage tube using a sterile stick and streaking immediately on appropriate agar media.

2.3 Bacterial growth media

All media used for bacterial growth were sterilized by autoclaving before use.

Luria-Bertani broth (LB) was routinely used for the growth of *E. coli*. LB contained 1% (w/v) Bacto™ tryptone, 0.5% (w/v) Bacto™ yeast extract, and 0.5% (w/v) sodium chloride in water. For rapid growth or reviving, *E. coli* cells were grown in nutrient-rich SOB media. SOB contained 2% (w/v) Bacto™ tryptone, 1% (w/v) Bacto™ yeast extract, and 1% (w/v) sodium chloride in water.

LBmc, used for the growth of *S. meliloti*, was LB media supplemented with 2.5 mM $MgSO_4$ and $CaCl_2$. When growing *S. meliloti* strains lacking pSymA, 38 µM $FeCl_3$ was added to the media.

M9 medium, used for both *S. meliloti* and *E. coli*, was made of commercially available M9 salts (Difco^{TM} M9 Minimal Salts, yielding 3.39% (w/v) Na_2HPO_4, 1.5% (w/v) KH_2PO_4, 0.25% (w/v) NaCl, and 0.5% NH_4Cl in the final M9 medium) supplemented with 1 mM $MgSO_4$, 0.25 mM $CaCl_2$, 10 ng/mL $CoCl_2$, 0.3 µg/ml biotin and 10 mM of carbon source (glucose, trigonelline or hydroxyproline, etc.) in water.

Antibiotics were added to the media to concentrations listed in Table 2. The concentrations of the antibiotics were sometimes reduced to half when cells were grown in liquid media, as indicated.

Table 2. Antibiotic Concentrations (µg/mL) Used in the study.

Antibiotic	Concentration (µg/mL) for *E. coli*	Concentration (µg/mL) for *S. meliloti*
Chloramphenicol (Cm)	5	-
Gentamycin (Gm)	10	60
Kanamycin (Km)	25	-
Neomycin (Nm)	-	200
Rifampicin (Rif)	50	20

Spectinomycin (Sp)	100	100
Streptomycin (Sm)	-	200
Tetracycline (Tc)	5	5

2.4 Transduction in *S. meliloti* using ΦM12 bacteriophage

S. meliloti bacteriophage ΦM12 was employed as the DNA carrier in the transduction experiments as described previously (Finan et al., 1984). Overnight cultures of donors grown in LBmc without antibiotics were subcultured in LBmc to an OD_{600} of 0.05. This culture was incubated until the OD_{600} reached ~ 0.4, and ~100 μL of *S. meliloti* RmP110 ΦM12 phage (10^{10}-10^{11} PFU/mL) was added. These cultures had generally cleared after overnight incubation at 30°C. Chloroform (200 μL) was then added to kill any remaining cells and the lysates were stored in glass tubes at 4°C.

Recipient *S. meliloti* cells were grown overnight in LBmc. An aliquot (500 μL) of the culture was mixed with an equal volume of 1:25 diluted donor lysate. After incubation for 20 min at 30°C, cells were spun down, washed twice with 1 mL of 0.85% (w/v) NaCl and resuspended in 500 μL of 0.85% (w/v) NaCl. Transductants were selected on appropriate selective media. The recipient culture and the lysate were

spread on the same selective media as controls. 10^{-5} to 10^{-7} dilutions of the recipient culture were also plated for the estimation of transduction frequency.

When necessary, phage titers were determined via plaque assays. 100 μL of 10^{-5} to 10^{-9} dilutions of the phage were mixed with 100 μL of cells from an overnight RmP110 *S. meliloti* culture (OD_{600}~1). The phages were allowed to absorb to the cells for 20 min and 2 mL of 0.7% (w/v) soft agar was added and gently mixed. The resulting mixture was poured onto LBmc plates, and the plates were incubated at 30°C. Plaque numbers were counted for each plate and the phage titer was calculated.

2.5 Bacterial conjugation and Transformation of *E. coli*

Bacterial conjugation experiments were all done via triparental mating employing donor, recipient and a helper *E. coli* strain MT616 (carrying the mobilizing plasmid pRK600). Aliquots from overnight liquid cultures containing appropriate antibiotics were mixed and 25 μL was incubated on LBmc agar medium overnight. The resulting mating spot was suspended in 1mL 0.85% NaCl(w/v) solution and serially diluted to 10^{0}~10^{-6}. While 10^{0} to 10^{-3} dilutions were plated on selective media to select transconjugants, 10^{-4} to 10^{-6} dilutions were put on the media selecting for recipients. Donor plus helper mixture and recipient were also plated on selective media as controls, respectively.

E. coli DH5α was used as the recipient for transformation. Cells were made competent by first subculturing an overnight DH5α culture into LB liquid media (100 mL) to an OD_{600} of ~0.05. The subculture was grown until its OD_{600} reached 0.4, after which the culture was put on ice for 10 min in pre-chilled tubes. After chilling, cells were centrifuged at 4000 x g for 10 min at 4°C and resuspended in 25 mL of ice-cold, sterile 100 mM $CaCl_2$ solution, followed by overnight incubation on ice (kept in a 4°C cold room). Cells were then collected by centrifugation and resuspended in ice-cold, sterile 100 mM $CaCl_2$ + 15% (w/v) glycerol solution. Fresh competent cells were flash frozen in liquid nitrogen and aliquots (200 μL) were stored at -80°C in 1.5 mL microcentrifuge tubes. Competent cells were taken out from the -80°C freezer and given ~20 min to thaw on ice. Next, 20-100 ng of DNA was pipetted into each tube and incubated on ice for 30 min, followed by heat shocking cells in the 42°C water bath for 45 seconds. Cell were put back on ice for 2 min before adding 900 μL of SOB into each tube. Cultures were incubated for 1 hr at 37˚C to allow the antibiotic resistance genes to be expressed. After incubation, cells were collected by centrifugation and resuspended in saline solution [0.85% (w/v) NaCl]. Transformants were selected on appropriate selective media.

2.6 Molecular biology techniques

2.6.1 Plasmid and genomic DNA preparation

When a small scale of *E. coli* plasmid DNA preparation was prepared, the EZ-10 Spin Column Plasmid DNA Miniprep Kit (Biobasic, Canada) was used according to the manufacturer's recommendations. 30-40 μL of elution buffer (10 mM Tris-HCl, pH 8.0) or ddH_2O was passed through the column to elute the plasmid DNA. When extracting plasmid larger than 40 kb or preparing a large amount of plasmid DNA, alkaline lysis was performed with 500 mL-2 L of a grown culture. In short, a 500 mL-2 L culture was first grown to a cell density of OD_{600}~0.4 and incubated for additional 12-16 hr at 37°C. Cells were centrifuged (2700 x g for 10 min at 4°C) and resuspended in 20 mL of Alkaline Lysis Solution I [50 mM glucose, 25 mM Tris-Cl, pH 8.0, 10 mM EDTA (pH 8.0) in ddH_2O], followed by adding 2 mL of a freshly prepared solution of 10 mg/mL lysozyme. Next, 40 mL of freshly prepared Alkaline lysis solution II [0.2 g/L NaOH, 1% (w/v) SDS in ddH_2O] was mixed into the cell suspension in a tube. The sample was gently mixed by inverting the tube several times. After incubation for 5-10 min at room temperature, 30 mL of ice-cold Alkaline lysis solution III was added to the mixture. The mixture was again gently mixed by inverting the tube several times. The mixture was kept on ice for 10 min and centrifuged at >20,000 x g for 30 min. The clear supernatant was transferred to a fresh tube. The volume of the supernatant was measured, and 0.6 volume of isopropanol was added. The sample was mixed, left for 10 min at room temperature, and

centrifuged at 12,000 x g for 10 min. The pellet was washed with 70% ethanol, and the resulting plasmid DNA was dissolved in 10 mM Tris-HCl, pH 8, 1 mM EDTA (TE) after ethanol drained completely. For better purity, DNA was cleaned up again using phenol chloroform extraction. Genomic DNA was prepared using genomic DNA extraction kit from Geneaid™. DNA concentrations were determined by UV spectrophotometry at absorbances of 260/280nm (Biotek Cytation 3 96-well plate Spectrophotometer).

2.6.2 PCR, gel electrophoresis, DNA purification from PCR mixtures and agarose gel

Polymerase chain reaction was employed to amplify DNA fragments for various purposes. For each reaction, PCR was performed in a total volume between 10~25 µL. A PCR *Taq* mastermix (2X PCR BestTaq MasterMix, Applied Biological Materials, Inc.) was used to supply PCR buffer, dNTPs, $MgCl_2$, and *Taq* polymerase. Forward and reverse primers were used at a final concentration of 0.3 µM. Based on the experiment, genomic DNA, plasmid DNA or a bacterial colony was used as template DNA. Genomic DNA template was used at a concentration of 20~40 ng/µL and plasmid DNA template was used at 5~20 ng/µL. For bacterial colony PCR, a colony was first picked up from a plate and suspended in 15 µL of ddH_20, following by

boiling for 15 min at 95°C. 1 μL of the boiled cell suspension was used as the template DNA. dd$H_2$0 was also added to the mixture to the desired final volume.

PCR programs began with a 5-min 94°C initial denaturation and ended with a 5-min 72°C final extension. Each cycle of PCR was consisted of 15 sec of denaturation at 94°C, 30 sec of annealing at 5°C lower than the melting temperature of the primers, and ~1-min extension time for every 2 kb of the PCR product. Usually 30~40 cycles were performed.

The PCR primers utilized in this study are listed in Table 3. Primers were designed manually or by using the Geneious 8 Prime or Benchling (online software). Primer melting temperatures (T_m) were determined using the UC Berkeley Oligo Calculator (http://mcb.berkeley.edu/labs/krantz/tools/oligocalc.html) or by using the above mentioned software. All oligonucleotides were synthesized by Integrated DNA Technologies Inc. (Coralville, Iowa, USA). Primers were supplied as lyophilized powder, which was resuspended in ddH_2O to a concentration of 100 μM and stocked at -20°C. Before use, working stocks of primers were made by diluting stocking solution to a final concentration of 10 μM.

Table 3. All primers utilized in this study.

Primer Name	Sequence (5’ to 3’)	Reference
3432OF	CAAAGGACGGCACTTATTCAACG	(Situ and Finan, unpublished)

3432OR	GATCAAGCTACTCGATCAGGTGC	(Situ and Finan, unpublished)
A118_check-II-F	AGGCTTTCCTCCATGTCCATCAGC	(Muhammed and Finan, unpublished)
A121_check-II-R	CGACAGGAACTTCCGGGGATTGG	(Muhammed and Finan, unpublished)
Left-F	AAGGTCTCAGGAGTAGGGATAACAGGGTAATCGAACGGCATCAACATGACC	This work
Left-R	AAGGTCTCACCTCTGGAAGACTTCGGATGTCGC	This work
Right-F	TTGGTCTCTAATGGCTATTCGTCGTAAAGGCAG	This work
Right-R	TTGGTCTCTAGCGGCAACACCCAATCTCATCG	This work
pBGMM027	TTTGGTCTCAGAGGTGCCGAATTCGGATCCGGA	This work
pBGMM028	TTTGGTCTCTCATTCGCAACGTTCAAATCCGCTC	This work
C-P-F	GCTACTGCAGACGAACGGCATCAACATGACC	This work
C-P-R	ATCGTCTAGAAGGCAACACCCAATCTCATCG	This work
C-P2-F	GCTACTCGAGACGAACGGCATCAACATGACC	This work
V1-F	AAGCTCGTCAAGGAAGTGGG	This work
G-V1-R	TTCCCACCAACGCTGATCAA	This work
C-V1-R	CACCAGACCAGCAGTAACGA	This work
G-V2-F	GAGAGGAGAGCAGGGTCG	This work

G-V2-R	TCGTCTATTGTGGCGATCAGG	This work
G-V3-F	AACAATGAGCTTCCCAGGCA	This work
G-v3-R	GATCCGAAGCGAACACAAGC	This work
C-V3-R	GGGAGCGGATTTGAACGTTG	This work

Agarose gels were made by dissolving agarose powder in Tris-Borate-EDTA (TBE) buffer (5.4 g/L Tris base, 27.5 g/L boric acid, 0.05 M EDTA) to a concentration of 0.7%~1% (w/v) through heating and was then solidified in a plastic cast with a comb placed in the gel to create wells for loading samples. Ethidium bromide (EB), an intercalating dye, was used either by adding directly into the agarose gel before solidification of the gel or soaking the gel in an EB solution (0.5 μg/mL in ddH_2O) for 20 min after electrophoresis. DNA samples were mixed with TriTrack DNA Loading Dye (6X) (Thermo Scientific™) prior to loading into wells. In order to analyze the sizes of DNA samples, a 100-bp or 1-kb ladder (GeneRuler, Thermo Scientific) was always loaded in a well and used for comparison. Electrophoresis was carried out under a constant voltage between 60-100V for 60-120 min. Gel images were visualized with an UV illuminator at 302 nm.

When indicated, DNA fragments produced by PCR were purified using an EZ-10 Spin Column PCR Products Purification Kit (Biobasic, Canada). After gel electrophoresis, the gel piece containing the desired DNA fragment (<250 mg) was

cut off from the rest of gel. DNA was extracted from the gel using QIAEX II Gel Extraction Kit (QIAGEN) according to the manufacturer's recommendations. DNA was usually dissolved in 20-25 μL of ddH_2O.

2.6.3 Restriction enzyme digestion, Ligation and Golden gate cloning

Restriction enzymes and buffers were obtained from New England Biolabs (NEB). Digestion reactions were carried out according to manufacturer's recommendations and protocols. Digestion products were purified using an EZ-10 Spin Column PCR Products Purification Kit (Biobasic, Canada) when required. Ligation reactions were usually performed with an insert-to-vector molar ratio of three in a total volume of 10 μL. Amounts of insert and vector DNA used in reactions were calculated and maximized with an online ligation calculator software (http://www.insilico.uni-duesseldorf.de/Lig_Input.html). Reactions were incubated at room temperature for 2 hr to permit ligation of the fragments. Half of the reaction mixture was then used in transformation. Putative transformants formed on the selective media were checked by PCR on the second day. If no colonies or false positive colonies appeared, the other half of the ligation mixture was transformed into the recipient cells.

The following protocol was taken from Geddes et al. (2019). For maximum efficiency, each DNA component used in golden gate cloning was added in equimolar

concentration of 40 fmol. Beside DNA fragments, 1 μL (20 units) of appropriate high fidelity Type IIS restriction enzymes (BsaI), 1.5 μL of BSA (bovine serum albumin, 2 mg/mL), 1 μL of T4-DNA ligase (highly concentrated, 2,000,000 units/ml) and 1.5 μL of T4 DNA ligase buffer were added in the mixture. Reaction volume was brought to 15 μL with ddH_2O. All enzymes and buffers were ordered from New England Biolabs (NEB). Tubes were prepared on ice before 25 cycles in a thermocycler: 3 min at 37°C then 4 min at 16°C, followed by 5 min at 50°C and 5 min at 80°C. The mixture was then transformed into *E. coli* DH5α. Usually, four putative transformants were used to grow overnight cultures, from which plasmids were extracted. Plasmid sequences were confirmed by gel electrophoresis of plasmid digestion products.

2.7 Plant growth

2.7.1 Seed scarification, sterilization and germination

For *M. truncatula*, and *M. officinalis* a seed chemical scarification assay was employed to encourage germination. Seeds were soaked in 5-10 mL of concentrated, anhydrous sulfuric acid (H_2SO_4) in a 50 mL flask with intermittent agitation for 8-12 min. The progress of the scarification was monitored by observing the gradual appearance of small black spots on the tegument surface. The H_2SO_4 was then removed with a glass pipette and the seeds were immersed in a large amount of

pre-chilled water. The water was decanted, and the seeds were rinsed an additional six times with water.

For seed sterilization of *M. truncatula*, *M. officinalis*, scarified seeds were submerged with 2.5% sodium hypochlorite (bleach) for about 2 min. After the removal of the bleach, the seeds were rinsed 5-6 times using sterile water. For *M. sativa* and *M. alba*, seeds were first washed with 95% ethanol for 5 min, followed by soaking in 2.5% sodium hypochlorite for 20 min with shaking. The seeds were then washed with ddH_2O for one hr, with water replaced every 15 min.

After sterilization, seeds were spread on 1% (w/v) sterile water agar plates evenly with distance and allowed to germinate for two days in the dark at room temperature.

2.7.2 Plant growth condition

Jensen's nitrogen deficient medium (diCenzo et al., 2015) was used for plant growth experiments. It contained (per litre) 1 g of $CaHPO_4$, 0.2 g of K_2HPO_4, 0.2 g of $MgSO_4{\cdot}7H_2O$, 0.2 g of NaCl, 0.1 g of $FeCl_3$, and 1 mL of 1000 x trace mineral solution. The pH of the medium was then adjusted to 7 with 5 M NaOH. For Jensen's agar, the above-mentioned solution was solidified with 1% sterile water agar. The 1000 x trace mineral solution contained (per litre) 1g of H_3BO_3, 1g of $ZnSO_4$, 0.5g of $CuSO_4$, 0.5g of $MnCl_2{\cdot}4H_2O$, 1g of Na_2MoO_4, 10g of Na_2EDTA, 2g of NaFeEDTA, and 0.4g of biotin.

Leonard assemblies were routinely used as the container for plant growth. Each assembly consisted of a plastic jar and a plastic beaker that can be connected to the bottom of the jar. In order to drain water from the bottom beaker, a strip of cotton batting was put into the drainage hole at the bottom of the jar. The assemblies were filled with a 1: 1 (w/w) mixture of quartz sand/vermiculite in the top jar, and 250 ml of Jensen's medium was also poured into the jar. Assemblies were autoclaved before use.

Alternatively, plants were grown in glass tubes. 20 mL of autoclaved Jensen's agar was poured into sterile glass tubes and solidified as a slope. Tubes were closed with plastic foam plugs and the bottom half parts were covered in black paper. After two days of germinating on agar plates, healthy germinated seedlings were moved into Leonard assemblies/tubes. Plants were grown in Conviron growth chamber with a day (18 h, 21°C, 70% humidity) and night (6 h, 17°C, 70% humidity) cycle. Plants were watered with double distilled water roughly every 3-4 days.

Plants were inoculated with *S. meliloti* strains 2 days after seedlings were grown in the Leonard assemblies or glass tubes. Inoculants were first grown for overnight in liquid media without antibiotics the day before inoculation. For Leonard assemblies, 10^8~10^9 cells in 10 mL ddH_2O were inoculated to one assembly, unless otherwise stated. For glass tubes, 10^7 cells in 100 μL ddH_2O were inoculated to one tube.

2.8 Determination of symbiotic nitrogen fixation effectiveness

2.8.1 Plant shoot dry weight assay and acetylene reduction assay

By measuring the shoot dry weights of the plants grown on nitrogen deficient media and inoculated with a desired *S. meliloti* strain, the accumulated effect of the SNF ability of the strain was determined. Following growth for 28-35 days, plant shoots from each jar were separated and put into an envelope made by paper towel. The shoots were kept at 55°C for 10 days before the determination of the shoot dry weight.

Besides reducing N_2 into ammonia, nitrogenase is also known for being able to reduce acetylene (C_2H_2) to ethylene (C_2H_4). The production of ethylene can be measured using gas chromatography (GC). In this assay, three or four roots, which had been grown for 28-35 days after inoculation with a desired *S. meliloti* strain, were collected from each jar (9-12 plants/strain) randomly, placed in a 40-ml glass tube and sealed with a cap. A 5-ml volume of the air inside the tube was replaced by 5-ml acetylene using a syringe and mixed well. A sample of the gas (50 µL) was withdrawn from the tube at 6, 12 and 18 min after the addition of acetylene and injected into the GC through injector port. The amount of ethylene production was detected via the flame ionization detector (FID) and determined by comparing the peak area of the ethylene to a standard curve generated by using standard ethylene gas. Nitrogenase activity was calculated and expressed as nmoles ethylene produced per plant per hour.

When necessary, nodules were taken off from the roots for wet weight measuring and number counting after the acetylene reduction assay. To make sure that no cross contamination between jars occurred, three nodules were washed with distilled water and soaked in 1 % hypochlorite for 15 min. These nodules were then washed twice with LB liquid medium and once with distilled water. The nodules were eventually crushed in 20 μL of distilled water and streaked on LBmc plates. Colonies formed on the plates were screened on selective media or by PCR, if needed.

2.8.2 Nodule kinetic assay

Nodule kinetic assay was carried out mostly as described by (diCenzo et al., 2015) with few changes. Through this method, nodulation efficiency of a strain can be scored by visualizing nodulation process of the strain. In order to observe nodulation, germinated alfalfa seedlings were put on a slope of Jensen's agar in clean glass tubes (1 seedling/tube). All tubes were closed with a plastic foam plug and the bottom parts were covered in black paper as described above. Three independent cultures of each test strain were first grown overnight and then inoculated to 15 plants (6 plants/independent culture, ~10^7 cells per plant). The number of nodules in all tubes was counted every two days for 3 weeks. Results were expressed as percentage of plants inoculated and the number of nodules per plant.

2.8.3 Nodule competition assay, integration of the *gusA* or *celB* marker gene

The *S. meliloti* partner preference of alfalfa was tested by using a newly established method described in this study. The test *S. meliloti* strains were first used as the recipients for transduction through which a *pheA* deletion was introduced by the ΦM12 phage lysate of a *S. meliloti* L-phenylalanine auxotroph (RmG430 *phe-232*::Gm^RSp^R) strain. The resulting transductants were selected on media containing Sp or Gm and patched on M9 glucose minimal medium with and without L-phenylalanine (1 mM). The correct transductants should fail to grow on M9 glucose minimal medium, unless it was supplemented with L-phenylalanine. Those transductants (*phe232*:: Gm^RSp^R) from the first transduction experiment were then again used as the recipient to be transduced with the lysate of RmP4505 (*gusA* marker gene integrated in the chromosome of RmP110, and 1098 bp downstream of the *pheA* gene) or RmP4506 (*celB* marker gene at the same location as *gusA* in RmP4505). Transductants were selected on M9 glucose minimal medium without L-phenylalanine and subsequently screened for Sp sensitivity. Presence of marker genes in the final constructed strains was also confirmed by colony PCR.

In order to minimize the effect of the use of different marker genes, the two marker genes were individually integrated into each of the two competitor strains, thus both forward and reverse marking systems were tested (i.e., *gusA* marked strain A

compared to *celB* marked strain B, *celB* marked strain A compared to *gusA* marked strain B). The two strains carrying different marker genes were grown for overnight in liquid LBmc and mixed at the ratios of 1:5, 5:1 and 1:1. The mixed culture at each ratio was inoculated to 20 alfalfas (4 plants/Leonard jar) with a total cell number of 10^4 per jar.

The following protocol was mostly taken from Westhoek et al., 2017 with a few changes. After 28-35 days of growth, alfalfa roots were separated from shoots and briefly cleaned. In tin foil covered 50 mL Falcon tubes, roots were soaked in about 40 mL phosphate buffer (7 g /L NaH_2PO_4, 7.2 g/L Na_2HPO_4, 1 mM EDTA, pH8, 1% Sarkosyl and 1 mL/L Triton) supplemented with 0.2 mg/ml Magenta-glcA (5-bromo-6-chloro-3-indoxyl-β-D-glucuronic acid) and incubated overnight at 28°C. On the second day, roots were put into new tubes with new phosphate buffer and incubated at 70°C for 1 hour 45 mins to destroy endogenous β-galactosidases. After cooling down, X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) was added to the tubes to a concentration of 0.25 mg/ml. Samples were incubated at 37°C overnight. Nodules formed by *gusA* marked strains showed a deep purple/red color and *celB* marked strains showed a light green/blue color. The numbers of nodules of different colors were counted and recorded.

Chapter 3 – Results

3.1 Cloning of symbiotic essential pSymA regions into a *S. meliloti* strain lacking pSymA

3.1.1 Flp/FRT site-specific recombination

Flp/FRT recombination is a site-directed recombination system. Flp refers to the Flp recombinase, also known as the flippase, which is a tyrosine family site-specific recombinase (Ma et al., 2007). The enzyme was first discovered in *Saccharomyces cerevisiae* (Senecoff et al., 1985). An FRT site is a 34-bp DNA sequence, i.e., 5'-GAAGTTCCTATTCtctagaaaGTATAGGAACTTC-3', in which the two capitalized 13-bp arms flank the 8-bp spacer shown in lower case. When catalyzed by Flp recombinase, the relative position and direction of the two FRT sites will be the determinant of the combining outcome, which could be insertion, deletion, inversion or translocation, as shown in Figure 2. In addition, these combination events may proceed in both directions.

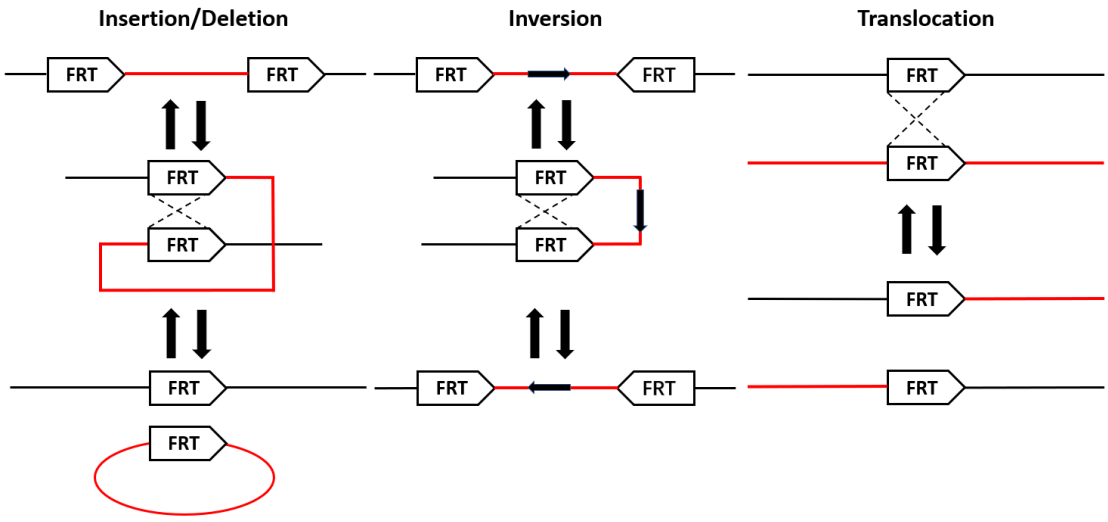


Figure 2. Schematic map of different types of Flp/FRT site-specific recombination as resulted from the orientation and position of FRTs. The relative position and direction of the two core spacers in FRTs determine the outcome of the recombination between the recombining DNAs. If the two spacers are in the same direction on the same sequence, the sequence between the two FRT sites will be excised. If the two spacers are in the opposite direction on the same sequence, the intervening region will be inversed. If the two spacers are located on two separate DNA molecules, translocation will occur. All the above types of recombination may take place in both forward and reverse directions.

Flp/FRT site-specific recombination was previously employed to make large deletions on pSymA and pSymB in *S. meliloti* (Milunovic et al., 2014). This was achieved by integrating, through single crossover recombination, two plasmids pTH1937 (Milunovic et al., 2014) and pTH1522 (Cowie et al., 2006), each of which carried an FRT site, into target DNA flanking the sequence to be deleted. When the two FRT sites were integrated in the same orientation, the region between the two FRT sites was excised upon expression of Flp recombinase, which was introduced into the cells by conjugating a plasmid encoding flippase (pTH2505). In addition to serving as an efficient and quick deletion tool, this Flp/FRT recombination system

leaves an FRT site and sometimes an antibiotic resistance gene that marks the deletion (Milunovic et al., 2014). These markers are useful for further genome editing, such as combining deletions or transduction (recombination) of deletions into desired recipient cells. On the other hand, the FRT site, and antibiotics or other markers left at the deletion site, can also be problematic as they can limit subsequent manipulations.

3.1.2 Deletion of the scar region between A118 and A121

As shown in Fig. 1, three regions, i.e., A117, A118 and A121, on pSymA are required for symbiotic nitrogen fixation. Since A117 is adjacent to A118, the three regions may be viewed as two fragments. Instead of cloning the two fragments separately, a more efficient cloning strategy was designed in which the region in between A117-A118 and A121 was deleted first and the resulting cluster of the three regions would then be manipulated as a single fragment. During the construction of ΔA301 deletion mutant strain, much of the region between A117-A118 and A121 was deleted (diCenzo et al., 2016; Milunovic et al., 2014). However, as a result of using the Flp/FRT site-specific recombination system, a ~11-kb 'scar' sequence that included the *aacC4* gene (Gm^R) and an FRT site was left between A118 and A121. To build the minimized SNF genome, it was desirable to remove these scar sequences for subsequent manipulations and to reduce the total size of A117-A118-A121 region. The scar sequence was removed by *sacB*-mediated double-crossover (Muhammed and

Finan, unpublished work) and the resulting strain, termed RmP4219, was constructed (Fig. 3).

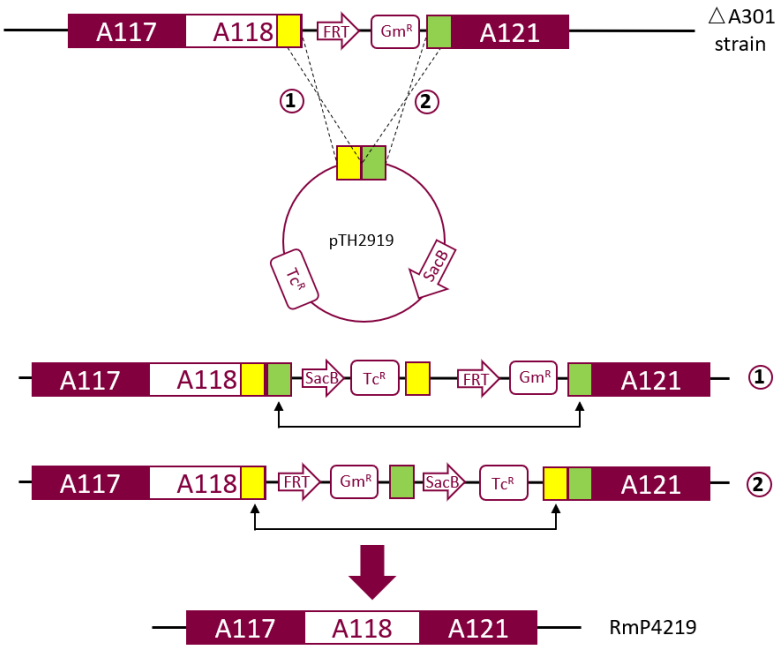


Figure 3. A strategy for the deletion of the "scar" between two A-SNF essential regions using *sacB*. The ΔA301 strain, RmP4218, was previously made using the FLP/FRT recombination system. The resulting scar sequences (including FRT and Gm^R) that mark the ΔA301 region are shown above. These were subsequently removed via double homologous recombination with recombinants selected using the *sacB*/sucrose selection method (Muhammed and Finan, unpublished). The resulting strain was designated RmP4219.

3.1.3 Flanking the A-SNF regions with FRT sites

Since the two A-SNF regions (A117-A118, and A121) are adjacent to each other in strain RmP4219, we sought to capture these regions by using the Flp/FRT recombination system. To do so, we first flanked the A-SNF regions with two FRT sites by transduction. As described above, all of the deletion library strains were made

by using the Flp/FRT recombination system. Among these strains, RmP938, which was used to make the A117 strain, has two FRT plasmids (i.e. pTH1522 and pTH1937) flanking the A117 region. Similarly, RmP946, which was used to prepare A121, contains the same two FRT plasmids flanking the A121 region. If pTH1522 in RmP938 and pTH1937 in RmP946 are both transduced into RmP4219, the excision and capture of A117-A118-A121 region by the Flp/FRT recombination system would become possible (Fig. 4). Thus, RmP938 and RmP946 were employed as donor strains in the transduction experiments with RmP4219 as the recipient.

For transduction with the RmP938 φM12 lysate, transductants were selected on LBmc Gm, since pTH1522 in the donor strain carried Gm^R. The transductants were then screened on LBmc Nm to avoid cotransducing pTH1937. 3 out of 36 transductants were able to grow on LBmc Nm and a single Gm^R Nm^S colony was purified and designated RmP4307. For transduction with the RmP946 lysate, transductants were selected on LBmc Nm since pTH1937 in the donor strain carried the Nm^R marker. There was no need to worry about co-transducing pTH1522 in this case because the size of the DNA to be packaged for the co-tranduction of both pTH1522 and pTH1937 (A117-A118-A119-A120-pTH1522-A121-pTH1937) would be over 260 kb and, therefore, well beyond the packaging limit of phage φM12 (~160kb). Indeed, no growth was observed when screening transductants on LBmc Gm, supporting the above prediction. A single Nm^R Gm^S colony was purified and

designated RmP4308. To obtain the target construct in which the A117-A118-A121 region was flanked with FRT sites, phage grown on RmP4307 were used to transduce Nm^R into RmP4308. Transductants were selected on LBmc Gm Nm. Single colonies of transductants were again streak purified on corresponding selective media, and the purified strain was named as RmP4309.

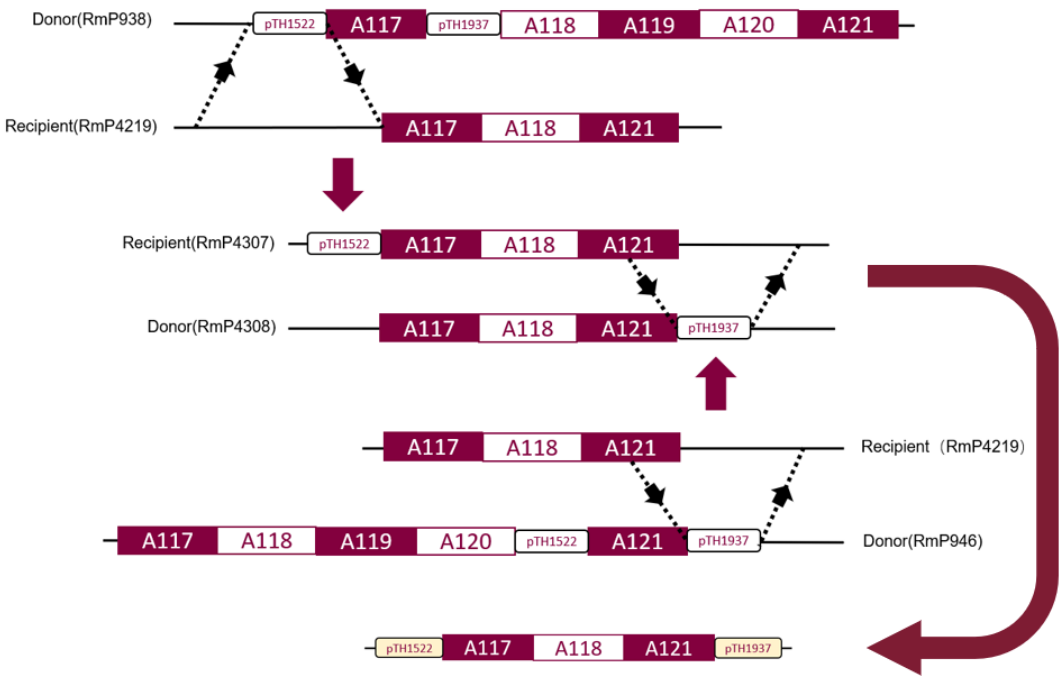


Figure 4. Schematic map showing the integration of two FRT plasmids flanking A117-A118-A121 on pSymA. Map is not drawn to scale. Phage lysates were used to transduce the FRT sites into the scarless strain RmP4219 (see Figure 3) As a result, the pTH1522 from RmP938 (upstream of A117) and pTH1937 from RmP946 (upstream of A121) were transduced into RmP4219. The resulting strain was named RmP4309.

3.1.4 Cloning of the A-SNF region (A117-A118-A121) in *Escherichia coli*

Taking advantage of the fact that the FRT-A117-A118-A121-FRT (A-SNF) region contains two copies of *oriV* (p15A from pTH1937 and pMB1 from pTH1522), a replication origin functional in *E. coli*, two conjugation experiments were carried out to capture the A-SNF in *E. coli* (Fig. 5).

Plasmid pTH2505, a derivative of pRK7813 carrying a Flp recombinase gene and the tetracycline resistant marker (Tc^R) (White et al., 2012) was conjugated from *E. coli* M1449 into *S. meliloti* strain RmP4309 by triparental mating using the helper strain MT616 (carrying the mobilizer plasmid pRK600). The mating was spotted on LBmc and, after incubation overnight, spread on LBmc Tc Gm Nm to select for transconjugants. The resulting RmP4309(pTH2505) strain was named RmP4310.

In a second experiment, the A-SNF region (A117-A118-A121) was excised and then rescued upon transfer into *E. coli*. This was performed by triparental mating of the RmP4310 (donor), *E. coli* MT616 (helper) and *E. coli* M928 (*E. coli* DH5α, the recipient strain carrying the rifampicin resistant marker or Rif^R). Since the transcription of the Flp recombinase gene on pTH2505 was driven from a protocatechuic acid (PCA)-inducible promoter, the mating of this conjugation was spotted on LBmc supplemented with 2.5 mM PCA to induce the expression of Flp recombinase. Following an overnight incubation, the mating spot was then spread on

LBmc Rif Km to select for transconjugants (conjugation frequency: 6.04 x 10^{-3}). After purifying transconjugants on selective media, they were selected again on LBmc Gm Km. A total of 19 out of 25 colonies were able to grow on LBmc with both Gm and Km.

After streak purification on selective media and cure of pRK600 (helper plasmid), polymerase chain reaction (PCR) was performed on colonies to confirm the presence of the A-SNF plasmid in *E. coli* M928 as well as the excision of the A-SNF region from pSymA (Appendix fig. A1 & A2). The *E. coli* M928 derivative carrying A-SNF plasmid was denoted M2608, and the A-SNF plasmid was named pTH3255 (Fig. 6).

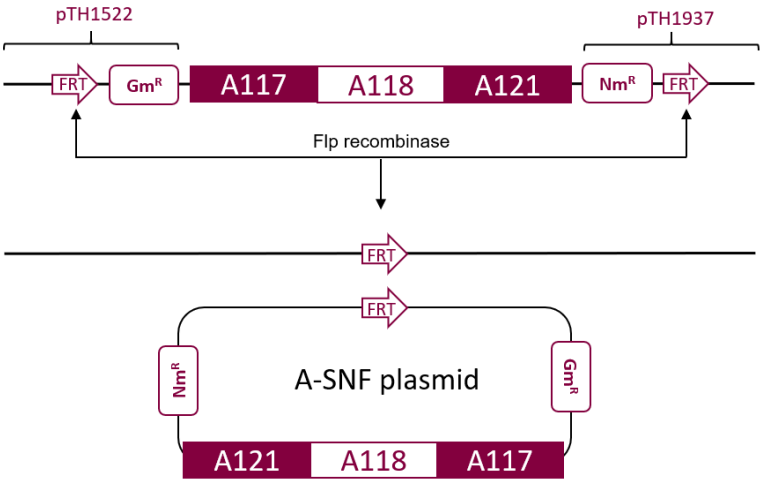


Figure 5. Schematic map showing the rescuing of the SNF region as a plasmid in *E. coli*. Flp recombinase was expressed in *S. meliloti* strain carrying pSymA with the SNF region flanked by FRT sites (FRT-A117-A118-A121-FRT). The SNF region was excised and transferred by conjugation to *E. coli* where it replicated as a plasmid (termed A-SNF plasmid, also named as pTH3255).

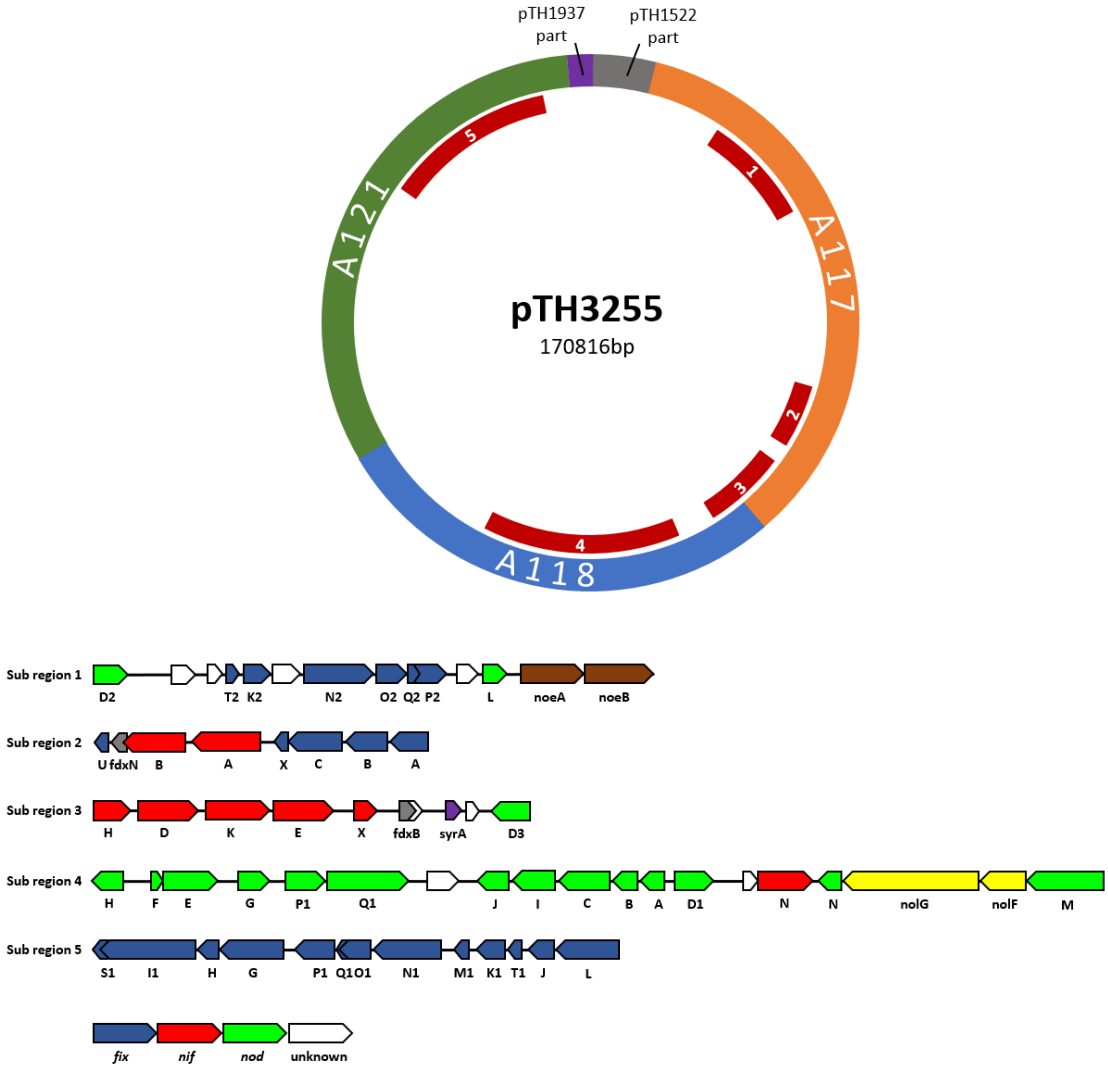


Figure 6. Schematic map of A-SNF containing plasmid pTH3255. The 170-kb sequence of A-SNF plasmid consists of: part of the pTH1522 backbone (PMB1 *oriV* - *aacC4* - *RFP* - *gusA*, shown in purple on the map), part of the pTH1937 backbone (P15A *oriV* - *nptII*, shown in grey on the map) and three putative essential SNF regions of pSymA (A117-A118-A121, shown in orange, blue, and green on the map, respectively). One FRT site is located at the border of pTH1937 and pTH1522, pointing to the left. Five symbiotic sub-regions are also shown inside the circle on the map, whose detailed compositions were shown below. *fix* genes were colored in dark blue, *nif* genes were colored in light red and *nod* genes were marked light green. White boxes represent genes encoding for hypothetical proteins.

3.1.5 Integration of A-SNF into pSymB of a *S. meliloti* mutant strain lacking pSymA (ΔpSymA)

To test whether the rescued A117-A118-A121 plasmid, pTH3255, carried all of the pSymA SNF genes, we sought to transfer pTH3255 into a ΔpSymA *S. meliloti* strain. For these experiments, we used strain RmP4253 (Situ and Finan, unpublished), a *S. meliloti* RmP110 Δ*hypRE* ΔpSymA derivative containing the Flp plasmid pTH2505. The *hypRE* gene is located on pSymB and encodes 4-hydroxyproline 2-epimerase which is required for *S. meliloti* to grow on L-hydroxyproline as a source of carbon (White et al., 2012). The transfer of pTH3255 into RmP4253 results in the integration of A117-A118-A121 into pSymB *hypRE* site through Flp/FRT recombination.

A triparental mating was performed with *E. coli* M2608 (containing pTH3255), *E. coli* MT616 and *S. meliloti* RmP4253 (Fig. 7). The mating was spotted on LBmc supplemented with 2.5 mM PCA to induce the expression of Flp recombinase and following an overnight incubation, transconjugants were selected on M9 minimal medium containing $FeCl_3$, Sm and 10 mM trigonelline. Genes for the trigonelline catabolic pathway are present in the A-SNF region. Of 25 transconjugants examined, all were able to grow on LBmc Gm Nm. The presence of the A-SNF region in the transconjugants was also confirmed by colony PCR (Appendix fig. A3). To verify that pTH3255 integrated into the *hypRE*::FRT site, the purified integrant strain RmP4291

was transduced with phage grown on the wild type Hyp^+ strain RmP110. Transductants selected on M9-Hydroxyproline were screened for their ability to grow on LBmc Gm Nm, and all failed to grow. This is consistent with loss of the integrated pTH3255 and hence that pTH3255 integrated at the *hypRE* locus. Later, the successful construction of RmP4291 was also confirmed by genome sequencing.

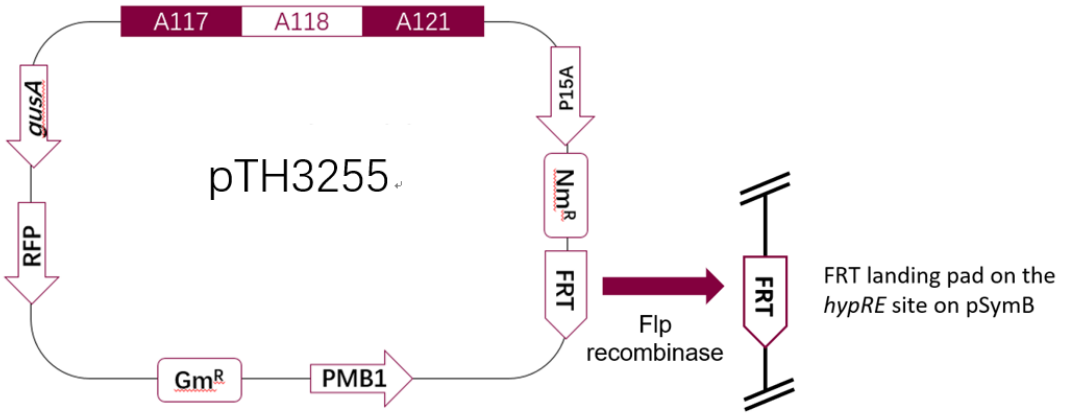


Figure 7. Integration of the A-SNF plasmid into pSymB of a △pSymA strain through Flp/FRT recombination. Flp-dependent integration of pTH3255 into the △*hypRE*::FRT landing pad in strain RmP4253 (△pSymA). The A-SNF plasmid was thus integrated into the pSymB of RmP4253. The resulting strain was designated RmP4291 – this strain was Tc^S due to loss of the Flp plasmid pTH2505.

3.2 Symbiotic phenotypes of *S. meliloti* harboring the minimal symbiotic pSymA

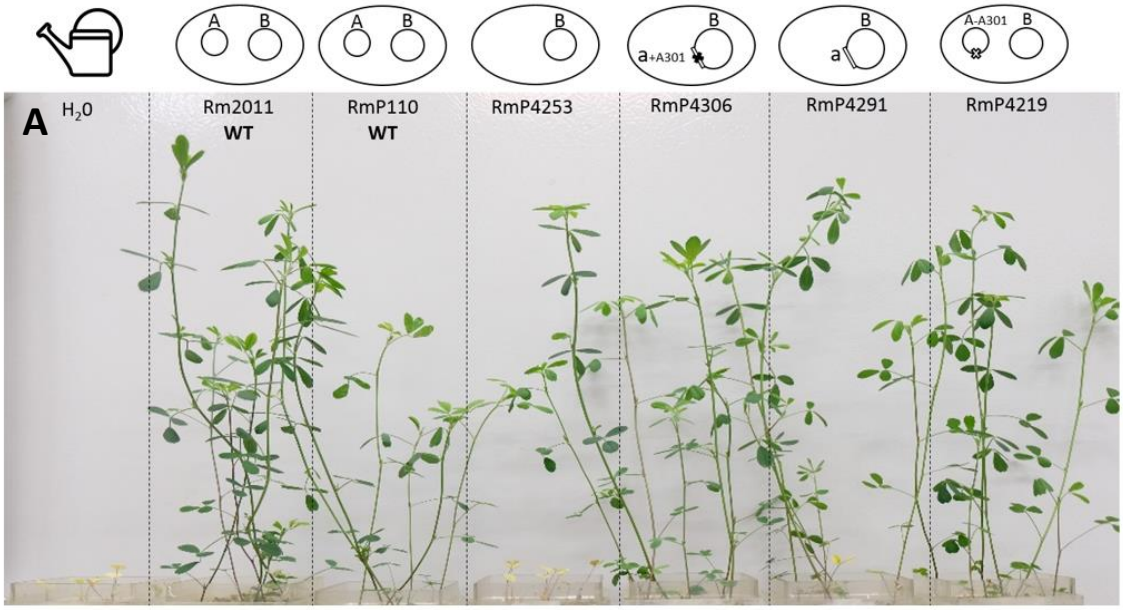
3.2.1 Minimal symbiotic pSymA was sufficient to support ~90% of the wild type symbiotic nitrogen fixation on *Medicago sativa*

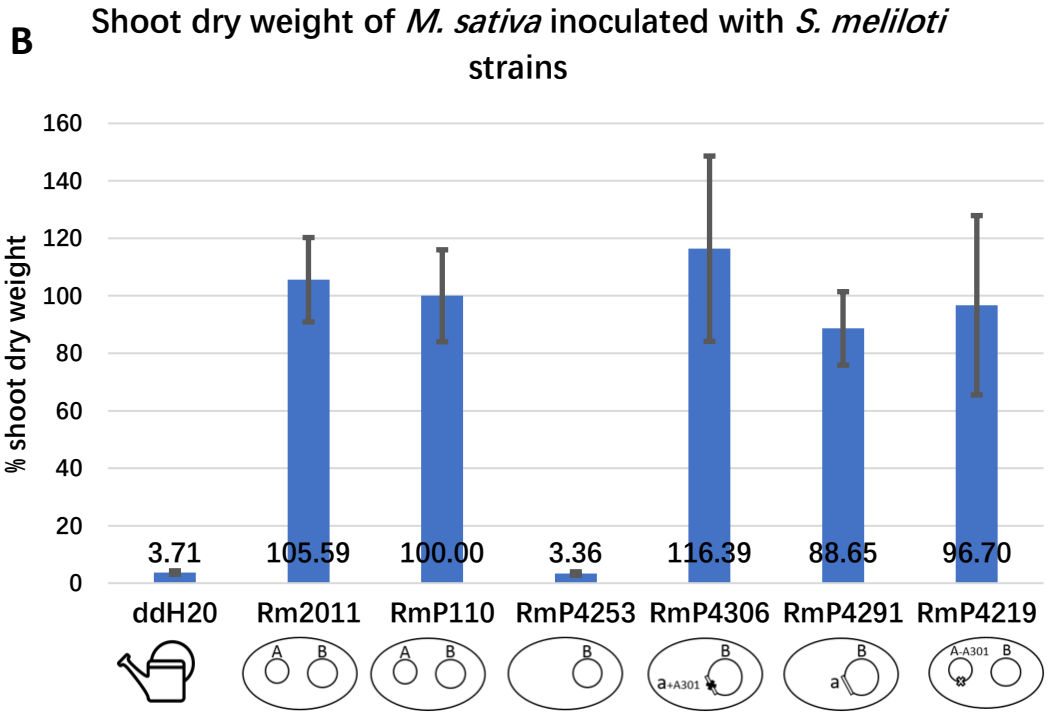
Upon construction of RmP4291 (RmP110ΔpSymA, *hypRE*::FRT-A117-A118-A121-FRT), we sought to determine if the A-SNF region

was sufficient to restore SNF to the ΔpSymA strain with alfalfa as host plant. Strains Rm2011 (wt), RmP110 (wt), RmP4219 (RmP110ΔA301), RmP4253 (RmP110ΔpSymA, *hypRE*::FRT), RmP4306 (RmP110ΔpSymA, *hypRE*::FRT-A117-A118-A301-A121-FRT), a *S. meliloti* derivative strain identical to RmP4291 except for the presence of A301 in the A-SNF region (Situ and Finan, unpublished), and RmP4291. Experiments were performed as described in the Materials and Methods. Three pots containing six plants per pot were inoculated with each strain and the plants were harvested and analyzed 35 days after inoculation. Visual examination of the plants revealed clear differences that were quantified via shoot-dry weight determinations and acetylene reduction assays (Figure 8A, 8B and 8C). As shown in Fig. 8A, the integration of the A-SNF region into pSymB of substantially restored the SNF activity to the ΔpSymA mutant strain. The plants inoculated with the ΔpSymA strain RmP4253 were similar in appearance and dry weight to the uninoculated control – both showing little growth. The dry weight of alfalfa inoculated with RmP4291 was only slightly less than that inoculated with the wild-type strains (~91% and ~86% of the dry weights of the alfalfa inoculated with RmP110 and Rm2011, respectively, Fig 8B). These observations also indicate that all of the SNF-essential genes on the two symbiotic replicons (pSymA and pSymB) can be placed on a single replicon, and therefore, the genomic separation of these genes is not required for SNF. Plants inoculated with RmP4306 (identical to RmP4291 except

for the presence of the A301 region; Situ and Finan, unpublished) or RmP4219 had the same shoot dry weight as that of the wildtype strain (~105% and ~100% of the shoot dry weights of plants inoculated with RmP110 and Rm2011, respectively, for RmP4306, or ~102% and ~97% of the shoot dry weights of plants inoculated with RmP110 and Rm2011, respectively, for RmP4219). As compared to RmP4219 and RmP4306, RmP4291 lacks the A301 region. However, it was shown previously that the deletion of A301 did not affect the symbiotic phenotype of the organism (diCenzo et al., 2016). Therefore, it is possible that the small difference between the shoot dry weight of the alfalfa inoculated with RmP4291 and that with the wild type strains resulted from experimental error or other reasons.

Next, we tested the nodules on the plants for SNF activity by gas chromatography (GC). The amount of acetylene reduced was determined by comparing the peak area of the ethylene in a sample to a standard curve generated by using standard ethylene gas (see Appendix Figure A4 for a standard ethylene curve). As shown in Figure 8C, nodules formed on plants inoculated with RmP4291 were as active in acetylene reduction as those inoculated with the wild type strains, supporting the notion that the introduction of the A-SNF region can restore the SNF ability to the *S. meliloti* strain lacking pSymA. In addition, the nodules of the alfalfas inoculated with RmP4291 display no significant difference from those inoculated with either RmP4219 or RmP4306 in nitrogen fixing activity.


A
H2O
Rm2011
WT
RmP110
WT
RmP4253
RmP4306
RmP4291
RmP4219
A
B
a+A301
a
A-A301


B
Shoot dry weight of M. sativa inoculated with S. meliloti strains
% shoot dry weight
160
140
120
100
80
60
40
20
0
3.71
105.59
100.00
3.36
116.39
88.65
96.70
ddH20
Rm2011
RmP110
RmP4253
RmP4306
RmP4291
RmP4219

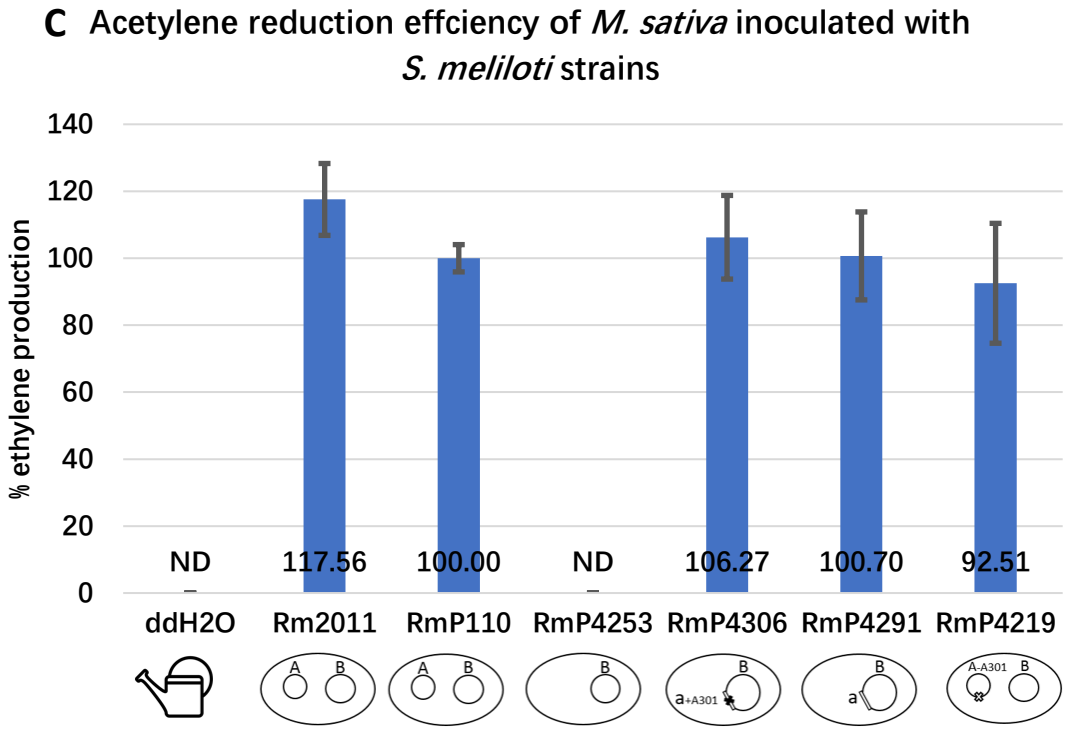


Figure 8. Symbiotic phenotypes of *M. sativa* inoculated with *S. meliloti*. (A) *M. sativa* plants (6 plants/pot) were grown for 35 days after inoculation with *S. meliloti* strains. From left to right, the plants were inoculated with Rm2011 (wt), RmP110 (wt) RmP4219 (RmP110ΔA301), RmP4253 (RmP110ΔpSymA, *hypRE*::FRT), RmP4306 (RmP110ΔpSymA, *hypRE*::FRT-A117-A118-A301-A121-FRT) and RmP4291 (RmP110ΔpSymA, *hypRE*::FRT-A117-A118-A121-FRT). (B) Average shoot dry weights of *M. sativa* inoculated with various *S. meliloti* strains. The average shoot dry weight of *M. sativa* inoculated with RmP110 was 0.1665 g and it was designated as 100%. (C) The average acetylene reduction efficiencies of *M. sativa* inoculated with various *S. meliloti* strains. The average acetylene production of *M. sativa* inoculated with RmP110 was 696.4 nmoles ethylene/plant/hour and this was designated as 100%. Values in (B) and (C) represent an average of triplicate samples with the standard deviation presented as an error bar.

3.2.2 *S. meliloti* containing minimal symbiotic pSymA formed nodules more than wildtype *S. meliloti* on Alfalfa

To study the ability of RmP4291 (RmP110ΔpSymA, *hypRE*::FRT-A117-A118-A121-FRT) to form nodules on *M. sativa*, we performed nodule kinetic assays. RmP110 (wt) was employed for comparison, and RmP4253 (RmP110ΔpSymA, *hypRE*::FRT)) was used as a negative control in the experiments. As shown in Fig. 9, no visible nodules appeared on any plants within the first 5 days after inoculation. Starting from day 6, nodulation became apparent on plants inoculated with RmP4291 and those on RmP110. Strikingly, RmP4291 formed more nodules than wildtype RmP110 at each time point after initial nodules were observed during the period of the experiments. Therefore, RmP4291 appears to be more efficient in nodulation than the wildtype strain.

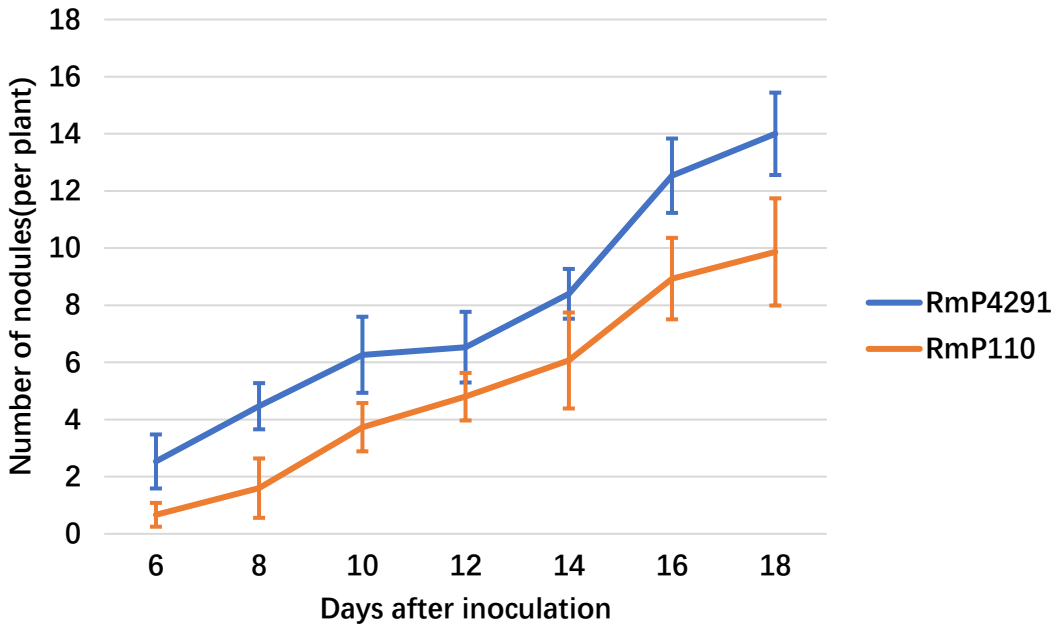
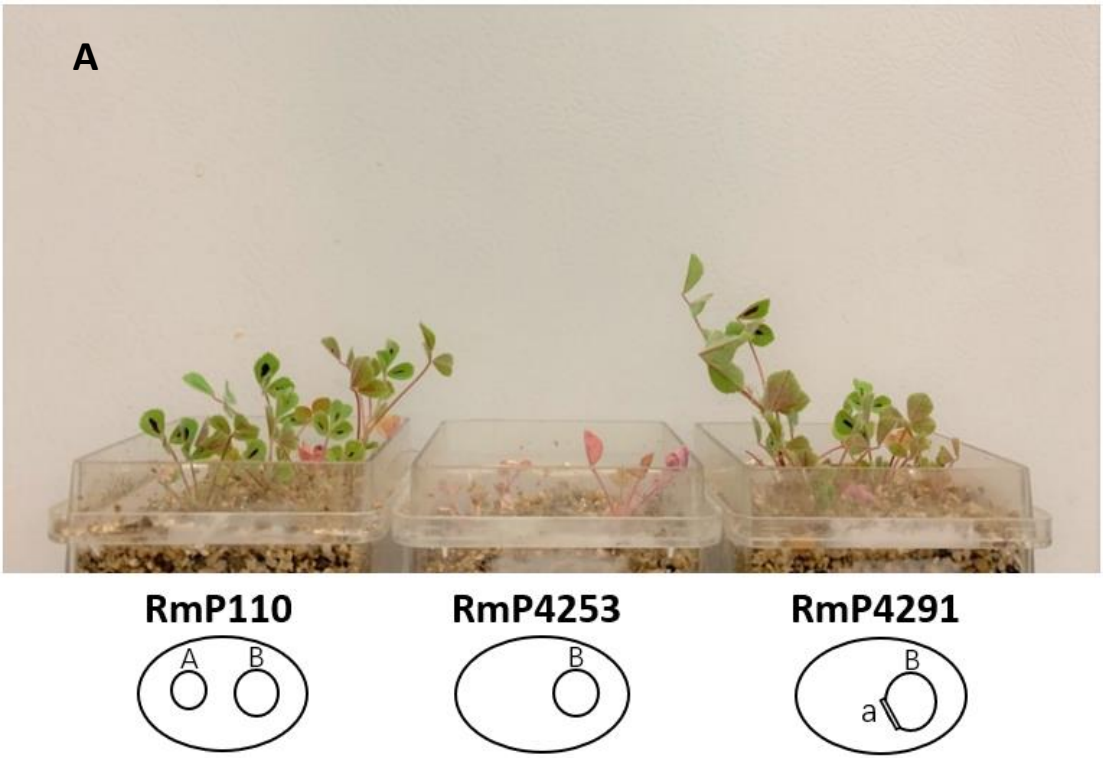


Figure 9. Kinetics of nodule formation by *S. meliloti* containing minimal symbiotic pSymA on *M. sativa*. 3 independent cultures of each test strain were first grown for overnight and then inoculated to 15 plants (5 plants/independent culture, ~10^7 cells per plant). Nodule formation of all inoculated plants were studied for a period of 18 days. Values represent the average of triplicate samples with the standard deviation presented as an error bar.

3.2.3 Minimal pSymA was capable of replacing wildtype pSymA in supporting symbiotic nitrogen fixation by *S. meliloti* on *Medicago truncatula*

After confirming that *S. meliloti* containing only the pTH3255(A117-A118-A121) region of pSymA, instead of wildtype pSymA, was capable of SNF on *M. sativa*, we wanted to learn if the strain had the symbiosis ability on other host plants. So we inoculated *M. truncatula*, another model host plant from the *Medicago* family, with RmP4253 (RmP110ΔpSymA, *hypRE*::FRT), RmP4291 (RmP110ΔpSymA,

hypRE::FRT-A117-A118-A121-FRT) or RmP110 (wt). We found that, while RmP4253 (RmP110ΔpSymA) failed to form nodules or fix nitrogen with *M. truncatula*, the plants inoculated with RmP4291 had a symbiotic phenotype similar to those inoculated with RmP110 (Fig.10A). Furthermore, the shoot dry weights of and acetylene reduction by the plants inoculated with RmP4291 were both over 90 % of those of the plants inoculated with wildtype *S. meliloti* (Figs. 10B, C). Therefore, we conclude that the minimal symbiotic pSymA was capable of substituting for wildtype pSymA in supporting symbiotic nitrogen fixation by *S. meliloti* on *M. truncatula*.



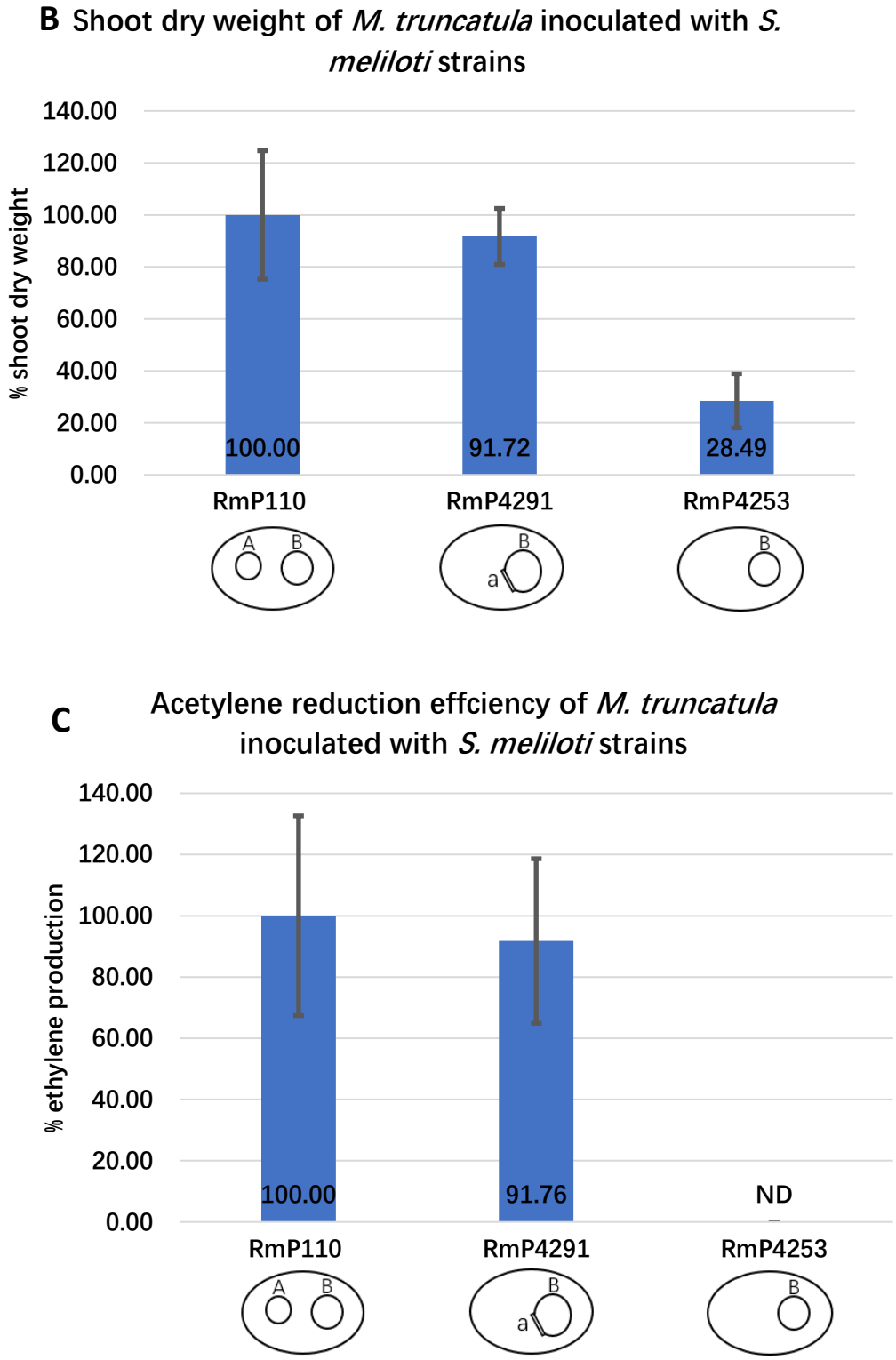
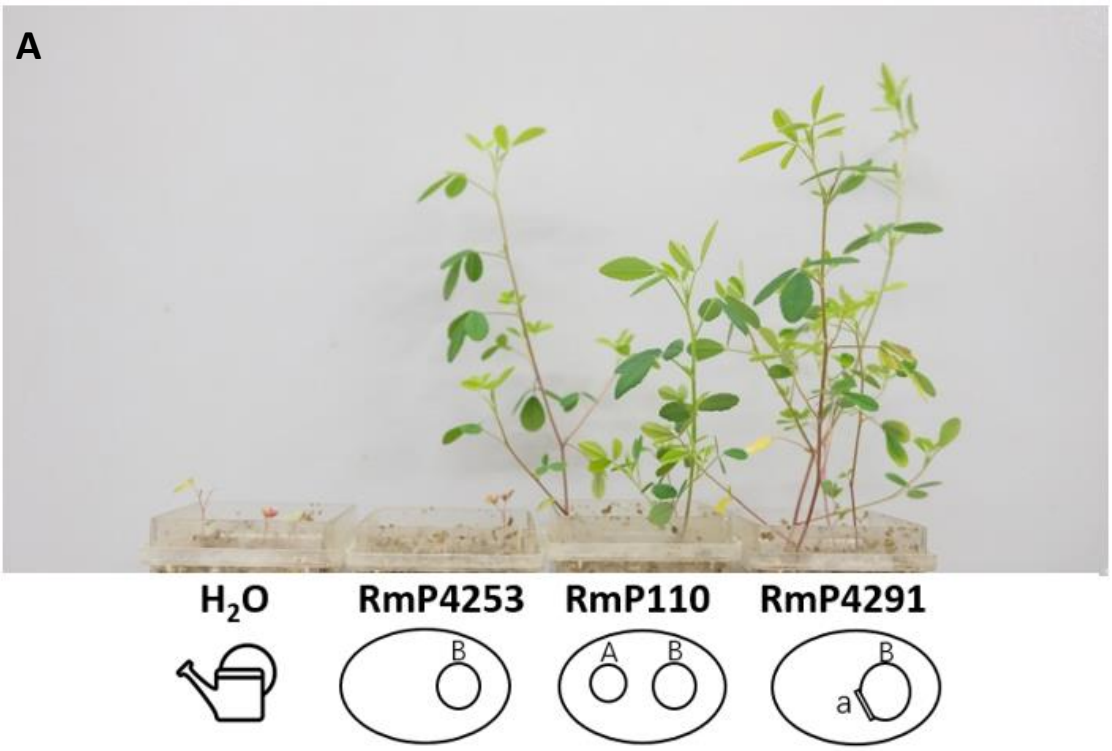


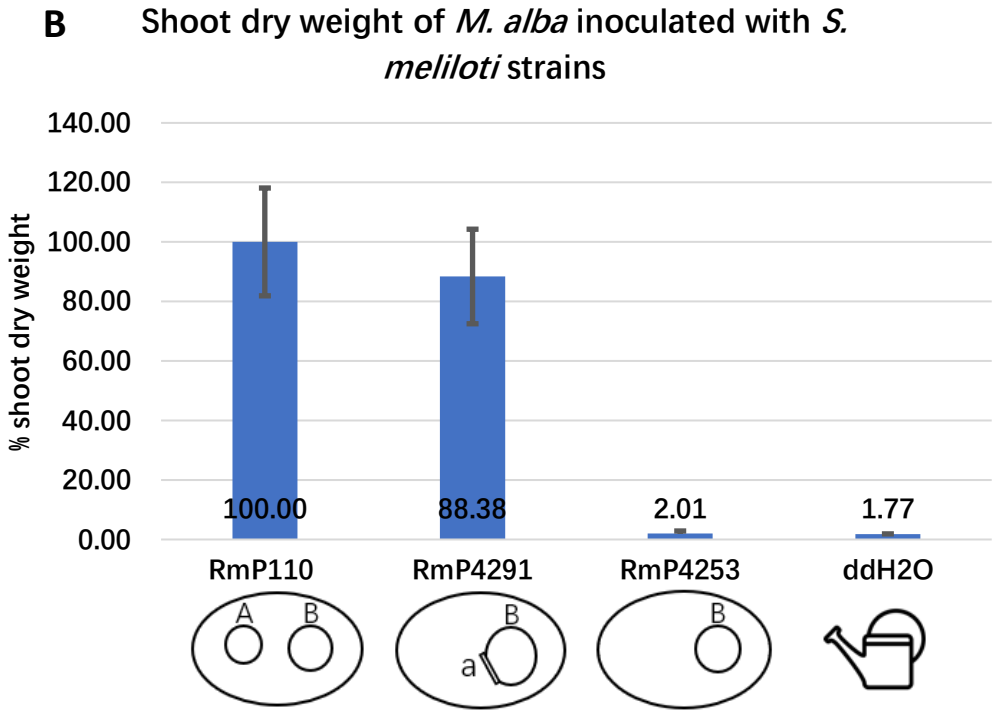
Figure 10. Symbiotic phenotypes of *M. truncatula* inoculated with various *S. meliloti* strains. (A) *M. truncatula* plants that had been grown for 35 days after inoculation with indicated *S. meliloti* strains. From left to right, the plants were

inoculated with RmP110 (wt), RmP4253 (RmP110ΔpSymA, *hypRE*::FRT), and RmP4291 (RmP110ΔpSymA, *hypRE*::FRT-A117-A118-A121-FRT). (B) The shoot dry weights of *M. truncatula* inoculated with various *S. meliloti* strains. The average shoot dry weight of *M. truncatula* inoculated with RmP110 was 0.0638 g and it was designated as 100%. (C) Acetylene reduction by *M. truncatula* inoculated with various *S. meliloti* strains. The average acetylene reduction of *M. truncatula* inoculated with RmP110 was 397.26 nmoles ethylene/plant/hour and it was designated as 100% in the figure. For (B) and (C), values represent an average of triplicate samples with the standard deviation presented as an error bar.

3.2.4 Minimal pSymA restored the wildtype SNF ability back to *S. meliloti* strain lacking pSymA on *Melilotus alba*

Melilotus alba (white sweetclover) is another symbiotic partner of *S. meliloti*. We tested RmP4253(RmP110ΔpSymA, *hypRE*::FRT), RmP4291(RmP110ΔpSymA, *hypRE*::FRT-A117-A118-A121-FRT) and RmP110(wt) in symbiotic nitrogen fixation assays with *M. alba*. We observed that RmP4291 was able to establish symbiotic relationship with the plant, and the shoot dry weight and acetylene reduction efficiency of the plant inoculated with RmP4291 were about 90% of those of the plants inoculated with RmP110 (Fig. 11).


A
H2O
RmP4253
RmP110
RmP4291
B
A
B
a


B
Shoot dry weight of M. alba inoculated with S. meliloti strains
% shoot dry weight
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
100.00
88.38
2.01
1.77
RmP110
RmP4291
RmP4253
ddH2O
A
B
a

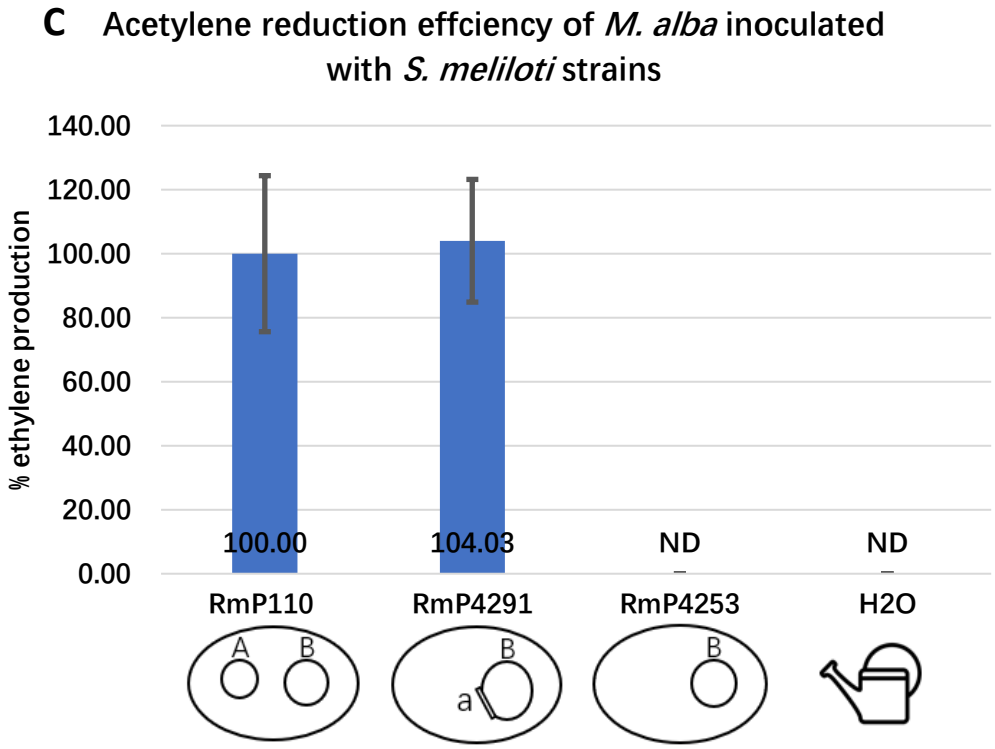


Figure 11. Symbiotic phenotypes of *M. alba* inoculated with *S. meliloti*. (A) *M. truncatula* that had been grown for 35 days after inoculation with various *S. meliloti* strains. From left to right, the plants were inoculated with ddH_2O, RmP4253 (RmP110ΔpSymA, *hypRE*::FRT), RmP110(wt), and RmP4291 (RmP110ΔpSymA, *hypRE*::FRT-A117-A118-A121-FRT). (B) The shoot dry weights of *M. alba* inoculated with various *S. meliloti* strains. The average shoot dry weight of *M. alba* inoculated with RmP110 was 0.24 g and it was designated as 100%. (C) Acetylene reduction by *M. alba* inoculated with various *S. meliloti* strains. The average acetylene reduction of *M. alba* inoculated with RmP110 was 480.25 nmoles ethylene/plant/hour and it was designated as 100% in the figure. For (B) and (C), values represent the average of triplicate samples with the standard deviation presented as an error bar.

3.2.5 Minimal pSymA restored about 60% of the wildtype SNF ability to *S. meliloti* strain lacking pSymA on *M. officinalis*

The same symbiotic assays were also performed on *M. officinalis* using RmP4291 (RmP110ΔpSymA, *hypRE*::FRT-A117-A118-A121-FRT), RmP110 (wt) and

RmP4253 (RmP110ΔpSymA, *hypRE*::FRT) as inoculants. However, unlike *M. sativa*, *M. truncatula* and *M. alba*, inoculated with RmP4291, *M. officinalis* inoculated with the *S. meliloti* strain containing minimal pSymA exhibited significantly reduced SNF. As shown in Fig. 12A, although *M officinalis* inoculated with neither RmP4291 nor RmP110 displayed disease symptoms such as mosaic, yellowed or crinkled leaves, a notable size difference was apparent between them. Moreover, quantitative analyses (Figs. 12B and C) indicate that the plant inoculated with *S. meliloti* containing minimal pSymA had about 63% of the wildtype dry weight and 65% of the wildtype acetylene reduction level of that inoculated with wildtype *S. meliloti*. These results indicate that, while RmP4291 was still able to establish symbiotic relationship with *M. officinalis* and fix significant amounts of nitrogen as compared to the negative controls (i.e., the plant treated with ddH_2O or inoculated with RmP4253), the relationship was somehow impaired. The observed decrease in SNF was unexpected.

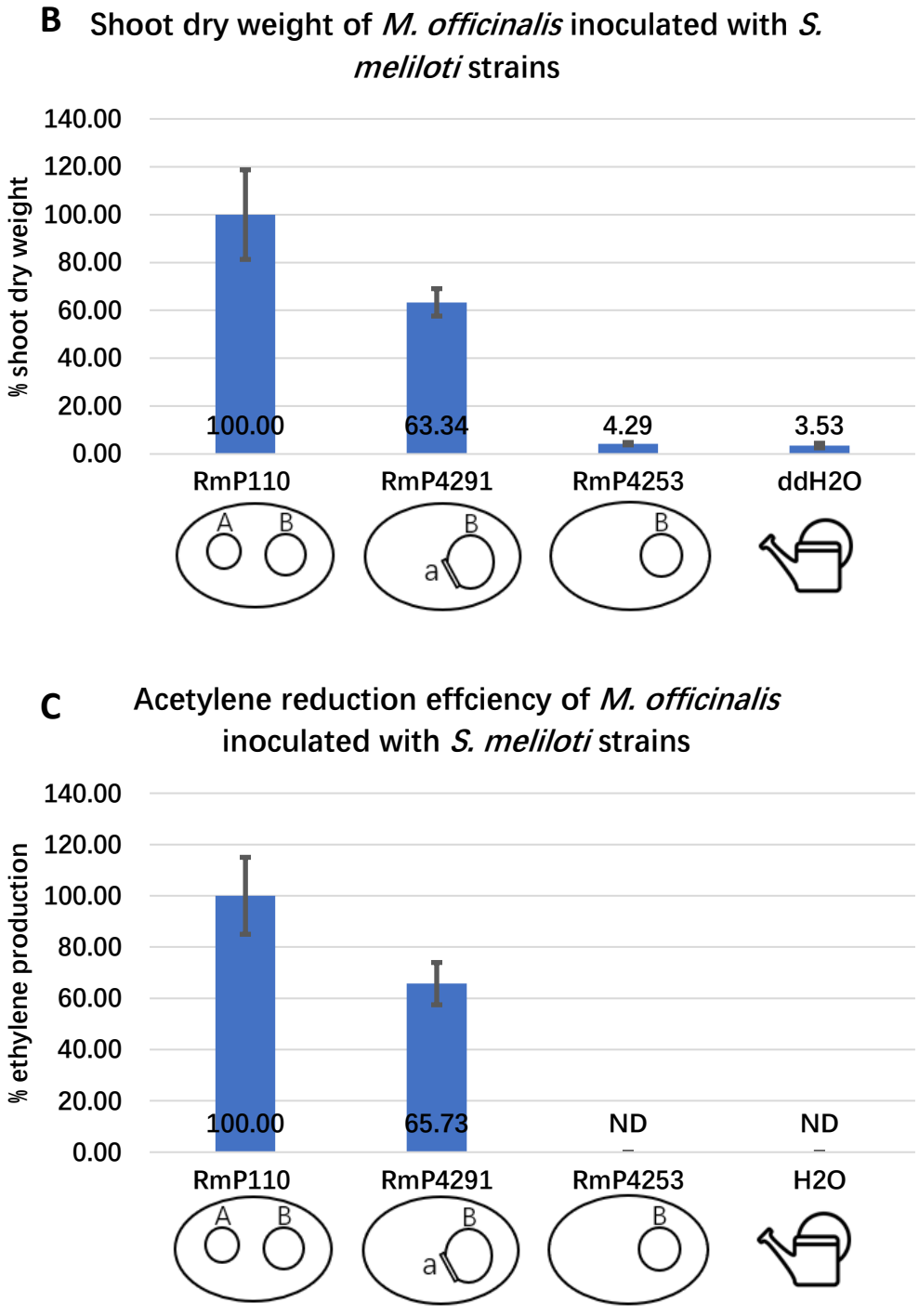


Figure 12. Symbiotic phenotypes of *M. officinalis* inoculated with *S. meliloti*. (A) *M. officinalis* that had been grown for 35 days after inoculation with various *S. meliloti* strains. From left to right, the plants were inoculated with ddH_2O, RmP4253 (RmP110ΔpSymA, *hypRE*::FRT), RmP4291 (RmP110ΔpSymA, *hypRE*::FRT-A117-A118-A121-FRT), RmP110(wt). (B) The shoot dry weights of *M. officinalis* inoculated with various *S. meliloti* strains. The average shoot dry weight of

M. officinalis inoculated with RmP110 was 0.1393 g and it was designated as 100%. (C) Acetylene reduction by *M. officinalis* inoculated with various *S. meliloti* strains. The average acetylene reduction of *M. officinalis* inoculated with RmP110 was 917.24 nmoles ethylene/plant/hour and it was designated as 100% in the figure. For (B) and (C), values represent the average of triplicate samples with the standard deviation presented as an error bar.

3.3 Establishment of a histochemical staining assay on nodules based on glucosidase and glucuronidase reactions in *S. meliloti*

3.3.1 Overall strategy of the integration of the *gusA* and *celB* genes into the *S. meliloti* genome

Introduction of the glucuronidase gene *gusA* and the thermostable glucosidase gene *celB* into *S. meliloti* may be achieved by transferring a plasmid that replicates in *S. meliloti* or by integrating the genes into the *S. meliloti* genome. Although the former approach is simple and efficient, plasmids can be lost in non-selection environments, e.g., nodules. Conceivably, when a partner plant is inoculated with *S. meliloti* containing the plasmid, a portion of nodules formed may become unstained. In addition, conjugation of plasmids usually comes with the introduction of antibiotic selectable markers into the strains, an unwanted consequence. On the other hand, integration of a gene into a genome is usually more stable, and introduction of extra DNA can be minimized by appropriate design. In this thesis, I will describe primarily the introduction of the *gusA* and *celB* gene into *S. meliloti* by genome integration.

Conjugation of two plasmids containing either *gusA* or *celB* into *S. meliloti* and examination of the stability of these plasmids in nodules will also be reported.

Since the minimal symbiotic genome project entails the manipulation and minimization of pSymA and pSymB, the marker genes should ideally be inserted into the host chromosome. After analyzing the sequence of the *S. meliloti* chromosome, we decided to place the maker genes 360 bp downstream of the *pheA* gene (255981 ~ 256835nt), which is involved in L-phenylalanine biosynthesis (Fig. 13). A *S. meliloti* strain lacking this gene would be a phenylalanine auxotroph (Oresnik et al., 1994) that can only survive in media containing L-phenylalanine, since the amino acid is an essential amino acid for the organism.

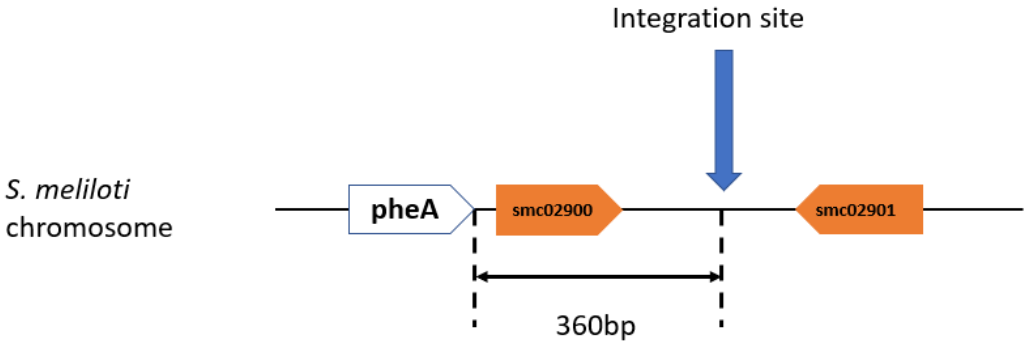


Figure 13. Diagram showing the chromosomal location of integration of *gusA* or *celB* marker genes. *pheA* gene is located at 255981 ~ 256835nt on *S. meliloti* chromosome. Marker genes would be inserted 360 bp downstream of *pheA*, between *smc02900* and *smc02901*.

Taking advantage of the upstream gene *pheA*, we designed a two-step transduction strategy to transduce the *gusA*/*celB* marker genes to *S. meliloti* (Fig. 14). As shown in the diagram (Fig. 14), the target strain will first be transduced with a φM12 lysate of a *S. meliloti* phenylalanine auxotroph (RmG430, *pheA::*Gm^RSp^R) (Oresnik et al., 1994). Transductants will be selected on media containing Gm^R or Sp^R and the resulting strain will then be transduced to prototrophy with a lysate prepared on the *gusA*/*celB* integrated strain. Because *pheA* and *gusA*/*celB* are genomically close to each other, the frequency of co-transduction of the two genomic regions would be high, or close to 100% by Wu's formula (Wu, 1966). The presence of the marker genes in the transductants could readily be verified by colony PCR. Therefore, without the need for repeated integration steps, *gusA* and *celB* genes can be easily transduced to a target strain using this procedure.

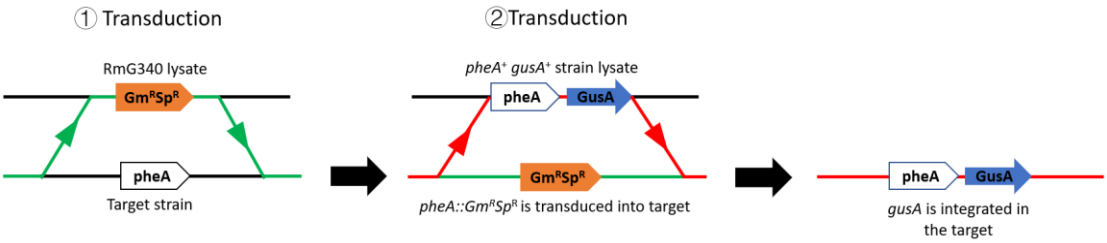


Figure 14. Schematic map showing the two-step transduction strategy of recombining *gusA/celB* into target strains. In the figure, a *gusA* integrant is used as an example. The first transduction occurs between the lysate of the donor RmG340 (*pheA::*Gm^RSp^R) and the target recipient strain. Transductants are selected on media containing appropriate Gm Sp. Resulting *pheA*$^-$ strain serves as the recipient in the second transduction with a *gusA* integrated strain as the donor. Transductants are selected on minimal media without L-phenylalanine, and then checked by colony PCR to verify the presence of the marker gene.

3.3.2 Construction of plasmids containing symbiosis inducible *gusA* and *celB* using golden gate assembly

Golden gate cloning is an efficient molecular cloning method known for its ability to assemble multiple DNA fragments simultaneously at the same time in desired order (Lee et al., 1996). The efficiency of golden gate assembly stems from the use of type IIS restriction enzymes. Unlike the standard type II restriction enzymes, these enzymes cleave outside of their recognition sequence. For example, the recognition site of the type IIS restriction enzyme BsaI is GGTCTC, but instead of cutting at the 6-bp site, it cleaves the four base pairs that are one nucleotide downstream of the recognition site (*5'-GGTCTC(N1)/(N5)-3')*. This feature allows the cut site to be freely customized, on the one hand, and on the other hand the recognition site may be eliminated from the ligated product. Therefore, with appropriate design, several fragments may be simultaneously digested and ligated by putting a unique cut site between every two adjacent fragments and using a single type IIS restriction enzyme and T4 DNA ligase *in vitro*. In this thesis, golden gate assembly was employed to construct two plasmids containing the marker genes *gusA* and *celB*, respectively. For each target plasmid, four fragments with special cut sites were prepared (Fig. 15A) (only fragments for the assembly of the *gusA* plasmid are shown). The fragments for

the construction of *celB* plasmid were the same as those for the preparation of *gusA* plasmid except that the *gusA* fragment was replaced by a *celB* fragment.

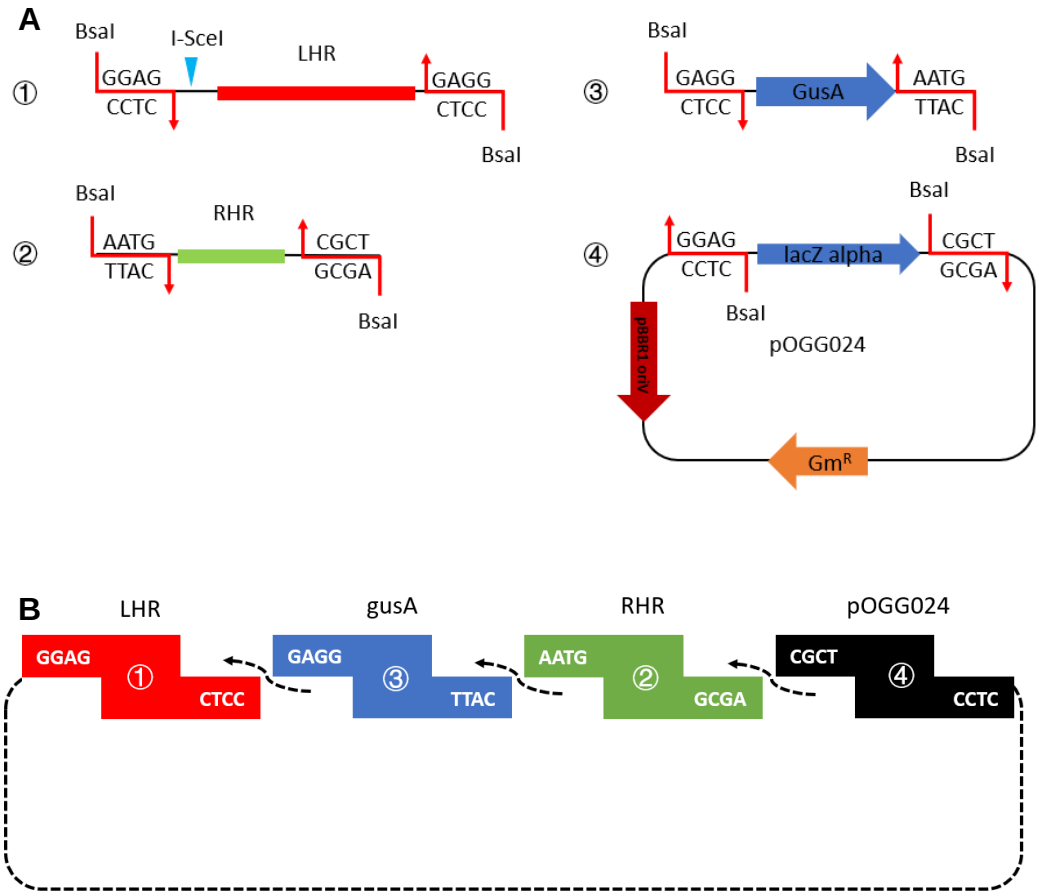


Figure 15. (A) Four fragments used in the construction of *gusA* plasmid by golden gate assembly and (B) schematic representation of predicted ligation process in the assembly of *gusA* plasmid. (A) Golden gate fragments LHR, RHR, *gusA* and pOGG024 with their unique 4 bp-long BsaI cutting sites are shown. Red arrows indicate the cutting direction of BsaI. An I-sceI cut site was added to LHR (Left-F), whose location is shown. The same sets of the fragments were used in the construction of *celB* plasmid except that the *gusA* fragment was replaced by the *celB* fragment (*gusA* and *celB* fragments were amplified by the same primer sets (pBGMM027/pBGMM028)). (B) The correct ligation order of fragments for *gusA* plasmid is shown. During the process of golden gate cloning, BsaI recognition sites were eliminated in the correct ligation products, while the undesired rejoins between the pieces of each fragment still had BsaI recognition sites which would be cut again

in the next cycle. Thus, more and more correct constructs would be produced and accumulated with increasing cycles.

The first and second fragments (i.e., LHR and RHR), used for both plasmids, were amplified from the genomic DNA of *S. meliloti* using the primer sets Left-F/Left-R and Right-F/Right-R, respectively. LHR (557 bp) and RHR (367 bp) are two homologous regions flanking the chromosomal integration site. Other than the BsaI cut site, an I-SceI cut site was also introduced into the LHR fragment by including it on the forward primer downstream of the BsaI site. The third fragment is the *gusA* or *celB* gene along with its *nifH* promoter, amplified from pOPS0253 or pOPS0254 (Fig. A5) with the same primer set (pBGMM027/pBGMM028). pOPS0253 and pOPS0254 were made previously by (Geddes et al., 2019). The *gusA* and *celB* genes in these plasmids are under the control of the *nifH* promoter from *Rhizobium leguminosarum*. The *nifH* promoter is known for being activated by the NifA protein which is expressed only in the nodule environment. Thus, the *gusA*/*celB* genes amplified from these plasmids would potentially be symbiosis inducible and not be expressed in free-living cells. The fourth part, POGG024 (Fig. A5) is one of the golden gate vectors also constructed by (Geddes et al., 2019) It has a broad-host range *oriV* (pBBR) and is a medium copy number vector used for gene expression. The plasmid is capable of replicating in *S. meliloti*. As a vector designed especially for

golden gate cloning, it carries two BsaI sites, both of which would leave unique sticky ends after digestion with BsaI. Meanwhile, a *lacZ alpha* gene, which encodes the α-peptide of β-galactosidase, resides between the two BsaI sites, making the vector a suitable backbone for the construction of our plasmids since the colony containing the successful ligation can be detected by blue-white screening.

After the required fragments were prepared by PCR amplification and the vector was extracted, two golden gate cloning reactions were performed to assemble both *gusA* and *celB* containing plasmids, as described in the Materials and Methods. Fig. 15B shows the predicted ligation process of the *gusA* plasmid. The resulting cloning reaction mixture was then used to transform *E. coli* DH5α cells. Putative transformants were shown as white colonies formed on selective medium (i.e., LB Gm supplemented with 40 µg/mL X-gal). Six of them were then picked up and purified by streaking on the selective media. One single colony from each streaking plate was inoculated into LB Gm liquid medium and grown overnight. Plasmids were extracted from these cultures and verified either by restriction digestion or PCR (see Figs. A6 and A7). The *gusA* containing plasmid was named pTH3362 while the *celB* plasmid was called pTH3363. The maps of the two plasmids are shown in Fig. 16.

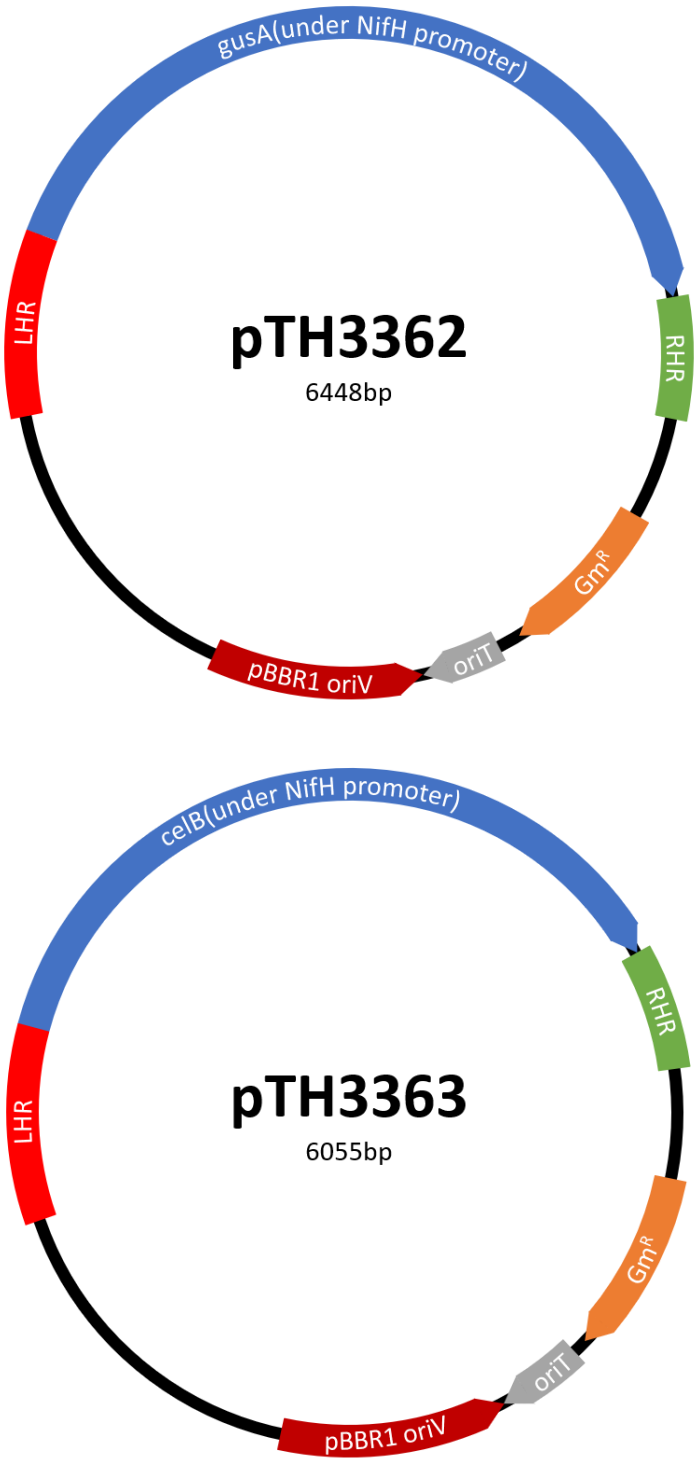


Figure 16. Schematic maps of plasmids pTH3362 (*gusA*) and pTH3363 (*celB*).

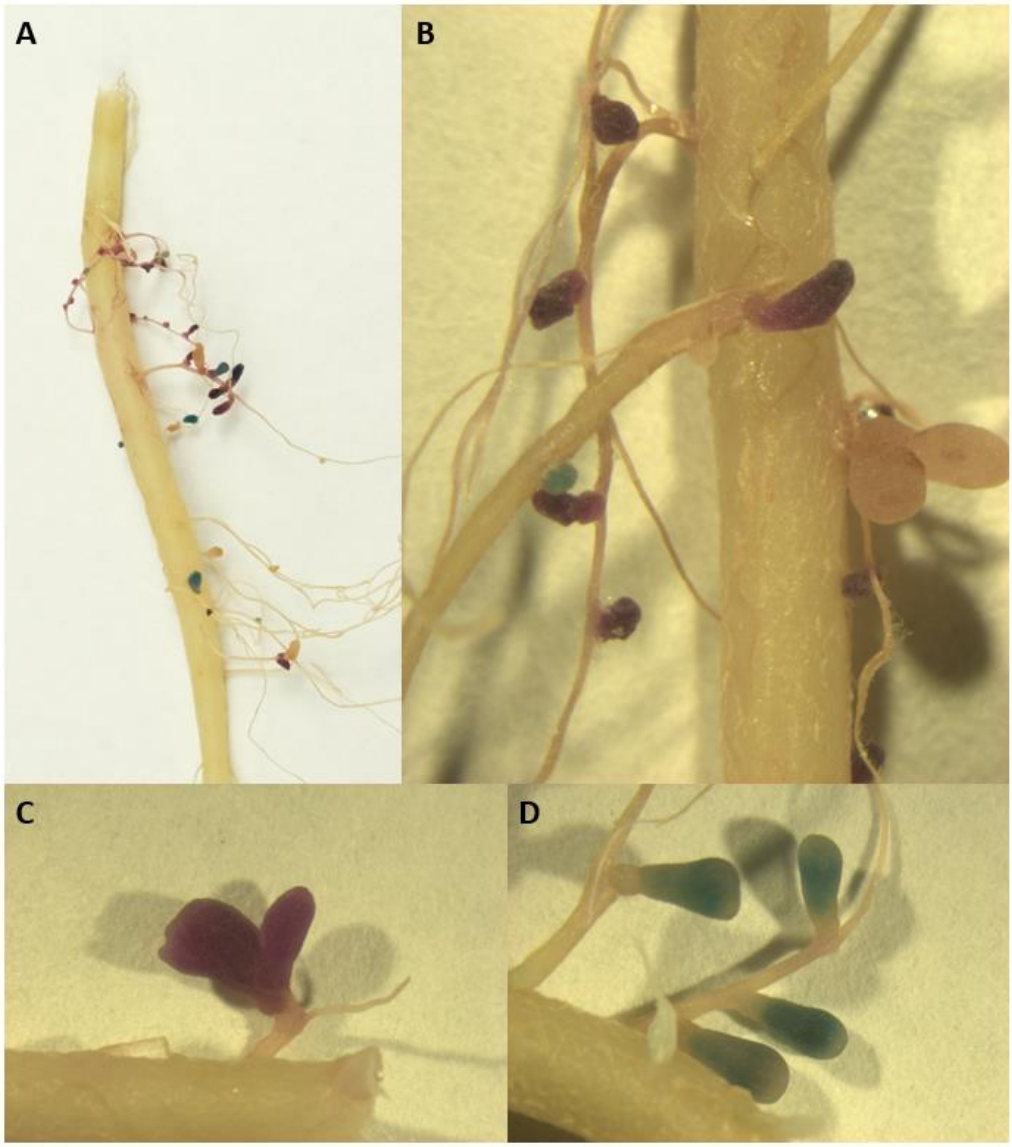
3.3.3 Conjugation of *gusA* and *celB* marker plasmids into *S. meliloti*

Since the origin of replication in both pTH3362 and pTH3363 was derived from pBBR1, the two constructed plasmids can replicate in *S. meliloti*. Therefore, we conjugated the two plasmids into wildtype *S. meliloti* RmP110 through tri-parental mating (MT616 was used as the helper) and transconjugants were selected on LBmc Gm. Single colonies were purified once by streaking. The presence of the plasmid in the transconjugants was confirmed by colony PCR. The two resulting strains were named RmP4501 (pTH3362) and RmP4502 (pTH3363), respectively.

3.3.4 Staining of alfalfa nodules formed by *S. meliloti* with marker plasmids

To determine whether the *gusA* or *celB* gene under the control of the *nifH* promoter from *Rhizobium leguminosarum* could be expressed and whether plasmids pTH3362 and pTH3363 could be stably maintained in *S. meliloti* during symbiosis with alfalfa, a symbiotic assay was carried out between RmP4501 and RmP4502. Overnight cultures the strains were mixed at the ratios of 1:5, 5:1, and 1:1 and the mixtures were inoculated onto 8 alfalfa seedlings (4 plants/Leonard jar) with a total cell number of 10^4 bacterial cells per jar. Alfalfa inoculated with RmP4501, RmP4502 or the wildtype RmP110 was used as a control. After 35 days of growth, plant roots were collected, and stained first with Magenta-glcA

(5-Bromo-6-chloro-3-indoxyl-β-D- glucuronic acid). The treated roots were then heated at 70°C for 1 hr and stained with X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside). After staining, nodules of different colors were counted.



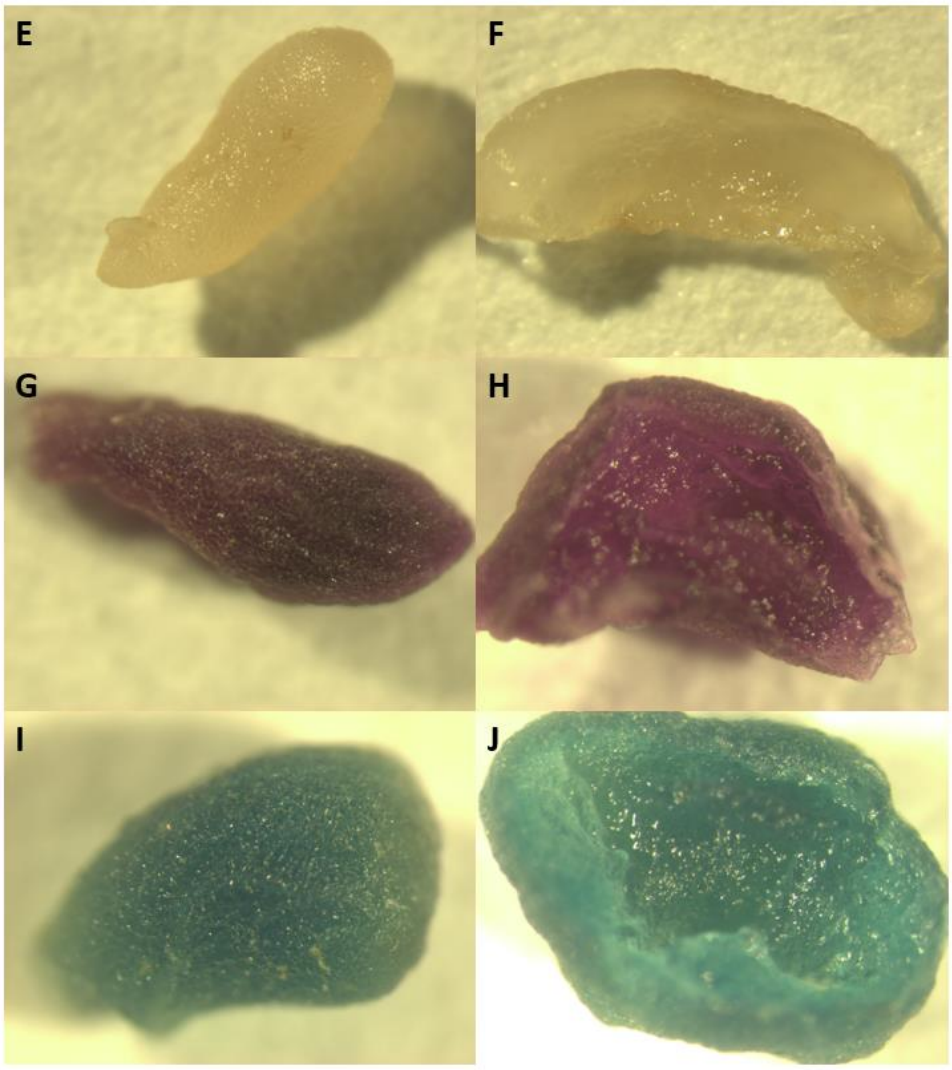


Figure 17. Alfalfa root nodules formed by RmP4501 (*gusA* plasmid) and RmP4502 (*celB* plasmid) after staining. (A) Nodules formed on one root inoculated with both RmP4501 and RmP4502. Nodules were seen in three colors: magenta (RmP4501), green (RmP4502) and white (unstained). (B) An enlarged image of a section of the root in (A). (C) A purple nodule formed by RmP4501. (D) Green nodules formed by RmP4502. (E) The surface of a nodule formed by the wild type RmP110. (F) The cross section of the nodule shown in (E). (G) The surface of a nodule formed by RmP4501. (H) The cross section of the nodule shown in (G). (I) The surface of a nodule formed by RmP4502. (J) The cross section of the nodule shown in (I).

As shown in Fig. 17, staining of nodules formed by *S. meliloti* carrying either plasmid appeared to be successful. Nodules formed by RmP4501 showed a deep magenta color after staining with Magenta-glcA (the first stain), while those formed by RmP4502 turned green/blue only after staining with X-gal (the second stain). It is very easy to distinguish visually nodules stained with the two different colors. In addition, nodules formed by RmP4501 or RmP4502 were stained both on the surface and in the interior. On the other hand, nodules formed by wildtype RmP110 as well as the host plant roots were not stained in our two-step staining procedure. In fact, pink nodules formed by wildtype RmP110 turned white following heat treatment (70°C for 1 hr), a step in the staining procedure, presumably because the leghaemoglobin was denatured by heating.

However, a significant number of unstained nodules (~20% of the total nodules), were found on all plant roots inoculated with RmP4501 and/or RmP4502 (Figs. 17 A and B). Fig. 18 depicts the relationship between the fraction of RmP4501 or RmP4502 in the total inoculum and that of nodules formed by RmP4501 or RmP4502 on alfalfa. RmP4501 and RmP4502 were identical to wildtype RmP110 except for the presence of the marker plasmids. So, it was expected that the two strains would be equally efficient in nodulation. In other words, the fraction of a strain in the inoculum would be similar to that of nodules generated by the strain. (i.e., the curves in Fig. 18 should be y = x). However, the fraction of nodules formed by either RmP4501 or RmP4502

were influenced by the presence of unstained nodules to varying degrees. The presence of these colorless nodules was taken as the evidence for the loss of marker plasmid (pTH3362 and pTH3363) during symbiosis. We also noticed that, although the two marker plasmids were essentially identical at the nucleotide sequence level except for the presence of different marker genes (i.e., *gusA* and *celB*), RmP4502(RmP110 (pTH3363-*celB*)) seemed to lose the marker plasmid more readily than RmP4501 (RmP110 (pTH3362 - *gusA*)). In general, the higher portion of RmP4502 in the inoculum, the more unstained nodules were found on the roots. On plants inoculated only with RmP4502, nearly 30% of the nodules were white. In comparison, ~13% of the nodules on plants inoculated with only RmP4501 were unstained. Therefore, we conclude that the *gusA* and *celB* genes in our constructs were expressed and functionally active in *S. meliloti*, and plasmids pTH3362 and pTH3363 were not stably maintained in *S. meliloti* nodules on alfalfa and thus were not appropriate for use in the histochemical staining assay of nodules.

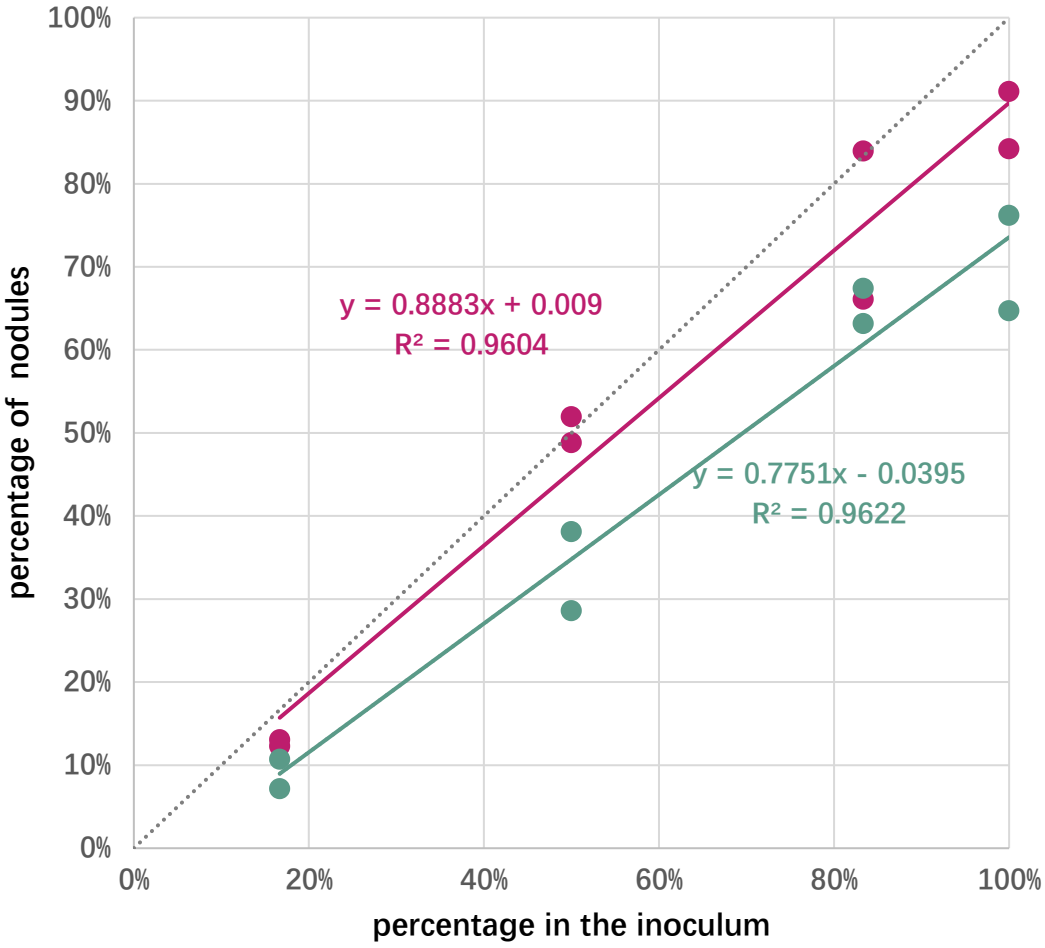


Figure 18. The relationship between the proportion of RmP4501 and RmP4502 in the total inoculum and the nodule percentage formed by them. Data obtained from nodules formed on alfalfas co-inoculated with RmP4501 (*gusA* marked RmP110) and RmP4502 (*celB* marked RmP110). The magenta dots represent the percentage of RmP4501 in the inoculum compared to the percentage of magenta nodules in total nodules. While the green dots represent the percentage of RmP4502 in the inoculum compared to the percentage of green nodules in total nodules. The dotted line represents the relationship of Y = X. 8 plants and 506 nodules were represented in the data.

3.3.5 Construction of the *gusA* and *celB* plasmids for integration into the *S. meliloti* genome

The pTH3362 and pTH3363 marker plasmids are self-replicating in *S. meliloti* and thus could not be used directly for recombination of *gusA* or *celB* into the *S. meliloti* genome. Therefore, we employed the suicide vector PJQ200SK as the new backbone for the plasmid reconstruction for the following considerations. Plasmid pJQ200SK is unable to replicate in *S. meliloti* and the *sacB* gene on this plasmid facilitates the selection of double crossover recombinants. The LHR-*gusA*(*celB*)-RHR regions from pTH3362 and pTH3363 were amplified by PCR using the primer set C-P-F/C-P-R or C-P2-F/C-P-R, respectively. The resulting LHR-*gusA*-RHR PCR product was cloned into PJQ200SK via XbaI and PstI restriction sites, while the LHR-*celB*-RHR PCR product was inserted into PJQ200SK via XhoI and XbaI restriction sites. The two ligated plasmids were transformed into *E. coli* DH5α. and PJQ200SK containing LHR-*gusA*-RHR or LHR-*celB*-RHR were obtained and named pTH3364 or pTH3365, respectively (evidences of the correct construction of plasmids were provided in Appendix, Figs. A8 and A9). Maps of the plasmids are shown in Fig. 19. Plasmids pTH3364 and pTH3365 were then introduced into wildtype *S. meliloti* RmP110 by tri-parental mating from *E. coli* donor strains M2735(pTH3364) or M2736(pTH3365) as the donor and single cross-over homologous recombinants were selected on LBmc Gm in both experiments (Fig. 20).

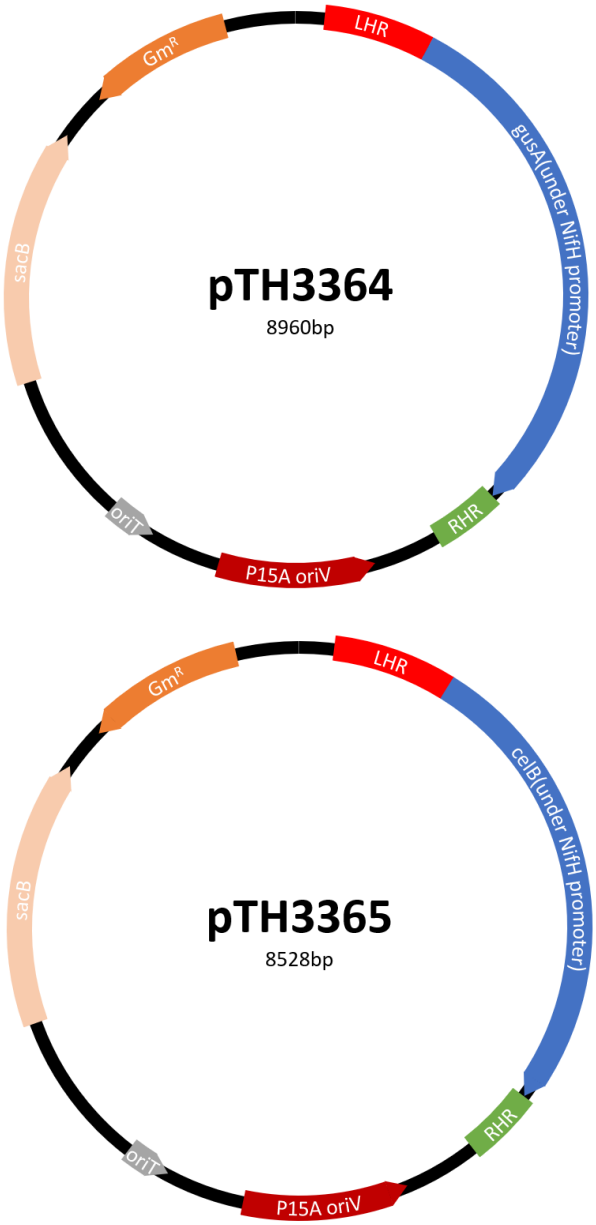


Figure 19. Schematic maps of plasmids pTH3364 and pTH3365. The plasmids were constructed through ligation between pJQ200SK and LHR-*gusA*-RHR (pTH3364) or LHR-*celB*-RHR (pTH3365).

3.3.6 Integration of the *gusA* and *celB* genes into the *S. meliloti* chromosome

Once pTH3364 and pTH3365 were available, we introduced them into wildtype *S. meliloti* RmP110 by tri-parental mating between M2735(pTH3364) or M2736(pTH3365) as the donor, RmP110 as the recipient and MT616 as the helper. Single homologous recombinants were selected on LBmc Gm in both experiments (Fig. 20). To select double homologous recombinants, both the *sacB*/sucrose and the I-SceI selection systems were employed in both cases. For *sacB*/sucrose selection, transconjugants were first streaked on LBmc supplemented with 5% sucrose. After 3 days of growth, colonies appearing on the sucrose media were then screened on LBmc Gm and LBmc sucrose. Colonies that failed to grow on LBmc Gm but grew on LBmc sucrose were then checked for the successful recombination by colony PCR (Appendix, Figs. A10 and A11). For I-SceI selection, plasmid pJG592, which carries an I-SceI endonuclease gene under the control of an IPTG inducible promoter, was used. pJG592 was conjugated into the single homologous recombinant cell with the help of pRK600. Transconjugants were screened on media supplemented with 0.5 mM IPTG and later confirmed by colony PCR. We succeeded in both recombination experiments. However, only the recombinants obtained through *sacB*/sucrose selection were maintained and will be used in future experiments to avoid the step of curing pJG592. As the final constructs, RmP110 with *gusA* integrated in the

chromosome was named RmP4505, and that with *celB* in the genome was named RmP4506.

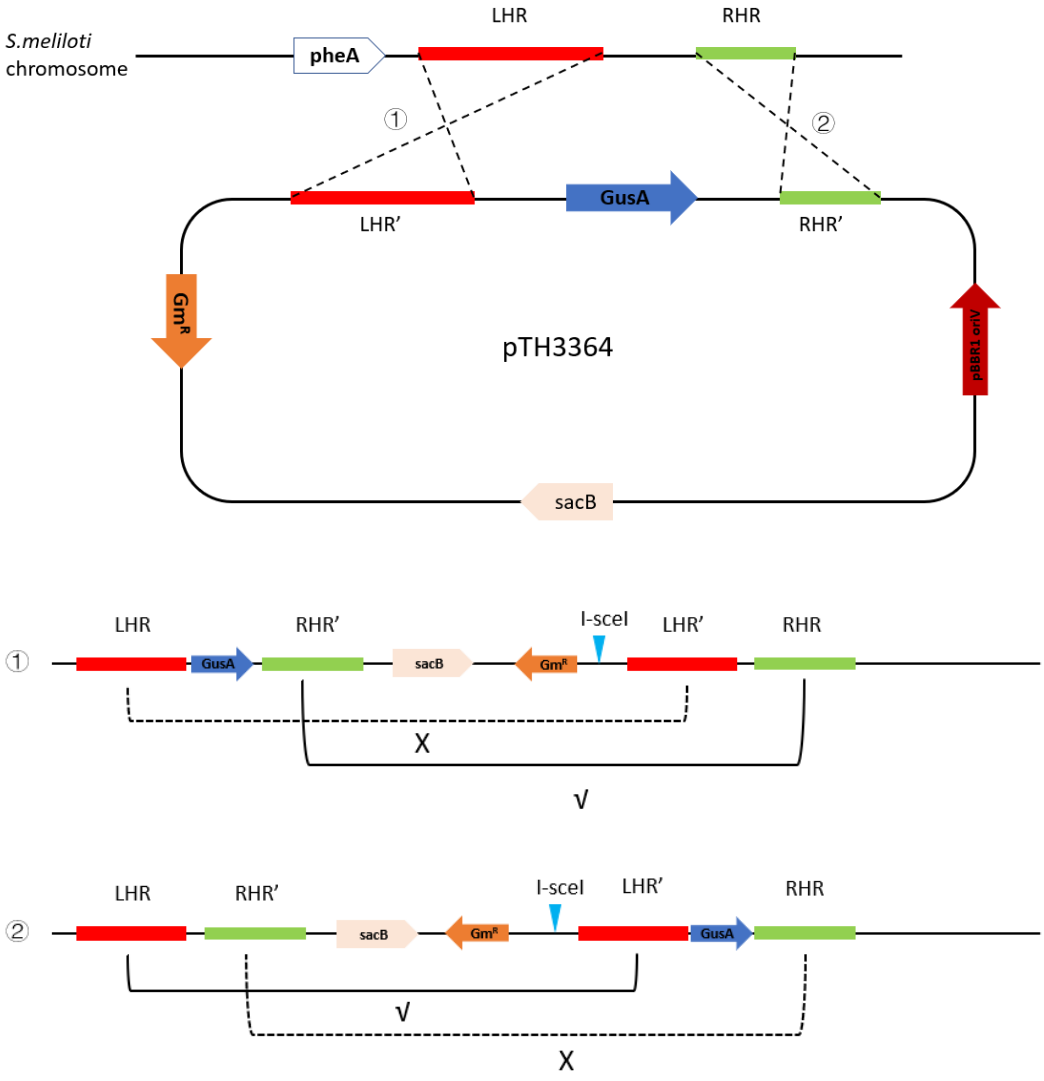


Figure 20. Schematic representation of double homologous recombination between pTH3364 and RmP110 *S. meliloti* chromosome. As shown in the diagram, the first recombination could happen between the two LHR or RHR homologous regions, indicated as ① or ②, respectively. Similarly, depending on the sites of recombination, the second recombination could also give rise to two possible recombinants in both scenarios ① and ②. Among the two possible recombination outcomes in each case, the desired recombination is shown in solid line and the undesired one is indicated in dotted line. The recombinants in the second round of

recombination were selected with the help of either the *sacB*/sucrose system or the I-SceI system.

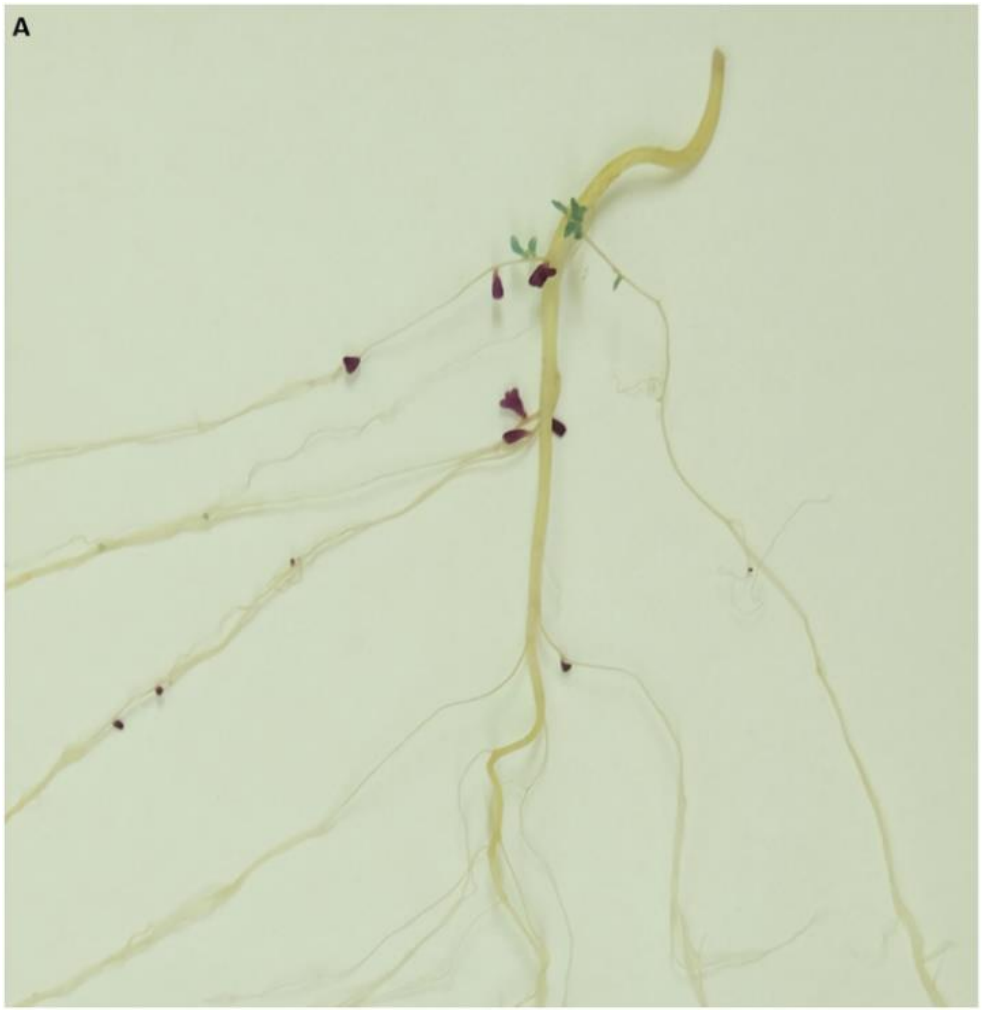
3.3.7 Nodule competition between RmP4291 and RmP110 using chromosome-integrated *gusA* and *celB*

Having successfully inserting *gusA* and *celB* into the *S. meliloti* chromosome, we wanted to perform a nodule competition assay between the minimal symbiotic pSymA strain RmP4291 and wildtype RmP110 using the marker-staining system that we had established. Since the markers were integrated into the genome of wildtype RmP110, the transfer of markers between *S. meliloti* strains could be carried out by transduction. RmP4291 was first transduced by the φM12 lysate of a *S. meliloti* phenylalanine auxotroph (RmG340, *pheA::*$Gm^R Sp^R$) (Oresnik et al., 1994). Transductants were selected on media containing Sp^R. The resulting transductants were again transduced by the lysate of RmP4505(*gusA*) or RmP4506(*celB*). The secondary transductants were selected on M9 minimal media without the supplement of L-phenylalanine, and also confirmed later by colony PCR. The resulting *gusA* marked RmP4291 was named RmP4508 and the *celB* marked RmP4291 was named RmP4509.

In order to minimize the effect of the difference of the two marker genes, both forward and reverse marking arrangements were tested (i.e., *gusA* marked RmP110 versus *celB* marked RmP4291 in one arrangement, and *celB* marked RmP110 versus *gusA* marked RmP4291 in another). Two test strains carrying different marker genes

were mixed at the ratios of 1:5, 5:1, and 1:1. A culture mixed at each ratio was inoculated to 20 alfalfas (4 plants/Leonard jar) with a total cell number of 10^4 per jar. Plants were harvested 28 days after inoculation.

As shown in Fig. 21, staining of the nodules was successful. Nodules formed by RmP4505 (Fig. 21D) and RmP4506 (Fig. 21E) were magenta and green, respectively, as expected. The nodules were stained from the inside to the surface. Furthermore, no colorless nodules were found on plants inoculated with strains carrying a marker gene.



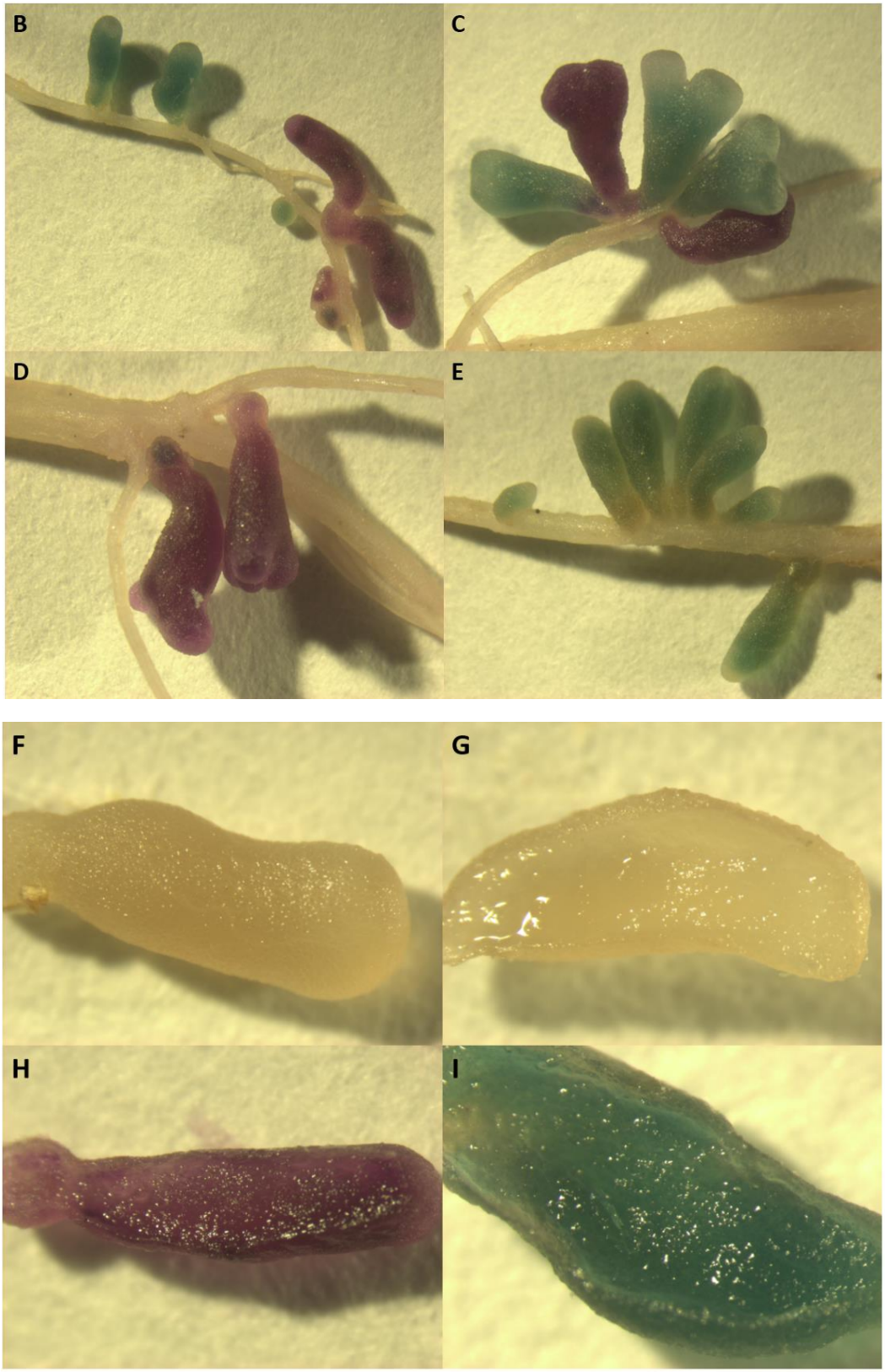


Figure 21. Stained alfalfa nodules formed by *S. meliloti* strains with *gusA* or *celB* inserted in the chromosome. Root nodules formed by RmP4505(*gusA* marked

RmP110), RmP4508(*gusA* marked RmP4291), RmP4506(*celB* marked RmP110) and RmP4509(*celB* marked RmP4291) are shown. (A) Nodules formed on one root inoculated with both RmP4505 (magenta nodules) and RmP4506 (green nodules). (B) Magenta and green nodules formed by RmP4505 (magenta nodules) and RmP4509 (green nodules). (C) Green and magenta nodules formed by RmP4506 (green nodules) and RmP4508 (magenta nodules). (D) The surface of stained nodules formed by RmP4505. (E) The surface of stained nodules formed by RmP4506. (F) The surface of a nodule formed by wild type RmP110 after going through the two-step staining procedure. (G) The cross section of the nodule shown in (F). (H) The cross section of a nodule shown in (D). (I) The cross section of a nodule shown in (E).

It has been reported that a single nodule may sometime be formed by more than one rhizobium strain if they are all captured when the root hair curling occurs. Indeed, we did observe nodules with mixed colors by using our staining assay (Fig. 22). Each of these nodules was apparently formed by the two *S. meliloti* strains used in the assay. Therefore, our staining method allows detection of nodules formed by multiple *Sinorhizobium* strains.

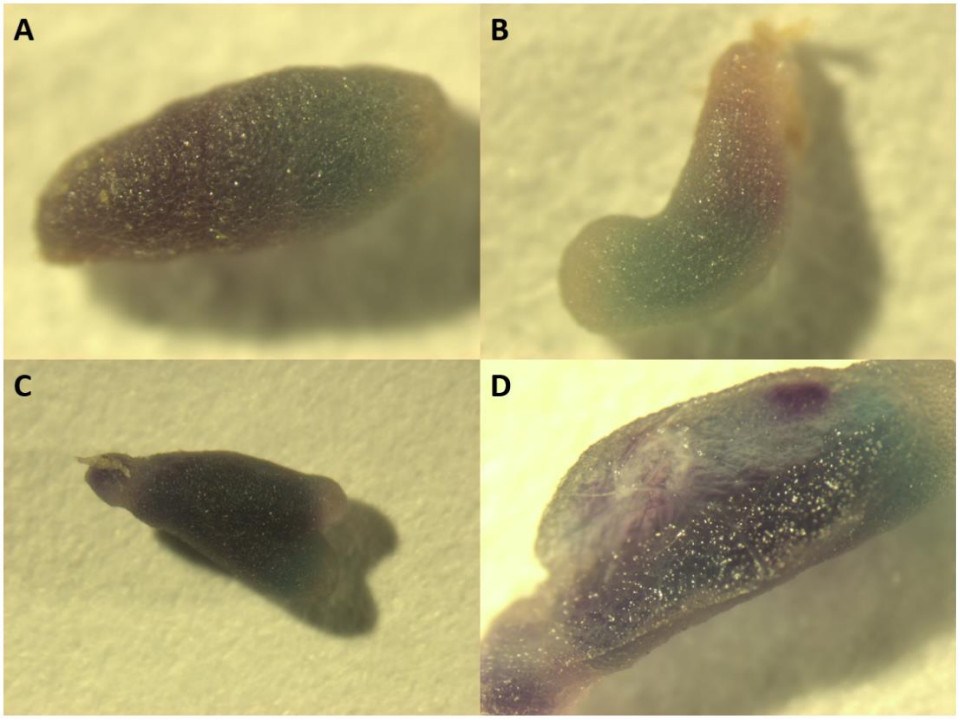


Figure 22. Individual nodules formed by two *S. meliloti* strains with different markers. (A~C) Single nodules formed by strains with both *gusA* (red) and *celB* (green) markers. (D) The cross section of the nodule in (C).

Having validated the effectiveness of the nodule staining method, we then investigated the competitiveness of the minimalized symbiotic pSymA strain RmP4291 compared to wildtype RmP110 for nodulation. The two strains marked with either gene were analyzed in pairs in the assay. Therefore, either RmP4508 (*gusA* marked RmP4291) and RmP4506 (*celB* marked RmP110), or RmP4509 (*celB* marked RmP4291) and RmP4505 (*gusA* marked RmP110) were used to co-inoculate alfalfa. As expected, both magenta nodules formed by RmP4505 or RmP4508 and green nodules formed by RmP4506 or RmP4509 were observed (Fig. 21). Intriguingly, when the plants were co-inoculated with RmP110 and RmP4291 at a 1:1 ratio, most of the nodules were formed by RmP110 and only about 15% of the nodules were formed by RmP4291 (Fig. 23). Even when RmP4291 accounted for ~83.3% of the total *S. meliloti* cells in the inoculum, the majority (~75%) of the nodules were occupied by RmP110. These results show that RmP4291 was far less competitive than RmP110 in nodulation. Therefore, we conclude that sequences other than the A117-A118-A121 region on pSymA contribute significantly to the nodulation ability of *S. meliloti*.

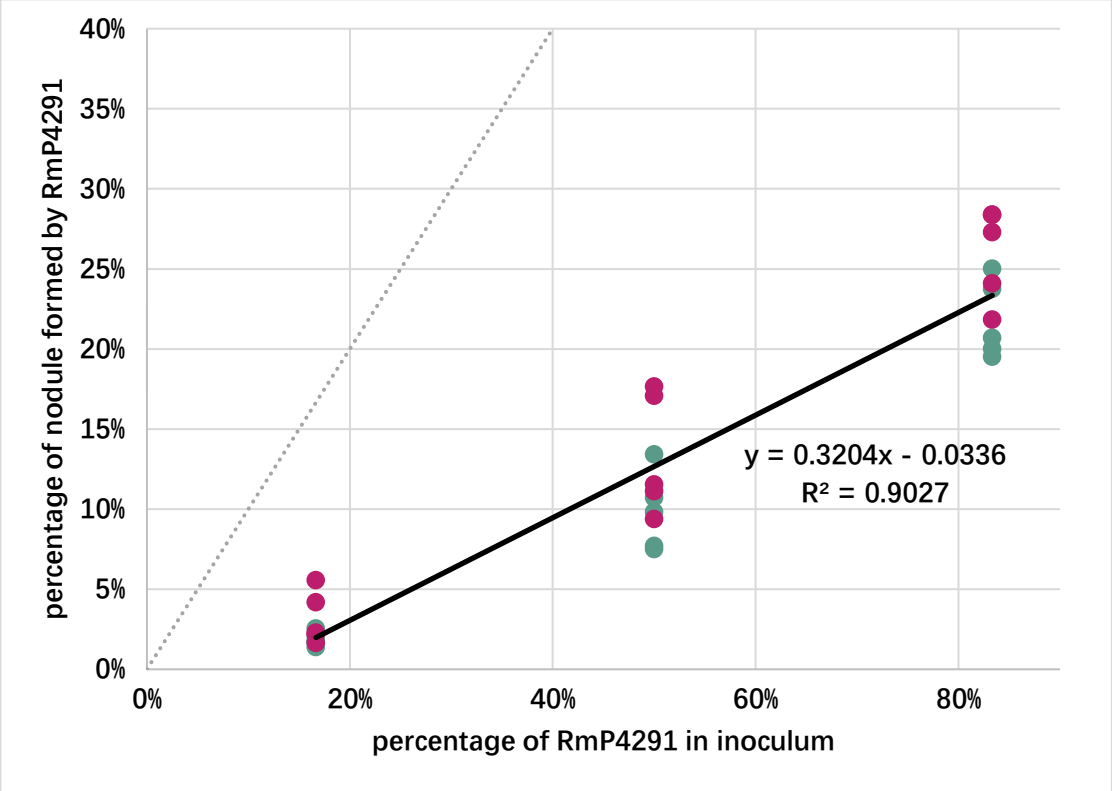


Figure 23. The relationship between the fraction of RmP4291 in the total inoculum and the fraction of the nodules formed by RmP4291 on alfalfa. Data obtained from nodules formed on alfalfas co-inoculated with RmP4508 (*gusA* marked RmP4291) and RmP4506 (*celB* marked RmP110) are indicated by magenta dots, while those from nodules formed on alfalfas co-inoculated with RmP4509 (*celB* marked RmP4291) and RmP4505 (*gusA* marked RmP110) are shown by green dots. The dotted line represents the relationship of Y = X. 30 plants and 1849 nodules were represented in the data.

Chapter 4 – Discussion

4.1 Symbiotic characteristics of the minimal symbiotic pSymA *S. meliloti* RmP4291

In this study, we have successfully constructed a *S. meliloti* derivative strain, denoted RmP4291, which lacks pSymA but contains the A-SNF region integrated into pSymB. Several symbiotic assays were performed on this strain to examine the ability of the minimized symbiotic pSymA to support SNF, compared to that of wildtype pSymA. The initial tests were carried out with *M. sativa*, since it has been used as the model plant host in our minimal symbiotic genome project. We showed that the integration of the A-SNF region into pSymB restored ~90% of the SNF capabilities of the wildtype strain on alfalfa to the *S. meliloti* mutant strain lacking pSymA through shoot dry weight assays and acetylene reduction assays. These results confirm that the nitrogen fixation ability of *S. meliloti* on *M. sativa* is preserved in the minimal symbiotic pSymA (Fig. 8). Subsequently, nodulation, another essential symbiotic characteristic of rhizobia, was studied by performing nodule kinetic assays on alfalfa inoculated with *S. meliloti* strains. Interestingly, while RmP4291 maintained a plant level of SNF similar to that of wildtype *S. meliloti*, nodules formed on plants inoculated with RmP4291 were ~172% on average, as many as those on plants inoculated with wildtype RmP110 at the same concentration as that of RmP4291 at all time points (Fig. 9). Therefore, inoculation of the plant with RmP4291 appears to have

a higher nodule forming efficiency compared to that with the wildtype strain. To further compare the ability of RmP4291 and RmP110 in nodulation, a nodule competition assay was performed on alfalfa. The data show that RmP4291 was significantly less competitive than wildtype *S. meliloti* in nodulation since most (85%) of the nodules were formed by wildtype *S. meliloti* in the experiments in which each plant was inoculated with same amounts of the two strains (Fig. 23). Lastly, to find out if deletion of a substantial portion of pSymA in the construction of the minimal symbiotic pSymA would result in a change in host range of *S. meliloti*, preliminary shoot dry weight assays and acetylene reduction assays were carried out with *M. truncatula* (Fig. 10), *M. alba* (Fig. 11) and *M. officinalis* (Fig. 12). Surprisingly, while RmP4291 was able to establish symbiosis with all the tested hosts, reduced SNF by the bacterium on *M. officinalis* was observed. Taken together, our data from the *M. sativa* experiments may suggest that, while the phenotype of the plant inoculated with *S. meliloti* was only minimally altered by replacing pSymA with the minimal symbiotic pSymA in the bacterium, the number of nodules on the plants inoculated with RmP4291 was significantly higher than that inoculated with RmP110, pointing to a possibility that SNF by RmP4291 was substantially lower than that by RmP110 at the single nodule level. Therefore, it seems that the extent of reduction in SNF by strain RmP4291 compared to that by RmP110 was greater than suggested by the shoot dry weight or acetylene reduction assays because the results of both of the assays were

calculated per plant (Figure 8B, 8C). In fact, an increase in the number of nodules may represent a mechanism in which the plant host compensates for the low SNF level within a single nodule, as the process of nodulation is known to be inhibited by the fixed nitrogen provide to plants (Harper and Gibson, 1984; Streeter and Wong, 1988). We found that, unlike the mostly deep pink/red nodules formed by wildtype *S. meliloti*, the nodules formed by RmP4291 were generally light pink/red, and occasionally even white (no quantitative data included). The color of a nodule usually reflects the nitrogen fixation efficiency within the nodule (Downie, 2005). The high oxygen-affinity leghaemoglobin, which is the source of red color, in the cell cytoplasm of a nodule is inhibited or not expressed if the bacteria inside of a nodule are not fixing nitrogen or have a low nitrogen fixing efficiency (Downie, 2005). There appear to be correlated development of the bacteria and the host in the symbiosis process. The mechanism of the correlation, however, is not well understood. Therefore, if the color of RmP110 and Rmp4291 nodules was indeed different, it might be indicative of the lower SNF activity in nodules formed by RmP4291 than those formed by RmP110. It should be pointed out, however, that the nodulation kinetics of RmP4291 and wildtype RmP110 were measured in a period of 18 days in this study. The maximum numbers of nodules that may form on plants inoculated with the two rhizobium strains remain to be determined by performing the assays for a longer time.

Westhoek et al., (2017) found that the number of nodules decreased as the percentage of nitrogen fixing bacteria increased when a plant host was co-inoculated with both nitrogen-fixing and non-fixing bacteria. In our competition experiments, we observed a similar trend when *Medicago sativa* was co-inoculated with RmP110 and RmP4291. With the increasing percentage of wildtype *S. meliloti* in the inoculum, the fewer nodules formed on the roots (data not shown). The trend was not as dramatic as reported by (Westhoek et al., 2017), presumably because unlike the non-fixing strain used in their experiment, RmP4291 was still a nitrogen fixer although it was significantly lower in competitiveness, as compared to RmP110, in nodulation. By comparison, the two competitor strains were similarly competitive in nodulation in their experiment. RmP4291 was capable of establishing robust symbiosis with alfalfa when it was the only symbiont in the rhizosphere but, when RmP4291 and RmP110 were used as co-inoculants, nodules formed by RmP4291 were rarely found (Figure 23, ~15% RmP4291 at a 1:1 ratio). The reduction of the competitiveness of RmP4291 was unlikely due to the loss of genes crucially involved in symbiotic nitrogen fixation, such as *nif*, *nod*, *fix*. It is known that loss of some of the SNF essential genes, *nod* genes in particular, could result in a delay in nodule formation (*nodL*) or a decrease in root hair deformation (*nodN*, *nolG* and *nolF*), which might directly or indirectly cause the low nodule number phenotype (Baev et al., 1991). But no significant defects in nodulation were seen when RmP4291 was the solo inoculant. Thus, some other genes

specifically involved in nodule competitiveness may exist on pSymA outside the minimal symbiotic regions.

Intriguingly, in our host range experiments, *M. officinalis* inoculated with RmP4291 exhibited a reduced symbiotic phenotype by ~40% compared to that of RmP110. No similar observations were made in previous research. However, the experiment needs to be repeated to verify this observation. Further investigation would be needed to shed light on this finding if the result is validated.

Despite the reduction in SNF efficiency of RmP4291, the nodulation and nitrogen fixation capabilities of the wildtype strain were largely preserved in RmP4291, indicating that all essential components for the SNF process on pSymA in *S. meliloti* should have been packaged in the minimal pSymA. Therefore, we have successfully achieved the preliminary goal of defining the minimal pSymA region for symbiotic nitrogen fixation. The availability of RmP4291 as well as the A-SNF plasmid pTH3255 offers a new start point for the identification of genes required for SNF and the construction of the minimal genome that only contains the minimal set of genes crucially required for successful SNF by *S. meliloti*.

4.2 Analysis of genes within the A117-A118-A121 region

All the known *nif* as well as *nod, noe* and *nol* genes on pSymA are packaged in the A117-A118-A121 region, and only a *fix* gene cluster *fixNOQP*$_3$*fixI*$_2$*fixS*$_2$ is located

outside of the region. In fact, the symbiotic genes on pSymA were previously shown to be clustered in a 275-kb region (Barnett et al., 2001), which overlaps with the A117-A118-A121 region. It should be noted, however, that a number of genes with functions unrelated to SNF, as well as many hypothetic genes, reside on pTH3255. Therefore, although 88% of the pSymA sequence has been deleted during the construction of the minimal symbiotic pSymA, the genome could be further reduced. The ultimate minimal symbiotic pSymA is expected to be 50~60 kb. There are many difficulties in further minimizing the pSymA genome, among which gene redundancy poses one of the major challenges.

The A117-A118-A121 region in plasmid pTH3255 which is 170,816 bp in size (Figure 6). All genes within the A117-A118-A121 region are listed in Table 4 with descriptions, including SNF relatedness and references. A total of 155 genes reside in this region, including 8 *nif* genes, 21 *nod*, *noe* and *nol* genes, and 25 *fix* genes.

Table 4. All genes within the A117-A118-A121 region.

Number	Gene	Description	Associate with SNF	Need for SNF	Reference
1	*GroES2*	Chaperone products help to refold denatured proteins and facilitate folding of newly synthesized proteins.	X	X	(Krol and Becker, 2004a)
2	*Sma0747*	Hypothetical protein.	-	-	
3	*Sma0748*	Transcriptional regulator (not classified).	-	-	
4	*Sma5003*	Hypothetical protein.	-	-	

5	*Sma0750*	Transcriptional regulator (not classified).	-	-	
6	*Sma0751-Sma0753*	Involved in utilization of stachydrine.	X	X	(diCenzo et al., 2015)
7	*Sma0754*	Hypothetical protein.	-	-	
8	*nodD2*	Nod box-dependent transcription activator. Activates *nod* genes in the presence of plant betaines, such as trigonelline and stachydrine.	√	√	(Honma and Ausubel, 1987a) (Honma et al., 1990)
9	*Sma0758*	Putative oxidoreductase.	X	X	(Cosseau et al., 2002)
10	*Sma0759*	Hypothetical protein.	-	-	
11	*fixT2*	Antikinase protein, which is a repressor of *fixK, nifA, fixL* expression.	√	X	(Foussard et al., 1997)
12	*fixK2*	Transcriptional regulator (activating expression of the *fix* operons).	√	√	(Batut et al., 1989)
13	*Sma0763*	Hypothetical protein	-	-	
14	*fixN2*	Subunit of a high-affinity *cbb*$_3$-type cytochrome c oxidase, which supports bacterial respiration in bacteroids.	√	√	(Preisig et al., 1993)
15	*fixO2*	Subunit of a high-affinity *cbb*$_3$-type cytochrome c oxidase	√	√	(Preisig et al., 1993)
16	*fixQ2*	Subunit of a high-affinity *cbb*$_3$-type cytochrome c oxidase	√	√	(Preisig et al., 1993)
17	*fixP2*	Subunit of a high-affinity *cbb*$_3$-type cytochrome c oxidase	√	√	(Preisig et al., 1993)
18	*Sma0771*	Hypothetical protein.	-	-	
19	*nodL*	Nod factor acetyltransferase; adds an acetyl group at the non-reducing terminal sugar residue on the Nod facto	√	√	(Truchet et al., 1991) (Ardourel et al., 1994,

					1995) (Geurts and Bisseling, 2002)
20	*noeA*	Host specific nodulation protein. *noeA* mutant delayed nodulation on *M. lupulina*.	√	√	(Ardourel et al., 1995)
21	*noeB*	Host specific nodulation protein. *noeB* delayed nodulation on *M. lupulina*.	√	√	(Ardourel et al., 1995)
22	*Sma0775*	Hypothetical protein.	-	-	
Insertion sequences *Sma4004-Sma4005*					
23	*Sma0789*	Transcriptional regulator (Not classified).	-	-	(Arnold et al., 2017)
24	*Sma0791-Sma0805*	Trigonelline degradation genes. Sma0794, encodes putative luciferase-like monooxygenase and mutants cause alteration of capsular polysaccharides, which will then affect the nodulation competitiveness.	X	X	(Onishchuk et al., 2005; Perchat et al., 2018)
25	*syrB3*	Repressor of *syrM*.	√	?	(Barnett and Long, 1997)
26	*Sma0809*	Hypothetical protein.	-	-	
27	*fixU*	Unknown.	?	?	(Barnett et al., 2001)
28	*fdxN*	Encoding a ferredoxin-like protein.	√	√	(Klipp et al., 1989a)
29	*nifB*	Nitrogenase cofactor biosynthesis protein (nitrogenase ironmolybdenum cofactor).	√	√	(Buikema et al., 1987)
30	*nifA*	Transcriptional activator (activation of *nif* operons).	√	√	(Szeto et al., 1984)
31	*fixX*	Ferredoxin-like protein.	√	√	(Dusha et al., 1987a; Earl et al., 1987)

32	*fixC*	Involved in electron transport for nitrogenase s	√	√	(Earl et al., 1987)
33	*fixB*	Electron transfer flavoprotein alpha chain. Involved in electron transport for nitrogenase	√	√	(Earl et al., 1987)
34	*fixA*	Electron transfer flavoprotein beta chain. Involved in electron transport for nitrogenase	√	√	(Earl et al., 1987)
35	*Sma0824*	Hypothetical protein.	-	-	
36	*nifH*	Encodes nitrogenase reductase (Fe protein).	√	√	(Török and Kondorosi, 1981)
37	*nifD*	Encodes for the alpha subunit of nitrogenase.	√	√	(Hirsch et al., 1983)
38	*nifK*	Encodes for the beta subunit of nitrogenase.	√	√	(Hirsch et al., 1983)
39	*nifE*	Nitrogenase molybdenum-cofactor biosynthesis protein.	√	√	(Barnett et al., 2001)
40	*nifX*	Iron-molybdenum cluster-binding protein.	√	?	(Moreno-Vivian et al., 1989a, 1989b)
41	*Sma0833*	Hypothetical protein.	-	-	
42	*fdxB*	Ferredoxin-like protein.	√	?	(Barnett et al., 2001)
43	*Sma5036*	Hypothetical protein.	-	-	
44	*Sma0835*	Hypothetical protein.	-	-	
45	*syrA*	Regulator of Exopolysaccharide.	√	X	(Barnett et al., 1998)
46	*nodD3*	Nod box-dependent transcription activator.	√	√	(Honma and Ausubel, 1987a) (Honma et al.,

					1990) (Swanson et al., 1993)
Insertion sequences *Sma0841-Sma0846*					
47	*Sma0848*	Hypothetical protein.	-	-	
48	*SyrM*	Transcriptional regulator activates expression of the nodD3 gene.	√	X	(Swanson et al., 1993)
49	*Sma0850*	Hypothetical protein.	-	-	
50	*nodH*	Sulfotransferase adds a sulfate group at the reducing terminal sugar residue on the Nod factor.	√	√	(Ehrhardt et al., 1995), (Lerouge et al., 1990)
51	*nodF*	Encodes for an acyl carrier protein;involved in the formation of an acyl chain of 16 C-atoms in length with two unsaturated bonds.	√	√	(Demont et al., 1993) (Geurts and Bisseling, 2002)
52	*nodE*	Beta-ketoacyl synthase, together with *nodF*, involved in the formation of an acyl chain of 16 C-atoms in length with two unsaturated bonds.	√	√	(Demont et al., 1993) (Geurts and Bisseling, 2002)
53	*nodG*	3-oxoacyl- (acyl carrier protein) reductase.	√	?	(López-Lara and Geiger, 2001; Ogawa et al., 1991; Slabas et al., 1992)
54	*nodP1*	Sulphate adenylyltransferase, involved in the formation of a sulfate group at the reducing terminal sugar residue on the Nod factor.	√	√	(Keating et al., 2002; Schwedock et al., 1994), (Lerouge et

					al., 1990)
55	*nodQ1*	Adenylylsulphate kinase, together with *nodP*, involved in the formation of a sulfate group at the reducing terminal sugar residue on the Nod factor.	√	√	(Keating et al., 2002; Schwedock et al., 1994), (Lerouge et al., 1990)
Insertion sequences *Sma4007, Sma0861*					
56	*nodJ*	ABC transporter gene, involved in exportation of nod factors.	√	√	(Barran et al., 2002)
57	*nodI*	ABC transporter gene, involved in exportation of nod factors	√	√	(Barran et al., 2002; Carlson et al., 1994),
58	*nodC*	N-acetylglucosaminyltransferase. It is a common nodulation gene, involved in the synthesis of lipochitooligosaccharide.	√	√	(Geremia et al., 1994)
59	*nodB*	Chitooligosaccharide deacetylase. It is a common nodulation gene, involved in the synthesis of lipochitooligosaccharide.	√	√	(John et al., 1993)
60	*nodA*	N-acylation of the aminosugar backbone - a common nod gene, involved in synthesis of lipochitooligosaccharide.	√	√	(Debellé et al., 1996)
61	*nodD1*	nod box-dependent transcription activator.	√	√	(Honma and Ausubel, 1987a) (Ardourel et al., 1995)
62	*Sma0872*	Hypothetical protein.	-	-	
63	*nifN*	biosynthesis of the iron molybdenum cofactor	√	√	(Aguilar et al., 1987)
64	*nodN*	Dehydratase. Involved in the production of the root hair deformation factor.	√	√	(Baev et al., 1991)

65	*nolG*	Efflux transporter. Potentially involved in the transportation of root hair deformation factor.	√	√	(Baev et al., 1991)
66	*nolF*	Secretion protein. Potentially involved in the secretion of root hair deformation factor.	√	√	(Baev et al., 1991)
67	*nodM*	D-glucosamine synthetase. Involved in the production of the root hair deformation factor.	√	√	(Baev et al., 1991)
68	*Sma0882-Sma0890*	Hypothetical protein.	-	-	
69	*dnaE3*	Encodes for subunit α of DNA polymerase III.	X	X	(Zhao et al., 2006)
70	*Sma0894-Sma0907*	Hypothetical protein.	-	-	
71	*Sma1146-Sma1154*	Hypothetical protein.	-	-	
72	*Sma1155*	Encodes for a hypothetical cation transport P-type ATPase.	X	X	(Meilhoc et al., 2010)
73	*Sma1156*	Alcohol dehydrogenase.	X	X	(Puskás et al., 2004)
74	*Sma1158-Sma1162*	Hypothetical protein.	-	-	
75	*Sma1163*	P1B-5-ATPase.	X	X	(L. Zielazinski et al., 2013)
76	*Sma1166*	Putative hydrolase.	X	X	(L. Zielazinski et al., 2013)
77	*Sma1168, Sma1169*	Hypothetical protein.	-	-	
78	*Sma1170*	Conserved hypothetical protein.	X	X	(Meilhoc et al., 2010)
79	*Sma1171-Sma11*	Hypothetical protein.	-	-	

	78				
80	*nosR*	Nitrous oxide reductase expression regulator.	X	X	(Holloway et al., 1996)
81	*nosZ*	Nitrous oxide reductase.	X	X	(Holloway et al., 1996)
82	*nosD*	Nitrous oxidase accessory protein.	X	X	(Holloway et al., 1996)
83	*nosF*	ATPase (copper-processing gene).	X	X	(Holloway et al., 1996)
84	*nosY*	permease（copper-processing gene）	X	X	(Holloway et al., 1996)
85	*nosL*	Putative outer-membrane lipoprotein.	X	X	(Chan et al., 1997)
86	*nosX*	Putative accessory protein.	X	X	(Chan et al., 1997)
87	*hmp*	Encoding a flavohaemoglobin able to scavenge Nitric oxide.	X	X	(del Giudice et al., 2011)
88	*Sma1192*	Hypothetical protein.	-	-	
89	*nnrS2*	Resistance gene to nitrosative stress, encodes for a copper containing membrane protein.	X	X	(Blanquet et al., 2015)
90	*Sma1195, Sma1197*	Hypothetical protein.	-	-	
91	*Sma1198*	Copper export protein.	X	X	(Barnett et al., 2001)
92	*Sma1200, Sma1201*	Hypothetical protein.	-	-	
93	*Sma1207*	FixK-like regulator.	X	X	(Bobik et al., 2006)
94	*fixS1*	Transmembrane protein. Involved in the formation of a specific cation pump in	√	√	(Kahn et al., 1989)

		SNF.			
95	*fixI1*	ATPase. Involved in the formation of a specific cation pump in SNF.	√	√	(Kahn et al., 1989)
96	*fixH*	Transmembrane proteins. Involved in the formation of a specific cation pump in SNF.	√	√	(Kahn et al., 1989)
97	*fixG*	Iron sulfur membrane protein. Involved in the formation of a specific cation pump in SNF.	√	√	(Kahn et al., 1989)
98	*fixP1*	Subunit of a high-affinity *cbb*$_3$-type cytochrome c oxidase, which supports bacterial respiration in endosymbiosis.	√	√	(Preisig et al., 1993)
99	*fixQ1*	Subunit of a high-affinity *cbb*$_3$-type cytochrome c oxidase, which supports bacterial respiration in endosymbiosis.	√	√	(Preisig et al., 1993)
100	*fixO1*	Subunit of a high-affinity *cbb*$_3$-type cytochrome c oxidase, which supports bacterial respiration in endosymbiosis.	√	√	(Preisig et al., 1993)
101	*fixN1*	Subunit of a high-affinity *cbb*$_3$-type cytochrome c oxidase, which supports bacterial respiration in endosymbiosis.	√	√	(Preisig et al., 1993)
102	*fixM*	Flavoprotein oxidoreductase.	√	X	(Cosseau et al., 2002)
103	*fixK1*	Transcriptional regulator (activating expression of *fixNOQP* operon).	√	√	(Batut et al., 1989)
104	*fixT1*	Antikinase protein, which is a repressor of *fixK, nifA, fixL* expression.	√	X	(Foussard et al., 1997)
105	*fixJ*	Response regulator (transcriptional activator of *nif* and *fix* genes).	√	√	(David et al., 1988)
106	*fixL*	Oxygen-regulated histidine kinase which activates *fixJ*.	√	√	(David et al., 1988)
107	*Sma1231*	Hypothetical protein.	-	-	

108	*napC*	Membrane protein, electron donor to periplasmic nitrate reductase.	X	X	(Perrine et al., 2005)
109	*napB*	Periplasmic nitrate reductase.	X	X	(Perrine et al., 2005)

There are five symbiotic sub-regions in the A117-A118-A121 region (Figure 6). Those sub-regions were symbiotic gene clusters that were mostly similar to the previously described gene locus: *fix-i*, *fix-ii*, *fix-iii*, *nod-i*, *nod-ii*, and *nif* (diCenzo et al., 2016). The reason that symbiotic regions are not named in gene categories here is that each functional group of genes (*nif*, *nod*, *fix*) is not perfectly and exclusively clustered together in some sub-regions. The sub-regions are expected to be the target of the construction of the next level of minimal symbiotic pSymA. Symbiotic genes with known or similar putative functions, such as *fixX*, *fdxN* and *fdxB*, and duplicated symbiotic genes, such as *fixTKNOQP*, *fixIS* and *nodD*, were identified in A117-A118-A121.

4.3 The redundancy of symbiotic genes on the minimal symbiotic pSymA

As mentioned above, the minimal symbiotic pSymA contains most of the genes that are known to be essential or related to symbiosis. Among them are three copies of *nodD*, two copies of *fixIS*, and two copies of *fixNOQP*. Other than these genes, *fixX*, *fdxN*, and *fdxB* outside of the A117-A118-A121 region of pSymA, and two ORFs (open reading frames) on the chromosome (*fdx*, *Y03875*) were found to encode

ferredoxin-like proteins (Klipp et al., 2006). A second copy of *nodPQ* is on pSymB (Galibert et al., 2001). The functions of most of these genes have been studied previously, but it remains to be understood why or whether *S. meliloti* needs these genes in multiple copies. To construct a minimal symbiotic genome for *S. meliloti*, is it necessary to keep these genes in multiple copies? Here, we discuss *nodD*, *fixNOQP* and three ferredoxin genes *fixX*, *fdxN* and *fdxB.*

The *nodD* gene encodes a LysR-type transcriptional regulator, which is responsible for the activation of *nod*, *nol* and *noe* genes in the presence of signals, usually a flavonoid, produced by the plant host (Long, 1996). *S. meliloti* possesses three copies of *nod*, which are all located on pSymA. They share over 77% identity at the amino acid sequence level (Peck et al., 2006). It is believed that the three copies of *nodD* regulate nodulation process under different circumstances. Honma and Ausubel, 1987a constructed all possible combinations among the three copies of *nodD* genes (by using mutation combination of the nodD genes) and determined the symbiotic phenotype of those mutants on alfalfa and sweet clover. They found that $nodD_1$ and $nodD_3$ had equivalent roles in nodulating sweet clover, and $nodD_1$ played a more important role than $nodD_3$ in nodulating alfalfa. The $nodD_2$ gene didn't show a strong influence on the nodulation on either host. It is also well known that $nodD_1$ is induced in the presence of luteolin and methoxychalcone (Hartwig et al., 1990; Mulligan and Long, 1985; Peck et al., 2006), and $nodD_2$ is induced if plant betaines, such as

trigonelline and stachydrine, are provided (Phillips et al., 1992). Unlike *nodD*$_1$ and *nodD*$_2$, *nodD*$_3$ is an inducer-independent gene activated by the LysR family transcriptional regulator SyrM (Swanson et al., 1993). SyrM also activates the expression of *syrA*, which codes for a regulator for exopolysaccharide production (Barnett and Long, 2015). Exopolysaccharides are essential for bacterial invasion of plant roots, but the mechanisms as how they function are unknown (Cheng and Walker, 1998; Downie, 2010). A SyrM-NodD$_3$-SyrA regulatory circuit model was proposed by Barnett and Long, 2015. All in all, the above studies may suggest that *nodD*$_1$ and *nodD*$_2$ are required for the host specificity of *S. meliloti*. As for the role of *nodD*$_3$, it is speculated that *nodD*$_3$ serves as a second *nod* gene activator (Smith and Long, 1998), and is additionally involved in the regulation of the synthesis of substances of importance to SNF. Therefore, in the construction of minimal symbiotic pSymA, *nodD*$_3$ should always be included, whereas one or both of the other two *nodD* genes may be included in the minimal genome depending on host plant for which the minimal genome is built.

Nitrogenase is known for being inactivated by oxygen. The leghaemoglobin on the nodule surface makes nodule an ideal environment (7~11 nM oxygen concentration inside nodules) for nitrogenase, since leghaemoglobin has an extremely fast O_2 association rate and a relatively slow O_2 dissociation rate (Appleby, 1984). But on the other hand, nitrogenase requires a high level of ATP and reducing power to

carry out the reaction. To overcome this conflict, rhizobia utilize a high affinity cytochrome cbb3-type oxidase, encoded by *fixNOQP*, to achieve a high rate of oxygen respiration. There are three copies of *fixNOQP* in *S. meliloti*, all of which are located on pSymA. Of the three *fixNOQP* copies, $fixNOQP_1$ is positioned at nt 667360-669146 on pSymA inside of a *fix* cluster containing *fixLJ*, $fixT_1$, $fixK_1$, *fixM*, *fixGHIS*, etc. The $fixNOQP_2$ cluster is at nt 414187- 417580 on pSymA downstream of $nodD_2$, $fixT_2$ and $fixK_2$, and $fixNOQP_3$, which were deleted in the construction of the A117-A118-A121 minimized symbiotic pSymA, are located at nt 326972-330449 on pSymA with no symbiotic gene, except for $fixIS_2$, in close vicinity. Besides, $fixNOQP_1$ and $fixNOQP_2$ share 95% similarity and $fixNOQP_3$ was ~61% similar to the other two copies of *fixNOQP* at the amino acid sequence level. The *fixNOQP* genes in rhizobia are homologous to the *ccoNOQP* genes, which also encodes a cytochrome c oxidase complex, of other Gram-negative proteobacteria, e.g., *Cytophaga*, *Flexibacter*, etc., that do not fix nitrogen (Cosseau and Batut, 2004). Thus, from an evolutionary point of view, Kopat et al., 2017 hypothesized that the ancestral *fixNOQP* of proteobacteria emerged initially to adapt to an environment characterized with a low concentration of oxygen in the atmosphere, and then disseminated to the modern members of α-, β-, and γ-proteobacteria. The first duplication and divergence of the *fixNOQP* operon (forming of a *fix* group) was thought to occur in the ancestor of Rhizobiales to allow the bacteria to better adapt to

symbiosis with plants. The operon was duplicated again with further divergence in some of the most specific symbionts including *Sinorhizobium*. It was believed that the three copies of *fixNOQP* underwent differential regulation in *S. meliloti* (Bobik et al., 2006). The *fixNOQP*$_1$ operon was regulated by FixJ under both microoxic free-living and symbiotic conditions, while the products of *fixNOQP*$_2$ were only detected in bacteriods. Interestingly, *fixNOQP*$_3$ was partially regulated by *phoB* and was not induced under either microoxic or symbiotic conditions (Krol and Becker, 2004b). However, a *S. meliloti* mutant (ΔA116) missing *fixNOQP*$_3$ did not show a reduction in SNF compared to the wildtype strain (diCenzo et al., 2016). Those observations support the hypothesis of Kopat et al., 2017 to a certain extent, as one might suggest that the *fixNOQP*$_3$ was the original copy of *fixNOQP* that encourages the bacteria to adapt to low oxygen environment, while *fixNOQP*$_1$ was the first duplicate to allow the symbiosis to happen and *fixNOQP*$_2$ could be the second duplicate which helps *Sinorhizobium* to further differentiate into a specialized nitrogen fixer. Renalier et al., 1987 found that maintaining either *fixNOQP*$_1$ or *fixNOQP*$_2$ was essential and sufficient for SNF in *S. meliloti*. Reyes-González et al., 2016 showed that both *fixN*$_1$ and *fixN*$_2$ were important for optimal nitrogen fixation since lacking either of them would result in a reduction in SNF by two-fold. Similar results were obtained in a recent preliminary attempt to further reduce the minimal symbiotic pSymA, in which the 4 sub-regions, sub-region 2, 3, 4 and 5 of pTH3255 (Figure 6) were assembled in

our lab (Kearsley, Geddes and Finan, unpublished). The resulting strain only contains *fixNOQP*$_1$ without *fixNOQP*$_2$ and *fixNOQP*$_3$.The symbiotic phenotype of this strain had ~60% of the shoot dry weight of wildtype RmP110 28 days after inoculation, while no large reduction was not seen in RmP4291. The wildtype performance can be largely restored by introducing region 1, which contains *fixNOQP*$_1$, back into the genome (Geddes and Finan, unpublished). *fixNOQP*$_1$ was thought to play a major role in this restoration. In addition, a mutation at *fixN*$_1$ *S. meliloti* was found to be associated with a delayed symbiotic phenotype (Torres et al., 2013). Interestingly, these authors found that an initial reduction in symbiosis performance was overcome following a longer plant growth period. To sum up, while a single copy of either *fixNOQP*$_1$ or *fixNOQP*$_2$ is capable of supporting SNF activities to certain extents, both *fixNOQP*$_1$ and *fixNOQP*$_2$ are required for a wildtype level of SNF by *S. meliloti*. The *fixNOQP*$_1$ gene alone would result in the lack of oxygen provided to the bacteria in nodules. On the other hand, *fixNOQP*$_2$ alone is also inadequate for optimal SNF probably because the operon is only induced in the bacteroid stage. The consequence of the late expression of *fixNOQP*$_2$ would be the lack of ATP for the bacteria in newly formed nodules and thus the delay in plant performance at the starting stage of symbiosis. The defect might eventually be overcome by the increasing number of bacteriods in mature nodules since the expression of *fixNOQP*$_2$ in bacteriods is relatively high (Torres et al., 2013). *fixNOQP*$_2$ may thus be the most important

contributor to SNF among the three homologous operons. The *fixNOQP*$_3$ operon is regulated differently (Bobik et al., 2006). The expression of *fixNOQP*$_3$ increases the cell viability in an anaerobic environment. But the deletion of *fixNOQP*$_3$ does not significantly affect SNF, although it might contribute to the slight reduction of SNF level in RmP4291. It would be of interest to determine if an *S. meliloti* mutant containing *fixNOQP*$_3$ and lacking both *fixNOQP*$_1$ and *fixNOQP*$_2$ shows detectable SNF activity. To better understand the maintenance of multiple copies of *fixNOQP* in *S. meliloti*, it might be helpful to investigate the dynamics of overall expression of *fixNOQP* since the three homologous operons are expressed at different levels under different circumstances. The coordinated expression of the three operons appears to support optimally cell respiration at all stages in symbiosis. Based on the above analyses, a single copy of *fixNOQP*$_2$ should be kept in a minimized SNF genome in *S. meliloti* to minimize the loss in SNF efficiency.

An electron transport chain is a system consisting of a series of electron carriers in the order of increasing electron affinity. Electrons from electron donors are delivered through the chain and finally to oxygen to produce water and ATP. For nitrogen fixing bacteria, the electron transport chain provides reducing power and ATP to nitrogenase and is, therefore, one of the determinants of nitrogen fixation efficiency. Ferredoxin proteins act as the electron donor in nitrogen fixation in *S. meliloti*. On pSymA, three genes were found to encode ferredoxin protein, i.e., *fixX*,

fdxN and *fdxB.* A ISRm2 insertion mutant of *fixX* was previously found to have a *fix*$^-$ phenotype (Dusha et al., 1987b). In the meantime, *fdxN* was also shown to be essential for SNF (Klipp et al., 1989b). *fdxN* is co-transcribed with *nifA*, while *fixX* is part of the *fixABCX* operon, which contains a promoter homologous to other *nifA*-activated promoters. *fdxB* is located downstream of the *nifHDK* on pSymA (Barnett et al., 2001). However, whether it is involved in or essential for SNF is still unknown. Why different ferredoxin genes are included and SNF-essential in *S. meliloti* has yet to be understood.

4.4 Development of a new nodule competition assay for *S. meliloti*

A new nodule competition assay based on histochemical staining was developed for *S. meliloti* in this study. This assay employed the two marker genes, *gusA* and *celB* as previously described for *R. tropici* and for *R. leguminosarum* (Sessitsch et al., 1996, Sánchez-Cañizares and Palacios 2013). To identify bacteria inside of nodules by staining, two marker genes, *gusA* and *celB*, were integrated into the *S. meliloti* chromosome. Taking advantage of the thermostability of β-galactosidase, encoded by *celB* and the absence of background β-glucuronidase activity in both plant tissues and *S. meliloti*, *gusA* and *celB* marked strains can be easily detected by staining (magenta for *gusA* and green for *celB*).

The results showed that marking systems, in which *gusA* or *celB* were on a plasmid or integrated in the *S. meliloti* genome allowed nodules to be stained. The nodules formed by either *gusA*-strain or *celB*-strain gave a recognizable color. Transcription of the *gusA* and *celB* genes was driven by a *nifH* promoter from *Rhizobium leguminosarum*, and the results indicate that the *nifH* promoter functioned well in *S. meliloti* induced nodules Figure 17 & 21. It should be noted that this assay permits more accurate and reliable analysis of nodules than conventional competition assays based on antibiotic resistance screening, which are often labor- and time-consuming.

We encountered several problems in the process of setting up this assay. When plants were inoculated with strains containing a marker gene on plasmid, some of the nodules failed to stain presumably because of loss of the marker-bearing plasmid. Because of the instability of the strain with respect to the maintenance of the marker plasmid in nodules, this plasmid-based marking system appeared to be unreliable. In comparison, (Geddes et al., 2019) also constructed two marker plasmids (pOPSO253 and pOPSO254, Appendix, Figure A5) for the detection of nodules formed by *Rhizobium leguminosarum* on pea. Interestingly, they stated that no loss of the plasmid was found in their experiments. The observed plasmid stability might be attributed to the presence of the *par* genes in their plasmids. The *par* locus, originally taken from pJP2, was shown to be able to maintain plasmid in the absence of

antibiotic selection (Prell et al., 2002). Therefore, if a plasmid-based marking system is desired, cloning of the *par* genes into the plasmid may improve the stability of the assay. When the marker gene is integrated into the genome, no unstained nodules were found, but a slight reduction in the proportion of *celB*-marked strains was observed not only in the competition between RmP4291 and RmP110 (Fig. 24) but also in control experiments, e.g., competition between two RmP110 strains with different markers and that between marked RmP110 and wildtype RmP110 (data not shown). A similar observation was also made previously (Sánchez-Cañizares and Palacios, 2013), who prepared a similar marking system in *Rhizobium leguminosarum*. There is no evidence suggesting that *gusA* or *celB* is related to the process of symbiotic nitrogen fixation either directly or indirectly. There are human factors that might contribute to undercounting of *celB*-marked nodules. For instance, the magenta color of a nodule stained by magenta-glcA was more readily recognized than the light green color of nodule treated with X-gal, especially when the nodule was newly formed and small. To improve the assay, one may try to employ better substrates for *celB*-based detection and/or extend the staining time for the detection.

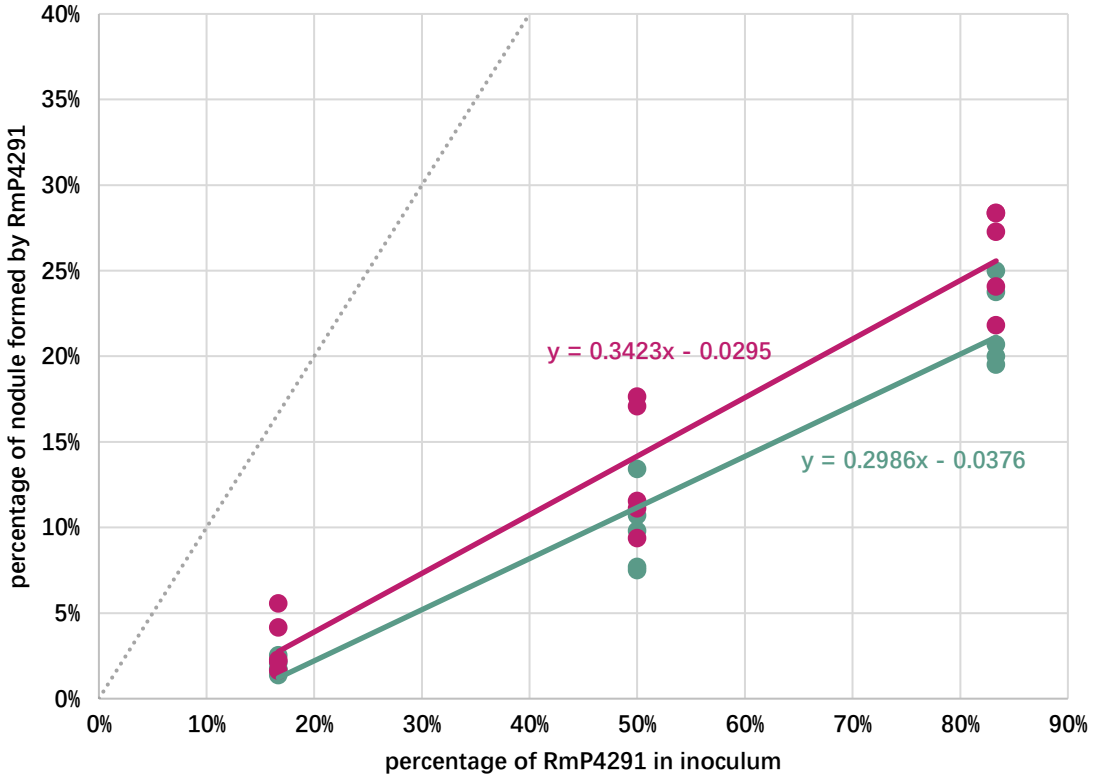


Figure 24. Relationship between the percentage of RmP4291 in the total inoculum and the percentage of nodules formed by RmP4291 on alfalfa. Measurements made with RmP4508 (*gusA*-marked RmP4291) and RmP4506 (*celB*-marked RmP110) are shown in magenta, whereas those made with RmP4509 (*celB*-marked RmP4291) and RmP4505 (*gusA*-marked RmP110) are shown in green. The grey dot line represents the identical competitiveness of two competitors (y = x).

4.5 Minimal genomes and minimal symbiotic genomes

Streamlining a genome or constructing a minimal genome is a major undertaking in synthetic biology and has attracted substantial research interest in the past two decades. A minimal genome was initially referred to the smallest number of genetic elements sufficient to build a modern-type free-living cellular organism (Mushegian, 1999). Researchers have been trying to design and construct a minimal genome that

includes only genes essential for life for several organisms, including *Mycoplasma mycoides* (Hutchison et al., 2016), *Escherichia coli* (Smalley et al., 2003), etc. With the evolving concept of the minimal genome, a minimal genome now also refers to a minimal set of genes required for a defined biological function. Construction of a minimal genome for nitrogen fixation in nitrogen fixing organisms is one such example. Like our laboratory, others have also made great efforts minimizing the number of genes required for nitrogen fixation in other organisms. For example, (Yang et al., 2018) took the *nif* genes from *Klebsiella oxytoca* and regrouped them into gene clusters either by fusing genes, or by expressing polyproteins that were subsequently cleaved *in vivo* with Tobacco Etch Virus protease. After selection based on protein expression levels, only five giant genes were chosen to keep were able to encode nitrogenase activity and support diazotrophic growth of *Escherichia coli*.

In our project, the minimal genome is a more complicated term since it concerns the *S. meliloti* genes required for nitrogen fixation which takes place only in a symbiotic environment. For the ultimate goal of building a rhizobial minimal symbiotic genome, both the bacteria and the plant host involved in the symbiosis need to be taken into consideration. The minimal genome for the symbiosis would include not only genes required for the 'core' nitrogen fixation process itself, but also those required for other aspects of symbiosis at all stages. The issue is further complicated by the fact that requirements for those other genes are known to differ with different

hosts. Moreover, the ‘minimal genome’ may vary depending on how efficient a symbiotic phenotype is desired, as demonstrated by the on-going research in our lab. Therefore, it would be necessary to define various versions of a minimal genome for different purposes. To simplify the issue, however, one might start with the construction of a genome which includes only genes required for minimal SNF by one host. The initial minimal genome could then be modified as required.

4.6 Future directions for the symbiotic minimal genome project

So far, we have deleted 88% of pSymA and the resulting *S. meliloti* mutant is still able to fix a significant amount of nitrogen on at least four host plants. Further pSymA genome minimizing is ongoing in our lab and the preliminary assembly of the five sub-regions has been completed. Later, the SNF essentiality of individual gene in the five sub-regions will be determined by using tools such as transposon mutagenesis. As for pSymB, more efforts will be needed to revaluate the effect of each deletion mutation in pSymB, prepared previously, on symbiosis (diCenzo et al., 2016). Since the current minimal pSymB (B108-B109-B123) can’t fully replace pSymB for the role in SNF (Kearsley and Finan, unpublished), other pSymB regions must be screened on the minimal pSymB strain background to identify the missing regions required to restore the SNF level achieved with wildtype pSymB. On the other hand, despite the lack of efficiency in symbiotic nitrogen fixation by *S. meliloti* containing the minimal pSymB, it is still worthwhile to test if the combination of the current version of the

minimal pSymA and that of the minimal pSymB would permit nodulation and/or detectable symbiotic nitrogen fixation by the *S. meliloti* mutant. This test will help understand the minimal set of genes required for SNF by the organism.

Once the minimal set of 'core' genes that allow SNF to occur is obtained, one could then try to improve the SNF performance of the strain by putting back individually genes/regions that had been deleted. In this way, new genes related to the SNF process may be discovered, and new functions/correlations of the known symbiotic genes may be understood. Besides, the availability of the new histochemical staining method developed in this study would allow the analysis of the contribution of a gene/gene region to nodule competitiveness of the strain. Taking advantage of the deletion library available in the lab, a screen for loss in nodule competitiveness would be a valid start. The identified genes can then be added back to the minimal symbiotic genome. Moreover, symbiotic genes from other rhizobia could be tested and introduced into the minimal symbiotic pSymA of *S. meliloti* for possible improvement in symbiotic phenotype. For instance, it would be interesting to determine if the integration of the unique *nod* genes from other rhizobia with different host preference could increase the host range of *S. meliloti*.

4.7 Potential applications

The construction of the minimal symbiotic genome in *S. meliloti* will potentially help in many aspects in agricultural activities. By carrying out this project, we will obtain valuable insights into the process of natural symbiotic nitrogen fixation, understand the function and co-ordination of genes involved in the symbiosis. With the new knowledge, we will be able to improve the legume-rhizobia symbiosis performance by increasing the plant productivity, widening host range of the symbionts, making efficient nitrogen fixers more competitive in nodulation, etc. Furthermore, this study will also contribute to the ultimate goal of the establishment of symbiotic nitrogen fixation in other crops. On the other hand, the demonstration of the histochemical nodule staining assay provides a useful tool in research on nodule competitiveness in *S. meliloti*. The assay enables a large-scale screen of strains for nodule competitiveness, thereby helping to achieve the goal of constructing the minimal symbiotic genome by identifying pSymA and pSymB genes of importance to nodule competitiveness.

Appendix

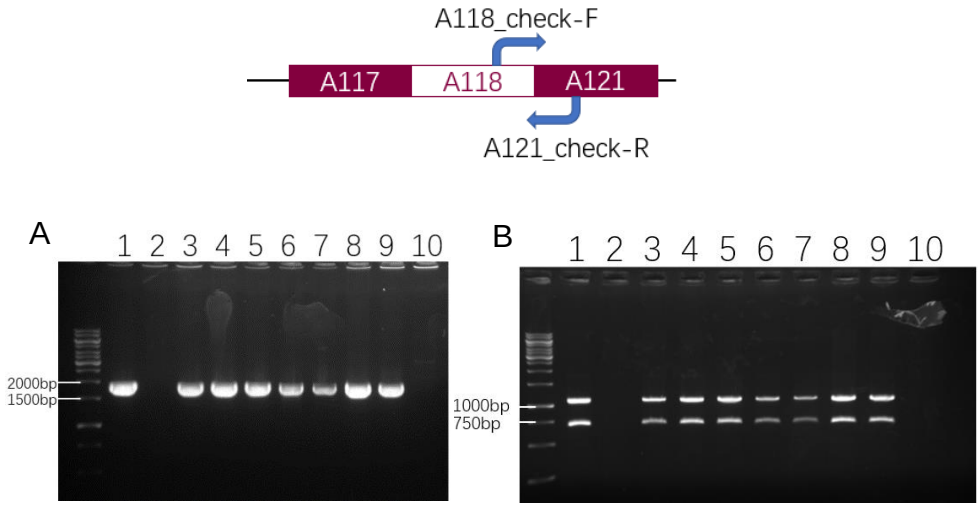


Figure A1. Analysis of the PCR products by agarose gel electrophoresis. (A) Colony PCR analysis using the primer set A118_check-II-F/A121_check-II-R with an expected product size of 1778 bp. The gel pattern confirms the presence of the A-SNF plasmid in *E. coli* M3255. (B) Agarose gel analysis of the cleavage of the PCR product from A with EcoRV. The gel pattern confirms the presence of the expected PCR fragment. For both (A) and (B), lanes 1~8, with RmP as donor different *E. coli* transconjugants; lane 9, the positive control RmP4219; lane 10, the negative control E. coli M928. As shown in both (A) and (B), products in all the sample lanes, except for lane 2, are of the expected sizes.

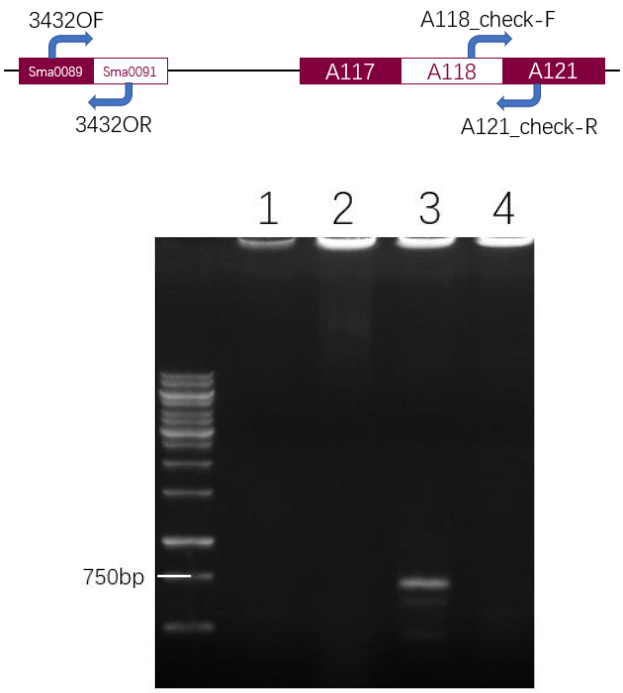


Figure A2. Colony PCR results confirming the excision of the A-SNF region from pSymA in M2608. Primers 3432OF/34320R annealed to the DNA sequences flanking the A-SNF region and the expected product would be 705 bp in size. Lanes 1 and 2, two random transconjugants from the 8 transconjugants in figure S1; lane 3, RmP4219 as the positive control; lane 4, M928 as the negative control. A 705-bp PCR product appeared only in the lane containing the positive control, indicating the absence of the A-SNF region in the sample, as expected.

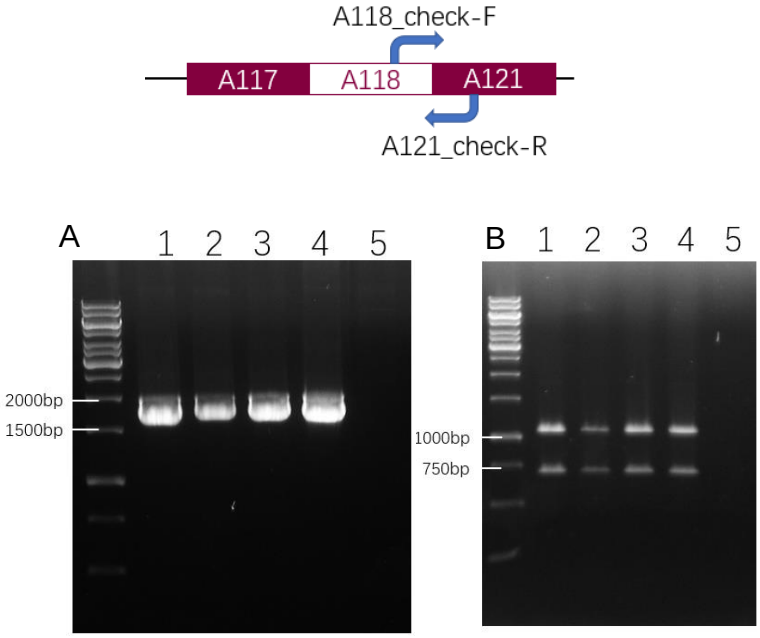


Figure A3. Confirming the presence of the A-SNF region in RmP4291 by PCR and agarose gel electrophoresis. A, Colony PCR analysis using the primer set A118_check-II-F/A121_check-II-R with an expected product size of 1778 bp. The gel pattern confirms the presence of the A-SNF plasmid in RmP4291. B, Agarose gel analysis of the cleavage of the PCR product from A with *Eco*RV with two expected bands at ~1.1 kb and ~700 bp. The gel pattern confirms the presence of the expected PCR fragment. For both A and B, lanes 1~3, different transconjugants; lane 4, the positive control RmP4219; lane 5, the negative control RmP4253.

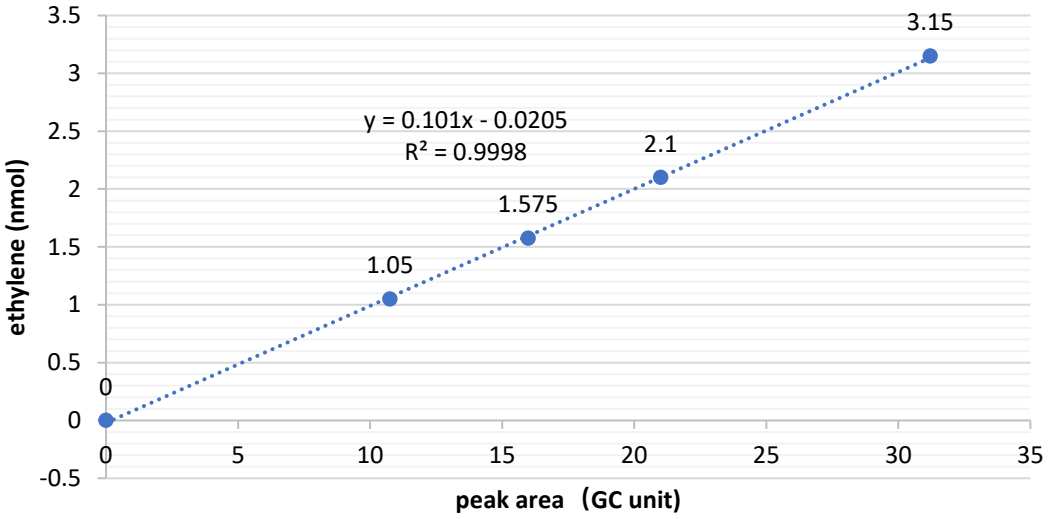


Figure A4. Standard Curve showing the relationship between the amount of ethylene (in nmol) injected and the peak area shown on the gas chromatograph (in gas chromatography units). The ethylene gas standard contained 22.6 PPM ethylene. Ethylene was injected with volumes of 0 uL, 50 uL, 75 uL, 100 uL, 150 uL. Values represent an average of five replicates.

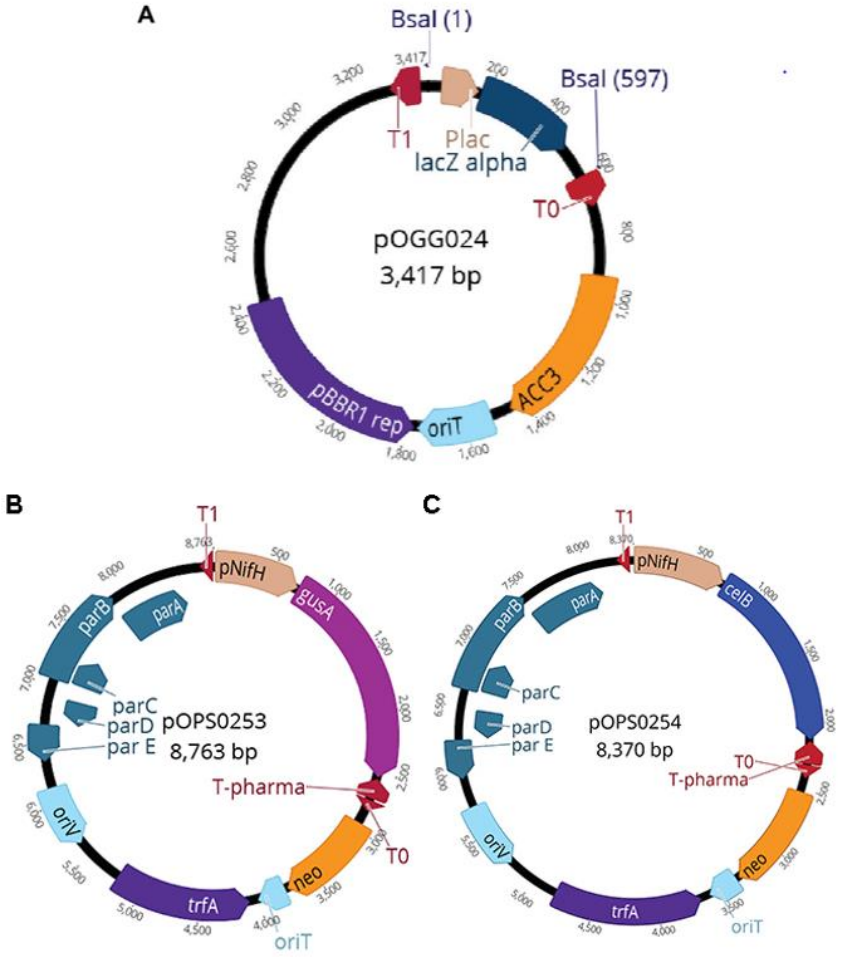


Figure A5. Schematic maps of plasmids pOGG024 (A), pOPS0253 (B) and pOPS0254 (C). In the golden gate assembly of the two marker (*gusA* and *celB*) plasmids, pOGG024 was used as the vector backbone. The *gusA* and *celB* fragments were amplified from pOPS0253 and pOPS0254, respectively. In the figure, T0 is the rho-independent transcription terminator in lambda phage while T1 represents *E. coli* T1 *rrnBT1* terminator. T-pharma is Pharmacia terminator from Invitrogen™. *parABCDE* are partition genes for plasmid stability originally from plasmid pJP2. P*lac* is the Lac promoter from *E. coli* and pNifH is the *nifH* promoter from *Rhizobium leguminosarum*. The figure is adapted from (Geddes et al., 2019).

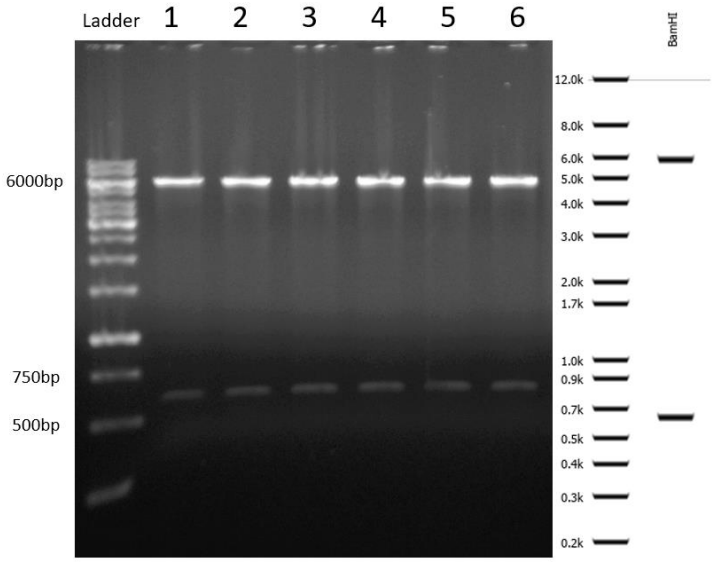


Figure A6. Verification of plasmid pTH3362 by restriction digestion. Plasmid DNAs extracted from six independent tranformants (lanes 1~6) were digested with BamHI o/n at 37°C, and the resulting digestion products were analyzed by agarose gel electrophoresis. Two bands were observed in all 6 lanes on the gel, as expected from the predicted gel pattern shown on the right (5853 bp and 603 bp). The BamHI digestion pattern of pT3362 was obtained using sequence analysis software Geneious 8 Prime. pTH3362 was thus confirmed to be successfully assembled as predicted.

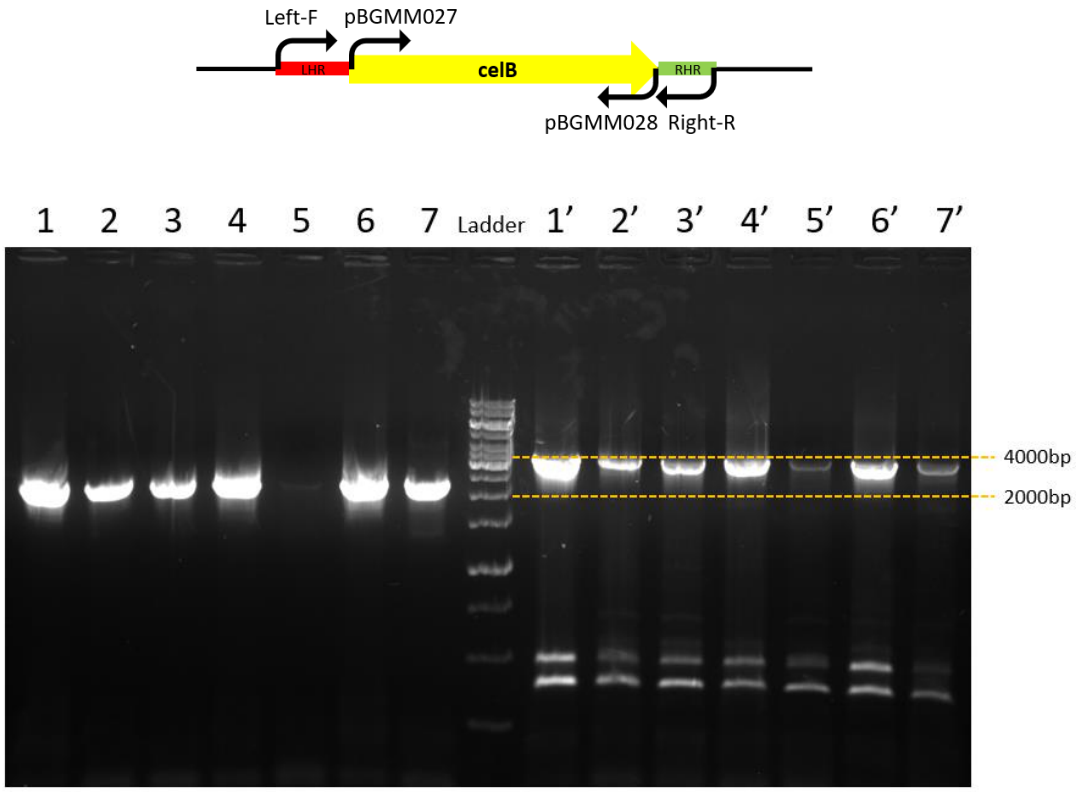


Figure A7. Verification of plasmid pTH3363 by PCR. Two set of primers (pBGMM027/pBGMM028 and Left-F/Right-R) were used in the PCR, in which putative pTH3363 plasmid samples from 7 independent transformants (lanes 1~7 or lanes 1'~7') were used as the template. The predicted primer annealing sites are shown in the schematic map on the top. Lanes 1~7, PCR products amplified by primer set pBGMM027/pBGMM028 with an expected product size of 2308 bp. Lanes 1'~7', the corresponding plasmid samples amplified by primer set Left-F/Right-R with an expected product size of 3256 bp. The gel pattern confirms that pTH3363 was assembled as designed.

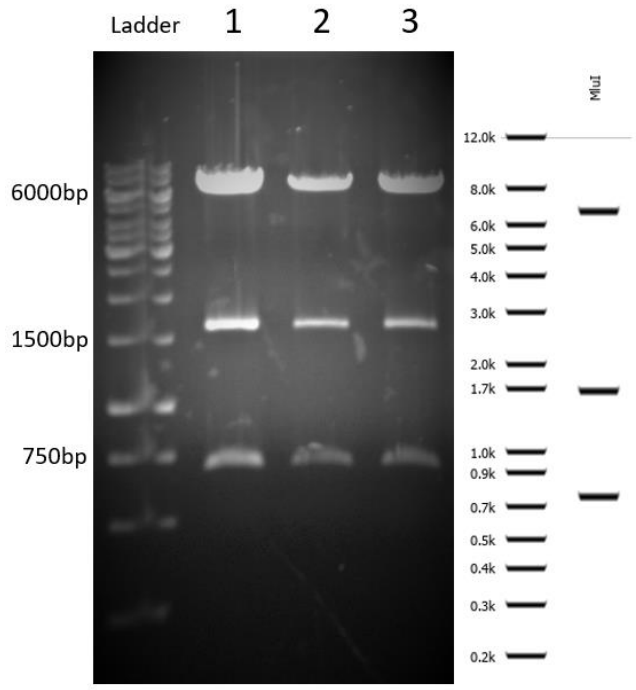


Figure A8. Verification of plasmid pTH3364 by restriction digestion. Plasmid DNAs from three independent tranformants (lanes 1~3) were digested with MluI o/n at 37°C, and the resulting digestion products were analyzed by agarose gel electrophoresis. Three bands were observed in all three lanes, as expected from the predicted pattern of pTH3364 digestion shown on the right (from top to bottom: 6631, 1621 and 702 bp). The predicted gel pattern of the plasmid digestion was obtained using sequence analysis software Geneious 8 Prime. pTH3364 was thus confirmed to be successfully assembled as designed.

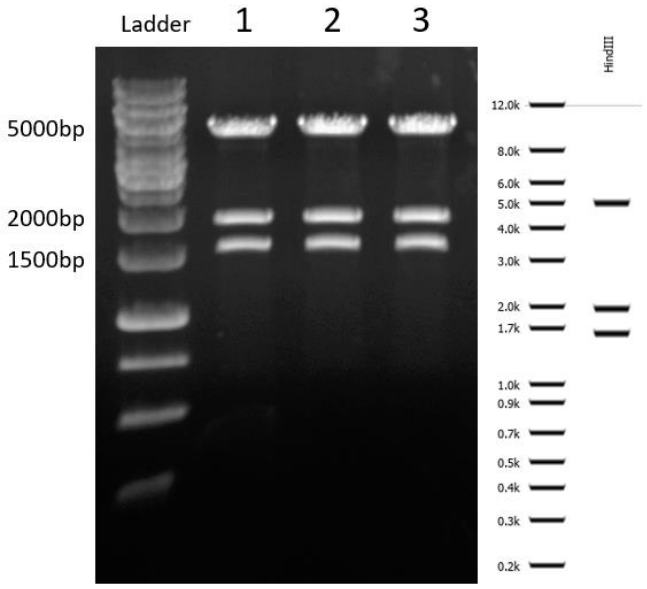


Figure A9. Verification of plasmid pTH3365 by restriction digestion. Plasmid DNAs from three independent tranformants (lanes 1~3) were digested with HindIII o/n at 37°C, and the resulting digestion products were analyzed by agarose gel electrophoresis. Three bands were observed in all three lanes, as expected from the predicted pattern of pTH3365 digestion shown on the right (From top to bottom: 5013, 1962 and 1565bp). The predicted gel pattern of the plasmid digestion was obtained using sequence analysis software Geneious 8 Prime. pTH3365 was thus confirmed to be successfully assembled as designed.

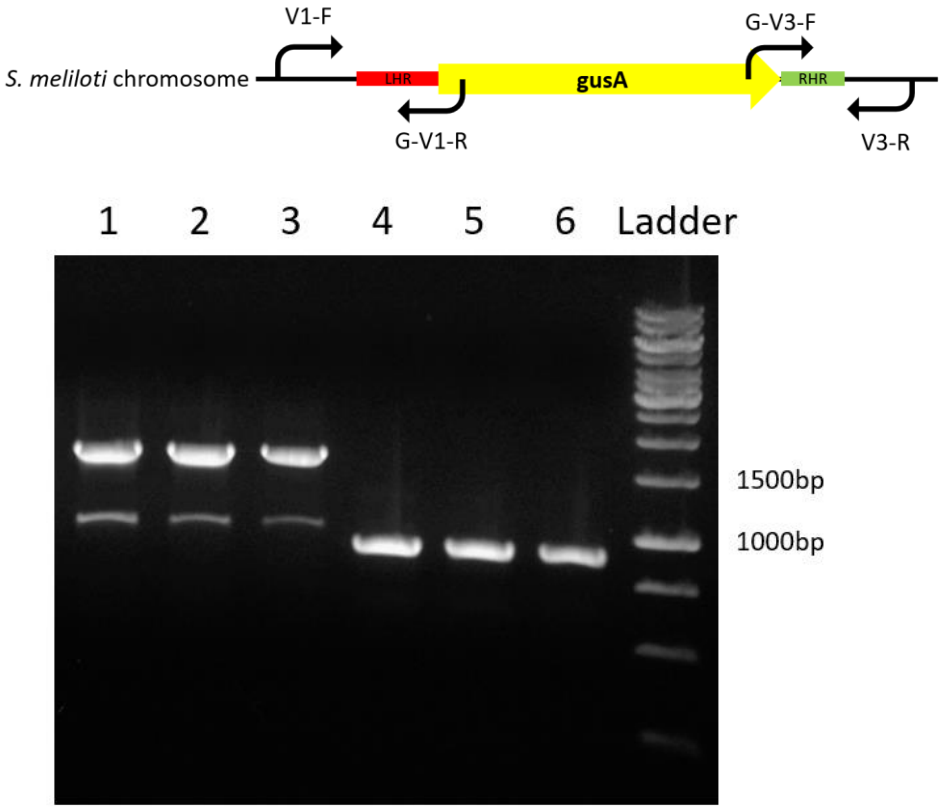


Figure A10. Confirmation of the integration of *gusA* into the chromosome of *S. meliloti* RmP110 by colony PCR. Three putative recombinants were subjected to PCR amplification (lanes 1/4, 2/5 and 3/6, respectively). RmP110 was included as a negative control (not shown). The predicted primer annealing sites are shown schematically on the top. Lanes 1~3, Primer set V1-F/G-V1-R was used with an expected product size of 1656 bp. Lanes 4~6, Primer set G-V3-F/V3-R was used with an expected product size of 866 bp. The gel pattern confirms that the *gusA* gene was successfully integrated into the chromosome of RmP110 at the designed site.

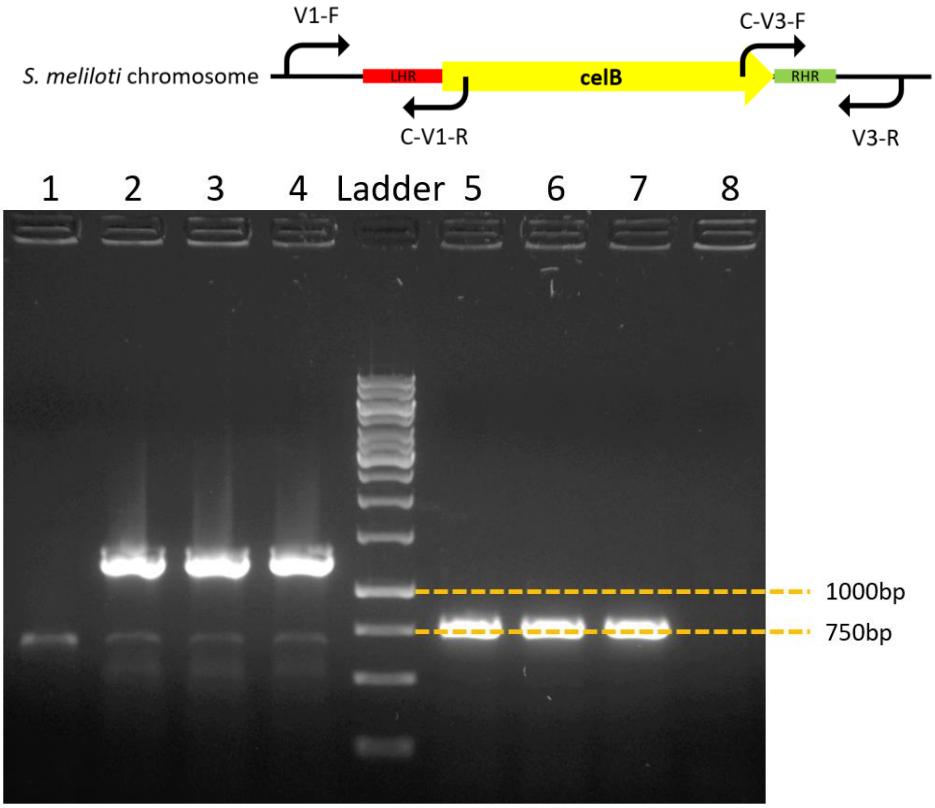


Figure A11. Confirmation of the integration of *celB* into the chromosome of *S. meliloti* RmP110 by colony PCR. Three putative recombinants were subjected to PCR amplification (lanes 2/5, 3/6 and 4/7, respectively). RmP110 was used as a negative control in lanes 1 and 8. The predicted primer annealing sites are shown schematically on the top. Lanes 1~4, Primer set V1-F/C-V1-R was used with an expected product size of 1248 bp. Lanes 5~8, primer set C-V3-F/V3-R was used with an expected product size of 753 bp. The gel pattern confirms that the *celB* gene was successfully integrated into the chromosome of RmP110 at the designed site.

References

Aguilar, O.M., Reiländer, H., Arnold, W., and Pühler, A. (1987). Rhizobium meliloti nifN (fixF) gene is part of an operon regulated by a nifA-dependent promoter and codes for a polypeptide homologous to the nifK gene product. Journal of Bacteriology *169*, 5393–5400.

Anzai, Y., Kim, H., Park, J.Y., Wakabayashi, H., and Oyaizu, H. (2000). Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence. Int. J. Syst. Evol. Microbiol. *50 Pt 4*, 1563–1589.

Appleby, C.A. (1984). Leghemoglobin and Rhizobium Respiration. Annual Review of Plant Physiology *35*, 443–478.

Ardourel, M., Demont, N., Debellé, F., Maillet, F., Billy, F. de, Promé, J.C., Dénarié, J., and Truchet, G. (1994). Rhizobium meliloti lipooligosaccharide nodulation factors: different structural requirements for bacterial entry into target root hair cells and induction of plant symbiotic developmental responses. The Plant Cell *6*, 1357–1374.

Ardourel, M., Lortet, G., Maillet, F., Roche, P., Truchet, G., Promé, J.C., and Rosenberg, C. (1995). In Rhizobium meliloti, the operon associated with the nod box n5 comprises nodL, noeA and noeB, three host-range genes specifically required for the nodulation of particular Medicago species. Mol. Microbiol. *17*, 687–699.

Arnold, M.F.F., Shabab, M., Penterman, J., Boehme, K.L., Griffitts, J.S., and Walker, G.C. (2017). Genome-Wide Sensitivity Analysis of the Microsymbiont Sinorhizobium meliloti to Symbiotically Important, Defensin-Like Host Peptides. MBio *8*, e01060-17.

Baev, N., Endre, G., Petrovics, G., Banfalvi, Z., and Kondorosi, A. (1991). Six nodulation genes of nod box locus 4 in Rhizobium meliloti are involved in nodulation signal production: nodM codes for d-glucosamine synthetase. Molec. Gen. Genet. *228*, 113–124.

Barnett, M.J., and Long, S.R. (1997). Identification and Characterization of a Gene on Rhizobium meliloti pSyma, syrB, That Negatively Affects syrM Expression. MPMI *10*, 550–559.

Barnett, M.J., and Long, S.R. (2015). The Sinorhizobium meliloti SyrM regulon: effects on global gene expression are mediated by syrA and nodD3. J. Bacteriol. *197*, 1792–1806.

Barnett, M.J., Swanson, J.A., and Long, S.R. (1998). Multiple Genetic Controls on Rhizobium meliloti syrA, a Regulator of Exopolysaccharide Abundance. Genetics *148*, 19–32.

Barnett, M.J., Fisher, R.F., Jones, T., Komp, C., Abola, A.P., Barloy-Hubler, F., Bowser, L., Capela, D., Galibert, F., Gouzy, J., et al. (2001). Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti pSymA megaplasmid. Proc Natl Acad Sci U S A *98*, 9883–9888.

Barran, L.R., Bromfield, E.S.P., and Brown, D.C.W. (2002). Identification and cloning of the bacterial nodulation specificity gene in the Sinorhizobium meliloti – Medicago laciniata symbiosis. Can. J. Microbiol. *48*, 765–771.

Batut, J., Daveran-Mingot, M.L., David, M., Jacobs, J., Garnerone, A.M., and Kahn, D. (1989). fixK, a gene homologous with fnr and crp from Escherichia coli, regulates nitrogen fixation genes both positively and negatively in Rhizobium meliloti. The EMBO Journal *8*, 1279–1286.

Bauer, M.W., Bylina, E.J., Swanson, R.V., and Kelly, R.M. (1996). Comparison of a beta-glucosidase and a beta-mannosidase from the hyperthermophilic archaeon Pyrococcus furiosus. Purification, characterization, gene cloning, and sequence analysis. J. Biol. Chem. *271*, 23749–23755.

Blanquet, P., Silva, L., Catrice, O., Bruand, C., Carvalho, H., and Meilhoc, E. (2015). Sinorhizobium meliloti Controls Nitric Oxide–Mediated Post-Translational Modification of a Medicago truncatula Nodule Protein. MPMI *28*, 1353–1363.

Bobik, C., Meilhoc, E., and Batut, J. (2006). FixJ: a Major Regulator of the Oxygen Limitation Response and Late Symbiotic Functions of Sinorhizobium meliloti. Journal of Bacteriology *188*, 4890–4902.

Buikema, W.J., Klingensmith, J.A., Gibbons, S.L., and Ausubel, F.M. (1987). Conservation of structure and location of Rhizobium meliloti and Klebsiella pneumoniae nifB genes. Journal of Bacteriology *169*, 1120–1126.

Burén, S., and Rubio, L.M. (2018). State of the art in eukaryotic nitrogenase engineering. FEMS Microbiol. Lett. *365*.

Canfield, D.E., Glazer, A.N., and Falkowski, P.G. (2010). The evolution and future of Earth's nitrogen cycle. Science *330*, 192–196.

Carlson, R.W., Price, N.P., and Stacey, G. (1994). The biosynthesis of rhizobial lipo-oligosaccharide nodulation signal molecules. Mol. Plant Microbe Interact. *7*, 684–695.

Chan, Y.K., McCormick, W.A., and Watson, R.J. (1997). A new nos gene downstream from nosDFY is essential for dissimilatory reduction of nitrous oxide by Rhizobium (Sinorhizobium) meliloti. Microbiology (Reading, Engl.) *143 (Pt 8)*, 2817–2824.

Cheng, H.P., and Walker, G.C. (1998). Succinoglycan is required for initiation and elongation of infection threads during nodulation of alfalfa by Rhizobium meliloti. J. Bacteriol. *180*, 5183–5191.

Cheng, Q., Day, A., Dowson-Day, M., Shen, G.-F., and Dixon, R. (2005). The Klebsiella pneumoniae nitrogenase Fe protein gene (nifH) functionally substitutes for the chlL gene in Chlamydomonas reinhardtii. Biochemical and Biophysical Research Communications *329*, 966–975.

Cosseau, C., and Batut, J. (2004). Genomics of the ccoNOQP-encoded cbb3 oxidase complex in bacteria. Arch Microbiol *181*, 89–96.

Cosseau, C., Garnerone, A.M., and Batut, J. (2002). The FixM Flavoprotein Modulates Inhibition by AICAR or 5′AMP of Respiratory and Nitrogen Fixation Gene Expression in Sinorhizobium meliloti. MPMI *15*, 598–607.

Cowie, A., Cheng, J., Sibley, C.D., Fong, Y., Zaheer, R., Patten, C.L., Morton, R.M., Golding, G.B., and Finan, T.M. (2006). An Integrated Approach to Functional Genomics: Construction of a Novel Reporter Gene Fusion Library for Sinorhizobium meliloti. Appl. Environ. Microbiol. *72*, 7156–7167.

Crans, D.C., Smee, J.J., Gaidamauskas, E., and Yang, L. (2004). The Chemistry and Biochemistry of Vanadium and the Biological Activities Exerted by Vanadium Compounds. Chem. Rev. *104*, 849–902.

David, M., Daveran, M.-L., Batut, J., Dedieu, A., Domergue, O., Ghai, J., Hertig, C., Boistard, P., and Kahn, D. (1988). Cascade regulation of nif gene expression in Rhizobium meliloti. Cell *54*, 671–683.

Debellé, F., Plazanet, C., Roche, P., Pujol, C., Savagnac, A., Rosenberg, C., Promé, J.C., and Dénarié, J. (1996). The NodA proteins of Rhizobium meliloti and Rhizobium tropici specify the N-acylation of Nod factors by different fatty acids. Mol. Microbiol. *22*, 303–314.

Demont, N., Debellé, F., Aurelle, H., Dénarié, J., and Promé, J.C. (1993). Role of the Rhizobium meliloti nodF and nodE genes in the biosynthesis of lipo-oligosaccharidic nodulation factors. J. Biol. Chem. *268*, 20134–20142.

diCenzo, G., Milunovic, B., Cheng, J., and Finan, T.M. (2013). The tRNAarg gene and engA are essential genes on the 1.7-Mb pSymB megaplasmid of Sinorhizobium meliloti and were translocated together from the chromosome in an ancestral strain. J. Bacteriol. *195*, 202–212.

diCenzo, G.C., MacLean, A.M., Milunovic, B., Golding, G.B., and Finan, T.M. (2014). Examination of Prokaryotic Multipartite Genome Evolution through Experimental Genome Reduction. PLOS Genetics *10*, e1004742.

diCenzo, G.C., Zamani, M., Cowie, A., and Finan, T.M. (2015). Proline auxotrophy in Sinorhizobium meliloti results in a plant-specific symbiotic phenotype. Microbiology *161*, 2341–2351.

diCenzo, G.C., Zamani, M., Milunovic, B., and Finan, T.M. (2016). Genomic resources for identification of the minimal N2 -fixing symbiotic genome. Environ. Microbiol. *18*, 2534–2547.

diCenzo, G.C., Zamani, M., Ludwig, H.N., and Finan, T.M. (2017). Heterologous Complementation Reveals a Specialized Activity for BacA in the Medicago–Sinorhizobium meliloti Symbiosis. MPMI *30*, 312–324.

van Dillewijn, P., Soto, M.J., Villadas, P.J., and Toro, N. (2001). Construction and Environmental Release of a Sinorhizobium meliloti Strain Genetically Modified To Be More Competitive for Alfalfa Nodulation. Appl Environ Microbiol *67*, 3860–3865.

Dixon, R., and Kahn, D. (2004). Genetic regulation of biological nitrogen fixation. Nat Rev Microbiol *2*, 621–631.

Downie, J.A. (2005). Legume Haemoglobins: Symbiotic Nitrogen Fixation Needs Bloody Nodules. Current Biology *15*, R196–R198.

Downie, J.A. (2010). The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. FEMS Microbiol Rev *34*, 150–170.

Dusha, I., Kovalenko, S., Banfalvi, Z., and Kondorosi, A. (1987a). Rhizobium meliloti insertion element ISRm2 and its use for identification of the fixX gene. J. Bacteriol. *169*, 1403–1409.

Dusha, I., Kovalenko, S., Banfalvi, Z., and Kondorosi, A. (1987b). Rhizobium meliloti insertion element ISRm2 and its use for identification of the fixX gene. Journal of Bacteriology *169*, 1403–1409.

Earl, C.D., Ronson, C.W., and Ausubel, F.M. (1987). Genetic and structural analysis of the Rhizobium meliloti fixA, fixB, fixC, and fixX genes. J. Bacteriol. *169*, 1127–1136.

Ehrhardt, D.W., Atkinson, E.M., Faull, K.F., Freedberg, D.I., Sutherlin, D.P., Armstrong, R., and Long, S.R. (1995). In vitro sulfotransferase activity of NodH, a nodulation protein of Rhizobium meliloti required for host-specific nodulation. Journal of Bacteriology *177*, 6237–6245.

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.

Finan, T.M., Hartweig, E., LeMieux, K., Bergman, K., Walker, G.C., and Signer, E.R. (1984). General transduction in Rhizobium meliloti. Journal of Bacteriology *159*, 120–124.

Finan, T.M., Weidner, S., Wong, K., Buhrmester, J., Chain, P., Vorhölter, F.J., Hernandez-Lucas, I., Becker, A., Cowie, A., Gouzy, J., et al. (2001). The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endosymbiont Sinorhizobium meliloti. Proc Natl Acad Sci U S A *98*, 9889–9894.

Foussard, M., Garnerone, A.M., Ni, F., Soupène, E., Boistard, P., and Batut, J. (1997). Negative autoregulation of the Rhizobium meliloti fixK gene is indirect and requires a newly identified regulator, FixT. Mol. Microbiol. *25*, 27–37.

Fox, A.R., Soto, G., Valverde, C., Russo, D., Lagares, A., Zorreguieta, Á., Alleva, K., Pascuan, C., Frare, R., Mercado-Blanco, J., et al. (2016). Major cereal crops benefit from biological nitrogen fixation when inoculated with the nitrogen-fixing bacterium Pseudomonas protegens Pf-5 X940. Environ. Microbiol. *18*, 3522–3534.

Galibert, F., Finan, T.M., Long, S.R., Puhler, A., Abola, P., Ampe, F., Barloy-Hubler, F., Barnett, M.J., Becker, A., Boistard, P., et al. (2001). The composite genome of the legume symbiont Sinorhizobium meliloti. Science *293*, 668–672.

Galloway, J.N., Townsend, A.R., Erisman, J.W., Bekunda, M., Cai, Z., Freney, J.R., Martinelli, L.A., Seitzinger, S.P., and Sutton, M.A. (2008). Transformation of the

nitrogen cycle: recent trends, questions, and potential solutions. Science *320*, 889–892.

Geddes, B.A., Mendoza-Suárez, M.A., and Poole, P.S. (2019). A Bacterial Expression Vector Archive (BEVA) for Flexible Modular Assembly of Golden Gate-Compatible Vectors. Front. Microbiol. *9*.

Geremia, R.A., Mergaert, P., Geelen, D., Montagu, M.V., and Holsters, M. (1994). The NodC protein of Azorhizobium caulinodans is an N-acetylglucosaminyltransferase. PNAS *91*, 2669–2673.

Geurts, R., and Bisseling, T. (2002). Rhizobium Nod Factor Perception and Signalling. Plant Cell *14*, s239–s249.

Geurts, R., Fedorova, E., and Bisseling, T. (2005). Nod factor signaling genes and their function in the early stages of Rhizobium infection. Current Opinion in Plant Biology *8*, 346–352.

del Giudice, J., Cam, Y., Damiani, I., Fung-Chat, F., Meilhoc, E., Bruand, C., Brouquisse, R., Puppo, A., and Boscari, A. (2011). Nitric oxide is required for an optimal establishment of the Medicago truncatula-Sinorhizobium meliloti symbiosis. New Phytol. *191*, 405–417.

Harper, J.E., and Gibson, A.H. (1984). Differential Nodulation Tolerance to Nitrate Among Legume Species. Crop Science *24*, 797–801.

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 in Microbiology *18*, 141–148.

Hartwig, U.A., Maxwell, C.A., Joseph, C.M., and Phillips, D.A. (1990). Effects of alfalfa nod gene-inducing flavonoids on nodABC transcription in Rhizobium meliloti strains containing different nodD genes. J Bacteriol *172*, 2769–2773.

Hirsch, A.M., Bang, M., and Ausubel, F.M. (1983). Ultrastructural analysis of ineffective alfalfa nodules formed by nif::Tn5 mutants of Rhizobium meliloti. Journal of Bacteriology *155*, 367–380.

Holloway, P., McCormick, W., Watson, R.J., and Chan, Y.K. (1996). Identification and analysis of the dissimilatory nitrous oxide reduction genes, nosRZDFY, of Rhizobium meliloti. Journal of Bacteriology *178*, 1505–1514.

Honma, M.A., and Ausubel, F.M. (1987a). Rhizobium meliloti has three functional copies of the nodD symbiotic regulatory gene. PNAS *84*, 8558–8562.

Honma, M.A., and Ausubel, F.M. (1987b). Rhizobium meliloti has three functional copies of the nodD symbiotic regulatory gene. Proc. Natl. Acad. Sci. U.S.A. *84*, 8558–8562.

Honma, M.A., Asomaning, M., and Ausubel, F.M. (1990). Rhizobium meliloti nodD genes mediate host-specific activation of nodABC. J. Bacteriol. *172*, 901–911.

Hu, Y., and Ribbe, M.W. (2013). Nitrogenase assembly. Biochim. Biophys. Acta *1827*, 1112–1122.

Hutchison, C.A., Chuang, R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., Ma, L., et al. (2016). Design and synthesis of a minimal bacterial genome. Science *351*, aad6253.

Ivleva, N.B., Groat, J., Staub, J.M., and Stephens, M. (2016). Expression of Active Subunit of Nitrogenase via Integration into Plant Organelle Genome. PLoS ONE *11*, e0160951.

John, M., Röhrig, H., Schmidt, J., Wieneke, U., and Schell, J. (1993). Rhizobium NodB protein involved in nodulation signal synthesis is a chitooligosaccharide deacetylase. PNAS *90*, 625–629.

Kahn, D., David, M., Domergue, O., Daveran, M.L., Ghai, J., Hirsch, P.R., and Batut, J. (1989). Rhizobium meliloti fixGHI sequence predicts involvement of a specific cation pump in symbiotic nitrogen fixation. Journal of Bacteriology *171*, 929–939.

Keating, D.H., Willits, M.G., and Long, S.R. (2002). A Sinorhizobium meliloti Lipopolysaccharide Mutant Altered in Cell Surface Sulfation. Journal of Bacteriology *184*, 6681–6689.

Klipp, W., Reiländer, H., Schlüter, A., Krey, R., and Pühler, A. (1989a). The Rhizobium meliloti fdxN gene encoding a ferredoxin-like protein is necessary for nitrogen fixation and is cotranscribed with nifA and nifB. Mol Gen Genet *216*, 293–302.

Klipp, W., Reiländer, H., Schlüter, A., Krey, R., and Pühler, A. (1989b). The Rhizobium meliloti fdxN gene encoding a ferredoxin-like protein is necessary for nitrogen fixation and is cotranscribed with nifA and nifB. Mol Gen Genet *216*, 293–302.

Klipp, W., Masepohl, B., Gallon, J.R., and Newton, W.E. (2006). Genetics and Regulation of Nitrogen Fixation in Free-Living Bacteria (Springer Science & Business Media).

Kohler, P.R.A., Zheng, J.Y., Schoffers, E., and Rossbach, S. (2010). Inositol catabolism, a key pathway in sinorhizobium meliloti for competitive host nodulation. Appl. Environ. Microbiol. *76*, 7972–7980.

Kopat, V.V., Chirak, E.R., Kimeklis, A.K., Safronova, V.I., Belimov, A.A., Kabilov, M.R., Andronov, E.E., and Provorov, N.A. (2017). Evolution of fixNOQP genes encoding cytochrome oxidase with high affinity to oxygen in rhizobia and related bacteria. Russ J Genet *53*, 766–774.

Krol, E., and Becker, A. (2004a). Global transcriptional analysis of the phosphate starvation response in Sinorhizobium meliloti strains 1021 and 2011. Molecular Genetics and Genomics *272*.

Krol, E., and Becker, A. (2004b). Global transcriptional analysis of the phosphate starvation response in Sinorhizobium meliloti strains 1021 and 2011. Mol Genet Genomics *272*, 1–17.

Lee, J.H., Skowron, P.M., Rutkowska, S.M., Hong, S.S., and Kim, S.C. (1996). Sequential amplification of cloned DNA as tandem multimers using class-IIS restriction enzymes. Genetic Analysis: Biomolecular Engineering *13*, 139–145.

Lerouge, P., Roche, P., Faucher, C., Maillet, F., Truchet, G., Promé, J.C., and Dénarié, J. (1990). Symbiotic host-specificity of Rhizobium meliloti is determined by a sulphated and acylated glucosamine oligosaccharide signal. Nature *344*, 781–784.

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

López-Lara, I.M., and Geiger, O. (2001). The Nodulation Protein NodG Shows the Enzymatic Activity of an 3-Oxoacyl-Acyl Carrier Protein Reductase. MPMI *14*, 349–357.

López-Torrejón, G., Jiménez-Vicente, E., Buesa, J.M., Hernandez, J.A., Verma, H.K., and Rubio, L.M. (2016). Expression of a functional oxygen-labile nitrogenase component in the mitochondrial matrix of aerobically grown yeast. Nat Commun *7*, 1–6.

L. Zielazinski, E., González-Guerrero, M., Subramanian, P., L. Stemmler, T., M. Argüello, J., and C. Rosenzweig, A. (2013). Sinorhizobium meliloti Nia is a P

1B-5 -ATPase expressed in the nodule during plant symbiosis and is involved in Ni and Fe transport. Metallomics *5*, 1614–1623.

Ma, C.-H., Kwiatek, A., Bolusani, S., Voziyanov, Y., and Jayaram, M. (2007). Unveiling hidden catalytic contributions of the conserved His/Trp-III in tyrosine recombinases: assembly of a novel active site in Flp recombinase harboring alanine at this position. J Mol Biol *368*, 183–196.

Maillet, F., Poinsot, V., André, O., Puech-Pagès, V., Haouy, A., Gueunier, M., Cromer, L., Giraudet, D., Formey, D., Niebel, A., et al. (2011). Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. Nature *469*, 58–63.

Meilhoc, E., Cam, Y., Skapski, A., and Bruand, C. (2010). The Response to Nitric Oxide of the Nitrogen-Fixing Symbiont Sinorhizobium meliloti. MPMI *23*, 748–759.

Milunovic, B., diCenzo, G.C., Morton, R.A., and Finan, T.M. (2014). Cell Growth Inhibition upon Deletion of Four Toxin-Antitoxin Loci from the Megaplasmids of Sinorhizobium meliloti. J. Bacteriol. *196*, 811–824.

Moreno-Vivian, C., Schmehl, M., Masepohl, B., Arnold, W., and Klipp, W. (1989a). DNA sequence and genetic analysis of the Rhodobacter capsulatus nifENX gene region: Homology between NifX and NifB suggests involvement of NifX in processing of the iron-molybdenum cofactor. Mol Gen Genet *216*, 353–363.

Moreno-Vivian, C., Hennecke, S., Pühler, A., and Klipp, W. (1989b). Open reading frame 5 (ORF5), encoding a ferredoxinlike protein, and nifQ are cotranscribed with nifE, nifN, nifX, and ORF4 in Rhodobacter capsulatus. Journal of Bacteriology *171*, 2591–2598.

Mulligan, J.T., and Long, S.R. (1985). Induction of Rhizobium meliloti nodC expression by plant exudate requires nodD. PNAS *82*, 6609–6613.

Mus, F., Crook, M.B., Garcia, K., Garcia Costas, A., Geddes, B.A., Kouri, E.D., Paramasivan, P., Ryu, M.-H., Oldroyd, G.E.D., Poole, P.S., et al. (2016). Symbiotic Nitrogen Fixation and the Challenges to Its Extension to Nonlegumes. Appl. Environ. Microbiol. *82*, 3698–3710.

Mushegian, A. (1999). The minimal genome concept. Current Opinion in Genetics & Development *9*, 709–714.

Ogawa, J., Brierley, H.L., and Long, S.R. (1991). Analysis of Rhizobium meliloti nodulation mutant WL131: novel insertion sequence ISRm3 in nodG and altered nodH protein product. Journal of Bacteriology *173*, 3060–3065.

Onishchuk, O.P., Sharypova, L.A., Kurchak, O.N., Becker, A., and Simarov, B.V. (2005). Identification of Sinorhizobium meliloti Genes Influencing Synthesis of Surface Polysaccharides and Competitiveness. Russ J Genet *41*, 1337–1342.

Oresnik, I.J., Charles, T.C., and Finan, T.M. (1994). Second site mutations specifically suppress the Fix- phenotype of Rhizobium meliloti ndvF mutations on alfalfa: identification of a conditional ndvF-dependent mucoid colony phenotype. Genetics *136*, 1233–1243.

Oresnik, I.J., Liu, S.-L., Yost, C.K., and Hynes, M.F. (2000). Megaplasmid pRme2011a of Sinorhizobium meliloti Is Not Required for Viability. Journal of Bacteriology *182*, 3582–3586.

Patankar, A.V., and González, J.E. (2009). An Orphan LuxR Homolog of Sinorhizobium meliloti Affects Stress Adaptation and Competition for Nodulation. Appl Environ Microbiol *75*, 946–955.

Paulsen, I.T., Press, C.M., Ravel, J., Kobayashi, D.Y., Myers, G.S.A., Mavrodi, D.V., DeBoy, R.T., Seshadri, R., Ren, Q., Madupu, R., et al. (2005). Complete genome sequence of the plant commensal Pseudomonas fluorescens Pf-5. Nat. Biotechnol. *23*, 873–878.

Peck, M.C., Fisher, R.F., and Long, S.R. (2006). Diverse Flavonoids Stimulate NodD1 Binding to nod Gene Promoters in Sinorhizobium meliloti. Journal of Bacteriology *188*, 5417–5427.

Peoples, M.B., Herridge, D.F., and Ladha, J.K. (1995). Biological nitrogen fixation: An efficient source of nitrogen for sustainable agricultural production? In Management of Biological Nitrogen Fixation for the Development of More Productive and Sustainable Agricultural Systems: Extended Versions of Papers Presented at the Symposium on Biological Nitrogen Fixation for Sustainable Agriculture at the 15th Congress of Soil Science, Acapulco, Mexico, 1994, J.K. Ladha, and M.B. Peoples, eds. (Dordrecht: Springer Netherlands), pp. 3–28.

Perchat, N., Saaidi, P.-L., Darii, E., Pellé, C., Petit, J.-L., Besnard-Gonnet, M., Berardinis, V. de, Dupont, M., Gimbernat, A., Salanoubat, M., et al. (2018). Elucidation of the trigonelline degradation pathway reveals previously undescribed enzymes and metabolites. PNAS *115*, E4358–E4367.

Perrine, F.M., Hocart, C.H., Hynes, M.F., and Rolfe, B.G. (2005). Plasmid-associated genes in the model micro-symbiont Sinorhizobium meliloti 1021 affect the growth and development of young rice seedlings. Environmental Microbiology *7*, 1826–1838.

Phillips, D.A., Joseph, C.M., and Maxwell, C.A. (1992). Trigonelline and Stachydrine Released from Alfalfa Seeds Activate NodD2 Protein in Rhizobium meliloti. Plant Physiology *99*, 1526–1531.

Preisig, O., Anthamatten, D., and Hennecke, H. (1993). Genes for a microaerobically induced oxidase complex in Bradyrhizobium japonicum are essential for a nitrogen-fixing endosymbiosis. PNAS *90*, 3309–3313.

Prell, J., Boesten, B., Poole, P., and Priefer, U.B. (2002). The Rhizobium leguminosarum bv. viciae VF39 gamma-aminobutyrate (GABA) aminotransferase gene (gabT) is induced by GABA and highly expressed in bacteroids. Microbiology (Reading, Engl.) *148*, 615–623.

Puskás, L.G., Nagy, Z.B., Kelemen, J.Z., Rüberg, S., Bodogai, M., Becker, A., and Dusha, I. (2004). Wide-range transcriptional modulating effect of ntrR under microaerobiosis in Sinorhizobium meliloti. Mol Genet Genomics *272*, 275–289.

Quandt, J., and Hynes, M.F. (1993). Versatile suicide vectors which allow direct selection for gene replacement in Gram-negative bacteria. Gene *127*, 15–21.

Renalier, M.H., Batut, J., Ghai, J., Terzaghi, B., Gherardi, M., David, M., Garnerone, A.M., Vasse, J., Truchet, G., and Huguet, T. (1987). A new symbiotic cluster on the pSym megaplasmid of Rhizobium meliloti 2011 carries a functional fix gene repeat and a nod locus. Journal of Bacteriology *169*, 2231–2238.

Reyes-González, A., Talbi, C., Rodríguez, S., Rivera, P., Zamorano-Sánchez, D., and Girard, L. (2016). Expanding the regulatory network that controls nitrogen fixation in Sinorhizobium meliloti: elucidating the role of the two-component system hFixL-FxkR. Microbiology, *162*, 979–988.

Rogers, C., and Oldroyd, G.E.D. (2014). Synthetic biology approaches to engineering the nitrogen symbiosis in cereals. J Exp Bot *65*, 1939–1946.

Sánchez-Cañizares, C., and Palacios, J. (2013). Construction of a marker system for the evaluation of competitiveness for legume nodulation in Rhizobium strains. J. Microbiol. Methods *92*, 246–249.

Scharff, L.B., and Bock, R. (2014). Synthetic biology in plastids. The Plant Journal *78*, 783–798.

Schindelin, H., Kisker, C., Schlessman, J.L., Howard, J.B., and Rees, D.C. (1997). Structure of ADP x AIF4(-)-stabilized nitrogenase complex and its implications for signal transduction. Nature *387*, 370–376.

Schlessman, J.L., Woo, D., Joshua-Tor, L., Howard, J.B., and Rees, D.C. (1998). Conformational variability in structures of the nitrogenase iron proteins from Azotobacter vinelandii and Clostridium pasteurianum11Edited by I. A. Wilson. Journal of Molecular Biology *280*, 669–685.

Schwedock, J.S., Liu, C., Leyh, T.S., and Long, S.R. (1994). Rhizobium meliloti NodP and NodQ form a multifunctional sulfate-activating complex requiring GTP for activity. Journal of Bacteriology *176*, 7055–7064.

Seefeldt, L.C., Hoffman, B.M., and Dean, D.R. (2009). Mechanism of Mo-Dependent Nitrogenase. Annual Review of Biochemistry *78*, 701–722.

Senecoff, J.F., Bruckner, R.C., and Cox, M.M. (1985). The FLP recombinase of the yeast 2-micron plasmid: characterization of its recombination site. Proc. Natl. Acad. Sci. U.S.A. *82*, 7270–7274.

Sessitsch, A., Wilson, K.J., Akkermans, A.D., and Vos, W.M. de (1996). Simultaneous detection of different Rhizobium strains marked with either the Escherichia coli gusA gene or the Pyrococcus furiosus celB gene. Appl. Environ. Microbiol. *62*, 4191–4194.

Setten, L., Soto, G., Mozzicafreddo, M., Fox, A.R., Lisi, C., Cuccioloni, M., Angeletti, M., Pagano, E., Díaz-Paleo, A., and Ayub, N.D. (2013). Engineering Pseudomonas protegens Pf-5 for Nitrogen Fixation and its Application to Improve Plant Growth under Nitrogen-Deficient Conditions. PLoS One *8*.

Slabas, A.R., Chase, D., Nishida, I., Murata, N., Sidebottom, C., Safford, R., Sheldon, P.S., Kekwick, R.G., Hardie, D.G., and Mackintosh, R.W. (1992). Molecular cloning of higher-plant 3-oxoacyl-(acyl carrier protein) reductase. Sequence identities with the nodG-gene product of the nitrogen-fixing soil bacterium Rhizobium meliloti. Biochemical Journal *283*, 321–326.

Smalley, D.J., Whiteley, M., and Conway, T. (2003). In search of the minimal Escherichia coli genome. Trends in Microbiology *11*, 6–8.

Smith, L.S., and Long, S.R. (1998). Requirements for syrM and nodD Genes in the Nodulation of Medicago truncatula by Rhizobium meliloti 1021. MPMI *11*, 937–940.

Streeter, J., and Wong, P.P. (1988). Inhibition of legume nodule formation and N2 fixation by nitrate. Critical Reviews in Plant Sciences *7*, 1–23.

Swanson, J.A., Mulligan, J.T., and Long, S.R. (1993). Regulation of syrM and nodD3 in Rhizobium meliloti. Genetics *134*, 435–444.

Szeto, W.W., Lynn Zimmerman, J., Sundaresan, V., and Ausubel, F.M. (1984). A Rhizobium meliloti symbiotic regulatory gene. Cell *36*, 1035–1043.

Tezcan, F.A., Kaiser, J.T., Mustafi, D., Walton, M.Y., Howard, J.B., and Rees, D.C. (2005). Nitrogenase complexes: multiple docking sites for a nucleotide switch protein. Science *309*, 1377–1380.

Török, I., and Kondorosi, A. (1981). Nucleotide sequence of the R.meliloti nitrogenase reductase (nifH) gene. Nucleic Acids Res. *9*, 5711–5723.

Torres, M.J., Hidalgo-García, A., Bedmar, E.J., and Delgado, M.J. (2013). Functional analysis of the copy 1 of the fixNOQP operon of Ensifer meliloti under free-living micro-oxic and symbiotic conditions. Journal of Applied Microbiology *114*, 1772–1781.

Triplett, E.W., and Sadowsky, M.J. (1992). Genetics of Competition for Nodulation of Legumes. Annual Review of Microbiology *46*, 399–422.

Truchet, G., Roche, P., Lerouge, P., Vasse, J., Camut, S., Billy, F. de, Promé, J.-C., and Dénarié, J. (1991). Sulphated lipo-oligosaccharide signals of Rhizobium meliloti elicit root nodule organogenesis in alfalfa. Nature *351*, 670–673.

Vicente, E.J., and Dean, D.R. (2017). Keeping the nitrogen-fixation dream alive. Proc. Natl. Acad. Sci. U.S.A. *114*, 3009–3011.

Vitousek, P.M., Aber, J.D., Howarth, R.W., Likens, G.E., Matson, P.A., Schindler, D.W., Schlesinger, W.H., and Tilman, D.G. (1997). Human Alteration of the Global Nitrogen Cycle: Sources and Consequences. Ecological Applications *7*, 737–750.

Westhoek, A., Field, E., Rehling, F., Mulley, G., Webb, I., Poole, P.S., and Turnbull, L.A. (2017). Policing the legume-Rhizobium symbiosis: a critical test of partner choice. Sci Rep *7*, 1–10.

White, C.E., Gavina, J.M.A., Morton, R., Britz-McKibbin, P., and Finan, T.M. (2012). Control of hydroxyproline catabolism in Sinorhizobium meliloti. Mol. Microbiol. *85*, 1133–1147.

Wu, T.T. (1966). A Model for Three-Point Analysis of Random General Transduction. Genetics *54*, 405–410.

Yang, J., Xie, X., Yang, M., Dixon, R., and Wang, Y.-P. (2017). Modular electron-transport chains from eukaryotic organelles function to support nitrogenase activity. Proc. Natl. Acad. Sci. U.S.A. *114*, E2460–E2465.

Yang, J., Xie, X., Xiang, N., Tian, Z.-X., Dixon, R., and Wang, Y.-P. (2018). Polyprotein strategy for stoichiometric assembly of nitrogen fixation components for synthetic biology. Proc. Natl. Acad. Sci. U.S.A. *115*, E8509–E8517.

Yuan, Z.-C., Zaheer, R., and Finan, T.M. (2006). Regulation and Properties of PstSCAB, a High-Affinity, High-Velocity Phosphate Transport System of Sinorhizobium meliloti. J. Bacteriol. *188*, 1089–1102.

Zhao, X.-Q., Hu, J.-F., and Yu, J. (2006). Comparative Analysis of Eubacterial DNA Polymerase III Alpha Subunits. Genomics, Proteomics & Bioinformatics *4*, 203–211.