Incompatible symbiotic interactions between *Sinorhizobium meliloti* strain Rm41 and the ecotypes of the host *Medicago truncatula*

PhD THESIS

by

Ting WANG

Supervised by Attila KERESZT, BRC, Szeged, HUNGARY

SUBMITTED TO THE
UNIVERSITY OF SZEGED, DOCTORAL SCHOOLOF BIOLOGY, HUNGARY

Abbreviations

CCaMK: calcium and calmodulin-dependent protein kinase

CNGC: cyclic nucleotide gated channels

CRE1: cytokinin response1

CSP: common symbiosis pathway

CTAB: cetyltrimethylammonium bromide

DMI: does not make infections

EPR3: exopolysaccharide receptor 3

EPS: exopolysaccharide

EPS I: succinoglycan

ETI: effector-triggered immunity

ERN: ethylene responsive factor required for nodulation

FST: flanking sequence tag

GFP: green fluorescent protein

IPTG: Isopropyl β-D-1-thiogalactopyranoside

IRLC: inverted repeat lacking clade

IPD3: interacting protein of DMI3

IT: infection thread

kbp: kilobase pair

KPS: K antigen, capsular polysaccharide

LHK1: Lotus histidine kinase1

LPS: lipopolysaccharide

LYK: LysM receptor kinase

LysM: lysin motif

LysM-RK: LysM receptor-like kinase

MIN: minimal medium

MNNG: N-methyl-N'-nitro-N-nitrosoguanidine

NCR: nodule-specific cysteine-rich

NF: Nod factor

NFP: Nod factor perception

NFR: Nod factor receptor

NFS: nitrogen fixation specificity

NIN: nodule inception

Nod: nodulation

NS1: nodulation specificity 1

NSP: nodulation signaling pathway

NUP: nucleoporin

OD: optical density

ORF: open reading frame

PBS: phosphate buffered saline

PCR: polymerase chain reaction

PIT: pre-infection thread

PM: plasma membrane

PVDF: polyvinylidene difluoride

R: resistance

rpm: revolutions per minute

SDS-PAGE: Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

SNF: symbiotic nitrogen fixation

SYMRK: symbiosis receptor-like kinase

T1SS: type I secretion system

T3SS: type III secretion system

TEMED: Tetramethylethylenediamine

TF: transcription factor

TIR-NBS-LRR: Toll-interleukin receptor/nucleotide-binding site/leucine-rich repeat

VLCFA: very-long-chain fatty acid

Table of contents

Abstract

Legumes develop a symbiosis with nitrogen-fixing soil bacteria called rhizobia. In this relationship, legumes provide an optimal environment and energy source for rhizobia. In return, rhizobia fix atmospheric nitrogen into ammonia for the plants. A significant characteristic of legume-rhizobia symbiosis is its high level of specificity. Each rhizobial species/strain can form successful symbiosis with only a few or even single legume species/genotype(s), and vice versa. Such specificity can exist at multiple stages during the nodule development and functioning and failure to suit the requirements results in different phenotypes. It can occur at early stages associated with bacterial infection and nodulation, leading to Nod^- phenotype. It also may occur at later stages of nodule development associated with nitrogen fixation, leading to Fix^- phenotype. Even when nitrogen-fixing symbiosis is formed successfully, the efficiency of nitrogen fixation differs between different legume-rhizobia pairings.

Our research group screened several ecotypes from *Medicago truncatula* using different *Sinorhizobium meliloti* and *Sinorhizobium medicae* strains as inoculants and identified incompatibility between *S. meliloti* strain Rm41 and *M. truncatula* ecotypes F83005 and Jemalong. To identify those existing or missing bacterial gene(s) that caused the incompatibilities, we created an ORFeome library carrying all predicted genes of *S. meliloti* strain 1021, a large insert size genomic library from *S. meliloti* strain FSM-MA, insertion and chemically induced mutant populations of strain Rm41 and inoculated transconjugant and mutant pools of strain Rm41 on F83005 and Jemalong. We identified 6 insertion mutants which restored the compatibility between F83005 and Rm41 and carried the insertion mutation in the same gene annotated as BN406_06091. The gene *BN406_06091* is located on the second symbiotic megaplasmid and encodes a protein of unknown function. It is surrounded by a number of genes coding for proteins involved in sugar and polysaccharide synthesis and modifications, implicating its role in polysaccharide production. We suppose that the *BN406_06091* gene containing region is responsible for the production of the strain-specific O-antigen of

lipopolysaccharide (LPS) and created mutants that might be affected in the LPS structure. Unfortunately, none of the created mutants could restore the compatibility with F83005. In future work, we will delete the whole strain-specific and alternatively the whole genus-specific LPS regions and check the phenotype of the mutants as well as that of the mutants carrying the *BN406_06091* gene. We will also explore the natural variations in *Sinorhizobium*-specific polysaccharide production related genes: we identified strains with some missing genes in the region with or without a *BN406_06091* homologue and also with different core structure. We will introduce either the whole or delimited regions with or without mutation in the *BN406_06091* gene and the phenotype of the strains and their derivatives will be investigated.

1. Introduction

1.1. Symbiotic nitrogen fixation

Nitrogen is an essential macronutrient for plants and is required for the synthesis of nucleic acids, amino acids and many other important metabolites. Although dinitrogen (N_2) accounts for a large proportion (around 78%) of Earth's atmosphere, its strong chemical stability makes it inaccessible for plants. Thus, nitrogen becomes one of the most limiting elements for plant growth.

Leguminous plants could grow in nitrogen poor soils because they enter into nitrogen-fixing symbiosis with a wide range of Gram-negative α- and β-Proteobacteria, referred to as rhizobia (Haag et al., 2013). In this relationship, both partners benefit from each other. Rhizobia invade the root of legumes and are released into newly formed specialized organs, the root (and occasionally, stem) nodules. Within root nodules, rhizobia differentiate into bacteroids and convert atmospheric dinitrogen (N_2) into ammonia, a form which is available for plants (Mergaert et al., 2006). In return, legume hosts provide proper environment and carbon sources to a large population of saprophytic and symbiotic rhizobia for their survival (White et al., 2007).

Symbiotic nitrogen fixation (SNF) enables legumes to grow in soils lacking fixed nitrogen and enrich soil nitrogen. Legumes inject approximately 40 to 60 million tons of fixed nitrogen into soils per annum (Smil, 1999). However, Canfield et al. (2010) estimated that 40 to 60 million tons of nitrogen was just a small part (less than half) compared to the nitrogen that is produced by chemical fertilizer. Unlike chemical nitrogen fertilizer, which is unfriendly to both economy and environment, SNF in legumes is considered as a sustainable nitrogen source for plant growth and agriculture (Sutton et al., 2011). Researchers have been working for decades on revealing the mechanisms of symbiosis and engineering nodulation-the miniature nitrogen fixation factory-to non-nodulating plants.

1.2. Development of a novel plant organ

The development of the symbiosis between legumes and rhizobia is a complex program including several developmental processes (bacterial infection and nodule organogenesis), multiple exchanges of signals and regulating the expression of numerous genes coordinately. Legumes perceive the lack of nitrogen in soils and secrete flavonoids into the rhizosphere (Peters et al., 1986; Redmond et al., 1986). Flavonoids are recognized as signals by rhizobia and interact with the NodD transcription factors (TFs), which induce the expression of nodulation (*nod*) genes (Göttfert et al. 1986; Kondorosi et al. 1989). The proteins encoded by *nod* genes are essential for the synthesis and export of lipochitooligosaccharides (LCOs) called nodulation (Nod) factors that have a core structure conserved in all different species of rhizobia (Lerouge et al., 1990). Several modifications on the side groups of Nod factors (NFs) are mediated by enzymes encoded by strain-specific *nod* genes and these modifications contribute to the specificity of the symbiosis between legumes and rhizobia (Long, 1996; Dénarié et al., 1996; Downie, 1998). NF is sufficient to induce nodule organogenesis even in the absence of rhizobia (Truchet et al., 1991). The recognition of NFs induces quick ion fluxes through the membrane of root hairs (Felle et al., 1998) then oscillations in calcium concentrations (calcium spiking) in the nuclei of epidermal cells. Calcium spiking may lead to expression changes of genes involved in the initiation of bacterial infection and nodule organogenesis (Charpentier and Oldroyd, 2013).

After the perception of NF, root hair polar growth is disrupted for a short time and sometimes accompanied with swelling of the root hair tip (Figure 1.1; Suzaki et al., 2019). Subsequently, root hair tip grows bending back upon itself, resulting in root hair curling formation (Esselinget al., 2003). The rhizobia that attach to the root hair are entrapped in the curling, forming an infection pocket in which rhizobia continue to divide and establish a microcolony (Geurts et al., 2005). From the infection pocket, plasma membrane invagination and new cell wall growing lead to the formation of a tubular structure called infection thread (IT). IT extends from root hair tip to root hair cell body, continues growth to cortical cells by degradation of cell wall (Van brussel et

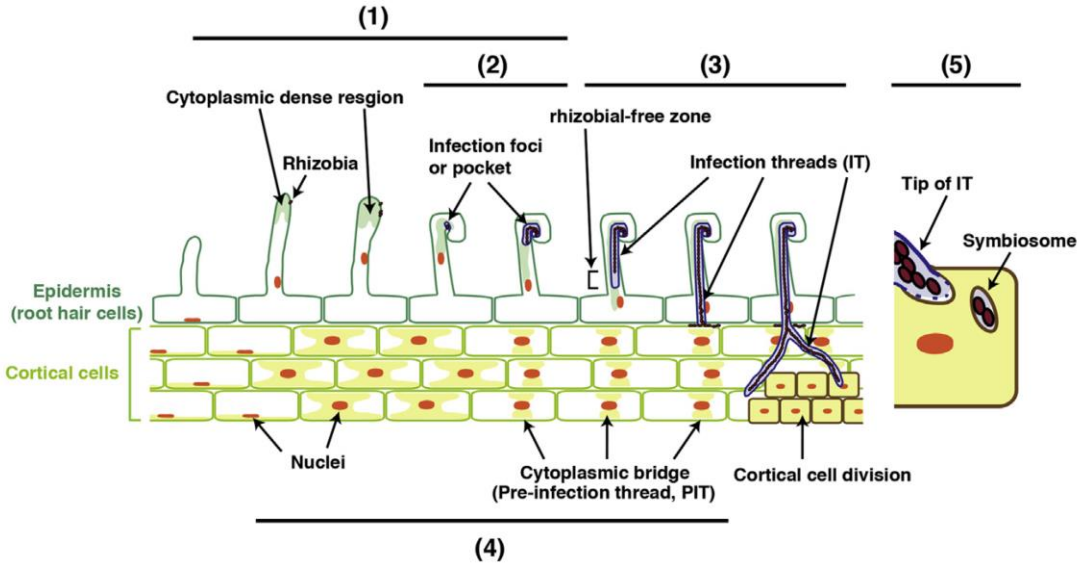

Figure 1.1 Rhizobial infection and invasion in root epidermis and cortical cells. (1) Rhizobial attachment to root hair and accumulated Nod factor induce initial responses involving root hair deformation and curling. **(2)** At the tightly curled root hair, cell wall degradation occurs and invagination of the IT membrane in the root hair cell is induced. **(3)** IT elongation occurs, which is accompanied by cytoplasmic streaming and nuclear movement. **(4)** During rhizobial infection [**(1)-(3)**], a cytoplasmic bridge or pre-infection thread (PIT) is formed to guide elongating ITs in the root cortical cells. **(5)** When an IT reaches the newly divided cortical cell, the IT membrane collapses and rhizobia are released into the cortical cell to form a specialized nitrogen-fixing organelle, the symbiosome. From Suzaki et al., 2019.

al., 1992; Van Spronsen et al., 1994) and develop through multiple cortical cells until reaching the dividing cortical cells called nodule primordium. Rhizobia multiplying inside IT are delivered to the primordium and released into the cytoplasm through an endocytotic process. In this process, rhizobia are surrounded by a plant-derived membrane, forming a specialized cellular organelle, referred to as symbiosome. In both ITs and symbiosomes, rhizobia are topologically in the extracellular space. Inside the plant cells, symbiosomes and the rhizobia in them divide until the cytoplasm of the plant cells are completely filled and differentiate into their nitrogen-fixing form called bacteroids (Brewin, 1991). The formation of ITs in root hair cells is the most common mode of rhizobial infection in legumes. However, the infection can also occur via cracks in the epidermis and via intercellular spaces between the cells of roots (Gage, 2004).

In parallel with IT development, root pericycle and cortical cells re-start cell division after NF recognition, leading to the formation of nodule primordium, which then

differentiate into nitrogen-fixing nodule (Crespi and Frugier, 2008). After division, nodule cells infected with rhizobia undergo endoreduplication cycles-genomic DNA replication without cell division-and differentiate into large, polyploid cells filled with thousands of symbiosomes (Kondorosi et al., 2000; Jones et al., 2007). Endoreduplication is essential for the formation of functional, mature nodules (Roy et al., 2020). The central region of nodule contains mostly infected cells and surrounded by cortex, endodermis, parenchyma and vascular bundles (Brewin, 1991). Mature root nodule provides an environment with rich nutrient and low free oxygen level that is optimal for nitrogen-fixation (Oldroyd et al., 2011). After four-six weeks of functioning, a senescence process starts in the oldest part of the nodule where both plant and bacterial cells disintegrate and their material is recycled (Dupont et al., 2012). Besides aging, nodule senescence can also be induced by stress or particular physiological conditions, for instance, drought stress, photosynthesis reduction, or nutrient transfer for seed production (Mergaert et al., 2020).

Leguminous root nodules can be classified into two types based on their mode of development: determinate and indeterminate nodules (Figure 1.2; Franssen et al., 1992; Larrainzar et al., 2020). Legumes such as *Phaseolus vulgaris*, *Lotus japonicus* or *Glycine max* form determinate nodules, which originate from the middle and outer

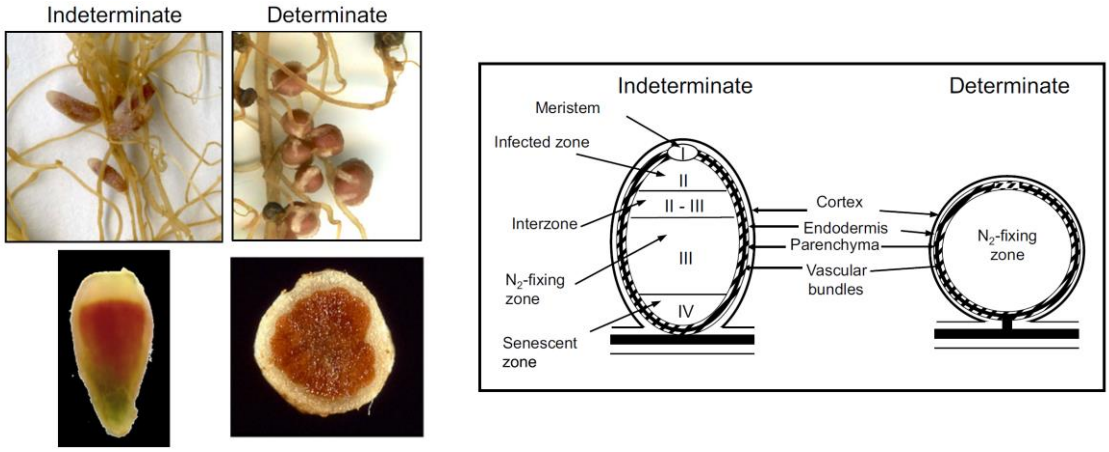

Figure 1.2 Major characteristics differentiating classical indeterminate and determinate nodules. Some features should not be considered universal because there are exceptions. From Larrainzar et al., 2020

cortex of roots. Determinate nodules have a transient meristem and develop into a spherical shape and their final size is the result of cell enlargement. In mature nodules, the central region contains only nitrogen-fixing zone which eventually becomes senescence zone (Hirsch, 1992). In contrast, legumes such as *Medicago truncatula*, *Pisum sativum* or *Trifolium* form indeterminate nodules, which originate from inner cortical cells. Indeterminate nodules have a persistent meristem and generate elongated nodules because new cells are added to the apex of this organ. In mature nodules of this type, five histological zone can be observed: the apical meristem (zone I), which is free from bacteria; the infection zone (zone II), where rhizobia are released from ITs; the interzone (zone II-III), where both bacterial and plant cells develop into their nitrogen-fixing form; the nitrogen-fixation zone (zone III), where bacteroids are able to fix nitrogen; and the senescence zone (zone IV), where bacteroids and nodule cells are degraded (Hirsch, 1992).

Bacteroids-depending on the host-are also different in morphology, cytology and cell fate. In the determinate and indeterminate nodules of *Lotus* and *Leucaena* species, respectively, symbiosomes contain two or more bacteroids, which retain a chromosome number of one or two and similar cell size as free-living bacteria. The membranes of these bacteroids are impermeable to propidium iodide and the cells can be regrown from nodules, i.e. they are not terminally differentiated (Mergaert et al., 2006). In contrast, symbiosomes from determinate and indeterminate nodules of *Aeschynomene* and legumes (such as *Medicago*, *Pisum*, *Trifolium* species) belonging to the Inverted Repeat Lacking Clade (IRLC) contain only one spherical or elongated bacteroids, respectively (Oono et al., 2010). Differentiated bacteroids possess a chromosome number of up to 24 because they undergo endoreduplication like their host symbiotic cells. They are permeable to propidium iodide and have lost their capacity to reproduce when isolated from nodules, thus they can be considered as terminally differentiated cells (Mergaert et al., 2006).

1.3. Molecular determinants of the partner selection in SNF

In natural soils, leguminous plants' roots are surrounded by mixed species/strains of rhizobia in the rhizosphere, but not all of them are able to enter symbiosis with host plants. A significant property of legume-rhizobia symbiosis is its high level of specificity. Each rhizobial species/strain can form successful symbiosis with only a few or even single legume species/genotype(s), and vice versa, however, there are exceptions such as *Sinorhizobium fredii* strain NGR234, which can establish symbioses with species from 112 genera (Pueppke and Broughton, 1999). An additional layer of specificity also exists, when certain genotypes of the partners (wild-type bacterial

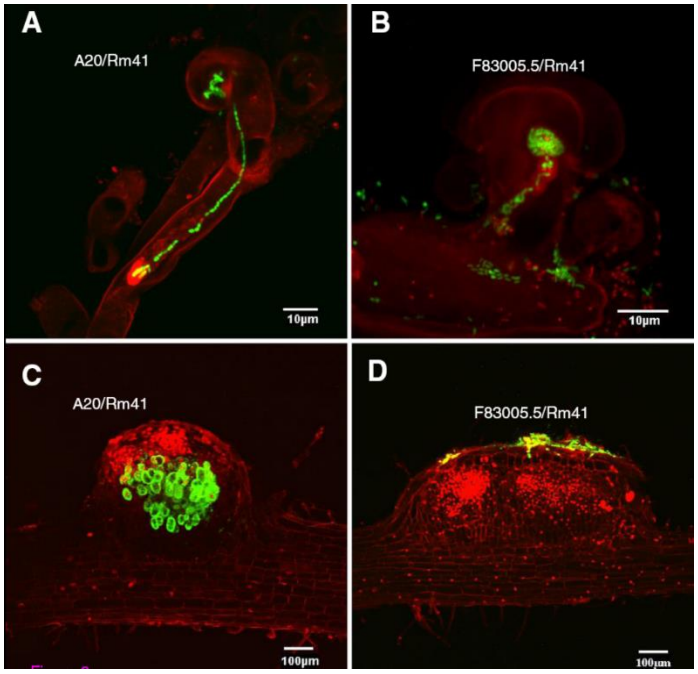

Figure 1.3 Fluorescence microscopy analyses of infection process of compatible and incompatible interactions between *M. truncatula* plants and Rm41. All images are composite images of GFP-expressing Rm41 cells (green) and root cells (red). **A**, Typical infection thread formed by compatible interaction between A20 and Rm41. The infection thread extends from the colonized, curled root hair to the base of the root hair cell. **B**, In the incompatible interaction between F83005.5 and Rm41, the bacteria can normally colonize the curled root hairs but typical infection threads cannot be detected. Occasionally, we can detect aborted, aberrant infection threads present on the F83005.5 roots. **C**, The nodule primordium on the A20 roots contained bacteria, while the nodule primordia on the F83005.5 roots **(D)** contained no bacteria despite frequent presence of bacterial colonies on the epidermal surface of the nodule primordia. From Liu et al., 2014.

strains and wild-type plant varieties often called ecotypes) that form effective symbiosis with other genotypes of the partner species, cannot establish nitrogen-fixing symbiosis with each other, i.e. they enter into incompatible interaction. Such specificity can take place at multiple stages during the nodule development and functioning and results in different phenotypes. It can occur at early stages associated with bacterial infection and nodulation, leading to Nod$^-$ or Inf$^-$ phenotype. For example, *Sinorhizobium meliloti* strain Rm41 could colonize the curled root hairs of *Medicago truncatula* ecotype F83005.5 but normal ITs were not detected, the nodule primordia contained no bacteria (Figure 1.3; Liu et al., 2014). Specificity also may occur at later stages of nodule development associated with nitrogen fixation, leading to Fix$^-$ phenotype. In the interaction of *Sinorhizobium meliloti* strain Rm41 and *Medicago truncatula* Jemalong, for instance, bacteria were able to infect nodule cells and to differentiate into elongated bacteroids, however, bacteroids could not persist and reduce nitrogen in the nodule, instead lysis and nodule senescence can be observed (Figure 1.4; Wang et al., 2017;

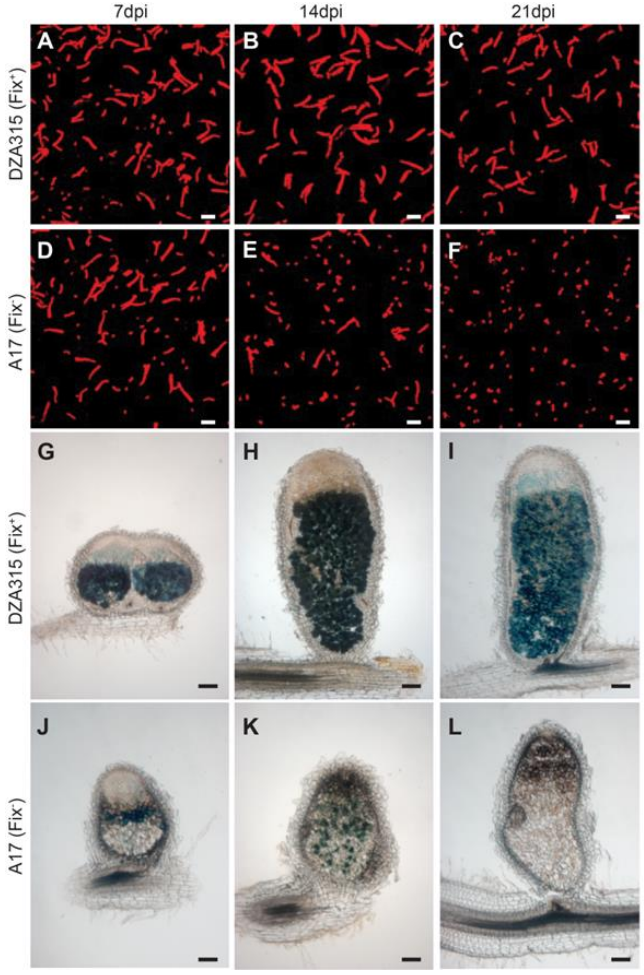

Figure 1.4 Differentiation and intracellular survival of the Rm41 bacteroids in the DZA315 (Fix$^+$) andA17 (Fix$^-$) nodules. (A–F) Confocal microscopy images of the bacterial populations isolated from the Fix$^+$ **(A–C)** and Fix$^-$ **(D–F)** nodules showing the gradual loss of elongated bacteroids and the accumulation of saprophytic bacteria in the Fix$^-$ nodules. **(G–L)** Expression of a GUS reporter gene driven by the *nifH* promoter in the Fix$^+$ **(G–I)** and Fix$^-$ **(J–L)** nodules showing that the *nifH* gene was expressed at the early developmental stages of the Fix$^-$ nodules, but its expression was abolished at later time points. [Scale bars, 5 μm **(A–F)** and 200 μm **(G–L)**.] From Wang et al., 2017.

Yang et al., 2017). Even when nitrogen-fixing symbiosis is formed successfully, the efficiency of nitrogen fixation differs between different legume-rhizobia pairings (Kazmierczak et al., 2017). A successful symbiosis requires signal recognitions and communications from both partners involving multiple molecular mechanisms, such as the interaction of flavonoids and NodD, the perception of NFs and rhizobial surface polysaccharides, the suppression of plant immunity or selection against incompatible bacteria mediated by nodule-specific cysteine-rich (NCR) peptides (Figure 1.5; Wang et al., 2018a). The understanding of genetic mechanisms underlying symbiotic specificity will allow researchers to manipulate genetic factors of legumes and/or rhizobia in order to enhance nitrogen fixation efficiency.

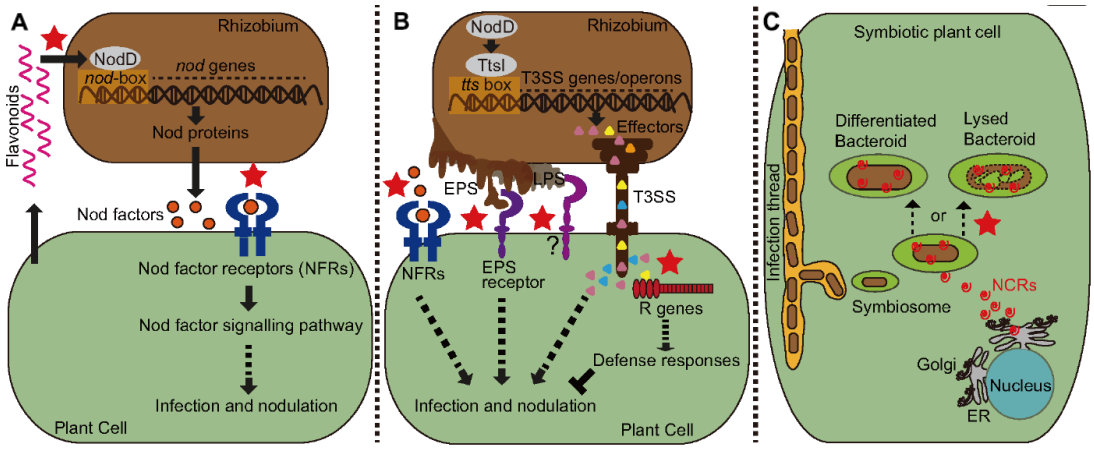

Figure 1.5 Symbiosis signaling and plant immunity involved in recognition specificity in the legume-rhizobial interactions (indicated by the red stars). A, The host secretes flavonoids to induce the expression of bacterial *nod* genes through the activation of NodD proteins. The enzymes encoded by the *nod* genes lead to the synthesis of NFs that are recognized by NFRs. Recognition specificity occurs both between flavonoids and NodDs and between NFs and NFRs. **B,** In addition to NF signaling, bacteria also produce extracellular polysaccharides (EPS, LPS, KPS) and type III effectors to facilitate their infection in compatible interactions, but these molecules may also induce immune responses causing resistance to infection in incompatible interactions. **C,** Certain legumes, such as *Medicago*, produce antimicrobial nodule-specific cysteine-rich (NCR) peptides to drive their bacterial partners to terminal differentiation that is required for nitrogen fixation. However, some rhizobial strains cannot survive the antibacterial activity of certain peptide isoforms, leading to formation of nodules defective in nitrogen fixation. From Wang et al., 2018a.

1.3.1. The first checkpoint: flavonoid - NodD recognition

The legume-rhizobia symbiosis begins with the interaction of flavonoids and NodD proteins that provides the first checkpoint of symbiosis. Legume roots secrete a complex mixture of flavonoids, but not all of them are involved in symbiosis. Some flavonoids act as an infection signal and stimulate the expression of *nod* genes. For example, luteolin is a primary infection signal in *Medicago sativa* and *Medicago truncatula*, while genistein and daidzein are infection signals in the *Glycine max* - *Bradyrhizobium* symbiosis (Liu and Murray, 2016). Some other flavonoids may also function as phytoalexins and contribute to symbiotic specificity. For example, *Sinorhizobium meliloti* shows tolerance to isoflavonoid medicarpin, which is bactericidal to *Bradyrhizobium japonicum* (Pankhurst and Biggs, 1980). *Bradyrhizobium japonicum* and *Sinorhizobium fredii* are sensitive to glyceollin, however, they become resistant to glyceollin in the presence of genistein and daidzein (Parniske et al., 1991).

Although there is no direct biochemical evidence for binding of flavonoids to NodD proteins, it has been shown that flavonoids could increase the binding of NodD1 to *nod* gene promoters in *Sinorhizobium meliloti* (Peck et al., 2006). NodD proteins produced by different rhizobial strains recognize different types of flavonoids (Györgypal et al., 1991). The important role of NodD in flavonoid recognition has been demonstrated by several experiments. For example, the expression of *nodD1* from broad host range *Rhizobium sp.* strain NGR234 in narrow host range *Rhizobium leguminosarum biovar trifolii* strain ANU834 makes the recipient strain to nodulate the non-legume *Parasponia* (Bender et al., 1988). Point mutations in *nodD* of *Rhizobium leguminosarum biovar trifolii* resulted in gaining the ability to activate *nod* gene expression in the presence of non-natural hosts and, thus, extending host range and allowing the nodulation of the non-legume *Parasponia* (McIver et al., 1989).

1.3.2. The NF signal

After the first checkpoint, the NFs are produced and they are the most important factors that determine which bacteria can nodulate a certain legume plant. NFs secreted by different rhizobia have the same backbone of 1,4-linked *N*-acetyl-glucosamine residues that carry a fatty acyl chain at the non-reducing end (Figure 1.6; Hirsch, 1992).

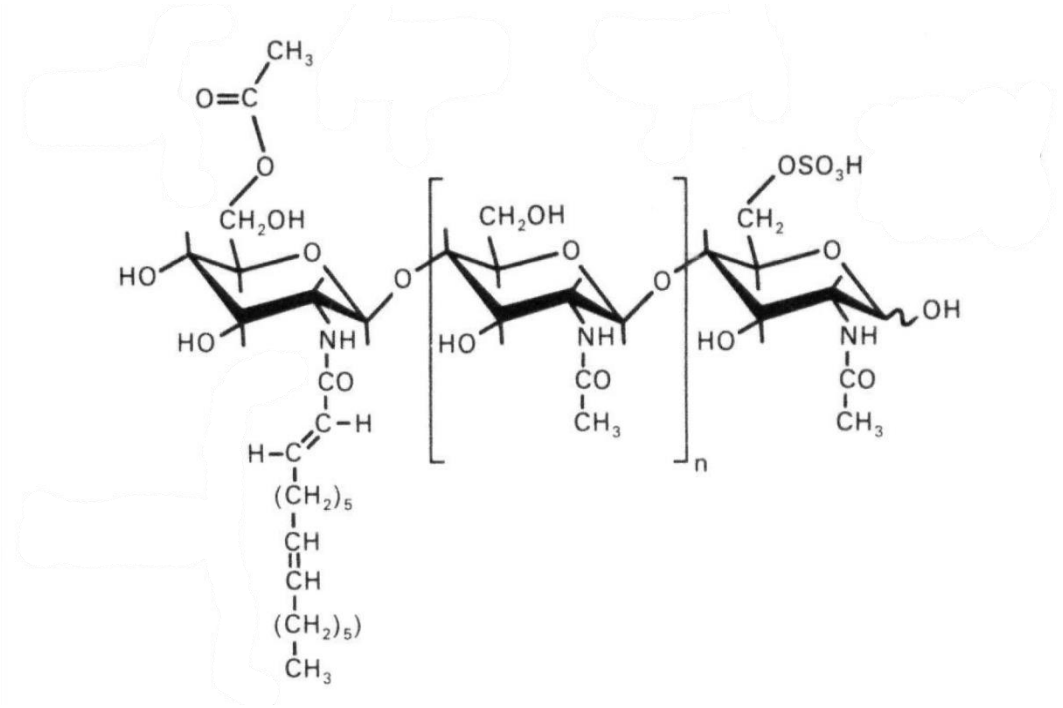

Figure 1.6 Structures of the *S. meliloti* NF. n refers to the number of glucosamine residues. Modified from Hirsch, 1992.

However, the length of the backbone, the size of the fatty acyl chain and the chemical decorations such as sulfation at the reducing and non-reducing ends are different (Long et al., 1996; Wang et al., 2012). The *nodABC* genes contribute to the synthesis of the backbone, while the other *nod* genes are involved in the strain specific modifications, which are crucial in determining the compatibility with legume species. For example, transfer of the plasmid carrying *Sinorhizobium meliloti* strain RCR2011 *nodPQ*, *nodFEG* and *nodH* genes (required for the sulfate modification of NFs) to *Rhizobium leguminosarum biovar viciae* 248 enables the recipient strain to nodulate alfalfa (Faucher et al., 1989). Introducing the *nodZ* gene coding for a fucosyl-transferase from *Bradyrhizobium japonicum* into *Rhizobium leguminosarum biovar viciae* results in the

formation of Fix$^-$ nodules on the roots of *Macroptilium atropurpureum*, *Glycine soja*, *Vigna unguiculata* and *Leucaena leucocephala* (López-Lara et al., 1996) that are normally not nodulated by this bacterium.

The perception of NFs secreted by rhizobia requires Nod factor receptors, LjNFR1 (NOD FACTOR RECEPTOR1)/MtLYK3 (LysM RECEPTOR KINASE3) and LjNFR5/MtNFP (NOD FACTOR PERCEPTION), which belong to the LysM receptor-like kinase (LysM-RK) family (Limpens et al., 2003; Madsen et al., 2003; Radutoiu et al., 2003; Arrighi et al., 2006). These receptors contain three extracellular lysin motif (LysM) receptor domains and an intracellular kinase domain. It was shown that LjNFR5 lacks intracellular kinase activity, which is essential for transmitting the signals to downstream symbiotic signaling pathways (Madsen et al., 2003). LjNFR1 and LjNFR5 form heterodimeric complexes at the plasma membrane (PM) of plant cells (Figure 1.7;

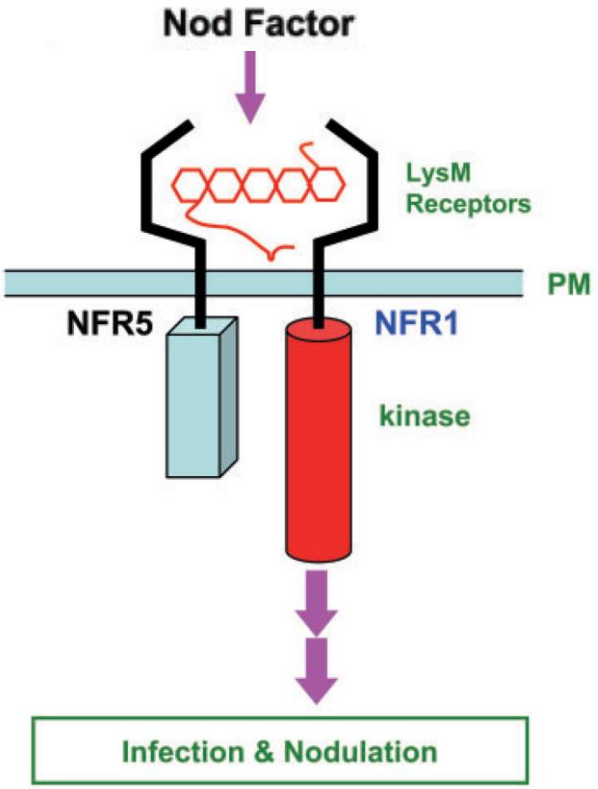

Figure 1.7 Recognition of Nod factors by NFR1 and NFR5 in *Lotus japonicus*. Extracellular LysM domains are thought to be responsible for binding Nod factors, and then transduce the symbiotic signals through the intracellular kinase domain of NFR1 to downstream signaling cascades, leading to rhizobial infection and nodulation. PM, plasma membrane. Modified from Kouchi et al., 2010

Kouchi et al., 2010). The extracellular LysM receptor domains are responsible for the recognition and binding of NFs (Broghammer et al., 2012; Moling et al., 2014). The role of NFRs in symbiosis specificity has been demonstrated by several experiments: Expression of *Nfr1* and *Nfr5* genes from *Lotus japonicus* in *Medicago truncatula* results in the initiation of symbiosis between transformants and *Mesorhizobium loti*, the natural symbiont of *Lotus japonicus*. In this symbiosis, the expression of early nodulin genes is activated, ITs are normally formed and Fix^- nodules are developed on roots (Radutoiu et al., 2007). LjNFR1 and LjNFR5 are demonstrated to be able to bind NFs directly with high affinity (Broghammer et al., 2012). Sørensen et al. (2014) studied the interactions between NFs and three LysM domains of LjNFR5, and revealed that the LysM2 domain binds to NFs, thus causing the conformational change in the LysM2 domain. In the LysM1 domain of LjNFR1 and MtLYK3, two regions are proved to be essential for NF signaling specificity (Bozsoki et al., 2020). Following the perception of NFs, complexes of LysM-RKs interact with a plasma membrane leucine-rich repeat receptor kinase LjSYMRK (SYMBIOSIS RECEPTOR-LIKE KINASE)/MtDMI2 (DOES NOT MAKE INFECTIONS2) (Endre et al., 2002; Stracke et al., 2002; Limpens et al., 2005; Antolín-Llovera et al., 2014), thereby activating the common symbiosis pathway (CSP) and leading to oscillations in calcium concentrations (calcium spiking) in the nuclei of root hair cells (Figure 1.8; Roy et al., 2020). Several nuclear envelope proteins are required for generating calcium spiking, such as cation (most probably potassium) channels LjCASTOR and LjPOLLUX/MtDMI1 (Ané et al., 2004; Imaizumi-Anraku et al., 2005; Charpentier et at., 2008); nucleoporin subunits LjNUP85 (NUCLEOPORIN85) and LjNUP133 (Kanamori et al., 2006; Saito et al., 2007); and calcium channels MtCNGC (CYCLIC NUCLEOTIDE GATED CHANNELS) a/b/c (Charpentier et al., 2016). Intriguingly, LjCASTOR and LjPOLLUX co-operate to perform their functions although they have similar structures, however, MtDMI1 seems to have the ability alone to accomplish the role, which needs the above two proteins. Calcium oscillations are decoded by a calcium and calmodulin dependent protein kinase CCaMK/MtDMI3 (Levy et al., 2004; Tirichine et al., 2006), which phosphorylates a

TF LjCYCLOPS/MtIPD3 (INTERACTING PROTEIN OF DMI3) (Messinese et al., 2007; Yano et al., 2008; Singh et al., 2014).

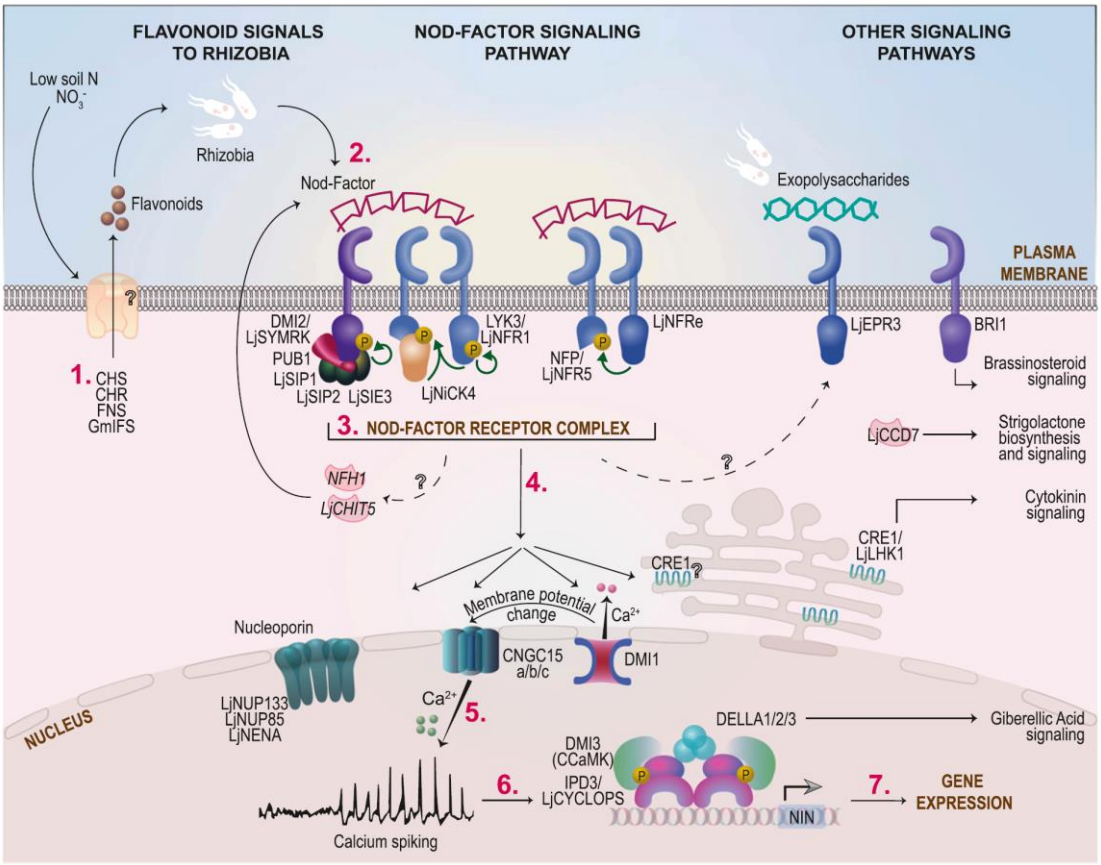

Figure 1.8 Proteins and processes involved in early signaling during nodulation. (1) (Iso)flavonoids produced under low soil N **(2)** trigger the production of bacterial Nod factors **(3)** that, together with other signals, are perceived by receptors at the plasma membrane of epidermal cells. **(4–6)** This triggers biochemical and physiological responses **(7)** that lead to changes in nuclear gene expression. See main text for key to gene/protein names and further explanation of early signaling pathways and components. *Medicago. truncatula* protein names are provided unless otherwise specified. From Roy et al., 2020.

Several TFs downstream of CSP have been identified, such as NIN (NODULE INCEPTION), ERN1/ERN2 (ETHYLENE RESPONSIVE FACTOR REQUIRED FOR NODULATION1 and 2), MtNSP1 (NODULATION SIGNALLING PATHWAY1), MtNSP2, etc. These TFs regulate NF-induced gene expressions. NIN and ERN1, induced by the CCaMK/CYCLOPS complex, are essential for both IT formation and nodule organogenesis (Yano et al., 2008; Schauser et al., 1999; Middleton et al., 2007; Marsh et al., 2007; Cerri et al., 2017; Kawaharada et al., 2017a). The expression of NIN can also be induced by the heterodimeric complex consisting of NSP1 and NSP2 that

activates NF-responsive genes (Hirsch et al., 2009). NSP2 acts downstream of cytokinin perception by LjLHK1/MtCRE1 and is required for nodulation (Kaló et al., 2005; Tirichine et al., 2007). Cytokinin receptor LjLHK1 (LOTUS HISTIDINE KINASE1)/MtCRE1 (CYTOKININ RESPONSE1) is also positioned downstream of LjCYCLOPS/MtIPD. Cytokinin signaling regulates cortical cell divisions, suppresses polar auxin transport, and is required for the activation of downstream TFs MtNSP2, MtERN1, NIN and leads to nodule organogenesis (Gonzalez-Rizzo et al., 2006; Murray et al., 2007; Tirichine et al., 2007; Plet et al., 2011).

1.3.3. The perception of rhizobial surface polysaccharides

In addition to NFs, rhizobial surface polysaccharides, including lipopolysaccharide (LPS), capsular polysaccharide (KPS or K antigen) and exopolysaccharide (EPS) are also required during the establishment of legume-rhizobia symbiosis. As components of or attached to the bacterial cell wall, surface polysaccharides play important roles in root attachment, signaling and the suppression of plant defense (Downie, 2010). Succinoglycan (EPS I), the major EPS produced by *Sinorhizobium meliloti*, is a polymer of a repeating octasaccharide unit that consists of seven glucoses and one galactose with succinyl, acetyl, and pyruvyl modifications (Reinhold et al., 1994). It is reported that the *Sinorhizobium meliloti* mutants (Exo^-), losing the ability to produce succinoglycan, form ineffective nodules on *Medicago sativa* and these nodules contain no ITs or bacteroids (Finan et al., 1985; Leigh et al., 1985). Succinoglycan is required for the initiation and elongation of ITs during nodulation on *Medicago sativa* (Cheng and Walker et al., 1998), and increased production of succinoglycan enhances symbiotic capacity with *Medicago truncatula* (Jones, 2012). LPS is composed of three moieties: the lipid A membrane anchor, the core oligosaccharide and the O-chain (O-antigen) polysaccharide. The rhizobia lipid A moiety is most often modified with a very-long-chain fatty acid (VLCFA) that has sufficient length to span the entire lipid bilayer. The O-antigen polysaccharide moiety contains a strain-specific oligosaccharide unit but in contrast to enteric bacteria, it is a monomer and does not form repeats (Kannenberg and

Carlson, 2001; Gibson et al., 2008). LPS is involved in the establishment of symbiosis evidenced by certain mutants defective in LPS synthesis, mainly in interactions with legumes forming determinate nodules, where it has similar functions as EPS in *Medicago* (Carlson et al., 1987; Lagares et al., 1992; Niehaus et al., 1998).

The structure of surface polysaccharides produced by different strains varies widely and there is a correlation between the structure and infection specificity. That is one of the possible reasons why (engineered) strains producing the correct NF required by a non-host legume cannot establish effective symbiosis with this host although nodule formation takes place (for example, Geddes et al., 2021). Succinoglycan produced by *Sinorhizobium meliloti* NGR185, which is compatible with *Medicago truncatula* A17, are almost fully succinylated. In contrast, succinoglycan produced by *Sinorhizobium meliloti* Rm41, which is incompatible with A17, contain many nonsuccinylated subunits (Simsek et al., 2007). Transfer of *exo* gene region from A17-compatible *Sinorhizobium meliloti* Rm1021 to Rm41 was reported to enable the recipient strain to produce similar succinoglycan oligosaccharides to donor strain and to form Fix^+ nodules on A17 roots (Simsek et al., 2007), however, we could not repeat this result.

Recently, the *Epr3* gene coding for a LysM-type receptor-like kinase was identified in *Lotus japonicus*. EPR3 was shown to bind EPS of compatible symbiont directly and to be involved in the checking of bacterial entry. The EPR3 protein is required to recognize and to exclude those bacteria that produce EPS in a truncated form caused by a mutation in the *exoU* gene, while bacteria with normal or no EPS can infect nodule tissues. Interestingly, the expression of *Epr3* is dependent on the perception of NFs, suggesting that the infection process of a compatible symbiosis is regulated by perception of two signals successively, namely, NFs and EPS produced by rhizobia (Kawaharada et al., 2015, 2017b). The LysM-receptor-like kinase MtLYK10 from *Medicago truncatula* was identified as the ortholog of LjEPR3. The symbiotic infection phenotypes between *Sinorhizobium meliloti exo* mutants and *Mtlyk10* mutant plants suggested that MtLYK10 plays an important role in symbiotic process. However, unlike LjEPR3, MtLYK10 and succinoglycan participate independently in symbiosis (Maillet

et al., 2020).

1.3.4. Effector-triggered host immunity as a specificity factor

Effector-triggered immunity (ETI) is the second layer of plant innate immunity (Jones and Dangl, 2006). Bacteria use a secretion system to deliver effectors into the cytoplasm of target host cells. The host plant resistance (R) proteins recognize these effectors and activate ETI (Cui et al., 2015). In legume-rhizobia interaction, ETI is another determinant of symbiosis specificity.

Many rhizobial strains use type III secretion system (T3SS) to translocate effectors into host cells. T3SS and T3 effectors are regulated by flavonoids and NodD, which induces the expression of *ttsI*. *TtsI* gene codes for a transcriptional regulator, which activate the genes involved in T3SS and the synthesis of T3 effectors by binding to highly conserved promoter elements termed *tts* boxes. It has been shown that rhizobial T3SS and T3 effectors play important roles in the control of host range. Soybean cultivar Hardee is incompatible with wild-type *Bradyrhizobium japonicum* USDA122, while the *ttsI* knock out mutant of USDA122 which is deficient in T3SS induced Fix^+ nodules on Hardee, suggesting T3SS mediates incompatibility between these two partners (Tsukui et al., 2013). The inactivation of *innB*, encoding a T3 effector, abolishes the incompatibility between *Vigna radiata cv.* KSP1 and *Bradyrhizobium elkanii* USDA61 (Nguyen et al., 2017), implying that InnB negatively regulates nodulation of KSP1. *Vigna mungo* (L.) Hepper cv. PI173934 inoculated with wild-type USDA61 formed more nodules than with *innB*-deficient mutant, suggesting that InnB promotes nodulation of PI173934 (Nguyen et al., 2018).

Several dominant genes involved in strain specific restriction of nodulation have been identified in soybeans, such as *Rj2, Rfg1* and *Rj4* (Yang et al. 2010; Tang et al., 2016; Fan et al., 2017). *Rj2* and *Rfg1* are allelic genes that encode classical Toll-interleukin receptor/nucleotide-binding site/leucine-rich repeat (TIR-NBS-LRR) R proteins. *Rj2* restricts nodulation by *Bradyrhizobium japonicum* strain USDA122 and *Rfg1* restricts nodulation by *Sinorhizobium fredii* strains USDA257, USDA193 and

USDA205 (Yang et al. 2010; Fan et al., 2017). Interestingly, Rj4 encodes a thaumatin-like protein rather than a R protein and restricts nodulation by *Bradyrhizobium elkanii* strain USDA61 (Tang et al., 2016). These genes function in nodulation specificity depending on T3SS. However, some sequenced rhizobial strains showed the lack of genes encoding T3SS. Liu et al. (2014) identified a naturally occurring dominant gene, named *MtNS1* (*Medicago truncatula* nodulation specificity 1) that restricts nodulation by *Sinorhizobium meliloti* strain Rm41. Unfortunately, *MtNS1* is not a classical R gene and the genetic mechanism of this phenotype has been unknown.

1.3.5. Nitrogen fixation specificity mediated by nodule-specific cysteine-rich (NCR) peptides

In addition to early recognition, bacterial infection and nodulation, symbiotic specificity also occurs at later stage associated with nitrogen fixation, leading to the formation of Fix^- nodules. Recently, Wang et al. (2017) and Yang et al. (2017) isolated two genes from *Medicago truncatula*, named *NFS1* (nitrogen fixation specificity 1) and *NFS2*, which both encode NCR peptides. NCR peptides were shown to positively regulate symbiosis, mediate bacterial differentiation into nitrogen-fixing bacteroids and maintain bacterial survival (Van de Velde et al., 2010; Horváth et al., 2015; Kim et al., 2015). However, *NFS1* and *NFS2* were demonstrated to be negative effectors on symbiotic persistence. They provoke bacterial lysis shortly after differentiation into bacteroids and early nodule senescence in the interaction between *Medicago truncatula* Jemalong A17 and *Sinorhizobium meliloti* strain Rm41. *NFS1* and *NFS2* were also demonstrated to negatively regulate the nitrogen fixation by *Sinorhizobium meliloti* strain A145 (Wang et al., 2018b). Therefore, some NCR peptides are host determinants of nitrogen fixation specificity.

1.4. Preliminary results grounding the current work

One of the main interest of our research group is to identify those bacterial and plant

genes, gene variants (alleles) that cause incompatible interactions between the ecotypes of the model legume *Medicago truncatula* and different strains of *Sinorhizobium meliloti* and *S. medicae*. In these incompatible interactions, the otherwise effective organisms-i.e. the bacteria and the plants establish nitrogen-fixing symbiosis with other *M. truncatula* ecotypes and *Sinorhizobium* strains-fail to develop any or efficient nodules. The dependence of the success/failure of these partnerships on the combination of certain gene varieties/alleles in the two organisms resembles the gene-for-gene interactions described for the relations of plants with pathogenic bacteria and fungi. The investigations of these incompatibilities might shed light on the fine-tuning of the symbiotic interaction, for example, how the symbionts select their partners from the available potential interactors. The existence of incompatible interactions necessitates practical considerations too because they might prevent to achieve the yield potential of a legume crop: If an elite variety is incompatible with rhizobia present in the field or in the commercial inoculant applied, its growth will be not supported by symbiotic nitrogen fixation.

It was earlier shown by our group and our collaborating partner at the University of Kentucky (Lexington, Kentucky, USA) that *Sinorhizobium meliloti* strain Rm41 forms incompatible interactions with *M. truncatula* ecotypes Jemalong and F83005. Strain Rm41 induces the early responses to NFs in F83005 such as root hair curling and cortical cell divisions (Liu et al., 2014) but IT and nodule meristem formation after these initial steps were blocked. Map-based cloning of the plant determinant revealed that alleles of two neighboring genes coding for receptor-like kinases are responsible for the incompatibility (unpublished results, Hongyan Zhu, University of Kentucky). In its interaction with *M. truncatula* cv. Jemalong, Rm41 bacteria are able to invade and colonize A17 nodule cells, but undergo lysis shortly after differentiation into elongated bacteroids (Wang et al., 2017). It was shown that dominant alleles of two genes coding for NCR peptides in cultivar Jemalong are required for the elimination of strain Rm41 (Yang et al., 2017; Wang et al., 2017) and A145 (Wang et al., 2018b).

2. Objectives

The collaboration between our research group and the team lead by Prof. Hongyan Zhu at the University of Kentucky revealed what are the genetic determinants of the incompatibility of *Sinorhizobium meliloti* strain Rm41 with *Medicago truncatula* ecotypes Jemalong and F83005 and are described in the previous chapter. On the bacterial side, there are two possible causes of the incompatibility: 1) there is a biological activity encoded in the genome, for example, for the production of a surface polysaccharides or a protein recognized as enemy or an enzyme/protein interfering with the process of nodule/bacteroid development or functioning or 2) a gene or genes coding for a biological activity is/are missing from the genome that prevent the normal development and functioning of the symbiotic nodule.

To facilitate the identification of the detrimental or missing functions, we aimed to create tools and biological materials that make possible to achieve these goals. For this purpose, the following tasks were set up:

*to create an ORFeome library carrying all predicted genes of *S. meliloti* strain 1021 compatible with the *M. truncatula* ecotypes Jemalong and F83005 behind a nodule-specific promoter in a broad host-range vector;

*to create a genomic library from *S. meliloti* strain FSM-MA, which forms effective symbiosis with most *M. truncatula* ecotypes investigated;

*to create insertion and chemically induced mutant populations of *S. meliloti* strain Rm41

*to create transconjugant populations of *S. meliloti* strain Rm41 by introducing the genomic libraries

*to test whether mutations or introduced genes can restore the symbiotic effectivity of *S. meliloti* strain Rm41 with *M. truncatula* ecotypes Jemalong and F83005

*to characterize the identified bacterial genes

3. Materials and Methods

3.1. Bacterial strains and growth conditions

Escherichia coli strains MDS42 ΔrecA Blue, HB101 and their derivatives were grown at 37 °C in LB medium (10 g/l tryptone; 5 g/l yeast extract; 5 g/l NaCl) supplemented with appropriate antibiotics and 100 µM IPTG and 20 µg/ml X-gal, when the detection of β-galactosidase activity was needed. *Sinorhizobium meliloti* strains were grown at 30 °C in TA (10 g/l tryptone; 1 g/l yeast extract; 5 g/l NaCl; 1 mM $MgSO_4$; 1 mM $CaCl_2$) or in minimal (MIN) medium (0.5 g/l K_2HPO_4; 0.2 g/l $MgSO_4 \cdot 7H_2O$; 0.1 g/l NaCl; 10 g/l Mannitol; 1 g/l $(NH_4)_2SO_4$; 1.34 g/l Malic acid; 1mg/l Biotin; Microelements) supplemented with appropriate antibiotics (Table 3.1).

Table 3.1 Antibiotic used in this study

Name of antibiotics	**Concentrations used in this study**	
	***E. coli* strains**	***Sinorhizobium meliloti* strains**
Streptomycin	100 µg/ml	100 µg/ml
Kanamycin	100 µg/ml	400 µg/ml
Spectinomycin	100 µg/ml	100 µg/ml
Gentamicin	25 µg/ml	200 µg/ml
Tetracycline	15 µg/ml	15 µg/ml
Ampicillin	100 µg/ml	100 µg/ml
Rifampicin	100 µg/ml	100 µg/ml

3.2. Plasmids

Plasmids used in this study are summarized in Table 3.2.

Mariner based transposons pSAM_R1 (Perry and Yost, 2014), pTPM1L71K and pTPM1L71KMCS carrying recognition sites of restriction endonucleases (*Sbf*I, *Nhe*I, *Bsr*GI, *Bst*1107I, *Acl*I, *Dra*I, *Pme*I, *Pml*I, *Hpa*I, *AsiS*I, *Not*I) cleaving the *Sinorhizobium* genomes infrequently (Attila Kereszt, unpublished development) were used for the

random insertion mutagenesis of the *S. meliloti* strains Rm41 and FSM-MA, respectively. Plasmids carrying the transposons and the transposase gene can be mobilized from *E. coli* into rhizobia, where they cannot replicate, thus, selection for the transposon bored kanamycin resistance allowed the isolation of mutant strains.

Table 3.2 Plasmids used in this study

Plasmid	Relevant characteristics	Reference or source
pSAM_RI	Kanamycin, Ampicillin; RP4-oriT, R6K γ ori R; transposon and transposase	Perry and Yost, 2014
pTPM1L71K	Kanamycin; pM1ori; transposon and transposase	Attila Kereszt, unpublished development
pTPM1L71KMCS	Kanamycin; pM1ori; transposon and transposase; a polylinker with RE recognition sites rare in rhizobia	Attila Kereszt, unpublished development
pK18mobsacB	Kanamycin; RP4-oriT, pMB1-oriV; *sacB*; *lacZ*	Schäfer et al., 1994
pK19mob	Kanamycin; RP4-oriT, pMB1-oriV; *lacZα*	Schäfer et al., 1994
pRK2013	Kanamycin; mobilization helper plasmid	Figurski and Helinski, 1979
pHC60	Tetracycline; GFP	Cheng and Walker, 1998
pBlueScriptII KS+	Ampicillin; *lacZ*; cloning vector	Alting-Mess and Short, 1989
pPR97	Kanamycin; binary vector	Szabados et al., 1995
pGREEN	Kanamycin; *lacZ*	Hellens et al., 2000
pPRlacZ	Kanamycin; *lacZ*	Attila Kereszt, unpublished development
pPZPgmSK2	Gentamycin; *lacZ*	Attila Kereszt, unpublished development
pSP-KS2	Spectinomycin; *lacZ*; cloning vector	Attila Kereszt, unpublished development
pSParaC-mCherry	Spectinomycin; mCherry	Attila Kereszt, unpublished development
pSPxcm3G_HA	Spectinomycin; HA tag	Attila Kereszt, unpublished development

Plasmid pK18mobsacB was used as a backbone to clone the two flanking fragments for the creation of deletion and resistance cassette insertion mutants in strain Rm41 as described by Schäfer et al. (1994). The plasmid can be transferred into rhizobia when

the transfer functions are provided on a helper plasmid pRK2013 (Figurski and Helinski, 1979) or are integrated into the chromosome of rhizobia and cannot replicate in rhizobia. It also harbors the *sacB* gene coding for the levansucrase protein, which could convert sucrose into a toxic product for bacteria. Thus, after integrating the construct into the rhizobial genome through one of the homologous sequences, it can be selected for the loss of the plasmid through a second homologous recombination that can occur through the second homologous sequence, thus, leaving the mutation in the genome-with a theoretical chance of 50%.

For plasmid integration mutagenesis, an internal fragment (without the translational start and stop codons) of the target gene was cloned into the vector pK19mob (Schäfer et al., 1994) and then, the construct was mobilized into Rm41, where it could not replicate, thus, the strains with integrated plasmid could be selected based on their kanamycin resistance.

For the protein localization studies, gene-fusion constructs were cloned into the broad host-range vector pHC60 (Cheng and Walker, 1998), which is stably maintained in rhizobial cells even without antibiotic selection.

3.3. Triparental mating

Triparental mating was performed to transfer plasmids from *E. coli* strains into *S. meliloti* strains. The integrative vectors, transposon sources and library clones were transformed into auxotrophic *E. coli* strain HB101 which could not grow on MIN medium. Recipient *S. meliloti* strains were cultured for 48 hours in liquid TA supplemented with antibiotics if necessary. Donor strains carrying the different constructs, transposon sources and library clones as well as helper strain HB101(pRK2013) were cultured for overnight in liquid LB medium supplemented with appropriate antibiotics. The mixture of 2 ml of recipient, 1 ml of donor and 1 ml of helper strains was centrifuged at 4000 rpm for 2 minutes and washed once with sterile distilled water to remove residual antibiotics. The mixed bacterial pellets were then re-suspended in 100 µl of TA medium and dropped onto TA agar plate. After incubation at

30 °C for overnight, bacterial cells were re-suspended in 1ml of sterile distilled water (10^0 dilution) and ten-fold serial dilutions (10^{-1}, 10^{-2}) were prepared. 100 µl of all three dilutions were plated on MIN plates supplemented with appropriate antibiotics and incubated at 30 °C for 3-4 days until single colonies appeared.

3.4. Isolation and manipulation of DNA

3.4.1. Genomic DNA isolation

For genomic DNA isolation, rhizobia were cultured at 30 °C for 48 hours in liquid TA supplemented with appropriate antibiotics. Bacterial cells were collected from 2 ml of the culture by centrifugation at 10000 rpm for 2 minutes and re-suspended in 550 µl of STET buffer (8 % sucrose, 0.5 % TritonX-100, 50 mM TRIS, 50 mM EDTA pH=8.0) containing 10 µg/ml RNase and 0.2 mg/ml lysozyme. Cell suspensions were vortexed and incubated at 30 °C in a rotator for 1 hour. 30 µl of 10 % SDS and 10 µl of proteinase K solution (20 mg/ml) were added and suspensions were incubated at 37 °C for 1 hour. 150 µl of 5 M NaCl, 80 µl of 700 mM NaCl/10 % CTAB solution (pre-warmed to 65 °C) were added, followed by the incubation at 65 °C for 40 minutes. After adding 400 µl of chloroform and isoamyl alcohol (24:1) mixture, the phases were mixed thoroughly and centrifuged at maximum speed for 10 minutes. Supernatant was transferred into a new microfuge tube and 0.54 volumes of isopropanol were added with mixing. The DNA precipitate was collected by centrifugation at maximum speed for 10 minutes. Then DNA pellets were washed with 70% ethanol for three times, dried under air and dissolved in 100 µl of sterile distilled water at 4 °C for overnight.

3.4.2. Plasmid DNA isolation

Single colonies were inoculated in 3-4 ml of LB supplemented with appropriate antibiotics and were cultured at 37 °C for overnight. The plasmids were isolated using EZ-10 Spin Column Plasmid DNA MiniPreps Kit (BIO BASIC INC.). Bacterial cells were collected from 3 ml of culture by centrifugation at 10000 rpm for 2 minutes and

re-suspended in 100 µl of Solution I containing 0.2 mg/ml RNase. Following incubation at room temperature for 1 minute, 200 µl of Solution II was added into the suspension, mixed gently by inverting the tube 4-6 times and kept for 1 minute. 350 µl of Solution III was added, mixed gently and the tube was incubated at room temperature for 1 minute. After the centrifugation at 12000 rpm for 5 minutes, the supernatant was transferred in EZ-10 column and centrifuged at 10000 rpm for 2 minutes. The flow-through in the tube were discarded. 750 µl of Wash Solution was added to the column to wash plasmid DNA twice and centrifuged at 10000 rpm for an additional minute to remove any residual Wash Solution. The column was then placed in a new 1.5 ml microfuge tube and 50 µl of Elution Buffer was added to the center of the column. After 2 minutes of incubation at room temperature, plasmid DNA was collected by the centrifugation at 12000 rpm for 2 minutes and stored at -20 °C.

3.4.3. Agarose gel electrophoresis, DNA purification from gels, PCR and enzymatic reactions, and DNA precipitation

Agarose gel electrophoresis was performed to check the results of PCR reactions or DNA digestions. 10 µl of DNA sample was mixed with 2 µl of 6X DNA Loading Dye (Thermo Fisher Scientific) and loaded into agarose gel (10 mM $Na_2B_4O_7$, 1% agarose, 0.5 µg/ml ethidium bromide). The electrophoresis was run in 10 mM $Na_2B_4O_7$ running buffer at 90 V for 25 minutes using PSE300/300 Power Supply (BIOCENTER). The gel was then exposed to UV light for imaging.

For DNA purification from agarose gels, EZ-10 Spin Column DNA Gel Extraction Kit (BIO BASIC INC.) was used. All of the samples were loaded into agarose gel and target fragment was excised. 400 µl of Binding Buffer II was added to each 100 mg of gel, followed by incubation at 55 °C for 10 minutes. The above mixture was then transferred to EZ-10 column and kept for 2 minutes. After the centrifugation at 10000 rpm for 2 minutes, the flow-through in the tube were discarded. 750 µl of Wash Solution was added to the column to wash DNA twice and centrifuged at 10000 rpm for an additional minute to remove any residual Wash Solution. The column was then placed

in a new 1.5 ml microfuge tube and 30 µl of Elution Buffer was added to the center of the column. After 2 minutes of incubation at room temperature, DNA was collected by the centrifugation at 12000 rpm for 2 minutes and stored at -20 °C.

For DNA purification from PCR or enzymatic reactions, DNA Clean & Concentrator^{TM}-5 Kit (ZYMO RESEARCH) was used. 1 volume of PCR or enzymatic reaction was mixed with 5 volumes of DNA Binding Buffer. The above mixture was transferred to the Zymo-Spin^{TM} Column. After the centrifugation at 10000 rpm for 30 seconds, the flow-through in the tube were discarded. 200 µl of DNA Wash Buffer was added to the column to wash DNA twice. The column was then placed in a new 1.5 ml microfuge tube and 20 µl of DNA Elution Buffer was added to the center of the column. After 1 minute of incubation at room temperature, DNA was collected by the centrifugation at 12000 rpm for 1 minute and stored at -20 °C.

For DNA precipitation, 1 volume of DNA solution was mixed with 0.1 volume of 3 M ammonium-acetate and 2 volumes of EtOH. After the incubation at room temperature for 15 minutes, the DNA precipitate was collected by centrifugation at maximum speed for 10 minutes. Then DNA pellet was washed with 70% ethanol twice, dried under air and dissolved in 50 µl of sterile distilled water.

3.4.4. DNA cloning methods

DNA digestions were performed using FastDigest enzymes (Thermo Fisher Scientific) in this study. Each reaction (20 µl) contained 1X FastDigest Buffer, 1 µg of DNA and 1 µl of FastDigest enzyme. The mixture was incubated at 37 °C for 30 minutes.

FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific) could catalyze the release of 5'- and 3'-phosphate groups from DNA and was used to prevent digested plasmids from self-ligation. FastAP was active in FastDigest Buffer and could be used together with FastDigest enzymes. 1 µl of FastAP was added into digestion reaction (20 µl), followed by the incubation at 37°C for 30 minutes.

Klenow Fragment (Thermo Fisher Scientific) was used for blunting the sticky end of DNA from enzymatic reaction. Each reaction (20 µl) contained 1X reaction buffer for

Klenow Fragment, 0.05 mM of each dNTP and 5 U of Klenow Fragment. The mixture was incubated at 37 °C for 30 minutes.

T4 DNA Ligase (Thermo Fisher Scientific) was used for DNA ligation. For stick-end ligation, each reaction (20 µl) contained 1X T4 DNA Ligase Buffer, 100 ng of linearized vector DNA, 200 ng of insert DNA and 1 Weiss U of T4 DNA Ligase. For blunt-end ligation, each reaction (20 µl) contained 1X T4 DNA Ligase Buffer, 100 ng of linearized vector DNA, 200 ng of insert DNA, 2 µl of 50% PEG 4000 Solution and 5 Weiss U of T4 DNA Ligase. The reaction was incubated at 16 °C for overnight.

In-Fusion HD Cloning Kit (Takara) was used for fast, directional DNA ligation, especially for the vector which did not contain appropriate restriction enzyme recognition site. In-Fusion enzyme could fuse DNA fragment and linearized vector DNA by a 15 bp overlap at their ends. Each reaction (10 µl) contained 1X In-Fusion HD Enzyme Premix, 50 ng of insert DNA and 100 ng of linearized vector DNA. The reaction was incubated at 50 °C for 15 minutes.

3.4.5. Transformation of DNA

To prepare chemical competent cells, *Mix & Go E. coli* Transformation Kit (ZYMO RESEARCH) was used. 0.5 ml of overnight *E. coli* strain MDS42 ΔrecA Blue or HB101 cultures were inoculated into 50 ml of liquid LB supplemented with appropriate antibiotics and cultured at 37 °C until the optical density at 600 nm (OD_{600nm}) was 0.4-0.6. After the incubation on ice for 10 minutes, bacterial cells were collected by centrifugation at 4000 rpm for 10 minutes at 4 °C. The cells were re-suspended in 5 ml 1X Wash Buffer gently, followed by the centrifugation at 4000 rpm for 10 minutes at 4 °C. The supernatant was removed completely and the cells were re-suspended in 5 ml 1X Competent Buffer. The competent cells were divided into 100 µl aliquots and stored at -80 °C.

For heat shock transformation, 10 µl of ligation or 1 µl of plasmid was mixed with 100 µl of chemical competent cells by gently flicking. After incubation on ice for 30 minutes, the mixture was placed in 42 °C water bath for 60 seconds and cooled down

on ice for 2 minutes. 900 µl of LB medium was added and the bacterial cells were cultured at 37 °C for 1 hour. Then the cells were plated on LB medium supplemented with appropriate antibiotics and incubated at 37 °C for overnight.

To prepare electro-competent cells, 0.2 ml of overnight *E. coli* strain MDS42 ΔrecA Blue or HB101 cultures were inoculated into 20 ml of liquid LB supplemented with appropriate antibiotics and cultured at 37 °C until the OD_{600nm} was 0.4-0.6. After the incubation on ice for 10 minutes, bacterial cells were collected by centrifugation at 4000 rpm for 10 minutes at 4 °C. The cells were washed with ice-cold sterile distilled water for three times and re-suspended in 50 ml of ice-cold sterile 20% glycerol. The competent cells were divided into 100 µl aliquots and stored in -80 °C.

For electroporation, 10 µl of ligation or 1 µl of plasmid was mixed with 100 µl of electro-competent cells by gently flicking. After incubation on ice for 30 minutes, the mixture was transferred into an ice-cold electroporation cuvette (1mm) without introducing bubbles. Electroporation was performed on Gene Pulser Xcell Electroporation Systems (BIO-RAD) using 1.8 kV, 200 Ω and 25 µF. 900 µl of LB medium was added into cuvette immediately. The mixture was transferred into a new sterile microfuge tube and cultured at 37 °C for 1 hour. Then the cells were plated on LB medium supplemented with appropriate antibiotics and incubated at 37 °C for overnight.

3.4.6. Polymerase Chain Reaction (PCR)

DreamTaq DNA Polymerase (Thermo Fisher Scientific) was used for colony PCR. Each reaction (20 µl) contained 1X DreamTaq Buffer, 200 µM of each dNTP, 0.5 µM of each primer, 0.5 U of DreamTaq DNA Polymerase. The single colonies were picked up by sterile toothpicks and placed into the reaction mixture as template DNA. The PCR reactions were performed using Applied Biosystems Veriti 96-well Thermal Cycler (Thermo Fisher Scientific). After an initial denaturation at 95 °C for 5 minutes, 30 cycles of amplification (95 °C for 30 seconds, 30 s at annealing temperature, and 72 °C for 1 minute/2kb) were performed, followed by a final extension at 72 °C for 10

minutes.

Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) and Q5 High-Fidelity 2X Master Mix (NEW ENGLAND Biolabs) were used for the amplification of fragments. For Phusion High-Fidelity DNA Polymerase, each reaction (50 µl) contained 1X Phusion HF/GC Buffer, 200 µM of each dNTP, 0.5 µM of each primer, moderate template DNA (1 pg-10 ng plasmid DNA or 50 ng-250 ng genomic DNA) and 1 U of Phusion High-Fidelity DNA Polymerase. For Q5 High-Fidelity 2X Master Mix, each reaction (50 µl) contained 1X Q5 High-Fidelity 2X Master Mix, 0.5 µM of each primer and moderate template DNA (1 pg-10 ng plasmid DNA or 50 ng-250 ng genomic DNA). The PCR reactions were performed using Applied Biosystems Veriti 96-well Thermal Cycler (Thermo Fisher Scientific). After an initial denaturation at 98 °C for 30 seconds, 33 cycles of amplification (98 °C for 20 seconds, 20 s at annealing temperature, and 72 °C for 30 seconds/kb) were performed, followed by a final extension at 72 °C for 10 minutes.

3.4.7. Sequencing library construction for the detection of transposon insertions

Genomic DNA isolated from transposon insertion pools were digested with *Mme*I. To create the Illumina adaptor, 10 µl-10 µl of 100 µM oligos IllMme1U-IllMme1L (Table S1) were mixed with 3 µl of T4 ligase buffer and 2 µl of 0.5 M NaCl. After 5 minutes incubation at 95 °C, the mixture was cooled down to ~30 °C. The precipitated *Mme*I digested genomic DNA was dissolved by Illumina adaptor and 5 µl of 5 mM ATP and 1.5 µl of T4 ligase (1 Weiss U/µl) was added, followed by ligation at 16 °C for overnight. The transposon flanking sequences were amplified from ligated product (adaptor ligated genomic DNA) with Q5 High-Fidelity 2X Master Mix using TpnIRill0-TpnIRill1-TpnIRill2-TpnIRill3-Mme1pcrF primers (Table S1). After agarose gel electrophoresis, the 125 bp-128 bp fragments were purified from the gel.

In a second round of a 4 cycle PCR, the adaptors and barcodes for Illumina sequencing were added using the Nextera Index Kit (Illumina), then the libraries were

sequenced on an Illumina NextSeq550 system. The sequencing was performed by Delta Bio 2000 Ltd.

3.4.8. Determination of the transposon insertion sites in Rm41S

Genomic DNA was isolated from transposon insertion mutants of Rm41S. Part of the transposon that carried *nptII* gene with the flanking genomic sequence was digested from the genomic DNA with *Xba*I-*Eco*RI and ligated into *Xba*I-FastAP and *Eco*RI-*Xba*I-FastAP digested pSP-KS2. The ligated products were transformed into MDS42 ΔrecA Blue and plated on LB supplemented with 100 μg/ml spectinomycin and 100 μg/ml kanamycin. The plasmids were isolated and sequenced to determine the insertion site. The primer Tn5prBam (Table S1) was used for sequencing.

3.5. Construction of mutant *Sinorhizobium* strains

3.5.1. Random insertion mutagenesis

S. meliloti strain Rm41S, the streptomycin resistant derivative of Rm41, was cultured in liquid TA medium supplemented with 100 μg/ml streptomycin while *E. coli* strains carrying the transposon sources and the pRK2013 helper plasmid were grown in LB medium supplemented with 100 μg/ml kanamycin. The transposon sources were transferred into Rm41 by triparental mating and the bacteria with transposon insertion were selected on MIN plates supplemented with 100 μg/ml streptomycin and 400 μg/ml kanamycin. All of the colonies were collected as one insertion mutant pool and 105+96 independent pools were created in the same way.

3.5.2. Chemical mutagenesis of Rm41S

S. meliloti strain Rm41S was cultured in liquid TA supplemented with 100 μg/ml streptomycin at 30 °C. Bacterial cells from 1 ml of log-phase culture ($OD_{600nm}$was 0.2) were collected by centrifugation at 4000 rpm for 2 minutes. The pellets were washed

and re-suspended in 1 ml of 0.1 M sodium citrate (pH 5.5). N-methyl-N'-nitro-N-nitrosoguanidine (MNNG) was added to a final concentration of 200 µg/ml, and the cells were incubated at 30 °C for 60 minutes. They were collected by centrifugation at 4000 rpm for 2 minutes and washed with 0.1 M potassium phosphate (pH 7.0). The pellets were re-suspended in 5 ml of liquid TA medium and incubated at 30 °C with shaking at 160 rpm. Second day, all of bacterial cells were pelleted by centrifugation, re-suspended in 100 µl of TA medium and plated on TA agar plate supplemented with 100 µg/ml streptomycin. After three days of incubation at 30 °C, all of the colonies which grew on TA plate were collected as one chemical mutant pool and 40 pools were created in the same way.

3.5.3. Construction of targeted mutants in strain Rm41

Fragments for plasmid integration mutagenesis were amplified from Rm41 genomic DNA using specific primer pairs (Table S1) which include the appropriate restriction enzyme sites. Following digestion, both PCR fragments were purified from agarose gel, ligated into purified digested pK19mob plasmid and the ligated product was transformed into MDS42 ΔrecA Blue chemically competent cells. The transformed cells were then plated on LB agar plates supplemented with 100 µg/ml kanamycin, 100 µM IPTG and 20 µg/ml X-gal and incubated at 37 °C for overnight. Colony PCR were performed using primer pair OPSEQ-UNIV (Table S1) and the plasmids were isolated from putative positive clones, verified by digestion and sequenced for further confirmation. Correct clones were transformed into HB101 chemically competent cells, transferred into strain Rm41 by triparental mating and the strains with integrated plasmid were selected on kanamycin containing MIN plates.

The resistance cassette insertion mutants of *BN406_06091* and *BN406_06087* were created in the similar way. *BN406_06091* and *BN406_06087* (LPS biosynthesis protein) were amplified from Rm41 genomic DNA using primer pairs BN406_06091accF-BN406_06091sacR and LPSbiosyntAccF-LPSbiosyntSacR (Table S1). The PCR fragments were purified from agarose gel, digested with *Acc*65I-*Sac*I and ligated into

*Acc*65I-*Sac*I-FastAP digested pBlueScriptII KS+ (Alting-Mees and Short, 1989). The ligated products (pBS::BN406_06091 and pBS::LPSbiosynt) were transformed into MDS42 ΔrecA Blue and plated on LB supplemented with 100 µg/ml ampicillin, 100 µM IPTG and 20 µg/ml X-gal. The *aad2* gene providing spectinomycin resistance was amplified from plasmid pSP-KS2 using primer pair SpCmPst-SpCmNhe (Table S1) and ligated into *Sma*I digested pBS::BN406_06091 and *Mlu*I-Klenow Fragment digested pBS::LPSbiosynt. The ligated products (pBS::BN406_06091-SpcR and pBS::LPSbiosynt-SpcR) were transformed into MDS42 ΔrecA Blue and plated on LB agar plates supplemented with 100 µg/ml ampicillin and 100 µg/ml spectinomycin. BN406_06091-SpcR and LPSbiosynt-SpcR fragments were digested from pBS::BN406_06091-SpcR and pBS::LPSbiosynt-SpcR with *Acc*65I-*Ecl*136II, the sticky ends were filled in with Klenow Fragment and ligated into *Sma*I digested pK18mobsacB. The ligated products (pK18mobsacB::BN406_06091-SpcR and pK18mobsacB::LPSbiosynt-SpcR) were transformed into MDS42 ΔrecA Blue and plated on LB agar plates supplemented with 100 µg/ml kanamycin and 100 µg/ml spectinomycin. Colony PCR were performed using primer pair OPSEQ-UNIV (Table S1) and the plasmids were isolated from putative positive clones, verified by digestion and sequenced for further confirmation. Correct plasmids were transformed into HB101 chemically competent cells.

To create deletion derivatives in other LPS related genes, two fragments upstream and downstream of the target genes were amplified from Rm41 genomic DNA using specific primer pairs (Table S1) which include the appropriate restriction enzyme sites. Following digestion, both PCR fragments were ligated into digested pK18mobsacB plasmid and the ligated product was transformed into MDS42 ΔrecA Blue. The transformed cells were then plated on LB agar plates supplemented with 100 µg/ml kanamycin, 100 µM IPTG and 20 µg/ml X-gal and incubated at 37 °C for overnight. Colony PCR were performed using primer pair OPSEQ-UNIV (Table S1) and the plasmids were isolated from putative positive clones, verified by digestion and sequenced for further confirmation. The correct plasmids were introduced into HB101

by heat shock transformation.

The constructs were transferred from HB101 into Rm41S by triparental mating. Two colonies grown on selective medium were streaked onto new MIN plates supplemented with 100 µg/ml streptomycin, 400 µg/ml kanamycin (and in the case of pK18mobsacB::BN406_06091-SpcR, 100 µg/ml spectinomycin) for purification. Following 3-4 days incubation, two single colonies were chosen randomly and inoculated into 1 ml of liquid TA medium (10^0 dilution). The next day ten-fold serial dilutions (10^{-1}, 10^{-2}) were made in liquid TA and 100 µl of all three dilutions were plated onto TA plates containing 10% sucrose. Single colonies appeared in 10% sucrose selection were then streaked on TA plates supplemented with 100 µg/ml streptomycin, 400 µg/ml kanamycin (and 100 µg/ml spectinomycin) and TA containing 100 µg/ml streptomycin (and 100 µg/ml spectinomycin). After 2 days incubation at 30 °C, the colonies which could grow only on TA plates supplemented with 100 µg/ml streptomycin (and 100 µg/ml spectinomycin) were tested by colony PCR for the presence of the deletion or the antibiotic cassette insertion.

3.6. Constructing libraries for genetic complementation

3.6.1. Construction of a nodule-expressed *Sinorhizobium meliloti* 1021 ORFeome library

3.6.1.1. Construction of Gateway compatible destination vectors

To create moderate size broad host range cloning vectors for rhizobia, the *Bgl*II fragment carrying the T-DNA part of binary vector pPR97 (Szabados et al., 1995) was replaced by the PCR fragment coding for the LacZα fragment amplified from vector pGREEN (Hellens et al., 2000). The *nptIII* gene providing kanamycin resistance in both *E. coli* and rhizobia/agrobacteria in the resulting plasmid termed pPRlacZ was replaced by the *aacC1* (coding for a gentamycin acetyltransferase) and *tetA* (coding for a tetracyclin/cation antiporter) genes to create pPRgmLacZ and pPRtcLacZ, respectively.

To create the destination vectors for cloning the *S. meliloti* 1021 ORFeome library (Schroeder et al., 2005) via Gateway recombination behind a nodule-specific promoter, a 537 bp preceding the *nifH* gene of *S. meliloti* was amplified with primers nifHprAccF and nifHprHindR (Table S1) and cloned into the *Acc*65I and *Hin*dIII sites of pPRtcLacZ to obtain pPRtcLacNifHpr. The Gateway cassette with reading frame B (Gateway Vector Conversion System, Invitrogen) was amplified with primers GWhindF and pKTgwXbaR (Table S1) and was cloned as a *Hin*dIII-*Xba*I fragment behind the *nifH* promoter sequence to obtain the pPRtcLacNifHprGW vector.

To create destination vectors to express the ORFs both in free-living state and in the nodule, the promoters of the *cycHJKL* operon (Kereszt et al., 1995) and of the *bacA* gene (Glazebrook and Walker, 1993) were amplified (Table S1) as *Eco*RV-*Bam*HI and *Acc*65I-*Bam*HI fragments, respectively, while the matching Gateway sequences were amplified (Table S1) as *Bgl*II-*Acc*65I and *Bgl*II-*Xba*I fragments. The promoter and Gateway cassette sequences were cloned into the *Ecl*136II-*Acc*65I and *Xba*I-*Acc*65I digested pPRtcLacZ vectors via tri-fragment ligations to obtain the pPRtcLacCycHprGW and the pPRtcLacBacAprGW vectors.

3.6.1.2. Transfer of the ORFeome library into the destination vector by in vivo recombination

To transfer the ORFeome library into pPRtclacNifHprGW destination vector, pentaparental matings were performed as described by House et al. (2004). Recipient strain HB101Rif, helper strain HB101(pRK2013), 67 pools of donor strains containing *Sinorhizobium meliloti* 1021 ORFs, DB3.1(pPRtclacNifHprGW) and DH5α λpir (pXINT29) were cultured for overnight in liquid LB supplemented with 100 µg/ml rifampicin, 100 µg/ml kanamycin, 100 µg/ml kanamycin, 15 µg/ml tetracycline and 100 µg/ml kanamycin, respectively. The mixture of 2 ml of HB101Rif and 1 ml of each of other four strains were centrifuged at 4000 rpm for 2 minutes and washed once with sterile distilled water to remove residual antibiotics. The mixed bacterial pellets were then re-suspended in 100 µl of LB medium and dropped onto LB agar plates

supplemented with 100 μM IPTG. After incubation at 30 °C for overnight, bacterial cells were re-suspended in 100 μl of sterile distilled water and plated on LB plates supplemented with 100 μg/ml rifampicin and 15 μg/ml tetracycline. After incubation for overnight at 37 °C, all of the colonies were collected.

3.6.1.3. Transfer of the ORFeome library into strain Rm41

Recipient strain Rm41 was cultured in liquid TA for 48 hours. Donor strains HB101Rif/pPRtclacNifHprGWSm1021ORFeome (67 pools) and helper strain HB101/pRK2013 were cultured for overnight in liquid LB supplemented with 100 μg/ml rifampicin, 15 μg/ml tetracycline and LB supplemented with 100 μg/ml kanamycin, respectively. The HB101Rif/pPRtclacNifHprGWSm1021ORFeome pools were introduced into Rm41 by triparental mating and the bacteria with ORFs were selected on MIN plates supplemented with 15 μg/ml tetracycline. After three days of incubation at 30 °C, all of the colonies were collected.

3.6.2. Construction of a large insert size genomic library from *Sinorhizobium meliloti* strain FSM-MA

Genomic DNA was isolated from superpools (10 pools/superpool) of *S. meliloti* strain FSM-MA mutants obtained after random insertion mutagenesis with transposon TPM1L71KMCS carrying a number of restriction enzyme recognition sites that are present in the *Sinorhizobium* genome only at low frequency. 5 μg genomic DNA from each superpools was digested with enzymes *Sbf*I, *Bsr*GI, *Acl*I, *Bst*1107I, *Dra*I, *Nhe*I, *Nhe*I-*Xba*I, *Nhe*I-*Spe*I and fragments between 15 and 33 kbp (the two fragments of λ phage DNA digested with *Xho*I) were isolated from 0.5% agarose gel. The DNA fragments were ligated into the pPZPgmSK2 vector (Attila Kereszt, unpublished development) digested with *Pst*I (compatible with *Sbf*I fragments), *Xba*I (compatible with fragments after *Nhe*I, *Xba*I, *Spe*I digestions), *Cla*I (compatible with *Acl*I fragments), *Acc*65I (compatible with *Bsr*GI fragments), *Ecl*136II (compatible with

fragments with blunt ends such as generated by *Dra*I and *Bst*1107I) and dephosphorilated by the FastAP enzyme. After overnight ligation at 16 °C, the ligation reactions were desalted by gel filtration through Sephadex G50 columns, divided into 7 µL aliquots and each aliquot was transformed into MDS42recAblue cells by electroporation and plated on LB medium supplemented with 25 µg/mL gentamycin, 100 µM IPTG and 20 µg/ml X-gal. 25 colonies from the transformation were streaked and grown for overnight at 37 °C on LB plates with 25 µg/mL gentamycin, 100 µM IPTG and 20 µg/ml X-gal, then the cells were washed down with liquid LB medium containing 20% glycerol to create a pool. 250 such pools were collected and stored at -80 °C. The pools were introduced into strain Rm41S by triparental mating.

3.7. Plant related techniques

3.7.1. Plant materials and seed germination

Medicago truncatula F83005 and Jemalong seeds were scarified in 5 ml of concentrated H_2SO_4 (96%) for 8 minutes in order to break the seed coat and washed with ice-cold sterilized water for three times. The scarified seeds were then surface sterilized in 5 ml of 0.1% (weight/volume) $HgCl_2$ for 3 minutes and washed with sterilized water for five times. The sterilized seeds were kept on 1% agar plates at 4 °C for 48 hours in dark and germinated at 22 °C for 12 hours in dark.

3.7.2. Nodulation assay

After germination, seedlings were grown in vermiculite in a greenhouse programmed for 16 hours light and 8 hours dark at 22 °C. For nodulation assay, individual wild-type and mutant strains as well as pools of transposon insertion mutants of Rm41S were used as inoculums on the roots of one-week-old F83005 seedlings. Similarly, individual strains, pools of transposon insertion and chemically induced mutants as well as transconjugants carrying Sm1021 ORFeome library and FSM-MA genomic library were used on the roots of one-week-old Jemalong seedlings. The whole process was

done under sterile and nitrogen-free conditions and nodulation phenotypes were examined four weeks post inoculation.

3.7.3. Isolation and identification of bacteria from Fix$^+$ nodules

Fix$^+$ nodules were cut from the roots and sterilized with 70% ethanol for 2 minutes followed by washing three times with sterile distilled water. The nodules were then put into 1.5 ml microfuge tubes containing 1 ml of liquid TA and crushed with the help of grinding rod (10^0 dilution). Ten-fold serial dilutions (10^{-1}, 10^{-2}) were made in liquid TA and 100 µl of all three dilutions were plated on TA agar plates. After three days incubation at 30 °C single colonies were streaked on TA and TA plates supplemented with 100 µg/ml streptomycin (in case of chemical mutagenesis of Rm41S), 15 µg/ml tetracycline (in case of overexpression of Sm1021 ORFs in Rm41), 100 µg/ml streptomycin and 400 µg/ml kanamycin plates (in case of random insertion mutagenesis of Rm41S) or 100 µg/ml streptomycin and 200 µg/mL gentamycin (in case of expression of FSM-MA genomic DNA in Rm41S). The Rm41 origin of bacteria that grew on both plates were confirmed by phage test. 500 µl of bacterial culture was mixed with 6 ml of top agar (TA, 0.7% agar) and poured on TA plate. When the top agar was solidified, 5 µl of bacteriophage 16-3 (a temperate phage of *S. meliloti* strain Rm41) was dropped on the plate. The plate was incubated at 30 °C for overnight. The identified bacteria were then re-tested individually in nodulation assay. Those mutants which formed effective symbiosis were stored in -80 °C.

3.8. Localization of the protein encoded by the BN406_06091 gene

3.8.1. Creating gene constructs coding for proteins fused to mCherry

To study the localization of the protein encoded by the *BN406_06091* gene, a series of constructs were created using a modified pHC60 vector as a backbone. To remove

the second *Nde*I site from the GFP coding sequence, GFP1 and GFP2 fragments were amplified from pHC60 using primer pairs U2-mCHmNdeR, OPSEQ-mCHmNdeF (Table S1). An overlap PCR was performed to obtain the *Nde*I (in GFP) deleted GFP amplicon using primer pair OPSEQ-UNIV (Table S1). The PCR fragment (GFPΔNdeI) was purified from agarose gel, digested with *Acc*65I-*Cla*I and ligated into *Acc*65I-*Cla*I-FastAP digested pSP-KS2 plasmid. The ligated product (pSP-KS2::GFP) was transformed into MDS42 ΔrecA Blue and plated on LB agar plates supplemented with 100 µg/ml spectinomycin, 100 µM IPTG and 20 µg/ml X-gal.

To create the construct coding for C-terminal mCherry fusion protein, *BN406_06091* was amplified from Rm41 genomic DNA using primer pair 06091sacF-06091nheR (Table S1). The *mCherry* gene was amplified from pSParaC-mCherry using primer pair mCherryNheF-T7promoter (Table S1). The two PCR fragments were purified from agarose gel, digested with *Sac*I-*Nhe*I and *Acc*65I-*Nhe*I respectively, and ligated into *Acc*65I-*Sac*I digested pBlueScriptII KS+. The ligated product (pBS::06091-mCherry) was transformed into MDS42 ΔrecA Blue and plated on LB supplemented with 100 µg/ml ampicillin, 100 µM IPTG and 20 µg/ml X-gal. Then 06091-mCherry fragment was amplified from pBS::06091-mCherry using primer pair 06091pHCinFF-mCHpHCinFR (Table S1) and purified from agarose gel.

To create the construct coding for N-terminal mCherry fusion protein, the *BN406_06091* promoter (06091pr) was amplified from Rm41 genomic DNA using primer pair 06091sacF-06091prR (Table S1). The *mCherry* gene was amplified from pSParaC-mCherry using primer pair 06091prCherryF-T7promoter (Table S1). An overlap PCR was performed to obtain the 06091prmCherry fragment using primer pair 06091sacF-T7promoter (Table S1). The PCR fragment was purified from agarose gel, digested with *Acc*65I-*Sac*I and ligated into *Acc*65I-*Sac*I-FastAP digested pBlueScriptII KS+. The ligated product (pBS::06091prmCherry) was transformed into MDS42 ΔrecA Blue and plated on LB supplemented with 100 µg/ml ampicillin, 100 µM IPTG and 20 µg/ml X-gal. The *BN406_06091* fragment was amplified from Rm41 genomic DNA using primer pair 06091bglF-06091accR (Table S1), purified from agarose gel, digested

with *Acc*65I-*Bgl*II and ligated into *Acc*65I-*Bgl*II-FastAP digested pBS::06091prmCherry. The ligated product (pBS::mCherry-06091) was transformed into MDS42 ΔrecA Blue and plated on LB supplemented with 100 μg/ml ampicillin. Then mCherry-06091 fragment was amplified from pBS::mCherry-06091 using primer pair mCHpHCinFF-06091pHCinFR (Table S1) and purified from agarose gel.

To create the construct coding for C-terminal HA fusion protein, the *BN406_06091* gene was amplified from Rm41 genomic DNA using primer pair 06091sacF-06091_3GlyHAsmaR (Table S1), purified from agarose gel and In-Fusion ligated into *Ecl*136II-*Sma*I digested pSPxcm3G_HA. The ligated product (pSPxcm3G::06091-HA) was transformed into MDS42 ΔrecA Blue and plated on LB supplemented with 100 μg/ml spectinomycin. Then 06091-HA fragment was amplified from pSPxcm3G::06091-HA using primer pair 06091pHCinFF-HApHCinFR (Table S1) and purified from agarose gel.

The *BN406_06091* gene was amplified from Rm41 genomic DNA using primer pair 06091pHCinFF-06091pHCinFR (Table S1) and purified from agarose gel.

A periplasmic protein EhuB (Hanekop et al., 2007) was used as a control. The signal peptide of *EhuB* gene was amplified from Rm41 genomic DNA using primer pair EhuBpagF-EhuBnheR (Table S1). The *mCherry* gene was amplified from pSParaC-mCherry using primer pair mCherryNheF-mCHecoRIr (Table S1). Both PCR fragments were purified from agarose gel, digested with *Nhe*I and ligated together. The EhuBsp-mCherry fragment was amplified from ligated product using primer pair EhuBpHCinFF-mCHpHCinFR (Table S1) and purified from agarose gel.

All of the above five PCR fragments (06091-mCherry, mCherry-06091, 06091-HA, 06091 and EhuBsp-mCherry) were In-Fusion ligated into *Nde*I digested pSP-KS2::GFP, transformed into MDS42 ΔrecA Blue and plated on LB supplemented with 100 μg/ml spectinomycin. The 06091-mCherry-GFP, mCherry-06091-GFP, 06091-HA-GFP, 06091-GFP, EhuBsp-mCherry-GFP and GFP fragments were digested from pSP-KS2::06091-mCherry-GFP, pSP-KS2::mCherry-06091-GFP, pSP-KS2::06091-HA-GFP, pSP-KS2::06091-GFP, pSP-KS2::EhuBsp-mCherry-GFP and pSP-KS2::GFP

with *Acc*65I-*Cla*I, ligated into *Acc*65I-*Cla*I-FastAP digested pHC60, transformed into MDS42 ΔrecA Blue and plated on LB supplemented with 15 µg/ml tetracycline.

All of the six plasmids (pHC60::06091-mCherry, pHC60::mCherry-06091, pHC60::06091-HA, pHC60::06091, pHC60::EhuB-mCherry and pHC60::GFP) were verified by digestion and sequencing, and introduced into the mutant strain Rm41S_*mBN406_06091* by triparental mating.

3.8.2. Confocal microscopy

The three rhizobia strains Rm41S_*mBN406_06091*/pHC60::*EhuB-mCherry*, Rm41S_*mBN406_06091*/pHC60::*06091-mCherry* and Rm41S_*mBN406_06091*/pHC60:: *mCherry-06091* were cultured in liquid TA supplemented with 100 µg/ml streptomycin, 200 µg/ml kanamycin and 15 µg/ml tetracycline at 30 °C for 48 hours. Fluorescence confocal images were taken by FV1000 confocal laser scanning biological microscope (Olympus). The GFP signals were excited by a 488 nm laser line and detected using a 500-550 nm emission filter. For the mCherry signals, a 575 nm laser line and a 588-650 nm emission filter were used.

3.8.3. Isolation of extracellular, periplasmic and cytoplasmic proteins

The two rhizobia strains Rm41S_*mBN406_06091*/pHC60::*06091-HA* and Rm41S_*mBN406_06091*/pHC60::*GFP* were cultured in 50 ml of liquid TA supplemented with 100 µg/ml streptomycin, 200 µg/ml kanamycin and 15 µg/ml tetracycline at 30 °C until OD_{600nm} was 0.5-0.6. After centrifugation at 4000 rpm for 10 minutes, the supernatant was collected and the extracellular proteins were precipitated by trichloroacetic acid (TCA) as follows. 9 volumes of supernatant were mixed with 1 volume of 100% (w/v) TCA and incubated on ice for 30 minutes. The precipitate was collected by centrifugation at 4000 rpm for 20 minutes at 4 °C. The supernatant was removed and the precipitate was washed with ice-cold acetone for three times and

dissolved in SDS PAGE loading buffer (250 mM Tris-HCl (pH 6.8), 10% SDS, 30% glycerol, 5% β-mercaptoethanol, 0.02% bromophenol blue).

The periplasmic and cytoplasmic proteins were isolated as described by de Maagd and Lugtenberg (1986). The rhizobial cells were re-suspended in 5 ml of lysis buffer (50 mM Tris (pH 8.0), 20% (w/v) sucrose, 2 mM EDTA, 0.2 mg of lysozyme per ml) and incubated at room temperature for 30 minutes, followed by centrifugation at 4000 rpm for 20 minutes at 4 °C. The periplasmic proteins were precipitated by TCA from the supernatant and dissolved in SDS PAGE loading buffer. For cytoplasmic proteins, the remaining cells were suspended in SDS PAGE loading buffer. All of the protein samples were stored in -80 °C.

3.8.4. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed to separate and analyse proteins. 10% resolving gel (Table 3.3) and 5% stacking gel (Table 3.3) were prepared successively. 15 µl of protein samples were incubated at 95 °C for 10 minutes and loaded into stacking gel. The electrophoresis was run in SDS running buffer (3 g/l Tris base, 14.4 g/l glycine, 1 g/l SDS) at 120 V for 90 minutes using PowerPac HV Power Supply (BIO-RAD).

Table 3.3 Preparation of resolving gel and stacking gel for SDS-PAGE

Resolving gel		**Stacking gel**	
Distilled water	5.9 ml	Distilled water	2.7 ml
30% Acrylamide mix	5.0 ml	30% Acrylamide mix	670 µl
1.5 M Tris-HCl (pH 8.8)	3.8 ml	1.5 M Tris-HCl (pH 6.8)	500 µl
10% SDS	150 µl	10% SDS	40 µl
10% Ammonium persulfate	150 µl	10% Ammonium persulfate	40 µl
TEMED	6 µl	TEMED	4 µl

3.8.5. Western blot analysis

After the electrophoresis, proteins were transferred to the PVDF (polyvinylidene difluoride) membrane which was pre-wet in 100% methanol. The transfer was performed in ice-cold transfer buffer (10% ethanol, 3 g/l Tris base, 14.4 g/l glycine) at 4 °C, 20 V for overnight using PowerPac HV Power Supply (BIO-RAD).

To detect GFP protein, anti-GFP antibody (ROCHE) was used in Western blot analysis. The membrane was blocked in PBS (1 mM KH_2PO_4, 10 mM Na_2HPO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.0) containing 5% milk powder with gentle rotation for 1 hour and incubated in working-strength anti-GFP reagent (PBS containing 5% milk powder and 0.4 µg/ml anti-GFP antibody) for 1 hour with gentle rotation. After washing with PBS containing 0.1% Tween 20 (PBST) for 3 times, 10 minutes per wash, the membrane was incubated in a 1:3000 dilution of anti-Mouse IgG-peroxidase antibody (Sigma-Aldrich) in PBS for 1 hour with gentle rotation, followed by 3 times washes with PBST. SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific) was then used to detect signals.

To detect HA fusion proteins, HA Tag Monoclonal Antibody (Invitrogen) was used in Western blot analysis. The membrane was blocked in TBST (150 mM NaCl, 20 mM Tris-HCl, 0.1% Tween-20, pH 7.6) containing 5% milk powder with gentle rotation for 1 hour and incubated in HA antibody reagent (TBST containing 5% milk powder and 0.3 µg/ml HA antibody) for 3 hours with gentle rotation. After washing with TBST for 3 times, 10 minutes per wash, the membrane was incubated in a 1:3000 dilution of anti-Mouse IgG-peroxidase antibody (Sigma-Aldrich) in PBS for 1 hour with gentle rotation, followed by 3 times washes with TBST. SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific) was then used to detect signals.

3.9. Bioinformatic analysis

To determine how many transposon insertion mutants can be found in the pools, sequence data obtained on the libraries described in 3.4.7. were analyzed by the

Geneious Prime software package (Biomatters Ltd, New Zealand). Reverse reads from the paired-end sequencing were trimmed to 30 nucleotides from the 3' end and mapped to the targeted end of the transposon extended by 16 random nucleotides (N) corresponding to the transposon flanking sequence tag (FST) obtained after the *Mme*I digestion. Those reads that were mapped to this reference sequence were trimmed to obtain the FSTs. FSTs were mapped to the latest version of the *S. meliloti* strain Rm41 genome (Accession numbers: CP021808-CP021811) and the consensus sequences at the mapped reads were listed and automatically counted by the software (Solaimanpour et al., 2015).

4. Results and Discussion

4.1 Identification of incompatible interactions between *Sinorhizobium meliloti/medicae* strains and the ecotypes of the model legume *Medicago truncatula*: strain Rm41 is not compatible with ecotypes F83005 and Jemalong

Most of the genetic studies to identify genetic determinants required for the development of indeterminate nodules has used *M. truncatula* ecotypes Jemalong and R108 as well as *S. meliloti* strain 1021 and *S. medicae* strain WSM419 although high number of symbiotic mutants is available only in strain 1021. Moreover, none of the two strains are fully effective with the two dominantly used ecotypes (Terpolilli et al., 2008; Kazmierczak et al., 2017) indicating the existence of genetic differences affecting the symbiotic compatibility and efficiency between strains and ecotypes.

To investigate how frequent the incompatibility is between the strains and ecotypes of different geographical origin, our research group and two other research teams (Crook et al., 2012; Liu et al., 2014) screened several ecotypes of *Medicago truncatula* using different *Sinorhizobium meliloti* and *Sinorhizobium medicae* strains as inoculants (Table 4.1). We identified more than 10 incompatible host-strain interactions in which the host plants showed yellowish leaves and smaller shoots, characteristic phenotypes associated with nitrogen starvation. In particular, we noticed that *S. meliloti* strain Rm41 of Hungarian origin and intensively used in our institute was incompatible with *Medicago truncatula* ecotypes F83005 and Jemalong, while the same strain could nodulate and fix nitrogen normally in association with other *Medicago truncatula* lines (such as A20, DZA315.16, DZA045, SA023859, …). Although the final outcome of the two incompatible interactions was the same, i.e. nitrogen starved plants, symbiotic development was arrested at different stages in the two plants: In ecotype F83005, strain Rm41 induces root hair curling and cortical cell division but infection thread and-as a consequence-nodule formation is ceased thus only bumps or occasionally empty

nodules are formed on the roots. In contrast, nodule formation, bacterial infection and bacteroid development take place in line Jemalong but bacteroids are eliminated from the nodule cells. From these differences we could conclude that the incompatibilities are caused and controlled by different genetic determinants/genes in strain Rm41.

Table 4.1 Symbiosis specificity in different *Medicago truncatula* - *Sinorhizobium* pairs

Sinorhizobium	***Medicago truncatula***					
strains	**Jemalong**	**A20**	**DZA315.16**	**DZA220H**	**F83005**	**L746**
S. meliloti FSM-MA	Fix+	Fix+	Fix+	Fix+	Fix+	Fix+
S. meliloti 1021	Fix+ (-)	Fix+	Fix+	Fix+	Fix+	**Fix-**
S. meliloti Rm41	**Fix-**	Fix+	Fix+	Fix+	**Fix-**	Fix+
S. meliloti GR4	**Fix-**	Fix+	Fix+	Fix+	Fix+	Fix+
S. meliloti SM11	**Fix-**	**Fix- (+)**	**Fix-**	**Fix-**	**Fix-**	**Fix-**
S. medicae WSM419	Fix+	Fix+	Fix+	Fix+	Fix+	Fix+
S. medicae ABS7	Fix+	Fix+	Fix+	**Fix-**	Fix+	Fix+

Fix+, effective nitrogen-fixing pairs; Fix-, ineffective nitrogen-fixing pairs

4.2 Establishment of genomic resources to identify the bacterial genes that control the fate of the interactions

Theoretically, there are two different explanations for an ineffective phenotype of a rhizobial strain: it can be caused by an activity that is either absent from or present in the bacterium meaning that two different strategies are needed to achieve the restoration of the ineffective interaction. To identify an interaction-detrimental function, the disruption/impairments of the encoding gene is needed. i.e. random insertion or point mutations in the incompatible strain has to be created. To introduce a gene missing from or inactive in the ineffective strains, genomic libraries from compatible strain(s) have to be constructed.

4.2.1 Creation of mutant pools of strain Rm41

There are multiple ways to create bacterial mutants that may lead to different “genetic outcomes”. Transposon insertions result in gene disruption and loss of protein function.

Often, the insertion has polar effect meaning that the transcription of downstream genes in the operon is also blocked, thus, the activity of multiple proteins is lost. There are cases, however, when complete loss of protein function is undesirable because it affects multiple biological pathways and/or the viability of the cells. Point mutations, however, can introduce subtle (amino acid) changes that might not be detrimental for the essential activity of the protein, rather, they might affect for example, its interaction(s) with other proteins or substrates. Because of this reasoning, we decided to perform both chemical mutagenesis and random transposon insertion mutagenesis.

4.2.1.1 Chemical mutagenesis of strain Rm41

To introduce point mutations into the Rm41 genome, we used N-methyl-N'-nitro-N-nitrosoguanidine (MNNG) acting by adding alkyl groups to the O^6 and O^4 atoms of guanine and thymine, respectively, as described by Glazebrook et al. (1996). We have not measured the efficiency of the mutagenesis but according to their published results, this mutagenesis protocol resulted in 90 to 95% lethality, and 1 to 2% of the survivors were auxotrophs. As our mutagenesis was performed with 40 populations of 1 mL bacteria at OD600 ≈ 0.2 (~2x 10^7 cells), there might have been ~10^6 survivors per population, i.e. altogether ~4x10^7 mutagenized cells. Although these cells might carry multiple mutations, if we calculate by one mutation per cell and with an approximately 7 Mbp genome size, each nucleotide might be targeted six times on average.

4.2.1.2 Random transposon insertion mutagenesis of strain Rm41

For the insertion mutagenesis of strain Rm41, the *mariner*-based transposon vector, pSAM_Rl (Perry and Yost, 2014) was used. The vector backbone contains an ampicillin resistance gene, an *RK6γ oriR* origin of replication allowing plasmid maintenance only in cells with a *λpir* function, an origin of transfer of the RP4 plasmid (oriT) as well as the transposase of the *himar1C*9 transposon preceded by the promoter and 5 'UTR sequence of the *Rhizobium leguminosarum rpoD* gene. The transposon contains the

*Mme*I-adapted mariner inverse repeats (IR_R, IR_L) and the *nptII* gene of transposon Tn5 (providing kanamycin resistance) followed Rho-independent terminators (rrnB_T1, rrnB_T2). This vector, which cannot replicate in rhizobia, was conjugated into strain Rm41 by tri-parental mating and rhizobia with insertion events were selected on minimal medium (where the auxotroph *E. coli* could not grow) supplemented with kanamycin. In two rounds, 105 plus 96, altogether 201 independent mutant pools were generated.

To estimate how many independent mutants were collected after the large scale mutagenesis, an Illumina technique-based transposon sequencing approach was applied to selected pools. To create Illumina sequencing libraries, genomic DNA was isolated from the cells of the pools and digested with the restriction enzyme *Mme*I, which cleaves the DNA 20/18 nucleotides after the recognition sequence resulting in a two nucleotide 3' overhang. A splinkerette (Devon et al., 1995) Illumina adaptor was ligated to the digested DNA then transposon flanking sequences were amplified using a transposon specific primer, which also carries Illumina-specific sequences for library production, and the splinkerette-specific Illumina primer. The 125-128 bp fragments were isolated from agarose gel, then Illumina adaptors and barcodes were added in a second PCR of few cycles, finally, approximately 250.000 reads were obtained from each library. After trimming the Illumina-and transposon-specific sequences, the 16 nucleotide long reads were mapped to the reference genome and the consensus sequences were generated and counted by the Geneious software. The summary of the sequencing results is shown in Table 4.2.

As it is shown in Table 4.2, each pools contains high number of independent transposon insertion mutants, much more than we expected (~2.000 clones per pool). If we calculate with the average number of independent mutants, our 201 populations may contain more than 1.5 million independently obtained insertion mutants. As this transposon integrates at TA sites into the DNA, there are only 2-300.000 target sites in the genome, thus, each site in genes non-essential for survival could have been targeted five-times in average. It is also worth to note that the distribution of insertions among

the replicons (chromosome, symbiotic plasmid pSymA, symbiotic plasmid pSymB and Plasmid3) correlates well with the size of the replicons without bias.

Table 4.2 Analysis of the results of transposon sequencing from the selected pools

Library	Number of insertions				
	chromosome	**pSymA**	**pSymB**	**Plasmid3**	**Total**
1st round pool 40	1897	908	1051	127	3983
1st round pool 41	2626	1356	1469	167	5618
2nd round pool 40	5292	2625	3040	289	11246
2nd round pool 41	4484	2284	2519	271	9558
Total	14299	7173	8079	854	30405
Average	3574.75	1793.25	2019.75	213.5	7601.25
Percentage of insertions	47.03%	23.59%	26.57%	2.81%	100.00%
Percentage of genome	51.52%	21.70%	23.32%	3.46%	100.00%
Genome size	3678504	1549593	1665079	247065	7140241

4.2.2 Construction of genomic libraries for complementation

A missing function can be caused by the absence of or a mutation in a gene or operon that can be complemented by genes/operons from a compatible strain. An ORFeome library carrying all predicted ORFs of the reference strain *S. meliloti* strain 1021 in a Gateway entry vector (Schroeder et al., 2005) was made available for us that could be used for the construction of a library. However, we saw two drawbacks of this library: 1) it contained only individual genes that cannot complement strains where whole operons are missing; 2) strain 1021 is not fully compatible with the *M. truncatula* cultivar Jemalong (see Table 4.1) and is incompatible with a number of other ecotypes. That is why, we decided to create a large-insert size genomic library from strain FSM-MA (see Table 4.1), which was shown to form effective symbiosis with the two most commonly used ecotypes, Jemalong and R108 (Kazmierczak et al., 2017) as well as with the most ecotypes we use in our laboratory.

4.2.2.1 Construction of the strain 1021 ORFeome library

To use the ORFeome library/libraries to complement an incompatibility or a mutation, we had to transfer the ORFs from the entry clones into destination vector(s) that are able to replicate in rhizobia and contain promoter sequences, which ensure the expression of the genes at the proper developmental stage such as during infection or bacteroid development or nitrogen fixation. To create such vectors, first, we constructed three general purpose broad host-range cloning vectors from the relatively small and moderate copy number binary vector pPR97 (Szabados et al., 1995) that contain the lacZα fragment with the pBlueScript multiple cloning site for blue-white selection and provide resistance against kanamycin/neomycin, gentamycin and tetracycline and termed as pPR97lacZ, pPRgmLacZ and pPRtcLacZ, respectively.

In the next step, we selected three rhizobial promoters with different expression pattern (Roux et al., 2014). The nodule specific *nifH* gene is expressed only in the symbiotic nodule and is activated in the Interzone and reaches its highest activity in the nitrogen-fixing zone (Zone III). The *bacA* gene is expressed both in free-living state and in the nodule but its expression is enhanced in Zone II (infection zone) and is highest in the Interzone where bacteria develop into bacteroids. The *cycHJKL* operon is expressed at lower level than the other two both in cultured and nodule bacteria but constitutively in every zone without any spatial preference. These promoters and the Gateway cassettes were cloned into the vector pPRtcLacZ to obtain the pPRtcLacNifHprGW, the pPRtcLacCycHprGW and the pPRtcLacBacAprGW vectors.

As our earlier results (Wang et al., 2017) showed that the *nifH* promoter is activated during the incompatible interaction between *M. truncatula* cv. Jemalong and strain Rm41, the ORFeome library containing all 6317 annotated ORFs of strain 1021 arranged in 67 pools were transferred from the entry clones into the pPRtcLacNifHprGW vector with the *in vivo* recombination method of House et al. (2004). The 67 pools of the library in the broad host-range vector were then introduced into strain Rm41.

4.2.2.2 Construction of the large insert size genomic library from strain FSM-MA

In our group, strain FSM-MA was mutagenized with such a modified version of the *himar* transposon we used for the mutagenesis of strain Rm41 that contains a number of recognition sites for restriction enzymes cutting rarely the rhizobial genomes. Genomic DNA from the mutagenized pools were isolated and digested with restriction enzymes *Sbf*I (end compatible with the end generated by *Pst*I), *Bsr*GI (end compatible with the end generated by *Acc*65I), *Acl*I (end compatible with the end generated by *Cla*I), *Xba*I, *Spe*I, *Nhe*I (ends compatible with the end generated by *Xba*I) and *Hpa*I, *Dra*I and *Bst*1107I generating blunt end. The resulting DNA fragments were separated by gel electrophoresis in 0. 5% agarose gels and the region corresponding to 15-33 kbp were isolated and ligated into the pPZPgmSK2 broad host-range cloning vector constructed in our lab. Approximately 400 pools of 25-30 clones were created that can be mobilized into rhizobia.

To get insight into the quality of the library, 16 clones from the *Sbf*I cloning were isolated and the ends were sequenced with M13 reverse and forward primers (Table S1). Blast searches against the FSM-MA genome sequences with the obtained reads allowed the positioning of the endpoints on replicons of the genome and the calculation of the insert sizes as shown in Table 4.3.

As we can see, there are clones from all three replicons of the strain, although pSymA sequences are underrepresented while pSymB fragments are overrepresented in this small sample. It is satisfactory that the average insert size is larger than 10 kbp, thus, we might expect a good representation of full operons.

Table 4.3 Summary of the sequencing results from randomly chosen SbfI clones

Library clone	Sequence from replicon	Start of sequence with primer		Length of insert (bp)
		M13 reverse	M13 forward	
1	FSMchr	3005784	3019321	13537
2	FSMchr	3284676	3294904	10228
3	FSMpSymB	156807	143195	13612
4	FSMchr	3019321	3005784	13537
5	FSMpSymA	930730	920896	9834
6	FSMchr	1605508	1613739	8231
7	FSMpSymB	1164628	1170604	5976
8	FSMpSymB	782166	789462	7296
9	FSMchr	1432719	1414827	17892
10	FSMchr	67044	83560	16516
11	FSMpSymB	894665	912837	18172
12	FSMpSymB	894671	878748	15923
13	FSMpSymB	878753	864543	14210
14	FSMpSymB	1381174	1366566	14608
15	FSMchr	1465461	1445484	19977
16	FSMchr	1490679	1472843	17836
Average				**13587**

4.3 Towards the identification of the genetic determinants responsible for the incompatible interactions of strain Rm41

4.3.1 Attempts to identify Rm41 derivative(s) compatible with Jemalong plants

To identify those bacterial genes that are required for the effective symbiosis or cause the incompatibility with Jemalong, the individual pools of the insertion mutants as well as those of the chemical mutants and the transconjugants with the ORFeome library were inoculated on Jemalong in plant nodulation assay.

In the Jemalong-transposon insertion mutants assay, all of the plants showed nitrogen starvation phenotype. The leaves were yellow and small, and the shoots were much shorter than ones in compatible host-strain pairs. After inoculating Jemalong plants with

the chemical mutants assay we found 7 nitrogen-fixing plants (Table 4.4), evidenced by

Table 4.4 Identification of mutants compatible with F83005 and Jemalong

	Strains from nodules	Antibiotic selection	Phage test	Re-test	Identified gene
Jemalong X Chemical mutants	Rm41S_A17MC1	-			
	Rm41S_A17MC2	-			
	Rm41S_A17MC3	-			
	Rm41S_A17MC4	-			
	Rm41S_A17MC5	-			
	Rm41S_A17MC6	+	+	-	
	Rm41S_A17MC7	+	+	-	
Jemalong X ORFeome library	Rm41::ORFeome_A17C1	+	+	-	
	Rm41::ORFeome_A17C2	+	+	-	
	Rm41::ORFeome_A17C3	+	+	-	
	Rm41::ORFeome_A17C4	+	+	-	
	Rm41::ORFeome_A17C5	+	+	-	
	Rm41::ORFeome_A17C6	+	+	-	
	Rm41::ORFeome_A17C7	-			
	Rm41::ORFeome_A17C8	-			
	Rm41::ORFeome_A17C9	-			
	Rm41::ORFeome_A17C10	-			
	Rm41::ORFeome_A17C11	-			
	Rm41::ORFeome_A17C12	-			
	Rm41::ORFeome_A17C13	-			
	Rm41::ORFeome_A17C14	-			
	Rm41::ORFeome_A17C15	-			
F83005 X Insertion mutants	Rm41S::SAM_F83005C1	+	+	+	*BN406_06091*
	Rm41S::SAM_F83005C2	+	+	+	*BN406_06091*
	Rm41S::SAM_F83005C3	+	+	+	*BN406_06091*
	Rm41S::SAM_F83005C4	+	+	+	*BN406_06091*
	Rm41S::SAM_F83005C5	+	+	+	*BN406_06091*
	Rm41S::SAM_F83005C6	+	+	+	*BN406_06091*
	Rm41S::SAM_F83005C7	+	+	-	
	Rm41S::SAM_F83005C8	+	+	-	
	Rm41S::SAM_F83005C9	+	+	-	
	Rm41S::SAM_F83005C10	+	+	-	
	Rm41S::SAM_F83005C11	+	+	-	
	Rm41S::SAM_F83005C12	+	+	-	
	Rm41S::SAM_F83005C13	+	+	-	
	Rm41S::SAM_F83005C14	+	+	-	

green, bigger shoots and elongated, pink nodules. The bacteria were isolated from the Fix^+ nodules and named as Rm41S_A17MC1-7. To ensure they originated from our chemical mutant pools, the bacteria were investigated with antibiotic selection and phage test, and only 2 of them showed positive results meaning Rm41 origin. Then we re-tested these 2 Rm41 strains individually in Jemalong nodulation assay. Unfortunately, all of the plants showed chlorotic shoots and both strains failed to induce functional root nodules on Jemalong plants indicating that the Fix^+ phenotype was caused by contaminating, compatible bacteria. After inoculating Jemalong plants with ORFeome library harboring rhizobia, we found 15 Fix^+ plants (Table 4.4) and isolated bacteria from their nodules. We named these bacteria as Rm41::ORFeome_A17C1-15 and only 6 of them showed positive results in both antibiotic selection and phage test, thus, confirming their Rm41 origin. In the re-test, none of them was able to establish successful symbiosis with Jemalong plants.

These results indicate, that the ability to tolerate the NCR peptide variants produced by Jemalong plants is determined by a function, which is missing from strain Rm41 because no Fix^+ mutant could be isolated that would be indicatory for an activity preventing compatibility. The function, which is absent in strain Rm41, might be determined by not a single gene, rather by an operon, that is why, the large insert library is being introduced into strain Rm41, then, the transconjugants will be tested on Jemalong plants.

4.3.2 Identification of an Rm41 gene causing incompatibility with F83005 plants

When we screened the transposon insertion mutant populations on F83005 plants, we found 14 plants, which grew bigger and were greener compared to the others and the Rm41 inoculated individuals (Table 4.4). We checked these 14 plants and all of them developed nitrogen-fixing nodules. The bacteria isolated from the Fix^+ nodules were named as Rm41S::SAM_F83005C1-14 and strains Rm41S::SAM_F83005C1-5 came from the same mutant pool. All of these 14 strains proved to be Rm41 derivative

based on antibiotic selection and phage test. When we re-tested them in nodulation assay, 6 strains (Rm41S::SAM_F83005C1-C6) could establish an effective symbiosis with F83005 plants (Figure 4.1).

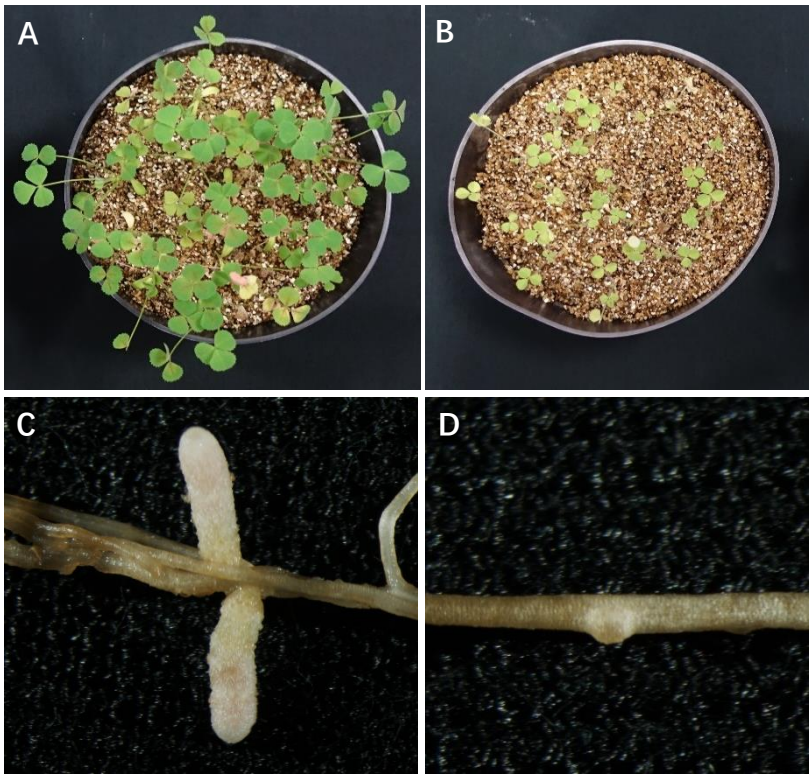

Figure 4.1 Nodulation and growth phenotypes of F83005 in re-test. A and **C**, F83005 showed Fix^+ phenotype after inoculation with Rm41S::SAM_F83005C1, evidenced by green leaves **(A)** and pink, elongated nodules **(C)** on roots. **B** and **D**, F83005 developed ineffective symbiosis with Rm41S, evidenced by yellow leaves **(B)** and nodule primordia **(D)** on roots. The phenotypes were checked four weeks after inoculation.

To find out the bacterial gene which caused incompatibility with F83005, we isolated the flanking sequence of the transposon insertions for sequencing (Figure 4.2). The

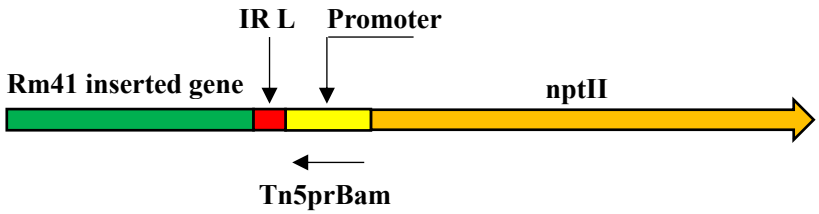

Figure 4.2 Flanking sequencing of inserted genes in Rm41S. The fragments containing *nptII* gene and part of inserted gene were digested from Rm41S::SAM_F83005C1-C6 genomic DNA, ligated into vector pSP-KS2 and sequenced using the primer (Tn5prBam) located in the *nptII* promoter.

Rm41S::SAM_F83005C1-C6 genomic DNA samples were digested with *Xba*I (one recognition site in the transposon next to the *nptII* gene)-*Eco*RI (recognition site only in the genome) and the *Xba*I and *Xba*I-*Eco*RI fragments putatively containing the *nptII* gene and part of the genome were ligated into vector pSP-KS2 and the transformants were selected for the presence of the *nptII* gene (kanamycin resistance). The obtained clones were sequenced using the primer (Tn5prBam) located in the *nptII* promoter. The sequencing results showed that all of these 6 Fix^+ mutants carried the transposon insertion in the same gene annotated as BN406_06091 (Table 4.4). In Rm41S::SAM_F83005C1-C5 mutants, the insertion sites were at the same location in the gene, while in Rm41S::SAM_F83005C6 mutant, the insertion site was at a different position (Figure 4.3). Although we inoculated all the mutant populations on F83005 plants, no insertion in other gene could restore the compatibility of strain Rm41 with this ecotype.

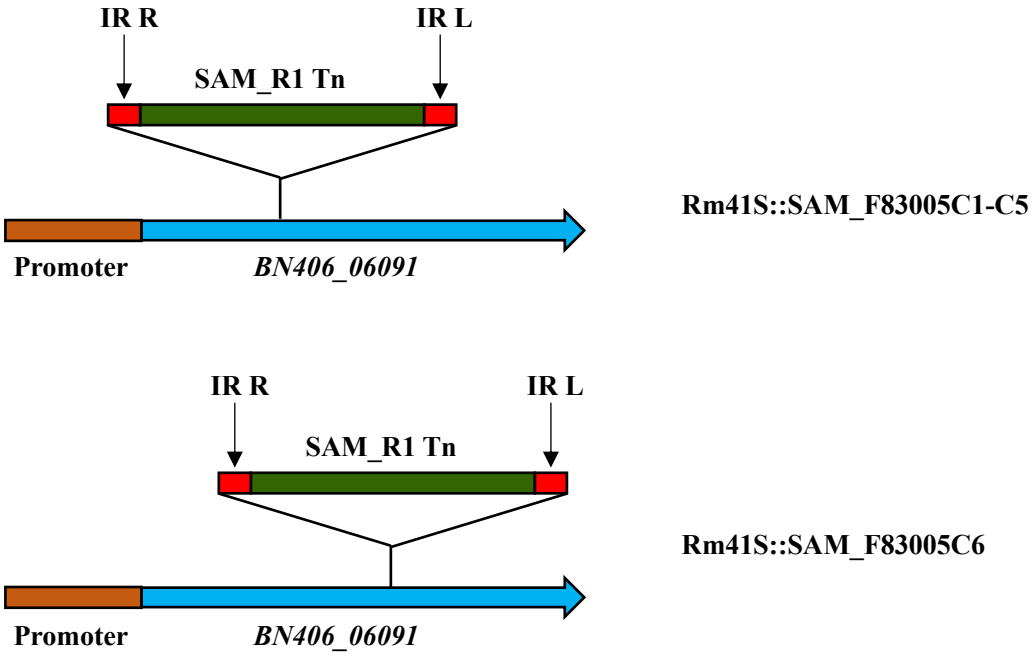

Figure 4.3 Different insertion sites of transposon insertion mutants compatible with F83005. All of 6 Fix^+ mutants were insertion mutated in the same gene, *BN406_06091*. In Rm41S::SAM_F83005C1-C5 mutants, the insertion sites were at the same location in the gene (**up**), while in Rm41S::SAM_F83005C6 mutant, the insertion site was at a different position (**below**).

Although the two independent mutations in the same gene indicated that this gene caused the incompatibility, we created a targeted *BN406_06091* mutant of Rm41S to confirm that this gene makes Rm41 strain incompatible with the F83005 plants. *BN406_06091* gene disrupted by *SpcR* cassette was introduced into pK18mobsacB (Figure 4.4) and mated into Rm41S strain by tri-parental mating. The plasmid

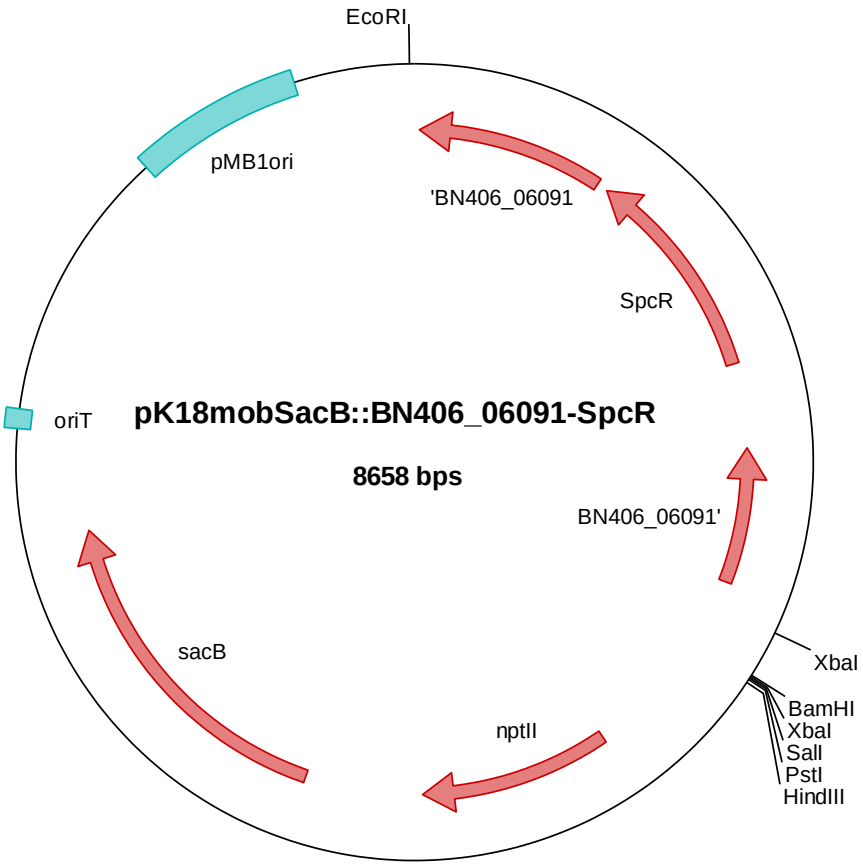

Figure 4.4 pK18mobsacB::BN406_06091-SpcR plasmid map. *BN406_06091* gene was disrupted by *SpcR* cassette and introduced into vector pK18mobsacB. *nptII* gene provided construction kanamycin resistance and *sacB* was used for 10% sucrose selection. Plasmid map was produced using Clone Manager software.

pK18mobsacB::BN406_06091-SpcR integrated into Rm41 genome via homologous recombination and those bacteria that lost the plasmid after second recombination survived in 10% sucrose selection (Figure 4.5). The putative mutant colonies retaining the *SpcR* casette after the vector loss were then checked by colony PCR and phage test. Then mutant Rm41S::*BN406_06091-SpcR* was tested in F83005 nodulation assay. Four weeks post inoculation, the plants inoculated with Rm41S::*BN406_06091-SpcR* were

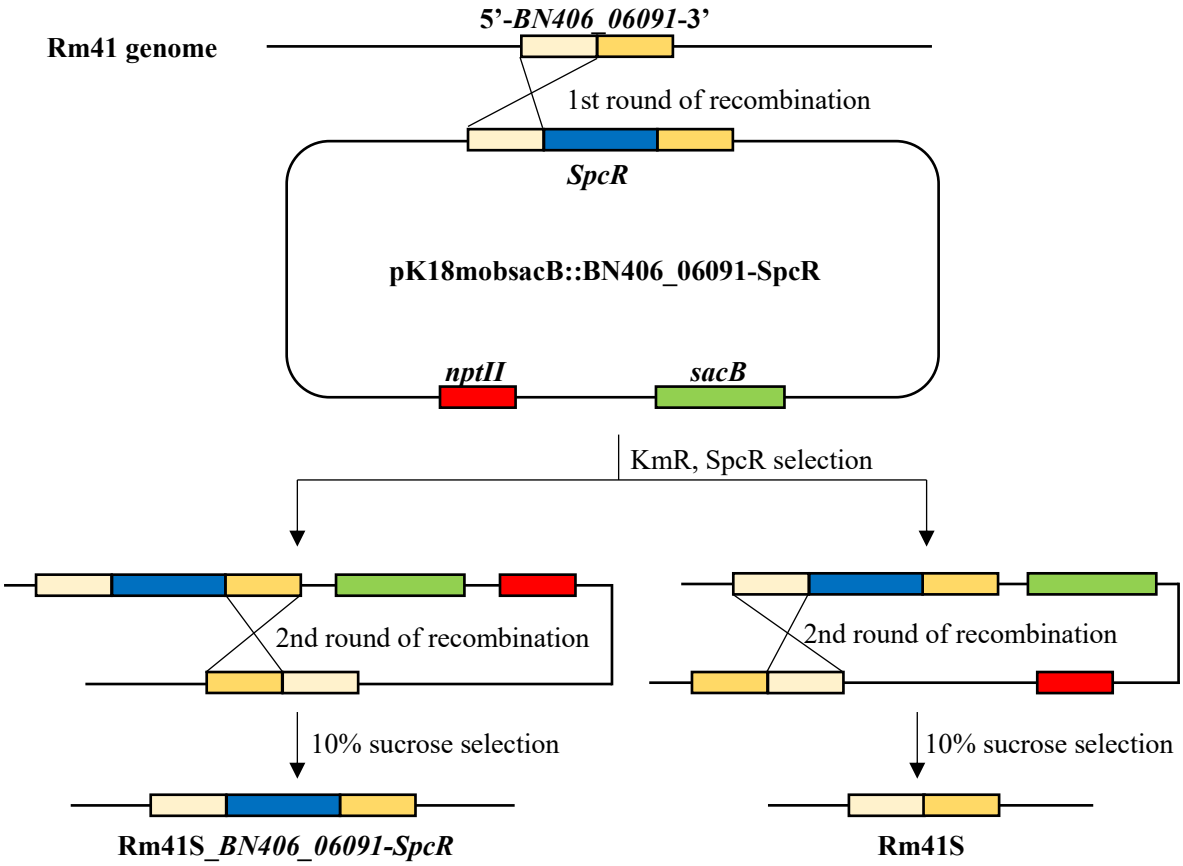

Figure 4.5 Creation of Rm41S_*BN406_06091-SpcR* by homologous recombination. The whole pK18mobsacB::BN406_06091-SpcR integrated into chromosome via first round of recombination and the recombinants were selected by KmR and SpcR. After second round of recombination and 10% sucrose selection, we obtained either mutants or wild-type Rm41S.

green, healthy and elongated, pink nodules were observed on roots (Figure 4.6), confirming that *BN406_06091* gene we found was indeed responsible for the ineffective symbiosis in the F83005-Rm41 interaction.

4.3.2.1 Characterization of the *BN406_06091* gene

The 1350 base pairs long *BN406_06091* gene is located on the second symbiotic megaplasmid. The gene encodes a protein of unknown function containing 449 amino acids. Searching for conserved protein domains using the IUPred2A (Erdős and Dosztányi, 2020) and InterPro (Blum et al., 2021) online prediction programs (https://iupred2a.elte.hu/; https://www.ebi.ac.uk/interpro/) revealed that the first half of

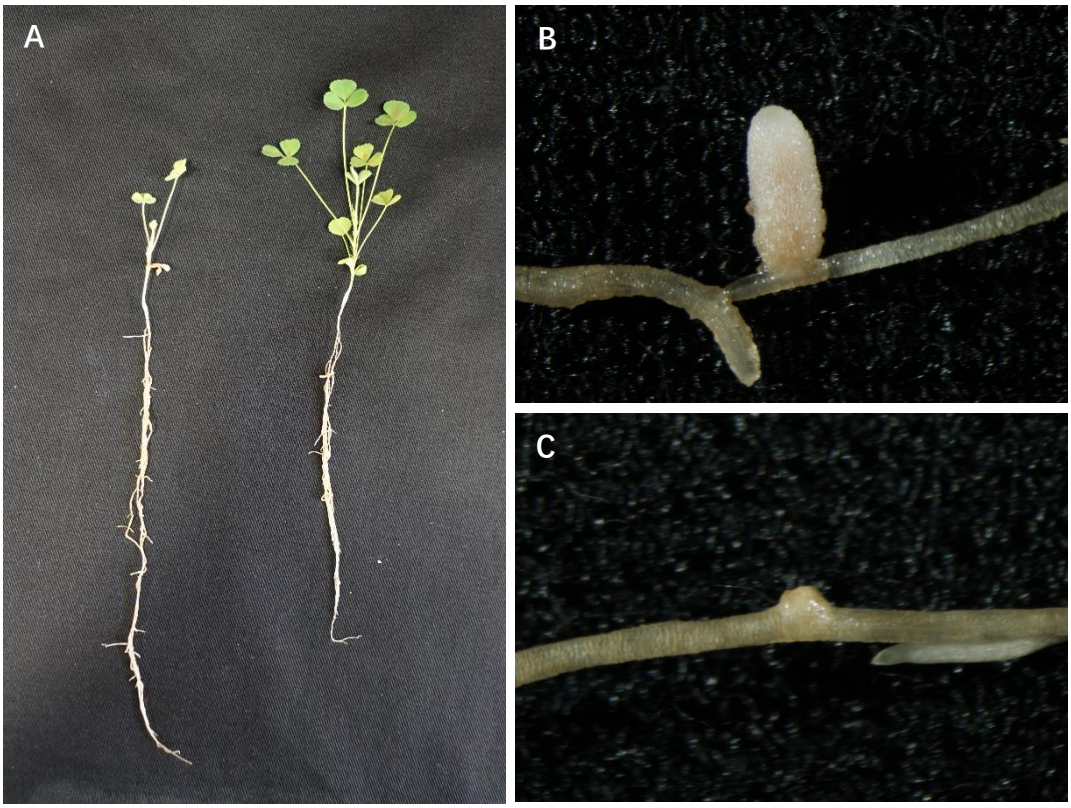

Figure 4.6 Nodulation and growth phenotypes of F83005 after inoculation with Rm41S and BN406_06091 mutant. A, F83005 inoculated with Rm41S::*BN406_06091-SpcR* (**right**) was green and well developed, whereas F83005 inoculated with wild-type Rm41S (**left**) showed nitrogen starvation phenotype. **B**, Rm41S::*BN406_06091-SpcR* induced pink and elongated nodules. **C**, Wild-type Rm41S induced nodule primordia on roots. The phenotypes were checked four weeks after inoculation.

the protein contains a right handed beta helix region, which is followed by a disordered region (Figure 4.7).

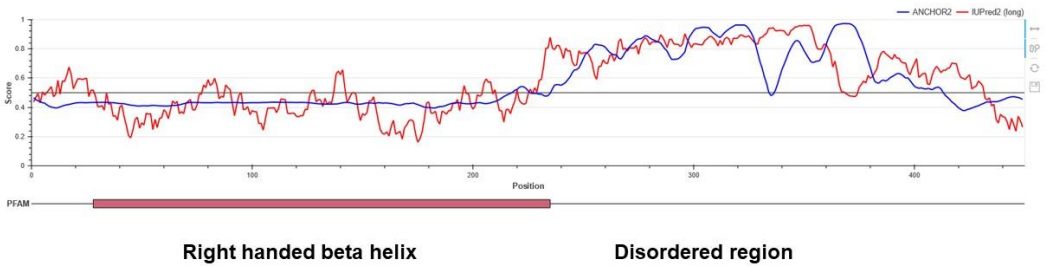

Figure 4.7 Prediction of the protein encoded by gene *BN406_06091*. The protein was predicted to contain a right handed beta helix and a disordered region. The picture was produced using the IUPred2A program.

Right handed beta helixes were first recognized in pectate lyases involved in the degradation of the pectic components (polysaccharides containing high amount of galacturonic acid) of the plant cell wall, thus, the structure prediction indicated a possible sugar/polysaccharide associated role of the protein. The disordered part might be involved in protein-protein interaction(s) when its structure might be stabilized. Neither cleavable N-terminal signal peptide nor transmembrane helices could be detected in the protein sequence by bioinformatics tools, but interestingly, the PSORTb web-server predicted an extracellular localization for the protein.

To examine the localization of the protein encoded by the *BN406_06091* gene, we created the constructs coding for BN406_06091 protein fused with mCherry at both N- and C-terminal (mCherry-06091 and 06091-mCherry) and the signal peptide of periplasmic protein EhuB fused with mCherry (EhuBsp-mCherry) as a localization control. The confocal images (Figure 4.8) showed that both of EhuB and the protein

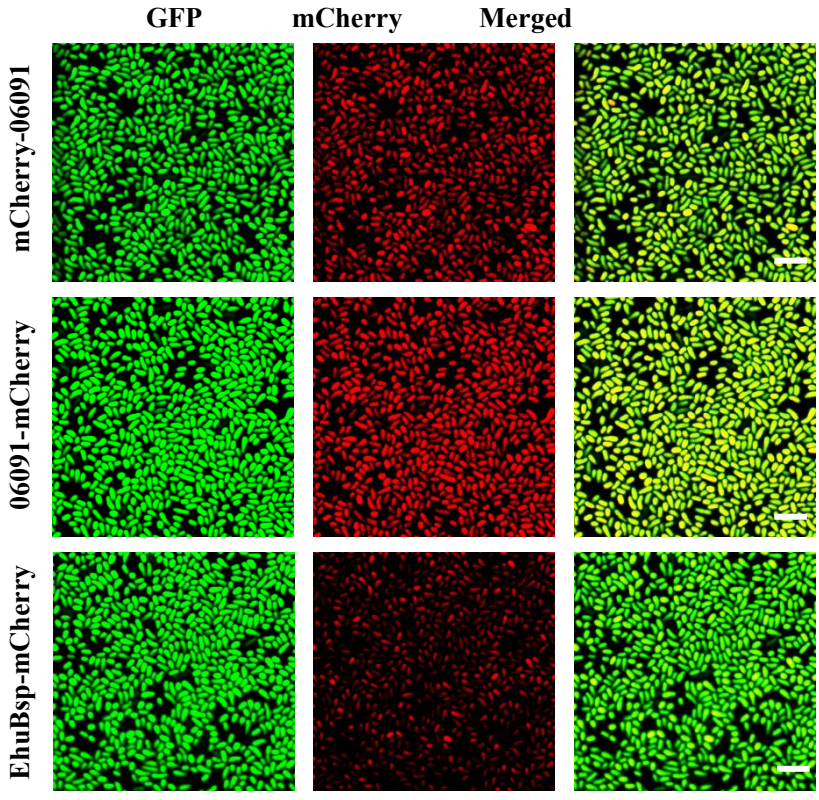

Figure 4.8 Localization of the protein encoded by gene *BN406_06091* by fluorescence analysis. Confocal images of strains Rm41S_*mBN406_06091*/pHC60::*mCherry-06091* (upper line), Rm41S_*mBN406_06091*/pHC60::*06091-mCherry* (middle line) and Rm41S_*mBN406_06091*/pHC60::*EhuBsp-mCherry* (bottom line). Bars, 10 µm.

encoded by gene *BN406_06091* were localized in cytoplasm. It surprised us that the localization of the periplasmic protein control EhuB was in cytoplasm instead of periplasmic membrane and that made us suspect if the mCherry tag or the overexpression might affect the localization.

To confirm our result from confocal microscopy, we also created the construct coding for BN406_06091 protein fused with HA tag (06091-HA) and analyzed rhizobia cell protein fractions using Western blotting with anti-GFP antibody and HA tag antibody. In the immunoblotting analysis of GFP protein (Figure 4.9), positive signals were detected in the cytoplasmic fractions of both Rm41S_*mBN406_06091*/pHC60::*GFP* (as the GFP positive control) and Rm41S_*mBN406_06091*/pHC60::*06091-HA*, suggesting that the GFP protein was localized in cytoplasm in both strains. In the immunoblotting analysis of HA tagged protein (Figure 4.9), positive signal of HA tag was detected only in the cytoplasmic fraction of Rm41S_*mBN406_06091*/pHC60::*06091-HA*. This result indicated that the protein encoded by the gene *BN406_06091* was localized in cytoplasm of Rm41 strain and was consistent with our previous result from confocal microscopy.

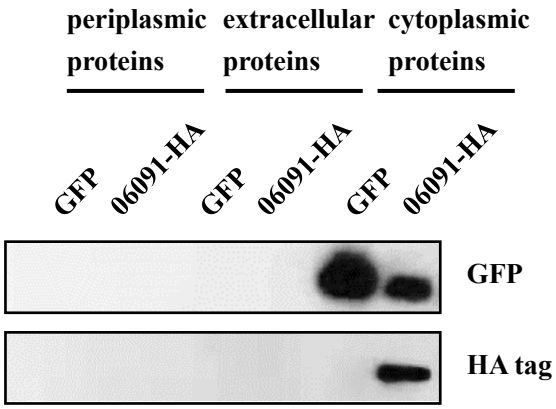

Figure 4.9 Localization of the protein encoded by gene *BN406_06091* by Western blot analysis. Periplasmic proteins (left), extracellular proteins (middle) and cytoplasmic proteins (right) isolated from strains Rm41S_*mBN406_06091*/pHC60::*GFP* and Rm41S_*mBN406_06091*/pHC60::*06091-HA* were detected using anti-GFP antibody and HA tag antibody.

We could not make a conclusion on localization of the protein encoded by gene

BN406_06091 from these fluorescence and immunological experiments, since the control protein EhuB signal peptide was not localized in periplasmic membrane as we expected. One possible reason might be that the overexpression system of pHC60 vector and/or the mCherry tag affect the expression and fate of the protein somehow. One of our future plans on this work is to place the fusion clone 06091-HA into an integrative vector and integrate them into Rm41 carrying a transposon insertion in order to drive the constructs by its native promoter.

To place the *BN406_06091* gene into a genomic context, we investigated its neighboring genes (Figure 4.10). The gene is surrounded by a number of genes coding for proteins involved in sugar and polysaccharide synthesis and modifications implicating its role in polysaccharide production. The immediate neighboring region is specific for only certain *S. meliloti* strains such as Rm41, M270, AK83 and SM11 indicating a strain-specific role. On one side of this region, the *rkp-3* gene cluster (Kereszt et al., 1998) responsible for the production of the sugar precursors as well as for the assembly and export of the strain-specific capsular polysaccharide (KPS/K-

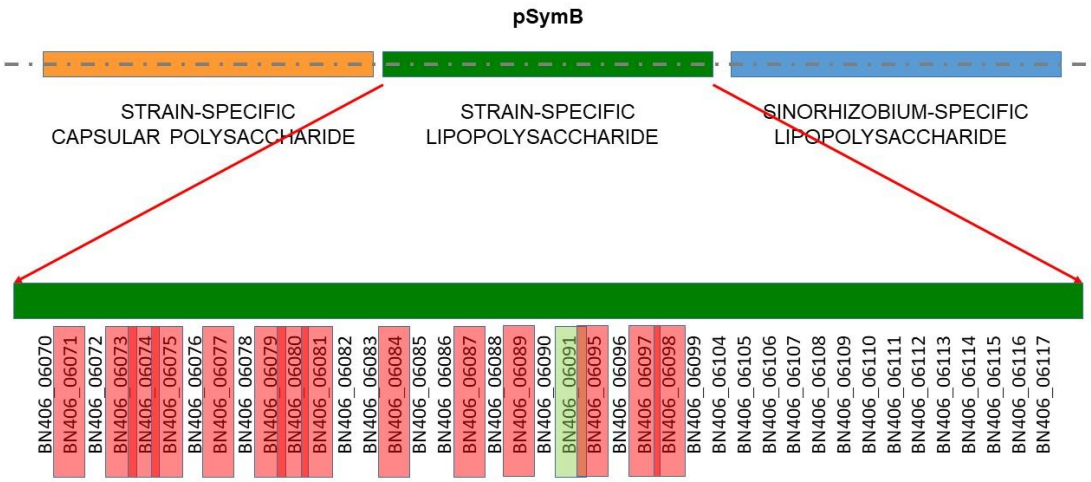

Figure 4.10 Schematic representation of the investigated part of the Rm41 pSymB megaplasmid with gene clusters potentially involved in surface polysaccharide production. In the upper part, the three regions involved in polysaccharide production are shown. In the lower part, genes in the strain-specific region coding for proteins predicted to be involved in sugar/polysaccharide synthesis and modification are listed. Selected genes mutated with plasmid integration are boxed. Green box: mutation in the gene restored compatibility. Red box: mutations could not restore compatibility.

antigen) was located (Kiss et al., 2001). On the other side of *BN406_06091* containing region, we identified a region gene cluster being present in a number of *S. meliloti* and *S. medicae* strains that contains genes coding for glycosyl-transferases, sugar activating and modifying enzymes indicating their role in polysaccharide production (Table S2).

Based on our knowledge on rhizobial surface polysaccharides, we suppose that the *BN406_06091* gene containing region is responsible for the production of the strain-specific O-antigen of LPS while the other region might contribute to the synthesis of the genus-specific LPS core oligosaccharide, which links the lipid A part in the outer membrane and the O-antigen on the surface. This idea is supported by both the phenotype observed during the interaction and by the plant proteins responsible for the incompatibility: The incompatibility between F83005 and Rm41 is due to a block of bacterial infection (Liu et al., 2014), i.e. bacteria (Nod factors) induce root hair curling and microcolony formation in the infection chamber but the plant arrests the formation of the infection threads. The incompatibility, i.e. infection restriction is caused by two receptor kinases termed NS1a and NS1b (Nodulation Specificity 1) induced by Nod factors, that contain malectin domain implicated in sugar binding in their extracellular part. According to our model, this receptor pair recognizes the surface of Rm41 as that of an “enemy” and as a result, the plant blocks the entry root. Changing or modifying or destroying the surface polysaccharide or the receptors will stop the recognition and the blocking of the infection.

Although we could not isolate other mutants establishing effective symbiosis with F83005 despite the high number of the independent insertion events, we created other mutants that might be affected in the LPS structure: 1) We targeted a number of genes in the vicinity of *BN406_06091* as well as other genes (*lpsL*, *rkpKm lpsB*) that were also implicated in LPS production that might contribute to the O-antigen or core oligosaccharide synthesis (Figure 4.10); 2) Our collaborating partners at the University of Kentucky performed binding assays with the purified extracellular domain (produced *in vitro* in a heterologous system) of the receptors and detected preference towards sulfated sugars, that is why, we mutated genes that are known or thought to be involved

in sulfate activation and sulfate modifications of sugars/polysaccharides. In *S. meliloti*, three sulfate activation systems (CysHDN, $NodP_1Q_1$, $NodP_2Q_2$) produce first adenosine 5'-phosphosulfate (APS) from ATP and sulfate ions, then the APS kinase activity synthetizes 3'-phosphoadenosine 5'-phosphosulfate (PAPS) from ATP and APS that is used in the sulfating reactions (Schwedock & Long, 1990; Roche et al., 1991; Schwedock et al., 1994; Abola et al., 1999). In addition, it was shown (Keating, 2007; Cronan and Keating, 2004) that the LpsS protein of *S. meliloti* is involved in the sulfation of its LPS, that is why, we selected these coding genes along with *BN406_02129* (encoding a hypothetical protein sulfotransferase) for mutagenesis. In one approach, we amplified internal fragments of the genes and cloned them into the pK19mob vector (Schäfer et al., 1994), then mobilized the clones into *S. meliloti* strain Rm41. As the vector cannot replicate in rhizobia, kanamycin resistant colonies can be obtained after the clones integrated into the genome via homologous recombination through the inserts. These integration events disrupt the genes and the plasmid integration mutants were used in the nodulation assays. In the case of other genes, we deleted the coding sequences using the pK18mobsacB vector and sucrose selection as described by Schäfer et al. (1994). These mutants were then tested in F83005 nodulation assay. Four weeks after inoculation, all of the plants showed Fix^- phenotype (Table 4.5), evidenced by chlorotic shoots and smaller, yellowish leaves.

We also tried to introduce the *BN406_06091* gene alone into *S. meliloti* strain Sm1021 and *S. medicae* strain WSM419 to test whether its expression in compatible bacteria can modify their cell surface to make them incompatible. Although the clone could "complement" the *BN406_06091* mutant, i.e. the transconjugant became incompatible again, the two compatible strain carrying the *BN406_06091* gene remained compatible with F83005.

Table 4.5 Nodulation phenotypes of F83005 with Rm41 mutants

Mutants	Phenotype
Rm41S_*LPSbiosynProtein-SpcR*	Fix-
Rm41S_*mExoY*	Fix-
Rm41S_*mRkpM*	Fix-
Rm41S_*mBN406_06091_mExoY*	Fix-
Rm41S_*BN406_06091-SpcR_mExoY*	Fix-
Rm41S_*mExoY_mRkpM*	Fix-
Rm41S_*BN406_06091-SpcR_mRkpM*	Fix-
Rm41S_*BN406_06091-SpcR_mRkpM_mExoY*	Fix-
Rm41S_*mLpsS*	Fix-
Rm41_*mNodPQ1*	Fix-
Rm41_*mNodPQ2*	Fix-
Rm41_*mNodH*	Fix-
Rm41_*mSulfoTr*	Fix-
Rm41S_*mLpsS_mSulfoTr*	Fix-
Rm41_*mNodPQ1_mNodPQ2*	Fix-
Rm41S_*mcysDN*	Fix-
Rm41_*mNodPQ1_mNodPQ2_mcysDN*	Fix-
Rm41S_*mNodFE*	Fix-
Rm41S_*mBN406_06071*	Fix-
Rm41S_*mBN406_06073*	Fix-
Rm41S_*mBN406_06074*	Fix-
Rm41S_*mBN406_06075*	Fix-
Rm41S_*mBN406_06077*	Fix-
Rm41S_*mBN406_06079*	Fix-
Rm41S_*mBN406_06080*	Fix-
Rm41S_*mBN406_06081*	Fix-
Rm41S_*mBN406_06084*	Fix-
Rm41S_*mLPSbiosythProtein*	Fix-
Rm41S_*mBN406_06089*	Fix-
Rm41S_*mOantigenPolymerase*	Fix-
Rm41S_*mBN406_06097*	Fix-
Rm41S_*mBN406_06098*	Fix-

It is difficult to explain all these negative results, however, there are a few ideas about how the BN406_06091 protein contributes to the incompatibility: 1) The protein is not a direct target of, i.e. not recognized by the NS1 receptor complex. Rather, it functions as a polysaccharide modifying enzyme, such as a glycanase that degrade/depolymerize/modify the polysaccharide, which is exported to the

extracellular space and is the direct activator of the NS1 receptors. In the compatible strains, the structure of the polysaccharide is so different that the molecule is not a substrate of the protein or part of the polysaccharide molecule masks/covers the protein's target. In the mutants of the region we created, the changes in the polysaccharide structure did not affect the target of the protein. 2) According to the second but less likely scenario, the protein is a direct target of the NS1 receptor complex but when it is expressed in the compatible strains, it is not exported because of the lack or not proper expression of a specific pair of type I secretion system (T1SS) proteins encoded by the *BN406_06116* and *BN406_06117* genes found in the region.

4.4 Conclusions and perspectives

As a result of this work, we have created a number of tools that are facilitating the discovery of rhizobial genes, which are involved in the determination of compatibility with the legume host, in the fine-tuning of the symbiotic interaction:

The vectors to express the coding sequences of the ORFeome library under different physiological and developmental stages allow the identification of ORFs if the lack or the mutation of individual genes inhibits the establishment or the functioning of the symbiosis. There are cases, when a missing function is determined by the concerted action of proteins encoded by genes organized into transcriptional units, operons. The large insert library we created from strain FSM-MA, which is very effective with most *M. truncatula* ecotypes tested, allows the identification of such gene clusters. We have started to use these libraries for the identification of the function missing from Rm41 to cope with the action of the *M. truncatula* cv. Jemalong type NCR peptides that eliminate bacteria from the nodule. The failure to achieve compatibility with the ORFeome library indicates that the function is encoded by more than one gene. Other members of the group started to use these resources to identify functions missing from other incompatible interactions such as those between *M. truncatula* cv. Jemalong and strains GR4 or SM11.

The libraries can be transferred into a number of strains, however, mutant populations

have to be created in all investigated strains individually. First, we created the mutant populations in strains Rm41 and FSM-MA that were used successfully to identify a function that prevents the effective interaction of strain Rm41 with ecotype F83005, but they can be used by a number of other projects. We hardly explored whether there are other ecotypes that form incompatible interaction with these strains but it would not be surprising to find such lines. Moreover, there are plant mutants in which rhizobia induce defense reactions during nodule invasion. The mutant populations will be tested with these plants whether we can identify the gene(s) that are involved in the production of the signal inducing the defense reactions. Similarly, with the help of the mutant populations, genes/proteins required for bacteriophage infection can be identified.

By studying the incompatible interaction between *M. truncatula* cv. Jemalong and strain Rm41, we revealed that the NCR peptides participate not only in the cell fate determination and directing bacteroid development but they contribute to partner selection too. Our main future task is to determine which bacterial proteins defend Rm41 bacteria from the activity of the NCR variants produced by the host Jemalong and what is their mode of action. To achieve this goal, the introduction of the large insert size genomics library into strain Rm41 was initiated.

Investigation of the incompatibility between strain Rm41 and ecotype F83005 has revealed a novel level of symbiosis specificity, the presence of a novel checkpoint for the selection of the bacterial partner by the plant host. In indeterminate nodules of *Medicago* plants, the presence of exopolysaccharides was shown to be essential for the infection process, i.e. for infection thread initiation and growth. The role of LPS in this type of interactions has been postulated to the later stages when the proper structure and thus its ability might be required for the countermeasures against the antimicrobial effects of the NCR peptides. Here, we show that plants check the LPS structure of the infecting bacteria at the initiation of the invasion process and arrests the process if it is not appropriate. This control process resembles a little bit the one described in *Lotus japonicus* for the EPS and the EPR3 receptor. EPS is not required for the infection per se, but a truncated form is worse than the complete lack of this polysaccharide. LPS is

an essential component of the outer membrane, thus, rhizobia devoid of LPS cannot be obtained but our results show that by changing most probably the structure of polysaccharide, the “enemy” may become “friend”. Although the structure of LPS is very important for this interaction, no LPS structure has been determined for any strains of *S. meliloti*. Our collaborating partners at the Complex Carbohydrate Research Center (Athens, Georgia, USA) are struggling with this task. Meantime, we try to find out that why none of the created mutants could restore the compatibility with F83005. One approach will be that we delete the whole strain-specific and alternatively the whole genus-specific LPS regions and check the phenotype of the mutants as well as that of the mutants carrying the *BN406_06091* gene. We will also explore the natural variations shown in Table S2: we identified strains with some missing genes in the region with or without a *BN406_06091* homologue and also with different core structure. We will introduce either the whole or delimited regions with or without mutation in the *BN406_06091* gene and the phenotype of the strains and their derivatives will be investigated.

However, when we reevaluated our transposon sequencing data of the four randomly selected insertion mutant pools (see Table 4.2), we realized some shortcomings of our data. If our estimation based on the number of insertions detected in the four selected pools and the specificity of the transposon insertions (between T and A nucleotides) is correct, we should have isolated much more compatible insertion mutant strains on F83005. Indeed, when we checked the transposon insertion sites from the four pools, we found 21 insertions in the gene *BN406_06091*, and five of them appeared in two pools. It means that we used too many mutant bacteria and too few plants during the screening. Thus, in the future work, we will use one pool with several hundred plants in the screening.

Summary

Nitrogen is an essential macronutrient for plants and is required for the synthesis of nucleic acids, amino acids and many other important metabolites. Although dinitrogen accounts for a large proportion (around 78%) of Earth's atmosphere, its strong chemical stability makes it inaccessible for plants.

Leguminous plants could grow in nitrogen poor soils because they enter into nitrogen-fixing symbiosis with a wide range of Gram-negative Proteobacteria, referred to as rhizobia (Haag et al., 2013). Rhizobia invade the root of legumes and are released into newly formed specialized organs, the root nodules. Within root nodules, rhizobia differentiate into bacteroids and convert atmospheric dinitrogen into ammonia, a form which is available for plants (Mergaert et al., 2006). In return, legume hosts provide proper environment and carbon sources to rhizobia for their survival (White et al., 2007).

A significant property of legume-rhizobia symbiosis is its high level of specificity. Each rhizobial species can form successful symbiosis with only a few or even single legume species, and vice versa. An additional layer of specificity also exists, when certain genotypes of the partners (wild-type bacterial strains and wild-type plant varieties often called ecotypes) that form effective symbiosis with other genotypes of the partner species, cannot establish nitrogen-fixing symbiosis with each other. Such specificity can take place at multiple stages during the nodule development and functioning and results in different phenotypes. It can occur at early stages associated with bacterial infection and nodulation, leading to Nod^- or Inf^- phenotype. For example, *Sinorhizobium meliloti* strain Rm41 could induce root hair curling on *Medicago truncatula* ecotype F83005.5 and form microcolonies in the curles but normal ITs were not detected, the nodule primordia contained no bacteria (Liu et al., 2014). Specificity also may occur at later stages of nodule development associated with nitrogen fixation, leading to Fix^- phenotype. In the interaction of *Sinorhizobium meliloti* strain Rm41 and *Medicago truncatula* Jemalong, for instance, bacteria were able to infect nodule cells and to differentiate into elongated bacteroids, however, bacteroids could not persist and

reduce nitrogen in the nodule, instead lysis and nodule senescence can be observed (Wang et al., 2017; Yang et al., 2017). Even when nitrogen-fixing symbiosis is formed successfully, the efficiency of nitrogen fixation differs between different legume-rhizobia pairings (Kazmierczak et al., 2017).

Our research group screened several ecotypes of *Medicago truncatula* using different *Sinorhizobium meliloti* and *Sinorhizobium medicae* strains as inoculants. We identified more than 10 incompatible host-strain interactions in which the host plants showed yellowish leaves and smaller shoots, characteristic phenotypes associated with nitrogen starvation. In particular, we noticed that *S. meliloti* strain Rm41 was incompatible with *Medicago truncatula* ecotypes F83005 and Jemalong, while the same strain could nodulate and fix nitrogen normally in association with other *Medicago truncatula* lines (such as A20, DZA315.16, DZA045, SA023859, …).

To identify those bacterial genes that are responsible for the ineffective symbiosis between *S. meliloti* strain Rm41 and *M. truncatula* ecotypes Jemalong and F83005, we created an ORFeome library carrying all predicted genes of *S. meliloti* strain 1021, a large insert size genomic library from *S. meliloti* strain FSM-MA, transposon insertion and chemically induced mutant populations of strain Rm41. The individual pools of the insertion mutants as well as those of the chemical mutants and the transconjugants with the ORFeome library were inoculated on Jemalong in plant nodulation assay. However, none of Rm41 derivatives was compatible with Jemalong plants, indicating that the incompatibility between Jemalong and Rm41 is determined by not a single gene, rather by an operon which is missing from Rm41. That is why, the large insert genomic library from FSM-MA is being introduced into Rm41, then, the transconjugants will be tested on Jemalong plants. When we screened the transposon insertion mutant populations on F83005 plants, we identified 6 mutants which could establish an effective symbiosis with F83005. We isolated the flanking sequence of the transposon insertions for sequencing and the results showed that all of these 6 mutants carried the transposon insertion in the same gene annotated as BN406_06091.

The gene *BN406_06091* is located on the second symbiotic megaplasmid and

encodes a protein of unknown function containing 449 amino acids. The structure prediction of the protein revealed that the first half of the protein contains a right handed beta helix domain, which is followed by a disordered region. The gene is surrounded by a number of strain-specific/genus-specific genes coding for proteins involved in sugar and polysaccharide synthesis and modifications, implicating its role in polysaccharide production. Based on our knowledge on rhizobial surface polysaccharides, we suppose that the *BN406_06091* gene containing region is responsible for the production of the strain-specific O-antigen of LPS.

Although we could not isolate other insertion mutants establishing effective symbiosis with F83005, we created other mutants that might be affected in the LPS structure. We targeted a number of genes in the vicinity of *BN406_06091*, some genes that were implicated in LPS production as well as other genes that were involved in sulfate activation and sulfate modifications of polysaccharides. These mutants were tested in F83005 nodulation assay, however, all of the plants showed Fix- phenotype.

In future work, we will try to find out why none of the created mutants could restore the compatibility with F83005. We will delete the whole strain-specific and alternatively the whole genus-specific LPS regions and check the phenotype of the mutants as well as that of the mutants carrying the *BN406_06091* gene. We will also explore the natural variations in genus-specific polysaccharide production related genes: we identified strains with some missing genes in the region with or without a *BN406_06091* homologue and also with different core structure. We will introduce either the whole or delimited regions with or without mutation in the *BN406_06091* gene and the phenotype of the strains and their derivatives will be investigated.

References

1. **Abola, A. P., Willits, M. G., Wang, R. C., & Long, S. R.** (1999). Reduction of adenosine-5′-phosphosulfate instead of 3′-phosphoadenosine-5′-phosphosulfate in cysteine biosynthesis by *Rhizobium meliloti* and other members of the family *Rhizobiaceae*. Journal of bacteriology, 181(17), 5280-5287.
2. **Alting-Mees, M. A., & Short, J. M.** (1989). pBluescript II: gene mapping vectors. Nucleic acids research, 17(22), 9494.
3. **Ané, J. M., Kiss, G. B., Riely, B. K., Penmetsa, R. V., Oldroyd, G. E., Ayax, C., Julien, L., Debellé, F., Baek, J-M., Kalo, P., Rosenberg, C., Roe, B. A., Long, S. R., Dénarié, J., & Cook, D. R.** (2004). *Medicago truncatula DMI1* required for bacterial and fungal symbioses in legumes. Science, 303(5662), 1364-1367.
4. **Antolín-Llovera, M., Ried, M. K., & Parniske, M.** (2014). Cleavage of the SYMBIOSIS RECEPTOR-LIKE KINASE ectodomain promotes complex formation with Nod factor receptor 5. Current Biology, 24(4), 422-427.
5. **Arrighi, J. F., Barre, A., Ben Amor, B., Bersoult, A., Soriano, L. C., Mirabella, R., de Carvalho-Niebel, F., Journet, E-P., Ghérardi, M., Huguet, T., Geurts, R., Dénarié, J., Rougé, P., & Gough, C.** (2006). The *Medicago truncatula* lysine motif-receptor-like kinase gene family includes *NFP* and new nodule-expressed genes. Plant physiology, 142(1), 265-279.
6. **Bender, G. L., Nayudu, M., Le Strange, K. K., & Rolfe, B. G.** (1988). The *nod D1* gene from *Rhizobium* strain NGR234 is a key determinant in the extension of host range to the nonlegume *Parasponia*. Molecular plant-microbe interactions, 1(7), 259-266.
7. **Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G. A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D. H., Letunic, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Necci, M., Orengo, C.**

A., Pandurangan, A. P., Rivoire, C., Sigrist, C. J. A., Sillitoe, I., Thanki, N., Thomas, P. D., Tosatto, S. C. E., Wu, C. H., Bateman, A., & Finn, R. D. (2021). The InterPro protein families and domains database: 20 years on. Nucleic acids research, 49(D1), D344-D354.

8. **Brewin, N. J.** (1991). Development of the legume root nodule. Annual review of cell biology, 7(1), 191-226.
9. **Broghammer, A., Krusell, L., Blaise, M., Sauer, J., Sullivan, J. T., Maolanon, N., Vinther, M., Lorentzen, A., Madsen, E. B., Jensen, K. J., Roepstorff, P., Thirup, S., Ronson, C. W., Thygesen, M. B., & Stougaard, J.** (2012). Legume receptors perceive the rhizobial lipochitin oligosaccharide signal molecules by direct binding. Proceedings of the National Academy of Sciences, 109(34), 13859-13864.
10. **Bozsoki, Z., Gysel, K., Hansen, S. B., Lironi, D., Krönauer, C., Feng, F., de Jong, N., Vinther, M., Kamble, M., Thygesen, M. B., Engholm, E., Kofoed, C., Fort, S., Sullivan, J. T., Ronson, C. W., Jensen, K. J., Blaise, M., Oldroye, G., Stougaard, J., Andersen, K. R., & Radutoiu, S.** (2020). Ligand-recognizing motifs in plant LysM receptors are major determinants of specificity. Science, 369(6504), 663-670.
11. **Canfield, D. E., Glazer, A. N., & Falkowski, P. G.** (2010). The evolution and future of Earth's nitrogen cycle. Science, 330(6001), 192-196.
12. **Cerri, M. R., Wang, Q., Stolz, P., Folgmann, J., Frances, L., Katzer, K., Li, X., Heckmann, A. B., Wang, T. L., Downie, J. A., Klingl, A., de Carvalho-Niebel, F., Xie, F., & Parniske, M.** (2017). The *ERN1* transcription factor gene is a target of the CC a MK/CYCLOPS complex and controls rhizobial infection in *Lotus japonicus*. New Phytologist, 215(1), 323-337.
13. **Carlson, R. W., Kalembasa, S., Turowski, D., Pachori, P., & Noel, K. D.** (1987). Characterization of the lipopolysaccharide from a *Rhizobium phaseoli* mutant that is defective in infection thread development. Journal of Bacteriology, 169(11), 4923-4928.

14. **Charpentier, M., Bredemeier, R., Wanner, G., Takeda, N., Schleiff, E., & Parniske, M.** (2008). *Lotus japonicus* CASTOR and POLLUX are ion channels essential for perinuclear calcium spiking in legume root endosymbiosis. The Plant Cell, 20(12), 3467-3479.
15. **Charpentier, M., & Oldroyd, G. E.** (2013). Nuclear calcium signaling in plants. Plant Physiology, 163(2), 496-503.
16. **Charpentier, M., Sun, J., Martins, T. V., Radhakrishnan, G. V., Findlay, K., Soumpourou, E., Thouin, J., Véry, A-A., Sanders, D., Morris, R. J., & Oldroyd, G. E.** (2016). Nuclear-localized cyclic nucleotide–gated channels mediate symbiotic calcium oscillations. Science, 352(6289), 1102-1105.
17. **Cheng, H. P., & Walker, G. C.** (1998). Succinoglycan is required for initiation and elongation of infection threads during nodulation of alfalfa by *Rhizobium meliloti*. Journal of bacteriology, 180(19), 5183-5191.
18. **Crespi, M., & Frugier, F.** (2008). De novo organ formation from differentiated cells: root nodule organogenesis. Science Signaling, 1(49), re11-re11.
19. **Cronan, G. E., & Keating, D. H.** (2004). *Sinorhizobium meliloti* sulfotransferase that modifies lipopolysaccharide. Journal of bacteriology, 186(13), 4168-4176.
20. **Crook, M. B., Lindsay, D. P., Biggs, M. B., Bentley, J. S., Price, J. C., Clement, S. C., Long, S. R., & Griffitts, J. S.** (2012). Rhizobial plasmids that cause impaired symbiotic nitrogen fixation and enhanced host invasion. Molecular plant-microbe interactions, 25(8), 1026-1033.
21. **Cui, H., Tsuda, K., & Parker, J. E.** (2015). Effector-triggered immunity: from pathogen perception to robust defense. Annual review of plant biology, 66, 487-511.
22. **de Maagd, R. A., & Lugtenberg, B.** (1986). Fractionation of *Rhizobium leguminosarum* cells into outer membrane, cytoplasmic membrane, periplasmic, and cytoplasmic components. Journal of bacteriology, 167(3), 1083-1085.
23. **Dénarié, J., Debellé, F., & Promé, J. C.** (1996). *Rhizobium* lipo-

chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. Annual review of biochemistry, 65(1), 503-535.

24. **Devon, R. S., Porteous, D. J., & Brookes, A. J.** (1995). Splinkerettes--improved vectorettes for greater efficiency in PCR walking. Nucleic acids research, 23(9), 1644.
25. **Downie, J. A.** (1998). Functions of rhizobial nodulation genes. In The Rhizobiaceae (pp. 387-402). Springer, Dordrecht.
26. **Downie, J. A.** (2010). The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. FEMS microbiology reviews, 34(2), 150-170.
27. **Dupont, L., Hérouart, D., Alloing, G., Hopkins, J., Pierre, O., Frendo, P., & El Msehli, S.** (2012). The legume root nodule: from symbiotic nitrogen fixation to senescence (pp. 137-157). London: INTECH Open Access Publisher.
28. **Endre, G., Kereszt, A., Kevei, Z., Mihacea, S., Kaló, P., & Kiss, G. B.** (2002). A receptor kinase gene regulating symbiotic nodule development. Nature, 417(6892), 962-966.
29. **Erdős, G., & Dosztányi, Z.** (2020). Analyzing protein disorder with IUPred2A. Current protocols in bioinformatics, 70(1), e99.
30. **Esseling, J. J., Lhuissier, F. G., & Emons, A. M. C.** (2003). Nod factor-induced root hair curling: continuous polar growth towards the point of nod factor application. Plant physiology, 132(4), 1982-1988.
31. **Fan, Y., Liu, J., Lyu, S., Wang, Q., Yang, S., & Zhu, H.** (2017). The soybean *Rfg1* gene restricts nodulation by *Sinorhizobium fredii* USDA193. Frontiers in plant science, 8, 1548.
32. **Faucher, C., Camut, S., Dénarié, J., & Truchet, G.** (1989). The *nodH* and *nodQ* host range genes of *Rhizobium meliloti* behave as avirulence genes in *R. leguminosarum* bv. *viciae* and determine changes in the production of plant-specific extracellular signals. Mol. Plant-Microbe Interact, 2(6), 291-300.
33. **Figurski, D. H., & Helinski, D. R.** (1979). Replication of an origin-containing

derivative of plasmid RK2 dependent on a plasmid function provided in trans. Proceedings of the National Academy of Sciences, 76(4), 1648-1652.

34. **Finan, T. M., Hirsch, A. M., Leigh, J. A., Johansen, E., Kuldau, G. A., Deegan, S., Walker, G. C., & Signer, E. R.** (1985). Symbiotic mutants of *Rhizobium meliloti* that uncouple plant from bacterial differentiation. Cell, 40(4), 869-877.
35. **Franssen, H. J., Vijn, I., Yang, W. C., & Bisseling, T.** (1992). Developmental aspects of the *Rhizobium*-legume symbiosis. Plant molecular biology, 19(1), 89-107.
36. **Felle, H. H., Kondorosi, É., Kondorosi, Á., & Schultze, M.** (1998). The role of ion fluxes in Nod factor signalling in *Medicago sativa*. The Plant Journal, 13(4), 455-463.
37. **Gage, D. J.** (2004). Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. Microbiology and molecular biology reviews, 68(2), 280-300.
38. **Geddes, B. A., Kearsley, J. V., Huang, J., Zamani, M., Muhammed, Z., Sather, L., Panchal, A. K., diCenzo, G. C., & Finan, T. M.** (2021). Minimal gene set from *Sinorhizobium* (*Ensifer*) *meliloti* pSymA required for efficient symbiosis with *Medicago*. Proceedings of the National Academy of Sciences, 118(2).
39. **Geurts, R., Fedorova, E., & Bisseling, T.** (2005). Nod factor signaling genes and their function in the early stages of *Rhizobium* infection. Current opinion in plant biology, 8(4), 346-352.
40. **Gibson, K. E., Kobayashi, H., & Walker, G. C.** (2008). Molecular determinants of a symbiotic chronic infection. Annual review of genetics, 42, 413-441.
41. **Glazebrook, J., Ichige, A., & Walker, G. C.** (1993). A *Rhizobium meliloti* homolog of the *Escherichia coli* peptide-antibiotic transport protein SbmA is essential for bacteroid development. Genes & development, 7(8), 1485-1497.

42. **Glazebrook, J., Ichige, A., & Walker, G. C.** (1996). Genetic analysis of *Rhizobium meliloti bacA-phoA* fusion results in identification of *degP*: two loci required for symbiosis are closely linked to *degP*. Journal of bacteriology, 178(3), 745-752.
43. **Gonzalez-Rizzo, S., Crespi, M., & Frugier, F.** (2006). The *Medicago truncatula* CRE1 cytokinin receptor regulates lateral root development and early symbiotic interaction with *Sinorhizobium meliloti*. The Plant Cell, 18(10), 2680-2693.
44. **Göttfert, M., Horvath, B., Kondorosi, E., Putnoky, P., Rodriguez-Quiñones, F., & Kondorosi, A.** (1986). At least two *nodD* genes are necessary for efficient nodulation of alfalfa by *Rhizobium meliloti*. Journal of molecular biology, 191(3), 411-420.
45. **Györgypal, Z., Kiss, G. B., & Kondorosi, A.** (1991). Transduction of plant signal molecules by the *Rhizobium* NodD proteins. BioEssays, 13(11), 575-581.
46. **Haag, A. F., Arnold, M. F. F., Myka, K. K., Kerscher, B., Dall'Angelo, S., Zanda, M., Mergaert, P., & Ferguson, G. P.** (2013). Molecular insights into bacteroid development during *Rhizobium*–legume symbiosis. FEMS microbiology reviews, 37(3), 364-383.
47. **Hanekop, N., Höing, M., Sohn-Bösser, L., Jebbar, M., Schmitt, L., & Bremer, E.** (2007). Crystal structure of the ligand-binding protein EhuB from *Sinorhizobium meliloti* reveals substrate recognition of the compatible solutes ectoine and hydroxyectoine. Journal of molecular biology, 374(5), 1237-1250.
48. **Hellens, R. P., Edwards, E. A., Leyland, N. R., Bean, S., & Mullineaux, P. M.** (2000). pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. Plant molecular biology, 42(6), 819-832.
49. **Hirsch, A. M.** (1992). Developmental biology of legume nodulation. New Phytologist, 122(2), 211-237.
50. **Hirsch, S., Kim, J., Muñoz, A., Heckmann, A. B., Downie, J. A., & Oldroyd, G. E.** (2009). GRAS proteins form a DNA binding complex to induce gene

expression during nodulation signaling in *Medicago truncatula*. The Plant Cell, 21(2), 545-557.

51. **Horváth, B., Domonkos, Á., Kereszt, A., Szűcs, A., Ábrahám, E., Ayaydin, F., Bóka, K., Chen, Y., Chen, R., Murray, J. D., Udvardi, M. K., Kondorosi, É., & Kaló, P.** (2015). Loss of the nodule-specific cysteine rich peptide, NCR169, abolishes symbiotic nitrogen fixation in the *Medicago truncatula dnf7* mutant. Proceedings of the National Academy of Sciences, 112(49), 15232-15237.
52. **House, B. L., Mortimer, M. W., & Kahn, M. L.** (2004). New recombination methods for *Sinorhizobium meliloti* genetics. Applied and environmental microbiology, 70(5), 2806-2815.
53. **Imaizumi-Anraku, H., Takeda, N., Charpentier, M., Perry, J., Miwa, H., Umehara, Y., Kouchi, H., Murakami, Y., Mulder, L., Vickers, K., Pike, J., Downie, J. A., Wang, T., Sato, S., Asamizu, E., Tabata, S., Yoshikawa, M., Murooka, Y., Wu, G-J., Kawaguchi, M., Kawasaki, S., Parniske, M., & Hayashi, M.** (2005). Plastid proteins crucial for symbiotic fungal and bacterial entry into plant roots. Nature, 433(7025), 527-531.
54. **Jones, J. D., & Dangl, J. L.** (2006). The plant immune system. nature, 444(7117), 323-329.
55. **Jones, K. M.** (2012). Increased production of the exopolysaccharide succinoglycan enhances *Sinorhizobium meliloti* 1021 symbiosis with the host plant *Medicago truncatula*. Journal of bacteriology, 194(16), 4322-4331.
56. **Jones, K. M., Kobayashi, H., Davies, B. W., Taga, M. E., & Walker, G. C.** (2007). How rhizobial symbionts invade plants: the *Sinorhizobium–Medicago* model. Nature Reviews Microbiology, 5(8), 619-633.
57. **Kaló, P., Gleason, C., Edwards, A., Marsh, J., Mitra, R. M., Hirsch, S., Jakab, J., Sims, S., Long, S. R., Rogers, J., Kiss, G. B., Downie, J. A., & Oldroyd, G. E.** (2005). Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. Science, 308(5729),

1786-1789.

58. **Kanamori, N., Madsen, L. H., Radutoiu, S., Frantescu, M., Quistgaard, E. M., Miwa, H., Downie, J. A., James, E. K., Felle, H. H., Haaning, L. L., Jensen, T. H., Sato, S., Nakamura, Y., Tabata, S., Sandal, N., & Stougaard, J.** (2006). A nucleoporin is required for induction of Ca^{2+} spiking in legume nodule development and essential for rhizobial and fungal symbiosis. Proceedings of the National Academy of Sciences, 103(2), 359-364.
59. **Kannenberg, E. L., & Carlson, R. W.** (2001). Lipid A and O-chain modifications cause *Rhizobium* lipopolysaccharides to become hydrophobic during bacteroid development. Molecular microbiology, 39(2), 379-392.
60. **Kawaharada, Y., James, E. K., Kelly, S., Sandal, N., & Stougaard, J.** (2017a). The ethylene responsive factor required for nodulation 1 (ERN1) transcription factor is required for infection-thread formation in *Lotus japonicus*. Molecular plant-microbe interactions, 30(3), 194-204.
61. **Kawaharada, Y., Kelly, S., Nielsen, M. W., Hjuler, C. T., Gysel, K., Muszyński, A., Carlson, R. W., Thygesen, M. B., Sandal, N., Asmussen, M. H., Vinther, M., Andersen, S. U., Krusell, L., Thirup, S., Jensen, K. J., Ronson, C. W., Blaise, M., Radutoiu, S., & Stougaard, J.** (2015). Receptor-mediated exopolysaccharide perception controls bacterial infection. Nature, 523(7560), 308-312.
62. **Kawaharada, Y., Nielsen, M. W., Kelly, S., James, E. K., Andersen, K. R., Rasmussen, S. R., Füchtbauer, W., Madsen, L. H., Heckmann, A. B., Radutoiu, S., & Stougaard, J.** (2017b). Differential regulation of the *Epr3* receptor coordinates membrane-restricted rhizobial colonization of root nodule primordia. Nature communications, 8(1), 1-11.
63. **Kazmierczak, T., Nagymihály, M., Lamouche, F., Barrière, Q., Guefrachi, I., Alunni, B., Ouadghiri, M., Ibijbijen, J., Kondorosi, É., Mergaert, P., & Gruber, V.** (2017). Specific host-responsive associations between Medicago truncatula accessions and Sinorhizobium strains. Molecular Plant-Microbe

Interactions, 30(5), 399-409.

64. **Keating, D. H.** (2007). The *Sinorhizobium meliloti* ExoR protein is required for the downregulation of *lpsS* transcription and succinoglycan biosynthesis in response to divalent cations. FEMS microbiology letters, 267(1), 23-29.
65. **Kereszt, A., Kiss, E., Reuhs, B. L., Carlson, R. W., Kondorosi, Á., & Putnoky, P.** (1998). Novel *rkp* gene clusters of *Sinorhizobium meliloti* involved in capsular polysaccharide production and invasion of the symbiotic nodule: the *rkpK* gene encodes a UDP-glucose dehydrogenase. Journal of Bacteriology, 180(20), 5426-5431.
66. **Kereszt, A., Slaska-Kiss, K., Putnoky, P., Banfalvi, Z., & Kondorosi, A.** (1995). The *cycHJKL* genes of *Rhizobium meliloti* involved in cytochrome *c* biogenesis are required for "respiratory" nitrate reduction *ex planta* and for nitrogen fixation during symbiosis. Molecular and General Genetics MGG, 247(1), 39-47.
67. **Kim, M., Chen, Y., Xi, J., Waters, C., Chen, R., & Wang, D.** (2015). An antimicrobial peptide essential for bacterial survival in the nitrogen-fixing symbiosis. Proceedings of the National Academy of Sciences, 112(49), 15238-15243.
68. **Kiss, E., Kereszt, A., Barta, F., Stephens, S., Reuhs, B. L., Kondorosi, Á., & Putnoky, P.** (2001). The *rkp-3* gene region of *Sinorhizobium meliloti* Rm41 contains strain-specific genes that determine K antigen structure. Molecular plant-microbe interactions, 14(12), 1395-1403.
69. **Kondorosi, E., Gyuris, J., Schmidt, J., John, M., Duda, E., Hoffmann, B., Schell, J., & Kondorosi, A.** (1989). Positive and negative control of *nod* gene expression in *Rhizobium meliloti* is required for optimal nodulation. The EMBO journal, 8(5), 1331-1340.
70. **Kondorosi, E., Roudier, F., & Gendreau, E.** (2000). Plant cell-size control: growing by ploidy?. Current opinion in plant biology, 3(6), 488-492.
71. **Kouchi, H., Imaizumi-Anraku, H., Hayashi, M., Hakoyama, T., Nakagawa,**

T., Umehara, Y., Suganuma, N., & Kawaguchi, M. (2010). How many peas in a pod? Legume genes responsible for mutualistic symbioses underground. Plant and Cell Physiology, 51(9), 1381-1397.

72. **Lagares, A., Caetano-Anollés, G., Niehaus, K., Lorenzen, J., Ljunggren, H. D., Pühler, A., & Favelukes, G.** (1992). A *Rhizobium meliloti* lipopolysaccharide mutant altered in competitiveness for nodulation of alfalfa. Journal of bacteriology, 174(18), 5941-5952.
73. **Larrainzar, E., Villar, I., Rubio, M. C., Pérez-Rontomé, C., Huertas, R., Sato, S., Mun, J. H., & Becana, M.** (2020). Hemoglobins in the legume–*Rhizobium* symbiosis. New Phytologist, 228(2), 472-484.
74. **Leigh, J. A., Signer, E. R., & Walker, G. C.** (1985). Exopolysaccharide-deficient mutants of *Rhizobium meliloti* that form ineffective nodules. Proceedings of the National Academy of Sciences, 82(18), 6231-6235.
75. **Lerouge, P., Roche, P., Faucher, C., Maillet, F., Truchet, G., Promé, J. C., & Dénarié, J.** (1990). Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. Nature, 344(6268), 781-784.
76. **Lévy, J., Bres, C., Geurts, R., Chalhoub, B., Kulikova, O., Duc, G., Journet, E-P., Ané, J-M., Lauber, E., Bisseling, T., Dénarié, J., Rosenberg, C., & Debellé, F.** (2004). A putative Ca^{2+} and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. Science, 303(5662), 1361-1364.
77. **Limpens, E., Franken, C., Smit, P., Willemse, J., Bisseling, T., & Geurts, R.** (2003). LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. Science, 302(5645), 630-633.
78. **Limpens, E., Mirabella, R., Fedorova, E., Franken, C., Franssen, H., Bisseling, T., & Geurts, R.** (2005). Formation of organelle-like N2-fixing symbiosomes in legume root nodules is controlled by *DMI2*. Proceedings of the National Academy of Sciences, 102(29), 10375-10380.
79. **Liu, C. W., & Murray, J. D.** (2016). The role of flavonoids in nodulation host-

range specificity: an update. Plants, 5(3), 33.

80. **Liu, J., Yang, S., Zheng, Q., & Zhu, H.** (2014). Identification of a dominant gene in *Medicago truncatula* that restricts nodulation by *Sinorhizobium meliloti* strain Rm41. BMC plant biology, 14(1), 1-9.
81. **Long, S. R.** (1996). *Rhizobium* symbiosis: Nod factors in perspective. The Plant Cell, 8(10), 1885.
82. **López-Lara, I. M., Blok-Tip, L., Quinto, C., Garcia, M. L., Stacey, G., Bloemberg, G. V., Lamers, G. E. M., Ligtenberg, B. J. J., Thomas-Oates, J. E., & Spaink, H. P.** (1996). NodZ of *Bradyrhizobium* extends the nodulation host range of *Rhizobium* by adding a fucosyl residue to nodulation signals. Molecular microbiology, 21(2), 397-408.
83. **Madsen, E. B., Madsen, L. H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., Sato, S., Kaneko, T., Tabata, S., Sandal, N., & Stougaard, J.** (2003). A receptor kinase gene of the LysM type is involved in legumeperception of rhizobial signals. Nature, 425(6958), 637-640.
84. **Maillet, F., Fournier, J., Mendis, H. C., Tadege, M., Wen, J., Ratet, P., Mysore, K. S., Gough, C., & Jones, K. M.** (2020). *Sinorhizobium meliloti* succinylated high-molecular-weight succinoglycan and the *Medicago truncatula* LysM receptor-like kinase MtLYK10 participate independently in symbiotic infection. The Plant Journal, 102(2), 311-326.
85. **Marsh, J. F., Rakocevic, A., Mitra, R. M., Brocard, L., Sun, J., Eschstruth, A., Long, S. R., Schultze, M., Ratet, P., & Oldroyd, G. E.** (2007). *Medicago truncatula NIN* is essential for rhizobial-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase. Plant physiology, 144(1), 324-335.
86. **McIver, J., Djordjevic, M. A., Weinman, J. J., Bender, G. L., & Rolfe, B. G.** (1989). Extension of host range of *Rhizobium leguminosarum* bv. *trifolii* caused by point mutations in *nodD* that result in alterations in regulatory function and recognition of inducer molecules. Mol. Plant-Microbe Interact, 2(3), 97-106.

87. **Mergaert, P., Kereszt, A., & Kondorosi, E.** (2020). Gene expression in nitrogen-fixing symbiotic nodule cells in *Medicago truncatula* and other nodulating plants. The Plant Cell, 32(1), 42-68.

88. **Mergaert, P., Uchiumi, T., Alunni, B., Evanno, G., Cheron, A., Catrice, O., Mausset, A-E., Barloy-Hubler, F., Galibert, F., Kondorosi, A., & Kondorosi, E.** (2006). Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*–legume symbiosis. Proceedings of the National Academy of Sciences, 103(13), 5230-5235.

89. **Messinese, E., Mun, J. H., Yeun, L. H., Jayaraman, D., Rougé, P., Barre, A., Lougnon, G., Schornack, S., Bono, J. J., Cook, D. R., & Ané, J. M.** (2007). A novel nuclear protein interacts with the symbiotic DMI3 calcium-and calmodulin-dependent protein kinase of *Medicago truncatula*. Molecular plant-microbe interactions, 20(8), 912-921.

90. **Middleton, P. H., Jakab, J., Penmetsa, R. V., Starker, C. G., Doll, J., Kaló, P., Prabhu, R., Marsh, J. F., Mitra, R. M., Kereszt, A., Dudas, B., VandenBosch, K., Long, S. R., Cook, D. R., Kiss, G. B., & Oldroyd, G. E.** (2007). An ERF transcription factor in *Medicago truncatula* that is essential for Nod factor signal transduction. The Plant Cell, 19(4), 1221-1234.

91. **Moling, S., Pietraszewska-Bogiel, A., Postma, M., Fedorova, E., Hink, M. A., Limpens, E., Gadella, T. W. J., & Bisseling, T.** (2014). Nod factor receptors form heteromeric complexes and are essential for intracellular infection in *Medicago* nodules. The Plant Cell, 26(10), 4188-4199.

92. **Murray, J. D., Karas, B. J., Sato, S., Tabata, S., Amyot, L., & Szczyglowski, K.** (2007). A cytokinin perception mutant colonized by *Rhizobium* in the absence of nodule organogenesis. Science, 315(5808), 101-104.

93. **Niehaus, K., Lagares, A., & Pühler, A.** (1998). A *Sinorhizobium meliloti* lipopolysaccharide mutant induces effective nodules on the host plant *Medicago sativa* (alfalfa) but fails to establish a symbiosis with *Medicago truncatula*. Molecular plant-microbe interactions, 11(9), 906-914.

94. **Nguyen, H. P., Miwa, H., Kaneko, T., Sato, S., & Okazaki, S.** (2017). Identification of *Bradyrhizobium elkanii* genes involved in incompatibility with *Vigna radiata*. Genes, 8(12), 374.

95. **Nguyen, H. P., Ratu, S. T., Yasuda, M., Göttfert, M., & Okazaki, S.** (2018). InnB, a novel type III effector of *Bradyrhizobium elkanii* USDA61, controls symbiosis with *Vigna* species. Frontiers in microbiology, 9, 3155.

96. **Oldroyd, G. E., Murray, J. D., Poole, P. S., & Downie, J. A.** (2011). The rules of engagement in the legume-rhizobial symbiosis. Annual review of genetics, 45, 119-144.

97. **Oono, R., Schmitt, I., Sprent, J. I., & Denison, R. F.** (2010). Multiple evolutionary origins of legume traits leading to extreme rhizobial differentiation. New Phytologist, 187(2), 508-520.

98. **Pankhurst, C. E., & Biggs, D. R.** (1980). Sensitivity of *Rhizobium* to selected isoflavonoids. Canadian Journal of Microbiology, 26(4), 542-545.

99. **Parniske, M., Ahlborn, B., & Werner, D.** (1991). Isoflavonoid-inducible resistance to the phytoalexin glyceollin in soybean rhizobia. Journal of Bacteriology, 173(11), 3432-3439.

100. **Peck, M. C., Fisher, R. F., & Long, S. R.** (2006). Diverse flavonoids stimulate NodD1 binding to *nod* gene promoters in *Sinorhizobium meliloti*. Journal of bacteriology, 188(15), 5417-5427.

101. **Perry, B. J., & Yost, C. K.** (2014). Construction of a mariner-based transposon vector for use in insertion sequence mutagenesis in selected members of the *Rhizobiaceae*. BMC microbiology, 14(1), 1-11.

102. **Peters, N. K., Frost, J. W., & Long, S. R.** (1986). A plant flavone, luteolin, induces expression of *Rhizobium meliloti* nodulation genes. Science, 233(4767), 977-980.

103. **Plet, J., Wasson, A., Ariel, F., Le Signor, C., Baker, D., Mathesius, U., Crespi, M., & Frugier, F.** (2011). MtCRE1-dependent cytokinin signaling integrates bacterial and plant cues to coordinate symbiotic nodule organogenesis

in *Medicago truncatula*. The Plant Journal, 65(4), 622-633.

104. **Pueppke, S. G., & Broughton, W. J.** (1999). *Rhizobium* sp. strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges. Molecular Plant-Microbe Interactions, 12(4), 293-318.

105. **Radutoiu, S., Madsen, L. H., Madsen, E. B., Felle, H. H., Umehara, Y., Grønlund, M., Sato, S., Nakamura, Y., Tabata, S., Sandal, N., & Stougaard, J.** (2003). Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. Nature, 425(6958), 585-592.

106. **Radutoiu, S., Madsen, L. H., Madsen, E. B., Jurkiewicz, A., Fukai, E., Quistgaard, E. M., Albrektsen, A. S., James, E. K., Thirup, S., & Stougaard, J.** (2007). LysM domains mediate lipochitin–oligosaccharide recognition and *Nfr* genes extend the symbiotic host range. The EMBO journal, 26(17), 3923-3935.

107. **Redmond, J. W., Batley, M., Djordjevic, M. A., Innes, R. W., Kuempel, P. L., & Rolfe, B. G.** (1986). Flavones induce expression of nodulation genes in *Rhizobium*. Nature, 323(6089), 632-635.

108. **Reinhold, B. B., Chan, S. Y., Reuber, T. L., Marra, A., Walker, G. C., & Reinhold, V. N.** (1994). Detailed structural characterization of succinoglycan, the major exopolysaccharide of *Rhizobium meliloti* Rm1021. Journal of bacteriology, 176(7), 1997-2002.

109. **Roche, P., Debellé, F., Maillet, F., Lerouge, P., Faucher, C., Truchet, G., Dénarié, J., & Promé, J. C.** (1991). Molecular basis of symbiotic host specificity in Rhizobium meliloti: *nodH* and *nodPQ* genes encode the sulfation of lipo-oligosaccharide signals. Cell, 67(6), 1131-1143.

110. **Roux, B., Rodde, N., Jardinaud, M. F., Timmers, T., Sauviac, L., Cottret, L., Carrère, S., Sallet, E., Courcelle, E., Moreau, S., Debellé, F., Capela, D., de Carvalho-Niebel, F., Gouzy, J., Bruand, C., & Gamas, P.** (2014). An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. The

Plant Journal, 77(6), 817-837.

111. **Roy, S., Liu, W., Nandety, R. S., Crook, A., Mysore, K. S., Pislariu, C. I., Frugoli, J., Dickstein, R., & Udvardi, M. K.** (2020). Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. The Plant Cell, 32(1), 15-41.

112. **Saito, K., Yoshikawa, M., Yano, K., Miwa, H., Uchida, H., Asamizu, E., Sato, S., Tabata, S., Imaizumi-Anraku, H., Umehara, Y., Kouchi, H., Murooka, Y., Szczyglowski, K., Downie, J. A., Parniske, M., Hayashi, M., & Kawaguchi, M.** (2007). NUCLEOPORIN85 is required for calcium spiking, fungal and bacterial symbioses, and seed production in *Lotus japonicus*. The Plant Cell, 19(2), 610-624.

113. **Schäfer, A., Tauch, A., Jäger, W., Kalinowski, J., Thierbach, G., & Pühler, A.** (1994). Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. Gene, 145(1), 69-73.

114. **Schauser, L., Roussis, A., Stiller, J., & Stougaard, J.** (1999). A plant regulator controlling development of symbiotic root nodules. Nature, 402(6758), 191-195.

115. **Schroeder, B. K., House, B. L., Mortimer, M. W., Yurgel, S. N., Maloney, S. C., Ward, K. L., & Kahn, M. L.** (2005). Development of a functional genomics platform for *Sinorhizobium meliloti*: construction of an ORFeome. Applied and environmental microbiology, 71(10), 5858-5864.

116. **Schwedock, J., & Long, S. R.** (1990). ATP sulphurylase activity of the *nodP* and *nodQ* gene products of *Rhizobium meliloti*. Nature, 348(6302), 644-647.

117. **Schwedock, J. S., Liu, C., Leyh, T. S., & Long, S. R.** (1994). *Rhizobium meliloti* NodP and NodQ form a multifunctional sulfate-activating complex requiring GTP for activity. Journal of Bacteriology, 176(22), 7055-7064.

118. **Simsek, S., Ojanen-Reuhs, T., Stephens, S. B., & Reuhs, B. L.** (2007). Strain-ecotype specificity in *Sinorhizobium meliloti-Medicago truncatula*

symbiosis is correlated to succinoglycan oligosaccharide structure. Journal of Bacteriology, 189(21), 7733-7740.

119. **Singh, S., Katzer, K., Lambert, J., Cerri, M., & Parniske, M.** (2014). CYCLOPS, a DNA-binding transcriptional activator, orchestrates symbiotic root nodule development. Cell host & microbe, 15(2), 139-152.

120. **Solaimanpour, S., Sarmiento, F., & Mrazek, J.** (2015). Tn-seq explorer: a tool for analysis of high-throughput sequencing data of transposon mutant libraries. PLoS One, 10(5), e0126070.

121. **Sørensen, K. K., Simonsen, J. B., Maolanon, N. N., Stougaard, J., & Jensen, K. J.** (2014). Chemically Synthesized 58-mer LysM Domain Binds Lipochitin Oligosaccharide. Chembiochem, 15(14), 2097-2105.

122. **Stracke, S., Kistner, C., Yoshida, S., Mulder, L., Sato, S., Kaneko, T., Tabata, S., Sandal, N., Stougaard, J., Szczyglowski, K., & Parniske, M.** (2002). A plant receptor-like kinase required for both bacterial and fungal symbiosis. Nature, 417(6892), 959-962.

123. **Sutton, M. A., Oenema, O., Erisman, J. W., Leip, A., van Grinsven, H., & Winiwarter, W.** (2011). Too much of a good thing. Nature, 472(7342), 159-161.

124. **Smil, V.** (1999). Nitrogen in crop production: An account of global flows. Global biogeochemical cycles, 13(2), 647-662.

125. **Suzaki, T., Yoro, E., & Kawaguchi, M.** (2015). Leguminous plants: inventors of root nodules to accommodate symbiotic bacteria. International review of cell and molecular biology, 316, 111-158.

126. **Szabados, L., Charrier, B., Kondorosi, A., de Bruijn, F. J., & Ratet, P.** (1995). New plant promoter and enhancer testing vectors. Molecular Breeding, 1(4), 419-423.

127. **Tang, F., Yang, S., Liu, J., & Zhu, H.** (2016). *Rj4*, a gene controlling nodulation specificity in soybeans, encodes a thaumatin-like protein but not the one previously reported. Plant Physiology, 170(1), 26-32.

128. **Terpolilli, J. J., O'Hara, G. W., Tiwari, R. P., Dilworth, M. J., & Howieson,**

J. G. (2008). The model legume *Medicago truncatula* A17 is poorly matched for N2 fixation with the sequenced microsymbiont *Sinorhizobium meliloti* 1021. New Phytologist, 179(1), 62-66.

129. **Tirichine, L., Imaizumi-Anraku, H., Yoshida, S., Murakami, Y., Madsen, L. H., Miwa, H., Nakagawa, T., Sandal, N., Albrektsen, A. S., Kawaguchi, M., Downie, A., Sato, S., Tabata, S., Kouchi, H., Parniske, M., Kawasaki, S., & Stougaard, J.** (2006). Deregulation of a Ca^{2+}/calmodulin-dependent kinase leads to spontaneous nodule development. Nature, 441(7097), 1153-1156.

130. **Tirichine, L., Sandal, N., Madsen, L. H., Radutoiu, S., Albrektsen, A. S., Sato, S., Asamizu, E., Tabata, S., & Stougaard, J.** (2007). A gain-of-function mutation in a cytokinin receptor triggers spontaneous root nodule organogenesis. Science, 315(5808), 104-107.

131. **Truchet, G., Roche, P., Lerouge, P., Vasse, J., Camut, S., de Billy, F., Promé, J. C., & Dénarié, J.** (1991). Sulphated lipo-oligosaccharide signals of *Rhizobium meliloti* elicit root nodule organogenesis in alfalfa. Nature, 351(6328), 670-673.

132. **Tsukui, T., Eda, S., Kaneko, T., Sato, S., Okazaki, S., Kakizaki-Chiba, K., Itakura, M., Mitsui, H., Yamashita, A., Terasawa, K., & Minamisawa, K.** (2013). The type III secretion system of *Bradyrhizobium japonicum* USDA122 mediates symbiotic incompatibility with *Rj2* soybean plants. Applied and environmental microbiology, 79(3), 1048-1051.

133. **van Brussel, A. A. N, Bakhuizen, R., van Spronsen, P. C., Spaink, H. P., Tak, T., Lugtenberg, B. J., & Kijne, J. W.** (1992). Induction of pre-infection thread structures in the leguminous host plant by mitogenic lipo-oligosaccharides of *Rhizobium*. Science, 257(5066), 70-72.

134. **Van Spronsen, P. C., Bakhuizen, R., Van Brussel, A. A., & Kijne, J. W.** (1994). Cell wall degradation during infection thread formation by the root nodule bacterium *Rhizobium leguminosarum* is a two-step process. European

journal of cell biology, 64(1), 88-94.

135. **Van de Velde, W., Zehirov, G., Szatmari, A., Debreczeny, M., Ishihara, H., Kevei, Z., Farkas, A., Mikulass, K., Nagy, A., Tiricz, H., Satiat-Jeunemaître, B., Alunni, B., Bourge, M., Kucho, K. I., Abe, M., Kereszt, A., Maroti, G., Uchiumi, T., Kondorosi, E., & Mergaert, P.** (2010). Plant peptides govern terminal differentiation of bacteria in symbiosis. Science, 327(5969), 1122-1126.

136. **Wang, Q., Yang, S., Liu, J., Terecskei, K., Ábrahám, E., Gombár, A., Domonkos, Á., Szűcs, A., Körmöczi, P., Wang, T., Fodor, L., Mao, L., Fei, Z., Kondorosi, É., Kaló, P., Kereszt, A., & Zhu, H.** (2017). Host-secreted antimicrobial peptide enforces symbiotic selectivity in Medicago truncatula. Proceedings of the National Academy of Sciences, 114(26), 6854-6859.

137. **Wang, D., Yang, S., Tang, F., & Zhu, H.** (2012). Symbiosis specificity in the legume–rhizobial mutualism. Cellular microbiology, 14(3), 334-342.

138. **Wang, Q., Liu, J., & Zhu, H.** (2018a). Genetic and molecular mechanisms underlying symbiotic specificity in legume-rhizobium interactions. Frontiers in Plant Science, 9, 313.

139. **Wang, Q., Liu, J., Li, H., Yang, S., Körmöczi, P., Kereszt, A., & Zhu, H.** (2018b). Nodule-specific cysteine-rich peptides negatively regulate nitrogen-fixing symbiosis in a strain-specific manner in *Medicago truncatula*. Molecular Plant-Microbe Interactions, 31(2), 240-248.

140. **White, J., Prell, J., James, E. K., & Poole, P.** (2007). Nutrient sharing between symbionts. Plant Physiology, 144(2), 604-614.

141. **Yang, S., Tang, F., Gao, M., Krishnan, H. B., & Zhu, H.** (2010). R gene-controlled host specificity in the legume–rhizobia symbiosis. Proceedings of the National Academy of Sciences, 107(43), 18735-18740.

142. **Yang, S., Wang, Q., Fedorova, E., Liu, J., Qin, Q., Zheng, Q., Price, P. A., Pan, H., Wang, D., Griffitts, J. S., Bisseling, T., & Zhu, H.** (2017). Microsymbiont discrimination mediated by a host-secreted peptide in Medicago truncatula. Proceedings of the National Academy of Sciences, 114(26), 6848-

6853.

143. **Yano**, **K., Yoshida, S., Müller, J., Singh, S., Banba, M., Vickers, K., Markmann, K., White, C., Schuller, B., Sato, S., Asamizu, E., Tabata, S., Murooka, Y., Perry, J., Wang, T. L., Kawaguchi, M., Imaizumi-Anraku, H., Hayashi, M., & Parniske, M.** (2008). CYCLOPS, a mediator of symbiotic intracellular accommodation. Proceedings of the National Academy of Sciences, 105(51), 20540-20545.

Supplemental materials

Table S1 Primers used in this study

Purpose	Primer Name	Sequence
Sequencing library construction for the detection of transposon insertions	IllMme1U	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNN
	IllMme1L	CTGTCTCTTATACACATCTTTTTTTTTTTCAAAAAAA
	TpnIRill0	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAAGACGGCATACGAAGACC
	TpnIRill1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNGAAGACGGCATACGAAGACC
	TpnIRill2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNGAAGACGGCATACGAAGACC
	TpnIRill3	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNGAAGACGGCATACGAAGACC
	Mme1pcrF	GTCTCGTGGGCTCGGAGA
Plasmid integration mutagenesis	06071accF	GCGGTACCCGCTAACAATCGCGGTC
	06071bamR	GCGGATCCTGCTCCAGATAGAAGTTG
	06073ecoF	AAGAATTCATCGAGACGAACATCGTCGGCA
	06073hindR	CTCGAAGCTTTCCAACGGTTTCCA
	06074ecoF	CCGAATTCGATCTCACGAGACCGA
	06074hindR	AAAAGCTTTCGTTGGGTAATCGGCGGTG
	06075ecoF	AGGAATTCGCAACAAGGCGATAGCGA
	06075pstR	GCCTGCAGGATGAAACCGGCATAG
	06077ecoF	TTGAATTCGCACATTGCCGATGTCGTC
	06077nheR	CCGCTAGCGACGTTGAACACCTCG
	06079ecoF	CAGCTCGAACGAGAATTCGCGGAA
	06079accR	ATGGTACCGGCGTTCCTGTCCCCA
Plasmid integration mutagenesis	06080bamF	CGGGATCCGCGGATAGACGCCGTC
	06080xhoR	GCTCGGCTCGAGATCATCGTAGAG
	06081pstF	CGCTGCATGACGGTGTAGTCATCT
	06081accR	CGGGTACGCCGGCAACGATCGC
	06084munF	CAACAATTGCACCGCATCGCCAC
	06084bamR	AGAGGATCCACAGATCCAGGCTGA

	LPSbioMunF	GGCAATTGCCGTGGCCGTAGCATTC
	LPSbioPstR	GGTTGAGCTGCAGGAGCATTTGCTG
	06089xhoF	TGCTCGAGATGGCAAGCACCTTTCC
	06089accR	AAGGTACCCCCCCATCGAAAGGTGG
	OantAccF	GGGGTACCGATTTGTTCTCGTCGGCAC
	OantSalR	TTGTCGACGCTGGCAGATCCCATCC
	06097munF	AGCAATTGGCGGAATGGTTCAAGG
	06097hindR	GGAAGCTTGAGGGGTATTCATACGTGAG
	06098salF	TTGTCGACAACAAAAGTACCGATGG
	06098ecoR	GAGAATTCAACGTCGTCCAACCAG
	OPSEQ	TGTGTGGAATTGTGAGCGG
	UNIV	GCCAGGGTTTTCCCAGTCACGA
Resistance cassette insertion mutants	BN406_06091accF	TCGGTACCTGCGCTACCAGCAGACATGG
	BN406_06091sacR	ACGAGCTCCCGCAGTTCCCAGTTCTTA
	SpCmPst	CCCTGCAGTCCGGGACGGCGTCAGC
	SpCmNhe	ACGCTAGCGAAGCGGCGTCGGCTTGA
	LPSbiosyntAccF	CCGGTACCGGTGCGTGACGATGGCATCCTG
	LPSbiosyntSacR	GGGAGCTCCCGCCGGGATGCTTCGTAATG
Deletion mutants of LPS related genes	dNodPQ1sacR	TTTTGAGCTCCAGGTCTTCGCTGAGCATTC
Deletion mutants of LPS related genes	ndPQ1xbaF	GCGATCTAGACGATTGCATTGAACATATTGG
	ndPQ1xbaR	TTTTCTAGACATTTGTGGCCTTGTTCTATGT
	dNodPQ1salF	AGGCGTCGACATCCTGGTCAACAATGC
	dNodPQ2sacF	AAAAGAGCTCGACGTGCGCCTCCAGAACGAC
	dNodPQ2xbaR	CCCTCTAGACATATCGCAATCCCTCTGA
	dNodPQ2xbaF	CTCTCTAGACGCCCCTTTGCGAAATGCTG
	dNodPQ2accR	GAGAGGTACCGCTGTTCACGACAGCGCGTAAG
	dNodHsacR	CCCGGAGCTCACCGGCCCGGAGGCCCTGAAC
	dNodHhindF	AAAAAAGCTTCTCGGCATAATCCAATTCAGC
	dNodHhindR	GTCAAAGCTTGACGTTTGCTAGTTTCGTG
	dNodHaccF	TTTGGTACCTTGATCTCAGCGCCCACTATG
	SulfoTsacR	CTCGAGCTCTGGTAGGCCATATCCA
	SulfoTnheF	AAAGCTAGCCGACGCCACCAATCA

	SulfoTnheR	TTTGCTAGCAAGTGCGTGAACTCTGGTTG
	SulfoTaccF	ATCGGTACCCGCTTGCAAATCACAGGA
	dCysDNxbaF	TTTTCTAGATATTTGCTCTGATCC
	dCysDNpstR	GCGCTGCAGTTCCGTTTCCGGAAGAGTATG
	dCysDNaccR	GCTGGTACCGGAGTGTAGCGCAGTCTG
	dCysDNpstF	AAACTGCAGACGCGATTGACGGATGACCTGA
	dNodFExbaF	CGCTCTAGAAGCTCTCGATCGCTGAG
	dNodFEhindR	TATAAGCTTGTGTTCGACGCACGGCGGCTC
	dNodFEaccR	AAAGGTACCAGTCGGGGTCTGCCATGTGCA
	dNodFEhindF2	GCGTAAGCTTCGCGTTGCCATGAGCAAC
	rkpMhindF	CCACCGACAAGGCGAGCAGTCCGA
	rkpMecoR	TCCTCAATGATCCTGAACCCGTACC
Deletion mutants of LPS related genes	dRkpMecoF2	TCTGAATTCGCAACGTGAATCTCGCCATCATTC
	dRkpMxbaR	AAGTCTAGACGTGCATCACCAATCGCCGACGCAGTC
	LpsS5xhoR	CTCGCTCGAGCGCCCCATATTACCGGCGCCA
	LpsS5hindF	TTCCAAGCTTTTGAAAGTACGCGCTTCCGCAT
	LpsS3bamR	AAAGGATCCGTCGTCGCTGTCGAGCGTGGCA
	LpsS3xhoF	TTTTCTCGAGCTCGCCGCCGATCCGATTCCATTCG
	exoY5F	TTGAATTCTTGCCCTCTCCTCCTGCCT
	dExoY5R	GTCTATTCGATTGCATAGAGGTGACTCCATTTACGTTG
	dExoY3F	ATGCAATCGAATAGACGATCAGGGA
	dExoYbglR	AAGAGATCTCAGATCGGCGAGCGCA
Localization of the protein encoded by BN406_06091	U2	GAAAGGGGGATGTGCTGCAAGG
	mCHmNdeR	GTTTCATGTGATCCGGGTATCTTG
	mCHmNdeF	GATCACATGAAACGGCATGACTTTTTC
	06091sacF	GGAACAAAAGCTGGAGCTCTTCGTGATAGCAACGAC
	06091nheR	TTTGCTAGCGGCGAAGTCCCAATTTGC
	mCherryNheF	AAAGCTAGCAAGGGCGAGGAGGATAAC
	T7promoter	TAATACGACTCACTATAGGG
	06091pHCinFF	GAAGGAGATATACATATGACAATAACAACATATCCTAATGCAAC
	mCHpHCinFR	TCCTTTGCTAGCCATACGTATATCTCCTTCTCAAGATCTGTACAGCTCGTCCATG
	06091prR	CACTTTCGAAGTACCTTTCTTTCTC

	06091prCherry	AGGTACTTCGAAAGTGAGCAAGGGCGAGGAGGATAAC
	06091bglF	AAAAGATCTGTGACAATAACAACATATCCTAATGCAAC
	06091accR	TTTGGTACCAGTCGCCATTGATTATGATCTG
	mCHpHCinFF	GAAGGAGATATACATATGGCTAGCAAGGGCGAGGAGGATAAC
	06091pHCinFR	TCCTTTGCTAGCCATACGTATATCTCCTTCTTAAGCGGCGAAGTCCCAATTTGC
Localization of the protein encoded by BN406_06091	06091_3GlyHAsmaR	ATGGATAGCCACCCCCAGCGGCGAAGTCCCAATTTGCA
	HApHCinFR	TCCTTTGCTAGCCATACGTATATCTCCTTCCTAAGCATAATCTGGAACATCGTATG
	EhuBpagF	TTTGAATTCATGAAACTGAGAGATTTTATGGC
	EhuBnheR	TTTGCTAGCGAAGCCCTGCTCTTTCAGCTC
	mCHecoRIr	TTGAATTCAAGATCTGTACAGCTCGTC
	EhuBpHCinFF	GAAGGAGATATACATATGAAACTGAGAGATTTTATGGC
Determine the transposon insertion site	Tn5prBam	GGGGATCCTCATCCTGTCTCTTGATCA
Construction of Gateway compatible destination vectors for Sm1021 ORFeome library	nifHprAccF	AAAGGTACCGGACTCGGAGGGATGCCGGCA
	nifHprHindR	TTAAGCTTTTTCGGTTGTTCGGACA
	GWhindF	AAAAAGCTTACAAGTTTGTACAAAAAAGCTG
	pKTgwXbaR	CCGGGGATCCTCTAGATCAACCACTTTGTACAAGAAAG
	cycHprF	AAAAGCTTGATATCTAAGCGCCAACCGGATAGGA
	cycHprBamR	GGGGATCCGTTTGGGGAATACTATGAGG
	BacAprAccF	TTTGGTACCGGCGGCCACAGCGTCGAGTTC
	BacAprR	TTGGATCCGGGGAGGTCAAGGACTTGCCTCCTG
	attR1bglF	GGAGATCTGGCTCTTCAATGACAAGTTTGTACAAAAAAGCTG
	attR2accR	GGGGTACCATCACCACTTTGTACAAGAAAG
Test of FSM-MA genome library's quality	M13 reverse	CAGGAAACAGCTATGAC
	M13 forward	GTAAAACGACGGCCAG

Table S2 Sinorhizobium-specific region gene cluster

Locus tag	Product	*Sinorhizobium meliloti* M270	*Sinorhizobium meliloti* KH46c	*Sinorhizobium meliloti* USDA1021	*Sinorhizobium medicae* WSM419	*Rhizobium favelukesii* LPU83
BN406_06118	hypothetical protein sulfotransferase (?)	+	+	+	+	-
BN406_06119	ABC-type multidrug transport system, lipid A export permease/ATP-binding protein MsbA	+	+	+	+	+
BN406_06120	hypothetical protein possibly membrane-anchored, pyruvyl transferase (?)	+	+	+	+	-
BN406_06121	hypothetical protein	+	+	+	+	-
BN406_06122	Peptidoglycan/LPS O-acetylase	+	+	+	+	-
BN406_06123	D-arabinose 5-phosphate isomerase	+	+	+	+	-
BN406_06124	putative membrane-anchored protein	+	+	+	+	+
BN406_06125	4-amino-4-deoxy-L-arabinose transferase or related glycosyltransferase of PMT family	+	+	+	+	-
BN406_06126	glycosyltransferase (?)	+	+	+	+	+
BN406_06127	glycosyltransferase (?)	+	+	+	+	-
BN406_06128	hypothetical protein	+	+	+	+	+
BN406_06129	3-deoxy-manno-octulosonate cytidylyltransferase (?)	+	+	+	+	-
BN406_06130	glycosyltransferase (?)	+	+	+	+	-
BN406_06131	epimerase (?)	+	+	+	+	-
BN406_06132	glycosyltransferase (?)	+	+	+	+	-

Acknowledgements

I would like to express my deep gratitude to my supervisor, Dr. Attila Kereszt, for his education, guidance and financial support during my PhD research. He is not only my supervisor who imparted knowledge to me, but also a good friend, sharing life experiences with me. I would also like to thank Professor Éva Kondorosi, for accepting me to work in the lab and providing financial support. I highly appreciate the time and energy they spent on me.

I am grateful to my colleagues Dr. Péter Körmöczi, Rui Dániel Lima, Helga Edina Vadasi, Benedikta Balla, Csaba Gellért and Sándor Jenei, for their assistance in my PhD research. I also thank all the lab members for their acceptance and friendliness.

I would like to thank China Scholarship Council for providing me financial support during 2015. 09-2018. 08.

Last but not least, I thank my parents for their understanding and support to my scientific research. I also take this opportunity to thank my husband, Dr. Senlei Zhang, for his company and advice on both my research and my life.