July 2020

Origin of Gene Specificity in the Nitrogen-Fixing Symbiosis

Christina Marie Stonoha-Arther
University of Massachusetts Amherst

ORIGIN OF GENE SPECIFICITY IN THE NITROGEN-FIXING SYMBIOSIS

A Dissertation Presented

By

CHRISTINA MARIE STONOHA-ARTHER

Submitted to the Graduate School of the

University of Massachusetts Amherst in partial fulfillment

of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2020

Plant Biology

ORIGIN OF GENE SPECIFICITY IN THE NITROGEN-FIXING SYMBIOSIS

A Dissertation Presented

By

CHRISTINA STONOHA-ARTHER

Approved as to style and content by:

_____
Dong Wang, Chair

_____
Li-Jun Ma, Member

_____
Jeanne Harris, Member

_____
Sam Hazen, Member

_____
Patricia Wadsworth, Director,
Interdepartmental Graduate Programs, CNS

DEDICATION

To Ryan and Jess

ACKNOWLEDGMENTS

Thank you to everyone who helped guide me through this entire process. Members of the Wang lab, past and present, were always helpful and gave excellent advice. Minsoo Kim, Huairong Pan, Onur Oztas, Xiaoyi Wu, Julie Sun, Abby Wells, Kwan Yoon, Milo Ray, and Dave Ahlberg were especially helpful for a lot of the data in this dissertation and for papers that either have been published or are in the works. Milo and Dave were instrumental in setting up the NCR assays, and they both collected a lot of data. Dave collected much of the data in chapter 4 and was really good at working with the bacteria. Milo, I am going to miss conferring with my colleague on very important issues.

My PI, Dr. Dong Wang, has been absolutely amazing and supportive. I could not have asked for a better mentor, and I feel fortunate to have had you as my PI. I had very little idea of what graduate school would be like, and you have always been so patient with me. I also really appreciate how generous you have been with your time, energy, resources, and ideas. I have learned a lot from your creativity as a scientist.

Thank you also to my committee, who has been extremely helpful and encouraging. Dr. Jeanne Harris has always had really great suggestions for experiments and has been a positive source of ideas. I am happy we will continue to work in the same system. Dr. Li-Jun Ma, thank you for allowing me to rotate in your lab all those years ago. I had very little experience with any kind of bioinformatics, and you gave me an excellent foundation that I have used in my graduate work. I also really appreciate the time you have taken to write reference letters and generally support me. Dr. Sam Hazen,

thank you for being on my committee and pushing me to be better. I really appreciate your feedback and suggestions.

Dr. Ertao Wang was extremely generous and allowed me to hang in his lab for two months. I really appreciated how welcome everyone in his lab made me feel and how much fun I had working with the AM symbiosis. Of course, I need to thank Xiaowei for being the best host and a great friend. She did more for me in those two months than I could ever repay her for and was always up for getting some noodles with me.

I also got amazing teaching opportunities at UMass, and I am grateful for the experience. I was a TA for some of the most thoughtful science teachers, and they have been very willing to help me and give me advice. Dr. Healey, Dr. Rounds, and Dr. Tyler have been amazing role models for me.

Thank you also to the PB grad students who were always there for me, even after they left UMass. Colby, Andrea, and Celina, I couldn't have gotten through the first few years without you. I miss you all.

I am very thankful for the undergraduate education I got at SHU. This has been the standard by which I compare other institutions and my own teaching. I would also like to apologize for often being the last student to finish an exam and for writing way, way too much. I finally understand how annoying I was. Special thanks to Dr. Kirk Bartholomew and Dr. Nicole Roy for being so supportive during and after my time at SHU. I can't begin to explain how much you influenced me and how much your support has meant, even from far away. I am incredibly lucky to have you rooting for me.

I would like to thank my mom, dad, sister, and especially my husband for supporting me through all of this and for such a long time. I know it wasn't always easy.

My husband has been absolutely amazing and has provided me with constant encouragement. My family has been super understanding and has helped me stay grounded. A lot happened while I was in graduate school; a lot of it was good, but a lot of it was hard. I hope you already know how much I appreciate the love and support that helped see me through all of this.

My grandma always asked me what I was working on, and I wish I had shared more with her because she was really interested in the microscopes that I was using. She always wanted me to be an actual doctor, but I told her I didn't really want to deal with patients. She knew more than anyone how important bedside manner was because she had it herself while she was nursing both patients at the hospital and members of her family. I think about her often.

My uncle was also an incredible person for following his passions in nursing and increasing the quality of care for people with AIDS. He was a pioneer for men in nursing, and he stepped up during the AIDS crisis to do what he did best for other people.

Nanny and Poppy were instrumental in getting me interested in plants, even though they may not have known it. They were the only people I knew with an extensive garden that could really feed us. I was eating kale before it was trendy because we would come home with bags full of the tough leaves to put into soup. I wrote my admission essay for UMass based on the time I used to spend in Poppy's garden- I am so grateful to have had that experience.

ABSTRACT

ORIGIN OF GENE SPECIFICITY IN THE NITROGEN-FIXING SYMBIOSIS

MAY 2020

CHRISTINA MARIE STONOHA-ARTHER, B.S., SACRED HEART UNIVERSITY

PH.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Dong Wang

Many legumes form a symbiosis with nitrogen-fixing bacteria found in the soil. This relationship is beneficial to both the plant and the bacteria; the plant receives nitrogen that is otherwise limited, and the bacteria receive fixed carbon. Upon sensing the bacteria, the plant forms a new organ (the nodule) where the bacteria are housed within the cells. Many genes are required for the proper formation and function of nodules; this dissertation is broadly focused on how genes required for nitrogen-fixing symbiosis are co-opted from other cellular processes and how they are specialized for symbiosis.

Protein trafficking from the plant to the intracellular bacteria is critical to the success of the symbiosis. This protein trafficking is the cellular anterograde secretory pathway repurposed toward a new intracellular compartment. In the model legume *Medicago truncatula*, a deletion in *DNF1*, which encodes the nodule-specific 22kDa signal peptidase complex (SPC) subunit, causes the nodules to be unable to fix nitrogen because the nodule-specific protein trafficking machinery disrupted. Here, we have shown that *DNF1* became specialized in symbiosis through nodule-specific expression, and we identify *cis*-elements that are crucial for that transcriptional control in the *DNF1* promoter and other SPC subunit genes. Furthermore, we have found that another protein

trafficking protein, SYP132A, was co-opted from arbuscular mycorrhizal symbiosis for its role in nitrogen-fixing symbiosis. This t-SNARE is localized to the symbiotic membranes in both symbioses and is required for proper arbuscule and bacteroid development within host cells.

One class of proteins that are specifically targeted to the symbiosome are nodule-specific cysteine-rich (NCR) peptides, which are involved in the differentiation of the intracellular bacteria. In general, NCR peptides are required for nitrogen fixation in Medicago, but their specific modes of action are largely unknown. Here, we describe and characterize two NCR motifs that are found in many different peptides, pointing to a common, conserved amino acid sequence, possibly contributing to the efficacy of these peptides.

Finally, we show that the receptor, DMI2, which is required for rhizobia infection, is regulated at the protein level in Medicago. *DMI2* is constitutively expressed, but in the absence of rhizobia infection, it is degraded by the proteasome. During infection, however, it is protected from degradation, and the protein accumulates.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

A modified version of this chapter has been published:

Stonoha-Arther C, Wang D. (2018) Tough love: accommodating intracellular bacteria through directed secretion of antimicrobial peptides during the nitrogen-fixing symbiosis. *Current Opinion in Plant Biology* (44) 155-163.

Opening

Some legume plants form a mutualistic relationship with soil bacteria to utilize atmospheric nitrogen, which is otherwise inaccessible to plants. The main benefit of nitrogen-fixing symbiosis for both organisms is the acquisition of nutrients. The plant host receives fixed nitrogen from the bacterial symbiont in the form of ammonia, and the bacteria receive photosynthetic carbon (Udvardi and Poole, 2013).

While the exact molecular dialogue between the plant and symbiont differs between species, the bacteria are ultimately internalized by the host and housed inside nodule cells. In all nodules, the internalized bacteria differentiate into bacteroids before they can fix nitrogen (Franssen et al., 1992; Mergaert et al., 2006a). Some legumes, such as ones in the inverted repeat-lacking clade (IRLC), including the model plant *Medicago truncatula*, cause the bacteroids to differentiate terminally (Lavin et al., 1990; Mergaert et al., 2006a).

In IRLC legumes, bacteroids are coaxed into more profound differentiation by the plant. This terminal differentiation consists of cell elongation, membrane permeabilization, cell cycle changes, and endoreduplication of the bacteria. This

differentiation is irreversible, as it drastically reduces the ability of the bacteria to reproduce (hence the term “terminal,” Mergaert et al., 2006). Bacteroids undergo this transformation surrounded by a plant-derived symbiotic membrane. This organelle, containing the bacteroid and the enveloping membrane (including the space between them) are collectively named the symbiosome. Symbiosome formation is the result of bacterial release from the infection thread, a long and narrow tubular invagination of the plasma membrane. In some less evolved systems of the symbiosis, such as in basal legumes and the non-legume Parasponia, rhizobia are not released but remain in a modified infection thread called the fixation thread to fix nitrogen (Sprent, 2007).

Bacteroid differentiation is dependent on the ability of the plant to traffic proteins to the membrane-bound bacteria inside the nodule cells (Wang et al., 2010). This can be tricky because the newly formed membrane that surrounds the bacteria is derived from the plasma membrane but requires a distinct identity. The work of differentiation is performed by nodule-specific cysteine-rich (NCR) peptides (Van de Velde et al., 2010), of which *M. truncatula* has a predicted 700, that are delivered to the symbiosome membrane. The exact role of every NCR peptide is unknown, but they seem to have antimicrobial effects in vitro while being necessary for successful symbiosis in planta.

The ability of these NCR peptides to differentiate the symbionts is dependent on their accurate and successful delivery to the symbiosome by the plant’s protein trafficking machinery (Van de Velde et al., 2010). Without the protein trafficking machinery or NCR peptides, the bacteria will not differentiate and will, therefore, not fix nitrogen for the plant (Wang et al., 2010). This review will focus on the protein

trafficking of NCRs within the nodules and emerging roles for NCR peptides beyond bacteroid differentiation.

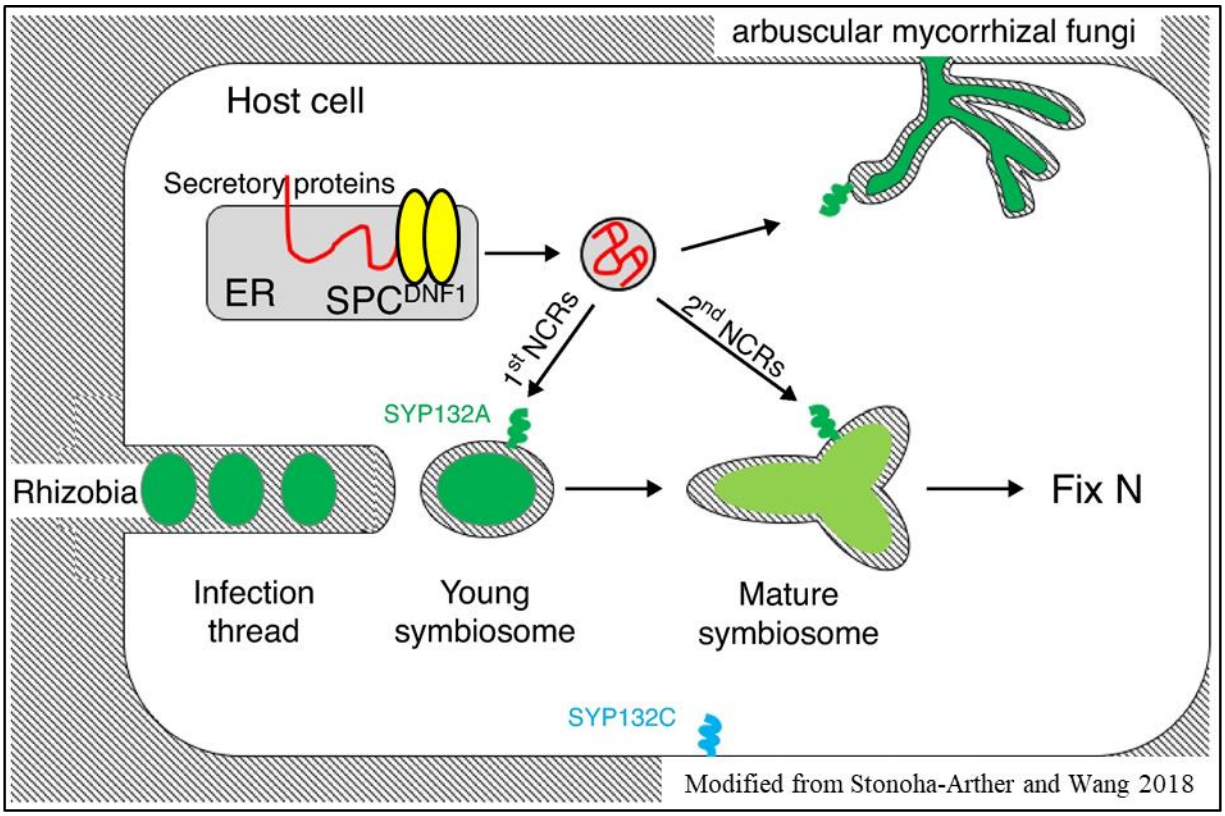


Figure 1.1 Protein secretion to accommodate intracellular rhizobia in the nitrogen-fixing symbiosis. In the *M. truncatula* nodule, rhizobia enter host cells through the infection thread, an invagination of the plasma membrane. After internalization, bacterial cells are individually enveloped in a host membrane to form symbiosomes. Maturation of naive symbiosomes is dictated by protein secretion directed toward this compartment, using components co-opted from the machinery supporting mycorrhizal arbuscules, such as the symbiotic t-SNARE SYP132A. Among the proteins delivered to the symbiosome are a group of NCR peptides, which are capable of interacting directly with the symbiont to drive their terminal differentiation. Some NCR peptides trigger this differentiation, while others maintain bacterial viability in its differentiated state.

Protein trafficking during the nitrogen-fixing symbiosis

Protein trafficking genes involved in multiple steps of the anterograde protein trafficking pathway have been implicated in the symbiosis. A component of the signal peptidase complex (SPC), DNF1, was identified in *M. truncatula* through a forward genetics screen. Without DNF1, the nodules are small and nonfunctional because the

bacteria in symbiosomes of this mutant do not differentiate into nitrogen-fixing forms (Wang et al., 2010). The SPC is responsible for cleaving the signal peptide sequence from secretory proteins (including NCR peptides) entering the endoplasmic reticulum (ER). *DNF1* encodes SPC22, an essential subunit of the complex. In the *dnf1* mutant, these proteins retain their signal peptides and are trapped in the ER instead of being trafficked through the ER to reach the symbiosome. This suggests that proper secretion of host proteins to the symbiosome is crucial for the functioning of the nitrogen-fixing symbiosis. Several other protein-trafficking genes, including ones encoding the other SPC subunits, are similarly upregulated in the nodule (Maunoury et al., 2010; Wang et al., 2010), further emphasizing the importance of host protein secretion to the intracellular bacteria.

Protein secretion is a universal cellular function, and SPC22, in particular, is essential for viability in eukaryotes. However, the *dnf1* mutant shows no other defects, that is, its effect is specific to symbiosome maturation. How could this be? A closer look at the *M. truncatula* genome provided the likely answer, in a close homolog sharing over 80% protein sequence identity with DNF1 (Wang et al., 2010). This protein, DNF1L, is apparently sufficient to handle house-keeping levels of secretion to the extracellular space, allowing DNF1 to specialize in symbiosis. DNF1 and DNF1L may have evolved to function in different processes. However, given how similar they are at the protein level, it is more likely that DNF1 and DNF1L are functionally analogous, and that the actual role of *DNF1* is to simply serve as a second gene copy of *SPC22*, to increase the secretory capacity of cells hosting symbioses. In other words, the specificity of the *DNF1* gene may lie in its transcriptional regulation rather than its protein sequence. Plant

SPCs are made of four subunits (SPC12, SPC18, SPC22, and SPC25), each usually encoded by small families (Paetzel et al., 2002). Similar to *SPC22*, one member each of *SPC12* and *SPC25* is also upregulated in *M. truncatula* nodules (Wang et al., 2010).

If this hypothesis is correct, then in symbiosome-containing cells, one single (albeit enhanced) secretory system handles cargoes destined for two destinations, the regular extracellular space, and the new symbiosome. These are two functionally dissimilar compartments. For example, compared with the extracellular space, the symbiosome lacks a plant cell wall; its membrane contains a plethora of channels and pumps for molecular exchange between the partners in both directions; it is home to a vast amount of proteins interacting with the bacteria (Brewin, 2004). Symbiotic cells may redirect their flow of secretory proteins, making the symbiosome membrane the default target, in a manner similar to cells hosting arbuscular mycorrhizal fungi (Pumplin et al., 2012). At the same time, symbiosomes are long-lived organelles that degrade only when their host cells senesce. If proteins are actively trafficked to both membrane compartments, cargo proteins in these cells must be steered to the correct compartments in a highly reliable manner. How are the separate flow of molecules to the extracellular space and the symbiosome accomplished in the same cell?

At the cellular level, this issue of specificity must be addressed at multiple stages of protein secretion. On the one hand, cargoes must be packaged into different vesicles. On the other hand, the two target membranes -- one marking the cell boundary and the other surrounding intracellular bacteria -- must somehow be distinguished. Solutions to the latter issue are partially accomplished with a t-SNARE protein, an isoform of SYP132, that distinguishes the symbiosome membrane from the plasma membrane

(Huisman et al., 2016; Pan et al., 2016). The SYP13 sub-family of t-SNAREs are well established markers for the plasma membrane, where they mediate the fusion with secretory vesicles in the final step of the secretory pathway (Blatt et al., 1999; Lipka et al., 2007). Therefore, the original discovery of SYP132 on the symbiosome membrane, through proteomic approaches, was a bit of a puzzle with unknown significance (Catalano et al., 2007, 2004). However, subsequent studies revealed that the SYP132 protein on the plasma membrane and the one on the symbiosome membrane are two different protein isoforms derived from the same gene. A transcriptional regulatory process called alternative cleavage and polyadenylation controls the production of these two forms, SYP132C and SYP132A. Usually, only the plasma membrane-localized protein isoform is produced. However, under symbiotic conditions, an additional form, called SYP132A, is produced, localizes to the symbiosome membrane, and performs indispensable functions for a successful symbiosis. *SYP132A* RNAi plants mimic a Fix- phenotype with small, white nodules that fail to fix nitrogen. Presumably, this phenotype is due to the lack of successful protein delivery to the developing bacteroids.

Interestingly, SYP132A is also required for the arbuscular mycorrhizal symbiosis, demonstrating the deep evolutionary root for the requirement of specialized protein secretion in accommodating beneficial microbes (Figure 2.1, Huisman et al., 2016; Pan et al., 2016). The more basal nitrogen-fixing structure, the fixation thread, is a modified infection thread and resembles, at least morphologically, fungal arbuscules. It will not be surprising if fixation threads are also sustained by targeted protein secretion.

The identity of the symbiosome membrane has been further characterized. Even though bacteria are internalized in a process superficially similar to endocytosis, the

ensemble of membrane markers on the symbiosome membrane is distinct from endosomes. In addition to the symbiosis-specific SYP132A discussed above, symbiosomes also contain the small GTPase Rab7, a late endosome marker. It is acquired once the bacteria stop dividing and start differentiating. However, symbiosomes never obtain the small GTPase Rab5, which is also present on late-stage endosomes.

Furthermore, the early endosome (and trans-Golgi network) syntaxin, SYP4, is also absent from the symbiosome membrane, indicating that the membrane that envelopes the bacteria is different than endosomal membranes. Gene silencing studies suggest that *Rab7* is necessary for symbiosome maturation, although the mechanistic basis for this requirement is unclear (Limpens et al., 2009). Finally, symbiosomes also acquire the vacuolar marker TIP1g, which is necessary for symbiosome development (Gavrin et al., 2014). As symbiosomes senesce, more vacuolar markers accumulate on the symbiosome membrane, such as the SNARE proteins SYP22 and VT111 (Limpens et al., 2009).

In summary, our characterization of the symbiosome membrane so far points to a unique combination of membrane markers. On the vesicular side, less is known about how proteins destined to the symbiosome are processed and packaged. Even though it is clear that these proteins transit through the ER, how much other components of the endomembrane system are involved in processing these proteins remains to be established. We do, however, have some information on how vesicles are targeted to the symbiosome membrane.

Two closely related v-SNAREs, VAMP721d and VAMP721e in *M. truncatula*, mark the vesicles targeted to the symbiosome and periarbuscular membranes. These v-

SNAREs are required for symbiosome membrane development, as RNAi knockdowns of both genes halt symbiosome maturation. In these plants, the infection thread can penetrate the plant cell but fails to release the bacteria to form symbiosomes (Ivanov et al., 2012). This is similar to nodules that most strongly silence *SYP132A* (Huisman et al., 2016; Pan et al., 2016). The non-symbiotic v-SNARE, *VAMP721a*, is negatively regulated at the transcriptional level by the C2H2 transcription factor RSD, the lack of which arrests symbiosome development; indicating that *VAMP721a* repression is required for successful nitrogen-fixing symbiosis (Sinharoy et al., 2013).

Recently, it was found that vesicles lacking VAMP721d in soybean nodules were unable to deliver pectin lyase to the tip of the infection thread, pointing to a mechanism by which VAMP721d-labeled vesicles aid in bacterial release from the infection thread (Gavrin et al., 2016). In *Lotus japonicus*, a pectate lyase is required for the formation of infection threads (Xie et al., 2012), probably in a step to remodeling the plant cell wall at the tip of the root hair. It is conceivable that a similar function is required to build the wall-less infection droplet as a prerequisite for bacterial release. Again, similar to SYP132A, deficiencies in VAMP721d/e impair the development of the arbuscular mycorrhizal symbiosis. Such overlapping phenotypes imply that VAMP721d/e and SYP132A may form a complex to mediate vesicle fusion to the symbiosome target membrane. However, this has yet to be shown directly.

Additional components of the exocytotic system are important for protein secretion in symbiosis. Recently, Gavrin and colleagues (2017) showed that synaptotagmins, Ca^{2+} sensors interacting with SNARE proteins, are recruited to the sites of rapid membrane expansion in nodule cells (infection threads, unwalled droplets, and

symbiosomes), most likely as a membrane repair mechanism. Pairwise RNAi knockdown of the three synaptotagmins used in this study prevented bacteroid differentiation, supporting the hypothesis that these membrane proteins are required for the development of the symbiotic membrane (Gavrin et al., 2017).

What are the cargo proteins of the protein secretion pathway in the nodule? Several nodule-expressed secretory proteins, such as Nodulin25 and glycine-rich proteins, have been shown to localize to the symbiosome (Hohnjec et al., 2009a; Mergaert et al., 2003), but their exact roles in the symbiosis are unknown. Recently, a few proteins crucial for a productive symbiosis have been identified, including DNF2 and DNF5 (also called SymCRK, Berrabah et al., 2014; Bourcy et al., 2013). Although both contain signal peptides, their localization patterns have not been demonstrated. They are reasonable candidates for cargoes of the protein secretion pathway.

The massive induction of the protein secretory pathway in infected *M. truncatula* nodule cells – for example, a drastic expansion of the ER (Maunoury et al., 2010) -- is, in part, needed to process the many defensin-like peptides that the plant delivers to the intracellular bacteria through this system (Durgo et al., 2015; Van de Velde et al., 2010). Called nodule-specific cysteine-rich (NCR) peptides, they play instrumental roles in bacteroid differentiation and shape the outcome of this symbiosis.

NCR peptides and bacteroid differentiation

NCR peptides were first discovered in the *M. truncatula* genome (Mergaert et al., 2003), which is predicted to encode about 700 of them (Graham et al., 2004; Maróti et al., 2015; Silverstein et al., 2007). Structurally, they resemble defensins, ancient

eukaryotic peptides used to defend against microbial pathogens (Carvalho and Gomes, 2009; Ganz, 2003; Mergaert et al., 2003; Zasloff, 2002). NCR peptides are highly polymorphic in amino acid sequence (Nallu et al., 2014); however, they share a few key characteristics with plant defensins. Both are short, cysteine-rich polypeptides (about 30-50 amino acids for mature NCR peptides) with a signal peptide that targets them to the secretory pathway (Carvalho and Gomes, 2009). NCRs often have 4 or 6 regularly spaced cysteines, which is distinct from defensins (Mergaert et al., 2003). Functionally, defensins are antimicrobial peptides, able to form pores in microbial cell membranes. Because bacteria have negatively charged membrane lipids, cationic defensins are more potent and can lead to bacterial lysis (Mikuláss et al., 2016).

Although it is difficult to prove the evolutionary origins of NCR peptides, they probably have activities similar to defensins, such as pore formation on membranes, which can explain the antibacterial actions of NCR peptides on free-living bacteria. Such antimicrobial activities can be broad-spectrum (Tiricz et al., 2013). For example, NCR247 and NCR335 can disrupt the membrane potential of *Salmonella enterica* (another gram-negative bacterial species) and the gram-positive species Listeria monocytogenes in addition to *S. meliloti* (Farkas et al., 2017).

However, the *in planta* function of NCR peptides, unlike defensins, is not to eliminate bacteria, but to induce terminal bacteroid differentiation in IRLC legumes. They do so by disrupting the membrane potential and altering the surface appearance of *S. meliloti* (with similar effects on *E. coli* and artificial membranes (Mikuláss et al., 2016; Nagy et al., 2015). NCR247 is one of the most well-studied NCR peptides and is often used as a model for their modes of action on cultured bacteria. For example, at low

concentrations, NCR247 can inhibit cell division, transcription, and translation (Farkas et al., 2014; Penterman et al., 2014). At higher concentrations, NCR247 has bactericidal activity (Van de Velde et al., 2010). NCR247 has four cysteines and, therefore, the ability to form different disulfide bonds when oxidized, or no disulfide bond when fully reduced. Shebab *et al.* (2016) used three such regioisomers of NCR247 to investigate which of these conformations are responsible for its range of effects on *S. meliloti*. Interestingly, the result depends on the cellular activity being investigated. While all forms of NCR247 can inhibit cell division at relatively low concentrations, the reduced form is more effective at inhibiting bacterial translation (Haag et al., 2012; Shabab et al., 2016).

The exact redox state of NCR peptides *in planta* is unknown, a significant gap in our knowledge on how they affect intracellular bacteroids. However, some information can be inferred from other genes; one nodule-specific thioredoxin (*Trx s1*) in *M. truncatula* (Alkhalfioui et al., 2008) was recently described. Unique among thioredoxins, Trx s1 contains a signal peptide and is localized to the symbiosome. Trx s1 can interact directly with NCR247 and NCR335 *in vitro* and can additionally reduce the former. When Trx s1 was ectopically expressed in free-living *S. meliloti*, the bacteria became more sensitive to NCR335, presumably because the peptide reduced by thioredoxin has increased activities (Ribeiro et al., 2017).

On the other hand, Trx s1-expressing bacteria were no more susceptible to treatment with NCR247, even though the reduced form of this peptide is more potent at inhibiting translation (Shabab et al., 2016). The involvement of a thioredoxin suggests that the reduced form is the relevant conformation of NCR peptides in this symbiosis.

However, more work is needed to ascertain which activities of the NCRs are more important *in planta*.

Complicating the matter is the immense size and high sequence divergence of the NCR peptide family in *M. truncatula*, in which NCR247 is among the most cationic (Montiel et al., 2016). Therefore, it remains to be seen how much of what we know about NCR247 is applicable to other family members. However, we do know that NCR peptides are not only important in manipulating bacteroid development toward differentiation but are also crucial for the persistence of the intracellular bacteria in planta (Figure 2.1). Two symbiotically deficient mutations, *dnf7* and *dnf4*, are caused by deletions of specific NCR peptides. *DNF7* encodes NCR169; without it, the nodule does not develop fully, with both living and dead bacteria in their cells (Horváth et al., 2015).

Similarly, *DNF4*, which encodes NCR211, is important for bacterial survival. In the *dnf4* mutant, bacteria differentiate but subsequently die (Kim et al., 2015). These results indicate that NCR peptides can both induce bacteroid differentiation and promote bacterial survival in the nodule. Given the large size of the family in *M. truncatula*, it is reasonable to speculate that NCRs may have diverse modes of action. It is unknown whether different NCRs interact with each other directly, but it has been proposed that NCR peptides with opposite properties, such as surface charges, may counteract each other (Mergaert, 2018).

Consistent with the idea of diverse functions in symbiosis, some NCR genes have very distinct expression patterns. Although they are all exclusively expressed in nodules, certain NCR peptides are transcribed only in particular zones, sometimes in a narrow band of cells of similar age (Guefrachi et al., 2014). NCR genes expressed early in the

symbiotic program may be more critical for inducing differentiation features in rhizobia, whereas late-expressed peptides may primarily function to control differentiated bacteria. NCR peptides that interact with each other could have similar expression patterns or could be expressed in adjacent zones of the nodule, which form a developmental gradient.

Most work on NCR peptides is done in *M. truncatula*, where the genes were first described (Mergaert et al., 2003). *M. truncatula* is one of the IRLC legumes, which impose terminal differentiation on their bacteroids (Mergaert et al., 2006a). Investigation on the prevalence of NCR peptides in other IRLC members generated essential insights into the evolutionary trajectory of these peptides. All IRLC legumes examined to date contain NCR peptides, but their numbers seem to vary greatly. Interestingly, the degree of bacteroid differentiation, judged by the extent of morphological changes, is also highly variable between host species. For example, *Cicer arietinum* houses elongated bacteroids, but without the branching morphology seen on *S. meliloti* in *M. truncatula* symbiosomes, while *Glycyrrhiza uralensis* contains bacteroids similar in size to free-living bacteria (Montiel et al., 2016). The number of NCR peptides among IRLC legumes appears to correlate with the degree of morphological changes in bacteroids. The more NCR peptides expressed in nodules, the more profoundly the bacteria change their morphology when differentiating (Montiel et al., 2017).

Interestingly, there is evidence of NCR-like peptides arising independently outside the IRLC. Within the more basal Dalbergioid legumes, species of the Aeschynomene genus house bacteroids with characteristics similar to those found in the IRLC legumes, and a class of cysteine-rich peptides specifically expressed in the nodule

were linked to this bacteroid differentiation. This is most likely a result of convergent evolution, showing that this effective method of controlling intracellular bacteria has been adopted repeatedly (Czernic et al., 2015).

That host antimicrobial peptides evolved more than once in the legume-rhizobia symbiosis suggests that these molecules provide an advantage to the plant. A comparative study showed plants that induce terminal differentiation in their rhizobial partners enjoy higher symbiotic efficiency (Oono and Denison, 2010). This may be because analogous to polyploid eukaryotic cells, terminally differentiated bacterial cells have more active metabolism concomitant with the onset of polyploidy (Van de Velde et al., 2010). It has also been shown that once internalized by bacteroids, NCR peptides interact with components of bacterial metabolism, target protein translation, and trigger global changes in the transcriptome (Farkas et al., 2014; Penterman et al., 2014). Regardless of the exact mechanism, terminally differentiated bacteria degrade during senescence and are absorbed by the host cell. Thus, any carbon not spent on nitrogen fixation can be recovered by the plant host.

Emerging roles for NCR peptides

It is likely that the two in planta roles of NCR peptides, inducing bacterial differentiation and ensuring the survival of differentiated bacteroids, are both based on these peptides' ability to manipulate bacterial physiology, such as membrane permeabilization. Collectively, these peptides steer bacteria away from proliferation, and convert the microbes into organelles reliant on the host cell for survival. Because the

same activity can also be lethal on free-living rhizobia, it is conceivable that the same bactericidal effect can manifest itself in nature as well.

Recent companion papers describe two NCR peptides (NFS1 and NFS2) important for strain specificity in *M. truncatula* during the nitrogen-fixing symbiosis. Both NFS1 and NFS2 contribute to the incompatibility between the bacterial strain RM41 and the *M. truncatula* accession Jemalong A17, while other accessions (specifically DZA) form functional nodules with the same bacterial strain. The A17 versions of these peptides negatively affect the intracellular bacteria; bacteroids fully differentiate but then lyse, causing premature nodule senescence (Wang et al., 2017; Yang et al., 2017).

The A17 version of NFS2 has antimicrobial activity on free-living RM41 (as well as on another compatible strain), while the DZA version does not exhibit antibacterial activity. This could explain the contribution of NFS2 to the incompatibility between A17 and RM41 (Wang et al., 2017). However, for the NFS1 peptide, both the A17 and DZA versions can restrict RM41 growth in culture, suggesting that the role of NFS1 in this incompatibility is more nuanced (Yang et al., 2017). This highlights the contrast between the complexity of NCR peptide actions *in planta* and the practice of studying individual family members on bacteria in culture. Potentially the effect of NFS1 in planta is modulated by other differences between A17 and DZA, such as expression levels.

FS1 and NFS2 are also responsible for the incompatibility between *M. truncatula* accessions and another bacterial strain, *S. meliloti* A145. Similar to the previously described studies, A145 forms functional nodules on DZA, but nonfunctional ones on A17. It was discovered that NFS1 and NSF2 also contribute to this incompatibility. Indicative of possible bacterial factors involved incompatibility, A145 was found to be

more sensitive than Rm41 to NFS1 (Wang et al., 2018). These two genes have shown that NCR peptides are not only important for bacteroid differentiation and subsequent survival but can contribute to symbiont specificity. If NCR peptides ultimately work like defensins and disrupt bacterial membrane integrity, this general mechanism can be co-opted for bacterial differentiation and for ensuring that the preferred bacterial strains have colonized the nodule.

Most of the demonstrated roles of NCR peptides ultimately restrict the proliferation of rhizobia in the nodule, and sometimes they even block bacterial access to the nodule in the first place. Therefore, it is not surprising that rhizobia have evolved countermeasures to resist the action of NCR peptides. These measures range from mild tolerance (for example by removing NCR peptides from bacterial cell surface via the BacA protein) to active degradation (in the case of a peptidase known as HrrP (Haag et al., 2011; Pan and Wang, 2017; Price et al., 2015). In most cases, we are still ignorant of the bacteria's self-preservation mechanism.

Closing

Recent investigations into the cellular mechanisms of the nitrogen-fixing symbiosis in model legumes revealed the prominence of the protein secretory pathway toward the microbial compartment, the symbiosome. This prominence can appear counterintuitive, as the symbiosome is an intracellular structure. Yet another valuable lesson learned is the impact of evolution in shaping this mutualism. It was discovered, sometimes, by tracing evolution backward, that much of this bacterial symbiosis is borrowed from the more ancient symbiosis with mycorrhizal fungi. In both symbioses,

the microbe and the host cell form an intimate membranous interface, across which nutrients and signals are exchanged. In developing the nitrogen-fixing symbiosis, legume cells enveloped the microbe, but retained the topological arrangement of segregating it with a membrane interface. At the subcellular level, the challenges of maintaining a distinct destination for vesicle trafficking remain the same. So would the likely solutions, one can surmise. We are only scratching the surface of this pathway. Vesicle trafficking is not only crucial for delivering protein factors to the other partner but probably instrumental in constructing this interfacial membrane in the first place, such as the deposition of membrane proteins and chemically distinct lipids. For example, how does the newly synthesized SYP132A protein localize to the symbiosome membrane, while its housekeeping counterpart still localizes to the plasma membrane? Although the underlying mechanism is currently unknown, it is reasonable to predict that a similar process functions in the arbuscular mycorrhizal symbiosis as well.

Although the emergence of the legume-rhizobia symbiosis is rooted in evolutionary conservation, the story is also one of evolutionary innovation. While the requirement for protein secretion is shared between mycorrhizal and rhizobial symbiosis, the functions of the secreted protein are necessarily specific to each microbial partner. This is best understood with NCR peptides, found in legumes that induce terminal differentiation in their rhizobial symbionts. By inducing terminal differentiation, these host "effectors" confer the host more control over the bacteria. It is conceivable that rhizobia evolved mechanisms to counter the actions of these peptides. The explosion of the NCR family in *M. truncatula* could well be the result of multiple rounds of host actions and rhizobial counteractions, in a mode that parallels the "zig-zag" model for

plant-pathogen interactions (Jones and Dangl, 2006). Sometimes the host's arsenal of NCR peptides may prove too overpowering, resulting in bacteria death. In other cases, the bacteria may effectively degrade these peptides to resist differentiation. Macroscopically, both could result in a collapse of the symbiosis, manifesting itself as nonperforming nodules. The productive symbiosis that we are accustomed to may be the result of a precarious balance between two species vying for an upper hand in a complex relationship.

Protein trafficking is known to be essential for both housekeeping functions and specifically for nitrogen-fixing symbiosis. However, the origin of the components of the pathway that are upregulated in the nodule is currently unknown. It is also unclear if the protein signaling pathway in the nodule is performing a function that is distinct from that which it does in non-symbiotic cells in the plant. It is known that NCR peptides are trafficked through the endomembrane system to the rhizobia in the nodule cells. However, the properties of NCR peptides that are critical for bacterial differentiation are mostly unknown except for a handful of well-studied peptides. Here, we will address several of these questions, with a general focus on the protein secretory pathway and its involvement in nitrogen-fixing symbiosis.

CHAPTER 2

TRANSCRIPTIONAL REGULATION OF PROTEIN TRAFFICKING IN NITROGEN-FIXING SYMBIOSIS

Introduction

The model legume *Medicago truncatula* forms indeterminate nodules with rhizobia that undergo terminal differentiation within nodule cells. This plant-directed bacterial differentiation is irreversible and required for nitrogen fixation (Mergaert et al., 2006b). Protein trafficking plays an important role in the communication between the plant host and rhizobia during nitrogen-fixing (NF) symbiosis. Without proper trafficking, fully differentiated, functional bacteroids do not form, and nitrogen fixation does not take place. This is because the plant supplies the bacteria with nodule-specific cysteine-rich (NCR) peptides and other proteins that make the nodule environment conducive to nitrogen fixation (Van de Velde et al., 2010). Many NCR peptides have antimicrobial properties that induce bacterial differentiation.

The anterograde protein secretion pathway is required for targeting proteins to different cellular compartments, including the symbiosome in nodule cells. One component that is essential for proper protein trafficking is the signal peptidase complex (SPC). In plants, this complex is composed of four subunits that are embedded in the membrane of the endoplasmic reticulum (ER). When proteins that are targeted to the ER for trafficking are translated, the SPC cleaves the signal peptide from the nascent polypeptide chain to allow proper folding of the mature protein (Paetzel et al., 2002).

DEFECTIVE IN NITROGEN FIXATION 1 (*DNF1*) is a gene necessary for protein trafficking in the nodule. Null mutations in this gene in *M. truncatula* result in plants that form many small, white nodules incapable of fixing nitrogen, due to a failure of the bacteroids to differentiate sufficiently (Wang et al., 2010). *DNF1* encodes the nodule-specific 22kD subunit of the SPC and is, therefore, essential for protein trafficking through the endomembrane system to the bacteria. *DNF1* is specifically upregulated in the nodule after the initiation of infection, with some basal transcription in the root (Wang et al., 2010). Furthermore, microarray and RNAseq data show that expression of DNF1 is significantly induced upon rhizobia inoculation. On the other hand, the *DNF1* paralog, *DNF1-like* (*DNF1L*), is expressed basally in all tissues, and not significantly upregulated in the nodule (Benedito et al., 2008; Roux et al., 2014).

Among the client proteins of the SPC^{DNF1} complex are over 700 NCR peptides encoded in the Medicago genome. Their expression is extremely specific to the nodule and therefore are clearly under tight transcriptional control. NCR genes are probably among the most tightly regulated of all Medicago genes; for most of them, there is virtually no expression in any other organ except the nodule, and their expression in the nodule can be extremely high (Guefrachi et al., 2014). NCR peptides are also expressed in various zones of the nodule throughout development, these NCRs are generally distinguished as "early" and "late," depending on when (and in which zones) they are expressed. The regulation of these NCR peptides involves both epigenetic changes, such as demethylation, and transcription factor binding to *cis*-elements in the promoters (Nallu et al., 2013; Satgé et al., 2016). Multiple known *cis*-elements occur upstream of the transcription start sites of NCR genes, such as an Auxin Response Factor, MADS box,

and Dof binding site. Furthermore, a nodule-specific motif found in leghemoglobin promoters of various species (including those that do not use NCR peptides in the nodule) is found in both early and late NCR promoters, suggesting that this may be a general nodule-specific *cis*-element (Nallu et al., 2013).

Many genes important for rhizobia infection and nodule development or function have specific expression patterns spatially and temporally. Leghemoglobins were some of the first genes investigated for nodule-specific promoter elements, because of their clear induction -- and importance -- in the nodule. The *Glycine max* leghemoglobin gene *Lbc3* was found to contain a strong positive element, a weak positive element, a negative element, and a motif responsible for nodule-specific expression, named the OSE for Organ Specific Element (Stougaard et al., 1987). Subsequently, the OSE has been found in leghemoglobin genes of other species such as *M. truncatula* and *Sesbania rostrata*, suggesting that this is a conserved element that recruits an unknown *trans*-factor (Metz et al., 1988; Ramlov et al., 1993; Stougaard et al., 1990; Szczyglowski et al., 1994). Finally, the OSE has also been found in promoters of other nodule-specific genes, including NCR peptides, as mentioned above (Nallu et al., 2013).

Regulation of other genes required for nitrogen-fixing symbiosis has also been investigated; many of them are involved in the early steps of nodulation (i.e., rhizobial infection and organogenesis; reviewed by (Mergaert, 2018). For example, *Nodule INception* (*NIN*) was discovered to be indispensable for nodulation, and it is under tight transcriptional control. NIN expression in the infection thread is induced by *CYCLOPS/ IPD3,* and it is also induced later by a heterodimer composed of NSP1 and NSP2 (reviewed by Roy et al., 2019). Interestingly, the cytokinin response elements required

for *NIN* expression in the pericycle are located 18kb upstream of the coding region (Liu et al., 2019).

Only limited information is available regarding the transcriptional regulation of protein trafficking genes during nitrogen-fixing symbiosis. For example, many protein trafficking genes were highlighted as being co-expressed with *DNF1*, one of which was *SYP132A* (Wang et al., 2010). *SYP132A* has been shown to be important for both nitrogen-fixing symbiosis and arbuscular mycorrhizal symbiosis. Expression of *SYP132A* is controlled by alternative transcriptional termination, and not at the promoter (Huisman et al., 2016; Pan et al., 2016). Furthermore, a transcriptional factor, RSD1, acts as a negative regulator of the canonical v-SNARE, VAMP72a, in the nodules. This downregulation is required for nitrogen fixation (Sinharoy et al., 2013). On the other hand, *VAMP72d* and *VAMP72e* are upregulated in the nodule and are required for vesicle delivery to the symbiosome and arbuscule (Ivanov et al., 2012).

Results

Transcriptional regulation of DNF1 is key to its role in symbiosis

Protein secretion and, in particular, SPC activity is essential for viability. Even though the SPC22 subunit is required for SPC function, the *dnf1* mutant phenotype is restricted to nodulation. Presumably, this is because signal peptide processing in other circumstances is carried about by the homologous DNF1L, clues about this are found in their expression profiles (Figure 2.1). However, DNF1 and DNF1L are highly similar at the protein level, sharing 82% amino acid identity (Figure 2.2). Because the similarity is observed throughout the polypeptide chain, it is less likely that the two proteins have

drastically divergent activities. To ascertain any potential functional consequence of the minor sequence differences between them, we asked whether the DNF1L protein can functionally replace DNF1.

To address this, we attempted to rescue *dnf1* with *DNF1L* by driving the expression of *DNF1L* with the ~3,000bp *DNF1* promoter, which shows high levels of activity in the nodule and was sufficient to allow the rescue of *dnf1* by a *DNF1* genomic construct (Wang et al., 2010). Indeed, *DNF1L* was able to rescue the nodule phenotype of *dnf1* when expressed highly in nodules (Figure 2.2B-E). This DNF1L construct was able to prevent symptoms of nitrogen starvation in the above-ground tissues, indicating that the nodules are indeed fixing nitrogen (Figure 2.3). This result shows that the two SPC22 proteins in *M. truncatula* are functionally equivalent and that the transcriptional regulation of *DNF1* imparts its nodule specificity.

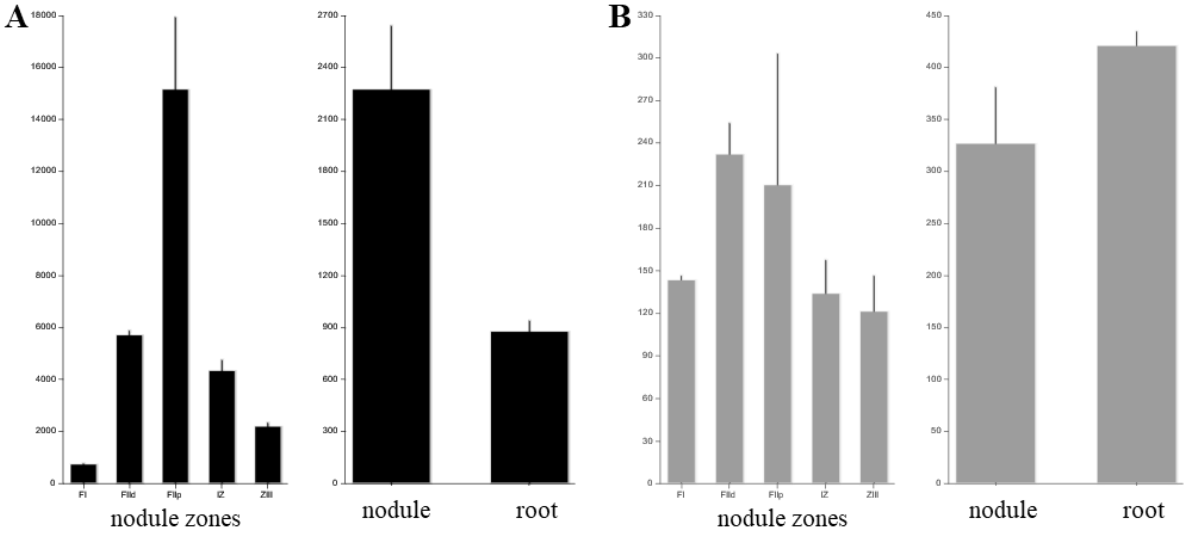


Figure 2.1. *DNF1* and *DNF1L* expression in nodules, roots and nodule zones. A. RNAseq expression data of *DNF1*. **B.** RNAseq expression data of *DNF1L*. Data obtained from Symbimics (https://iant.toulouse.inra.fr/symbimics/).

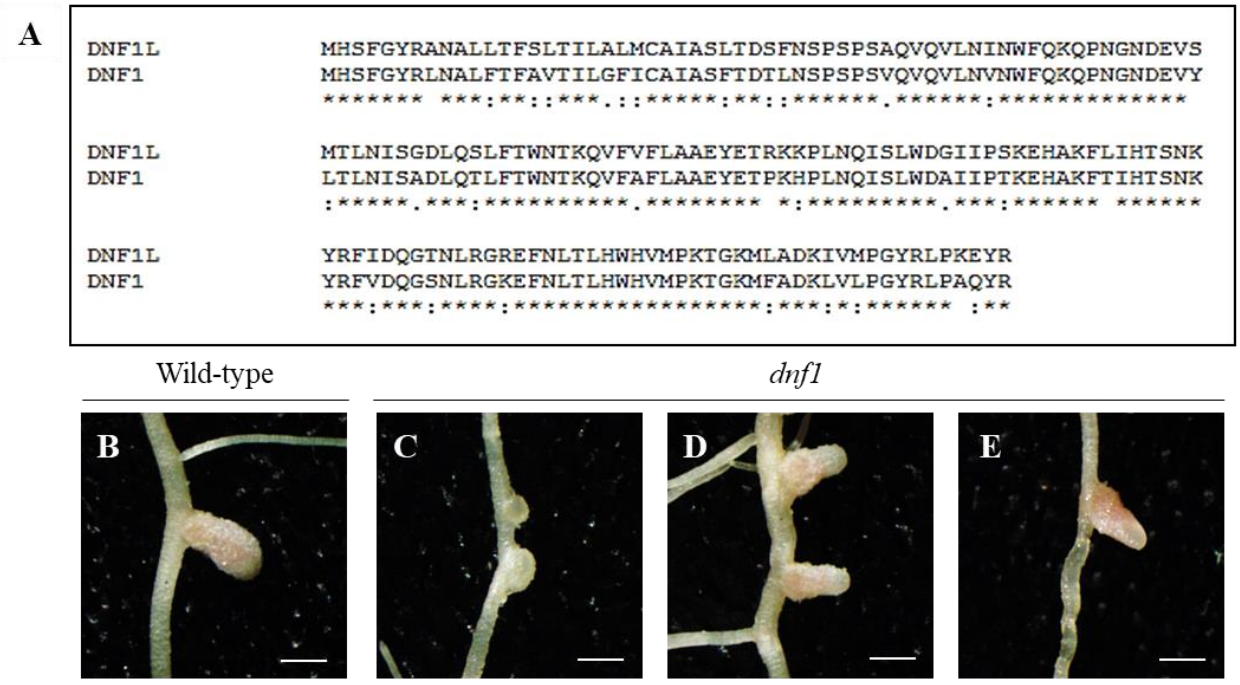


Figure 2.2. *DNF1* and *DNF1L* are paralogs in *Medicago truncatula*. A. MUSCLE alignment of DNF1L and DNF1 peptide sequences. **B.-E.** Nodules on *M. truncatula* hairy roots for complementation of the *dnf1* mutant. Nodules on Jemalong wild-type hairy roots transformed with empty vector as shown in (B). Nodules on hairy roots of *dnf1* transformed with empty vector (C). Nodules on *dnf1* hairy roots transformed with *DNF1* driven by the full-length (~3000bp) *DNF1* promoter (D). Nodules on *dnf1* hairy roots transformed with *DNF1L* driven by the full-length (~3000bp) *DNF1* promoter (E). Plants were inoculated with *S. medicae* strain ABS7 and harvested 3wpi (weeks post inoculation). Scale bars= 1mm.

Figure 2.3. Aboveground phenotype of complemented, composite *dnf1* plants. A. Aboveground portion of Jemalong wild-type composite plants with roots transformed with empty vector. **B.** Aboveground portion of *dnf1* wild-type composite plants with roots transformed with empty vector. **C.** Aboveground portion of *dnf1* wild-type composite plants with roots transformed with *DNF1* driven by the full-length (~3000bp) *DNF1* promoter. **D.** Aboveground portion of *dnf1* wild-type composite plants with roots transformed with *DNF1L* driven by the full-length (~3000bp) *DNF1* promoter. Plants were inoculated with *S. medicae* strain ABS7. Plant were harvested 3wpi (weeks post inoculation). Scale bar= 2cm.

Putative *trans*-factor binding motifs exist in the *DNF1* promoter

The 3kb DNF1 promoter drives the expression of a GUS reporter to high levels inside the nodule and That the 3kb DNF1 promoter can rescue the *dnf1* mutant when driving either *DNF1* or *DNF1L* expression suggests this region could contain important transcriptional regulatory sequences. In order to delineate regulatory regions in the *DNF1* promoter important for nodule expression, we successively removed sequences from its 5' end, creating fragments of approximately 1300, 650, and 420 bp upstream of the start codon. The full-length (3kb) DNF1 promoter drives the expression of a GUS reporter to high levels inside the nodule (Wang et al., 2010). Similar to the full-length promoter, all truncated *DNF1* promoter-*GUS* constructs resulted in reporter activity in the nodules when transiently transformed into Jemalong A17 hairy roots (Figure 2.4). This indicates that *cis*-elements within the 422 bp fragment of the promoter are sufficient to confer nodule-specific expression.

We then analyzed this 422 bp fragment for possible motifs that may be imparting nodule-specific expression to *DNF1*, using the Plant *Cis*-Acting regulatory DNA Elements (PLACE) database (Higo et al., 1999). We found two instances of a sequence of interest that could be potential regulatory motifs. These sequences resemble part of the Organ Specific Element (OSE), previously discovered in leghemoglobin promoters of several species (Ramlov et al., 1993; Stougaard et al., 1990; Szczyglowski et al., 1994). Generally, the OSE is made of two partially palindromic sequences' AAAGAT' and 'CTCTT' (that is, '[A]AAGA[T/G]' in opposite directions), separated by about six base-pairs. In comparison, the *DNF1* promoter contains two copies of 'AAAGAT' in the same direction, separated by 27 base-pairs (Table 2.1). Since the OSE has been shown to be

responsible for nodule-specific expression in various leghemoglobin promoters, we hypothesized that the similar motifs found in the *DNF1* promoter are at least partially responsible for its nodule-specific expression, as such, we have named this motif SOLE (Semi OSE-Like Element).

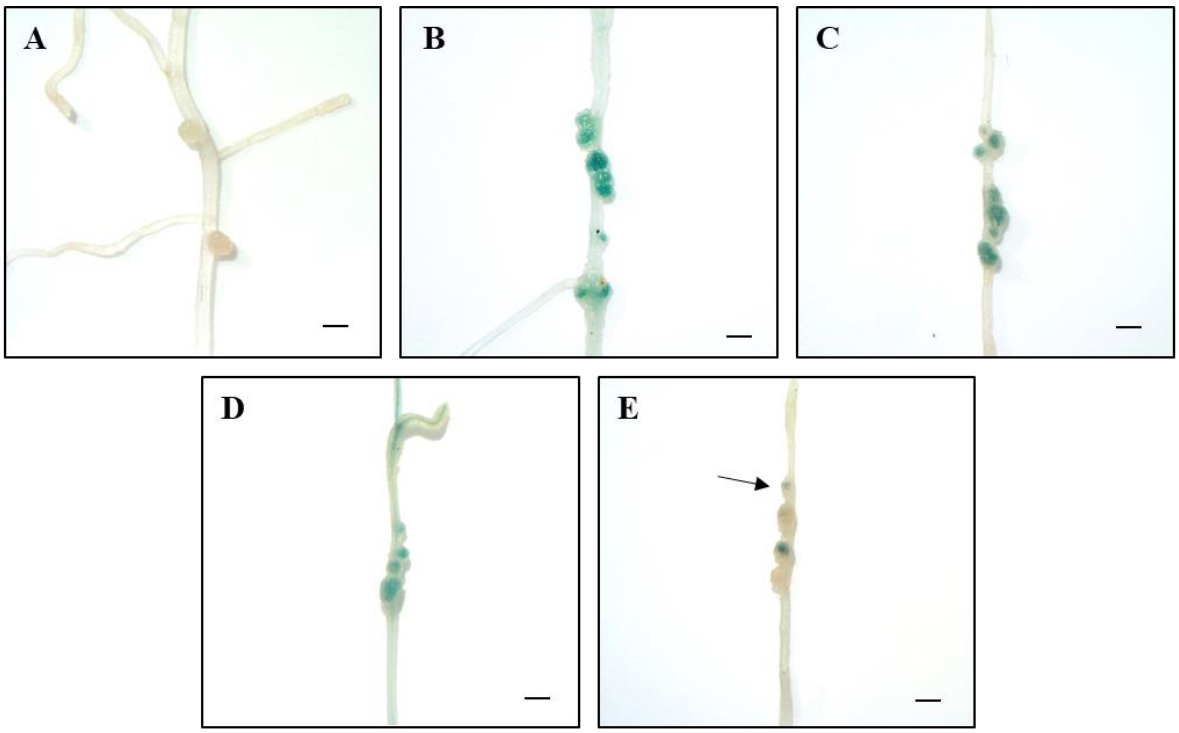


Figure 2.4. Beta-glucuronidase (GUS) staining of hairy roots transformed with *DNF1* promoter deletion constructs. A. Hairy roots transformed with empty vector control. **B.** Hairy roots transformed with full-length (approximately 3000bp upstream of the start codon) *DNF1* promoter construct. **C.** Hairy roots transformed with the first promoter-deletion construct (approximately 1300bp upstream of the start codon). **D.** Hairy roots transformed with the second promoter-deletion construct (approximately 650bp upstream of the start codon). **E.** Hairy roots transformed with the third promoter-deletion construct (approximately 420bp upstream of the start codon). Arrows indicate GUS staining. All constructs were transformed into Jemalong A17 hairy roots. Plants were inoculated with *Sinorhizobium medicae* strain ABS7 and harvest 3 weeks post inoculation. Arrows indicate GUS staining. Scale bar= 1mm.

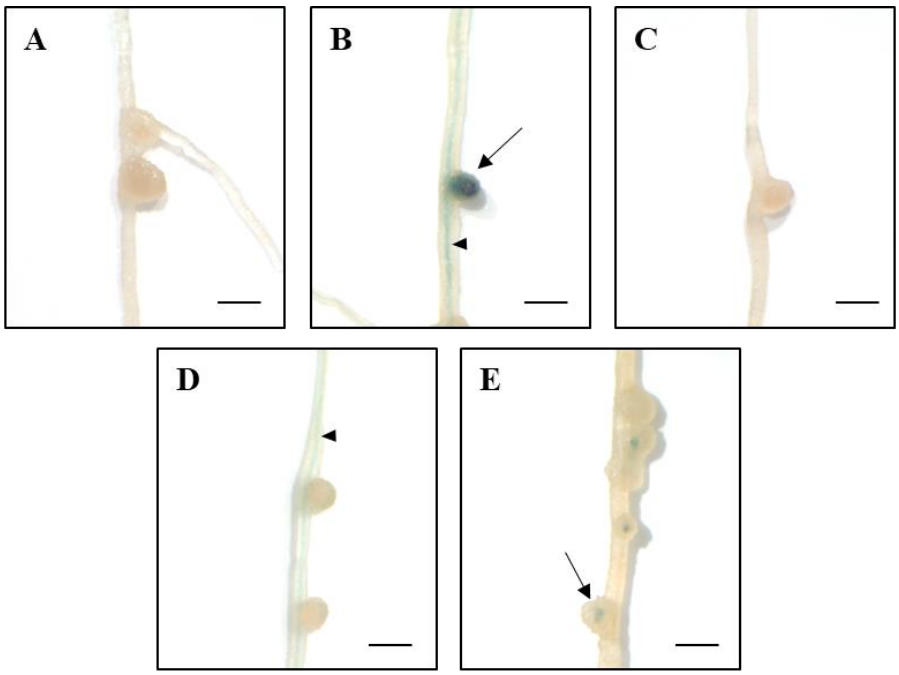


Figure 2.5. GUS staining of *M. truncatula* root nodules transformed with nodule-specific SPC gene promoter-*GUS* constructs. A. GUS staining of nodules transformed with the empty vector. **B.** GUS staining of nodules transformed with the full-length *DNF1* promoter-*GUS* construct. **C.** GUS staining of nodules transformed with *SPC12* promoter-*GUS* construct. **D.** GUS staining of nodules transformed with *SPC18* promoter-*GUS* construct. **E.** GUS staining of nodules transformed with *SPC25* promoter-*GUS* construct. Arrowhead indicates GUS staining in the roots. Arrows indicate GUS staining in the nodule. Scale bars= 1mm.

Table 2.1. Organ Specific Element (OSE) and OSE-like motifs in the promoters of various genes with nodule-specific expression.

Second element	Bases between elements	First element	Bases upstream of the start of the transcript	Gene	Species
AAAGAT	6	TTCTT	107	*LBA*	*Glycine max*
AAAGAT	6	CTCTT	102	*LBC1*	*Glycine max*
AAAGAT	7	CTCTT	102	*LBC2*	*Glycine max*
AAAGAT	6	CTCTT	115	*LBC3*	*Glycine max*
AAAGTT	6	CTCTT	170	*LB1*	*Medicago truncatula*
AATGAT	6	CTCTT	178	*LB2*	*Medicago truncatula*
AAAGAT	27	AAAGAT	273	*DNF1*	*Medicago truncatula*

A *trans*-factor binds to the OSE-like motifs in the *DNF1* promoter

If these OSE-like motifs in the *DNF1* promoter are involved in the transcriptional regulation of *DNF1*, then a *trans* factor should bind them. We performed an electrophoretic mobility shift assay (EMSA) with a 49-bp probe that contains both SOLE motifs identified from the 420bp fragment of the *DNF1* promoter. We detected binding when the probe was incubated with total protein extract from nodules, indicated by a radiolabeled band of lower mobility (Figure 2.6A). This binding suggests that these motifs could indeed be functional *cis*-elements.

To understand if these motifs are specifically involved in *DNF1* expression in the nodule, we qualitatively compared the binding pattern with protein extracts from roots and nodules. We found that the shift caused by nodule protein extracts was greater in intensity than that by root protein. Furthermore, the intensity of the shifted band increased with an increasing amount of nodule protein, indicating that a *trans*-factor that is enriched in the nodule can bind this element (Figure 2.6A).

We also performed a competition assay and found the radiolabeled probe can be competed away with a 25 and 50 molar excess of unlabeled probe, suggesting that recognition of the *DNF1* probe is sequence-specific (Figure 2.6B). These two experiments provide evidence for the existence of a nodule *trans*-factor that specifically binds to the *DNF1* promoter fragment, which contains the SOLE motifs.

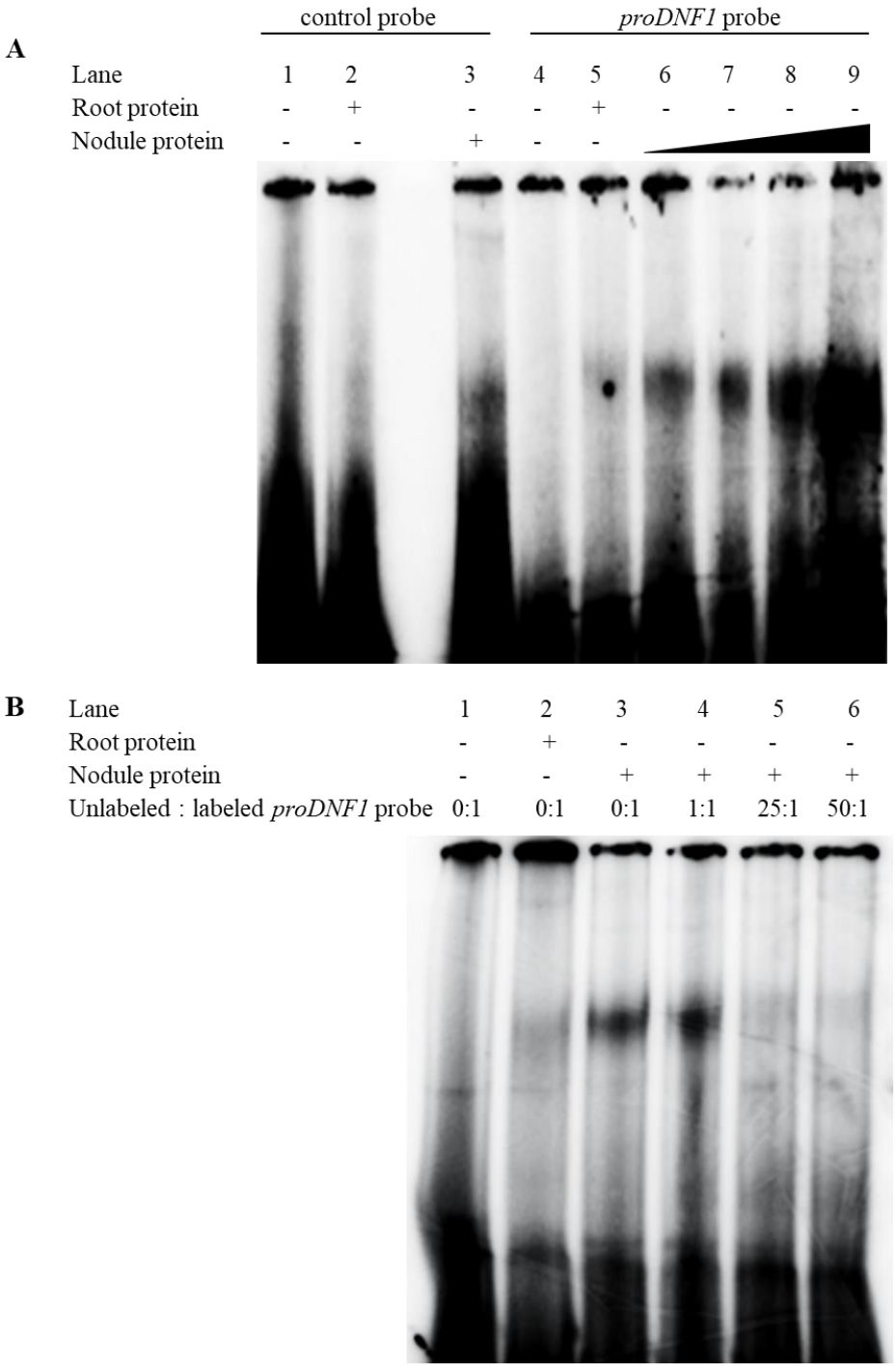


Figure 2.6. Gel shift assays using probe from the *DNF1* promoter. A. EMSA of *DNF1* promoter probe incubated with root or nodule protein. Row 1 indicates the labeled probe used. Row 2 shows the lane number. Row 3 indicates presence (+) or absence (-) of root protein. Row 4 indicates presence, absence, and relative amount of nodule protein. **B. Competition assay using labeled and unlabeled *DNF1* promoter probe.** Row 1 indicates the lane number. Row 2 indicates presence (+) or absence (-) of root protein. Row 3 indicates presence or absence of nodule protein. Row 4 indicates ratio of unlabeled to labeled probe.

The two SOLEs in the *DNF1* promoter are important for nodule-specific expression

To evaluate the biological importance of this OSE-like motif, we mutated the two 'AAAGAT' sequences in the 420bp *DNF1* promoter-*GUS* construct. We mutated both 'AAAGAT' sequences to account for the possibility that they function as a single unit. A random sequence generator ensured that every position of the SOLE was changed, turning each 'AAAGAT' sequence to 'CCTACA.' We transiently transformed both the wild-type 420bp promoter-*GUS* construct and the mutated version into *M. truncatula* hairy roots. Among about 40 *M. truncatula* plants (two sets of 20), each with multiple independently transformed hairy roots, we detected no discernable *GUS* expression in the nodules transformed with the mutant promoter, although both sets of plants had some staining in the roots (Figure 2.7). The results suggest that the OSE-like motif in the *DNF1* promoter is required for its expression in the nodule.

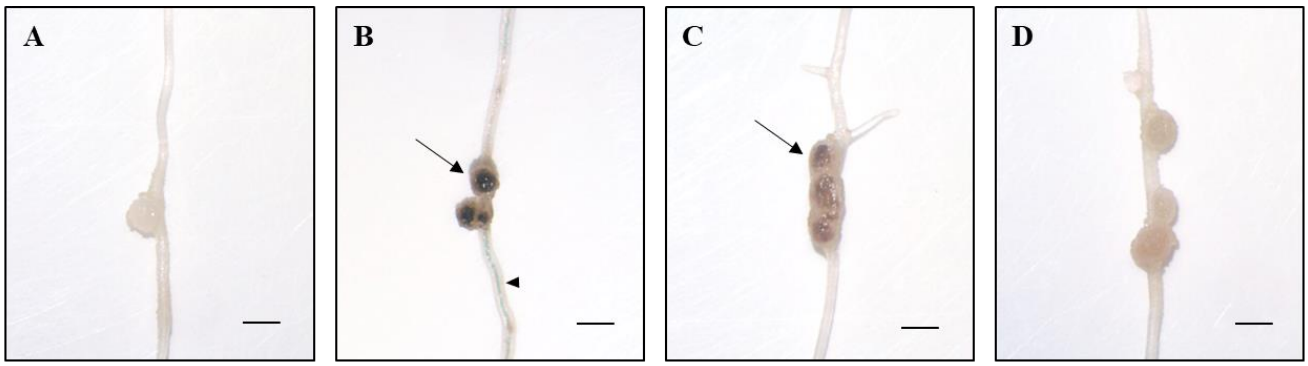


Figure 2.7. *DNF1* promoter fragments drive *GUS* expression in the nodules. A. *M. truncatula* nodule transformed with empty vector. **B.** Nodules transformed with the full length *DNF1* promoter *GUS* construct. **C.** GUS staining of nodules transformed with a ~400bp *GUS* construct. **D.** GUS staining of nodules transformed with a mutated version of the ~400bp *GUS* construct, where both 'AAAGAT' sequences were replaced with 'CCTACA'. Arrow heads show GUS staining in the roots, arrows indicate GUS staining in the nodules. Scale bar represents 1mm.

The SOLE is common among symbiotic SPC genes

Specialized promoter motifs may underlie the different expression patterns between *DNF1* and *DNF1L* (Figure 2.1). When we analyzed the *DNF1L* promoter for the SOLE, we found no instance of this sequence up to 3,000 bp upstream of the transcriptional start site. The lack of the SOLE in the promoter of the housekeeping *DNF1L* prompted us to hypothesize that this sequence may similarly be present in the promoters of other symbiotic SPC genes but absent in the promoters of their nonsymbiotic counterparts. These "symbiotic" SPC genes, including *DNF1*, are coordinately upregulated during nodulation (Wang et al., 2010).

In addition to $SPC22^{DNF1}$, the SPC contains three other subunits, SPC12 (DAS12), SPC25 (DAS25), and SPC18 (DAS18). We found the SOLE in the promoters of all the symbiotic genes that are co-regulated with *DNF1*. Uniquely for *SPC18*, we also saw this element downstream of the transcriptional termination site (Figure 2.8). For *SPC25*, the promoter was able to induce GUS staining in a nodule-specific manner. However, *SPC12* and *SPC18* promoter-*GUS* constructs were not able to induce nodule-specific expression (Figure 2.5), although the presence/ absence of GUS staining in nodules transformed with the *SPC12-GUS* construct was inconsistent.

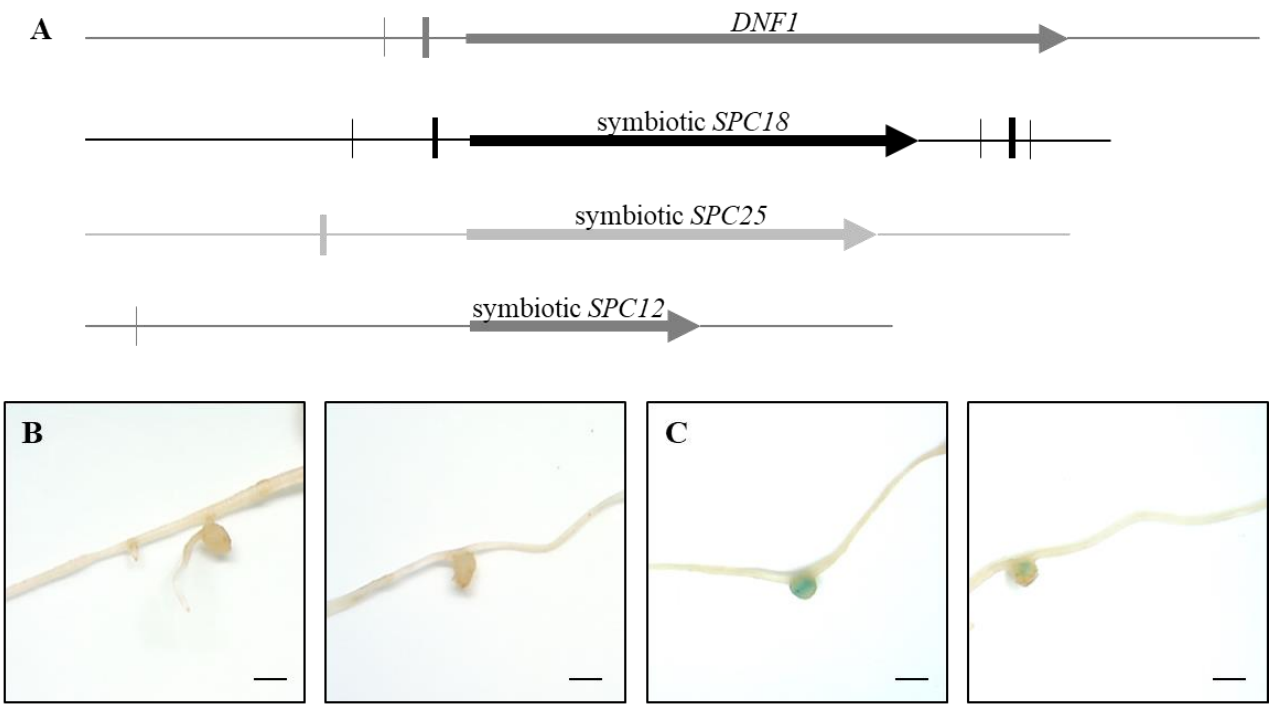


Figure 2.8. The region downstream of SPC18 is required for nodule-specific expression. B. GUS staining of *M. truncatula* roots transformed with a 2750bp *SPC18* promoter GUS construct. **C.** GUS staining of *M. truncatula* roots transformed with a 2750bp *SPC18* promoter GUS construct with 1300bp of the region downstream of *SPC18* on the 3' end of GUS. The arrows point to the GUS staining. Scale bars= 1mm.

SPC18 regulation

We wanted to take a closer look at the *SPC18* promoter because it is the catalytic subunit of the SPC and likely required for signal peptide cleavage in the nodule. We made a promoter-*GUS* construct with the symbiotic *SPC18* promoter, using the sequence approximately 2,750 bp upstream of the transcriptional start site. When this construct was tested in transformed roots, it produced no discernable GUS staining in the nodules, despite the presence of OSE-like motifs in the promoter region that are reminiscent of the *DNF1* promoter. The presence of GUS activity in transgenic root vasculature indicated that our transformation was successful. The experiment was repeated five times with the same result.

Since regulatory sequences are sometimes found in the first introns of genes, we made a second construct that included the first exon and first intron of the *SPC18* gene in addition to the promoter region. This second construct was about 4.5 kilobases long and, like the promoter-only construct, did not result in GUS staining in the nodules of composite plants. One SOLE is found in the sixth intron, but we did not examine its functional significance.

Unlike *DNF1*, symbiotic *SPC18* has four copies of the SOLE within 1,300 bp downstream of the transcriptional termination site. Using the *SPC18* promoter-*GUS* construct, we inserted this region downstream of *SPC18* behind the *GUS* gene. This new construct was able to induce nodule-specific GUS staining in composite plants (Figure 2.8). This result indicates that the downstream region of SPC18 is important for module-specific expression, which is distinct from the other symbiotic SPC genes.

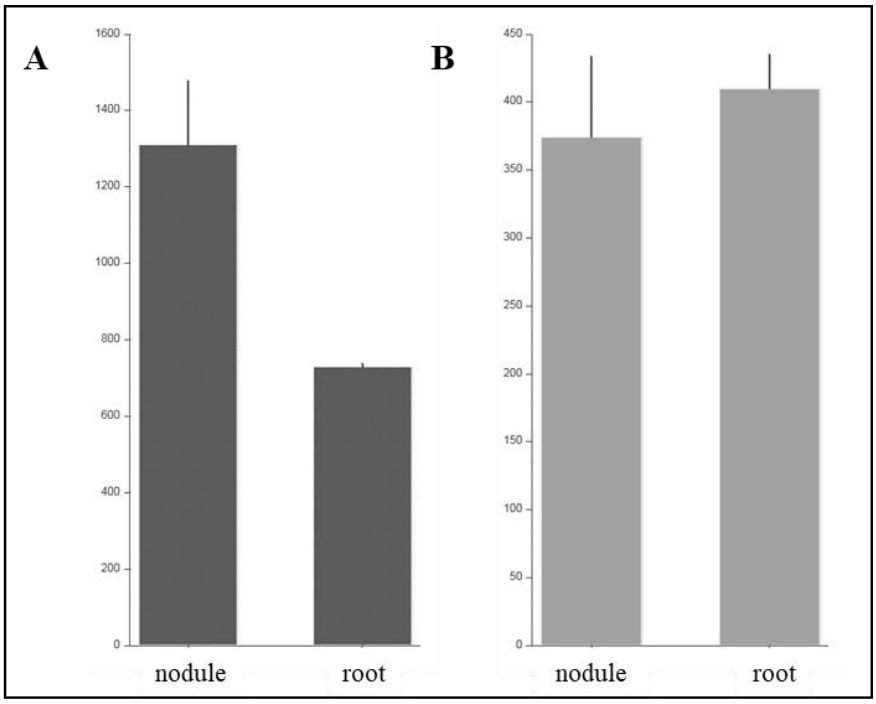


Figure 2.9. *SPC18 symbiotic* and *SPC18 housekeeping* expression in nodules and roots. A. RNAseq expression data of *SPC18* that is co-regulated with *DNF1*. **B.** RNAseq expression data of *SPC18* that is absent from the *M. truncatula* v. 4.0 genome. Data obtained from Symbimics (https://iant.toulouse.inra.fr/symbimics/).

Table 2.2. Loci of SPC subunit genes in two versions of the *M. truncatula* genome

SPC subunit	***M. truncatula* locus (v. 4.0)**	***M. truncatula* locus (v. 5.0)**
SPC22	*Medtr3g027890*	*MtrunA17_Chr3g0087791*
SPC22	*Medtr7g113740*	*MtrunA17_Chr7g0273011*
SPC18	*Medtr2g103570*	*MtrunA17_Chr2g0332921*
SPC18	N/A	*MtrunA17_Chr2g0307071*
SPC12	*Medtr3g085510*	*MtrunA17_Chr3g0122651*
SPC12	*Medtr8g080380*	*MtrunA17_Chr8g0377251*
SPC12	*Medtr3g074500*	*MtrunA17_Chr3g0115811*
SPC25	*Medtr5g081900*	*MtrunA17_Chr5g0436541* *MtrunA17_Chr5g0436561*
SPC25	*Medtr8g093950*	*MtrunA17_Chr8g0384471*

Functional confirmation of *SPC18* through CRISPR knockout

SPC18 is the catalytic subunit of the signal peptidase complex. The multitude of *cis*-elements, both up- and downstream of the transcribed region, suggests that it is under robust regulation, which hints toward a critical function of this gene. When we initially investigated the *SPC18* gene(s) using the Medicago v. 4.0 genome, we discovered only the gene studied here (i.e., the *SPC18* co-regulated with *DNF1*, Krishnakumar et al., 2015; Young et al., 2011). If the genome indeed contains a single gene encoding the catalytic subunit of SPC, it should be essential for viability and would thus preclude a functional investigation of its role in the symbiosis.

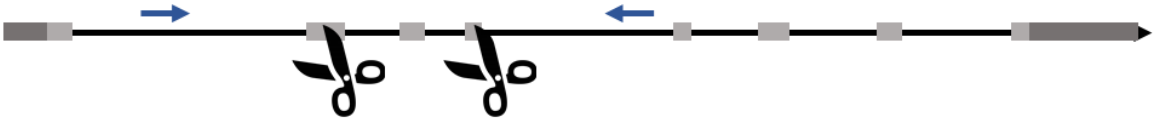
As all other SPC subunits are encoded by small gene families, we found it intriguing that the most critical subunit, SPC18, should be encoded by a single gene. We noticed that the microarray used for the Medicago Gene Expression Atlas contains probes matching a second transcript annotated as *SPC18,* even though the corresponding genomic sequence is absent in Mt. 4.0 (Benedito et al., 2008). This second *SPC18* transcript is also present in the Symbimics RNAseq database, and its expression pattern is both distinct from its paralog and reminiscent of *DNF1L* (Roux et al., 2014, Figure 2.9). This raised the possibility of an unannotated or missing *SPC18* gene in the 4.0 assembly.

To further investigate the possible presence of this second copy of *SPC18*, we used the *M. truncatula* R108 assembly and also found two copies of *SPC18*. This further supported the idea the A17 genome indeed encodes two copies of SPC18, but one is absent from the v. 4.0 assembly. As the newly released Medicago v. 5.0 genome made extensive use of long-read sequencing technology such as PacBio, we sought to ask whether the more complete assembly of Medicago v. 5.0 can resolve the number of

SPC18 genes (and their positions) in A17 (Pecrix et al., 2018). We indeed found two SPC18-encoding genes in the v. 5.0 genome, the one whose promoter was characterized above, and a new one, *MtrunA17_Chr2g0307071* (Table 2.2).

Based on its expression pattern, we conclude that this second *SPC18* gene likely performs housekeeping functions as opposed to a specialized role in symbiosis, akin to the functional divergence between *DNF1* and *DNF1L*. The presence of a housekeeping copy of *SPC18* implies the symbiotically induced *SPC18* gene would be dispensable for viability, which should allow us to validate its function in the symbiosis.

A



B

GCATGATTGTTACGTCTGCACTGATAATATGGAAGGCATTGATGTGCATCACTGGCAGTGAATCT
CCTGTTGTCGTTGTTCTTTCTGGAAGTATGGAACCTGGCTTCAAACGGGTAAGCAAATCTTATCA
TCGAAGCCTCGGTAATTCTGTATCTAATATTTTTCTACAACAACAGTACAGTGTCGTCAAGAAGA
GACAAGCTGCATATATAGATTCAAATCATTTGTCAATATAAACTGTCTGTTTCTTTTTGTTGCTA
ATGAATGTTGGTTTTTGAAGGGGGACATCTTGTTTTTACACATGAGTAAGGATCCAATTCGCGCA
GGAGAGATTGTTGTATTTAATATAGATGTAAGTTCACCTCGCTGTTTTATATCCTTTATGTTTTG
CTTACTTCGATTCATATCTATATATCAGCCGATTCCAATTCTTTGTATTTTGCTTGATACTTTCC
GTACTGACAAAGTTTGTTGGAATGTACCCTTTTTTTTCTCTAGGGACGAGAGATTCCAATCGTGC
ATCGTGTAATTAAGG

Figure 2.10. Overview of guide RNA design for the symbiotic *SPC18* CRISPR construct in hairy roots. A. Representation of the symbiotic *SPC18* gene in *M. truncatula*. The boxes represent exons, the lines are the introns. The light gray boxes indicate the coding region; the dark gray boxes are the untranslated regions. The scissors show approximately where the Cas9 cuts. The blue arrows denote the genotyping primers. **B.** Partial genomic sequence of the symbiotic *SPC18* gene in *M. truncatula*. The gray highlighted regions indicate the second, third, and fourth exons. The red highlighted sequences are the guide RNA sequences; the PAM is shown in red letters.

SPC18 mutagenesis

Mutations in the symbiotic *SPC18* as insertion/deletion lines in the Noble Research Institute's *Tnt1* and FNB databases were not available at the time of this study. However, we hypothesized that a mutant of this gene should be viable, and thus decided to knock it out via CRISPR/Cas9, as gene silencing might affect both *SPC18* genes.

To improve the applicability of the CRISPR/Cas9 technology in hairy root transformation, we opted for a multiplex approach, using two guide RNAs to induce an easily detectable deletion within the target gene, using the direct cloning vectors described in (Čermák et al., 2017). We targeted exons 2 and 4 of *SPC18* with guide RNA sites that are 483bp apart in the genome (Figure 2.10). DNA was first extracted from three pools of hairy roots, one transformed with the empty vector control and two with the *SPC18* CRISPR construct. Using genotyping primers that flank the two gRNA target sites, we detected a PCR product consistent with the expected deletion *SPC18* from SPC*18* CRISPR Pool #1 DNA (Figure 2.11). Sequencing of this product revealed that the sequence between the two gRNA sites was indeed removed. Pool #1 consists of hairy roots with white, small (Fix-) nodules (Figure 2.11B). Pool #2 is from SPC*18* CRISPR-transformed roots with pink nodules (Figure 2.11C). This and the empty vector pool only amplified the WT-sized band.

We subsequently genotyped a total of 45 independently transformed hairy roots. Among them, four roots clearly lacked the WT-sized PCR product (Figure 2.12). Of these, root #6 predominantly amplified a band with the size expected of a deletion between the two gRNA sites. Sequencing the PCR product confirmed the nature of the deletion (Figure 2.13). However, we found only one plant with roots containing this

expected deletion. This is most likely because the second guide RNA was less efficient at inducing mutations compared to the first gRNA. We found two examples of roots that had mutations at the first gRNA (root #21 and root #24), but no mutations at the second, a result revealed by sequencing the corresponding PCR products. We found no roots with mutations at only the second gRNA. Furthermore, we saw two different large deletions that were presumably caused by breaks at both gRNA binding sites (root #6 and root #10).

After sequencing the *SPC18* PCR product from root #21 (which had a mutation at the first gRNA), we realized that the mutation was an in-frame, three-base-pair deletion. Yet this root had only small, white nodules resembling a null mutation (such as the *dnf1* phenotype, Figure 2.14). To confirm that this was the only mutation in root #21, we cloned the PCR product(s) into pMD19 (Takara Bio), sequenced plasmids from eight individual colonies, and found only the same three-base-pair deletion in each one. This indicates that this root is homozygous for this allele.

The in-frame deletion found in root #21 caused the deletion of a single isoleucine in the SPC18 protein. A closer look at this amino acid in the SPC18 proteins from other species such as Saccharomyces, Limulus, Chlamydomonas, and Oryza showed that this residue is a conserved hydrophobic amino acid, although it is methionine in many species including Chlamydomonas. Interestingly, the Chara, Physcomitrella, Oryza, Arabidopsis, Medicago, and Glycine proteins have a conserved isoleucine at that position (Figure 2.14). This example of a random mutation in the symbiotic-specific *SPC18* in Medicago may shed some light on a critical residue important for its proper function. Because this

gene is symbiotic-specific and not required for the viability of the organism, we may be able to use this system to study the SPC in more detail.

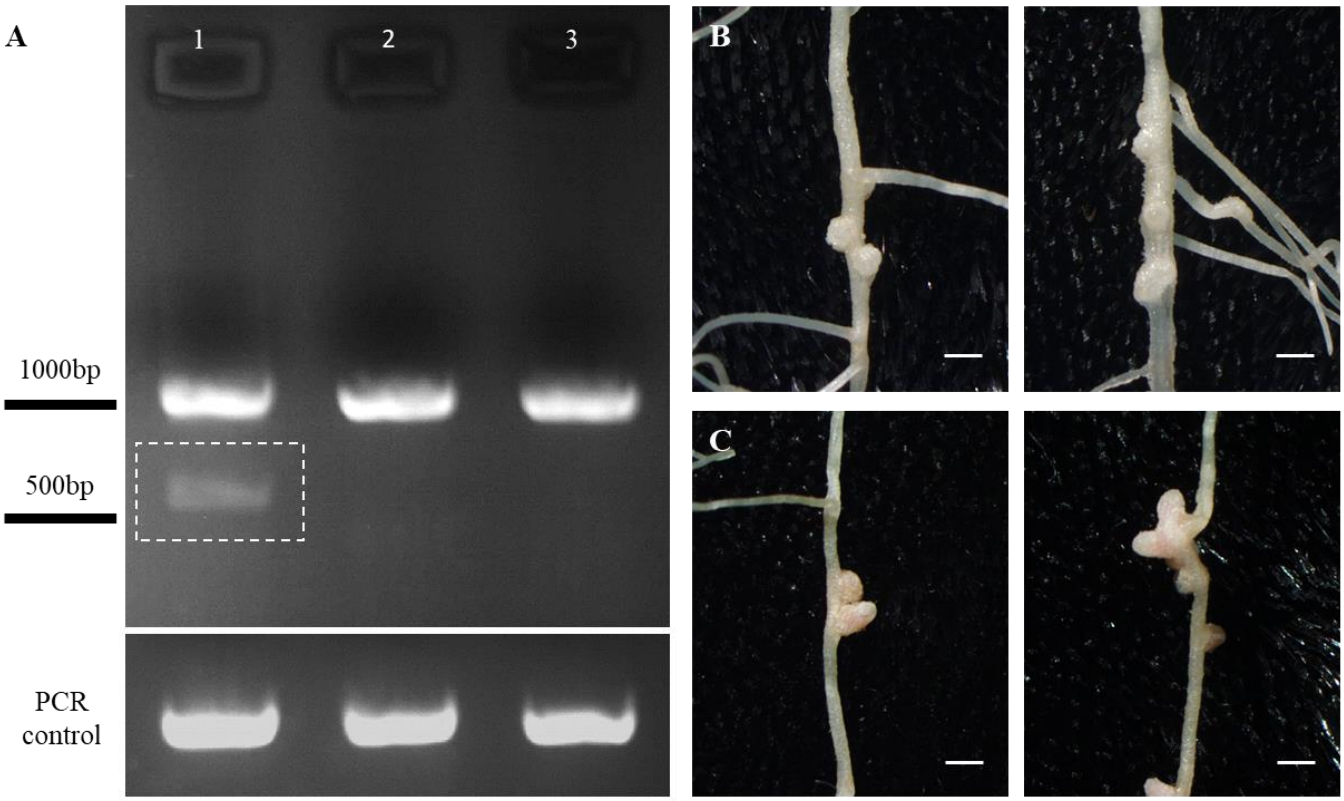


Figure 2.11. *M. truncatula* hairy roots transformed with *SPC18* CRISPR gRNA construct. A. Genotyping PCR of pooled hairy roots. Lane 1 is PCR of *SPC18* using genotyping primers with *SPC18* CRISPR transformed hairy root DNA from Pool #1. Lane 2 is PCR from Pool #2 of *SPC18* CRISPR transformed hairy roots. Lane 3 is PCR from empty vector transformed hairy roots. Dashed box indicates the mutant *SPC18* band. Bottom portion shows PCR of a genomic region as a control, the DNA used in this PCR is the same that was used in lanes 1-3. **B.** Examples of nodules on hairy roots from Pool #1. **C.** Examples od nodules on hairy roots from Pool #2. Scale bars= 1mm.

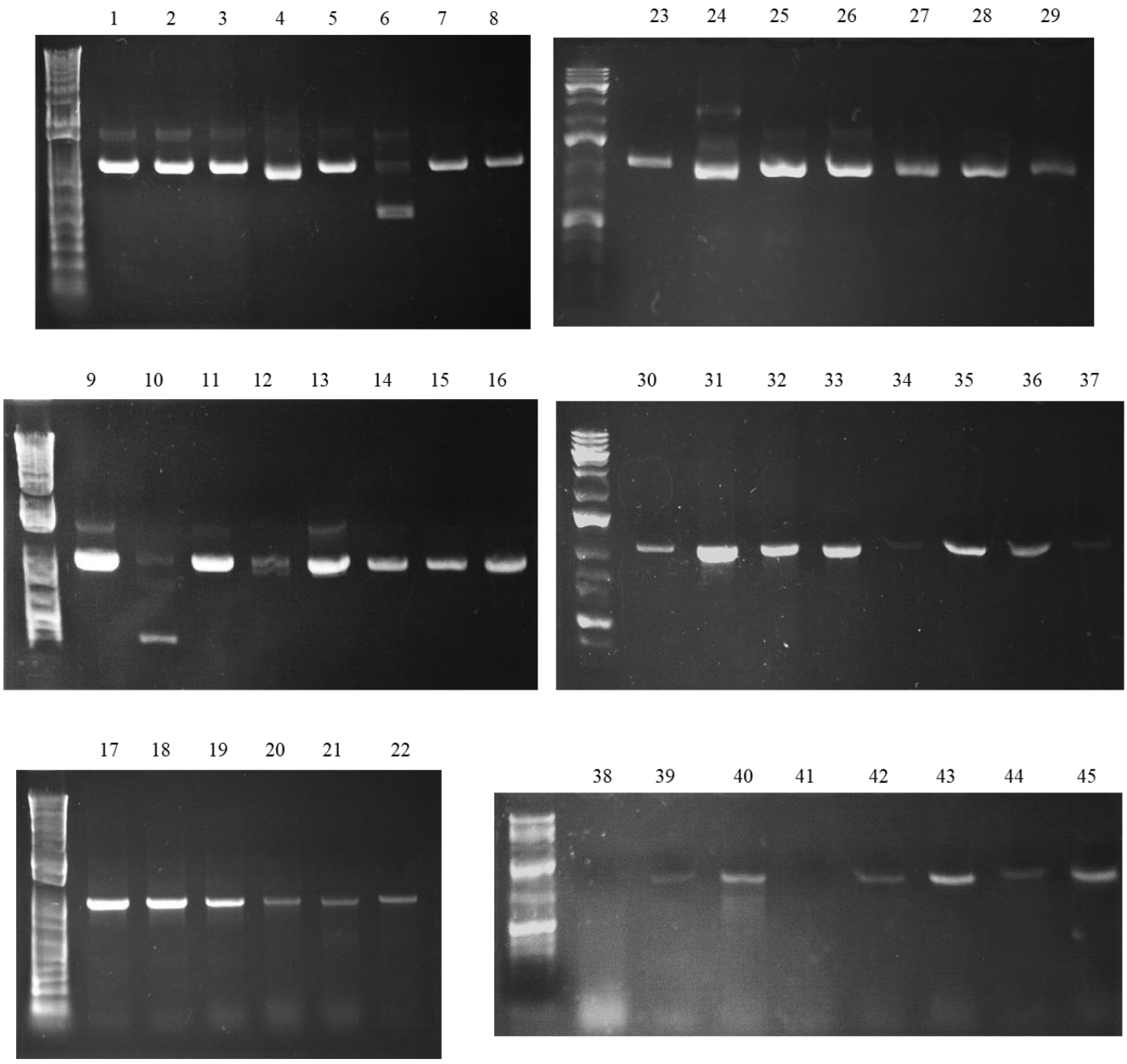


Figure 2.12. *SPC18* PCR of independently transformed *M. truncatula* hairy roots. The first lane of each DNA gel is a 1kb DNA ladder. The numbers indicate an independently transformed root.

Figure 2.13. Partial DNA sequence MUSCLE alignment of root #4, transformed with the *SPC18* CRISPR construct. The top sequence is the reference and the bottom sequence is from root #4. Red outlined boxes are the guide RNA sequences bind; the yellow highlighted boxes denote the PAM sequence.

SPC18 is required for the nitrogen-fixing symbiosis

Roots lacking the WT-sized PCR product all produced white, unelongated nodules, which is reminiscent of the *dnf1* mutant (Wang et al., 2010). These roots represent several different alleles of *SPC18,* and some were shown to be homozygous for the given mutation. To confirm our hypothesis that *SPC18* is required for bacteroid differentiation, we used SYTO9 staining to observe bacteroids inside WT and mutant nodules with confocal microscopy. Nodules with a functional *SPC18* have large, elongated bacteroids that fill nodule cells around a central vacuole. In comparison, the mutant nodules are not only small and white, but the bacteroids contained within host cells are small, round, and sparsely populated (Figure 2.15). Indeed, the bacteroids in the mutant nodules exhibit a morphology similar to those of *dnf1* mutant nodules. This implies that the bacteria in *spc18* roots are not differentiated and, therefore, cannot fix nitrogen (Mergaert et al., 2006b; Van de Velde et al., 2010; Wang et al., 2010).

Figure 2.14. A three base-pair deletion in symbiotic *SPC18*, resulting in the deletion of an isoleucine, causes nonfunctional nodules in *M. truncatula*. A. Amino acid alignment of SPC18 in various species. The amino acid deletion in one of the CRISPR-mutated roots (#22) is indicated in bold and by the orange arrowhead. The arrow indicates the catalytic site. **B.** Picture of a nodule on a wild-type root of *M. truncatula* (root #21). **C.** Nodules on CRISPR-mutated root (root #22) with the three base-pair deletion resulting in a deletion of an isoleucine. Nodules were imaged 40 days post inoculation. Scale bars= 1mm. **D**. Sequencing results and alignment of PCR from root #22 showing the three base-pair deletion when aligned with the reference *SPC18* sequence (third row).

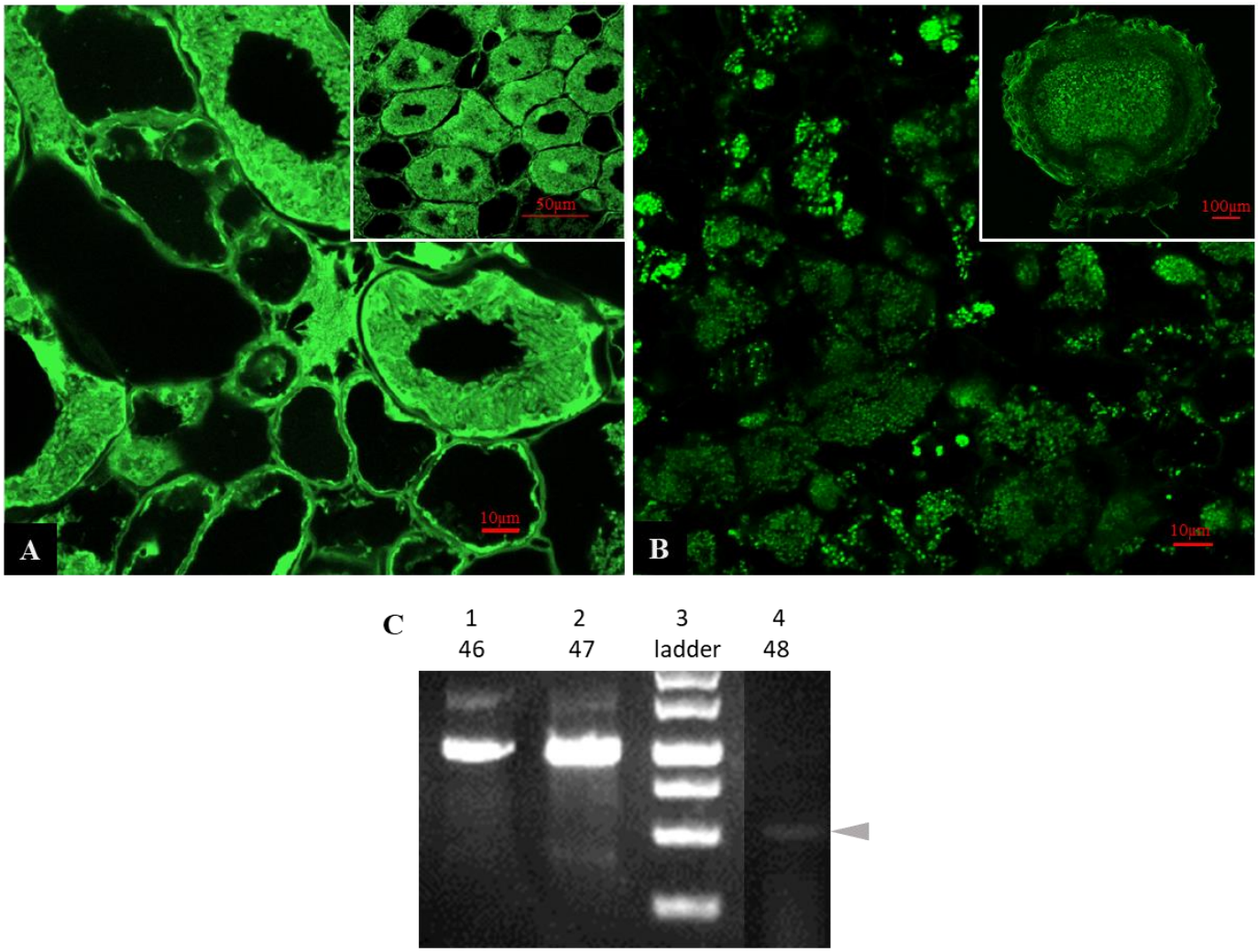


Figure 2.15 Intracellular bacteria are not differentiated in *SPC18* knockout nodules. A. Confocal image of SYTO9 stained nodule cells on a wild-type *SPC18* root (formed on a CRISPR-transformed hairy root), scale bar= 10µm. Inset shows the same nodule at a lower magnification, scale bar= 50µm. **B.** Nodule cells from an *SPC18* knockout root with the expected deletion between the two gRNAs, scale bar= 50µm. Inset shows a cross-section of the whole nodule, scale bar= 100µm. **C.** *SPC18* PCR from transformed roots. Lane 1 is an empty vector control transformed root (root #46), this shows the wild-type size of the PCR product. Lane 2 is PCR with root DNA from root #47, the root used in (A). Lane 3 is a 1kb DNA ladder. From top to bottom the sizes are as follows: 2000bp, 1500bp, 1000bp, 750bp, 500bp, and 250bp. Lane 4 is the PCR from root #48, the root used in (B). The grey arrow indicates the mutant *SPC18* band. Two lanes were cropped between lanes 3 and 4 for clarity, but the position of the band relative to the ladder was not changed.

Discussion

Plant-microbe symbioses evolve by frequently recruiting genes from other processes and co-opting them to meet specific requirements in cross-kingdom interactions. As such, there are many examples of legume-specific and even lineage-specific clades of genes involved in nitrogen-fixing symbiosis. One classic example of a legume-specific gene clade is leghemoglobin. Hemoglobin genes are found in plants in general and possibly serve to regulate oxygen potentials. Hemoglobins expanded in legumes and are necessary for nitrogen-fixing symbiosis by sequestering free oxygen within the nodule to allow proper nitrogenase activity (Ott et al., 2005). Expansion and diversification of the hemoglobin clade led to the neofunctionalization of leghemoglobin in symbiosis.

Retention of duplicated genes and their subsequent neo- or subfunctionalization for symbiosis is also common within specific legume lineages. A well-studied example of this is the NCR peptides in the IRLC legumes, including Cicer and Medicago, although there is great variability in both the number of these genes between species, as well as their sequence (Montiel et al., 2017, 2016). Recently, (Trujillo et al., 2019)colleagues (2019) identified several groups of genes with nodule-specific expression that underwent a lineage-specific expansion in certain legumes, such as PLAT domain proteins in Medicago and NCR-like peptides in Aeschynomene, indicating that symbiotic-specific genes are continuously being recruited.

One common feature shared by most legumes seems to be the redirection of secretory proteins to the symbiosome. Many nodule-specific proteins are targeted to the symbiosome through the protein secretory pathway, including NCR peptides and Nodulin

25 (Hohnjec et al., 2009b; Van de Velde et al., 2010). Many other nodule-specific proteins are also assumed to be trafficked to the symbiosome based on the presence of a signal peptide. Most likely due to the increased capacity required to traffic more proteins through the endomembrane system, the ER dramatically expands in nodule cells. One important aspect related to this ER expansion involves transcriptional upregulation of genes involved in protein secretion, including *VAMP72d/e*, *SYP132A* and the SPC complex (Huisman et al., 2016; Ivanov et al., 2012; Maunoury et al., 2010; Pan et al., 2016; Wang et al., 2010).

We have investigated the mechanism involved in the transcriptional upregulation of SPC genes in the nodule. Based on this study, we have identified a *cis*-element reminiscent of the OSE in the promoters of nodule-specific SPC genes, which we call the SOLE. The SOLE was first identified in the promoter of *DNF1* and was found to be important for its nodule-specific expression. This motif was also present in the promoters of the other nodule-specific SPC genes, namely, *SPC12*, *SPC25*, and *SPC18*. Interestingly, the *SPC18* promoter alone was not sufficient to drive *GUS* expression in hairy root nodules, and we discovered that sequences downstream of the transcriptional termination site in SPC18, which contain an additional SOLE, found, was required for nodule-specific expression of *SPC18*. At present, we do not know whether this downstream sequence is sufficient to confer a nodule-specific expression pattern. Given that SPC18 is the catalytic subunit of the complex, it would not be surprising if its expression is under tighter control, and that elements both up- and downstream of the transcribed region have regulatory roles. The SOLE thus points to a conserved DNA element important for the nodule-specific expression to SPC genes. It is likely related to

the OSE, which is necessary for nodule-specific expression in leghemoglobin genes across various legumes.

We speculate that the *trans*-factor that binds the SOLE (and also the OSE) is NIN, which is a central regulator of nodulation present in diverse plant species that form nitrogen-fixing symbiosis. NIN has been shown to bind promoters of several genes important for rhizobia infection and nodule organogenesis, such as *RINRK1*, *CRE1*, and *NF-YA1* (Li et al., 2019; Soyano et al., 2013; Vernié et al., 2015). Furthermore, NIN binding sites have been defined, and one such site in the *LjNF-YB1* promoter (in *Lotus japonicus*) is 'ATCTTT,' the reverse complement of the SOLE (Soyano et al., 2013). Similarly, in Medicago, NIN was shown to bind the NF-box, which is 'AAG(A/C)T,' originally found in the *ENOD11* promoter (Vernié et al., 2015).

Furthermore, the related NIN-Like Proteins (NLPs) in Arabidopsis bind the Nitrate Responsive *cis*-Element (NRE) present in the promoter of genes such as nitrite reductase. The NRE contains two pseudo-palindromic sequences, including 'AAGAG,' which is reminiscent of the SOLE (AAAGAT) and the exact reverse complement of part of the OSE ('CTCTT;' Konishi and Yanagisawa, 2013). The similarity between the NIN binding site(s) and the OSE and SOLE indicates that these *cis*-elements may be variants of a core sequence that is recognized by protein complexes containing NIN or the closely related NLPs. Furthermore, the repeated acquisition of NIN binding sites in the promoters of multiple genes suggests it is part of a common path of neofunctionalization for genes to be recruited for nitrogen-fixing symbiosis.

The evolutionary path of gene duplication followed by functional diversification of *DNF1* is similarly observed in other SPC genes. Here we have shown that paralleling

the evolution of SPC22 genes, two copies of *SPC18* genes exist in the *M. truncatula* genome. Of the two paralogs, one has an expression profile that is correlated with *DNF1* (Wang et al., 2010), and the other paralog has an expression profile that points to a housekeeping function, similar to *DNF1L*. However, unlike *DNF1*, we found that the induction of the symbiotic *SPC18* requires regions downstream of the gene, and it is likely that the sequences both upstream and downstream of the coding region are needed for proper expression due to the presence of SOLEs in those areas. The importance of the downstream region in *SPC18* is reminiscent of the transcriptional regulation of the symbiotic t-SNARE, *SYP132A*. The *SYP132A* expression is highly correlated with *DNF1*, but it is produced by alternative termination and polyadenylation, which indicates that sequences close to the termination site are required for its genesis (Huisman et al., 2016; Pan, Oztas et al., 2016). It is possible that certain unknown features of the chromatin near the transcriptional termination site underlies the regulation of both *SPC18* and *SYP132A*.

For the functional study of the symbiotic *SPC18*, we decided to adapt CRISPR/Cas9 for Medicago hairy roots in order to examine any phenotype caused by a null mutation. CRISPR/Cas9 has been used for reverse genetics in Lotus and Medicago since 2016 and 2017, respectively (reviewed by Roy et al., 2019). For example, the function of a PLAT domain-encoding gene, *NPD1*, was investigated using a CRISPR/Cas9 knockout. This gene was part of a family that underwent a lineage-specific expansion in Medicago, and it was found that *NPD1* is required for efficient nitrogen fixation, but the nodule phenotype is strain-dependent (Pislariu et al., 2019). The drawback to using CRISPR in this way, however, is that stable transformation takes months, and it takes even longer to evaluate the symbiotic phenotype (if there is one).

Existing methods for using CRISPR/Cas9 in hairy roots also has its shortcomings because genotyping heterogeneous hairy roots can be tedious: single guide RNAs tend to cause unpredictable, small mutations. Establishing that one sample has lost both WT copies of the gene often requires processing the PCR product (by annealing and digestion or subcloning followed by sequencing) and finding multiple homozygous/bi-allelic roots among a large number of hairy roots is often labor-intensive (Cai et al., 2015; Wang et al., 2017; Yang et al., 2017). Given the short life span of *M. truncatula*, it is often not feasible to carry out genotyping and phenotyping sequentially, which necessitates the parallel characterization of roots that subsequently proven to contain WT alleles. The simplest approach makes use of restriction enzyme sites affected by Cas9 cleavage, which constrains the choice of gRNA target sequences (Cai et al., 2015)**.** Sometimes roots are screened first for phenotypes when the effect of a loss-of-function mutation can be easily predicted (Wang et al., 2016), but the suitability of such an approach in reverse genetics is clearly limited.

Our rationale for determining the function of SPC*18* via CRISPR/Cas9 in hairy roots was to identify roots with homozygous or biallelic mutations first and then evaluate their phenotype. Therefore, we utilized paired guide RNAs, which, when working together, should result in large deletions of predictable sizes, and make the absence of wild-type alleles easier to ascertain (Čermák et al., 2017). We applied this strategy in hairy roots, reasoning that roots carrying such bi-allelic deletions should be considered functional nulls, even when the exact mutations are heterogeneous. This genotyping approach is efficient enough to allow subsequent phenotyping on confirmed functionally null samples.

Methods

Plant growth and inoculation

M. truncatula seeds (Jemalong or Jemalong A17 accessions) were germinated on water agar plates at 30ºC overnight in the dark. They were placed on half sand/half perlite and grown in a Percival growth chamber at 20ºC 16-hour/8-hour light/dark cycle. Plants were inoculated with *Sinorhizobium medicae* strain ABS7 suspended in ½ buffered nodulation media (BNM; 2mM $Ca(SO_4)_2$, 0.5mM KH_2PO_4, 0.5mM $Mg(SO_4)_2$, 50µM Na_2EDTA, 50µM $FeSO_4$ x $7H_20$, 16µM $ZnSO_4$ x $7H_20$, 50µM H_3BO_3, 50 µM MnSO4, 1 µM Na_2MoO_4 x H_20, 0.1µM $CuSO_4$, 0.1µM $CoCl_2$ x $6H_20$, and 2mM MES hydrate, pH = 6.5) with a final optical density (600nm) of .05. Each plant received approximately 10mL of the rhizobia solution. Plants and nodules were harvested at approximately three weeks post-inoculation unless otherwise indicated.

Plasmids and vector construction

The promoter-*GUS* plasmid constructs were all made using the Gateway compatible plant vector, pMDC163 (Curtis and Grossniklaus, 2003). The SPC gene promoters were amplified from Jemalong A17 genomic DNA and first cloned into either pCR8/GW/TOPO vector (Invitrogen) through TOPO TA cloning, or pDONR/Zeo vector (Invitrogen) through BP clonase Gateway recombination. From either pCR8/GW/TOPO or pDONR/Zeo, the promoter sequences were introduced into pMDC163 via an LR clonase reaction (Invitrogen). For the mutant version of the 420bp *DNF1* promoter-GUS construct, the sequence was synthesized by Genscript into pUC57. Primers were then

used to amplify the sequence and add the recombination sites to clone into pDONR/Zeo, as described above.

The *dnf1* mutant complementation plasmid constructs, with either *DNF1* or *DNF1L* genomic sequences, were inserted into pMDC100 (Curtis and Grossniklaus, 2003) via Gateway cloning, through the pCR8/GW/TOPO vector as described above. The pMDC100 promoter (CaMV35S) was replaced by the full-length *DNF1* promoter through restriction enzyme cloning.

The *SPC18* CRISPR construct was made using Golden Gate Assembly into pDIRECT_22C, which was gifted to us from the Voytas lab (Čermák et al., 2017). Two guide RNAs targeting exon two and four of symbiotic *SPC18* were designed using ATUM design software and assembled into the vector using the guidelines from the Voytas lab website (http://crispr-multiplex.cbs.umn.edu/). The primers, including the gRNA sequences, were designed using the Voytas lab website, and PCR was carried out using the pDIRECT_22C as the template DNA. The SapI restriction enzyme (New England Biolabs) was used to linearize the plasmid backbone and were added to two of the PCR products through the primers. Esp3I was used to ligate the other PCR products together. Golden Gate Assembly was carried out with pDIRECT_22C, the PCR products, T4 DNA ligase, Cutsmart restriction enzyme buffer (New England Biolabs), and ATP for 30 cycles using the protocol from New England Biolabs.

All of the constructs were first transformed into DH5a chemically competent *Escherichia coli.* Resulting colonies were screened for the correct plasmid using colony PCR and subsequent sequencing through Psomagen. The constructs were transformed

into *Agrobacterium rhizogenes* strain ARqua1 through electroporation. Those colonies were screened for the constructs using colony PCR.

For genotyping some *SPC18* CRISPR- induced mutations, PCR products from transformed roots were cloned into the TA vector, pMD19 (Takara Bio). Individual colonies were chosen for sequencing via blue-white screening. Plasmids from individual colonies were sequenced through Psomagen. A list of the primers used in this chapter can be found in Table 2.3.

Hairy root transformation

The ARqua1 *A. rhizogenes* strain carrying the desired plasmid was used to transform *M. truncatula* Jemalong A17 or Jemalong seedlings. The seeds were surface sterilized using 5% bleach and a few drops of TWEEN-20 after mechanical scarification and imbibition for a few hours at room temperature. After sterilization, the seeds were imbibed again in sterile water at 4 degrees celsius overnight. After imbibition, the seeds were plated on 7% water agar plates and incubated at 30 degrees celsius overnight in the dark. After germination, the seedlings were transformed following the previously published protocol from Boisson-Dernier et al. (2001)**.** At least 40 seedlings for each construct were transformed and analyzed (usually, 20 seedlings were transformed in two separate experiments).

GUS staining

GUS staining was carried out on all roots from promoter-*GUS* transformed plants. Fresh roots with nodules (three weeks post-inoculation) were pooled in 6-well plates in

50mM sodium phosphate buffer (pH 7.2) containing 2mM 5-bromo-4-chloro-3-indoxyl-β-D-glucuronide cyclohexylammonium salt (X-Gluc), 0.2% Triton X-100, and 2mM potassium ferricyanide, and 2mM ferrocyanide. The roots were infiltrated with the solution under vacuum for 15 minutes, the vacuum was released, and reapplied for another 15 minutes. The roots were then incubated at room temperature overnight. Before imaging, the roots were transferred to 4 degrees celsius, and in most cases, the roots were destained in distilled water overnight at 4 degrees celsius.

Microscopy

Whole nodules (both GUS-stained and fresh) were imaged using an Olympus SZ61 dissecting microscope and SPOT Imaging camera and software.

Roots from *SPC18* CRISPR transiently transformed plants were harvested 67 days post-inoculation. Whole roots with nodules were fixed in 25% (v/v) glutaraldehyde and placed at 4 degrees celsius. Nodule cells were imaged using an A1R-SIMe confocal microscope (Nikon) through the IALS Light Microscope Facility (UMass Amherst). Before imaging, roots were rinsed with water and embedded in 5% (w/v) low-melt agarose. 60-70 µM nodule slices were dissected using a vibrotome (Leica VT1000 S) and stained with 5µM SYTO9 (Thermo Fisher) in 50mM Tris-HCl (pH 7.0) and 25mg/mL sucrose for 10-15 minutes. Sections were then placed in water for a few minutes before imaging. A 488nm laser and 525nm emission were used for confocal imaging.

Electrophoretic mobility shift assay

The EMSAs were carried out using a protocol similar to the one used in Soyano et al. (2013). In brief, crude protein extracts were taken from fresh *M. truncatula* roots and nodules using whole-cell extraction buffer (50mM Tris, 200μM EDTA, 10% v/v glycerol, 150mM NaCl, 2mM $MgCl_2$, 1mM DTT, protease inhibitor), and the protein content was approximated using a Bradford assay. The two complementary strands of the 49bp *DNF1* promoter probe were annealed together and diluted to 6μM. The probe was then labeled with ɣ32P-ATP using T4 PNK (New England Biolabs). The probe was incubated at room temperature for 40 minutes with the crude protein extract (or buffer for the negative control) and 50mM KCl, 10% v/v glycerol, 1x binding buffer (10mM HEPES, 2mg/mL BSA, 10mM DTT, .2mg/mL sonicated herring sperm DNA). The probe-binding reactions were loaded onto a 5.5% polyacrylamide native gel and run at 120 volts in a cold room and on ice. The gel was then dried, and a magnetic sheet was used in a GE Amersham Typhoon Gel and Blot Imaging System to visualize the radiolabeled bands.

CRISPR genotyping and phenotyping

After nodules formed on the CRISPR-transformed roots, the roots were harvested for subsequent genotyping and phenotyping. Notes and pictures of the appearance of nodules were taken, and a specific note was made of any roots that had only small, white nodules. Whole roots were frozen at -20 degrees celsius in 2x CTAB buffer (2% w/v CTAB, 1.4M NaCl, 100mM Tris-HCl pH 8.0, 20mM EDTA) for later DNA extraction.

The first round of genotyping was done by pooling some root tissue from transformed roots with white nodules (Pool #1), pooling tissue from roots with pink nodules (Pool #2), and pooling root tissue from empty vector transformed roots (Pool #3), DNA was extracted using a CTAB DNA extraction protocol, similar to many plant DNA extraction protocols (Clarke, 2009). PCR was done on each of these pools using *SPC18* genotyping primers.

After the pools of DNA were tested, PCR of *SPC18* was done on each root individually, evaluated for possible CRISPR-induced mutations via DNA gel electrophoresis. Those roots with clear mutations in *SPC18* (due to larger deletions) were then evaluated for nodule phenotype through the original pictures and notes. Some roots that had a clear Fix- nodule phenotype, but no obvious large deletion in *SPC18* were genotyped through sequencing the *SPC18* gene.

Table 2.3. Primers and oligos used in this chapter.

Name	5' to 3' sequence	Use
DNF1_A	GACCAAAGCGTCACTGAATTGCTG	DNF1 full-length promoter
DNF1_B	CGTTCAATCTGTAACCGAATGAATGC	DNF1 full-length promoter
pDNF1_Seq2F	GCACCATTATCCTAATTTCATCAAACTC	*DNF1 ~1300bp promoter*
pDNF1_Seq3F	GGTTCACCATATGCATTTGGGG	*DNF1 ~600bp promoter*
SPC18_A	GACCAAACAAAACAAGAGAAAGCG	SPC18 promoter
SPC18_B	ACAATCTCTGCTAAATTATACAACCG	SPC18 promoter
AttB1_pDNF1_F	GGGGACAAGTTTGTACAAAAAAAGCAGGCTAAGTTCACGAGATCAAGGACCAA	DNF1 420bp promoter primers
AttB1_pDNF1_R	GGGGACCACTTTGTACAAGAAAGCTGGGTTTTGATTCAATTGATGAAAATT	DNF1 420bp promoter primers
3pSPC18_sac_F	TAAGCAGAGCTCTTGTTGGTTGATTGAAACCCCC	SPC18 downstream of coding region
3pSPC18_sac_F	ATACCGGAGCTCTGACACTCTTGCTGTCTTGC	SPC18 downstream of coding region
symSPC18_F	AGGGGCATGTTTGGATTGGCTT	forward SPC18 CRISPR genotyping
symSPC18_R	CCATCCAGCACCGTGAACAA	reverse SPC18 CRISPR genotyping
TC320	CTAGAAGTAGTCAAGGCGGC	CRISPR construct sequencing
M13F	GTAAAACGACGGCCAGT	CRISPR construct sequencing
oCmYLCV	TGCTCTTCGCGCTGGCAGACATACTGTCCCAC	SPC18 CRISPR golden gate assembly
CSY_gRNA_1	TCGTCTCCGTGCAGACGTAACTGCCTATACGGCAGTGAAC	SPC18 CRISPR golden gate assembly
REP_gRNA_1	TCGTCTCAGCACTGATAATAGTTTTAGAGCTAGAAATAGC	SPC18 CRISPR golden gate assembly
CSY_gRNA_2	TCGTCTCCCGATGCACGATTCTGCCTATACGGCAGTGAAC	SPC18 CRISPR golden gate assembly
REP_gRNA_2	TCGTCTCAATCGTGTAATTAGTTTTAGAGCTAGAAATAGC	SPC18 CRISPR golden gate assembly
CSY_term	TGCTCTTCTGACCTGCCTATACGGCAGTGAAC	SPC18 CRISPR golden gate assembly
pDNF1_F	GAATCAAAGATAAATATATATATATTTTTTTTAGATAAAAAGATTAATA	probe for EMSA
pDNF1_R	TATTAATCTTTTTATCTAAAAAAAATATATATATATTTATCTTTGATTC	probe for EMSA

CHAPTER 3

SYNTAXIN132A WAS CO-OPTED FROM ARBUSCULAR MYCORRHIZAL SYMBIOSIS FOR ITS ROLE IN NITROGEN-FIXING SYMBIOSIS

A modified version of this chapter has been published:

Pan H, Oztas O, Zhang X, Wu X, **Stonoha C**, Wang E, Wang B, and Wang D. (2016) A symbiotic SNARE protein generated by alternative termination of transcription. *Nature Plants* (2) 15197.

I am listed as a middle author on this paper because of the role that I played in generating data and figures, as well as my writing contributions. Specifically, I was responsible for much of the evolutionary analysis showing that *SYP132A* is present plant species that form mycorrhizal symbiosis and absent from the Brassicaceae, which has lost the symbiosis. I contributed Figures 3.5A and 3.6, as well as Table 3.1. I also contributed to writing the manuscript.

Here is the author contribution list from the published paper (modified for clarity): D.W., H.P., O.O., E.W., and B.W. designed the experiments; H.P., O.O., X.W., and X.Z. performed the experiments; D.W., **C.S.**, B.W., X.W. performed alignments and evolutionary analysis; D.W., H.P., O.O. and **C.S.** wrote the manuscript.

Introduction

In biotic interactions, how organisms communicate across species boundaries is a central question. In intimate host–microbe interactions, the interface is often defined by a specialized membrane. Since this interfacial membrane frequently originates from a pre-existing membranous compartment (such as the plasma membrane) of the host cell, one

major challenge to understanding host–microbe communication is to elucidate how a new, specialized membrane is distinguished from its progenitor. In the legume–rhizobia symbiosis, intracellular bacteroids are individually surrounded by a peribacteroid membrane, which is derived from the host plasma membrane and functions as a destination for protein secretion from the host (Ivanov et al., 2012; Wang et al., 2010). Thus, the host cell must clearly define the peribacteroid membrane to ensure the proper targeting of secretory vesicles. In plants, syntaxin proteins target membrane-localized receptors (also known as t-SNARE) for secretory vesicles (Zheng et al., 1999). Previously, we have found that a transcript annotated as *SYNTAXIN 132* (or *SYP132*) is co-expressed with *DNF1*, which encodes a component of the nodule-specific signal peptidase (Wang et al., 2010). In *Medicago truncatula* nodules, SYP132 protein has been shown to localize to both the plasma and the peribacteroid membranes (Catalano et al., 2007, 2004; Limpens et al., 2009). How secretory vesicles distinguish these two membranes is currently unknown.

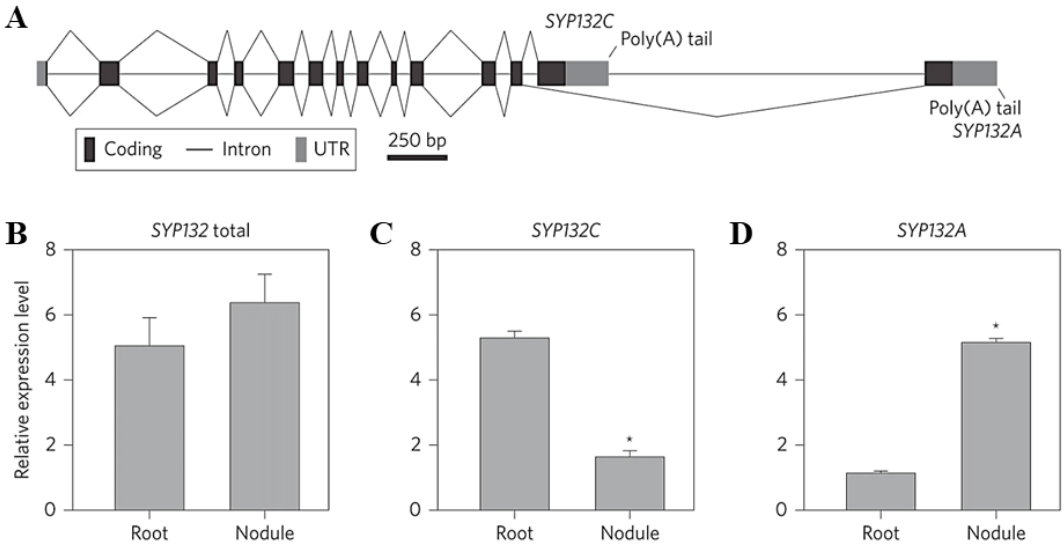


Figure 3.1. The nodule-specific SYP132A is produced through alternative cleavage and polyadenylation. A. Schematic illustration of the *SYP132* gene, with jagged lines representing splicing events. *SYP132C*, canonical *SYP132*; *SYP132A*, alternative *SYP132*; UTR, untranslated region. **B-D.** Quantitative RT–PCR of *SYP132* transcripts in roots and nodules. '*SYP132* total' shows the result using a pair of primers common to all *SPY132* transcripts. The expression level is normalized to that of the *PDF* gene. Error bars represent standard deviation from three biological replicates. $*P < 0.05$, Student's *t*-test.

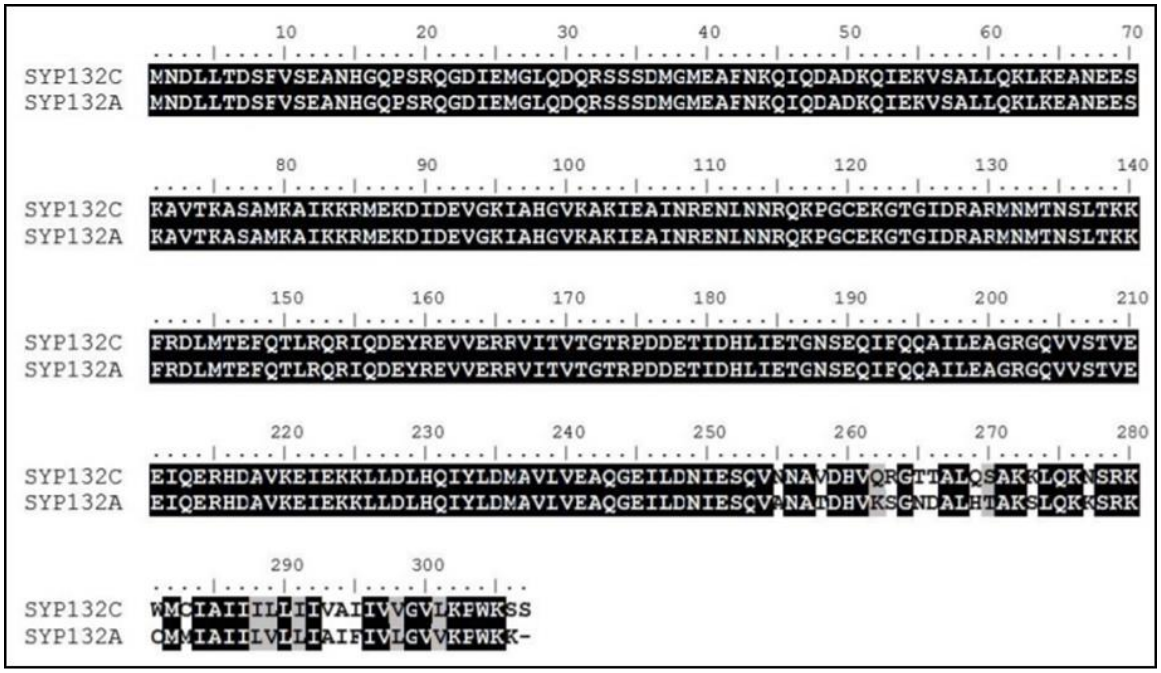


Figure 3.2. Alignment of full-length protein sequences between SYP132C and SYP132A. Numbers indicate positions along the alignment. Identical and similar residues are in black and grey background, respectively.

Results

Medicago produces two SYP132 isoforms

In interrogating the transcriptional regulation of protein secretory genes in nitrogen-fixing symbiosis, we noticed that the *SYP132* gene is represented on the *Medicago truncatula* Gene Expression Atlas by two probe sets, corresponding to two distinct transcripts (Benedito et al., 2008). We discovered that these two *SYP132* transcripts are generated by an alternative cleavage and polyadenylation (APA) mechanism, where transcription terminates either at the canonical terminal exon (exon 13C) or an alternative exon (exon 13A) about 2,000 base pairs further downstream (Figure 3.1A). When exon 13A is produced, it replaces exon 13C through alternative splicing. The results are two mature transcripts differing in their coding sequences at the 3′ end (Figure 3.2). We named these two isoforms *SYP132C* (for canonical) and *SYP132A* (for alternative). Although the transcriptional initiation of *SYP132* is constitutively active (Figure 3.1B), the termination of its transcription is regulated. In non-nodulated roots, *SYP132C* is the predominant form. In contrast, nodules produce primarily *SYP132A*, with a concomitant decrease in *SYP132C* (Figure 3.1C, D).

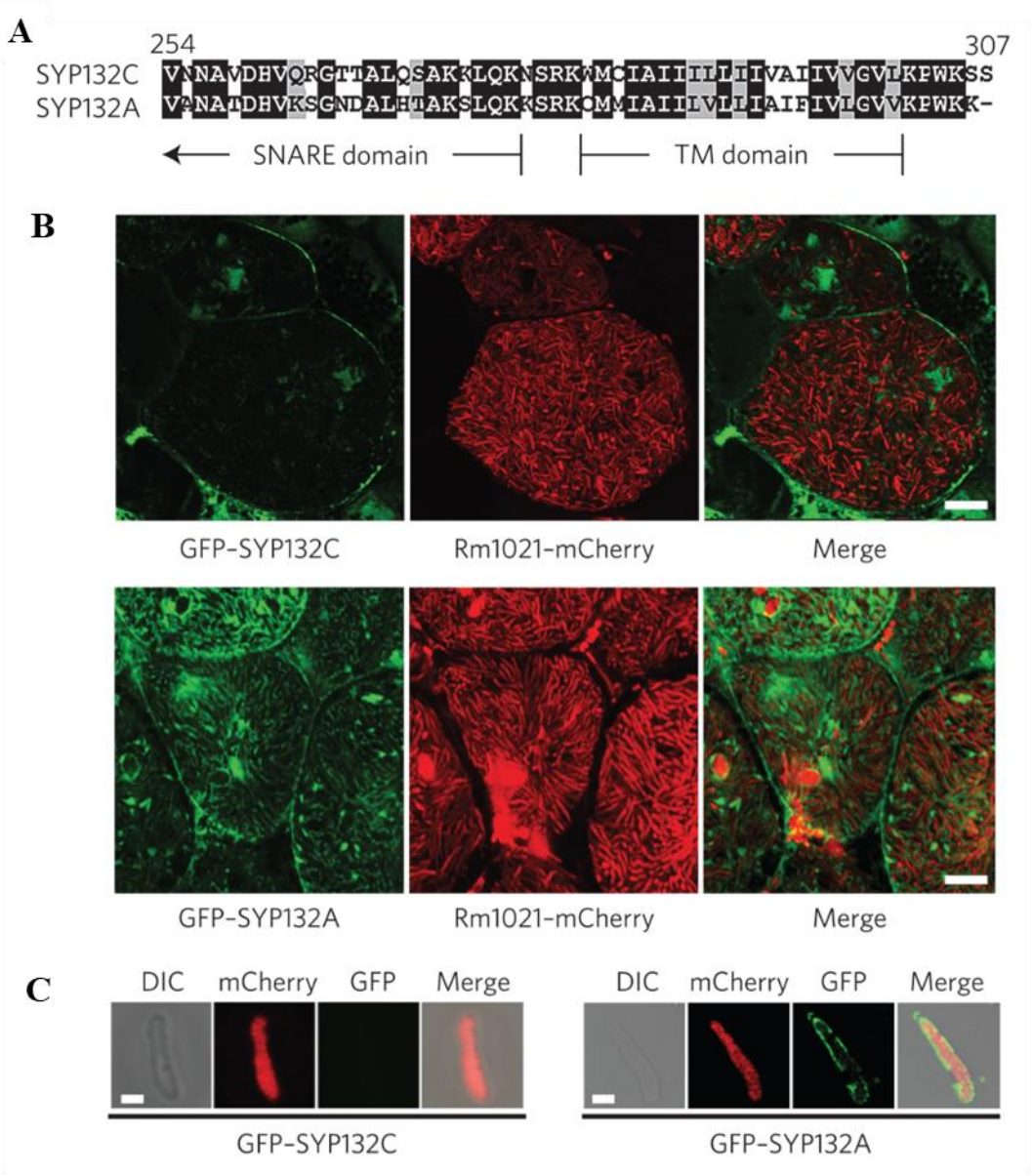


Figure 3.3. Differential subcellular localization of SYP132 protein isoforms. A. Alignment of amino acid sequences encoded by the two terminal exons (exons 13C and 13A). Numbers refer to positions in the full-length proteins. The boundaries of the two affected functional domains are indicated. Identical residues are repeated in a separate line. Both identical residues and residues with similar structures have a black background. **B.** Subcellular localization, by confocal microscopy, of full-length SYP132C and SYP132A, each fused with GFP, in nodule cells infected with Rm1021 expressing an mCherry marker. Both SYP132C and SYP132A were driven from the DNF1 promoter. Scale bars, 20 µm. **C.** GFP–SYP132C and GFP–SYP132A signal on individual symbiosomes isolated from transgenic nodules. Symbiosomes were isolated from mCherry–Rm1021 infected nodules and checked under confocal microscope immediately. Scale bars, 2µm.

SYP132A can localize to the symbiosome membrane

APA most commonly alters the 3′ untranslated region or truncates a transcript prematurely, both of which affect the abundance of the protein product8. APA at the *SYP132* locus is unusual for a plant gene in that it can code for two full-length syntaxin proteins, which differ in their SNARE domains and the whole transmembrane domains at their carboxy-termini (Figure 3.3A). To test whether such differences influence protein localization, we fused each isoform to green fluorescent protein (GFP) and expressed each construct in rhizobia-containing nodule cells. Both proteins were detected at their expected sizes (data not shown). Consistent with reports on other species, SYP132C is mostly found on the cell periphery, with little overlap observed between the GFP signal and mCherry-expressing bacteroids inside symbiosomes (Figure 3.3B, upper panel). In contrast, SYP132A was found to localize to the symbiosome membrane surrounding the bacteroids (Figure 3.3B, lower panel). In isolated symbiosomes, fluorescence from GFP–SYP132A (but not GFP–SYP132C) is clearly visible on the periphery, surrounding bacteroids with mCherry fluorescence (Figure 3.3C). In some cells, SYP132A also accumulates on the plasma membrane, possibly as a consequence of ectopic expression.

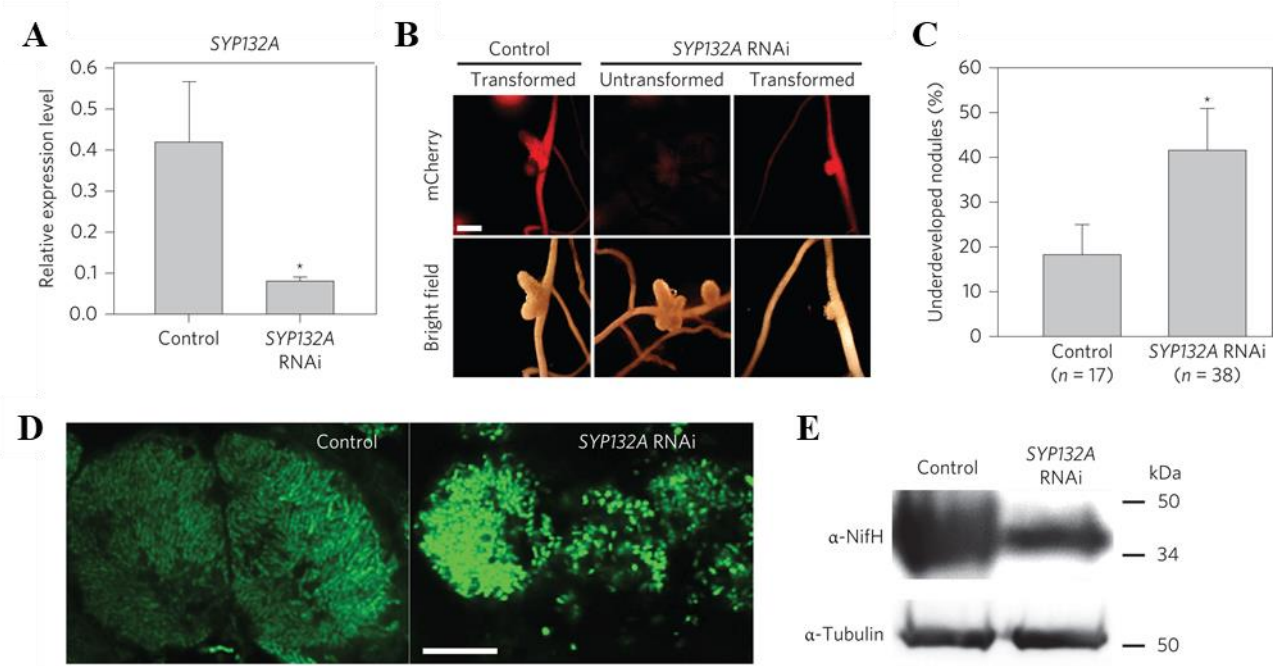


Figure 3.4. SYP132A is required for the maturation of symbiosomes. A. Quantitative RT–PCR showing isoform-specific silencing of *SYP132A.* Control, empty vector. Expression levels are normalized to that of *PDF*. *$P < 0.05$, Student's *t*-test. **B.** Morphology of *SYP132A*-silenced nodules and the control. The upper panel shows the fluorescence of the mCherry reporter encoded in the *pHellsgate8* RNAi vector. The lower panel shows the morphology of *SYP132A*-silenced nodules and control nodules. Scale bar, 2 mm. **C.** Statistical analysis to show that *SYP132 RNAi* plants have more white nodules comparing to the empty vector control. ***$P < 0.001$, Student's *t*-test. *n* = the number of transformed plants for each vector. **D.** SYTO9 staining of bacteroids in control and *SYP132A*-silenced nodules. Scale bar, 40 μm. **E.** NifH protein levels in *SYP132A*-silenced nodules and control nodules by Western blot. Tubulin was used as a loading control.

SYP132A is required for the maturation of bacteroids

To ascertain whether the differential localization of the SYP132 isoforms is functionally relevant, we generated RNAi constructs specifically targeting the terminal exon of *SYP132C* or *SYP132A*, as well as one targeting all *SYP132* transcripts (*SYP132* RNAi). Roots in which *SYP132* and *SYP132C* were silenced stopped growing, indicating that the canonical protein isoform, similar to its counterparts in other flowering plants, performs essential functions in secreting to the extracellular space (Kalde et al., 2007).

Roots where the *SYP132A* transcript level is reduced by 80%—but in which *SYP132C* was only moderately affected—grew normally (Figure 3.4A), suggesting that SYP132A on the plasma membrane is dispensable. However, the nodules produced on these roots were small (Figure 3.4B). By calculation, almost half of the nodules from SYP132A RNAi-transformed roots were white; in contrast, less than 20% nodules from roots expressing the control vector were white (Figure 3.4C). Although *SYP132A* RNAi host cells were occupied by bacteria to an extent similar to empty vector control, these cells seem smaller. Confocal microscopy showed that bacteria in control cells elongated, a hallmark of differentiation, whereas those that infected *SYP132A* RNAi host cells remained small and undifferentiated (Figure 3.4D). Consistent with their morphology, bacteroids in *SYP132A*-silenced nodules produced much less nitrogenase iron protein (NifH) than in control nodules (Figure 3.4E). This phenotype is reminiscent of the defect seen when the DNF1 signal peptidase is removed (Wang et al., 2010), suggesting that SYP132A is an essential symbiosome t-SNARE for the trafficking of host secretory proteins important for bacterial differentiation (Kondorosi et al., 2013; Van de Velde et al., 2010).

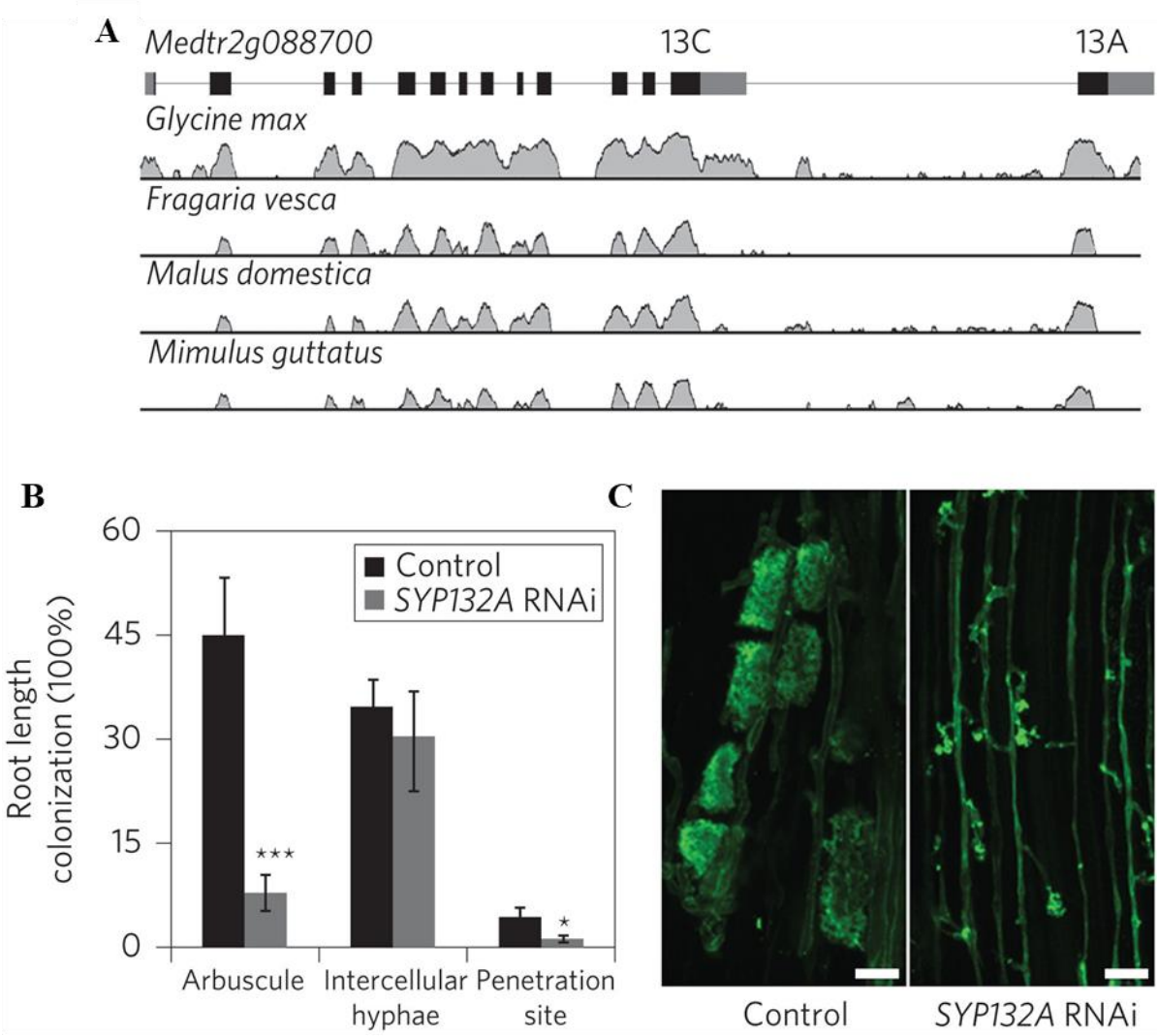


Figure 3.5. SYP132A is necessary for arbuscular mycorrhizal symbiosis. A. The *M. truncatula SYP132* locus (base genome, *M. truncatula* Mt4.0; position, chromosome 2: 37, 401, 493–37, 407, 300; exons 13A and C are indicated) aligned with syntenic regions in *Glycine max* (soybean), *Malus domestica* (apple) and *Mimulus guttatus* (Seep monkeyflower) using the VISTA browser. **B.** Quantification of *Rhizophagus irregularis* colonization levels in control (empty vector) and *SYP132A*-silenced *M. truncatula* plants. Nine transformed plants were obtained for each transformation. ***$P < 0.01$ and *$P < 0.05$, Student's *t*-test. **C.** Representative confocal images of *R. irregularis* arbuscules in control (empty vector) and *SYP132A*-silenced *M. truncatula* roots. The arbuscules were stained with wheat germ agglutinin (WGA)-conjugated Alexa Fluor 488. The experiments were done at 8 weeks post inoculation. Scale bars, 20 μm.

Figure 3.6. ClustalW protein alignment of homologues of MtSYP132C and MtSYP132A. Amino acid sequences from homologues of both SYP132C and SYP132A from multiple plant species are aligned. *Arabidopsis thaliana* has a homolog of MtSYP132C, while missing SYP132A. The box delineates the predicted transmembrane domain.

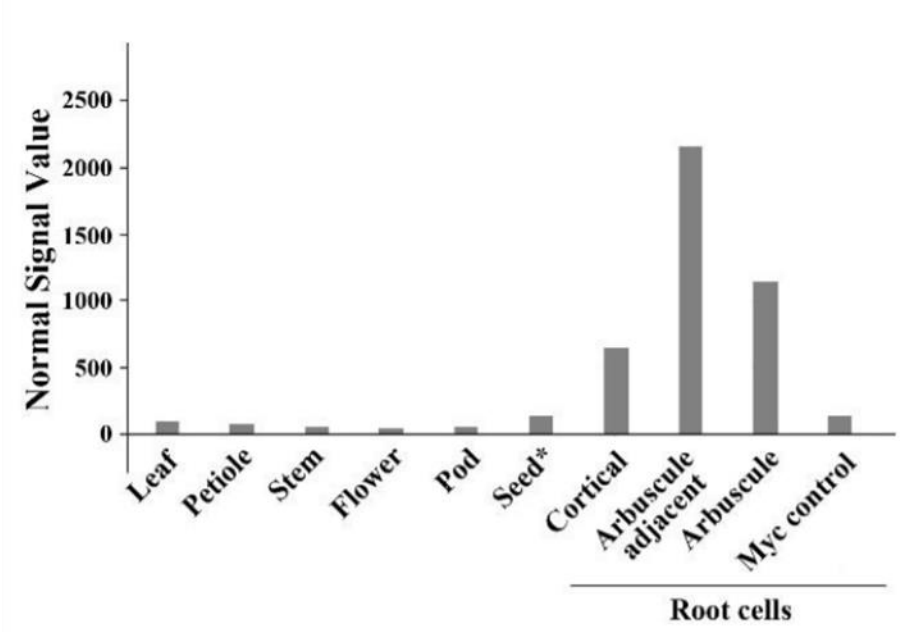


Figure 3.7. Expression data for *MtSYP132A* from the *Medicago truncatula* Gene Expression Atlas. The normalized expression data are shown, the highest signal values were from roots colonized with arbuscular mycorrhizal fungi.

SYP132A is conserved in advanced angiosperms and functions in arbuscular mycorrhizal symbiosis

The unique biogenesis of the *M. truncatula SYP132* isoforms prompted us to investigate the structure and expression of this gene in other species. When the *M. truncatula SYP132* gene is aligned with syntenic regions of multiple dicot angiosperms, including non-legumes, exon 13A, as well as the rest of the gene, appears well conserved (Figure 3.5A), indicating that the APA of *SYP132* may have been widely adapted by dicotyledonous plants before the emergence of legumes. Indeed, transcripts encoding the SYP132A protein were detected in most dicotyledonous plants we examined (Figure 3.6 and Table 3.1), where a single gene encodes SYP132C and SYP132A through APA. In monocotyledonous species, sequences homologous to SYP132C and SYP132A are encoded by separate genes (Figure 3.6 and Table 3.1). However, both monocotyledonous and dicotyledonous SYP132A-like sequences are clearly distinct from SYP132C-like ones at the amino acid level (Figure 3.6), suggesting a conserved functional difference. Furthermore, in both monocotyledonous and dicotyledonous species, another SYP13 protein—which can be named SYP131, is encoded by a separate gene.

One notable exception is the Brassicaceae family, which includes Arabidopsis (Figure 3.6 and Table 3.1). Arabidopsis seems to have lost the *SYP132* gene orthologous with other species. The *AtSYP132* gene is likely to be a recent duplication of *AtSYP131*, analogous to the scenario in monocotyledonous species. The presence of *SYP132A* in angiosperms correlates tightly with a plant's ability to form arbuscular mycorrhizal symbiosis, an ancient association with soil fungi that later gave rise to rhizobial symbiosis12 (the Brassicaceae family has subsequently lost this ability). Furthermore, *M.*

truncatula SYP132A is induced in roots inoculated with arbuscular mycorrhizal fungi, specifically in cells closely interacting with the microbe (Figure 3.7, Benedito et al., 2008). We also observed the induction of the *SYP132A* homologue in the roots of *Sorghum bicolor*, a grass, following inoculation with arbuscular mycorrhizal fungi spores (Figure 3.8).

It is, therefore, likely that the SYP132A protein is broadly involved in arbuscular mycorrhizal symbiosis as well. To test whether SYP132A is required for arbuscular mycorrhizal symbiosis, we inoculated control and SYP132A-silenced *M. truncatula* plants with *Rhizophagus irregularis* spores. Compared with the control, the fungus produced fewer penetration sites on *SYP132A* RNAi roots, and arbuscules were drastically underdeveloped (Figure 3.5B, C). Hyphae growth between plant cells was unaffected, indicating that SYP132A specifically controls the development of the peri-arbuscular membrane, another host–microbe interface. Taken together with the nodule phenotype of *SYP132A*-silenced roots, our data show that SYP132A is a common determinant of protein secretion toward the interface between the plant cell and a beneficial microbe. The SYP13 clade of t-SNAREs carry out housekeeping and symbiosis-specific secretion functions via distinct members (Figure 3.9). Advanced angiosperms seem to require two housekeeping SYP13 proteins and, with the exception of the Brassicaceae family, one symbiotic SYP13. Monocotyledonous species evolved a conventional solution of designating one gene for each protein. However, in most dicotyledonous species, the symbiotic SYP132 protein shares one gene with a housekeeping SYP132 protein, a novel use of the widespread phenomenon of APA.

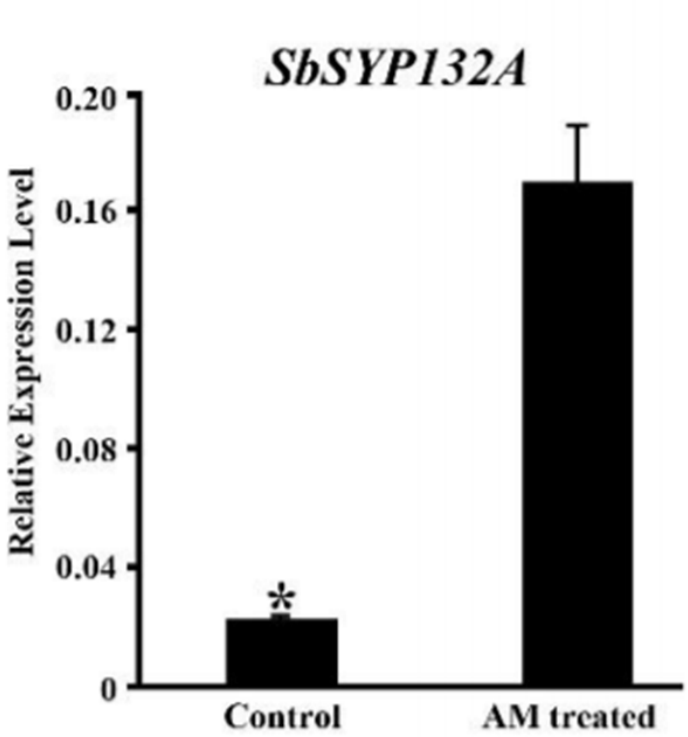


Figure 3.8. The induction of the *SYP132A* homologue in *Sorghum bicolor* by AM fungus in the root. Quantitative RT-PCR shows the level of *Sb2g035120* transcript in the root with or without *Rhizophagus irregularis* inoculation. The expression level is normalised to that of the GTP Binding Protein gene GTPB. "*" means $P<0.05$ (Student's t-test).

Table 3.1. SYP132 genes in various plant species with their proposed function based on homology to *M. truncatula* genes

Species	Proposed function	Gene locus	Transcript name	Genome version	NCBI reference sequence
Arabidopsis thaliana	Housekeeping	*AT5G08080*	*AT5G08080.3*	TAIR 10	NM_001203327.1
Medicago truncatula	Housekeeping	*Medtr2g088700*	*Medtr2g088700.2*	Mt4.0v1	XM_003596968.1
Medicago truncatula	Symbiotic	*Medtr2g088700*	*Medtr2g088700.1*	Mt4.0v1	N/A
Glycine max	Housekeeping	*Glyma13g3870*	*Glyma13g3870.1*	JGI v1.1	NM_001254606.1
Glycine max	Symbiotic	*Glyma13g3870*	*Glyma13g3870.2*	JGI v1.1	N/A
Fragaria vesca	Housekeeping	*gene23921-v1.0-hybrid*	*mrna23921.1-v1.0-hybrid*	v1.1	N/A
Fragaria vesca	Symbiotic	*gene23921-v1.0-hybrid*	*N/A*	v1.1	XM_004287883.1
Populus trichocarpa	Housekeeping	*Potri.019G036700*	*Potri.019G036700.1*	JGI v3.0	XM_002325810.2
Populus trichocarpa	Symbiotic	*Potri.007G123000*	*Potri.007G123000.1*	JGI v3.0	XM_006380247.1
Citrus sinesis	Housekeeping	*orange1.1g038457m.g*	*orange1.1g038457m*	v1.1	XM_006469638.1
Citrus sinesis	Symbiotic	*orange1.1g038457m.g*	*N/A*	v1.1	XM_006469639.1
Oryza sativa	Housekeeping	*LOC_Os07g07000*	*LOC_Os07g07000.1*	v7.0	NM_001065494.1
Oryza sativa	Symbiotic	*LOC_Os07g34510**	*LOC_Os07g34510.1**	v7.0	N/A
Brachypodium distachyon	Housekeeping	*Bradi1g56720*	*Bradi1g56720.1*	JGI v2.0	XM_003557507.2
Brachypodium distachyon	Symbiotic	*Bradi1g25880*	*Bradi1g25880.1*	JGI v2.0	XM_003562952.2

Discussion

Here we describe a novel mechanism to specify a target membrane for protein secretion through APA of a t-SNARE gene. This parallels a recent report on symbiosis-specific v-SNAREs, where distinct *M. truncatula VAMP72* genes (*VAMP721d* and *VAMP721e*) are upregulated during both nitrogen fixing and arbuscular mycorrhizal symbiosis, presumably to label secretory vesicles destined for the host–microbe interface (Ivanov et al., 2012; Sinharoy et al., 2013). SYP132A is likely to form a complex with VAMP721d/e in these symbioses to deliver specialized cargo molecules to the symbionts. Targeted secretion is not restricted to symbiotic interactions. Arabidopsis utilizes the PEN1/SYP122 protein to block penetration attempts from non-host pathogens (Assaad et al., 2004; Collins et al., 2003; Kwon et al., 2008; Meyer et al., 2009). Employing specialized SNARE proteins to delineate a dedicated membrane interface appears to be a general mechanism in intimate plant–microbe interactions.

In gene regulation, much effort has been focused on how promoters drive transcriptional initiation in particular patterns, often as a result of gene duplication through evolution. Our findings demonstrate that transcriptional termination can also be regulated to alter gene function, in this case, by generating new coding capacities. The molecular machinery involved in terminating transcription may provide powerful means to modify gene function in general.

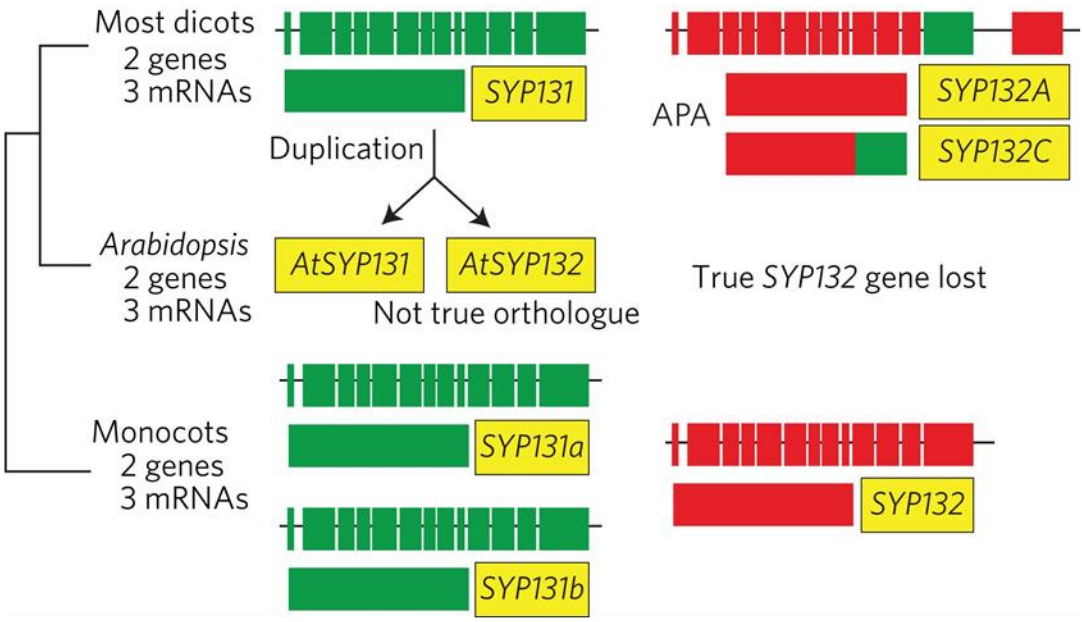


Figure 3.9. Biogenesis of SYP13 proteins in higher plants. Most advanced angiosperms encode two housekeeping SYP13 proteins and one symbiotic SYP13. In monocotyledonous species, the symbiotic SYP132 is encoded by its own gene. In most dicotyledonous plants, the symbiotic SYP132A and housekeeping SYP132C are two isoforms generated from a single gene via APA. In Brassicaceae such as *Arabidopsis*, the original SYP132 was lost, and gene duplication analogous to that in grasses produced AtSYP131 and AtSYP132. Boxes with connecting lines represent gene structures; filled boxes represent the resulting transcripts, with green representing likely housekeeping genes/transcripts and red likely symbiotic ones; yellow boxes show proposed names for the transcripts.

Methods

Plasmids and vectors

For the GFP–SYP132 fusions, the protein-coding sequences of SYP132A and SYP132C were first cloned into the pCR8/GW/TOPO vector (Life Technologies) and introduced into pDNF1–pMDC43 using Gateway recombination facilitated by the LR clonase (Life Technologies). The original pMDC43 vector has a 35S promoter, which was replaced by a 3 kb DNF1 promoter (Curtis and Grossniklaus, 2003; Wang et al., 2010).

For RNA interference, the regions between 3049–3474 bp and 5410–5659 bp of the SYP132 genomic sequence (downstream of the start codon), as well as the region between 345–594 bp of the SYP132 coding sequence (for SYP132C, SYP132A, and SYP132, respectively) were first cloned into the pCR8/GW/TOPO vector and recombined into the pHellsGate8 vector modified with a 35S promoter-driven mCherry fluorescent marker for plant transformation (Helliwell and Waterhouse, 2005).

All constructs were transformed into the ARqua1 strain of Agrobacterium rhizogenes through electroporation. The presence of the target plasmids was checked by PCR and sequencing. Primers used in this chapter can be found in Table 3.2.

Plant growth conditions, transformation, and inoculation

All experiments were performed on *M. truncatula* ecotype A17. Plants were grown under a 16 h/22 °C–8 h/18 °C light/dark cycle. Hairy root transformation of plant seedlings using *A. rhizogenes* ARqua1 was performed as described (A. Boisson-Dernier et al., 2001). Transgenic roots were selected based on antibiotic resistance or mCherry fluorescence.

For nodulation, *Sinorhizobium medicae* strain ABS7 *pHemA::LacZ* or *Sinorhizobium meliloti* strain Rm1021 *pHemA::mCherry* was used (Leong et al., 1985). The bacteria cells used for inoculation were suspended in ½ BNM liquid medium to an optical density (600 nm) of 0.05. Five milliliters of the liquid suspension was used for each plant. Phenotypes were checked 21 days post inoculation (dpi).

For arbuscular mycorrhizal symbiosis, *M. truncatula* plants harboring empty or *SYP132A* RNAi vectors were inoculated with *Rhizophagus irregularis*. Four hundred

commercial spores of *R. irregularis* were used for each plant. Colonization was analyzed 8 weeks post inoculation; colonized roots were collected and treated with 10% KOH at 95 °C for 6 min, followed by 3 min in ink. Root length colonization was calculated using the gridline intersect method and imaged under a Nikon Eclipse 800 light microscope (Giovannetti and Mosse, 1980). In brief, the inoculated roots were spread out on a Petri dish with a grid of 0.5 cm squares drawn on the bottom. Every root section that was intersected by a square was categorized as either having a fully formed arbuscule, intercellular hyphae, a penetration site, or no fungal structure.

Gene expression analysis with quantitative RT–PCR

To compare the levels of *SYP132C, SYP132A*, and total *SYP132* transcripts between roots and nodules, wild-type A17 plants were inoculated with ABS7 *pHemA::LacZ*. Nodules of all developmental stages were collected at 21 dpi, and roots were collected from un-inoculated plants at the same time. To analyze transcript levels in the *SYP132A* RNAi experiment, transformed hairy roots (in the A17 background) were identified under a fluorescence stereoscope using mCherry fluorescence. Nodules on fluorescent roots were collected at 21 dpi. In the *SbSYP132A* gene expression assay, root samples were taken from *Sorghum bicolor* plants inoculated with *R. irregularis* for 3 months and the control.

Total RNA was extracted using TRIzol (Life Technologies) following the manufacturer's instructions. To eliminate potential contamination from genomic DNA, the extracted RNA was treated with the Turbo DNA-free kit (Life Technologies) according to the manufacturer's instructions. cDNA synthesis was done using the iScript

cDNA synthesis kit (Bio-Rad). Synthesized cDNA was amplified using ExTaq polymerase (Takara) with SYBR Green dye (Thermo Fisher Scientific) on a Mastercycler ep Realplex system (Eppendorf) with a volume of 20 µl for each reaction. Following PCR amplification, melting curves were analysed to rule out potential non-specific amplification. Results were represented as the average threshold cycle (CT) value of three replicates. The expression of target transcripts was normalized using the PDF gene, which is considered to be one of the best reference genes (Kakar et al., 2008). The gene locus of *PDF* is *Medtr6g084690*.

Microscopy

In confocal microscopy, 21 dpi transgenic nodules were hand-sectioned using double-edged razor blades and mounted on microscope slides. For SYTO9, nodule slices were stained in 5µM SYTO9 (Life Technologies) with 50 mM Tris-HCl (pH 7.0) and 25mg ml–1 sucrose for 15 min.

For single symbiosome confocal microscopy, young nodules were ground in 10% freshly depolymerized paraformaldehyde in 1 × TBS (pH 7.5) and put on ice for 10 min. The suspension containing symbiosomes was mounted onto a microscope slide and observed immediately.

Samples were observed under an Olympus FLUOVIEW FV1000 confocal laser scanning microscope. GFP/SYTO9 signal was detected using excitation with a 485 nm laser and emission with a 490–540 nm band pass filter. The mCherry signal was detected using excitation with a 587 nm laser and emission with a 575–675 nm band pass filter.

For arbuscular mycorrhization, *R. irregularis* inoculated roots were stained with WGA-Alexa Flouor 488 (Life Technologies) using the same buffer as SYTO9 staining. Images of roots stained with WGA-Alexa Flouor 488 were taken with an Olympus Fluoview FV1000 microscope.

To visualize X-gal staining results, nodules inoculated with ABS7 pHemA::LacZ were rinsed in Z′ buffer (16.1 g Na2HPO4•7H2O, 5.5 g NaH2PO4•H2O, 0.75 g KCl, 0.246 g MgSO4•7H2O, 2.7 ml β-mercaptoethanol, and adjusted to pH 7.0) for 15 min. Nodules were first treated with 1.25% glutaraldehyde for 1h to inactivate endogenous plant β-galactosidase and then rinsed twice in Z′ buffer for 15min each. Afterwards, nodules were stained with 0.8 mg ml–1 X-gal in Z′ buffer containing 5mM each K3Fe(CN)6 and K4Fe(CN)6 overnight at 37 °C. Stained nodules were embedded in 6% low-melting-temperature agarose (Sigma) and sectioned into 100μm slices on a VIBRATOME 1500 tissue-sectioning system (Intracel).

Nodule protein analysis

Nodules from transformed roots were harvested into a precooled extract solution containing 0.5 M sucrose, 10 mM DTT, 1% V/V protease inhibitor cocktail (P8340, Sigma-Aldrich), and 50 mM Tris-HCl, pH 7.4. Tissues were ground by hand using pestles, filtered through two layers of Micracloth (Calbiochem), and then centrifuged to obtain protein extracts. These extracts were boiled in 4× Laemmli buffer, and proteins were separated on SDS–PAGE gels. Proteins were then transferred to nitrocellulose membranes (Adventec) for immunoblotting analysis. The membrane was incubated with an anti-tubulin antibody (at 1:4,000 dilution, Sigma), anti-NifH antibody (1:4,000,

AgriSera) or anti-GFP antibody (1:2,000, Life Technologies). A secondary antibody (anti-mouse or anti-rabbit) conjugated with horseradish peroxidase was incubated with the membrane, and signals were visualized by enhanced chemiluminescence (Thermo Scientific) on a G-box (New England Biolabs).

Phylogenetic analysis

Homologues of either MtSYP132C or MtSYP132A in different plant species were determined using the Basic Local Alignment Search Tool (BLAST) against various databases. BLASTP was used to search for homologues of both proteins in Phytozome v.10, with the default parameters (http://phytozome.jgi.doe.gov). To verify the results from Phytozome, BLASTP and TBLASTN were used at GenBank, again with the default parameters (Altschul et al., 1990). The *Oryza sativa* MtSYP132A homologue was obtained from PlantGDB by using BLASTP with the default parameters (www.plantgdb.org/). The protein sequences of the MtSYP132C and MtSYP132A homologues were obtained from either Phytozome or GenBank and aligned to the *M. truncatula* proteins using BioEdit version 7.2.5 (Hall, 1999).

Table 3.2. Primers used in this chapter. Obtained from: Oztas, Onur, "THE KEY QUESTION IN SYMBIOTIC NITROGEN FIXATION: HOW DOES HOST MAINTAIN A BACTERIAL SYMBIONT?" (2017). Doctoral Dissertations. 945.

Name	5' to 3' sequence	Use
Syp_RNAi_F	ACAAAAGCCTGGCTGTGAGA	to target both isoforms
Syp_RNAi_R	AAGAATTGCCTGCTGGAAGA	to target both isoforms
SypC_RNAi_F	CAATGCAGTCGATCATGTCC	to target *SYPC* isoform
SypC_RNAi_R	AATGGCAACAAAACCAGCTT	to target *SYPC* isoform
SypA_RNAi_F	GAAGTCGGGTAACGATGCTC	to target *SYPA* isoform
SypA_RNAi_R	CCATGAACCTTGTTAAACTGC	to target *SYPA* isoform
Syp_Start	ATGAACGACCTTCTCACTGAT	to amplify *SYPA* and *SYPC* CDS without signal sequence
SypA_Last	CTTCTTCCATGGTTTCACAAC	to amplify *SYPA* CDS without signal sequence
SypC_Last	AGAACTCTTCCAAGGTTTGAG	to amplify *SYPC* CDS without signal sequence

CHAPTER 4

THE ROLE OF TWO CONSERVED MOTIFS DISCOVERED IN NODULE-SPECIFIC CYSTEINE-RICH PEPTIDES

Introduction

Medicago truncatula is predicted to have over 700 nodule-specific cysteine-rich (NCR) peptides encoded in its genome. Out of the species in the inverted repeat-lacking clade of legumes with available genomic information, *M. truncatula* seems to have the most NCR peptides, and this correlates with highly elongated, terminally differentiated symbionts (Montiel et al., 2017). Much of the information we have about the role of *M. truncatula* NCR peptides in nitrogen-fixing (NF) symbiosis is from a general viewpoint: as a group, the NCRs are required for bacteroid differentiation and, in turn, required for the symbiosis (Mergaert et al., 2006a; Van de Velde et al., 2010). On the other hand, we have information about activity of a few, specific NCR peptides on free-living bacteria, and the role of even fewer NCRs *in planta* (Farkas et al., 2014; Horváth et al., 2015; Kim et al., 2015; Tiricz et al., 2013; Wang et al., 2017; Yang et al., 2017).

NCR247 is one of the most well-studied NCR peptides, and its effects on free-living bacteria are varied and well documented. For example, at high concentrations, NCR247 has bactericidal effects, and at low concentrations, it can inhibit essential cellular processes (Farkas et al., 2014; Penterman et al., 2014; Van de Velde et al., 2010). Despite all the information we have about NCR247, it seems to have fairly unique characteristics and may not be representative of many NCR peptides. For example, the mature peptide is among the smallest: it is made up of 24 amino acids and has a

molecular weight of 3010 g/mol. In addition, it has one of the highest isoelectric points of all the NCRs in *M. truncatula* at 10.17 (Farkas et al., 2014; Montiel et al., 2017). In fact, NCR247 was originally chosen for study because of the extreme characteristics that set it apart from other NCR peptides (Farkas et al., 2014).

Due to the number of NCR peptides in *M. truncatula*, it has been difficult to connect the characteristics of the peptides with effects on bacteria (free-living or intracellular). Furthermore, we need to move towards a more medium- or high-throughput assay format in order to identify patterns and describe more peptides. Here, we test the effects of 23 peptides on free-living *Sinorhizobium meliloti* and describe possible conserved sequence motifs that may be important for their function.

Results

NCR peptide discovery and characterization

Only a small number of NCR peptides have been investigated for their effects on free-living bacteria, and fewer have been investigated *in planta*. Because bacteroid differentiation in *M. truncatula* is controlled by NCR peptides, we wanted to investigate greater numbers of NCR peptides and their activities on rhizobia (Van de Velde et al., 2010). The peptides were chosen for several reasons. Some were reported to be active on *S. meliloti* or other microbes. Others were chosen for their cationic properties and ease of synthesis. Together, we assembled 23 peptides, 18 of which, to the best of our knowledge, have not been described in the literature as active against *S. meliloti*. Using the established direct colony count assay, we found new peptides with high activity against free-living rhizobia (Figure 4.1).

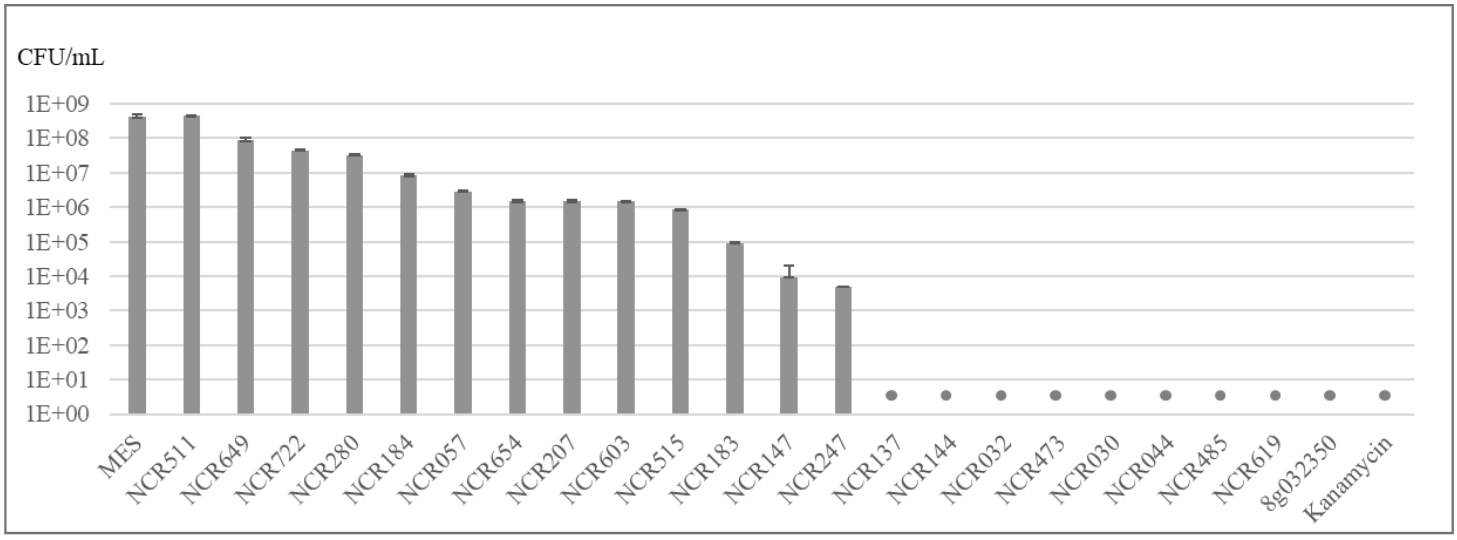


Figure 4.1. Colony-forming units of *Sinorhizobium meliloti* Rm1021 treated with various nodule-specific cysteine-rich peptides at 20μM. *S. meliloti* colony-forming units (CFU) per milliliter when incubated with 20μM of various NCR peptides indicated on the x-axis. 'MES' indicates the negative control, *S. meliloti* incubated with buffer. The dots mark CFU counts that are zero. Error bars are the standard deviation.

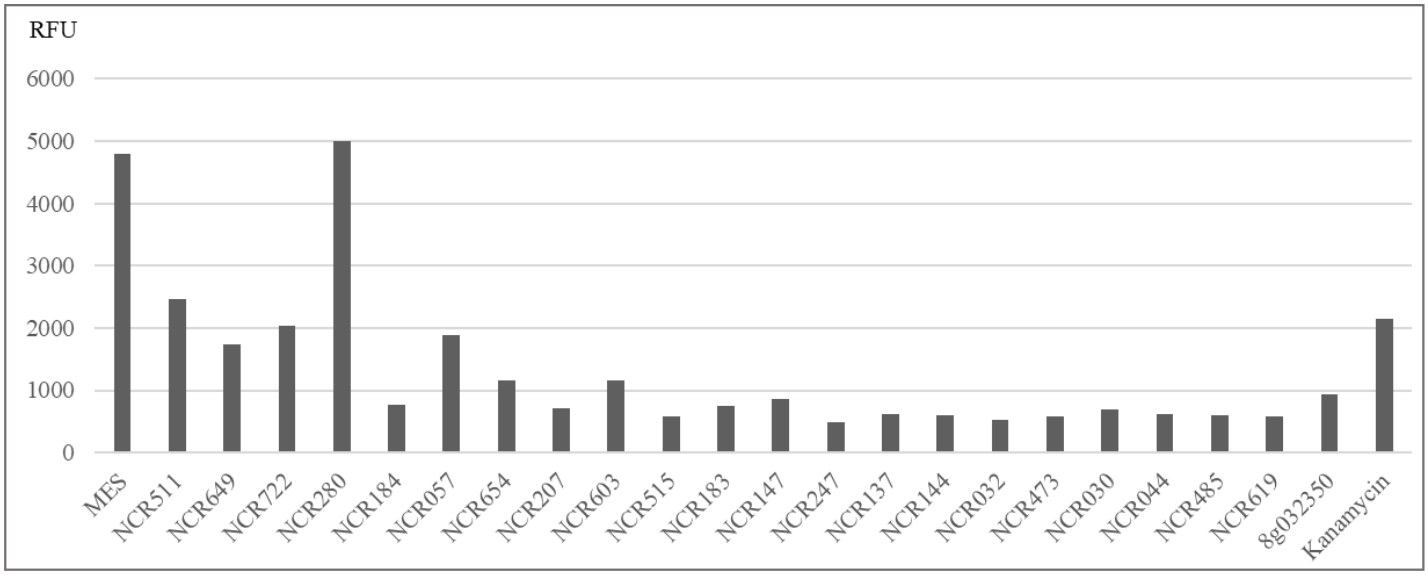


Figure 4.2. Resazurin relative fluorescence units of *Sinorhizobium meliloti* incubated with various nodule-specific cysteine-rich peptides at 20μM. *S. meliloti* relative fluorescence units (RFU) when incubated with resazurin and 20μM of various NCR peptides indicated on the x-axis. 'MES' indicates the negative control, *S. meliloti* incubated with buffer.

Medium-throughput plate-based assay for testing new peptides

With this relatively large collection of peptides, we sought to establish a more medium-throughput assay instead of directly measuring bacterial growth on solid media. We designed a plate-based assay dependent on the conversion of resazurin to resafluorin by bacterial metabolism (reduction). This results in a fluorescence readout related to the uptake and reduction of resazurin by the bacteria. This assay was carried out in two phases: first, the bacteria were incubated with the NCR peptides for three hours. Then, in the same 96-well plate, liquid growth media and resazurin were added. After resazurin was added, measurements were taken immediately (at the zero time point) and then at regular intervals until the negative control (no NCR peptides) turned visibly pink. We developed this assay to run in parallel with growth screens to differentiate between peptides that exhibit bactericidal and/or bacteriostatic effects on *S. meliloti*.

The two phases closely follow the procedure for direct colony counts in our lab. During the optimization of the plate-based assay, we noticed that the bacteria need to be treated with the peptides in the absence of growth media in order to detect the effect of those peptides. This is similar to how the colony count assay is performed: bacteria are treated with the NCR peptides in the buffer before they are plated on solid growth media.

We first established the reliability and the parameters of this metabolism-based assay in comparison with direct colony-count methods using the antibiotic kanamycin and NCR247 as a standard, well-studied antimicrobial peptide. In general, these screening methods were fairly consistent, bar a few outliers that could be due to unknown effects of those peptides (Figure 4.2).

Although these assays tested different aspects of the NCR peptides' antimicrobial effects, they generally agree with each other. NCR247 and kanamycin were used as positive controls as kanamycin and NCR247 have both been shown to be effective in killing *S. meliloti*. On media plates, these both show decreased growth, and in the resazurin screen both show low fluorescence, near the baseline level (no bacteria). This agrees with what was expected of these antimicrobial compounds, as killing many bacteria directly decreases overall metabolism in that sample. MES negative controls also agree across assays with a large amount of growth on solid media and a large difference in fluorescence over time.

Fluorescence was measured at multiple time points over a period of approximately six hours in order to capture the time point with the biggest difference in fluorescence between the negative control (MES buffer) and positive control (kanamycin), which ended up being between four and six hours. This is the time point that we used for the analysis of the NCR peptides on the metabolism of *S. meliloti* (Figure 4.2).

The resazurin screen revealed several peptides with negative effects on bacterial growth. NCR247 controls established a baseline fluorescence of approximately 480-600 RFU, and several peptides fall within or near that range, such as NCR515, NCR032, and NCR044 (Figure 4.2).

Colony counts and resazurin data generally agree, but there are a few outliers such as NCR183, NCR184, and NCR207 (Figure 4.1, 4.2). We believe these could be due to bacteriostatic but not bactericidal effects, such as limiting metabolic processes. This would account for higher colony counts but low resazurin fluorescence. Other

discrepancies, such as kanamycin samples fluorescing in the absence of CFUs could be cellular responses to specific stresses caused by kanamycin before cell death. While the resazurin assay was designed to reflect the conditions used for the direct colony counts, it also measures the changes in the bacteria over a period of hours after treatment, while the colony counts are done days after treatment. This may also contribute to some of the differences in bacterial growth and metabolism observed between the two assays.

The treatment of free-living bacteria with NCR peptides in this study and others can be variable for a few reasons. The redox state of the cysteine-rich peptides is known to be important for folding and activity (Shabab et al. 2016), and the synthesized peptides that we use are usually a mix of conformations. Furthermore, it has been shown that the exact effects of some NCR peptides changes depending on the assay conditions (Farkas et al. 2018). Therefore, some variation in our results and previously tested NCR peptides, such as NCR247, is expected.

By characterizing the activities of 18 novel NCR peptides against free-living *S. meliloti*, we expanded the repertoire of active peptides, nine of which were able to eliminate the bacteria at 20μM (Figure 4.1).

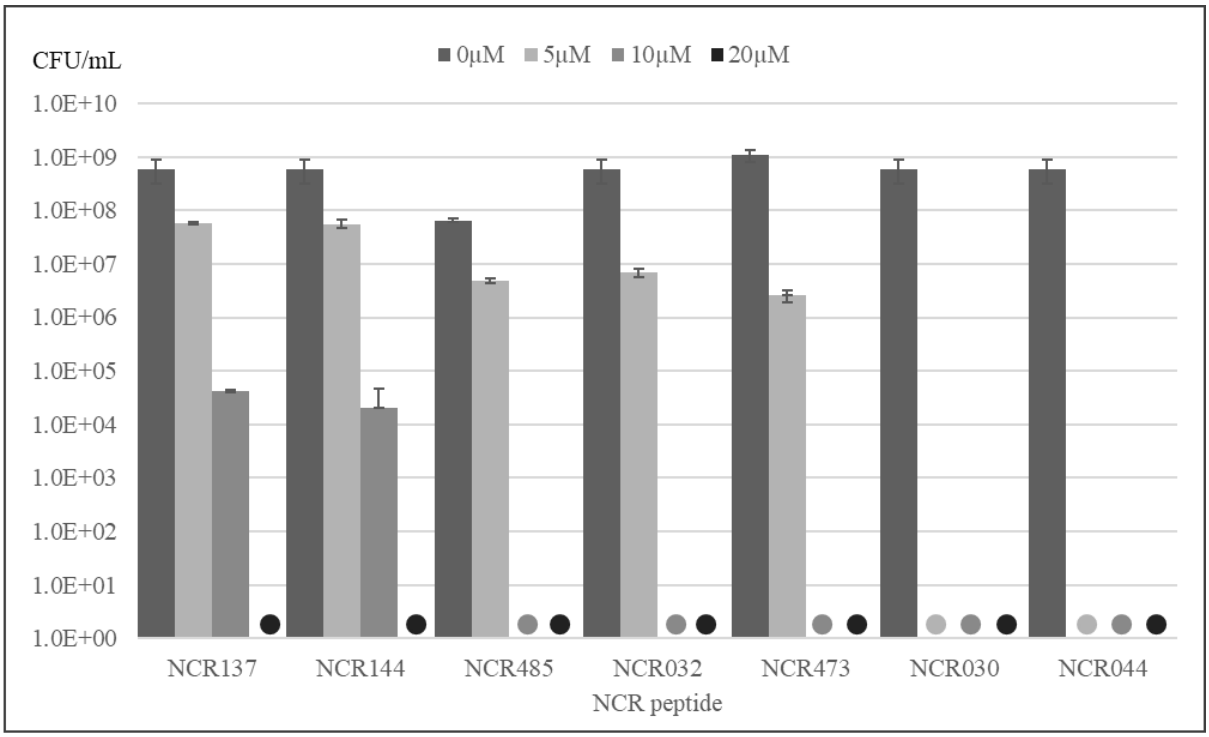


Figure 4.3. Colony-forming units per milliliter (CFU/mL) of *Sinorhizobium meliloti* treated with increasing concentrations of selected NCR peptides. Dots indicate CFU counts that are zero. The "0" concentration of the NCR peptides indicates that the bacteria were not treated with NCR peptides, only MES buffer.

Dose response of *S. meliloti* to active peptides

Based on the results of both the colony count and resazurin assays, we chose seven peptides that eliminated the bacteria at 20μM for a dose-response assay. NCR030 and NCR044 had the strongest activity against *S. meliloti* with minimum inhibitory concentrations (MIC) below 5μM. NCR032, NCR473, and NCR485 have a MIC between 10 and 5μM, and NCR137 and NCR144 have a MIC between 20 and 10μM (Figure 4.3).

Figure 4.4. Sequence alignment of the NCR (mature) peptides with the greatest effect on the viability of *S. meliloti*. These are the predicted mature peptides (signal peptide removed). The dashed boxes mark the common motifs.

NCR peptide motif discovery

After we characterized a number of active peptides, we set out to search for potential shared sequence features. We also looked for other common properties, such as charge, but found no correlation.

MUSCLE sequence alignment of active NCR peptides (all with four cysteines) revealed two possibly conserved amino acid sequences: 'DKDC,' containing the second cysteine residue and 'RCRK,' which contains the third cysteine (Figure 4.4). These sequences were immediately interesting due to their high charge-density and because they encompass two of the four cysteine residues. When the peptide is in its oxidized form, the cysteines form disulfide bonds, which are important for the protein structure. In particular, the alternation of negative and positive residues in the first motif suggests these residues could be involved in protein folding through electrostatic interactions. These hypothesized interactions could serve to stabilize the NCR peptides *en route* to the symbiosome membrane, where they may be reduced to a more active form (Shabab et al., 2016).

A cursory analysis of the list of NCR peptides in the *M. truncatula* genome as curated by Montiel et al. (2017) indicates that over 10% of them contain exactly 'DKDC' and/or 'RCRK.' Furthermore, these motifs are also found in NCRs of other IRLC legumes including species in the subclades Hedysaroid, Astragalean, and other Vicioid legumes (including species such as *Galega orientalis*, *Cicer arietinum*, *Ononis spinosa*, *Pisum sativum*, and *Medicago sativa*). Both motifs, or variations of these motifs (e.g., 'DKEC'), seem to be overrepresented in NCR peptides, but the 'DKDC' motif seems to be present in more species. Due to the frequency of these motifs in *M. truncatula* and other IRLC legumes, these sequences are conserved and, therefore, potentially important for function.

Only a handful of NCR peptides to date have been directly identified through forward genetics studies to be necessary for symbiosis *in planta.* DNF4 (NCR211) and DNF7 (NCR169) are both short peptides with four cysteines (Horváth et al., 2015; Kim et al., 2015). They are also much less active than the peptides that we tested that had MICs less than 20μM. NSF1 and NSF2 are both NCR peptides implicated in symbiont selection in *M. truncatula*. NSF2 is a relatively short peptide with four cysteines and 44 total residues (Yang et al., 2017). NSF1 is a longer peptide with six cysteines, and it has a low pI of 3.79 (Wang et al., 2017). These four peptides, along with NCR247, do not contain either of the motifs described here (Figure 4.5). Since some of the most well-studied peptides do not contain these common motifs, we have very little information on the possible role they play in symbiosis or in the antimicrobial action of the peptides.

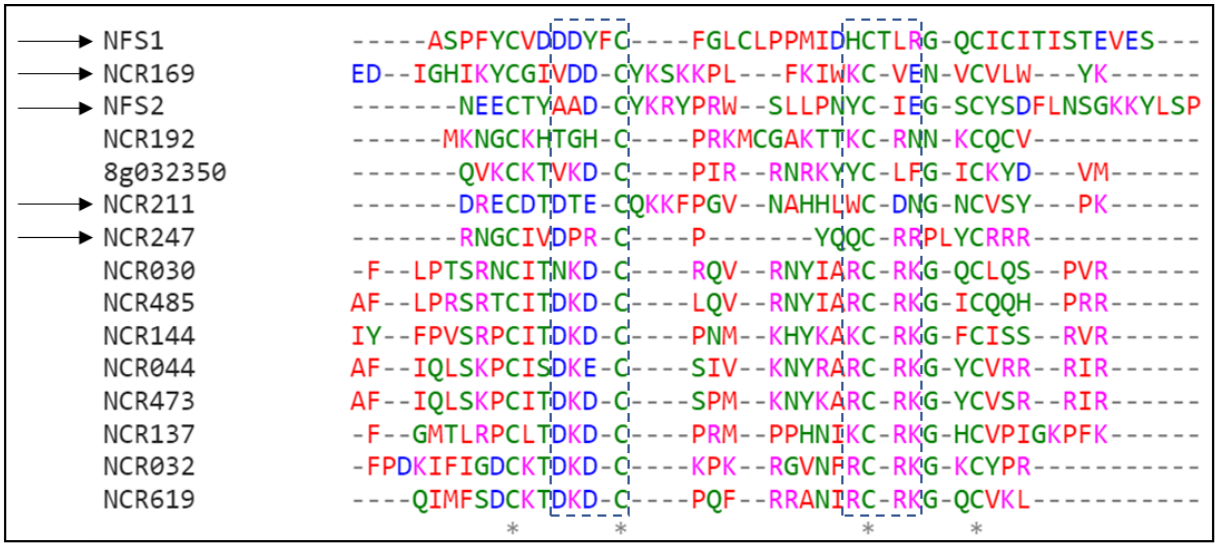


Figure 4.5. Amino acid sequence alignment of various NCR peptides used in this study and others. The dashed boxes indicate the location of the two common motifs 'DKDC' and "RCRK.' The asterisks denote the four conserved cysteines. The horizontal arrows are pointing to five well-studied NCR peptides, four of which were discovered through forward genetic studies (NFS1, NCR169, NFS2, NCR211). NCR192 is on example of a peptide with six cysteines, like NFS1.

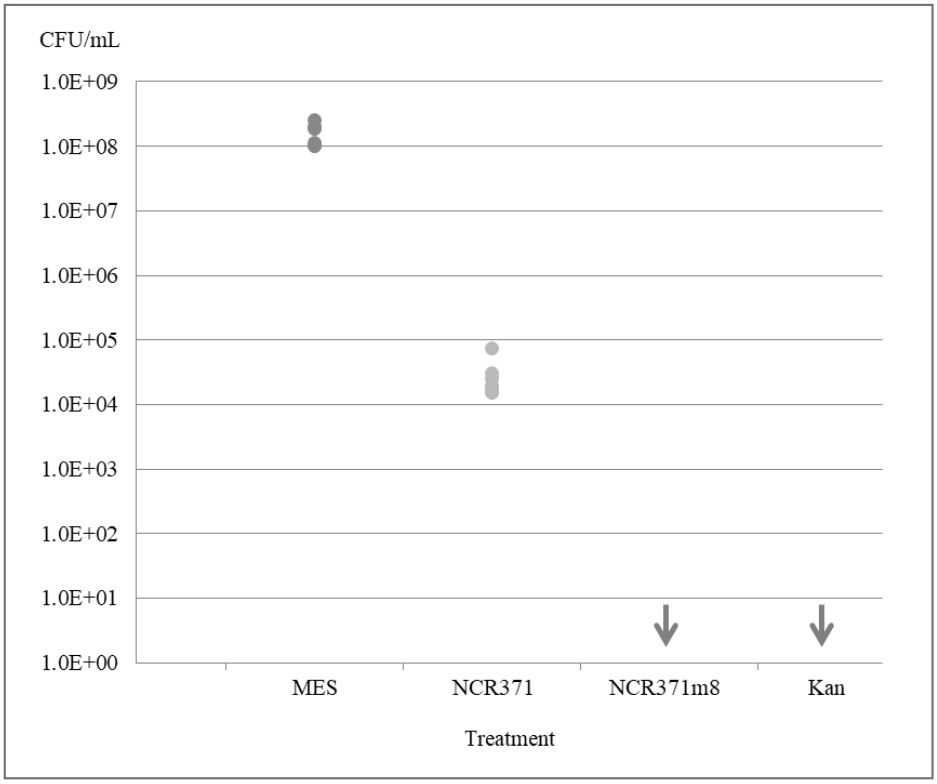


Figure 4.6. Colony-forming units of *Sinorhizobium meliloti* Rm1021 treated with wild-type and mutant versions of NCR371 at 20μM. *S. meliloti* colony-forming units (CFU) per milliliter when incubated with 20μM of various NCR peptides indicated on the x-axis. 'NCR371m8' indicates a mutated version of the peptide. 'MES' indicates the negative control, *S. meliloti* incubated with buffer. The arrows mark CFU counts that are zero.

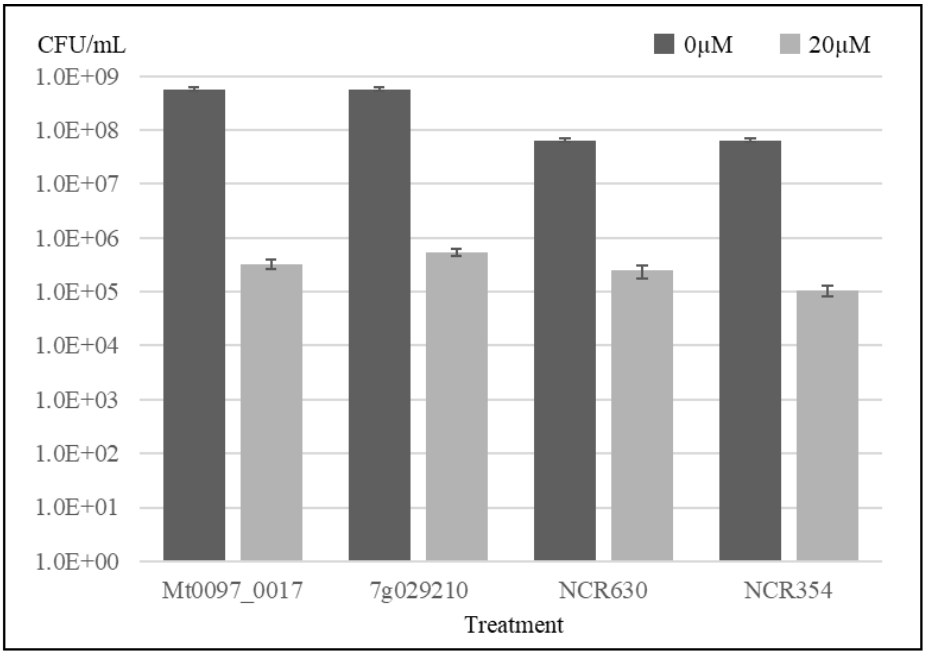


Figure 4.7. Colony-forming units of *Sinorhizobium meliloti* Sm1021 treated with NCR peptides with the motifs present, but low relative pI. *S. meliloti* colony-forming units (CFU) per milliliter when incubated with 20µM of various NCR peptides indicated on the x-axis, organized from higher to lower pI. Mt0097_0017, 7g029210, NCR630, and NCR354 have isoelectric points of 8.69, 8.68, 8.68, and 8.48, respectively. The "0" concentration of the NCR peptides indicates that the bacteria were not treated with NCR peptides, only MES buffer. Kanamycin treatment during the peptide treatment resulted in 0 CFU/mL. Error bars are the standard deviation.

Implications for activity

If the conserved sequence has functional importance, we predict it can confer activity. We noticed that NCR371 contains a small deviation from the consensus sequence, which gave us an opportunity to test our prediction. We synthesized both the wild-type version of NCR371 and a mutant version (NCR371m8) that contained both conserved motifs. At 20µM, the wild-type peptide has mild antimicrobial activity, but the mutant is active at a concentration as low as 10µM. Depending on the concentration tested, the mutant is 200-1,000 times more potent (Figure 4.6).

We next asked whether we can identify NCR peptides with more potential to exhibit antimicrobial activities based on the presence of the motif. To test the predictive power of this motif, we chose two un-numbered NCR peptides, *Medtr7g029210* (locus ID) and *Mt0097_00017* (transcriptome ID), as well as NCR630 and NCR354. These four peptides have low pI values (<8.7) compared to many of the described active peptides, they contain no features found in known antimicrobial NCR peptides other than the motifs. In fact, among peptides containing the "DKDC...RCRK" sequences, they are among the less cationic ones.

At 20μM, these four peptides reduced bacterial colony formation by ~100-1,000 fold. This suggests that the presence of these two motifs allows us to predict that these peptides, despite their lower relative pI, possess antimicrobial activity against *S. meliloti* (Figure 4.7). The predictive power of these motifs indicates that the sequences need to be considered when assessing the activities of NCR peptides in addition to other characteristics such as charge.

Discussion

NCRs are a large class of peptides encoded in the *M. truncatula* genome; they are believed to be very diverse through duplication and diversification events. This diversity has made it difficult to draw general conclusions about these peptides, but we have uncovered two motifs that seem to be conserved in many of the NCRs in *M. truncatula* and also present in other IRLC legumes.

These common sequences seem to have an effect on the antimicrobial activity of the peptides on free-living bacteria because restoring these motifs in NCR371 increases

the effect on rhizobia in culture. Furthermore, we were able to use the presence of these motifs to accurately predict the antimicrobial activity of NCRs with lower pI values, which has been the best indicator of activity. The correlation between the presence of the motif and antibacterial effect in culture could be an avenue for engineering stronger, possibly specific, antimicrobial peptides. It will be important to understand the mode-of-action of these peptides and how these sequences play a role.

The discovery of these sequence motifs around two of the cysteines within some of the NCR peptides may be helpful with protein folding, possibly involving the disulfide bonds that form between the cysteine residues when the peptide is oxidized. It is speculated that NCR peptides are fully reduced when they reach the intracellular bacteria in nodule cells. This is because NCR247, the most well-studied NCR peptide, is more active against bacteria growth in culture when it is in its reduced form compared to most of the other regioisomers (Shabab et al., 2016). Furthermore, *M. truncatula* has a nodule-specific thioredoxin that has been shown to work to reduce NCR peptides and increase their bactericidal effects (Ribeiro et al., 2017). The disulfide bonds, therefore, may be important for stabilizing the peptides and protecting them from degradation by the bacteria. For instance, Shabab et al., (2016) found that the reduced form of NCR247 is more vulnerable to the bacterial peptidase, HrrP, while other isomers of NCR247 are partially protected. Therefore, the presence of these sequence motifs may be important for peptide stability *in planta*. However, this hypothesis does not necessarily explain why these motifs would have increased cytotoxic effects on the bacteria in culture, which remains to be explored.

The motifs described here, ‘DKDC’ and ‘RCRK,’ are not found in well-studied NCR peptides such as NCR247, NCR211, NCR169, NFS1, and NFS2. Therefore, it will be intriguing to study the functions of these motifs in more detail. We have shown a general pattern of antimicrobial activity that correlates with these motifs, but their mode-of-action is unknown.

So far, the results have been acquired from synthetic peptides applied to free-living bacteria. We do not know whether the sequence signature/motif has similar functional implications when synthesized biologically, or whether the inhibitory effect/ antimicrobial activity observed on free-living bacteria is directly relevant to the biological functions of these peptides in the symbiosis. This is currently under investigation, and it will be important to connect the patterns we observe on free-living bacteria with the effects of these NCR peptides on the symbiosis.

Methods

Peptides and Colony count assay

The peptides were chosen from a previously published list (Montiel et al., 2017), and the predicted mature peptides were synthesized by Genscript and New England Peptide.

For the colony count assay, the protocol from Kim *et al.* (2014) was used, with one exception. Instead of ½ NaCl LB, the bacteria were grown in tryptone-yeast (TY) media (3 grams tryptone, 1.5 grams yeast extract, .25 grams $CaCl_2$ in 1L water) and plated on TY-agar plates. This was done to reduce the amount of NaCl in the assay.

Resazurin assay

Overnight cultures of *S. meliloti* in TY liquid media were selected for $O.D._{600}$ 0.3-0.6 and were pelleted at 3000xg at room temperature for 10 minutes. They were then washed in 5mM MES monohydrate buffer (pH 5.7) and spun again, then resuspended in MES to $O.D._{600}$= 0.2. 49μL of this suspension was added to each well used in a 96-well plate as well as 1μL MES (negative control lane), 1μL 100X kanamycin (positive control lane) or an NCR peptide at a final concentration of 20μM. The plate was then incubated at 30ºC for three hours on a shaker incubator. After incubation, 40μL of TY liquid medium and 10μL 10x resazurin dye was added to each well, and fluorescence was measured for t= 0 (fluorescence 590, excitation 550, auto-cutoff 590). The plate was then incubated again for 2-6 hours under the same conditions, checked periodically for color change in the MES lane. Once the MES lane changed to pink in color, the plate was read again with the same settings. We developed this assay to run in parallel with growth screens to differentiate between peptides which exhibit bactericidal and/or bacteriostatic effects on *S. meliloti*. This assay was done three separate times with similar results.

CHAPTER 5

REGULATION OF DMI2 PROTEIN LEVELS IN NITROGEN-FIXING SYMBIOSIS

A slightly modified version of this chapter has been published:

Pan H, **Stonoha-Arther C**, Wang D. (2018) Medicago Plants Control Nodulation by Regulating Proteolysis of the Receptor-Like Kinase DMI2. *Plant physiology* 177(2), 792–802.

I am listed as a middle author on this paper because of my contribution to one of the larger experiments in this paper. I carried out the hairy root transformations of all the DMI2 amino acid mutation constructs into the *dmi2* background. I also inoculated these transformants in preparation for phenotyping. The work I did contributed to Table 5.1, Figure 5.12, Figure 5.13, and Figure 5.14.

Introduction

Plants form symbiotic relationships with surrounding microbes to gain access to nutrients in natural environments. Most land plants form beneficial interactions with arbuscular mycorrhizal (AM) fungi, developing mycorrhized roots, which provide phosphorous and micronutrients to plants in exchange for fixed carbon. A subset of plants, including plants of the legume family, develop nitrogen-fixing symbioses in specialized organs, or root nodules. This sophisticated symbiosis with rhizobia provides legumes with nitrogen fixed by rhizobia hosted inside plant cells of the nodules (Oldroyd et al., 2011).

Plants use receptors to discriminate between symbiotic and pathogenic microbes. In both AM and nitrogen-fixing symbiosis, plant roots detect the existence of beneficial microbes by a group of receptor-like kinases (RLKs). In nitrogen-fixing symbiosis, rhizobia secrete a group of lipochitooligosaccharides, or Nod factors, to initiate nodule development and symbiosis. In *Medicago truncatula* (hereafter Medicago), Nod factors are perceived by two RLKs: NFP (Nod Factor Perception) and LYK3 (LysM Domain-Containing Receptor-Like Kinase3), which are named NFR1 (Nod Factor Receptor1) and NFR5 in *Lotus japonicus*, respectively. The perception of Nod factors activates the common symbiosis signaling pathway, including root hair-associated calcium spiking, early nodulation gene activation, and cortical cell division (Schauser et al., 1999). DMI2 (DOES NOT MAKE INFECTIONS2, the name in *M. truncatula*)/SYMRK (the name in *L. japonicus*)/Nodulation Receptor Kinase (the name in *Medicago sativa*) is believed to interact with Nod factor receptors (Antolín-Llovera et al., 2014a). Mutations in *DMI2* lead to the abortion of rhizobial infection at a very early stage (Endre et al., 2002; Stracke et al., 2002). DMI2 also is indispensable for AM symbiosis and Frankia symbiosis, another type of nitrogen-fixing symbiosis between certain plants and Frankia bacteria (Endre et al., 2002; Gherbi et al., 2008; Stracke et al., 2002), indicating a conserved role of DMI2 throughout the evolution of plant-microbe symbioses.

The DMI2 protein contains an intracellular kinase domain, a transmembrane domain, and the extracellular portion, including a region with antolinthree leucine-rich repeats (LRRs) and a malectin-like domain (MLD). In human cells, the single-domain protein Malectin functions in the endoplasmic reticulum (ER) lumen in protein quality control by binding to diglucosylated Glc2Man9GlcNAc2, a glycan composed of three

glucoses, nine mannoses, and two GlcNAcs (Schallus et al., 2008). Utilizing its carbohydrate-binding activity, Malectin interacts directly with misfolded glycoproteins and inhibits their secretion (Qin et al., 2012).

While Malectin-like sequences are widespread among biological kingdoms, two features make MLD-containing proteins in plants unique: (1) their gene families are greatly expanded in plants, and (2) MLDs mostly occur in the extracellular portion of RLKs. A few MLD-containing RLKs have been shown to play vital roles in plant development, male-female interaction, disease resistance, and plant-microbe symbiosis (Aurélien Boisson-Dernier et al., 2001; Endre et al., 2002; Haruta et al., 2014; Hok et al., 2011). Although the functions of MLDs have yet to be revealed, the position of MLDs in the extracellular portions of proteins points to the possibility that MLDs may be necessary for activating or deactivating the intracellular kinase domain through binding extracellular ligands. Interestingly, it has been reported that the MLD of SYMRK/DMI2 is cleaved constitutively, with or without rhizobia, and that the MLD-cleaved SYMRK/DMI2 protein outcompetes the full-length SYMRK/DMI2 in the interaction with Nod factor receptors (Antolín-Llovera et al., 2014b).

During nitrogen-fixing symbiosis, the host is required to provide nutrients to the rhizobia (Oldroyd et al., 2011). This burden on the plant has promoted the evolution of a sophisticated regulatory network controlling the scale and timing of nodule development (Oldroyd et al., 2011). Several reports show that overexpressing the full-length SYMRK/DMI2 or the intracellular kinase domain of SYMRK/DMI2 leads to spontaneous nodule formation even in the absence of rhizobia (Ried et al., 2014; Saha et al., 2014), suggesting that the protein level of DMI2 needs to be regulated precisely.

SYMRK/DMI2 also has been reported to interact with two E3 ligases: SINA4 (SEVEN IN ABSENTIA4) and SIE3 (SYMRK-INTERACTING E3 UBIQUITIN LIGASE; Den Herder et al., 2012; Yuan et al., 2012). However, direct genetic evidence illustrating that these two E3 ligases affect the level of SYMRK/DMI2 *in planta* is still missing, and the dynamics of DMI2 protein levels during nitrogen-fixing symbiosis remain unknown.

Here, we report that DMI2 protein levels are tightly regulated by legume hosts to properly respond to rhizobia infection. We find that, without rhizobia infection, DMI2 protein is constitutively degraded through the proteasome apparatus; during rhizobia infection, the DMI2 protein level is induced by blocking proteasome-mediated degradation. Meanwhile, if key amino acid residues in the DMI2-MLD are mutated, DMI2 is degraded constitutively, suggesting a crucial role of MLD in the regulation of DMI2 protein homeostasis. Taken together with the reports that overexpression of the DMI2/SYMRK kinase domain causes spontaneous nodulation (Ried et al., 2014; Saha et al., 2014), fine-tuning the protein level of DMI2 is critical for legumes to maximize the profit of nitrogen-fixing symbiosis at the lowest cost.

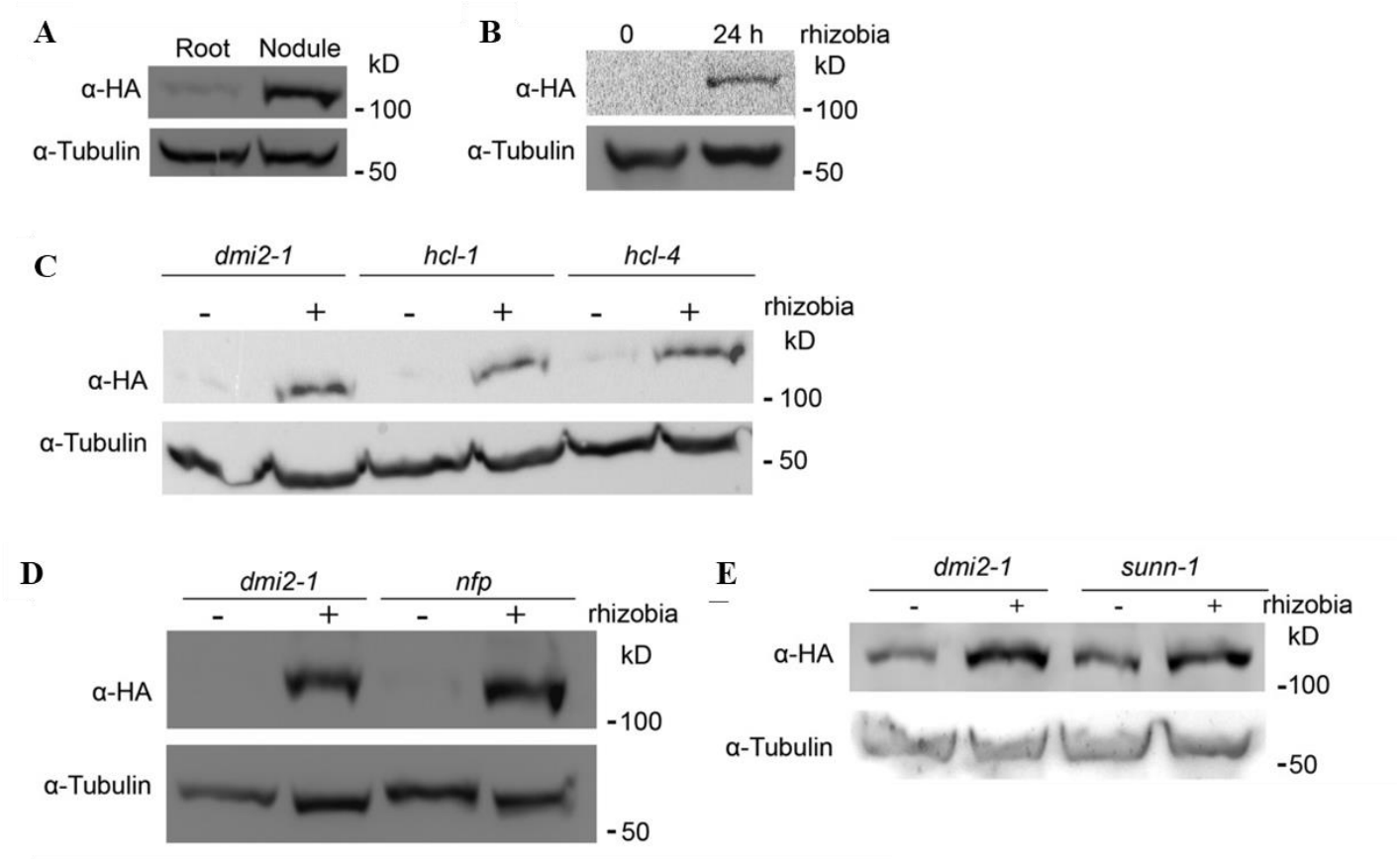


Figure 5.1. Rhizobia treatment could induce the protein level of DMI2 in a Nod factor receptor-independent manner. A. DMI2-HAST protein accumulated in the nodules compared with uninoculated roots. Protein samples were taken from 14-d-old nodules and 4-week-old uninoculated roots. **B.** DMI2-HAST protein was induced to a much higher level 24h after rhizobia treatment. Four-week-old *Medicago* roots were treated with *S. meliloti* strain ABS7 *hemA:LacZ* at the concentration of $O.D._{600}$ = 0.05 for 24h. **C.** and **D.** DMI2-HAST protein was induced to a similar level by rhizobia treatment in wild-type and Nod factor receptor mutant backgrounds. The *gDMI2-HAST* construct was transiently expressed in *dmi2-1*, *hcl-1*, and *hcl-4*, and transformed roots were treated with or without *S. meliloti* strain ABS7 *hemA:LacZ*. **E.** DMI2-HAST protein could be induced to a similar level by rhizobia in *dmi2-1* and *sunn-1* plants stably transformed with g*DMI2-HAST*. α-HA was used to detect the DMI2-HAST protein level, and tubulin was used as the loading control. The experiments were repeated five times with similar results.

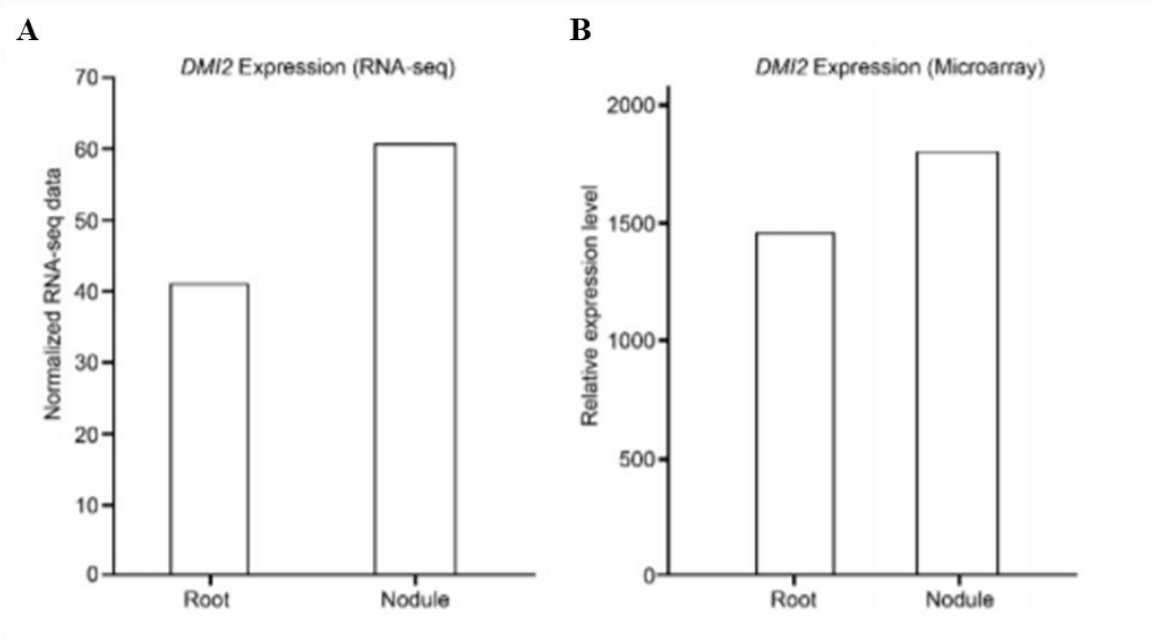


Figure 5.2. The expression of *DMI2* is not induced to a significantly higher level in nodules comparing with roots. A and **B.** data of *DMI2* expression from RNA-Seq and microarray databases respectively. Numbers in the vertical axis indicate normalized expression data from different databases.

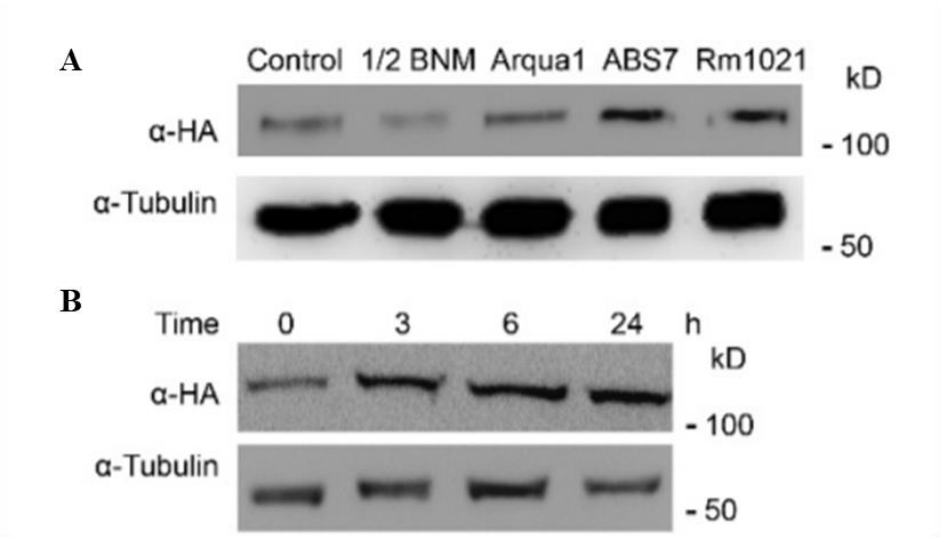


Figure 5.3. The induction of DMI2 protein by rhizobia is strain specific and the responsive time is less than 3 hours. A. DMI2-HAST protein level after inoculation with ½ BNM media, *Agrobacterium rhizogenes* strain Arqua1, *Sinorhizobium meliloti* strain ABS7 and Rm1021. ½ BNM and Arqua1 could not induce the protein level of DMI2-HAST. Medicago plants were treated with indicated media or bacteria for 24 hours and the protein level of DMI2-HAST was tested. **B.** DMI2 protein could be induced by rhizobia in 3 hours. The protein level of DMI2-HAST protein in rhizobia-treated Medicago roots at different time points post *S. meliloti* ABS7 treatment were shown. "h", hours post inoculation. Tubulin protein was used as the internal control.

Results

Rhizobia induce DMI2 protein at the posttranscriptional level

To determine the dynamics of DMI2 protein levels during nitrogen-fixing symbiosis, a stable transgenic line in the Medicago *dmi2*-1 background was used. This line expresses the *DMI2* genomic sequence driven by its native promoter; it is also fused to a dual affinity tag containing three copies of the hemagglutinin epitope (HA) and a single StrepII (ST); therefore, the resulting protein is named DMI2-HAST. The *gDMI2:HAST* construct complements the phenotype of *dmi2* mutants, and the protein is easy to detect (Riely et al., 2013). After inoculating the transgenic lines with *Sinorhizobium meliloti* strain ABS7, we compared the protein accumulation of DMI2-HAST between nodules and untreated roots. The result showed that, before rhizobia treatment, the DMI2-HAST protein level in the roots was very low (Figure 5.1A), which is consistent with previous reports that DMI2-HAST protein is almost undetectable (Riely et al., 2013). In nodules, the protein level of DMI2-HAST was much higher compared with rhizobia-free roots (Figure 5.1A). Furthermore, by analyzing the expression of *DMI2* in nodules and uninoculated roots in the *Medicago truncatula* Gene Expression Atlas database (Benedito et al., 2008), we found that the transcription of *DMI2* is not highly activated in whole nodules (Figure 5.2A and B), which also is consistent with a previous report using northern-blot assays (Bersoult et al., 2005). These results suggest that the DMI2-HAST protein accumulates in the nodules through posttranscriptional regulation.

To further investigate the protein level variation of DMI2 during rhizobia inoculation, we treated *dmi2-1 gDMI2:HAST* plant roots with rhizobia strain ABS7 and

checked protein abundance during rhizobia infection at the earliest stage. Twenty-four hours post ABS7 strain treatment, the protein level of DMI2-HAST increased dramatically compared with untreated roots (Figure 5.1B). Treating *dmi2-1 gDMI2:HAST* plants with one-half-strength basic nodulation medium, the liquid medium for rhizobia inoculation, *Agrobacterium rhizogenes* strain Arqua1, and *S. meliloti* strains Rm1021 and ABS7 showed that the DMI2-HAST protein level increased only during rhizobia inoculation (Figure 5.3A), indicating that the DMI2-HAST induction effect is specific to rhizobia. To establish how quickly the protein accumulates, we analyzed the protein level of DMI2-HAST at different time points post rhizobia treatment and found that, as early as 3h after ABS7 inoculation, the DMI2-HAST protein level was already induced (Figure 5.3B). To rule out the possibility that the accumulation of DMI2-HAST was the result of transcriptional activation, we checked the expression level of *DMI2* by reverse transcription-quantitative PCR (RT-qPCR) as well as by browsing the *Medicago truncatula* Gene Expression Atlas and the *Medicago truncatula* Genome Project version 4.0 database (Benedito et al., 2008; Krishnakumar et al., 2015). The RT-qPCR assay showed that the transcription of *DMI2* was not induced substantially at 3, 6, 12, and 24 h after rhizobia inoculation, with microarray and RNA sequencing data showing similar results (Figure 5.4), which is consistent with previous reports that *DMI2* transcripts are not induced significantly by rhizobia treatment in a northern-blot assay (Mirabella et al., 2005). Taking these results together, we conclude that rhizobia treatment induces the abundance of DMI2 protein by affecting posttranscriptional regulation at a very early stage of symbiosis.

It is reported that DMI2/SYMRK interacts directly with the Nod factor receptors to perceive Nod factors (Antolín-Llovera et al., 2014b). To determine whether the accumulation of DMI2-HAST protein in the presence of rhizobia is dependent on Nod factor receptors, we expressed *gDMI2:HAST* in *hcl-1* and *hcl-4* (two mutant alleles of the Medicago Nod factor receptor gene *LYK3*; Smit et al., 2007) as well as in *dmi2-1*. Similar to *dmi2-1*, 24 h post rhizobia treatment, the protein level of DMI2-HAST in *hcl-1* and *hcl-4* mutants accumulated to a much higher level compared with the control (Figure 5.1C). We also performed this experiment with the mutant of another Nod factor receptor, NFP, and found similar results (Figure 5.1D). In various experiments detecting the protein level changes of DMI2-HAST, the intensity of the DMI2-HAST bands in uninoculated Medicago roots ranged from very weak to difficult-to-detect. It has been reported that, in *hcl-1*, *hcl-4*, and *nfp* mutants, the expression of *DMI2* transcripts was identical to that in the wild-type (Mirabella et al., 2005), indicating that the accumulation of DMI2-HAST protein in *hcl-1*, *hcl-4*, and *nfp* also is posttranscriptionally controlled.

To further confirm that the induction of DMI2 protein level by rhizobia was independent of Nod factor perception, we inoculated Medicago plants with wild-type *S. meliloti* RM1021 and two rhizobium mutant strains, RJW14 and JT210, which have defects in Nod factor synthesis (Wais et al., 2002). The protein level of DMI2 was induced to a similar level by RJW14 and JT210 compared with the wild-type strain (Figure 5.5). These results show that DMI2 protein accumulation induced by rhizobia is independent of Nod factor perception, suggesting the existence of another as-yet-unknown rhizobia signal.

SUNN (SUPER NUMERIC NODULES) is an LRR receptor kinase that functions in the shoot to regulate nodule numbers, and *sunn* mutants display a supernodulation phenotype (Penmetsa et al., 2003; Schnabel et al., 2005). Surprisingly, it was reported that overexpressing the *DMI2/SYMRK* kinase domain in *sunn* mutants decreases the number of nodules formed in these supernodulation mutants (Saha and DasGupta, 2015). To analyze whether SUNN could affect the induction of DMI2-HAST protein by rhizobia, we transformed *gDMI2-HAST* into *dmi2-1* and *sunn-1* backgrounds and checked the protein level of DMI2-HAST during ABS7 strain infection. The result showed that DMI2-HAST was induced to a similar level in *sunn-1* and *dmi2-1* mutant backgrounds (Figure 5.1E), suggesting that the induction of DMI2 protein by rhizobia also is independent of SUNN.

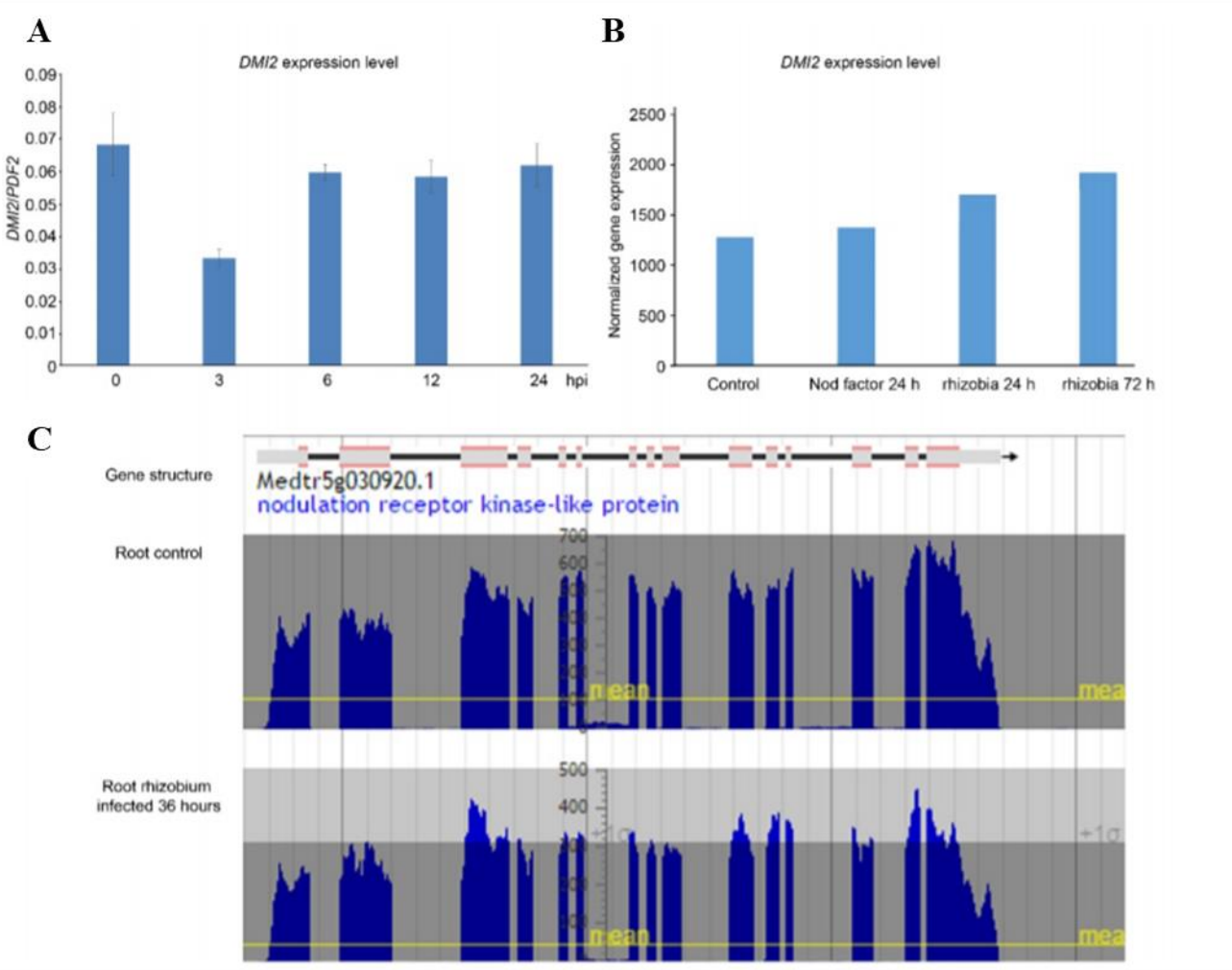


Figure 5.4. The transcript of *DMI2* is not dramatically induced by rhizobia or Nod factor treatment. A. In quantitative real-time PCR assay, the expression level of *DMI2* was not induced by rhizobia at the early stage post inoculation. Wild-type Medicago plants were treated with *S. meliloti* ABS7 strain and plant samples were taken at 0, 3, 6, 12, 24 hours post inoculation. The expression of *DMI2* was normalized to that of *PDF2* gene. Experiments were repeated two times with similar results. "hpi" means hours post inoculation. **B.** Microarray data show that the expression of *DMI2* was not induced significantly by rhizobia treatment. Data were obtained from Medicago Gene Expression Atlas. The horizontal axis shows treatment type and time. "h" means hours after treatment. **C.** RNA-Seq data show the expression of *DMI2* in Medicago roots is not induced to a much higher level with rhizobia treatment for 36 hours. Data were obtained from *Medicago truncatula* Genome Project v4.0

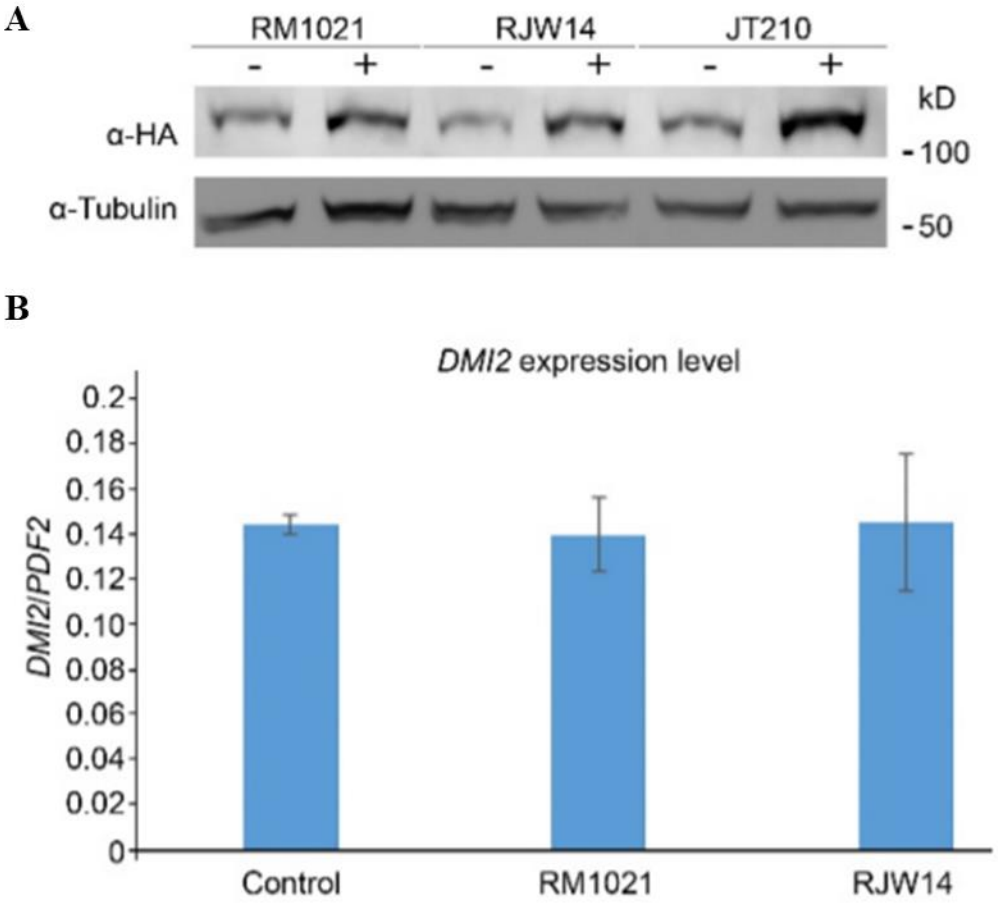


Figure 5.5. DMI2 protein level can be induced by rhizobia strains with defects in Nod factor synthesis. A. DMI2-HAST level was induced with *S. meliloti* Rm1021, RJW14 and JT210 treatment respectively. *dmi2-1 gDMI2-HAST* plants were treated with indicated rhizobia strains for 24 hours and the protein level of DMI2-HAST was tested. α-HA was used to detect the DMI2-HAST, Tubulin was used as the control. **B.** The expression of *DMI2* in Medicago roots was not induced to a much higher level by Rm1021 or RJW14 treatment for 24 hours. Quantitative Real-Time PCR assays were performed with samples from the untreated control plants, RM1021- and RJW14-treated plants. The expression of *DMI2* was normalized to that of *PDF2* gene. The statistic assay was done using Student's t-test. Experiments were repeated two times with similar results.

DMI2 protein is constitutively degraded in uninoculated Medicago roots

The accumulation of DMI2 protein by rhizobia is regulated at the posttranscriptional level, indicating that protein stability regulation may play a role. To test this hypothesis, we performed a cell-free degradation assay using protein samples from *dmi2-1 gDMI2-HAST* roots grown in a sterile environment. DMI2-HAST protein from uninoculated roots was completely degraded within 2h (Figure 5.6A). To study which mechanism was responsible for the degradation of DMI2-HAST, several different proteolysis inhibitors were tested. While the protease inhibitor phenylmethylsulfonyl fluoride and a plant-specific protease inhibitor cocktail failed to rescue the protein level of DMI2, the proteasome inhibitor MG132 largely prevented DMI2-HAST protein degradation (Figure 5.6A), showing that, in uninoculated roots, DMI2-HAST may be degraded in a proteasome-dependent manner.

The *in planta* protein degradation of DMI2-HAST was then examined by treating *dmi2-1 gDMI2:HAST* roots with MG132. Four hours post MG132 treatment, the accumulation of DMI2-HAST protein was enhanced significantly (Figure 5.6B), showing that the inhibition of proteasome activity mimics the effect of rhizobia treatment. When treating *hcl-1* and *hcl-4* with MG132, the protein level of DMI2-HAST was similar to that in *dmi2-1* plants, suggesting that the accumulation of DMI2 protein by MG132 treatment also is independent of Nod factor receptors (Figure 5.6C). To rule out that MG132 treatment promoted the transcriptional activation of *DMI2*, we analyzed the dynamics of *DMI2* transcripts during MG132 treatment by RT-qPCR. The transcript level of *gDMI2:HAST* was indistinguishable before and after MG132 treatment (Figure 5.7). These data show that MG132 can block the degradation of DMI2-HAST protein in

uninoculated Medicago roots, suggesting that the DMI2-HAST protein is degraded through a proteasome apparatus in the absence of rhizobia.

Recently, there have been several reports showing that overexpressing the kinase domain or the full-length DMI2/SYMRK protein can induce a spontaneous nodulation phenotype in legume plants (Ried et al., 2014; Saha et al., 2014). To evaluate whether MG132-induced DMI2 protein accumulation carries any biological significance, we checked the expression of *ENOD11* (*EARLY NODULIN11*) and *NIN* (*NODULE INCEPTION PROTEIN*), two marker genes of early nodule development (Schauser et al., 1999), in *dmi2-1* and *dmi2-1 gDMI2:HAST* roots after MG132 treatment. As shown in Figure 5.4D and E**,** 4 h after MG132 treatment in *dmi2-1 gDMI2:HAST* roots, the expression of *NIN* and *ENOD11* was induced to a higher level compared with *dmi2-1* mutants. We conclude that DMI2 protein accumulation may be able to partially activate downstream nodulation signaling.

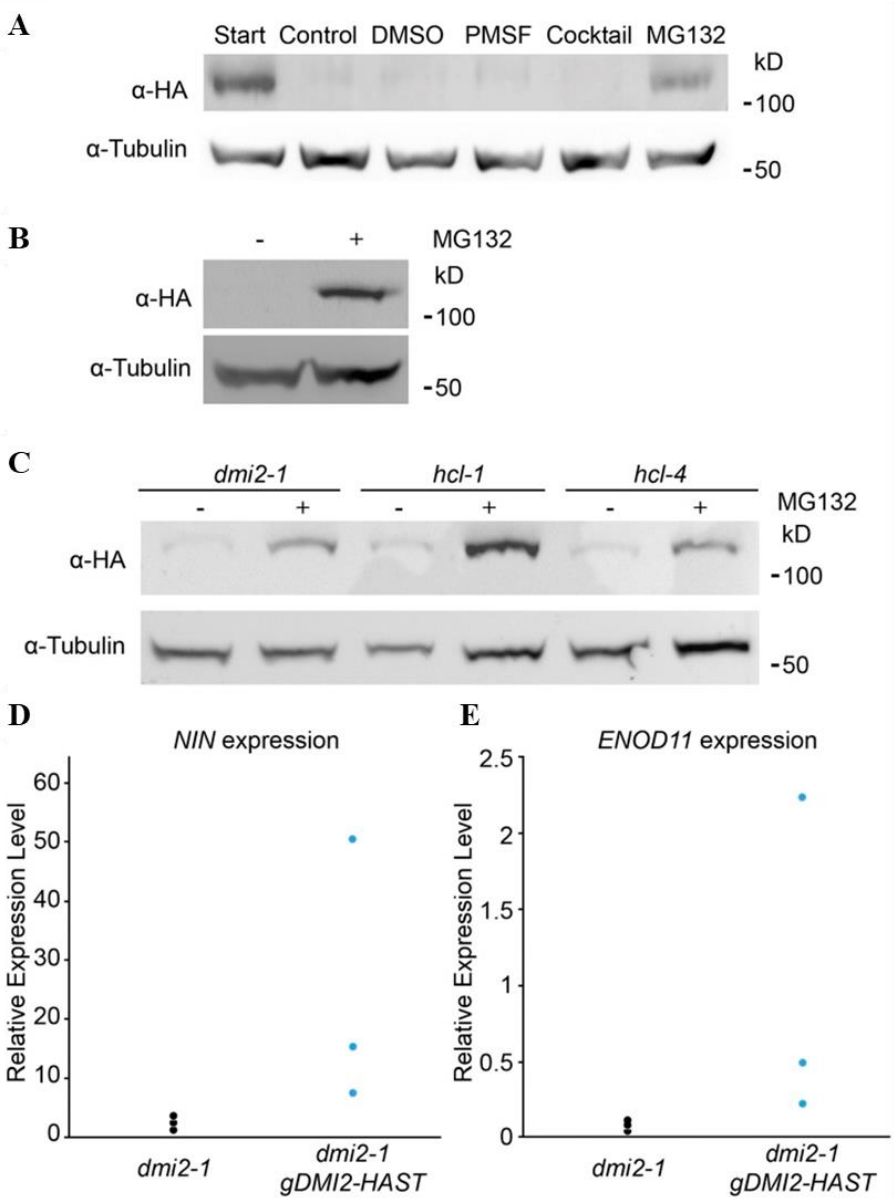


Figure 5.6. MG132 can block the degradation of DMI2 protein in uninoculated *Medicago* roots. A. MG132 could partially rescue DMI2-HAST protein level in a cell-free degradation assay. Cell-free DMI2-HAST protein samples were treated with different reagents for 2h and subjected to western blot using α-HA. Tubulin was used to show equal loading. Start, Freshly extracted protein sample; Cocktail, plant-specific protease inhibitor mixture. **B.** MG132 treatment could induce the protein level of DMI2-HAST *in vivo*. Uninoculated Medicago roots were treated with 100μM MG132, and the level of DMI2-HAST protein was detected. Experiments were repeated at least five times with similar results. **C.** MG132 could block the degradation of DMI2 independent of Nod factor reception. Samples were taken from *dmi2-1*, *hcl-1*, and *hcl-4* plants expressing *gDMI2-HAST* and treated with or without 100μM MG132 for 4h. **D.** and **E.** After MG132 treatment, the expression of *NIN* and *ENOD11* was modestly induced in 2-week-old *dmi2-1 gDMI2-HAST Medicago* roots. Each dot represents a biological replicate. α-HA was used to detect the DMI2-HAST protein level, and tubulin was used as the internal control. Experiments were repeated two times with similar results.

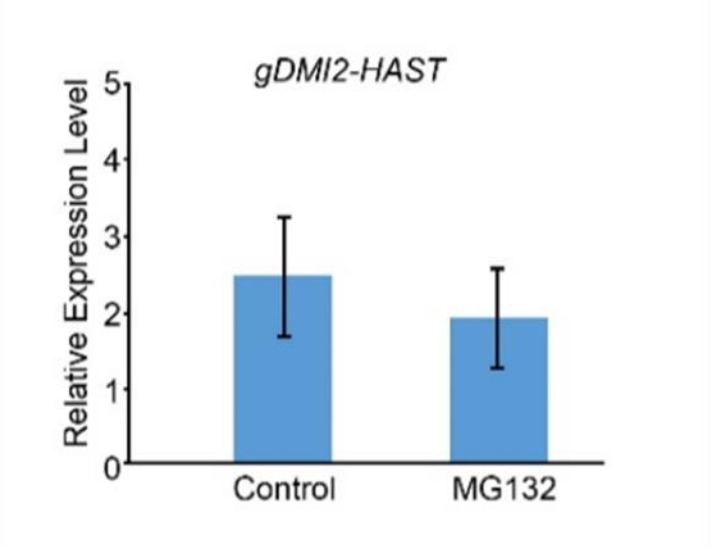


Figure 5.7. The transcriptional level of *DMI2-HAST* is not induced by MG132 treatment. Stable transformed *dmi2-1 gDMI2-HAST* plant roots with or without 100μM MG132 for 4 hours were collected for RNA extraction. In the quantitative Real Time PCR analysis, the expression of *DMI2-HAST* was normalized to that of *PDF2*. Error bar indicates standard deviation. Experiments were repeated 3 times with similar results.

DMI2 protein is protected from degradation in inoculated roots

To further test the effect of MG132 on DMI2 protein in inoculated Medicago roots, *S. meliloti* ABS7-inoculated *dmi2-1 gDMI2-HAST* roots were treated with MG132 for 4h. Compared with untreated roots, the protein level of DMI2-HAST was not increased further (Figure 5.8), showing that MG132 has little impact on DMI2 protein level in the presence of rhizobia, which indicates that, in inoculated Medicago roots, proteasome-mediated degradation of DMI2 has already been blocked (Figure 5.6B). Thus, during nitrogen-fixing symbiosis, DMI2 protein is protected from proteasome-mediated degradation.

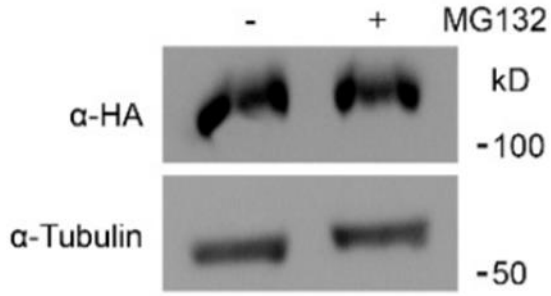


Figure 5.8. MG132 does not affect DMI2 protein level in rhizobia inoculated Medicago roots. Stable transformed *dmi2-1 gDMI2-HAST* roots were inoculated with ABS7 for 14 days and treated with or without 100μM MG132 for 4 hours. The protein level of DMI2-HAST was indistinguishable after MG132 treatment in the tissues. Experiments were repeated 5 times with similar results.

MLD is vital for DMI2 function

In the extracellular region, DMI2 has an LRR and an MLD. To gain more insights about the MLD, we aligned the MLD with human Malectin protein using a homology-modeling method (http://swissmodel.expasy.org/; Biasini et al., 2014). While DMI2-MLD was described originally as a Malectin-like sequence, we found that it actually contains two tandem matches to the human Malectin A domain, and each match has about 140 amino acids (Figure 5.9). The human Malectin protein is reported to bind Glc2Man9GlcNAc2 and regulate protein folding (Schallus et al., 2008). However, very little is known about the function of DMI2-MLD or any MLD in plants.

To study the function of DMI2-MLD, we aligned the amino acid sequence of DMI2-MLD with its close homologs in the plant kingdom. The results show that DMI2-

MLD is conserved among its homologs in dicots (Figure 5.11), suggesting that MLD may have a conserved role. To gain further insights into the function of DMI2-MLD, we performed site-directed mutagenesis. We analyzed the following point mutations: DMI2C39D, DMI2Y95A, DMI2F111A, DMI2Y164A, DMI2F271A, DMI2Y291A, and DMI2F294A, where the numbers indicate the locations of these residues in the full-length DMI2 protein, including the signal peptide sequence (Figure 5.11). We targeted these amino acids because they were conserved among the homologs included in the alignment (Figure 5.10), pointing to a greater possibility that these point mutations may affect the function of DMI2-MLD. Among them, Y95A, Y164A, and Y291A may be in the predicted ligand-binding pockets (Figure 5.10; Schallus et al., 2008)).

To investigate whether mutating the conserved amino acids in MLD would affect the function of the DMI2 protein, we introduced wild-type *gDMI2-HAST* and the seven MLD point-mutated versions into *dmi2-1* mutant plants using the hairy root transformation method. As shown in Table 5.1, 14d after rhizobia inoculation, *dmi2-1* plants expressing wild-type *gDMI2-HAST* generated many nodules, while *dmi2-1* roots expressing MLD point mutants had few nodules, and even fewer of which were pink, suggesting a failure in nitrogen fixation (Figure 5.12A). Some point mutations, such as DMI2C39D, had no nodules at all (Table 5.1). These results show that, regarding sufficient nodule development, MLD is vital for the proper function of DMI2.

While *dmi2-1* plants transformed with wild-type *gDMI2-HAST* generated many pink nodules, roots expressing *gDMI2-HAST* versions with amino acid substitutions in MLD were impaired in nodule development and generated very few pink nodules.

To enter legume roots, rhizobia normally have to penetrate plant cells through a plant-derived tubular structure at infected root hair cells, known as the infection thread (Oldroyd et al., 2011). It has been shown that, in *dmi2/symrk* mutants, the development of the infection thread is blocked (Endre et al., 2002; Stracke et al., 2002). To find out whether blocking the function of MLD could affect the role of DMI2 in infection thread formation, we checked the infection thread phenotype of *dmi2-1* mutants transformed with wild-type *gDMI2-HAST* and the seven MLD point-mutated versions of *gDMI2-HAST*. Three days after rhizobia inoculation, infection threads could be seen using the microscope in *dmi2-1* mutants transformed with wild-type *gDMI2-HAST*. In contrast, there were few infection thread-like structures in the plants transformed with *gDMI2-HAST* containing point mutations in MLD (Figure 5.13). Counting the numbers of infection threads, we found that *dmi2-1* roots transformed with wild-type *gDMI2-HAST* could produce large numbers of infection threads, but the roots transformed with *gDMI2-HAST* with point mutations in MLD had very few infection threads (Figure 5.12B). These results show that MLD is necessary for DMI2 protein to function properly in infection thread formation. Taken together, the proper function of MLD is necessary for DMI2 protein to play a fundamental role in nodule development at a very early stage.

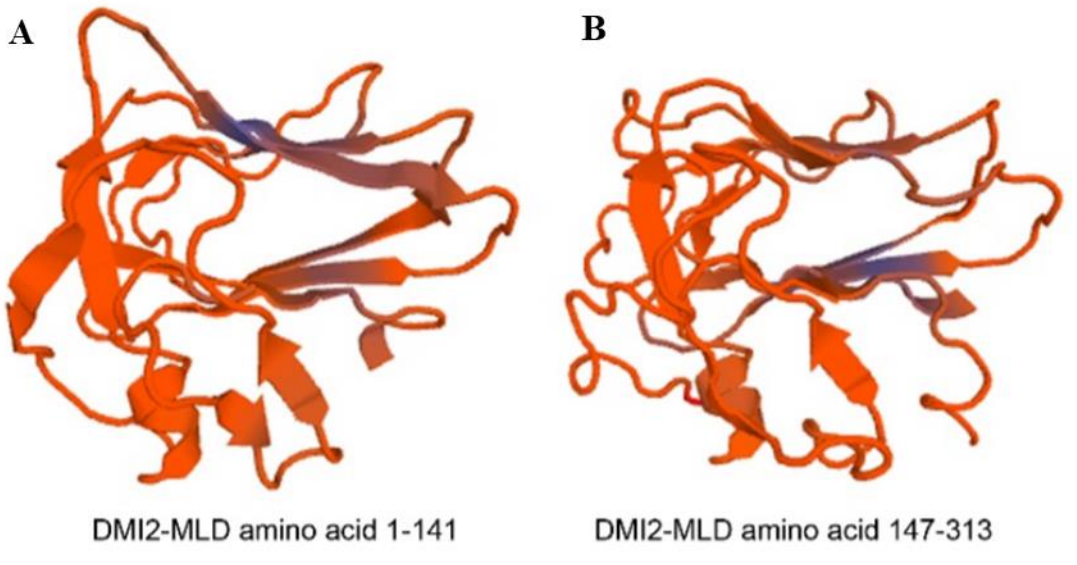


Figure 5.9. Predicted protein structure of two human MALA (PDB code: 2kr2.1.A) matches in DMI2-MLD. DMI2-MLD has 313 amino acids in total, the first MALA match (**A**) is 141 amino acids, and second match (**B**) is 166 amino acids. The broad bands in the diagram indicate predicted β-sheet structure. The predictive protein structure reconstruction was done using online tool SWISS-MODEL.

MLD Is Required for the Homeostasis of DMI2

Since the protein level of DMI2 is important for its proper function in nodule development, we studied whether the amino acid substitutions in MLD would affect the protein level of DMI2. In Medicago roots expressing either wild-type *gDMI2-HAST* or MLD point mutations of *gDMI2-HAST*, the protein could be detected in plants transformed with wild-type sequence, but the plants transformed with *gDMI2-HAST* containing amino acid substitutions in MLD could not generate detectable DMI2-HAST protein, with or without rhizobia inoculation (Figure 5.12C). To rule out the possibility that the disappearance of the MLD point-mutated protein is regulated at the transcriptional level, we checked the transcripts of wild-type *gDMI2-HAST* and the MLD point mutation versions using RT-qPCR. We found that transgenic plants could generate

comparable amounts of wild-type *gDMI2-HAST* and MLD point mutation transcripts (Figure 5.14), albeit the expression level varied due to variations in hairy root transformation. These results show that the proper function of MLD is critical for the precise regulation of DMI2 protein abundance at the posttranscriptional stage.

To find out whether the degradation of MLD amino acid substitution versions of DMI2 protein is dependent on the proteasome apparatus, we treated *dmi2-1* plants expressing wild-type *gDMI2-HAST* and MLD point mutation versions with MG132 and checked DMI2 levels before and after MG132 treatment. As shown in Figure 5.12D, 4h after MG132 treatment, the protein level of wild-type full-length DMI2-HAST was strongly induced, while MLD point mutation versions could not accumulate the protein. This result shows that the degradation of DMI2 containing point mutations in MLD could not be rescued by inhibiting proteasome activity with MG132. Considering that the human Malectin protein binds to carbohydrate and functions in ER quality control (Qin et al., 2012; Schallus et al., 2008), it is possible that the MLD is required for the proper folding of DMI2 in the ER; blocking MLD function may activate ER quality control signaling and result in the degradation of DMI2.

As MLD is indispensable for the function of DMI2 in Medicago, we further investigated the origin of MLD and whether it is conserved in the plant kingdom. We constructed a phylogenetic tree of DMI2 in dicots, monocots, and basal angiosperm species and found that MLD is not present in monocot homologs of DMI2, but the protein from *Amborella trichopoda*, a basal angiosperm species, has an MLD (Figure 5.15). On the other hand, we examined the phylogenetic tree of the closest DMI2 paralog in *M. truncatula*, *Medtr7g057170*. To our surprise, the orthologs of this gene from every

species had an MLD. These results suggest that MLD is important for the function of DMI2 in dicots and basal angiosperms; however, in monocot plants, the MLD is missing, indicating that there may be other proteins functioning together with DMI2 to perform the function of MLD (Figure 5.15).

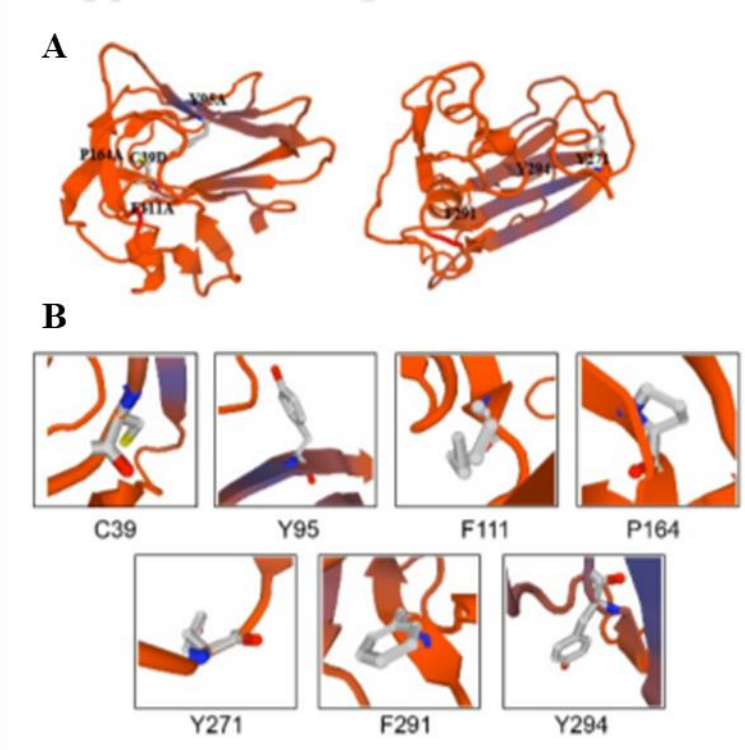


Figure 5.10. Positions of amino acid substitutions in DMI2-MLD protein sequence. A. The illustration of seven point mutations in the predicted DMI2-MLD structure. The first letter shows the original amino acid, number shows the position of the amino acid in DMI2 whole protein sequence, the last letter shows the amino acid used for point mutation. **B.** The putative local structures of seven conserved amino acids chosen for generating point mutations. The predictive protein structure reconstruction was done using online tool SWISS-MODEL.

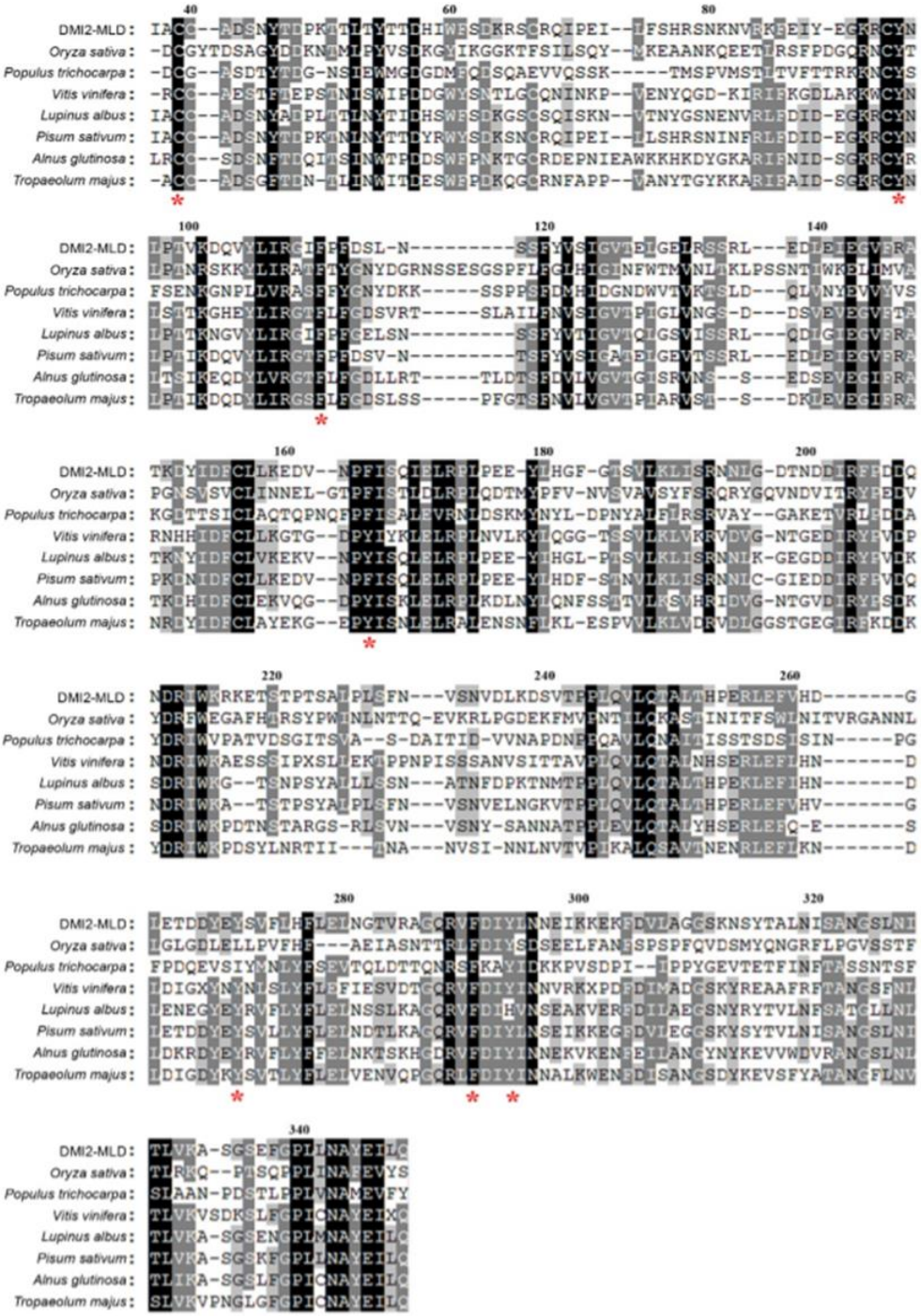


Figure 5.11. Alignment of DMI2-MLD with its closest plant homologs. Protein sequences were aligned using ClustalW Omega and the shaded using Gendoc software. Residues identical or with very similar character among all the homologs were shaded in black background; Residues conserved among part of tested homologs were shaded in grey. Red "*" indicate the residues picked for amino acid substitution analysis. Numbers indicate the position of the amino acids in DMI2 full length sequence. DMI2 has a signal peptide sequence at the N terminal which contains 36 amino acids, so numbering of the DMI2-MLD amino acids started from 37.

Table 5.1 Summary of the complementation assay using wild-type and MLD point mutation versions of *gDMI2-HAST*. While *dmi2-1* plants transformed with wild-type *gDMI2-HAST* could generate a large number of pink nodules, roots expressing *gDMI2-HAST* versions with amino acid substitutions in MLD were impaired in nodule development and generated very few pink nodules. Control, Plants transformed with empty vector. Transformed roots were selected based on the mCherry fluorescence marker of the *gDMI2-HAST* construct. Experiments were repeated more than three times with similar results.

Parameter	Wild Type	C39D	Y95A	F111A	P164A	Y271A	F291A	Y294A	Control
Plants in total	13	14	11	12	13	14	12	11	14
Transformed plants	13	13	11	12	13	14	12	11	14
Plants with nodules	13	0	9	4	9	1	5	3	0
Nodule numbers	>300	0	29	4	17	1	19	4	0
Pink nodules	>300	0	5	0	0	1	2	0	0

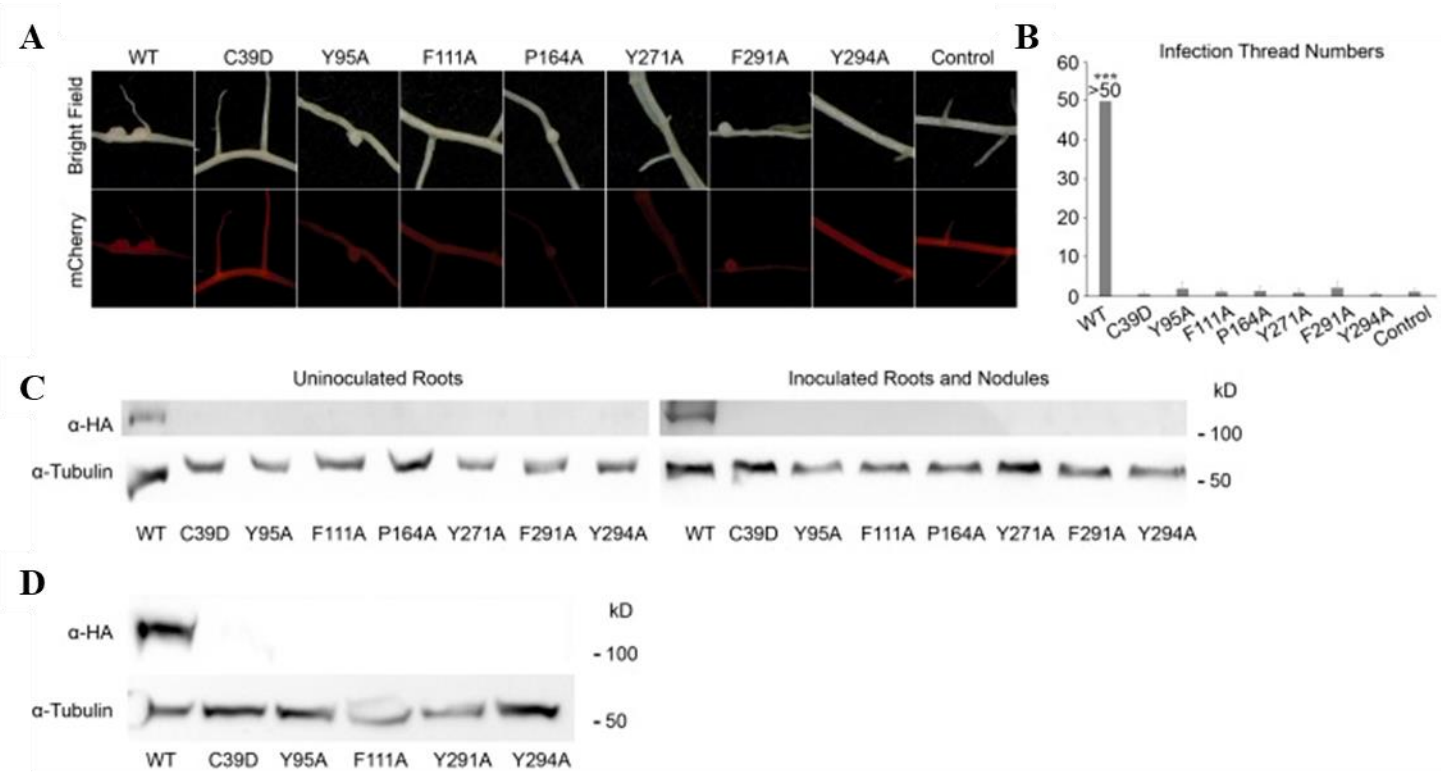


Figure 5.12. MLD is required for DMI2 protein to function properly in the early nodule development process and to stabilize full-length DMI2 proteins in plants. A. Images showing representative nodules in *dmi2-1* plants transforming wild-type (WT) *gDMI2-HAST* and MLD point mutation versions of *gDMI2-HAST*. Although some point mutation versions could generate nodules, most of them were white and small. mCherry fluorescence was used to select the transformed roots. **B.** *dmi2-1* plants expressing *gDMI2-HAST* containing point mutations in MLD had few infection threads in a quantitative assay. The number of infection threads in *dmi2-1* plants expressing wild-type *gDMI2-HAST* was estimated to be more than 50. More than 10 transformed plants were used for each line. Data represent means and SD. Asterisks show a significant difference (***, $P < 0.001$, Student's *t* test). **C.** DMI2 protein containing amino acid substitutions in MLD is unstable in *Medicago* plants, with or without rhizobia inoculation. In *dmi2-1 gDMI2-HAST* plants, DMI2-HAST protein could be seen, while the MLD amino acid substitution version proteins were totally undetectable. **D.** MG132 treatment could not rescue the constitutive degradation of DMI2 protein containing amino acid substitutions. *dmi2-1* plants expressing wild-type *gDMI2-HAST* and MLD point mutation versions were treated with 100μM MG132, and the protein level was tested. Experiments were repeated three times with similar results.

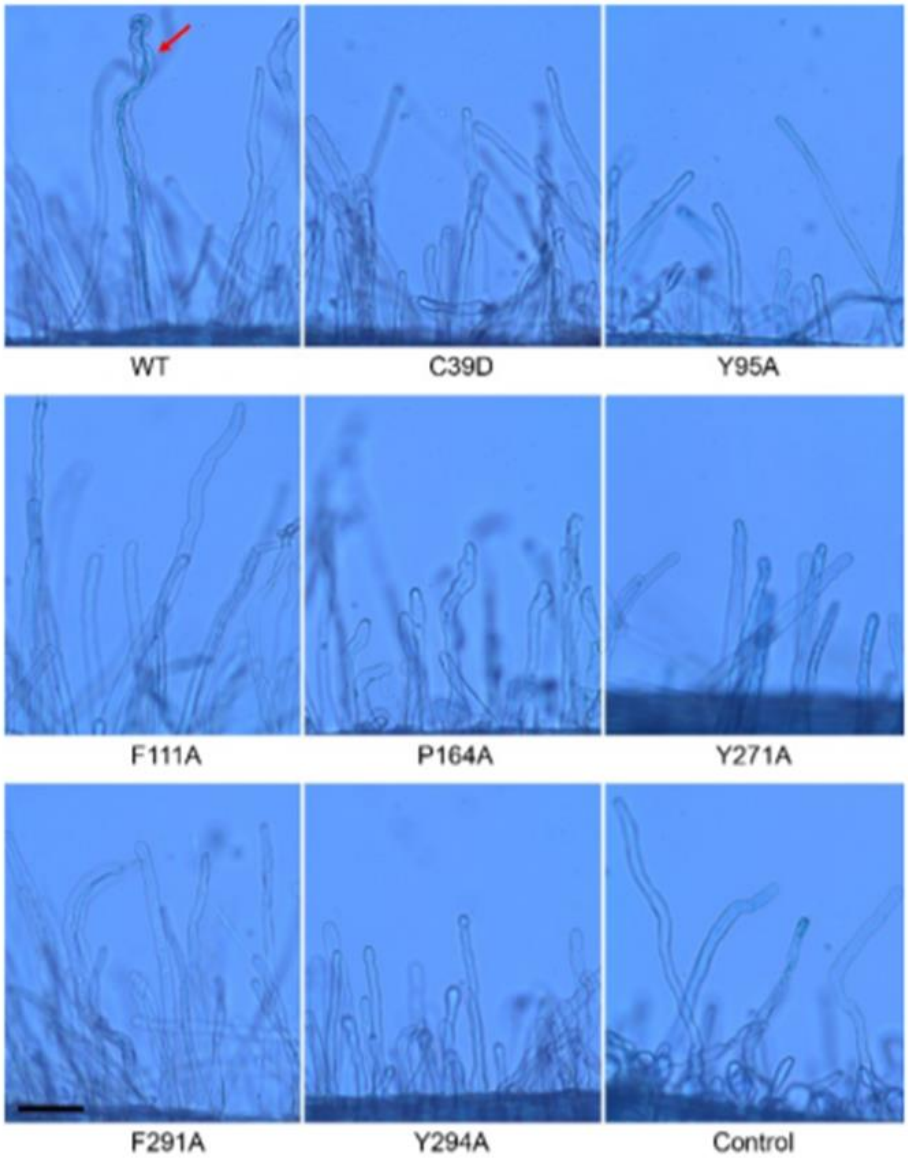


Figure 5.13. Representative pictures showing the infection thread development phenotype in *dmi2-1* plants expressing wild-type *gDMI2-HAST* or the versions containing point mutations in MLD. Red arrow indicates a wild type infection thread. Bar= 0.1 mm.

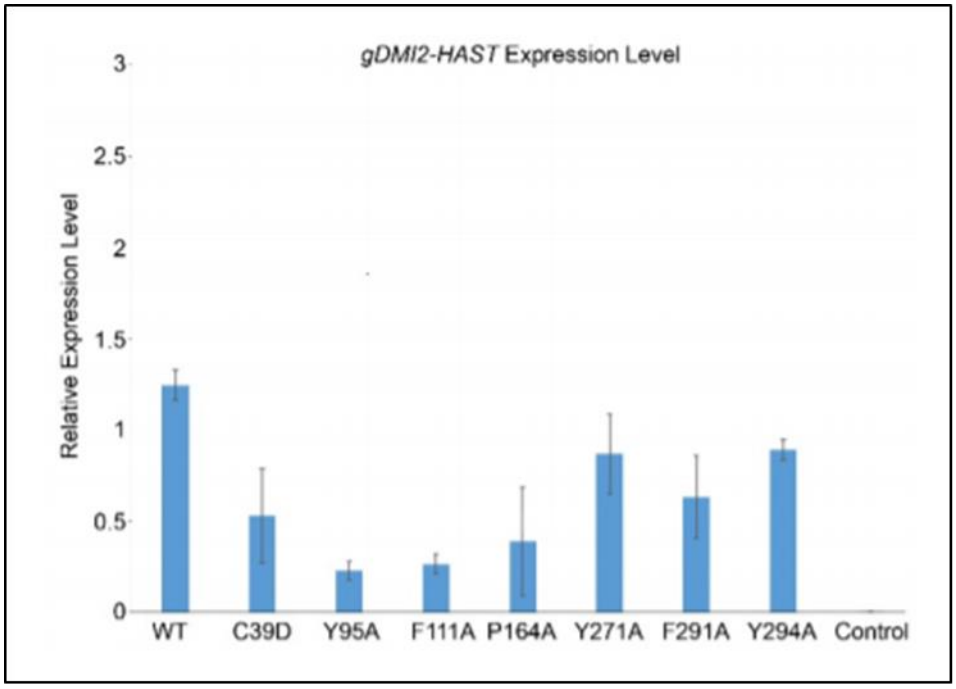


Figure 5.14. The expression of wild-type *DMI2-HAST* and the versions with amino acid substitutions in MLD were detectable in quantitative Real-Time PCR assay. The expression of *DMI2-HAST* was normalized to that of *PDF2* gene. Error bars indicate standard deviation. Experiments were repeated 3 times with similar results.

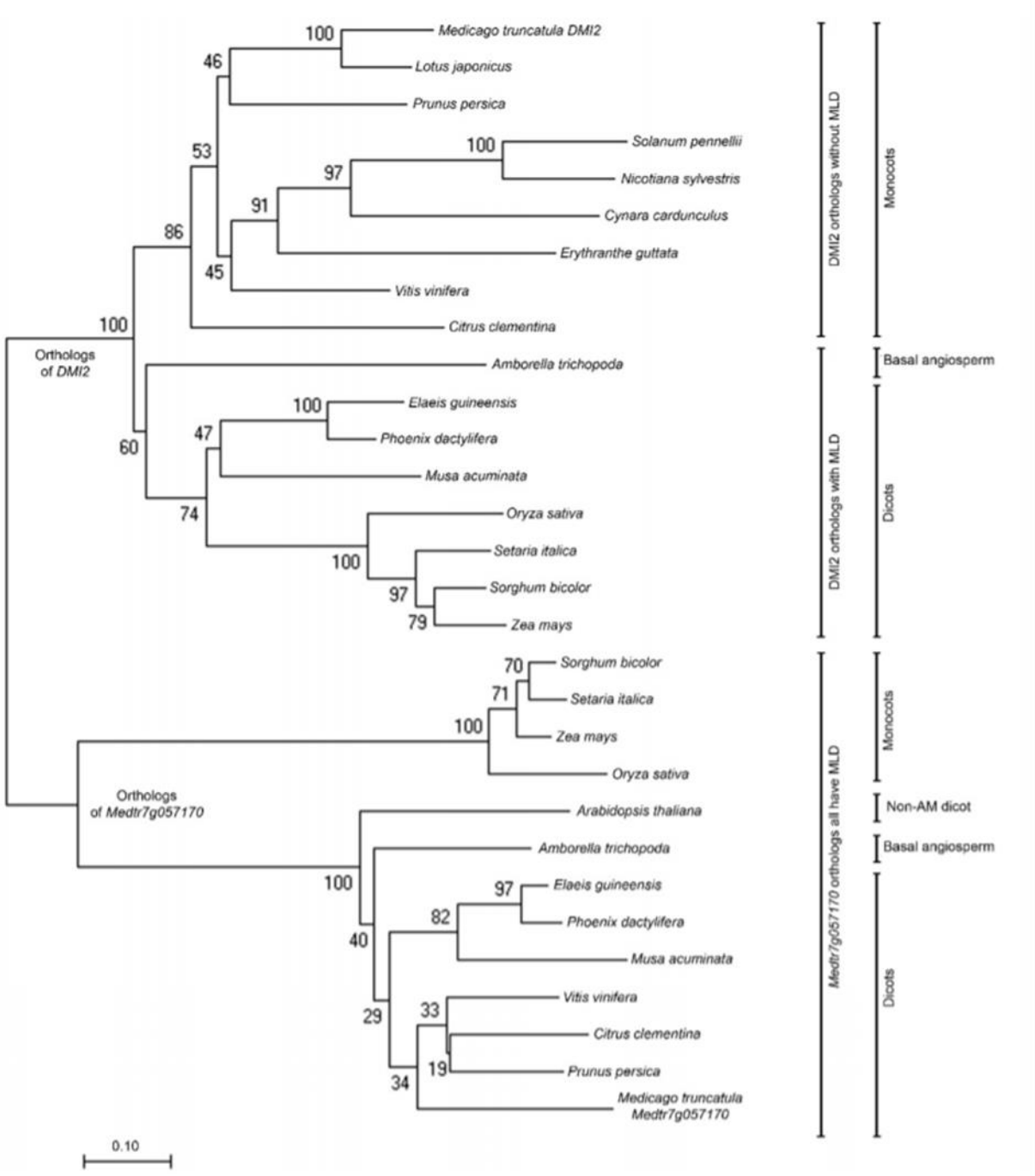


Figure 5.15. Phylogenetic tree of DMI2 and its closest Medicago homolog in various plant species. The orthologs of DMI2 from monocot species do not have MLD in their sequences, the rest of DMI2 orthologs from angiosperm species we checked have MLD, including the basal angiosperm *Amborella tricopoda*. Meanwhile in every angiosperm species we tested, orthologs of *Medtr7g057170* (the closest homolog of DMI2 in Medicago) have MLD in their sequences. Phylogenic tree was generated using the neighbor-joining method in MEGA 7. Bootstrap values (percent as numbers) were determined based on results for 1000 replicates. The scale bar indicates 0.1 nucleotide substitutions per sequence position.

Discussion

Rhizobia inoculation can block the protein degradation of DMI2

Localized to the plasma membrane, RLKs can detect environmental changes through extracellular domains and transduce the signals into cells through their intracellular catalytic domains (Lemmon and Schlessinger, 2010). To ensure the proper activation of RLKs during development and stress-response processes, the protein levels of RLKs must be kept in check by plants to avoid runaway intracellular responses. Here, we report that the protein level of the *M. truncatula* symbiosis receptor DMI2 is regulated at the posttranslational level to modulate the proper response to rhizobia inoculation.

DMI2 plays a key role in regulating plant-microbe symbiosis, as it is required for the symbiosis between plants and AM fungi, Frankia symbiosis, and legume plant-rhizobia nitrogen-fixing symbiosis (Gherbi et al., 2008). Furthermore, DMI2 displays protein level increases coinciding with rhizobia inoculation. When plants are grown in the absence of rhizobia, the DMI2 protein is kept at a very low level (Figure 5.1A and B); during rhizobia infection, through a currently unknown mechanism, rhizobia block the degradation of DMI2, resulting in increased protein levels. As a consequence, accumulated DMI2 induces plant roots to start the process of nodule development. This is consistent with previous reports that, when the DMI2/SYMRK kinase domain or full-length protein is overexpressed in legume roots, plants will generate spontaneous nodules without rhizobia infection (Ried et al., 2014; Saha et al., 2014).

We also found that MG132 treatment caused accumulation of the protein level of DMI2 in uninoculated Medicago roots (Figure 5.6A and B), indicating that the turnover of DMI2 in the absence of rhizobia infection is through proteasome-mediated protein

degradation. More importantly, MG132 treatment mimicked the effect of rhizobia inoculation with respect to activating *NIN* and *ENOD11* (Figure 5.6D and E). In inoculated roots, MG132 treatment did not increase the protein level of DMI2 (Figure 5.8), suggesting that, after rhizobia inoculation, DMI2 protein is already protected from proteasome degradation. The robust induction of DMI2 protein levels by MG132 also suggests that MG132 treatment can be used for further characterization of the DMI2 protein. For example, it may be used to find DMI2-interacting partners, especially the substrates of the DMI2 kinase domain.

Although the mechanisms coupling ligand recognition in the extracellular parts and the activation of intracellular catalytic domains are surprisingly diverse (Lemmon and Schlessinger, 2010), there are reports that ligand binding induces the protein levels of RLKs. For instance, it is widely known that, in human cells, insulin treatment can increase the intracellular protein abundance of insulin receptors (Lemmon and Schlessinger, 2010). In *Arabidopsis thaliana*, FLS2 (FLAGELLIN SENSING2) is an LRR RLK that functions as a receptor for bacterial flagellin protein or flg22, a 22-amino acid active peptide derivative (Gómez-Gómez and Boller, 2000). FLS2 can be ubiquitinated by two related U-box E3 ligases, PUB12 and PUB13, in the absence of the pathogen (Lu et al., 2011). The ubiquitination and subsequent degradation of FLS2 protects plant cells from the harm of the excessive activation of defense responses. These findings indicate that the induction of RLK protein levels by their ligands may be a widespread phenomenon.

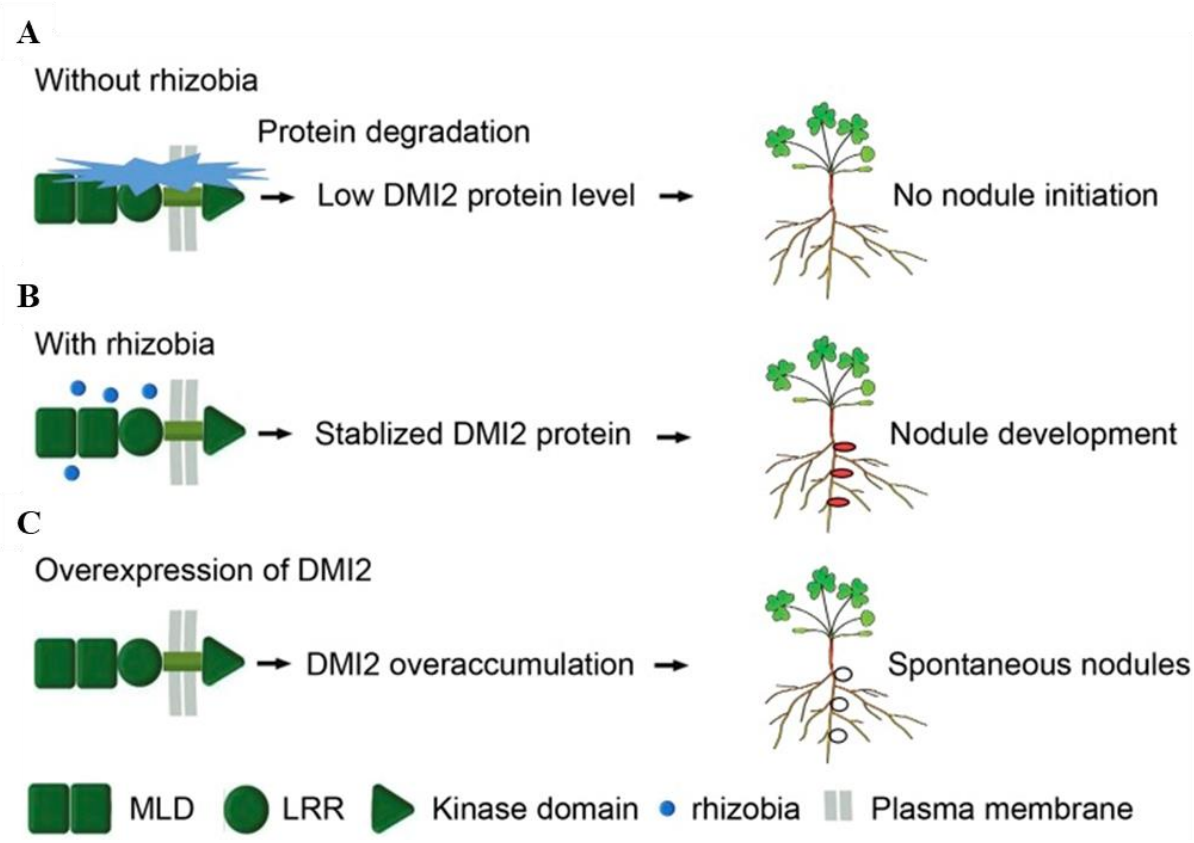


Figure 5.16. Plants control nodule development through fine regulation of DMI2 protein level. A. Without the presence of rhizobia, DMI2 protein is constitutively degraded by the proteasome apparatus and the nodule development signaling pathway is suppressed. **B.** Rhizobia inoculation blocks the degradation of DMI2 protein, and subsequently, stabilized DMI2 initiates nodule development signaling; thus, plants develop functional nodules. **C.** When full-length DMI2 protein or the kinase domain is overexpressed, overaccumulated DMI2 protein activates the nodule development signaling pathway in the absence of rhizobia, leading to a spontaneous nodulation phenotype. Red ovals in (**B**) indicate functional nodules, and the white ovals in (**C**) represent spontaneous nodules.

DMI2 protein level is at the center of regulating nodule development and rhizobial initiation

It is broadly known that, when growing in environments with abundant nitrogen, legume plants do not make nodules, despite the presence of rhizobia. Similarly, when nodulating plants are provided with bioavailable nitrogen, the plant will shut down the nodules immediately in order to save energy and resources. Hypernodulation mutants,

which generate many more nodules than the wild-type, are reported to have severe growth defects, suggesting that excessive nodules have negative effects on the fitness of plants (Schnabel et al., 2005). Understanding how plants maintain the ability to make the right number of nodules at the right time is a great interest in the research community. Our results show that the DMI2 protein accumulates during rhizobial inoculation. Taken together with reports that overexpressing the kinase domain or full-length DMI2/SYMRK results in spontaneous nodulation or a hypernodulation phenotype in legumes (Ried et al., 2014; Saha et al., 2014), we conclude that the protein level of DMI2/SYMRK is a master determinate signal of nodule development (Figure 5.16).

Our results show that the induction of DMI2 protein is independent of Nod factor reception (Figure 5.1C), pointing to the possibility that there is another hidden signaling pathway used by rhizobia to regulate DMI2 protein levels, likely through blocking protein degradation. It has been reported that DMI2/SYMRK interacts with several E3 ligases in *L. japonicus*, like SINA4 and SIE3 (Den Herder et al., 2012; Yuan et al., 2012). As these reports have shown that SINA4 and SIE3 could ubiquitinate DMI2/SYMRK, our results provide the necessary evidence that the protein level of DMI2/SYMRK is altered during rhizobia infection in a proteasome-dependent manner *in vivo*. To truly confirm that SINA4 and SIE3 are responsible for DMI2/SYMRK protein turnover, expressing DMI2/SYMRK in *sina4* and *sie3* mutants to investigate DMI2/SYMRK protein dynamics is necessary.

Since DMI2 also is required for the symbiosis between plants and AM fungi and Frankia bacteria, it will be interesting to find out whether the protein level of DMI2 is induced by AM treatment. We would predict that this is the case because we found that

the induction of DMI2 protein by rhizobia is independent of Nod factor perception (Figure 5.1C). Furthermore, any differences in DMI2 protein dynamics between rhizobia and AM infection might be related to how plants distinguish these two different symbionts, even though they utilize a shared symbiosis signaling pathway.

MLD is required for the proper folding of DMI2 protein

The extracellular part of DMI2 contains an MLD domain and an LRR. Considering that the MLDs reside at the extracellular domain of RLKs in plants, it is speculated that it may directly recognize some microbial signals or signals generated from the plant-microbe interaction and, subsequently, control the proper activation of the kinase domain. We found that the MLD is at least required for the proper folding of DMI2 protein because, after we introduced amino acid substitutions into the MLD, plants could not generate full-length DMI2 protein even in the presence of the proteasome inhibitor MG132 or rhizobia inoculation (Figure 5.12C and D). Human Malectin protein is reported to function in the ER (Schallus et al., 2008). Through its affinity to carbohydrate molecules, human Malectin protein can bind to misfolded proteins and activate ER quality control signaling (Qin et al., 2012). From our results, we speculate that DMI2-MLD also may function in the ER quality control process to guide the proper folding and successful secretion of the protein of which MLD itself is a part. The exact ligand for DMI2-MLD has yet to be found, but if this hypothesis is correct, the likely scenario is that DMI2-MLD could bind to some carbohydrate ligands similar to diglucosylated Glc2Man9GlcNAc2 during protein folding in the ER lumen (Schallus et

al., 2008). A carbohydrate microarray will be required to find the exact ligand of DMI2-MLD.

By showing the protein-level dynamics of DMI2 during rhizobia inoculation and the altered protein behavior of DMI2 with amino acid substitutions in the MLD, we show the probable function of MLD and present a previously unidentified signaling pathway of how rhizobia affect DMI2 protein levels.

Methods

Plant growth and inoculation

Medicago truncatula ecotype A17 was used for all the experiments in this study. Plants were grown under 16h of light/ 8h of dark at 22°C. *Agrobacterium rhizogenes* strain Arqua1 was used for plant hairy root transformation, and the procedure was conducted as described previously (Boisson-Dernier et al., 2001). Transgenic roots were selected based on antibiotic resistance against kanamycin at the concentration of 15µg mL^{-1} in Fahraeus medium for 10d.

Sinorhizobium meliloti strain ABS7 *hemA:LacZ* was used for rhizobia inoculation. Fresh overnight rhizobia culture was centrifuged and suspended in one-half-strength basic nodulation medium to a concentration of $O.D._{600}$ = 0.05. Five milliliters of liquid rhizobia culture were used per plant. Nodulation and infection thread phenotypes were checked 14 and 7d post inoculation, respectively; for protein sample collection, tissues were collected at different time points, as indicated in the figure legends.

Evaluation of symbiotic phenotypes

To analyze the nodule development phenotype, 14d after ABS7 *hemA:LacZ* inoculation, plants were harvested and total and pink nodules were counted. For infection thread analysis, 7d post ABS7 *hemA:LacZ* inoculation, Medicago roots were stained for β-galactosidase activity as described previously (Pan, Oztas et al., 2016). Root samples were filled with staining buffer (0.5M sodium phosphate buffer, pH 7.2, 10% (v/v) Triton X-100, 100mM potassium ferrocyannide, and 100mM potassium ferricyanide) containing 2mM 5-bromo-4-chloro-3-indolyl-β-glucuronic acid, then incubated at 37°C overnight. Infection threads were checked, and photographs were taken using a Nikon E200 microscope. The number of infection threads was counted on at least 10 individual transformed plants for each construct.

Plasmids and vectors

pKGW-RR::*gDMI2-HAST* was used to express wild-type gDMI2-HAST in plants stably and transiently (Riely et al., 2013). To make point mutations in the MLD, an overlapping PCR method was used to generate desired amino acid substitutions in pENTER::*gDMI2-HAST* (Heckman and Pease, 2007). After sequencing to confirm that the sequences were correct, the *gDMI2-HAST* sequences containing various point mutations in the MLD were introduced into the pKGW-RR vector using an LR recombination kit (Invitrogen). Then, the plasmids containing the point mutations were transformed into *A. rhizogenes* strain Arqua1 by electroporation. Primers used in this chapter can be found in Table 5.2.

Protein extraction and immunoblot

To extract proteins from uninoculated roots, inoculated roots, and nodules of Medicago plants, plant tissues were collected at the time points indicated in the figure legends and put into liquid nitrogen immediately. After grinding the tissues into a fine powder, proteins were extracted using native extraction buffer 1 (50mM Tris-MES, pH 8, 0.5M Suc, 1mM MgCl2, 10mM EDTA, and 5mM DTT) with or without the protease inhibitor cocktail (Sigma-Aldrich; Liu et al., 2010). Extracted proteins were boiled for 5 min with 6×SDS sampling buffer and stored at −20°C.

For immunoblots, various protein samples were loaded onto a 12% (w/v) acrylamide SDS running gel, and the protein samples were transferred from the gel to nitrocellulose membranes (Adventec) by electroblotting. After blocking the membrane with milk, the membrane was subjected to incubation with 1:500 diluted anti-HA antibody (New England Biolabs) and 1:4,000 diluted anti-α-tubulin antibody (Life Sciences). The secondary antibodies were anti-rat and anti-mouse, respectively (Life Sciences). As the secondary antibodies were conjugated with horseradish peroxidase, the membrane was treated with the detection reagent from Thermo Scientific (products 1859707 and 1859678). The bands on the membrane were visualized via a G-box machine (New England Biolabs).

Cell-free degradation assay and MG132 treatment

Cell-free degradation assays were performed following previous reports (Spoel et al., 2009). Extracted protein samples were split into individual centrifuge tubes using

equal amounts, different reagents were added into the tubes as indicated (MG132 concentration, 40μM mL^{-1}), and the tubes were shaken slowly at room temperature for 2h. Individual samples were boiled with 6×SDS sampling buffer and subjected to immunoblot with anti-HA and anti-tubulin antibodies.

For MG132 in vivo treatments, 2-week-old Medicago plants were washed clean and put into distilled water containing 100μM mL−1 MG132 at 22°C. Root samples were collected at different time points during MG132 treatment, as indicated.

RNA extraction and RT-qPCR

Total RNA extraction of rhizobia-inoculated or MG132-treated Medicago roots was carried out using TRIzol (Invitrogen) following the manufacturer's instructions. To eliminate possible DNA contamination, extracted RNA was treated with the Turbo DNA-free kit (Life Science Technologies) using the manufacturer's instructions. The iScript cDNA synthesis kit (Bio-Rad) was used for second chain synthesis following the manufacturer's instructions step by step.

RT-qPCR was performed on the Eppendorf Mastercycler ep Realplex system. PCR was conducted using ExTaq polymerase (Takara) with SYBR Green dye (Thermo Fisher Scientific) in 20μL total volume: distilled, deionized water, 9.9μL; 10×ExTaq buffer, 2μL; 2.5mM deoxyribonucleotide triphosphate mix, 2μL; 10μm primer mix, 2μL; SYBR Green dye, 2μL; template cDNA, 2μL; ExTaq polymerase, 0.1μL. For real-time PCR, annealing temperature was set at 60°C and elongation time was 30s. When the PCR was finished, the melting curve was analyzed to rule out possible nonspecific amplification, and the result represented was the mean threshold cycle value of three

technical replicates. Three biological replicates were used for each sample. *PROTODERMAL FACTOR2* (locus identifier *Medtr6g084690*) was used as the internal control.

Protein domain analysis and sequence alignments

The gene expression analysis using existing databases was divided into two groups: for microarray analysis, the data were obtained from the *Medicago truncatula* Gene Expression Atlas (http://mtgea.noble.org/v3/; Benedito et al., 2008); for RNA sequencing analysis, the *Medicago truncatula* Genome Project version 4.0 (http://medicago.jcvi.org/medicago/index.php) and the Symbimics website (https://iant.toulouse.inra.fr/symbimics/) were used (Roux et al., 2014; Young et al., 2011).

For protein structure prediction and conserved residue analysis, the SWISS-MODEL database (http://swissmodel.expasy.org/) was used, and the DMI2-MLD domain structure was predicted based on the structure of the human Malectin A domain. The positions of conserved amino acids were mapped to the predicted structure in the same program (Biasini et al., 2014).

The phylogenic tree was built using the MEGA 7.0 program under the developer's instruction.

Table 5.2 Primers used in this chapter.

Primers for *gDMI2HAST* real-time PCR	
forward primer	ACTCAATTGTTATGGACAAACGG
reverse primer	ACTTTTCGAATTGTGGATGAGAC
Primers for *NIN* real-time PCR	
forward primer	CCCACAAGCAGTAAAGGTATTTG
reverse primer	CTCCTCCTTGTTGTTGTTGTAGG
Primers for *ENOD11* real-time PCR	
forward primer	TAAATCTCCATCCCACAATATGC
reverse primer	ATTGGAGGCTTGTAAGTAGGAGG
Primers for *PDF2* real-time PCR	
forward primer	GTGTTTTGCTTCCGCCGTT
reverse primer	CCAAATCTTGCTCCCTCATCTG
Primers for *DMI2* real-time PCR	
forward primer	CAGTTATGGGAGATGCAAAGC
reverse primer	AATGCCTACTTGTTGCCATTG
Primer sequences for making point mutations	
$DMI2^{C39D}$	
forward primer	CATAGCAGATTGTGCTGATTCAAATTAC
reverse primer	GCACAATCTG CTATGCTCTC AAACCCTACA AGGA
$DMI2^{Y95A}$	
forward primer	GAGATGTGCTAATTTGCCAACAGTTAAGG A
reverse primer	CAAATTAGCACATCTCTTTCCTTCATATATTT
$DMI2^{F111A}$	
forward primer	GGCATAGCTCCCTTTGATAGTTTAAATTC
reverse primer	AAGGGAGCTATGCCCCTTATCAAATATACTTG
$DMI2^{P164A}$	
forward primer	AATGCCTTCATTTCTCAGATTGAATTG
reverse primer	AGGCATTGACATCCTCCTTCAATA
$DMI2^{Y271A}$	
forward primer	TTATGAAGCCTCTGTGTTTCTCCACTTTC
reverse primer	CACAGAGGCTTCATAATCATCGGTCTCGAGG
$DMI2^{F291A}$	
forward primer	AAGGGTGGCTGACATCTATCTAAACAATG
reverse primer	ATGTCAGCCACCCTTTGTCCTGCTCTGACAGTGC
$DMI2^{Y294A}$	
forward primer	GACATCGCTCTAAACAATGAGATTAA
reverse primer	TTAGAGCGATGTCAAACACCCTTTGTCCTGCTC

CHAPTER 6

CONCLUSIONS/ FUTURE DIRECTIONS

I am interested in how legumes, specifically Medicago, have co-opted genes to function specifically in the nitrogen-fixing (NF) symbiosis. Nodules are specialized organs that form *de novo* following rhizobia infection. Many genes are specific and important for nodule formation, function, and for maintaining the symbiosis (reviewed by Roy et al., 2019). It has been shown that many genes that are required for nitrogen-fixing symbiosis were co-opted from arbuscular mycorrhizal (AM) symbiosis (Oldroyd, 2013). Many of these genes are part of the common symbiosis signaling pathway and are involved in the early steps of the symbiosis. From an evolutionary perspective, this makes sense because both symbioses involve the symbionts signaling to the plant, entering root cells enclosed in a plant-derived membrane, and exchanging nutrients across that membrane. It has recently been shown that all intracellular symbiotic pathways have evolved from the more ancient AM pathway (Radhakrishnan et al., 2020). However, the later stages of AM and NF symbioses are different, and this requires the activation of different genes.

In chapter 2, the signal peptidase complex (SPC) was looked at in-depth for its role in NF symbiosis. The question is important because components of the SPC are extremely well-conserved even among disparate species. In Medicago, there is a dedicated SPC22 subunit (DNF1) that functions in the nodule and is not induced in any other organ (Wang et al., 2010). It would be extremely surprising if DNF1 performs a novel function in the nodule that the housekeeping version cannot. We found that, indeed,

the housekeeping SPC22 can perform the role of the nodule-specific version. Furthermore, we found that the expression pattern of *DNF1* confers its nodule specificity.

The same case is true for the catalytic SPC subunit, SPC18. There are two versions in Medicago: one presumably housekeeping version and one symbiotic version, as we have shown. Knocking out the symbiotic version causes a Fix- phenotype very similar to *dnf1*. This nodule-specific SPC is required for bacteroid differentiation and is involved in delivering NCR peptides to the intracellular bacteria (Van de Velde et al., 2010; Wang et al., 2010). Only some clades of legumes use peptides, like NCRs, to force the intracellular bacteria to differentiate. Therefore, only some legume species may require a dedicated nodule SPC. There is at least one example of convergent evolution of a symbiosis-dedicated SPC22 in a species that has independently evolved peptides that are used for bacteroid differentiation (Czernic et al., 2015).

The information gleaned from chapter 2 seems to indicate that the nodule-specific SPC complex has been co-opted from the housekeeping version and fundamentally performs the same function. This dedicated protein secretory pathway in species with NCR or NCR-like peptides may be necessary to accommodate the sheer number of peptides that need to be processed and delivered to the intracellular bacteria. It remains to be seen if other legume species have independently evolved a dedicated SPC to serve the same purpose in nodule cells.

In chapter 3, we found that SYP132A, a t-SNARE that is localized to the symbiotic membrane, is required for successful nitrogen fixation in Medicago. SYP132A is part of the protein secretory pathway, and it is also necessary for successful AM symbiosis. Due to its role in AM symbiosis and the presence of orthologs in many land

plant species, we conclude that it was co-opted for its role in NF symbiosis from the AM symbiosis.

Interestingly, the expression of *SYP132A* and *DNF1* are highly correlated (Wang et al., 2010), but they are not regulated in the same way. In chapter 2, we showed that *DNF1* and *SPC18* have *cis*-elements in their promoters (and downstream of the coding region) that confer nodule-specific expression. However, the *SYP132A* transcript is created by alternative cleavage and polyadenylation. One question, then, is what factor is responsible for the generation of this alternative transcript, and how is its action regulated?

Another major question involves how this SYP132A is specifically embedded in the symbiotic membrane, while the canonical version, SYP132C, is found on the plasma membrane. The transmembrane domains of these proteins are different, and presumably, these differences are responsible for their distinct locations and functions, but the mechanisms involved are unknown.

As mentioned above, NCR peptides are delivered to the intracellular bacteria through the protein secretion pathway in the nodule. In the IRLC, there are considerable differences in the number of NCRs encoded in the genomes of different species (Montiel et al., 2017). Medicago has about 700 NCR peptides, and as a group, they are involved in bacteroid differentiation, and many are thought to play a role in symbiont selection (Van de Velde et al., 2010; Wang et al., 2017; Yang et al., 2017). It has been challenging to study this group of peptides because they are thought to be incredibly diverse. In chapter 4, we describe two sequence motifs that are common in many NCR peptides in Medicago and other IRLC species. This is a step towards understanding how these peptides may

function as a group and what properties are essential for their function. One important question that we are currently investigating is if the presence of these motifs influences the survival or differentiation of the intracellular rhizobia. We are doing this by expressing a mutant version of NCR371 in Medicago hairy roots, which contains both of the motifs that we have identified. This mutant NCR371 was shown to have a greater antimicrobial effect on free-living rhizobia, and expressing the mutant and wild-type versions *in planta* will shed light on the effects of these motifs on the symbiosis. Several other questions remain surrounding these NCR sequence motifs, specifically, their mode of action on free-living bacteria.

Finally, in chapter 5, we show that DMI2, which is required for NF symbiosis, is regulated at the protein level. DMI2 is constitutively expressed and translated, but it is degraded in the absence of rhizobia. Once the rhizobia are present, DMI2 is protected from degradation. It seems that, unlike components that are involved in later stages of NF symbiosis (such as genes in the protein secretory pathway), DMI2 needs to be constitutively expressed and regulated at the protein level in order to participate in the early steps of the signaling pathway.

Interestingly, the protection of DMI2 from degradation is Nod-factor independent. DMI2 is also necessary for AM and Frankia symbioses, which begs the question of what is triggering the accumulation of DMI2, and it is the same in the other symbioses?

In general, I am interested in the genetic and molecular innovations that have given rise to the nitrogen-fixing symbiosis in Medicago. While many of these genes are shared between the AM and NF pathways such as *SYP132A* and *DMI2*, some are unique to NF symbiosis and even particular legume clades (i.e., symbiotic SPC genes and NCR

peptides). A deeper understanding of these innovations will help shed light on what is required and indispensable for successful NF symbiosis.

BIBLIOGRAPHY

Alkhalfioui, F., Renard, M., Frendo, P., Keichinger, C., Meyer, Y., Gelhaye, E., Hirasawa, M., Knaff, D.B., Ritzenthaler, C., Montrichard, F., 2008. A Novel Type of Thioredoxin Dedicated to Symbiosis in Legumes. Plant Physiol 148, 424–435. https://doi.org/10.1104/pp.108.123778

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of Molecular Biology 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Antolín-Llovera, M., Petutsching, E.K., Ried, M.K., Lipka, V., Nürnberger, T., Robatzek, S., Parniske, M., 2014a. Knowing your friends and foes--plant receptor-like kinases as initiators of symbiosis or defence. New Phytol. 204, 791–802. https://doi.org/10.1111/nph.13117

Antolín-Llovera, M., Ried, M.K., Parniske, M., 2014b. Cleavage of the SYMBIOSIS RECEPTOR-LIKE KINASE ectodomain promotes complex formation with Nod factor receptor 5. Curr. Biol. 24, 422–427. https://doi.org/10.1016/j.cub.2013.12.053

Assaad, F.F., Qiu, J.-L., Youngs, H., Ehrhardt, D., Zimmerli, L., Kalde, M., Wanner, G., Peck, S.C., Edwards, H., Ramonell, K., Somerville, C.R., Thordal-Christensen, H., 2004. The PEN1 Syntaxin Defines a Novel Cellular Compartment upon Fungal Attack and Is Required for the Timely Assembly of Papillae. MBoC 15, 5118–5129. https://doi.org/10.1091/mbc.e04-02-0140

Benedito, V.A., Torres-Jerez, I., Murray, J.D., Andriankaja, A., Allen, S., Kakar, K., Wandrey, M., Verdier, J., Zuber, H., Ott, T., Moreau, S., Niebel, A., Frickey, T., Weiller, G., He, J., Dai, X., Zhao, P.X., Tang, Y., Udvardi, M.K., 2008. A gene expression atlas of the model legume Medicago truncatula. Plant J. 55, 504–513. https://doi.org/10.1111/j.1365-313X.2008.03519.x

Berrabah, F., Bourcy, M., Eschstruth, A., Cayrel, A., Guefrachi, I., Mergaert, P., Wen, J., Jean, V., Mysore, K.S., Gourion, B., Ratet, P., 2014. A nonRD receptor-like kinase prevents nodule early senescence and defense-like reactions during symbiosis. New Phytologist 203, 1305–1314. https://doi.org/10.1111/nph.12881

Bersoult, A., Camut, S., Perhald, A., Kereszt, A., Kiss, G.B., Cullimore, J.V., 2005. Expression of the Medicago truncatula DM12 gene suggests roles of the symbiotic nodulation receptor kinase in nodules and during early nodule development. Mol. Plant Microbe Interact. 18, 869–876. https://doi.org/10.1094/MPMI-18-0869

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., Schwede, T., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary

information. Nucleic Acids Res. 42, W252-258. https://doi.org/10.1093/nar/gku340

Blatt, M.R., Leyman, B., Geelen, D., 1999. Tansley Review No. 108 Molecular events of vesicle trafficking and control by SNARE proteins in plants. The New Phytologist 144, 389–418. https://doi.org/10.1046/j.1469-8137.1999.00546.x

Boisson-Dernier, A., Chabaud, M., Garcia, F., Bécard, G., Rosenberg, C., Barker, D.G., 2001. Agrobacterium rhizogenes-transformed roots of Medicago truncatula for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. Mol. Plant Microbe Interact. 14, 695–700. https://doi.org/10.1094/MPMI.2001.14.6.695

Boisson-Dernier, Aurélien, Chabaud, M., Garcia, F., Bécard, G., Rosenberg, C., Barker, D.G., 2001. Agrobacterium rhizogenes-Transformed Roots of Medicago truncatula for the Study of Nitrogen-Fixing and Endomycorrhizal Symbiotic Associations. MPMI 14, 695–700. https://doi.org/10.1094/MPMI.2001.14.6.695

Bourcy, M., Brocard, L., Pislariu, C.I., Cosson, V., Mergaert, P., Tadege, M., Mysore, K.S., Udvardi, M.K., Gourion, B., Ratet, P., 2013. Medicago truncatula DNF2 is a PI-PLC-XD-containing protein required for bacteroid persistence and prevention of nodule early senescence and defense-like reactions. New Phytologist 197, 1250–1261. https://doi.org/10.1111/nph.12091

Brewin, N.J., 2004. Plant Cell Wall Remodelling in the Rhizobium–Legume Symbiosis. Critical Reviews in Plant Sciences 23, 293–316. https://doi.org/10.1080/07352680490480734

Cai, Y., Chen, L., Liu, X., Sun, S., Wu, C., Jiang, B., Han, T., Hou, W., 2015. CRISPR/Cas9-Mediated Genome Editing in Soybean Hairy Roots. PLOS ONE 10, e0136064. https://doi.org/10.1371/journal.pone.0136064

Carvalho, A. de O., Gomes, V.M., 2009. Plant defensins--prospects for the biological functions and biotechnological properties. Peptides 30, 1007–1020. https://doi.org/10.1016/j.peptides.2009.01.018

Catalano, C.M., Czymmek, K.J., Gann, J.G., Sherrier, D.J., 2007. Medicago truncatula syntaxin SYP132 defines the symbiosome membrane and infection droplet membrane in root nodules. Planta 225, 541–550. https://doi.org/10.1007/s00425-006-0369-y

Catalano, C.M., Lane, W.S., Sherrier, D.J., 2004. Biochemical characterization of symbiosome membrane proteins from Medicago truncatula root nodules. Electrophoresis 25, 519–531. https://doi.org/10.1002/elps.200305711

Čermák, T., Curtin, S.J., Gil-Humanes, J., Čegan, R., Kono, T.J.Y., Konečná, E., Belanto, J.J., Starker, C.G., Mathre, J.W., Greenstein, R.L., Voytas, D.F., 2017. A

Multipurpose Toolkit to Enable Advanced Genome Engineering in Plants. Plant Cell 29, 1196–1217. https://doi.org/10.1105/tpc.16.00922

Clarke, J.D., 2009. Cetyltrimethyl Ammonium Bromide (CTAB) DNA Miniprep for Plant DNA Isolation. Cold Spring Harb Protoc 2009, pdb.prot5177. https://doi.org/10.1101/pdb.prot5177

Collins, N.C., Thordal-Christensen, H., Lipka, V., Bau, S., Kombrink, E., Qiu, J.-L., Hückelhoven, R., Stein, M., Freialdenhoven, A., Somerville, S.C., Schulze-Lefert, P., 2003. SNARE-protein-mediated disease resistance at the plant cell wall. Nature 425, 973–977. https://doi.org/10.1038/nature02076

Curtis, M.D., Grossniklaus, U., 2003. A Gateway Cloning Vector Set for High-Throughput Functional Analysis of Genes in Planta. Plant Physiology 133, 462–469. https://doi.org/10.1104/pp.103.027979

Czernic, P., Gully, D., Cartieaux, F., Moulin, L., Guefrachi, I., Patrel, D., Pierre, O., Fardoux, J., Chaintreuil, C., Nguyen, P., Gressent, F., Da Silva, C., Poulain, J., Wincker, P., Rofidal, V., Hem, S., Barrière, Q., Arrighi, J.-F., Mergaert, P., Giraud, E., 2015. Convergent Evolution of Endosymbiont Differentiation in Dalbergioid and Inverted Repeat-Lacking Clade Legumes Mediated by Nodule-Specific Cysteine-Rich Peptides. Plant Physiol. 169, 1254–1265. https://doi.org/10.1104/pp.15.00584

Den Herder, G., Yoshida, S., Antolín-Llovera, M., Ried, M.K., Parniske, M., 2012. Lotus japonicus E3 ligase SEVEN IN ABSENTIA4 destabilizes the symbiosis receptor-like kinase SYMRK and negatively regulates rhizobial infection. Plant Cell 24, 1691–1707. https://doi.org/10.1105/tpc.110.082248

Durgo, H., Klement, E., Hunyadi-Gulyas, E., Szucs, A., Kereszt, A., Medzihradszky, K.F., Kondorosi, E., 2015. Identification of nodule-specific cysteine-rich plant peptides in endosymbiotic bacteria. PROTEOMICS 15, 2291–2295. https://doi.org/10.1002/pmic.201400385

Endre, G., Kereszt, A., Kevei, Z., Mihacea, S., Kaló, P., Kiss, G.B., 2002. A receptor kinase gene regulating symbiotic nodule development. Nature 417, 962–966. https://doi.org/10.1038/nature00842

Farkas, A., Maróti, G., Dürgő, H., Györgypál, Z., Lima, R.M., Medzihradszky, K.F., Kereszt, A., Mergaert, P., Kondorosi, É., 2014. Medicago truncatula symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms. PNAS 111, 5183–5188. https://doi.org/10.1073/pnas.1404169111

Farkas, A., Maróti, G., Kereszt, A., Kondorosi, É., 2017. Comparative Analysis of the Bacterial Membrane Disruption Effect of Two Natural Plant Antimicrobial Peptides. Front Microbiol 8, 51. https://doi.org/10.3389/fmicb.2017.00051

Franssen, H.J., Vijn, I., Yang, W.C., Bisseling, T., 1992. Developmental aspects of the Rhizobium-legume symbiosis. Plant Mol Biol 19, 89–107. https://doi.org/10.1007/BF00015608

Ganz, T., 2003. Defensins: antimicrobial peptides of innate immunity. Nat. Rev. Immunol. 3, 710–720. https://doi.org/10.1038/nri1180

Gavrin, A., Chiasson, D., Ovchinnikova, E., Kaiser, B.N., Bisseling, T., Fedorova, E.E., 2016. VAMP721a and VAMP721d are important for pectin dynamics and release of bacteria in soybean nodules. New Phytol. 210, 1011–1021. https://doi.org/10.1111/nph.13837

Gavrin, A., Kulikova, O., Bisseling, T., Fedorova, E.E., 2017. Interface Symbiotic Membrane Formation in Root Nodules of Medicago truncatula: the Role of Synaptotagmins MtSyt1, MtSyt2 and MtSyt3. Front Plant Sci 8, 201. https://doi.org/10.3389/fpls.2017.00201

Gherbi, H., Markmann, K., Svistoonoff, S., Estevan, J., Autran, D., Giczey, G., Auguy, F., Péret, B., Laplaze, L., Franche, C., Parniske, M., Bogusz, D., 2008. SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and Frankiabacteria. Proc. Natl. Acad. Sci. U.S.A. 105, 4928–4932. https://doi.org/10.1073/pnas.0710618105

Giovannetti, M., Mosse, B., 1980. An Evaluation of Techniques for Measuring Vesicular Arbuscular Mycorrhizal Infection in Roots. New Phytologist 84, 489–500. https://doi.org/10.1111/j.1469-8137.1980.tb04556.x

Gómez-Gómez, L., Boller, T., 2000. FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. Mol. Cell 5, 1003–1011. https://doi.org/10.1016/s1097-2765(00)80265-8

Graham, M.A., Silverstein, K.A.T., Cannon, S.B., VandenBosch, K.A., 2004. Computational identification and characterization of novel genes from legumes. Plant Physiol. 135, 1179–1197. https://doi.org/10.1104/pp.104.037531

Guefrachi, I., Nagymihaly, M., Pislariu, C.I., Van de Velde, W., Ratet, P., Mars, M., Udvardi, M.K., Kondorosi, E., Mergaert, P., Alunni, B., 2014. Extreme specificity of NCR gene expression in Medicago truncatula. BMC Genomics 15. https://doi.org/10.1186/1471-2164-15-712

Haag, A.F., Baloban, M., Sani, M., Kerscher, B., Pierre, O., Farkas, A., Longhi, R., Boncompagni, E., Hérouart, D., Dall'Angelo, S., Kondorosi, E., Zanda, M., Mergaert, P., Ferguson, G.P., 2011. Protection of Sinorhizobium against Host Cysteine-Rich Antimicrobial Peptides Is Critical for Symbiosis. PLOS Biology 9, e1001169. https://doi.org/10.1371/journal.pbio.1001169

Haag, A.F., Kerscher, B., Dall'Angelo, S., Sani, M., Longhi, R., Baloban, M., Wilson, H.M., Mergaert, P., Zanda, M., Ferguson, G.P., 2012. Role of cysteine residues

and disulfide bonds in the activity of a legume root nodule-specific, cysteine-rich peptide. J. Biol. Chem. 287, 10791–10798. https://doi.org/10.1074/jbc.M111.311316

Haruta, M., Sabat, G., Stecker, K., Minkoff, B.B., Sussman, M.R., 2014. A peptide hormone and its receptor protein kinase regulate plant cell expansion. Science 343, 408–411. https://doi.org/10.1126/science.1244454

Heckman, K.L., Pease, L.R., 2007. Gene splicing and mutagenesis by PCR-driven overlap extension. Nat Protoc 2, 924–932. https://doi.org/10.1038/nprot.2007.132

Helliwell, C.A., Waterhouse, P.M., 2005. Constructs and Methods for Hairpin RNA-Mediated Gene Silencing in Plants, in: Methods in Enzymology, RNA Interference. Academic Press, pp. 24–35. https://doi.org/10.1016/S0076-6879(04)92002-2

Higo, K., Ugawa, Y., Iwamoto, M., Korenaga, T., 1999. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res. 27, 297–300. https://doi.org/10.1093/nar/27.1.297

Hohnjec, N., Lenz, F., Fehlberg, V., Vieweg, M.F., Baier, M.C., Hause, B., Küster, H., 2009a. The Signal Peptide of the *Medicago truncatula* Modular Nodulin MtNOD25 Operates as an Address Label for the Specific Targeting of Proteins to Nitrogen-Fixing Symbiosomes. MPMI 22, 63–72. https://doi.org/10.1094/MPMI-22-1-0063

Hohnjec, N., Lenz, F., Fehlberg, V., Vieweg, M.F., Baier, M.C., Hause, B., Küster, H., 2009b. The signal peptide of the Medicago truncatula modular nodulin MtNOD25 operates as an address label for the specific targeting of proteins to nitrogen-fixing symbiosomes. Mol. Plant Microbe Interact. 22, 63–72. https://doi.org/10.1094/MPMI-22-1-0063

Hok, S., Danchin, E.G.J., Allasia, V., Panabières, F., Attard, A., Keller, H., 2011. An Arabidopsis (malectin-like) leucine-rich repeat receptor-like kinase contributes to downy mildew disease. Plant Cell Environ. 34, 1944–1957. https://doi.org/10.1111/j.1365-3040.2011.02390.x

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. Proc. Natl. Acad. Sci. U.S.A. 112, 15232–15237. https://doi.org/10.1073/pnas.1500777112

Huisman, R., Hontelez, J., Mysore, K.S., Wen, J., Bisseling, T., Limpens, E., 2016. A symbiosis-dedicated SYNTAXIN OF PLANTS 13II isoform controls the formation of a stable host–microbe interface in symbiosis. New Phytologist 211, 1338–1351. https://doi.org/10.1111/nph.13973

Ivanov, S., Fedorova, E.E., Limpens, E., De Mita, S., Genre, A., Bonfante, P., Bisseling, T., 2012. Rhizobium-legume symbiosis shares an exocytotic pathway required for arbuscule formation. Proc. Natl. Acad. Sci. U.S.A. 109, 8316–8321. https://doi.org/10.1073/pnas.1200407109

Jones, J.D.G., Dangl, J.L., 2006. The plant immune system. Nature 444, 323–329. https://doi.org/10.1038/nature05286

Kakar, K., Wandrey, M., Czechowski, T., Gaertner, T., Scheible, W.-R., Stitt, M., Torres-Jerez, I., Xiao, Y., Redman, J.C., Wu, H.C., Cheung, F., Town, C.D., Udvardi, M.K., 2008. A community resource for high-throughput quantitative RT-PCR analysis of transcription factor gene expression in Medicago truncatula. Plant Methods 4, 18. https://doi.org/10.1186/1746-4811-4-18

Kalde, M., Nühse, T.S., Findlay, K., Peck, S.C., 2007. The syntaxin SYP132 contributes to plant resistance against bacteria and secretion of pathogenesis-related protein 1. PNAS 104, 11850–11855. https://doi.org/10.1073/pnas.0701083104

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. Proc. Natl. Acad. Sci. U.S.A. 112, 15238–15243. https://doi.org/10.1073/pnas.1500123112

Kondorosi, E., Mergaert, P., Kereszt, A., 2013. A Paradigm for Endosymbiotic Life: Cell Differentiation of Rhizobium Bacteria Provoked by Host Plant Factors. Annu. Rev. Microbiol. 67, 611–628. https://doi.org/10.1146/annurev-micro-092412-155630

Konishi, M., Yanagisawa, S., 2013. Arabidopsis NIN-like transcription factors have a central role in nitrate signalling. Nature Communications 4, 1–9. https://doi.org/10.1038/ncomms2621

Krishnakumar, V., Kim, M., Rosen, B.D., Karamycheva, S., Bidwell, S.L., Tang, H., Town, C.D., 2015. MTGD: The Medicago truncatula genome database. Plant Cell Physiol. 56, e1. https://doi.org/10.1093/pcp/pcu179

Kwon, C., Neu, C., Pajonk, S., Yun, H.S., Lipka, U., Humphry, M., Bau, S., Straus, M., Kwaaitaal, M., Rampelt, H., Kasmi, F.E., Jürgens, G., Parker, J., Panstruga, R., Lipka, V., Schulze-Lefert, P., 2008. Co-option of a default secretory pathway for plant immune responses. Nature 451, 835–840. https://doi.org/10.1038/nature06545

Lavin, M., Doyle, J.J., Palmer, J.D., 1990. EVOLUTIONARY SIGNIFICANCE OF THE LOSS OF THE CHLOROPLAST-DNA INVERTED REPEAT IN THE LEGUMINOSAE SUBFAMILY PAPILIONOIDEAE. Evolution 44, 390–402. https://doi.org/10.1111/j.1558-5646.1990.tb05207.x

Lemmon, M.A., Schlessinger, J., 2010. Cell signaling by receptor tyrosine kinases. Cell 141, 1117–1134. https://doi.org/10.1016/j.cell.2010.06.011

Leong, S.A., Williams, P.H., Ditta, G.S., 1985. Analysis of the 5′ regulatory region of the gene for δ-aminolevulinic acid synthetase of Rhizobium meliloti. Nucleic Acids Res 13, 5965–5976. https://doi.org/10.1093/nar/13.16.5965

Li, X., Zheng, Z., Kong, X., Xu, J., Qiu, L., Sun, J., Reid, D.E., Jin, H., Andersen, S.U., Oldroyd, G.E.D., Stougaard, J., Downie, J.A., Xie, F., 2019. Atypical Receptor Kinase RINRK1 Required for Rhizobial Infection but not Nodule Development in Lotus japonicus. Plant Physiology pp.00509.2019. https://doi.org/10.1104/pp.19.00509

Limpens, E., Ivanov, S., Esse, W. van, Voets, G., Fedorova, E., Bisseling, T., 2009. Medicago N2-Fixing Symbiosomes Acquire the Endocytic Identity Marker Rab7 but Delay the Acquisition of Vacuolar Identity. The Plant Cell 21, 2811–2828. https://doi.org/10.1105/tpc.108.064410

Lipka, V., Kwon, C., Panstruga, R., 2007. SNARE-ware: the role of SNARE-domain proteins in plant biology. Annu. Rev. Cell Dev. Biol. 23, 147–174. https://doi.org/10.1146/annurev.cellbio.23.090506.123529

Liu, J., Rutten, L., Limpens, E., Molen, T. van der, Velzen, R. van, Chen, R., Chen, Y., Geurts, R., Kohlen, W., Kulikova, O., Bisseling, T., 2019. A Remote cis-Regulatory Region Is Required for NIN Expression in the Pericycle to Initiate Nodule Primordium Formation in Medicago truncatula. The Plant Cell 31, 68–83. https://doi.org/10.1105/tpc.18.00478

Liu, L., Zhang, Y., Tang, S., Zhao, Q., Zhang, Z., Zhang, H., Dong, L., Guo, H., Xie, Q., 2010. An efficient system to detect protein ubiquitination by agroinfiltration in Nicotiana benthamiana. Plant J. 61, 893–903. https://doi.org/10.1111/j.1365-313X.2009.04109.x

Lu, D., Lin, W., Gao, X., Wu, S., Cheng, C., Avila, J., Heese, A., Devarenne, T.P., He, P., Shan, L., 2011. Direct ubiquitination of pattern recognition receptor FLS2 attenuates plant innate immunity. Science 332, 1439–1442. https://doi.org/10.1126/science.1204903

Maróti, G., Downie, J.A., Kondorosi, É., 2015. Plant cysteine-rich peptides that inhibit pathogen growth and control rhizobial differentiation in legume nodules. Curr. Opin. Plant Biol. 26, 57–63. https://doi.org/10.1016/j.pbi.2015.05.031

Maunoury, N., Redondo-Nieto, M., Bourcy, M., Van de Velde, W., Alunni, B., Laporte, P., Durand, P., Agier, N., Marisa, L., Vaubert, D., Delacroix, H., Duc, G., Ratet, P., Aggerbeck, L., Kondorosi, E., Mergaert, P., 2010. Differentiation of symbiotic cells and endosymbionts in Medicago truncatula nodulation are coupled to two transcriptome-switches. PLoS ONE 5, e9519. https://doi.org/10.1371/journal.pone.0009519

Mergaert, P., 2018. Role of antimicrobial peptides in controlling symbiotic bacterial populations. Nat. Prod. Rep. 35, 336–356. https://doi.org/10.1039/C7NP00056A

Mergaert, P., Nikovics, K., Kelemen, Z., Maunoury, N., Vaubert, D., Kondorosi, A., Kondorosi, E., 2003. A Novel Family in Medicago truncatula Consisting of More Than 300 Nodule-Specific Genes Coding for Small, Secreted Polypeptides with Conserved Cysteine Motifs. Plant Physiol 132, 161–173. https://doi.org/10.1104/pp.102.018192

Mergaert, P., Uchiumi, T., Alunni, B., Evanno, G., Cheron, A., Catrice, O., Mausset, A.-E., Barloy-Hubler, F., Galibert, F., Kondorosi, A., Kondorosi, E., 2006a. Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium–legume symbiosis. PNAS 103, 5230–5235. https://doi.org/10.1073/pnas.0600912103

Mergaert, P., Uchiumi, T., Alunni, B., Evanno, G., Cheron, A., Catrice, O., Mausset, A.-E., Barloy-Hubler, F., Galibert, F., Kondorosi, A., Kondorosi, E., 2006b. Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium-legume symbiosis. Proc. Natl. Acad. Sci. U.S.A. 103, 5230–5235. https://doi.org/10.1073/pnas.0600912103

Metz, B.A., Welters, P., Hoffmann, H.J., Jensen, E.O., Schell, J., de Bruijn, F.J., 1988. Primary structure and promoter analysis of leghemoglobin genes of the stem-nodulated tropical legume Sesbania rostrata: conserved coding sequences, cis-elements and trans-acting factors. Mol. Gen. Genet. 214, 181–191. https://doi.org/10.1007/bf00337709

Meyer, D., Pajonk, S., Micali, C., O'Connell, R., Schulze-Lefert, P., 2009. Extracellular transport and integration of plant secretory proteins into pathogen-induced cell wall compartments. The Plant Journal 57, 986–999. https://doi.org/10.1111/j.1365-313X.2008.03743.x

Mikuláss, K.R., Nagy, K., Bogos, B., Szegletes, Z., Kovács, E., Farkas, A., Váró, G., Kondorosi, É., Kereszt, A., 2016. Antimicrobial nodule-specific cysteine-rich peptides disturb the integrity of bacterial outer and inner membranes and cause loss of membrane potential. Ann Clin Microbiol Antimicrob 15. https://doi.org/10.1186/s12941-016-0159-8

Mirabella, R., Hartog, M., Franken, C., Geurts, R., Bisseling, T., 2005. Expression Pattern of DMI Genes in Medicago Nodules, in: Wang, Y.-P., Lin, M., Tian, Z.-X., Elmerich, C., Newton, W.E. (Eds.), Biological Nitrogen Fixation, Sustainable Agriculture and the Environment, Current Plant Science and Biotechnology in Agriculture. Springer Netherlands, Dordrecht, pp. 153–155. https://doi.org/10.1007/1-4020-3570-5_36

Montiel, J., Downie, J.A., Farkas, A., Bihari, P., Herczeg, R., Bálint, B., Mergaert, P., Kereszt, A., Kondorosi, É., 2017. Morphotype of bacteroids in different legumes

correlates with the number and type of symbiotic NCR peptides. Proc. Natl. Acad. Sci. U.S.A. 114, 5041–5046. https://doi.org/10.1073/pnas.1704217114

Montiel, J., Szűcs, A., Boboescu, I.Z., Gherman, V.D., Kondorosi, É., Kereszt, A., 2016. Terminal Bacteroid Differentiation Is Associated With Variable Morphological Changes in Legume Species Belonging to the Inverted Repeat-Lacking Clade. Mol. Plant Microbe Interact. 29, 210–219. https://doi.org/10.1094/MPMI-09-15-0213-R

Nagy, K., Mikuláss, K.R., Végh, A.G., Kereszt, A., Kondorosi, É., Váró, G., Szegletes, Z., 2015. Interaction of cysteine-rich cationic antimicrobial peptides with intact bacteria and model membranes. Gen. Physiol. Biophys. 34, 135–144. https://doi.org/10.4149/gpb_2015002

Nallu, S., Silverstein, K.A.T., Samac, D.A., Bucciarelli, B., Vance, C.P., VandenBosch, K.A., 2013. Regulatory Patterns of a Large Family of Defensin-Like Genes Expressed in Nodules of Medicago truncatula. PLOS ONE 8, e60355. https://doi.org/10.1371/journal.pone.0060355

Nallu, S., Silverstein, K.A.T., Zhou, P., Young, N.D., Vandenbosch, K.A., 2014. Patterns of divergence of a large family of nodule cysteine-rich peptides in accessions of Medicago truncatula. Plant J. 78, 697–705. https://doi.org/10.1111/tpj.12506

Oldroyd, G.E.D., 2013. Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. Nat. Rev. Microbiol. 11, 252–263. https://doi.org/10.1038/nrmicro2990

Oldroyd, G.E.D., Murray, J.D., Poole, P.S., Downie, J.A., 2011. The rules of engagement in the legume-rhizobial symbiosis. Annu. Rev. Genet. 45, 119–144. https://doi.org/10.1146/annurev-genet-110410-132549

Oono, R., Denison, R.F., 2010. Comparing Symbiotic Efficiency between Swollen versus Nonswollen Rhizobial Bacteroids. Plant Physiology 154, 1541–1548. https://doi.org/10.1104/pp.110.163436

Ott, T., van Dongen, J.T., Gu¨nther, C., Krusell, L., Desbrosses, G., Vigeolas, H., Bock, V., Czechowski, T., Geigenberger, P., Udvardi, M.K., 2005. Symbiotic Leghemoglobins Are Crucial for Nitrogen Fixation in Legume Root Nodules but Not for General Plant Growth and Development. Current Biology 15, 531–535. https://doi.org/10.1016/j.cub.2005.01.042

Paetzel, M., Karla, A., Strynadka, N.C.J., Dalbey, R.E., 2002. Signal peptidases. Chem. Rev. 102, 4549–4580.

Pan, H., Oztas, O., Zhang, X., Wu, X., Stonoha, C., Wang, E., Wang, B., Wang, D., 2016. A symbiotic SNARE protein generated by alternative termination of transcription. Nature Plants 2, 1–5. https://doi.org/10.1038/nplants.2015.197

Pan, H., Wang, D., 2017. Nodule cysteine-rich peptides maintain a working balance during nitrogen-fixing symbiosis. Nature Plants 3, 1–6. https://doi.org/10.1038/nplants.2017.48

Pecrix, Y., Staton, S.E., Sallet, E., Lelandais-Brière, C., Moreau, S., Carrère, S., Blein, T., Jardinaud, M.-F., Latrasse, D., Zouine, M., Zahm, M., Kreplak, J., Mayjonade, B., Satgé, C., Perez, M., Cauet, S., Marande, W., Chantry-Darmon, C., Lopez-Roques, C., Bouchez, O., Bérard, A., Debellé, F., Muños, S., Bendahmane, A., Bergès, H., Niebel, A., Buitink, J., Frugier, F., Benhamed, M., Crespi, M., Gouzy, J., Gamas, P., 2018. Whole-genome landscape of Medicago truncatula symbiotic genes. Nat Plants 4, 1017–1025. https://doi.org/10.1038/s41477-018-0286-7

Penmetsa, R.V., Frugoli, J.A., Smith, L.S., Long, S.R., Cook, D.R., 2003. Dual genetic pathways controlling nodule number in Medicago truncatula. Plant Physiol. 131, 998–1008. https://doi.org/10.1104/pp.015677

Penterman, J., Abo, R.P., De Nisco, N.J., Arnold, M.F.F., Longhi, R., Zanda, M., Walker, G.C., 2014. Host plant peptides elicit a transcriptional response to control the Sinorhizobium meliloti cell cycle during symbiosis. Proc. Natl. Acad. Sci. U.S.A. 111, 3561–3566. https://doi.org/10.1073/pnas.1400450111

Pislariu, C.I., Sinharoy, S., Torres-Jerez, I., Nakashima, J., Blancaflor, E.B., Udvardi, M.K., 2019. The Nodule-Specific PLAT Domain Protein NPD1 Is Required for Nitrogen-Fixing Symbiosis. Plant Physiology 180, 1480–1497. https://doi.org/10.1104/pp.18.01613

Price, P.A., Tanner, H.R., Dillon, B.A., Shabab, M., Walker, G.C., Griffitts, J.S., 2015. Rhizobial peptidase HrrP cleaves host-encoded signaling peptides and mediates symbiotic compatibility. Proc Natl Acad Sci U S A 112, 15244–15249. https://doi.org/10.1073/pnas.1417797112

Pumplin, N., Zhang, X., Noar, R.D., Harrison, M.J., 2012. Polar localization of a symbiosis-specific phosphate transporter is mediated by a transient reorientation of secretion. PNAS 109, E665–E672. https://doi.org/10.1073/pnas.1110215109

Qin, S.-Y., Hu, D., Matsumoto, K., Takeda, K., Matsumoto, N., Yamaguchi, Y., Yamamoto, K., 2012. Malectin forms a complex with ribophorin I for enhanced association with misfolded glycoproteins. J. Biol. Chem. 287, 38080–38089. https://doi.org/10.1074/jbc.M112.394288

Radhakrishnan, G.V., Keller, J., Rich, M.K., Vernié, T., Mbadinga Mbadinga, D.L., Vigneron, N., Cottret, L., Clemente, H.S., Libourel, C., Cheema, J., Linde, A.-M., Eklund, D.M., Cheng, S., Wong, G.K.S., Lagercrantz, U., Li, F.-W., Oldroyd, G.E.D., Delaux, P.-M., 2020. An ancestral signalling pathway is conserved in intracellular symbioses-forming plant lineages. Nature Plants 1–10. https://doi.org/10.1038/s41477-020-0613-7

Ramlov, K.B., Laursen, N.B., Stougaard, J., Marcker, K.A., 1993. Site-directed mutagenesis of the organ-specific element in the soybean leghemoglobin lbc3 gene promoter. Plant J. 4, 577–580. https://doi.org/10.1046/j.1365-313x.1993.04030577.x

Ribeiro, C.W., Baldacci-Cresp, F., Pierre, O., Larousse, M., Benyamina, S., Lambert, A., Hopkins, J., Castella, C., Cazareth, J., Alloing, G., Boncompagni, E., Couturier, J., Mergaert, P., Gamas, P., Rouhier, N., Montrichard, F., Frendo, P., 2017. Regulation of Differentiation of Nitrogen-Fixing Bacteria by Microsymbiont Targeting of Plant Thioredoxin s1. Curr. Biol. 27, 250–256. https://doi.org/10.1016/j.cub.2016.11.013

Ried, M.K., Antolín-Llovera, M., Parniske, M., 2014. Spontaneous symbiotic reprogramming of plant roots triggered by receptor-like kinases. Elife 3. https://doi.org/10.7554/eLife.03891

Riely, B.K., Larrainzar, E., Haney, C.H., Mun, J.-H., Gil-Quintana, E., González, E.M., Yu, H.-J., Tricoli, D., Ehrhardt, D.W., Long, S.R., Cook, D.R., 2013. Development of tools for the biochemical characterization of the symbiotic receptor-like kinase DMI2. Mol. Plant Microbe Interact. 26, 216–226. https://doi.org/10.1094/MPMI-10-11-0276

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., Carvalho-Niebel, F. de, 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, 817–837. https://doi.org/10.1111/tpj.12442

Roy, S., Liu, W., Nandety, R.S., Crook, A.D., Mysore, K.S., Pislariu, C.I., Frugoli, J.A., Dickstein, R., Udvardi, M.K., 2019. Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. The Plant Cell. https://doi.org/10.1105/tpc.19.00279

Saha, S., DasGupta, M., 2015. Does SUNN-SYMRK Crosstalk occur in Medicago truncatula for regulating nodule organogenesis? Plant Signal Behav 10, e1028703. https://doi.org/10.1080/15592324.2015.1028703

Saha, S., Dutta, A., Bhattacharya, A., DasGupta, M., 2014. Intracellular catalytic domain of symbiosis receptor kinase hyperactivates spontaneous nodulation in absence of rhizobia. Plant Physiol. 166, 1699–1708. https://doi.org/10.1104/pp.114.250084

Satgé, C., Moreau, S., Sallet, E., Lefort, G., Auriac, M.-C., Remblière, C., Cottret, L., Gallardo, K., Noirot, C., Jardinaud, M.-F., Gamas, P., 2016. Reprogramming of DNA methylation is critical for nodule development in Medicago truncatula. Nature Plants 2, 1–10. https://doi.org/10.1038/nplants.2016.166

Schallus, T., Jaeckh, C., Fehér, K., Palma, A.S., Liu, Y., Simpson, J.C., Mackeen, M., Stier, G., Gibson, T.J., Feizi, T., Pieler, T., Muhle-Goll, C., 2008. Malectin: a novel carbohydrate-binding protein of the endoplasmic reticulum and a candidate player in the early steps of protein N-glycosylation. Mol. Biol. Cell 19, 3404–3414. https://doi.org/10.1091/mbc.e08-04-0354

Schauser, L., Roussis, A., Stiller, J., Stougaard, J., 1999. A plant regulator controlling development of symbiotic root nodules. Nature 402, 191–195. https://doi.org/10.1038/46058

Schnabel, E., Journet, E.-P., de Carvalho-Niebel, F., Duc, G., Frugoli, J., 2005. The Medicago truncatula SUNN gene encodes a CLV1-like leucine-rich repeat receptor kinase that regulates nodule number and root length. Plant Mol. Biol. 58, 809–822. https://doi.org/10.1007/s11103-005-8102-y

Shabab, M., Arnold, M.F.F., Penterman, J., Wommack, A.J., Bocker, H.T., Price, P.A., Griffitts, J.S., Nolan, E.M., Walker, G.C., 2016. Disulfide cross-linking influences symbiotic activities of nodule peptide NCR247. Proc. Natl. Acad. Sci. U.S.A. 113, 10157–10162. https://doi.org/10.1073/pnas.1610724113

Silverstein, K.A.T., Moskal, W.A., Wu, H.C., Underwood, B.A., Graham, M.A., Town, C.D., VandenBosch, K.A., 2007. Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. Plant J. 51, 262–280. https://doi.org/10.1111/j.1365-313X.2007.03136.x

Sinharoy, S., Torres-Jerez, I., Bandyopadhyay, K., Kereszt, A., Pislariu, C.I., Nakashima, J., Benedito, V.A., Kondorosi, E., Udvardi, M.K., 2013. The C2H2 transcription factor regulator of symbiosome differentiation represses transcription of the secretory pathway gene VAMP721a and promotes symbiosome development in Medicago truncatula. Plant Cell 25, 3584–3601. https://doi.org/10.1105/tpc.113.114017

Smit, P., Limpens, E., Geurts, R., Fedorova, E., Dolgikh, E., Gough, C., Bisseling, T., 2007. Medicago LYK3, an entry receptor in rhizobial nodulation factor signaling. Plant Physiol. 145, 183–191. https://doi.org/10.1104/pp.107.100495

Soyano, T., Kouchi, H., Hirota, A., Hayashi, M., 2013. NODULE INCEPTION Directly Targets NF-Y Subunit Genes to Regulate Essential Processes of Root Nodule Development in Lotus japonicus. PLOS Genetics 9, e1003352. https://doi.org/10.1371/journal.pgen.1003352

Spoel, S.H., Mou, Z., Tada, Y., Spivey, N.W., Genschik, P., Dong, X., 2009. Proteasome-mediated turnover of the transcription coactivator NPR1 plays dual roles in regulating plant immunity. Cell 137, 860–872. https://doi.org/10.1016/j.cell.2009.03.038

Sprent, J.I., 2007. Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation. New Phytol. 174, 11–25. https://doi.org/10.1111/j.1469-8137.2007.02015.x

Stougaard, J., Jørgensen, J.E., Christensen, T., Kühle, A., Marcker, K.A., 1990. Interdependence and nodule specificity of cis-acting regulatory elements in the soybean leghemoglobin lbc3 and N23 gene promoters. Mol. Gen. Genet. 220, 353–360. https://doi.org/10.1007/bf00391738

Stougaard, J., Sandal, N.N., Grøn, A., Kühle, A., Marcker, K.A., 1987. 5' Analysis of the soybean leghaemoglobin lbc3 gene: regulatory elements required for promoter activity and organ specificity. EMBO J 6, 3565–3569.

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, 959–962. https://doi.org/10.1038/nature00841

Szczyglowski, K., Szabados, L., Fujimoto, S.Y., Silver, D., de Bruijn, F.J., 1994. Site-specific mutagenesis of the nodule-infected cell expression (NICE) element and the AT-rich element ATRE-BS2* of the Sesbania rostrata leghemoglobin glb3 promoter. Plant Cell 6, 317–332. https://doi.org/10.1105/tpc.6.3.317

Tiricz, H., Szucs, A., Farkas, A., Pap, B., Lima, R.M., Maróti, G., Kondorosi, É., Kereszt, A., 2013. Antimicrobial nodule-specific cysteine-rich peptides induce membrane depolarization-associated changes in the transcriptome of Sinorhizobium meliloti. Appl. Environ. Microbiol. 79, 6737–6746. https://doi.org/10.1128/AEM.01791-13

Trujillo, D.I., Silverstein, K.A.T., Young, N.D., 2019. Nodule-specific PLAT domain proteins are expanded in the Medicago lineage and required for nodulation. New Phytologist 222, 1538–1550. https://doi.org/10.1111/nph.15697

Udvardi, M., Poole, P.S., 2013. Transport and metabolism in legume-rhizobia symbioses. Annu Rev Plant Biol 64, 781–805. https://doi.org/10.1146/annurev-arplant-050312-120235

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., Abe, M., Kereszt, A., Maroti, G., Uchiumi, T., Kondorosi, E., Mergaert, P., 2010. Plant peptides govern terminal differentiation of bacteria in symbiosis. Science 327, 1122–1126. https://doi.org/10.1126/science.1184057

Vernié, T., Kim, J., Frances, L., Ding, Y., Sun, J., Guan, D., Niebel, A., Gifford, M.L., Carvalho-Niebel, F. de, Oldroyd, G.E.D., 2015. The NIN Transcription Factor Coordinates Diverse Nodulation Programs in Different Tissues of the Medicago truncatula Root. The Plant Cell 27, 3410–3424. https://doi.org/10.1105/tpc.15.00461

Wais, R.J., Keating, D.H., Long, S.R., 2002. Structure-function analysis of nod factor-induced root hair calcium spiking in Rhizobium-legume symbiosis. Plant Physiol. 129, 211–224. https://doi.org/10.1104/pp.010690

Wang, D., Griffitts, J., Starker, C., Fedorova, E., Limpens, E., Ivanov, S., Bisseling, T., Long, S., 2010. A Nodule-Specific Protein Secretory Pathway Required for Nitrogen-Fixing Symbiosis. Science 327, 1126–1129. https://doi.org/10.1126/science.1184096

Wang, Longxiang, Wang, Longlong, Tan, Q., Fan, Q., Zhu, H., Hong, Z., Zhang, Z., Duanmu, D., 2016. Efficient Inactivation of Symbiotic Nitrogen Fixation Related Genes in Lotus japonicus Using CRISPR-Cas9. Front. Plant Sci. 7. https://doi.org/10.3389/fpls.2016.01333

Wang, Q., Liu, J., Li, H., Yang, S., Körmöczi, P., Kereszt, A., Zhu, H., 2018. Nodule-Specific Cysteine-Rich Peptides Negatively Regulate Nitrogen-Fixing Symbiosis in a Strain-Specific Manner in Medicago truncatula. Mol. Plant Microbe Interact. 31, 240–248. https://doi.org/10.1094/MPMI-08-17-0207-R

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. Proc Natl Acad Sci U S A 114, 6854–6859. https://doi.org/10.1073/pnas.1700715114

Xie, F., Murray, J.D., Kim, J., Heckmann, A.B., Edwards, A., Oldroyd, G.E.D., Downie, J.A., 2012. Legume pectate lyase required for root infection by rhizobia. PNAS 109, 633–638. https://doi.org/10.1073/pnas.1113992109

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. Proc. Natl. Acad. Sci. U.S.A. 114, 6848–6853. https://doi.org/10.1073/pnas.1700460114

Young, N.D., Debellé, F., Oldroyd, G.E.D., Geurts, R., Cannon, S.B., Udvardi, M.K., Benedito, V.A., Mayer, K.F.X., Gouzy, J., Schoof, H., Van de Peer, Y., Proost, S., Cook, D.R., Meyers, B.C., Spannagl, M., Cheung, F., De Mita, S., Krishnakumar, V., Gundlach, H., Zhou, S., Mudge, J., Bharti, A.K., Murray, J.D., Naoumkina, M.A., Rosen, B., Silverstein, K.A.T., Tang, H., Rombauts, S., Zhao, P.X., Zhou, P., Barbe, V., Bardou, P., Bechner, M., Bellec, A., Berger, A., Bergès, H., Bidwell, S., Bisseling, T., Choisne, N., Couloux, A., Denny, R., Deshpande, S., Dai, X., Doyle, J.J., Dudez, A.-M., Farmer, A.D., Fouteau, S., Franken, C., Gibelin, C., Gish, J., Goldstein, S., González, A.J., Green, P.J., Hallab, A., Hartog, M., Hua, A., Humphray, S.J., Jeong, D.-H., Jing, Y., Jöcker, A., Kenton, S.M., Kim, D.-J., Klee, K., Lai, H., Lang, C., Lin, S., Macmil, S.L., Magdelenat, G., Matthews, L., McCorrison, J., Monaghan, E.L., Mun, J.-H., Najar, F.Z., Nicholson, C., Noirot, C., O'Bleness, M., Paule, C.R., Poulain, J., Prion, F., Qin,

B., Qu, C., Retzel, E.F., Riddle, C., Sallet, E., Samain, S., Samson, N., Sanders, I., Saurat, O., Scarpelli, C., Schiex, T., Segurens, B., Severin, A.J., Sherrier, D.J., Shi, R., Sims, S., Singer, S.R., Sinharoy, S., Sterck, L., Viollet, A., Wang, B.-B., Wang, K., Wang, M., Wang, X., Warfsmann, J., Weissenbach, J., White, D.D., White, J.D., Wiley, G.B., Wincker, P., Xing, Y., Yang, L., Yao, Z., Ying, F., Zhai, J., Zhou, L., Zuber, A., Dénarié, J., Dixon, R.A., May, G.D., Schwartz, D.C., Rogers, J., Quétier, F., Town, C.D., Roe, B.A., 2011. The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature 480, 520–524. https://doi.org/10.1038/nature10625

Yuan, S., Zhu, H., Gou, H., Fu, W., Liu, L., Chen, T., Ke, D., Kang, H., Xie, Q., Hong, Z., Zhang, Z., 2012. A ubiquitin ligase of symbiosis receptor kinase involved in nodule organogenesis. Plant Physiol. 160, 106–117. https://doi.org/10.1104/pp.112.199000

Zasloff, M., 2002. Antimicrobial peptides of multicellular organisms. Nature 415, 389–395. https://doi.org/10.1038/415389a

Zheng, H., Bassham, D.C., Conceição, A. da S., Raikhel, N.V., 1999. The syntaxin family of proteins in Arabidopsis: a new syntaxin homologue shows polymorphism between two ecotypes. J Exp Bot 50, 915–924. https://doi.org/10.1093/jxb/50.Special_Issue.915