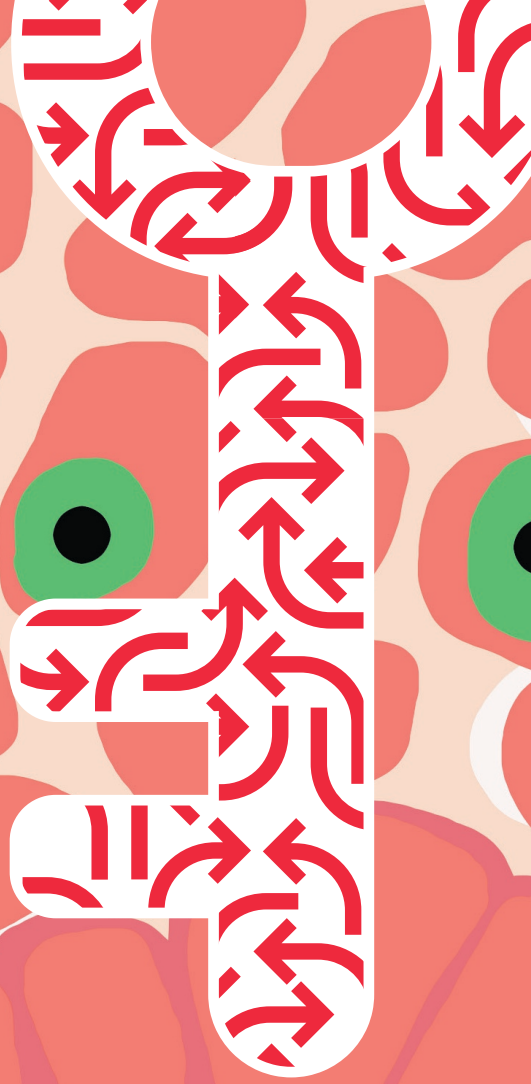Jieyu Liu

NIN:
the key to nodulation

Propositions

1. NIN (NODULE INCEPTION) plays a key role in all nodule developmental stages.
   (this thesis)

2. The function of NIN depends on its concentration.
   (this thesis)

3. Salt-tolerant microbiomes do not necessarily promote salt tolerance of plants
   (Santos et al., 2021).

4. Synthetic microbial communities are like traditional Chinese medicine: complex
   components produce obvious effects, while the mechanism remains a mystery.

5. Making progress in what is causing worries is the best way to relieve these
   worries.

6. Marriage for the purpose of symbiosis is fading out of human society, due to
   sufficient social supply.

Propositions belonging to the thesis, entitled

NIN: the key to nodulation

Jieyu Liu
Wageningen, 3 December 2021

NIN: the key to nodulation

Jieyu Liu

Thesis committee

Promotor

Prof. Dr T. Bisseling

Professor of Molecular Biology, Wageningen University & Research

Co-Promotors

Dr O. Kulikova

Researcher, Laboratory of Molecular Biology,

Wageningen University & Research

Dr W. Kohlen

Assistant professor, Laboratory of Molecular Biology,

Wageningen University & Research

Other members

Prof. Dr F. P. M. Govers, Wageningen University & Research

Prof. Dr D. Weijers, Wageningen University & Research

Prof. Dr G.E.D. Oldroyd, University of Cambridge, United Kingdom

Dr P. Ratet, Institute of Plant Sciences Paris-Saclay (IPS2); Centre National de la Recherche Scientifique (CNRS), France

NIN: the key to nodulation

Jieyu Liu

Thesis

submitted in fulfilment of the requirements for the degree of doctor

at Wageningen University

by the authority of the Rector Magnificus

Prof. Dr A. P. J. Mol,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Friday December 3, 2021

at 4 p.m. in the Aula.

CONTENTS

CHAPTER 1

General Introduction

Jieyu Liu

Laboratory of Molecular Biology, Department of Plant Sciences, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.

1. Introduction

As a major plant nutrient, nitrogen is essential for plant growth and development. Despite nitrogen gas is abundant in the atmosphere, the accessible forms of nitrogen for plants, such as nitrate and ammonium, are often limited in soil and restrict plant growth. To overcome this, about 110 million years ago, plants evolved a symbiosis with nitrogen fixing bacteria (Griesmann et al., 2018; van Velzen et al., 2018). In this symbiosis, bacteria are accommodated intracellularly in a specialized root organ named nodule. Inside the root nodules, bacteria convert the atmospheric nitrogen into ammonia, which is provided to the plant in exchange for carbon resources.

The nodule symbiosis only occurs in plant species which belong to the nitrogen fixation clade (NFC). This clade consists of four orders: Fabales, Rosales, Cucurbitales and Fagales. The ability to form nodules frequently occurs within species of the legume family (Fabales) (Soltis et al., 1995). This makes legumes important crop plants that are able to grow with low input of nitrate fertilizer and produce protein-rich seeds. In the other three orders, nodulation occurs in markedly fewer species. Legumes and Parasponia (Rosales) establish the nodule symbiosis with Gram-negative bacteria, which belong to different genera collectively called rhizobia (Peter et al., 1996). The other plant species that form the nodule symbiosis with Gram-positive *Frankia* bacteria are called actinorhizal plants (Santi et al., 2013).

Plant species in the different orders of the NFC use a conserved signaling pathway to initiate the nodule formation. This pathway is activated by lipo-chitooligosaccharides (LCOs, also called Nod factors (NFs)) that are secreted by rhizobia (Denarie and Debelle, 1996). Although the nature of the *Frankia*-secreted signal is still not clear, components of this signaling pathway have also been shown to be required for the nodule formation in the actinorhizal plants (Gherbi et al., 2008; Markmann et al., 2008; Svistoonoff et al., 2013; Den Camp et al., 2011). Interestingly, this signaling pathway is shared with a much older and far more widespread endosymbiosis: the arbuscular mycorrhizal (AM) symbiosis (Oldroyd, 2013). AM symbiosis is thought to have evolved ~450 million years ago (Remy et al., 1994; Bonfante and Genre, 2008). AM fungi are able to colonize 72%-80% of vascular plants and facilitate the uptake of micronutrients, phosphate and fixed form of nitrogen in exchange of plant carbon sources (Brundrett and Tedersoo, 2018). In addition to the common symbiotic signaling pathway, both microsymbionts from the AM and nodule symbiosis are hosted

intracellularly and surrounded by a specialized host membrane to avoid direct contact between the two partners and facilitate the exchange of, for example, nutrients in a controlled manner (Parniske, 2000). For both interactions, the formation of this symbiotic membrane interface require the specific exocytotic vesicle-associated membrane proteins (VAMP721d&e) of the host plant (Ivanov et al., 2012; Gavrin et al., 2016). As the AM symbiosis is more ancient, the common symbiotic signaling pathway as well as the VAMPs are most likely recruited from the AM symbiosis when nodulation evolved. However, although AM and nodule symbiosis share the common symbiotic signaling pathway, the downstream transcriptional network is different. In nodule symbiosis, this signaling pathway transcriptionally activates the key transcription factor NIN (NODULE INCEPTION) (Oldroyd, 2013; Singh et al., 2014). As a nodule specifically expressed gene, *NIN* distinguishes nodule symbiosis from AM symbiosis (Schauser et al., 1999; Marsh et al., 2007; Borisov et al., 2003; Wang et al., 2019; Bu et al., 2020; Clavijo et al., 2015; Kumar et al., 2020). NIN is essential for nodulation and this has been shown in legumes species, *Parasponia andersonii* (Parasponia) and in the actinorhizal species *Casuarina glauca* (Casuarina) (Schauser et al., 1999; Marsh et al., 2007; Borisov et al., 2003; Wang et al., 2019; Bu et al., 2020; Clavijo et al., 2015).

Recently, phylogenomic studies showed that the genes, which function in establishing nodule symbiosis are independently lost in the non-nodulating species of the NFC (Griesmann et al., 2018; Van Velzen et al., 2018). This supported the hypothesis that a single origin of nodulation was followed by massive loss of this special ability (van Velzen et al., 2019). Noticeably, all the (studied) nodulating species possess NIN, while the species that lost the nodulation ability in general lost NIN. This indicates that in addition to recruitment of the common symbiotic signaling pathway and symbiotic VAMPs from the AM symbiosis, nodule symbiosis specifically recruited NIN.

Root nodule initiation involves multiple developmental processes: root hair curling, infection thread formation, the start of nodule organogenesis, and a negative feedback loop that regulates nodule number (Kosslak and Bohlool, 1984; Downie, 2014). These processes occur at different time points and locations. Strikingly, NIN plays a key role in all these processes (Schauser et al., 1999; Marsh et al., 2007; Borisov et al., 2003; Wang et al., 2019; Bu et al., 2020; Clavijo et al., 2015; Soyano et al., 2014; Laffont et al., 2020). However, although the importance of NIN has been demonstrated in many aspects of nodulation, how it achieves these versatility remains elusive.

Understanding the molecular mechanisms by which NIN controls different steps is important in understanding nodulation. In this thesis, I focused on this. Here, I will first summarize the current knowledge concerning NIN.

2. Nod factor signaling pathway

Genetic studies on model legumes *Medicago truncatula* (Medicago) and *Lotus japonicus* (Lotus) identified a signaling pathway, which later was shown to be conserved within the NFC in establishing nodule symbiosis. In response to root exudates, including flavonoids, rhizobia synthesize and secrete the NFs. NFs can be perceived by the membrane located heteromeric complex of LysM-domain receptor kinases (LYK3/NFP in Medicago; NFR1/NFR5 in Lotus) and leucine-rich repeat receptor kinase (DMI2 in Medicago and SYMRK in Lotus) (Stracke et al., 2002; Limpens et al., 2005; Madsen et al., 2004; Radutoiu et al., 2003; Limpens et al., 2003; Arrighi et al., 2006; Mulder et al., 2006; Smit et al., 2007; Endre et al., 2002). The NF receptor complex activates the downstream cascade resulting in calcium oscillations in the nucleus. This requires HMGR1 (3-hydroxy-3-methylglutaryl CoA reductase 1), an enzyme involved in mevalonate production (Jayaraman et al., 2015; Kevei et al., 2007); nuclear membrane-located cation channels (DMI1 in Medicago; CASTOR and POLLUX in Lotus) , and calcium channels (CNGC15a-c in Medicago) (Ané et al., 2004; Kanamori et al., 2006; Charpentier et al., 2008, 2016). The calcium signal is decoded by a calcium calmodulin-dependent kinase (DMI3 in Medicago and CCaMK in Lotus) (Mitra et al., 2004; Tirichine et al., 2006). Further, CCaMK/DMI3 activates transcription factor CYCLOPS (IPD3 in Medicago) by phosphorylation after which, it is able to induce the transcription of downstream genes (Messinese et al., 2007; Yano et al., 2008).

The signaling pathway described above (except the LysM receptors) is shared with AM symbiosis. However, downstream of CYCLOPS, the signaling pathway required for AM and nodulation is diverged. In AM symbiosis, CYCLOPS directly induces the expression of *REDUCED ARBUSCULAR MYCORRHIZA1* (*RAM1*), a GRAS-type transcription factor that functions as a central regulator of arbuscule development (Gobbato et al., 2012; Pimprikar et al., 2016). In contrast, in nodule symbiosis, CYCLOPS induces the expression of *NIN* (Singh et al., 2014). As a direct target of CYCLOPS, *NIN* is among the first induced nodule symbiotic genes. NIN is specifically expressed during the nodulation process and essential for nodule development, but not for AM symbiosis

(Schauser et al., 1999; Marsh et al., 2007; Borisov et al., 2003; Wang et al., 2019; Bu et al., 2020; Clavijo et al., 2015; Kumar et al., 2020).

3. Function of NIN during nodule development

3.1 NIN is a network hub in Infection thread formation

To reach the developing nodule primordia, bacteria have to traverse the root epidermis, sometimes also cortical cell layers. There are several ways to achieve this such as: crack entry through the lateral root base, or via intercellular infection between epidermal cells and the most common mode by an intracellular infection thread (Sprent, 2007). Infection thread formation has been well-studied in Medicago and Lotus. Upon exposure to rhizobium secreted NF, the microtubule skeleton starts to rearrange within a few minutes, the root hair tip (or branch of the root hair) curls around the bacteria by which they become entrapped (Timmers et al., 1999). This curling stops when the tip reaches the shank of the root hair, forming an enclosed infection chamber. The entrapped bacteria divide within this chamber, forming a colony that is referred to as infection focus (Oldroyd et al., 2011). By targeting of vesicles to this site, local cell wall hydrolysis occurs and the plasma membrane starts to invaginate. It forms a tube-like structure and a new cell wall is formed inside this tube. This intracellular tube-like structure is named infection thread (IT) (Oldroyd et al., 2011). The IT leads bacteria inside the plant cell, traverses the root hair and grows towards the nodule primordium, where rhizobia are released (Daniel J. Gage, 2004; Oldroyd et al., 2011).

NIN is transcriptionally induced in the root epidermis within hours upon NF perception (Yoro et al., 2014; Vernié et al., 2015a; Van Zeijl et al., 2015; Breakspear et al., 2014). The role of NIN in IT formation has been studied in several legumes. Knockout mutants of *nin* showed excessive root hair curling in response to NF (Marsh et al., 2007; Schauser et al., 1999; Borisov et al., 2003). This improper root hair curling makes it impossible to form a closed infection chamber. In actinorhizal species Casuarina, *NIN* expression has been shown to be correlated with infection by Frankia and knockdown mutants were impaired in root hair deformation (Clavijo et al., 2015). So most likely, the function of NIN in controlling IT formation is conserved within the NFC.

NIN protein induces IT formation by triggering expression of genes required for this process. A recent root hair transcriptome study identified NIN as a transcriptional network hub controlling growth modules required for IT formation (Liu et al., 2019a). It has been shown that the expression of 54% of the induced genes involved in infection are NIN dependent. These genes are involved in multiple molecular processes, including cell cycle, cell wall remodeling, hormone metabolism, nutrient uptake and assimilation. For example, NIN directly binds to the promoter of the gene encoding the cell wall modifying enzyme NPL (Nodulation Pectate Lyase) (Xie et al., 2012). NIN also regulates the CCAAT box-binding transcription factor *NF-YA1* (*Nuclear Factor-Y Subunit A1*), and the nuclear-localized long coiled–coil protein *RPG* (*Rhizobium-directed Polar Growth*) (Soyano et al., 2013; Liu et al., 2019a; Soyano et al., 2014). Mutations in these direct target genes lead to the abnormal IT formation, underlining the importance of NIN in controlling IT formation.

3.2 NIN is required for the initiation of nodule organogenesis

Concomitant with the rhizobia infection in the epidermis, mitosis is activated in the pericycle/cortical cells below the site of infection to form a nodule primordium, which further develops into a nitrogen fixing nodule. This nodule organogenesis process is fully blocked in knockout mutants of *nin* in legumes and Parasponia (Schauser et al., 1999; Marsh et al., 2007; Borisov et al., 2003; Bu et al., 2020; Wang et al., 2019). In Casuarina, nodule organogenesis is impaired in knockdown mutants (Clavijo et al., 2015). In addition, overexpression of *NIN* can induce the formation of a nodule like structure in the absence of rhizobia (Soyano et al., 2013). Therefore, NIN plays a key role in nodule organogenesis and most likely this role is conserved within the NFC.

At 24 hours post inoculation with rhizobia, when IT formation has been initiated in the epidermis, cell divisions already occur in the pericycle of the Medicago roots (Xiao et al., 2014). At this stage, bacteria are only present in the root epidermis. NFs are immobile molecules, therefore, NF signaling could have only occurred in the epidermis (Goedhart et al., 2000). It is puzzling how NIN controls nodule organogenesis in the inner root cell layers. It has been proposed that *NIN* expression in the inner root cell layers is activated by a mobile signal which is generated by the NF signaling in the epidermis (Deinum et al., 2016). Alternatively, it has been postulated that the epidermal produced NIN protein might be transported to the inner root cell layers and activate cell division there (Figure et al., 2015).

In addition to NF, expression of *NIN* can also be induced by cytokinin. Application of the exogenous cytokinin induces *NIN* expression and can form a nodule like structure in the absence of rhizobia (Gauthier-Coles et al., 2019; Cooper and Long, 1994). Cytokinin is accumulated upon NF perception and the induction of *NIN* by either cytokinin or NFs requires the cytokinin receptor MtCRE1 (ortholog of LjLHK1) (van Zeijl et al., 2015; Gonzalez-Rizzo et al., 2006a; Plet et al., 2011). For these reasons cytokinin might be the mobile signal that triggers *NIN* expression in the inner layers.

Studies on the *nin* weak allele *daphne* of Lotus, provide a valuable insight in how NIN simultaneously controls IT formation and nodule organogenesis (Yoro et al., 2014). The *daphne* mutant forms an excessive number of ITs, but nodule organogenesis is blocked. This showed that the regulation of these two processes by NIN can be uncoupled. The *daphne* mutant phenotype is caused by a chromosomal translocation about 7kb upstream of the *NIN* start codon (Yoro et al., 2014). This suggests that the regulatory region required for NIN controlled nodule organogenesis is located upstream of this 7kb promoter.

The regulatory sequence of *NIN* in controlling nodule organogenesis has been a long standing mystery. In **Chapter 2**, we identified a remote *cis*-regulatory region which is required for *NIN* expression in the pericycle to initiate nodule primordium formation in Medicago.

3.3 The function of NIN in later developmental stages

Legume nodules have been divided into two types: indeterminate and determinate nodules (Franssen et al., 1992). The model legume Medicago, which the study in this thesis focused on, forms indeterminate nodules as they have a persistent meristem at their apex. The other model legume Lotus forms determinate nodule, as it has a meristem that is only transiently present.

Nodule development has been studied in detail in Medicago (Xiao et al., 2014). It starts with anticlinal cell divisions in pericycle cells opposite the protoxylem poles (Heidstra et al., 1997). Subsequently, cell divisions are induced in the most inner two cortical cell layers (C5 and C4). This is followed by cell division in the middle (third) cortical cell layer (C3) and endodermis. When C4/5 formed about eight cell layers and endodermis and pericycle formed six to eight cell layers, mitotic activity stopped in these cells. In contrast, C3-derived cells kept on dividing and developed into a nodule meristem which continuously adds cells to the nodule tissues. The C4/5 derived cells become penetrated by the

infection threads from which rhizobia are released. The released bacteria are surrounded by a host derived membrane, a process controlled by a nodule specific exocytotic pathway (Ivanov et al., 2012; Gavrin et al., 2016). These organelle-like structures are called symbiosomes (Roth and Stacey, 1989). The infection and release processes continue in the mature nodule in the cell layers adjacent to meristem. These one/two adjacent cell layers are also the place where NF receptors accumulated, so the place where NF signaling occurs (Moling et al., 2014). Upon release, symbiosomes divide, which results in massive colonization of the host cell. In addition, both the infected host cell and bacteria undergo endoreduplication and become enlarged (Vinardell et al., 2013). Bacterial differentiation is controlled by a number of nodule-specific cysteine-rich peptides (NCRs) which are produced by the host and delivered into bacteria (Van de Velde et al., 2010; Alunni and Gourion, 2016; Mergaert et al., 2003).

The meristem derived cells form a distinct zone called infection zone. The infection zone is followed, along the longitudinal axis, by the fixation zone where nitrogen fixation take place. The transition from infection zone to the fixation zone (interzone II/III according to Vasse et al., 1990) occurs rapidly. This transition is accompanied with many changes in both gene expression and cell morphology (Gavrin et al., 2014; Vasse et al., 1990). For example, induction of rhizobial *nif* genes which encode components of nitrogenase (the enzyme responsible for nitrogen fixation); the accumulation of starch in the non-infected cells and at the periphery of the infected cells; bacteria markedly enlarge and fill most space of the host cell; rhizobia become radially aligned; vacuoles collapse; and cytoskeleton rearranges (Gavrin et al., 2014; Vasse et al., 1990; Timmers et al., 1998; Kitaeva et al., 2016; Franssen et al., 1992). The cells derived from C4 and C5 of the nodule primordia form the basal part of the fixation zone. The nodule meristem adds cell to the central tissue and after differentiate they become part of the fixation zone. In the older nodules, a senescence zone is formed at the basal part of the nodule (Van De Velde et al., 2006). In this zone, both host cell and bacteria are degraded by highly induced proteolytic activities. So, along the longitudinal axis a developmental gradient is formed with the youngest cells in the apex (meristem) and the oldest cells in the basis, near the root attachment point.

Interestingly, although massive intracellular infection by the rhizobia, the defense response of the host plant is well suppressed. This is due to the induction of defense suppressing genes, such as *Symbiotic Cysteine-rich Receptor*

Kinase (*SymCRK*), *Regulator of Symbiosome Differentiation* (*RSD*), *Defective in Nitrogen Fixation 2* (*DNF2*) and *NODULESWITH ACTIVATED DEFENCE 1* (*NAD1*) (Bourcy et al., 2013; Sinharoy et al., 2013; Berrabah et al., 2014b; Wang et al., 2016; Domonkos et al., 2017). Mutants of these genes showed strong defense responses and/or a premature senescence phenotype.

In contrast to the intensive studies of the function of NIN during nodule initiation, the role of NIN in later nodule developmental stages has hardly been investigated. Two Medicago *nin* weak alleles, NF10547 (*nin-16*) and NF0440, have been reported to form the nitrogen fixing deficient nodules (Pislariu et al., 2012; Veerappan et al., 2016). However, it is not clear what caused the nitrogen fixing deficient phenotype. In **Chapter 4**, we studied these two mutants in detail.

3.4 NIN systemically controls nodule number through autoregulation of nodulation (AON)

Symbiotic nitrogen fixation is an energy consuming process. Therefore, plants must evaluate the costs and benefits and make an economically optimal decision. To do this, plants developed a mechanism called autoregulation of nodulation (AON) to control the nodule number (Kosslak and Bohlool, 1984; Krusell et al., 2002; Nishimura et al., 2002; Okamoto et al., 2013, 2009). AON involves communication between root and shoot. Along with nodule initiation, CLAVATA3/ ENDOSPERMSURROUNDING REGION (CLE) peptides (CLE12/13/34/35 in Medicago, CLE-RS1/2/3 in Lotus, GmRIC1/2 in *Glycine max*, and *PvRIC1/2* in common bean (*Phaseolus vulgaris*)) are produced in roots (Okamoto et al., 2013, 2009; Soyano et al., 2014; Mortier et al., 2010; Lim et al., 2011, 2014; Reid et al., 2011; Ferguson et al., 2014; Nishida et al., 2016; Mens et al., 2021). They act as mobile signals and are transported through xylem to the shoot where they are perceived by their receptor CLAVATA1 (CLV1)-like leucine-rich repeat receptor-like kinase (LjHAR1/MtSUNN/PsSYM29/GmNARK/PvNARK) (Krusell et al., 2002; Nishimura et al., 2002; Searle et al., 2003; Schnabel et al., 2005; Ferguson et al., 2014). Mutations in these receptor genes result in a hyper nodulation phenotype. Perception of CLE peptides by the receptors results in downregulation of the shoot produced signal miR2111 (Tsikou et al., 2018; Gautrat et al., 2020). miR2111 can move to the root where it cleaves the nodulation inhibitor TOO MUCH LOVE (TML). So, upon perception of CLE peptides, less miR2111 is transported to the root, enabling TML to block the formation of new nodules. In this way the number of nodules that are formed is controlled. In addition to miR2111, perception of CLE peptides activates the

biosynthesis of iP-type cytokinin which is transported from shoots to roots where it suppress nodulation also in a TML dependent manner (Sasaki et al., 2014). So, the robust negative feedback loops are formed to control the number of nodules.

Studies in legumes showed that during nodulation the expression of *CLE* genes are induced in a NIN-dependent manner. In Lotus and *Glycine max,* LjNIN and GmNINa directly bind to the NIN-binding sequence (NBS) in *LjCLE-RS1/2* and *GmRIC1/2* promoter and induce their expression, respectively (Wang et al., 2019; Soyano et al., 2014). In Medicago, the expression of Mt*CLE12/13* is also dependent on MtNIN (Mortier et al., 2010). In Lotus, constitutively expressed *NIN* systemically represses the induction of *NIN* expression in the non-transgenic roots, and this suppression is dependent on the Lotus CLE peptides receptor HAR1 (Soyano et al., 2014). This further results in systemic suppression of nodule formation in untransformed roots of the same plant. So NIN, as both inducer and responder has a central role in this feedback loop, which ensures that an optimal number of nodule is formed.

4. *NIN* has evolved from a *NIN-like Protein* (*NLP*)

As the first cloned nodulation gene, NIN was first identified in Lotus in 1999 (Schauser et al., 1999). With NIN as the founding member, NIN-like proteins (NLP) were later identified as master regulators of nitrate signaling. NLPs are wide spread in the plant kingdom and regulate multiple plant developmental processes in response to nitrate state (Mu and Luo, 2019a). NINs and NLPs share two conserved domains, the RWP-RK DNA binding domain and Phox and Bem1 (PB1) protein-protein interaction domain (Schauser et al., 2005). RWP-RK domain is highly conserved between NINs and NLPs, indicating they might bind to similar *cis*-regulatory elements. Indeed, it has been shown in Lotus that both LjNIN and NRSYM1 (LjNLP4) can bind to the NIN-binding sequence in the Lj*CLE-RS2* promoter (Nishida et al., 2018). Recently, a more detailed comparison was performed between LjNIN and LjNLP4 (Nishida et al., 2021). This showed that they bind to both identical and distinct *cis*-elements, indicating that changes in their DNA binding properties has been evolved. The PB1 domain can mediate the interaction between NIN and NLPs and it is necessary for the full induction of nitrate-dependent expression of target genes of NLPs (Konishi and Yanagisawa, 2019; Lin et al., 2018a; Nishida et al., 2021).

Unlike NINs, which are transcriptionally activated during nodule symbiosis, NLPs are constitutively expressed and are post transcriptionally regulated by nitrate signaling (Liu et al., 2017; Chardin et al., 2014; Lin et al., 2018a; Nishida et al., 2018; Marchive et al., 2013; Cao et al., 2017). Under low nitrate condition, most studied NLPs are retained in the cytoplasm, while they are translocated to the nuclei when a high nitrate level is sensed (Marchive et al., 2013; Guan et al., 2017; Lin et al., 2018a; Cao et al., 2017; Nishida et al., 2018). In some special cases, NLPs are constitutively located in the nucleus, but its transcriptional activity is still regulated by nitrate (Alfatih et al., 2020; Yan et al., 2016; Nishida et al., 2021). In response to nitrate, nuclear localized NLPs negatively regulate nodule symbiosis by (Lin et al., 2018a; Luo et al., 2021; Nishida et al., 2018, 2021). This can be mediated by competitive interaction with NIN through the PB1 domain. For example, under high nitrate conditions, co-transformation of MtNLP1 with MtNIN significantly reduced the MtNIN activation of *proCRE1:GUS* expression in tobacco leaves and the interaction between MtNLP1 and MtNIN has shown to be depended on the PB1 domain (Lin et al., 2018a).

RWP-RK and PB1 domain both are close to the C-terminus of the protein. Notably, the N-terminal region is not conserved between legume NINs and NLPs (Chardin et al., 2014; Suzuki et al., 2013b). A nitrate-responsive domain (NRD) is present in NLPs, but not in legume NIN (Suzuki et al., 2013b). Within this domain, there is a serine (S205 in AtNLP7) which can be phosphorylated in response to nitrate and this is essential to mediate translocation of NLP from cytoplasm to nucleus (Liu et al., 2017). This serine is conserved in NLPs, but lost in legume NINs. However, it is not known whether the change in this region was essential for the recruitment of NIN in the nodulation process. In **Chapter 5** and **6**, we focused on NIN and NLP in an evolutionary perspective. We showed that NLPs are not functional equivalence with Medicago NIN and studied the minimal changes required for NIN recruitment in nodulation in both its promoter and protein sequences.

Outline

In **Chapter 2**, we studied how NIN controls multiple processes during nodule initiation. We showed that the regulatory sequences sufficient for the infection process in epidermis are located within a 5kb region directly upstream of the *NIN* start codon in Medicago. Furthermore, we identif a remote upstream *cis*-regulatory region, which required for the expression of *NIN* in the pericycle to initiate nodule organogenesis. This remote region contains putative cytokinin response elements, and it is conserved in legume species. This together with cytokinin signaling genes are expressed in the pericycle in the non-inoculated roots strongly suggest that *NIN* expression is first triggered by cytokinin signalling in the pericycle to initiate nodule primordium formation.

In **Chapter 3**, we analysed NF signaling controlled genes in Medicago nodules. We made use of spontaneous nodules formed by overexpression of NF signaling gene. These spontaneous nodules showed wildtype (WT)-like tissue organization. Transcriptome analysis showed that many genes, which are especially expressed in infection and fixation zone of WT nodules, are induced in the spontaneous nodules. Further, to study late stage NF signaling controlled genes and distinguish them from genes that are activated upon bacterial release, we made use of the mutants involved (TE7) and independent (*Vamp721d&e RNAi*) of NF signaling. Combined with laser microdissection and transcriptome analysis, we identified the late stage NF signalling controlled set of genes. These genes turned to be markedly different from the genes activated in roots upon NF application.

In **Chapter 4**, we studied the functions of NIN in Medicago nodules. We found both NIN mRNA and protein accumulated at the highest level in the proximal part of the infection zone and suddenly dropped to a lower level when transition to fixation zone. Two *nin* weak allele mutants, *nin-13/16,* form a rather normal nodule infection zone, but failed in transition to fixation zone. Instead, a zone with defence responses and premature senescence occurred and symbiosome development gets arrested. Mutations in *nin-13/16* resulted in a truncated NIN lacking the conserved PB1 domain. However, this did not cause the nodule phenotype as *nin* mutants expressing $NIN_{\Delta PB1}$ formed wild-type-like nodule. The phenotype is likely caused by reduced *NIN* mRNA levels in the cytoplasm.

In **Chapter 5**, we studied changes in NIN related to its recruitment in nodulation. We showed that NIN is neither functionally equivalent with its orthologues

outside the NFC nor with its paralogues. The main reason for this could be that most NLPs are nuclear localized only when high nitrate is sensed, whereas NIN is constitutively nuclear localized. Further, we identified a few amino acid differences between NIN and its orthologues outside the NFC. These changes are important for the function of NIN in nodule formation. They probably were introduced in NIN when it was recruited in the nodulation process and might contribute to its constitutive nuclear localization.

In **Chapter 6**, we discuss how NIN has diverged from the ancestral NLP, what minimal changes would have been essential to become a key transcription controlling nodulation and which adaptations might have evolved later.

In **Chapter 7**, I integrated the findings described in this thesis and the knowledge from literature and proposed a model that can explain how NIN functions in multiple steps of the nodulation process.

CHAPTER 2

A remote *cis*-regulatory region is required for *NIN* expression in the pericycle to initiate Medicago nodule primordium formation

Jieyu Liu[a,1], Luuk Rutten[a,1], Erik Limpens[a], Tjitse van der Molen[a], Robin van Velzen[a], Rujin Chen[b,2], Yuhui Chen[b,2], Rene Geurts[a], Wouter Kohlen[a], Olga Kulikova[a,*], and Ton Bisseling[a,*]

[a] Laboratory of Molecular Biology, Department of Plant Sciences, Graduate School Experimental Plant Sciences, Wageningen University, 6708 PB Wageningen, The Netherlands

[b] Plant Biology Division, The Samuel Roberts Noble Foundation, Ardmore, OK 73401, USA.

[1] These authors contributed equally to this work

[2] Present address: School of Life Sciences, Lanzhou University, 730000 Lanzhou, China

*Correspondence to: ton.bisseling@wur.nl; olga.kulikova@wur.nl

Abstract

The legume-rhizobium symbiosis results in nitrogen fixing root nodules, and their formation involves both intracellular infection initiated in the epidermis and nodule organogenesis initiated in inner root cell layers. *NODULE INCEPTION* (*NIN*) is a nodule-specific transcription factor essential for both processes. These NIN-regulated processes occur at different times and locations in the root, demonstrating a complex pattern of spatiotemporal regulation. We show that regulatory sequences sufficient for the epidermal infection process are located within a 5 kb region directly upstream of the *NIN* start codon in *Medicago truncatula*. Furthermore, we identify a remote upstream *cis*-regulatory region required for the expression of *NIN* in the pericycle, and we show that this region is essential for nodule organogenesis. This region contains putative cytokinin response elements, and is conserved in eight more legume species. Both the cytokinin receptor *CRE1*, which is essential for nodule primordium formation, and the B-type response regulator *RR1* are expressed in the pericycle in the susceptible zone of the uninoculated root. This, together with the identification of the cytokinin responsive elements in the *NIN* promoter, strongly suggests that *NIN* expression is initially triggered by cytokinin signalling in the pericycle to initiate nodule primordium formation.

Introduction

The formation of nitrogen-fixing nodules is induced by rhizobium bacteria on the roots of legumes. It involves several processes: the induction of intracellular infection by rhizobia; nodule organogenesis; and a negative feedback loop that determines the number of nodules (Downie, 2014; Kosslak and Bohlool, 1984). Strikingly, the transcription factor NODULE INCEPTION (NIN), which is specifically expressed during nodulation, plays a key role in all of these processes (Schauser et al., 1999; Marsh et al., 2007; Soyano et al., 2014). These processes occur at different time points and locations, suggesting that *NIN* has a complex spatiotemporal regulation of expression that is regulated by distinct *cis*-regulatory sequences in its promoter. However, although *NIN* was identified almost two decades ago in *Lotus japonicus* (Lotus) (Schauser et al., 1999) and more than ten years ago in *Medicago truncatula* (Medicago) (Marsh et al., 2007), the promoter regions required for full complementation of *nin* knockout mutants have not been identified. Currently, it is unclear how NIN is involved in the multiple steps of the nodulation process.

2

In Medicago, nodule organogenesis starts with the local mitotic activation of pericycle cells, and subsequently cell division extends to the more outward located endodermis and cortex (Xiao et al., 2014). *Sinorhizobium meliloti* bacteria invade roots through tube-like structures called infection threads. Formation of infection threads in root hairs requires the prior induction of root hair curling. A tight curl is formed when the tip of the curling root hair touches the shank of the hair and the root hair stops growing and forms an infection chamber. Microcolonies of rhizobia then develop within these chambers, in which the rhizobia can induce formation of the infection thread.

In *nin* null mutants, extensive root hair curling and deformation are induced by bacteria, a proper infection chamber fails to be established and only few bacteria are present within curled root hairs (Fournier et al., 2015; Schauser et al., 1999; Marsh et al., 2007). Wild-type NIN induces infection thread formation by triggering expression of genes required for infection thread formation such as *NF-YA1*, which encodes a subunit of a nuclear factor Y complex, and *NPL*, which encodes a nodulation pectate lyase (Soyano et al., 2013; Laporte et al., 2014; Xie et al., 2012). Subsequently, infection threads grow towards the nodule primordia. There, rhizobia are released into nodule primordium cells derived from the cortex. These cells become infected cells which host thousands of nitrogen-fixing bacteria. NIN is also required for autoregulation of nodulation, a negative feedback system involving root-to-shoot communication to determine the optimal number of nodules. This autoregulation mechanism includes the induction of CLAVATA3/ESR-RELATED (*CLE*) genes by NIN, and the *CLE*-encoded peptides induce systemic signalling between root and shoot, suppressing the formation of new nodule primordia (Soyano et al., 2014).

Expression of *NIN* is induced in the epidermis upon perception of nodulation (Nod) factors, which are lipochitooligosaccharides(LCOs) secreted by rhizobia (Vernié et al., 2015; Van Zeijl et al., 2015). Nod-factor signalling induces Ca2+ spiking, which activates the nuclear-localized calcium and calmodulin-dependent kinase (CCaMK) (Ehrhardt et al., 1996; Mitra et al., 2004). CCaMK phosphorylates *CYCLOPS,* a transcription factor that activates *NIN* expression (Yano et al., 2008; Singh et al., 2014). At about 24 h post inoculation (hpi), formation of both infection threads and nodule primordium, are initiated in Medicago roots (Xiao et al., 2014). At this developmental stage, Nod factor signalling occurs exclusively in the epidermis because rhizobia are present only there and Nod factors are immobile molecules (Goedhart et al., 2000). Therefore, NIN can induce infection thread formation in a cell-autonomous way

in the epidermis, but it remains unclear how NIN can induce nodule primordium formation in inner root cell layers. It has been postulated that *NIN* expression can be induced in these root layers by a mobile signal that is generated upon Nod factor signalling in the epidermis (Hayashi et al., 2014). Alternatively, NIN proteins produced in the epidermis may be transported to the inner root layers (Vernié et al., 2015; Jardinaud et al., 2016).

In addition to regulation of *NIN* by CYCLOPS, *NIN* expression depends on cytokinin signalling. Exogenous application of cytokinin is sufficient to trigger *NIN* expression and also formation of structures resembling nodules (Heckmann et al., 2011; Gonzalez-Rizzo et al., 2006; Plet et al., 2011). Notably, Nod factor application results in the accumulation of cytokinin (Van Zeil et al., 2015). Furthermore, the induction of *NIN* expression by either Nod factors or cytokinin requires the cytokinin receptor CRE1, which plays a key role in nodule organogenesis (Van Zeijl et al., 2015; Gonzalez-Rizzo et al., 2006; Plet et al., 2011). Studies of the weak *nin* allele of the Lotus *daphne* mutant (Yoro et al., 2014) have provided valuable insight into the involvement of different *cis*-regulatory sequences in the *NIN* promoter that regulate infection and nodule organogenesis. In *daphne* roots, rhizobium infection and primordium formation are uncoupled. Formation of nodule primordia is completely absent, and increased numbers of infection threads are formed in the epidermis. The *daphne* mutation is caused by a large insertion that is ~7 kb upstream of the *NIN* start codon. This suggests that this 7-kb region includes essential *cis*-regulatory regions that are required for infection thread formation in the epidermis but are insufficient for the activation of cortical cell divisions. To determine how NIN induces these processes, we must identify the precise *cis*-regulatory regions in the *NIN* promoter that drive proper spatiotemporal *NIN* transcription.

Here, we have identified a conserved *NIN* promoter region that is essential for nodule organogenesis in Medicago. This region contains several putative cytokinin response elements and regulates *NIN* expression in the pericycle, where the cytokinin receptor (*CRE1*) and a B-type response regulator (*RR1*) are constitutively expressed. This reveals a key role for the pericycle in formation of nodule primordia, in which *NIN* expression is most likely activated by cytokinin signalling.

Results

Isolation of a Medicago *nin* Mutant in which Infection and Nodule Organogenesis Are Uncoupled

By screening a plant population obtained from Medicago seeds that were mutagenized by fast neutron bombardment (Noble Research Institute, LLC., Ardmore USA), we identified a Nod- mutant which we have named FN8113. Three weeks post inoculation (wpi) with *Sinorhizobium meliloti*, the FN8113 mutant formed excessive numbers of infection threads, but nodulation was strongly impaired (compare Figures 1A and 1B with Figures 1C and 1D). We quantified the infection thread number in FN8113 and wild type roots at 2 wpi. The number of infection threads in FN8113 roots was more than tenfold the number in wild type roots (Figure 1E). Root hair curling in FN8113 resembled that of wild type, as entrapped bacteria formed colonies and infection threads were formed (compare Figure 1F and 1G). The majority of infection threads were arrested in root hairs, but longitudinal sections of roots showed that a few infection threads could reach cortical cell layers (Figure 1H). Occasionally, some cortical cells had divided locally around infection threads. However, cell divisions were not induced in the inner root cell layers, where nodule primordia are initiated in wild type Medicago plants.

A segregating F2 population resulting from a cross between FN8113 (cv Jemalong A17) and Jemalong A20, showed an approximately 3:1 ratio of Nod+:Nod- plants (118 F2 plants; 84 Nod+: 34 Nod-). This indicates that FN8113 has a single recessive mutation that is responsible for its Nod- phenotype. Simple sequence repeat markers were used to determine the position of the mutation, which was localized to the end of chromosome 5, where *NIN* is located. Next, whole-genome sequencing was used to identify the mutation in this region, and this revealed a translocation of a ~2.49 Mbp region from chromosome 2 into chromosome 5. This was inserted 4120 bp upstream of the *NIN* start codon (–4120). In addition, a small deletion of 15 bp between –4121 and –4135 was detected (Figure 1I). No mutations were found in the *NIN* coding sequence. FN8113 was shown to be a *nin* mutant because its Nod- phenotype could be complemented with a biologically functional *NIN* promoter driving *NIN* (described below). Because the phenotype of FN8113, as well as the nature of its mutation, are strikingly similar to Lotus *daphne*, we named the FN8113 mutant *daphne-like*.
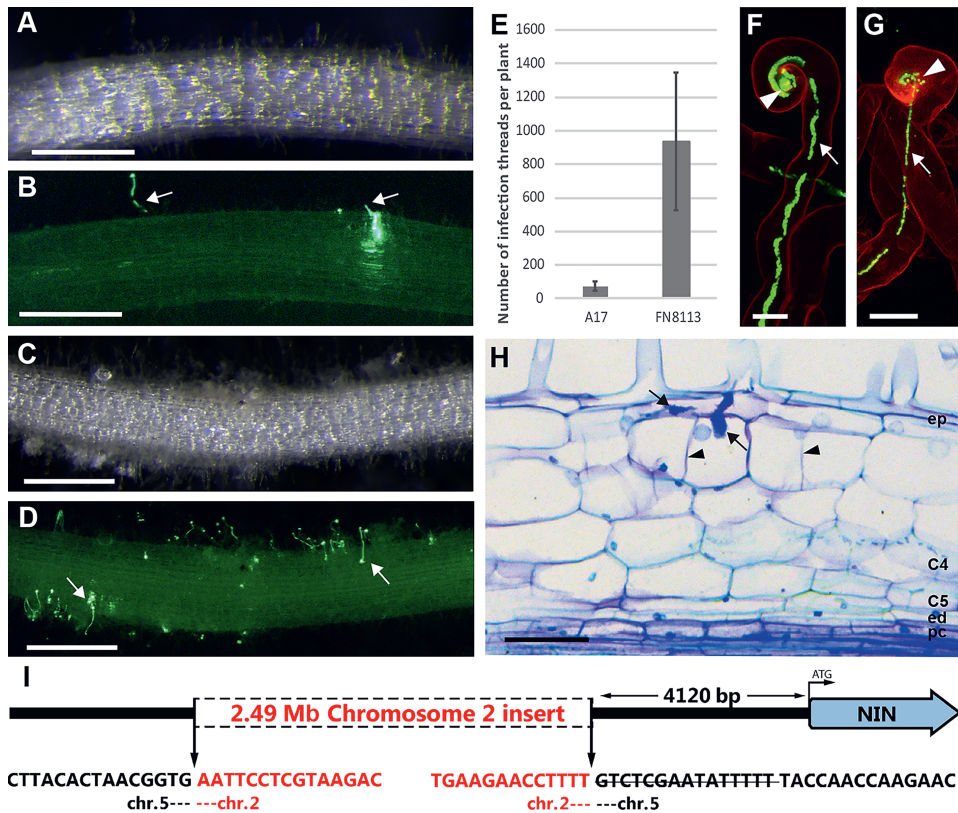
Figure 1. A new Medicago *nin* Mutant *daphne-like* in which Infection and Nodule Organogenesis Are Uncoupled.

(A) to (D) Images of wild-type and mutant roots. These transmitted light stereo macroscopy images (A) and (C) and corresponding green fluorescence stereo macroscopy images (B) and (D) were taken at 3 weeks post inoculation (wpi). Roots of *daphne-like* (FN8113) mutant plants (C) and (D) have an excessive number of infection threads in comparison to wild type A17 roots (A) and (B). Bars= 2 mm. (E) Quantification of infection thread number. The number of infection threads per root was counted at 2 wpi in both A17 roots (n = 12) and FN8113 roots (n = 12). Data are mean ± SD. (F) and (G) Infection thread formation in a mutant and a wild type roots. These confocal images of roots stained with propidium iodide at 1 wpi show that a bacterial colony (arrow head) is formed inside a *daphne-like* curled root hair and an infection thread (arrow) is initiated (F) like in a wild type root hair (G). *S. meliloti* containing constitutively expressed GFP was used as inoculum. ep, epidermis; C4, C5, cortical cell layers 4 and 5; ed, endodermis; pc, pericycle. Bars = 10 μm (H) Longitudinal plastic section of *daphne-like* root at 3 wpi. The section stained with toluidine blue displays infection thread (arrow). The infection threads in a mutant can occasionally reach cortical cell layers and induce some cell divisions (arrowhead). Bars = 50 μm. (I) Schematic representation of the chromosome translocation at the *NIN* locus in the *daphne-like* mutant. The strikethrough indicates a 15 bp deleted sequence.

28

The 5kb Upstream Region of Medicago *NIN* Contains Discrete Regulatory Sequences that Affect Root Hair Curling and Infection

The phenotype of *daphne-like* strongly suggests that *NIN* regulatory sequences required for primordium formation are located more than 4120 bp upstream of the *NIN* start codon. In addition, this phenotype indicates that the regulatory sequences located within this 4120-bp region are sufficient for proper root hair curling and infection thread formation. We tested this by using a construct containing the 5-kb region upstream of the start codon to drive expression of *NIN*. We introduced this construct, *ProNIN$_{5kb}$:NIN*, into Medicago *nin-1* (null mutant, Marsh et al., 2007) roots by *Agrobacterium rhizogenes*-mediated root transformation. At 4 wpi, 41 of 44 analyzed transgenic roots showed excessive infection thread formation (Figure 2B to 2D). Despite the numerous infections, these roots did not form nodules, except for one root on which four nodules were observed. As other transgenic roots of this composite plant have no nodules, we assume that this is caused by transgene insertion. Longitudinal sections of infected transgenic roots confirmed that cell divisions were not induced in the pericycle, endodermis and inner cortical cell layers (Figure 2A). Infection threads were arrested in the epidermis, but occasionally some of these reached the cortex. Thus, the 5 kb promoter region is sufficient for infection thread formation, but it lacks regulatory sequences for primordium formation.

Interestingly, a single putative CYCLOPS/IPD3 binding site is located about –3 kb upstream of the start codon (Figures 3 and 4A, Supplemental Figure 1, Supplemental Table 1) (Singh et al., 2014). We therefore checked whether the function of NIN in the epidermis fully depends on this putative CYCLOPS binding site by using the –2.2 kb region (Figure 4A) to drive *NIN* expression. The *ProNIN$_{2.2kb}$:NIN* construct was introduced into *nin-1* by *A. rhizogenes*-mediated root transformation. The *nin-1* null mutant has excessive root hair curling but fails to form infection threads (Figure 2H to 2J). Although all 37 analysed transgenic roots at 4 wpi showed tight root hair curls that enclosed bacterial colonies, infection threads were rare (Figure 2E to 2G). Of 298 curled root hairs in *ProNIN$_{2.2kb}$:NIN* transgenic roots containing a bacterial colony, only ~3% had an infection thread. This shows that root hair curling and establishment of infection chambers do not rely on the putative CYCLOPS binding site. By contrast, ~70% of these curled root hairs (n = 324) formed infection threads in *ProNIN$_{5kb}$:NIN* transgenic roots.
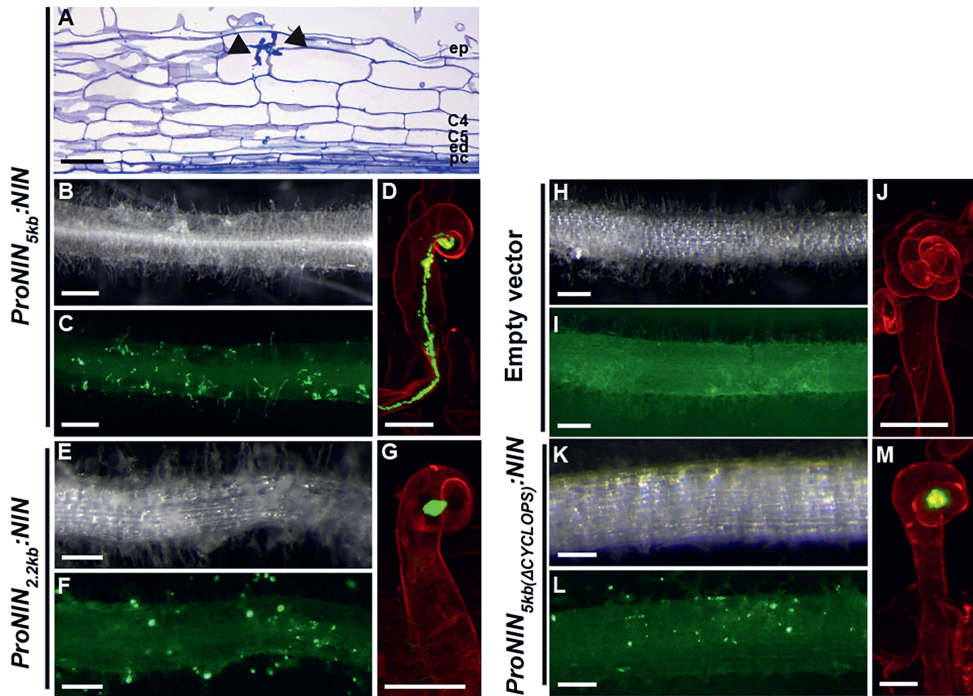
Figure 2. The Infection Process in Medicago *nin-1* Roots Is Partially Rescued by Introducing *ProNIN$_{5kb}$:NIN* or *ProNIN$_{2.2kb}$:NIN*.

(A) to **(D)** Phenotype of *nin-1* roots transformed with *ProNIN$_{5kb}$:NIN* at 4 weeks post inoculation (wpi) with *S. meliloti* constitutively expressing GFP. The transmitted light microscopy image of longitudinal plastic section of transgenic root stained with toluidine blue displays infection threads (arrow) that occasionally can reach cortical cell layers **(A).** Bar = 50 μm. The transmitted light stereo macroscopy image **(B)** and corresponding green fluorescence stereo macroscopy image **(C)** show excessive infection thread formation in transgenic root. Bars = 2mm. Confocal image of transgenic root stained with propidium iodide shows infection thread initiated in the curled root hair **(D).** Bar = 10 μm. **(E)** to **(G)** Phenotype of *nin-1* roots transformed with *ProNIN$_{2.2kb}$:NIN* at 4 wpi with *S. meliloti* constitutively expressing GFP. The transmitted light stereo macroscopy image **(E)** and corresponding green fluorescence stereo macroscopy image **(F)** display numerous curled root hairs with bacterial colonies in transgenic root. Bars = 2mm. Confocal image of transgenic root stained with propidium iodide **(G).** Colonies are formed inside the chamber of the root hair curl, but infection thread is not developed. Bar = 10 μm. **(H)** to **(J)** Phenotype of *nin-1* roots transformed with empty vector at 4 wpi with *S. meliloti* constitutively expressing GFP. The transmitted light stereo macroscopy image **(H)** and corresponding green fluorescence stereo macroscopy image **(I)** show that transgenic root forms neither infection threads nor bacterial colonies. Bars = 2mm. Transgenic root stained with propidium iodide shows excessive root hair curling **(J).**

These results indicate that the –5 kb to –2.2 kb region contains regulatory sequences that are critical for infection thread formation. The observed phenotype is reminiscent of that of Lotus and Medicago *cyclops-3/ipd3-2* mutants (Yano et al., 2008; Horváth et al., 2011), which do not form infection threads but show formation of bacterial colonies in tightly curled root hairs. Therefore, the –2.2 kb region can activate *NIN* expression in the epidermis, and the expression level is sufficient for tight root hair curling, allowing rhizobia to form a colony inside the curl. However, additional regulatory sequences located between –5 kb and –2.2 kb, probably involving the putative CYCLOPS binding site, are required for efficient infection thread formation. To test this, we analysed *nin-1* roots transformed with *NIN* driven by the –5 kb promoter in which the putative CYCLOPS binding site was deleted ($ProNIN_{5kb(\Delta cyclops)}$:*NIN*) (Figure 2K to 2M). Due to this mutation, the number of curled root hairs with a colony (similar in size to the one formed in wild type roots) that initiated an infection thread dropped from 70% to 7% (n = 434). This shows that the putative CYCLOPS-binding site within the *NIN* promoter is essential for efficient infection thread formation.

A Conserved Region with Putative Cytokinin Response Elements is Located ~18 kb Upstream of the *NIN* Coding Region in *Medicago truncatula*

The *daphne-like* mutant, as well as *nin-1* transformed with $ProNIN_{5kb}$:*NIN,* can induce formation of infection threads but not nodule primordia. Based on this, we hypothesized that the regulatory elements required for NIN-induced nodule primordium formation are located upstream of –5 kb. This resembles the Lotus *daphne* mutant, which contains a chromosomal insertion at ~7 kb upstream of the *NIN* start codon. Therefore, we expected that such remote regulatory regions would be conserved in Lotus and Medicago and probably in other legumes. To test this, we compared the genomic DNA sequences spanning from the *NIN* coding region to the first upstream gene in eight legume species (*Medicago*

Continued

Bar = 10 µm. **(K)** to **(M)** Phenotype of *nin-1* roots transformed with $ProNIN_{5kb(\Delta cyclops)}$:*NIN* at 4 wpi with *S. meliloti* constitutively expressing GFP. The transmitted light stereo macroscopy image **(K)** and corresponding green fluorescence stereo macroscopy image **(L)** display many bacterial colonies in transgenic root hairs. Bars = 2mm. Confocal image of transgenic root stained with propidium iodide shows that colony is formed inside the chamber of the root hair curl but infection thread is not initiated **(M)**. Bar = 10 µm.
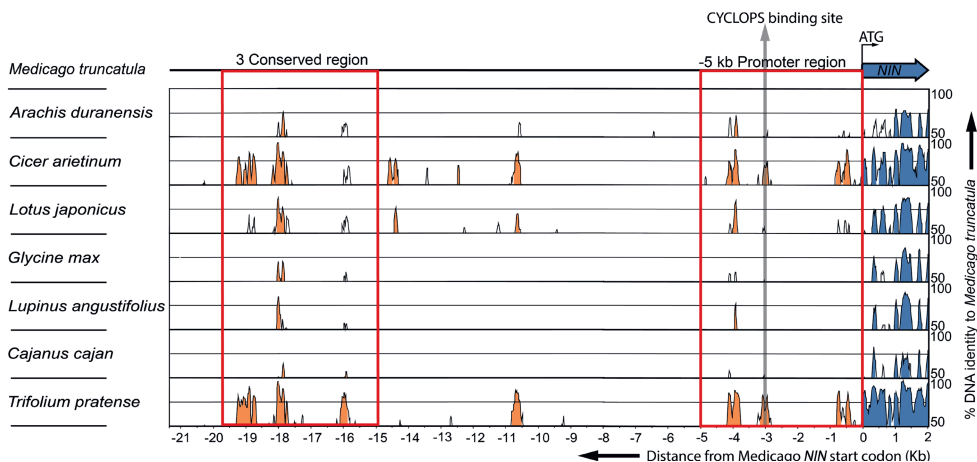
Figure 3. mVISTA Alignment of Genomic DNA Sequences Including 2 kb from the Start Codon of the *NIN* Gene and 5'-Upstream Regions from Eight Legume Species.

Peaks indicate the level of identity with Medicago on a scale of 50%- 100%. Identities lower than 50% were not scored. The figure represents 2 kb of each *NIN* sequence downstream of the start codon (blue) and 5'- upstream DNA sequence (orange). The red rectangle on the right next to the *NIN* coding sequence indicates the –5 kb promoter region. The red rectangle on the left indicates the 3 conserved regions (3C). The grey arrow shows the putative CYCLOPS binding site.

truncatula, *Lotus japonicus, Arachis duranensis*, *Cicer arietinum*, *Glycine max*, *Lupinus angustifolius*, *Cajanus cajan* and *Trifolium pratense*). Based on the high level of identity (50%- 100%) among all these species we identified DNA sequences with three conserved regions (3C) far upstream of the *NIN* start codon (Figure 3 and Supplemental Table 1). In Medicago, 3C is located 15 to 20 kb upstream of the *NIN* start codon, and in Lotus it is located between 42 to 49 kb upstream (Supplemental Table 1). The levels of identity in conserved regions of 3C are similar to that of the *NIN* coding region (Figure 3). The second region in 3C is most conserved and includes about ten putative B-type cytokinin signalling RESPONSE REGULATOR (RR) binding sites (Figure 3 and Supplemental Figure 2) (Sheen, 2002; Heyl and Schmülling, 2003; Hosoda et al., 2002; Imamura et al., 2003). Therefore, we named this middle region the CE region (for cytokinin response elements-containing). Because cytokinin signalling is essential for nodule organogenesis and induction of *NIN* expression, the CE region may be involved in regulation of *NIN* expression during initiation of nodule primordium formation.

The CE Region Contains Regulatory Elements Required for Nodule Organogenesis

To determine whether 3C (~4 kb) contains regulatory sequences necessary for nodule primordium formation, we fused it to the (upstream) –5 kb region (*ProNIN$_{3C-5kb}$:NIN*), as the latter is sufficient for infection. *ProNIN$_{3C-5kb}$:NIN* was introduced into *nin-1* by *A. rhizogenes*-mediated root transformation. 21 of 26 analysed transgenic roots (at 4 wpi) formed, on average, eight nodules per root (Figure 4B). As the CE region (~1 kb) contains several putative cytokinin response elements, we tested whether this region is sufficient to trigger primordium formation. To this end, we transformed *nin-1* with the CE region fused to the –5 kb region driving *NIN* (*ProNIN$_{CE-5kb}$:NIN*). This resulted in 18 out of 37 transgenic roots forming on average eight nodules per root (Figure 4B). This demonstrates that the CE region contains regulatory sequences that are required for primordium formation. Furthermore, the number of nodules formed on *ProNIN$_{CE-5kb}$:NIN* expressing roots was similar to the number on wild type roots transformed with an empty vector control (Figure 4B). This suggests that the autoregulation of nodulation mechanism is also activated (Soyano et al., 2014). In addition, 12 out of 22 *nin-1* roots transformed with *ProNIN$_{CE-5kb}$:NIN* displayed wild type like infection thread numbers, indicating that the excessive infection phenotype can be partially rescued by including the CE region in *NIN* promoter.

Normal nodules are pink due to the presence of leghemoglobin, which keeps the oxygen level low in the infected cells of the fixation zone so that the anaerobic process of nitrogen fixation can proceed. Pink nodules were formed on *nin-1* roots transformed with either *ProNIN$_{3C-5kb}$:NIN* or *ProNIN$_{CE-5kb}$:NIN*. Longitudinal sections of these nodules showed a zonation similar to wild type nodules: a meristem at the apex; the infection zone, where rhizobia are released from the infection thread and subsequently divide and begin to enlarge; and the fixation zone, where rhizobia have reached their fully enlarged shape and are able to fix nitrogen (Figure 5A to 5D). Nodules formed by inoculation with *S. meliloti* carrying the *PronifH:GFP* reporter, showed that nitrogenase *nifH* was expressed in the fixation zone, confirming that these nodules are functional (Supplemental Figure 3). Thus, CE in combination with the –5 kb region is sufficient to induce wild type-like nodule organogenesis.
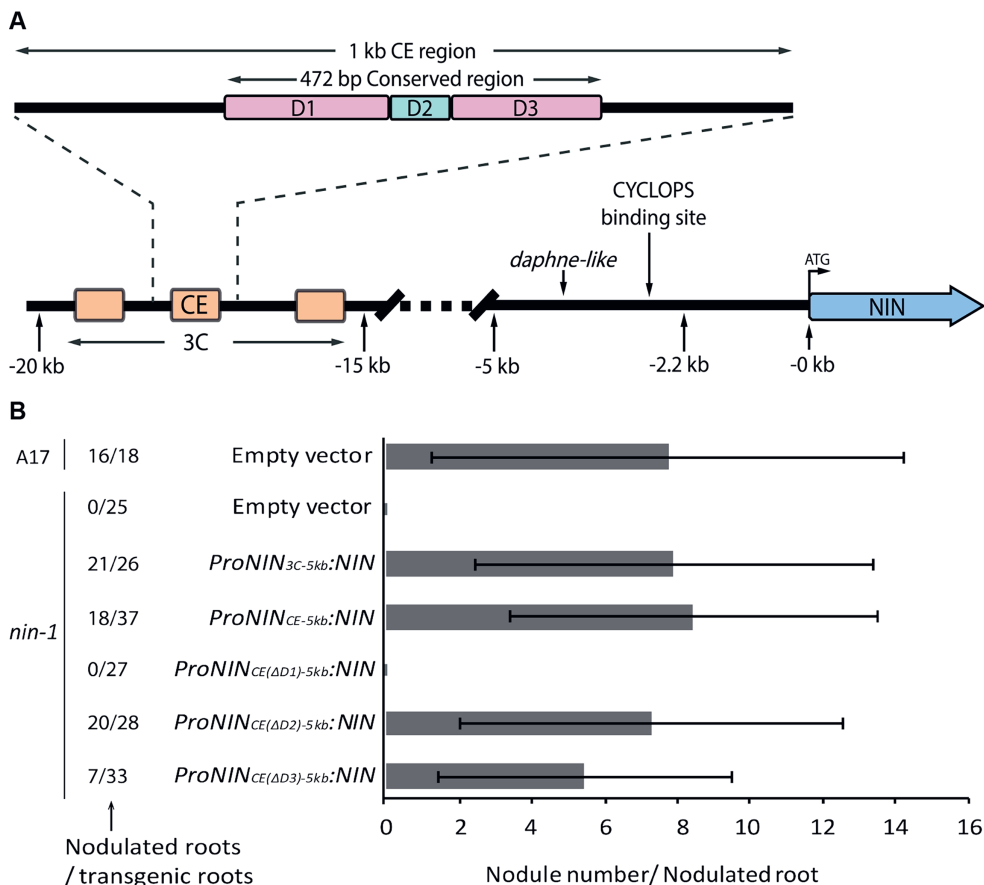
Figure 4. The CE Region in the *NIN* Promoter Is Essential for Nodule Organogenesis.

(A) Schematic illustration of the Medicago *truncatula NIN* promoter. Three conserved remote regions (3C) (orange) were identified among eight legume species studied here. The second region of 3C is most conserved, it includes about ten putative B-type cytokinin signalling RESPONSE REGULATOR (RR) binding sites and named the CE region (for cytokinin response elements-containing). CE region in its turn contains a highly conserved 472 bp sequence that was divided into three parts named domains D1, D2 and D3. D1 and D3 (purple) contain six and three putative cytokinin response elements, respectively, whereas D2 (green) contains a putative AP2-binding site and a single putative cytokinin response element. **(B)** Number of nodules formed on wild type (A17) roots transformed with empty vector and *nin-1* roots transformed with the constructs carrying the *NIN* gene driven by different parts of the *NIN* promoter as indicated. *S. meliloti* containing constitutively expressed GFP was used as inoculum. Nodule numbers per nodulated root were counted at 4 wpi. Data are means ± SD.
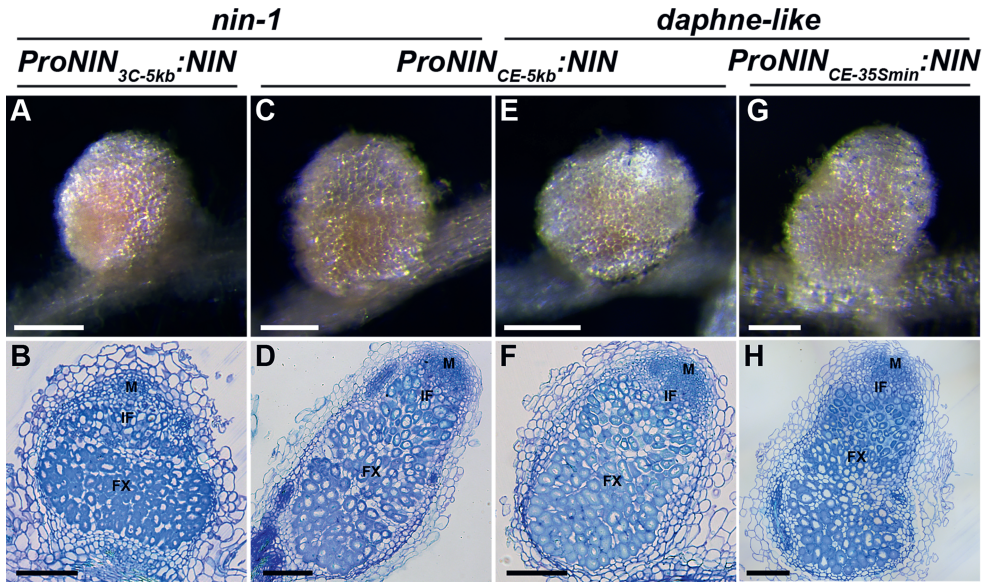
Figure 5. Non-Nodulating Phenotypes of *nin-1* and *daphne-like* are Rescued by *A. rhizogenes*-mediated Transformation with *ProNIN_{3C-5kb}:NIN*, *ProNIN_{CE-5kb}:NIN* and *ProNIN_{CE-35Smin}:NIN*.

(A) to **(D)** Nodules formed on *nin-1* roots transformed with the indicated constructs at 4 wpi with *S. meliloti*. Transmitted light macroscopy images of nodules **(A)** and **(C)**. Nodules are pink due to the presence of leghemoglobin. Bars = 2mm. Longitudinal plastic sections of these nodules stained with toluidine blue display normal zonation. **(B)** and **(D)**. M, meristem; IF, infection zone; FX, fixation zone. Bars = 200 µm. **(E)** *to* **(H)** Nodules formed on *daphne-like* roots transformed with the indicated constructs at 4 wpi with *S. meliloti*. Transmitted light macroscopy images of nodules **(E)** and **(G)**. Bars = 2mm. Longitudinal plastic sections of these nodules stained with toluidine blue display normal zonation **(F)** and **(H)**. M, meristem; IF, infection zone; FX, fixation zone. Bars = 200 µm.

NIN Expression Cannot be Induced by Cytokinin in the *daphne-like* Mutant

Because the *ProNIN_{CE-5kb}:NIN* construct can fully restore nodulation ability in *nin-1* roots, we used it to verify that *daphne-like* is indeed a *nin* allele. Therefore, *daphne-like* was transformed with *ProNIN_{CE-5kb}:NIN* (Figures 5E and 5F), and 15 of 17 transgenic roots analyzed at 4 wpi formed on average about seven nodules per root. The excessive infection phenotype in the *daphne-like* background was rescued by *ProNIN_{CE-5kb}:NIN* in 11 of these 17 transgenic roots. This result shows that *daphne-like* is a *nin* mutant. Its phenotype is most likely caused by the 2.49 Mbp insertion by which the CE region is positioned too far away from

the transcription start to contribute to the correct expression of *NIN* for nodule primordium formation.

To test whether the CE region is sufficient to complement nodule organogenesis in *daphne-like*, we used a minimal –46 bp CaMV 35S promoter (Benfey and Chua, 1990) fused to the CE region *(ProNIN$_{CE-35Smin}$:NIN)* (Figures 5G and 5H). We found that 37 out of 45 transgenic *daphne-like* roots had formed on average four nodules per root at 4 wpi. This indicates not only that the CE region is sufficient to induce nodule organogenesis, but also that, in combination with the –5 kb region, more nodules (about seven per root) can be formed. The ability to form nodules can be rescued in *daphne-like* by the CE region driving *NIN* expression. Therefore, it is likely that the CE region in *daphne-like* cannot regulate the expression of *NIN*. Because the CE region contains several putative response elements, we hypothesised that *NIN* expression would not be induced by cytokinin in *daphne-like*. To test this, we compared the induction of *NIN* expression by cytokinin vs. water (as control) in wild type (A17) and *daphne-like.* We found that 16 h after 10-7M benzylaminopurine (BAP) application, *NIN* expression level increased 37 fold compared with the control, and *NF-YA1* expression level increased over a hundred fold in wild-type, while both *NIN* and *NF-YA1* expression levels in *daphne-like* were not changed (Figures 6A and 6B). This suggests that the CE region is required for the induction of *NIN* expression by cytokinin.
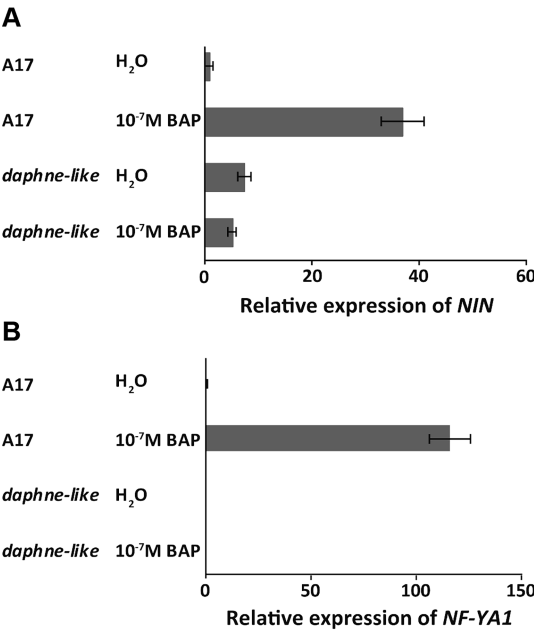
Figure 6. *NIN* and *NF-YA1* Expression Cannot Be Induced by Cytokinin Application in *daphne-like*.

Quantitative real-time RT-PCR analysis of *NIN* **(A)** and *NF-YA1* **(B)** expression in wild type (A17) and *daphne-like* roots after 10^{-7} M benzylaminopurine (BAP) or water application for 16 h. Data are means ± SEM of three biological replicates.

## A Domain with Six Putative Cytokinin Response Elements is Essential for Nodule Primordium Formation

Because cytokinin is known to be a positive regulator of nodule primordium formation (Suzaki et al., 2013), we tested whether the putative cytokinin response elements within the CE region are essential for primordium formation. To this end, we made several deletions in the CE region, which contains a 472-bp region that is highly conserved in all eight legume species studied here (Figures 3 and 4A and Supplemental Figure 2). We divided this 472-bp region into three parts named domains one to three (D1 to D3). D1 and D3 contain six and three putative cytokinin response elements, respectively, whereas domain 2 (D2) contains a putative AP2-binding site as well as a single cytokinin response element (Figure 4A and Supplemental Figure 2).

Several studies have shown that transcription factors of the AP2 family, including ERN (ethylene response factor required for nodulation) are involved in regulating nodulation (Andriankaja et al., 2007; Middleton et al., 2007; Wang et al., 2014). To investigate their respective contribution to nodule primordium formation, the D1, D2 or D3 regions were separately deleted from the 1-kb CE region (Figure 4A and Supplemental Figure 2), and the modified CE regions were fused to the –5 kb region to drive *NIN* expression. These three constructs were introduced into *nin-1* by *A. rhizogenes*-mediated root transformation. Our results show that deletion of D1 eliminated nodulation ability (Figure 4B and Supplemental Figure 4A), whereas deletion of D2 had no significant effect on nodulation (Figure 4B and Supplemental Figure 4B). Deletion of D3 caused a reduction of the relative number of roots with nodules from 49% to 21% and also reduced the average nodule number per root from eight to five (Figure 4B and Supplemental Figure 4C). These results show that regulatory sequences in D1 are essential for NIN-regulated nodule primordium formation and suggest that the putative cytokinin response elements within D1 are likely responsible. In contrast, the putative AP2 binding site in D2 is not essential for nodule organogenesis.

Induction of *NIN* expression in Inner Root Cell Layers Occurs in a Non-Cell-Autonomous Manner

It was shown that the 2.2 kb upstream region of Medicago *NIN* is activated in the epidermis, 24 h after Nod factor application (Vernié et al., 2015). Because this promoter region lacks the regulatory sequences required for nodule organogenesis, we wanted to determine whether expression of *NIN* is induced

in inner root cell layers during primordium formation. We studied the localization pattern of *NIN* mRNA in nodule primordia via *in situ* hybridization. Plants were grown on plates and spot inoculated with *S. meliloti*. We analyzed a primordial stage at 2 dpi in which the pericycle cells have divided and some anticlinal divisions have occurred in the inner cortical cell layers C4 and C5 (Figure 7A and 7B). We also analyzed cells at a slightly later stage at 3 dpi when cortical cells have divided more extensively (Figure 7C and 7D). At both stages, the infection thread had not yet reached the primordia. At the younger stage, *NIN*
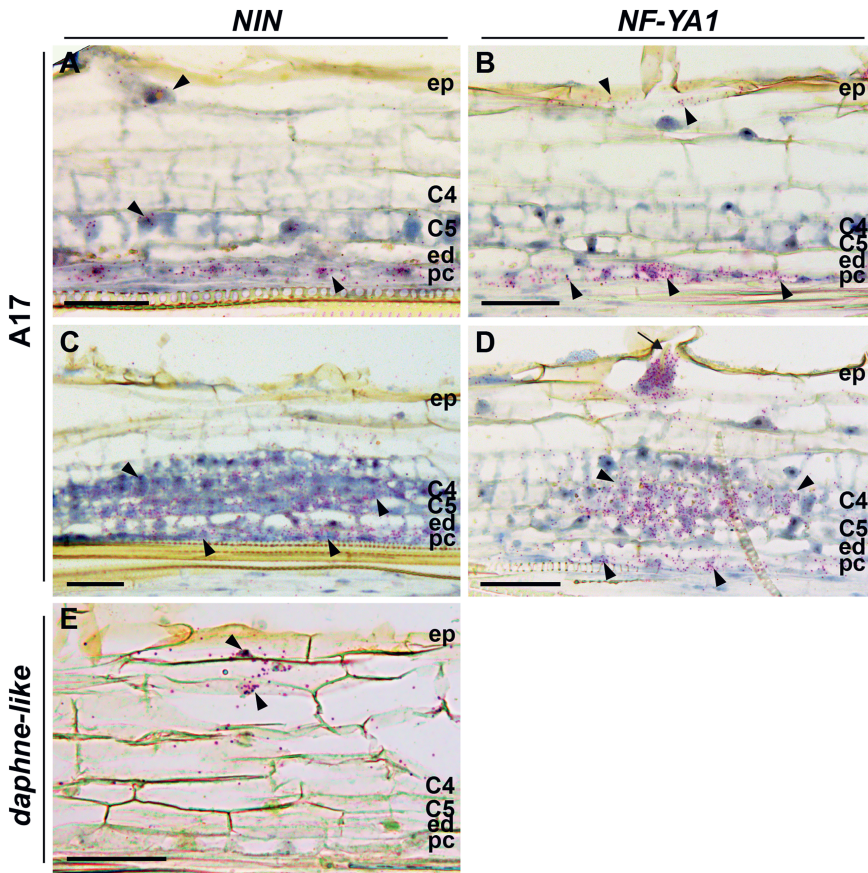
Figure 7. *NIN* and *NF-YA1* Expression Patterns in Medicago Wild Type (A17) Nodule Primordia and *daphne-like* Mutant.

(A) to **(D)** RNA *in situ* localization of *NIN* **(A)** and **(C)** and *NF-YA1* **(B)** and **(D) in nodule primordia** at 2 days post inoculation (dpi) **(A)** and **(B)** and at 3 dpi **(C)** and **(D).** The arrow indicates an infection thread. **(E)** RNA *in situ* localization of *NIN* in roots of the *daphne-like* mutant at 2 dpi. Hybridization signals are visible as red dots (arrowheads). ep, epidermis; C4, C5, cortical cell layers 4 and 5; ed, endodermis; pc, pericycle. Bars = 50 µm.

mRNA occurred in pericycle and epidermis, but it was hardly detectable in the divided cortical cells (Figure 7A). The highest expression level occurred in the pericycle-derived cells. At the stage, when cortical cells have divided more extensively, the expression level of *NIN* in cortex derived cells was similar to that in the pericycle (Figure 7C). This shows that expression of *NIN* was first strongly induced in the pericycle after which it extends to the other inner cell layers. *NF-YA1* is a known direct target of NIN (Soyano et al., 2013). Like *NIN,* it is expressed in the epidermis, where it regulates rhizobial infection (Laporte et al., 2014). To test whether NIN might also regulate *NF-YA1* expression in the primordia, we performed RNA *in situ* hybridization using *NF-YA1* as a probe. This analysis showed that *NF-YA1* expression is similar to *NIN* expression, as it is also first induced in pericycle and most likely cortical cell division precedes *NIN* and *NF-YA1* expression (Figure 7B and 7D). This suggest that *NF-YA1* is regulated by *NIN* in both pericycle and other nodule primordium cells.

Therefore, rhizobia present in the epidermis induce *NIN* and *NF-YA1* expression in the pericycle-derived cells. Furthermore, because Nod factors are immobile molecules (Goedhart et al., 2000) that do not diffuse to the inner cell layers, *NIN* and *NF-YA* expression in the inner cell layer is most likely induced by a mobile signal generated in the epidermis where Nod factor signaling takes place.

The CE Region is Required for Induction of *NIN* Expression in the Pericycle

We wanted to test whether the CE region is required for *NIN* expression in the inner cell layers, we compared expression patterns of *ProNIN$_{CE-5kb}$:GUS* and *ProNIN$_{5kb}$:GUS* in roots. We first introduced these constructs into wild type Medicago (A17) roots *by A. rhizogenes* mediated transformation. We analyzed an early stage of primordium development when pericycle cells have divided and some anticlinal divisions have occurred in the inner cortical cell layers similar to the early stage tested for *in situ*. Both constructs were expressed in epidermis, pericycle, and endodermis and a lower signal was detected in some cortical cells (Figure 8A and 8B). This result is surprising considering that *ProNIN$_{5kb}$:NIN* is not sufficient for primordium formation in the *nin-1* background. Therefore, we hypothesized that expression of *ProNIN$_{5kb}$:GUS* in inner cell layers is induced by endogenous NIN that is produced in the wild-type background. This implies that *NIN* expression in the inner layers is regulated by a positive feedback loop involving NIN itself, and that the essential *cis*-regulatory elements required for this are located in the –5 kb promoter region.
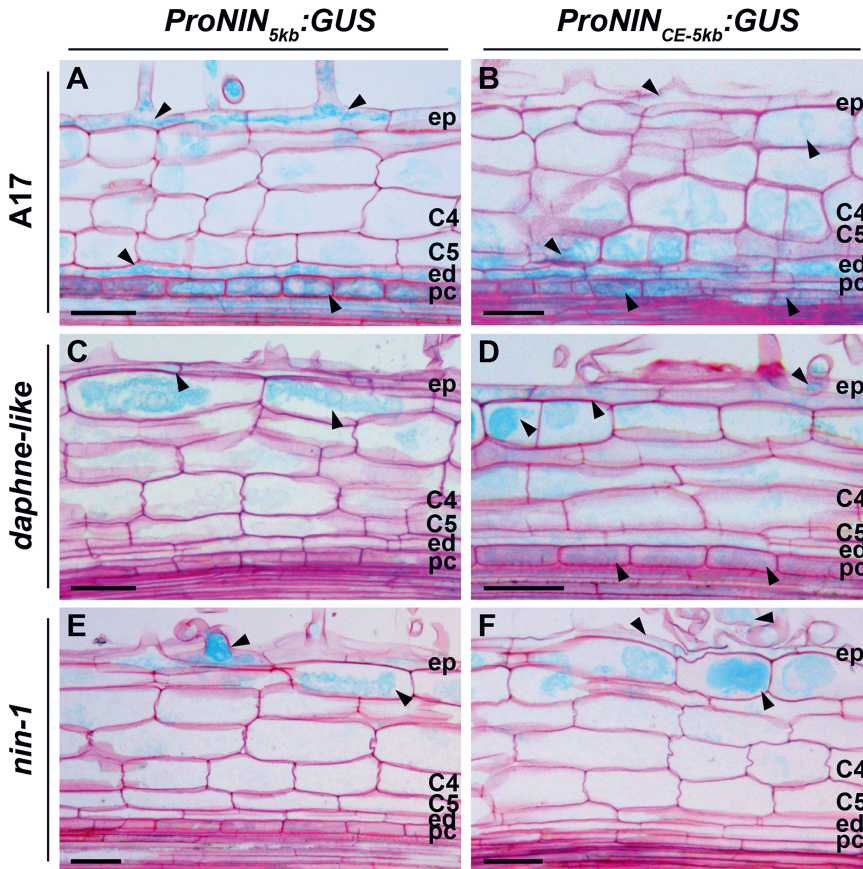
Figure 8. The CE Region Is Required for Rhizobium-Induced *NIN* Expression in the Pericycle.

(A) to **(F)** Tissue-specific *ProNIN$_{5kb}$:GUS* and *ProNIN$_{CE-5kb}$:GUS* expression patterns in wild type and *nin* mutants at 2 dpi. Arrowheads indicate GUS expression (light blue) in wild type **(A)** and **(B)**, in *daphne-like* **(C)** and **(D)** and in *nin-1* **(E)** and **(F)** roots. ep, epidermis; C4, C5, cortical cell layers 4 and 5; ed, endodermis; pc, pericycle. Bars = 50 µm.

To test this hypothesis, we introduced *ProNIN$_{CE-5kb}$:GUS* and *ProNIN$_{5kb}$:GUS* into *daphne-like* by *A. rhizogenes* mediated transformation. In *daphne-like*, infection threads can be formed indicating that *NIN* is induced in the epidermis and the production of the mobile signal might not be affected. However, nodule primordium formation is impaired, indicating there is no NIN production in the inner cell layers. Indeed, *ProNIN$_{5kb}$:GUS* transgenic roots showed *GUS* expression only in epidermis and outer cortex (Figure 8C), whereas no expression was observed in the pericycle cells. In contrast, *ProNIN$_{CE-5kb}$:GUS*

transgenic roots showed *GUS* expression in epidermis, outer cortex and in the pericycle (Figure 8D). In this case, cell division was not induced in the pericycle, due to the absence of NIN. Taken together, these results demonstrate that the CE region regulates *NIN* expression in the pericycle prior to cell division in wild type roots. This means that the CE region is required for the initial induction of *NIN* expression in the pericycle. In addition, the expression of *ProNIN$_{CE-5kb}$:GUS* in the pericycle of *daphne-like* is weak which is consistent with the involvement of NIN in a feedback loop by (directly or indirectly) positively regulating its own expression.

To further demonstrate that the CE region is required for *NIN* expression in the pericycle, we studied *NIN* expression in *daphne-like* roots using RNA *in situ* hybridization at 2 dpi with rhizobia. In contrast to wild type (Figure 7A), *NIN* is expressed in the epidermis and outer cortex but not in the pericycle (Figure 7E). This result supports the idea that that CE region is required for *NIN* expression in the pericycle.

Induction of *NIN* in the Pericycle Depends on *NIN* Expression in the Epidermis

It is likely that a mobile signal generated by Nod factor signaling in the epidermis induces *NIN* expression in the pericycle. If true, *NIN* expression in the pericycle would depend on *NIN* induction in the epidermis. To test this, we introduced *ProNIN$_{CE-5kb}$:GUS* and *ProNIN$_{5kb}$:GUS* into *nin-1* by hairy root transformation. In both cases, GUS was present only in the epidermis and outer cortex, and not in the pericycle at 3 dpi (Figure 8E and 8F). This suggests that NIN is required in the epidermis, probably for the generation of the mobile signal, in order to induce *NIN* expression in pericycle cells.

CRE1 and *RR1* are Expressed in the Pericycle of Uninoculated roots

Rhizobium-induced *NIN* expression in the pericycle is dependent on the CE region and formation of precedes nodule primordia. The occurrence of multiple B-type RR response regulatory elements in the CE region suggests that the cytokinin signalling machinery is important for *NIN* transcriptional activation in the pericycle. To examine this, we determined the expression pattern of the cytokinin receptor *CRE1* and its putative target the B-type RESPONSE REGULATOR RR1, which is known to be expressed during nodule formation (Gonzalez-Rizzo et al., 2006). Using RNA *in situ* hybridization, we found that
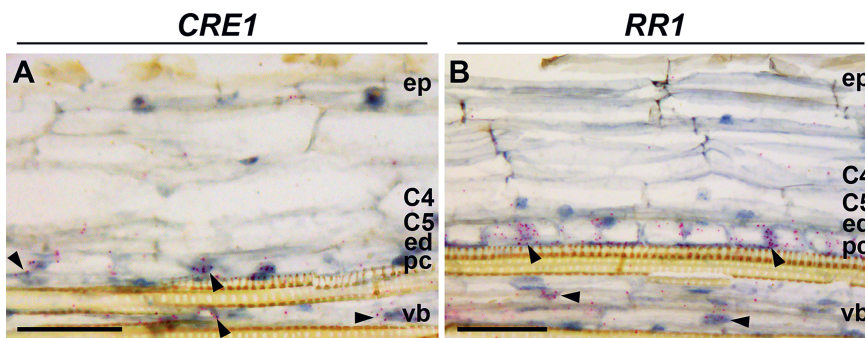
Figure 9. *CRE1* and *RR1* are Expressed in the Pericycle of Uninoculated roots.

RNA *in situ* localization of *CRE1* **(A)** and *RR1* **(B)** in susceptible zone of uninoculated roots. For *in situ* hybridization root tips of ~1 cm of 4 days old seedlings were used. Hybridization signals are visible as red dots (arrowhead). ep, epidermis; C4, C5, cortical cell layers 4 and 5; ed, endodermis; pc, pericycle; vb, vascular bundle. Bars = 50 μm

CRE1 is actively transcribed in pericycle and vasculature cells of uninoculated roots, but not in endodermal or cortical cells (Figure 9A). Also, mRNA of the B-type RR1 was present at the highest level in pericycle cells, and to a lower extent in root vasculature cells (Figure 9B). Therefore, both *CRE1* and *RR1* have been already expressed in the pericycle by the time rhizobial signalling starts, suggesting that, initially, only this layer is responsive to cytokinin.

Discussion

In this study, we show that remote upstream regulatory sequences (the CE region) are required for proper regulation of *NIN* expression and *Medicago truncatula* nodule organogenesis. By contrast, regulatory sequences required for the infection process are located within a 5 kb region directly upstream of the start codon. The CE region contains several putative cytokinin response elements and domain 1 (D1), which contains six of these elements, is essential for nodule primordia formation. The CE region appears to be important for cytokinin induced expression of *NIN,* as *daphne-like* has lost this ability. Formation of nodule primordium initiates with *NIN* induction in the pericycle, and subsequently it extends to the cortical cells. The fact that *CRE1* and *RR1* are expressed in the pericycle supports the idea that cytokinin perception is necessary for the induction of *NIN* at the start of primordium formation.

In animals, many genes have been identified that are regulated by remote *cis*-regulatory elements that can be megabases away from the transcription start site. By contrast, in plants only a few remote *cis*-regulatory sequences are known (Shlyueva et al., 2014; Weber et al., 2016; Symmons and Spitz, 2013). One of the best characterized remote *cis*-regulatory sequence is the enhancer of *booster1* (*b1*) in *Zea mays,* which is located 100 kb upstream of the gene (Stam et al., 2002). It has been shown that gene activation by remote enhancers can be associated with chromatin loop formation that brings the enhancer in close proximity to the promoter, a process that can facilitate assembly of transcription complexes (Cook, 2003; Nolis et al., 2009; Deng et al., 2012). The distance between the CE region and the transcription start site varies in the legume species studied here. In *L. angustifolius* it is about –7 kb, whereas in Lotus it is about –45 kb. We demonstrated that the CE region fused to the –5-kb promoter can rescue nodule organogenesis in Medicago. This shows that the sequences between the CE region and the –5 kb region are not essential for nodule organogenesis. However, we cannot exclude the possibility that in this region there are regulatory sequences required for fine tuning *NIN* expression.

During the infection process, NIN participates in a mechanism wherein root hair growth stops when a proper curl is formed. Regulatory sequences required for this process are located within the –2.2 kb promoter region. The fact that this region lacks the putative CYCLOPS binding site implies that in addition to CYCLOPS (IPD3 in Medicago), another transcription factor or factors is involved in regulating *NIN* expression in the epidermis. Because this –2.2 kb region is not sufficient for efficient infection thread formation, we assume that the expression level of *NIN* in the epidermis remains below the threshold level required for infection thread formation, whereas this level can be reached by the –5 kb promoter region which includes the putative CYCLOPS binding site (Figure 10).

We present a model for the regulation of expression of *NIN* in Figure 10. After the rapid induction of *NIN* in the epidermis, *NIN* is subsequently induced in the pericycle. The latter most likely precedes the mitotic activation of pericycle cells. The induction of *NIN* in the pericycle requires the presence of the CE region and involves a positive feedback loop including NIN itself. The proposed feedback loop was based on our observation that expression of *ProNIN$_{5kb}$:GUS* in the Medicago wild type background was induced in nodule primordia, despite the fact that this promoter region is not sufficient to trigger primordium formation. This result is similar to what was found in Lotus where a promoter region of *NIN*
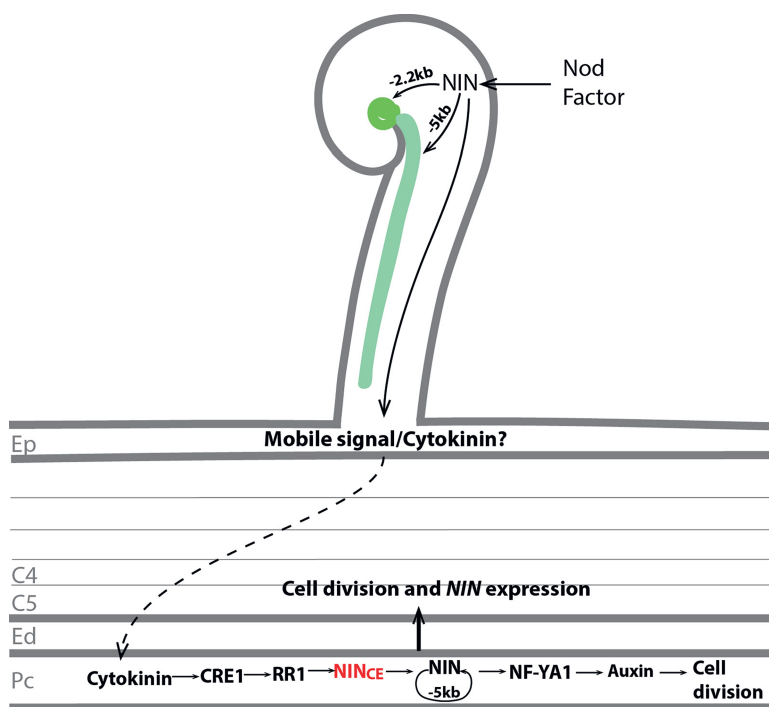
Figure 10. Proposed Model of *NIN* Function during Initiation of Nodule Primordia.

After perception of the Nod factor, *NIN* expression is induced in the epidermis. The −5-kb regulatory region of the *NIN* promoter is sufficient for both tight root hair curling and infection thread formation. By contrast, expression driven by the −2.2 kb region is sufficient only for the tight root hair curling and formation of bacterial colonies inside the curl. A mobile signal is generated in the epidermis in a NIN-dependent manner, and this signal translocates to the pericycle. Whether or not this mobile signal is cytokinin or an unknown signal, it causes cytokinin accumulation in the inner root cell layers. The CRE1 receptor in the pericycle perceives cytokinin and activates the B-type RR1, which further activates *NIN* expression. NIN directly or indirectly regulates its own expression via a positive feedback loop, and the −5 kb promoter region is sufficient for this. NIN directly activates *NF-YA1* expression and stimulates further cell divisions. Later, the NIN-induced response in pericycle contributes to cell division and *NIN* expression in the endodermis and cortical cells. ep, epidermis; C4, C5, cortical cell layers 4 and 5; ed, endodermis; pc, pericycle.

that does not trigger primordium formation was sufficient to drive expression of GUS in primordia (Yoro et al., 2014; Heckmann et al., 2011; Kosuta et al., 2011).

Our conclusion that nodule primordium formation requires the induction of *NIN* expression in inner root layers is consistent with the observation that nodule organogenesis is restored in the Lotus *daphne* mutant by *NIN* driven by a

heterologous Arabidopsis enhancer that is active in endodermis and cortex (Yoro et al., 2014). When we transformed the Medicago *nin* null mutant with the *ProNIN$_{2.2kb}$:NIN* construct, nodule organogenesis was not restored. By contrast, Vernie et al. 2015 reported the formation of nodules on a Medicago *nin* null mutant transformed with a similar construct. However, the nodule number was very low and nodules were observed a long time (50 d) after inoculation. To determine whether these structures are indeed genuine nodules and not simply modified lateral roots, analysis of sections is required.

Deletion of sequences within the CE region, which contains six putative cytokinin response elements, blocks primordium formation. We hypothesize that cytokinin signalling in the pericycle induces *NIN* expression. This hypothesis is supported by the fact that the expression of the cytokinin receptor (*CRE1*) and the B-type response regulator (*RR1*) is observed in the pericycle before rhizobial signalling is initiated. This agrees with a previous study showing that a *CRE1* promoter region driving *GUS* expression is specifically expressed in endodermis/pericycle cells opposite the protoxylem poles (Boivin et al., 2016), the sites where nodule primordia are formed (Heidstra et al., 1997). The involvement of the CE region in cytokinin induced *NIN* expression is indicated by the *daphne-like* mutant, which has lost this ability. However, it remains to be proven that the CE region is sufficient for the cytokinin induced *NIN* expression. Our attempts to show this in *Agrobacterium*-mediated hairy root transformation were inconclusive due to high basal levels of *NIN* expression in the absence of applied cytokinin in hairy roots.

The CE region is conserved in the eight legume species that we studied. They belong to different clades of the legume Papilionoideae subfamily, representing the Genistoids, IRLC, Robinioids, Milletioids and Dalbergioids clades. This suggests that the regulation of *NIN* expression by cytokinin is conserved in this subfamily. After the induction of *NIN* in the pericycle, *NIN* expression extends to the endodermis and inner cortex. In young nodule primordia in which cortical cells have divided anticlinally (Figure 7A and 7B), expression of both *NIN* and *NF-YA1* are highest in pericycle, and it is hardly detectable in the divided cortical and endodermal cells. This suggests that NIN-induced responses in the pericycle contribute to cell division in endodermis and cortical cells (Figure 10). At a later stage of development, *NIN* is expressed in the dividing cortical cells (Figure 7C and 7D). How *NIN* expression is regulated in these cells remains to be studied. Expression of cytokinin biosynthesis genes as well as bioactive cytokinin accumulation is induced by Nod factor signalling (Van Zeijl et al., 2015). As it is a

mobile molecule, it is likely that this results in the accumulation of cytokinin in (at least) the pericycle. Whether cytokinin biosynthesis in the pericycle is triggered by an as yet unknown mobile signal generated in epidermis or whether cytokinin itself is this mobile signal that accumulates in the pericycle is not known.

Cell division in nodule primordia correlates with auxin accumulation, and this occurs before the first cell division (Mathesius et al., 1998; Suzaki et al., 2012). Auxin accumulation (*DR5* expression) depends on NIN, as it does not occur in a *nin* null mutant (Suzaki et al., 2012). Furthermore, ectopic expression of both *NIN* and *NF-YA1* is sufficient to induce abnormal cell division during lateral root development (Soyano et al., 2013), suggesting that their expression causes the local accumulation of auxin. Therefore, we hypothesize that cytokinin signalling in the pericycle triggers *NIN* expression and that this results in the local accumulation of auxin, which subsequently triggers mitotic activity (Figure 10). This is supported by a previous study showing that *STY* genes are targets of NF-YA1 (Hossain et al., 2016). *STY* genes encode transcription factors that have been shown to regulate YUCCA auxin biosynthesis genes in *Arabidopsis* (Eklund et al., 2010; Sohlberg et al., 2006). If this is the case, then during nodule primordium formation, NIN induced *NF-YA* expression in the pericycle might induce the local production of auxin which subsequently induces cell division in pericycle, endodermis and cortex.

Materials and Methods

Plant Material and Growth, Hairy Root Transformation and Inoculation with Rhizobia

Medicago (*Medicago truncatula*) ecotype Jemalong A17 was used as the wild type. *Agrobacterium* msu 440-mediated hairy root transformation was performed according to Limpens et al., 2004. Medicago plants were grown in perlite saturated with low nitrate [0.25 mM Ca(NO3)2] Färhaeus (Fa) medium (Catoira et al., 2000) at 21°C under a 16h light/8h dark regime. After one week of growth, plants were inoculated with *S. meliloti 2011* constitutively expressing GFP or carrying the *PronifH:GFP* reporter (OD600 = 0.1, 1 mL per plant). Plants growing on Fä plates were spot inoculated with 0.5 µL of rhizobium suspension per root.

DNA Constructs

DNA fragments of *NIN* including the 3'UTR and promoter regions were generated by PCR using Medicago genomic DNA as a template and Phusion high-fidelity DNA polymerase (Finnzymes) with the specific primers listed in Supplemental Table 2. The DNA fragments used for pENTR-D-TOPO cloning (Invitrogen) were amplified with forward primers containing an extra 5'-CACC sequence. Forward primers containing an attB4 site (GGGGACAACTTTGTATAGAAAAGTTGNN) and reverse primers with an attB1 site (GGGGACTGCTTTTTTGTACAAACTTGN) were used to generate DNA fragments for cloning into pDONOR P4-P1 by BP recombination (Invitrogen). The forward primers with attB2 (GGGGACAGCTTTCTTGTACAAAGTGGAA) and reverse primers with attB3 (GGGGACAACTTTGTATAATAAAGTTGC) were used to amplify DNA fragments for cloning into pDONOR P2-P3. To generate deletions (D1/D2/D3) in the CE region and deletion of the putative CYCLOPS binding site in the –5 kb region, two rounds of PCR were performed. In the first round, two DNA fragments that are separated by the deletion were amplified with specific primers to introduce a 15 bp overhang (Supplemental Table 2). Subsequently, the PCR products were purified and mixed and 5 µL of this mixture was used as a template in a second round of PCR with ProNIN-CE-F and ProNIN-CE-R or ProNIN-5kb-F and ProNIN-0kb-R primers (Supplemental Table 2). This allowed creation of a single amplicon with a deletion in either the CE or the –5 kb regions. The Entry vectors were recombined into the modified Gateway binary vector pKGW-RR-MGW (Ovchinnikova et al., 2011) using Multisite LR recombination (Invitrogen).

Histological Analysis and Microscopy

Transgenic roots carrying the *ProNIN:GUS* constructs were incubated in GUS buffer [3% sucrose, 10 mM EDTA, 2 mM k-ferrocyanide, 2 mM k-ferricyanide, 0.5mg/mL X-Gluc in 0.1M phosphate buffer (pH = 7)] at 37°C for 1 to 2 h. Embedding of plant tissue in plastic, sectioning and tissue staining were performed as described in Xiao et al., 2014. Sections were analysed using a DM5500B microscope equipped with a DFC425C camera (Leica). Bright-field and fluorescence images of transgenic roots and nodules were taken using a stereo macroscope (M165 FC, Leica). Confocal images were taken with an SP8 (Leica) microscope, using excitation wavelengths of 488 nm and 543 nm for GFP and propidium iodide respectively.

RNA Isolation and qRT-PCR

RNA was isolated from one-week-old A17 and *daphne-like* roots using the EZNA Plant RNA mini kit (Omega Bio-tek, Norcross, GA, USA). For cDNA synthesis, 1 μg of this RNA was used with the iScript cDNA synthesis kit (Bio-Rad). Real-time qPCR was performed in 10 μL reactions using SYBR Green Supermix (Bio-Rad) and a CFX real-time system (Bio-Rad). Gene expression levels were determined using the primers listed in Supplemental Table 2. The gene expression was normalized using *ACTIN2* as a reference gene.

Quantification of Colonies, Infection Threads and Nodules

To quantify the number of curled root hairs containing colonies or infection threads, more than 20 transgenic roots (5 to 10 cm long) were cut into fragments of ~1 cm and randomly selected for counting. To quantify the nodule number per root, 5 to 10 cm long transgenic roots were selected.

RNA *in situ* Hybridization

Medicago roots were fixed with 4% paraformaldehyde mixed with 3% glutaraldehyde in 50 mM phosphate buffer (pH = 7.4) and were then embedded in paraffin (Paraplast X-tra, McCormick Scientific). Root sections (7 μm) were prepared using a RJ2035 microtome (Leica). RNA *in situ* hybridization was conducted using Invitrogen™ ViewRNA™ ISH Tissue 1- Plex Assay kits (ThermoFisher Scientific) and was performed according to the user manual, which can be accecced at https://cdn.panomics.com/. RNA ISH probe sets were designed and synthesized by request at ThermoFisher Scientific. Catalogue numbers of probes for Medicago genes: VF1-20312 for *NIN*, VF1-6000865 for *CRE1*, VF1-6000866 for *RR1* and VF-20311 for *NF-YA1*. A typical probe set consisted of ~20 pairs of oligonucleotide probes (20-nt long) that hybridize to specific regions across the target mRNA. Each probe was composed of a region of ~20 nucleotides, a short linker region, and a tail sequence. The two tail sequences (double Z) together form a site for signal amplification. This design controls increased background by reducing the chance of a nonspecific hybridization event being amplified. For the nodulation specific genes, we used uninoculated roots as a negative control. For ISH with *CRE1* and *RR1* performed on non-inoculated roots of 4-d-old seedlings, we used *ENOD2* (nodule-specific gene) probe set as a negative control. Images were taken with an AU5500B microscope equipped with a DFC425c camera (Leica).

Map-Based Cloning of *daphne-like*

A segregating F2 populations resulting from a cross between FN8113 (cv Jemalong A17) and Jemalong A20 (118 plants) was made. DNA was extracted using a standard CTAB DNA miniprep method (Taylor and Powell, 1982). Initially, simple sequence repeat markers based on Mun et al. 2006 (Mun et al., 2006) were used to determine the global chromosomal location of the FN8113 locus. Subsequently, additional SSR markers were developed for the FN8113 locus on chromosome 5, and were used for chromosome walking. PCR was performed using 100 ng of genomic DNA and was analysed on 2.5 % agarose gels. The SSR marker JH5.17 (Supplemental Table 2) on BAC clone CU424494 showed the closest linkage to the FN8113 locus. No crossovers were found at the distal end of chromosome 5. Next, whole genome sequencing (Illumina Hiseq2000, paired-end) was used to identify mutations in the genomic region identified from the genetic mapping. The genomic sequence of the mutated region is provided in Supplemental Data Set 1. Cleaned DNA sequence reads were mapped against the Medicago genome (Young et al., 2011) using the bwa_ mem algorithm (Li and Durbin, 2010). Clipped reads and mismapped mate pairs revealed an interchromosomal translocation, and this was further confirmed by aligning reads spanning the mutation to the genome using BLASTN (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

Alignment of Upstream Regions of *NIN*

Most of the alignment work used Geneious v8.1.9 (https://www.geneious.com) (Kearse et al., 2012). The Medicago NIN protein sequence was analyzed using custom BLAST databases and Geneious v8.1.9 (Altschul et al., 1990; Kearse et al., 2012). A diverse selection of legume species with a good quality of publicly available genomic sequences were used: *Medicago truncatula* (Young et al., 2011; Tang et al., 2014), *Lotus japonicus* (Sato et al., 2008), *Arachis duranensis* (Bertioli et al., 2016), *Cicer arietinum* (Varshney et al., 2013), *Glycine max* (Schmutz et al., 2010), *Lupinus angustifolius* (Hane et al., 2017), *Cajanus cajan* (Varshney et al., 2012) and *Trifolium pratense* (De Vega et al., 2015). Selected *NIN* scaffolds (Supplemental Table 1) and up to 80 kb of upstream sequence and 10 kb of downstream sequence of *NIN* were extracted. Selected sequences were custom aligned using the mVISTAs web-based alignment tool (http://genome.lbl.gov/vista/mvista) (Frazer et al., 2004). The alignment program selected was the shuffle-lagan global alignment program, which detects rearrangements (Brudno

et al., 2003). In addition to this larger scale alignment, individual alignments were made using MAUVE as a Geneious plugin (Darling et al., 2004) . This better allowed for more precise determination of conserved sequences relative to the *NIN* start codon in all species. A complete overview of detected conserved regions can be found in Supplemental Table 1.

Alignment of CE Regions and Prediction of Binding Sites

Detected conserved sequences of CE regions for selected scaffolds (Supplemental Table 1)were aligned using MAFFTv7.017 as Geneious plugin (Katoh, 2002). Conservedbinding sites were predicted by using PlantPAN2.0 (Chow et al., 2016). Some sites were manually added based on homology with known putative B-type RR binding sequences (Heyl and Schmülling, 2003; Hosoda et al., 2002b; Imamura et al., 2003).

Accession Numbers

Sequence data from this article can be found in the GenBank/EMBL or Mt4.0v1 databases under the following accession numbers: *NIN* (Medtr5g099060), *CRE1* (Medtr8g106150), *NF-YA1* (Medtr1g056530), and *RR1* (Medtr3g102600).

Acknowledgements

Author contributions

J.L., L.R., O.K., R.G. and T.B. designed the research. J.L., L.R., O.K., E.L., T.v.d.M., R.v.V.and W.K. performed research and analyzed data. E.L., R.v.V., R.C. and Y.C. identified *daphne-like* mutant. J.L., T.B. and O.K. wrote the article.

Supplemental Data

Supplemental Data Can be download from the ***The Plant Cell***.

Supplemental Figure 1. MAFFT Alignment of the Putative CYCLOPS Binding Site of Eight Legume Species.

Supplemental Figure 2. MAFFT Alignment of the 472-bp Conserved Region of Eight Legume Species.

Supplemental Figure 3. *nifH* Expression Is Induced in *ProNIN$_{CE-5kb}$:NIN* Transgenic *nin-1* Root Nodules.

Supplemental Figure 4. Phenotype of *nin-1* Transformed with *ProNIN$_{CE(\Delta D1/D2/D3)-5kb}$:NIN* Constructs.

Supplemental Table 1. Sequence Information of Aligned Species.

Supplemental Table 2. Primers Used in This Study.

Supplemental Data Set 1. Genomic Sequence of the Mutated Region in FN8113.

Supplemental Figure 1. MAFFT Alignment of the CYCLOPS Binding Site of Eight Legume Species. (Supports Figures 3 and 4A).

Conservation of nucleotides is indicated by shades; black - 100% identity, grey - 80-100% identity, light grey - 60-80% identity and white - less than 60% identity. Two red rectangles indicate the palindromic sequence (*CYC-box*) within CYCLOPS responsive cis element (*CYC-RE*).

2
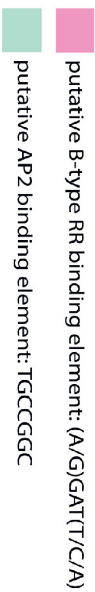
Supplemental Figure 2. MAFFT Alignment of the 472 bp Conserved Region of Eight Legume Species. (Support Figure 4A). Conservation of nucleotide is indicated by shades; black - 100% identity, grey - 80-100% identity, light grey - 60-80% identity and white - less than 60% identity. 472-bp region was divided into three parts named domains one to three (D1 to D3). D1 and D3 contain six and three putative cytokinin response elements, respectively, whereas domain 2 (D2) contains a putative AP2-binding site and a single cytokinin response element.
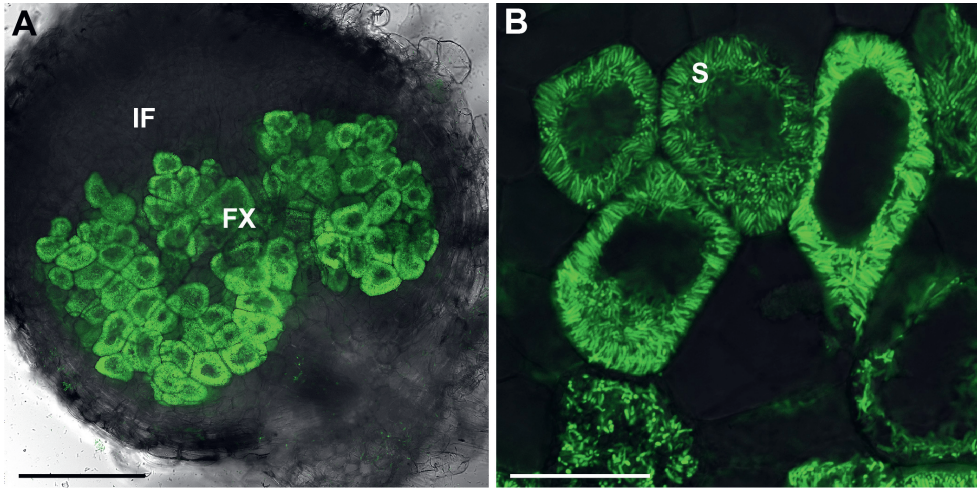
putative B-type RR binding element: (A/G)GAT(T/C/A)

putative AP2 binding element: TGCCGGC

Supplemental Figure 3. *nifH* **Expression Is Induced in** *ProNIN$_{CE-5kb}$:NIN* **Transgenic** *nin-1* **Root Nodules.** (Support Figure 5A and B).

(A) to **(B)** Confocal images of a nodule formed on the *ProNIN$_{CE-5kb}$:NIN* Transgenic *nin-1* root at four days post inoculation with rhizobial strain carrying nifH:GFP. Transgenic nodule showed that *nifH* is switched on in fixation zone **(A)**. Bar = 200 µm. IF, infection zone; FX, fixation zone. Magnified image of infected cells in fixation zone **(B)**. Bar = 50 µm. S, symbiosome.
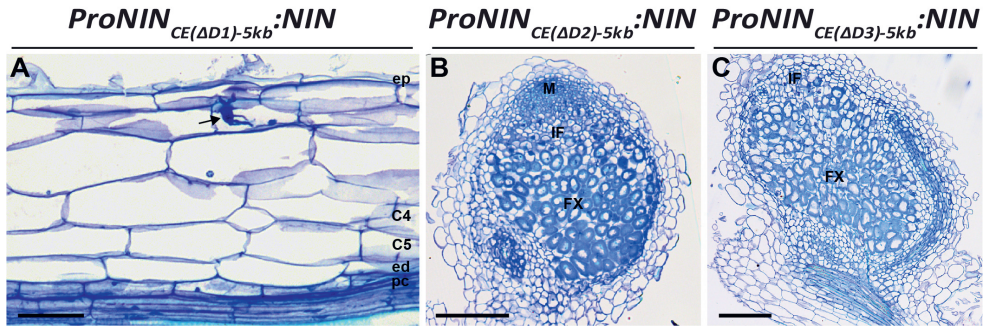
$ProNIN_{CE(\Delta D1)-5kb}$:NIN $ProNIN_{CE(\Delta D2)-5kb}$:NIN $ProNIN_{CE(\Delta D3)-5kb}$:NIN

Supplemental Figure 4. Phenotype of *nin-1* Transformed with $ProNIN_{CE(\Delta D1/D2/D3)-5kb}$:NIN Constructs. (Support Figure 4).

Longitudinal plastic sections of inoculated *nin-1* root transformed with $ProNIN_{CE(\Delta D1)-5kb}$:NIN **(A).** Bar =50 µm. Nodule plastic sections of *nin-1* transformed with $ProNIN_{CE(\Delta D2)-5kb}$:NIN **(B)** and $ProNIN_{CE(\Delta D3)-5kb}$:NIN **(C)**. Bar = 200 µm. Plastic sections are stained with toluidine blue.

Supplemental Table 1. Sequence Information of Aligned Species. (Support Figure 3).

Species	Genome version	Pseudochromose/ scaffold id	NIN gene annotation	Putative CYCLOPS binding site (-bp) †	Location of three conserved regions (3C) (- bp) †		
					Region 1	Region 2/ CE Region	Region 3
Arachis duranensis	Aradu_v1.0	Aradu.A07	Aradu.46M2Y	1102…1114	12755…12939	13822…14280	ND‡
Cicer arietenum	Cicer_arietinum_GA_v1.0	Ca2	Ca_09832	1467…1481	12144…12314	12503…13011	14593…15207
Lotus japonicus	Lj3.0	Chr2	Lj2g3v3373100	945…957	42126…42392	44723…45273	48459…48904
*Glycine max**	Wm82.a2.v1	Gm02	Glyma.02g311000.	640…652	9258…9544	11271…11783	ND‡
	Wm82.a2.v1	**Gm04**	**Glyma.04g000600**	**2241…2257**	**8536…8814**	**10738…11278**	**ND‡**
*Lupinus angustifolius**	Lupinus_angustifolius_pschrom_v1.0	NLL-02	Lup019181.1	886…900	4800…5162	6439…6774	ND‡
		NLL-011	**Lup029716.1**	**590..614**	**4111…4507**	**4617…4874**	**ND‡**
*Cajanus cajan**	PigeonPea.scafSeq. LG_V5.0	Scaffold133201	C.cajan_37712	1396..1410	18677…18772	23796…24273	ND‡
		Scaffold132542	**C.cajan_33924**	**7032..7046**	**17957…19089**	**23574…24225**	**26.361…26599**
Trifolium pratense	redclover_v2.1	Tp57577_TGAC_v2_LG4	Tp57577_TGAC_v2_gene18624	2216..2230	12254…12463	15497…16132	17392…18136
Medicago truncatula	JCVI.Medtr. v4.20130313	Chr5	Medtr5g099060	3026..3040	15769…15985	17611…18083	18654…19153

* *Glycine max*, *Lupinus angustifolius* and *Cajanus cajan* has two *NIN* genes, the one designated in boldface was not used for alignment in Figure 3 and Supplemental Figure 1 and 2.
† The relative position of conserved regions compared to Medicago was scored for each scaffold based on MAUVE alignments. Regions are annotated relative to *NIN* start codon.
‡ ND, not detected.

Supplemental Table 2. Primers Used in This Study.

Name	Sequence (5′→3′)
JH5.17-F	GACATCTTTCGTTGGTGGCAA
JH5.17-R	TCGATGTTTTTCGGGGGTGT
NINg-F	**GGGGACAGCTTTCTTGTACAAAGTGGAA**ATGGAATATGGTGGTG-GGTTAGTG
NINg-R	**GGGGACAACTTTGTATAATAAAGTTGC**GAACAAAATAGTTTATG-TAATCACAAAGAC
ProNIN-2.2kb-F	**CACC**GTGGTACCCACTCAATGGTA
ProNIN-5kb-F	**CACC**TACTCTATTAGTGCTACCTT
ProNIN-0kb-R	CCTTATAATTAAAGTTGTTTCTCAGATC
ProNIN-3C-F	**GGGGACAACTTTGTATAGAAAAGTTG**GTCATGGCTTGTCCAA-CAAC
ProNIN-3C-R	**GGGGACTGCTTTTTTGTACAAACTTG**CTTTCCCGCATGATACT-CAACG
ProNIN-CE-F	**GGGGACAACTTTGTATAGAAAAGTTG**GCAACAACGCACAACTCG-TAC
ProNIN-CE-R	**GGGGACTGCTTTTTTGTACAAACTTG**GTTGCTAACGAGTGCCT-TCATG
ProNIN-ΔD1-R	CAATCAGTGTTAACGTTCTATTATACTATA
ProNIN-ΔD1-F	TATAGTATAATAGAACGTTAACACTGATTG
ProNIN-ΔD2-R	GTCTCAAGAGCAGTGGATCTGCTTAAGTAA
ProNIN-ΔD2-F	TTACTTAAGCAGATCCACTGCTCTTGAGAC
ProNIN-ΔD3-R	TTTTAGTTATAATAGGAACATGTCTGATCA
ProNIN-ΔD3-F	TGATCAGACATGTTCCTATTATAACTAAAA
ProNIN-ΔCYCLOPS-R	GGGCCATCTCTCTGCTTCTACAAATTTTCT
ProNIN-ΔCYCLOPS-F	AGAAAATTTGTAGAAGCAGAGAGATGGCCC
NIN-qPCR-F	ATTGCAAGGCGATTTAACCTAACA
NIN-qPCR-R	GAGAGGGGAAGCTTGAAAAAGAGA
NF-YA1-qPCR-F	TATGGAGGAGACTCTTGTGG
NF-YA1-qPCR-R	GGTTGCTTGATGATTTGGTG
ACTIN2-qPCR-F	TGGCATCACTCAGTACCTTTCAACAG
ACTIN2-qPCR-R	ACCCAAAGCATCAAATAATAAGTCAACC

Sequences designated in boldface were added to primers for TOPO cloning or BP recombination

CHAPTER 3

Nod factor signaling controlled genes in Medicago nodule

Jieyu Liu[1], Defeng Shen[1], Wei Liu[1], Rik Huisman[1], Hetty Busink[3], Tian Zeng[1], Erik Limpens[1], Elio Schijlen[3], Olga Kulikova[1] and Ton Bisseling[1,2,*]

[1] Laboratory of Molecular Biology, Department of Plant Sciences, Wageningen University & Research, 6708 PB Wageningen, The Netherlands

[2] Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing University of Agriculture, 102206, Beijing, China

[3] Bioscience, Plant Research International, Wageningen University & Research, 6708 PB Wageningen, The Netherlands

*Correspondence to: ton.bisseling@wur.nl

3

Abstract

Legume nodule formation is induced by lipo-chito-oligosaccharides, known as Nod factors (NFs), that are secreted by the rhizobia. Transcriptional changes induced in roots upon application of NFs have been intensively studied. Although it has been shown that the NF receptors also accumulate in mature *Medicago truncatula* (Medicago) nodules, the transcriptional changes induced by this late stage NF signaling have never been investigated. To study the NF signaling controlled genes in nodules, we made use of spontaneous nodules which were formed by overexpression of an auto-active *CCaMK* (*Calcium/ calmodulin-dependent kinase*) gene. These spontaneous nodules showed wildtype (WT)-like tissue organization with a developmental gradient along their longitudinal axis. Transcriptome analysis showed that many genes which are especially expressed in infection and fixation zone of WT nodules are induced in the spontaneous nodules, such as some *NCR* (*Nodule-specific Cystein-Rich Peptide*) genes and *leghemoglobin* genes. Further, to study late stage NF signaling controlled genes and distinguish them from genes which are activated upon bacterial release, we made use of the mutants involved (TE7) and independent [*Vamp*(*Vesicle-associated membrane protein*s)*721d&e RNAi*] of NF signaling. Combined with laser microdissection and transcriptome analysis, we identified the late stage NF signalling controlled set of genes. These genes turned to be markedly different from the genes activated in roots upon NF application.

Introduction

By establishing a root nodule symbiosis with nitrogen-fixing rhizobia legumes can grow in nitrogen-poor soil. In most legumes the formation of these nodules are induced by lipo-chito-oligosaccharides, known as Nod factors (NFs) that are secreted by the rhizobia (Denarie and Debelle, 1996). NFs are perceived in the root epidermis by two LysM domain containing receptor like kinases, which activate a specific signal transduction cascade (reviewed in Downie, 2014). This sets in motion the nodule formation process. Transcriptional changes induced upon perception of NFs in the epidermis (early stage NF signaling) have been intensively studied. In *Medicago truncatula* (Medicago) it has been shown that the NF receptors also accumulate in the mature nodules (Moling et al., 2014). This strongly suggests that NF signaling also regulates gene expression in mature nodules. However, the transcriptional changes induced by this late

stage NF signaling have never been studied. In root nodules numerous genes are upregulated in comparison to roots. In this study we will focus on the genes that are induced in Medicago nodules as a result of late stage NF signaling.

After perception of NFs, a signaling cascade is activated that involves the following components: a leucine-rich repeat receptor kinase [DMI2 in Medicago and SYMRK in *Lotus japonicus* (Lotus)] (Stracke et al., 2002; Endre et al., 2002), an enzyme involved in mevalonate production (HMGR1) (Jayaraman et al., 2015; Kevei et al., 2007), nuclear membrane-located cation channels (MtDMI1, MtCNGC15a-c, LjCASTOR and LjPOLLUX) (Ané et al., 2004; Kanamori et al., 2006; Charpentier et al., 2008, 2016). These lead to nuclear calcium oscillation which can be decoded by a calcium calmodulin-dependent kinase (MtDMI3/LjCCaMK) (Mitra et al., 2004; Tirichine et al., 2006). Further, MtDMI3/LjCCaMK activates the transcription factor CYCLOPS (IPD3 in Medicago) by phosphorylation (Messinese et al., 2007; Yano et al., 2008). Upon phosphorylation, CYCLOPS induces the expression of *NIN* (Singh et al., 2014), the transcription factor that plays a key role during nodule formation.

NF signaling activates two main processes during nodule initiation: the infection process that starts in the epidermis and the mitotic activation of inner root cells resulting in the formation of a nodule primordium (Oldroyd and Downie, 2008). Infection often involves the formation of a tube-like structure, named infection thread, by which the rhizobia enter the plant. Infection threads grow towards the primordia and enter the primordium cells. Subsequently, rhizobia are released from the infection thread. Medicago forms indeterminate nodules which have a persistent apical meristem that adds cells to the nodule tissues (Hirsch, 1992; Sprent, 2007). This are the central tissue containing the infected cells and the peripheral nodule cortex and parenchyma. The latter contains the nodule vascular bundles. Due to the persistent activity of the meristem a developmental gradient is formed along the longitudinal axis of the nodule. Cells added by the meristem to the central tissue can be penetrated by infection threads after which rhizobia are released. This occurs in about two cell layers adjacent to the meristem and in these cell layers the two NF receptors accumulate (Moling et al., 2014).

The co-occurrence of NF receptors and rhizobial release indicate that NF signaling might control this process. Indeed, knockdown of the NF receptor gene *MtNFP* or downstream gene *MtDMI2* causes a block of bacteria release (Moling et al., 2014). Furthermore, the type II nodules formed on the TE7 mutant, a weak

allele of CYCLOPS/IPD3, also shows a block of rhizobia release (Ovchinnikova et al., 2011; Benaben et al., 1995). In addition to NF signaling, rhiziobial release requires processes independent of NF signaling. For example, rhizobial release is hampered in Medicago *Vamp721d&e RNAi* roots whereas NF signaling is not affected (Ivanov et al., 2012).

Upon release, bacteria divide, differentiate and subsequently colonize the infected cells. Bacterial differentiation involves endoreduplication by which they become elongated. This is controlled by a number of nodule-specific cysteine-rich peptides (NCRs) which are produced by the host and delivered to the rhizobia (Van de Velde et al., 2010; Alunni and Gourion, 2016; Mergaert et al., 2003). The infected host cells also undergo endoreduplication and become enlarged (Vinardell et al., 2013).

Due to the developmental gradient, indeterminate nodules can be divided in zones. Using laser capture microdissection it has been shown that these zones have distinct transcriptomes (Roux et al., 2014). Adjacent to the apical meristem (zone I) is the infection zone (zone II). This zone can be further divide into the distal fraction zone II (FIId) and proximal fraction zone II (FIIp), according to Roux *et al,* 2014. The infection zone is followed by the fixation zone (ZIII), where nitrogen fixation take place. The transition from infection zone to the distal part of the fixation zone (interzone IZ according to Roux *et al,* 2014 and Vasse *et al.,* 1990) occurs rapidly and is accompanied by many changes in both gene expression and cell morphology (Gavrin et al., 2014). For example, NIN accumulates to the highest level in the most proximal part of the infection zone and its level markedly drops at the transition to the fixation zone (Liu et al., 2021). Further, the rhizobial *nif* genes, encoding nitrogenase subunits are induced at this transition (Yang et al., 1991; Gavrin et al., 2014).

The crucial role of NF signaling in nodule formation is further underlined by spontaneous nodules formed by auto-activate NF signaling pathway genes. For example, in both Lotus and Medicago, mutations in the autoinhibitory domain of CCaMK resulted in the formation of spontaneous nodules, in the absence of rhizobia (Tirichine et al., 2006; Gleason et al., 2006). The peripheral vascular bundles and expression of some nodule specific genes in these spontaneous nodules support that they are genuine nodules.

As spontaneous nodules are the result of NF signaling, they are ideal material for studying the NF induced genes in nodules. Here, we analyzed the transcriptome of spontaneous nodules formed by overexpression of an auto-active *CCaMK*

gene. The tissue organization and a developmental gradient resembling wildtype nodules, and the transcriptome data support this observation. Further, by using laser microdissection coupled to RNA sequencing, we analysed the transcriptome from NF signaling related (TE7) and independent (*Vamp721d&e RNAi*) mutants nodule apexes. This allow us to identify late stage NF signalling controlled genes.

Results

Spontaneous nodules have a tissue organization and a developmental gradient resembling wildtype nodules

In root nodules numerous genes are upregulated in comparison to roots. A subset of these genes will be part of the nodule developmental program that is set in motion by NF signaling. To identify these genes, we made use of spontaneous nodule that are formed by overexpression of the auto-active form of NF signaling gene *CCaMK*,* in the absence of rhizobia. When spontaneous nodules have a similar tissue layout and a developmental gradient along their longitudinal axis as wildtype (WT) nodules, they likely have gone through a developmental program closest to that of WT nodules. To select such spontaneous nodules, we compared spontaneous nodules formed by overexpressing dominant active forms of *CCaMK**. Rice (*Oryza sativa*) OsCCaMK has been shown to be functionally equivalent to MtCCaMK (Chen et al., 2007). Therefore, we used a dominant active form of rice OsCCaMK* (1-303 aa) as well as of MtCCaMK* (1-311 aa) to test which construct induces spontaneous nodules most similar to WT nodules. The constructs were driven by an enhanced *CaMV35S* promoter and were introduced in Medicago *ccamk/dmi3* mutant (Lévy et al., 2004) through *Agrobacterium rhizogenes* mediated hairy-root transformation. Expression of *pCaMV35S:MtCCaMK** resulted in small, round shaped spontaneous nodules (Figure 1A). In contrast, the spontaneous nodules formed on roots transformed with *pCaMV35S:OsCCaMK** were more elongated (Figure 1B), indicating that a persistent meristem is present. Therefore, we decided to focus on the spontaneous nodules formed by the rice construct.

Semi-thin sections showed that these spontaneous nodules have a similar tissue layout as WT nodules (Figure 1C). The central tissue, displayed a zonation like WT nodules and it is flanked by peripheral tissues containing nodule vasculature. A meristem is present at the apex. Cells proximal to the meristem are relatively
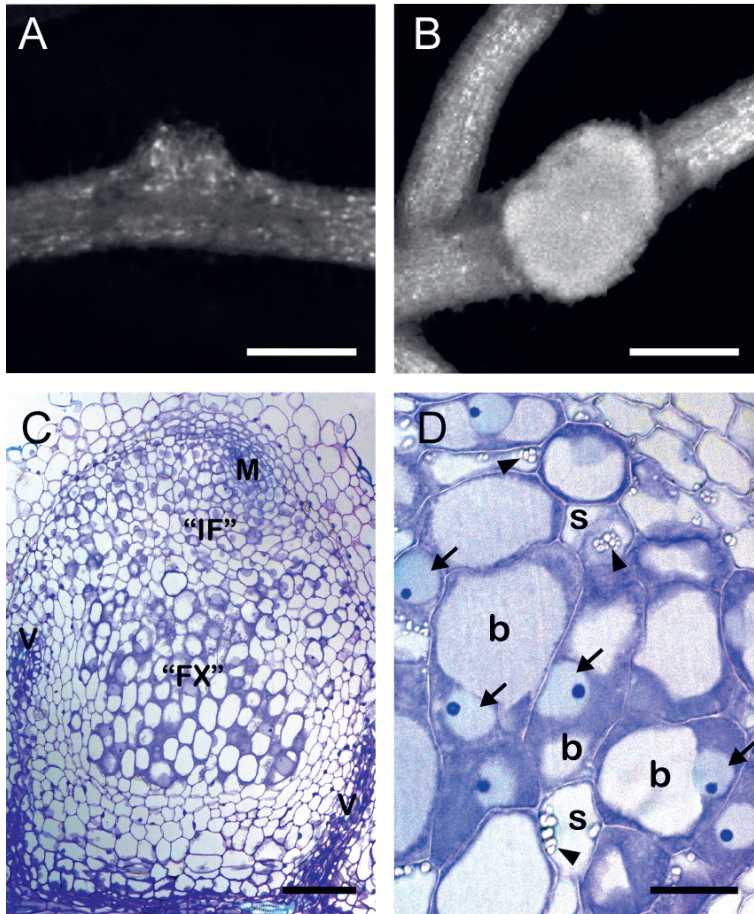
Figure 1. **Spontaneous nodules formed by overexpression of auto-active CCaMK* on Medicago *ccamk/dmi3* mutant roots.**

(A) Spontaneous nodules formed on *pCaMV35S:MtCCaMK** transgenic roots are small, round shaped.

(B) Spontaneous nodules formed on *pCaMV35S:OsCCaMK** transgenic roots are more elongated.

(C) Longitudinal section of spontaneous nodule induced by *pCaMV35S:OsCCaMK** shows a tissue organization and a developmental gradient resembling WT nodules. M, meristem; "IF", a zone resembling WT infection zone; "FX", a zone resembling WT fixation zone; V, vasculature.

(D) Magnification of "FX" in (C) shows large cells (L) containing enlarged nuclei (indicated by arrows) interspersed with smaller sized cells (S) which contain starch granules (indicated by arrowheads).

Bars: 2mm in A and B; 100 µm in C; 20 µm in D.

small. The basal part of the central tissue is composed of relatively large cells that are interspersed with smaller sized cells (Figure 1C and 1D). The big cells contain enlarged nuclei (Figure 1D) suggesting that these cells went through rounds of endoreduplication, similar as occurs in infected cells of WT nodules. The small cells contain starch granules (Figure 1D), similar to uninfected cells in WT nodules. The zone with relatively small cells close to the meristem could be similar to the infection zone, whereas the basal part with large cells interspersed with the smaller cells has some similarity with the fixation zone of WT nodules.

In WT nodules *NIN* mRNA as well as NIN protein gradually accumulate in the infection zone and their level suddenly drops at the transition from infection to fixation zone (Liu et al., 2021). Therefore, we tested whether the *NIN* expression pattern in spontaneous nodules could support the conclusion that they have a similar zonation as WT nodules. We analyzed *NIN* expression by RNA *in situ* hybridization in spontaneous nodules formed on J5 (WT) roots transformed with *pCaMV35S:OsCCaMK**. Similar to the rhizobia infected WT nodules, *NIN* is highest expressed in a zone at the apex of the nodule and its mRNA level increased towards the proximal part of this zone (Figure 2). Then the mRNA level suddenly decreased. This expression pattern supports the idea that an "infection zone" and a "fixation zone" were formed in the spontaneous nodules. A low level of *NIN* mRNA was also detected in the meristem and vasculature of the spontaneous nodule (Figure 2), similar as in WT nodules (Liu et al., 2021).

So, the spontaneous nodules formed by overexpression of auto-active *OsCCaMK** resemble WT Medicago nodules, albeit without rhizobia. Therefore, it seems to be a suitable material for identifying NF signaling induced genes.
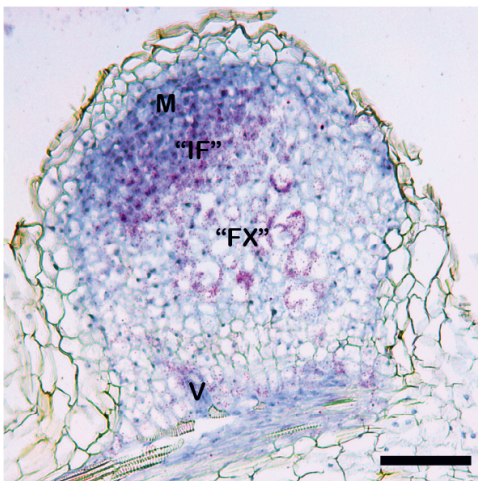
Figure 2. *NIN* **expression pattern in spontaneous nodules is similar to that in WT nodules.**

Hybridization signals are visible as red dots. *NIN* is highest expressed in a zone at the apex of the nodule and its mRNA level increased towards the proximal part of this zone. M, meristem; "IF", a zone resembling WT infection zone; "FX", a zone resembling WT fixation zone; V, vasculature. Scale bar: 100 μm.

Genes induced in infection and fixation zone of WT nodules are induced in the spontaneous nodules

To identify the genes that are upregulated in spontaneous nodules, we compared the transcriptome of spontaneous nodules, formed on Medicago J5 roots transformed with *pCaMV35S:OsCCaMK**, to that of transgenic root tips. To exclude upregulated genes that are not related to nodulation, the list of identified genes was compared with the list of genes upregulated in rhizobia infected WT nodules in comparison to roots. Genes that occurred in both lists were maintained. In this way, we identified 1923 genes which are upregulated in spontaneous nodules. In line with the *in situ* hybridization experiment, *NIN* is among these induced genes. In addition to *NIN*, there are many nodule specific genes related to different nodule developmental stages that are induced in the spontaneous nodules. This include 464 genes especially (more than 50% of the total number of reads) expressed in the infection zone of WT nodules (FIId and FIIp, according to Roux *et al,* 2014). Examples are nodule specific transcription factors *NF-YA1* and *EFD*; a component of the nodule-specific signal peptidase complex *DNF1*; nodule-specific remorin *SYMREM*; early nodulin genes *ENOD11* and *ENOD12;* and nodule specific peptides, including 13 GRPs and 76 NCRs. Among the genes upregulated in spontaneous nodules, 893 genes are especially expressed in the fixation zone of WT nodules (IZ and ZIII, according to Roux *et al,* 2014). Examples are six leghemoglobin genes, 66 *NCRs*, seven *GRPs*, seven papain family cysteine protease (*CPs*) and the defense suppressing genes *NAD1*, *RSD*, *SymCRK*, and *DNF2*. Induction of these genes supports the idea that a developmental gradient has been formed in spontaneous nodules. However, compared with WT nodules, which have 735 genes especially expressed in the infection zone and 2722 genes especially expressed in the fixation zone, there are markedly fewer genes upregulated in spontaneous nodules. For example, in WT nodules, there are 11 leghemoglobin genes induced, whereas only six leghemoglobin genes are induced in the spontaneous nodules and their induction level is markedly lower (Table 1). Similarly, only 143 out of 527 NCR genes, which induced in WT nodule, which are induced in spontaneous nodules, and induction level of most of these genes is also reduced. These data imply that the interaction with rhizobia might be required for a higher level of induction of nodule enhanced genes and to induce the expression of all WT infection and fixation zone related genes.

The genes upregulated in spontaneous nodules are the result of NF signalling. This can be NF signalling at an early stage which sets the nodule developmental

Table 1. Six out of 11 leghemoglobin genes are induced in spontaneous nodules.

Mt genome V.5.0 ID	WT nodule (AvTPM)	Spontaneous nodule (AvTPM)	Root tip (AvTPM)
MtrunA17Chr5g0435991	**3018**	**165**	**4**
MtrunA17Chr5g0427351	**23755**	**80**	**13**
MtrunA17Chr1g0197501	**394**	**57**	**1**
MtrunA17Chr1g0197531	**1471**	**23**	**1**
MtrunA17Chr1g0148751	**6880**	**22**	**3**
MtrunA17Chr5g0435621	**1441**	**20**	**1**
MtrunA17Chr5g0417631	1670	3	1
MtrunA17Chr5g0435611	644	2	0
MtrunA17Chr5g0435981	3412	1	1
MtrunA17Chr1g0170851	665	0	0
MtrunA17Chr7g0270491	61	0	0

Boldface character indicates that the leghemoglobin genes have a significantly enhanced expression in spontaneous nodules; AvTPM, average transcripts per million.

program in motion, we name it early stage NF signalling. Alternatively, this can be the result of NF signalling in the nodule which in WT Medicago nodules occurs in a few cell layers at the apex of the nodule, we name it late stage NF signalling. However, the transcriptome data obtained from spontaneous nodules do not allow to distinguish between early and late stage NF signalling. Therefore, we next made use of laser-capture microdissection, by which the nodule apex was specifically collected for RNA sequencing, to identify genes that are induced upon late stage NF signalling.

Identification of genes whose expression is the result of late stage NF signalling in Medicago nodules

NF receptors accumulate and form heterodimers in Medicago nodules, but only in two cell layers adjacent to the apical meristem (Moling et al., 2014). To identify late stage NF signalling controlled genes, we made use of laser microdissection to isolate nodule apices as this is the site where NF signalling occurs. Rhizobial release also occurs at the apex, therefore, to eliminate the genes which are regulated after release of rhizobia, we made use of three Medicago strains: WT (J5) (release+/NF signalling+), *TE7* (release-/NF signalling-) and *Vamp721d&e RNAi* (release-/NF signalling+). RNA was extracted from three biological replicates and amplified for library preparation and RNA-seq.

Table 2. Expression of defence suppressing genes.

Mt genome V.5.0 ID	Gene name	WT nodule apex (AvTPM)	Spontaneous nodule (AvTPM)	Root tip (AvTPM)	WT nodule apex (AvTPM)	TE7 nodule apex (AvTPM)	Vamp721d&e RNAi nodule apex (AvTPM)
MtrunA17Chr4g0044681	*MtDNF2*	170	**78**	1	315	**15**	310
MtrunA17Chr7g0239441	*MtRSD*	42	**12**	0	5	**0**	6
MtrunA17Chr7g0222521	*NAD1*	49	**19**	0	315	**2**	71
MtrunA17Chr3g0119041	*MtSymCRK*	167	**68**	3	44	44	43

Boldface character indicates that the either the expression level is significantly higher in spontaneous nodules in comparison with root tip or the expression level is significantly lower in TE7 nodule apex compared with J5 nodule apex and Vamp721d&e RNAi nodule apex; AvTPM, average transcripts per million.

We identified 3235 genes differentially expressed in *TE7* compared with WT. 1289 of these *TE7* differentially expressed genes (DEG) were shared with *Vamp721d&e RNAi* nodules. These DEGs were most likely regulated upon release of rhizobia. 1216 genes are differentially expressed in *TE7* in comparison to both WT and *Vamp721e&d RNAi* and so these genes are most likely late stage NF signaling controlled genes. Among these DEGs, 462 were upregulated and 754 were down regulated in TE7. This indicates that NF signaling induces gene expression, but also represses a substantial number of genes.

From these 754 late stage NF signaling induced genes, 206 were induced in spontaneous nodule. This included genes encoding key transcription factors such as NIN and EFD; nodule specific peptides, such as NCRs and GRPs; proteins involved in cytokinin biosynthesis, signaling or metabolism such as KNOX3, LOG2, RRA4 and CKX1; and defense suppressing proteins such as DNF2 and RSD.

Although infected cells of WT nodules accommodate massive amounts of bacteria, defence responses are well suppressed. Several defence suppressing genes contribute to this, including *RSD*, *NAD1*, *DNF2* and *SymCRK* (Bourcy *et al.*, 2013; Sinharoy *et al.*, 2013; Berrabah *et al.*, 2014; Wang *et al.*, 2016; Domonkos *et al.*, 2017). High induction all of these defence suppressing genes in spontaneous nodules shows that NF signalling is sufficient to induce these genes and even in the

absence of bacteria defence responses are actively suppressed (Table 2). Moreover, induction level of *DNF2*, *RSD* and *NAD1* is significantly lower in *TE7* in comparison with WT and *Vamp721d&e* RNAi nodules, indicating that they are regulated by the late stage NF signalling (Table 2). It is noteworthy that the expression level of *NAD1* in *Vamp721d&e* RNAi nodules is significantly lower compared with WT. This suggests that *NAD1* expression is induced by late stage NF signalling, but the higher level of expression depends on release of rhizobia. In contrast, *SymCRK* has a similar expression level in all three laser-captured nodule types, therefore, its expression most likely fully depends on the early stage NF signaling (Table 2).

Although some genes whose expression is induced by late stage NF signaling are also induced in the spontaneous nodules, the majority (548) are not. For example, many flavonoid biosynthesis genes such as chalcone synthase 4 (CHS4) were induced by late stage NF signalling according to laser capture data, but not in spontaneous nodules. This is also the case for 27 NCRs. An explanation might be that the induction of these genes requires the presence of rhizobia, in addition to late NF signaling.

When roots are treated with NFs a set of genes is induced within hours. We tested to what extend this set of genes is similar to the genes that are upregulated by late stage NF signaling nodules. We used the 754 late stage NF induced genes and compared these with NF treated roots (3h), in which 609 genes were induced. These two sets of NF signaling induced genes only shared 39 genes, including NIN, genes related to cytokinin biosynthesis/signaling and flavonoid biosynthesis genes. So, the set of NF signaling induced genes in nodules is markedly different from that in roots. To further verify this conclusion, we used other transcriptome data to perform similar analysis. This showed similar results. For example, in the transcriptome data of Medicago root epidermis cells after 4 or 24h NF treatment, 722 genes were induced, but only 30 of these genes are shared with late stage NF signaling induced genes.

It is noticeable that some genes, such as *NF-YA1*, were induced by NF signalling in both spontaneous nodules and in roots, but not shown to be induced by late stage NF signalling. These genes most likely fully depend on early stage NF signalling and are independent of late stage NF signalling. There are 83 of such genes, including *NF-YA1*, *Nip*, *NSP1*, *SymCRK*, *Noot1, VAPYRIN, ENOD12*, and *ENOD11*.

Discussion

Here we identified genes that are controlled by late stage NF signaling in Medicago root nodules. We showed that this set of genes is markedly different from that activated at early stage. Combined with the transcriptome of spontaneous nodules, we found that the genes induced in the nodules are the result of both early and late stage NF signaling. Surprisingly, several nodule enhanced genes that are especially expressed in the fixation zone of WT nodules are expressed in the spontaneous nodules, albeit at a markedly lower level. The induction of all nodule enhanced genes and at a high level appears to require the presence of rhizobia in root nodules.

We identified in total 1216 genes whose expression is regulated by late stage NF signaling in Medicago nodules. 754 of these genes are induced. The vast majority of these genes (>700) are not induced by application of NFs to roots. These nodule specific genes include for example some NCR genes and defense suppressing genes. It also includes many genes which function in nodule remains unknown, so they might be interesting for further functional studies. There are genes regulated by late stage NF signaling as well as in NF treated roots. This includes the gene encoding the key transcription factor NIN. These genes might play a role in both nodule initiation and in mature nodules. For example, NIN is required for not only early stage infection thread initiation and nodule organogenesis, but also essential for late stage rhizobial release (J. Liu unpublished), symbiosome development, suppression of defense and premature senescence in Medicago nodules (Liu et al., 2021).

In some species, application of NFs to roots results in the formation of nodule primordia in which nodule specific genes are expressed. This is for example the case in *Vicia sativa* (vetch) (Vijn et al., 1995). As NFs are immobile molecules, this implies that NF perception at the epidermis is sufficient to induce a developmental program that results in nodule primordium formation. The induction of nodule enhanced genes in such primordia cannot depend on the activation of NF receptors in these structures as this would require the presence of rhizobia inside these primordia. Therefore, it is possible that enhanced expression in WT nodules is the result of the nodule developmental program induced by early stage NF signaling and does not depend on late stage NF signaling. Such genes can have a similar level of expression in WT and TE7, but induced in spontaneous nodules. An experimental support for this is the expression pattern of nodule specific transcription factor *NF-YA1*. It is highly

expressed in the proximal part of nodule meristem of WT nodules. So in the absence the late stage NF signalling, and its expression does not increase in cells where late stage NF signalling occurs (transition to the infection zone) (Liu et al., 2021). We identified 83 genes that are specifically regulated by early stage NF signaling.

206 late stage NF signaling induced genes are induced also in spontaneous nodules. Therefore, it is probable that in spontaneous nodules the late stage NF signaling is induced by CCaMK*. In WT nodules we identified genes induced by late stage NF signaling by analyzing the transcriptomes of the apex of the nodules. Therefore, we cannot conclude whether NF signaling dependent gene expressions in for example fixation zone are the result of late or early stage NF signaling. It also cannot be concluded whether the genes induced in spontaneous nodules that are typical for the fixation zone, like the leghemoglobin genes, are induced by early or late stage NF signaling.

The majority (548) of genes induced by late stage NF signaling are not induced in spontaneous nodule. This indicate that in addition to NFs other signals might be involved in the regulation of host genes. This hypothesis is supported by some Medicago mutants in which NF signaling is present but bacterial differentiation is hampered and non-functional white nodules are formed. This indicates that bacterial differentiation is required to produce factors to further induce *leghemoglobin* genes expression. For example, the *dnf1* mutant of Medicago forms white nodule and bacterial differentiation is blocked (Wang et al., 2010; Van de Velde et al., 2010; Lang and Long, 2015). Its phenotype is caused by mutation in a gene encoding a subunit of a signal peptidase complex which required for processing NCR peptides, so NF signaling suppose not affected in this mutant. The phenotype of *dnf1* suggests that differentiation of rhizobia is required for the induction of nodule enhanced genes, for example the *leghemoglobin* genes.

Spontaneous nodules formed by *pCaMV35S:OsCCaMK** resemble the WT nodules. They have peripheral vasculature and a developmental gradient with recognizable zones. Transcriptionally, many infection zone or fixation zone specifically expressed genes are induced in these spontaneous nodules. Besides, the expression pattern of *NIN* in spontaneous nodules resembles its expression pattern in WT nodules. To induce the formation of such spontaneous nodules in mutant background can be useful to study the gene function independent of bacteria.

Many legume species such as Lotus, form determinate nodules, in which the meristem is only transiently present. In these nodules, developmental gradient is not formed along the longitudinal axis. It will be interesting to see whether NF receptors also accumulate in these determinate nodules and to what extend they control gene expression in this nodule type.

Materials and Methods

Plant material and growth conditions

Medicago (*Medicago truncatula*) Jemalong 5 (J5), and two mutants TE7 (Benaben et al., 1995) and *dmi3* (Lévy et al., 2004) were used in this study. Plants were grown in a growth chamber at 20°C with 16h/8h day/night regime. Seed sterilization, germination and *Agrobacterium* msu440 mediated hairy root transformation was performed according to Limpens *et al.* (2004). Medicago plants were grown in perlite saturated with low nitrate (0.25 mM Ca(NO3)$_2$) containing Farhaeus (Fa) medium (Catoira *et al.,* 2000) at 21°C and a 16 h : 8 h, light : dark regime. After 1 week of growth, plants were inoculated with *Sinorhizobium meliloti* 2011 wild-type (OD600 = 0.1, 2 ml per plant) or without inoculation in case of spontaneous nodule formation.

Constructs

For spontaneous nodule formation assay, the constructs *pCaMV35S:MtCCaMK** (1-311 aa) and *pCaMV35S:OsCCaMK** (1-303 aa) were generated in Liu (2013). For knock down exocytosis pathway the *p35S:VAMP721d&e-RNAi* construct was generated in Ivanov *et al.* (2012).

Microscopy

Spontaneous nodule formed after 3 weeks of growth in perlite was embedded in Technovit 7100 (Heraeus Kulzer), sectioning and staining were performed according to Xiao *et al.* (2014). Sections were analyzed with DM5500B microscope equipped with a DFC425C camera (Leica). Bright-field images were taken under a stereo macroscope (M165 FC, Leica).

RNA in situ hybridization

RNA *in situ* hybridization was conducted using Invitrogen ViewRNA ISH Tissue 1-Plex Assay kits (Thermo Fisher Scientific) according to manufacturer's user guide and optimized for Medicago root and nodules sections (Kulikova et al., 2018). *NIN* (Medtr5g099060) RNA ISH probe set was designed and synthesized at Thermo Fisher Scientific (Catalogue number: VF1-20312). As a negative control for each hybridization procedure any probe set was omitted, no hybridization signals were detected. The *in situ* images were taken with an AU5500B microscope equipped with a DFC425c camera (Leica).

Laser microdissection

Nodules formed on J5, TE7 and *Vamp721d&e RNAi* were collected two weeks after inoculation with rhizobia. Further fixation, dehydration, embedding, section and laser microdissection procedure were performed according to Zeng *et al.* (2018). Leica LMD7000 laser microdissection microscope was used for laser capture. About 2000 cells were collected for each replicate.

RNA isolation and quality check

RNA from spontaneous nodules (grown three weeks in perlite), ~3mm transgenic root tips (grown three weeks in perlite), and J5 nodules (two weeks post inoculation with rhizobia) were isolated using the Qiagen plant RNA mini kit according to manufacturer's protocol. Isolated RNA samples were quality checked using an Agilent 2100 Bioanalyzer at BGI (Hong Kong).

For samples collected by laser microdissection, RNA was isolated using the Qiagen RNeasy Micro kit according to manufacturer's protocol. The quality of isolated RNA samples was checked using both Agilent 2100 Bioanalyzer and qPCR. For quality check by qPCR, 2 µL of isolated RNA sample was reverse transcribed using iScript cDNA Synthesis kit (Bio-Rad) and pre-amplified using Bio-Rad Sso Advanced PreAmp Supermix. *NIN*, *NF-YA1*, *Vamp721d* and *Vamp721e* were used as marker genes checked by qPCR. RNA samples from the best three replicates of each cell type were amplified using the SMART-Seq v4 Ultra Low Input RNA Kit (Clontech, http://www.clontech.com) according to manufacturer's instructions. The number of amplification cycles (11–15 cycles) for each sample was adjusted based on the amount of RNA determined by bioanalyzer and qPCR analyses.

Library preparation and sequencing

For the spontaneous nodule, root tip and WT nodule samples, RNA-seq library was constructed and sequenced on an BGISEQ-500 platform with paired-end reads at BGI (Hong Kong). For laser micro-dissected samples, sequencing libraries were made using ThruPLEX DNA-seq Kit (Rubicon genomics, http://rubicongenomics.com/) and sequenced on an Illumina HiSeq 2500 platform (paired-end) at the Next-Generation Sequencing (NGS) facility of Bioscience, Wageningen University & Research.

RNA-seq analyses

About 4 Gb cleaned reads were obtained for each sequenced sample. Sequence reads were mapped to the *M. truncatula* genome v.5.0 (Pecrix *et al.,* 2018) using the CLC genomics workbench 10.0.1 (Qiagen). Length fraction and similarity fraction were set to 0.9 during mapping and only uniquely mapped reads were considered in the analysis. All other parameters were set as default. Transcripts per million (TPM) (Wagner *et al.,* 2012) and differential expression analyses were performed using the CLC genomic workbench 10.0.1. The criteria of fold change>4, P-value<0.05 and average TPM>3 were used to identify the differentially expressed genes.

Acknowledgements

3

CHAPTER 4

NIN is essential for development of symbiosomes, suppression of defence and premature senescence in *Medicago truncatula* nodules

Jieyu Liu[1], Menno Rasing[1], Tian Zeng[1], Joël Klein[1], Olga Kulikova[1], and Ton Bisseling[1,2*]

[1] Laboratory of Molecular Biology, Department of Plant Sciences, Graduate School Experimental Plant Sciences, Wageningen University & Research, 6708 PB, Wageningen, The Netherlands

[2] Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing University of Agriculture, 102206, Beijing, China

*Corresponding author: ton.bisseling@wur.nl.

SUMMARY

NIN (NODULE INCEPTION) is a transcription factor that plays a key role during root nodule initiation. However, it's role in later nodule developmental stages is unclear.

Both NIN mRNA and protein accumulated at the highest level in the proximal part of the infection zone in *Medicago truncatula* nodules. Two *nin* weak allele mutants, *nin-13/16*, form a rather normal nodule infection zone, whereas a fixation zone is not formed. Instead, a zone with defence responses and premature senescence occured and symbiosome development gets arrested.

Mutations in *nin-13/16* resulted in a truncated NIN lacking the conserved PB1 domain. However, this did not cause the nodule phenotype as *nin* mutants expressing $NIN_{\Delta PB1}$ formed wild-type-like nodule. The phenotype is likely caused by reduced *NIN* mRNA levels in the cytoplasm. Transcriptome analyses of *nin-16* nodules showed that expression level of defence/senescence-related genes is markedly increased, whereas this of defence suppressing genes is reduced. Although defence/senescence seems well suppressed in the infection zone, the transcriptome is already markedly changed in the proximal part of infection zone.

In addition to its function in infection and nodule organogenesis, NIN also plays a major role at the transition from infection to fixation zone in establishing a functional symbiosis.

Introduction

Legumes have the ability to establish a root nodule symbiosis with bacteria belonging to different genera collectively named rhizobia. Rhizobia are hosted inside specialised nodule cells, where they are able to reduce atmospheric nitrogen into ammonia. Root nodule formation involves two main processes, infection and nodule organogenesis (Oldroyd and Downie, 2008). Infection starts in root hairs with the formation of tube-like structures called infection threads by which bacteria enter root cells (Oldroyd et al., 2011). Nodule organogenesis begins with mitotic reactivation of fully differentiated root cells leading to the formation of a nodule primordium (Xiao et al., 2014). NIN (NODULE INCEPTION) is a key transcription factor that is essential for both processes (Marsh et al., 2007; Schauser et al., 1999). Weak alleles of *nin* have been identified that allow

the formation of nodules (Pislariu *et al.*, 2012; Veerappan *et al.*, 2016). These nodules are small and do not fix nitrogen, suggesting that NIN also plays a role at later stages of nodule development. However, it is not clear in which nodule developmental process NIN is involved. Here we studied the role of NIN during later nodule developmental stages in *Medicago truncatula* (Medicago).

At the start of nodulation, *NIN* is transcriptionally induced upon perception of rhizobia secreted lipo-chito-oligosaccharides, called Nod factors (NF). NF activate a signalling pathway which is shared with the more ancient arbuscular mycorrhizal symbiosis. NIN is the first induced transcription factor that distinguishes the rhizobium activated responses from that of arbuscular mycorrhizae (Oldroyd, 2013). NIN is not only essential for nodulation in legumes, but also in actinorhizal(-like) plants such as *Casuarina* and *Parasponia* (Bu et al., 2020; Clavijo et al., 2015). In both legumes as well as actinorhizal plants, *NIN* is nodule specifically expressed (Marsh et al., 2007; Schauser et al., 1999; Bu et al., 2020; Clavijo et al., 2015). During nodule initiation in legumes, *NIN* is induced in the epidermis, where it is required for infection thread formation (Vernié et al., 2015b). In Medicago, *NIN* is also induced in the pericycle and this marks the start of nodule primordium formation. After which, the expression of *NIN* extends to the dividing cortical cells (Liu *et al.*, 2019b).

Medicago forms indeterminate nodules which have a zonation representing successive developmental stages (Vasse et al., 1990). At the apex there is a meristem which continuously adds cells to nodule tissues. Proximal to the meristem is the infection zone, where cells are penetrated by infection threads from which rhizobia are released. During release, bacteria become surrounded by a plant-derived membrane, these organelle-like structures are called symbiosomes (Roth *et al*., 1988). Subsequently, symbiosomes divide and gradually enlarge. Bacterial differentiation involves endoreduplication and this is controlled by numerous antimicrobial peptides called nodule-specific cysteine-rich (NCR) peptides (Alunni and Gourion, 2016; Vinardell et al., 2013; Van de Velde et al., 2010). NCRs are defensin-like peptides which are targeted to symbiosomes (Mergaert et al., 2003; Van de Velde et al., 2010). Although the infected cells become massively infected by rhizobia, defence responses are well suppressed. Several host genes have been identified that contribute to this suppression, including *Symbiotic Cysteine-rich Receptor Kinase* (*SymCRK*), *Regulator of Symbiosome Differentiation* (*RSD*), *Defective in Nitrogen Fixation 2* (*DNF2*) and *NODULES WITH ACTIVATED DEFENCE 1* (*NAD1*) (Bourcy *et al.*, 2013; Sinharoy *et al.*, 2013; Berrabah *et al.*, 2014; Wang *et al.*, 2016; Domonkos

et al., 2017). Mutations in these genes activate strong defence responses in nodules leading to necrotic cell death and/or premature senescence. Proximal to infection zone is the fixation zone, where rhizobia fully differentiate and start to fix nitrogen. The switch from infection to the distal part of the fixation zone (interzone II/III according to Vasse *et al.*, 1990) occurs rapidly, and is marked by changes in cell morphology as well as gene expressions. For example, the rhizobial *nif* genes, encoding components of nitrogenase are induced, amyloplasts accumulate, bacteroids markedly enlarge and become radially aligned (Gavrin et al., 2014).

In older nodules, a senescent zone is formed at the basal part of the nodule, where both host cells and rhizobia are degraded (Van De Velde et al., 2006). A key feature of nodule senescence is the increase of proteolytic activities (Pierre et al., 2014). Cysteine protease genes are highly induced in senescent nodules (Pierre et al., 2014; Perez Guerra et al., 2010). Knockdown of cysteine proteinase genes delays nodule senescence, while their ectopic expression induces premature senescence (Pierre et al., 2014). This underlines a crucial role of cysteine proteases in nodule senescence.

NIN is the first gene of the *NIN-like protein* (*NLPs*) family that was identified (Schauser et al., 2005). Two domains are highly conserved in this gene family: a DNA-binding RWP-RK domain and a Phox and Bem1 (PB1) domain involved in protein-protein interaction (Schauser et al., 2005). A first clue that NIN also functions at later nodule developmental stages was gained from two weak *nin* alleles: NF10547 (*nin-16*) and NF0440 (Pislariu *et al.*, 2012; Veerappan *et al.*, 2016). Both mutants form nitrogen fixing deficient (Fix-) nodules. Here, we analysed these mutants in detail to uncover the role of NIN in later nodule developmental stages.

Materials and Methods

Plant material and growth conditions

Medicago (*Medicago truncatula*) ecotypes Jemalong A17 and R108, and two *nin* mutants (*nin-13* and *nin-16*) were used in this study. *nin-16* was provided by Rebecca Dickstein (University of North Texas); *nin-13* was obtained from the Noble Research Institute (https://medicago-mutant.noble.org/mutant/). *nin-16* were backcrossed with R108, and *nin-13* and *nin-16* were crossed (F1

hybrid) according to Chabaud *et al*. (2006). Seven *nin-13* and *nin-16* hybrid F1 plants were genotyped by PCR (primers listed in Table S1). The obtained PCR fragments were sequenced to confirm *Tnt1* insertion positions. Seed sterilization, germination and *Agrobacterium* msu440 mediated hairy root transformation was performed according to Limpens *et al*. (2004). Medicago plants were grown in perlite saturated with low nitrate [0.25 mM $Ca(NO_3)_2$] containing Färhaeus (Fä) medium (Catoira et al., 2000) at 21°C and 16h light/8h dark regime. After one week of growth, plants were inoculated with *Sinorhizobium meliloti 2011* wild-type or constitutively expressing GFP or carrying the *PronifH:GFP* reporter (OD_{600}=0.1, 2ml per plant).

RNA isolation and qRT-PCR

RNA was isolated from two-week old nodules using the EZNA Plant RNA mini kit (Omega Bio-tek, Norcross, GA, USA). 1 µg RNA was used for cDNA synthesis with the iScript cDNA synthesis kit (Bio-Rad). Real-time qPCR was performed in 10 µl reactions using SYBR Green Supermix (Bio-Rad) and a CFX real-time system (Bio-Rad). Primers are listed in Table S1. Gene expression was normalized using *ACTIN2* as a reference gene.

Constructs

NIN promotor (*ProNIN*$_{3C-5kb}$) and *ProNIN*$_{3C-5kb}$:*NIN* construct are described in Liu *et al*. (2019). To construct *ProNIN*$_{3C-5kb}$:*GFP-NIN, ProNIN*$_{3C-5kb}$:*NIN-GFP* and *ProNIN*$_{3C-5kb}$:*NIN*$_{ΔPB1}$, MultiSite Gateway (Thermo Fisher Schientific) was used. pENTR-4,1, and pENTR-1,2 contained the 3C and 5kb region of the *NIN* promoter, respectively, pENTR-2,3 was used to introduce coding sequences and 35S terminator (35Ster). *NIN* coding sequence was amplified (full length *NIN* or truncated *NIN* without PB1 domain, 3'-UTR region of *NIN* was not included) by PCR using Medicago A17 nodule cDNA as template, Phusion high-fidelity DNA polymerase (Finnzymes) and specific primers listed in Table S1. To make a N- or C-terminus GFP fusion with NIN, pENTR-TOPO-*NIN* together with pENTR-4,1-35S:GFP and pENTR-2,3-35Ster or pENTR-2,3-GFP-35Ster were recombined into Gateway binary vector pKGW-RR-MGW using LR Clonase II plus (Invitrogen) to generate *Pro35S:GFP-NIN-35Ster* or *Pro35S:GFP-NIN-GFP-35Ster*. To fuse *NIN*$_{ΔPB1}$ to 35S terminator, pENTR-TOPO-*NIN*$_{ΔPB1}$ together with pENTR-4,1-35S:GFP and pENTR-2,3-35Ster were recombined into a pKGW-RR-MGW to generate *Pro35S:GFP-NIN*$_{ΔPB1}$-35Ster. These constructs were used to amplify *GFP-NIN-35Ster/NIN-GFP-35Ster/NIN*$_{ΔPB1}$*-35Ster* by PCR using the forward

primers with attB2 (GGGGACAGCTTTCTTGTACAAAGTGGAA) and reverse primers with attB3 (GGGGACAACTTTGTATAATAAAGTTGC) (Table S1). By BP reaction these PCR fragments were introduced into pENTR-2,3. Finally, MultiSite Gateway reaction was performed to assemble pENTR-2,3-*GFP-NIN-35Ster/NIN-GFP35Ster/NIN$_{ΔPB1}$-35Ster* with *NIN* promoter regions in pENTR-4,1 and pENTR-1,2 (TOPO) into Gateway binary vector pKGW-RR-MGW.

Microscopy

Embedding of plant tissue in Technovit 7100 (Heraeus Kulzer), sectioning and staining were performed according to Xiao *et al.*, 2014. Sections were analysed with DM5500B microscope equipped with DFC425C camera (Leica). Bright-field images were taken under stereo macroscope (M165 FC, Leica). Hand-sectioned nodules were examined using Leica SP8 confocal microscope, excitation wavelengths of 488 nm and 543 nm were used for GFP/SYTO 9 and propidium iodide/DsRED, respectively. The accumulation of phenolic compounds was visualized by potassium permanganate staining (Bourcy *et al.* 2013). Live and dead staining of bacteria was performed as described by Haag *et al.* (2011)

Bacterial length measurement

R108 and *nin-13/16* were inoculated with *Sinorhizobium meliloti 2011* constitutively expressing GFP. Two weeks after inoculation, pictures of bacteria were taken with a Leica SP8 confocal microscope. The length of bacteria was measured using ImageJ (Schneider et al., 2012).

RNA *in situ* hybridization

RNA *in situ* hybridization was conducted using Invitrogen ViewRNA ISH Tissue 1-Plex Assay kits (Thermo Fisher Scientific) according to manufacturer's user guide and optimized for Medicago root and nodules sections (Kulikova et al., 2018). RNA ISH probe sets were designed and synthesized at Thermo Fisher Scientific. Catalogue numbers are VF1-20312 for *NIN* (Medtr5g099060), VP2W7MP for *CP2* (Medtr5g022560), VP7DPDU for *SymCRK* (Medtr3g079850), VPRWEMC for *NAD1* (Medtr7g022640) and VF-20311 for *NF-YA1* (Medtr1g056530). Each probe set was tested on tissue where it should not be present, in this case on non-inoculated roots. As a negative control for each hybridization procedure any probe set was omitted. In both cases no hybridization signals were detected.

The *in situ* images were taken with an AU5500B microscope equipped with a DFC425c camera (Leica).

RNA sequencing analyses

R108 and *nin-16* nodules were harvested two weeks post inoculation. Three biological replicates were prepared for RNA sequencing (RNA-seq). Total RNAs were isolated with Qiagen RNeasy mini kit according to manufacturer's protocol. Isolated RNA samples were quality checked using an Agilent 2100 Bioanalyzer. Subsequently, an RNA-seq library was constructed and sequenced on an BGISEQ-500 platform with paired-end reads at BGI (Hong Kong). About 4Gb cleaned reads were obtained for each sample. Sequence reads were mapped to the *M. truncatula* genome version 5.0 (Pecrix et al., 2018) using CLC genomics workbench 10.0.1 (Qiagen). Length fraction and similarity fraction was set to 0.9 during mapping and only uniquely mapped reads were considered in the analysis. All other parameters were set as default. TPM (transcripts per million) (Wagner et al., 2012) and differential expression analyses were performed using the CLC genomic workbench 10.0.1. The criteria of fold change ≥ 4, FDR (false discovery rate) P-value ≤ 0.01 and average TPM (transcripts per million) ≥ 5 were used to identify the differentially expressed genes.

Results

NIN is highly expressed in the nodule infection zone

According to the *M. truncatula* gene expression atlas, *NIN* is expressed in root nodules (Fig. S1). To determine in which stage of nodule development NIN could play a role, we performed RNA *in situ* hybridisation using a *NIN* specific probe set (Material and Methods). This showed that in two-week old nodules, *NIN* was highest expressed in the infected cells of the infection zone and the expression level was very low in the non-infected cells (Fig. 1a-c). *NIN* mRNA was also detected in the nodule vasculature (Fig. 1a). In the meristem and distal part of the infection zone the expression of *NIN* was rather low (Fig. 1b), whereas in the proximal cell layers of the infection zone, expression of *NIN* was highest (Fig. 1a,b). At the transition from infection to fixation zone *NIN* expression level suddenly dropped (Fig. 1a-c). This switch coincides with amyloplast accumulation at the periphery of the infected cells (Fig. 1c) (Gavrin et al., 2014). A very low level of *NIN* expression was maintained throughout the fixation zone
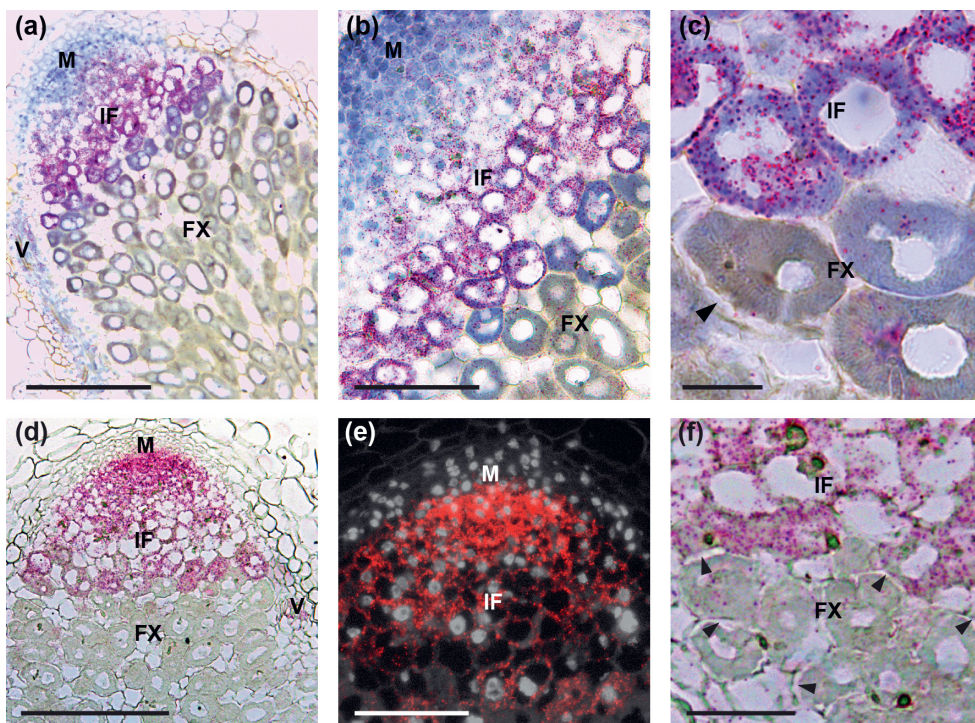
Fig. 1. *NIN* and *NF-YA1* expression pattern in Medicago A17 nodule.

An overview (a, d) and close-up pictures (b, c, e, f) of RNA *in situ* localization of *NIN* (a-c) or *NF-YA1* (d-f) in a nodule at two weeks post inoculation (wpi). Hybridization signals are visible as red dots. Arrowhead indicate amyloplast deposition. M, meristem; IF, infection zone; FX, fixation zone; V, vasculature. Bars: (a, d) 200 µm; (b, e) 100 µm, (c) 20 µm and (f) 50 µm.

(Fig. 1a-c). The results of RNA *in situ* hybridisation are consistent with laser capture transcriptome data (Table S2) (Roux et al., 2014). This expression pattern suggests that NIN plays a role in processes that occur in the infection zone and/or at the transition from infection to fixation zone.

NIN is nuclear located and is present at highest level in the proximal part of the infection zone

Studies on some NIN-like proteins (NLPs) have shown that they are located in the cytoplasm under low nitrate conditions and are translocated to the nucleus when the concentration of nitrate is high (Marchive et al., 2013; Cao et al., 2017; Lin et al., 2018a). To determine in which cells NIN is located in nuclei, where it can induce expression of its target genes, we studied its subcellular localization
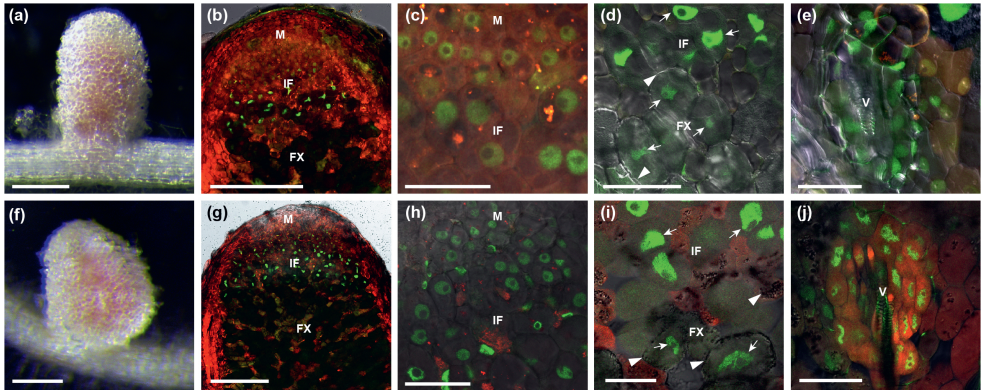
**Fig. 2. NIN protein is nuclear located in Medicago nodules.**

Pink nodules formed on *nin-1* roots transformed with *ProNIN:NIN-GFP* (a) and *ProNIN:GFP-NIN* (f) constructs at six wpi. Confocal images show that both NIN-GFP (b, c, d, e) and GFP-NIN (g, h, i, j) are located in the nucleus of cells in the meristem (c, h), infection zone (b, c, d, g, h, i), fixation zone (d, i) and vasculature (e, j). Transgenic roots were identified by *DsRED* fluorescent (red colour). Arrowhead indicate amyloplast deposition. Arrow indicate nucleus. M, meristem; IF, infection zone; FX, fixation zone; V, vasculature. Bars: (a, f) 2 mm; (b, g) 250 μm and (c, d, e, h, i, j) 40 μm.

in nodules. The functional promoter of *NIN* (Liu et al., 2019b) was used to express *NIN*-tagged with green fluorescent protein (*GFP*) at its C- or N-terminus (*ProNIN$_{3C-5kb}$:NIN-GFP* or *ProNIN$_{3C-5kb}$:GFP-NIN*) in the *nin* knockout mutant *nin-1* (Marsh et al., 2007). Four weeks after inoculation, pink nodules were formed in both cases (Fig. 2a,f), showing that both GFP fusions are biologically functional. Sections of these nodules showed that both C- and N-terminus NIN GFP fusions localized in nuclei in meristem, infection zone, fixation zone and nodule vasculature (Fig. 2b-e,g-j). So, in all cells where *NIN* is expressed, NIN is located in the nucleus. NIN occurred at a rather low level in the distal part of the infection zone and at markedly higher levels in the proximal cell layers of infection zone (Fig. 2b,g). At the transition to the fixation zone the NIN-GFP/GFP-NIN protein level dropped markedly similar to *NIN* mRNA (Fig. 2b,d,g,i).

The expression level of *NF-YA1* does not correlate with the level of NIN

NF-YA1 has been shown to be a direct target of NIN in Lotus (*Lotus japonicus*) (Soyano et al., 2013). To test whether the level of NIN correlates with the level of *NF-YA1* expression in Medicago nodules, we determined the expression pattern of *NF-YA1* by RNA *in situ* hybridization. This showed that *NF-YA1* was
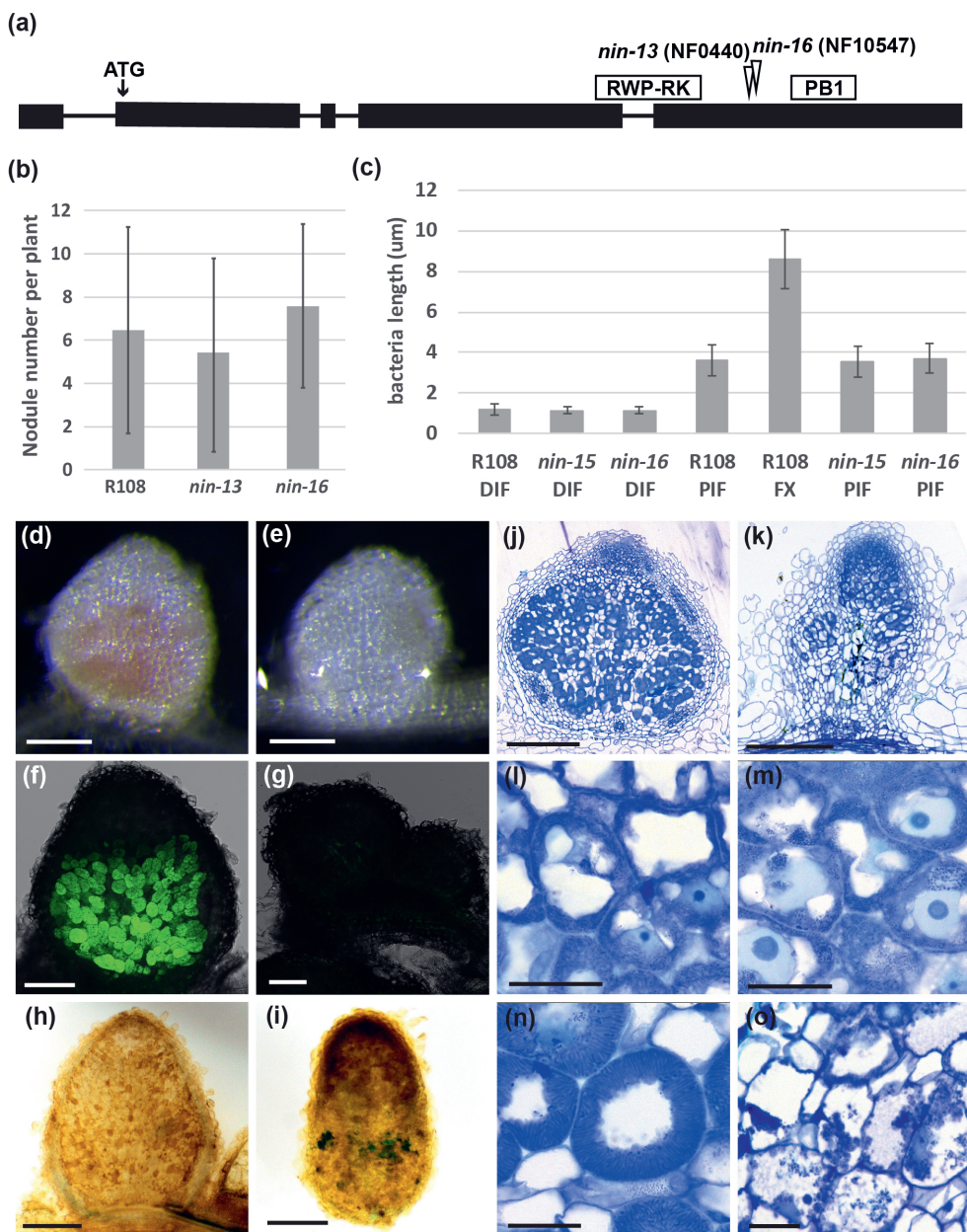
Fig. 3. Phenotype of Medicago *nin-13* and *nin-16* mutants.

(a) Schematic representation of the *Tnt1* insertion site in *nin-13* and *nin-16* alleles. Exons of *NIN* are shown as black boxes, and introns are represented by lines. (b) The number of nodules formed on R108 (wild-type), *nin-13* and *nin-16* are similar. Nodule numbers per plant (n=19) were counted at two wpi. Data are means ± SD. (c) Rhizobia in *nin-13/16* were able to elongate to the size that in wild-type plants

expressed in the entire infection zone like *NIN* (Fig. 1d). However, different from *NIN*, the expression level of *NF-YA1* was high in the proximal part but not detectable in the distal part of the meristem (Fig. 1e). *NF-YA1* was also highly expressed in the distal part of infection zone, after which it decreased (Fig. 1d,e). At the transition to fixation zone, its expression level markedly dropped like *NIN* (Fig. 1f). Like *NIN*, *NF-YA1* is also expressed in the nodule vasculature (Fig. 1d). So the spatial expression patterns of *NIN* and *NF-YA1* were similar, but the expression level of *NF-YA1* did not correlate with that of *NIN*. This difference indicates complexity of regulation of *NF-YA1* by NIN and suggests that additional transcription factors are involved.

nin-13 and *nin-16* nodules showed premature senescence, defence responses and defects in symbiosome differentiation

To study the function of NIN in nodules, we analysed two weak alleles of *nin* (NF10547 and NF0440) which have been reported to form non-functional nodules (Pislariu *et al.*, 2012; Veerappan *et al.*, 2016). NF10547 has been named *nin-16*, and we named NF0440 *nin-13*. *nin-16* is a monogenic recessive mutant (Veerappan et al., 2016), with a *Tnt1* insertion in the last exon, between the DNA binding domain (RWP-RK) and protein-protein interaction domain (PB1) (Fig. 3a). *nin-13* also has a *Tnt1* insertion between the RWP-RK and PB1 domain, but in *nin-13* it is located 22 bp upstream of that in *nin-16* (Fig. 3a). Two weeks after inoculation, *nin-16* had formed white nodules indicating that they were not functional (Fig. 3e). This was confirmed by inoculation with

---

Continued

reached before the transition to the fixation zone. Bacteria length are measured at two wpi. n>70, data are means ± SD. DIF: distal part of infection zone; PIF: proximal part of infection zone; FX: fixation zone. At two wpi with *Sinorhizobium meliloti* 2011, wild-type nodules were pink (d) whereas *nin-16* nodules remained white (e) which indicate the absence of leghemoglobin. Formation of non-functional nodules on *nin-16* was further confirmed by inoculation with *Sinorhizobium meliloti* 2011 carrying the *PronifH::GFP* reporter construct (f, g). *nifH* was highly induced in R108 (f) but not induced in *nin-16* nodules (g). Semi-thin sections stained with toluidine blue show that in comparison with wild-type nodules (j, l, n), an apical meristem was formed on *nin-16* mutant nodule (k), rhizobia were released and divided (m), however, their development was arrested, and premature senescence was induced (o). Potassium permanganate/methylene blue staining shows accumulation of phenolic compound in *nin-16* nodule (i), but not in wild-type nodule (h). *nin-13* shows similar phenotype as *nin-16*, figures shown in Fig. S3. Bars: (d, e) 2 mm; (f, g) 200 μm, (h, i) 500 μm; (j, k) 300 μm, and (l - o) 30 μm.

Sinorhizobium meliloti 2011 containing *PronifH:GFP*. *PronifH:GFP* was highly expressed in the fixation zone of wild-type nodules, but not in *nin-16* nodules (Fig. 3f,g). The number of nodules formed on *nin-16* was similar to that of wild-type (R108) (Fig. 3b), indicating that nodule initiation was not affected. Semi-thin sections of *nin-16* nodules showed that an apical meristem was formed (Fig. 3k), infection threads penetrated nodule cells and rhizobia were released and had divided (Fig. 3m). However, before the rhizobia were fully elongated nodule development was arrested and premature senescence was induced (Fig. 3o). This was also supported by live/dead staining of rhizobia showing many dead bacteria in infected cells (Fig. S2). *nin-13* displayed the same nodule phenotype as *nin-16* (Fig. S3). F1 plants obtained by crossing these two mutants had also the same nodule phenotype confirming that these mutants are allelic (Fig. S4). In the following studies we mainly focused on *nin-16*.

To determine when rhizobial development was blocked in the *nin-16* mutant nodules, we measured the length of the rhizobia (Fig. 3c). In both wild-type and mutant nodules the rhizobia in the infection thread, or just released from it, were about one micrometre (µm) long. In the most proximal part of infection zone of the wild-type nodules, the symbiosomes were about 3.5 µm in length. After the transition to fixation zone, the fully differentiated rhizobia are about 8 µm long. Rhizobia in *nin-13/16* nodules could reach a length of ~3.5 µm (Fig. 3c). This in combination with the lack of *nif* expression, indicated that the block of differentiation occurred at the stage when in wild-type the transition from infection to fixation zone takes place.

The block of rhizobia differentiation might be due to premature senescence and not sufficiently suppressed defence response. The accumulation of phenolic compounds is commonly used to identify a defence response in nodules (Bourcy et al., 2013; Berrabah et al., 2015; Wang et al., 2016; Domonkos et al., 2017; Berrabah et al., 2018). Potassium permanganate/methylene blue staining showed that phenolic compounds indeed accumulated in *nin-16* nodules (Fig. 3i). This means that a defense response was induced in mutant nodules.

PB1 domain of NIN is not essential for nodule development

The *Tnt1* insertion in both *nin-13* and *nin-16* will result in a truncated NIN protein lacking the protein-protein interaction domain (PB1). To test whether the *nin-13/16* phenotype is caused by the absence of the PB1 domain, we introduced a construct encoding a truncated *NIN* (deletion starts at the position of the *Tnt1*
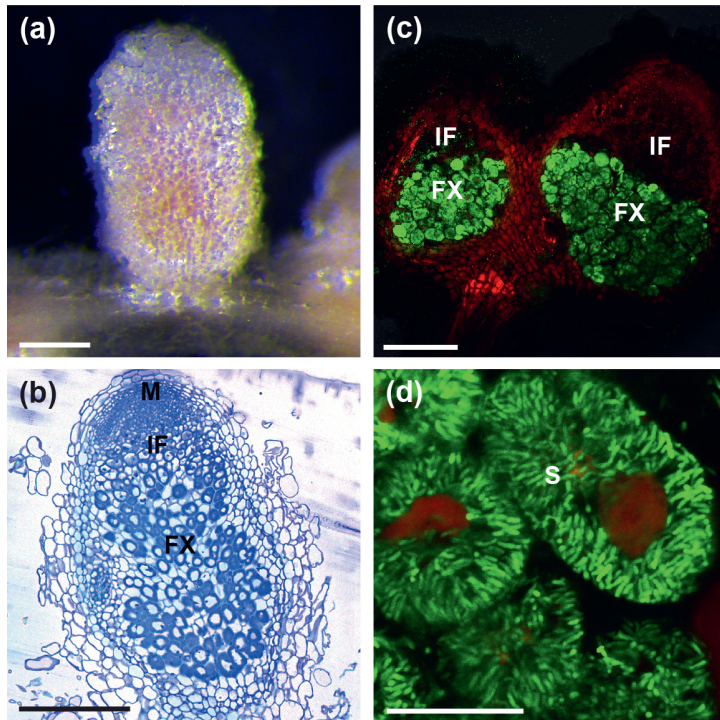
Fig. 4. Wild-type-like pink nodules formed on Medicago *nin* knockout mutant (*nin-1*) roots transformed with truncated *NIN*$_{\Delta PB1}$.

Pink nodules formed on *nin-1* roots transformed with *ProNIN*$_{3C-5kb}$*:NIN*$_{\Delta PB1}$ at four wpi (a). Semi-thin sections of the nodules stained with toluidine blue display normal zonation (b). Confocal image of nodules formed on the *ProNIN*$_{3C-5kb}$*:NIN*$_{\Delta PB1}$ transgenic *nin-1* roots at four wpi with a rhizobial strain carrying *nifH:GFP* showed that *nifH* is highly expressed in the fixation zone (c). Close-up of nodule cells in fixation zone, showing well-developed symbiosomes (d). M, meristem; IF, infection zone; FX, fixation zone; S, symbiosome. Bars: (a) 2 mm; (b) 300 μm; (c) 250 μm and (d) 30 μm.

insertion in *nin-16*, including the PB1 domain) (*ProNIN*$_{3C-5kb}$*:NIN*$_{\Delta PB1}$) into the knockout mutant *nin-1* by hairy root transformation. Surprisingly, pink nodules were formed on transgenic roots four weeks after inoculation (Fig. 4a). Next, we inoculated transgenic roots with rhizobia containing *PronifH:GFP*, and *nifH* was highly induced in the formed nodules (Fig. 4c,d). Semi-thin sections of these nodules revealed that 23 out of 67 (34.3%) were wild-type-like (Fig. 4b), comparable to 12 out of 38 (31.6%) wild-type-like nodules formed on *nin-1* roots transformed with full length *NIN* (*ProNIN*$_{3C-5kb}$*:NIN*). This suggests that the PB1 domain of NIN is not essential for the formation of functional nodules.

Reduced *NIN* mRNA level in the cytoplasm of *nin-13/16* nodule cells

The transition from infection to fixation zone does not occur in *nin-13/16*. As NIN accumulates at a high level at this transition site, we analysed the *NIN* expression pattern in *nin-13/16* nodules by RNA *in situ* hybridization. Like in wild-type (R108) nodules (Fig. 5a), *NIN* was highly expressed in the infection zone of the mutant nodules (Fig. 5b,c). However, there was a remarkable difference between *nin-13/16* and wild-type nodules. *NIN* RNA in *nin-13/16* nodules accumulated to higher levels in nuclei than in the cytoplasm, which is not the case in wild-type nodules (Fig. 5d,e,f). This indicated that the transport of *NIN* mRNA from nucleus to cytoplasm was hampered. In contrast, transport of *NF-YA1* RNA was not affected (Fig. S5), suggesting that the inhibition of mRNA transport was most likely specific for *NIN* in *nin-13/16*. By quantitative real-time PCR (qRT-PCR) we showed that *NIN-Tnt1* fusion transcripts were formed in *nin-13/16* nodules (Fig. S6). In *nin-13,* about 80% of *NIN* transcripts
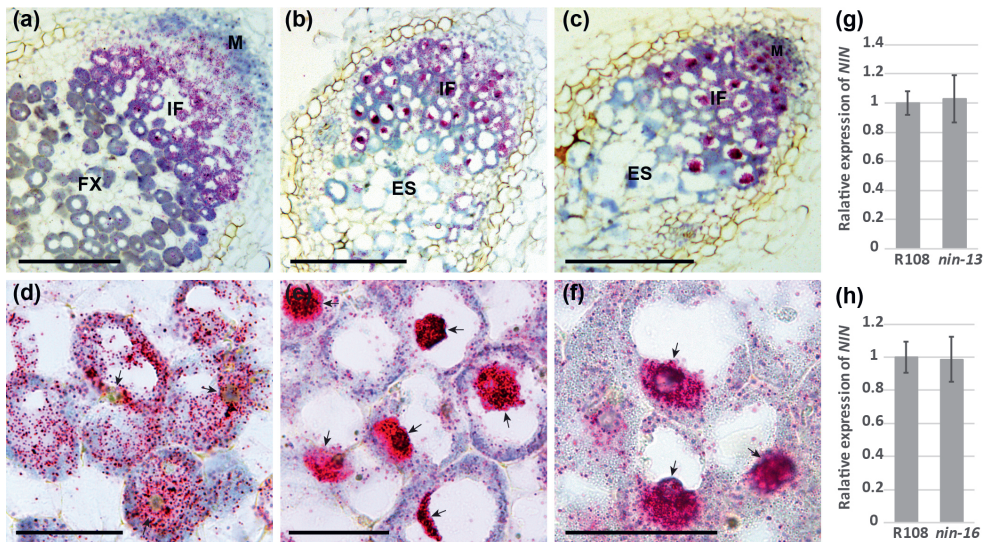
Fig. 5. *NIN* mRNA level reduced in the cytoplasm of nodule cells of Medicago *nin-16*.

RNA *in situ* localization of *NIN* shows substantial part of *NIN* mRNA accumulates in nuclei of *nin-13* (b, e) and *nin-16* nodule cells (c, f), compare with wildtype (R108) nodule section (a, d). Hybridization signals are visible as red dots (a-f). Arrow indicate the nuclei (d, e, f). M, meristem; IF, infection zone; FX, fixation zone; ES, early senescence zone. Bars: (a, b, c) 250 μm and (d, e, f) 40 μm. Quantitative reverse transcription polymerase chain reaction (qRT-PCR) shows that *NIN* expression level in wild-type (R108), *nin-13* and *nin-16* nodules were similar (g, h). Data are means ± SD of three biological replicates.

contained *Tnt1* sequences, and about 60% of *NIN* transcripts contained *Tnt1* sequences in *nin-16* (Fig. S6). Moreover, *nin-16* RNA sequencing data (see below) revealed that some reads were mapped to the intron and promoter region of *NIN* (Fig. S7). The frequency of aberrantly spliced *NIN* transcripts in *nin-16* was between 30.24% and 64.80%. This is markedly higher than in R108, in which this was between 1.39% and 4.37% (Fig. S7). This indicated that the *Tnt1* insertion affected *NIN* transcription and splicing. As export of unspliced and incompletely spliced transcripts from nuclei are prevented (Schmid and Jensen, 2008), we assumed that this might have caused the reduced transport of *NIN* RNA to cytoplasm. By using *NIN* specific primers (upstream of *Tnt1* insertion), we showed that the expression level of *NIN* in wild-type and mutant nodules was similar (Fig. 5g,h). As a substantial part of *NIN* mRNA accumulated in nuclei of *nin-13/16*, the level of *NIN* mRNA that can be translated (in the cytoplasm) was reduced, this might have caused the phenotype of *nin-13/16* nodules.

To test this hypothesis, we introduced *ProNIN$_{3C-5kb}$:NIN$_{ΔPB1}$* into *nin-16* by hairy root transformation. Semi-thin sections of the nodules formed on transgenic roots showed that symbiosome development was restored (symbiosomes reached the length ~8 µm) and the defense/premature senescence phenotype was markedly reduced (Fig. S8). This result supports the conclusion that the *nin-13/16* nodule phenotype is cased by a reduced *NIN* mRNA level, but not the absence of PB1 domain of NIN.

nin-16 nodule transcriptome switched from symbiosis to defence/senescence-related processes

To determine which *NIN*-controlled processes were disturbed in *nin-13/16* mutants, we compared the transcriptomes of two-week-old *nin-16* and wild-type nodules. 31841 gene transcripts were detected in these nodules (Table S2). We identified 2744 genes differentially expressed in *nin-16* nodules compared to wild-type (Table S3). 1281 (46.7%) genes were upregulated and 1463 (53.3%) genes were down regulated in *nin-16* (Table S3).

The histology of the *nin-13/16* nodules indicated that the infection zone was not much affected, whereas a fixation zone was not formed. Consistent with this is that only a small fraction of the genes that are especially expressed in meristem or distal part of the infection zone of wild-type nodule (according to Roux *et al.*, 2014) is differentially expressed in *nin-16*. In contrast, a markedly higher percentage of genes that are especially expressed in fixation zone of wild-type

Table 1. Analysis of genes specifically expressed in different Medicago wild-type nodule developmental zones (Roux et al., 2014) differentially expressed in *nin-16.*
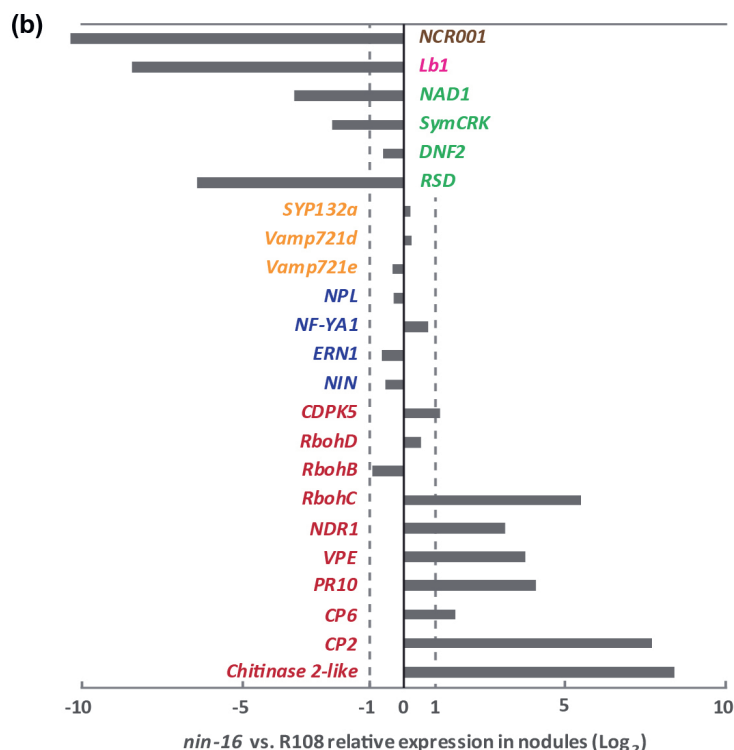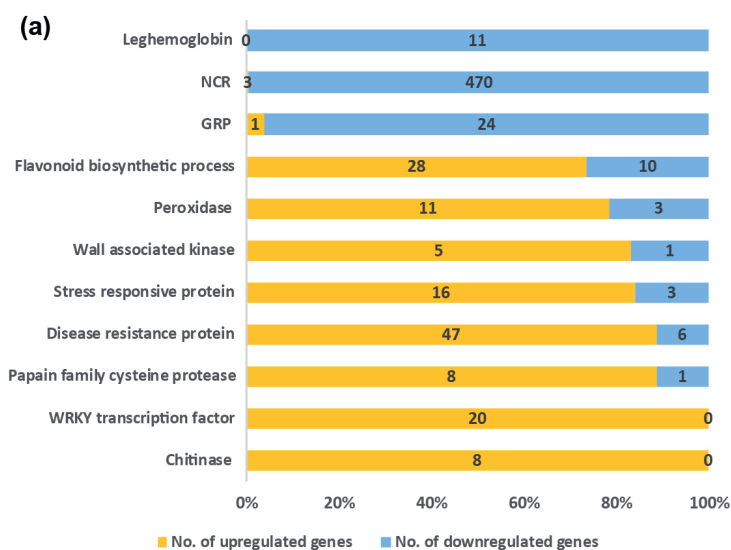
Genes 70% expressed in zone:	Meristem	Distal infection zone	Proximal infection zone	Distal fixation zone (inter-zone)	Proximal fixation zone (zone III)
Total Number of genes	1901	370	191	665	536
Number of genes differentially expressed in *nin-16*	126	24	51	184	82
Percentage of genes of differentially expressed in *nin-16*	6.63%	6.49%	26.70%	27.67%	15.30%
Number of genes up regulated in *nin-16*	101	21	1	2	35
Percentage of genes up regulated in *nin-16*	80.16%	87.50%	1.96%	1.09%	42.68%
Number of genes down regulated in *nin-16*	25	3	50	182	47
Percentage of genes down regulated in *nin-16*	19.84%	12.50%	98.04%	98.91%	57.32%

Darker shade indicate higher percentage of genes differentially expressed in *nin-16.* Higher percentage of genes up regulated in *nin-16* shown in darker red colour; Higher percentage of genes down regulated in *nin-16* shown in darker green colour,

nodules is differentially expressed (Table 1). However, for the genes especially expressed in the proximal part of the infection zone of wild-type nodules the percentage of differentially expressed genes *in nin-16* was also high (Table 1). The few differentially expressed genes which are especially expressed in distal part of infection zone were mainly up regulated (Table 1). Several of these up regulated genes are related to a defence response (Table S4). In contrast, differentially expressed genes that are especially expressed in the proximal part of infection zone or distal part of fixation zone [interzone II/III according to (Roux et al., 2014)] in wild-type nodules, were mainly down regulated (Table 1). Most of these genes are related to symbiotic functions like *leghemoglobin* and *NCR* genes (Table S4).For the genes especially expressed in proximal part of fixation zone [(zone III according to Roux *et al.*, 2014)], which expression level was changed, similar numbers were up and down regulated, respectively (Table 1). The up regulated genes were mainly related to senescence and defence and down regulated genes included symbiotic genes (Table S4). We do not know in which part of the *nin-13/16* nodules the genes are upregulated and this can be different from their expression in wildtype nodules. For example, the defense related genes that are expressed in the meristem of wildtype nodules might be induced in the early senescene zone of *nin-13/16* nodules.

Consistent with the fact that the early nodule developmental stages were not much affected, the expression level of genes required for intracellular infection were not significantly changed. This included *Ethylene response factor* (*ERF*), transcription factor *required for nodulation 1* (*ERN1*) and direct targets of NIN, *NF-YA1* and *nodulation pectate lyase* (*NPL*) (Fig. 6b) (Xie et al., 2012; Soyano et al., 2013; Laporte et al., 2014). Release of rhizobia from infection threads occurs in the distal part of the infection zone. The symbiotic exocytosis pathway related genes *Vamp721d, Vamp721e* and *SYP13II* are important for release (Ivanov et al., 2012; Huisman et al., 2016). Their level of expression in *nin-16* nodules was not significantly changed (Fig. 6b) which was consistent with the occurrence of release.

In *nin-16* nodules rhizobia elongated and reached a size that in wild-type nodules is characteristic for rhizobia in the most proximal part of the infection zone (Fig. 3c). The elongation of rhizobia is initiated in *nin-16* nodules (Fig. 3c) and this was consistent with an unaffected or even upregulated expression level of some *NCR* genes that were expressed in the infection zone (Fig. 6a, Table S2). The block in further differentiation might be caused by the lack of fixation zone formation, where other *NCR* genes are expressed. The accumulation of phenolic compounds and the induction of premature senescence suggested that defence/senescence was not sufficiently suppressed in *nin-13/16* nodules. Consistent with this, there were many defence/senescence- related genes significantly upregulated in *nin-16* nodules (Fig. 6a, Table S5) including well-known marker genes like *PR10* (16.9-fold upregulated) and *NDR1* (8.7-fold upregulated) (Fig. 6b; Table S3). The expression of flavonoid biosynthesis genes also supported the occurance of defense response. Among 38 differentially expressed genes, 28 were upregulated. Most of these genes have putative chalcone synthase activity or chalcone isomerase activity which are involved in formation of flavonoid scaffolds. In contrast, all the ten down regulated genes have putative glucosyltransferase activity which is related to the diversity of flavonoids (Figure 6a, Table S5) (Saito et al., 2013). *Cysteine proteases* (*CPs*) and *vacuolar processing enzyme* (*VPE*) have been shown to play a crucial role in the onset of nodule senescence and their expression can also be the consequence of plant defence responses (Hara-Nishimura et al., 2005; Sheokand and Brewin, 2003; Grudkowska and Zagdańska, 2004; Wang et al., 2016). In *nin-16* nodules, *CP2* and *VPE* are 202.6-fold and 13.6-fold upregulated, respectively (Fig. 6b; Table S3), and they were among the most highly expressed genes (Table S2). Moreover, genes involved in suppression of plant defence in nodules like *DNF2*,

(a)

Category	No. of upregulated genes	No. of downregulated genes
Leghemoglobin	0	11
NCR	3	470
GRP	1	24
Flavonoid biosynthetic process	28	10
Peroxidase	11	3
Wall associated kinase	5	1
Stress responsive protein	16	3
Disease resistance protein	47	6
Papain family cysteine protease	8	1
WRKY transcription factor	20	0
Chitinase	8	0

(b)

nin-16 vs. R108 relative expression in nodules (Log$_2$)

Fig. 6. Medicago *nin-16* transcriptome is consistent with the reduced suppression of defence and early senescence.

Transcriptional regulation of gene families/metabolism pathways in nodules (a). The classification of each gene in the families/metabolism pathways is based on

SymCRK, *NAD1* and *RSD* were 1.6-, 4.7-, 10.6- and 83.9-fold down regulated in *nin-16* (Fig. 6b; Table S2). Taken together, these data support the conclusion that *nin-16* transcriptome switched from symbiosis to defence/senescence-related processes.

Defence/senescence is suppressed in the infection zone of *nin-13/16* nodules

Many plant defence/senescence-related genes were highly induced in *nin-16* nodules. To determine whether the induction of these genes already occurred in the infection zone, we used *CP2* as a marker gene as it was the highest expressed defence/senescence-related gene in *nin-16*. *In situ* hybridisation showed that in two weeks old mutant nodules, *CP2* was extremely highly expressed in the infected cells of the premature senescence zone, but not in the infection zone or other nodule tissue (Fig. 7b,c). This suggested that the defence and early senescence response was well suppressed in the nodule infection zone, but the suppression was not sufficient at the stage when transition to fixation zone would occur.

symcrk and *nad1* mutants form nodules in which defence responses are induced and symbiosome differentiation gets arrested (Berrabah et al., 2014b; Wang et al., 2016; Yu et al., 2018; Domonkos et al., 2017). This is similar to the nodule phenotype of *nin-13/16*. *rsd* and *dnf2* mutants also have a similar nodule

---

Continued

the protein function indicated in the Mt 5.0 annotation or gene ontology terms. NCR, Nodule Cystein-Rich peptide; GRP, Nodule-specific Glycine Rich Peptide. Analysis of representative genes from the transcriptome data (b). Genes required for primordium initiation and infection formation (blue colour), including *NIN* (*Nodule inception*) itself, *Ethylene response factor required for nodulation 1* (*ERN1*), *Nuclear transcription factor Y subunit A-1* (*NF-YA1*) and *nodulation pectate lyase* (*NPL*). The symbiotic exocytosis pathway related genes (yellow colour) required for rhizobial release, including *VAMP721d* and *VAMP721e* which are members of the *vesicle-associated membrane protein 721* family*,* and *SYNTAXIN OF PLANTS 132* (*SYP132*). *Nodule-specific cysteine-rich* (*NCR*) *peptides* (brown colour) and *leghemoglobin* (*LB*) (pink colour), are known to be required for bacteria differentiation and nitrogen fixation. Genes related to defence and senescence (red colour), including *calcium- dependent protein kinase* (*CDPK*), *respiratory burst oxidase homologs* (*Rboh*), *cysteine proteinases* (*CPs*), *non-race-specific disease resistance 1* (*NDR1*), *vacuole processing enzyme* (*VPE*), *pathogenesis-related protein 10* (*PR10*) and Chitinase genes. See detailed information about selected genes in Table S1.
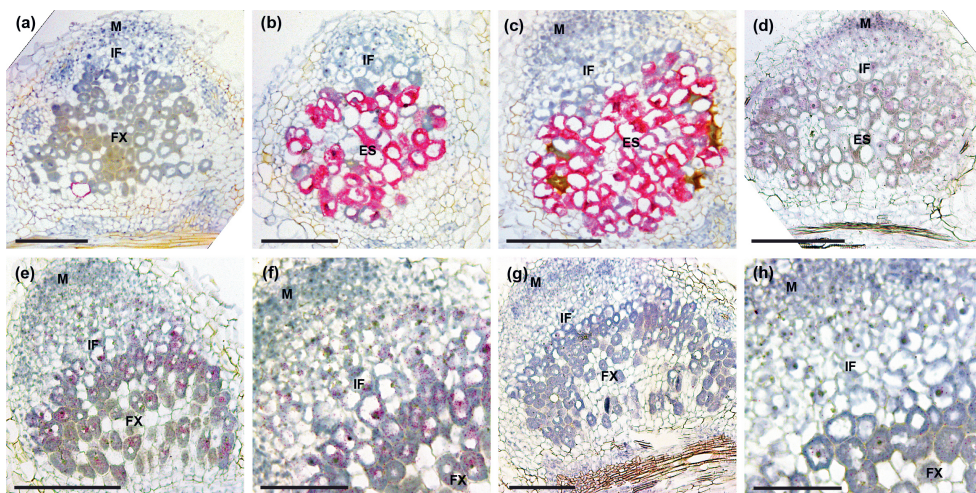
Fig. 7. Expression pattern of defence/senescence response and suppression genes.

RNA *in situ* localization of defence/senescence response gene *CP2* in Medicago R108 (a), *nin-13* (*b*) and *nin-16* (c) nodules at two wpi. Overview (e, g) and close-up (f, h) pictures of expression patterns of defence suppression genes *SymCRK* (e, f) and *NAD1* (g, h) in two wpi R108 nodules. (d) Expression pattern of *SymCRK* in two wpi *nin-16* nodule. Hybridization signals are visible as red dots. M, meristem; IF, infection zone; FX, fixation zone; ES, early senescence zone. Bars: (a, b, c, d, e, g) 200 μm and (f, h) 100 μm.

phenotype (Bourcy *et al.*, 2012; Sinharoy *et al.*, 2013). However, the expression level of *RSD* was very low and *DNF2* expression level in *nin-16* was not significantly changed (Table S2). To determine when plant defence suppressing genes were activated during nodule development, we checked the expression pattern of *SymCRK* and *NAD1*. The RNA *in situ* hybridization showed that both *SymCRK* and *NAD1* were expressed in the infected cells of the infection zone and their expression level increased from the distal to the proximal part (Fig. 7e-h) which is similar to the *NIN* expression pattern (Fig. 1a,5a). However, *SymCRK* and *NAD1* remained expressed at a relatively high level in the fixation zone (Fig. 7e-h), whereas *NIN* expression markedly dropped at the transition from infection to fixation zone (Fig. 1a-c,5a). To determine where in the nodule the expression of defence suppressing genes was affected, we studied *SymCRK* expression in *nin-16* nodules. RNA *in situ* hybridization showed that *SymCRK* was expressed in the infection zone (Fig. 7d). In the senescing cells, *SymCRK* mRNA was hardly detectable.

These results indicate that the expression level of defence suppressing genes in the infection zone is sufficient to suppress the defence response and early senescence, whereas at the transition to fixation zone, the suppression is not sufficient anymore.

Discussion

Here we showed that NIN plays, in addition to its previously identified role in infection and nodule organogenesis, a key role in late steps of nodule development leading to a functional symbiosis. *nin-13/16* nodules display a rather normal infection zone, whereas a fixation zone is not formed. Instead, a zone with defence responses and premature senescence occurs. The phenotype is most likely caused by reduced *NIN* mRNA levels in the cytoplasm. Consistent with this phenotype, transcriptome analyses of *nin-16* nodules showed that defence/senescence-related genes are highly induced and defence suppressing genes are reduced. However, although defence/senescence seems well suppressed in the infection zone, the transcriptome is most likely already markedly changed in the proximal part of the nodule infection zone. NIN mRNA and protein are accumulated at the highest level in the proximal part of the infection zone. At the transition to fixation zone, it suddenly dropped to a much lower level, suggesting that NIN is involved in the switch from infection to fixation zone.

In both *nin-13* and *nin-16, Tnt1* is inserted upstream of the region that encodes the PB1 domain which is located at the carboxy-terminus of NIN. Therefore, we expected that loss of this domain would have caused the mutant nodule phenotype. However, by introducing truncated $NIN_{\Delta PB1}$ into a *nin* knockout mutant (*nin-1*), we showed that PB1 domain is not necessary for the formation of wildtype-like nodules in Medicago A17 (Fig. 4). The importance of the PB1 domain is underlined by being conserved in NLP family, including NIN. The PB1 domain has been shown to be essential for the formation of homo- and hetero-dimers of Arabidopsis NLP6 and NLP7 (Guan et al., 2017). In addition, Medicago NLP1 is required for suppression of nodulation by nitrate, and this suppression most likely involves a physical interaction of NIN and NLP1 by their PB1 domains (Lin et al., 2018a). Therefore, the PB1 domain of NIN may be involved in regulating its activity in response to for example exogenous nitrate, whereas it is not essential for intrinsic nodule development.

nin-13/16 nodule phenotype most likely is caused by reduced *NIN* mRNA levels in the cytoplasm. *In situ* hybridization showed that a substantial part of *NIN* RNA accumulated in *nin-13/16* nuclei. Further, by introducing truncated *NIN* into *nin-16*, symbiosome development was restored, and the defence and early senescence phenotype was markedly reduced. This confirmed that the mutants nodule phenotype is caused by reduced *NIN* mRNA level, but not the absence of the PB1 domain of NIN. The *NIN* RNA transcribed from the *nin-16* allele was altered by the *Tnt1* insertion. Therefore, it might be that this improper nature of the *NIN* transcripts caused (partial) retainment of it in nuclei. The nuclear retention of transcripts might also occur in other mutants with *Tnt1* insertion. Although the qPCR data showed that *NIN-Tnt1* chimeric sequences were present, it did not prove that full length *NIN-Tnt1* chimeric sequences were formed, as promoter and terminator regions exist in the identical 5' and 3' long terminal repeats (LTR) of *Tnt1* (Grandbastien et al., 1989; Grandbastien, 2015). It has been shown that some LTR regulatory regions can induce trancritiption in opposite direction (Grandbastien et al., 1989; Grandbastien, 2015). This could explain the reads mapped to promoter region of *NIN* in *nin-16* as readout transcript of LTR.

In Pislariu *et al.*, (2012), more mutants with a *Tnt1* insertion in *NIN* were described. However, except *nin-13* (NF0440) they all showed Nod- phenotype, including two mutants (NF0825 and NF2700) which have *Tnt1* insertion at a position that is downstream of that in *nin-13/16*. However, it cannot yet be concluded that the truncated NIN proteins which are sligtly longer than in *nin-13* and *nin-16* are not functional. In both *nin-13* and *nin-16, NIN* mRNA is retained in the nucleus. It seems probable that the *Tnt1* insertions in NF0825 and NF2700 also cause such *NIN* mRNA retention. In case it is stronger than in *nin-13/16,* this could explain the Nod- phenotype of these mutants. It is also possible that other *Tnt1* insertions in these mutants caused Nod- phenotype, as it has not been studied yet whether the nodule phenotype co-segregates with *Tnt1* insertion in *NIN*.

Several NLPs are located in the cytoplasm under low nitrate condition and are translocated to the nucleus under high nitrate condition (Marchive et al., 2013; Cao et al., 2017; Lin et al., 2018a). Here we showed that NIN is located in nuclei in all nodule cells where it is expressed. For the nitrate-triggered translocation to nuclei Ser205 of AtNLP7 becomes phosphorylated. This serine is conserved in the several NLPs, but is lacking in NIN (Liu et al., 2017). So NIN might have lost the information required for retention in the cytoplasm.

In Medicago, *NIN* is expressed in the infection zone and meristem. This is consistent with a previous study on pea (*Pisum sativum*) (Borisov *et al.*, 2003). Both *NIN* mRNA and NIN protein had reached their highest level in the proximal part of infection zone in Medicago nodules, and this suddenly dropped at the transition to fixation zone. However, it is not clear whether in pea the decrease of *NIN* expression also occurs at the switch to interzone/distal fixation zone or somewhat later (Borisov *et al.*, 2003). The sudden change of the NIN levels at transition from infection to fixation zone coincides with several other rapid changes like the accumulation of starch, the induction of bacterial *nif* genes and the radial alignment of the symbiosomes (Gavrin et al., 2014). Rapid changes in protein/mRNA activity/level often involve negative and positive feedback mechanisms. For example, in the Arabidopsis stem cell niche an asymmetric cell division initiates the formation of root endodermis and cortex. This involves the SHORT ROOT (SHR) /SCARECROW (SCR) complex. This complex induces the expression of *cyclinD6,1* that inactivates RETINOBLASTOMA-RELATED, the repressor of SCR activity. By this positive feedback loop SHR-SCR complex activity increases and triggers asymmetric division. It is followed by a negative feedback involving degradation of the SHR-SCR complex (Cruz-Ramírez et al., 2012). So, the accumulation of NIN to a high level which is followed by a sudden drop is consistent with a role in a molecular switch. Whether and how NIN is involved in a molecular circuit controlling the switch from infection to fixation zone remains to be studied.

In the previous studies on *nin-13/16* (Veerappan et al., 2016; Pislariu et al., 2012), the accumulation of phenolic compounds was not described. The level of accumulation of phenolic compounds in these two mutants seems lower in comparison with some other mutants (Wang et al., 2016; Domonkos et al., 2017; Bourcy et al., 2013; Berrabah et al., 2014b), therefore, it might have been overlooked in previous studies. Further, it has been shown that in some mutants, for example *dnf2* (Berrabah et al., 2014a), the accumulation of phenolic compounds is conditional. Therefore, it can not be excluded that less phenolic compounds had accumulated in *nin-13/16* nodules due to different growth conditions. *nin-13* previously has been reported as Nod- at 21 days post inoculation with rhizobia (Pislariu et al., 2012). The reason for this difference is not clear, as in all our experiments the nodule number formed on the mutant is similar to that of the wild-type (R108).

Although *nin-13/16* has been reported before, the phenotypic characterization was not detailed (Veerappan et al., 2016; Pislariu et al., 2012). Here we analysed

nin-13/16 nodule phenotype explicitely. Both mutant nodules make an infection zone that at the cytological level is rather normal. Rhizobia are released from infection threads, they multiply and the symbiosomes start to elongate, reaching the size that is obtained in the proximal part of the infection zone in WT nodules. In contrast, processes that occur at the transition to fixation zone do not take place. For example, *nif* genes are not induced, symbiosomes do not further elongate and amyloplasts do not accumulate at the periphery of the infected cells. This phenotype is in part consistent with the transcriptome data. Genes that are especially expressed in the fixation zone are down regulated (except genes related to senescence and defence response). In contrast, expression levels of genes especially active in the distal infection zone are in most cases not affected, except defence-related genes which are upregulated. It is not known where they are upregulated in *nin-16* nodules. Further, the expression level of many genes that are especially expressed in the proximal part of the infection zone is already affected, the genes related to symbiotic functions are down regulated.

Symbiosome development gets arrested in *nin-13/16* nodules and defence and early senescence are induced. The defence suppressing genes *SymCRK* and *NAD1* are markedly down regulated in *nin-16* nodules. Loss of function mutations in these genes cause a similar symbiosome phenotype as *nin-16* (Berrabah et al., 2014b; Wang et al., 2016; Domonkos et al., 2017). Therefore, it is possible that symbiosome development was blocked due to the defence/ senescence-related responses that were not sufficiently suppressed. In addition, other studies have shown that arrest of symbiosome development does not induce defence responses (Berrabah et al., 2015) So, defence is more likely a cause than a result of blocking symbiosome development.

Laser capture transcriptome data suggested that *NF-YA1* is highly expressed in the nodule meristem and distal part of infection zone (Roux et al., 2014), Our *in situ* hybridisation showed that *NF-YA1* is not expressed in the distal part of meristem, but it is expressed at a high level in the proximal part of meristem and distal part of infection zone, where NIN is only present at a very low level. This supports that *NF-YA1* expressioin can be regulated not only by NIN but also other transcription factor(s).

SymCRK might also be a direct target of NIN as indicated by chromatin immunoprecipitation sequencing (ChIP-seq) analyses on Lotus roots ectopically expressing *NIN* (Soyano et al., 2014; Liu et al., 2019a). In the infection zone *NIN*

is expressed at a rather high level, so it could control *SymCRK* expression. In case NIN also directly controls *SymCRK* expression in the fixation zone, then a very low level of NIN might still be sufficient, considering that a low level of NIN is sufficient to induce a high level of *NF-YA1* expression. Alternatively, it has been shown that members of the NLP family can also bind to the promoter of NIN target gene *CLE-RS2* (Nishida et al., 2018). Moreover, NLPs are highly expressed in Medicago nodules, especially MtNLP1/2/3 (Lin et al., 2018a). Therefore, it is possible that they take over the function of NIN to maintain expression of this defence suppression gene in the fixation zone.

So, in addition to its function in nodule organogenesis and infection, NIN has a major function in establishing a functional symbiosis in Medicago nodules. It will be interesting to see whether this function is shared with other legumes or even actinorhizal plants. Such studies in an evolutionary context can provide first insight when this additional function of NIN evolved or whether it is related to a function of the ancester of NIN that was recruited when nodulation evolved.

Acknowledgements

Author Contribution

JL, OK, and TB designed the research. JL, MR, and OK performed experiments. JL, MR, TZ, JK, OK and TB analysed data. JL, OK, and TB wrote the manuscript.

Supporting Information

Supplemental Data Can be download from the **New Phytologist**.

https://doi.org/10.1111/nph.17215

Fig. S1 Expression profile of Medicago NIN/Medtr5g099060 (probe id. Mtr.28094.1.S1.st) based on *M. truncatula* Gene Atlas in various tissues.

Fig. S2 Live/dead staining shows prematurely death of rhizobia in Medicago *nin-16* nodules.

Fig. S3 Medicago *nin-13* mutant nodules show the same phenotype as *nin-16.*

Fig. S4 F1 plants obtained by crossing Medicago *nin-13* and *nin-16* showed the same nodule phenotype as the parental plants.

Fig. S5 *NF-YA1* expression pattern in Medicago *nin-16* nodule.

Fig. S6 *Tnt1* was transcribed in Medicago *nin-13* and *nin-16* mutant nodules.

Fig. S7 *NIN* RNA transcribed from the Medicago *nin-16* allele was altered by *Tnt1* insertion*.*

Fig. S8 Complementation of Medicago *nin-16* nodule phenotype with *ProNIN$_{3C-5kb}$:NIN$_{\Delta PB1}$* and *ProNIN$_{3C-5kb}$:NIN.*

Table S1 Primers used in this study.

Table S2 Genes with transcripts detected in Medicago *nin-16* or R108 (wild-type). (separate Excel file)

Table S3 Genes differentially expressed in Medicago *nin-16*. (separate Excel file)

Table S4 Genes specifically expressed in different Medicago wild-type nodule developmental zones differentially expressed in *nin-16*. (separate Excel file)
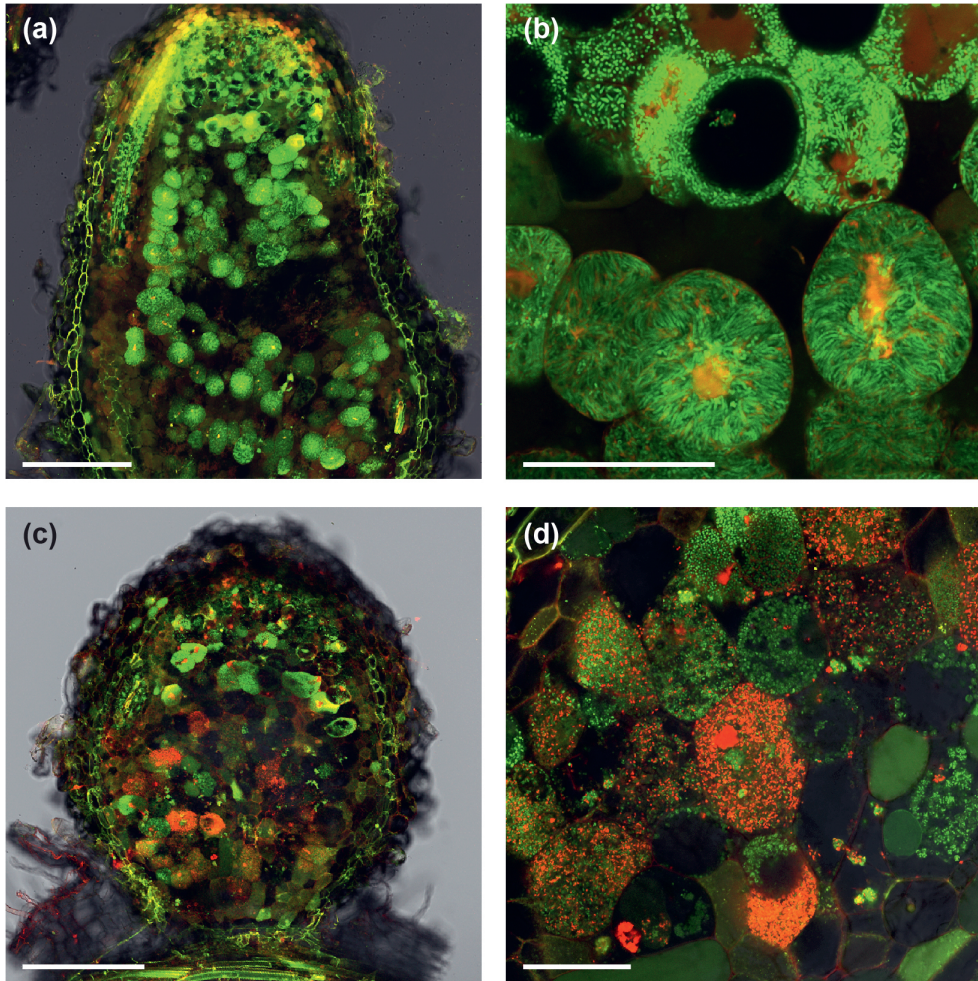
Table S5 Differentially expressed gene families/metabolism pathways in Medicago *nin-16.* (separate Excel file)
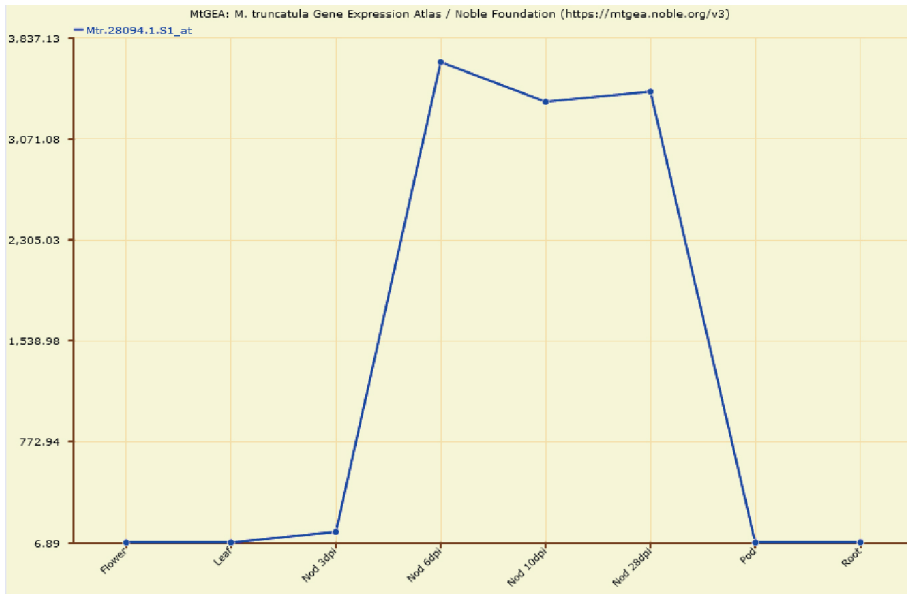
Fig. S1. Expression profile of Medicago NIN/Medtr5g099060 (probe id. Mtr.28094.1.S1.st) based on M. truncatula Gene Atlas in various tissues.

Mt *NIN* is specifically expressed in nodules.

Fig. S2. Live/dead staining shows prematurely death of rhizobia in Medicago *nin-16* nodules.

Sections of two weeks post inoculation (wpi) nodules, imaged by a confocal microscope after live/dead staining assay. Green (SYTO 9) and red (propidium iodide) stain alive and dead bacteria respectively. Compare with wildtype (a, b), the early death of the bacteria in the *nin-16* nodule was detected (c, d). Bars: (a, c) 250 μm and (b, d) 50 μm.
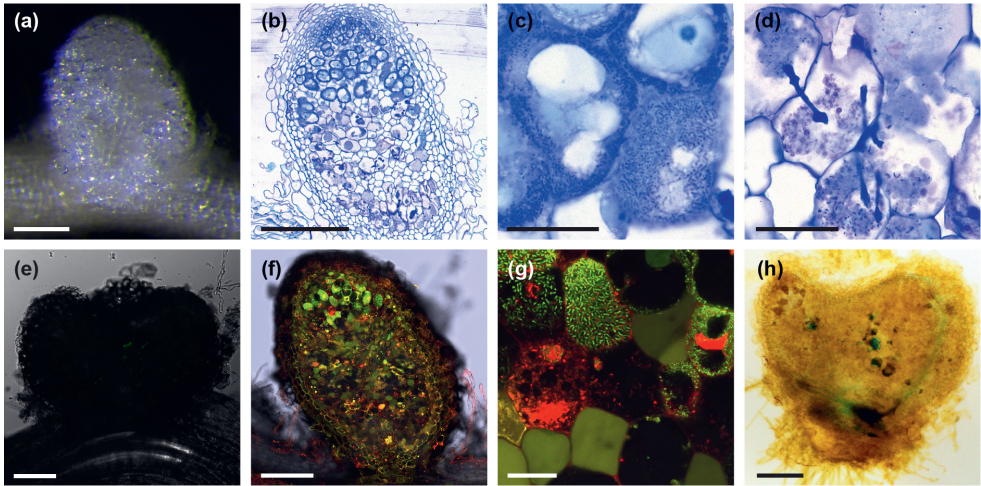
**Fig. S3. Medicago *nin-13* mutant nodules show the same phenotype as *nin-16*.**

Like *nin-16*, *nin-13* formed white nodules at two wpi (a). Inoculation with rhizobial carrying *nifH:GFP* showed that *nifH* was not induced in *nin-13* nodules (e). Sections of these nodules showed that meristem was formed (b), rhizobia were released and divided (c), but bacteria differentiation were arrested, and premature senescence was induced (d). The bacteria death was confirmed by live/dead staining (f, g). Green (SYTO 9) and red (propidium iodide) stain alive and dead bacteria respectively. Potassium permanganate/methylene blue staining shows accumulation of phenolic compound in *nin-13* nodules (h). Bars: (a) 2 mm; (b) 300 μm; (c, d, g) 30 μm; (e, f) 250 μm and (h) 500 μm.
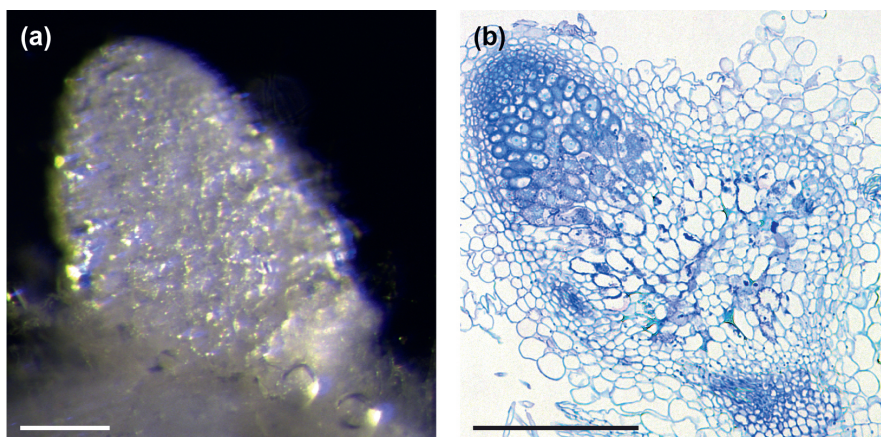
Fig. S4. F1 plants obtained by crossing Medicago *nin-13* and *nin-16* showed the same nodule phenotype as the parental plants.

Transmitted light macroscopy images of root nodules formed F1 plants obtained by crossing *nin-13* and *nin-16* (a). Semi-thin sections of these nodules stained with toluidine blue display the same nodule phenotype as the *nin-13* and *nin-16* (b). Bars: (a) 2 mm and (b) 300 µm.
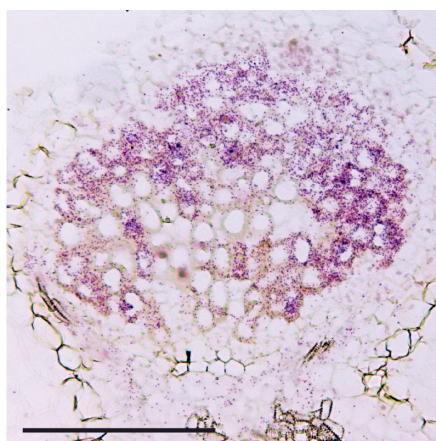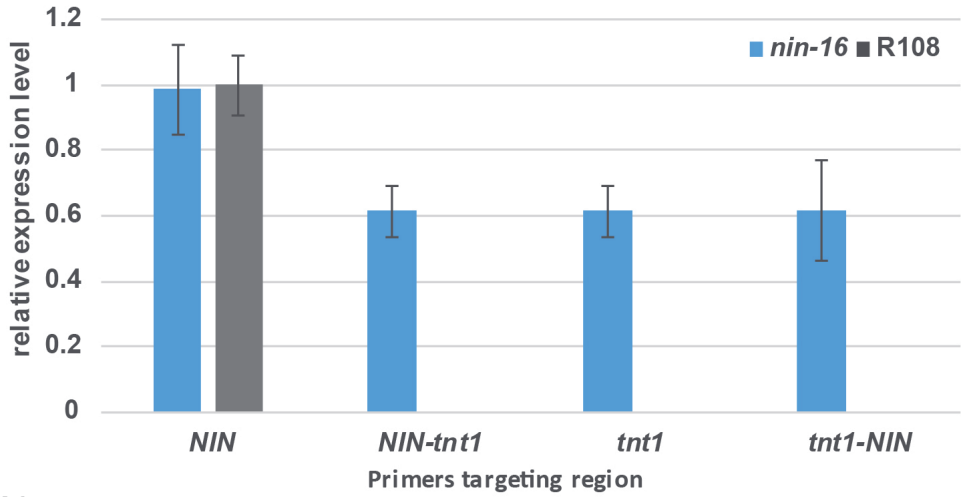

Fig. S5. *NF-YA1* expression pattern in Medicago *nin-16* nodule.

RNA *in situ* localization of *NF-YA1* in *nin-16* nodule at two wpi. Hybridization signals are visible as red dots. Scale bar: 200 µm.

(a)

(b)

4
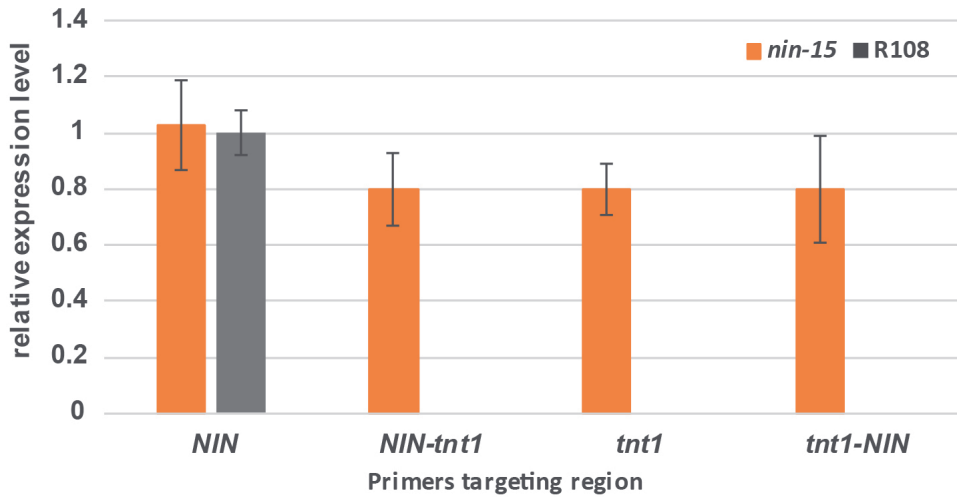
Fig. S6. ***Tnt1* was transcribed in Medicago *nin-13* and *nin-16* mutant nodules.**

Quantitative real-time (qRT-PCR) using different primer sets targeting *NIN*, upstream junction of *NIN* and *Tnt1* (*NIN-Tnt1*)*,* within *Tnt1* insertion (*Tnt1*) and downstream junction of *Tnt1* and *NIN* (*Tnt1-NIN*) show that *Tnt1* was transcribed in *nin-13* (a) and *nin-16* (b) nodules. Data are means ± SD of three biological replicates.

(a)

(b)
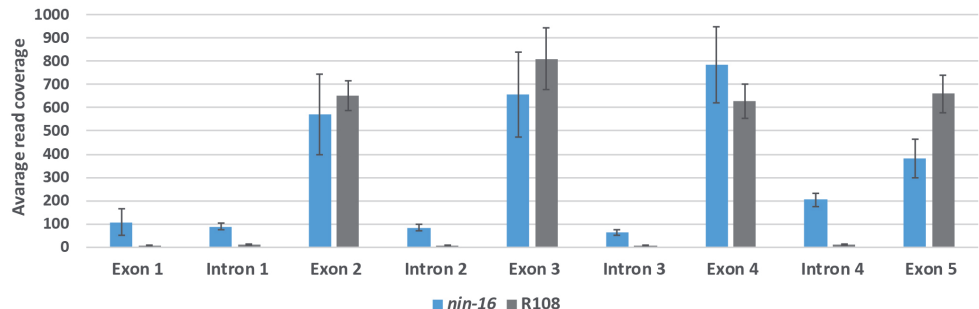
(c)

Non-spliced intron rate of *NIN* in R108 and *nin-16*

	Intron 1	Intron 2	Intron 3	Intron 4	Sum intron 1-4
R108	1.39%	0.66%	0.95%	1.38%	4.37%
nin-16	12.98%	12.26%	9.32%	30.24%	64.80%

Fig. S7. *NIN* RNA transcribed from the Medicago *nin-16* allele was altered by *Tnt1* insertion.

(a) Mapping of the reads at the *NIN* locus in R108 (above) and *nin-16* (below). Arrow heads indicate the reads that mapped in *NIN* intron regions and arrow indicates the much longer 5' UTR in *nin-16*. (b) Quantification of average read coverage for each exon/intron region of *NIN* in *nin-16* and R108. The read coverage was determined with mosdepht (version 0.3.0) using default settings (Pedersen and Quinlan,
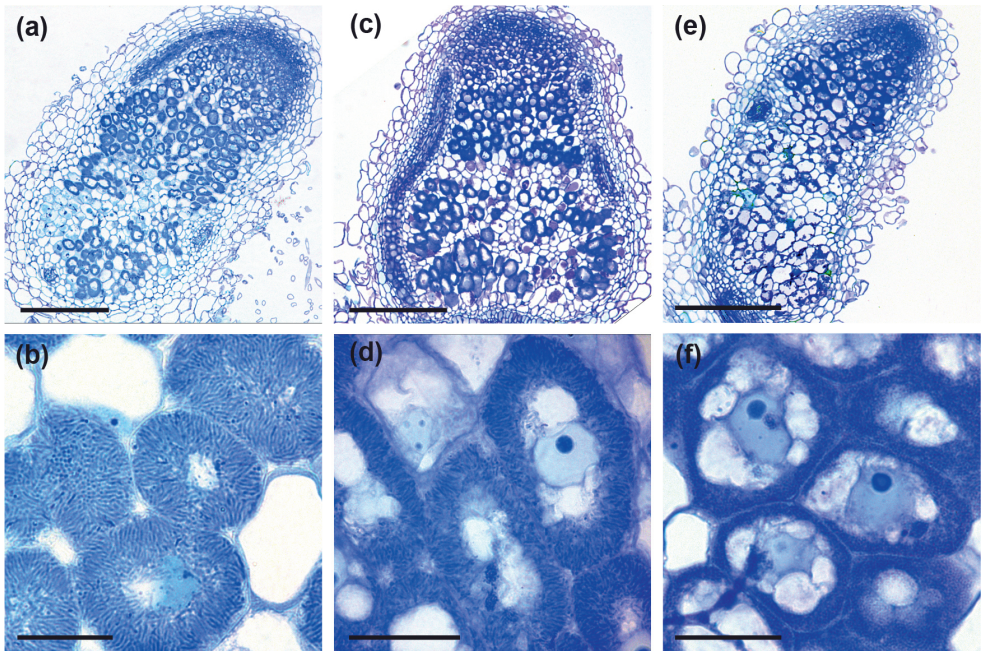
Fig. S8. Complementation of Medicago *nin-16* nodule phenotype with *ProNIN*_{3C-5kb}:*NIN*_{ΔPB1} and *ProNIN*_{3C-5kb}:*NIN*.

Overview (a, c, e) and close-up (b, d, f) images of semi-thin longitudinal sections of nodules formed on *nin-16* roots transformed with *ProNIN*$_{3C-5kb}$:*NIN*$_{ΔPB1}$ (a, b), *ProNIN*$_{3C-5kb}$:*NIN* (c, d) and empty vector (e, f). Fully elongated symbiosomes were observed in five out of 30 sectioned nodules formed on *nin-16* roots transformed with *ProNIN*$_{3C-5kb}$:*NIN*$_{ΔPB1}$ and five out of 60 nodules formed on *nin-16* roots transformed with *ProNIN*$_{3C-5kb}$:*NIN,* but not in nodules formed on the roots transformed with empty vector (n=55). In addition, the defence-related phenotype was markedly reduced in the complemented nodules. Bars: (a, c, e) 300 µm and (b, d, f) 30 µm.

Continued

2018). Data are means ± SD. (c) Non-spliced intron rate of *NIN* RNA in *nin-16* and R108. Non-spliced intron rate was calculated by intron reads coverage divided by the average exon reads coverage. Calculation of average read coverage in exon regions were based on the exons 2, 3 and 4. This is because *nin-16* contains a *Tnt1* insertion in exon 5 which leads to low read coverage, and exon 1 is in the non-coding region and there the read coverage is much lower than in the coding region. The average read coverage in exons of R108 and *nin-16* were similar, in R108 it is 697.23 and in *nin-16* it is 671.39. Theoretically it is possible that only one or all introns are aberrantly spliced. Therefore, we can only indicate a range within which the transcripts are wrongly spliced. The frequency of aberrantly spliced *NIN* transcripts in *nin-16* is between 30.24% (all introns are aberrantly spliced) and 64.80% (a single intron is aberrantly spliced). This is markedly higher than the frequency in R108 which is between 1.39% and 4.37%.

Table S1. Primers used in this study.

Name	Sequence (5′→3′)
qPCR-NIN-F	TACTTTGCCGGAAGCCTAAA
qPCR-NIN-R	ATCTGTATGGCACCCTCTGC
qPCR-NIN-Tnt1-F	GAGTTGATCATGCCTTCATGC
qPCR-NIN-Tnt1-R	GGTTGGCTACCAAACCAAAG
qPCR-Tnt1-F	GCGTTTGAAATCCCAGAGAG
qPCR-Tnt1-R	AACCGAACACCTTCAGATGC
qPCR-Tnt1-NIN-F	TCAGAAGGGTTTTCCACGTAA
qPCR-Tnt1-NIN-R	CCACAGTTGGTCTTGGAGGT
qPCR-ACTIN2-F	TGGCATCACTCAGTACCTTTCAACAG
qPCR-ACTIN2-R	ACCCAAAGCATCAAATAATAAGTCAACC
Genotyping-NIN-F	TGCTAATGGTGGTGATGGTAAT
Genotyping-NIN-R	GGTTAAATCGCCTTGCAATCTC
Genotyping-Tnt1-R	TGTAGCACCGAGATACGGTAATTAACAAGA
NIN$_{\Delta PB1}$-F	**CACC**ATGGAATATGGTGGTGGGTT
NIN$_{\Delta PB1}$-R	GTAGTCCTGGATATTAATATTAGATGCA
NIN$_{\Delta PB1}$-35Ster-BP-F	**GGGGACAGCTTTCTTGTACAAAGTGGAA**ATGGAATATGGTGGTGG-GTTAGTG
NIN$_{\Delta PB1}$-35Ster-BP-R	**GGGGACAACTTTGTATAATAAAGTTGC**TCACTGGATTTTGGTTTTAG-GAATTA
NIN-CDs-F	**CACC**ATGGAATATGGTGGTGGGTTAGTGG
NIN-CDs-R	GCTAGGAGGATGGACTGCTGCTGCT
NINGFP-35Ster-BP-F	**GGGGACAGCTTTCTTGTACAAAGTGGAA**ATGGAATATGGTGGTGG-GTTAGTG
NINGFP-35Ster-BP-R	**GGGGACAACTTTGTATAATAAAGTTGC**TCACTGGATTTTGGTTTTAG-GAATTA
GFPNIN-35Ster-BP-F	**GGGGACAGCTTTCTTGTACAAAGTGGAA**ATGGTGAGCAAGGGC-GAGGA
GFPNIN-35Ster-BP-R	**GGGGACAACTTTGTATAATAAAGTTGC**TCACTGGATTTTGGTTTTAG-GAATTA

Sequences designated in boldface are added to primers for TOPO cloning or BP recombination

4

CHAPTER 5

Changes in NIN related to its recruitment in nodulation

Jieyu Liu[1], Siqi Yan[1], Michaela Škoríková[3], Defeng Shen[1], Olga Kulikova[1], and Ton Bisseling[1,2,*]

[1] Laboratory of Molecular Biology, Department of Plant Sciences, Wageningen University & Research, 6708 PB Wageningen, The Netherlands

[2] Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing University of Agriculture, 102206, Beijing, China; ton.bisseling@wur.nl

[3] Department of Biochemistry, Palacký University Olomouc, Czech Republic

* Correspondence: ton.bisseling@wur.nl

Abstract

The key transcription factor NIN is essential for nodule formation in all studied species that belong to a single phylogenetic group, referred to as the nitrogen fixing clade (NFC). NIN arose by duplication from the NIN-like proteins (NLPs) and this occurred before nodulation evolved. Therefore, several plant species outside the NFC have NIN orthologous. A Comparison of *NIN* with *NIN* orthologous outside the NFC, led to the conclusion that essential adaption in its promoter occurred when NIN was recruited in nodulation. However, it is unclear whether changes at the protein level were also required for *NIN* recruitment in the nodulation process. Here, we showed that NIN is neither functionally equivalent with its orthologues outside the NFC nor with its paralogues. The main reason for this could be that most NLPs are nuclear localized only when high nitrate is sensed, whereas NIN is constitutively nuclear localized. Further, we identified a few amino acid differences between NIN and its orthologues outside the NFC. These changes are important for the function of NIN in nodule formation. They probably were introduced in NIN when it was recruited in the nodulation process and might contribute to its constitutive nuclear localization.

Introduction

Root nodule symbiosis exclusively occurs in plant species belonging to the nitrogen fixation clade (NFC). This NFC contains four orders: Fabales, Rosales, Cucurbitales and Fagales (Soltis et al., 1995). Legumes (order Fabales) and species of the non-legume genus Parasponia (order Rosales) form nodules with Gram-negative bacteria, collectively named rhizobium (Peter et al., 1996). The other nodulating plant species which form nodules with gram positive *Frankia* bacteria, are named actinorhizal plants (Santi et al., 2013). The majority of the legume species can form nodules, whereas nodulation is more rare in the three orders to which actinorhizal plants belong. Recently, both phylogenetic and histological studies indicated that a common ancestor of the NFC gained the nodulation ability, but later, many species lost it (Van Velzen et al., 2018; Griesmann et al., 2018; Shen et al., 2020).

The transcription factor NIN (NODULE INCEPTION) is indispensable for nodule development. This has been shown in different species of the NFC, including several legumes (Schauser et al., 1999; Marsh et al., 2007; Borisov et al., 2003; Wang et al., 2019), Parasponia (Bu et al., 2020) and the actinorhizal species

Casuarina glauca (Casuarina) (Fagalas) (Clavijo et al., 2015). Interestingly, phylogenomic analysis showed that all species capable of forming nodules contain *NIN*, while species that lost the nodulation ability in general lost *NIN* or became a pseudogene (Griesmann et al., 2018). These data strongly suggest that when the common ancestor gained the nodulation ability, it recruited NIN for this process. Some species outside the NFC have NIN orthologue(s) (Liu and Bisseling, 2020). This indicates that NIN had a non-symbiotic function before symbiotic nodulation has been evolved. In contemporary species, *NIN* is specifically expressed during nodulation and knock out mutants do not have a non-symbiotic phenotype (Schauser et al., 1999; Marsh et al., 2007; Borisov et al., 2003; Wang et al., 2019; Bu et al., 2020; Clavijo et al., 2015). So, after recruitment into the nodulation process, *NIN* most likely lost its function in non-symbiotic organs. The function of NIN orthologues in species outside the NFC is not yet known. Therefore, it is unclear whether the nodule specific expression is the only adaptation or changes at the protein level were also required for NIN recruitment in the nodulation process. Here, we focus on this question.

NIN was first identified in *Lotus japonicus* (Lotus) and it is the founding member of the *NIN-like protein* (NLP) gene family (Schauser et al., 2005, 1999). Phylogenetic analyses divided NIN and NLPs into three orthogroups (Mu and Luo, 2019a; Schauser et al., 2005; Liu and Bisseling, 2020). Group 1 contains the NIN subgroup and the NLPs that are orthologous to NIN. This group most likely was generated by a duplication which occurred after the separation of monocots and dicots, but before the NFC evolved. Therefore, species outside the NFC can also contain NIN orthologue(s), such as Arabidopsis NLP1/2/3 (Liu and Bisseling, 2020). NIN and NLPs share two conserved domains, the RWP-RK DNA binding domain and the PB1 protein-protein interaction domain (Schauser et al., 2005). Both are located close to the C-terminus of the protein. Notably, the N-terminal region is not conserved between legume NINs and NLPs (including NIN orthologues) (Chardin et al., 2014; Suzuki et al., 2013b). However, these analyses did not include Parasponia and actinorhizal plants. Therefore, it is not clear when the change occurred and whether it was essential for the recruitment of NIN in the nodulation process.

Studies on the model legumes *Medicago truncatula* (Medicago) and Lotus showed that NIN controls all the main nodule developmental steps. This includes infection thread formation, nodule organogenesis, autoregulation of nodule number, release of rhizobia from infection threads (J. Liu unpublished), development of symbiosomes and suppression of both defence and premature

senescence (Marsh et al., 2007; Schauser et al., 1999; Soyano et al., 2014; Liu et al., 2021). The function of *NLPs* that are orthologues to *NIN,* have not been studied. In contrast, some NLPs that are not orthologues to NIN have been intensively studied. These studies showed that they play a central role in regulating nitrate-induced responses during plant development (reviewed in Mu and Luo, 2019b). Interestingly, although NIN and NLPs regulate different processes, they both can bind to the nitrate-responsive *cis*-element (NRE) and have shared some target genes (Liu and Bisseling, 2020; Nishida et al., 2018). A recent study comparing the direct targets and binding sites of Lotus LjNLP4 and LjNIN revealed that they bind both identical and distinct *cis*-elements (Nishida et al., 2021). These indicating that changes in their DNA binding properties evolved. However, when these changes evolved is unclear.

One major difference between NIN and most of the studied non-orthologues NLPs is their subcellular localization. A recent study on Medicago showed that NIN is nuclear localized in nodules in all cells where it is expressed, suggesting that NIN cannot be retained in the cytoplasm (Liu et al., 2021). In contrast, most NLPs are localized in the cytoplasm under low nitrate condition and are translocated to the nucleus when high nitrate is sensed (Marchive et al., 2013; Guan et al., 2017; Lin et al., 2018a; Cao et al., 2017; Nishida et al., 2018). It has been reported that AtNLP7 translocation is mediated by phosphorylation of serine at position 205 (S205) (Liu et al., 2017). For convenience, we will refer to this serine in NLPs of different species as S205. This S205 is conserved in NLPs, including the NIN orthologues, but it is not present in Lotus NIN (Liu et al., 2017). It is not known when this serine was lost, and whether it was an essential change for NIN recruitment in the nodulation process.

Here, we show that NIN and NLPs, including NIN orthologues, are not functionally equivalent and loss of S205 is not a key change for NIN to function in nodulation. Further, we identified five amino acid differences between NIN and NLPs which probably were introduced in NIN when it was recruited in the nodulation process and one amino acid change occurred later. One or more of these changes are important for nodulation.

Results

NIN and NLPs are not functionally equivalent

Studies on NIN and NLPs have shown that they bind to similar cis-regulatory elements. Therefore, NIN and NLPs might be functionally equivalent when expressed at the same location. To test this, we used the coding region of *MtNLP1*, the closest Medicago homologue of *MtNIN*, driven by the Medicago *NIN* promoter. This construct was introduced into the Medicago *nin-1* knockout mutant by hairy root transformation. At three weeks post inoculation (3wpi), neither nodules, nor infection threads were observed (Figure 1). This showed that MtNIN and MtNLP1 are not functionally equivalent.

As NIN orthologues exist in non-nodulating plants such as Arabidopsis, we further studied whether the NIN orthologue AtNLP1, which is phylogenetically closer to MtNIN, is functionally equivalent to MtNIN. We introduced *AtNLP1* driven by the *MtNIN* promoter into Medicago *nin-1*. However, neither nodules nor infection threads were formed at 3wpi (Figure 1). So, a functional divergence also had occurred between MtNIN and one of the Arabidopsis NIN orthologues.
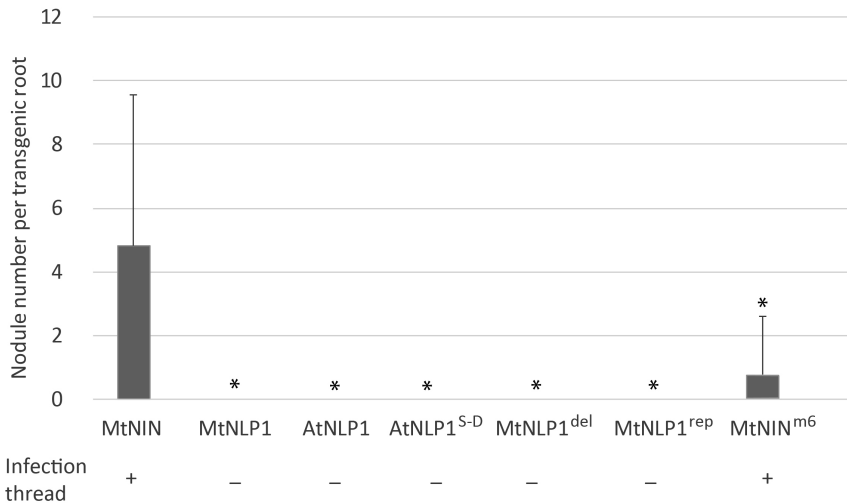
Figure 1. Number of nodule formed on *nin-1* knockout mutant roots transformed with *NIN* or *NLPs* or their modified version (as indicated).

Nodule number per root (n>12) was determined at 3 wpi. Data are means ± SD. * indicate significant mean difference compared with MtNIN (p<0.05). The deleted and replaced region in MtNLP1 and MtNIN to create MtNLP1[del] and MtNIN[rep] , are indicated by black rectangles in **Fig. 3**.

The disability of AtNLP1/MtNLP1 to complement Medicago *nin-1* might be caused by cytoplasmic retention

Previous studies on NLPs have shown that most of the NLPs are located in the cytoplasm under low nitrate condition, and they are translocated to the nucleus when sensing high levels of nitrate (Marchive et al., 2013; Guan et al., 2017; Lin et al., 2018a; Cao et al., 2017; Nishida et al., 2018). In contrast, in Medicago NIN is located in nuclei (Liu et al., 2021). This could mean that expression of the AtNLP1/MtNLP1 was induced, but the encoded proteins were retained in the cytoplasm. This could explain why they cannot complement the *nin-1* mutant. Previously, it has been shown that MtNLP1 indeed can shuttle between cytoplasm and nucleus (Lin et al., 2018a), while AtNLP1 subcellular localization has not been studied.

To study the subcellular localization of AtNLP1, we introduced the coding region of *AtNLP1* fused to that of *GFP* (*Pro35S::GFP-AtNLP1*) into tobacco (*Nicotiana benthamiana*) leaves. The *Pro35S::GFP-AtNLP1* transgenic cells showed a lower GFP signal in the nucleus than in the cytoplasm (Figure 2A). In comparison, cells transformed with *MtNIN-GFP* (*ProUbi::MtNIN-GFP*) showed a higher GFP signal in the nucleus than in the cytoplasm (Figure 2B). This indicated that one of the major changes required for recruitment of NIN in the nodulation process is the loss of the ability to be retained in the cytoplasm and to be (constitutively) nuclear localized.

Studies on AtNLP7 gave some insight in the mechanism controlling translocation from cytoplasm to nucleus. Upon nitrate sensing, serine 205 (S205) of AtNLP7 becomes phosphorylated, after which this protein is translocated from the cytoplasm to nucleus (Liu et al., 2017). This serine is conserved in NLPs, but does not occur in Lotus NIN (Liu et al., 2017). Previous studies have shown that in comparison with NLPs, Medicago and Lotus NIN have a deletion in their N-terminal region (Liu et al., 2017; Chardin et al., 2014; Suzuki et al., 2013b). Interestingly, S205 is located in this deleted region (Figure 3). This could be the reason why MtNIN is constitutively targeted to the nucleus. So, phosphorylation of S205 of MtNLP1 and AtNLP1 probably essential for their translocation to the nucleus. This might not have occurred at the start of the nodulation process as nodulation requires a low nitrate level. This could be the reason why MtNLP1 and AtNLP1 failed to complemented *nin-1*.
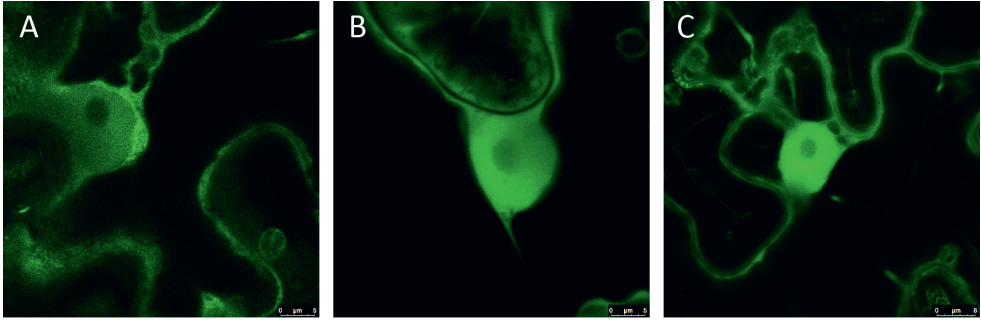
Figure 2. Confocal images of *Pro35S::GFP-AtNLP1* (A)*, ProUbi::MtNIN-GFP* (B) *and ProUbi::NIN^{m6}-GFP* (C) expressed in *Nicotiana benthamiana* leaves.

To test this, we made a point mutation in *AtNLP1* to modify the serine (S) codon into an aspartate (D) codon, to mimic the phosphorylated serine. The modified AtNLP1[S-D] was introduced into *nin-1* by hairy root transformation. However, neither nodules nor infection threads were formed at 3wpi (Figure 1). Further, we deleted the N terminal region, which is specific for NLPs, in MtNLP1 (MtNLP1[del]) or replaced that region by the corresponding N terminal region of MtNIN (MtNLP1[rep]). The regions that were deleted or replaced are indicated in Figure 3. However, none of these modified MtNLP1s were able to complement *nin-1* (Figure 1). This showed that the loss of S205 and surrounding region is not sufficient to create a NIN that can drive the nodulation process.

Did loss of S205 occurred when nodulation evolved?

To understand the importance of the loss of S205 and its surrounding region for recruitment of NIN in nodulation, we compared NIN protein sequences from species within NFC with NIN orthologues from species outside the NFC and some other NLPs (Figure 3). Consistent with previous studies, the S205 and its surrounding region is conserved in all NLPs, including the NIN orthologues (Figure 3). In contrast, legume NINs lost the S205 and its surrounding region. In actinorhizal species *Alnus glutinosa* (Alnus) and *Datisca glomerata* (Datisca) S205 was mutated into a cysteine and leucine, respectively, but they still contain the surrounding region. Surprisingly, Parasponia and actinorhizal species Casuarina, *Dryas drummondii* (Dryas), and *Discaria trinervis* (Discaria) did not lose the S205 and surrounding region. Parasponia NIN (PanNIN) and Casuarina NIN (CgNIN) have been shown to be essential for nodulation. This strongly suggests that the loss of S205 and its surrounding region did not occur when NIN was recruited in the nodulation process.

Figure 3. Sequence alignment of the N-terminal region of NIN/NLP proteins.

The alignment comprises 33 NINs/NLPs from *Oryza sativa* (Os), *Solanum lycopersicum* (Solyc), *Arabidopsis thaliana* (At), *Manihot esculenta* (Manes), *Alnus glutinosa* (Alngl), *Casuarina glauca* (Casgl/Cg), *Datisca glomerata* (Datgl), *Dryas drummondii* (Drydr), *Discaria trinervis* (Distr), *Parasponia andersonii* (Pan), *Chamaecrista fasciculata* (Chafa), *Glycine max* (Glyma), *Medicago truncatula* (Mt) and *Lotus japonicus* (Lj). The S205 (red shaded S) is conserved in NLPs, including both NIN paralogues (purple shade) and NIN orthologues (yellow shade). NIN of some actinorhizal species and Parasponia (green shade) have maintained this S205 and its surrounding region. Independent loss of the S205 of NIN (blue coloured) occurred in actinorhizal species Alnus (*Alnus glutinosa*/Alngl) and Datisca (*Datisca glomerata*/Datgl), but the region surrounding it was maintained. Legumes (pink shade) lost the S205 and corresponding region. The deleted and replaced regions in MtNLP1 and MtNIN to create MtNLP1^del and MtNLP1^rep (see Fig. 1) are indicated by the black rectangles.

Amino acid changes in NIN correlate with the recruitment of NIN in nodulation

To identify amino acid changes that potentially occurred when NIN was recruited in the nodulation process, we aligned the NIN and NLP protein sequences and manually searched for the conserved amino acid changes that most likely occurred when nodulation evolved in the common ancestor of the NFC (Figure 4). We found five positions (position 1, 3-6) that meet this criterion (Figure 4). There is one change present in the Fabales (legumes) and Fagales (Alnus and Casuarina) NINs, but it occurred later as Cucurbitales (Datisca) and Rosales (Dryas, Discaria and Parasponia) NINs maintained the same amino acids as the NIN orthologues outside the NFC (Figure 4, position 2). However, these changes might explain the functional inequivalence of AtNLP1/MtNLP1 and MtNIN. Therefore, we included it in our functional analysis (Figure 4, position 2).

To test whether the identified six amino acid changes are essential for the function of NIN, we mutated these amino acids in MtNIN to the corresponding amino acids in AtNLP1. This mutated *MtNIN^{m6}*, driven by the *MtNIN* promoter, was introduced in roots of Medicago *nin-1*. In comparison to the roots transformed with MtNIN, nodule number was markedly reduced in the roots transformed with *MtNIN^{m6}* at 3 wpi, demonstrating that MtNIN^{m6} can only partially complement *nin-1* (Figure 1). It implies that one or more amino acid changes within these six positions are important for nodule formation.

Mutations in the six identified amino acid changes do not affect the nuclear localization of MtNIN

Among the identified six amino acid changes, the one at the fourth position is particularly of interest for the potential in affecting nuclear localization of NIN. At this position two non-conserved amino acids were changed to RR/KR. The RR/KR is located in the vicinity of the previously identified monopartite (SV20-type) nuclear localization signal (NLS) (Schauser et al., 1999). This change resulted in the formation of a classical bipartite NLS which has been defined as $(K/R)(K/R)X_{10-12}(K/R)_{3/5}$ (Robbins et al., 1991; Dingwall C and Laskey RA, 1991; Kosugi et al., 2009). In species such as Medicago and Lotus, the linker region (*X*) consists ten to twelve amino acids. Thus, the amino acid change at position 4 could lead to a stronger binding by the importin α, which mediates the nuclear import of proteins together with importin β (Weis, 2003; Pemberton and Paschal, 2005) Therefore, this amino acid change might promote constitutive nuclear localization of NIN.
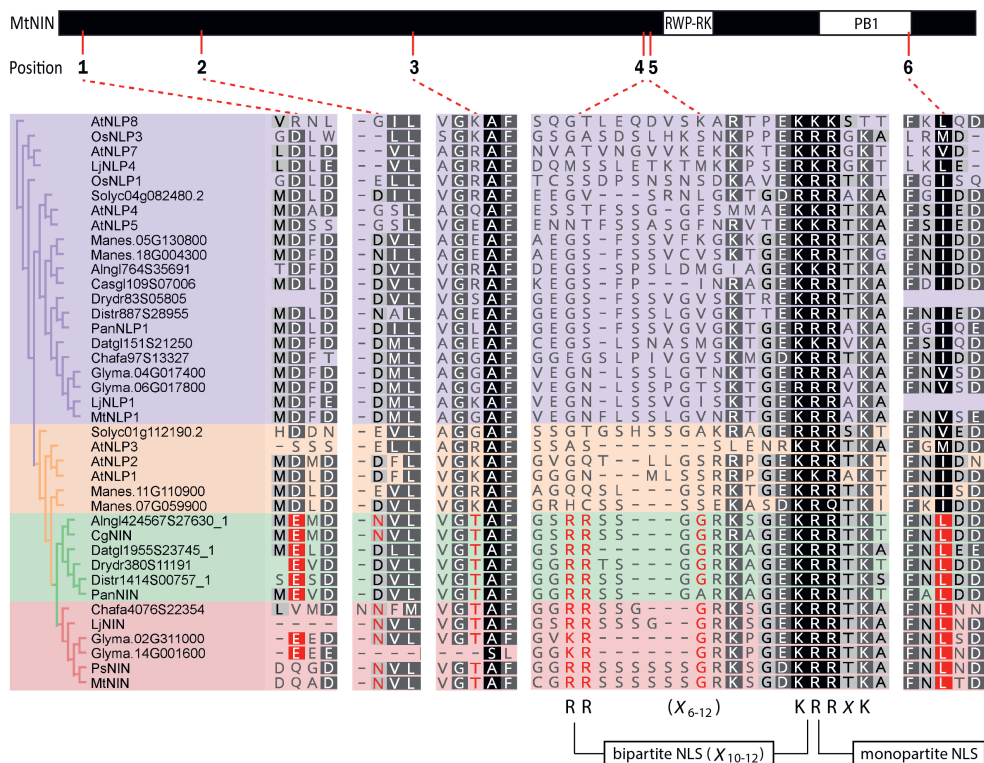
Figure 4. **Illustration of the six amino acid changes.**

Position 1 and 3-6 are putatively related to the recruitment of NIN into the nodulation process. Change at position 2 occurred later. Red colored or red shaded amino acid are specific for NIN (in NFC). Purple shade background marked NIN paralogues, yellow shade background marked NIN orthologues in species outside the NFC, green shade background marked actinorhizal plants and Parasponia NIN, pink shade background marked legume NIN. *Oryza sativa* (Os), *Solanum lycopersicum* (Solyc), *Arabidopsis thaliana* (At), *Manihot esculenta* (Manes), *Alnus glutinosa* (Alngl), *Casuarina glauca* (Casgl/Cg), *Datisca glomerata* (Datgl), *Dryas drummondii* (Drydr), *Discaria trinervis* (Distr), *Parasponia andersonii* (Pan), *Chamaecrista fasciculata* (Chafa), *Glycine max* (Glyma), *Medicago truncatula* (Mt) and *Lotus japonicus* (Lj).

To check whether the identified six amino acid changes affect the subcellular localization of NIN, we introduced *MtNIN^{m6}-GFP* into tobacco leaves. This showed that MtNIN^{m6}-GFP is localized in the nucleus at a higher level than in the cytoplasm, similar to MtNIN-GFP (Figure 2C). We also introduced *MtNIN^{m6}-GFP* or *MtNIN-GFP,* driven by the Medicago *NIN* promoter, into *nin* weak allele *daphne-like* by hairy root transformation. In the nodules formed on the transgenic roots, we observed a strong GFP signal in the nucleus, whereas
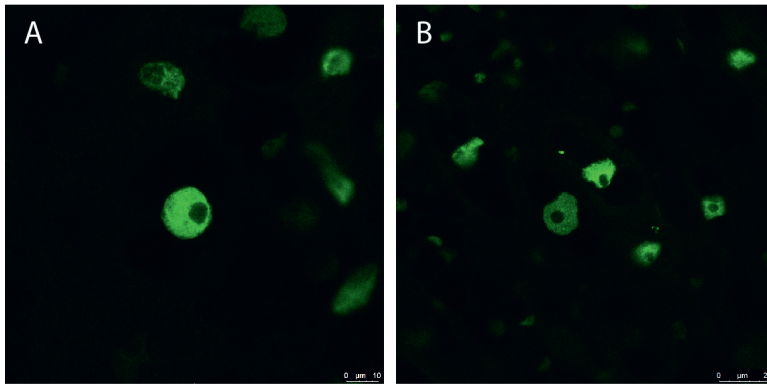
Figure 5. Subcellular localizations of MtNIN-GFP (A) and MtNIN^m6-GFP (B) in nodules formed on *daphne-like* roots transformed with corresponding constructs.

it was hardly detectable in the cytoplasm (Figure 5B). This is similar to the subcellular localization of MtNIN-GFP in nodules (Figure 5A). These results showed that mutations in these six positions do not significantly affect the subcellular localization of NIN.

Discussion

Here we showed that Medicago NIN is not functionally equivalent with its orthologue AtNLP1 and paralogue MtNLP1. To determine the potential cause, amino acid changes in NIN that might correlate with recruitment of NIN were identified. One or more of these changes are essential for the function of NIN in nodulation. However, mutation in the identified amino acid do not significantly affect the nuclear localization of Medicago NIN.

Most of the studied NLPs including MtNLP1 are located in the cytoplasm under low nitrate condition and they shuttle to the nucleus only when high nitrate level is sensed (Marchive et al., 2013; Guan et al., 2017; Lin et al., 2018a; Cao et al., 2017; Nishida et al., 2018). Here we showed that AtNLP1 is also localized in the cytoplasm in tobacco leaf epidermal cells. In contrast, MtNIN is nuclear localized in both Medicago root nodule cells and tobacco leaf epidermal cells. Therefore, one major adaption of NIN upon recruitment in nodulation is to become constitutively nuclear localized.

A study in rice gave an indication how NLPs are maintained in the cytoplasm (Hu et al., 2019). The nitrate-induced translocation of OsNLP3 (orthologue of

AtNLP7) from cytoplasm to nuclei is hindered by co-expression of OsSPX4 in rice protoplasts. Moreover, OsSPX4 interacts with OsNLP3 and nitrate promotes the ubiquitination and subsequent degradation of OsSPX4. This suggests that nitrate promotes the release of OsNLP3 from OsSPX4.

Based on these studies it can be hypothesized that the youngest common ancestor of NIN was most likely a transcription factor that was maintained in the cytoplasm by interaction with another protein. When an environmental factor, most likely nitrate, was sensed, S205 became phosphorylated and the binding protein could no longer bind and was degraded. Subsequently, the ancestral NIN was translocated to the nucleus.

One of the possible way to create a constitutively nuclear localized NIN is the "loss" of the binding protein that retains it in the cytoplasm. In Medicago and Parasponia, none of the *SPX* family genes showed a strong induction, like *NIN,* in nodules compared to roots (Roux et al., 2014; Van Velzen et al., 2018).

A major adaption in *NIN* promoter when it was recruited in nodulation is the gain of a CYCLOPS binding site (Liu and Bisseling, 2020). This enables *NIN* induction upon perception of lipo-chito-oligosaccharides (Nod factors) secreted by the microsymbiont (Singh et al., 2014). However, currently no known SPX genes that are also induced by Nod factor signaling. This might have caused a constitutive nuclear localization of NIN.

We showed that most likely at the start of nodulation NIN obtained five amino acid changes of which one could have created an extra NLS (position 4). However, when we mutated the amino acid change by which a putative NLS was formed, MtNIN maintained its constitutive nuclear localization. We postulate that this might be due to loss of sequences (in legume NINs) that are required for cytoplasmic retention. Actinorhizal plants and Parasponia still have S205 and its surrounding region showing that its presence do not block the function of NIN in nodulation. It would be interesting to see whether these NINs still can be retained in the cytoplasm when the putative NLS (position 4) is mutated. It would also be interesting to study whether AtNLP1 provided with the extra NLS (position 4) becomes constitutively located in the nucleus.

Although most studied NLPs are translocated from cytoplasm to nucleus when high nitrate is sensed, there are some exceptions. For example, AtNLP8, OsNLP1 and LjNLP1 (Alfatih et al., 2020; Yan et al., 2016; Nishida et al., 2021). These three NLPs have been shown to be constitutively nuclear localized,

but still regulate gene transcription in response to nitrate level. This indicates that activation of these NLPs by nitrate signalling occurs, although they are constitutively nuclear localized. There are no indications that this is also the case for the NLPs that are shuttling between cytoplasm and nucleus.

Transcriptome data and binding studies both showed that NIN and NLPs have shared downstream targets such as *CLE-RS2, NIGT1, LBD38* and *NIR1* in Lotus (Nishida et al., 2018, 2021). However, they also regulate specific targets. For example, Lj*NF-YB* and Lj*EPR3* are specifically regulated by LjNIN, but not by LjNLP4 (Nishida et al., 2021). These indicate that the binding specificity of NIN was also evolved. However, it is not known whether it is essential for NIN function in nodulation and when the change occurred.

Although we could not prove that the five amino acid changes that were introduced upon recruitment, they most likely are important for the function of NIN in nodulation. Mutation back of these five changes and one additional change that occurred later in nodule evolution markedly reduced the nodule number formed on transgenic roots. It remains to be studied which amino acid change is important. At position 1, 2 and 6, the changes resulted in amino acids with similar characteristic. In contrast, the changes at position 3, 4 and 5 resulted in amino acids with very different biochemical properties. Therefore, the changes at position 3, 4 and 5 are the promising candidates to be important for the recruitment of NIN in nodulation.

Material and Methods

Plant and bacterial material

Medicago *nin* knockout mutant *nin-1* (Marsh et al., 2007) and weak allele *daphne-like* (Liu et al., 2019b) were used in this study. Seed sterilization, germination and Agrobacterium msu440 mediated hairy root transformation was performed according to Limpens et al. (2004). Medicago plants were grown in perlite saturated with Färhaeus (Fä) medium (Catoira et al., 2000) containing low nitrate (0.25 mM Ca(NO3)$_2$ at 21°C and 16h light/8h dark regime. After one week of growth, plants were inoculated with *Sinorhizobium meliloti* constitutively expressing green fluorescent protein (GFP) (OD600 = 0.1, 2 ml per plant). About five weeks old *Nicotiana benthamiana* plants were used for *Agrobacterium tumefaciens*-mediated transient expression.

Constructs

NIN promoters ($ProNIN_{3C-5kb}$ and $ProNIN_{CE-5kb}$) and $ProNIN_{3C-5kb}$:NIN construct were described in *Liu et al.* (2019). $ProNIN_{3C-5kb}$:NIN-GFP construct has been described in *Liu et al.* (2021). The other constructs are either build by Golden Gate (Engler et al., 2014) or MultiSite Gateway (Thermo Fisher Scientific, Waltham, MA, USA) cloning system. The cloning system, constructs information and the primers used in cloning NIN/NLP or their modified versions are listed in Table S1. In both cloning systems, constructs were recombined into the modified binary vectors. These vectors contain kanamycin resistant gene and DsRed/mCherry fluorescent reporter gene driving by constitutive promoter for selection and identification of transgenic roots.

Agroinfiltration of *Nicotiana benthamiana*

Transient Agrobacterium-mediated transformation of *Nicotiana benthamiana* leaves was performed by co-infiltration of *Agrobacterium tumefaciens* strain C58 containing GFP-tagged constructs and the silencing suppressor p19. Transgenic bacteria were grown in 10ml LB medium with appropriate antibiotics at 28°C overnight. The bacteria were collected by centrifugation and resuspended in MMAi medium (20 g/L sucrose, 5 g/L MS basal salts, 2 g/L MES, 200 µM acetosyringone, pH 5.6). The OD_{600} was measured on spectrophotometer and the suspensions were diluted to final OD_{600}=1.0. The bacteria were incubated for 1 hour at room temperature. The bacterial suspensions with GFP-tagged constructs and P19 were mixed in a 1:1 ratio. The mixed suspensions were injected into the leaves of 5-week-old *Nicotiana benthamiana* plants. Two days post infiltration, the infiltrated area were analyzed by Leica SP8 confocal microscope.

Alignment and phylogenetic analysis of NIN/NLP sequences

The alignment and phylogenetic analysis of NIN and NLP sequences in this study used Geneious v9.1.8 (https://www.geneious.com) (Kearse et al., 2012). Alignment was created by ClustalW (Thompson et al., 1994) with default parameter settings. Created alignment was used for tree building using FastTree 2.1.5 (Price et al., 2010) with default parameter settings. Corresponding NIN/ NLP accession numbers and protein sequences are listed in the Supplementary data S1.

Acknowledgements

5

CHAPTER 6

Evolution of NIN and NIN-like genes in relation to nodule symbiosis

Jieyu Liu[1] and Ton Bisseling[1,2,*]

[1] Laboratory of Molecular Biology, Department of Plant Sciences, Graduate School Experimental Plant Sciences, Wageningen University & Research, 6708 PB, Wageningen, The Netherlands; jieyu.liu@wur.nl; ton.bisseling@wur.nl

[2] Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing University of Agriculture, 102206, Beijing, China; ton.bisseling@wur.nl

* Correspondence: ton.bisseling@wur.nl; Tel.: +31 317 482036

6

Abstract

Legumes and actinorhizal plants are capable to form root nodules symbiosis with rhizobia and Frankia bacteria. All these nodulating species belong to the nitrogen fixation clade. Most likely nodulation evolved once in the last common ancestor of this clade. NIN (NODULE INCEPTION) is a transcription factor that is essential for nodulation in all studied species. Therefore, it seems probable that it was recruited at the start when nodulation evolved. NIN is the founding member of the NIN-like protein (NLP) family. It arose by duplication and this occurred before nodulation evolved. Therefore, several plant species outside the nitrogen fixation clade have NLP(s) that is orthologous to NIN. In this review we discuss how NIN has diverged from the ancestral NLP, what minimal changes would have been essential to become a key transcription controlling nodulation and which adaptations might have evolved later.

1. Introduction

Nitrogen fixing root nodule formation is a special property of some plant species. All these species belong to the nitrogen fixation clade (NFC), which is composed by four orders: Fabales, Rosales, Cucurbitales, and Fagales (Soltis et al., 1995). Nodulation occurs abundantly within the Fabales (legumes). Legumes establish the nodule symbiosis with Gram-negative bacteria belonging to different genera, that collectively are named rhizobium (Peter et al., 1996). In the other three orders nodulation is more rare. In general nodulation in these orders is induced by Gram-positive Frankia bacteria (Santi et al., 2013). The plants that form nodules with Frankia are named actinorhizal plants. The exception is the genus Parasponia (order Rosales) that forms nodules with rhizobia (Lancelle and Torrey, 1984).

Nodule types and mode of infection within the NFC are diverse. Nevertheless, recent studies indicate that a common ancestor of the NFC gained the nodulation ability (Griesmann et al., 2018; Van Velzen et al., 2018; Shen et al., 2020). Later during evolution, many species within the NFC lost the nodulation ability. Interestingly, phylogenomic analyses revealed that the transcription factor NIN (NODULE INCEPTION) that is essential for nodulation became a pseudogene or was lost in most non-nodulating species of the NFC (Griesmann et al., 2018; Van Velzen et al., 2018). In this review, we will discuss the evolution of NIN and its related NIN-like proteins in relation to nodule symbiosis.

Root nodule formation involves intracellular infection, nodule organogenesis; and a negative feedback mechanism that controls the number of nodules. Strikingly, the transcription factor NIN has been shown to play an indispensable role in all these processes (Schauser et al., 1999; Marsh et al., 2007; Batagov et al., 2003; Clavijo et al., 2015; Bu et al., 2020; Soyano et al., 2014). For infection, the most common and advanced way is when bacteria attach to the root hair tip and stimulate the root hair to curl (Oldroyd et al., 2011). In this way the bacteria are entrapped in an enclosed cavity. By deposition of new plant cell wall material and invagination and growth of plasma membrane, a tube like structure, the infection thread, is formed which guides bacteria into the plant (Brewin, 2004; Oldroyd et al., 2011). Alternatively, the bacteria can enter the roots without forming such infection threads. For example through intercellular infection or crack entry (González-Sama et al., 2004; Sprent, 2007).

In several legumes the role of NIN in forming infection threads has been well studied. Loss of function of *NIN* leads to excessive root hair curling and infection thread formation is blocked (Marsh et al., 2007; Batagov et al., 2003; Schauser et al., 1999). Similarly, NIN also most likely plays a role in infection of the actinorhizal plant as it has been shown to be required for Frankia induced root hair deformation in Casuarina (Clavijo et al., 2015). This suggests that the role of NIN in infection thread formation is conserved in both legumes and actinorhizal plants.

In the model legume Medicago, nodule formation starts with mitotic activation of pericycle cells and this is followed by divisions in cortical and endodermal cells (Xiao et al., 2014). The divisions in pericycle and endodermis derived cells stop at an early stage of nodule primordium formation, whereas divisions in the cortical cells persist. The cells derived from the cortex become infected and form the central tissue with infected cells. Cells derived from the cortex also form the peripheral tissue , including nodule vascular bundles (Xiao et al., 2014).

The formation of actinorhizal nodules also starts with mitotic activation of the pericycle and cortical cells (Pawlowski and Demchenko, 2012; Pawlowski and Bisseling, 1996). During nodule primordium formation the pericycle derived cells remain mitotically active. Previously, it has been described that these cells form the (complete) nodule (Pawlowski and Bisseling, 1996; Pawlowski and Demchenko, 2012). However, it has been shown recently that these pericycle derived cells only form the nodule vasculature (Shen et al., 2020). Also during actinorhizal nodule primordium formation the cortex derived cells

6

form the tissue with infected cells. Nodule formation in Parasponia is similar to that of actinorhizal plants (actinorhizal-like) (Shen et al., 2020). So the major difference between legume and actinorhizal(-like) nodule organogenesis is the origin of the vascular bundle. In Medicago it has been shown that a mutation in *NOOT1* causes a homeotic switch in the formation of the nodule vasculature as it becomes actinorhizal-like since it is formed from pericycle cells that remain mitotically active (Shen et al., 2020; Couzigou et al., 2012).

Legume nodules can be divided in indeterminate and determinate nodules. Indeterminate nodules have a persistent meristem at their apex. This is similar to actinorhizal(-like) nodules. Due to their indeterminate growth their tissues are of grade age with the youngest cells near the meristem and the oldest in the part proximal to the root. NIN has been shown to be essential for both determinate (e.g. Lotus) (Schauser et al., 1999) and indeterminate (e.g. Medicago and pea) (Marsh et al., 2007; Batagov et al., 2003) nodules as well as actinorhizal(-like) (e.g. Casuarina and Parasponia) (Clavijo et al., 2015; Bu et al., 2020) nodules. This indicates a common role of NIN in the formation of different types of nodules.

Both infection and nodule organogenesis are initiated upon perception of signal molecules from the bacteria (Denarie and Debelle, 1996). Most rhizobia secrete lipo-chito-oligosaccharides, called Nod factors (NFs), whereas the nature of the signals secreted by Frankia is not known. NFs activate a signalling pathway which is shared with the more ancient arbuscular mycorrhizal symbiosis and NIN is the first induced transcription factor that distinguishes the rhizobium activated responses from that of arbuscular mycorrhizae (Oldroyd, 2013). Although the nature of the Frankia secreted signal is not clear, this common signaling pathway has also been shown to be required for actinorhizal(-like) nodule formation (Svistoonoff et al., 2014).

To balance costs and benefit during nodule symbiosis, plant developed a mechanism called autoregulation of nodulation (AON) by which nodule number is controlled (Kosslak and Bohlool, 1984; Krusell et al., 2002; Nishimura et al., 2002; Okamoto et al., 2013, 2009). It involves a communication between root and shoot. The signals that are send to the shoot are CLE peptides (Okamoto et al., 2013, 2009; Soyano et al., 2014; Mortier et al., 2010). For example, in Lotus this are CLE-RS1 and CLE-RS2 and NIN directly induces their expression by binding to NIN-binding sequence (NBS) in their promoters (Soyano et al., 2014). Upon perception of CLE peptides in the shoot, signals are send to the root resulting in reduced *NIN* expression and so expression of targets will be

reduced (Soyano et al., 2014). So NIN plays a central role in the feedback loop which ensures the formation of an optimal number of nodules.

NIN is a founding member of a small gene family called NIN-like protein (NLP) (Schauser et al., 2005). The studies on paralogues of NIN showed that they play an essential role in regulating nitrate-induced responses(reviewed in (Mu and Luo, 2019a)). Interestingly, studies in Lotus showed that the expression of *CLE* genes is induced not only by rhizobia, but also by application of nitrate (Okamoto et al., 2009; Nishida et al., 2018). One NLP (NRSYM1) can directly activate *CLE-RS2* expression in response to nitrate (Nishida et al., 2018). This suggests that nitrate-induced block of nodulation shares common elements with AON and NIN is partially functionally equivalent with NLPs. NIN orthologues have also been shown to be present in species outside the NFC like in Tomato and Arabidopsis. This indicate NIN might recruited in nodulation based on its original function.

In this review, we will discuss possible evolutionary events underlying the recruitment of *NIN* in nodule symbiosis, based on comparing NIN and NLPs of legumes, Actinorhizal-like plants and non-nodulating species.

2. Phylogenetic analysis of NIN

In Figure 1 a phylogenetic tree is shown, including NIN and NLPs from legume, actinorhizal plant species as well as mono- and dicot species outside the NFC. NIN and NLPs are divided in 3 orthogroups. Group 1 contains NIN and NLPs of dicots and is divided in two subgroups. One contains orthologues of NIN and is named NIN subgroup. The other contains orthologues of MtNLP1 and is named NLP1 subgroup. Most likely these two subgroups are the result of a duplication that occurred before the NFC evolved, but after dicots and monocots separated. As a result NIN orthologues occur in dicot species that are not within the NFC. Tomato and Arabidopsis NIN orthologues have been included in Figure 1. As these species have maintained a NIN orthologue it strongly suggests that they have an essential non symbiotic function. When NIN was recruited in the nodulation process it most likely lost this non-symbiotic function. This is supported by the fact that most species within NFC that have lost nodulation also lost a functional NIN (Griesmann et al., 2018; Van Velzen et al., 2018).

A comparison of *NIN* with their orthologues inside and outside the NFC as well as with *NLPs* in the *NLP1* subgroup will provide insight in the changes in the
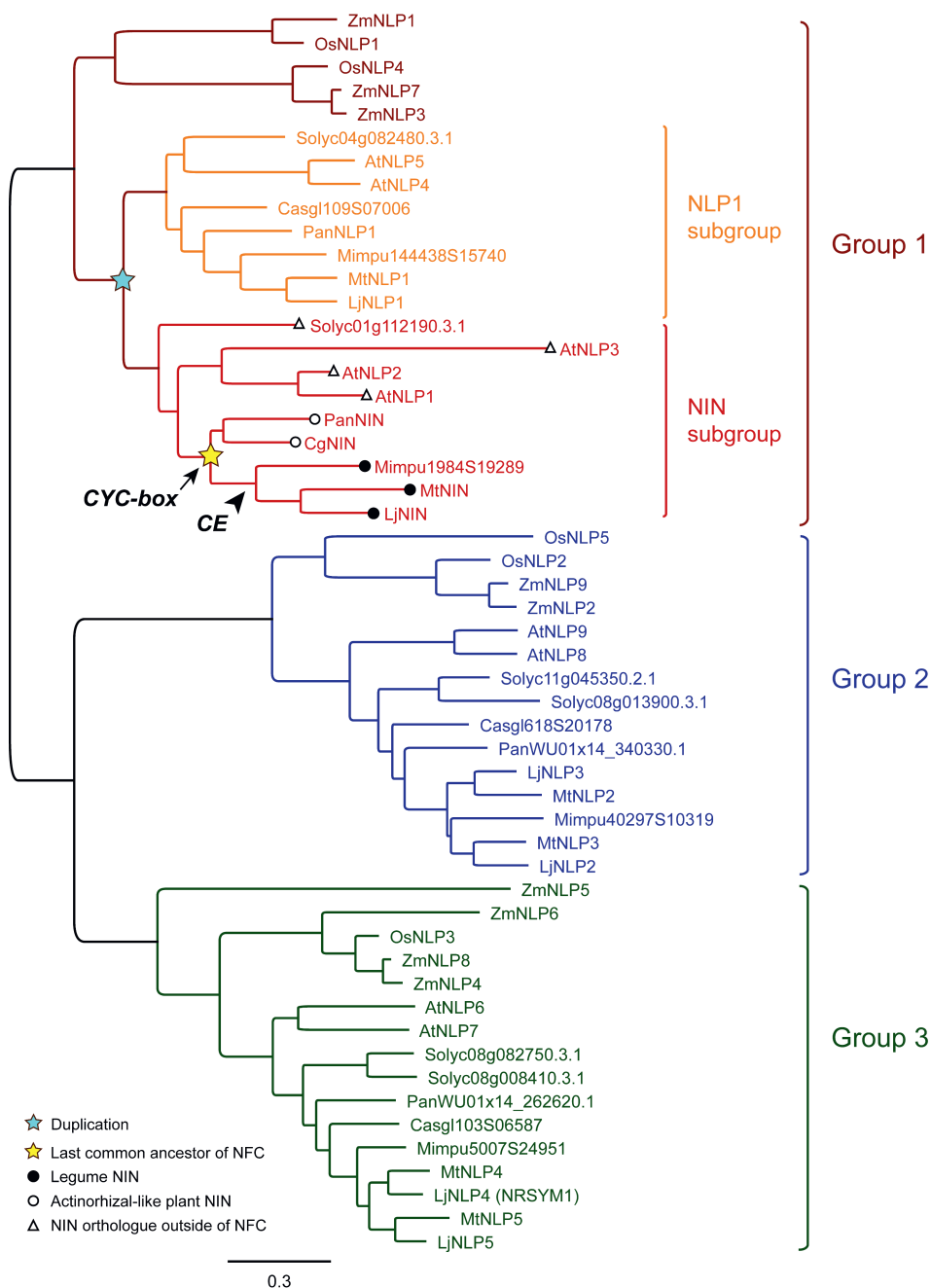
Figure 1. Phylogenetic tree of NIN and NLPs.

The tree comprises 53 NIN/NLPs from Maize (Zm), Rice (Os), Tomato (Solyc), Arabidopsis (At), Casuarina (Casgl/Cg), Parasponia (Pan), Mimosa (Mimpu) Medicago (Mt) and Lotus (Lj). These NIN/NLPs are divided in three orthogroups

ancestral *NIN* that were essential to become a key regulator in nodulation. We will discuss these in the following paragraphs and will focus on the properties of the proteins as well as the regulation of expression.

3. Evolutionary adaptations in the NIN promoter to serve in nodule formation

NIN is nodule specifically expressed in all studied species (Marsh et al., 2007; Batagov et al., 2003; Schauser et al., 1999; Clavijo et al., 2015; Bu et al., 2020; Van Velzen et al., 2018; Demina et al., 2013). In contrast, most *NLPs* are constitutively expressed (Cao et al., 2017; Chardin et al., 2014; Lin et al., 2018a). For example in Rice, Maize, Arabidopsis and Medicago *NLPs* are expressed in almost all organs, although with preferential expression in certain stages and tissues (Cao et al., 2017; Chardin et al., 2014; Lin et al., 2018a). In legumes and actinorhizal (-like) plants, nodulation requires the common signaling pathway (see above) and in some species it has been shown that it induces the expression of *NIN* (Marsh et al., 2007; Batagov et al., 2003; Schauser et al., 1999; Clavijo et al., 2015; Bu et al., 2020; Van Velzen et al., 2018; Demina et al., 2013). Therefore, we hypothesize that when *NIN* was recruited in root nodule formation, it became under control of the common signaling pathway to express in nodules. Further, at a certain moment it also lost its constitutive expression.

The spatiotemporal regulation of *NIN* has been studied in detail in Medicago and Lotus and it is highly complex. We will first summarize this and then discuss whether such regulation also occurs in other species. Upon inoculation, *NIN* is induced in the epidermis where it is required for infection. This was demonstrated by in situ hybridization, promoter GUS reporter constructs as well as root hair transcriptome analyses (Yoro et al., 2014; Liu et al., 2019b; Breakspear et al., 2014; Vernié et al., 2015b). The expression of *NIN* in epidermis could

6

Continued

as indicated. Dicots of Group 1 most likely undergo duplication (blue star) which generate NIN and NLP1 subgroups. NIN subgroup comprises symbiotic NIN from NFC species (yellow star indicate the last common ancestor of NFC), including NIN of actinorhizal-like plant NIN (hollow circle) and legume NIN (filled circle), and putative NIN orthologues in non-nodulating species outside of NFC (hollow triangle). CYCLOPs binding site (*CYC-box*) in *NIN* promoter likely gained by the ancestor of NFC (arrow), whereas cytokinin responsive elements (CE) in the *NIN* promoter specifically evolved in legume branch (arrow head). Corresponding accession numbers and protein sequences are listed in Supplementary data S1.

be cell autonomously regulated upon perception of NFs which induce the common signaling pathway, resulting in the activation of the transcription factor CYCLOPS by phosphorylation (Singh et al., 2014). It was first shown in Lotus that phosphorylated CYCLOPS binds to *NIN* promoter in a sequence-specific manner to regulate *NIN* expression (Singh et al., 2014). In Medicago, the CYCLOPS binding site is located about 3kb upstream of the *NIN* start codon (Liu et al., 2019b). A 5kb long promoter including the CYCLOPS binding site driving *NIN* expression can restore infection thread formation in a *nin* knockout mutant (Liu et al., 2019b). However, it does not complement for nodule organogenesis, showing that additional *cis*-regulatory elements are required for this (see below). Deletion of the CYCLOPS binding site dramatically reduced the infection thread forming efficiency (Liu et al., 2019b). However, it still complements for tight root hair curling and bacterial colonies are formed in these curl. This resembles the Lotus *cyclops* and Medicago *ipd3-2 (cyclops)* mutant phenotype (Horváth et al., 2011a; Yano et al., 2008). This indicates that in addition to the CYCLOPS binding site other *cis*-regulatory element(s) must be present which is sufficient to induce *NIN* expression that leads to tight root hair curling. It was assumed that the CYCLOPS binding site results in a higher expression level of *NIN* and the initiation of infection thread formation requires a higher threshold level than curling (Liu et al., 2019b).

Interestingly, CYCLOPS binding sites are conserved in legume *NIN* promoters (Liu et al., 2019b) and also occurs in the Parasponia *NIN* promoter. However, it does not occur in the *MtNLP1* promoter. This suggests that the gain of the CYCLOPS binding site occurred after the duplication that resulted in the NIN and NLP1 subgroups (Figure 1). This CYCLOPS binding site also does not occur in *NIN* orthologues of Arabidopsis (*NLP1/2/3*). Taken together, it can be hypothesized that the gain of the CYCLOPS binding site had occurred when *NIN* was recruited into the nodule formation process and was present in *NIN* of the ancestor of the NFC. This gain of the CYCLOPS binding site seems an essential step for its expression during nodule formation.

At an early stage of Medicago nodule development, when rhizobia have only colonized the epidermis, *NIN* expression and cell divisions are induced in the pericycle (Liu et al., 2019b). Subsequently, both extend to inner cortex and endodermis (Liu et al., 2019b). Because NFs are immobile molecules (Goedhart et al., 2000a), most likely a mobile signal was generated upon perception of NFs in the epidermis. This mobile signal was then translocated to the pericycle to activate *NIN* expression and cell division there (Liu et al., 2019b).

The first insight that other, more remote, promoter elements are involved in the induction of *NIN* in inner layers came from studies with the *nin* weak alleles *daphne* (Lotus) and *daphne-like* (Medicago) in which the infection process is induced, but nodule organogenesis is blocked (Yoro et al., 2014; Liu et al., 2019b). *NIN* expression was shown to be induced in epidermis in *daphne-like*, but not in pericycle (Liu et al., 2019b). Both mutants contain a large insertion in the promoter region of *NIN* due to a chromosome translocation (Yoro et al., 2014; Liu et al., 2019b). In both cases these are located upstream of the CYCLOPS binding site (Liu et al., 2019b). This strongly suggested that the regulatory region required for *NIN* expression in the pericycle and *NIN* controlled cell division is located upstream of the insertion. Recently, a remote *cis*-regulatory region named CE region was identified which is essential for NIN controlled nodule organogenesis and expression of *NIN* in the pericycle (Liu et al., 2019b). This CE region contains many putative cytokinin response regulator (RR) binding site (Liu et al., 2019b). Studies in which cytokinin is applied exogenously did not induce *NIN* expression in either *daphne* or *daphne-like* (Yoro et al., 2014; Liu et al., 2019b), supporting that that this region is important for cytokinin induced *NIN* expression. In addition, both the cytokinin receptor *CRE1* and B-type cytokinin response regulator *RR1* are expressed in the pericycle prior to *NIN* expression and cell division there (Liu et al., 2019b). Taken together, it is very likely that cytokinin signaling via the CE region induces *NIN* expression in the pericycle and this leads to nodule organogenesis. It has been shown in Medicago that *YUCCA* genes, involved in auxin biosynthesis, are induced in an inner cell layer of the root (most likely pericycle), in a NIN dependent manner (Schiessl et al., 2019a). Therefore, it is probable that subsequent cell divisions are induced by local auxin production.

CE region is conserved in legumes and is located far upstream from the *NIN* start codon (Liu et al., 2019b). The distance varies between species. For example, in Medicago it is about 18kb upstream and in Lotus it is about 45kb upstream of the *NIN* start codon (Liu et al., 2019b). So it is very probable that in legumes the regulation of *NIN* expression in inner layers is controlled by cytokinin signaling.

In Actinorhizal (-like) plants, NIN is required for nodule organogenesis and this involves divisions in inner root cell layers (Chabaud et al., 2016; Bu et al., 2020). So is the CE region conserved in actinorhizal *NIN* genes? So far only Parasponia *NIN* (*PanNIN*) has been analyzed and the CE region was not identified. Nodule formation in Parasponia is induced by Nod factors and so like in legumes a mobile signal seems essential to trigger cell division in inner layers. Whether *NIN*

6

is expressed in the Parasponia pericycle cells when cell division is induced has not been studied. However, this seems probable as a putative target of PanNIN, namely *PanNF-YA1* (*NUCLEAR TRANSCRIPTION FACTOR Y SUBUNIT A-1*) is induced there (Bu et al., 2020). *NF-YA1* has been shown to be a direct target of NIN in Lotus (Soyano et al., 2013). In Medicago the expression of *NIN* and *NF-YA1* coincides in the dividing pericycle cells and other nodule primordium cells (Liu et al., 2019b). Further, *PanNF-YA1* expression has been shown to be PanNIN dependent (Bu et al., 2020). As the CE region does not occur in the Parasponia *NIN* promoter, the mobile signal most likely does not regulate *NIN* expression through cytokinin signaling. This conclusion is in line with a recent study in which cytokinin is applied to different plant species (Gauthier-Coles et al., 2019). It was shown that exogenous application of cytokinin induces nodule-like structures on nodulating legume species and for example in Lotus this depends on NIN (Heckmann et al., 2011). However, nodule-like structures were not induced by cytokinin on either non-nodulating legumes or on actinorhizal species (Gauthier-Coles et al., 2019). The non-nodulating legumes most likely have lost the ability to respond to cytokinin due to the loss of *NIN*, whereas in actinorhizal plants this might be due to the absence of a CE region in their *NIN* promoter. Therefore, the gain of cytokinin responsive elements in the *NIN* promoter seems specific for the legumes and might not have already occurred in the ancestor of the NFC.

It cannot be excluded that more changes in the *NIN* promoter evolved during nodule evolution. For example NSP1/2, IPN2 and ERN1 have been shown to play a role in Medicago and Lotus *NIN* expression (Liu et al., 2019c; Kang et al., 2015; Murakami et al., 2007; Hirsch et al., 2009; Xiao et al., 2020). However, it is not clear that this involved specific evolutionary adaptations in the promoter of *NIN* related to its recruitment for nodulation.

Taken together, there seem to be (at least) three main evolutionary changes for regulation of *NIN* expression during nodule evolution. First, it gained the CYCLOPS binding site most likely in the ancestor of the NFC. Second, at an early stage of nodulation the constitutive expression is lost by which in nodulating plants *NIN* became nodule specific. Third it gained cytokinin regulated *NIN* expression during nodule organogenesis which most likely occurred the legume branches.

4. NIN controlled epidermis-pericycle communication; a conserved module?

It seems probable that the activation of *NIN* in inner root layers of legumes and Parasponia requires a mobile signal. What is the probability that this is a general property of nodulating plants? It has been proposed that the last common ancestor of the NFC formed a nodule symbiosis with Frankia (van Velzen et al., 2019). Although it has not been demonstrated that Frankia can produce Nod factors, basal Frankia strains do have genes that are homologous to rhizobial nod genes which are involved in Nod factor biosynthesis. Therefore, it has been hypothesized that Frankia induced nodulation in the last common ancestor of NFC by secreting Nod factors (van Velzen et al., 2019). If so, as has been argued for legumes and Parasponia, a mobile signal might also be required to induce *NIN* in inner root layers in the common ancestor as well as in current actinorhizal plants that interact with Frankia that secrete Nod factors. It has been shown in legumes that the production of the mobile signal depends on *NIN* expression in the epidermis (Liu et al., 2019b). So *NIN*, expressed in the epidermis, might also induce the formation of a mobile signal in actinorhizal plants that interact with Nod factor producing Frankia. Subsequently, that mobile signal triggers *NIN* expression and cell division in the pericycle. Thus we hypothesize that in legumes and these actinorhizal plants a conserved module, involved in communication between epidermis and pericycle, is active and this was already present in the common ancestor of the NFC: Nod factor signaling induces *NIN* expression in the epidermis– NIN activates the production of a mobile signal – this mobile signal induces *NIN* expression and cell division in the pericycle. Is such a module maintained in all actinorhizal plants? Some Actinorhizal plants interact with Frankia strains that do not produce NFs, such as Ccl3 which secretes a hydrophilic symbiotic signal which is resistant to chitinase (Chabaud et al., 2016). So it is chemically distinct from the amphiphilic and chitin-based NFs. This symbiotic signal is able to induce *NIN* expression in the Casuarina root epidermis and so NIN might still be able to induce the production of a mobile signal which subsequently induces cell division in the pericycle. Alternatively, the mobile signal might not be required in these Actinorhizal plants if the symbiotic signal itself is mobile. However, it has been proposed that an immobile NF may facilitate detection and localization of the bacteria by which redirection of root hair growth is induced and a curl is formed that entraps the bacteria (Goedhart et al., 2000a). Ccl3 infect Casuarina roots through infection threads (Chabaud et al., 2016), therefore, it seems likely that this symbiotic signal is immobile as NFs.

In case cytokinin is the mobile signal in legumes, then actinorhizal plants will form most likely a different mobile signal. Alternatively, the nature of the mobile signal is conserved and it induces cytokinin signaling in legumes and a different pathway in actinorhizal(-like) plants. The induction of the formation of a mobile signal might be an ancestral/original function of NIN in the epidermis. As described above, NIN is essential for infection thread formation. However, it seems likely that for more primitive infection modes like crack entry, NIN is not required. Therefore, the ancestral function of NIN in the epidermis is probably the induction of the production of the mobile signal, whereas its role in infection thread formation evolved later.

5. Function of NIN: acquired upon recruitment or adopted from NLP controlled process?

NIN is essential for nodule initiation, including infection thread formation, nodule organogenesis and AON. We will discuss the functions of NIN in these different processes, including some of its target genes. Further, we will discuss to what extent they are adopted from processes in which NLPs function or whether they are innovations after its recruitment.

5.1. NF-Ys

NF-Ys are transcription factors and some of them play a role in NIN induced nodulation processes. NF-Y transcription factor complexes are composed of three subunits: NF-YA, NF-YB and NF-YC (Baudin et al., 2015). *NF-YA1* and *NF-YB1* have been shown to be direct targets of NIN in Lotus (Soyano et al., 2013). Knockout mutations in *NF-YA1* cause aberrant infection thread formation in Medicago (Laporte et al., 2014) and a block of intracellular infection in Parasponia (Bu et al., 2020). In *Phaseolus vulgaris* (common bean) knockdown of *NF-YC1* also causes aberrant infection thread formation (Zanetti et al., 2010). Further, NF-Ys play a role in nodule organogenesis. In Lotus, Medicago and Parasponia *NF-YA1* mutations lead to the formation of small nodules and/or reduced nodule number (Combier et al., 2006; Soyano et al., 2013; Bu et al., 2020). Further, in both Medicago and Parasponia nodule primordia and nodules, *NF-YA1* is expressed at sites where *NIN* has been shown to be expressed (Liu et al., 2019b; Bu et al., 2020). This suggests that the involvement a NIN-NF-Y module in intracellular infection and nodule organogenesis is conserved within the NFC and it evolved at an early moment during nodule evolution. This

raises the question whether NF-Ys are involved in NLP controlled processes (in species outside NFC) and are these processes related to nodule organogenesis and/or intracellular infection?

In Lotus (Soyano et al., 2019) and Parasponia (Bu et al., 2020) a mutation in *NF-YA1* also has a non-symbiotic phenotype, namely reduced lateral root formation. Moreover, in Lotus over expression of *NIN* or *NF-YA1* induces extra cell division in pericycle and this leads to the formation of lateral roots with a malformed tip (Soyano et al., 2013). This suggests that an ancestral function of NF-YA1 could be related to lateral root formation, a process very similar to nodule organogenesis. This hypothesis is further supported by studies on the *NF-YA1* orthologues in Arabidopsis as AtNF-YA2/AtNF-YA10 are both involved in primary and lateral root growth (Sorin et al., 2014). In addition, they are expressed in pericycle cells where lateral root as well as nodule organogenesis are initiated.

Interestingly, transcriptome analysis of Arabidopsis *nlp7* (from a different orthogroup than *NIN*) mutant has altered expression levels of several genes encoding NF-YA subunits, including *AtNF-YA2* and *AtNF-YA10* (Zhao et al., 2018; Alvarez et al., 2020). This suggests that (some) NLPs can regulate the expression of *NF-Y(A1)*. So, NIN-NF-Y module might have been adopted from a NLP-NF-Y module which already occurred before nodulation evolved.

How can NF-Y be involved in both infection and nodule organogenesis? In mammals, NF-Y regulates expression of cell cycle genes (Caretti et al., 2003). During root nodule development, not only nodule organogenesis requires expression of cell cycle genes, but also passage of infection threads through cells requires entry into cell cycle (Breakspear et al., 2014). Therefore, NF-Y might regulate infection as well as organogenesis through activation of cell cycle genes. This is supported by knockdown of NF-YC in common bean which results in reduced expression of cell cycle genes, whereas over expression of *NF-YC1* causes higher expression levels of these genes (Zanetti et al., 2010). Further, in Lotus, ectopic expression of *NIN* or *NF-YA1* and *NF-YB1* results in ectopic expression of a cyclin gene (Soyano et al., 2013).

Accumulation of the phytohormone auxin is correlated with mitotic activity. For example when lateral root formation is initiated auxin has accumulated in pericycle cells (Dubrovsky et al., 2008). In nodulation, auxin signaling is required for both nodule organogenesis and initiation of infection threads (Mathesius et al., 1998; Liu et al., 2019a). Further, it has been shown that NF-YA1 directly

6

regulates the expression of *STY* genes which encode transcription factors that regulate expression of *YUCCA* auxin biosynthesis genes in Arabidopsis (Hossain et al., 2016; Sohlberg et al., 2006; Eklund et al., 2010). So, NF-Y might by regulating auxin biosynthesis genes stimulate entry into the cell cycle which is involved in both infection and organogenesis and this also holds for NIN which regulates *NF-Y* expression.

5.2. LBD16

Recent studies show that NIN directly regulates the expression of *LBD16* (*lob-domain protein 16*), which is known to be essential for lateral root formation (Soyano et al., 2019; Schiessl et al., 2019a). In Medicago an *lbd16* mutant shows similar defects in nodule and lateral root initiation (Schiessl et al., 2019a). Like NF-YA1, LBD16 promotes auxin biosynthesis via transcriptional induction of *STY* and *YUCCA* genes (Schiessl et al., 2019a). In Lotus, it was shown that LBD16 and NF-Y in an additive way regulate nodule organogenesis, as mutation in *LBD16* enhances nodulation phenotypes of *nf-y* subunit mutants (Soyano et al., 2019). In addition, co-expression of *LBD16* and *NF-Y* subunit genes can partially replace *NIN* as it can rescue nodule organogenesis in the weak *nin* allele *daphne* (Soyano et al., 2019). In Parasponia the expression level of *LBD16* is about six-times increased at initial stages of nodule formation (Van Velzen et al., 2018). This level of induction is comparable to that in Lotus and Medicago (Soyano et al., 2019; Schiessl et al., 2019a). In Arabidopsis, *LBD16* has been shown to be a direct target of NLP7 (Alvarez et al., 2020). Therefore, NIN controlled *LBD16* expression most likely is adopted from a non-symbiotic module in which its expression is controlled by an NLP and this module evolved before nodulation evolved.

5.3. NPL1

Cell wall remodeling is required during rhizobial infection (Brewin, 2004). In Medicago root hairs, cell wall modification genes are induced upon rhizobial infection and the expression of many of them depends on NIN (Liu et al., 2019a). Among these genes, only *nodulation pectate lyase 1* (*NPL1*) shows nodule specific expression (Liu et al., 2019a). In both Lotus and Medicago, NPL1 is essential for infection thread formation (Xie et al., 2012; Liu et al., 2019a). Its expression is regulated by NIN and it is highly induced at infection sites (Xie et al., 2012; Liu et al., 2019a). NPL seems the result of a tandem gene duplication and its orthologues occur in Soybean (*Glycine max*), Lupin (*Lupinus albus*)

and *Arachis ipaensis* (Liu et al., 2019a). So the NIN-NPL module is most likely conserved in the papilionoid legume sub-family.

Parasponia has three putative *NPL* orthologs namely *PanPLL8*, *PanPLL9* and *PanPLL10* (Van Velzen et al., 2018). However, the most closely related *PanPLL8* is not nodule specifically expressed and *PanPLL9* and *PanPLL10* are only 4-fold induced at initial stages of colonization and in later developmental stages the relative fold change is even lower (Van Velzen et al., 2018). This is different from legumes in which *NPL1* is highly induced (Xie et al., 2012; Liu et al., 2019a). This suggests that the NIN controlled nodule specific expression of NPL might have been gained within the legume branch.

5.4. RPG

RPG (rhizobium-directed polar growth) is a gene of which the expression is controlled by NIN (Liu et al., 2019a). RPG is a long coiled-coil protein that is nuclear localized (Arrighi et al., 2008). In Medicago, it has been shown to be essential for normal root hair curling and infection thread formation (Arrighi et al., 2008). RPG is nodule specifically expressed in Medicago, Lotus and Parasponia (Arrighi et al., 2008; Van Velzen et al., 2018; Mun et al., 2016), which is consistent with the loss of a functional *RPG* gene in several non-nodulating NFC species (Van Velzen et al., 2018; Griesmann et al., 2018). This strongly supports that it only has a symbiotic function in the NFC. In Lotus, it has been shown by CHIP-seq that *RPG* is a direct target of NIN (Liu et al., 2019a; Soyano et al., 2014). Together, this strongly suggests that a NIN controlled *RPG* expression evolved early in the NFC or it is already controlled by a NLP before nodulation evolved. However, neither NLP7 CHIP-chip nor in *nlp7* transcriptome indicated that the expression of a *RPG* homologue is controlled (Marchive et al., 2013; Zhao et al., 2018). To further support that NIN controlled RPG expression evolved early in NFC, the RPG regulation in other *nlp* mutants, especially of *NIN* orthologues remains to be analyzed.

5.5. CLEs

As described above, AON is essential for the regulation of nodule number to balance the gain and costs of nodulation. In Medicago and Lotus *CLE* genes that are involved in AON are nodule specifically expressed in a NIN dependent manner (Soyano et al., 2014; Liu et al., 2019a; Mortier et al., 2010). In addition, PanCLE9 which is the putative Parasponia orthologue of Medicago CLE12/13

is also expressed at enhanced levels in nodules (Van Velzen et al., 2018). This suggest that NIN controlled AON is conserved in NFC. However, the nodule specific expression of CLE genes is not only controlled by NIN, but also by NLP whereas in response nitrate (Nishida et al., 2018). In Lotus, it has been shown that both NRSYM1 (one of the NLPs belonging to the same orthogroup as AtNLP7) and NIN can bind to the Lotus CLE-RS2 promoter and NRSYM1 even has a higher affinity (Nishida et al., 2018). In addition, Arabidopsis *nlp7* mutants have altered expression of several *CLE* genes including *AtCLE5/6/7* which are putative orthologues of *LjCLE-RS1/2* (Zhao et al., 2018). Therefore, the function of NIN in controlling AON/CLE genes expression is most likely adopted from an ancestral NLP module.

6. Are NIN and NLPs functionally equivalent?

Several of the processes and genes regulated by NIN appear to be adopted from those controlled by NLPs. Therefore, it seems probable that NIN and NLPs are to some extend functionally equivalent. To test this, *MtNLP1*, the closest Medicago *NIN* paralogue, and *AtNLP1*, an Arabidopsis *NIN* orthologue, driven by the Medicago *NIN* promoter were introduced into the Medicago *nin-1* knockout mutant. However, neither nodules nor infection threads were formed. This shows that NIN and these two NLPs are not functionally equivalent. So their protein sequence has diverged and most likely NIN obtained amino acid changes that are essential to function in nodulation.

Most NLPs that have been studied are located in the cytoplasm under nitrate starvation and when high nitrate is sensed they are transported to the nucleus (Guan et al., 2017; Nishida et al., 2018; Cao et al., 2017; Lin et al., 2018b; Marchive et al., 2013). In contrast, in Medicago nodules, NIN is located in the nucleus of all cells where it is expressed and this became independent of nitrate sensing (Liu *et al.* unpublished data) (Liu et al.). So a major difference between NLPs and NIN is the constitutive nuclear localization of NIN and this might be the reason that NLPs cannot complement the *nin* mutant.

In Arabidopsis, it has been shown that upon nitrate sensing Serine 205 (S205) in AtNLP7 becomes phosphorylated after which it is translocation from cytoplasm to nucleus (Liu et al., 2017). In addition, S205 in AtNLP7 is located in the region previously predicted as NLP specific and it is absent in Medicago and Lotus NIN, whereas MtNLP1 and AtNLP1 have it (Liu et al., 2017; Chardin et al., 2014; Suzuki et al., 2013a).

This serine might be the cause that MtNLP1/AtNLP1 fail to complement *nin-1*. As nodulated plants are grown under low nitrate (0.5mM) conditions. However, phosphomimic versions of MtNLP1/AtNLP1, serine modified into aspartate, also did not complement *nin-1*. This indicates that in addition to this serine other changes are introduced in NIN that are required to be functional in nodulation. Identifying such critical changes will provide insight in the evolution of NIN.

7. Concluding remarks

Both functional and phylogenomic analysis underpin the conserved key position of NIN in root nodule symbiosis. Further, the recruitment of NIN in the nodulation process seems the key step in the birth of this symbiosis. Therefore, understanding which minimal changes were required will provide major insight in the evolutionary trajectory leading to nodule formation. NIN is closely related to NLPs with whom it shares some downstream targets, pointing to partial functional equivalence. Two major changes are the regulation of expression by the microbe via the common signalling pathway and the constitutive nuclear localization. Further adaptations will have occurred within the NFC. An example is the regulation of expression by cytokinin in the legume branch. Further studies on NIN orthologues inside and outside of NFC will help to refinement the picture. The puzzle in understanding the evolution of NIN comes from its complicated spaciotemporal expression pattern and its multifunctionality. Until now, almost all studies on NIN are focused on its role during nodule initiation. However, *NIN* is also expressed in the mature nodule (Roux et al., 2014; Batagov et al., 2003) where it most likely plays a different role than during the initial stages. Further studies on NIN in later nodule developmental stages will complete the picture of evolution of NIN.

Supplementary Materials

The following are available online at https://doi.org/10.3390/genes11070777, Supplementary data S1: NIN and NLP sequences used in the phylogenetic analysis.

Author Contributions

J.L. and T.B. wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding

Acknowledgments

6

CHAPTER 7

General Discussion

Jieyu Liu

Laboratory of Molecular Biology, Department of Plant Sciences, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.

Introduction

In this thesis, I focused on NIN, a transcription factor that plays a key role in almost all steps of the formation of nitrogen fixing root nodules. Its recruitment was also a major step in the evolution of the nodule symbiosis. We discovered that its complicated spatio-temporal expression pattern allows it to function in multiple processes of nodule development (Chapter 2, 3, and 4). We identified the *cis-* regulatory region which is required for *NIN* expression in the pericycle to initiate nodule organogenesis (Chapter 2). Next, by studying the weak alleles *nin-13* and *nin-16*, we showed that NIN is also indispensable for late stages of nodule development (Chapter 4). Further, from an evolutionary perspective we discussed how *NIN* diverged from *NIN-like proteins* (*NLPs*), and what were the minimal changes that might be essential for its recruitment in nodulation in both its promoter and protein sequences (Chapter 5 and 6).

In this discussion, I will integrate the findings described in this thesis and the recently reported developments concerning NIN. Based on this, I propose a model that explains how NIN functions in multiple steps of the nodulation process. The recruitment of NIN during the evolution of nodulation has been discussed in detail in Chapter 6.

2. NIN controlled steps in nodulation

2.1 NIN controls two different steps of the infection process in root hair containing cells

There are two main steps in the initiation of intracellular infection in root hairs: root hair curling and infection thread formation. Root hair curling requires redirection of root hair growth, followed by stopping of growth when its tip touches the shank of the root hair. In this way a closed infection chamber is formed. Arrest of root hair growth means that deposition of membrane and cell wall material to the root hair tip stops. In contrast, the formation of an infection thread requires deposition of membrane and wall material at the site where its formation is initiated (Oldroyd et al., 2011). This is preceded by a local weakening of the cell wall in the infection chamber. The nodule specific pectic lyase (NPL) , a direct target of NIN, plays a role in this (Xie et al., 2012; Liu et al., 2019a). The arrest of root hair growth and initiation of infection thread formation seem like two completely different processes. However, they have to be initiated in the

right order in a single root hair cell and both are controlled by NIN. It is puzzling how this can be achieved. The dissection of the *NIN* promoter (Chapter 2) gives a possible explanation for this. We showed that the -2.2kb promoter region of *NIN* is sufficient to allow NIN to stop root hair growth and the formation of an enclosed curl. However, the initiation of infection thread formation requires a -5kb promoter region which includes the CYCLOPS binding site. As CYCLOPS is the transcription factor that is activated by Nod Factor (NF) signaling and is known to activate *NIN* expression (Singh et al., 2014), this promoter region most likely leads to a higher expression level of *NIN*. This suggests that the way NIN controls these two different processes in one cell is by its concentration level. The simplest putative mechanism might be the accumulation of the NIN protein over time. Therefore, the following model can be proposed: when *NIN* expression is induced in the epidermis initially it will accumulate at a low level. This low level is sufficient to activate a mechanism by which root hair growth stops when the tip touches the shank. This might be a mechanism related to mechanical sensing. Subsequently, NIN accumulates over time and reaches a certain threshold by which it is able to induce sufficient expression of genes required for infection thread initiation such as its direct target *NPL*. To reach this threshold not only requires sufficient time but also strong expression induction level of *NIN*, as the -2.2kb promoter is not sufficient to induce infection thread formation in *nin* knock-out mutants even after 4 weeks of inoculation.

2.2 Cytokinin is probably the mobile signal that moves from epidermis to pericycle

An important function of NIN in the epidermis seems to be the activation of a "mechanism" by which mitotic activity in the inner cell layers is induced. This is described in Chapter 2, as without expression of *NIN* in the epidermis, the induction of *NIN* in the pericycle and the subsequent mitotic activity do not occur.

Previously, it has been proposed that NIN might be transported from epidermis to inner root cell layers and activates cell division there (Vernié et al., 2015b). However, as we showed that *NIN* is transcriptionally activated in the pericycle, it is not likely that NIN protein functions as a mobile signal. Further, we showed that the promoter region enriched in cytokinin response elements is essential for expression of *NIN* in the pericycle. Therefore, it is very likely that cytokinin triggers expression of *NIN* in the pericycle. This is further supported by the constitutive expression of the cytokinin receptor *CRE1* and the transcription factor *RR1* in the pericycle. As cytokinin is a small diffusible molecule, it

raises the question whether cytokinin is the mobile molecule produced in the epidermis. There are several observations to support this. First, NF perception induces accumulation of bioactive cytokinin and cytokinin biosynthesis genes are activated in the root epidermis (Jardinaud et al., 2016; Van Zeijl et al., 2015; Gühl et al., 2021). Second, application of NF induces cytokinin accumulation in cortex, endodermis and pericycle, suggesting that cytokinin acts as a non-cell-autonomous signal during nodule initiation (Van Zeijl et al., 2015). Third, mutation in ATP-binding cassette (ABC) transporter ABCG56, which can export bioactive cytokinin, causes a significant reduction of nodulation (Jarzyniak et al., 2021). Fourth, in case cytokinin is the mobile signal, this can explain why first the pericycle is mitotically activated, but not the cortex. This because cytokinin signaling genes such as *CRE1* and *RR1* are constitutively expressed in the pericycle, but not in the cortex.

Based on these observations the following model can be proposed: In the epidermis, NF signaling induces the biosynthesis of cytokinin which is then transported to the inner root cell layers by the ABCG56 transporter. This results in cytokinin accumulation in the pericycle where it activates the constitutively present cytokinin signaling cascade which subsequently activates *NIN* expression.

2.3 NIN activates cell division by induction of auxin accumulation

A function of NIN in the pericycle is to initiate cell division. This is most likely caused by accumulation of auxin. It has been shown that auxin biosynthesis genes *YUC2* and *YUC8* are induced in an inner root layer within 12 hours upon inoculation with rhizobia (Schiessl et al., 2019b). We showed by *in situ* hybridization that this inner layer is the pericycle (O. Kulikova unpublished). The induction of these auxin biosynthesis genes precedes cell divisions and depends on NIN.

Recent studies gave insight in the mechanism by which NIN induces auxin biosynthesis genes. It is shown that NIN directly induces the expression of the lateral root developmental program gene *LBD16* (*lob-domain protein 16*) as well as nodule specific transcription factors *NF-YA1* and *NF-YB1* (Soyano et al., 2019, 2013). Both NF-YA1 and LBD16 induce the expression of *STYLISH* (*STY*) transcription factor genes that are supposed to regulate the *YUC* genes (Shrestha et al., 2020; Schiessl et al., 2019b; Hossain et al., 2016). Co-expression of *LBD16*, *NF-YA1* and *NF-YB1* can partially complement the

nodule organogenesis deficient phenotype of *nin* weak allele *daphne* (Soyano et al., 2019). So, NIN activates nodule specific transcription factors as well as a transcription factor from the lateral root developmental program to induce auxin biosynthesis.

When the nodule primordium develops, the expression of auxin biosynthesis genes remains restricted to the pericycle derived cells (O. Kulikova unpublished). However, the cell division in the cortex by which nodule primordia further develop correlates with auxin accumulation (Xiao, 2015). Studies on auxin efflux carriers PIN proteins give an explanation for this. It has been shown that during initial stages (I-III) of nodule primordium formation, MtPIN2 and MtPIN10 are accumulate in all nodule primordium cells (Xiao, 2015). At later nodule primordium developmental stages (IV-VI), MtPIN10 mainly accumulates at the nodule periphery and MtPIN2 mainly accumulates at the future meristem (Xiao, 2015). Therefore, the auxin synthesized in the pericycle is probably transported to the cortical cells and induced division in these cells.

In addition to auxin biosynthesis, local reduction of the polar auxin transport in response to rhizobia inoculation has been proposed to play a role in auxin accumulation in nodule primordia. This is on one hand predicted by a modeling study and has also some experimental support (Deinum et al., 2016; Ng et al., 2015). The model showed that a local reduction of PIN proteins leads to accumulation of auxin in the inner cortex and pericycle (Deinum et al., 2016). Flavonoid accumulation most likely causes this local reduction of polar auxin transport and subsequent auxin accumulation (Ng et al., 2015). So, auxin biosynthesis and transport inhibition work together to establish the auxin maximum in the nodule primordium.

We have observed that at the stage when cortical cells have divided, both *NIN* and *NF-YA1* are expressed in these cells in a level similar to that in the pericycle (Chapter 2). Whether this is the result of auxin accumulation in these cells or the result of other factors, and whether NIN also has a role here, still require further study.

Taken together, the following model can be proposed: NIN activates the expression of *NF-YA1*, *NF-YB1*, and *LBD16* in the pericycle which induce the expression of *YUC* genes. This results in local auxin biosynthesis which induce local cell divisions and the auxin is further transported to the cortical cells. In addition to this transport of locally produced auxin, inhibition of both acropetal (from root base to root tip) and basipetal (from the root tip upwards) polar auxin

transport contributes to the formation of an auxin maximum in the cortex and the further development of the nodule primordium.

2.4 NIN controls two main transitions in the indeterminate nodule

During the development of indeterminate nodule there are two main transitions: from meristem to infection zone and from infection zone to fixation zone. Late stage Nod factor (NF) signaling and release of bacteria both occur at the transition from meristem to infection zone (Moling et al., 2014). In Chapter 3 we have shown that the induction of *NIN* expression at this transition depends on the late stage NF signaling. Furthermore, knockdown *NIN* in the apical part of Medicago nodules block the rhizobial release, suggesting the NF signaling controlled rhizobial release is depend on NIN (J. Liu unpublished). So most likely, late stage NF signaling induces expression of *NIN,* which further activates genes required for rhizobial release.

In addition to these release related genes, late stage NF signaling leads to expression of some other genes in a NIN dependent manner. Examples are defense suppressing genes which are essential to establish a successful symbiosis. The expression of these genes is significantly lower in the *nin* weak alleles *nin-13* and *nin-16* (Chapter 4) and two of them (*DNF2* and *SymCRK*) have been indicated as direct targets of NIN according to ChIP-seq analysis (Soyano et al., 2014; Liu et al., 2019a). This strongly suggests that NF signaling induces expression of defense suppressing genes in a *NIN* dependent manner.

The NF signaling receptors only accumulate in one/two cell layers at the transition from meristem to infection zone, but the level of NIN gradually accumulates from distal to proximal part of the infection zone (Chapter 4). Probable explanations for this are: 1) the accumulation of NIN depends on a positive feedback loop including NIN itself (proposed in Chapter 2); 2) the active form of CYCLOPS can be maintained in the nodule infection zone; 3) NF signaling might activate other factors to induce *NIN* in the infection zone.

The accumulation of NIN reaches its highest level at the proximal part of the infection zone and after transition to the fixation zone suddenly it is reduced to a much lower level (Chapter 4). This transition is accompanied by many other morphological and transcriptional changes (Chapter 1). The *nin* weak alleles *nin-13* and *nin-16* are unable to pass this transition and have a defense and premature senescence phenotype (Chapter 4). This is most likely due to a too low level of NIN which in turn leads to insufficient suppression of defense

responses. This underlines the importance of a high NIN level in passing the transition from infection to fixation zone. The sudden reduction of NIN protein after transition to fixation zone might be regulated by an active protein degradation mechanism. Whether it indeed exist and how important is this sudden reduction are interesting questions for future studies.

Based on these observations the following model can be proposed: late stage NF signaling induces the expression of *NIN*. NIN on one hand activate genes required for rhizobial release, on the other hand activate defense suppressing genes which can ensure symbiosome maintenance. Further, NIN gradually accumulates in the infection zone and reaches its highest level at the proximal part. This high level is required to activate sufficient target genes which ensure the successful transition to fixation zone. After transition to fixation zone, the level of NIN is maintained at a low level.

3. Concluding remarks

Research described in this thesis and several other studies showed that NIN plays an essential role in all nodule developmental stages. We discovered that the *cis*-regulatory regions in *NIN* promoter can induce its expression at different loci during nodule development. This contributes to its ability to function in multiple processes, like infection thread formation in the epidermis and cell division in pericycle (Chapter 2). Also the function of NIN might depend on its concentration. For example in root hair containing cells, a shorter promoter can only induce proper root hair curling, whereas the initiation of infection thread depends on the presence of transcription activator (CYCLOPS) binding site in the *NIN* promoter (Chapter 2); in the nodule, release of rhizobia correlates with a low level of NIN whereas the successful transition to the fixation zone requires a high level (Chapter 4).

As NIN recruitment was an important step in evolution of the nodulation, we identified changes which might contribute to its recruitment (Chapter 5 and 6). The function of NIN diverged from its non-symbiotic orthologues whose function still remain unclear. Further studies on these orthologues genes might provide more insight in the evolution and the mode of action of NIN.

References

Alfatih, A., Wu, J., Zhang, Z.S., Xia, J.Q., Jan, S.U., Yu, L.H., and Xiang, C. Bin (2020). Rice NIN-LIKE PROTEIN 1 rapidly responds to nitrogen deficiency and improves yield and nitrogen use efficiency. J. Exp. Bot. **71**: 6032–6042.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. **215**: 403–410.

Alunni, B. and Gourion, B. (2016). Terminal bacteroid differentiation in the legume−rhizobium symbiosis: nodule-specific cysteine-rich peptides and beyond. New Phytol. **211**: 411–417.

Alvarez, J.M., Schinke, A.L., Brooks, M.D., Pasquino, A., Leonelli, L., Varala, K., Safi, A., Krouk, G., Krapp, A., and Coruzzi, G.M. (2020). Transient genome-wide interactions of the master transcription factor NLP7 initiate a rapid nitrogen-response cascade. Nat. Commun. **11**: 1–13.

Andriankaja, A., Boisson-Dernier, A., Frances, L., Sauviac, L., Jauneau, A., Barker, D.G., and de Carvalho-Niebel, F. (2007). AP2-ERF transcription factors mediate Nod factor dependent Mt ENOD11 activation in root hairs via a novel cis-regulatory motif. Plant Cell **19**: 2866–2885.

Ané, J.M. et al. (2004). Medicago truncatula DMI1 Required for Bacterial and Fungal Symbioses in Legumes. Science (80-.). **303**: 1364–1367.

Arrighi, J.-F., Godfroy, O., de Billy, F., Saurat, O., Jauneau, A., and Gough, C. (2008). The RPG gene of Medicago truncatula controls Rhizobium-directed polar growth during infection. Proc. Natl. Acad. Sci. **105**: 9817–9822.

Arrighi, J.F. et al. (2006). The Medicago truncatula lysine motif-receptor-like kinase gene family includes NFP and new nodule-expressed genes. Plant Physiol. **142**: 265–279.

Batagov, A.O., Sandal, N., Mortensen, A., Schauser, L., Ellis, N., Tikhonovich, I. a, and Stougaard, J. (2003). The Sym35 Gene Required for Root Nodule Development in Pea Is an Ortholog of Nin from Lotus japonicus. plant Physiol. **131**: 1009–1017.

Baudin, M., Laloum, T., Lepage, A., Rípodas, C., Ariel, F., Frances, L., Crespi, M., Gamas, P., Blanco, F.A., Zanetti, M.E., de Carvalho-Niebel, F., and Niebel, A. (2015). A phylogenetically conserved group of nuclear factor-Y transcription factors interact to control nodulation in legumes. Plant Physiol. **169**: 2761–2773.

Benaben, V., Duc, G., Lefebvre, V., and Huguet, T. (1995). TE7, an inefficient symbiotic mutant of Medicago truncatula Gaertn. cv Jemalong. Plant Physiol **107**: 53–62.

Benfey, P.N. and Chua, N.-H. (1990). The Cauliflower Mosaic Virus 35S Promoter: Combinatorial Regulation of Transcription in Plants. Science (80-.). **250**: 959–966.

Berrabah, F., Balliau, T., Aït-Salem, E.H., George, J., Zivy, M., Ratet, P., and Gourion, B. (2018). Control of the ethylene signaling pathway prevents plant defenses during intracellular accommodation of the rhizobia. New Phytol. **219**: 310–323.

Berrabah, F., Bourcy, M., Cayrel, A., Eschstruth, A., Mondy, S., Ratet, P., and Gourion, B. (2014a). Growth conditions determine the DNF2 requirement for symbiosis. PLoS One **9**: e91866.

Berrabah, F., Bourcy, M., Eschstruth, A., Cayrel, A., Guefrachi, I., Mergaert, P., Wen, J., Jean, V., Mysore, K.S., Gourion, B., and Ratet, P. (2014b). A nonRD receptor-like kinase prevents nodule early senescence and defense-like reactions during symbiosis. New Phytol. **203**: 1305–1314.

Berrabah, F., Ratet, P., and Gourion, B. (2015). Multiple steps control immunity during the intracellular accommodation of rhizobia. J. Exp. Bot. **66**: 1977–1985.

Bertioli, D.J. et al. (2016). The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. Nat. Genet. **48**: 438–446.

Boivin, S., Kazmierczak, T., Brault, M., Wen, J., Gamas, P., Mysore, K.S., and Frugier, F. (2016). Different cytokinin histidine kinase receptors regulate nodule initiation as well as later nodule developmental stages in Medicago truncatula. Plant Cell Environ. **39**: 2198–2209.

Bonfante, P. and Genre, A. (2008). Plants and arbuscular mycorrhizal fungi: an evolutionary-develop-

mental perspective. Trends Plant Sci. **13**: 492–498.

Borisov, A.Y., Madsen, L.H., Tsyganov, V.E., Umehara, Y., Voroshilova, V.A., Batagov, A.O., Sandal, N., Mortensen, A., Schauser, L., Ellis, N., Tikhonovich, I.A., and Stougaard, J. (2003). The Sym35 Gene Required for Root Nodule Development in Pea Is an Ortholog of Nin from Lotus japonicus. Plant Physiol. **131**: 1009–1017.

Bourcy, M., Brocard, L., Pislariu, C.I., Cosson, V., Mergaert, P., Tadege, M., Mysore, K.S., Udvardi, M.K., Gourion, B., and 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 Phytol. **197**: 1250–1261.

Breakspear, A., Liu, C., Roy, S., Stacey, N., Rogers, C., Trick, M., Morieri, G., Mysore, K.S., Wen, J., Oldroyd, G.E.D., Downie, J.A., and Murray, J.D. (2014). The Root Hair "Infectome" of *Medicago truncatula* Uncovers Changes in Cell Cycle Genes and Reveals a Requirement for Auxin Signaling in Rhizobial Infection. Plant Cell Online **26**: 4680–4701.

Brewin, N.J. (2004). Plant cell wall remodelling in the rhizobium-legume symbiosis. CRC. Crit. Rev. Plant Sci. **23**: 293–316.

Brudno, M., Malde, S., Poliakov, A., Do, C.B., Couronne, O., Dubchak, I., and Batzoglou, S. (2003). Glocal alignment: Finding rearrangements during alignment. Bioinformatics **19**.

Brundrett, M.C. and Tedersoo, L. (2018). Evolutionary history of mycorrhizal symbioses and global host plant diversity. New Phytol. **220**: 1108–1115.

Bu, F., Rutten, L., Roswanjaya, Y.P., Kulikova, O., Rodriguez-Franco, M., Ott, T., Bisseling, T., van Zeijl, A., and Geurts, R. (2020). Mutant analysis in the nonlegume Parasponia andersonii identifies NIN and NF-YA1 transcription factors as a core genetic network in nitrogen-fixing nodule symbioses. New Phytol. **226**: 541–554.

Den Camp, R.O., Streng, A., De Mita, S., Cao, Q., Polone, E., Liu, W., Ammiraju, J.S.S., Kudrna, D., Wing, R., Untergasser, A., Bisseling, T., and Geurts, R. (2011). LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume Parasponia. Science (80-.). **331**: 909–912.

Cao, H., Qi, S., Sun, M., Li, Z., Yang, Y., Crawford, N.M., and Wang, Y. (2017). Overexpression of the maize ZmNLP6 and ZmNLP8 can complement the arabidopsis nitrate regulatory mutant nlp7 by restoring nitrate signaling and assimilation. Front. Plant Sci. **8**: 1–14.

Caretti, G., Salsi, V., Vecchi, C., Imbriano, C., and Mantovani, R. (2003). Dynamic recruitment of NF-Y and histone acetyltransferases on cell-cycle promoters. J. Biol. Chem. **278**: 30435–30440.

Catoira, R., Galera, C., De Billy, F., Penmetsa, R. V., Journet, E.P., Maillet, F., Rosenberg, C., Cook, D., Gough, C., and Denarie, J. (2000). Four genes of Medicago truncatula controlling components of a Nod factor transduction pathway. Plant Cell **12**: 1647–1665.

Chabaud, M., Gherbi, H., Pirolles, E., Vaissayre, V., Fournier, J., Moukouanga, D., Franche, C., Bogusz, D., Tisa, L.S., Barker, D.G., and Svistoonoff, S. (2016). Chitinase-resistant hydrophilic symbiotic factors secreted by Frankia activate both Ca2+ spiking and NIN gene expression in the actinorhizal plant Casuarina glauca. New Phytol. **209**: 86–93.

Chabaud, M., Lichtenzveig, J., Ellwood, S., Pfaff, T., and Journet, P. (2006). Vernalization , crossings and testing for pollen viability. Medicago truncatula Handbook. Mathesius U, Journet EP, Sumner LW: 1–13.

Chardin, C., Girin, T., Roudier, F., Meyer, C., and Krapp, A. (2014). The plant RWP-RK transcription factors: Key regulators of nitrogen responses and of gametophyte development. J. Exp. Bot. **65**: 5577–5587.

Charpentier, M., Bredemeier, R., Wanner, G., Takeda, N., Schleiff, E., and Parniske, M. (2008). Lotus japonicus Castor and Pollux are ion channels essential for perinuclear calcium spiking in legume root endosymbiosis. Plant Cell **20**: 3467–3479.

Charpentier, M., Sun, J., Martins, T.V., Radhakrishnan, G. V, Findlay, K., Soumpourou, E., Thouin, J., Véry, A., Sanders, D., Morris, R.J., and Oldroyd, G.E.D. (2016). Nuclear-localized cyclic nucleotide–gated channels mediate symbiotic calcium oscillations. Science (80-.). **352**: 1102–5.

Chen, C., Gao, M., Liu, J., and Zhu, H. (2007). Fungal symbiosis in rice requires an ortholog of a legume common symbiosis gene encoding a Ca2+/calmodulin-dependent protein kinase. Plant Physiol. **145**: 1619–1628.

Chow, C.N., Zheng, H.Q., Wu, N.Y., Chien, C.H., Huang, H. Da, Lee, T.Y., Chiang-Hsieh, Y.F., Hou, P.F., Yang, T.Y., and Chang, W.C. (2016). PlantPAN 2.0: An update of Plant Promoter Analysis Navigator for reconstructing transcriptional regulatory networks in plants. Nucleic Acids Res. **44**: D1154–D1164.

Clavijo, F. et al. (2015). The Casuarina NIN gene is transcriptionally activated throughout Frankia root infection as well as in response to bacterial diffusible signals. New Phytol. **208**: 887–903.

Combier, J., Frugier, F., Billy, F. De, Boualem, A., El-yahyaoui, F., Moreau, S., Vernié, T., Ott, T., Gamas, P., Crespi, M., and Niebel, A. (2006). MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in Medicago truncatula service MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in Medicago truncatula. Res. Commun.: 3084–3088.

Cooper, J.B. and Long, S.R. (1994). Morphogenetic rescue of Rhizobium meliloti nodulation mutants by trans-zeatin secretion. Plant Cell **6**: 215–225.

Couzigou, J.-M. et al. (2012). NODULE ROOT and COCHLEATA maintain nodule development and are legume orthologs of Arabidopsis BLADE-ON-PETIOLE genes. Plant Cell **24**: 4498–510.

Cruz-Ramírez, A. et al. (2012). A bistable circuit involving SCARECROW-RETINOBLASTOMA integrates cues to inform asymmetric stem cell division. Cell **150**: 1002–1015.

Daniel J. Gage (2004). Infection and Invasion of Roots by Symbiotic, Nitrogen-Fixing Rhizobia during Nodulation of Temperate Legumes. Microbiol. Mol. Biol. Rev. **68**: 280.

Darling, A.C.E., Mau, B., Blattner, F.R., and Perna, N.T. (2004). Mauve : Multiple Alignment of Conserved Genomic Sequence With Rearrangements Mauve : Multiple Alignment of Conserved Genomic Sequence With Rearrangements. Genome Res. **14**: 1394–1403.

Deinum, E.E., Kohlen, W., and Geurts, R. (2016). Quantitative modelling of legume root nodule primordium induction by a diffusive signal of epidermal origin that inhibits auxin efflux. BMC Plant Biol. **16**: 1–14.

Demina, I. V., Persson, T., Santos, P., Plaszczyca, M., and Pawlowski, K. (2013). Comparison of the Nodule vs. Root Transcriptome of the Actinorhizal Plant Datisca glomerata: Actinorhizal Nodules Contain a Specific Class of Defensins. PLoS One **8**: e72442.

Denarie, J. and Debelle, F. (1996). Rhizobium Lipo-Chitooligosaccharide Nodulation Factors: Signaling Molecules Mediating Recognition and Morphogenesis. Annu. Rev. Biochem. **65**: 503–535.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell **149**: 1233–1244.

Dingwall C and Laskey RA (1991). Nuclear target sequences - a consensus? Trends Biochem. Sci. **16**: 478–481.

Domonkos, Á. et al. (2017). NAD1 controls defense-like responses in Medicago truncatula symbiotic nitrogen fixing nodules following rhizobial colonization in a BacA-independent manner. Genes (Basel). **8**: 1–21.

Downie, J.A. (2014). Legume nodulation. Curr. Biol. **24**: R184–R190.

Dubrovsky, J.G., Sauer, M., Napsucialy-Mendivil, S., Ivanchenko, M.G., Friml, J., Shishkova, S., Celenza, J., and Benková, E. (2008). Auxin acts as a local morphogenetic trigger to specify

lateral root founder cells. Proc. Natl. Acad. Sci. U. S. A. **105**: 8790–8794.

Ehrhardt, D.W., Wais, R., and Long, S.R. (1996). Calcium spiking in plant root hairs responding to rhizobium modulation signals. Cell **85**: 673–681.

Eklund, D.M., Staldal, V., Valsecchi, I., Cierlik, I., Eriksson, C., Hiratsu, K., Ohme-Takagi, M., Sundstrom, J.F., Thelander, M., Ezcurra, I., and Sundberg, E. (2010). The Arabidopsis thaliana STYLISH1 Protein Acts as a Transcriptional Activator Regulating Auxin Biosynthesis. Plant Cell **22**: 349–363.

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

Engler, C., Youles, M., Gruetzner, R., Ehnert, T.M., Werner, S., Jones, J.D.G., Patron, N.J., and Marillonnet, S. (2014). A Golden Gate modular cloning toolbox for plants. ACS Synth. Biol. **3**: 839–843.

Ferguson, B.J., Li, D., Hastwell, A.H., Reid, D.E., Li, Y., Jackson, S.A., and Gresshoff, P.M. (2014). The soybean (Glycine max) nodulation-suppressive CLE peptide, GmRIC1, functions interspecifically in common white bean (Phaseolus vulgaris), but not in a supernodulating line mutated in the receptor PvNARK. Plant Biotechnol. J. **12**: 1085–1097.

Figure, S., Gus, D., and Factors, N. (2015). The NIN Transcription Factor Coordinates Diverse Nodulation Programs in Different Tissues of the Medicago truncatula Root (supplemental data).

Fournier, J., Teillet, A., Chabaud, M., Ivanov, S., Genre, A., Limpens, E., Carvalho-niebel, F. De, and Barker, D.G. (2015). Remodeling of the Infection Chamber before Infection Thread Formation Reveals a Two-Step Mechanism for Rhizobial Entry into the Host Legume Root Hair. Plant Physiol. **167**: 1233–1242.

Franssen, H.J., Vijn, I., Yang, W.C., and Bisseling, T. (1992). Developmental aspects of the Rhizobium-legume symbiosis. Plant Mol. Biol. **19**: 89–107.

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I. (2004). VISTA: Computational tools for comparative genomics. Nucleic Acids Res. **32**: 273–279.

Gauthier-Coles, C., White, R.G., and Mathesius, U. (2019). Nodulating legumes are distinguished by a sensitivity to cytokinin in the root cortex leading to pseudonodule development. Front. Plant Sci. **9**: 1–14.

Gautrat, P., Laffont, C., and Frugier, F. (2020). Compact Root Architecture 2 Promotes Root Competence for Nodulation through the miR2111 Systemic Effector. Curr. Biol.: 1–7.

Gavrin, A., Chiasson, D., Ovchinnikova, E., Kaiser, B.N., Bisseling, T., and Fedorova, E.E. (2016). VAMP721a and VAMP721d are important for pectin dynamics and release of bacteria in soybean nodules. New Phytol.

Gavrin, A., Kaiser, B.N., Geiger, D., Tyerman, S.D., Wen, Z., Bisseling, T., and Fedorova, E.E. (2014). Adjustment of Host Cells for Accommodation of Symbiotic Bacteria: Vacuole Defunctionalization, HOPS Suppression, and TIP1g Retargeting in Medicago. Plant Cell **26**: 3809–3822.

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

Gleason, C., Chaudhuri, S., Yang, T., and Mun, A. (2006). Nodulation independent of rhizobia induced by a calcium-activated kinase lacking autoinhibition. **441**: 1149–1152.

Gobbato, E. et al. (2012). A GRAS-type transcription factor with a specific function in mycorrhizal signaling. Curr. Biol. **22**: 2236–2241.

Goedhart, J., Hink, M.A., Visser, A.J.W.G., Bisseling, T., and Gadella, T.W.J. (2000). In vivo fluorescence correlation microscopy (FCM) reveals accumulation and immobilization of Nod factors in

R

root hair cell walls. Plant J. **21**: 109–119.

Gonzalez-Rizzo, S., Crespi, M., and Frugier, F. (2006a). The Medicago truncatula CRE1 cytokinin receptor regulates lateral root development and early symbiotic interaction with Sinorhizobium meliloti. Plant Cell **18**: 2680–2693.

González-Sama, A., Lucas, M.M., De Felipe, M.R., and Pueyo, J.J. (2004). An unusual infection mechanism and nodule morphogenesis in white lupin (Lupinus albus). New Phytol. **163**: 371–380.

Grandbastien, M.-A., Spielmann, A., and Caboche, M. (1989). Tntl, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature **337**: 376–380.

Grandbastien, M.A. (2015). LTR retrotransposons, handy hitchhikers of plant regulation and stress response. Biochim. Biophys. Acta - Gene Regul. Mech. **1849**: 403–416.

Griesmann, M. et al. (2018). Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. Science (80-.). **361**: eaat1743.

Grudkowska, M. and Zagdańska, B. (2004). Multifunctional role of plant cysteine proteinases. Acta Biochim. Pol. **51**: 609–624.

Guan, P., Ripoll, J.-J., Wang, R., Vuong, L., Bailey-Steinitz, L.J., Ye, D., and Crawford, N.M. (2017). Interacting TCP and NLP transcription factors control plant responses to nitrate availability. Proc. Natl. Acad. Sci. **114**: 2419–2424.

Gühl, K., Holmer, R., Xiao, T.T., Shen, D., Wardhani, T.A.K., Geurts, R., Zeijl, A. van, and Kohlen, W. (2021). The Effect of Exogenous Nitrate on LCO Signalling, Cytokinin Accumulation, and Nodule Initiation in Medicago truncatula. Genes (Basel). **12**: 988.

Haag, A.F. et al. (2011). Protection of sinorhizobium against host cysteine-rich antimicrobial peptides is critical for symbiosis. PLoS Biol. **9**: e1001169.

Hane, J.K. et al. (2017). A comprehensive draft genome sequence for lupin (Lupinus angustifolius), an emerging health food: insights into plant–microbe interactions and legume evolution. Plant Biotechnol. J. **15**: 318–330.

Hara-Nishimura, I., Hatsugai, N., Nakaune, S., Kuroyanagi, M., and Nishimura, M. (2005). Vacuolar processing enzyme: An executor of plant cell death. Curr. Opin. Plant Biol. **8**: 404–408.

Heckmann, A.B., Sandal, N., Bek, A.S., Madsen, L.H., Jurkiewicz, A., Nielsen, M.W., Tirichine, L., and Stougaard, J. (2011). Cytokinin Induction of Root Nodule Primordia in *Lotus japonicus* Is Regulated by a Mechanism Operating in the Root Cortex. Mol. Plant-Microbe Interact. **24**: 1385–1395.

Heidstra, R., Yang, W.C., Yalcin, Y., Peck, S., Emons, A., van Kammen, A., and Bisseling, T. (1997). Ethylene provides positional information on cortical cell division but is notinvolved in Nod factor-induced root hair tip growth in Rhizobium-legumeinteraction. Development **124**: 1781–1787.

Heyl, A. and Schmülling, T. (2003). Cytokinin signal perception and transduction. Curr. Opin. Plant Biol. **6**: 480–488.

Hirsch, A.M. (1992). Developmental biology of legume nodulation. New Phytol. **122**: 211–237.

Hirsch, S., Kim, J., Muñoz, A., Heckmann, A.B., Downie, J.A., and Oldroyd, G.E.D. (2009). GRAS proteins form a DNA binding complex to induce gene expression during nodulation signaling in Medicago truncatula. Plant Cell **21**: 545–57.

Horváth, B. et al. (2011a). Medicago truncatula IPD3 is a member of the common symbiotic signaling pathway required for rhizobial and mycorrhizal symbioses. Mol. Plant-Microbe Interact. **24**: 1345–1358.

Horváth, B. et al. (2011b). Medicago truncatula IPD3 Is a Member of the Common Symbiotic Signaling Pathway Required for Rhizobial and Mycorrhizal Symbioses. Mol. plant-microbe Interact. **24**: 1345–1358.

Hosoda, K., Imamura, A., Katoh, E., Hatta, T., Tachiki, M., Yamada, H., Mizuno, T., and Yamazaki, T.

(2002). Molecular Structure of the GARP Family of Plant Myb-Related DNA Binding Motifs of the Arabidopsis Response Regulators. Plant Cell **14**: 2015–2029.

Hossain, M.S. et al. (2016). Lotus japonicus NF-YA1 Plays an Essential Role During Nodule Differentiation and Targets Members of the SHI/STY Gene Family. Mpmi **29**: 950–964.

Hu, B. et al. (2019). Nitrate–NRT1.1B–SPX4 cascade integrates nitrogen and phosphorus signalling networks in plants. Nat. Plants **5**: 401–413.

Huisman, R., Hontelez, J., Mysore, K.S., Wen, J., Bisseling, T., and Limpens, E. (2016). A symbiosis-dedicated SYNTAXIN OF PLANTS 13II isoform controls the formation of a stable host-microbe interface in symbiosis. New Phytol. **211**: 1338–1351.

Imamura, A., Kiba, T., Tajima, Y., Yamashino, T., and Mizuno, T. (2003). In vivo and in vitro characterization of the ARR11 response regulator implicated in the His-to-Asp phosphorelay signal transduction in Arabidopsis thaliana. Plant Cell Physiol. **44**: 122–131.

Ivanov, S., Fedorova, E.E., Limpens, E., De Mita, S., Genre, a., Bonfante, P., and Bisseling, T. (2012). Rhizobium-legume symbiosis shares an exocytotic pathway required for arbuscule formation. Proc. Natl. Acad. Sci. **109**: 8316–8321.

Jardinaud, M.-F. et al. (2016). A Laser Dissection-RNAseq Analysis Highlights the Activation of Cytokinin Pathways by Nod Factors in the *Medicago truncatula* Root Epidermis. Plant Physiol. **171**: 2256–2276.

Jarzyniak, K., Banasiak, J., Jamruszka, T., Pawela, A., Di Donato, M., Novák, O., Geisler, M., and Jasiński, M. (2021). Early stages of legume–rhizobia symbiosis are controlled by ABCG-mediated transport of active cytokinins. Nat. Plants **7**: 428–436.

Jayaraman, D. et al. (2015). A role for the mevalonate pathway in early plant symbiotic signaling. Proc. Natl. Acad. Sci. **112**: E5378–E5378.

Kanamori, N. et al. (2006). A nucleoporin is required for induction of Ca2+ spiking in legume nodule development and essential for rhizobial and fungal symbiosis. Proc. Natl. Acad. Sci. U. S. A. **103**: 359–364.

Kang, H., Hong, Z., and Zhang, Z. (2015). A MYB Transcription Factor Interacts with NSP2 and Is Involved in Nodulation in Lotus japonicus. Biol. Nitrogen Fixat. **2**–**2**: 599–607.

Katoh, K. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. **30**: 3059–3066.

Kearse, M. et al. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics **28**: 1647–1649.

Kevei, Z. et al. (2007). 3-Hydroxy-3-methylglutaryl coenzyme A reductase1 interacts with NORK and is crucial for nodulation in Medicago truncatula. Plant Cell **19**: 3974–3989.

Kitaeva, A.B., Demchenko, K.N., Tikhonovich, I.A., Timmers, A.C.J., and Tsyganov, V.E. (2016). Comparative analysis of the tubulin cytoskeleton organization in nodules of Medicago truncatula and Pisum sativum: Bacterial release and bacteroid positioning correlate with characteristic microtubule rearrangements. New Phytol. **210**: 168–183.

Konishi, M. and Yanagisawa, S. (2019). The role of protein-protein interactions mediated by the PB1 domain of NLP transcription factors in nitrate-inducible gene expression. BMC Plant Biol. **19**: 1–12.

Kosslak, R.M. and Bohlool, B.B. (1984). Suppression of nodule development of one side of a split-root system of soybeans caused by prior inoculation of the other side. Plant Physiol. **75**: 125–130.

Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M., and Yanagawa, H. (2009). Six classes of nuclear localization signals specific to different binding grooves of importinα. J. Biol. Chem. **284**: 478–485.

Kosuta, S., Held, M., Hossain, M.S., Morieri, G., MacGillivary, A., Johansen, C., Antolín-Llovera,

M., Parniske, M., Oldroyd, G.E.D., Downie, A.J., Karas, B., and Szczyglowski, K. (2011). Lotus japonicus symRK-14 uncouples the cortical and epidermal symbiotic program. Plant J. **67**: 929–940.

Krusell, L. et al. (2002). Shoot control of root development and nodulation is mediated by a receptor-like kinase. Nature **420**: 422–426.

Kulikova, O., Franken, C., and Bisseling, T. (2018). In Situ Hybridization Method for Localization of mRNA Molecules in Medicago Tissue Sections. Methods Mol. Biol. **1822**: 145–159.

Kumar, A., Cousins, D.R., Liu, C.W., Xu, P., and Murray, J.D. (2020). Nodule inception is not required for arbuscular mycorrhizal colonization of Medicago truncatula. Plants **9**: 1–9.

Laffont, C., Ivanovici, A., Gautrat, P., Brault, M., Djordjevic, M.A., and Frugier, F. (2020). The NIN transcription factor coordinates CEP and CLE signaling peptides that regulate nodulation antagonistically. Nat. Commun. **11**: 1–13.

Lancelle, S.A. and Torrey, J.G. (1984). Early development of Rhizobium-induced root nodules of Parasponia rigida. I. Infection and early nodule initiation. Protoplasma **123**: 26–37.

Lang, C. and Long, S.R. (2015). Transcriptomic Analysis of *Sinorhizobium meliloti* and *Medicago truncatula* Symbiosis Using Nitrogen Fixation–Deficient Nodules. Mol. Plant-Microbe Interact. **28**: 856–868.

Laporte, P., Lepage, A., Fournier, J., Catrice, O., Moreau, S., Jardinaud, M.-F., Mun, J.-H., Larrainzar, E., Cook, D.R., Gamas, P., and Niebel, A. (2014). The CCAAT box-binding transcription factor NF-YA1 controls rhizobial infection. J. Exp. Bot. **65**: 481–494.

Lévy, J. et al. (2004). A Putative Ca 2 and Calmodulin- Dependent Protein Kinase Required. Science **303**: 1361–4.

Li, H. and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics **26**: 589–595.

Lim, C.W., Lee, Y.W., and Hwang, C.H. (2011). Soybean nodule-enhanced CLE peptides in roots act as signals in gmnark-mediated nodulation suppression. Plant Cell Physiol. **52**: 1613–1627.

Lim, C.W., Lee, Y.W., Lee, S.C., and Hwang, C.H. (2014). Nitrate inhibits soybean nodulation by regulating expression of CLE genes. Plant Sci. **229**: 1–9.

Limpens, E., Franken, C., Smit, P., Willemse, J., Bisseling, T., and Geurts, R. (2003). LysM Domain Receptor Kinases Regulating Rhizobial Nod Factor-Induced Infection. Science (80-.). **302**: 630–633.

Limpens, E., Mirabella, R., Fedorova, E., Franken, C., Franssen, H., Bisseling, T., and Geurts, R. (2005). Formation of organelle-like N2-fixing symbiosomes in legume root nodules is controlled by DMI2. Proc. Natl. Acad. Sci. U. S. A. **102**: 10375–80.

Limpens, E., Ramos, J., Franken, C., Raz, V., Compaan, B., Franssen, H., Bisseling, T., and Geurts, R. (2004). RNA interference in Agrobacterium rhizogenes-transformed roots of Arabidopsis and Medicago truncatula. J. Exp. Bot. **55**: 983–992.

Lin, J., Li, X., Luo, Z.L., Mysore, K.S., Wen, J., and Xie, F. (2018). NIN interacts with NLPs to mediate nitrate inhibition of nodulation in Medicago truncatula. Nat. Plants **4**: 942–952.

Liu, C. et al. (2019a). NIN Acts as a Network Hub Controlling a Growth Module Required for Rhizobial Infection. **179**: 1704–1722.

Liu, J. and Bisseling, T. (2020). Evolution of NIN and NIN-like Genes in Relation to Nodule Symbiosis. Genes (Basel). **11**: 777.

Liu, J., Rasing, M., Zeng, T., Klein, J., Kulikova, O., and Bisseling, T. (2021). NIN is essential for development of symbiosomes, suppression of defence and premature senescence in Medicago truncatula nodules. New Phytol. **230**: 290–303.

Liu, J., Rutten, L., Limpens, E., Molen, T. van der, Velzen, R. van, Chen, R., Chen, Y., Geurts, R.,

Kohlen, W., Kulikova, O., and Bisseling, T. (2019b). A Remote cis-Regulatory Region Is Required for NIN Expression in the Pericycle to Initiate Nodule Primordium Formation in Medicago truncatula. Plant Cell **31**: 68–83.

Liu, K. et al. (2017). Discovery of nitrate – CPK – NLP signalling in central nutrient – growth networks. Nature **545**: 311–316.

Liu, M., Soyano, T., Yano, K., Hayashi, M., and Kawaguchi, M. (2019c). ERN1 and CYCLOPS coordinately activate NIN signaling to promote infection thread formation in Lotus japonicus. J. Plant Res. **132**: 641–653.

Liu, W. (2013). Comparative and functional analysis of NODULATION SIGNALING PATHWAY 1 (NSP1) and NSP2 in rice and Medicago.

Luo, Z., Lin, J., Zhu, Y., Fu, M., Li, X., and Xie, F. (2021). NLP1 reciprocally regulates nitrate inhibition of nodulation through SUNN-CRA2 signaling in Medicago truncatula. Plant Commun. **2**: 100183.

Madsen, E.B., Madsen, L.H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., Sato, S., Kaneko, T., Tabata, S., Sandal, N., and Stougaard, J. (2004). A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. Chemtracts **17**: 508–514.

Marchive, C., Roudier, F., Castaings, L., Bréhaut, V., Blondet, E., Colot, V., Meyer, C., and Krapp, A. (2013). Nuclear retention of the transcription factor NLP7 orchestrates the early response to nitrate in plants. Nat. Commun. **4**: 1–9.

Markmann, K., Giczey, G., and Parniske, M. (2008). Functional adaptation of a plant receptor-kinase paved the way for the evolution of intracellular root symbioses with bacteria. PLoS Biol. **6**: 0497–0506.

Marsh, J.F., Rakocevic, A., Mitra, R.M., Brocard, L., Sun, J., Eschstruth, A., Long, S.R., Schultze, M., Ratet, P., and Oldroyd, G.E.D. (2007). Medicago truncatula NIN Is Essential for Rhizobial-Independent Nodule Organogenesis Induced by Autoactive Calcium/Calmodulin-Dependent Protein Kinase. Plant Physiol. **144**: 324–335.

Mathesius, U., Schlaman, H.R.M., Spaink, H.P., Sautter, C., Rolfe, B.G., and Djordjevic, M.A. (1998). Auxin transport inhibition precedes root nodule formation in white clover roots and is regulated by flavonoids and derivatives of chitin oligosaccharides. Plant J. **14**: 23–34.

Mens, C., Hastwell, A.H., Su, H., Gresshoff, P.M., Mathesius, U., and Ferguson, B.J. (2021). Characterisation of Medicago truncatula CLE34 and CLE35 in nitrate and rhizobia regulation of nodulation. New Phytol. **229**: 2525–2534.

Mergaert, P., Nikovics, K., Kelemen, Z., Maunoury, N., Vaubert, D., Kondorosi, A., and 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.

Messinese, E., Mun, J.H., Yeun, L.H., Jayaraman, D., Rouge, P., Barre, A., Lougnon, G., Schornack, S., Bono, J.J., Cook, D.R., and Ane, J.M. (2007). A novel nuclear protein interacts with the symbiotic DMI3 calcium- and calmodulin-dependent protein kinase of Medicago truncatula. Mol. Plant-Microbe Interact. **20**: 912–921.

Middleton, P.H. et al. (2007). An ERF Transcription Factor in Medicago truncatula That Is Essential for Nod Factor Signal Transduction. Plant Cell Online **19**: 1221–1234.

Mitra, R.M., Gleason, C. a, Edwards, A., Hadfield, J., Downie, J.A., Oldroyd, G.E.D., and Long, S.R. (2004). A Ca2+/calmodulin-dependent protein kinase required for symbiotic nodule development: Gene identification by transcript-based cloning. Proc. Natl. Acad. Sci. U. S. A. **101**: 4701–4705.

Moling, S., Pietraszewska-Bogiel, A., Postma, M., Fedorova, E., Hink, M. a, Limpens, E., Gadella, T.W.J., and Bisseling, T. (2014). Nod Factor Receptors Form Heteromeric Complexes and Are Essential for Intracellular Infection in Medicago Nodules. Plant Cell **26**: 4188–4199.

Mortier, V., Den Herder, G., Whitford, R., Van de Velde, W., Rombauts, S., D'Haeseleer, K., Holsters, M., and Goormachtig, S. (2010). CLE peptides control Medicago truncatula nodulation locally and systemically. Plant Physiol. **153**: 222–237.

Mu, X. and Luo, J. (2019). Evolutionary analyses of NIN - like proteins in plants and their roles in nitrate signaling. Cell. Mol. Life Sci.: 18–23.

Mulder, L., Lefebvre, B., Cullimore, J., and Imberty, A. (2006). LysM domains of Medicago truncatula NFP protein involved in Nod factor perception. Glycosylation state, molecular modeling and docking of chitooligosaccharides and Nod factors. Glycobiology **16**: 801–809.

Mun, J.H. et al. (2006). Distribution of microsatellites in the genome of Medicago truncatula: A resource of genetic markers that integrate genetic and physical maps. Genetics **172**: 2541–2555.

Mun, T., Bachmann, A., Gupta, V., Stougaard, J., and Andersen, S.U. (2016). Lotus Base: An integrated information portal for the model legume Lotus japonicus. Sci. Rep. **6**: 1–18.

Murakami, Y., Miwa, H., Imaizumi-Anraku, H., Kouchi, H., Downie, J.A., Kawaguchi, M., and Kawasaki, S. (2007). Positional cloning identifies Lotus japonicus NSP2, a putative transcription factor of the GRAS family, required for NIN and ENOD40 gene expression in nodule initiation. DNA Res. **13**: 255–265.

Ng, J.L.P., Hassan, S., Truong, T.T., Hocart, C.H., Laffont, C., Frugier, F., and Mathesius, U. (2015). Flavonoids and Auxin Transport Inhibitors Rescue Symbiotic Nodulation in the Medicago truncatula Cytokinin Perception Mutant cre1. Plant Cell **27**: 2210–2226.

Nishida, H., Handa, Y., Tanaka, S., Suzaki, T., and Kawaguchi, M. (2016). Expression of the CLE-RS3 gene suppresses root nodulation in Lotus japonicus. J. Plant Res. **129**: 909–919.

Nishida, H., Nosaki, S., Suzuki, T., Ito, M., Miyakawa, T., Nomoto, M., Tada, Y., Miura, K., Tanokura, M., Kawaguchi, M., and Suzaki, T. (2021). Different DNA-binding specificities of NLP and NIN transcription factors underlie nitrate-induced control of root nodulation. Plant Cell: 1–20.

Nishida, H., Tanaka, S., Handa, Y., Ito, M., Sakamoto, Y., Matsunaga, S., Betsuyaku, S., Miura, K., Soyano, T., Kawaguchi, M., and Suzaki, T. (2018). A NIN-LIKE PROTEIN mediates nitrate-induced control of root nodule symbiosis in Lotus japonicus. Nat. Commun. **9**: 499.

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

Nolis, I.K., McKay, D.J., Mantouvalou, E., Lomvardas, S., Merika, M., and Thanos, D. (2009). Transcription factors mediate long-range enhancer-promoter interactions. Proc. Natl. Acad. Sci. **106**: 20222–20227.

Okamoto, S., Ohnishi, E., Sato, S., Takahashi, H., Nakazono, M., Tabata, S., and Kawaguchi, M. (2009). Nod factor/nitrate-induced CLE genes that drive HAR1-mediated systemic regulation of nodulation. Plant Cell Physiol. **50**: 67–77.

Okamoto, S., Shinohara, H., Mori, T., Matsubayashi, Y., and Kawaguchi, M. (2013). Root-derived CLE glycopeptides control nodulation by direct binding to HAR1 receptor kinase. Nat. Commun. **4**: 1–7.

Oldroyd, G.E.D. (2013). Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. Nat. Rev. Microbiol. **11**: 252–263.

Oldroyd, G.E.D. and Downie, J.A. (2008). Coordinating Nodule Morphogenesis with Rhizobial Infection in Legumes. Annu. Rev. Plant Biol. **59**: 519–546.

Oldroyd, G.E.D., Murray, J.D., Poole, P.S., and Downie, J.A. (2011). The rules of engagement in the legume-rhizobial symbiosis. Annu. Rev. Genet. **45**: 119–144.

Ovchinnikova, E. et al. (2011). IPD3 controls the formation of nitrogen-fixing symbiosomes in pea and Medicago Spp. Mol. Plant. Microbe. Interact. **24**: 1333–44.

Parniske, M. (2000). Intracellular accommodation of microbes by plants: A common developmental program for symbiosis and disease? Curr. Opin. Plant Biol. **3**: 320–328.

Pawlowski, K. and Bisseling, T. (1996). Rhizobial and Actinorhizal Symbioses: What Are the Shared Features? Plant Cell **8**: 1899–1913.

Pawlowski, K. and Demchenko, K.N. (2012). The diversity of actinorhizal symbiosis. Protoplasma **249**: 967–979.

Pecrix, Y. et al. (2018). Whole-genome landscape of Medicago truncatula symbiotic genes. Nat. Plants **4**: 1017–1025.

Pedersen, B.S. and Quinlan, A.R. (2018). Mosdepth: Quick coverage calculation for genomes and exomes. Bioinformatics **34**: 867–868.

Pemberton, L.F. and Paschal, B.M. (2005). Mechanisms of receptor-mediated nuclear import and nuclear export. Traffic **6**: 187–198.

Perez Guerra, J.C., Coussens, G., De Keyser, A., De Rycke, R., De Bodt, S., Van De Velde, W., Goormachtig, S., and Holsters, M. (2010). Comparison of Developmental and Stress-Induced Nodule Senescence in Medicago truncatula. Plant Physiol. **152**: 1574–1584.

Peter, J., Young, W., and Haukka, K.E. (1996). Diversity and phylogeny of rhizobia. New Phytol. **133**: 87–94.

Pierre, O., Hopkins, J., Combier, M., Baldacci, F., Engler, G., Brouquisse, R., Hérouart, D., and Boncompagni, E. (2014). Involvement of papain and legumain proteinase in the senescence process of Medicago truncatula nodules. New Phytol. **202**: 849–863.

Pimprikar, P., Carbonnel, S., Paries, M., Katzer, K., Klingl, V., Bohmer, M.J., Karl, L., Floss, D.S., Harrison, M.J., Parniske, M., and Gutjahr, C. (2016). A CCaMK-CYCLOPS-DELLA complex activates transcription of RAM1 to regulate arbuscule branching. Curr. Biol. **26**: 987–998.

Pislariu, C.I. et al. (2012). A Medicago truncatula tobacco retrotransposon insertion mutant collection with defects in nodule development and symbiotic nitrogen fixation. Plant Physiol **159**: 1686–1699.

Plet, J., Wasson, A., Ariel, F., Le Signor, C., Baker, D., Mathesius, U., Crespi, M., and Frugier, F. (2011). MtCRE1-dependent cytokinin signaling integrates bacterial and plant cues to coordinate symbiotic nodule organogenesis in Medicago truncatula. Plant J. **65**: 622–633.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS One **5**.

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

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

Remy, W., Taylor, T.N., Hass, H., and Kerp, H. (1994). Four hundred-million-year-old vesicular arbuscular mycorrhizae. Proc. Natl. Acad. Sci. USA **91**: 11841–11843.

Robbins, J., Dilwortht, S.M., Laskey, R.A., and Dingwall, C. (1991). Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: Identification of a class of bipartite nuclear targeting sequence. Cell **64**: 615–623.

Roth, L.E. and Stacey, G. (1989). Bacterium release into host cells of nitrogen-fixing soybean nodules: the symbiosome membrane comes from three sources. Eur. J. Cell Biol. **49**: 13–23.

Roux, B. et al. (2014). An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. Plant J. **77**: 817–837.

Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., and Fernie, A.R. (2013). The flavonoid biosynthetic pathway in Arabidopsis: Structural and genetic diversity. Plant Physiol. Biochem. **72**: 21–34.

Santi, C., Bogusz, D., and Franche, C. (2013). Biological nitrogen fixation in non-legume plants. Ann. Bot. **111**: 743–767.

Sasaki, T., Suzaki, T., Soyano, T., Kojima, M., Sakakibara, H., and Kawaguchi, M. (2014). Shoot-derived cytokinins systemically regulate root nodulation. Nat. Commun. **5**: 4983.

Sato, S. et al. (2008). Genome structure of the legume, Lotus japonicus. DNA Res. **15**: 227–239.

Schauser, L., Roussis, a, Stiller, J., and Stougaard, J. (1999). A plant regulator controlling development of symbiotic root nodules. Nature **402**: 191–195.

Schauser, L., Wieloch, W., and Stougaard, J. (2005). Evolution of NIN-like proteins in Arabidopsis, Rice, and Lotus japonicus. J. Mol. Evol. **60**: 229–237.

Schiessl, K. et al. (2019a). NODULE INCEPTION Recruits the Lateral Root Developmental Program for Symbiotic Nodule Organogenesis in Medicago truncatula. Curr. Biol. **29**: 3657-3668.e5.

Schiessl, K. et al. (2019b). NODULE INCEPTION Recruits the Lateral Root Developmental Program for Symbiotic Nodule Organogenesis in Medicago truncatula. Curr. Biol. **29**: 3657–3668.

Schmid, M. and Jensen, T.H. (2008). Quality control of mRNP in the nucleus. Chromosoma **117**: 419–429.

Schmutz, J. et al. (2010). Genome sequence of the palaeopolyploid soybean. Nature **463**: 178–183.

Schnabel, E., Journet, E.P., De Carvalho-Niebel, F., Duc, G., and Frugoli, J. (2005). The Medicago truncatula SUNN gene encodes a CLV1-like leucine-rich repeat receptor kinase that regulates nodule number and root length. Plant Mol. Biol. **58**: 809–822.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods **9**: 671–675.

Searle, I.R., Men, A.E., Laniya, T.S., Buzas, D.M., Iturbe-Ormaetxe, I., Carroll, B.J., and Gresshoff, P.M. (2003). Long-distance signaling in nodulation directed by a CLAVATA1-like receptor kinase. Science (80-.). **299**: 109–112.

Sheen, J. (2002). Phosphorelay and transcription control in cytokinin signal transduction. Science (80-.). **296**: 1650–1652.

Shen, D., Xiao, T.T., van Velzen, R., Kulikova, O., Gong, X., Geurts, R., Pawlowski, K., and Bisseling, T. (2020). A Homeotic Mutation Changes Legume Nodule Ontogeny into Actinorhizal-Type Ontogeny. Plant Cell **32**: 1868–1885.

Sheokand, S. and Brewin, N.J. (2003). Cysteine proteases in nodulation and nitrogen fixation. Indian J. Exp. Biol. **41**: 1124–1132.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: From properties to genome-wide predictions. Nat. Rev. Genet. **15**: 272–286.

Shrestha, A., Zhong, S., Therrien, J., Huebert, T., Sato, S., Mun, T., Andersen, S.U., Stougaard, J., Lepage, A., Niebel, A., Ross, L., and Szczyglowski, K. (2020). Lotus japonicus Nuclear Factor YA1, a nodule emergence stage-specific regulator of auxin signalling . New Phytol.

Singh, S., Katzer, K., Lambert, J., Cerri, M., and Parniske, M. (2014). CYCLOPS, A DNA-binding transcriptional activator, orchestrates symbiotic root nodule development. Cell Host Microbe **15**: 139–152.

Sinharoy, S., Torres-Jerez, I., Bandyopadhyay, K., Kereszt, A., Pislariu, C.I., Nakashima, J., Benedito, V.A., Kondorosi, E., and 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.

Smit, P., Limpens, E., Geurts, R., Fedorova, E., Dolgikh, E., Gough, C., and Bisseling, T. (2007). Medicago LYK3, an Entry Receptor in Rhizobial Nodulation Factor Signaling. Plant Physiol. **145**: 183–191.

Sohlberg, J.J., Myrenås, M., Kuusk, S., Lagercrantz, U., Kowalczyk, M., Sandberg, G., and Sundberg, E. (2006). STY1 regulates auxin homeostasis and affects apical-basal patterning of the Arabidopsis gynoecium. Plant J. **47**: 112–123.

Soltis, D.E. et al. (1995). Chloroplast Gene Sequence Data Suggest a Single Origin of the Predisposition for Symbiotic Nitrogen Fixation in Angiosperms. Proc. Natl. Acad. Sci. **92**: 2647–2651.

Sorin, C., Declerck, M., Christ, A., Blein, T., Ma, L., Lelandais-Briere, C., Njo, M.F., Beeckman, T., Crespi, M., and Hartmann, C. (2014). A miR169 isoform regulates specific NF-YA targets and root architecture in Arabidopsis. New Phytol. **202**: 1197–1211.

Soyano, T., Hirakawa, H., Sato, S., Hayashi, M., and Kawaguchi, M. (2014). Nodule Inception creates a long-distance negative feedback loop involved in homeostatic regulation of nodule organ production. Proc. Natl. Acad. Sci. U. S. A. **111**: 14607–14612.

Soyano, T., Kouchi, H., Hirota, A., and Hayashi, M. (2013). NODULE INCEPTION Directly Targets NF-Y Subunit Genes to Regulate Essential Processes of Root Nodule Development in Lotus japonicus. PLoS Genet. **9**: e1003352.

Soyano, T., Shimoda, Y., Kawaguchi, M., and Hayashi, M. (2019). A shared gene drives lateral root development and root nodule symbiosis pathways in Lotus. Science **366**: 1021–1023.

Sprent, J.I. (2007). Evolving ideas of legume evolution and diversity: A taxonomic perspective on the occurrence of nodulation: Tansley review. New Phytol. **174**: 11–25.

Stam, M., Belele, C., Dorweiler, J.E., and Chandler, V.L. (2002). Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation. Genes Dev. **16**: 1906–1918.

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

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

Suzaki, T., Yano, K., Ito, M., Umehara, Y., and Suganuma, N. (2012). Positive and negative regulation of cortical cell division during root nodule development in Lotus japonicus is accompanied by auxin response. Development **4006**: 3997–4006.

Suzuki, W., Konishi, M., and Yanagisawa, S. (2013). The evolutionary events necessary for the emergence of symbiotic nitrogen fixation in legumes may involve a loss of nitrate responsiveness of the NIN transcription factor. Plant Signal. Behav. **8**: 6–10.

Svistoonoff, S. et al. (2013). The Independent Acquisition of Plant Root Nitrogen-Fixing Symbiosis in Fabids Recruited the Same Genetic Pathway for Nodule Organogenesis. PLoS One **8**.

Svistoonoff, S., Hocher, V., and Gherbi, H. (2014). Actinorhizal root nodule symbioses: What is signalling telling on the origins of nodulation? Curr. Opin. Plant Biol. **20**: 11–18.

Symmons, O. and Spitz, F. (2013). From remote enhancers to gene regulation: Charting the genome's regulatory landscapes. Philos. Trans. R. Soc. B Biol. Sci. **368**.

Tang, H. et al. (2014). An improved genome release (version Mt4.0) for the model legume Medicago truncatula. BMC Genomics **15**: 1–14.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**: 4673–4680.

Timmers, A.C., Auriac, M.C., and Truchet, G. (1999). Refined analysis of early symbiotic steps of the Rhizobium-Medicago interaction in relationship with microtubular cytoskeleton rearrangements. Development **126**: 3617–3628.

Timmers, A.C.J., Auriac, M.C., De Billy, F., and Truchet, G. (1998). Nod factor internalization and

microtubular cytoskeleton changes occur concomitantly during nodule differentiation in alfalfa. Development **125**: 339–349.

Tirichine, L. et al. (2006). Deregulation of a Ca2+/calmodulin-dependent kinase leads to spontaneous nodule development. Nature **441**: 1153–6.

Tsikou, D., Yan, Z., Holt, D.B., Abel, N.B., Reid, D.E., Madsen, L.H., Bhasin, H., Sexauer, M., Stougaard, J., and Markmann, K. (2018). Systemic control of legume susceptibility to rhizobial infection by a mobile microRNA. Science (80-.). **362**: 233–236.

Varshney, R.K. et al. (2013). Draft genome sequence of chickpea (Cicer arietinum) provides a resource for trait improvement. Nat Biotechnol **31**: 240–246.

Varshney, R.K. et al. (2012). Draft genome sequence of pigeonpea (Cajanus cajan), an orphan legume crop of resource-poor farmers. Nat. Biotechnol. **30**: 83–89.

Vasse, J., De Billy, F., Camut, S., and Truchet, G. (1990). Correlation between ultrastructural differentiation of bacteriods and nitrogen fixation in alfalfa nodules. J. Bacteriol. **172**: 4295–4306.

Veerappan, V., Jani, M., Kadel, K., Troiani, T., Gale, R., Mayes, T., Shulaev, E., Wen, J., Mysore, K.S., Azad, R.K., and Dickstein, R. (2016). Rapid identification of causative insertions underlying Medicago truncatula Tnt1 mutants defective in symbiotic nitrogen fixation from a forward genetic screen by whole genome sequencing. BMC Genomics **17**: 141.

De Vega, J.J. et al. (2015). Red clover (Trifolium pratense L.) draft genome provides a platform for trait improvement. Sci. Rep. **5**.

Van de Velde, W. et al. (2010). Plant Peptides Govern Terminal Differentiation of Bacteria in Symbiosis. Science (80-.). **328**: 25–28.

Van De Velde, W., Guerra, J.C.P., De Keyser, A., De Rycke, R., Rombauts, S., Maunoury, N., Mergaert, P., Kondorosi, E., Holsters, M., and Goormachtig, S. (2006). Aging in legume symbiosis. A molecular view on nodule senescence in Medicago truncatula. Plant Physiol. **141**: 711–720.

Van Velzen, R. et al. (2018). Comparative genomics of the nonlegume Parasponia reveals insights into evolution of nitrogen-fixing rhizobium symbioses. Proc. Natl. Acad. Sci. U. S. A. **115**: E4700–E4709.

van Velzen, R., Doyle, J.J., and Geurts, R. (2019). A Resurrected Scenario: Single Gain and Massive Loss of Nitrogen-Fixing Nodulation. Trends Plant Sci. **24**: 49–57.

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

Vijn, I., Martinez-Abarca, F., Yang, W.-C., Neves, L. das, Brussel, A. van, Kammen, A. van, and Bisseling, T. (1995). Early nodulin gene expression during Nod factor-induced processes in Vicia sativa.: 8: 111-119.

Vinardell, J.M., Fedorova, E., Cebolla, A., Kevei, Z., Horvath, G., Kelemen, Z., Tarayre, S., Roudier, F., Mergaert, P., Kondorosi, A., and Kondorosi, E. (2013). Endoreduplication Mediated by the Anaphase-Promoting Complex Activator CCS52A Is Required for Symbiotic Cell Differentiation in Medicago truncatula Nodules. Plant Cell Online **23**: 2093–2105.

Wagner, G.P., Kin, K., and Lynch, V.J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. **131**: 281–285.

Wang, C. et al. (2016). NODULES WITH ACTIVATED DEFENSE 1 is required for maintenance of rhizobial endosymbiosis in Medicago truncatula. New Phytol. **212**: 176–191.

Wang, D., Griffitts, J., Starker, C., Fedorova, E., Limpens, E., Ivanov, S., Bisseling, T., and Long, S. (2010). A nodule-specific protein secretory pathway required for nitrogen-fixing symbiosis. Science (80-.). **327**: 1126–1130.

Wang, L., Sun, Z., Su, C., Wang, Y., Yan, Q., Chen, J., Ott, T., and Li, X. (2019). A GmNI-

Na-miR172c-NNC1 Regulatory Network Coordinates the Nodulation and Autoregulation of Nodulation Pathways in Soybean. Mol. Plant **12**: 1211–1226.

Wang, Y., Wang, L., Zou, Y., Chen, L., Cai, Z., Zhang, S., Zhao, F., Tian, Y., Jiang, Q., Ferguson, B.J., Gresshoff, P.M., and Li, X. (2014). Soybean miR172c targets the repressive AP2 transcription factor NNC1 to activate ENOD40 expression and regulate nodule initiation. Plant Cell **26**: 4782–4801.

Weber, B., Zicola, J., Oka, R., and Stam, M. (2016). Plant Enhancers: A Call for Discovery. Trends Plant Sci. **21**: 974–987.

Weis, K. (2003). Regulating access to the genome: Nucleocytoplasmic transport throughout the cell cycle. Cell **112**: 441–451.

Xiao, A., Yu, H., Fan, Y., Kang, H., Ren, Y., Huang, X., Gao, X., Wang, C., Zhang, Z., Zhu, H., and Cao, Y. (2020). Transcriptional regulation of NIN expression by IPN2 is required for root nodule symbiosis in Lotus japonicus. New Phytol. **1**: 513–528.

Xiao, T.T. (2015). Root and Nodule Lateral Organ Development in N2-Fixing Plants.

Xiao, T.T., Schilderink, S., Moling, S., Deinum, E.E., Kondorosi, E., Franssen, H., Kulikova, O., Niebel, A., and Bisseling, T. (2014). Fate map of Medicago truncatula root nodules. Development **141**: 3517–3528.

Xie, F., Murray, J.D., Kim, J., Heckmann, A.B., Edwards, A., Oldroyd, G.E.D., and Downie, J.A. (2012). Legume pectate lyase required for root infection by rhizobia. Proc. Natl. Acad. Sci. U. S. A. **109**: 633–638.

Yan, D. et al. (2016). NIN-like protein 8 is a master regulator of nitrate-promoted seed germination in Arabidopsis. Nat. Commun. **7**: 1–11.

Yang, W.-C., Horváth, B., Hontelez, J., Kammen, A. Van, and Bisseling, T. (1991). In situ Localization of Rhizobium mRNAs in Pea Root Nodules: nif A and nif H Localization. Mol. Plant-Microbe Interact. **4**: 464.

Yano, K. et al. (2008). CYCLOPS, a mediator of symbiotic intracellular accommodation. Proc. Natl. Acad. Sci. U. S. A. **105**: 20540–20545.

Yoro, E., Suzaki, T., Toyokura, K., Miyazawa, H., Fukaki, H., and Kawaguchi, M. (2014). A positive regulator of nodule organogenesis, NODULE INCEPTION, acts as a negative regulator of rhizobial infection in Lotus japonicus. Plant Physiol. **165**: 747–58.

Young, N.D. et al. (2011). The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature **480**: 520–524.

Yu, H., Xiao, A., Dong, R., Fan, Y., Zhang, X., Liu, C., Wang, C., Zhu, H., Duanmu, D., Cao, Y., and Zhang, Z. (2018). Suppression of innate immunity mediated by the CDPK-Rboh complex is required for rhizobial colonization in Medicago truncatula nodules. New Phytol. **220**: 425–434.

Zanetti, M.E., Blanco, F.A., Beker, M.P., Battaglia, M., and Aguilar, O.M. (2010). A C subunit of the plant nuclear factor NF-Y required for rhizobial infection and nodule development affects partner selection in the common bean-Rhizobium etli Symbiosis. Plant Cell **22**: 4142–4157.

Van Zeijl, A., Op Den Camp, R.H.M., Deinum, E.E., Charnikhova, T., Franssen, H., Op Den Camp, H.J.M., Bouwmeester, H., Kohlen, W., Bisseling, T., and Geurts, R. (2015). Rhizobium Lipo-chitooligosaccharide Signaling Triggers Accumulation of Cytokinins in Medicago truncatula Roots. Mol. Plant **8**: 1213–1226.

Zeng, T., Holmer, R., Hontelez, J., te Lintel-Hekkert, B., Marufu, L., de Zeeuw, T., Wu, F., Schijlen, E., Bisseling, T., and Limpens, E. (2018). Host- and stage-dependent secretome of the arbuscular mycorrhizal fungus *Rhizophagus irregularis*. Plant J.: 411–425.

Zhao, L., Zhang, W., Yang, Y., Li, Z., Li, N., Qi, S., Crawford, N.M., and Wang, Y. (2018). The Arabidopsis NLP7 gene regulates nitrate signaling via NRT1.1-dependent pathway in the presence of ammonium. Sci. Rep. **8**: 1–13.

SUMMARY

Nitrogen fixing root nodule symbiosis only occurs in plant species that belong to a single phylogenetic group, referred to as the nitrogen fixing clade (NFC). These species use a conserved signaling pathway to initiate nodule symbiosis. As this signaling pathway is shared with the much older and far more widespread arbuscular mycorrhizal (AM) symbiosis, it is most likely recruited from the AM symbiosis when nodulation evolved. *NIN* is a transcription factor and is among the first genes that are specifically induced during nodule symbiosis by this common symbiotic signaling pathway. The key role of NIN in the nodule symbiosis has been demonstrated by both evolutionary analyses and functional studies. All (studied) nodulating species possess *NIN*, while the species that lost the nodulation ability in general lost *NIN*. NIN is essential in multiple developmental processes during root nodule formation. However, although the importance of NIN has been demonstrated, how it achieves this versatility remains elusive.

In this thesis, I focused on NIN and tried to answer this question by analyzing its regulation of expression, functional studies in different nodule developmental stages and identifying evolutionary changes during its recruitment in the nodulation process.

During nodule initiation, one of the main puzzles was how NIN controls both infection in the epidermis and mitotic activity in the inner root layers at the same time. We solved this cold case by discovering distinct regulatory regions in its promoter. The regulatory sequences sufficient for the epidermal infection process are located within a 5 kb region directly upstream of the NIN start codon in *Medicago truncatula* (Medicago). Furthermore, we have identified a remote upstream cis-regulatory region required for the expression of *NIN* in the pericycle and we showed that this region is essential for nodule organogenesis. This region contains putative cytokinin response elements and is shown to be conserved in other legume species. Both the cytokinin receptor 1, which is essential for nodule primordium formation and the B-type response regulator *RR1* are expressed in the pericycle in the susceptible zone of the uninoculated roots. This, together with the identification of the cytokinin-responsive elements in the *NIN* promoter, strongly suggests that *NIN* expression is initially triggered by cytokinin signaling in the pericycle to initiate nodule primordium formation.

The expression of *NIN* is induced by rhizobia secreted Nod factors (NFs). Transcriptional changes induced in roots upon application of NFs have been

extensively studied. Although it has been shown that the NF receptors also accumulate in mature Medicago nodules, the transcriptional changes induced during this late stage NF signaling have never been investigated. To identify NF signaling controlled genes that are expressed in nodules, we made use of spontaneous nodules which were formed by overexpression of an auto-active *CCaMK* gene. These spontaneous nodules showed a wildtype (WT)-like tissue organization and a developmental gradient along their longitudinal axis. Transcriptome analysis showed that many genes, which are especially expressed in infection and fixation zone of WT nodules, are induced in the spontaneous nodules. Further, to study late stage NF signaling controlled genes and distinguish these from genes which are activated upon release of rhizobia from the infection threads, we made use of a mutant blocked in release due to disturbed of NF signaling (TE7) and nodules formed on *Vamp721d&e RNAi* roots in which release is also blocked, but NF signaling is not affected. Combined with laser microdissection and transcriptome analysis, we identified the set of late stage NF signalling controlled genes and showed that this is markedly different from that induced in roots treated with NFs. It is noteworthy that *NIN* is among the few genes which is induced by NF signaling in both roots and nodules.

The essential role of NIN during root nodule initiation has been well demonstrated, but it's role in late nodule developmental stages remained unclear. By investigating the *NIN* mRNA as well as NIN protein accumulation pattern in Medicago nodules, we found that the expression of *NIN* is highly induced in the infection zone and reached the highest level in the proximal part of this zone. Further, by detailed studies of two *nin* weak allele mutants, *nin-13/16*, we showed that NIN plays a key role in the successful transition from infection zone to fixation zone, where nitrogen fixation take place. Based on cytological as well as transcriptional analysis, we concluded that suppression of defence/senescence-related genes by NIN is necessary for this transition. Therefore, NIN also plays a vital role in late stages of the formation of a functional nodule. Interestingly, we found that the level of NIN in the nodule is likely critical for the successful transition from infection zone to fixation zone, but the PB1 domain, one of the two most conserved domains of NIN, is not essential for the formation of functional nodules.

NIN is the founding member of the NIN-like protein (NLP) family. It arose by duplication and this occurred before nodulation evolved. Therefore, several plant species outside the NFC have NIN orthologous. Compared with its orthologues outside of the NFC, at least two major changes occurred when NIN was

recruited in nodulation: the specific induction of expression by the NF signalling pathway and a constitutive nuclear localization. Concerning the regulation of *NIN* expression during nodule evolution, it first gained the CYCLOPS binding site, most likely in the ancestor of the NFC. Second, at an early stage of nodulation, the constitutive expression is lost by which, in nodulating plants, *NIN* became nodule-specifically expressed. Third, it developed cytokinin-regulated expression during nodule organogenesis, which most likely occurred in the legume branch. Concerning the protein function, although NIN and NLPs share a conserved DNA binding domain and many downstream targets, they are not functionally equivalent. The main reason for this could be that most NLPs are only nuclear localized when high nitrate is sensed, whereas nodule is formed under low nitrate condition. In contrast, NIN is constitutively nuclear localized. Further, we identified a few amino acid differences between NIN and its NLP orthologues. These changes probably were introduced in NIN when it was recruited in the nodulation process and might contribute to its constitutive nuclear localization.

In the final chapter, I integrated the findings described in this thesis and the knowledge from literature and proposed a model that can explain how NIN functions in multiple steps of the nodulation process.

ACKNOWLEDGMENTS

Andrea, Tijs, Peter, Jeroen, Norbert, Wilma, Menno, Jorge, Renan, Kavya, Michiel, Merijn, Vera, Jana, Kiki, Iris ...

Many thanks to the Chinese community who have been so supportive to me: Xu, Wenkun, DuDu, Fengjiao, Guiling, Jianyong & Liu, Peng, Zhichun Zhizhi Lulu, Jing & Lili, Yinshan, Fuxi, Mengmeng, Jundi, Yueyang, Honglei, Zhuang, Xun Qian, Jun, Qingqin, Haolin, Li Qian, Fang, Zhiyong, Qi, Ma Yuanyuan, Henan, Tang Han, Lu & Brother Biao, Yanting, Zhang Yuanyuan, Lu Yuan, Baojian, Rufang, Xu Wei, Tiantian, Zhou Libin, Zhang Wei, Bai Bing, Zhang Yanxia, Song Wei, Song Shuang, Ji Yongran, Wu Jinbing, Xiong Wei, Jin Wenqing, Jiang Bin, Xiaowei, Xinping, Mengran ...

Jing, Peng and Siqi, special thanks for always bringing lunch for me.

Tian, thanks for the happy times we spend together, and the great foods you cooked.

Thanks to the orange cat[Tiger] who frequently visit me during my PhD, and carefully checking whether everything is ok in my room.

Thanks Prof. Dr. Xie Xiaodong, for choosing me to work in your lab and get in touch with molecular biology. Thanks Prof. Dr. Song Weining for giving me the opportunity to continue MSc in your lab, and also thank you, Nie Xiaojun and Tong Wei for your support in applying for CSC.

感谢我的朋友们一直以来对我的关心和支持：娇，帅，静，盼妮，伟，豪爽，小黑，臣...

感谢我亲爱的家人。妈妈，感谢你总是支持我，给我追求的自由。姐姐，是你的陪伴、鼓励、安慰和信任，让我有努力实现这一切的力量。爸爸，感谢你一直在我梦和梦想的边缘徘徊，指引我方向，不曾离开。

Jieyu 洁宇

Wageningen, 3rd Nov 2021

ABOUT THE AUTHOR

Jieyu Liu (刘洁宇) was born on the 4th of August 1992 in Huadian, Jilin, China. She started her bachelor education at Tianjin Agricultural University (2009-2013), majored in Seed Science and Engineering. There she worked in the laboratory of Prof. Dr. Xiaodong Xie and get in touch with molecular biology experiments. Her study on the effects of mRNA N^6-adenosine methylation on cell morphogenesis in *Arabidopsis thaliana* was awarded as the outstanding thesis of Tianjin Agricultural University in 2013.

After graduation, she continued academic education with MSc at Northwest A&F University (2013-2015), majored in Seed Industry. Under the supervision of Prof. Dr. Weining Song, she studied the function of wheat Tryptophan Decarboxylase (TDC) gene which catalyze the first committed step for melatonin biosynthesis in plants. In addition, she joined the evaluation of wild emmer as the germplasm resources of bread wheat.

In 2015, with the financial support from China Scholarship Council (CSC), she joined the laboratory of Molecular Biology at Wageningen University to perform her PhD research. Under supervision of Prof. Dr. Ton Bisseling, Dr. Olga Kulikova and Dr. Wouter Kohlen, she studied the role of key transcription factor NIN in nodulation. Main findings obtained during her doctoral study are presented in this thesis.

LIST OF PUBLICATIONS

Liu, J.*, Rutten, L.*, Limpens, E., Molen, T. van der, Velzen, R. van, Chen, R., Chen, Y., Geurts, R., Kohlen, W., Kulikova, O., and Bisseling, T. (2019). A Remote cis-Regulatory Region Is Required for *NIN* Expression in the Pericycle to Initiate Nodule Primordium Formation in *Medicago truncatula*. **Plant Cell**. 31: 68–83. (*Equally contribution)

Liu, J., Rasing, M., Tian, Z., Klein, J., Kulikova, O., and Bisseling, T. (2021). NIN is Essential for Development of Symbiosomes, Suppression of Defence and Premature Senescence in *Medicago truncatula* Nodules. **New Phytologist**. 230: 290–303.

Liu, J. and Bisseling, T. (2020). Evolution of NIN and NIN-like Genes in Relation to Nodule Symbiosis. **Genes**. 11: 777.

Wang, P., Snijders, R., Kohlen, W., **Liu, J**., Bisseling, T., and Limpens, E. (2021). Medicago SPX1 and SPX3 regulate phosphate homeostasis, mycorrhizal colonization, and arbuscule degradation. **Plant Cell**: 0:1–17.

Liu, J., Shen, D., Liu, W., Huisman, R., Limpens, E., Fedorova, E., Hetty, B., Zeng, T., Elio, S., Kulikova, O., and Bisseling, T. Nod Factor Signaling Controlled Genes in *Medicago truncatula* Nodule. In preparation.

Liu, J., Yan, S., Škoríková, M., Kulikova, O., and Bisseling, T. Changes in NIN related to its recruitment in nodulation. In preparation.

Education Statement of the Graduate School

Experimental Plant Sciences

Issued to: **Jieyu Liu**
Date: **03 December 2021**
Group: **Laboratory of Molecular Biology**
University: **Wageningen University & Research**

1) Start-Up Phase	*date*	*cp*
► **First presentation of your project**		
NIN is Essential for Release and Differentiation of rhizobia in Medicago truncatula Nodules	19 February, 2016	1.5
► **Writing or rewriting a project proposal**		
► **Writing a review or book chapter**		
Liu J. & Bisseling T. Evolution of NIN and NIN-like Genes in Relation to Nodule Symbiosis. Genes. 2020; 11(7):777. https://doi.org/10.3390/genes11070777	11 July, 2020	6.0
► **MSc courses**		
Subtotal Start-Up Phase		7.5

2) Scientific Exposure	*date*	*cp*
► **EPS PhD student days**		
EPS Get2Gether, Soest, NL	28-29 January, 2016	0.6
EPS Get2Gether, Soest, NL	09-10 February, 2017	0.6
► **EPS theme symposia**		
EPS theme 1 'Developmental Biology of Plants', Wageningen, NL	30 January, 2018	0.3
EPS theme 3 'Metabolism and Adaptation', Wageningen, NL	13 March, 2018	0.3
EPS theme 1 'Developmental Biology of Plants', Wageningen, NL	05 February, 2020	0.3
► **Lunteren Days and other national platforms**		
Annual meeting 'Experimental Plant Sciences', Lunteren, NL	11-12 April, 2016	0.6
Annual meeting 'Experimental Plant Sciences', Lunteren, NL	10-11 April, 2017	0.6
Annual meeting 'Experimental Plant Sciences', Lunteren, NL	09-10 April, 2018	0.6
Annual meeting 'Experimental Plant Sciences', Lunteren, NL	08-09 April, 2019	0.6
Annual meeting 'Experimental Plant Sciences', online	12-13 April, 2019	0.5
► **Seminars (series), workshops and symposia**		
Lecture Jean-Francois Arrighi; 'Evolution of Nod factor-independent rhizobium symbiosis'	18 October, 2017	0.1
Lecture Giles Oldroyd; 'Recognition of symbiotic microorganisms by plants'	19 October, 2017	0.1
Symposium: Lunch and Meet with Janet I. Sprent, Euan James & Sofie Goormachtig	23 January, 2018	0.2
Public lecture Prof. Alga Zuccaro	10 September, 2019	0.1
Public lecture Prof. Ikram Blilou	10 September, 2019	0.1
Lecture Jian Xu; 'The root of single cell CAPability'	02 December, 2019	0.1
Lecture Pascal Ratet; 'Suppression of defense during Medicago root nodule formation'	03 December, 2019	0.1
Virtual Seminars in Symbiosis: Uta Paszkowski	27 April, 2020	0.1
Virtual Seminars in Symbiosis: Martin Parniske	27 April, 2020	0.1
Virtual Seminars in Symbiosis: Dugald Reid	25 May, 2020	0.1
Virtual Seminars in Symbiosis: Chao Su	25 May, 2020	0.1
Virtual Seminars in Symbiosis: Maria Harrison	29 June, 2020	0.1
Virtual Seminars in Symbiosis: Benoit Lefebvre	29 June, 2020	0.1
Virtual Seminars in Symbiosis: Takashi Soyano	27 July, 2020	0.1
Virtual Seminars in Symbiosis: Mélanie Rich	27 July, 2020	0.1
Virtual Seminars in Symbiosis: Nadia Radzman	31 August, 2020	0.1
Virtual Seminars in Symbiosis: Florian Frugier	28 September, 2020	0.1
Virtual Seminars in Symbiosis: Chloe Cathebras	28 September, 2020	0.1
► **Seminar plus**		
► **International symposia and congresses**		
13th European Nitrogen Fixation Conference (ENFC), Stockholm, Sweden	18-21 August, 2018	1.2
21st International Congress on Nitrogen Fixation (ICNF), Wuhan, China	10-15 October, 2019	1.5
4th Adam Kondorosi Symposium "Beneficial Plant-Microbe Interactions", Gif-sur-Yvette, France	26-27 November, 2019	0.6
► **Presentations**		
Poster: "NIN: The Key to Successful Nodulation" at Annual meeting 'Experimental Plant Sciences', Lunteren, NL	08-09 April, 2019	1.0
Poster: "NIN is essential for development of symbiosomes and suppression of defense-related response in Medicago	10-15 October, 2019	1.0
Poster: "NIN is essential for development of symbiosomes and suppression of defense-related response in Medicago	26-27 November, 2019	1.0
Talk: "NIN: The Key to Successful Nodulation", Beijing University of Agriculture, China	23 October, 2019	1.0
Talk: "NIN: The Key to Nodulation" at EPS Theme 1, Wageningen, NL	05 February, 2020	1.0
► **IAB interview**		
► **Excursions**		
Subtotal Scientific Exposure		15.2

3) In-Depth Studies	*date*	*cp*
► **Advanced scientific courses & workshops**		
EPS course "Transcription Factors and Transcriptional Regulation", Wageningen, NL	12-14 December, 2016	1.0
Electron Microscopy Course "The Basics: from Amsterdam to Wageningen", Amsterdam & Wageningen, NL	12-16 November, 2018	1.6
PE&RC/WIMEK course "Introduction to R and R Studio", Wageningen, NL	09 February - 02 March, 2021	0.9
EPS/ELIXIR Belgium course "Gentle hands-on introduction to Python programming", online	18 March - 01 April, 2021	0.9
► **Journal club**		
Member of literature discussion group at Laboratory of Molecular Biology	2015-2020	3.0
► **Individual research training**		
Subtotal In-Depth Studies		7.4

4) Personal Development	*date*	*cp*
► **General skill training courses**		
EPS Introduction Course, Wageningen, NL	11 February, 2016	0.3
WGS course "Project and Time Management", Wageningen, NL	10 May - 21 June, 2016	1.5
WGS course "Techniques for Writing and Presenting a Scientific Paper", Wageningen, NL	6-9 September, 2016	1.2
WGS course "Brain Training", Wageningen, NL	19 November, 2019	0.3
EPS Writing Support Group, Hertz-Training for Scientists, online	19 August - 16 September, 2020	1.0
EPS Workshop "Applying for a Marie Skłodowska-Curie Fellowship: from proposal to project", online	10 December, 2020	0.2
► **Organisation of meetings, PhD courses or outreach activities**		
Cowriting project proposal 'Molecular networks controlling development of rhizobium infected cells in Medicago	12 February, 2019	1.5
► **Membership of EPS PhD Council**		
Subtotal Personal Development		6.0

TOTAL NUMBER OF CREDIT POINTS*	**36.1**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS with a minimum

* A credit represents a normative study load of 28 hours of study.

182