Developing tools for the functional analysis of NCR peptides in *Medicago truncatula*

PhD THESIS

By

Senlei ZHANG

Supervised by Attila KERESZT, BRC, Szeged, HUNGARY

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

Abbreviations

ABC: ATP-binding cassette

AD: activation domain

APS: ammonium persulfate

AM: arbuscular mycorrhiza

AON: autoregulation of nodulation

BacA: bacterial development factor A

BclA: *Bradyrhizobium* BacA-like

C: haploid DNA content

CCaMK: calcium-and calmodulin-dependent protein kinase

Ccs52: cell cycle switch gene

CDS: coding sequence

CSSP: common symbiosis signaling pathway

CLE: CLAVATA3/embryo-surrounding region

CRE: cytokinin response

CRISPR: clustered regularly interspaced short palindromic repeats

dpi: days post inoculation

DMI: does not make infection

DNF: does not fix nitrogen

DUF: domain of unknown function

E: elongated

ENOD: early nodulin

EM: ectomycorrhiza

EMSA: electrophoretic mobility shift assay

ER: endoplasmic reticulum

ERN: ethylene responsive factor required for nodulation

EST: expressed sequence tag

GFP: green fluorescent protein

GRP: glycine-rich peptide

GUS: β-glucuronidase reporter gene

HMG: high mobility group protein

HPLC: high performance liquid chromatography

HrrP: host range restriction peptidase

IRLC: inverted repeat lacking clade of legumes

IT: infection thread

K: lysine

kb: kilobase

LAP: legume anthocyanin production

Lb: leghemoglobin

LCO: lipo-chitooligosaccharides

LHK: Lotus histidine kinase

LRR: leucine rich repeat

MCS: multiple cloning site

MOA: mode of action

MS: mass spectrometry

N: nitrogen

NCR: nodule-specific cysteine-rich peptide

NF: nod factor

NFR: nod factor receptor

NFS: nitrogen fixation specificity

NF-Y: nuclear factor-Y

NGS: next generation sequencing

NH_4^+: ammonium

NIB: nuclear isolation buffer

NIN: nodule inception

NLP: NIN-like protein

NO_3-: nitrate

NSP: nodulation signaling pathway

PBS: phosphate buffered saline

PCR: polymerase chain reaction

PTM: post translation modification

R: arginine

RAM: required for Arbuscular Mycorrhization

ROS: reactive oxygen species

RT-PCR: reverse transcription PCR

S: swollen/spherical

SD: synthetic defined medium

SNARP: small nodulin acidic RNA-binding protein

SNF: symbiotic nitrogen fixation

SPC: signal peptidase complex

T3SS: type III secretion system

TEMED: Tetramethylethylenediamine

TF: transcription factor

U: unmodified

Y1H: yeast one hybrid

ZI: nodule zone I, meristem

ZIId: distal nodule zone II, invasion zone

ZIIp: proximal nodule zone II, invasion zone

IZ: nodule zone II-III, interzone

ZIII: nodule zone III, nitrogen-fixing zone

ZIV: nodule zone IV, senescence zone

Table of contents

Abstract

The *Rhizobium*-legume symbiosis is initiated via the primary signal exchanging process between the two partners which ensures that they recognize each other. As a result, host plants form root nodules to be infected by rhizobia that differentiate into the nitrogen fixing endosymbiotic form termed as bacteroids. Because the nitrogenase that reduces nitrogen to ammonium is very sensitive to oxygen, plants produce a large amount of the oxygen-binding leghemoglobin to restrict free-oxygen level and to protect the enzyme. Beside leghemoglobins, IRLC legumes like *Medicago* and *Pisum* also express NCR and GRP peptides to govern the differentiation of bacteroids. The model legume, *Medicago truncatula* expresses more than 700 *NCR* genes and many of the produced peptides have been proven to be targeted to the symbiosomes. The expression of *NCR* genes is extremely specific, that their transcription is restricted to the infected symbiotic cells of root nodules. Surprisingly, among these more than 700 *NCR* genes, it was found that the *NCR169* and *NCR211* genes are essential for the symbiotic nitrogen fixation despite the large number and expected redundant functions of the *NCR*s in *Medicago truncatula*. In this study, we developed two new reporter systems, which are suitable for the large-scale functional analysis and the identification of other essential *NCR*s. The anthocyanin reporter system can provide a directly visible signal for distinguishing transgenic roots from the untransformed ones. The NSP2 reporter system on the other hand, ensures that nodules can be formed only on the transgenic roots and the effect of the transgene can be immediately investigated. Besides these two reporter system, I also focused my efforts on the transcriptional regulation of *NCR* genes. I surprisingly found that the *NCR169* gene can be activated in soybean nodule where no *NCR* gene is expressed, meaning that soybean and *Medicago truncatula* share common transcription activator(s) of *NCR169* gene. To identify transcription factors contributing to *NCR169* gene expression, we combined DNA pull-down, Y1H screening and EMSA techniques applied to both *Medicago* and *Glycine* genes/proteins that resulted in the identification of several potential transcription regulators of *NCR169*.

1.Introduction

1.1 Major beneficial symbioses of plants to provide nutrients

Symbiosis (from Greek συμβίωσις "living together") can be defined the simplest term as a relationship between two types of organisms in which each provides for the other the conditions necessary for its continued existence. Being organism without the ability to move from one site to another, i.e. from an unfavorable, for example, nutrient poor place to an advantageous one, such interactions are almost indispensable for plants.

Plant growth and development rely on the accessibility of nutritional elements, such as carbon (C), hydrogen (H), oxygen (O), nitrogen (N), phosphorus (P), sulfur (S) from the environment. These elements are essential for the plant to synthesize carbohydrates, lipids, amino acids and nucleic acids, which are fundamental compounds that make up the plant itself and maintain its metabolism. Carbon and oxygen supply are not limited for plants, because they are able to convert CO_2 from the air into carbohydrates and to produce O_2 from water through photosynthesis, however, the acquirement of other elements, especially nitrogen, phosphorous and certain metals can be difficult. On the other hand, the situation for many rhizosphere microbes is the opposite, in that their diverse functions provide them easy access to N, P and S, but not to carbon sources. This complementarity in element acquisition provides the foundation for plant and rhizosphere microbes to establish symbiosis in which they can exchange certain nutrients with each other. There are many types of well-known plant symbioses but the most beneficial ones are the nitrogen-fixing and mycorrhizal symbioses.

Mycorrhizal fungi can develop association with most plant species and it is believed that this type of interactions made it possible for plants to colonize land. There are two major types of mycorrhizal symbioses: Ectomycorrhizal (EM) fungi colonize the root tissues of host plants extracellularly and form a mantle around the roots. In contrast, hyphae of endomycorrhizal mycorrhizal fungi penetrate the roots

intercellularly, and in the case of the most common arbuscular mycorrhiza (AM), fungi penetrate root cells. It is reported that 72% of vascular plants are arbuscular mycorrhizal (in which fungus from Glomeromycotina phylum form inter-/intracellular hyphal networks within the roots), the rest, ~15%, can get mycorrhiza in other forms (ectomycorrhizal, ericoid mycorrhizal and orchid mycorrhizal), leaving only 8% plants completely nonmycorrhizal (Brundrett & Tedersoo 2018). AM fungi can significantly contribute to the uptake of nutrients, increase plant biomass and improve resistance to stress and pathogens (Smith & Read 2008). Plants often suffer from phosphorus deficiency because of the low phosphate abundancy and solubility in the soil. Absorption of inorganic phosphate (Pi) by plants leads to the formation of depletion zones around the roots, which rapidly limits its further uptake and plant growth (Schachtman et al., 1998; Vance, 2001). Rhizosphere AM fungus can mobilize and accumulate Pi from the soil through the activity of extracellular enzymes and Pi transporters, and then transfer it to the host plant in the form of polyphosphate (Harrison et al. 1995; Hijikata et al. 2010). With their long hyphae growing outside roots, AMs provide plants access to Pi further away from the root system (Smith et al. 2011). AMs can also help plants with the uptake of nitrogen in a similar manner; NH_4^+, nitrate, urea and amino acid can first be acquired by AMs through specific transporters and then transferred to plants (Hodge et al. 2001; Tobar et al. 1994; Jin et al. 2012; Lopez-Pedrosa et al. 2006; Cappellazzp et al. 2008).

While the AM symbiosis helps plants to uptake elements that exist in the soil in limited quantities, the nitrogen fixing symbiosis links plants with a nearly infinite, but for most organisms unaccessible nitrogen source, the air, in which the content of nitrogen is 78%. Gram-negative bacteria from the α- and β-subgroups of the phylum *Proteobacteria*, collectively called rhizobia, establish symbiosis with legumes (family *Fabaceae*) and the non-legume *Parasponia* species (family *Cannabaceae*), while the Gram-positive filamentous bacteria of the *Frankia* genus from the *Actinobacteria* phylum associate with a broad spectrum of actinorhizal plants. Nitrogen-fixing microorganisms evolved an enzyme, nitrogenase that can break the extremely strong

triple bond in nitrogen (N_2) and reduce it to the form of ammonia, thus making the atmospheric nitrogen accessible for other organisms.

Combined nitrogen in the soil is very limited. Nitrate (NO_3^-) and ammonium (NH_4^+) ion quantities in the soil are not enough for maintaining plant growth, so nitrogen fertilizers are widely used in the agriculture to achieve higher crop yield. At the beginning of the 20th century, an industrial scale technology, the Haber-Bosch process was developed to produce nitrogen fertilizers. To break the very strong triple bond in nitrogen molecules, extreme conditions of high pressure and high temperature are applied, making this process costly and detrimental to the environment (due to the fossil energy used to generate the pressure and temperature). Application of synthetic fertilizers has indeed greatly increased world crop production feeding billions of people, but it also causes a lot of environmental problems. It is estimated that nearly half of the nitrogen fertilizer input cannot be absorbed by crops and is lost to the environment (Smil, 1999; Cassman et al., 2002; Tilman et al., 2002; Ciampitti & Vyn, 2014). The excessive use of nitrogen fertilizers reduces biodiversity, endangers the quality of drinking water, leads to the formation of marine algal blooms and contributes to air pollution as well as climate change, affecting the quality of life and the general health of the human population. Moreover, driven by population growth and the global shift towards a more protein-rich diet in developing countries, world demand for nitrogen fertilizer is projected to grow annually with 1.5% from 2015 to 2020, reaching 118.7 Tg/year in 2020 (FAO 2015; Lassaletta et al. 2016; Herrero et al. 2017). However, many evidence have shown that the total crop yield in many intensive farming systems has failed to improve in proportion to the application of nitrogen fertilizers (Shen et al. 2013; Ray et al. 2012), revealing that the increase of fertilizer use might not be a sustainable solution for the production of more food in the long term. These data underscore the challenges and potentials of increasing global food production while implementing new strategies to replace the application of nitrogen fertilizers in improving crop yield.

Unlike industrial nitrogen fixation, the biological process of symbiotic nitrogen

fixation (SNF) takes place under normal environmental (temperature and pressure) conditions using carbohydrates originated from photosynthesis as energy source. This process does not generate any pollution, which makes it quite suitable for developing sustainable agriculture in the future. No wonder that – with the ability to form nitrogen-fixing symbiosis – legumes that are the second most important crops after grasses have been important elements of cropping systems for a long time. Scientists have been working on the *Rhizobium*-legume symbiosis subject for decades in order to understand SNF and apply it to other crops like rice, wheat and corn to reduce the use of nitrogen fertilizers.

1.2 The *Rhizobium*-legume symbiosis

There are a number of requirements for the efficient use of atmospheric nitrogen via symbiotic nitrogen fixation by legume plants: The reduction of nitrogen gas (N_2) has high energy demand as the breakage of the triple bond and the formation of two ammonium (and one hidrogen) molecules are accompanied by the hydrolysis of 16 ATP molecules. Thus, bacteria must produce large amounts of ATP via respiratory processes and at the same time, the nitrogenase enzyme complex catalyzing the reduction of atmospheric nitrogen must be protected from oxygen to maintain its activity (Shah & Brill 1977). As ammonium (NH_4^+) is toxic for cells, it must be assimilated into organic molecules, mainly into amino acids, thus carbon skeletons should be provided in a sufficient amount (Mus et al. 2016; Lodwig et al. 2003). To cope with these demands, a specific niche – in the form of a novel organ called nodule (usually on the roots of the plants) – formed, in which, specialized forms of bacteria called bacteroids fix nitrogen. The nodule tissues provide and maintain a low free oxygen concentration via physical oxygen barriers and high amount of oxygen-binding proteins, leghemoglobins. To produce enough ATP in the microaerobic environment, bacteria use terminal oxidases with very high affinity for oxygen. To provide energy and carbon sources for nitrogen fixation and assimilation, the nodule serves as a sink organ to where sugars derived from photosynthesis are loaded and after their metabolism, dicarboxilic acids are transported to the bacteroids.

The assimilated nitrogen is transported from the nodule to plant in the form of amino acids or ureids.

1.2.1 Development of nitrogen fixing root nodules

The development of nodules involves a series of physiological and morphological changes and starts with the cross talk between the plant root and the specific rhizobium partner in the rhizosphere. Nitrogen starvation of legumes induces the secretion of aromatic compounds, flavonoids or isoflavonoids, to the rhizosphere that can be recognized by their rhizobium partners and activate the bacterial NodD transcription factors, which induce the expression of the rhizobium *nod* genes (Kondorosi et al. 1989; Gottfert et al. 1986). Expression of *nod* genes is necessary for the synthesis of lipochitooligosaccharides (LCOs) called Nod factors (NFs) that are the bacterial signals and are the primary determinants to control the compatibility with their legume hosts. Nod factors, even in the absence of rhizobia, are able to trigger the first physiological and morphological changes that can be observed upon rhizobial inoculation. This chemical communication between legume and rhizobium is essential for both sideslooking for its compatible partner in order to have successful infection and symbiosis (Oldroyd & Downie 2008; Denarie et al. 1996; Schultze & Kondorosi 1998). Rhizobial infection starts with the attachment of bacteria to the root hair, which changes the direction of its polar growth as a consequence of Nod factor perception and curls like a shepherd's crook around the dividing bacteria (Esseling et al. 2003). Before the visible morphological changes, a Ca^{2+}-ion influx and depolarization of the root hair cell membrane can be observed, which is followed by the oscillation of the intracellular Ca^{2+} concentration (Ca-spiking) after a few minutes. Curled root hair surrounds the microcolony formed by the offspring of the attached rhizobium in the so-called infection chamber (Verma et al. 1992). From the infection chamber, a tubular invagination of cell wall and plasma membrane leads to the development of the structure called “infection thread”, which extends by polar growth and delivers rhizobia to the deeper layers of the root cortex. Meanwhile, inner cortical cells

underlying the infected root hair will start dividing that leads to the formation of nodule primordia (Dudley et al. 1987). These newly divided cells will then be infected by rhizobia when infection threads reach them and release bacteria into the cytoplasm (Oldroyd & Downie 2008; Downie & Walker 1999). During their release, which is an endocytosis like process, rhizobia will get surrounded by a membrane of plant origin and an organelle-like structure known as symbiosome will be formed (Roth et al. 1988). Inside the symbiosome, rhizobia undergo a series of changes to adapt to the symbiotic condition and to be prepared for nitrogen fixation (Brewin 1991).

Parallel to the infection process, the nodule organogenesis takes place. After the perception of NFs, cells from the inner parts of the root start to divide first, to form the nodule primordium from which the nitrogen fixing nodule develops. It was shown (Xiao et al. 2014) that the nodule meristem in *Medicago* originates from the third cortical layer, while several cell layers of the base of the nodule are directly formed from cells of the inner cortical layers, root endodermis and pericycle. After cell divisions, submeristematic cells undergo drastic cellular modifications, involving cell enlargement and polyploidization of the genome through endoreduplication cycles, which are mainly controlled by the mitotic inhibitor Cell Cycle Switch protein (CCS52) (Cebolla et al. 1999). The formation of large polyploid symbiotic cells is essential for nodule organogenesis in all legumes that have been tested (Cebolla et al. 1999; Vinardell et al. 2003; Gonzalez-Sama et al. 2006). During the development and endoreduplication of cells, the released bacteria develop into bacteroids. Root nodules on different plant species differ from each other in their shape depending on plant species and can be grouped into two major types: determinate and indeterminate (Franssen et al. 1992; Maunoury et al. 2008). Meristematic cells in the determinate nodules on legumes like soybean and *Lotus japonicus* are not persistent, giving them a round shape with no zonation inside. Symbiotic cells differentiate synchronously to nitrogen fixing cells with a radial gradient of development and later deteriorate similarly, senescence beginning at the center (marked as “S” in Figure 1.1) and spreading outwards (Szczyglowski et al. 1998). The indeterminate nodules on

legumes like *Medicago* have an elongated or cylindrical shape and nodule can be divided into different developmental zones as shown below because of the presence of a persistent meristem region (Sutton 1983). Active meristem at the nodule apex provide a continuous source for newly divided uninfected nodule cells, which will get infected afterwards and enter the nodule differentiation program, leading to layers of cells with different extents of differentiation and infection. In a mature nodule, five distinct zones can be observed: the meristem (ZI), the invasion zone (ZII), the interzone (IZ), the nitrogen-fixing zone (ZIII), and in older nodules, the senescence zone (ZIV) (Fig 1.1, right).

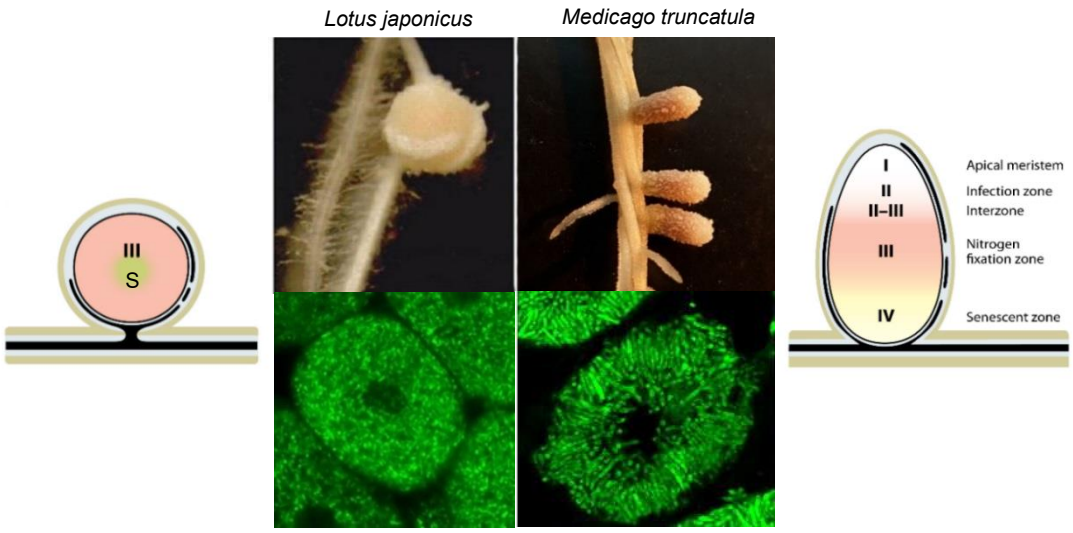


Figure 1.1. A comparison between the nodules and bacteroids from *L. japonicus* and *M. truncatula*. Determinate nodules on *Lotus* roots are round-shaped and bacteroids inside display a morphology similar to free-living rhizobia (***left***). On *Medicago*, nodules have an elongated shapeand are composed of different zones harboring cells at different developmental stages and with different extents of infection. Bacteroids in *Medicago* display elongated morphology (***right***). Modified from Kondorosi et al. 2013)

During the development of symbiosis, the rhizobial symbionts also undergo multiple cytological and morphological changes to adapt to the *in planta* life-style as bacteroids. The extent of morphological changes mainly depends on the host plants. Bacteroids from *L. japonicus* and soybean are usually termed as U-morphotype since they are unmodified in their shape compared to free-living rhizobia. U type bacteroids are able to revert to free-living state when released from nodule cells, meaning that

experimentally they are culturable to form colonies on plate. In other cases, bacteroids are terminally differentiated as they lose their reproductive capacity and they cannot return to their free-living stage. These bacteroids have modified morphology and are classified as S-morphotype bacteroids which are swollen/spherical and E-morphotype with elongated, sometimes branched "Y" shape. S type bacteroids can be found in legumes like *Ononis* and *Cicer* species in the IRLC (Inverted Repeat Lacking Clade) legumes as well as in *Aeschynomene* and *Arachis* species in the *Dalbergoid* clade, and E type bacteroids are quite common in *Medicago*, *Pisum*, and *Trifolium* species also in the IRLC (Mergaert et al. 2006; Montiel et al. 2016; Czernic et al. 2015; Guefrachi et al. 2014; Oono et al. 2010; Montiel et al. 2017).

In addition to the difference in morphology, these terminally differentiated bacteroids also carry altered cytological features, including an amplified genome as result of endoreduplication and modifications in their membranes that might contribute to their unculturable nature. The DNA content can range from 4 to 32 folds higher in differentiated bacteroids, depending on the host plants and their rhizobial partners (Mergaert et al. 2006; Montiel et al. 2016; Czernic et al. 2015; Guefrachi et al. 2014). The increase of bacteroid membrane permeability was revealed in an *in vitro* test by showing that propidium iodide, which is normally excluded from living cells, slowly penetrates the cytosol. The membrane permeability change might facilitate exchange of metabolites between the partners (Mergaert et al 2006).

Since symbiotic nitrogen fixation consumes extremely large amounts of energy, i.e. the production of each gram of fixed nitrogen costs more than ten grams of carbohydrates, this process is tightly controlled by the plant to balance nitrogen acquisition and plant development (Gibson 1966; Silsbury 1977; Mahon 1977). Legumes have evolved mechanisms inhibiting nodule formation and function when enough ammonia or nitrate is available from the nodules or the rhizosphere, when they need to channel energy to the reproductive organs during flowering, when the plants encounter challenges of harmful conditions, such as unfavorable light intensity, mineral nutrition starvation, low temperature, salinity condition and drought, where

attainment of nitrogen is no longer a priority (Hernandez et al. 2009; Zhang et al. 2014; Soussi et al. 1999; Serraj et al. 1999; Rice et al. 1977; Lie 1974). Development of nodules triggers a feedback regulatory circuit called autoregulation of nodulation (AON), which involves both root to shoot and shoot to root signaling and prevents nodule initiation if the development of enough nodules is in progress. Nitrate strongly inhibits symbiotic nitrogen fixation. Application of nitrate to a plant leads to a dramatical decrease in nodule number, nodule mass, nitrogen fixing activity and accelerated nodule senescence, and this process was shown to be closely related to AON (Carroll & Gresshoff 1983; Carroll et al. 1985; Streeter, 1988). It has been proven that nitrate and nodule development can induce the production of CLE (CLAVATA3/EMBRYO SURROUNDING REGION-RELATED) peptides in the roots/nodule primordia that can interact with their receptor HAR1/NARK/SUNN in the shoot, which mediates the nodulation inhibition on root (Okamoto et al 2008; Reid et al. 2011; Okamoto & Kawaguchi 2015).

1.2.2 Genetic determinants of symbiosis

At molecular level, Nod Factor Receptors (NFR), members of the Lysine Motif (LysM) domain receptor kinase family, are responsible for the perception of NFs (Radutoiu et al. 2003; Esseling et al. 2003). The two receptors, MtNFP/LjNFR5 without kinase activity and MtLYK3/LjNFR1 with a functional kinase domain form a complex and can bind the specific NFs with their extracellular domain (Broghammer et al., 2012). While these two receptors are required only for symbiotic nodule initiation and development, a third receptor-like kinase (SYMRK/NORK/DMI2), with several other elements in the signal transduction cascade, participates not only in nodulation, but also in mycorrhization. The existence of this so-called common symbiosis signaling pathway (CSSP) indicates that rhizobia hijacked the infection mechanism developed during/for the more ancient mycorrhiza-plant interactions. A lectin-like apyrase, LjNLP is required for both the initial calcium influx and the following Ca-spiking around the nucleus Ca (Roberts et al., 2003). A number of

proteins in the nuclear and the attached endoplasmatic reticulum (ER) membranes, such as the nucleoporins LjNENA (Groth et al., 2010), LjNUP133 (Kanamori et al., 2006) and LjNUP85 (Saito et al., 2007), the potassium channels DMI1 in *Medicago* (Ané et al., 2004), CASTOR and POLLUX in *Lotus* (Charpentier et al., 2008), the CNGC15 calcium channel (Charpentier et al., 2016) and the MCA8 calcium pump are required for Ca-spiking. The calcium-oscillations are deciphered by the DMI3 calcium/calmodulin dependent serine/threonine protein kinase, which interacts with the MtIPD3/LjCYCLOPS transcription factor. The transduction of this signal leads to the activation of transcriptional regulatory complexes in a hierarchical way. Several transcription factors such as NSP1, NSP2, NIN, ERN1 and ERN2, NF-YA1 and NF-YB1 that control the expression of early symbiotic genes have been identified (Singh & Parniske 2012; Singh et al. 2014; Parniske et al. 2008; Schauser et al. 1999; Kaló et al, 2005; Smit et al. 2005; Marsh et al. 2007; Middleton et al. 2007; Soyano et al. 2013; Combier et al. 2006). Cytokinin biosynthesis is induced in the cortex downstream of DMI3 and its local accumulation synchronizes cortical cell division with the progress of rhizobial infection (Frugier et al. 2008; van Zeijl et al. 2015; Jardinaud et al. 2016; Reid et al. 2017). Cytokinin and gibberelic acid through their effectors, RR1/4 and DELLA-type transcription factors, respectively, contribute to Nod factor induced transcriptional changes (Gonzalez-Rizzo et al. 2006; Jin et al. 2016). A high number of genes – called nodulins – are essential for or accompanying the infection and later, the invasion processes or the development and functioning of bacteroids and the assimilation of the fixed nitrogen is activated in the developing nodule. The best known and most abundant nodulins are the leghemoglobins (Lbs) (Appleby 1984; Stougaard et al. 1987). Lbs are essential for the maintainance of the low and constant free oxygen concentration that still cause no damage to the oxygen-sensitive nitrogenase and allow the functioning of the high affinity terminal oxidase of bacteria to support the production of enough ATP required for nitrogen reduction (Wittenberg et al. 1974; Ott et al. 2005).

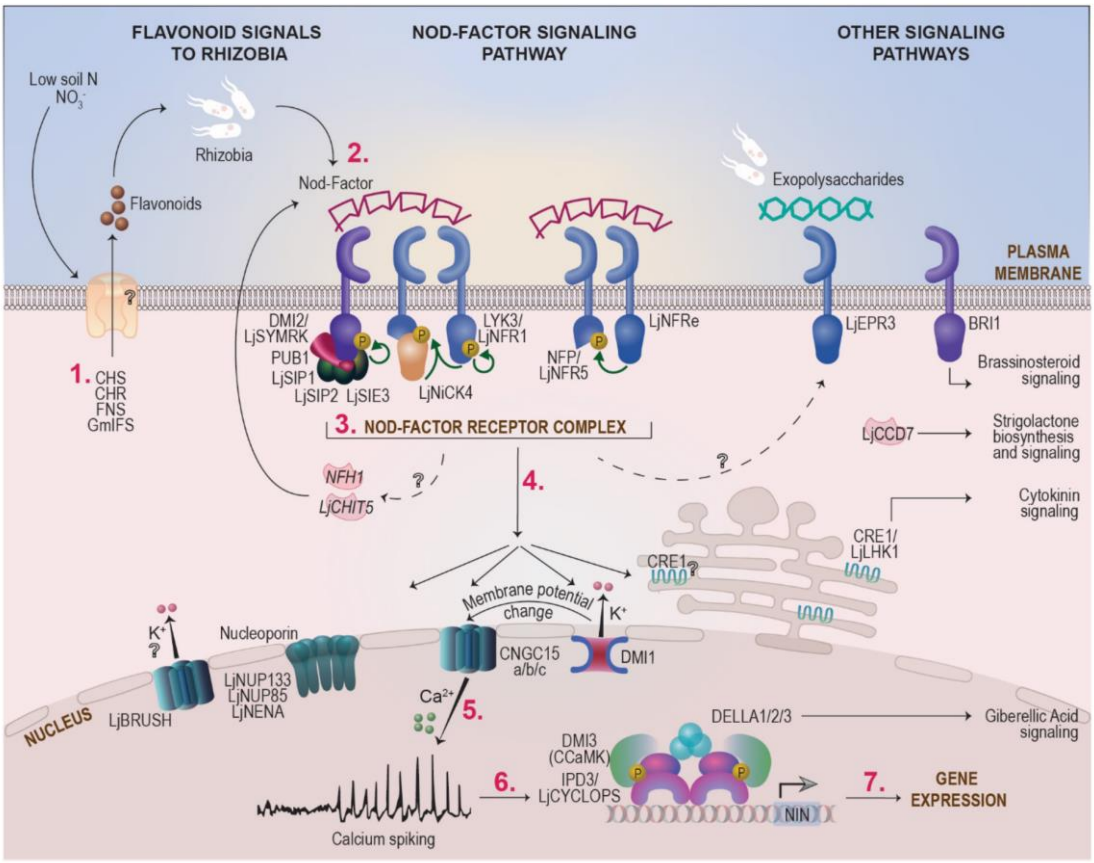


Figure 1.2. Genes and processes involved in early signaling during nodulation. (Iso)Flavonoids produced under low soil N (1) trigger the production of bacterial Nod factors (2) that, together with other signals, are perceived by receptors at the plasma membrane of epidermal cells (3). This triggers biochemical and physiological responses (4,5,6) that lead to changes in nuclear gene expression (7). Symbiotic gene regulation acts in a NIN centered manner, that – together with ERN1, ERN2 and maybe some other unknown TFs – controls the massive expression network of downstream genes, including NF-YA2, IPD3L, ENOD40 and other symbiotic genes. *M. truncatula* protein names are provided unless otherwise specified. From Sonali Roy et al. 2019.

1.3 Nodule-specific Cysteing-Rich (NCR) peptides

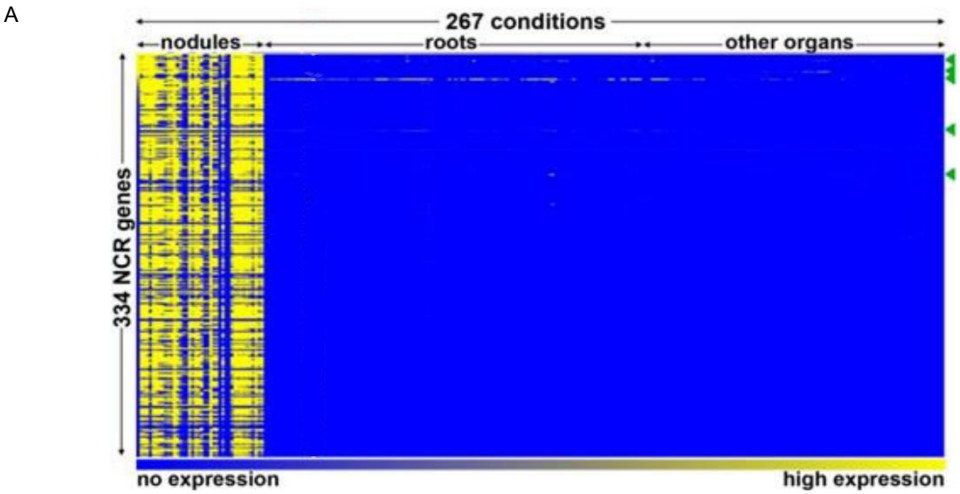
1.3.1 Plant effectors directing the irreversible terminal differentiation of bacteroids

Neither the nodule type, i.e. determinate or indeterminate, nor the rhizobium species can fully reflect the morphotype of bacteroids, because bacteroids in the indeterminate nodules of *Medicago* and *Leucaena* are E and U morphotypes,

respectively (Ishihara et al., 2011) while in some cases, such as *Bradyrhizobium* sp. 32H1, the same rhizobium strain can behave distinctly when inoculated to different host plants (Sen & Weaver 1980; Bonaldi et al. 2011). With the use of nearly isogenic rhizobial strains on different hosts, Mergaert et al. (2006) proved that the differentiation of bacteroids is under the control of the plant host. Recent studies – including comparative transcriptomic studies between species with E and U morphotype bacteroids – have provided more and more evidences that Nodule Specific Cysteine Rich peptides (NCRs), and probably glycine-rich peptides (GRPs) and small nodulin acidic RNA-binding protein (SNARP, or LEED..PEED) as well, all expressed in symbiotic cells of the IRLC and Dalbergioid legumes play a very important role in this process and are the plant factors that direct terminal bacteroid differentiation (Mergaert et al. 2003; Mergaert et al. 2006; Alunni et al. 2007; Kereszt et al. 2011; Velde et al. 2010; Laporte et al. 2010; Trujillo et al. 2014; Kondorosi et al. 2013, Montiel et al. 2016, Montiel et al. 2017, Czernic et al. 2015).

In *Medicago truncatula*, more than 700 *NCR* genes were identified while no such sequence could be found in the genome of *Lotus* and *Glycine* forming nodules containing U morphotype bacteroids (Mergaert et al. 2003). Usually, the *NCR* genes are composed of two exons with coding sequences, however, in a number of cases, there is a second intron after the translational stop codon and a third exon with the 3' untranslated region. The more than 700 genes in *Medicago truncatula* are spread on all eight chromosomes. Such high number of genes in a gene family suggests functional redundancy, but there are examples that the loss of a single *NCR* like *NCR169* and *NCR211* in the *dnf7* and in the *dnf4* mutants, respectively, results in the arrest of symbiotic nodule development (Horvath et al. 2015; Kim et al. 2015).

Analysis of expression data from 267 different experimental conditions (Figure 1.3.A) assembled in the Medicago Gene Expression Atlas (Vagner B et al. 2008; Ji H et al. 2009) revealed that *NCRs* (with only a few exceptions, like *NCR122* and *NCR218*) are expressed and only expressed in nodule cells, not in other plant tissues, and moreover, this expression is not inducible by treatment conditions other than the -



cluster	%FI	%FIId	%FIIp	%IZ	%ZIII
1	81	13	3	2	1
2	57	24	8	6	5
4	41	33	13	5	7
5	31	25	19	12	13
3	16	48	22	7	7
11	13	17	29	28	14
10	1	5	61	28	4
9	1	1	10	68	20
12	6	7	15	37	35
6	1	2	4	41	52
7	3	3	3	12	79
8	10	11	11	15	53
13	15	15	16	21	33

Figure 1.3. Expression pattern of *NCR* genes. (A) Majority of the 334 *NCR* genes investigated with Affymetrix chip is only expressed in nodules samples among the 267 treatment conditions assembled in the Medicago Gene expression Atlas (Guefrachi et al. 2014). (B) Localization of the *NCR001* (right) and *NCR084* transcripts (right) by *in situ* hybridization using antisense probes (Mergaert et al. 2003). (C) Gene expression categories based on the distribution of RNA-Seq reads in different nodule zones sample after laser-captured microdissection. *NCR*s mainly belong to the clusters 6, 9 and 12 (Roux et al. 2014).

-inoculation of rhizobia (Guefrachi et al. 2014). Localization of the mRNA (by in situ hybridization and reporter genes, Fig 1.4 B) as well as the peptides (by immunological methods) produced from the *NCR* genes showed that the genes are active only in the infected cells of the nodules (Van de Velde et al., 2010), thus, further strengthening their putative role in directing bacteroid differentiation. This investigation, as well as new RNA-seq studies using laser-captured microdissection samples (Roux et al. 2014) demonstrated that expression of *NCR*s in nodules is not

even in the same zones, they are transcribed in different waves. Early *NCR*s are induced in the proximal part of the invasion zone (ZII) and in the interzone (IZ) and their expression drops in the nitrogen fixation zone (ZIII), while other *NCR* genes are switched on only in Zone III. In general, most of the *NCR*s showed high expression in the inter-zone and zone III, while low or no expression could be observed in zone II and zone I (meristem), respectively.

So far, little is known about the strict transcriptional regulation of *NCR* genes. Recently, it was shown that the transcriptional waves of nodule-specific (including *NCR*) gene expression observed correlate with growing ploidy levels of the developing nodule cells (Nagymihály et al. 2017). Differential DNA methylation was found in only a small subset of symbiotic nodule-specific genes, including more than half of the *NCR* genes, whereas for most of the genes, DNA methylation was unaffected by the ploidy levels and was independent of the active or repressed state of genes. On the other hand, expression of the nodule-specific genes correlated with ploidy-dependent opening of the chromatin as well as, in a subset of tested genes, with reduced tri-methylation levels of histon H3 on lysin 27 (H3K27me3) combined with enhanced histon H3 lysin 9 acetylation (H3K9ac) levels. There is no knowledge available on the transcription factors contributing to the induction/repression of the *NCR* genes (Nagymihály et al. 2017).

The translated NCR polypeptides contain a secretory signal peptide sequence, which will be cleaved to produce mature peptides of 24-50 amino acids in length, with 4 or 6 cysteines in conserved positions (Figure 1.4), that would form 2 or 3 disulfide bonds to ensure and stabilize the active structure that might be critical for their proper activity. These features make NCRs quite like defensins, while differ in the number of cysteine residues and length of mature peptides (Fig 1.4) (Maróti et al. 2015; Kereszt et al. 2011; Kondorosi et al. 2013).

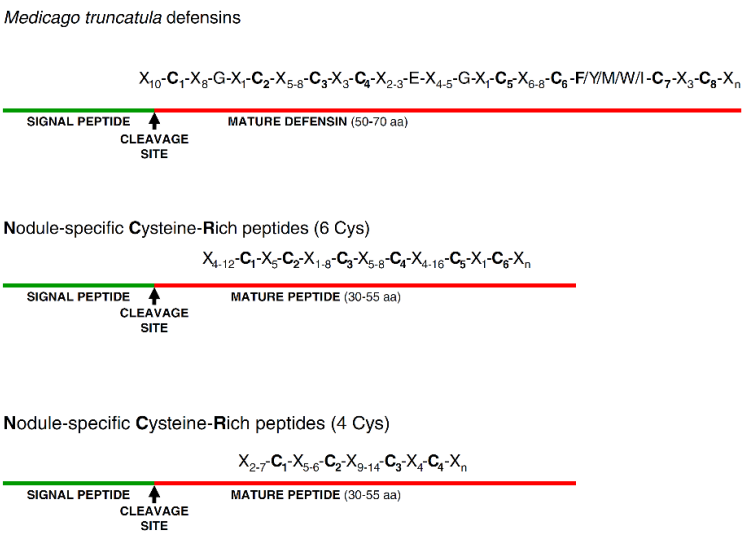


Figure 1.4. Comparison of the backbones of NCRs and defensin from *Medicago*. NCRs are generally shorter (30- 55 amino acids) than defensins (50-70 amino acids), but these two types of peptides share common features such as containing a signal peptide and conserved cysteine residues. Conserved cysteines in defensins and in the two major types of NCR peptides (containing 4 and 6 cysteines, respectively) are demonstrated in bold and numbered, the length of conserved spacings between cysteines are indicated (Xn). From Maróti et al. 2015.

As the majority of the infected nodule cells' membrane system are around rhizobia, NCRs are targeted by the protein secretion system to the symbiosomes. Immunological techniques (Van de Velde et al., 2010) and proteomic investigations (Dürgő et al. 2015, Marx et al. 2016) showed that the peptides are indeed targeted to the bacteroids. Impairment of a nodule-specific element of the secretory machinery (signal peptidase complex (SPC)) in the *dnf1* mutant (Wang et al., 2010) results in NCRs stacked in the ER with their signal peptides not removed and in bacteroid development arrested just after the release from the ITs (Van de Velde et al., 2010). The importance of the cystein residues and disulphide bridges was shown for the NCR169 peptide in complementation experiments, where mutant versions with one and any cystein changed to alanine could not complement the *dnf7* mutants (Horvath et al., 2015). Similarly, research on the role of plant thioredoxins in symbiosis confirmed the importance of the redox state of NCR peptides (Ribeiro et al., 2017). Thioredoxin s1 (Trx s1) production was shown to be induced in the infection zone with the protein targeted to the symbiosomes. Both silencing and overexpression of

Trx s1 impairs bacteroid growth and endoreduplication, two features of terminal bacteroid differentiation, highlighting the importance of the balance between reduced and oxidized states of NCRs in the bacteroids. Despite the conserved signal peptide and the cysteines in conserved position, the peptides are very diverse in amino acid composition and sequence that are mirrored in their charge, the isolecetric point (pI) ranging from 4 to 10.5. This high variability indicates that NCRs might function under different cellular conditions (Maróti et al. 2015) and might be involved in several steps during bacteroid development.

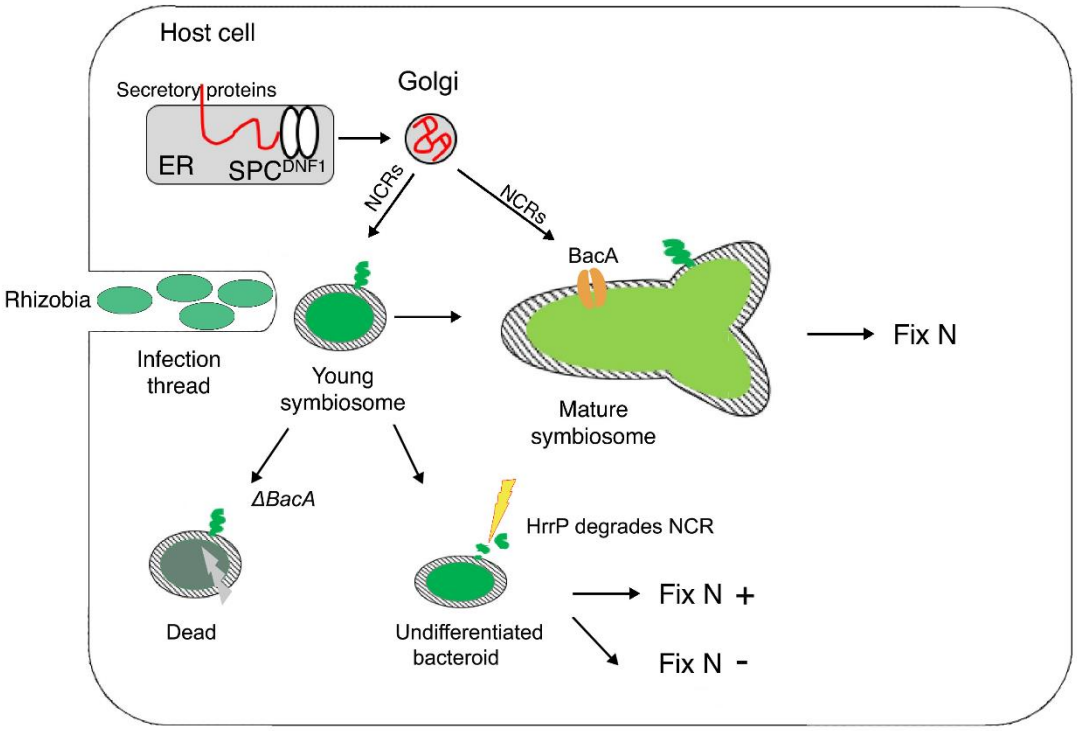


Figure 1.5. Secretion of NCR peptides and their targeting to symbiosomes. NCRs are secreted from ER with the help of the SPC that cleaves signal peptides, and mature peptides are targeted to the symbiosomes. The BacA protein was shown to protect rhizobia from the lethal effect of NCRs. A *ΔBacA* mutant cannot survive the NCRs' challenge after getting released from infection thread to the cytosol. HrrP (host range restriction protein) protein carries peptidase activity and might help rhizobia to antagonize NCRs by degrading them. The effects of HrrP on symbiosis depend on the genetic background of both host and symbiont, manifesting as a host range restriction phenotype. Adapted from Stonoha-Arther and Wang 2018 and Haag et al. 2011, refering to also Price et al. 2015.

1.3.2 Functions of NCR peptides

Despite a lot of work on NCR peptides been done, little is known about their functions and mode of actions in symbiosis. Based on their similarity to defensins, the effector molecules of innate immunity, their antimicrobial activity was investigated. Cationic NCRs were characterized by their broad range of bactericidal and fungicidal activities, being toxic to various Gram-negative and Gram-positive bacteria and both yeast-like and filamentous fungi (Maróti et al. 2011; Maróti et al. 2015; Ördögh et al. 2014; Tiricz et al. 2013). They were shown to interact with the membranes and to disturb the integrity of the fungal and bacterial membranes and thus destroying the membrane potential (Ördögh et al. 2014; Nagy et al. 2015; Mikuláss et al. 2016). It is not known whether the NCRs' antibacterial activity contributes to bacteroid development, however, the bacterial peptide transporter BacA protein is essential for bacteroid development in those legumes that produce NCRs. The *Sinorhizobium meliloti* BacA protein's transporter activity was shown to protect rhizobia from the lethal effect of NCRs in both symbiotic and free-living conditions. The BacA-deficient mutant strains exposed to the antimicrobial activity of NCRs after getting released from the infenction thread to the cytosol are killed, which arrests further nodule developmment (Haag et al. 2011). These observations indicate that rhizobia released into the NCR-loaded nodule cells are balanced between death and terminal differentiation inflicted on them by the host plant. Recent work suggested that the BacA homologue from *Bradyrhizobium* spp, BclA (*Bradyrhizobium* BacA-like) is required for the survival and differentiation in the symbiosis with *Aeschynomene* legumes (Guefrachi et al. 2015). A hypothesis for the role of BacA(-like) proteins is that they import (for example, for degradation) the NCR peptides in order to remove them from the bacterial membrane, thus, preventing their membrane damaging activity (Pan and Wang 2017; Haag et al, 2011; Marlow et al. 2009; Guefrachi et al. 2015).

Most of the investigations on NCR biology have been performed with NCR247, which is the smallest member of the NCR family, with its mature peptide composed

of only 24 amino acids. The *NCR247* gene is expressed in the older cell layers of nodule zone II, where the endosymbionts stop dividing and cell elongation begins, and in the interzone II-III, where dramatic growth of endosymbionts occurs. Transcriptome, peptide-protein interaction and physiological investigations (Tiricz et. al, 2013; Penterman et al. 2014; Farkas et al. 2014) revealed that NCR247 affects multiple biological processes in bacteria. The peptide penetrates the membranes of both free-living cells and bacteroids and accumulates in the cytosol (Farkas et al. 2014). In synchronized cell cultures, it could be shown that sublethal concentration of the peptide inhibited cell division but not DNA replication that are the features of endoredupliation taking place in the nodule infection and interzones where NCR247 is produced. On transcriptional level, NCR247 treatments of exponentially growing *S. meliloti* cells resulted in a quick down-regulation of genes involved in basic cellular functions, such as transcription-translation and energy production, as well as upregulation of genes involved in stress and oxidative stress responses and membrane transport (Tiricz et al. 2013). Similar changes provoked mainly in Gram-positive bacteria by antimicrobial agents were coupled with the destruction of membrane potential that could be observed during NCR247 treatment of *S. meliloti* (Mikuláss et al. 2016). In synchronized cell cultures, low concentration of NCR247 down-regulated the expression of cell-cycle-regulated genes that could contribute to its cell division inhibiting ability (Penterman et al. 2014). The expression of many hypothetical genes and genes involved in nitrate and nitric oxide metabolism was also attenuated. In contrast, a subset of genes regulated by global stress responses through the FecI and RpoH1 sigma factors as well as 153 genes in three regulons of ExoS-ChvI, RirA, and FeuP-FeuQ affecting the exopolysaccharide, cyclic glucan production and iron acquisition, respectively, could be induced by NCR247. The NCR247 peptide binds to a number of rhizobial proteins. Its interaction with the FtsZ protein involved in cell division by marking the position of septum formation via forming the so-called Z-ring was shown by pull-down experiments (Farkas et al. 2014). This interaction prevented the formation of the Z-ring that could be visualized

by both tagging FtsZ with GFP (Penterman et al., 2014) and by its interaction with another, fluorescently labeled peptide, FITC-NCR035 (Farkas et al. 2014), i.e. no fluorescent signal could be detected at midcell upon NCR247 treatment. Thus, NCR247 possesses another way to inhibit cell division and induce endoreduplication. NCR247 was shown to bind to different subunits of the ribosome, and then its ability to inhibit translation was demonstrated both *in vitro* and *in vivo* (Farkas et al. 2014) suggesting that NCRs might influence the bacteroid proteome and contribute to the altered pattern and reduced complexity of bacteroid proteins. The chaperon GroEL protein was also proven to be an interacting partner of NCR247, however, the relevance of this interaction is not known: it is possible that GroEL might be necessary for correct folding of the NCR peptides, or NCR247 by its interaction with GroEL affects GroEL-dependent functions (Farkas et al. 2014).

At present, two peptides, NCR169 and NCR211 are known to be essential for symbiotic nitrogen fixation in *Medicago truncatula* (Horvath et al. 2015; Kim et al. 2015). Deletion of *NCR169* in mutant *dnf7* and *NCR211* in *dnf4* causes impaired symbiotic nodule development. Microscopic analysis demonstrated that in both mutants, bacteroid differentiation and the expression of bacterial *nifH* gene coding for a subunit of the nitrogenase enzyme complex was initiated, however, the rhizobium cells lost their viability later and disappeared from the nodules. Thus, both peptides are required for the differentiation and persistence of the bacteroids during nodule development. Expression analysis demonstrated that *NCR211* possesses a primary expression pattern in the infection zone and interzone, while *NCR169* is expressed in the inter zone and nitrogen fixing zone, zone III. These data may indicate that NCR211 has a function earlier than NCR169 in the bacteroid differentiation process.

Figure 1.7. Fix- phenotype of *Medicago truncatula* mutants *dnf-7* and *dnf-4*. Mutation in the *NCR169* and *NCR211* genes lead to the formation of functionally defective nodules. Mutant plants demonstrate nitrogen starvation symptom manifested in having yellowish leaves and retarded growth. From Horvath et al. 2015 and Kim et al. 2015

Considering all the information available on the NCR peptides, it appears quite contradictory that NCRs that have been proved to possess antimicrobial activity, or potentially have such activity, turned out to be essential for the development and/or the survival of bacteroids in symbiosis, i.e. these peptides exert negative effect via their antimicrobial activity *in vitro* but have positive effect via their differentiation inducing activity *in planta*. A possible explanation for this dual "behavior" is that the actual peptide concentration in the symbiosomes is lower than the applied one *in vitro* and/or the peptides act as cocktails containg both cationic ones with antimicrobial activity and neutral/anioinic ones that might counteract the effects of the antimicrobial peptides.

Recently some new findings broadened our knowledge on how diverse the function of NCRs can be. Yang et al (2017) and Wang et al (2017; 2018) reported that the *NFS1* and *NFS2* genes both encoding NCR peptides control the compatibility between rhizobium strains (*S. meliloti* RM41 and A145) and *Medicago truncatula* ecotypes. Plants carrying the *NFS* alleles of Jemalong A17 were incompatible, i.e. Fix^- with the

two *S. meliloti* strains while lines carrying the *NFS* alleles of DZA315.16 or A20 formed normal, Fix^+ symbiosis with these rhizobia. In the incompatible interaction, bacteroid development and *nifH* gene expression were initiated but bacteroids were eliminated from the nodule tissues. These *NCR* genes were not essential for symbiosis because if mutations were induced in the genes, the mutant plants formed effective, Fix^+ symbiosis with the rhizobia. The Jemalong alleles were dominant over the DZA315.16 and A20 alleles because the heterozygous plants were also incompatible with the Rm41 and A145 strains. These results revealed that beside the Nod factor recognition mechanism, plants also developed other selection method(s) to ensure that an optimal partner invades the nodule. NCRs might contribute to this selection, while the detailed mechanism is still not known.

1.3.3 NCRs in other legumes

Expression of *NCR*s is not restricted to *Medicago truncatula*, they were identified also in other species of the IRLC, such as *Vicia faba*, *Galega orientalis*, *Glycyrrhiza uralensis*, *Astragalus canadensis*, *Ononis spinosa*, *Onobrychis viciifolia*, *Medicago sativa*, *Trifolium repens* and *Pisum sativum* (Chou et al. 2006; Montiel et al. 2016; Montiel et al. 2017; Crockard et al. 2002; Fruhling et al. 2000, Jimenez-Zurdo et al. 2000, Kaijalainen et al. 2002; Kato et al. 2002; Mergaert et al. 2003). More focused analyses (Montiel et al., 2016; 2017) showed that the numbers of *NCR* genes are highly variable and expanded independently in different lineages of IRLC legumes ranging from 7 in *G. uralensis* to over 700 in *M. truncatula*. Phylogenetic investigations revealed that each of the seven *G. uralensis* NCRs had at least one highly similar homolog in another legume, whereas no NCR with such high similarity was found in the other five Vicioid legume species investigated. These results suggested common origin and conservation of a few NCRs, coupled with the emergence of many species-specific peptides. The investigations showed that bacteroids in all tested IRLC legumes were larger and had more DNA than cultured cells, but the degree of cell elongation was rather variable in different species. There

was shown a positive correlation between the degree of bacteroid elongation and the number of the expressed *NCR*s.

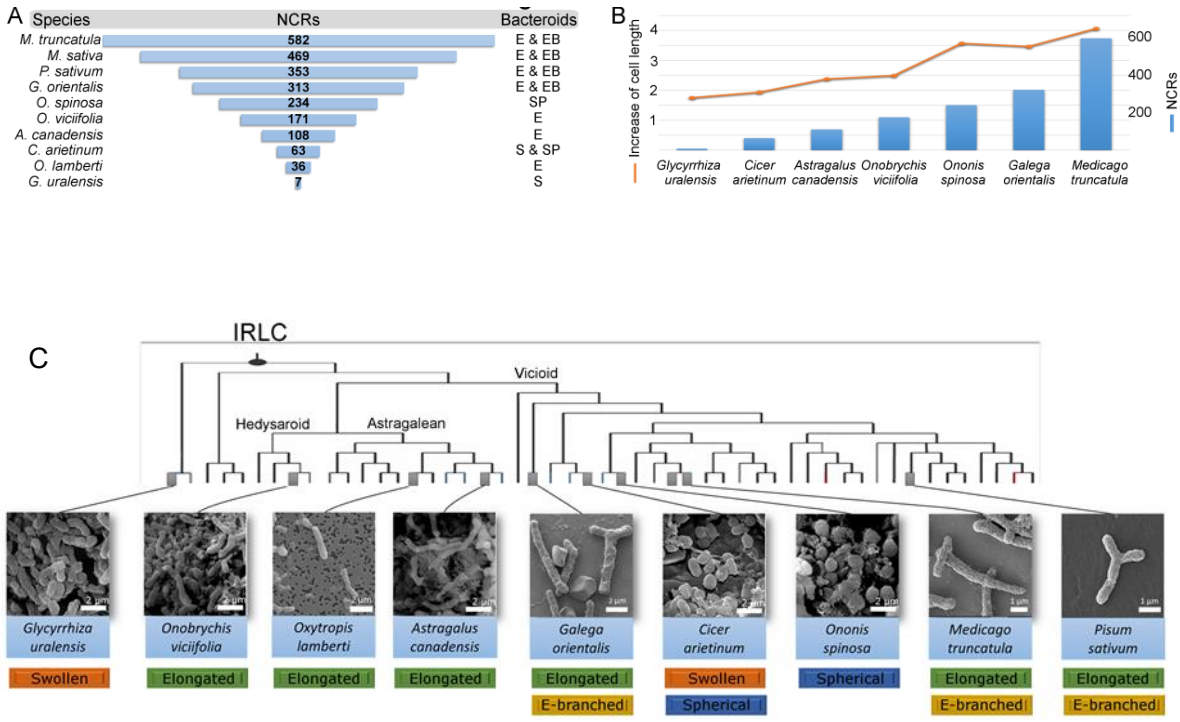


Figure 1.8. Number of NCR peptides in different IRLC legumes correlates with bacteroid morphology. Numbers of NCR peptides predicted from nodule transcriptomes or genome sequences of 10 IRLC legumes are shown in relation to the morphotype of the bacteroids (A). There is a positive correlation between average bacteroid length and the size of the NCR family (B and C). Pearson correlation coefficient: 0.90 (P value > 0.001). From Montiel et al. 2017.

Terminal bacteroid differentiation is also not restricted to the IRLC legumes: Phylogenetic analysis and ancestral state reconstruction suggested that the ancestral morphotype of bacteroids was the undifferentiated U-morphotype and that the differentiated E- and S-morphotypes appeared several times independently in the legume family (Oono et al., 2010). For example, in the Dalbergioid legume clade, bacteroids can be of the E- or S-morphotype: Within the *Aeschynomene* genus, *A. afraspera* has E-type bacteroids but *A. indica* or *A. evenia* have S-type bacteroids (Bonaldi et al., 2011; Arrighi et al., 2012). Transcriptome, *in situ* hybridization, and proteome analyses demonstrated that the symbiotic cells in the *Aeschynomene* spp.

nodules also produce a large diversity of NCR-like peptides, which are transported to the bacteroids (Czernic et al., 2015). These results support the view that bacteroid differentiation in the Dalbergioid clade, which likely evolved independently from the bacteroid differentiation in the IRLC clade, is based on very similar mechanisms used by IRLC legumes.

1.4 Genes considered for developing reporter

Hairy root transformation of legumes (Stiller et al. 1997; Boisson-Dernier et al. 2001; Kereszt et al. 2007; Estrada-Navarrete et al. 2007; Clemow et al. 2011) has been widely used in symbiosis research, for example, to complement symbiotic mutants (Endre et al. 2002; Madsen et al. 2003; Indrasumunar et al. 2010; 2011), to reveal the spatial and temporal aspects of gene expression and to identify promoter elements (Bersoult et al. 2005; Gavrilovic et al. 2016; Liu et al. 2019), to determine the cellular localization of proteins (Limpens et al. 2009; Gavrin et al. 2014; 2017), to overexpress (Indrasumunar et al. 2011; Reid et al. 2011), to silence (Limpens et al. 2005; Sinharoy et al. 2015; Sogawa et al. 2019) or to knock-out genes (Michno et al. 2015; Wang et al. 2017; 2018). Despite its many advantages, the efficiency of hairy root transformation is not 100% even with antibiotic or herbicide selection, i.e not all the roots formed on transformed plants are transgenic. The identification of transgenic tissues via the detection of the proteins produced by the currently used reporter genes coding for β-glucuronidase/GUS in vectors such as pBI121 or the pCAMBIA series (Jefferson et al. 1986; 1987;) and fluorescent proteins (GFP, YFP, DsRed, ...) in plasmids like the pUB series or the pHairyRed (Maekawa et al. 2008; Lin et al. 2011) requires destructive techniques and fluorescent microscopes, respectively, making impossible or uncomfortable the screening if further tissue growth is needed subsequently. Moreover, the detection of signals in these reporter systems is either time- or labor-consuming, which will be troublesome when large scale gene function analysis needs to be conducted, for example, to identify other essential *NCR* genes via genome engineering. Therefore, easy, non-destructive methods for the identification of transgenic roots are highly needed.

Anthocyanins are secondary metabolites, which - if accumulated in tissues - provide red, blue, or purple coloration. This feature has been applied for the color modification of fruits and flowers to achieve new traits for higher economic benefit. Biosynthesis of anthocyanin uses phenylalanine as a precursor and involves catalysis by several enzymes of the general phenylpropanoid and flavonoid pathways (Holton A, and Cornish C, 1995; Shi Z et al. 2014). Recently, it was shown that transient (Picard et al. 2013) or ectopic (Peel et al. 2009) expression of *MtLAP1* - coding for a member of the largest and plant/yeast specific R2R3 subfamily of the MYB transcription factors - results in anthocyanin production and accumulation in *Medicago truncatula*, *M. sativa* and *Trifolium repens* as well as with a lower intensity and homogeneity in tobacco. As the accumulation of anthocyanins of the transgenic plants resulted in purple colored plants, it could be assumed that using the *MtLAP1* gene as reporter may facilitate the selection of transgenic roots by naked eyes.

Adventitious root development can be initiated and leads to the formation of un-tranformed roots at the same time when transgenic hairy root development starts. And these adventitious roots can form nodules as efficinently as the hairy roots do. In those experiments that the activity of the introduced gene(s) can be obviously revealled, for example, via GUS staining or by restoration of normal symbiotic nodule development in complementation experiments manifested in green foliage and high biomass, the presence of non-transgenic roots is not a problem. However, when the role of genes in symbiotic nodule development and functioning is to be investigated by creating mutations in the genes, for example, with the help of the CRISPR/CAS9 system, the deleterious effect of the mutation in the transgenic nodules on the plant vigor will be masked by the presence of functional nodules on the adventitious roots. Thus, in such an experiment, the genotype (i.e. whether it is transgenic and whether there is/are mutation/s in the targeted gene) and the phenotype of all roots/nodules must be checked individually. To reduce the time and labour needs of such studies, a possible approach is based on the use of genes as reporters that are essential for nodule initiation, i.e. their mutation results in the failure of nodule formation. Nodules

formed after the transformation on plants carrying mutation in such a gene with vectors carrying the wild-type gene as reporter to complement the nodulation phenotype would be all transgenic. In such a system, only the phenotype and genotype of those roots that formed nodules need to be checked.

To develop the system with the characteristics mentioned above, a gene involved in nodule initiation with small size to avoid complicated cloning steps and to reduce the vector size as well as with limited number of commonly used restriction enzyme recognition sites is needed. The *nsp2* gene of *M. truncatula* (Kaló et al., 2005) coding for a GRAS-type transcriptional regulator satisfies these criteria: i) the *nsp2* mutant plants – although exhibit Ca-spiking in response to Nod factors and rhizobia – do not show any morphological changes and do not develop nodules; ii) the gene comprises a single exon and codes for a protein of 508 amino acids, i.e. a 1527 bp coding sequence between a promoter and a poly-adenylation signal has to be cloned.

2. Objectives

To facilitate the functional analysis as well as studies on the regulation of the *NCR* genes, our aim was to create and test several tools such as:

* to develop binary plant transformation vectors with a reporter marker that can be detected with naked eyes without the use of staining techniques or fluorescent microscopy

*to develop a hairy root transformation system that ensures that all the developing symbiotic root nodules are transgenic and thus, the the effect of the transgene on the nodule phenotype can be investigated without sorting and separating of the transgenic/non-transgenic tissues

*to establish systems including EMSA (electrophoretic mobility shift assay), DNA pull-down, yeast one-hybrid screening to identify cis- and trans-acting regulatory elements of *NCR* genes.

3. Materials and Methods

3.1. Biological materials and growth conditions

Escherichia coli strain MDS™42 ΔrecA Blue (Scarab Genomics, USA) was used for cloning purposes and grown at 37°C in LB medium (10 g/l tryptone; 5 g/l yeast extract; 5 g/l NaCl). *Agrobacterium rhizogenes* strain ARqua-1 (Quandt et al. 1993), K599 (Savka et al. 1990) and *Agrobacterium tumefaciens* strain AGL-1 (Lazo et al. 1991) harboring the different binary vectors was grown in LB medium at 30 °C. Wild-type *Medicago truncatula* cv. Jemalong (Register of Australian Herbage Plant Cultivars: Reg. No. B-9a-2) as well as *dnf-4* (Kim et al. 2015), *dnf7-2* (Horváth et al. 2015) and *nsp2*-2 (Oldroyd et al. 2003) mutant plants were inoculated with *Sinorhizobium medicae* strain WSM419 (Reeve et al. 2010) grown in TA medium (10 g/l tryptone; 1 g/l yeast extract; 5 g/l NaCl; 1 mM $MgSO_4$; 1 mM $CaCl_2$) at 30 °C for two days. Hairy roots of soybean (Glycine max cv. Williams 82) were inoculated with Bradyrhizobium japonicum strain CB1809 grown in YEM medium (0,5 g/l K_2HPO_4, 0,2 g/l $MgSO_4$ x $7H_2O$, 0,1 g/l NaCl, 10 g/l mannitol, 0,4 g/l yeast-extract).

M. truncatula plants were scarified with H_2SO_4 for 8 minutes and washed with large volume of ice-cold H_2O for five times. Scarified seeds were surface sterilized with 0.1 % (weight/volume) $HgCl_2$ for 2 minutes followed by three washes with sterile water, vernalized at 4 °C for two days in dark, then germinated on water-agar plates at 24 °C in dark. Seeds of soybean cultivar Williams 82 were surface sterilized with 3% of H_2O_2 in 70 % ethanol and germinated in wet vermiculite or on wet filter paper. *Arabidopsis thaliana* ecotype Columbia (Rédei 1992) was suface sterilized with 70 % ethanol and 0.1 % (weight/volume) $HgCl_2$, vernalized at 4 °C for two days in dark, then germinate on 1/2 MS (Duchefa, Haarlem, the Netherlands) argar medium supplied with 0.5 g/L MES (Duchefa, Haarlem, the Netherlands) and 0.5% sucrose, pH 5.7.

Plants were grown in vermiculite and assayed for nitrogen fixation ability in a glasshouse at 22 °C with 16/8 hours light/dark cycles. Nitrogen fixation efficiency of

all *Medicago* plants was assessed by both the phenotype (green healthy Fix^+ plants versus nitrogen starved yellow Fix^- plants) and the dry weight of the plants six weeks after inoculation. Dry weight measurement was performed after drying the samples at 80 °C for 48 hours.

Sacharomyces cerevisiae strains AH109 (MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2::$GAL1_{UAS}$-$GAL1_{TATA}$-HIS3, $GAL2_{UAS}$-$GAL2_{TATA}$-ADE2, URA3::$MEL1_{UAS}$-$MEL1_{TATA}$-lacZ, MEL1) and Y187 (MATα, ura3-52, his3-200, ade2-101, trp1-901, leu2-3, 112, gal4Δ, met^-, gal80Δ, URA3::$GAL1_{UAS}$-$GAL1_{TATA}$-lacZ) were grown in YPDA medium (Ausubel et al. 1989). For the selection of transformants and interactions synthetic media SD/ -Leu -Trp and SD/ -Leu -Trp -His (Clontech Laboratories), respectively, were used.

3.2 Hairy root transformation

Hairy root transformation of *M. truncatula* was performed according to the protocol described by Boisson-Dernier et al. (2001) with some modifications. To prepare the bacterial inoculum, a single colony of *A. rhizogenes* with binary vector was inoculated into 4 mL LB liquid medium supplemented with the appropriate antibiotics and grown at 30 °C with shaking for 2 days. 500 µL of the starter culture was added to 50 mL fresh LB with antibiotics and bacteria were grown till OD_{600}= ~0.8. Cells were harvested with centrifugation at 4000 rpm for 5 minutes, washed 3 times with sterilized water to remove nutrients that might reduce virulence, and finally, re-suspended in 1 ml sterile water.

Seedlings with 0.7-1 cm radicle length were cut under the cotyledonary node and all cut surfaces were covered with 10 µL of the bacterial suspension. The infected seedlings were kept at 22 °C in dark for three hours. After the initial co-cultivation, two methods were used for hairy root development and assay: In the case of the "agar only" approach, infected seedlings were placed onto half-strength Murashige-Skoog (1/2 MS) buffered with 0.5g/L MES (pH=6.5) or Fahreus medium both supplied with the appropriate selection if needed. After hairy roots were formed, the transformed plants were either directly inoculated with rhizoba on the Fahreus medium or were

transferred into vermiculite and inoculated with the symbiotic partner after 1 week of recovery and accomodation to the new environment.

The “soil plug” method was designed to avoid harming roots that may occur when tranformed plants are transferred from the agar plates into vermiculite. In this approach, infected seedlings were transferred into “soil plugs” after the 3 hours co-cultivation period, that were prepared by filling a 1 cm long tube (diameter 0.82 cm) with soil and sterilization. In this way, hairy roots emerge and develop under the protection of the tube in the soil. After two weeks of hairy root development, planta can be transferred into vermiculite together with the tube and continue growth without any disruption.

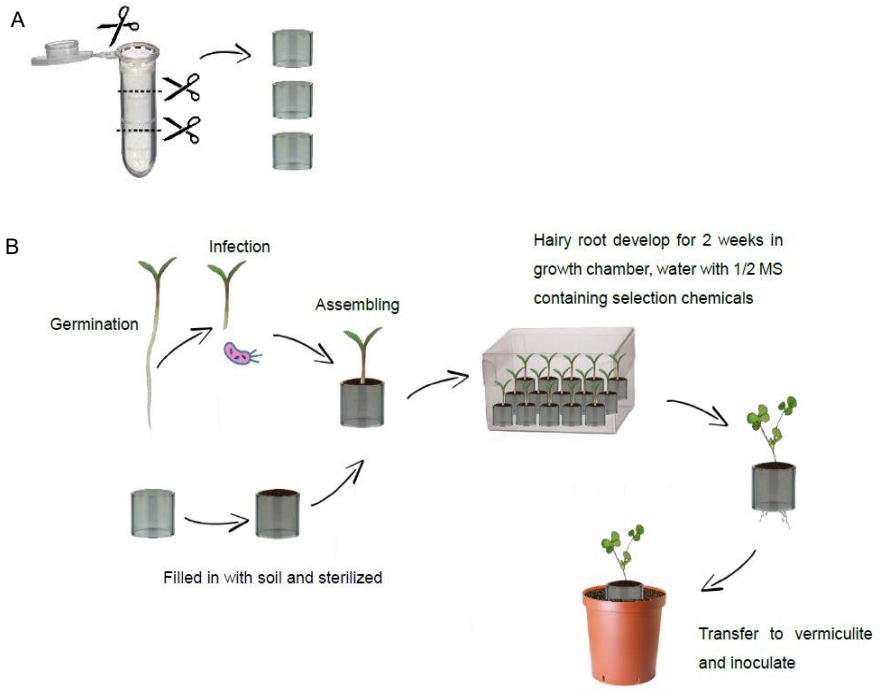


Figure 3.1. Procedure for conducting “soil plug” hairy root transformation. Tube can be prepared as shown in ***A*** from Eppendorf tube or other sources. Before use, fill soil into the tubes and sterilize to have the “soil plug” (do not fill too tight, which may affect hairy root development). After the initial co-cultivation, infected seedling is placed into the “soil plug” (***B***). During this process, 1/2 MS (pH 6.5, supplied with 0.5g/L MES and with antibiotics if selection is needed) can be used to water the “soil plug” to provide moisture. After assembly, keep the plants in a transparent box and place in a growth chamber with 16 hours light and 8 hours darkness under 22 degree for ~2 weeks. After hairy roots develop well, plants can be transferred into vermiculite and keep in a green house and inoculate with rhizobium.

3.3 *Arabidopsis thaliana* floral-dip transformation

Arabidopsis thaliana was grown in the glasshouse in soil at at 22 °C with 16/8 hours light/dark cycles till flowers developed. *Agrobacterium tumefaciens* strain AGL-1 was streaked on LB agar plate supplied with desired antibiotics and incubated at 30 °C for two days till colonies were formed. Colonies were inoculated into 5 mL LB liquid culture, allowed to grow for another two days. 0.5 mL culture was inoculated into fresh 50 mL LB medium and grow for 16-24 hours before transformation. For the floral-dip transformation, agrobactruim was collected by centrifugation at 4000 rpm for 10 min, washed with 50 mL water once and resuspended in 1/2 MS, supplied with 5% sucrose and 0.03% Silwet-L77 to an optical density nm of ~0.8 at 600 (Chung et al. 2000; Clough and Bent 1998). After removing the developing siliques and opened flowers, the rest of the inflorescence was inmmersed in the bacterial suspension for 10 sencends. After dipping, the plants were covered to keep high huminity and keep off light for 16-24 hours and then kept growing in the green house till siliques matured.

3.4 DNA extraction

For plant genomic DNA isolation, lysis of plant tissues was performed by grinding them into fine powder in liquid nitrogen with mortar and pestle. Approximately 100 µL powder was transformed into an Eppendorf tube and 650 µL CTAB extraction buffer (2% w/v CTAB, 0.1 M TRIS-HCl (pH=8.0), 20 mM EDTA (pH=8.0), 1.4 M NaCl) were added. The suspension was incubated at 65 °C for 1 hour with ocasional mixing. To remove chlorophyll and cell debris, 200 µL chloroform: isoamyl alcohol (24:1) were added and the mixture was incubated at RT for 15 minutes. After centrifugation at 10000 rpm for 10 minutes, 500 µL of supernatant was transferred to a new Eppendorf tube and the DNA was precipitated by adding two volumes of 96% ethanol. After 1 hour at -20 °C, DNA was pelleted by centrifugation at 13000 rpm for 10 minutes. The pellet was washed for 3 times with 70% ethanol, dried and the DNA was dissolved in 100 µL water.

For plasmid DNA isolation, a kit from Bio Basic inc was used. Bacteria from 4 mL of overnight culture was collected by centrifugation, resuspended in 100 µL of Solution I, lysedby adding 200 µL Solution II, then the proteins and genomic DNA were precipitated by adding 350 µL Solution III. The mixture was centrifuged for 5 minutes at 12, 000 rpm and the supernant was transferred onto the column and after 1-minute incubation, it was centrifuged for 2 minutes at 10, 000 rpm. Plasmid DNA was washed twice with the Wash solution and eluted with 50 µL DNAase free water or Elution buffer.

Plasmid DNA from yeast was released by resuspending yeast cells in 20 µL DNAase free water or TE buffer followed by 3 rounds of freeze-thaw cycles using liquid nitrogen. The release plasmid could be used for PCR reaction or transformation of *E. coli*.

3.5. PCR conditions and primers used

In gernal, DreamTaq DNA Polymerase (Fisher-Thermo Scientific) was used for colony PCR and Phusion High-Fidelity DNA Polymerase (Thermo-Fisher Scientific) was used to amplify fragments for vector construction or mutation detection after CRISPR-CAS9 mediated genome editing. PCR reactions were conducted by following the manufacturers' instructions.

Primers used in this study were synethized by Thermo-Fisher Scientific and are listed in table S1.

3.6. Vector and gene constructions

The *MtLAP1* gene was amplified with Phusion High-Fidelity DNA Polymerase (ThermoFisher Scientific) using *M. truncatula* cv. Jemalong genomic DNA as template and the MtLAP1_pCncoF-MtLap1_pCeheR primer pair (Table S1) and the PCR fragment was cloned into the *Nco*I-*Ehe*I digested pCAMBIA2201 vector with the help of the In-Fusion Ligation Kit (TaKaRa) to replace the *gusA* gene coding for β-glucuronidase. To provide other regulatory sequences, first, the Cauliflower Mosaic

Virus 35S promoter was removed and the promoter sequences of the *At2g37950* (*pAtE47*) and the *At5g24800* (*pAtS5*) genes showing tissue-specific expression in both *Arabidopsis thaliana* (Lee et al. 2006) and *Lotus japonicus* (Gavrilovic et al. 2016) were amplified using *Arabidopsis thaliana* ecotype Columbia (Rédei 1992) genomic DNA as template and primer pairs 2g37950prF-2g37950prR and 5g24800prF-5g24800prR were cloned at the NcoI site with the help of the In-Fusion Ligation Kit (TaKaRa). The vectors harboring the CaMV 35S, pAtE47 and pAtS5 promoters in front of the MtLAP1 gene were named pPurpleRootC, pPurpleRootE and pPurpleRootP, respectively.

The same MtLAP1 fragment generated with MtLAP1_pCncoF-MtLap1_pCeheR primer pair was cloned into the *Nco*I-*Ehe*I digested pCAMBIA3301 vector to replace the *gusA* gene, thus generate the pPurpleShoot vector which can be used for transformation of *Arabidopsis* or *Medicago* if transferred into *Agrobacterium tumefaciens* strain AGL-1.

The primer pair MaNCR169salF-MaNCR169HindR was used to amplify the *NCR169* gene of *Melilotus albus*, and then the *Sal*I-*Hin*dIII fragment replaced the *Medicago* sequence in the *M. truncatula NCR169-Strep* construct (Horvath et al. 2015). To introduce the K40R mutation into the *MtNCR169* gene, overlapping fragments were amplified with the help of the NCR169sacF-NCR169K40Rrev and the NCR169K40Rfw-NCR169hindR2 primer pairs that were joined by overlapping PCR. The mutant fragment replaced the wild-type fragment in the *NCR169-Strep* construct.

3.7 RNA extraction and cDNA synthesis

For RNA extraction, plant tissues were frozen in liquid nitrogen immediately after collection, ground into fine powder under liquid nitrogen and were immediately extracted or kept at -80 °C until used. To ~100 µL tissue powder, 1 mL of Tri-Reagent (Sigma) was added, followed by vortexing for 30 seconds and incubating at RT for 5 minutes. The extract was mixed with 200 µL of phenol: chloroform: isoamyl alcohol (25:24:1, v/v) and incubated for 2 minutes. After centrifugation for 10 minutes at

12000 rpm at 4 °C, equal volume of isopropanol was added to the supernatantand the RNA was precipitated at -20 °C for 1 hour. The RNA was pelleted with centrifugation for 10 minutes at 12000 rpm at 4 °C and then washed with prechilled 70% ethanol for 3 times. The RNA was dissolved in suitable volume of nuclease free water and stored at -80 °C.

DNase treatment was carried out by TURBO DNase (Thermo Fisher) to remove contaminating genomic DNA from the RNA sample prior to cDNA transcription. Reaction mixtures were composed as follows: 10x DNase Buffer 6 µL; RNase inhibitor (2U) 1 µL; TURBO DNase (2U) 1 µL; Nuclear RNA 50 µL; Nuclease-free water 2 µL in a total volume of 60 µL. Samples were incubated for 30 min at 37°C. To remove the DNAse from the RNA sample, 290 µL nuclease-free sterile water and 350 µL phenol/chloroform/isoamyl alcohol (25:24:1 v/v) were added and were mixed with vortexing and centrifuged at 12000 rpm for 15 min at 4°C. To the aqueous phase (350 µL) 35 µL 3 M Na-acetate (1:10 v/v), 20 µg glycogen and 875 µL (1:2.5 v/v) ice cold ethanol were added and the mixture was kept at -20°C for 2 hours and then centrifuged at 16.000 x g for 30 min at 4°C. The RNA pellet was washed with 500 µL prechilled 70% ethanol, centrifuged at 12000 rpm for 5 min at 4 °C, air dried at RT and resuspended in 20 µL nuclease-free water.

cDNA was synthesized by using 500 ng total RNA (measured by NanoDrop2000 Spectrophotometer) with SuperScript® IV VILO Master Mix (Thermo Fisher). The reaction mixture was composed as follows: Total RNA 20 µL; nuclease-free water 18 µL; SuperScript IV Vilo Master Mix 2 µL in a total volume of 40 µL. The samples were incubated at 25 °C for 10 min, at 42 °C for 120 min and at 85°C for 5 min. cDNA was stored at -20 °C.

3.8 Extraction of nuclei from nodules

If small amount of cell nuclei was needed, nodule tissues and cells were broken by chopping nodules into small particles in a Petri dish with a sharp razorblade under pre-chilled nuclear isolation buffer (NIB) containing 20 mM Tris-HCl (pH 7.4), 25% glycerol and 2.5mM $MgCl_2$. If large number of nuclei was needed, nodules were

collected and ground into fine powder in liquid nitrogen, then the tissue was suspended in chilled NIB. The suspensions from both sources were filtered through a nylon mesh with a pore size of 100 μM first, then through a nylon mesh with a pore size of 50 μM to remove cell debris. After filtration, nuclei were collected by centrifugation at 1500 x g for 10 minutes at 4 °C and were immediately used or stored at -80 °C.

Nuclear proteins were extracted by resuspending the isolated nuclei in BS/THES washing buffer composed of 44.3 % THES buffer (50 mM Tris HCl (pH=7.5), 10 mM EDTA, 20% sucrose (w/v), 140 mM NaCl), 20% 5x BS buffer (50 mM HEPES, 25 mM $CaCl_2$, 250 mM KCl, 60% glycerol) and 35.7% water, supplied with 0.3% cOmplete protease inhibitor cocktail (Roche Diagnostics Ltd, Mannheim, Germany), followed by vortexing for 5 seconds every 3 minutes for 15 minutes. To increase extraction rate, the extract was sonicated 3 times for 10 seconds.

3.9 DNA pull-down assay

To isolate and identify proteins binding to the *NCR169* gene promoter, DNA pull-down assay was carried out with a modified protocol of Jutras et al. (2012). To create the baits for the pull-down, the 907 bp promoter sequence was amplified as three 400 bp fragments overlapping with 100 bp tiling. For each bait, two PCR reactions in 1.5 ml volume were set up: in one reaction the forward primer was biotin-labeled while in the other one, the reverse primer carried the biotin tag. PCR fragments were precipitated by adding 1/10 volume of 3 M sodium-acetate and 1 volume of isopropanol at -80 °C overnight. Precipitated DNA was collected by centrifugation at 13000 rpm for 10 minutes and washed 3 times with 70% ethanol, then dissolved in nuclearase free H_2O. The purity and concentration of the DNA fragments were checked by agarose gel electrophoresis and optical density measurement using Nanodrop.

200 μL of Dynabeads M280 (which can bind 40-80 μg of DNA) was used to bind each bait fragment pairs. The beads were washed 3 times with 2x B/W buffer (10 mM Tris HCl (pH=7.5), 1 mM EDTA, 2 M NaCl) and then incubated with 200 μL

DNA probe (containing 40-80 µg of DNA in total) for 20 minutes. To saturate beads, they were removed from the probe with the help of a magnet, then the binding reaction was repeated. Finally, beads were washed three times with TE buffer (10 mM Tris HCl, 1 mM EDTA, pH=8.0) and two times with BS/THES washing buffer composed of 44.3 % THES buffer (50 mM Tri HCl (pH=7.5), 10 mM EDTA, 20% sucrose (w/v), 140 mM NaCl, 20% 5x BS buffer (50 mM HEPES, 25 mM $CaCl_2$, 250 mM KCl, 60% glycerol) and 35.7% water. One microgram of protein extract from nodule nuclei were added to the bait-bead complexes along with sonicated *E. coli* genomic DNA as competitor to reduce unspecific binding of proteins to the bait DNAs and incubated at room temperature for 30 minutes. After binding, the bead DNA complexes were washed five times with BS/THES buffer supplemented with sonicated *E. coli* genomic DNA and five times with PBS. Trypsin digestion of the proteins bound to the baits were performed directly on the beads to reduce loss of samples. Digested samples were then used for mass spectrometry analysis.

3.10 Yeast One Hybrid cDNA library screening

3.10.1 Generation of the bait strains

Yeast strains Y187 and AH109 were transformed using the modified protocol of Bass et al. 2016. A loop of yeast colony was inoculated into 10 mL of liquid YPDA and the culture was grown overnight. Next day, the starter culture was diluted to OD_{600}~0.2-0.3 in 50 mL fresh YPDA and grown for ~4 hours till OD_{600} reached 0.4-0.6. Cells were harvested by centrifugation at 4000 rpm for 1 min and washed once with 50 mL sterilized water and resuspended in 1 ml 1xTE/LiAc (100 mM lithium actate, 10 mM Tris-Hcl, 1 mM EDTA, pH 7.5) to genaerate competent cells. 100 ng linearized pINT1 vector (Meijer et al, 1998) carrying the *NCR169promoter-HIS3* or the *NCR211promoter-HIS3* gene construct and 25 µg denatured salmon sperm DNA were mixed with 50 µl yeast competent cell suspension and 300 µl fresh 40% PEG 3350 in 1xTE/LiAc and incubated at 30 °C with shaking

for 30 minutes then at 42 °C for 20 minutes (mix every 10 minute by inversion). Cells were collected with centrifugation at top speed for 30s and resuspended in 1 mL YPDA, then incubated at 30 °C for 3-6 hours and finally spread on YPDA plates containing 150 μg /mL G418. Colonies were formed after 2-3 days.

3.10.2 Transformation Based Screening

Competent cells of yeast strains carrying the bait constructs were prepared as described in the previous section. To transform the prey clones, 25 μg Yeastmaker carrier DNA (Clontech Laboratories) and 0.5-1 μg plasmid DNA from a *M. truncatula* nodule ETS library (Györgyey J et al. 2000) were mixed with 50 μL yeast competent cell and 300 μL freshly made 40% PEG in 1xLiAc/TE, and incubated at 30 °C with shaking for 30 minutes then at 42 °C for 20 minutes (mix every 10 minute by inversion). To increase the coverage of the library in the screening, ten transformations were conducted. Transformed cells from the ten transformation reactions were spread on 20 SD -Leucine -Histidine plates (diameter: 150 mm) and grown for 6 to 14 days at 30 °C. Faster growing putative positive colonies were transfered to new SD -Leu-His medium plate containing 50 mM 3-AT (3-amino-1, 2, 4, -triazole, Sigma) in “96-spot format”. Those colonies that could grow again were treated as positive ones expressing proteins binding to the bait DNA.

To test the binding specificity, β-Galactosidase activity was determined because specific binding proteins can only activate the *HIS3* gene, while unspecific ones can also activate the *lacZ* gene expression. Newly formed colonies were transferred to YPDA plates and grown for another two days. Colonies were suspended in 20 μL PBS containg 1 mg/mL X-gal and cell were disrupted by three rounds of freeze-thaw cycles using liquid nitrogen and then incubated at 37 °C for 3 hours till the blue color appeared. For the positive yeast colonies (do not change to blue), after the incubation, cell debris was removed by centrifugation, then 10 μL of the supernatant was used for transformation of *E.coli* MDS42 recA Blue and the insert of the clones was sequenced with primer 3’AD and 5’AD (seuquences listed in Table S1).

3.10.3 Mating based screening

Nodule cDNA library in yeast strain Y187 was generated by transforming 100 μg DNA from the library plasmid into yeast competent cells made from 1 L cell culture as described above. After transformation, cells were spread on 50 SD-Leu plates (150 mm). A total number of at least 1 million colonies had to be collected for a sufficient screening. Library colonies were harvested by washing the plates with 5 mL YPDA medium containing 25% glycerol and were pooled. The library was stored in 1 mL aliquots at -80 °C.

For screening, overnight culture of AH109 carrying the pINT1-NCR169promoter-HIS3 construct was diluted in 50 mL YPDA to OD_{600}~0.2, grown to OD_{600}=0.8 and harvested by centrifugation for 2 minutes at 4000 rpm. Cells were resuspended in 50 mL 2x YPDA medium supplied with 50 μg/mL kanamycin (for getting rid of contamination) and transfered into a 2 L flask. One mL of the cDNA library stock was added and the mixed culture was incubated at 30 °C with shaking at low speed (30-50 rpm) for 24 hours. Mating status was checked by microscopy and when the majority of the observed cells seemed to be diploid, the culture was plated onto SD-His-Leu medium and incubated at 30 °C for 2-7 days till colonies formed. Positive colonies were examninated in the same method described for transformation-based screening.

3.11 EMSA assay

For EMSA assay, a 400 bp promoter region before the ATG translation start codon of the *NCR169* gene (able to drive the expression of a gene construct that could complement the *dnf7* mutant, data not shown) was chosen and amplified with seven 100 bp long fragments with 50 bp tiling using primers carrying the recognition sequence of the EcoRI restriction enzyme.

Each fragment was digested with FastDigest *Eco*RI (Thermo Fisher Scientific) enzyme and radioactively labeled with ^{32}P-dATP while the ends were filled in using the Klenow Fragment of DNA polymerae I (Thermo Fisher Scientific). Labeled

probes were purified with precipitatation by adding 1/10 volume of 3 M sodium-acetate and 1 volume of isopropanol at -80 °C overnight. Precipitated DNA was collected by centrifugation at 13000 rpm for 10 minutes and washed 3 times with 70% ethanol, then dissolved in nuclearase free H_2O.

6 µL of the nuclear protein extracts (~30 µg protein) were incubated with 2 µL of labelled probes in a reaction mixture containing 2 µL 100x BSA, 2 µL (3 µg) sonicated *E. coli* genomic DNA and 3 µL 5x binding buffer (50 mM Tris HCl (pH 8.0), 750 mM KCl, 2.5 mM EDTA, 0.5% Triton-X 100, 62.5% glycerol (v/v), 1 mM DTT).

The labeled DNA fragments complexed with nuclear proteins were loaded onto 5% polyacrylamide native gels (prepared by mixing 63 mL H_2O, 4 mL 10x TBE, 13 mL 30% acrylamide/bis-acrylamide stock (29:1), 800 µL 10% APS and 50 µL TEMED) that were pre-run for 30 minutes before loading. Gels were run for ~2 hours at 150V in 0.5x TBE buffer (44.5 mM Tris, 44.5 mM boric acid, 1 mM EDTA), then wraped with saran wrap and exposed for 2 hours on X-ray film.

For non-radioactive EMSA, non-labeled fragments were complexed with the nuclear protein extracts as described for the radioactively labeled fragments, then the gels were stained in suitable volume of 10000-times diluted SYBR gold (Thermo Fisher) in TBE buffer for 30 minutes. After washing the stained gels for 30 seconds with water, imaging was done with a MiniLumi gel imaging system (DNR Bio Imaging Systems) applying a HOYA X1 green UV filter.

4. Results and Discussion

4.1 Optimization of hairy root transformation for large-scale reverse genetic studies

4.1.1 Anthocyanin as a new reporter

Based on its ability to induce the production of anthocyanins, the *MtLAP1* gene coding for a MYB transcription factor was chosen to use as reporter in a new set of vector system designated as pPurpleRoot vectors.

4.1.1.1 Construction of pPurpleRoot vectors

The pCAMBIA3301 and pCAMBIA2201 vectors were chosen to be the backbone of the pPurpleRoot plasmids because of their advantageous features such as high copy number in *E. coli* and and a number of restriction enzyme cloning sites in the *lacZα* fragment providing the possibility for blue-white selection of the recombinant clones. First, the *MtLAP1* gene was amplified from the *Medicago truncatula* genome and cloned into pCAMBIA3301 to replace the *GUS* gene which is driven by the constitutive *CaMV35S* promoter resulting in pPurpleShoot. In the course of *Agrobacterium tumefaciens* mediated stable transformation of plants with pPurpleShoot, transgenic tissues and plants can be selected using the herbicide glufosinate ammonium if needed. To construct vectors with kanamycin resistance, which is generally used in hairy root transformation experiments, the *35S promoter-MtLAP1* fragment was transferred from pPurpleShoot to pCAMBIA2201 to generate the pPurpleRootC (Continuous expression) vector. As there are reports on the antimicrobial activity of anthocyanins that might affect the nodulation and nitrogen fixation processes, promoters with different tissue specifities were tested to find ones that facilitate anthocyanin production in cells that are not in contact with the infecting rhizobia. We constructed two other plasmids in which the promoters that replaced the

35S promoter restrict the expression of the *MtLAP1* gene to the central cylinder of the root. The *pAtE47* promoter from gene *At2g37950*, which has an endodermis and pericycle specific expression pattern in both *Arabidopsis thaliana* (Lee et al. 2006) and *Lotus japonicus* (Gavrilovic et al. 2016), was used to generate pPurpleRootE (specific expression in Endodermis and pericycle), while pericycle and phloem specific promoter *pAtS5* from gene *At5g24800* was used to construct pPurpleRootP (specific expression in Pericycle and phloem) (Fig 4.1).

4.1.1.2 Anthocyanin reporter allows the easy visual selection of transgenic roots

To test the usability of the anthocyanin reporter system, the pPurpleRoot vectors were transformed into *A. rhizogenes* strain ARqua-1, and then the bacteria were used to induce the formation of transgenic hairy roots on *M. truncatula* cv. Jemalong A17 plants. The purple color of roots due to anthocyanin production makes it easy to distinguish the transgenic roots from the non-transformed ones (Fig. 4.2). Roots transformed with the pPurpleRootC construct showed strong coloration in the whole root that was even more enhanced in the vascular tissues (Fig. 4.2 a, d). Coloration of the transgenic roots emerging after transformation with pPurpleRootE (Fig. 4.2 b) and pPurpleRootP (Fig. 4.2 c) vectors agreed with the expected expression pattern in the endodermis/pericycle and pericycle/phloem tissues, respectively.

To investigate whether the production of anthocyanins affects the development of symbiotic nodules, the transgenic roots were inoculated on agar plates with *Sinorhizobium medicae* strain WSM419. Although these environmental conditions allowed the formation of only low numbers of nodules that have deep purple coloration (most probably, as the mixture of leghemoglobin and anthocyanin colors) or purple vascular bundles on the roots transformed with pPurpleRootC (Fig. 4.2 d, g) and pPurpleRootP (Fig. 4.2 f, i) vectors, respectively, while no coloration of nodules could be observed on pPurpleRootE transformed roots (Fig. 4.2 b, e, h).

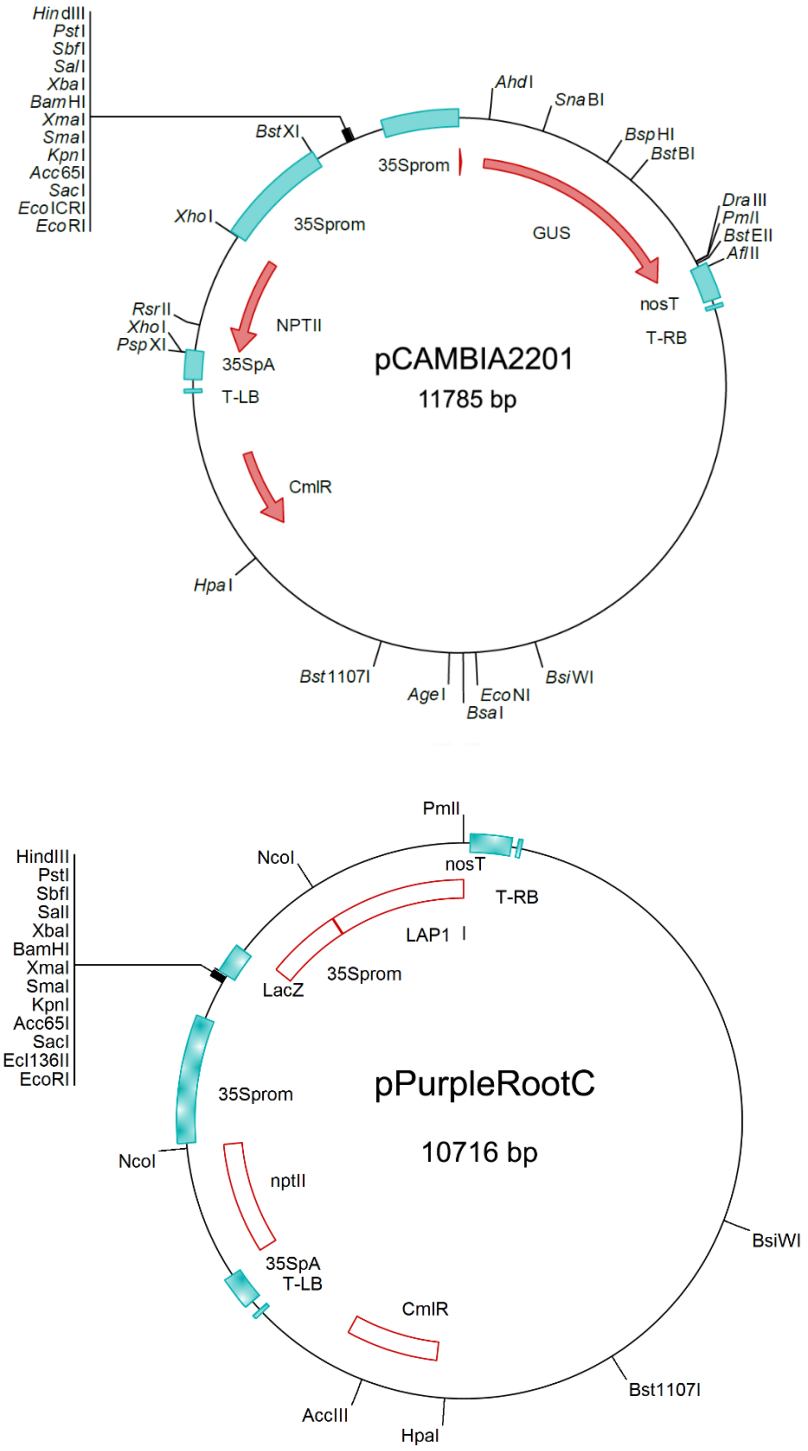


Figure 4.1. Schematic map of pPurpleRootC vector in comparison with the original pCAMBIA2201 vector. The vector was modified from pCAMBIA2201 (***up***) by replacing *GUS* gene with *MtLAP1*, while keeping the *CaMV35S* promoter to drive the expression of the reporter gene in the pPurpleRootC vector (***below***). For vascular tissue specific expression, the *CaMV35S* promoter was replaced with promoters *pAtE47* or *pAtS5* to create pPurpleRootE and pPurpleRootP respectively. Chloramphenicol is used for maintenance and selection of the plasmid in *E. coli* and *Agrobacterium*, while the *nptII* gene allows the selection of transgenic roots after *Agrobacterium* mediated transformation.

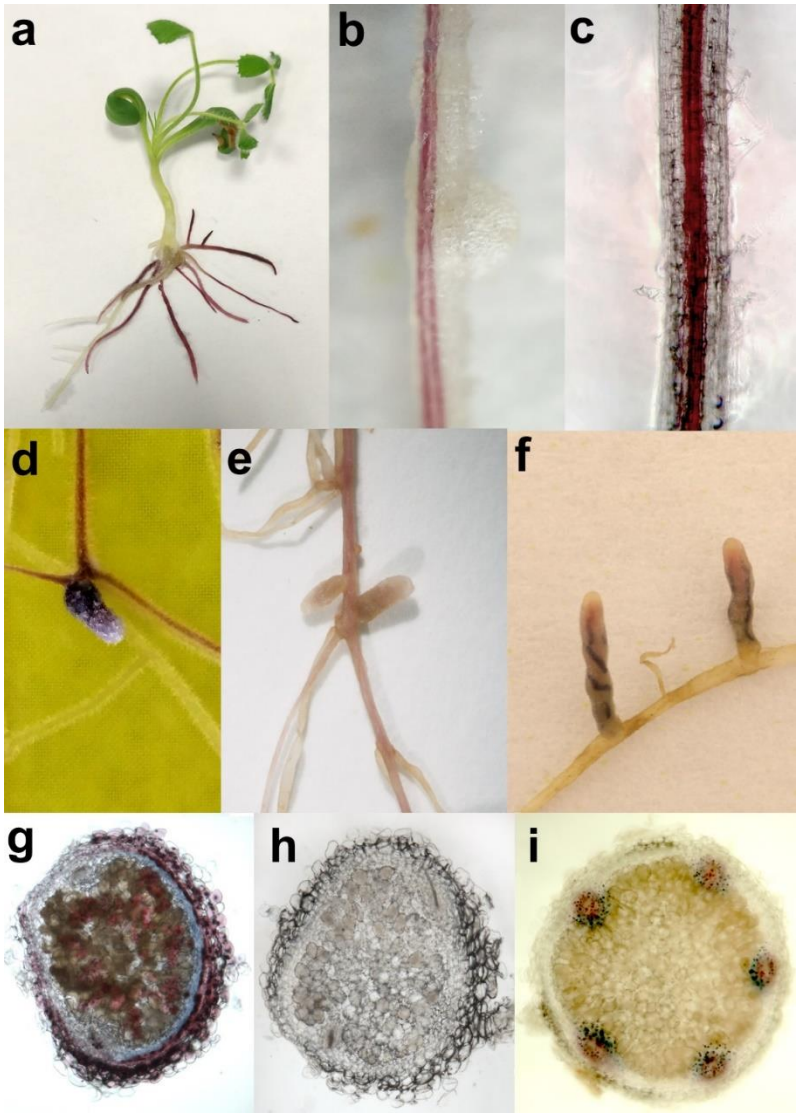


Figure 4.2. Anthocyanin accumulation in roots and nodules after hairy root transformation. **a** Transgenic hairy root formed on *M. truncatula* shoots display purple coloration. **b** Endodermis/pericycle and **c** Pericycle/phloem specific production of anthocyanins after transformation with pPurpleRootE and pPurpleRootP, respectively. **d, g** Strong coloration of the vascular and nodule tissues of roots formed after transformation with pPurpleRootC. **e, h** Nodules formed on pPurpleRootE transformed roots do not produce anthocyanins in the nodule (vasculature). Note that no coloration can be observed in young developing nodules in panel **b** either. **f, i** Anthocyanin accumulation in the nodule vasculature on hairy roots transformed with pPurpleRootP. Note that weak or no coloration can be observed in the root vasculature of older plants (at least 6 weeks after inoculation) shown in panels **e** and **f**.

4.1.1.3 The new reporter system can be used for functional investigation of genes required for the formation of nitrogen-fixing root symbiotic nodules

Hairy root transformation performed on the wild-type A17 plants demonstrated that the *MtLAP1* reporter gene driven by *CaMV 35S*, *pAtE47* and *pAtS5* promoters can be used for the easy identification of transgenic hairy roots, however, we wanted to show that the anthocyanin accumulation does not interfere with normal nodule development and function.

The usability of this new reporter/vector system in symbiosis was tested by complementation of the *M. truncatula dnf7-2* mutant (Table 2, Fig. 2) carrying a deletion in the *NCR169* gene and resulting in Fix^- phenotype (Horváth et al. 2015). To compare the efficiency of the vectors, the *NCR169* gene was cloned both into the initial pCAMBIA2201 and into the pPurpleRoot vectors. The constructs were introduced into the mutant with the two different transformation protocols described in the Materials and Methods section in details, and then the transformed plants were transferred into vermiculite where nodulation assays were performed. The plants were inoculated with *S. medicae* strain WSM419 and their root/nodule and symbiotic phenotypes were evaluated 6 weeks post-inoculation. In the "agar only" experiment, reporter gene activity driven by the constitutive CaMV35S promoter was observed (Table 4.1) in at least on one root of almost all plant (pCAMBIA2201: 90-95%; pPurpleRootC: 100%). In the case of pPurpleRootE and pPurpleRootP transformed roots, anthocyanin accumulation indicated lower numbers of transformed plants corresponding to ~58% and ~44% transformation frequency, respectively. However, there were pPurpleRootE and pPurpleRootP transformed plants in which anthocyanin could not be seen in the roots, though they were complemented as manifested in green foliage and high biomass (Fig. 4.3). One possible explanation for this observation is that reporter gene activity might have been silenced via post-transcriptional gene silencing (Depicker and Van Montague 1997) resulting in no or too low anthocyanin

accumulation to be detected in these tissues. The complementation efficiency of the pPurpleRootE and pPurpleRootP constructs was comparable to that of the pCAMBIA2201 harbouring the *NCR169* gene. In contrast, despite the high transformation rate, the complementation efficiency of the pPurpleRootC derived construct was very low (~20%). Although antimicrobial activity and the mode of actions of anthocyanins has been studied and shown mostly in relation to human health (reviewed in Smeriglio et al. 2016), these results indicate that too high level of anthocyanins may negatively affect the interaction of the plant roots with rhizobia. Interestingly, in the "soil plug" experiment, the ratio of plants with roots expressing the transgene from the 35S promoter was lower than when the "agar only" method was used (Table 4.1), but all the plants scored to be transgenic were complemented. The explanation for this observation requires further experiments.

Table 4.1. Transformation and complementation efficiency by using the pPurleRoot vectors.

[a] Co-cultivation and hairy root development on agar plates. [b] Hairy root development in soil plugs.

Construct	number of plants with visible reporter gene activity (total number of plants)	number of plants with effective nodules showing reporter gene activity	number of plants with effective nodules lacking reporter gene activity
[a]pCAMBIA2201	19 (20)	0	0
[a]pCAMBIA2201:: NCR169	18 (20)	12	0
[a]pPurpleRootC:: NCR169	20 (20)	5	0
[a]pPurpleRootE:: NCR169	11 (19)	11	5
[a]pPurpleRootP:: NCR169	7 (16)	7	3
[b]pCAMBIA2201	74 (80)	0	0
[b]pCAMBIA2201:: NCR169	76 (78)	38	0
[b]pPurpleRootC:: NCR169	20 (80)	20	0
[b]pPurpleRootE:: NCR169	36 (79)	36	ND
[b]pPurpleRootP:: NCR169	29 (77)	29	ND

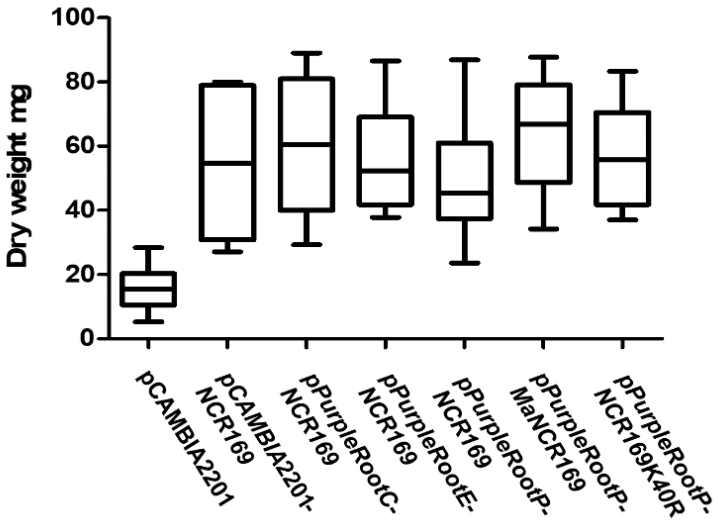


Figure 4.3. Dry weigth of *dnf7-2* mutant plants carrying transgenic roots after transformation with the indicated vector constructs. Empty vector pCAMBIA2201 was used as a negative control and pCAMBIA2201 carrying the *NCR169* gene served as the positive control. For each transformation, ~30 plants were collected for dry weight measurement.

In addition, we performed functional studies with the *NCR169* gene. Allelic variations and post-translational modifications (PTMs) in proteins may affect their biological activity. Indeed, allelic variations of certain NCR peptides have been shown to affect the interaction of the host with certain strains of their bacterial partners (Yang et al. 2017; Wang et al. 2017; 2018). Similarly, it was shown that the NCR169 peptide essential for bacteroid development and symbiotic nitrogen fixation (Horváth et al. 2015) carries an acetyl modification at lysine 40 in *M. truncatula* nodules (Marx et al. 2016). This reversible type of PTM changes the charge of proteins and their interactions with other macromolecules as best known in the case of histones (Drazic et al. 2016). The mature NCR peptides of *M. truncatula* and *Melilotus albus* share 68% amino acid identity (Horváth et al. 2015). One of the differing residues is an asparagine in the *Melilotus* peptide in place of the lysine (K40) acetylated in *Medicago truncatula*, however, this amino acid has similar characteristics (polar, uncharged side chain) as the acetylated lysine, thus, it may not affect the activity of the peptide. To investigate whether the other amino acid differences in the *Melilotus NCR169* allele affect the complementation of the *M.*

truncatula mutant and whether the lysine acetylation of NCR169 is required for its biological activity, *MaNCR169* and *NCR169K40R* were cloned into pPurpleRootP and were used to complement the *dnf7-2* mutant. Based on dry weight data, both constructs were able to restore the symbiotic nitrogen fixation capability of the mutant (Fig. 4.3) indicating that NCR169 of *M. albus* have the same biological activity as that of *M. truncatula* cv. Jemalong during the interaction with *S. medicae* strain WSM419 and that reversible lysine acetylation, i.e. the charge of the peptide at that position does not affect its function and activity.

4.1.1.4 Anthocyanin might be used as a plant stable transformation marker

During construction of pPurpleRoot vectors, pPurpleShoot (pCAMBIA3301 vector carrying *CaMV 35S-MtLAP1*) was also generated which can be used for stable transformation on plant. In this study, for testing the system, *Agrobacterium tumefaciens* strain AGL-1 was transformed with pPurpleShoot and later *Arabidopsis thaliana* floral dip transformation was conducted. Transgenic plant was selected with Basta till T3 generation for homozygotes.

Based on observation, there is no obvious difference in growth and development between the transgenic *Arabidopsis* carrying *MtLAP1* gene and wild type plant, only the purple color from anthocyanin accumulation can be seen on leaves, flowers and siliques (Fig. 4.4). This result demonstrates the possibility to utilize *MtLAP1* gene as an indicator for transgenic *Arabidopsis* in the lab, or even for other plants in the future.

MtLAP1 gene encodes a R2R3-MYB family transcription factor, which can regulate the production of anthocyanin. Based on previous study, *Arabidopsis Production of Anthocyanin Pigment1* (*AtPAP1*) which is a closely related gene of *MtLAP1* cannot induce the biosynthesis of anthocyanin in *Medicago truncatula* or in *Medicago sativa*, even their share quite conserved R2R3 domain in their amino acid sequences (Peel et al., 2009), but in this study the *Medicago truncatula MtLAP1* gene

can triger anthocyanin accumulation in *Arabidopsis*, indicating that *MtLAP1* gene might possess the potential to be more widely used in other plant species as a reporter for transformation. If this reporter system can be developed, it will make the detection and selection procedure of plant transformation much easier. However, more work needed to be done to optimize the system since a high number of genes are involved in the process of anthocyanin synthesis, thus, the expression level of these genes must be tightly controlled to reduce side effects.

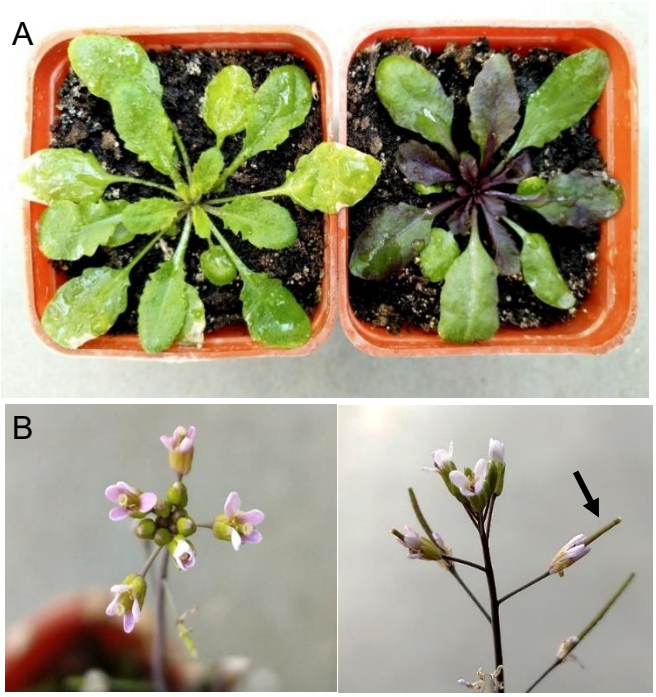


Figure 4.4. Anthocyanin accumulation pattern in transgenic *Arabidopsis thaliana*. A. Growth of T3 transgenic *Arabidopsis* homozygous plant over-expressing *MtLAP1* and wild type Col-0 was compared and showed no developmental difference. Purple color result from anthocyanin accumulation on rosette leaves can be observed. B. Expression of *MtLAP1* leads to coloration of anthocyanin on flowers (***left***) and siliques (***right***).

4.1.2 A hairy root selection system based on the *NSP2* gene essential for nodule initiation for gene knock-out experiments to ensure the formation of nodules only on transgenic roots

As described in the Introduction, the *NSP2* gene of *Medicago truncatula* fulfills the requirements set for the hairy root selection system that ensures the formation of only transgenic nodules: It is essential for nodule initiation because the *nsp2* mutant plants – although exhibit Ca-spiking in response to Nod factors and rhizobia – do not show any morphological changes and do not develop nodules. Moreover, it is small sized and does not increase the size of the vectors too much.

To facilitate CRISPR/CAS9 based editing of genes in hairy roots that might be involved in nodule development, we constructed a vector with the *CaMV35S* promoter driven *NSP2* gene as reporter and the soybean ubiquitin promoter (*GmUBIpr*) driven *CAS9* gene that was codon optimized for soybean (Curtin et al., 2017). The map of the vector is shown in Fig. 4.4. To validate the system, the *MtNFS1* gene (Yang et al., 2017) coding for an NCR peptide and causing incompatibility between *M. truncatula* cv. Jemalong and *S. meliloti* strain Rm41 was chosen. If the Jemalong allele is expressed in the nodules, the developing bacteroids of Rm41 are eliminated from the nodules, which, thus, do not provide reduced nitrogen for the plant (Fix^- interaction), however, mutating the *NFS1* gene via CRISPR/CAS9 in hairy roots results in compatible interaction between the two organism, i.e. in functional nitrogen fixing (Fix^+) nodules. We cloned the same guide RNA casette assembled from oligonucleotides shown in Table S1. into the vector that was used by Wang et al. (2017) and used the construct for the transformation of *nsp2* mutants of *M. truncatula* cv. Jemalong.

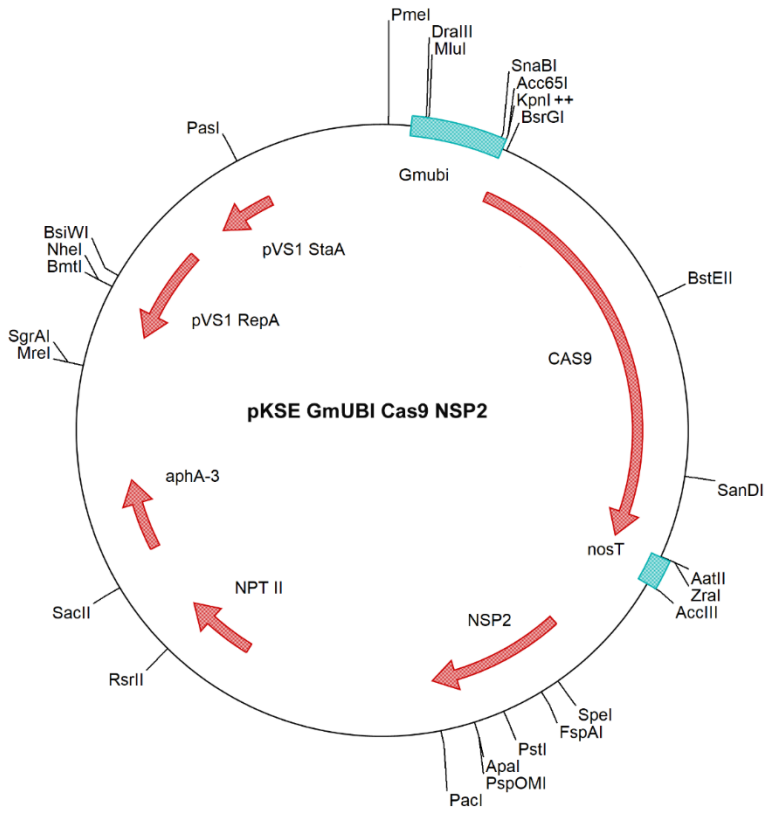


Figure 4.4. Schematic map of the vector carrying the *NSP2* reporter gene for CRISPR-Cas gene modification. The pKSE401 vector (Xing HL et al. 2014) was modified by replacing the *CaMV 35S* promoter driven, human codon optimized *Cas9* gene with a soybean codon optimized Cas9 for legume plant expression that is driven by the soybean ubiquitin (*GmUBI*) promoter. The *35S* promoter-*NSP2* cassette was introduced to complement the Nod$^-$ phenotype of the *nsp2* mutant as a reporter for transformation. At the PmeI site, it is possible to insert by In-Fusion ligation technology one or more gRNA elements for single or multiple gene modifications.

To detect the gene modification events obtained by the CRISPR/CAS9 technique, restriction enzyme digestion and/or PCR fragment sequencing were applied. The diseappearance of a restriction enzyme recognition site in close vicinity of the target site may indicate succesful editing, however, does not provide information about the nature of the mutation and the ratio of the mutant cells. Direct Sanger sequencing sequencing of the PCR products may show by the appearance of mixed peaks in the chromatograms where a mutation (insertion/deletion) starts but the extent of the mutation cannot be predicted and the presence of additional mutation(s) cannot be detected. Cloning of the PCR products and subsequent Sanger sequencing of multiple

clones from a single cloning allow the detection of multiple mutations and determination of their extent, but this approach is laborous and expensive. To circumvent these problems we developed an approach to detect editing events by next-generation sequencing (NGS): In the first PCR reaction, the edited region is amplified from the hairy root genome by using primers carrying gene-specific sequences in their 3' part and general adaptor (for example, M13 forward and reverse) sequences at the 5' end. In the second round of PCR, primers hybridizing to the general adaptor sequences on the amplification products of the first PCR and carrying barcode sequences to distinguish individual samples as well as sequencer-specific (for example, the Ion Torrent A and P) adaptors are used. After purification and quantification of the PCR products, they can be used directly in the sequencing reactions. In this way, we could detect multiple mutations in single hairy roots in agreement with earlier observations on the mosaic nature of edited hairy roots (Yang et al., 2017, Wang et al., 2017; 2018) and determine the sequnce and ratio of the mutant alleles using a few thousand of sequence reads per sample.

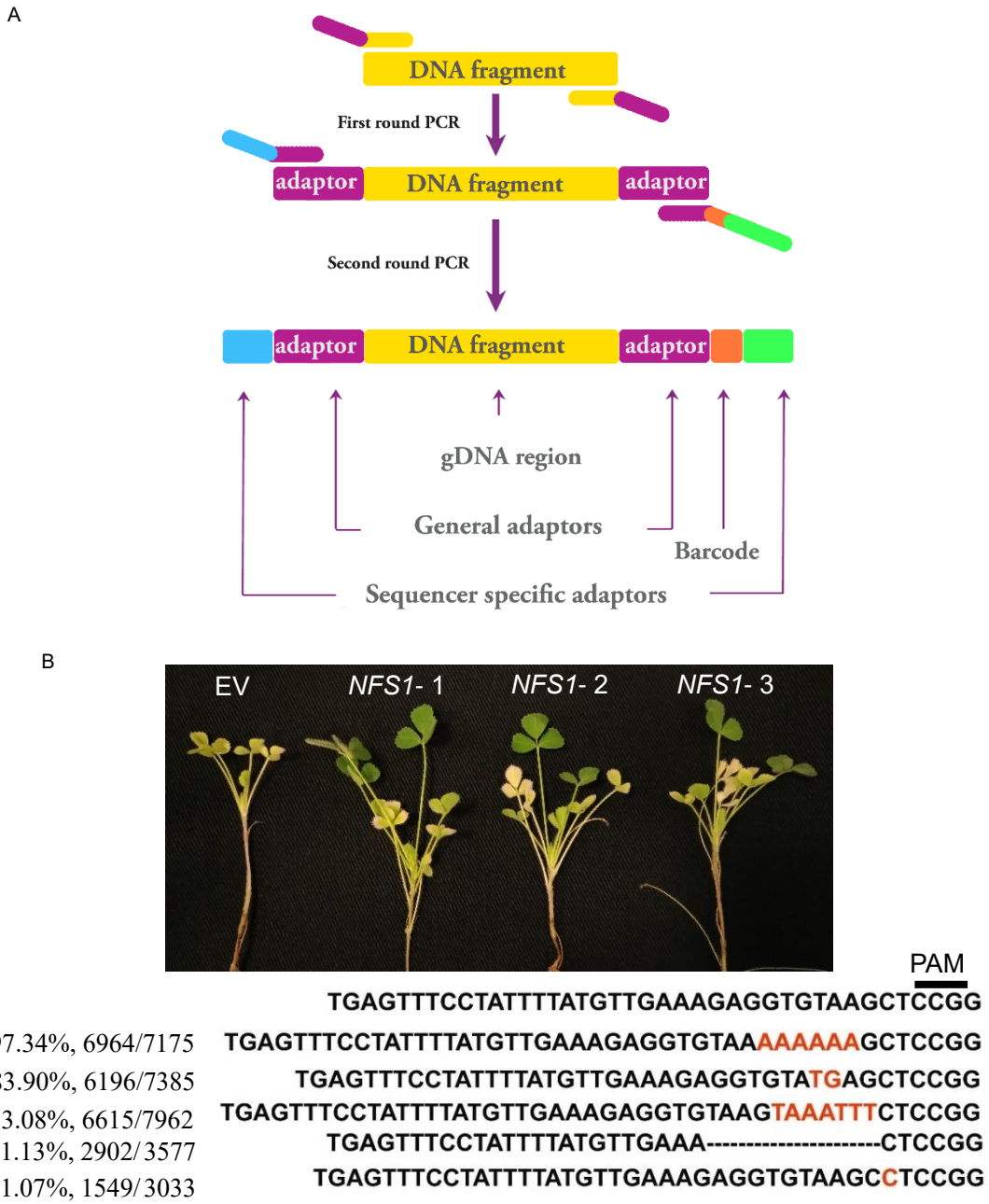


Figure 4.5. Next-generation sequencing based detection of gene editing events. A. Work flow of library construction for NGS analysis. First, the edited DNA region is amplified with gene specific primers that carry general adaptor sequences at their 5' end. In the second round of PCR, primers hybridizing to the general adaptor sequences on the amplification products of the first PCR and carrying barcode sequences to distinguish individual samples as well as sequencer-specific (for example, the Ion Torrent A and P) adaptors are used. Purified and the quantified PCR products can be used directly in the sequencing reactions. B. Representative results from *NFS1* knock-out plants and sequencing of the edited *NFS1* region from individual hairy roots. Few thousands reads are enough to determine the frequency of mutations even in mosaic roots, i.e. when the genotype/sequnce of cell populations are different.

4.2 Towards the identification of cis- and trans-acting regulatory elements of *NCR* genes

During the past 20 years, many transcription factors important for symbiotic nitrogen fixation have been identified and characterized. These transcription factors, including IPD3 (CYCLOPS), IPD3L, NIN, NSP1, NSP2, ERN1, ERN2, DELLA proteins and NF-YA1 are activated at the early stage of symbiosis and form a network responding to Nod factors that activates the infection process, nodule organogenesis and autoregulation of nodulation. In contrast, very little is known about the cis- (promoter) and trans-acting (transcription factors) elements that regulate those genes – like the *NCR* genes – that are (almost) exclusively expressed in the developing and/or functioning symbiotic cells and contribute to the differentiation of both prokaryotic and eukaryotic cells as well as the reduction and assimilation of nitrogen.

To identify those cis- and trans-acting elements that are important for the expression of *NCR169* and *NCR211* genes essential for bacteroid and nodule development, we initiated the study applying multiple approaches.

4.2.1 Activation of the *NCR169* promoter in soybean nodule

Despite the big differences in their organogenesis (maintained vs. ceased meristematic activity), structure (developmental gradient of cells vs. synchronized cell population) and genetic repertoire (presence vs. absence of *NCR* genes), the nodules of *Medicago* and *Glycine* (soybean) or *Lotus* (bird's-foot trefoil) plants have a number of common characteristics such as nodule cell differentiation via endoreduplication or expression of high number of genes with very specific and restricted (to nodule) pattern. To investigate whether the cis-acting elements of *NCR* genes are recognized by *Glycine* transcription factors, *NCR* promoter - *GUS* reporter constructs – exampled by the *NCR169* gene – were introduced into this legume with the help of *A. rhizogenes* mediated hairy root transformation.

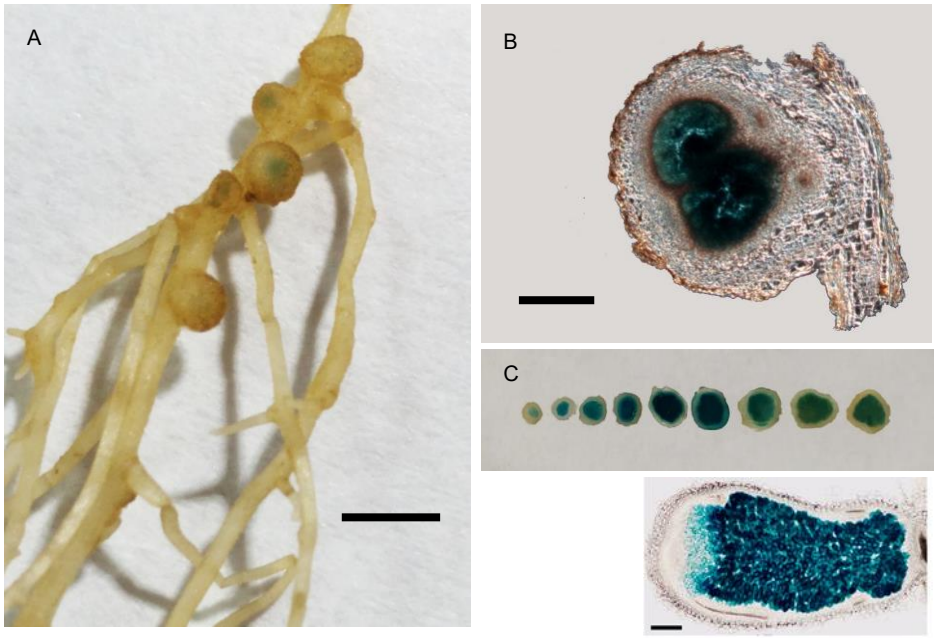


Figure 4.6. *NCR169* promoter activation in soybean nodule demonstrated by GUS staining. A. Hairy root and B. nodule showing GUS activity in the central nitrogen-fixing cells obtained after transforming soybean with vector pKM43eGFP carrying the *NCR169* promoter driven GUS reporter gene. C. GUS staining of transgenic soybean nodules at different development stages (***u***p). In the *Medicago* nodule (***below***), the gene is activated in the infected cells of the interzone and of the mature nitrogen fixation zone. Transgenic roots and nodules were first selected based on GFP fluorescence and stained with X-Gluc solution. No GUS activity could be detected in the empty vector control (not shown).

The emerged hairy roots were inculated with rhizobium strain *Bradyrhizobium japonicum* CB1809. GUS activity in roots and nodules was checked at different stages of nodule development. Surprisingly, although soybean does not have *NCR* genes, it could activate the *Medicago NCR169* promoters and this activation was also restricted to the infected cells of the nodules (Fig. 4.6). Moreover, the expression pattern in soybean showed by GUS staining is quite similar with the one in *Medicago truncatula*: the expression is initiated in the endoreduplicating cells and reaches its maximum in the nitrogen fixing cells. This result implicates that a(the) transcription factor(s) in the soybean nodule can bind to the cis-acting element(s) of the *NCR169* promoter.

4.2.2 *cis*-element prediction on *NCR169* promoter

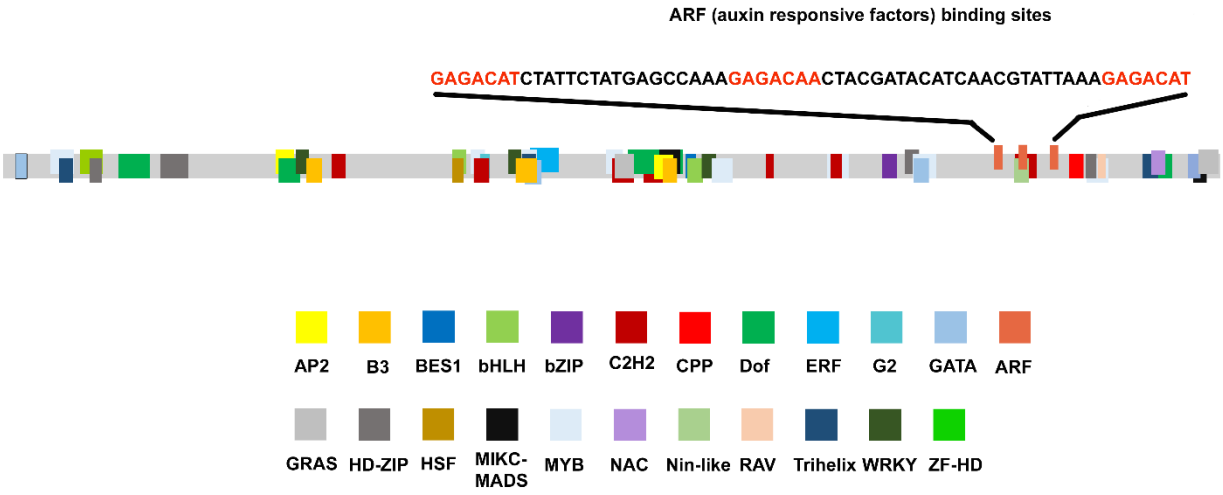


Figure 4.7. Schematic diagram of elements presenting on the 1187 bp promoter region of *NCR169* gene. Promoter sequence was analyzed on the website of PlantRegMap: http://plantregmap.cbi.pku.edu.cn/binding_site_prediction.php for TF binding sites prediction. Different cis elements are presented with corresponding colors. Region containing ARF (auxin responsive factor) binding site is emphasized with the binding sites marked in red letters.

By analyzing the promoter region of *NCR169* for putative cis-elments on PlantRegMap (http://plantregmap.cbi.pku.edu.cn/binding_site_prediction.php), many stress responding elements were found (Fig. 4.7), such as bZIP, ERF, NAC, and WRKY, which indicated that the *NCR169* gene might be induced by stress treatment. Inside nodules, under symbiotic conditions, plant cells might be challenged by stress conditions like unphysical pH, salinity, osmotic changes and altered, i.e. decreased oxygen level because symbiotic nitrogen fixation takes place in a strictly micro-aerobic environment that have to be maintained to ensure the functioning of the oxygen-sensitive nitrogenase and to satisfy the need for the high rate of respiration of bacteria. However, there was no other experimental condition reported where the *NCR* gene expression was induced than the nodule infected by rhizobia, it can be speculated that either the condition applied is not strict enough as in the nodule, or

other factors participate in the regulation of gene expression, like auxin, as auxin responsive elements can also be found in the promoter region of *NCR169*. This bioinformatic analysis for cis-acting elements offers the information for potential binding sites of transcription factors and can be used as a reference for probe design and for analyzing experimental results from other assays.

4.2.3 DNA pull-down assay with *NCR169* promoter

DNA pull-down assay is a widely used method to study DNA-protein interactions. By incubating protein samples with a DNA fragment of interest, proteins that can interact with sequences in this DNA will bind to it and thus can be isolated for further analysis and identification. We implemented this method to identify those DNA binding proteins from the nuclei of nodules that can interact with the *NCR169* promoter. As this promoter can be activated both in *Medicago* and *Glycine* nodules, protein extracts were used from both type of nodules with the hope that orthologous proteins will be identified. A 1187 bp part of the *NCR169* promoter upstream of the start codon was chosen and diveded into three overlapping fragments were used as baits (see Materials and Methods).

After pull-down assay, proteins bound to the probes were eluted and analyzed with HPLC/MS/MS. Mass spectrum data were analyzed on the UniProt website (https://www.uniprot.org/). Unknown proteins and proteins without any annotation were analyzed for functional information using protein BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Data from two sets of DNA pull-down experiments were combined and detected proteins from soybean and *Medicago truncatula* were compared, probable homologues were searched for and identified through BLAST. It should be noted that the DNA baits were not cross-linked with the interacting proteins, thus, besides sequence non-specific and sequence specific DNA binding proteins such as histones and transcription factors, respectively, as well as co-factors binding to the interacting proteins, lots of false positive hits (for example, abundant proteins that could not be removed) were all identified by the mass

spectrometry. It made the analysis on the result difficult, but also made it possible to identify proteins of DNA-binding activity and others, which – acting as partners of transcription factors and gene expression regulators – do not bind directly to the promoter region.

The first category we looked for was transcription factors or proteins predicted to have transcription factor activity (Table 4.2) and a number of such proteins was found by the DNA pull-down experiment. Among them are AT hook motif proteins, MYB family proteins, WD-40 repeat containing proteins, zinc finger CCCH domain proteins, storekeeper proteins, NF-Y transcription factor, transcription factor Pur-alpha-like protein and THO proteins etc. These proteins have either not been studied or not been reported as transcription factors involved in nodule development and functioning, however, they showed up in both soybean and *Medicago* samples. There are also some other transcription factor proteins that were not present in both samples but also can be promising as gene regulators of *NCR169*. These proteins include transcription factor jumonji (JmjC) domain protein from the *Medicago* sample and transcription factor IIIA-like, trihelix transcription factor, SHOOT2, and AP2/B3-like transcriptional factor family protein from the soybean sample.

The second category of proteins identified were the sequence non-specific DNA binding proteins and other nuclear components like histones, DNA-directed RNA polymerase or RNA splicing factors. These proteins are abundant nuclear proteins, some of them are component of chromatin, while others are important for DNA structure maintenance, DNA replication and transcription, as well as for the processing and export of mRNAs out from the nucleus. Some other proteins responsible for DNA epigenetic modification, like methyltransferase, acetylesterase and histone deacetylase, were also identified and listed in this group.

The third category contains proteins that were reported to be co-factors (or corepressor) contributing to transcription regulation: A number of such proteins showed up in both samples including HMG proteins, topless protein, TOM protein, clathrin, calnexin, calmodulin, DEK protein and GRF (General Regulator Factor)

protein. Their putative function is to change the structure of DNA to help the binding of TFs to DNA (HMG proteins, calnexin and DEK proteins), to modify the conformation of TF proteins (chaperones), to activate or deactivate TFs by phosphorylation or dephosphorylation (for example, calmodulin proteins). Some of these proteins can directly bind to DNA probes while the others were identified in the pull-down experiments because of their interaction with DNA binding proteins and all of them might be subunits of transcription complexes, which control the expression of the *NCR169* gene and compose the gene regulation network.

Table 4.2. List of *Medicago truncatula* proteins identified in the DNA pull-down experiments that are predicted to be transcription factors or can potentially interact with transcription factors. Binding of the proteins to probe 1, probe 2 and probe 3 is indicated on the right.

Access ID in UniProt	Protein annotation	Probe binding		
		1	2	3
A0A072VXL7	AT hook motif DNA-binding family protein	-	+	+
G7JK75	AT hook motif DNA-binding family protein	-	+	-
A0A072TPB2	AT hook motif DNA-binding family protein	-	+	-
G7JGG2	AT-hook motif nuclear-localized protein	+	+	+
G7KSI4	AT-hook motif nuclear-localized protein	+	+	+
G7K181	High mobility group (HMG)-box protein	-	-	+
G7KS90	High mobility group (HMG)-box protein	+	+	+
A0A072TSA9	High mobility group (HMG)-box protein	+	+	+
Q1SN01	H15 domain protein	-	+	+
A0A072VR11	Jumonji (JmjC) domain protein	+	+	-
G7IS32	MYB transcription factor MYB91	+	-	-
G7KSV7	Myb/SANT-like domain protein	+	+	+
A0A072U4N3	Myb/SANT-like domain protein	+	+	+
A0A072VDQ0	PHD finger alfin-like protein	-	+	+
A0A072UGM4	PHD finger alfin-like protein	+	+	+
A0A072UTR5	PHD finger alfin-like protein	+	+	-
Q2HU65	Probable histone H2A.2	+	+	+
Q1S9I9	Probable histone H2B.1	+	+	+
G7IK18	STOREKEEPER protein	+	+	+
G7JHJ7	Transcription factor Pur-alpha-like protein	+	+	+
G7J609	Zinc finger CCCH domain protein	+	-	+
A0A072VIQ3	Zinc finger CCCH type family protein	+	+	-

Table 4.3. List of soybean proteins identified in the DNA pull-down experiments that are predicted to be transcription factors or can potentially interact with transcription factors. Binding of the proteins to probe 1, probe 2 and probe 3 is indicated on the right.

Access ID in UniProt	Protein annotation	Probe binding		
		1	2	3
C6TMY6	AT-hook motif nuclear-localized protein	+	+	+
I1K793	AT-hook motif nuclear-localized protein	+	+	+
I1J9J5	AT-hook motif nuclear-localized protein	+	-	-
C6TF47	AT-hook motif nuclear-localized protein 1	-	+	-
A0A0R0F4F8	AT-hook motif nuclear-localized protein 14	+	+	+
C6TMY4	AT-hook motif nuclear-localized protein 23-like	-	-	+
I1NIC3	AT-hook motif nuclear-localized protein 6	+	+	-
A0A0R0FSY5	AT-hook motif nuclear-localized protein 7	+	+	-
A5JUQ4	C2H2 zinc finger protein	+	-	-
A0A0R0J669	CCR4-NOT transcription complex subunit 1	-	-	+
I1JSU0	HMG box domain-containing protein	+	+	+
Q93YH8	HMG I/Y like protein	+	+	+
Q10370	HMG-Y-related protein B	+	+	+
A0A0R0KFD0	HMG box domain-containing protein	+	+	+
I1MBW3	H15 domain protein	+	+	+
K7L454	H15 domain protein	+	+	+
A0A0K2CT26	Homeodomain/HOMEOBOX transcription factor	-	+	-
I1LFM1	MYB-like domain protein	+	+	+
I1N4Y2	MYB-like domain protein	+	+	+
C6T7X8	PHD finger protein ALFIN-LIKE 2	+	+	+
C6TCB0	PHD finger protein ALFIN-LIKE 2	+	+	+
Q06A73	PHD1	-	+	-
Q06A76	PHD4	+	+	-
C6TI90	SHOOT2, MYB domain protein	-	+	+
K7M1T5	STOREKEEPER protein	+	+	-
I1KMC5	Transcription factor Pur-alpha 1	+	+	+
A0A0R0FEQ4	Trihelix transcription factor ASIL2-like	+	+	-
K7M034	Zinc finger BED domain protein DAYSLEEPER	+	+	+
C6TA41	Zinc finger CCCH domain-containing protein 14	+	+	-
I1KRM9	zinc finger CCCH domain-containing protein 17	+	-	-

As shown in Tables 4.2 and 4.3, we identified many potential transcription regulators targeted to the *NCR169* promoter, which might help understand the mechanism of *NCR* gene expression regulation. We could pull-down with the DNA

probes a number of homologous proteins from the nuclear protein extracts of soybean and *Medicago truncatula* nodules. This observation and the ability of soybean to correctly express *NCR* genes in its nodules indicate that although soybean contains and expresses no *NCR* gene, a gene expression regulatory network conserved in soybean and *Medicago* nodules might exist and controls the expression of *NCR* genes (besides other nodulins). It might mean also that during evolution *NCR* genes might be simply added to the the existing strictly nodule-specific gene repertoire without the invention of a specific regulatory network for them.

4.2.4 EMSA screening for gene expression regulators with *NCR169* promoter as probes

Parallel to the DNA pull-down experiments, we also tried to identify proteins that bind to the *NCR169* promoter based on their ability to cause electrophoretic mobility shift of bait fragments. For this purpose, an ~400 bp minimal promoter of the *NCR169* gene, which provides sufficient expression to complement the Fix^- phenotype of the *dnf-7* mutant (unpublished data from Kaló Péter's group), was amplified as ~100 bp long fragments with ~50 bp tilings to avoid the disruption of potential TF recognition sequences at the fragment ends (Fig. 4.8). In the experiments, the same soybean and *Medicago* nuclear protein extracts were added to the probes that were used in the DNA pull-down assays. After the binding reaction, samples were loaded into native polyacrylamide gels and signals were collected by detecting the radioactivity on an X-ray film.

For probes A, B, C, F and G, retarted bands could be observed (Fig. 4.8 B) with both *Medicago* and soybean nodule cell nuclear protein extracts. These results indicate that there might be at least three DNA sequences (unique, 5' part of fragment A; overlap of fragments B and F; and overlap of fragments C and G) where protein binding takes place. The third candidate, the overlap of fragments C and G, synthetized as fragment H was shown to be shifted by the nuclear extracts, thus proving the presence of a protein binding site on it. By aligning these EMSA assay

results with the cis-element prediction on the 1187 bp promoter region (Fig 4.7), the members of the following transcription factor families might bind to the *NCR169* promoter: a MYB family protein and a C2H2 transcription factor might bind to the 5' part of fragment A; the overlap region of fragments B and F might contain the binding site(s) of a MYB protein and a GATA protein; fragment H might interact with CPP, HD-ZIP and MYB proteins. To confirm these predictions, the shifted bands have to be isolated from gel and the associated proteins should be detected and identified by MS analysis.

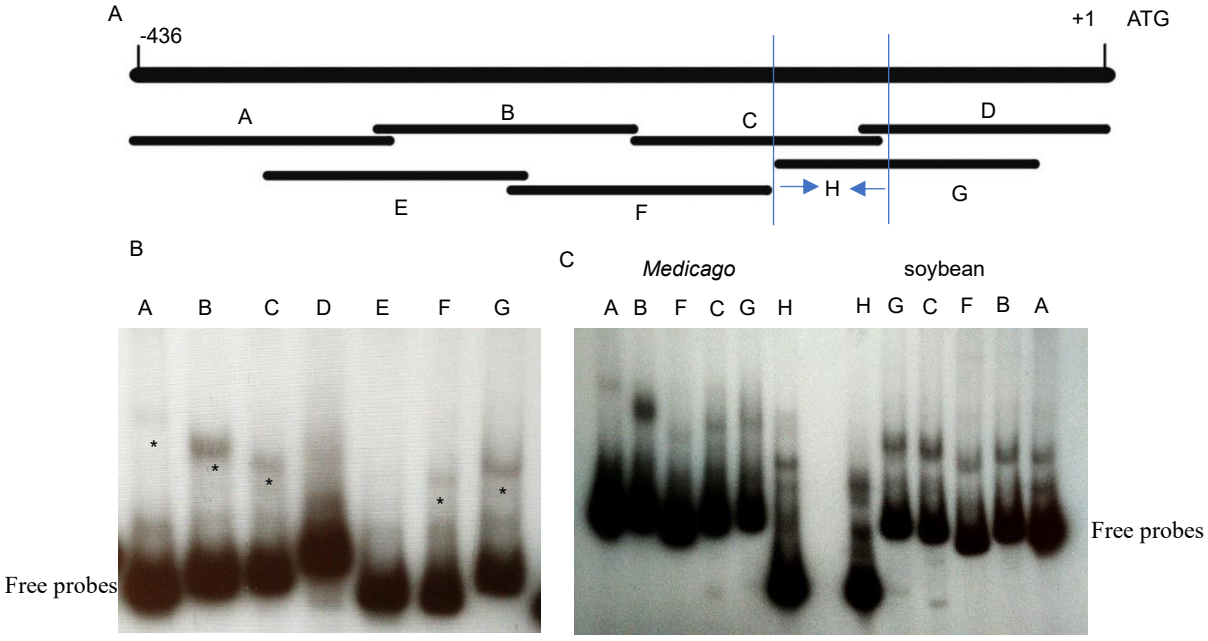


Figure 4.8. EMSA with 436 bp of the *NCR169* promoter region. A. 7 overlapping probes (A, B, C, D, E, F and G, ~100 bp each in length) were designed to cover the *NCR169* gene promoter region 436 bp upstream of the translational start codon. Probe H covers the 72 bp overlap between probes C and G. **B**. EMSA with the 7 probes shown in panel A using protein extract from *Medicago* nodule cell nuclei. Shifted bands are marked with asterisks. **C**. Probes shifted in panel B as well as probe H were re-tested with both *Medicago* (left) and soybean (right) nuclear extracts.

The high sensitivity of radioactive isotope lablling makes it possible to visualize trace amount of DNA in the gel. By optimizing the conditions, like gel size, comb type, running time, and the amount of protein extract loaded, we demonstrated that with a given DNA fragment of interest, it is possible to find interacting proteins

through EMSA assay from nuclear protein extract. However, it was a pity that in the host institute, it is forbidden to analyze protein samples carrying radioactivity, so we cannot provide further information on what kind of interacting protein we found for this study. To circumvent this restriction, we try to optimize the method for non-radioactively labeled (for example, biotinylated) DNA fragments.

4.2.5 Yeast one-hybrid screen to identify proteins interacting with the *NCR169* and *NCR211* promoters

As a straightforward approach for investigating the interaction between DNA and gene expression regulators, especially transcription factors, the yeast one hybrid technique has been widely used. The interaction can be detected by the activation of reporter gene(s), for example, the yeast *HIS3* gene, that allows the growth of a mutant yeast strain on restrictive media in the case of interaction. For this purpose, the DNA fragment of interest, i.e. the investigated promoter is cloned upstream of one (sometimes more) reporter gene(s) and the (library of) protein(s) of interest is/are fused to the activation domain (AD) of the yeast transcription factor GAL4 and expressed constitutively from the yeast *ADH1* promoter.

In this study, we used a *Medicago truncatula* nodule cDNA library (Györgyey et al. 2000) for Y1H screening in order to identify gene expression regulators for *NCR169* and *NCR211*. By working together with our collaborators (Xia Li's group, from Huazhong Agriculture University, China), we could manage to conduct Y1H screening also with a soybean nodule cDNA library, from which we expected to acquire information on the shared common transcription factor(s) that induce the expression of the *NCR169* gene in the soybean nodule.

According to our preliminary results, like in the DNA pull-down experiments, most of the identified proteins are non-specific DNA binding proteins, such as histones, linker histones, DNA methyladenine glycosylase, methyltransferase or HMG proteins. We identified some proteins that are transcription factors or predicted to possess transcription factor activity including a AT hook protein (Medtr4g098450),

TCP transicription factor (Medtr7g028160), transcription factor VOZ1-like protein (Medtr4g088125), BHLH domain class transcription factor (Medtr4g081370), and MYB family transcription factor (Medtr5g037080). Being at the early phase of this approach, i.e. the Y1H work on *Medicago* and soybean nodule cDNA libraries is still ongoing, so far only one AT hook protein (Medtr4g098450) was found till now to be identified by both DNA pull-down and Y1H screening. Further investigation on this protein have been initiated.

Since the gene annotation on *Medicago truncatula* and soybean genes is not very complete, in Y1H screening, a lot of proteins with unknown function and hypothetical proteins were found similarly to the DNA pull-down results. Among them are many Nodulins (MtN25 from *Medicago* and early nodulin 93 with several other nodulins in soybean). For those proteins, further analysis needed to be done to confirm their binding to the probes applying other approaches. Those potential transcription factors will be further tested to confirm their correlation with *NCR* gene expression.

Table 4.4. List of *Medicago* proteins from yeast one hybrid screening with *NCR169* promoter and *Medicago truncatula* cDNA library. Plasmids from positive yeast colonies were amplified in E.coli and their inserts were sequenced. BLAST analysis was conducted on Phytozome (https://phytozome.jgi.doe.gov/) to identify the gene.

Number	Gene ID	Gene annotation
1	Medtr4g098450	AT hook motif DNA-binding family protein
2	Medtr6g079520	Linker histone H1 and h5 family protein
3	Medtr4g081370	bHLH domain class transcription factor
4	Medtr5g037080	MYB-like transcription factor family protein
5	Medtr7g028160	TCP family transcription factor
6	Medtr4g088125	Transcription factor VOZ1-like protein

4.2.6 Bioinformatic prediction of *NCR* gene regulators and library construction

The expression of *NCR*s is extremely nodule-specific, and most of them are considered as late nodulation genes. *NCR* gene transcripts are mainly detected in the interzone and zone III of *Medicago truncatula* nodules based on laser dissection-RNAseq analysis. Taking advantage of *Medicago truncatula* Gene Expression Atlas and plant transcription factor databases (MtGEA and PlantTFDB) and searching for those transcription factors that have a matched expression pattern with *NCR* genes, potentially a library of TFs that might regulate *NCR* genes can be constructed. The detailed expression pattern of the genes selected this way were further analyzed based on the RNA-seq results obtained after laser-capture microdissection of *Medicago truncatula* nodules (Roux et al., 2014). Using this strategy, 150 genes belonging to clusters 6, 9 and 12, which showed high expression level in the interzone and zone III of the nodule, while lowly expressed in zone I and zone II, were chosen. From the coding sequences of these genes and selected proteins identified in the DNA pull-down assay, a library was constructed into a Gateway entry vector. The individual clones of the library will be transferred into a bacterial protein expression system as well as into a destination vector that can replicate in yeast. The proteins produced in bacteria will be tested in EMSA experiments while the clones in the yeast vector will be used for one-hybrid investigations.

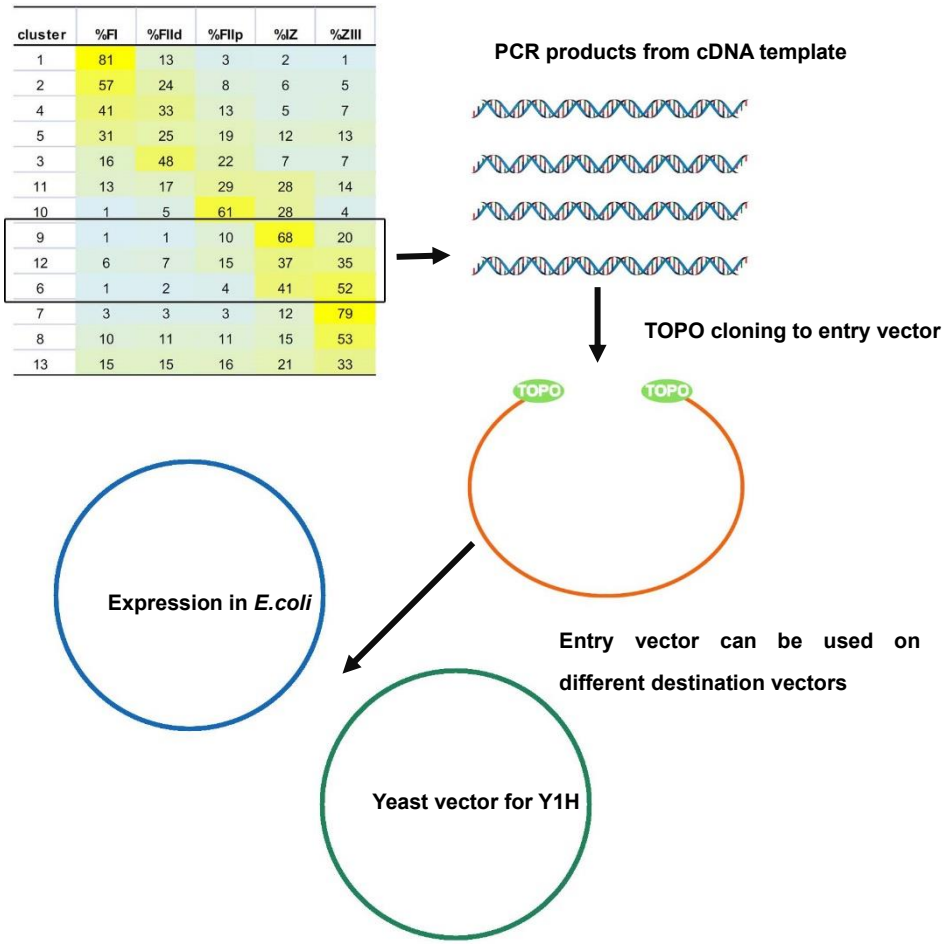


Figure 4.9. Schematic diagram of work flow for the selection and construction of potential *NCR* gene expression regulators for bacterial protein expression and Y1H. Predicted transcription factors were analyzed for their expression pattern in different zones of *Medicago truncatula* based on the RNA-seq data published by Roux et al. (2014) and the ones with high expression level in inter zone and zone III while lowly expressed in zone I and zone II were selected, mainly from the clusters 6, 9 and 12. Primers with adaptors suitable for TOPO cloning were designed to amplify these genes using *Medicago truncatula* nodule cDNA as template. PCR products will be cloned into Gateway entry vector via TOPO cloning, and then transferred into destination vectors to be used for the expression of proteins in a bacterial system (expressed protein can be applied to EMSA experiment to validify their interaction with DNA region of interest) or to investigate DNA-protein interactions using the Y1H system.

5. Conclusion

In this study, we developed two novel reporter systems for hairy root transformation in *Medicago truncatula*: One of them uses anthocyanin as reporter. The over-expression or vascular tissue specific expression of a MYB-R2R3 family transcription factor, *MtLAP1*, transgenic hairy roots and/or nodules results in visible purple color from anthocyanin accumulation and, thus, the transgenic tissues can be identified by naked eyes. The second one is based on the NSP2 transcription factor, which is essential for the initiation of nodule development, and on *nsp2* mutant plants. When hairy root transformation is conducted on the Nod^- *nsp2* mutant plants, all the formed nodules are transgenic and their as well as the plants' phenotype depends on the nodule genotype, for example, after CRISPR/CAS9 gene editing. These two systems can significantly reduce the labor in transgenic root detection and the second system also provides a mean to annihilate the contribution of non-transgenic tissues to the symbiotic phenotype and vigor of the plant.

Through searching for transcription factors (in this study, we focus on *NCR169* and *NCR211*), we are trying to decode the gene regulation network of *NCR* genes. The finding that *NCR169* promoter can get activated in soybean nodule indicates that the nodules of soybean, in which *NCR* genes are not present, share similar transcription activation factor(s) with *Mediccago truncatula*. Durning the screening for transcription factors, we did find some conserved regulators that are responsible for *NCR169* gene activation in both species. Our result and ongoing works may not only shed light on how IRLC legumes control the massive expression of *NCR* genes, but also provide informations on the conserved regulatory network in non-IRLC legumes.

Summary

Plant growth and development rely on the accessibility of nutritional elements, such as carbon (C), hydrogen (H), oxygen (O), nitrogen (N), phosphorus (P), sulfur (S) from the environment. Carbon and oxygen supply are not limited for plants, however, the acquirement of other elements, especially nitrogen can be difficult. Nitrogen is essential for the plant to synthesize amino acids, which are the basic units of proteins and peptides. On the other hand, the situation for rhizobia is the opposite, the nitrogenase expressed by rhizobia can fix nitrogen from the atmosphere but carbon source limits the growth of them in soil. This complementarity in element acquisition provides the foundation for legume plants and rhizobia to establish symbiosis in which they can exchange carbon and nitrogen with each other.

Legume plants develop root nodules to accommodate their rhizobia symbionts. Root nodule not just acts as the place where exchange of nitrogen and carbon source happens, but also provides the suitable cellular condition for the rhizobia to fix nitrogen. The nitrogenase enzyme complex catalyzing the reduction of atmospheric nitrogen is extremely sensitive to free oxygen (Shah & Brill 1977). The nodule tissues provide and maintain a low free oxygen concentration via physical oxygen barriers and the expression of high amount of oxygen-binding proteins, leghemoglobins.

Root nodules on different plant species differ from each other in their shape depending on plant species and can be grouped into two major types: determinate and indeterminate (Franssen et al. 1992; Maunoury et al. 2008). Meristematic cells in the determinate nodules on legumes like soybean and *Lotus japonicus* are not persistent, giving them a round shape with no zonation inside. The indeterminate nodules on legumes like *Medicago* have an elongated or cylindrical shape and nodule can be divided into different developmental zones because of the presence of a persistent meristem region (Sutton 1983). In a mature nodule, five distinct zones can be observed: the meristem (ZI), the invasion zone (ZII), the interzone (IZ), the nitrogen-fixing zone (ZIII), and in older nodules, the senescence zone (ZIV).

The fate for the nitrogen fixing bacteria in nodule cells can also be divided into two types: reversible and irreversible/terminal differentiation. With the use of nearly isogenic rhizobial strains on different hosts, Mergaert et al. (2006) proved that the differentiation of bacteroids is under the control of the plant host. In IRLC legumes, like *Medicago*, NCR peptides as well as GRPs have been proven to be responsible for the terminal differentiation. In *Medicago truncatula*, more than 700 *NCR* genes spread on all eight chromosomes were identified while no such sequence could be found in the genome of *Lotus* and *Glycine* (Mergaert et al. 2003). Such high number of genes in a gene family suggests functional redundancy but there are examples that the loss of a single *NCR* like *NCR169* and *NCR211* in the *dnf7* and in the *dnf4* mutants, respectively, results in the arrest of symbiotic nodule development (Horvath et al. 2015; Kim et al. 2015). This observation provokes the idea to search for the other essential *NCR*s, which will help us gain knowledge on which types of *NCR*s are essential and understand why plant needs to produce so many of them.

The main difficulty in studying *NCR*s is that there are too many of them: if we want to identify other essential *NCR* genes, a very large-scale gene knock-out experiment targeting all *NCR* genes one by one will be needed. We started this project with the optimization of the hairy root transformation system we use, mainly by developing new reporters.

Hairy root transformation is a widely used method for studying the molecular biology of the interaction of beneficial (rhizobia, mycorrhiza fungi) and detrimental (plant pathogenic bacteria and fungi) microbes with roots in legume plants. But the commonly used reporters in hairy root system, such as GFP and GUS, are not suitable for large scale analysis, because in these systems, fluorescent microscope or chemical treatment is needed to visualize the reporter signal, which take a lot of labor and time. In my study, I developed two new reporter systems that are based on the accumulation of the purple colored anthocyanins and the complementation of the non-nodulating *nsp2* mutant, respectively. In the anthocyanin system, the over-expression or vascular tissue specific expression of the *MtLAP1* gene in transgenic hairy roots and/or nodules results in purple coloration from anthocyanin accumulation and, thus, the transgenic tissues can be identified by naked eyes. In the case of the *NSP2* reporter,

which codes for a transcription factor essential for the initiation of nodule development, hairy root transformation is performed on the Nod$^-$ *nsp2* mutant plants. As a result, all the formed nodules are transgenic and their as well as the plants' phenotype depends on the nodule genotype, for example, after CRISPR/CAS9 gene editing. These two systems can significantly reduce the labor in transgenic root detection and the second system also provides a mean to annihilate the contribution of non-transgenic tissues to the symbiotic phenotype and vigor of the plant.

Beside developing these two reporter systems, the regulation of *NCR* gene expression also drew my research interest. The expression of the more than 700 *NCR* genes in *Medicago truncatula* are extremely specific, their transcription is restricted to the infected symbiotic cells of root nodules while all the other experimental conditions tested cannot induce their activity. It would be quite meaningful to explore how the strict expression pattern of *NCR* genes is achieved. Previously, it was suspected that the *NCR* genes are controlled by IRLC legume specific transcription factors, or even *NCR* specific ones. However, by testing *NCR169* gene expression with the help of a promoter-GUS fusion in soybean, I found that the *NCR169* gene promoter is active in soybean nodule, where no *NCR* gene is expressed, meaning that soybean and *Medicago truncatula* share common transcription activator(s) of the *NCR169* gene. This result also indicated that during evolution, plant did not acquire *NCR* genes through *de novo* process. To look for the supposed conserved transcription regulators, I combined the DNA pull-down, Y1H screening and EMSA techniques and used materials and cDNA libraries from both *Medicago* and soybean nodules. Those potential *NCR169* promoter interactors were selected for further work that were identified in both the DNA pull-down and Y1H screenings of both the *Medicago* and the soybean samples and were co-expressed with *NCR169* in the interzone and the nitrogen fixing zone of the nodule. The combined screening strategy provided us several interesting candidates that are under investigation.

Összefoglalás

A növények növekedése és fejlődése nagymértékben függ a környezetben található és a hozzáférhető elemek jelenlététől, úgymint a széntől (C), a hidrogéntől (H), oxigéntől (O), nitrogéntől (N), foszfortól (P) és kéntől (S). A szén és az oxigén mennyisége nem limitáló tényező a növény számára, de más elemek, elsősorban a nitrogén, igen. A nitrogén elengedhetetlen a nukleotidok és az aminosavak szintéziséhez, amelyek az alap építőelemei a nukleinsavaknak illetve a fehérjéknek/peptideknek. A rhizóbiumok számára fordított a helyzet, hiszen a baktérium a nitrogenáz segítségével képes a légköri nitrogén megkötésére, de a talajban található szén alacsony mennyisége gyakran a növekedést gátló tényező. Az elemek megszerzésének ezen egymást kiegészítő volta biztosítja az alapot a pillangósvirágú növények és a rhizóbium baktériumok által kialakított szimbiózishoz, ami lehetővé teszi a szén és a nitrogén kölcsönös megosztását.

A pillangósvirágú növények gyökérgümőket hoznak létre, ahol a rhizóbiumok megtelepedhetnek. A gümő nem csupán lehetővé teszi a nitrogén és szén cseréjét, de biztosítja a megfelelő sejten belüli környezet is a nitrogén megkötésére. A légköri nitrogén redukálását végző nitrogenáz enzim komplex rendkívül érzékeny az oxigén jelenlétére (Shah és Brill, 1977). A gümő szövetei azonban biztosítanak és fenntartanak egy alacsony oxigén-szintet különböző fizikai gátak létrehozásával és nagy mennyiségű oxigén-kötő fehérje, leghemoglobin, termelésével.

A különböző növényfajokon található gümők alakja jellemző az adott fajra és két fő típusba oszthatók: determinált és indeterminált gümők (Franssen és mtsai., 1992; Maunoury és mtsai., 2008). A szójababon és a *Lotus japonicus*-on található determinált gümők merisztémája egy idő után eltűnik, miáltal a gümő kerekded formájú lesz, fejlődési zónák nem alakulnak ki. A pl. a *Medicago* fajokon fejlődő indeterminált gümő hosszúkás, hengeres alakú és különböző fejlődési zónákra osztható fel az aktív merisztematikus régió jelenléte miatt (Sutton, 1983). Az érett gümőben öt egyértelmű zóna különíthető el: a merisztematikus zóna (ZI), az inváziós zóna (ZII), az interzóna (IZ), a nitrogén-kötő zóna (ZIII) és az idősebb

gümőkben az öregedési zóna (ZIV).

A gümőben a nitrogénkötő baktérium sorsa is két típusba sorolható: a bakteroidok vagy reverzibilisen, vagy irreverzibilisen/terminálisan differenciálódnak. Közel izogén rhizóbiumok használatával Mergaert és munkatársai (2006) bizonyították, hogy a bakteroid differenciációját a gazdanövény irányítja. Az IRLC hüvelyesek, mint a *Medicago*, esetében ezért a terminális bakteroid differenciációért a gümőspecifikus ciszteinben gazdag (NCR) és a glicinben gazdag (GRP) peptidek tehetők felelőssé. A *Medicago truncatula* több mint 700 *NCR* génje a növény mind a nyolc kromoszómáján szétszórva helyezkedik el, míg a *Lotus* és a *Glycine* genomban egy ilyen szekvencia sem található (Mergaert és mtsai., 2003). Ilyen sok génnel rendelkező géncsaládokra jellemző a magas fokú redundancia, de az *NCR*-ek esetében bizonyított, hogy csupán egy gén elvesztése, mint az *NCR169* a *dnf7* és az *NCR211* a *dnf4* mutánsok esetében a gümőfejlődés és szimbiózis megakadásához vezethet (Horváth és mtsai., 2015, Kim és mtsai., 2015). Ez a megfigyelés adta az ötletet, hogy további esszenciális *NCR*-eket kutassunk fel, ami megmutathatja milyen típusú *NCR*-ek elengedhetetlenek a gümőfejlődéshez és talán azt is megérthetjük, hogy a növény miért termel belőlük ennyifélét.

A legnagyobb probléma az *NCR*-ek tanulmányozásában az, hogy túl sok van belőlük: ha további elengedhetetlenül fontos *NCR* géneket szeretnénk találni, ahhoz egy rendkívül nagymértékű géninaktiválási kísérlet elvégzése lenne szükséges, ahol az összes *NCR* gént egyesével kéne kiütni. Meg is kezdtük ezt a kutatást az általunk használt „hairy root" transzformációs rendszer optimalizálásával, elsősorban új jelzőrendszerek kifejlesztésével.

A „hairy root" transzformáció széleskörűen használt módszer a hasznos (rhizóbiumok, mikorrhizális gombák), illetve a káros (növénypatogén baktériumok és gombák) mikrobák és a pillangósvirágú növények gyökere közti kapcsolat tanulmányozására. Az általánosan használt jelzőrendszerek azonban, mint a GFP és a GUS, nem a legalkalmasabbak nagyléptékű vizsgálatokra, mivel ezeknél fluoreszcens mikroszkópra vagy kémiai kezelésre van szükség a jel érzékeléséhez, ami munka- és időigényes. Munkánk során kifejlesztettünk két új jelzőrendszert, ami lila színű antocianinok felhalmozódásán, illetve a gümőképzésre képtelen *nsp2* mutáns növény komplementálásán alapul. Az antocianin alapú rendszernél a

túltermeltetett vagy a specifikusan az edénynyalábokban kifejeztetett *MtLAP1* gén hatására a transzgenikus gyökerek és/vagy gümők esetében lila elszíneződés figyelhető meg az antocianinok felhalmozódása miatt és így a transzgenikus szövetek szabad szemmel is felismerhetőek. Az *NSP2* jelzőrendszer esetében, ahol a gén által kódolt transzkripciós faktor elengedhetetlen a gümőfejlődés elindulásához, a „hairy root" transzformálást a gümőzésre képtelen *nsp2* mutáns növényeken végezzük el. Az eljárás során kapott gümők minden esetben transzgenikusak és a fenotípusuk, illetve a növény fenotípusa a gümő genotípusától függ, például CRISPR/CAS9 génszerkesztés után. Ez a két rendszer szignifikánsan le tudja csökkenteni a transzgenikus gyökerek azonosításához szükséges munkaidőt, ráadásul a második rendszer alkalmazása során a nem-transzgenikus szövetek nem járulnak hozzá a szimbiotikus fenotípushoz és a növény életképességéhez, fitnesszéhez.

A két jelzőrendszer kifejlesztése mellett az *NCR* gének szabályzása is felkeltette az érdeklődésünket. A több mint 700 *Medicago truncatula NCR* gén kifejeződése rendkívül specifikus, az átíródásuk a gyökérgümő szimbiotikus sejtjeire korlátozódik és eddig egyetlen más tesztelt kísérleti körülmény sem tudta kiváltani az aktivitásukat. Vélekedésünk szerint, igazán hasznos lenne felderíteni, hogy az *NCR* gének ezen szigorú szabályzása hogyan is valósul meg. Korábban azt feltételezték, hogy az *NCR* gének kifejeződését IRLC- vagy akár *NCR* specifikus transzkripciós faktorok irányítják. Az *NCR169* gén szójabababban, amelyben nicsenek *NCR* és *GRP* gének, történő kifejeződését promóter-GUS fúziós rendszerrel vizsgálva azt találtuk, hogy a gén promótere aktív a szója gümőben, tehát a szójában és a *M. truncatula*-ban egyaránt jelen van az *NCR169* gén transzkripciós aktivátora. Ez az eredmény arra is utalhat, hogy az evolúció során az *NCR* gének kialakulása nem *de novo* történt. A feltételezett transzkripciós szabályozó elemek felkutatásának céljából ötvöztük a DNS affinitás kromatográfia, az élesztő-egy-hibrid (Y1H) és az EMSA módszereket, valamint felhasználtunk mind *Medicago*, mind szója gümőből származó biológiai anyagokat és cDNS könyvtárakat. Azokat a lehetséges *NCR169* promóter interakciós partnereket választottuk ki a további kísérletekhez, melyeket mind a DNS affinitás kromatográfia és az Y1H kísérletek esetében, valamint mind a *Medicago* és mind a szója mintáknál felbukkantak és együtt

fejeződtek ki az *NCR169*-el a gümő interzónájában és a nitrogén-kötő zónájában. Ezek az együttesen használt stratégiák több ígéretes jelöltet is adtak nekünk, melyeket jelenleg is vizsgálunk.

Reference

Albus U, Baier R, Holst O, Pühler A, Niehaus K, 2001. Suppression of an elicitor-induced oxidative burst reaction in *Medicago sativa* cell cultures by *Sinorhizobium meliloti* lipopolysaccharides. *New Phytologist*, 151:597-606.

Alkhalfioui F et al., 2008. A novel type of thioredoxin dedicated to symbiosis in legumes *Plant Physiology*, 148:424-435.

Alkhalfioui F et al., 2008. A novel type of thioredoxin dedicated to symbiosis in legumes. *Plant Physiology*, 148:424-435.

Alonso JM, Ecker JR, 2006. Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in *Arabidopsis. Nature Reviews Genetics*, 7:524.

Alunni B, Kevei Z, Redondo-Nieto M, Kondorosi A, Mergaert P, Kondorosi E, 2007. Genomic organization and evolutionary insights on *GRP* and *NCR* genes, two large nodule-specific gene families in *Medicago truncatula. Molecular Plant-Microbe Interactions*, 20:1138-1148.

Appleby CA, 1984. Leghemoglobin and *Rhizobium* respiration. *Annual Review of Plant Physiology*, 35:443-478.

Aslam SN et al., 2008. Bacterial polysaccharides suppress induced innate immunity by calcium chelation. *Current Biology*, 18:1078-1083.

Baird GS, Zacharias DA, Tsien RY, 2000. Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proceedings of the National Academy of Sciences*, 97:11984-11989.

Bass JIF, Reece-Hoyes JS, Walhout AJ, 2016. Generating bait strains for yeast one-hybrid assays. *Cold Spring Harbor Protocols*, 2016:pdb. prot088948.

Boisson-Dernier A, Chabaud M, Garcia F, Bécard G, Rosenberg C, Barker DG, 2001.

Agrobacterium rhizogenes-transformed roots of *Medicago truncatula* for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. *Molecular Plant-Microbe Interactions*, 14:695-700.

Bonaldi K et al., 2011. Nodulation of *Aeschynomene afraspera* and *A. indica* by photosynthetic *Bradyrhizobium* sp. strain ORS285: the nod-dependent versus the nod-independent symbiotic interaction. *Molecular Plant-Microbe Interactions*, 24:1359-1371.

Boogerd FC, van Rossum D, 1997. Nodulation of groundnut by *Bradyrhizobium*: a simple infection process by crack entry. *FEMS Microbiology Reviews*, 21:5-27.

Bourcy M et al., 2013. *Medicago truncatula* DNF2 is a PI-PLC-XD-containing protein required for bacteroid persistence and prevention of nodule early senescence and defense-like reactions. *New Phytologist*, 197:1250-1261.

Brewin NJ, 1991. Development of the legume root nodule. *Annual Review of Cell Biology*, 7:191-226.

Brundrett MC, Tedersoo L, 2018. Evolutionary history of mycorrhizal symbioses and global host plant diversity. *New Phytologist*, 220:1108-1115.

Carroll BJ, Gresshoff PM, 1983. Nitrate inhibition of nodulation and nitrogen fixation in white clover. *Zeitschrift für Pflanzenphysiologie*, 110:77-88.

Cassman KG, Dobermann A, Walters DT, 2002. Agroecosystems, nitrogen-use efficiency, and nitrogen management. *AMBIO: A Journal of the Human Environment*, 31:132-141.

Cebolla A, Vinardell JM, Kiss E, Olah B, Roudier F, Kondorosi A, Kondorosi E, 1999. The mitotic inhibitor ccs52 is required for endoreduplication and ploidy-dependent cell enlargement in plants. *The EMBO Journal*, 18:4476-4484.

Chiu W-l, Niwa Y, Zeng W, Hirano T, Kobayashi H, Sheen J, 1996. Engineered GFP as a vital reporter in plants. *Current Biology*, 6:325-330.

Chung MH, Chen MK, Pan SM, 2000. Floral spray transformation can efficiently generate *Arabidopsis*. *Transgenic research*, 9(6), 471-486.

Ciampitti IA, Vyn TJ, 2014. Nitrogen use efficiency for old versus modern corn hybrids. *Better Crops*, 98:10-21.

Clough SJ, Bent AF, 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana. The Plant Journal*, 16:735-743.

Clúa J, Roda C, Zanetti M, Blanco F, 2018. Compatibility between legumes and rhizobia for the establishment of a successful nitrogen-fixing symbiosis. *Genes*, 9:125.

Combier J-P et al., 2006. MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula. Genes & Development*, 20:3084-3088.

Cong L et al., 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339:819-823.

Curtin SJ et al., 2017. Validating genome-wide association candidates controlling quantitative variation in nodulation. *Plant Physiology*, 173:921-931.

Czernic P et al., 2015. Convergent evolution of endosymbiont differentiation in Dalbergioid and IRLC legumes mediated by nodule-specific cysteine-rich peptides. *Plant Physiology*, pp. 00584.02015.

De Stasio E, Moazed D, Noller H, Dahlberg A, 1989. Mutations in 16S ribosomal RNA disrupt antibiotic–RNA interactions. *The EMBO Journal*, 8:1213-1216.

Denarie J, Debelle F, Prome J-C, 1996. *Rhizobium* lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. *Annual review of biochemistry*, 65:503-535.

Doudna JA, Charpentier E, 2014. The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346:1258096.

Downie JA, 2007. Infectious heresy. *Science*, 316:1296-1297.

Downie JA, Walker SA, 1999. Plant responses to nodulation factors. *Current Opinion in Plant Biology*, 2:483-489.

Dudley ME, Jacobs TW, Long SR, 1987. Microscopic studies of cell divisions induced in alfalfa roots by *Rhizobium meliloti. Planta*, 171:289-301.

Esseling JJ, Lhuissier FG, Emons AMC, 2003. Nod factor-induced root hair curling: continuous polar growth towards the point of nod factor application. *Plant Physiology*, 132:1982-1988.

Evans Roth KJ, Stacey G. Homology in endosymbiotic systems: the term 'symbiosome'. In: Molecular genetics of plant-microbe interactions, 1988: proceedings of the 4th International Symposium on Molecular Genetics of Plant-Microbe Interactions, Acapulco, Mexico, May 15-20, 1988. Amer Phytopathological Society, p 220.

FAO F, 2015. World fertilizer trends and outlook to 2018. Food and Agriculture Organization of the United Nations Reporte.

Farkas A et al., 2014. *Medicago truncatula* symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms. *Proceedings of the National Academy of Sciences*, 111:5183-5188.

Feldmann KA, Marks MD, Christianson ML, Quatrano RS, 1989. A dwarf mutant of *Arabidopsis* generated by T-DNA insertion mutagenesis. *Science*, 243:1351-1354.

Fitter AH, 2005. Darkness visible: reflections on underground ecology. *Journal of Ecology*, 93:231-243.

Fraley RT, Rogers SG, Horsch RB, Gelvin SB, 1986. Genetic transformation in higher plants. *Critical Reviews in Plant Sciences*, 4:1-46.

Franssen HJ, Vijn I, Yang WC, Bisseling T, 1992. Developmental aspects of the *Rhizobium*-legume symbiosis. *Plant Molecular Biology*, 19:89-107.

Frugier F, Kosuta S, Murray JD, Crespi M, Szczyglowski K, 2008. Cytokinin: secret agent of symbiosis. *Trends in Plant Science*, 13:115-120.

Galindo-Trigo S, Gray JE, Smith LM, 2016. Conserved roles of CrRLK1L receptor-like kinases in cell expansion and reproduction from algae to angiosperms. *Frontiers in Plant Science*, 7:1269.

Gavrilovic S, Yan Z, Jurkiewicz AM, Stougaard J, Markmann K, 2016. Inoculation insensitive promoters for cell type enriched gene expression in legume roots and nodules. *Plant Methods*, 12:4

Gibson A, 1966. The carbohydrate requirements for symbiotic nitrogen fixation: a" whole-plant" growth analysis approach. *Australian Journal of Biological Sciences*, 19:499-516.

Giraud E et al., 2007. Legumes symbioses: absence of *Nod* genes in photosynthetic *Bradyrhizobia. Science*, 316:1307-1312.

Gonzalez-Rizzo S, Crespi M, Frugier F., 2006. The *Medicago truncatula* CRE1 cytokinin receptor regulates lateral root development and early symbiotic interaction with *Sinorhizobium meliloti*. *The Plant Cell*, 18(10), 2680-2693.

González-Sama A et al., 2006. Nuclear DNA endoreduplication and expression of the mitotic inhibitor Ccs52 associated to determinate and lupinoid nodule organogenesis. *Molecular Plant-Microbe Interactions*, 19:173-180.

Göttfert M, Horvath B, Kondorosi E, Putnoky P, Rodriguez-Quiñones F, Kondorosi A, 1986. At least two *nodD* genes are necessary for efficient nodulation of alfalfa by *Rhizobium meliloti. Journal of Molecular Biology*, 191:411-420.

Gourion B, Berrabah F, Ratet P, Stacey G, 2015. *Rhizobium*–legume symbioses: the crucial role of plant immunity. *Trends in Plant Science*, 20:186-194.

Guefrachi I et al., 2014. Extreme specificity of *NCR* gene expression in *Medicago truncatula.*

BMC Genomics, 15:712.

Guefrachi I et al., 2015. *Bradyrhizobium* BclA is a peptide transporter required for bacterial differentiation in symbiosis with *Aeschynomene* legumes. *Molecular Plant-Microbe Interactions*, 28:1155-1166.

Györgyey J, Vaubert D, Jiménez-Zurdo JI, Charon C, Troussard L, Kondorosi Á, Kondorosi É, 2000. Analysis of *Medicago truncatula* nodule expressed sequence tags. *Molecular Plant-Microbe Interactions*, 13(1., 62-71.

Haag AF et al., 2011. Protection of *Sinorhizobium* against host cysteine-rich antimicrobial peptides is critical for symbiosis. *PLoS Biology*, 9:e1001169.

Hammami R, Ben Hamida J, Vergoten G, Fliss I, 2008. PhytAMP: a database dedicated to antimicrobial plant peptides. *Nucleic Acids Research*, 37:D963-D968.

Hancock RE, 2001. Cationic peptides: effectors in innate immunity and novel antimicrobials. *The Lancet Infectious Diseases*, 1:156-164.

Hancock RE, Diamond G, 2000. The role of cationic antimicrobial peptides in innate host defences. *Trends in Microbiology*, 8:402-410.

Harrison MJ, van Buuren ML, 1995. A phosphate transporter from the mycorrhizal fungus *Glomus versiforme*. *Nature*, 378:626.

Hernández G et al., 2009. Global changes in the transcript and metabolic profiles during symbiotic nitrogen fixation in phosphorus-stressed common bean plants. *Plant Physiology*, 151:1221-1238.

Herrero M et al., 2017. Farming and the geography of nutrient production for human use: a transdisciplinary analysis. *The Lancet Planetary Health*, 1:e33-e42.

Hibbett DS et al., 2007. A higher-level phylogenetic classification of the Fungi. *Mycological Research*, 111:509-547.

Hijikata N, Murase M, Tani C, Ohtomo R, Osaki M, Ezawa T, 2010. Polyphosphate has a central role in the rapid and massive accumulation of phosphorus in extraradical mycelium of an arbuscular mycorrhizal fungus. *New Phytologist*, 186:285-289.

Hodge A, Campbell CD, Fitter AH, 2001. An arbuscular mycorrhizal fungus accelerates decomposition and acquires nitrogen directly from organic material. *Nature*, 413:297.

Holton A, Cornish C., 1995. Genetics and biochemistry of anthocyanin biosynthesis. *The Plant Cell*, 7(7), 1071.

Horsch R, Rogers S, Fraley R Transgenic plants. In: Cold Spring Harbor Symposia on Quantitative Biology, 1985. Cold Spring Harbor Laboratory Press, pp 433-437.

Horváth B et al., 2015. Loss of the nodule-specific cysteine rich peptide, NCR169, abolishes symbiotic nitrogen fixation in the *Medicago truncatula dnf7* mutant. *Proceedings of the National Academy of Sciences*, 112:15232-15237.

Jefferson RA, 1987. Assaying chimeric genes in plants: the *GUS* gene fusion system. *Plant Molecular Biology Reporter*, 5:387-405.

Jefferson RA, Burgess SM, Hirsh D, 1986. beta-Glucuronidase from *Escherichia coli* as a gene-fusion marker. *Proceedings of the National Academy of Sciences*, 83:8447-8451.

Ji H, Vagner A, Mingyi W, Jeremy M, Patrick Z, Yuhong T, Michael U. The *Medicago truncatula* gene expression atlas web server. *BMC Bioinformatics*, 2009, 10:441.

Jin H, Liu J, Liu J, Huang X, 2012. Forms of nitrogen uptake, translocation, and transfer via arbuscular mycorrhizal fungi: a review. *Science China Life Sciences*, 55:474-482.

Jin Y et al., 2016. DELLA proteins are common components of symbiotic rhizobial and mycorrhizal signalling pathways. *Nature Communications*, 7, 12433.

Jones JD, Dangl JL, 2006. The plant immune system nature 444:323.

Jones KM, Walker GC, 2008. Responses of the model legume *Medicago truncatula* to the

rhizobial exopolysaccharide succinoglycan. *Plant Signaling & Behavior*, 3:888-890.

Jutras BL, Verma A, Stevenson B, 2012. Identification of novel DNA-Binding proteins using DNA-Affinity Chromatography/Pull-down. *Current Protocols in Microbiology*, 24:1F. 1.1-1F. 1.13.

Kaló P et al., 2005. Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. *Science*, 308:1786-1789.

Kereszt A, Li D, Indrasumunar A, Nguyen CD, Nontachaiyapoom S, Kinkema M, Gresshoff PM, 2007. *Agrobacterium rhizogenes*-mediated transformation of soybean to study root biology. *Nature Protocols*, 2:948.

Kereszt A, Mergaert P, Maróti G, Kondorosi É, 2011. Innate immunity effectors and virulence factors in symbiosis. *Current Opinion in Microbiology*, 14:76-81.

Kim M, Chen Y, Xi J, Waters C, Chen R, Wang D, 2015. An antimicrobial peptide essential for bacterial survival in the nitrogen-fixing symbiosis. *Proceedings of the National Academy of Sciences*, 112:15238-15243.

Kolkman JM et al., 2005. Distribution of Activator (Ac) throughout the maize genome for use in regional mutagenesis. *Genetics*, 169:981-995.

Kondorosi E et al., 1989. Positive and negative control of nod gene expression in *Rhizobium meliloti* is required for optimal nodulation. *The EMBO Journal*, 8:1331-1340.

Kondorosi E, Mergaert P, Kereszt A, 2013. A paradigm for endosymbiotic life: cell differentiation of *Rhizobium* bacteria provoked by host plant factors. *Annual Review of Microbiology*, 67:611-628.

Kouchi H et al., 2004. Large-scale analysis of gene expression profiles during early stages of root nodule formation in a model legume, *Lotus japonicus. DNA Research*, 11:263-274.

Lamouche F, Bonade-Bottino N, Mergaert P, Alunni B, 2019. Symbiotic efficiency of spherical and elongated bacteroids in the *Aeschynomene-Bradyrhizobium* symbiosis.

Frontiers in Plant Science, 10:377.

Laporte P et al., 2010. A novel RNA-binding peptide regulates the establishment of the *Medicago truncatula–Sinorhizobium meliloti* nitrogen-fixing symbiosis. *The Plant Journal*, 62:24-38.

Lassaletta L, Billen G, Garnier J, Bouwman L, Velazquez E, Mueller ND, Gerber JS, 2016. Nitrogen use in the global food system: past trends and future trajectories of agronomic performance, pollution, trade, and dietary demand. *Environmental Research Letters*,11:095007.

Lazo GR, Stein PA, Ludwig RA, 1991. A DNA transformation–competent *Arabidopsis* genomic library in *Agrobacterium*. *Bio/technology*, 9(10), 963-967.

Li T, Liu B, Spalding MH, Weeks DP, Yang B, 2012. High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nature Biotechnology*, 30:390.

Libault M et al., 2010. An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *The Plant Journal*, 63:86-99.

Lie T, 1974. Environmental effects on nodulation and symbiotic nitrogen fixation. The biology of Nitrogen fixation. *Frontiers of Biology*, 33:555-582.

Lin M-H, Gresshoff PM, Indrasumunar A, Ferguson BJ, 2011. pHairyRed: a novel binary vector containing the *DsRed2* reporter gene for visual selection of transgenic hairy roots. *Molecular Plant*, 4:537-545.

Lloyd A, Plaisier CL, Carroll D, Drews GN, 2005. Targeted mutagenesis using zinc-finger nucleases in *Arabidopsis. Proceedings of the National Academy of Sciences*, 102:2232-2237.

Lodwig EM et al., 2003. Amino-acid cycling drives nitrogen fixation in the legume–*Rhizobium* symbiosis. *Nature*, 422:722.

Lohar DP et al., 2006. Transcript analysis of early nodulation events in *Medicago truncatula. Plant Physiology*, 140:221-234.

Maekawa T, Kusakabe M, Shimoda Y, Sato S, Tabata S, Murooka Y, Hayashi M, 2008. Polyubiquitin promoter-based binary vectors for overexpression and gene silencing in *Lotus japonicus. Molecular Plant-Microbe Interactions*, 21:375-382.

Mahon JD, 1977. Respiration and the energy requirement for nitrogen fixation in nodulated pea roots. *Plant Physiology*, 60:817-821.

Mali P et al., 2013. RNA-guided human genome engineering via Cas9. *Science*, 339:823-826

Marlow VL, Haag AF, Kobayashi H, Fletcher V, Scocchi M, Walker GC, Ferguson GP, 2009. Essential role for the BacA protein in the uptake of a truncated eukaryotic peptide in *Sinorhizobium meliloti. Journal of Bacteriology*, 191:1519-1527.

Maróti G, Downie JA, Kondorosi É, 2015. Plant cysteine-rich peptides that inhibit pathogen growth and control rhizobial differentiation in legume nodules. *Current Opinion in Plant Biology*, 26:57-63.

Marsh JF et al., 2007. *Medicago truncatula* NIN is essential for rhizobial-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase. *Plant Physiology*, 144:324-335.

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

Masukawa H, 1969. Localization of sensitivity to kanamycin and streptomycin in 30S ribosomal proteins of *Escherichia coli*. *The Journal of Antibiotics*, 22:612-623.

Maunoury N, Kondorosi A, Kondorosi E, Mergaert P, 2008. Cell biology of nodule infection and development. In: Nitrogen-fixing Leguminous Symbioses. Springer, pp 153-189.

Meijer, H, Ouwerkerk B, Hoge C, 1998. Vectors for transcription factor cloning and target site identification by means of genetic selection in yeast. *Yeast*, 14(15), 1407-1416.

Mergaert P, Nikovics K, Kelemen Z, Maunoury N, Vaubert D, Kondorosi A, Kondorosi E, 2003. A novel family in *Medicago truncatula* consisting of more than 300 nodule-specific

genes coding for small, secreted polypeptides with conserved cysteine motifs. *Plant Physiology*, 132:161-173.

Mergaert P et al., 2006. Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*–legume symbiosis. *Proceedings of the National Academy of Sciences*, 103:5230-5235.

Miao Y, Jiang L, 2007. Transient expression of fluorescent fusion proteins in protoplasts of suspension cultured cells. *Nature Protocols*, 2:2348.

Middleton PH et al., 2007. An ERF transcription factor in *Medicago truncatula* that is essential for Nod factor signal transduction. *The Plant Cell*, 19:1221-1234.

Mikuláss KR et al., 2016. Antimicrobial nodule-specific cysteine-rich peptides disturb the integrity of bacterial outer and inner membranes and cause loss of membrane potential. *Annals of Clinical Microbiology and Antimicrobials*, 15:43.

Misumi M, Nishimura T, Komai T, Tanaka N, 1978. Interaction of kanamycin and related antibiotics with the large subunit of ribosomes and the inhibition of translocation. *Biochemical and Biophysical Research Communications*, 84:358-365.

Montiel J et al., 2017. Morphotype of bacteroids in different legumes correlates with the number and type of symbiotic NCR peptides. *Proceedings of the National Academy of Sciences*, 114:5041-5046.

Montiel J, Szűcs A, Boboescu IZ, Gherman VD, Kondorosi É, Kereszt A, 2016. Terminal bacteroid differentiation is associated with variable morphological changes in legume species belonging to the inverted repeat-lacking clade. *Molecular Plant-Microbe Interactions*, 29:210-219.

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

Mus F et al., 2016. Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. *Applied and Environmental Microbiology*, 82:3698-3710.

Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K, Miyawaki A, 2002. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnology*, 20:87.

Nagymihály M et al, 2017. Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of gene expression. *Proceedings of the National Academy of Sciences*, 114(17), 4543-4548.

Okamoto S, Kawaguchi M, 2015. Shoot HAR1 mediates nitrate inhibition of nodulation in *Lotus japonicus. Plant Signaling & Behavior*, 10:e1000138.

Okamoto S, Ohnishi E, Sato S, Takahashi H, Nakazono M, Tabata S, Kawaguchi M, 2008. Nod factor/nitrate-induced *CLE* genes that drive HAR1-mediated systemic regulation of nodulation. *Plant and Cell Physiology*, 50:67-77.

Oldroyd GE, Downie JA, 2008. Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annual Reviews in Plant Biolology*, 59:519-546.

Oldroyd GE, Long SR, 2003. Identification and characterization of *Nodulation-Signaling Pathway 2*, a gene of *Medicago truncatula* involved in Nod factor signaling. *Plant Physiology*, 131:1027-1032.

Oono R, Schmitt I, Sprent JI, Denison RF, 2010. Multiple evolutionary origins of legume traits leading to extreme rhizobial differentiation. *New Phytologist*, 187:508-520.

Ördögh L et al., 2014. Symbiotic plant peptides eliminate *Candida albicans* both *in vitro* and in an epithelial infection model and inhibit the proliferation of immortalized human cells. *BioMed Research International*, 2014.

Ott T et al., 2005. Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. *Current Biology*, 15:531-535.

Pan H, Wang D, 2017. Nodule cysteine-rich peptides maintain a working balance during nitrogen-fixing symbiosis. *Nature Plants*, 3:17048.

Parniske M, 2008. Arbuscular mycorrhiza: the mother of plant root endosymbiosis. *Nature Reviews Microbiology*, 6:763.

Parniske M, 2008. Arbuscular mycorrhiza: the mother of plant root endosymbiosis. *Nature Reviews Microbiology*, 6:763.

Peel GJ, Pang Y, Modolo LV, Dixon RA, 2009. The LAP1 MYB transcription factor orchestrates anthocyanidin biosynthesis and glycosylation in *Medicago. The Plant Journal*, 59:136-149.

Penterman J, Abo RP, De Nisco NJ, Arnold MF, Longhi R, Zanda M, Walker GC, 2014. Host plant peptides elicit a transcriptional response to control the *Sinorhizobium meliloti* cell cycle during symbiosis. *Proceedings of the National Academy of Sciences*, 111:3561-3566.

Petit A, Stougaard J, Kühle A, Marcker KA, Tempé J, 1987. Transformation and regeneration of the legume *Lotus corniculatus*: a system for molecular studies of symbiotic nitrogen fixation. *Molecular and General Genetics*, 207:245-250.

Price PA, Tanner HR, Dillon BA, Shabab M, Walker GC, Griffitts JS, 2015. Rhizobial peptidase HrrP cleaves host-encoded signaling peptides and mediates symbiotic compatibility. *Proceedings of the National Academy of Sciences*, 112:15244-15249.

Quandt HJ, Pühler A, Broer I, 1993. Transgenic root nodules of *Vicia hirsuta*: a fast and efficient system for the study of gene expression in indeterminate-type nodules. *Molecular Plant Microbe Interactions*, 6:699-706.

Radutoiu S et al., 2003. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature*, 425:585.

Radutoiu S et al., 2007. LysM domains mediate lipochitin–oligosaccharide recognition and *NFR* genes extend the symbiotic host range. *The EMBO Journal*, 26:3923-3935.

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F, 2013. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, 8:2281.

Ray DK, Ramankutty N, Mueller ND, West PC, Foley JA, 2012. Recent patterns of crop yield growth and stagnation. *Nature Communications*, 3:1293.

Reeve W et al., 2010. Complete genome sequence of the *Medicago* microsymbiont *Ensifer*, *Sinorhizobium. medicae* strain WSM419. *Standards in Genomic Sciences*, 2:77.

Rey T, Jacquet C, 2018. Symbiosis genes for immunity and vice versa. *Current Opinion in Plant Biology*, 44:64-71.

Ribeiro CW et al., 2017. Regulation of differentiation of nitrogen-fixing bacteria by microsymbiont targeting of plant thioredoxin s1. *Current Biology*, 27:250-256.

Rice W, Penney D, Nyborg M, 1977. Effects of soil acidity on rhizobia numbers, nodulation and nitrogen fixation by alfalfa and red clover. *Canadian Journal of Soil Science*, 57:197-203.

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. *The Plant Journal*, 77:817-837.

Savka MA, Ravillion B, Noel GR, Farrand SK., 1990. Induction of hairy roots on cultivated soybean genotypes and their use to propagate the soybean cyst nematode. *Phytopathology*, 80(5), 503-508.

Schauser L, Roussis A, Stiller J, Stougaard J, 1999. A plant regulator controlling development of symbiotic root nodules. *Nature*, 402:191.

Scheidle H, Groß A, Niehaus K, 2005. The lipid A substructure of the S*inorhizobium meliloti* lipopolysaccharides is sufficient to suppress the oxidative burst in host plants. *New Phytologist*, 165:559-566.

Schultze M, Kondorosi A, 1998. Regulation of symbiotic root nodule development. *Annual*

Review of Genetics, 32:33-57.

Sen D, Weaver R, 1980. Nitrogen fixing activity of rhizobial strain 32H1 in peanut and cowpea nodules. *Plant Science Letters*, 18:315-318.

Serraj R, Sinclair TR, Purcell LC, 1999. Symbiotic N_2 fixation response to drought. *Journal of Experimental Botany*, 50:143-155.

Shabab M et al., 2016. Disulfide cross-linking influences symbiotic activities of nodule peptide NCR247. *Proceedings of the National Academy of Sciences*, 113:10157-10162.

Shah VK, Brill WJ, 1977. Isolation of an iron-molybdenum cofactor from nitrogenase. *Proceedings of the National Academy of Sciences*, 74:3249-3253.

Shen J, Li C, Mi G, Li L, Yuan L, Jiang R, Zhang F, 2012. Maximizing root/rhizosphere efficiency to improve crop productivity and nutrient use efficiency in intensive agriculture of China. *Journal of Experimental Botany*, 64:1181-1192.

Shi Z, Xie Y, 2014. Biosynthesis and metabolic engineering of anthocyanins in *Arabidopsis thaliana*. *Recent Patents on Biotechnology*, 8(1), 47-60.

Singh S, 2014. The calcium signature decoding CCaMK/CYCLOPS complex activates the transcription of symbiosis associated genes. Doctoral dissertation.

Singh S, Parniske M, 2012. Activation of calcium-and calmodulin-dependent protein kinase, CCaMK., the central regulator of plant root endosymbiosis. *Current Opinion in Plant Biology*, 15:444-453.

Smil V, 1999. Nitrogen in crop production: An account of global flows. *Global Biogeochemical Cycles*, 13:647-662.

Smit P, Raedts J, Portyanko V, Debellé F, Gough C, Bisseling T, Geurts R, 2005. NSP1 of the GRAS protein family is essential for rhizobial Nod factor-induced transcription. *Science*, 308:1789-1791.

Smith SE, Jakobsen I, Grønlund M, Smith FA, 2011. Roles of arbuscular mycorrhizas in plant phosphorus nutrition: interactions between pathways of phosphorus uptake in arbuscular mycorrhizal roots have important implications for understanding and manipulating plant phosphorus acquisition. *Plant Physiology*, 156:1050-1057.

Smith SE, Read DJ, 2010. Mycorrhizal symbiosis. *Academic Press*.

Soussi M, Lluch C, Ocana A, Norero A, 1999. Comparative study of nitrogen fixation and carbon metabolism in two chick-pea (*Cicer arietinum* L.) cultivars under salt stress. *Journal of Experimental Botany*, 50:1701-1708.

Spano L, Mariotti D, Pezzotti M, Damiani F, Arcioni S, 1987. Hairy root transformation in alfalfa, *Medicago sativa* L.. *Theoretical and Applied Genetics*, 73:523-530.

Sparkes IA, Runions J, Kearns A, Hawes C, 2006. Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nature Protocols*, 1:2019.

Sprent JI, James EK, 2007. Legume evolution: where do nodules and mycorrhizas fit in? *Plant Physiology*, 144:575-581.

Starker CG, Parra-Colmenares AL, Smith L, Mitra RM, Long SR, 2006. Nitrogen fixation mutants of *Medicago truncatula* fail to support plant and bacterial symbiotic gene expression. *Plant Physiology*, 140:671-680.

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

Stougaard J, Sandal NN, Grøn A, Kühle A, Marcker KA, 1987. 5′ analysis of the soybean leghemoglobin *lbc3* gene: regulatory elements required for promoter activity and organ specificity. *The EMBO Journal*, 6:3565-3569.

Streeter J, Wong PP, 1988. Inhibition of legume nodule formation and N_2 fixation by nitrate.

Critical Reviews in Plant Sciences, 7:1-23.

Sutton R, Tinus RW, 1983. Root and root system terminology. *Forest Science*, 29:a0001-z0001.

Suzuki J, KUNIMOTO T, HORI M, 1970. Effects of kanamycin on protein synthesis: inhibition of elongation of peptide chains. *The Journal of Antibiotics*, 23:99-101.

Szczyglowski K et al., 1998. Nodule organogenesis and symbiotic mutants of the model legume *Lotus japonicus. Molecular Plant-Microbe Interactions*, 11:684-697.

Takahashi F, Hanada K, Kondo T, Shinozaki K, 2019. Hormone-like peptides and small coding genes in plant stress signaling and development. *Current Opinion in Plant Biology*, 51:88-95.

Tilman D, Cassman KG, Matson PA, Naylor R, Polasky S, 2002. Agricultural sustainability and intensive production practices. *Nature*, 418:671.

Tirichine L et al., 2007. A gain-of-function mutation in a cytokinin receptor triggers spontaneous root nodule organogenesis. *Science*, 315:104-107.

Tiricz H et al., 2013. Antimicrobial nodule-specific cysteine-rich peptides induce membrane depolarization-associated changes in the transcriptome of *Sinorhizobium meliloti. Applied Environmental Microbiology*, 79:6737-6746.

Trujillo DI, Silverstein KA, Young ND, 2014. Genomic characterization of the *LEED.. PEED*s, a gene family unique to the *Medicago* lineage. *G3: Genes*, *Genomes*, *Genetics*, 4:2003-2012.

Uroz S, Courty PE, Oger P, 2019. Plant symbionts are engineers of the plant-associated microbiome. *Trends in plant science.*

Vagner B et al, 2008. A gene expression atlas of the model legume *Medicago truncatula*. *The Plant Journal*, 55(3):504--513.

Van de Velde W et al., 2010. Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science*, 327:1122-1126.

Verma D, 1992. Signals in root nodule organogenesis and endocytosis of *Rhizobium. The Plant Cell*, 4:373.

Vinardell JM et al., 2003. Endoreduplication mediated by the anaphase-promoting complex activator CCS52A is required for symbiotic cell differentiation in *Medicago truncatula* nodules. *The Plant Cell*, 15:2093-2105.

Wang D et al., 2010. A nodule-specific protein secretory pathway required for nitrogen-fixing symbiosis. *Science*, 327:1126-1129.

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

Wang Q et al., 2017. Host-secreted antimicrobial peptide enforces symbiotic selectivity in *Medicago truncatula. Proceedings of the National Academy of Sciences*, 114:6854-6859.

Wittenberg JB, Bergersen FJ, Appleby CA, Turner GL, 1974. Facilitated oxygen diffusion the role of leghemoglobin in nitrogen fixation by bacteroids isolated from soybean root nodules. *Journal of Biological Chemistry*, 249:4057-4066.

Xie K, Minkenberg B, Yang Y, 2015. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proceedings of the National Academy of Sciences*, 112:3570-3575.

Xing HL et al., 2014. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biology*, 14:327.

Yang S et al., 2017. Microsymbiont discrimination mediated by a host-secreted peptide in *Medicago truncatula. Proceedings of the National Academy of Sciences*, 114:6848-6853.

Yeaman MR, Yount NY, 2007. Unifying themes in host defence effector polypeptides. *Nature Reviews Microbiology*, 5:727.

Yount NY, Yeaman MR, 2004. Multidimensional signatures in antimicrobial peptides. *Proceedings of the National Academy of Sciences*, 101:7363-7368.

Zhang S, Wang Y, Li K, Zou Y, Chen L, Li X, 2014. Identification of cold-responsive miRNAs and their target genes in nitrogen-fixing nodules of soybean. *International Journal of Molecular Sciences*, 15:13596-13614.

Zhang Z, Mao Y, Ha S, Liu W, Botella JR, Zhu J-K, 2016. A multiplex CRISPR/Cas9 platform for fast and efficient editing of multiple genes in *Arabidopsis. Plant Cell Reports*, 35:1519-1533.

Supplemental material

Table S1. List for primers used in this study.

Purpose	Name	Sequence
pPurpleRoot and pPurpleShoot vectors construction	MtLAP1_pCncoF	GGACTCTTGACCATGGAGAATACCGGAGGTGTGAG
	MtLap1_pCeheR	ACCTGTAATTCACACGTGTCAAGGTAGATCCCAAAG
	MaNCR169salF	AAGTCGACAAGATGGTTTAGTACATC
	MaNCR169HindR	CCAAGCTTATACCAGAGAACGCAAATATTTTC
	2g37950prF	AGCAGCTTGACCATGGGCCACCAGCCAAATGTTTCTG
	2g37950prR	TCCGGTATTCTCCATGATTTTTGCCTAATGAATGTTTCTTTTTG
	5g24800prF	AGCAGCTTGACCATGGCTACGTATAGTGGATATACGTCGTTCC
	5g24800prR	TCCGGTATTCTCCATGTTCTTTGAATGTGAACACACAAGAAAGA
pKSE401-NSP2	mNSP2r	CGATGGGGGACCAGTGGTTCCAG
	mNSP2f	CTGGAACCACTGGTCCCCCATCG
	35Snsp2ecoF	GATTGACAACGAATTCTCATCAAGACGATCTACCCGA
	35Snsp2ecoR	TTACGAATTCGAATTCATCTTTTGCTCCCATAGTTTTTTCAG
NFS1 knock-out and detection	NFS1gRNAf	ATTGTTGAAAGAGGTGTAAGCTC
	NFS1gRNAr	AAACGAGCTTACACCTCTTTCAA
	PmeOPfw	CAAACACTGATAGTTTAAACTGTGTGGAATTGTGAGCGG
	UnivPmeR1	TTCCCGCCTTCAGTTTGCCAGGGTTTTCCCAGTCACGA
	UnivOpPmeR2	CAATTCCACACAGTTTGCCAGGGTTTTCCCAGTCACGA
Y1H screening on *NCR169* and	NCR211prXbaF	CCCTCTAGACCAGGTAAGGTCCCCACATCCAACTC
	NCR211prMunR	TTTCAATTGAATTGTTTGATCCCGTTATATAG

NCR211. Bait generation and prey gene sequencing.	NCR169prXbaF	AAATCTAGACCAAAACCGCTACAATCAGATCG
	NCR169prEcoR	TGTGAATTCCCCTTTGCGTGAAATAAGTG
	PDC6aF	CTTATTGTTTGGCATTTGTAGCGGCAGTC
	5’ AD	AGGGATGTTTAATACCACTAC
	3’ AD	GCACAGTTGAAGTGAACTTGC
Prone generation for DNA pull-down with *NCR169* promoter	NCR69 prom 1F	CATTTTTTCTATTTGACTTGTC
	NCR69 prom 1R	GGATACTTTGCACATGTAACATAGTG
	NCR69 prom 2F	TCATATCATGAGTTCATTACTCCTGC
	NCR69 prom 2R	GAGCTCATTATTGTTTAACAACC
	NCR69 prom 3F	GTGATTCCACACCCACTCATATC
Probe generation for EMSA assay on *NCR169* promoter	NCR69 prom 3R	CCCATTGTTTTTCCCCTTTGCGTG
	NCR169a F	CACGTTTTATTTAACCACAAGAC
	NCR169a R	GAGATAGGAAAACCAATATTA
	NCR169b F	CTCTATTATGAAAGGCTGGC
	NCR169b R	CTTTCTTTTATACAAATCATAAAAT
	NCR169c F	GAAAGTAATAATTTTGAGACATCTATTC
	NCR169c R	GTAATTAATACTTTTTGAAAATT
	NCR169d F	GTATTAATTACAACAACATAATTATTA
	NCR169d R	CCCATTGTTTTTCCCCTTTGC
	NCR169e F	TGATTCCACACCCACTCATATC
	NCR169e R	GTTTAACAACCTTTCATTATT
	NCR169f F	GTTAAACAATAATGAGCTCTATT
	NCR169f R	GTTGTCTCTTTGGCTCATAGA
	NCR169g F	GACAACTACGATACATCAACGTATT
	NCR169g R	GATATGTTTACATCAACAAAAC

Detection of gene modification on NFS1 via NGS. Gene specific primers with adaptor, and NGS primers with barcode (BC).	NTS1_OP1	TGTGTGGAATTGTGAGCGGGACATCAAAGAGAGAAAAAAATATGAG
	NTS1_UN1	GCCAGGGTTTTCCCAGTCACGATCATCAACGCAATAAAAAGGAGATGC
	IonP_UNIV	CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATGCCAGGGTTTTCCCAGTCACGA
	IonA_OP_BC85	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCAGCCTCAACGATTGTGTGGAATTGTGAGCGG
	IonA_OP_BC86	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTGGTTATTCGATTGTGTGGAATTGTGAGCGG
	IonA_OP_BC87	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGGCTGGACGATTGTGTGGAATTGTGAGCGG
	IonA_OP_BC88	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCGAACACTTCGATTGTGTGGAATTGTGAGCGG
	IonA_OP_BC89	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTGAATCTCGATTGTGTGGAATTGTGAGCGG
	IonA_OP_BC90	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAACCACGGCGATTGTGTGGAATTGTGAGCGG
	IonA_OP_BC91	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGGAAGGATGCGATTGTGTGGAATTGTGAGCGG
	IonA_OP_BC92	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAGGAACCGCGATTGTGTGGAATTGTGAGCGG
	IonA_OP_BC93	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTGTCCAATCGATTGTGTGGAATTGTGAGCGG
	IonA_OP_BC94	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCGACAAGCGATTGTGTGGAATTGTGAGCGG
	IonA_OP_BC95	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGGACAGATCGATTGTGTGGAATTGTGAGCGG
	IonA_OP_BC96	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTAAGCGGTCGATTGTGTGGAATTGTGAGCGG

Table S2. Predicted transcription factors with specific expression in the nodule and highly expressed in nodule Inter Zone and Zone III. Gene function annotation information from PlantTFDB (http://planttfdb.cbi.pku.edu.cn/), and expression data refered to MtGEA (https://mtgea.noble.org/v2/) and Roux et al., 2014. Zone II D, distal part of Zone II; Zone II P, proximal part of Zone II.

Number	Gene ID	Annotation from database	Total reads	Expression distribution in different nodule zones				
				Zone I	Zone II D	Zone II P	Interzone	Zone III
1	Medtr0249s0070	A20-like	9567.9	6.6	8.9	19.7	28.8	36.1
2	Medtr0558s0020	FAR	75.0	0.0	0.0	0.0	39.9	60.1
3	Medtr1g008990	bZIP	7796.2	1.3	7.7	23.1	36.3	31.6
4	Medtr1g012960	Homobox-WOX	9920.0	2.5	8.8	23.5	33.1	32.1
5	Medtr1g013790	WRKY	1652.1	0.0	0.2	0.0	28.9	71.0
6	Medtr1g023050	Homobox-WOX	3083.8	7.3	10.2	14.1	28.1	40.2
7	Medtr1g026070	C2H2	3000.5	0.0	0.0	0.7	66.4	32.9
8	Medtr1g032440	C2H2	3446.9	6.8	8.4	12.0	38.6	34.2
9	Medtr1g050502	bZIP	1667.7	3.4	7.0	24.0	34.2	31.3
10	Medtr1g055240	FAR	354.9	0.1	0.0	1.4	39.8	58.7
11	Medtr1g057790	Homobox-WOX	6340.2	3.7	7.3	14.2	46.9	27.9
12	Medtr1g061130	C2H2	2728.8	7.6	12.6	13.8	35.0	31.1
13	Medtr1g064350	GRF	71297.3	0.1	1.2	11.6	64.8	22.4
14	Medtr1g064680	C2H2	2130.7	5.2	4.6	8.7	46.2	35.4
15	Medtr1g070250	C2H2	9020.8	2.6	14.0	11.6	27.3	44.4
16	Medtr1g073840	RR-A-type	2425.7	2.5	4.9	15.1	30.9	46.7
17	Medtr1g075430	C2H2	393.2	0.2	5.1	11.2	57.5	26.0
18	Medtr1g084160	bHLH	822.0	9.5	11.5	23.0	32.9	23.1
19	Medtr1g087920	AP2-EREBP	22143.1	0.9	2.8	13.2	39.2	43.9
20	Medtr1g088660	C2H2	1443.9	2.8	7.9	28.3	24.8	36.2
21	Medtr1g088875	C3H	2464.7	15.5	12.1	8.5	28.3	35.6
22	Medtr1g089290	WD40-like	1346.7	5.7	6.0	13.2	29.1	46.1

23	Medtr1g090450	C2H2	9402.9	0.1	0.7	6.4	45.3	47.4
24	Medtr1g090723	NAM	6277.3	9.7	13.9	15.0	29.6	31.8
25	Medtr1g090803	C2H2	2901.8	6.5	8.4	16.8	27.9	40.3
26	Medtr1g101810	TCP	1577.6	10.8	6.4	11.8	31.5	39.5
27	Medtr1g102710	WD40-like	2223.1	7.1	9.9	20.7	36.4	25.9
28	Medtr1g106420	AS2-LOB	457.3	5.0	3.3	1.6	26.6	63.6
29	Medtr1g109280	C2H2	758.4	4.2	6.9	9.1	35.2	44.6
30	Medtr1g114150	JmjC	5058.2	11.2	3.5	2.7	39.0	43.6
31	Medtr1g116790	bHLH	2101.7	10.3	10.9	13.4	32.0	33.4
32	Medtr1g116940	C3H	51.1	6.8	18.9	4.3	29.1	40.9
33	Medtr2g009310	C2H2	490.1	1.1	3.5	7.6	40.0	47.7
34	Medtr2g012830	C2H2	2968.1	12.7	11.7	11.9	29.9	33.8
35	Medtr2g014680	NAM	328.8	0.0	1.0	4.9	57.1	37.0
36	Medtr2g020280	TUBBY	4559.2	5.6	6.6	10.7	41.5	35.7
37	Medtr2g020870	C2H2	1400.3	11.2	11.0	13.9	30.9	33.0
38	Medtr2g042420	WD40-like	10160.1	0.1	0.2	4.4	37.9	57.3
39	Medtr2g044910	TIFY	25483.7	5.1	9.6	20.6	39.7	24.9
40	Medtr2g047130	JmjC	533.5	8.6	8.3	11.0	35.3	36.8
41	Medtr2g078450	C2H2	1380.4	9.2	7.7	11.2	43.3	28.6
42	Medtr2g082640	bHLH	5469.3	8.1	8.5	9.2	30.5	43.7
43	Medtr2g082770	WD40-like	5430.0	3.2	2.7	4.5	23.9	65.8
44	Medtr2g090295	C2H2	1441.3	5.0	4.1	12.9	48.4	29.7
45	Medtr2g090440	bHLH	393.4	0.4	0.6	1.4	58.8	38.8
46	Medtr2g101410	WD40-like	4057.1	0.5	7.9	20.6	33.7	37.2
47	Medtr2g104490	bHLH	3841.4	0.0	0.3	0.2	44.7	54.8
48	Medtr3g005790	C2H2	175.7	3.5	1.7	0.0	34.8	60.1
49	Medtr3g007760	PHD	3365.1	9.3	13.5	23.5	30.7	23.0

50	Medtr3g055430	C2H2	425.5	0.2	0.4	15.0	40.1	44.3
51	Medtr3g056110	GRAS	2728.4	9.9	7.0	7.7	34.8	40.6
52	Medtr3g056160	C3H	1728.1	0.8	8.7	3.4	33.5	53.6
53	Medtr3g060942	CCHC(Zn)	3721.5	0.1	0.1	0.4	26.8	72.6
54	Medtr3g060950	C2H2	4339.2	0.1	0.0	0.0	44.4	55.5
55	Medtr3g069000	BTB-POZ	688.4	0.4	2.4	8.9	64.4	23.9
56	Medtr3g070780	PHD	768.4	14.1	11.4	10.1	31.4	32.9
57	Medtr3g075530	C2H2	5530.2	6.4	4.9	12.4	49.4	26.9
58	Medtr3g083480	C2H2	4873.4	3.4	4.4	12.2	34.5	45.5
59	Medtr3g088560	BTB-POZ	1012.1	8.3	10.2	17.7	28.8	35.0
60	Medtr3g089055	GRAS	2514.0	2.7	6.5	8.6	35.5	46.7
61	Medtr3g091510	C2H2	1516.8	13.9	14.1	9.7	28.0	34.2
62	Medtr3g096520	C2H2	1286.6	1.6	2.8	6.7	68.0	21.0
63	Medtr3g099370	JmjC	1584.8	9.0	12.9	17.2	28.5	32.4
64	Medtr3g099620	bHLH	2844.5	5.6	0.8	5.7	61.1	26.9
65	Medtr3g102100	AP2-EREBP	93.2	0.0	2.0	1.2	31.2	65.6
66	Medtr3g109310	bZIP	1454.4	2.6	9.8	20.0	42.6	25.0
67	Medtr3g112220	bZIP	108.3	3.0	2.4	1.6	53.6	39.4
68	Medtr3g114080	PHD	6100.2	9.7	15.4	15.7	27.6	31.7
69	Medtr3g115560	Bromodomain	827.7	7.6	12.1	21.6	33.6	25.0
70	Medtr4g008090	C2C2-CO-like	90.2	2.4	1.2	0.0	60.8	35.6
71	Medtr4g015600	C2H2	1411.4	0.4	7.9	28.7	38.9	24.0
72	Medtr4g026485	GRAS	9865.6	6.2	12.3	19.1	35.9	26.5
73	Medtr4g056600	C2H2	2125.2	7.3	7.0	2.6	24.0	59.1
74	Medtr4g063735	Homobox-WOX	771.0	0.1	0.0	2.1	54.8	43.1
75	Medtr4g064873	C2H2	276.4	0.0	6.1	23.6	36.8	33.5
76	Medtr4g068000	Nin-like	62296.8	0.1	0.1	0.3	46.4	53.2

77	Medtr4g075480	bHLH	1200.6	9.9	12.1	12.7	32.4	32.8
78	Medtr4g077970	HSF-typ	1574.2	8.3	10.9	16.2	42.7	21.9
79	Medtr4g078170	Znf-B	952.9	6.5	9.6	13.4	33.6	36.8
80	Medtr4g082060	C2C2-Dof	382.2	12.6	6.4	9.0	35.4	36.6
81	Medtr4g087300	bHLH	29041.9	7.4	8.8	20.1	36.8	26.9
82	Medtr4g090460	Bromodomain	1638.1	9.0	9.8	20.3	37.8	23.1
83	Medtr4g094292	C3H	5633.1	5.1	6.2	5.9	38.9	44.0
84	Medtr4g094420	C2H2	2089.0	8.2	8.3	24.9	34.8	23.7
85	Medtr4g094588	WRKY	3669.3	3.0	0.8	2.3	45.5	48.5
86	Medtr4g097440	bZIP	2090.4	3.0	7.0	7.1	39.2	43.6
87	Medtr4g098560	C2H2	3927.3	8.4	13.6	22.5	30.0	25.4
88	Medtr4g107230	MYB-HB-like	14986.4	6.2	6.4	5.4	31.6	50.3
89	Medtr4g110040	bHLH	894.9	10.1	16.9	8.1	28.7	36.2
90	Medtr4g113200	PAZ-Argonaute	629.8	0.7	0.9	0.2	31.6	66.6
91	Medtr4g114970	A20-like	1149.8	6.3	11.0	18.4	43.1	21.2
92	Medtr4g130190	WD40-like	1451.0	3.1	9.5	22.1	42.4	23.0
93	Medtr4g132540	JmjC	484.8	7.9	12.7	8.5	41.2	29.8
94	Medtr5g010470	MYB-HB-like	87.9	10.1	7.8	3.7	32.6	45.7
95	Medtr5g013690	C2H2	3203.7	5.5	6.9	19.4	41.0	27.2
96	Medtr5g015090	bZIP	1827.8	0.9	11.0	12.6	35.6	39.9
97	Medtr5g015680	BTB-POZ	7145.0	1.8	2.8	6.5	33.3	55.5
98	Medtr5g024730	C2H2	9202.6	3.3	4.8	13.9	35.2	42.8
99	Medtr5g026530	C2H2	13325.1	7.1	8.6	19.9	31.0	33.4
100	Medtr5g027440	MYB-HB-like	2479.1	0.7	1.9	1.4	49.6	46.4
101	Medtr5g041530	C2C2-Dof	110.7	4.1	7.7	3.8	30.8	53.6
102	Medtr5g041940	NAM	217.4	3.0	3.6	27.1	35.6	30.8
103	Medtr5g046670	SBP	1104.8	11.6	7.9	13.2	31.5	35.8

104	Medtr5g064990	C2H2	1874.0	0.1	5.5	21.4	41.2	31.9
105	Medtr5g066730	BTB-POZ	1126.5	9.6	11.4	17.5	24.6	36.9
106	Medtr5g069490	Bromodomain	274.3	0.4	0.0	13.7	53.4	32.5
107	Medtr5g070310	CCHC(Zn)	6783.0	5.2	7.1	16.5	39.2	32.0
108	Medtr5g083030	C2H2	2662.3	4.5	17.1	11.4	33.4	33.6
109	Medtr5g083340	AP2-EREBP	370.9	11.1	6.5	3.8	28.9	49.8
110	Medtr5g085100	AP2-EREBP	111.2	12.4	9.0	12.7	32.0	34.0
111	Medtr6g015060	PHD	6883.3	10.7	10.0	13.0	30.3	36.0
112	Medtr6g021730	C2H2	4796.7	2.8	2.7	11.8	41.2	41.5
113	Medtr6g047750	GRAS	14251.8	4.2	4.6	18.0	43.3	29.9
114	Medtr6g083900	C2H2	11493.2	2.6	2.7	5.9	37.8	51.0
115	Medtr6g090505	FAR	68.2	0.0	2.4	0.0	34.5	63.2
116	Medtr6g453220	MYB-HB-like	146.7	0.0	0.0	13.5	51.1	35.4
117	Medtr7g005280	NAM	13331.4	1.7	5.6	7.2	46.2	39.3
118	Medtr7g007010	AS2-LOB	40267.9	0.7	2.5	23.7	40.1	33.0
119	Medtr7g010950	C2C2-Dof	125.8	8.5	6.9	3.2	20.8	60.7
120	Medtr7g017100	WD40-like	13311.3	7.6	8.5	18.5	23.9	41.5
121	Medtr7g020870	MYB	12482.1	1.7	10.4	26.0	32.3	29.6
122	Medtr7g022190	WD40-like	895.2	9.5	9.9	9.2	29.6	41.8
123	Medtr7g059360	C2H2	271.0	4.4	10.2	5.5	48.5	31.4
124	Medtr7g061980	FAR	992.0	0.0	0.5	0.5	70.9	28.2
125	Medtr7g068770	MYB	561.0	0.7	0.6	3.3	56.3	39.1
126	Medtr7g074450	WD40-like	11801.4	4.0	6.7	10.7	40.5	38.2
127	Medtr7g078150	C2H2	1379.6	4.4	4.9	11.7	36.2	42.7
128	Medtr7g079840	C3H	2178.9	6.4	8.7	9.7	27.6	47.7
129	Medtr7g081815	AP2-EREBP	17118.5	4.6	10.1	15.7	48.0	21.7
130	Medtr7g084810	WD40-like	6060.0	10.6	11.2	10.4	31.7	36.0

131	Medtr7g093600	C2H2	18626.0	2.9	5.2	6.0	49.7	36.2
132	Medtr7g101110	WD40-like	3848.0	9.5	10.6	15.1	27.5	37.4
133	Medtr7g101150	C2H2	1131.0	9.6	11.0	8.4	35.8	35.3
134	Medtr7g102990	C2H2	1913.3	2.4	0.4	4.0	55.9	37.3
135	Medtr7g106320	Homobox-WOX	3727.9	0.2	0.2	0.2	30.0	69.4
136	Medtr7g108470	Homeodomain-LIKE	941.2	4.5	7.7	3.7	26.8	57.3
137	Medtr7g451400	Znf-LSD	3188.3	11.1	10.1	6.2	29.0	43.6
138	Medtr8g020480	Hap3/NF-YB	3589.2	9.2	11.8	12.2	29.7	37.0
139	Medtr8g032690	C2H2	2181.6	0.2	5.1	25.6	33.5	35.5
140	Medtr8g037260	PHD	6156.0	3.5	5.9	16.7	32.3	41.6
141	Medtr8g038620	RR-A-type	2565.8	3.9	9.3	7.5	47.6	31.6
142	Medtr8g073335	WD40-like	6760.1	0.8	5.8	4.2	34.5	54.8
143	Medtr8g078410	LIM	1928.6	6.2	7.1	14.4	37.6	34.8
144	Medtr8g079910	C2H2	1029.2	6.0	9.2	5.8	30.5	48.5
145	Medtr8g089310	Znf-LSD	3785.2	8.0	8.2	17.5	39.8	26.5
146	Medtr8g091650	bZIP	562.5	3.6	5.0	12.9	40.5	38.0
147	Medtr8g098665	C2H2	1203.9	10.0	11.4	14.3	32.0	32.3
148	Medtr8g107350	HMG	98.1	9.1	12.3	12.8	34.1	31.8
149	Medtr8g432590	WD40-like	366.6	10.0	10.3	11.2	27.1	41.4
150	Medtr8g447220	C2H2	891.5	14.7	3.8	5.7	28.6	47.2
151	Medtr8g465910	C2H2	55.6	9.7	13.4	0.0	35.2	41.7

Publication list

Senlei Zhang, Éva Kondorosi, and Attila Kereszt. **2019.** An anthocyanin marker for direct visualization of plant transformation and its use to study nitrogen-fixing nodule development. Journal of plant research. 132(5), 695-703.

Zhengxi Sun, Youning Wang, Fupeng Mou, Yinping Tian, Liang Chen, **Senlei Zhang**, Qiong Jiang and Xia Li. **2016**. Soybean reveals auxin-responsive microRNAs that are differentially expressed in response to salt stress in root apex. Frontiers in plant science. 6, 1273

Acknowledgements

Deep gratitude must be expressed to Dr. Éva Kondorosi, for her generous acceptance and support towards me to conduct my PhD research in the labor, and for the encouragement given to me to explore the unknown. I must say thanks to my supervisor, Dr. Attila Kereszt, for his guidance during this five-year research experience. Attila impressed me with his active mind and his passion for science. He is not only a supervisor to me, but also a great colleague and a friend. I highly appreciate the time and energy that Éva and Attila spent on me.

I am extremely grateful to work with Dr Jesús Montiel, Dr. Edit Ábrahám, Rui Lima, Attila Farkas and Máté Makra. I thank them for the help they provided. The discussion and debate we have had keeps inspiring me on my research. I thank them also for being accommodating towards me for their tolerance on my shortcomings as friends.

I must thank my mother and my elder brother for their support and for their understanding of my pursuit of doing scientific research.

Many thanks must be addressed to my wife, Ting Wang, for being always supportive and for not just being an amazing life mate but also a wonderful research collaborator.