HUNGARIAN UNIVERSITY OF AGRICULTURE AND LIFE SCIENCES

CLONING AND FUNCTIONAL ANALYSIS OF THE GENE RESPONSIBLE FOR THE INEFFECTIVE SYMBIOTIC PHENOTYPE OF *MEDICAGO TRUNCATULA* MUTANT *SYM20* (TRV43)

DOI: 10.54598/004920

BERIVAN GÜNGÖR

GÖDÖLLŐ

2024

The PhD School

Name: Doctoral School of Biological Sciences

Discipline: Biology

Head: Prof. Dr. Zoltán Nagy
Professor, DSc
MATE, Institute of Agronomy
Department of Plant Physiology and Plant Ecology

Supervisor(s): Dr. Péter Kaló
PhD, DSc
MATE, Institute of Genetics and Biotechnology
Department of Microbiology and Applied Biotechnology

.....

Approval of the Head of Doctoral School
Dr. Zoltán Nagy

.....

Approval of the Supervisor
Dr. Péter Kaló

TABLE OF CONTENTS

1. ABBREVIATIONS

AON: Autoregulation of Nodulation

BAC: Bacterial Artificial Chromosome

cDNA: Complementary DNA

CEP7: Cysteine-Rich Peptide 7

Chr: Chromosome

CK: Cytokinin

CLE13: Clavata3/Esr-Related 13

cM: Centimorgan

CCaMK: Calcium Calmodulin dependent protein kinase

CNV: Copy Number Variation

CO2: Carbon Dioxide

CRISPR/Cas9: Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR:associated protein 9

CV: Cultivar

DEFLs: Defensin-like peptides

DELLA1/2/3: DELLA Protein 1/2/3

DMI3: Does not Make Infection3 (Calcium/Calmodulin dependent Protein Kinase in *M. truncatula*)

DNF: Does Not Fix nitrogen

dpi: days post inoculation

dsDNA: Double:Stranded DNA

dsRed: Discosoma sp. red fluorescent protein

E: Elongated bacteroids

EB: Elongated-Branched Bacteroids

eGFP: Enhanced green fluorescent protein

ENOD11: Early Nodulin 11

ERN1: Ethylene Responsive Factor Required for Nodulation 1

ESTs: Expressed Sequence Tags

Fix- : Ineffective nitrogen -fixing mutant

FSC: Forward Scatter

GBS: Genotyping-By-Sequencing

gDNA: Genomic DNA

GFP: Green Fluorescent Protein

GRPs: Glycine-Rich Peptides

GUS: β-Glucuronidase

INDELs: Insertions/Deletions

IPD3: Interacting Protein of DMI3

IPTG: Isopropyl β:D:1-thiogalactopyranoside

IRLC: Inverted Repeat Lacking Clade

IT: Infection thread

Km: Kanamycin

LCM: Laser Capture Microdissection

LCO: Lipochitooligosaccharide

LG: Linkage group

LysM: Lysin Motif

Mbp: Million base pairs

mCherry: A red fluorescent protein

MtGEA: Medicago truncatula Gene Expression Atlas

MtLYK3: Medicago truncatula Lysin Motif Receptor-like Kinase 3

MtNFP: Medicago truncatula Nod Factor Perception

N_2: Nitrogen

NCR: Nodule-specific Cysteine-Rich

NF: Nod Factor

NF-YA: Nuclear Factor YA

NF-YB16: Nuclear Factor YB16

NF-YC2: Nuclear Factor YC2

NGS: Next Generation Sequencing

NH_3: Ammonia

NH_4NO_3: Ammonium Nitrate

NIN: Nodule Inception

NO_3^-: Nitrate

Nod -: non-nodulating mutant

Nod +: effective symbiotic mutant

NSP: Nodulation Signaling Pathway

NSS: N-terminal Secretion Signal

OD: Optical Density

PAM: Protospacer Adjacent Motif

PBS: Phosphate Buffered Saline

PCR: Polymerase chain reaction

PI: Propidium Iodide

pI: Isoelectric point

PPA: Paraformaldehyde

RirA: Rhizobium Iron-Regulated protein A

RNAseq: RNA Sequencing

RR1/4: Response Regulator 1/4

RT: Room Temperature

S: Swollen bacteroids

SEM: Scanning Electron Microscopy

sgRNA: Single guide RNA

Sm: Streptomycin

Sm/SpR: Streptomycin/Spectinomycin Resistance

SNP: Single Nucleotide Polymorphism

SOB: Super optimal broth

SpCas9: Streptococcus pyogenes Cas9

SP: Spherical bacteroids

Spr: Spectinomycin

SSC: Side Scatter

SSCP: Single-Strand Conformation Polymorphism

ssDNA: Single-Stranded DNA

SSR: Simple Sequence Repeat

STY: Symbiosis Transcription Factor Y

SV40: Simian virus 40

TFs: Transcription Factors

Trx: Thioredoxin

UBQ10: Ubiquitin 10

wpi: weeks post inoculation

YUCCAs: YUCCA Genes

ZI: nodule zone I, meristem

ZIId: distal nodule zone II, invasion zone

ZIIp: proximal nodule zone II, invasion zone

ZII-ZIII: nodule zone II-III, interzone

IZ: nodule zone II-III, interzone

ZIII: nodule zone III, nitrogen-fixing zone

ZIV: nodule zone IV, senescence zone

2. INTRODUCTION

Nitrogen is a fundamental element for biomolecules, and its bioavailable forms are essential for plant health and ecosystem balance. The nitrogen fertilizers produced by a process known as Haber-Bosch, has improved crop production in recent decades, however fertilizers cause pollution and eutrophication and therefore they can be harmful for the environment. As an ecological alternative, organic matter decomposition presents a sustainable nitrogen fixation method, however application of compost is a slow process and requires labor moreover it tightly depends on specific biological and environmental conditions.

A group of free-living soil bacteria, collectively termed as rhizobia, can fix nitrogen, by establishing symbiosis with a certain group of plants, specifically legumes. The studies conducted with model organism *Medicago truncatula*, reveals the molecular and genetic mechanisms of the nitrogen fixing symbiosis. In *M. truncatula* and in other legumes, the symbiotic interaction starts with a signal exchange. Host plants release flavonoids and following their perception, rhizobia produce the Nod factors in response. The Nod factors prompt the root hair cells of the plants to form infection threads (ITs), through which rhizobia can enter to the root hair cells and colonize the cells of the emerging nodule. The nodules provide a microaerobic environment and produce leghemoglobin to maintain low oxygen levels to provide optimal working conditions for nitrogenase enzyme.

In plant cells, bacteria get encapsulated by a plant derived membrane and form organelle like structures, known as symbiosomes. In certain legumes (Inverted Repeat Lacking Clade legumes and Dalbergoids) Nodule specific Cysteine-Rich peptides (NCRs), which are expressed exclusively in nodule cells, mediate the terminal differentiation of bacteria into nitrogen fixing bacteroids. This process is characterized by endoreduplication and elongation of bacteria in the symbiosomes and results in irreversible differentiation of plant cells.

The genetic background of symbiotic nitrogen fixation has extensively studied for more than 50 years, starting from the genetic mapping of rhizobium strains and leading to the discovery of a complex network of genetic pathways that determine the details of legume-rhizobium interactions. Large mutagenesis screens and subsequent mapping of symbiotic genes on the chromosomes of *M. truncatula*, results in the identification of many crucial genes involved in symbiotic interaction.

The family of *NCR* genes encode over 700 NCR peptides in *M. truncatula* suggesting redundant and collective roles. However, their diversity suggests that specific NCR peptides may have different or non-redundant functions in the differentiation processes.

A number of studies using forward and reverse genetic approaches have demonstrated that *NCR*s are essential for nitrogen fixing symbiosis of *M. truncatula*, and even the absence of a single specific *NCR* can cause a substantial decrease in nitrogen fixing efficiency. The transformation of NCR-mediated symbiosis capabilities to non-leguminous crops would make significant contribution to nitrogen production capabilities of crops. Therefore, understanding the genetic and molecular working mechanism of NCR peptides is important to enhance nitrogen fixation both in legumes and non-leguminous plants.

3. OBJECTIVES

This study extends our knowledge regarding to complexities of symbiotic nitrogen fixation, focusing on the nitrogen fixing-deficient *Mtsym20* mutant of *Medicago truncatula*. The study aims to characterize the symbiotic phenotype and genetic features of the *Mtsym20* mutant, identify the gene responsible for the defective nitrogen fixation, and examine role of the identified gene in symbiosis. In order to achieve the presumed results, the following objectives are set:

- Characterization of the symbiotic nitrogen-fixing phenotype of the *Mtsym20* mutant compared to its wild-type counterpart with microscopic examinations, to determine the specific stage at which the symbiotic nodule formation is blocked.
- Comparison of *Mtsym20* to other nitrogen fixation-deficient mutants to identify common and distinct phenotypic features concerning their nitrogen fixing ability
- Localization of the *Mtsym20* locus in *M. truncatula* genome using genetic mapping
- Identification of the impaired gene within the *Mtsym20* locus which is responsible for the nitrogen fixation defective phenotype of *Mtsym20* and confirming the gene identity with genetic complementation assay.
- Examination of the function of the identified gene required for nitrogen fixing symbiosis in *M. truncatula*

4. LITERATURE BACKGROUND

4.1. The significance of symbiosis in nitrogen fixation

Nitrogen (N_2) is a key element in maintaining the balance of the ecosystem and it is a crucial component of the fundamental molecular structures of organisms, such as nucleic acids, amino acids, vitamins, signaling molecules, etc. Additionally, plants contain nitrogen in their chlorophyll and ATP which are essential factors for photosynthesis. Thus, a deficiency of nitrogen in plants can significantly affect plant growth and development in negative manner, leading to reduced crop yields and disruptions in ecological balance. Nitrogen is the most abundant element in the atmosphere, comprising approximately 78% of the air, and it is transferred to soil and water in a process known as nitrogen cycle. However, plants can only uptake the nitrogen in combined forms such as ammonia (NH_3) or nitrate (NO_3^-) (Crawford and Forde, 2002). The reduction of N_2 requires a very high level of energy to break the covalent triple bond between two nitrogen atoms (Postgate, 1982). Such energy can be obtained naturally from lightening. However it is not a practical solution for nitrogen fixation as it is sporadically occur depending on the geographical conditions and yet produce insufficient amount of fixed nitrogen for agricultural use.

The Haber-Bosch process is a widely used method in modern industry for producing nitrogen rich fertilizers to supply the nitrogen requirements of global crop production. In the Haber-Bosch process, the production of ammonia and/or nitrate is catalyzed under high temperatures, varying between 400 °C to 500 °C, to create the energy necessary for breaking the nitrogen bonds and facilitating the combination of nitrogen atoms with the hydrogen. This method provides significant amount of combined nitrogen and thus contribute largely to crop production and stable food supply. However, Haber-Bosch process is an exothermic reaction releasing a significant amount of carbon dioxide (CO_2) to the atmosphere. Additionally, hydrogen production is utilized largely from fossil fuels which increase the level of carbon emission. Consequently, the large-scale use of these chemical fertilizers cause a significant air pollution and speed up the climate change. Additionally, the extensive use of these fertilizers leads to amounts of N that cannot be fully absorbed by plants, resulting in nitrogen accumulation in surface and groundwater. This accumulation results in overproduction of green algae and the death of algae leads to reduced oxygen levels in waters. This phenomenon which is known as eutrophication endangers the natural life in water (Helsel, 1987; Galloway et al., 1995).

Decomposition of organic matter is an environmentally friendly alternative way to produce fixed nitrogen. In this process the free-living heterotrophic organisms, such as Azotobacter, Bacillus,

Clostridium, and Klebsiella, oxidize the organic matter to use it as a source of carbon and energy and produce fixed nitrogen without interaction with any other organism (ammonification). However, the neutral decomposition process requires availability of optimal organic matter, optimal soil moisture and temperature as well as the presence of decomposing bacteria. Even when conditions are optimized for decomposition, the process is slow and its application to plants is demanding additional time and labor. Moreover, the decomposition in large scale in agriculture potentially can cause release of disease due to accumulation of pest and plant pathogens (Vadakattu and Paterson, 2006). Therefore the contribution of decomposition to nitrogen requirements of crop production is limited by biological and environmental constraints.

Certain microorganisms are able to fix nitrogen and can obtain organic supply from plants in a process known as associative nitrogen fixation. In this process, soil bacteria colonize the rhizosphere or the outer layer of the root in a niche where it can benefit from sugars produced by host plants in exchange to fixed nitrogen. The association of Azospirillum with various grasses including rice, wheat, corn, oats, and barley, is the most widespread example of associative nitrogen fixation (Vlassak and Reynders, 1979).

Unlike to associative nitrogen fixing bacteria, some microorganisms establish nitrogen fixing symbiosis with plants by invading their root cells and inducing the development of nodules. Symbiotic nodule cells provide a microaerobic environment facilating the nitrogen fixation of bacteria and utilize the fixed nitrogen meanwhile bacteria obtain carbon sources from the plant. An example of symbiotic nitrogen fixation is established between Frankia and a large variety of non-legumnous plants containing 8 angiosperm family with over 250 species including trees and shrubs, such as Alder (*Alnus sp.)* and sea buckthorns (Hippophae) (Benson and Silvester, 1993).

The associative nitrogen fixation and Frankia symbiosis importantly contribute the ecological nitrogen fixation. However, the symbiotic relationship between rhizobia and legumes is the most widespread and effective form of symbiotic nitrogen fixation in context of both crop production and ecological management. Rhizobia is a name used for soil delving, rod shaped, Gram negative group of Alphaproteobacteria and Betaproteobacteria and includes several genera such as *Rhizobium*, *Bradyrhizobium*, *Sinorhizobium*, *Mesorhizobium* and *Allorhizobium. Rhizobium* is the largest genus of rhizobia containing 112 species (Sprent, 2008; Mousavi et al., 2016; de Lajudie et al., 2019). Different rhizobia may adapt different environmental conditions or establish symbiosis with different legumes enlarging the extent of the nitrogen-fixing symbiosis.

4.2. Symbiotic nitrogen fixation in legumes

The soil contains large variety of microorganisms containing both harmful and beneficiary bacteria. The root cells of the plants form a strong barrier in order to prevent invasion of adverse microorganisms. However, rhizobia can invade the root cells of legumes following a molecular dialogue of an initial signal exchange between the two symbiotic partners (Perret et al., 2000; Gough and Cullimore, 2011; Oldroyd et al., 2011).

The symbiosis between rhizobia and legumes is initiated when rhizobia recognize the flavonoids secreted by the host roots in response to nitrogen-deficient conditions (Fig. 4.2). Flavonoids are intensively released around root tips and in concentrations ideal for symbiosis in the emerging root hair zone. Nodulation factor (Nod factor, NF) production is initiated upon flavonoids secreted by *Medicago* root cells are perceived by rhizobia (Fig. 4.2 step 1). The presence of flavonoids activates rhizobial NodD transcription factors (TFs) which induce the expression of *nod* genes. The *nod* genes encode proteins essential for synthesis and extracellular transportation of bacterial nodulation (Nod) Factor (NF) (Downie and Walker, 1999; Geurts and Bisseling, 2002; Cooper, 2004; Gage, 2004; Oldroyd et al., 2011). The recognition of NFs by host plant induces rapid increase of intracellular calcium ion (Ca+) levels in the cytoplasm of root hair cells, a phenomenon known as calcium spiking (Felle et al., 1998). The mechanism and effects of calcium spiking are not yet fully understood; it possibly plays a role in the subsequent activation of signaling pathways during the bacterial infection of host cells. This activation induces a series of structural alterations in the root hair cytoskeleton, enabling the entry of rhizobia into the root hair cells. The rearrangement of actin filaments and microtubules in the cytoskeleton of the cells results in curling of the root hairs (Oldroyd and Downie, 2004; Stacey et al., 2006). The curled root hairs entrap the rhizobia into an infection pocket. The rhizobia-produced NF stimulates cell division of surrounding cortical cells (either inner cortical cells in temperate legumes like *Medicago truncatula* or outer cells in tropical legumes like *Lotus japonicas*) leading to formation of the nodule primordium (Fig. 4.2-step 2) The cells covering the infection pocket degrade and form a tube-like structure, known as infection thread (IT), in which rhizobia can enter the root tissue (Fig. 4.2-step 3). The infection thread passes rhizobia to invade the cortical cells of the root trough crossing cellular boundaries and reaching the nodule primordium (Fig. 4.2-step 4). The rhizobia in the infection thread continuously grow and divide. Simultaneously, nodule primordium develops a meristem at the apex of the emerging nodule primordia. The infection threads ramify in the nodule primordium and penetrate the cells derived from meristem of nodule primordium. Penetrated cells allow bacteria to exit the IT and ensure that sufficient number of host cells become

colonized. As infection threads progress, rhizobia are released into the inner cells of the root. The rhizobia get encapsulated by a plant derived membrane known as peribacteroid membrane, through an endocytotic-like process (Fig. 4.2-step 5). (Heidstra and Bisseling, 1996; Gage, 2004; Jones et al., 2007; Oldroyd and Downie, 2008; Murray, 2011). The peribacteroid membrane forms a barrier between the bacterium and the host cell which prevents potential pathogenic interactions and enables the nutrient exchange between two symbiotic partners (Udvardi and Day, 1997; White et al., 2007). The bacteria, encapsulated with peribacteroid membrane, transform into organelle-like structures, known as symbiosomes (Roth and Stacey, 1989; Brewin, 2004). Bacteria begin to divide in symbiosomes and differentiate into bacteroids to start biological nitrogen fixation within the plant cells (Fig. 4.2-step 6). The differentiation process in root cortical, pericyclic, and endodermal cells results in development of a novel organ-like structures 'Nodule' (Fig. 4.2-step 7).

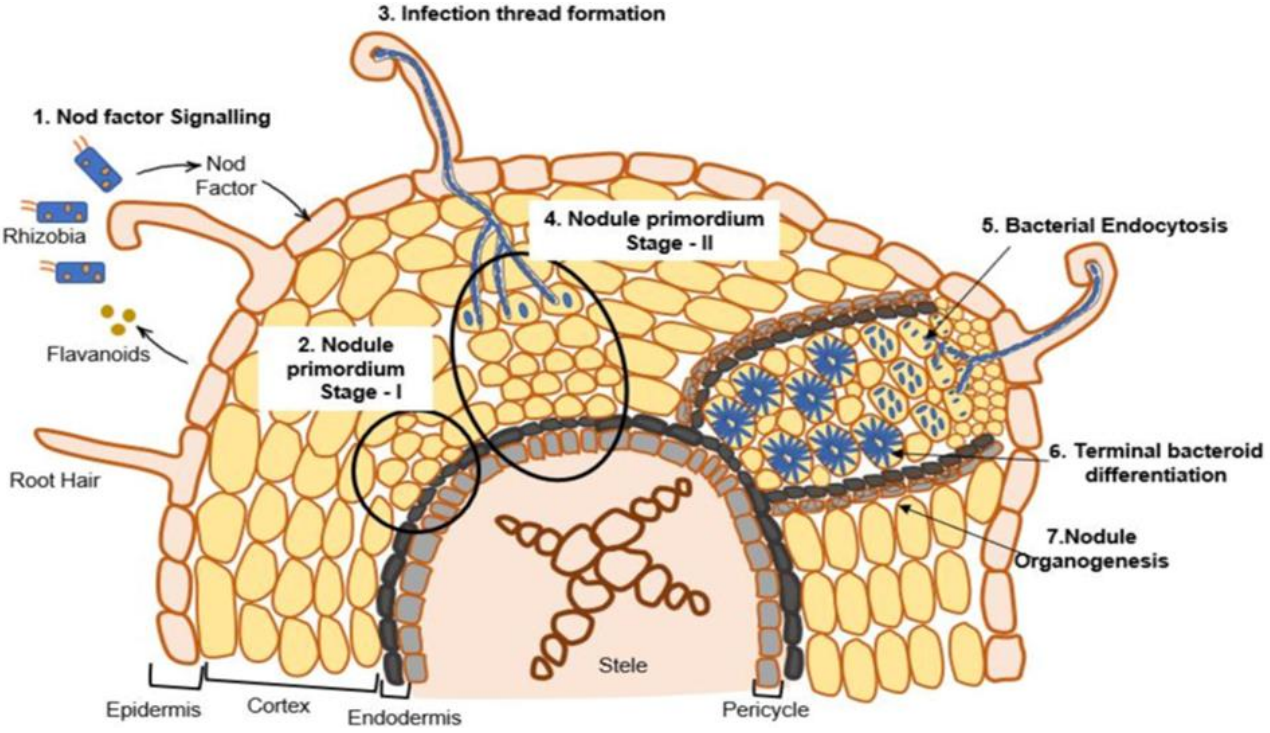


Fig. 4.2 Nodule - a *de novo* root organ development

The *Medicago* root nodule development is shown in 7 steps (From Bhardwaj and Sinharoy, 2022)

4.2.1. Nod factor signaling pathway

The NFs are lipochitooligosaccharide (LCO) signal molecules which possess different, generally 4-5, numbers of N-acetyl glucosamine residues consisting of an acyl chain. The *nodABC* operon genes of rhizobia encode for synthesis of the core structure of Nod factors, while other *nod*

genes determine the host specificity through the addition of residues, modifications to the length of the acyl chain or alterations to the backbone structure (Jones et al., 2007; Crespi and Frugier, 2008).

The NFs are recognized by specific receptor proteins located on the membrane of the root hair cells (Fig. 4.2.1a). These receptors are the members of the LysM (Lysin Motif) receptor-like kinase family and includes MtNFP and MtLYK3 in *M. truncatula* (Arrighi et al., 2006; Smit et al., 2007). NF perception induces calcium spiking in the root hair and epidermal cell nuclei and in the periphery of nuclei. The calcium oscillations are interpreted by the plant via a specific protein kinase, DMI3, which is a calcium/calmodulin-dependent protein kinase (CCaMK). DMI3 phosphorylates and activates IPD3 (the interacting protein of DMI3) transcription factor which is involved in IT formation and the release of bacteria (Fig. 4.2.1a) (Horváth et al., 2011; Ovchinnikova et al., 2011). The interaction between DMI3 and the MtIPD3 transcription factor leads to the activation of IT guiding transcriptional regulatory complexes such as *Nodulation Signaling Pathway 1* and *2* (NSP1 and NSP2) (Kaló et al., 2005; Smit et al., 2005). This is followed by interactions of the NSP2 and MtIPD3 with DELLA1/2/3 proteins (Fonouni-Farde et al., 2016). The resulting protein complexes upregulate the expression of *NIN* (*Nodule Inception*) transcription factor (Singh et al., 2014; Jin et al., 2016). The activated NIN proteins bind to the promoter regions of symbiotic genes, including those involved in IT formation and nodule organogenesis, and induce their expression. *NIN* upregulation activates the *NF-YA (Nuclear Factor Y)* gene leading to formation of heterotrimeric complexes such as NF-YA1, NF-YB16, and NF-YC2 (Laporte et al., 2014; Baudin et al., 2015). The ERN1 (Ethylene Responsive Factor Required for Nodulation 1) transcription factor is activated by IPD3, NSP1/NSP2, NF-YA1/A2 and initiate expression of early nodulation genes such as *ENOD11*. *NIN* simultaneously repress the expression of *ENOD11* (cell wall-associated protein) and activates the CEP7 and CLE13 peptides involved in autoregulation of nodulation (AON) which coordinates root-shoot signals to control nodule formation (Fig. 4.2.1a) (Cerri et al., 2012; Liu et al., 2019; Laffont et al., 2020).

Meanwhile, cytokinin biosynthesis is induced to promote cell division in the root cortex (Fig. 4.2.1b). Cytokinin and auxin, through their respective effectors RR1/4 and DELLA-type transcription factors, play a significant role in initiating transcriptional changes in response to Nod factor signaling. This process activates a large number of genes (*NF-YA1* and *LBD16*, *STY* and *YUCCA*s), known as nodulins. Nodulins are crucial for the infection and invasion processes, the functioning of bacteroids, and the assimilation of fixed nitrogen within the developing nodule (Fig. 4.2.1b) (Vernié et al., 2008, 2015; Laporte et al., 2014; Liu et al., 2019; Schiessl et al., 2019; Dong et al., 2021; Jarzyniak et al., 2021).

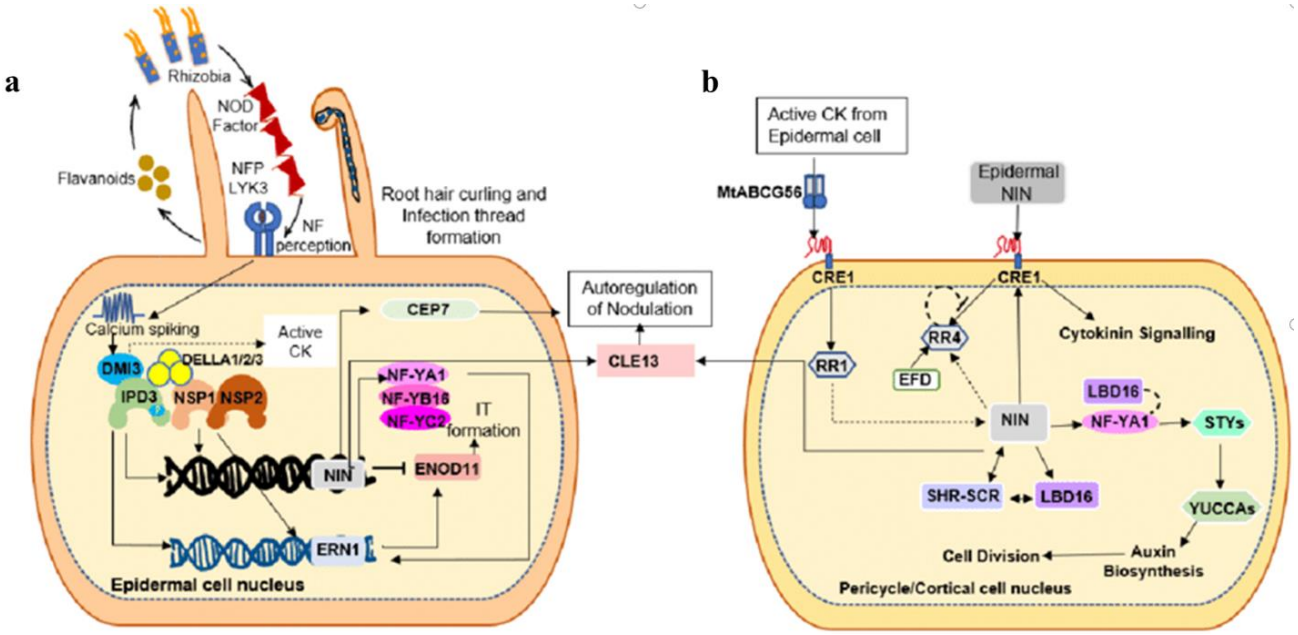


Fig. 4.2.1 Schematic representation of the NF signaling pathway in root epidermal cells and cortical cells during nodule development.

Transcription factors and signaling events occur in the epidermal cells and root hair cells during rhizobial infection and root hair curling, IT formation, and propagation (a). Transcription factors and signaling events occur during the division of pericycle or cortical cells to form nodule primordia (b). (Bhardwaj and Sinharoy, 2022)

4.2.2. Nodule organogenesis

The nodules developed on leguminous roots can be distinguished into two groups based on their morphological characteristic as determinate and indeterminate nodules (Fig. 4.2.2a) (Franssen et al., 1992; Maunoury et al., 2008, 2010). Determinate nodules are found mainly on the roots of tropical legumes such as *Glycine* (soybean), *Phaseolus* (common bean), *Vigna* and Lotus. Cell division in the cortex of determinate nodules ceases following the formation of nodule primordium, thus determinate nodules do not include a persistent meristem. Consequently, the size and shape of the nodules is determined in the early stage of the development. The growth of determinate nodules is achieved by the expansion of the divided cells and the mature nodule cells contain a homogenous population of nitrogen-fixing bacteroids (Fig. 4.2.2a).

The indeterminate nodules are found in a large group of temperate legumes including, *Medicago* (alfalfa), *Trifolium* (clover), *Vicia* (vetch), *Pisum* (pea) and all mimosoid legumes. Indeterminate nodules are characterized by an active apical meristem harboring continuously proliferating cells leading to formation of cylindrically shaped nodules which, in some cases, form extensive branches (Xiao et al., 2014). The continuous division of meristematic cells result in zonation

in indeterminate nodules including meristematic zone (ZI), infection zone (ZII), interzone (ZII-ZIII), nitrogen fixation zone (ZIII) and senescence zone (ZIV) (Fig. 4.2.2b). At the apical of nodule, following the meristematic zone downwards, bacteria are released from the infection threads and begin to invade the plant cells, forming the infection zone (ZII). The cells in the infection zone are notably larger comparing to those found in meristematic zone and their cell division is halted. The infection zone is followed by the interzone/transition zone (ZII-ZIII) where bacteria begin to transform into terminally differentiated form of nitrogen-fixing bacteroids. This is followed by nitrogen fixation zone (ZIII) containing large central vacuoles surrounded with fully differentiated nitrogen fixing bacteroids. The presence of plant-derived leghemoglobin can be detected by pink color in the nitrogen fixing zone of matured nodules. The leghemoglobin binds oxygen and provides low-oxygen conditions essential for the proper operation of the nitrogenase enzyme. At the basal part of the matured indeterminate nodule which is known as senescent zone (ZIV), degraded plant cells and bacteroid are found therefore in this zone nitrogen fixation does not occur and the bacteria become saprophytic (Timmers et al., 2000; Pérez Guerra et al., 2010). The presence of senescent is easily observable in matured nodules due to yellowish-greenish color which derives from the degradation of heme component of the leghemoglobin.

In determinate nodules, bacteroids preserve their morphology which resembles to the rod shape of free-living rhizobia. These bacteroids are named as U type bacteroids, and their differentiation is reversible, meaning that they can revert to their free-living state once released from nodule tissue. In indeterminate nodules, nitrogen fixing bacteroids lose their reproduction ability and cannot colonize once isolated from nodules suggesting that bacteria are terminally differentiated (Mergaert et al., 2006). The membrane permeability is increased in these bacteroids allowing more effective metabolite exchange between the host cell and bacteroid. The terminal differentiation of bacteria includes an elongation step where 1- to 2-μm-long bacteria elongate into 5 to 10 μm long bacteroids (Mergaert et al., 2006; Van de Velde et al., 2010; Montiel et al., 2017). Any failure in differentiation process results in death of bacteria and prevents the subsequent nodule developmental steps (Werner et al., 1984). The terminally differentiated bacteroids are classified in different groups based on their morphology such as swollen (S), elongated (E), spherical (SP), and elongated–branched (EB), depending on their host-species (Vasse et al., 1990; Mergaert et al., 2006; Bonaldi et al., 2011; Montiel et al., 2016).

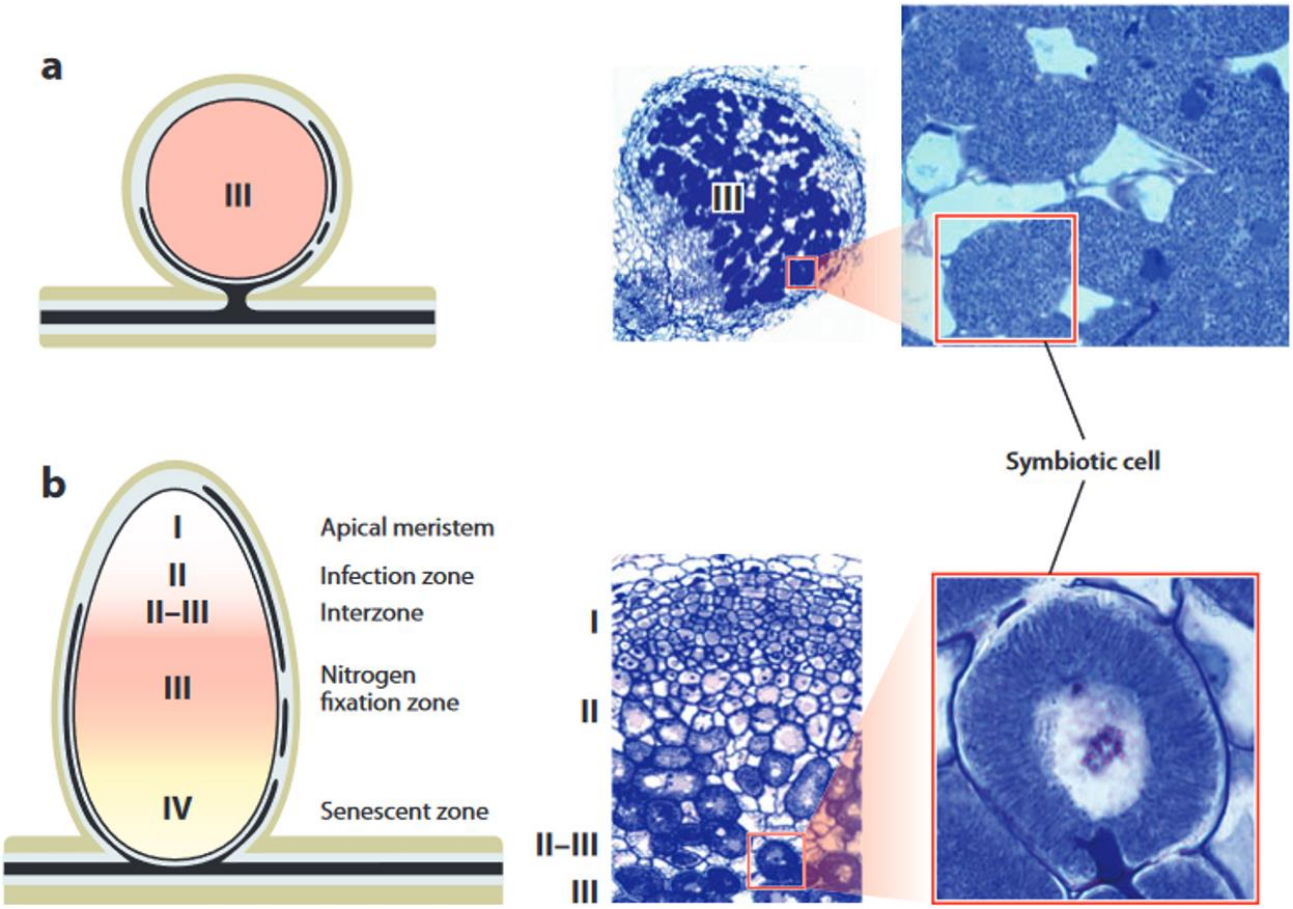


Fig. 4.2.2 Structure of determinate and indeterminate root nodules

Determinate nodules are spherical, and the central region (III) contains uninfected cells and nitrogen-fixing symbiotic cells. Enlargement of the symbiotic cells shows the presence of endosymbionts in the host cytoplasm (a). Indeterminate nodules have an apical meristem (I), an infection zone (II), an interzone (II–III), the nitrogen fixation zone (III), and, in older nodules, a senescent zone (IV). Symbiotic cells differentiate along zone II and interzone II–III reaching the final state in zone III. The nucleus and cell volume of infected plant cells grow gradually with repeated endoreduplication cycles, and the nitrogen-fixing cells are fully packed with the endosymbionts (b). (Kondorosi et al., 2013)

Infected cells of the indeterminate nodules undergo extreme levels of enlargement resulting in growth of nodules. In these nodule cells, level of DNA content varies between 64C or 128C (C is referring to haploid DNA content) (Cebolla et al., 1999; Kondorosi et al., 2000; González-Sama et al., 2006). The DNA content of the U-type bacteroids is 1C/2C similar to free-living cultured rhizobia, whereas differentiated bacteroids, similar to nodule cells, show an increased ploidy level reaching up to 32C (González-Sama et al., 2006; Guefrachi et al., 2014; Czernic et al., 2015; Montiel et al., 2016). This growth is mediated with a modified cell cycle where repeating genome replications take place without mitosis and cytokinesis. This modified cell cycle is known as endoreduplication. The level of endoreduplication is tightly linked to elongation of the bacteroids (Mergaert et al., 2006).

4.3. *Medicago truncatula*: A model organism for rhizobia-legume symbiosis

Medicago species can establish symbiotic relationship with various rhizobia strains. *M. truncatula* in particular has a various genotypes or ecotypes enabling different specific symbiotic interactions with different rhizobia strains. This diversity makes the *M. truncatula* an important species to explore host specific symbiosis. Some of these genotypes and ecotypes are highly compatible for genetic transformation. Additionally, *M. truncatula* has a genome as small as ~500 Mbp (2n=16) (Bennett and Leitch, 1995) and its genome sequence is extensively progressed (Pecrix et al., 2018). Assessing the impact of genetic alterations on the symbiotic phenotype of *M. truncatula* is relatively easy due to the short generation time and high seed yield. For these reasons, *M. truncatula* has been selected as a model organism to understand symbiotic process occurring between the rhizobia and leguminous plants.

4.3.1. *Medicago truncatula* database

The first draft genome assembly (Mt3.5) of *M. truncatula* A17 was sequenced with conventional BAC-tiling path using Sanger sequencing, high-quality optical mapping methodologies and Illumina shotgun sequencing (Young et al., 2011). The Mt3.5 version of genome assembly spanned to 418Mb containing 60,143 sequence contigs. The gaps in this earliest genome assembly were attempted to be filled with expressed sequence tags (ESTs) (Bell et al., 2001) and gene expression datasets.

The Mt3.5 genome assembly version of A17 genome was updated to Mt4.0 in 2014 representing significant improvements. The *M. truncatula* genome (Mt4.0) was sequenced mostly based on *de novo* whole genome shotgun assembly of the reads sequenced on Illumina and 454 platforms. These reads were assembled using ALLPATHS-LG, a *de novo* assembly method specifically used for large genomes, resulting in creation of large segments of sequences known as scaffolds. Scaffolds were positioned within the genome considering their comprehensive alignments with the genomic maps such as the optical map and the genotyping-by-sequencing (GBS) map (Tang et al., 2014). *M. truncatula* genome (Mt4.0) combined the sequences in 10,160 contigs spanning 412 Mbp.

M. truncatula genome (Mt4.0) offered more genomic content comparing to earlier genome assembly (Mt3.5), covering significant sequence of all 8 chromosomes. However, some of the sequenced genes or genetic regions, although sequenced successfully, could not be positioned in a

precise location within the chromosomal structures and thus located on separate scaffolds. Therefore an important portion of the genome remained unresolved, resulting in gaps on the genome database.

The symbiotic functions of the genes from *M. truncatula* genome assembly Mt4.0 was investigated in a study using the RNA sequencing data simultaneously obtained from whole uninfected roots and nodule tissues of *M. truncatula.* Additionally zone specific expressions of the genes were evaluated using laser-capture-microdissection of nodules (LCM) (Roux et al., 2014). The data collected from the study of Roux and co-workers (2014) is presented on the Symbimics data gateway (iant.toulouse.inra.fr/symbimics/). These comparative RNAseq data analysis revealed a total of 5,513 genes, which are strongly upregulated in nodules compared to roots of *M. truncatula*.

Following the release of *M. truncatula* A17 genome assembly Mt4.0, the reference genome sequence of *M. truncatula* R108 genotype - recently renamed as *Medicago littoralis* R108 (Choi et al., 2022) - was released in in 2017 (Young et al., 2011; Tang et al., 2014; Moll et al., 2017). Beside these genome assemblies, a variety of *M. truncatula* accessions spanning to 384 lines were sequenced using Illumnia Next Generation Sequencing Technology (NGS) to discover sequence variation across *M. truncatula* ecotypes and genotypes in an international cooperation program named as Medicago Hapmap project (www.medicagohapmap.org). In this project several *M. truncatula* accessions were sequenced to identify Single Nucleotide Polymorphisms (SNPs), insertions/deletions (INDELs) and copy number variations (CNVs) (Kang et al., 2015). The data obtained from Medicago HapMap project provided insights in comparative genome analysis of different *M. truncatula* accessions.

The genome assembly of *M. truncatula* was further improved recently, using high-depth long-read (PacBio) sequencing in addition to BioNano technology optical maps. This most recently released genome database (Mt5.0) (https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR) combines the reference genome in only 64 sequence contigs spanning to 430Mb providing a highly contiguous reference sequence. The *M. truncatula* genome assembly Mt5.0 represents 44,623 protein-coding genes, ~25K transposable elements and ~4K non-coding RNAs enlarging the known gene content of the genome significantly (Pecrix et al., 2018).

4.4. Identification of chromosomal position of mutant loci by genetic mapping

Identification of the genes involved in symbiotic nitrogen fixation in *M. truncatula* was primarily conducted using forward genetic approaches, until recently. The forward genetic approach requires the induction or search for natural mutations in wild type plants and creating mutant

populations. The mutant populations of *M. truncatula* are obtained from a range of mutagenesis programs. The chemical mutagenesis such as Ethyl Methanesulfonate (EMS) mutagenesis alkylates the guanine bases in DNA leading to point mutations. Physical mutagenesis such as gamma irradiation mutagenesis results in a broad range of mutations including a varying scale of deletions, insertions and chromosomal rearrangements. Fast neutron bombardment mutagenesis induces different size of deletions. Additionally, the insertion mutagenesis applies *Tnt1* retrotransposon insertions (Benedito et al., 2008; Tadege et al., 2008). All the mentioned mutagenesis methods create random mutations in the genome.

Numerous mutant lines defective in symbiotic nitrogen fixation were identified in several large scale random mutagenesis screens of *M. truncatula* in co-operative projects of the legume community over decades (Starker et al., 2006; Domonkos et al., 2013; Pislariu et al., 2015). These mutants can be classified based on their symbiotic phenotype: some are unable to form nodules (Nod-), some do not support infection (Inf-), and some are able to develop nodules but do not support symbiotic nitrogen fixation (ineffective or Fix- mutants). Because these mutants have been generated by random mutations, the responsible gene(s) for the symbiotic defects can be investigated using classical genetic analyses, and the impaired gene can be effectively identified by positional cloning.

To identify complementation groups, an allelism test can be carried out by crossing mutant lines with similar phenotypes. The wild type phenotype of the F1 hybrids indicates that independent loci (genes) are responsible for their phenotype, and both parents carried recessive alleles of the mutant genes. The identification of the location of mutant genes traditionally requires extensive genetic mapping, but currently it can be facilitated by applying genome or transcriptome sequencing.

Genetic mapping is based on the principles of inheritance of traits discovered by Mendel and the studies of Morgan and Sturtevant in the early 20th century (Sturtevant, 1913; Morgan et al., 1923). The latter ones studied the "anomalies" they encountered when trying to match their experimental results with Mendel's "laws" of independent inheritance. Their research has established the concept of linkage and the theoretical foundation of genetic mapping still used today ((Weber, 1998)Marcel Weber 1998). However, the entire process became much faster and easier with the development of molecular genetic markers. The number of DNA-based molecular markers can be very high because theoretically every nucleotide can be polymorphic between the alleles. The variations in the DNA can be identified using various methods depending on the type of polymorphism (nucleotide exchange, INDELS) present, e.g.with SSCP (single strand conofrmatin polymorphisms) analysis, with RFLP

(restriction fragment length polymorphism) analysis, or with simple agarose gel electrophoresis following PCR amplification in the case of length polymorphisms (Kiss et al., 1998a). In *M. truncatula*, a plant species commonly used in genetic studies of legumes, numerous molecular markers have been developed (Choi et al., 2004; Mun et al., 2006). These markers are mostly designed based on small repeat regions that exhibit length variations among different ecotypes of *M. truncatula*. Practically speaking, this means amplifying the fragments from both paternal and maternal genomes and separating them in agarose gel which will result in length polymorphism of the amplified fragment. The PCR amplification of these type of genetic markers allows us to identify the genotypes (paternal or maternal homozygotes or heterozygotes) of the offspring in the mapping population (co-dominant markers). The linkage analysis of the genetic markers for the individuals of the mapping population reveals the relative position of the markes on the genetic map. One of the useful methods to define the order of the markers is employing color-coded maps known as colormaps, which provide a visual representation of recombination events along the chromosomes. This non-mathematical mapping technique simplifies the identification of linked markers, recombination events, and any discrepancies in the order of genetic markers. It has proven to be a valuable tool in genetic research, enabling scientists to better understand the organization and inheritance patterns of genes (Kiss et al., 1998b).

The segregation populations of the cross is used to analyze the inheritance of the mutation trait (Young et al., 2011). Based on the Mendelian inheritance patterns, the inheritance of mutation with a single dominant gene can be observed in F1 segregation population whereas in case of a recessive gene mutant phenotype can be observed in F2 segregation population as two copies of the mutant allele are required for homozygosity to express the recessive trait. The proportion of the mutant plants to wild type plants indicates whether a single gene or multiple genes are responsible for the defective symbiotic phenotype. For a single gene mutation 3:1 ratio is expected for dominant traits whereas for recessive traits 1:3 ratio is expected. Following the resolution of inheritance pattern, the genotype of each individual in segregation population is determined using molecular markers. The numbers of the genetic markers on the map as well as the size of the mapping population are crucial parameters for resolution of the genetic map. The closely located markers are more likely to inherit together in next generations indicating linkage. Recombination events are less likely to occur between closely linked markers. The frequency of recombination between the mutant locus and genetic markers is proportional to their physical distance. This distance is measured in centimorgans (cM) and refers to the map distance where 1 cM indicates a 1% recombination frequency.

The rough mapping, which positions the location of the mutation on one of the 8 chromosomes of *M. truncatula,* can be conducted using available genetic marker set of *M. truncatula*. This is followed by fine mapping, where co-segregation of mutant phenotype with chromosome specific linked markers are investigated. Identification of genetic markers showing recombination indicates the border of the genomic region of the mutant locus. The mapping population can be extended, and novel genetic markers can be introduced into genetic map to narrow down the genomic region. At the next step, the sequence and the gene content of genomic region is analyzed. If the genomic region has an incomplete sequence, gaps can be bridged with *de novo* assembly of Bacterial Artificial Chromosomes (BAC) clones or the scaffold sequences obtained from the shotgun sequencing of various other *M. truncatula* accession (www.medicagohapmap.org). The accuracy of the assembly can be further assessed with the positioning of genetic mapping markers on *de novo* assembled sequences, creating a contig that covers the genomic region. The identification of the mutated gene in the genomic region can be investigated by comparison of DNA sequences of the gene obtained from wild type and mutant alleles. In case of large number of candidate genes, alternatively, RNA sequence data obtained from mutant plants can be compared to expression of the genes in wild type plants.

4.5. Plant factors mediating terminal bacteroid differentiation

In the nodule cells of inverted repeat-lacking clade (IRLC) and Dalbergoids, the polyploidy state of the bacteroids, stem from multiple rounds of DNA replication without cell division. Several studies have demonstrated that interfering with the rhizobial cell division genes can lead to the transformation of bacteria into differentiated bacteroids. This suggests a strong link between the terminal bacteroid differentiation process and the rhizobial cell cycle (Latch and Margolin, 1997; Wright et al., 1997; Sibley et al., 2006; Cheng et al., 2007). Conversely, in the cells of determinate nodules of non IRLC legumes, bacteria preserve their free-living size and ploidy level indicating lack of terminal differentiation (Gresshoff and Rolfe, 1978; Mergaert et al., 2006). The role of host legume species in the morphological and genetic differentiation of bacteroids in nodule cells was investigated in a research conducted by Mergaert et al. (2006) by utilizing a genetically modified strain of *Rhizobium leguminosarum* capable to infect *P. sativum* (forming indeterminate nodules) as well as *L. japonicus* (forming determinate nodules). This study indicates that the same bacterial strain can transform into bacteroids, maintaining an undifferentiated state in non-galegoid legumes while undergoing elongation and differentiation in galegoid species. This indicates that bacteroid differentiation is controlled by the host plants (Mergaert et al., 2006).

4.5.1. Nodule-specific cysteine rich peptides

Nodule-specific Cysteine-Rich (NCR) peptides and Glycine-Rich peptides (GRPs) meditate the terminal differentiation of bacteroids in different legumes (Mergaert et al., 2003, 2006; Laporte et al., 2010; Van de Velde et al., 2010; Guefrachi et al., 2014; Kereszt et al., 2018). *NCR*s that are exclusively expressed in nodule cells are the major effectors in this peptide-mediated signaling processes in the IRLC legumes and in the *Aeschynomene* spp. belonging to more ancient lineage of the Dalbergioid clade (Czernic et al., 2015). *NCR* genes are found in the genome of all IRLC legumes in varying numbers from 7 to over 700 (7 in *Glycyrrhiza uralensis* (Chinese licorice), 63 in *Cicer arietinum* (chickpea), 353 in *Pisum sativum* (pea), 469 in *M. sativa* (alfalfa) and more than 700 in *M. truncatula*. Conversely, no *NCR* could be detected in the genome of non-IRLC legumes such as *Lotus japonicas* (Japanese trefoil), Glycine max (soybean), and *Phaseolus vulgaris* (common bean) (Kevei et al., 2002; Graham et al., 2004; Alunni et al., 2007; Montiel et al., 2017; Downie and Kondorosi, 2021). The extent of bacteroid differentiation observed in the IRLC legumes is significantly related to the quantity of NCR peptides (Montiel et al., 2016, 2017). The model organism *M. truncatula* contains the highest number of *NCR*s in its genome. This signifies the importance of NCR peptides in bacteroid differentiation and consequently nodule formation process of *M. truncatula*. The *NCR* genes in *M. truncatula* is spread across the genome and in many cases a few of them are found next to each other in clusters, suggesting rapid duplication events in evolutionary processes (Alunni et al., 2007; Pecrix et al., 2018). Gene structure of *NCR*s in *M. truncatula* genome is generally characterized with two exons both consisting of coding sequences. However, in less common instances, some *NCR*s may also contain a third exon following an intron region positioned after the stop codon at the 3'end of second exon.

Studies exploring the organ specific expression of 334 *NCR*s, using *M. truncatula* Gene Expression Atlas (MtGEA) and other available microarray data, revealed that all but five *NCR*s are specifically expressed only in nodules when tested under a large scale of differing environmental conditions (Fig. 4.5.1.1a,b) (Limpens et al., 2013; Guefrachi et al., 2014; Roux et al., 2014). Researchers demonstrated that, the expression of these *NCR*s in other plant organs are not inducible with any biotic or abiotic stress indicating the strict nodule specific expression of *NCR* genes. *NCR* genes are not activated in early stage of nodule formation (3 dpi) but only after the invasion of bacteria into nodule cells (4 dpi) whereas they reach the highest levels of expression at differentiation and elongation stage of bacteroids, generally observed at 10 dpi or 14 dpi in *M. truncatula* (Fig. 4.5.1.1b)

(Guefrachi et al., 2014). Few *NCR* genes that are activated early (at 4 or 6 dpi) predominantly express in the nodule apex, and those activated later are found in more proximal tissues (Fig. 4.5.1.1a,b). In a study conducted by Roux and co-workers (2014), the RNA sequence count of approximately 600 *NCR* genes was evaluated in 2 week-old laser-dissected nodule samples representing five different nodule regions; from apical to basal part of the nodule containing meristematic zone (FI/ ZI), the distal (FIId/ ZIId) and proximal (FIId/ ZIId) infection zone, the interzone (IZ/ZII-ZIII) and nitrogen fixation zone (ZIII) (Fig. 4.5.1a). Researchers demonstrated that, none of the *NCR* genes have expression in the meristem in accordance with previous studies indicating that *NCR*s are only expressed in the infected nodule cells (Van de Velde et al., 2010; Nallu et al., 2013). Additionally, results indicated that *NCR*s are expressed in four different waves; with read count of early expressed 83 genes predominantly observed in the distal or proximal infection zone, while RNA sequencing of 583 genes having accumulation predominantly in interzone and nitrogen fixation zone (Fig. 4.5.1.1a). The strong correlation between the spatial and temporal expression patterns of *NCR* genes in nodule development lead to categorization of *NCR* genes as early and late *NCR*s (Limpens et al., 2013; Nallu et al., 2013; Guefrachi et al., 2014; Roux et al., 2014).

NCR peptides are secreted into the plant-derived membrane of the symbiosome, through a secretion system that recognizes the highly conserved N-terminal secretion signal (NSS) of the NCR peptides (Wang et al., 2010; Stonoha-Arther and Wang, 2018) The NSS direct mature peptide to the bacteroids to initiate alterations associated with terminal differentiation such as endoreduplication, inhibition of cell division, cell elongation and increased membrane permeability (Mergaert et al., 2006; Van de Velde et al., 2010; Batut et al., 2011; Kereszt et al., 2011; Haag et al., 2013; Kondorosi et al., 2013; Tiricz et al., 2013; Farkas et al., 2014; Maróti and Kondorosi, 2014; Penterman et al., 2014; Alunni and Gourion, 2016). The mature peptides of NCRs generally contain 20-50 residues and characterized with four or six cysteine residues in conserved positions (Fig. 4.5.1.2a). The number of the genes coding for NCR peptides containing four cysteines or six cysteines in their mature peptide is similar. Although at a less conserved degree, an aspartic acid located in front of the second cysteine and a proline following the second cysteine residue are also observed in more than half of the NCRs. NCR peptides share common structural characteristics to small cysteine-rich antimicrobial defensins. However defensins differ from NCRs with the length of their mature peptide which contains 45-70 amino acids, and with the eight regularly spaced cysteine residues which form 4 disulfide bonds (Fig. 4.5.1.2b) (de Oliveira Carvalho and Gomes, 2009a; Farkas et al., 2018).

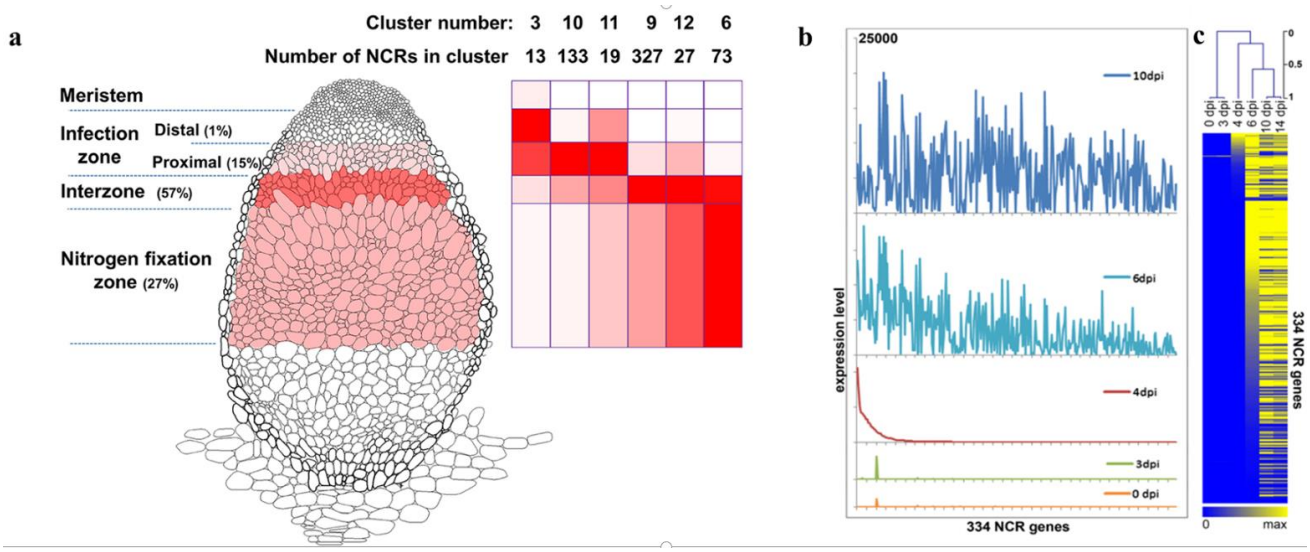


Fig. 4.5.1.1 Spatio-temporal expression patterns of *NCR* genes in *Medicago truncatula* nodules

The different nodule zones are named on the left. The intensity of color in the cells of these zones reflects the relative expression of all the *NCR*s based on the data from Roux et al. (2014). *NCR*s are mainly grouped in ZIIp (cluster 10), the ZII-ZIII (clusters 9, 6) and ZIII (ZIII (cluster 11, 12) based on the relative expression identified in the five laser capture microdissection fractions as shown on the right panel. The color intensity illustrates their average patterns of expression in the different developmental stages with which they are aligned (a) (From Downie and Kondorosi 2021). Successive activation of *NCR* genes during nodule development (b,c). Expression level of the 334 *NCR* probe-sets (x-axis) at 0, 3, 4, 6 and 10 dpi. The y-axis scale is the same for all graphs and is 25,000 at maximum (b). Heat map of the same expression data as in b (c). (From Guefrachi et al., 2014)

The presence of cysteines in NCRs is crucial due to their potential to form disulfide bonds. Disulfide bonds significantly affect peptide folding and are essential for the structural integrity. The studies regarding the role of disulfide bonds in NCR peptides is mainly carried on three NCRs; NCR247, NCR044 and NCR169, all containing 4 cysteines and forming 2 intramolecular disulfide bonds. In NCR247 disulfid bonds are formed between first two (C1-C2) and last two cysteines (C3-C4). The specific arrangement of disulfide linkages in NCR247 shown to be crucial for the *in vitro* antimicrobial activity as replacing any cysteine residue with a serin, exchanging the positions of disulfide bonds or breaking the disulfide bonds with reduction changes the stability of the peptide to varying extents (Haag et al., 2012; Shabab et al., 2016). Similar substitution of a single cysteine residue with serine in NCR169, causing the disruption of disulfide bonds between C1–C2 and C3–C4, results in loss of *in planta* function emphasizing the critical effect of cysteines on functionality (Horváth et al., 2015; Isozumi et al., 2021). The disulfide bonds are detected between C1–C4 and C2–C3 in NCR044, in the experiments where peptide is produced in the yeast *Pichia pastoris* (Velivelli

et al., 2020). The disordered structure of NCR044 may suggest functional stability in interactions with bacterial targets.

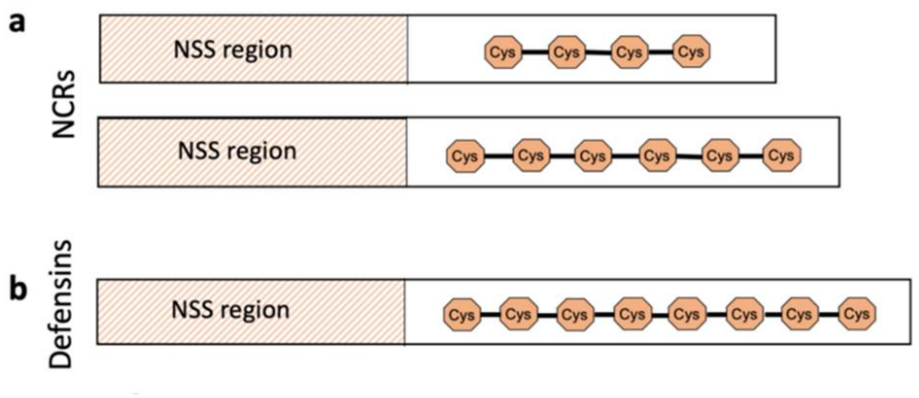


Fig. 4.5.1.2 Simplified precursor structure of NCR peptides with defensins for comparison

NCRs in indeterminate legumes such as *Medicago truncatula* and *Aeschynomene* are characterized by the presence of a signaling peptide at the N-termini of the peptides and four or six cysteines in conserved positions (a). Defensins similarly have signal peptides but have eight cysteine residues (b). (Modified from Roy et al., 2020)

The correct formation of disulfide bridges in NCR peptides is controlled by upregulation of protein disulfide isomerase and ER oxidoreductin 1 enzymes in the endoplasmic reticulum (Mergaert et al., 2003; Roux et al., 2014). Additionally, thioredoxins produced by symbiotic cells, are also targeted to the bacteroids together with NCRs. Thioredoxin (Trx) s1 plays a crucial role in fine-tuning of the antimicrobial function of NCR peptides by balancing their cytotoxic effects to prevent harmful effects to the bacterial partner (Ribeiro et al., 2017). Trx s1 interacts with specific NCRs, such as NCR247 and NCR335 and increase the cytotoxic effect of NCR335 against *Sinorhizobium meliloti* suggesting that Trx s1 modulates the activity of NCRs, probably through the reduction of disulfide bonds. The experiments carried on with Trx s1 revealed that silencing as well as overexpressing of the gene results in defects in elongation of bacteroids emphasizing the critical effect of reduced and oxidized states of NCRs in bacteroid differentiation (Ribeiro et al., 2017).

4.5.2. The diverse roles of nodule-specific cysteine-rich peptides

Functions and working mechanisms of NCRs remain to be explored despite their extensive studies. NCR peptides resemble with defensins due to their peptide sequence containing conserved cysteines. Some NCRs, in particular cationic NCR peptides, indeed perform antimicrobial activity against various bacteria and fungi in *in vitro* experiments (Farkas et al., 2017, 2018; Isozumi et al., 2021). Therefore, the isoelectronic point of NCRs is hypothesized to be a critical determinant in

antimicrobial function of the peptides. The antimicrobial activity of NCR peptides such as NCR035, NCR055, NCR211, NCR169 and NCR247 demonstrated the inhibition of the proliferation of *S. meliloti*. (Van de Velde et al., 2010; Kim et al., 2015; Mikuláss et al., 2016; Isozumi et al., 2021). However, *S. melloti* survives during the symbiotic processes which due to the fact that the *in planta* concentration of antimicrobial NCR peptides is much lower than the amounts used *in vitro* (Maróti et al., 2015; Kereszt et al., 2018).

The role of antimicrobial feature of NCRs in bacteroid differentiation is not well characterized, yet. However, researchers demonstrated that the bacterial peptide transporter protein BacA protects bacteroids against the *in vitro* antimicrobial activity of oxidized states of NCR247. Conversely, reduced state of NCR247 maintains its destructive effect. The reduced form of NCR247 shows most effective antimicrobial activity suggesting that redox state of the peptide is crucial for its antimicrobial function (Haag et al., 2011). Consequently, it is hypothesized that BacA may protect the rhizobia by promoting a less active oxidized state of NCR247 *in vivo*.

The capacity of NCR peptides to initiate bacteroid differentiation is demonstrated in a research where the ectopic expression of NCR peptides, in *L. japonicus*, a legume that naturally does not expresses NCR peptides, results in increased bacteroid size (Van de Velde et al., 2010). The result of the study suggests that NCR peptides, in particular NCR035, can induce bacteroid differentiation across different legume clades, including those outside of the Galegoid lineage (Fig. 4.5.2.1a-h). Flow cytometry measurements revealed an increased level (4C) of DNA content in the *S. meliloti* cells treated *in vitro* with NCR035 peptide which further supports the role of NCRs in elongation and endoreduplication processes (Fig. 4.5.2.1h) (Van de Velde et al., 2010). In another study conducted by Durgo and co-workers (2015), the accumulation of NCR peptides in the bacteroids is demonstrated by detecting 138 NCRs in the bacteroids.

Remarkably, localization studies utilizing the techniques such as SDS-PAGE analysis, Western blotting, and advanced imaging techniques, including immunofluorescence and immunogold transmission electron microscopy, have revealed that some NCRs such as NCR001, NCR035, NCR169, NCR211 and NCR247 are not only targeted to the bacteroids but also penetrate the bacterial cytosol localizing together with bacteroids in the symbiosome (Fig. 4.5.2.2d-g) (Van de Velde et al., 2010; Farkas et al., 2014; Horváth et al., 2015; Kim et al., 2015). This might indicate the activity of certain NCRs in the rhizobia cytosol and not only on their cell wall or membranes.

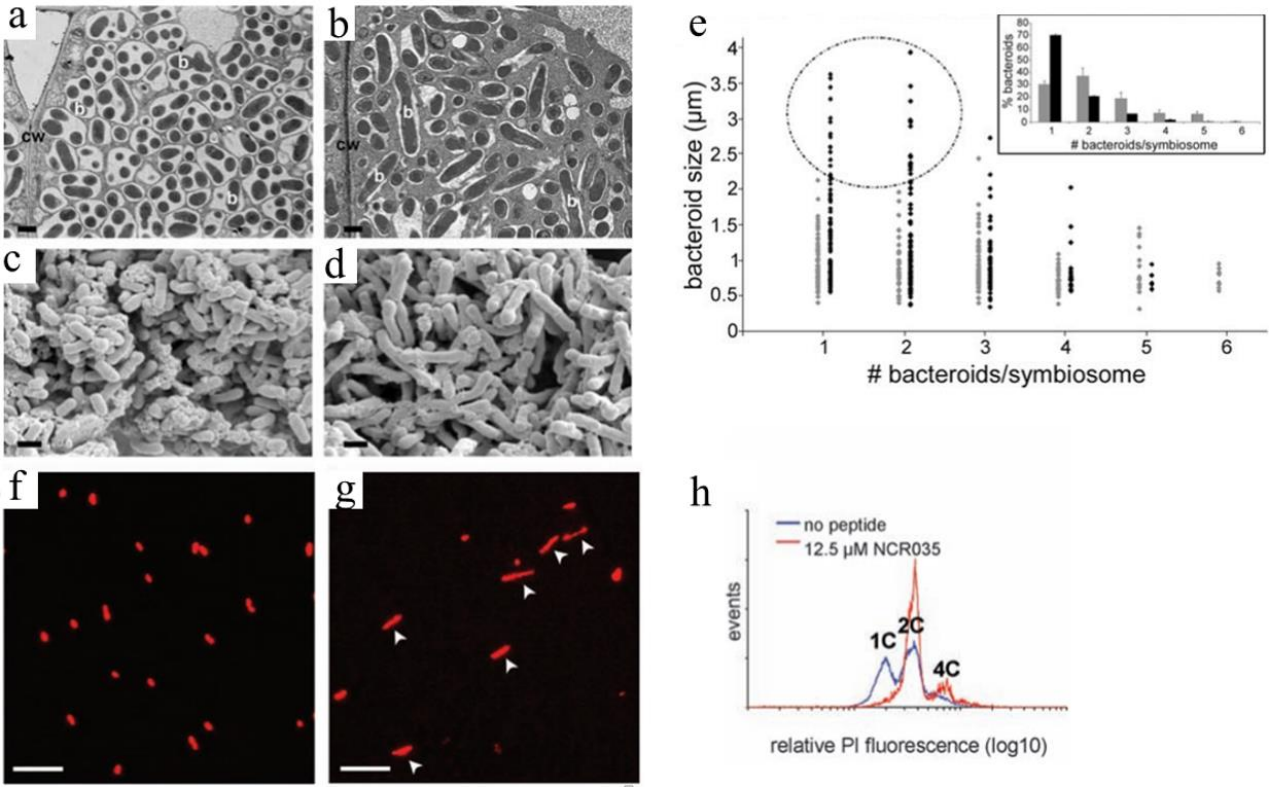


Fig. 4.5.2.1 Expression of *NCR035* in *L. japonicus* leads to features of terminal differentiation of Rhizobia

Transmission Electron Microscopy (TEM) of a symbiotic nodule cell expressing GUS (a) and *NCR035* (b) demonstrates an increase in bacteroid size in the case of *NCR035* expression. b, bacteroid; cw, cell wall (a and b). Scanning Electron Microscopy (SEM) of bacteroids in a symbiotic cell expressing GUS (c) and *NCR035* (d) visualizes the morphological differences of bacteroids upon *NCR035* expression (c and d). Plotting of bacteroid size as a function of bacteroid number per symbiosome shows the presence of bacteroid populations with enlarged size in *NCR035* transgenic lines (encircled). In the control nodules, most symbiosomes contain more than one bacteroid, whereas in the *NCR035*-expressing lines, there is most frequently one bacteroid per symbiosome. Gray is the quantification for control lines and black for NCR035-expressing lines (e). Control (f) and *NCR*-treated (g) *S. meliloti* cells showing an increase in size upon treatment with NCR peptides (arrowheads). Cells were stained with PI. Increased DNA content of *S. meliloti* cells after *NCR035* treatment in comparison to the control. (h) Scale bars: (a-d) 1 µm; (f,g) 10 µm (Modified from Van de Velde et al., 2010)

Although the impact of NCR peptides in differentiation process were demonstrated with these studies, it was assumed that NCRs have redundant roles because of their large numbers and spatio-temporal expressions. However, recent studies demonstrated that absence of some single *NCR* gene can cause inefficient nitrogen-fixing symbiosis.

The forward genetic approaches revealed that the lack of *NCR169* causes nitrogen fixation defective phenotype of *dnf7-2* (Horváth et al., 2015), while the loss of *NCR211* was found to be responsible for the nitrogen starvation phenotype of *dnf4-1* (Kim et al., 2015). Simultaneously with

our study, another nodule cysteine rich peptide, the *NCR343*, was proved to be essential for the symbiosis between *Medicago truncatula* and rhizobia based on the complementation experiment of the nitrogen fixation deficient phenotype of mutant NF-NF9363 (Xi et al., 2013), within a research project conducted in our research group, which was then combined with this study for comparative analysis (Horváth et al., 2023). Our group demonstrated that in *Mtdnf4-1*, *Mtdnf7-2* and NF9363 mutants, bacteria occupy the nodule cells at early developmental stage of the nodules. However, differentiated bacteroids are absent in nitrogen fixation zone of nodules 2 wpi with rhizobia underlying the essential role of *NCR169*, *NCR211* and *NCR343* for the differentiation and persistence of bacteroids in matured nodules. *NCR169* is expressed in interzone and nitrogen fixation zone while *NCR211* and *NCR343* are mainly expressed in infection zone and interzone of 2-week-old nodules validating the spatio-temporal expression patterns of these essential *NCR*s in nitrogen fixing symbiosis (Horváth et al., 2015, 2023; Kim et al., 2015). Collectively, these findings indicate the positive regulation of nitrogen fixing symbiosis with essential *NCR* genes.

However, other studies indicated that some NCRs can negatively regulate the symbiosis between *S. meliloti* and *M. tuncatula* in a strain and host-dependent manner (Yang et al., 2017; Wang et al., 2018a, 2018b). The *M. truncatula* accessions DZA315.16 (DZA315) or A20 form Fix+ nodules when inoculated with either *S. meliloti* RM41 or *S. meliloti* A145 (A145). Conversely Jemalong (A17) plants form Fix- nodules characterized by degradation of bacteria promptly following the colonization of nodule cells when inoculated with *S. meliloti* RM41 or *S. meliloti* A145 (Tirichine et al., 2000; Yang et al., 2017; Wang et al., 2018a). The incompatibility of A17 genotype with Rm41 strain was demonstrated to be controlled by multiple genes including *NFS1* and *NFS2* that code for NCR peptides. The observations obtained from the inheritance pattern of NFS1 and NFS2 alleles, complementation tests as well as reverse genetic approaches targeting *NFS1* or/and *NFS2* suggested that both genes are responsible for the Fix- phenotype of plants inoculated with A145 and *NFS1* is dominant over *NFS2* in blocking nitrogen fixation by *S. meliloti* A145. Alleles represent strain-specific nitrogen fixation and the expression of *NFS1* and *NFS2* possibly can be suppressed by other host genes (Pan and Wang, 2017; Yang et al., 2017; Wang et al., 2018b).

Studies suggested that NCR247 penetrates cell membranes, accumulates in the cytosol, and induces the endoreduplication process, leading to transcriptional changes in *S. meliloti* (Tiricz et al., 2013; Farkas et al., 2014). These transcriptional alterations involve the downregulation of basic cellular function genes and the upregulation of stress response genes (Tiricz et al., 2013). The

expression of *NCR247* is detected in the early stages of nodule development, in the ZII where plant cell division is arrested and in ZII-ZIII where bacterial elongation is initiated which supports the hypothesis that NCR247 is involved in endoreduplication process (Fig. 4.5.2.2c). NCR247 peptide was shown to interact with FtsZ, a conserved bacterial cell cycle protein required for septum formation and cell division, and inhibit FtsZ polymerization. Considering that NCR247 localizes to the symbiosome (Fig. 4.5.2.2d-g) and inhibit the activity of FtsZ, researchers conclude that NCR247 may be involved in inhibition of endosymbiont proliferation (Farkas et al., 2014). Subsequently, the effect of NCR247 peptide on multiple cellular processes was demonstrated with studies showing its interaction with the GroEL chaperone, which is essential for infection and terminal differentiation (Farkas et al., 2014). It was also suggested that NCR247 peptide interacts with ribosomal subunits, resulting in the inhibition of translation, and possibly influencing NCR peptide folding although exact mechanism has not been revealed (Farkas et al., 2014).

Recently, the essential role of *NCR247* in the symbiosis between the *M. truncatula* and the *S. meliloti* has been proven using *A. rhizogensis*-mediated CRISPR gene editing by Sankari et al., (2022) (Fig. 4.5.2.2a,b). This study describes the working mechanism of *NCR247* in detail by demonstrating its role in promoting iron uptake in symbiotic bacteria by the sequestration of heme. Sankari and co-workers initiated their study by examining transcriptional response of *S. meliloti* to NCR247. This experiment revealed that NCR247 treatment increase the expression of iron uptake genes, *Irr* and *RirA*, in *S. meliloti* even under iron-deplete conditions (Fig. 4.5.2.3). Both Irr and RirA can bind to DNA from upstream of the genes to repress their expression. Irr can sense the level of iron by the interaction with the heme, resulting in loss of ability to bind DNA, as well as RirA, which needs to interact with an Fe-S cluster to bind DNA. In case of the low heme concentration in the cells, Irr downregulate the genes responsible for iron storage, export, and the production of RirA. The absence of RirA activity results in an increased expression of iron uptake genes. When iron concentration is high, heme binds to Irr preventing it to repress the Rir and decrease the activity of iron uptake genes. However, if NCR247 is present under high iron concentration conditions, it sequesters the heme, and thus heme cannot repress Irr mediated repression of RirA. This mechanism stimulates rhizobia to import the iron that is needed for the nitrogenase enzyme complex even when the plant cells already have sufficient iron for their normal metabolic needs (Fig. 4.5.2.3). Although exploration of the working mechanism of NCR247 peptide provides significant information regarding to NCR peptides, its unique sequence which is associated to iron uptake mechanism suggests that *NCR247* has a specialized role in nitrogen fixation symbiosis.

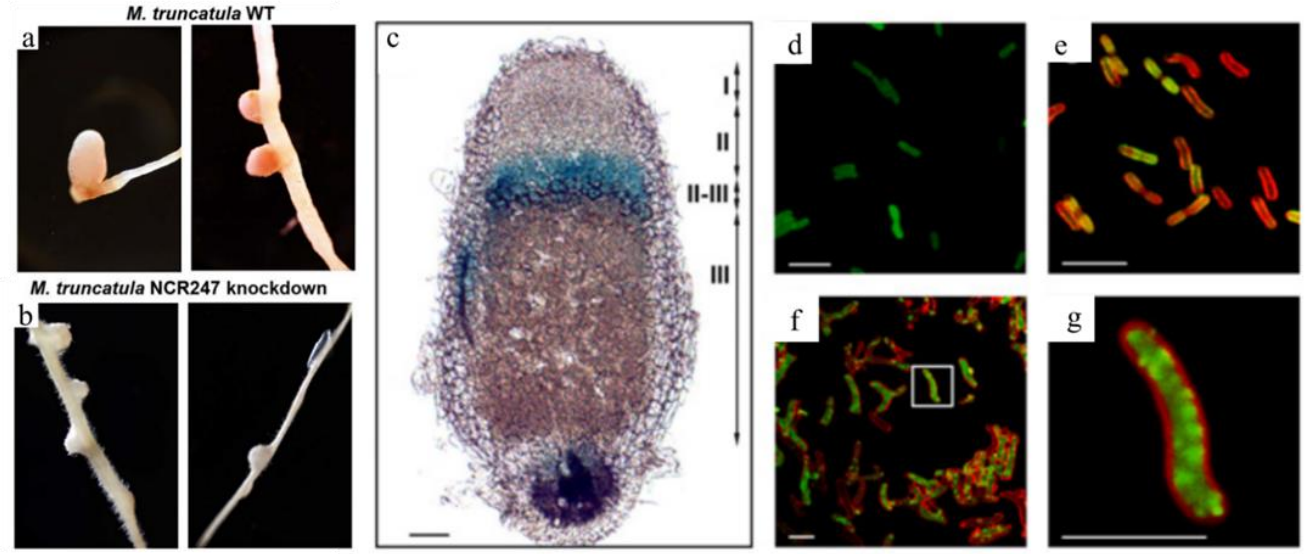


Fig. 4.5.2.2 Phenotype of *NCR247*-Deficient Nodules and Expression of *NCR247-GUS* in Nodules, Along with *In Situ* Localization of the NCR247 Peptide

Wild-type and *NCR247* knockdown of *M. truncatula* (A17) inoculated with wild-type *S. meliloti* (Rm1021) (a, b) (Sankari et al., 2022). *NCR247-GUS* expression in nodules (c). Blue staining in the older cells of zone II and in the interzone II-III corresponds to *NCR247-GUS* expression in transgenic *M. truncatula* nodules. I, nodule meristem; II, infection zone; II-III, interzone; III, nitrogen fixation zone (c). Localization of fluorescein isothiocyanate-NCR247 (FITC-NCR247) in *S. meliloti* bacteria and bacteroids (d-g). *S. meliloti* culture stained with FITC-NCR247 (green) and PI (red color of PI could not be observed) (d). *S. meliloti* culture stained with FITC-NCR247 (green) and FM 4-64 fluorescent dye (red) (e). *S. meliloti* bacteroids stained with FITC-NCR247 and FM 4-64 (f). Enlarged image of the boxed bacteroid (g) Scale bars: (c-g) 5 μm. (Farkas et al., 2014)

However, it is important to note that NCR247 is the most extensively studied NCR and there is almost no available data regarding to working mechanisms of the other NCR peptides. Nevertheless, data obtained from identified crucial *NCR*s for nitrogen fixation suggests that NCR peptides are responsible for diverse roles. Further large-scale research such as whole genome comparison of IRLC and non IRLC legumes as well as extensive targeted gene editing researches would release more information to broaden our understanding regarding to the role of NCRs in nitrogen-fixing symbiosis.

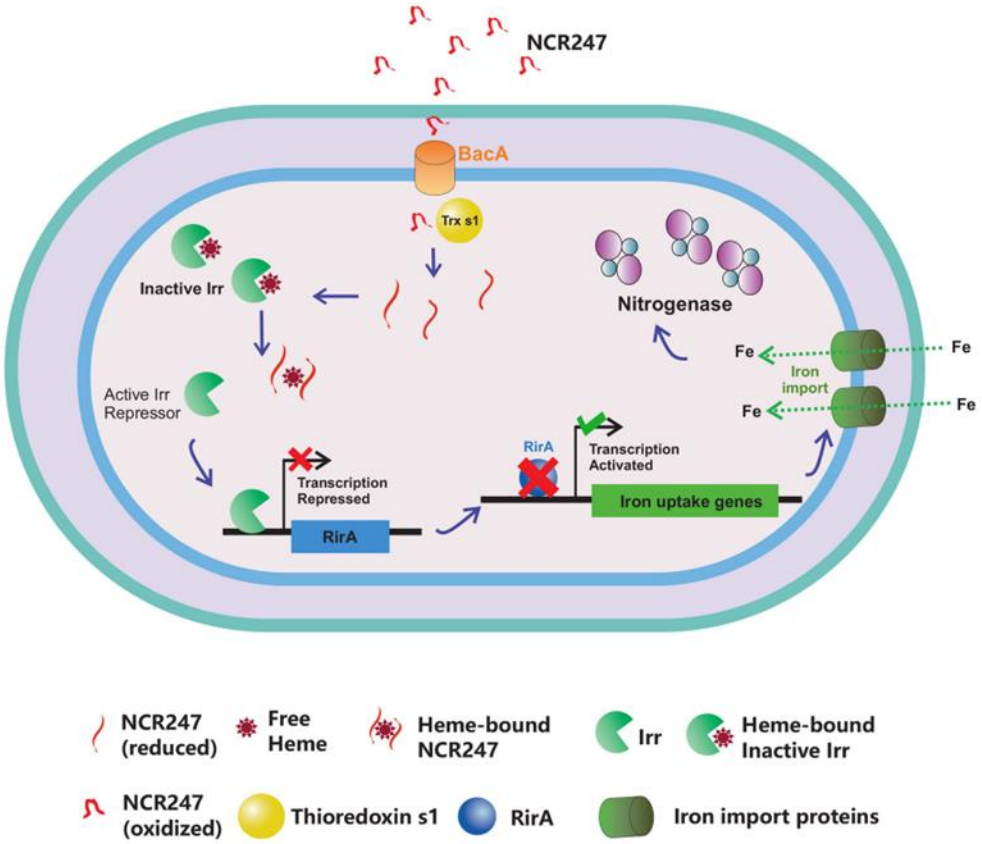


Fig. 4.5.2.3 NCR247 sequesters heme and overrides bacterial iron regulation to aid in symbiosis

Model for the proposed role of NCR247 in iron regulation of bacteria and symbiosis. NCR247 secreted by the Medicago plant enters the cytoplasm of *S. meliloti* through the inner membrane protein BacA. NCR247 could be then reduced by plant produced Thioredoxin s1. Reduced NCR247 sequesters heme tightly. This leads to unavailability of heme and stabilization of heme regulated transcriptional repressor Irr even under iron sufficient conditions. Active Irr represses RirA which is a transcriptional repressor of iron uptake genes. This leads to an increase in transcription of iron uptake genes and ultimately results in an increase in iron import into the cell. The need for iron in nodule increases during nitrogen fixation and nitrogenase (the key nitrogen fixing enzyme) requires numerous iron atoms structurally and functionally. Thus, NCR247 mediated boost in iron import could improve nitrogen fixation to ultimately benefit the plant. (Sankari et al., 2022)

5. MATERIALS and METHOD

5.1. Plant and bacterial materials and growth conditions

5.1.1. Plant material and growth conditions

The seeds of mutants *Mtsym20* (TRV43), *M(tsym19* (TR183), and *Mtsym18* (TR36) were were obtained from the mutant population of *Medicago truncatula* cv. Jemalong (J5) subjected to gamma ray exposure. (Sagan et al., 1998; Morandi et al., 2005). *M. truncatula* Jemalong accession and *M. truncatula* A17 genotype were used as wild type controls in the phenotypic characterization experiments. Additionally, ineffective nitrogen-fixing (Fix-) symbiotic mutants *Mtdnf7-2*, *Mtdnf4-1*, and NF-FN9363, which were obtained from symbiotic screens of fast neutron-bombarded populations of *Medicago truncatula* cv. Jemalong, cv. Gaertn., and *Medicago truncatula* cv. A17, respectively, were used for the comparative phenotypic analysis of *Mtsym20* mutants (Starker et al., 2006; Domonkos et al., 2013; Xi et al., 2013). *Mtsym20* mutant plants were crossed with *Medicago truncatula* A20 accession and the genotype and phenotype of each plant in the F2, F3, and F4 segregating populations were identified for genetic mapping. For the *Agrobacterium rhizogenes*-mediated CRISPR/Cas9 gene editing experiments, the *NCR-new35* gene was targeted in *Medicago truncatula* cv. 2HA and *Medicago truncatula* ssp. *littoralis* R108 ecotypes. The experiments were conducted at least with 30 plants for each plant line and repeated at least twice.

5.1.2. Germination of *M. truncatula* seeds and cultivation of seedlings

M. truncatula seeds were extracted from pods by trashing on corrugated a rubber mat. Seeds were incubated in H_2SO_4 for 10 minutes for surface sterilization and rinsed with sterile water 3 times. Sterilized seeds were vernalized on 0.8% inverted water agar plates at 4°C for 4 days. Vernalized seeds were then kept in dark at room temperature overnight for germination. The germinated seedlings were planted into zeolite substrate (Geoproduct Kft., Mád, Hungary) when radicles reach to 5-10 mm. The seedlings were grown in the zeolite substrate in the absence of nitrogen, under sterile conditions for 4 days before inoculation of rhizobia. Exceptionally, the seedlings transformed with CRISPR/CAS9 binary gene editing constructs, were grown in the zeolite supplemented with Fahraeus medium (Barker et al., 2006) containing 2 mM NH_4NO_3. These plants were further grown in plastic boxes for 2-3 weeks to preserve rhizobium-free conditions prior to inoculation with rhizobia.

5.1.3. Bacterial strains and plant inoculation

The *Sinorhizobium* (*Ensifer*) *medicae* WSM419 (Howieson and Ewing, 1986) carrying the *hemA::lacZ* marker construct on the pXLGD4 plasmid (Reeve et al., 2010) and *S. meliloti* FSM-MA (*E. meliloti* strain FSM-MA (first catalogued as *E. arboris strain* CCMM B554, also known as LMG-R33403) were grown in TA (Sambrook et al., 1989) medium containing 6 mM $CaCl_2$, or in tryptone-yeast (TY) media (pH 7.2) supplemented with 6 mM $CaCl_2$ (Beringer, 1974). The bacterial liquid culture was spun down and the pellet was suspended in Gibson liquid media (Gibson and Nutman, 1960) at 1:50 final dilution ratio, achieving an approximate optical density of 0.1 at 600 nm (~OD_{600} 0.1) and each plant was supplied with 2 ml Gibson-Rhizobia mixture. Following the inoculation of rhizobia, plants were grown under 16 hours day and 8 hours night photoperiod at 22 °C in growth chamber. Symbiotic phenotypes of the wild type control and *Mtsym20* mutant plants were evaluated at 2 weeks post inoculation for phenotypic characterization experiments and 3wpi for genetic mapping experiments.

5.2. Histological staining and microscopic analyses

5.2.1. Fixation and sectioning of *M. truncatula* nodule samples for symbiotic phenotypic characterization

For the characterization of the symbiotic phenotype, nodules were collected from rhizobia inoculated wild type and/or mutant plants and prepared for microscopic analysis following the fixation, sectioning, and staining procedures.

Nodules were initially collected into 1x phosphate buffered saline (PBS) solution (8g NaCl, 0,2g KCl, 1,44g Na2HPO4, 0,24g KH2HPO4, in 1000 ml deionized H_2O, pH 7,4) on ice. 1xPBS solution was replaced with 4% (wt/vol) paraformaldehyde (PPA) solution and nodules were fixed applying 3x30 second vacuum infiltration with the help of Eppendorf Concentrator 5301. Nodules in the fixation solution were shaken gently for 30 minutes on ice and then 30 minutes at room temperature (RT) to enhance the absorption of paraformaldehyde. Fixed nodules were rinsed 3 times with 1xPBS (pH 7.4) to eliminate the paraformaldehyde and then stored in 1xPBS at 4°C until use. Fixed nodules were embedded into 5% (wt/vol) agarose gel (SeaKem LE Agarose Lonza) and 65 µm and 70 µm longitudinal sections were prepared with the help of Leica VT1200S vibratome.

5.2.2. X-gal staining

To observe nodule occupancy, rhizobia carrying the pXLGD4 plasmid was used for inoculation. For the complementation experiments, transgenic roots were selected 4 wpi *S. medicae* WSM419 based on the dsRED fluorescence using a Leica MZ10F fluorescence stereomicroscope.

To detect β-galactosidase activity in the symbiotic cells, nodule sections of 65 µm were stained with X-Gal staining solution (0,05 M $[Fe(CN)_6]^3$; 0,05 M $[Fe(CN)_6]^4$; %2 X.Gal ; 0,001 $MgCl_2$ and 1x PBS) for 30 minutes at 37°C (Horváth et al., 2023). Following the removal of staining solutions nodule sections were rinsed with 1xPBS two times. The observation of stained nodule sections was conducted under the Olympus BX41M microscope (Olympus Life Science Europa GmbH, Hamburg, Germany) using 10x objective and images were taken with Olympus E-10 digital camera.

5.2.3. SYTO-13 staining

To investigate the morphology of bacteria and the cellular structure of host cells in different nodule zones, 70 µm nodule sections were stained with 5 µM SYTO13 (Thermo Fisher Scientific) dissolved in 1xPBS pH7.4 buffer and incubated on ice for 30 minutes under dark conditions. Staining solution was removed, and sections were rinsed with 1xPBS twice before the preparation of microscope slides. SYTO-13-stained nodule sections were observed under the Leica TCS SP8 confocal laser scanning microscope. The microscope was configured with the following settings: objective lenses HCX PL FLUOTAR 10x dry (NA: 0,3) and HC PL APO CS2 63x oil (NA: 1.4); zooms: 0.75 and 2.14; excitation: OPSL 488 nm laser (SYTO13, green) and OPSL 552 nm laser (red autofluorescence); spectral emission detector: 510-540 nm (SYTO13) and 558 nm – 800 nm (red autofluorescence).

For the *in-situ* localization experiment, the eGFP fluorescence signal of pK7m34GW-rolD: EGFP vector was used as transformation marker to select transformed nodules on *Mtsym20* and wild type roots 4 wpi with *S. medicae* WSM419 under Leica MZ10F fluorescence stereomicroscope. 65 µm thick longitudinal sections were prepared from transformed nodules as described in section 5.2.1 and then sections were stained with 5 µM SYTO-13 dye. The stained nodule sections were observed for florescent signal of mCherry tagged NCR-new35 peptide using a 561 nm laser excitation and 570-640 nm emission, and a 488 nm laser excitation with a 500-550 nm emission range for the visualization of SYTO13 under a Visitron spinning disk confocal microscope equipped with a Yokogawa CSU-W1 spinning disk unit, an Andor Zyla 4.2 Plus camera and an Olympus IX83 inverted

microscope (Visitron Ssystem GmBH, Puchheim, Germany). For low magnification of nodule imaging 10x UPlanFL N dry objective (NA: 0.30) and for closeups 100x UPLXAPOXO (NA: 1.45) oil objective was used. Composite images were generated using the ImageJ software, facilitating the comprehensive analysis and documentation of the fluorescence patterns.

5.2.4. GUS-staining

Histological staining based on β-glucuronidase (GUS) activity was performed using X-Gluc containing staining solution. The transgenic nodules were collected from transformed roots 3 wpi with *S. medicae* WSM419. Nodules were fixed and 65 µm thick sections were prepared as detailed in section 5.2.1. Nodule sections were then treated with X-Gluc staining solution (1 mM 5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid substrate, 50 mM sodium phosphate buffer (pH 7.4), 2.5 mM potassium ferrocyanide, 2.5 mM potassium ferricyanide, 20% methanol and Triton X-100 at 0.1 ml/100 ml) (Horváth et al., 2023). Nodule sections were added to GUS staining solution in the Eppendorf tubes and samples were vacuum infiltrated three times for 40 seconds, followed by incubation at 37°C in the dark for 1-3 hours to detect GUS activity. The images of GUS-stained nodule sections were captured with an Olympus E-10 digital camera under the Olympus BX41M light microscope (Olympus Life Science Europa GmbH, Hamburg, Germany) with 10x objective.

5.2.5. Analyzes of bacteroid elongation and endoreduplication

5.2.5.1. Measurements of length of rhizobia with confocal microscopy and Imagej software

Investigation of bacteroid elongation was carried out on bacteroids isolated from *Mtsym20*, *Mtdnf7-2*, *Mtdnf4-1*, NF-FN9363 mutant and wild type *Medicago truncatula* A17 nodules 16 days post inoculation (dpi) with *S. medicae* WSM419. Nodules were collected in 1xPBS (pH 7.4) solution and homogenized using sterile glass pestles. Ground nodule samples were centrifuged (2000 rpm (Revolutions Per Minute)/ 10 min./4°C) and the supernatant was filtered through 50 µm and 20 µm filters (Sysmex/Partec CellTrics filters), consecutively. Filtered samples were centrifuged again (7000 rpm/15 min./4°C) and pellet was diluted in 20 µl 1xPBS buffer. Samples were stained with 20 µM propidium iodide fluorescent dye (ThermoFisher) for 30 minutes and photographed under Leica TCS SP8 confocal laser scanning microscopy using the 40x objective. The microscope configuration was as follows: objective lens PL Fluotar 40x oil (NA: 1); zoom: 0,75; excitation: OPSL 552 nm laser;

spectral emission detector: 573 - 696 nm. Bacteroid lengths were measured using ImageJ program (Schneider et al., 2012).

5.2.5.2. Flow cytometry analysis

Flow cytometry analysis was employed to examine level of DNA content and the size of the bacteroid populations by a collaborating partner (Mickaël Bourge, Institut de Biologie Intégrative de la Cellule, CNRS, Gif-sur-Yvette, France). Nodules were collected from at least 10 individual plants 16 dpi with rhizobia. Collected nodules were ground in bacteroid extraction buffer (BEB: 125 mM KCl, 50 mM Na-succinate, 50 mM TES, 0.1% BSA, pH 7.0) to extract the bacteroid populations and then filtered through a 50-µm nylon mesh (Horváth et al., 2023). Subsequently, suspensions of bacteroids or free-living bacteria were subjected to a 30-minute heat treatment at 70°C. Following the heat treatment samples were treated with the fluorescent dye propidium iodide (PI) in 50 µg/ml concentration for 10 minutes. Forward scatter of the laser ray indicative of cell size and level of fluorescence of the PI as an indicative of DNA content level were determined on a MoFlo Astrios flow cytometer and the results were analyzed using the Cytoflex S cytometer (Beckman Coulter).

5.3. Molecular biology techniques

5.3.1. DNA isolation

5.3.1.1. Genomic DNA isolation from leaves

For genomic DNA isolation from *Medicago truncatula* leaves, the ZenoGene DNA Purification Kit was used, according to the protocol specified by the manufacturer available at www.zenonbio.hu.

5.3.1.2. DNA isolation from sections of fixed transgenic nodules

The transgenic nodules were collected from the *M. truncatula* 2HA and *M. truncatula* ssp. *littoralis* R108 roots, transformed with gene editing construct of *NCR-new35* gene (sgRNA:*NCR-new35*::pKSE466_RR), using the Leica MZ10F fluorescence microscope. The transgenic nodules displaying inefficient symbiotic nitrogen fixing phenotype were characterized for their size, color, and morphology prior to the DNA isolation. These phenotypic characterization experiments were performed following the fixation of nodules, preparation of nodule sections, staining nodule sections with SYTO13 and analyzing the colonization in the symbiotic cells under a confocal microscope as described in section 5.2.1 and section 5.2.3 respectively. To extract DNA, 2-8 sections from each nodule were collected into lysis buffer [100 mM Tris-HCl (pH 8), 0,5M EDTA (pH 8), 0.2 % SDS, 5

mM NaCl, Proteinase K 0.02 µg/µl] (Güngör et al., 2023). These samples were then homogenized using glass pestles and incubated overnight at 56 °C. The samples were spun down using a microcentrifuge and then allowed to cool to room temperature. Following this, 2 µL of RNase A (from a 10 mg/mL stock solution) was added and the mixture incubated at room temperature for 2 minutes. Subsequently, 200 µL of binding buffer (2.55 M Guanidium-HCl, 10% Triton X-100, 37% ethanol) was introduced and briefly vortexed before the addition of 200 µL of 96% ethanol, which was immediately vortexed to ensure thorough mixing. The samples were then centrifuged at 8000 rpm for 30 seconds to sediment the tissue debris. The clear supernatant was carefully transferred to QIAprep Spin Miniprep columns (Qiagen, Hilden, Germany). After a centrifugation step at 8000 rpm for 1 minute, the flow-through was discarded, and the columns were washed twice with 500 µL of washing solution (10 mM Tris-HCl pH 7.5, 80% ethanol), discarding the flow-through after each wash. A final centrifugation at 13,000 rpm for 3 minutes ensured removal of any residual washing solution. DNA was then eluted by adding 35-50 µL of de-ionized water to the columns, which were then incubated at 50 °C for 3 minutes. Eluted DNA was collected by centrifuging the columns at 13,000 rpm for 1 minute.

5.3.2. PCR amplification

PCR amplifications were carried out to identify the genotypes for genetic mapping using DreamTaq DNA Polymerase. The reaction contained 1 µl of ~100 ng/µl genomic DNA, 1 µl of each forward and reverse primers at 10 pmol/µl, 0.15 µl of Taq polymerase (5U/ µl), 10X DreamTaq buffer containing 2 mM $MgCl_2$ and 2.85 µl of sterile distilled water in a final volume of 12 µl reaction mix. The polymerase chain reaction was carried out in a C1000 Touch™ Thermal Cycler (Life Science Research, Bio-Rad) using the following program: 2 minutes of initial denaturation at 94°C; 35 cycles of 30 seconds at 94°C for denaturation, 1 minute of annealing optimized to the specific temperature of the primers (Supplementary Tables 1, 2, and 3) 1 minute/kb at 72°C for elongation, followed by a final extension of 10 minutes at 72°C and 5 minutes at 4°C. The primers were selected from a set of genetic markers designed for the *M. truncatula* genome (Choi et al., 2004; Mun et al., 2006) and novel primers were generated using either Primer3web (https://primer3.ut.ee/) or SnapGene software (www.snapgene.com).

The amplification of the fragments, for preparation of transformation constructs, with PCR experiments was conducted using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) and CloneAmp™ HiFi PCR Premix (Takara Bio) kits as specified by manufacturer.

5.3.3. Agarose gel electrophoresis, DNA purification from gels, cloning of PCR products and plasmids

Agarose gel electrophoresis was applied to separate the fragments obtained by PCR amplification. 10 µl of DNA samples was mixed with 2 µl of 6X DNA Loading Dye (Thermo Fisher Scientific) and samples were loaded into agarose gel (10 mM Na2B4O7, 1% agarose) supplemented with EtBr at a concentration of 0.5 µg/ml.

For genetic mapping purposes, amplicons were separated by electrophoresis in 2-3.5% agarose gels, while fragments required for cloning were run in 1% agarose gels. Electrophoresis was conducted in a tank filled with 1xTAE buffer (Tris-Acetate-EDTA buffer pH: 8.00) at 120V. For visualization of the fragments, a Syngene GeneGenius gel documentation system equipped with UV light illumination and high-resolution camera was used.

The PCR products used for cloning were isolated from 1% agarose gel and purified using the Promega PCR Purification Kit (Promega Corporation, Madison, WI, USA). Plasmid isolation from bacteria was carried out using the QIAGEN QIAprep Miniprep Kit (Invitrogen, Carlsbad, CA, USA).

5.3.4. SSCP analyses

Single-Strand Conformation Polymorphism (SSCP) analysis was applied to detect the polymorphism of single-nucleotide polymorphisms (SNPs) and small mutations for genetic markers Scaffold733 F3R3, Scaffold005 FR and Scaffol005 F4R4 (Supplementary Table 1).

The PCR reactions were prepared in a final volume of 24 µl. Following amplification, the PCR fragments were denatured to convert the double-stranded DNA (dsDNA) products into single-stranded DNA (ssDNA) molecules. These ssDNA molecules were immediately transferred to ice.

The presence of the fragment was examined by loading 5 µl of the reaction mix to agarose gel. Subsequently, 5 µl to 10 µl of PCR product was mixed with 5 µl of two-phase STOP dye consisting of 98% deionized formamide, 10 mM EDTA (pH 8.0), 0.2% bromophenol blue, and 0.2% xylene cyanol and the mixture was denatured at 95°C for 5 minutes followed by cooling to 4°C and samples were stored on ice. The samples were then loaded to polyacrylamide gel (acrylamide: bisacrylamide ratio 40:1) where conformational differences in the DNA strand, resulting from minor changes in the nucleotide sequence, were observed. Subsequently, the samples were run on an 8% or 10% native polyacrylamide at 4°C in 1xTBE buffer at a voltage of 150-200V for 16-21 hours, according to the expected fragment size.

Silver staining was carried out to visualize the fragments in SSCP gels. The polyacrylamide gel was first fixed in 200 ml of a 1% nitric acid solution for 4 minutes, followed by three rinses with sterile distilled water. Subsequently, it was stained in 200 ml of a 2% silver nitrate solution for 30 minutes. After removing the staining solution, the gel was rinsed three more times with distilled water. The DNA pattern was then observed using 200 ml of a 0.3 M sodium carbonate solution containing 60 µl of formaldehyde. Following three more rinses with sterile distilled water, the gel was fixed in 300 ml of a 10% acetic acid solution for 5 minutes, followed by soaking in 300 ml of a 10% glycerin solution for at least 30 minutes. Finally, the gel was positioned between two layers of cellophane without bubbles and stretched to dry.

5.3.5. DNA sequence analyses

Sequencing of PCR fragments and clones was performed using the Sanger sequencing method developed by Sanger et al. (1977) and analyzed on an ABI Prism 3100 Analyzer. Sequence data analysis was conducted using the SnapGene program (www.snapgene.com).

The sequences of the DNA isolated from fixed sections of nodules edited for the gene *NCR-new35*, were identified by Next-Generation Sequencing (NGS). The 200 base pairs (bp) long genomic region around the *NCR-new35* target site was amplified using the primers listed in Supplementary Table 3. This amplification was done using GoTaq™ G2 polymerase (Promega corp. Madison WI, USA) following the guidelines provided by the manufacturer. After amplification, the PCR products, which contain the targeted genomic regions of *NCR-new35*, were sequenced using an Illumina sequencing platform and each sample generated 10,000 reads. The sequencing data were then analyzed using the CRISPresso2 program to identify, quantify, and visualize the mutations introduced into the genome by gene editing techniques.

5.3.6. Genetic mapping and reconstruction of the genomic region of *Mtsym20* locus

5.3.6.1. Genetic mapping with color map

A genetic map was generated to identify the position of the *Mtsym20* locus causing deficiency in symbiotic nitrogen fixation. F2, F3 and F4 segregation populations derived from the cross between *Mtsym20* and *Medicago truncatula* A20 genotype were used for genetic mapping. To determine the order of the markers based on the defined genotypes, the color mapping method was applied (Kiss et al., 1998b). In the color map, each row corresponds to a genetic marker showing the genotypes of individual plants presented in the columns. Pink color denotes the homozygous A20 allele, green

signifies the heterozygous allele, and yellow indicates the homozygous *Mtsym20* mutant genotype (Fig.7 and Fig.10).

598 individual plants of the segregation populations were grown in zeolite substrate, inoculated with *S. medicae* strain WSM419 (pXLGD4) and scored for their symbiotic phenotype, based on their shoot, leaf and nodule development at 3 wpi. The Fix+ (wild type) symbiotic phenotype of the *Mtsym20* locus was shown by dark grey blue, yellow presented the Fix- (mutant) symbiotic phenotype.

The map position of *Mtsym20* was determined by examining the genotypes of the mapping population with a set of genetic markers designed for the *M. truncatula* genome (Choi et al., 2004; Mun et al., 2006). Novel genetic markers were generated based on the intron targeting approach to have higher chance for polymorphisms between the parental alleles.

5.3.6.2. De-novo sequence assembly

To assemble the genomic region of the *Mtsym20* mutant, a *de novo* assembly was performed. Scaffold sequences obtained from various *M. truncatula* accessions accessible on www.medicagohapmap.org were screened on Bioedit sequence alignment editor version 7.0.5.3 (Hall, 1999) and **o**verlaps between sequences were detected. Overlapping sequences were stitched together and organized into a layout that represents the consensus sequence using the Overlap-Layout-Consensus (OLC) method. Primer pairs were designed based on reconstructed reference sequence of *Mtsym20* genomic region using the SnapGene software (www.snapgene.com) version 5.1 and used to develop genetic markers (Supplementary Table 1). These genetic markers were integrated into the genetic map, by PCR amplifications and genotyping of the plants from the mapping populations.

5.3.7. Assessment of the allelism of *Mtsym20* and *Mtsym19*

Allelic relationship between *Mtsym20* and *Mtsym19* loci was examined with F1 allele test which was carried out with the cross of plants homozygous for *Mtsym19* and *Mtsym20* selected from the mapping populations. Initially, DNA was isolated from leaves of the selected plants derived from F3 segregation populations within the mapping populations of *Mtsym20* and *Mtsym19* mutants. The Mtb267 genetic marker located on chromosome 6 (MtrunA17Chr6:36722270-36722569: 301 bp), was selected for subsequent PCR amplifications to genotype F3 plants from the segregation populations and selected homozygous plants for the *Mtsym20* and *Mtsym19* loci, respectively. The seeds of selected parent plants were germinated and cultivated as described in section 5.1.2. Plants

were grown at 22 °C under controlled conditions in the greenhouse until flowering. To prevent the self-fertilization plants, flowers were emasculated by removing anthers under Leica MZ10 F stereo microscope at 2.5x magnification. The pollen-bearing anthers from the donor flowers were directly touched to the stigma of the acceptor flowers for pollination. Pollinated flowers were covered with a filtered bag to prevent accidental cross-pollination from other plants. The seeds were collected from matured pods and plants were grown for F1 allele tests. DNA was isolated from the F1 progeny of the crosses and PCR amplifications were carried out to genotype the progeny produced by the crosses for the genetic marker Mtb267. Successful crosses were evaluated for allelic relationship between *Mtsym20* and *Mtsym19* mutants based on their symbiotic nodule phenotype examined under the Leica MZ10 F stereo microscope at 2.5x standard magnification.

5.3.8. RNA sequencing analyses

Nodules from the roots of *Mtsym20*, *Mtsym19,* and *Mtsym18* mutant plants were harvested 2 wpi with *S. medicae* WSM419 (pXLGD4). Collected nodules were immediately frozen in liquid nitrogen to maintain RNA integrity. The total RNA was extracted from frozen nodules using the TRI ® Reagent (Sigma-Aldrich) and purified using the Direct-zol RNA MiniPrep Kit (Zymo Research, Irvine, CA, USA). Following the isolation, RNA samples were treated with DNaseI using Zymo-Spin columns (Zymo Research) according to the manufacturer's guidelines to ensure the removal of any contaminating genomic DNA. The integrity and quality of the total RNA were assessed using the Agilent 2100 bioanalyzer. This step was followed by the depletion of ribosomal RNA, to enrich messenger RNA (mRNA), using the Ribominus Plant Kit (Invitrogen). RNA libraries were prepared for RNA sequencing using 1.2 µg of rRNA-depleted RNA utilizing the Illumina TruSeq RNA Sample Preparation Kit v2 (Illumina, Inc., San Diego, CA, USA). The sequencing was performed on the Illumina NextSeq 500/550 platform (Seqomics Ltd., Mórahalom, Hungary), targeting minimum twenty-five million reads per sample in a 150 bp paired-end format. Subsequent analysis of the RNA sequences was performed utilizing the CLC Genomics Workbench 9.5.3 program with default parameters and 0.98 similarity fraction.

Thirty-six million raw reads from the RNAseq of *Mtsym20* and twenty-five million from *Mtsym19*, obtained from Illumina sequencing, were trimmed. This step, eliminate the adaptors and low-quality, ambiguous nucleotides (set with a quality limit of 0.05 and an ambiguous limit of 2). The cleaned and trimmed reads were aligned to specific mRNA reference sequences derived from the *Medicago truncatula* A17r5.0 genome assembly (MtrunA17r5.0-ANR-EGN-r1.6.cds)

(https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR/). The selected reads were mapped on the overlapping genomic region of the *Mtsym20* and *Mtsym19* deletions between the Chr4g0020111 and Chr4g0020631 genetic markers.

A list of primer pairs used for subsequent amplification of the genes in the deletion region and determination of the deletion borders in the genomic region of corresponding mutants are provided in Supplementary Table 2.

The whole organ analysis of deseq normalized RNA sequencing read counts for genes deleted in the genome of the *Mtsym20* mutant was obtained from the Symbimics Gene Atlas database (https://iant.toulouse.inra.fr/symbimics/; (Roux et al., 2014)).

5.3.9. DNA transformation techniques

5.3.9.1. Preparation and transformation of competent Escherichia coli DH5α cells

A single colony of *E. coli* DH5α was inoculated into SOB medium (Sambrook and Russell, 2001) and incubated at 37°C with shaking overnight. The following day, the overnight culture was diluted at a 1:100 ratio in fresh SOB medium and grown at 37°C with shaking until the optical density at 600 nm (OD600) reached 0.4 to 0.6, as measured with a NanoDrop spectrophotometer (Thermo Fisher). The culture was then chilled on ice for 15 minutes to slow down cellular processes and centrifuged at 3,300 rpm for 10 minutes at 4°C. The pellet was resuspended in an ice-cold 0.1 M $CaCl_2$ solution at a 1:10 ratio of the original culture volume. The resuspended cells were incubated on ice for 3 hours to increase cell wall permeability to DNA, followed by another round of centrifugation at (3,300 rpm/10 minutes/4°C). The pellet was then resuspended in 0.1 M $CaCl_2$, and 20% glycerol was added for cryoprotection. The competent cells were aliquoted into pre-chilled sterile tubes, frozen in liquid nitrogen, and stored at -80°C.

Constructs were introduced into *E. coli* strain DH5α competent cells using the heat shock method described by Sambrook and Russell (2001). Competent *E. coli* cells were thawed by incubating on ice for 15 minutes. 4 µl of reaction mix containing 150 ng/ul entry vector and ~50 fmol PCR product was combined with the thawed *E. coli* DH5α competent cells and incubated on ice for 30 minutes to facilitate DNA uptake. To enhance membrane permeability, samples were exposed to a 42°C thermo block for 1 minute and subsequently transferred to a 37°C shaking incubator for one hour to allow for recovery and expression of antibiotic resistance markers. The cells were centrifuged at 13,000 rpm for 1 minute to pellet the cells, which were then resuspended in 100 µL of the supernatant. This

suspension was spread onto solid LB agar plates (Luria-Bertani Broth agar media: 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, 20 g/L agar) containing the corresponding antibiotics for selection of the transferred vectors. Bacteria were cultured overnight at 37°C. Selected colonies were subsequently amplified using the vector and fragment specific PCR based primers to confirm the integration of the fragments into the target vectors (Supplementary Table.3). Confirmed colonies were cultured in liquid LB broth medium (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl) at 37°C with shaking overnight. The selected plasmids were purified using the Qiagen plasmid purification kit and plasmid concentrations were determined using NanoDrop spectrophotometry. Purified plasmids were then sequenced with Sanger sequencing to verify the correct orientation and integration of the fragment within the constructs.

5.3.9.2. Preperation and transformation of Agrobacterium rhizogenes Arqua-1 strain compotent cells

A single colony of the Arqua-1 strain of *A. rhizogenes* was inoculated into YEB liquid medium and incubated at 28°C in a shaking incubator until the OD_{600} reached approximately 0.5-1.0 identified with a NanoDrop spectrophotometer. The culture was subsequently cooled on ice for 15 minutes and centrifuged at 3,300 rpm for 10 minutes at 4°C. The pellet was gently resuspended in an ice-cold 50 mM $CaCl_2$ solution, using a volume equivalent to one-tenth of the original culture volume. To enhance the permeability of the cell membrane, the resuspended cells were incubated on ice for 2 hours. Following this incubation, the cells were centrifuged again under the same conditions, and the pellet was resuspended in 100 µL of 50 mM $CaCl_2$ solution (pH 8.0). Glycerol was added to a final concentration of 20%, and the samples were frozen in liquid nitrogen and stored at -80°C for future use.

Plasmids isolated from *E. coli* strains were introduced into *A. rhizogenes* Arqua-1 competent cells as described by Quandt et al., (Quandt et al., 1993). Arqua-1 competent cells, stored at -80°C were thawed on ice cautiously. Plasmids were mixed with Arqua-1 competent cells and incubated on ice for 5 minutes, then immersed into cold nitrogen vapor for 5 minutes, and finally kept at room temperature for another 5 minutes to facilitate DNA uptake. The samples were subsequently incubated at 30°C for 2 hours to allow for expression of antibiotic resistance genes.

Following incubation samples were centrifuged at 13,000 rpm for 1 minute. The 100 µL of the cell culture was plated onto solid TA media (15 g/L tryptone, 1.5 g/L yeast extract, 7.5 g/L NaCl, 24 g/L agar, and 1.5 mL of 5% Antifoam per 1.3 L), supplemented with the corresponding selective

antibiotics for transformed plasmids and incubated at 30°C for 48 hours. Arqua-1 colonies were subsequently verified with PCR based vector and fragment specific primers to confirm successful transformation (Supplementary Table.3).

To prepare a culture for hairy root transformation, an overnight liquid culture of *A. rhizogenes* strain Arqua-1, containing the relevant constructs and grown in TY/calcium medium was supplemented with corresponding antibiotics. Approximately 100 µL of this bacterial culture was plated on TY/calcium agar plates ([5 g/L Bacto-tryptone, 3 g/L yeast extract, 6 mM CaCl2 (added post-autoclaving), pH 7.2, 15 g/L agar]) containing the selective antibiotics and incubated overnight at 30°C which resulted in a dense bacterial lawn on the agar surface for subsequent experiments.

5.3.9.3. Agrobacterium rhizogenes-mediated hairy root transformation

Constructs were introduced into the roots of wild-type *Medicago truncatula* 2HA, *M. littoralis* R108, and the *Mtsym20* mutant seedlings with hairy root transformation mediated by *A. rhizogenes* using the slightly modified protocol described by Boisson-Dernier et al. (Boisson-Dernier et al., 2001). Each construct was transformed into at least 30 plants and experimets were repeated at least twice. *M. truncatula* seeds were sterilized and germinated as described in section 5.1 and grown until the root length reached approximately 1 cm. Root radicles were then cut approximately 2 mm from their tips using a sterile scalpel, and the wounded root surface was immersed into bacterial culture of *A. rhizogenes* Arqua-1 suspension. Following bacterial inoculation, seedlings were grown on square Petri dishes containing antibiotic-free Fahraeus medium (0.5 mM MgSO4·7H2O; 0.7 mM KH2PO4; 0.8 mM Na2HPO4·2H2O; 50 µM Fe-EDTA; 0.1 µg MnSO4; 0.1 µg CuSO4; 0.1 µg ZnSO4; 0.1 µg H3BO3; 0.1 µg Na2MoO4; 5 g/l agar) for one week to facilitate the transfer of the T-DNA into the plant genome and induce hairy roots. After one week, plants were positioned between two layers of Whatman filter paper on fresh selective Fahraeus medium supplemented with 25 mg/l kanamycin to inhibit the growth of non-transformed roots. Plants were then grown for an additional two weeks to induce the growth of transgenic roots.

Eventually, non-transgenic roots, identified based on the absence of fluorescence signal of markers (eGFP, dsRed, or mCherry) used for selection, were removed using sterile tweezers and a blade under a Leica MZ10F fluorescence stereomicroscope, and plants were transferred into zeolite. For plants transformed with CRISPR/Cas9-mediated gene editing constructs, newly emerged non-transgenic roots were excised again prior to inoculation with rhizobia to prevent the dominance of

non-transformed (wild-type) roots. Each construct was transformed into the roots of at least 30 plants in transformation and experiments were repeated at least twice.

5.3.10. Design and assembly of *M. truncatula* transformation constructs

5.3.10.1. Generating the complementation construct containing the NCR-new35 gene

For the complementation experiments, constructs were generated using Single Site Gateway Recombination Cloning Technology (Thermo Fisher Scientific). *NCR-new35* gene fragments, including the native promoter, the coding sequence and the 3' UTR, were amplified from *M. truncatula* cv. Jemalong genomic DNA using the corresponding primers (Supplementary Table.3). The attB1 and attB2 sequences were tagged to the 5' ends of the forward and reverse primers, respectively to recombine the amplified fragments into the Gateway vectors, carrying attP sites, with site specific recombination. To assess the impact of promoter length on gene activity, two distinct sequences were amplified. The first sequence comprised a 516 bp promoter region, the genomic sequence of the *NCR-new35* gene, and a 903 bp 3' end, resulting in a total length of 1733 bp. For the second fragment, the forward primer was positioned to incorporate a longer promoter region, resulting in 2036 bp long fragment, including 819 bp promoter region, genomic sequence of *NCR-new35* gene and 903 bp 3' end (Table 5.3.10). The amplifications were conducted using CloneAmp™ HiFi PCR Premix (Takara Bio).

The amplified fragments were cloned into pDONR^{TM}201 entry vector using BP Clonase™ II-mediated reactions including 50 fmol of attB-tagged PCR products, pDONR^{TM}201 vector at a concentration of 150 ng/µl, and a TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0), along with the Gateway™ BP Clonase™ II enzyme mix. Reactions were incubated overnight at room temperature. The following day 1 µl Proteinase K was added to each reaction which were incubated at 37°C for 10 minutes to terminate the reaction.

BP reaction mixtures were transformed into the *E. coli* DH5α competent cells and cells with the plasmids were selected on LB-Km^{50} selective media (LB Kanamaycin 50 µg/mL), as described in section 5.9.3.1.

pDONR^{TM}201 plasmids containing the *pNCR-new35-516bp::NCR-new35gDNA-3'UTR-902bp* and the *pNCR-new35-819bp::NCR-new35gDNA-3'UTR-902bp* fragments were transferred into pKGW-prUBQ10::dsRED binary destination vector in LR reactions. The LR reaction mixture contained the 50 fmol of pDONR^{TM}201 entry vector, the pKGW-prUBQ10::dsRED vector at a

concentration of 150 ng/µl, Tris-EDTA Buffer in a 10:1 ratio (pH 8.0), and LR Clonase II enzyme. LR reactions were incubated overnight at room temperature. The next morning, reactions were supplied with 1 µL of Proteinase K and incubated at 37°C for 10 minutes to stop the reactions. LR reactions were transformed into the *E. coli* DH5α competent cells and cells carrying the plasmids were selected on LB-Sm/Spr selective media (LB containing Streptomycin (100 µg/ml) or Spectinomycin (150 µg/ml)), as described in section 5.3.9.1. The recombination product was then transformed into the *A. rhizogenes* Arqua-1 competent cells (Quandt et al., 1993) and cultured as described in section 5.3.9.2. The introduction of complementation constructs into the mutant and wild type plants was conducted using *A. rhizogenes*-mediated hairy root transformation (Boisson-Dernier et al., 2001) as described in section 5.3.9.3.

5.3.10.2. In-situ localization of NCR-new35 peptide

The localization of NCR-new35 peptide in symbiotic nodule cells was investigated with NCR-new35::mCherry fusion proteins encoded by the construct generated with the MultiSite Gateway System (Thermo Fisher Scientific) enabling simultaneous insertion of multiple DNA fragments into a single vector. DNA fragments of 819 bp long native *NCR-new35* promoter and 314 bp long coding sequence region starting from START till the STOP codon of *NCR-new35*, were amplified from *Medicago truncatula* cv. Jemalong genomic DNA using the corresponding primers tagged with attB1 and attB2 recombination sides to facilitate the subsequent recombination reactions (Supplementary Table.3). The *mCherry* coding sequence containing the attB sides was amplified to visualize the sub cellular localization of *NCR-new35* peptide. The CloneAmp™ HiFi PCR Premix (Takara Bio) was used to generate all fragments.

The promoter region and the coding sequence of the *NCR-new35* gene were cloned into pDONR221 entry vectors and *mCherry* florescent tag was cloned into pDONRP2R-P3 entry vector, containing corresponding attP sites, in BP reactions mediated by Gateway™ BP Clonase™ II Enzyme mix. Each BP reaction contained 50 fmol of attB-tagged PCR products, entry vector at a concentration of 150 ng/µl, and a TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0), along with the Gateway™ BP Clonase™ II enzyme mix. Following a day of incubation at room temperature, reactions were terminated with Proteinase K (10 mins). Subsequently, the BP reaction mixtures were introduced into *E. coli* DH5α competent cells, and bacterial cells carrying the plasmids were grown on LB-Km^{50} selective media (section 5.3.9.1). Plasmids were selected as described in section 5.3.9.2 and selected plasmids were isolated as described in 1.3.3.

The DNA fragments inserted into the entry vectors pDONR221, and pDONRP2R-P3 were assembled into the pK7m34GW-rolD: EGFP destination vector using LR Clonase II Plus enzyme (Table 5.3.10). This reaction contained 10 fmol of the 819 bp long *NCR-new35* promoter cloned into pDONR221, 10 fmol of the *NCR-new35* coding sequence (314 bp, excluding the stop codon) cloned into pDONR221, 10 fmol of the *mCherry* sequence cloned into pDONRP2R-P3, and 10 fmol of the pK7m34GW-rolD: EGFP vector, which confers Spectinomycin/Streptomycin resistance (Sm/SpR) (Table 5.3.10). The components were mixed in a reaction containing a 10:1 ratio of TE Buffer to the total reaction volume. The LR reaction was terminated with Proteinase K after one day incubation at room temperature and transferred into *E. coli* and Arqua-1 cells respectively as described in sections 5.3.9.1 and 5.3.9.2.

5.3.10.3. Designing and generating constructs for gene expression analyses

The GUS (β-glucuronidase) reporter system was applied to analyze the promoter activity of selected *NCR* genes. The promoter regions of the *NCR-new35* (819bp), *NCR169* (1178bp), *NCR211* (2055bp) and *NCR343* (1620) genes were amplified from *M. truncatula* cv. A17 genomic DNA with Phusion DNA polymerase (ThermoFisher) using specific primers (Supplementary Table.3). The amplified promoter fragments contained attB sites at their ends, were cloned into the pDONR201 Gateway entry clone vector in a BP recombination reaction catalyzed by the BP Clonase enzyme mix (50 fmol of attB-tagged PCR products, 150 ng/µl pDONR^{TM}201 vector, TE buffer (pH 8.0), the Gateway™ BP Clonase™ II enzyme mix). The correct integration of the promoter fragments into the pDONR201 entry vector was verified following *E. coli* transformation and Sanger sequencing as described in section 5.3.9.2.

Subsequently, these fragments were fused upstream with the β-glucuronidase (*GUS*) reporter gene within the modified pKGWFS7 Gateway destination vector using LR Clonase enzyme mix (150 ng/µl pDONR201 carrying corresponding NCR promoter, 150 ng/µl pKGWFS7, TE Buffer (pH 8.0) and LR Clonase enzyme mix). pKGWFS7 destination vector contained a dsRED fluorescent transformation marker, controlled by the constitutively active promoter of the *Arabidopsis thaliana Ubiquitin10* gene (prUBQ10::dsRED) to allow the selection of transgenic nodules under the fluorescent microscope (Table 5.3.10). The resulting constructs were then transformed into *E. coli* DH5α and *A. rhizogenes*, Arqua-1 competent cells as described in section 5.3.9.1 and 5.3.9.2, respectively. The Arqua-1 competent cells carrying the promoter-*GUS* reporter constructs were

introduced into the roots of wild type *M. truncatula* A17 plants using *A. rhizogenes*-mediated hairy root transformation as described in section 5.3.9.3.

5.3.10.4. Construction of the gene editing construct targeting NCR-new35

Golden Gate cloning system was employed using the pKSE466_RR vector designed to express SpCas9 (*Streptococcus pyogenes* Cas9) under the SV40 promoter, the DsRed fluorescent protein under the control of the 830 bp promoter of the *UBQ10* gene from *A. thaliana*, and the MtU6.6 promoter for sgRNA expression (Güngör et al., 2023).

To design an optimal sgRNA targeting *NCR-new35* gene in *M. truncatula* A17 and *M. littoralis* R108, the genome databases MtrunA17r5.0 and R108 0.95 were screened, respectively using the CRISPOR online tool (http://crispor.tefor.net/) and the off targets having high similarity to the target sequence (less than 3 nucleotide mismatch) were eliminated. Eventually, a sgRNA (5' GGAAGGAAAGGAGGACCACC 3') that is matching the GN_{19} NGG motif, indicative of high functionality within the CRISPR/Cas9 system, was preferred among two available candidates. The off targets of selected sgRNA are listed together with their position in the *M. trucatula* A17 genome and mismatch score in the Supplementary Table 4.

Golden gate assembly of the selected sgRNA into the pKSE_466_RR vector achieved through simultaneous digestion of the vector with BsaI restriction endonuclease and ligation of the sgRNA fragment into the vector, utilizing the unique sticky ends generated by BsaI.

To prepare the sgRNA for cloning into the pKSE466_RR vector, BsaI-compatible overhangs were tagged to 5' ends of the sgRNA in both forward and reverse orientations. This procedure generated oligonucleotides with the sequences 5'^{CTTG}GGAAGGAAAGGAGGACCACC 3' and 3' GGAAGGAAAGGAGGACCACC^{CAAA} 5', respectively. The oligonucleotides were then annealed by incubating the mixture, consisting of 10 µl of 100 pmol/µl forward oligo and 10 µl of 100 pmol/µl reverse oligo, at 75°C for 10 minutes. This annealing process resulted in the formation of a double-stranded sgRNA fragment, carrying BsaI-compatible overhangs.

The prepared double-stranded sgRNA fragment was subsequently cloned into the pKSE466_RR vector using a Golden Gate Assembly reaction. The reaction mixture of the volume of 12 µl included the following components: G buffer (10x), Thermo Scientific T4 DNA ligase buffer (10x), BsaI (New England Bioabs) restriction enzyme, T4 DNA ligase, the annealed oligonucleotide mix, and 600 ng of the pKSE466_RR vector.

The golden gate assembly reaction was then transformed into the *E. coli* competent cells (section 5.3.9.1). Following transformation, the plasmid which carried the sg*NCR-new35*::pKSE466_RR gene editing construct was isolated from recombinant cells grown on LB agar plates containing kanamycin (50 μg/ml). The plasmid was introduced into *A. rhizogenes* Arqua-1 (section 5.3.9.2) which were used for hairy root transformation experiments (section 5.3.9.3).

Table 5.3.10 Description and composition of the *Medicago truncatula* transformation constructs

Entry Clones	**Complementation Constructs**
pDONR201::*prNCR-new35-516bp*	pKGW-*prUBQ10* ::dsRED::*prNCR-new35-516bp* ::*NCR-new35gDNA* -3'UTR-902bp
pDONR201::*prNCR-new35-819bp*	pKGW-*prUBQ10* ::dsRED::*prNCR-new35-819bp* ::*NCR-new35gDNA* -3'UTR-902bp
	In-situ Localization Construct
pDONR221::*prNCR-new35-819bp*	pK7m34GW-rolD::EGFP::*prNCR-new35-819bp* ::*NCR-new35* g*DNA-nostop* ::*mCherry*
pDONR221::*NCR-new35gDNA-nostop*	
pDONRP2R-P3::*mCherry*	
	GUS Reporter Assay Constructs
pDONR201::*prNCR-new35-819bp*	pKGWFS7-*prUBQ10* ::dsRED::*prNCR-new35-819bp*
pDONR201::*prNCR343-1620 bp*	pKGWFS7-*prUBQ10* ::dsRED::*prNCR343-1620 bp*
pDONR201::*prNCR169-1178bp*	pKGWFS7-*prUBQ10* ::dsRED::*prNCR169-1178bp*
pDONR201::*prNCR211-2055bp*	pKGWFS7-*prUBQ10* ::dsRED::*prNCR211-2055bp*
	CRISPR/CAS9 Gene Editing Construct
	pKSE466_RR-*prUBQ10* ::dsRED::*sgNCR-new35*

6. PRELIMINARY RESULTS

A pool of *Medicago truncatula* cv. Jemalong J5 mutants was generated by gamma-irradiation at The Institut National de la Recherche Agronomique (INRA, Dijon, France) (Sagan et al., 1998) Gamma ray mutagenesis generated symbiotic mutant plants were classified into three groups based on their nodulation phenotype: as non-nodulating mutants (Nod-) which did not form nodules, ineffective symbiotic mutants (Fix-) which were able to form nodules (Nod+) but nitrogen fixation did not occur in the mutant nodules (Fix-) and supernodulating mutants (Nod++) which formed increased number of nodules when compared to nodule number on wild-type roots (Sagan et al., 1998). Further screening of this mutant population by Morandi and co-workers (2005) revealed additional symbiotic mutants detected in M2 generation. New mutant populations were self-fertilized till the M5 generation in order to stabilize the mutant phenotype and mutants with uniformed phenotypes were coded either as TR if isolated from the distinct M2 progenies or as TRV if isolated from a pool of M2 progenies (Morandi et al., 2005). Mutants showing clear nitrogen fixation defective phenotype were further back-crossed with parental line *Medicago truncatula* cv. Jemalong J5 and F2 segregation populations were used to evaluate the inheritance of the mutations concluding that the phenotype of each mutant was controlled by a single gene which was inherited as a recessive allele. Mutants were crossed in every possible combination to detect allelic mutants. When a wild-type phenotype was observed in the progeny after crossing of two recessive mutants, it was considered as complementation and indicated that crossed mutants were not allelic. Contrary, if a mutant phenotype was observed in the progeny of the crossed mutants, these mutants were considered as allelic because the defect of the same gene caused the symbiotic mutant phenotype. Based on the F1 complementation tests, two newly described Fix- mutant lines; TRV43 and TRV54 were found to be allelic to each other and no other mutant line was found to be allelic. Mutants TRV43 and TRV54 were assigned to *Mtsym20* locus and termed *Mtsym20-1* and *Mtsym20-2*. The *Mtsym20* was found as a distinct allele from the previously found Fix- mutants TR3, TR36 and TR183 isolated by Sagan and co-workers (1998) which were assigned as *Mtsym17*, *Mtsym18* and *Mtsym19*, respectively (Morandi et al., 2005). In this PhD project, the *Mtsym20* mutant which can form nodules (Nod+) but cannot fix nitrogen (Fix-) after rhizobium inoculation was investigated in order to analyze its symbiotic phenotype and identify the impaired gene in the mutant *Mtsym20* locus. Previously, *Mtsym20* mutant plants were crossed with *M. truncatula* A20 ecotype and an F2 segregation population was generated and used for genetic mapping by our research group. Rough genetic mapping was carried out by the MSc student Róbert

Barassó (2017) on this F2 segregating population. He identified the chromosome location of the *Mtsym20* locus on the upper arm of chromosome 4 in *M. truncatula* and subsequently started the fine mapping using markers specific to chromosome 4 and found a 3 Mbp large genomic region to be further investigated. This region is hypothesized to contain candidate genes responsible for the nitrogen fixation defective phenotype observed in the *Mtsym20* mutant.

7. RESULTS AND DISCUSSION

7.1. *Mtsym20* mutant nodules are not colonized by rhizobia in the nitrogen fixation zone

Mtsym20 was identified as a Fix- mutant line based on the mutant screen series described in chapter 6 (Sagan et al., 1998; Morandi et al., 2005). To confirm the symbiotic phenotype of *Mtsym20* and further investigate the symbiotic deficiencies in *Mtsym20* mutants compared to their wild-type counterparts, wild-type Jemalong and *Mtsym20* mutant plants were inoculated with their symbiotic partner, *Sinorhizobium medicae* WSM419 carrying the *hemA::lacZ* reporter construct (pXLGD4). Plants were grown in the absence of combined nitrogen, a condition that allows for a clearer assessment of the mutant's ability to fix nitrogen in interaction with the compatible symbiotic partner and scored for symbiotic characteristics 2 wpi (week post inoculation).

The shoots of wild-type *M. truncatula* Jemalong control plants were developed vigorously and the leaves had dark green color 2 wpi with *S. medicae* WSM419 (7.1a). The healthy growth of the wild-type plants confirmed that plants were cultivated in optimal environmental and experimental conditions for nitrogen fixation.

While the wild type *M. truncatula* Jemalong control plants developed well, the *Mtsym20* mutant plants which were grown under the same conditions presented the symptoms of nitrogen starvation with retarded growth, having reduced shoot length and small, pale green leaves underlying the nitrogen deficiency of mutant plants (Fig. 7.1b). When the plant roots were removed from the substrate and further analyzed for nodulation phenotype, elongated and pink nodules were observed on the wild-type roots (7.1c,e). In contrary, small and white nodules were observed on *Mtsym20* roots indicating the lack of leghemoglobin, which is essential for maintaining reduced oxygen levels to optimize activity of nitrogenase enzyme (Fig. 7.1d,f). In the *Mtsym20* mutants, we observed an increased number of nodules on the roots compared to the wild-type plants (Fig. 7.1c,d). The increase in the number of nodules is probably due to the negative feedback of the N status of mutant plants. Consequently, the mutant plants produce a higher number of nodules to compensate for the ineffective nitrogen fixation.

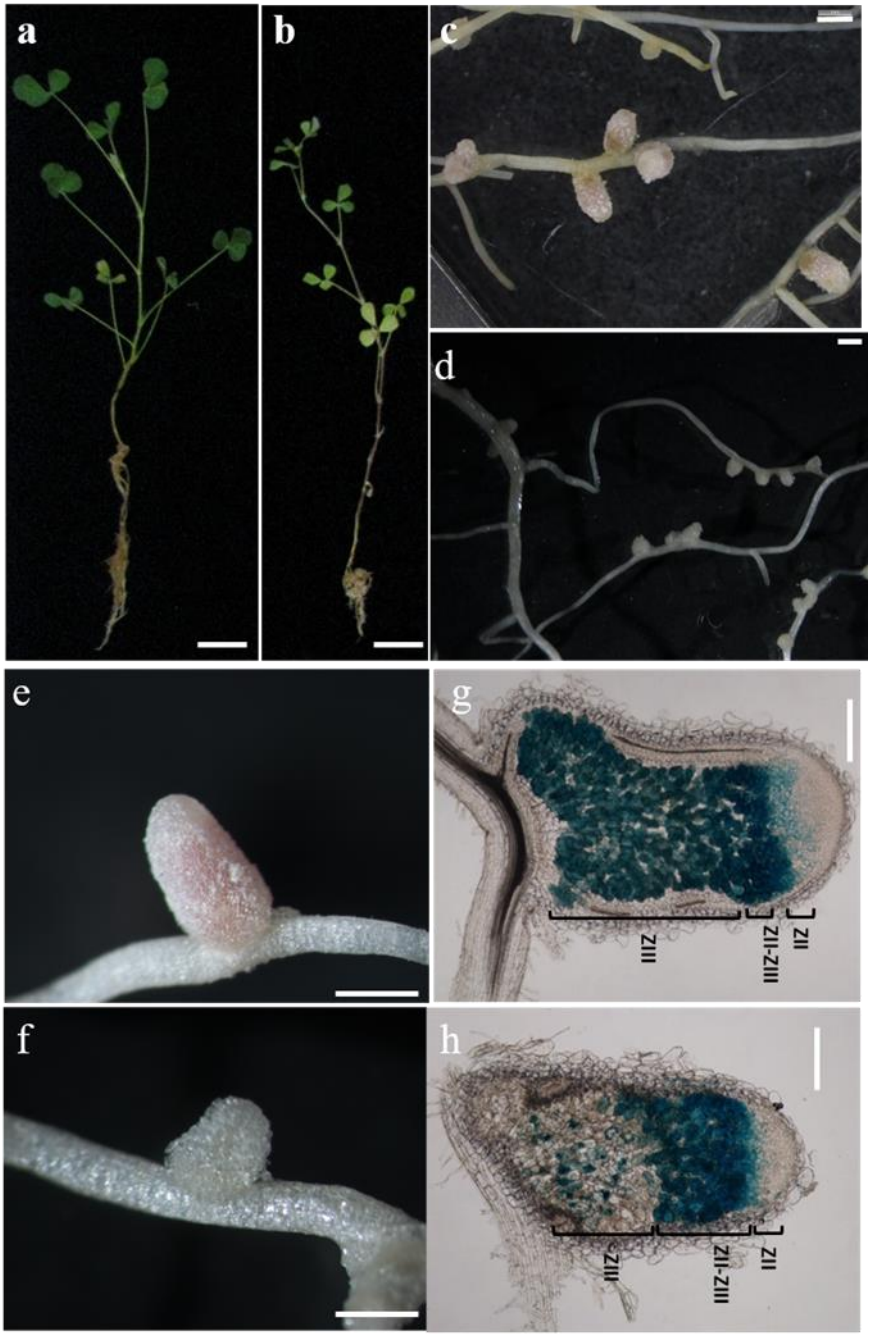


Fig. 7.1 Symbiotic phenotype of *Mtsym20* mutant compared to a wild-type plant

M. truncatula cv. Jemalong and *Mtsym20* plants were examined two weeks post inoculation with the *Sinorhizobium medicae* WSM419 strain harboring the *lacZ* reporter gene. The shoot growth of wild-type plants (a) and *Mtsym20* plants (b). Nodules developed on wild-type (c,e) and *Mtsym20* roots (d,f). (Longitudinal sections of the nodules stained for β-galactosidase activity in (g) wild-type, and (h) *Mtsym20* nodules. ZI: meristem; ZII: infection zone; ZII-ZIII: interzone ZIII: nitrogen fixation zone, Scale Bars: (a,b) 2 cm; (c,d,e,f) 1 mm; (g,h) 200 µm.

The nodule structure and colonization phenotype of *Mtsym20* nodules compared to *M. truncatula* Jemalong nodules was further analyzed with microscopy 2 wpi with *S. medicae* WSM419 expressing

the *lacZ* marker gene. Nodules were longitudinally sectioned and stained for β-galactosidase activity. The samples were examined using light microscopy to identify at what stage the symbiotic interaction was blocked. Longitudinal sections of wild-type nodules showed the characteristic zonation of indeterminate nodules. The presence of bacteria was detected in the infection zone (ZII), interzone (ZII-ZIII) and nitrogen-fixing zone (ZIII) indicating the normal colonization of wild-type nodules. The distribution of bacteria across the zones in these nodules reflected a typical progression of symbiotic colonization (Fig. 7.1g).

The nodules of the *Mtsym20* mutants were smaller in size and round-shaped or slightly elongated. However, beside the morphological differences, the characteristic zonation of indeterminate nodules was visible in the mutant nodules, too. In contrast to wild-type nodules, symbiotic cells in *Mtsym20* nodules were colonized by rhizobia only in infection (ZII) and inter-zones (ZII-ZIII). The region corresponding to the nitrogen fixation zone (ZIII) of mutant nodules was not colonized by rhizobia, but sporadic staining was detected, suggesting the release of rhizobia from infection threads having saprophytic life style. The restricted colonization of rhizobia in the mutant nodules suggests that the defective gene in *Mtsym20* is required for complete differentiation or persistence of rhizobia (Fig. 7.1h).

7.2. Rhizobial colonization in *Mtsym20* mutant nodules shows similarity to the colonization patterns observed in nodules of other fix- mutants

The symbiotic phenotype of *Mtsym20* mutant was analyzed along with a group of other Fix- mutants which have been studied intensively studied in our research group. This group of mutants included the *Mtsym19* (Sagan et al., 1998; Morandi et al., 2005) identified in the same mutant screening program with *Mtsym20* as described earlier in chapter 6, the *Mtdnf4-1* (Starker et al., 2006), a mutant observed during the genetic screens of fast neutron irradiated seeds of *M. truncatula* Gaertn. cv. which exhibits the Fix- phenotype due to the lack of *NCR211* gene (Kim et al., 2015), the *Mtdnf7-2* (Domonkos et al., 2013), identified during the symbiotic screen of a fast neutron-bombarded population of *M. truncatula* cv. Jemalong, known for its mutant phenotype resulting from the absence of *NCR169* (Horváth et al., 2015), and the NF-FN9363 (Xi et al., 2013), a mutant, isolated during genetic screens of fast neutron bombarded *M. truncatula* A17 plants, exhibiting mutant phenotype due to the lack of *NCR343* (Horváth et al., 2023).

Nodulation and growth phenotype of *Mtsym20* was compared to wild-type and ineffective mutant plants, including, *Mtsym19* (analyzed by Mónika Tünde Tóth), *Mtdnf4-1* (*dnf4-1*) and *Mtdnf7-*

2 (*dnf7-2*), NF-FN9363 (analyzed by Beatrix Horváth) two weeks after inoculation with *S. medicae* WSM419 (SmWSM419) (Fig. 7.2). The wild-type control plants showed robust shoot development, featuring dark green leaves (Fig. 7.2a) and elongated pink nodules on the roots (Fig. 7.2b), indicating effective nitrogen fixation. In contrast, all tested mutants displayed symptoms of nitrogen starvation, characterized by light green foliage and stunted growth (Fig. 7.2a). Distinct roundish or slightly cylindrical white nodules developed on the roots of all tested mutant plants indicating the absence of leghemoglobin (Fig. 7.2b). In order to assess rhizobia colonization, symbiotic nodules were subjected to longitudinal sectioning and stained for β-galactosidase activity 2 wpi with the SmWSM419 (pXLGD4) strain constitutively expressing the *lacZ* gene (*hemA::lacZ*). Wild-type nodules displayed the characteristic zonation of indeterminate nodules, with rhizobia colonization of symbiotic cells in all zones. Although mutant nodules were significantly less elongated compared to wild-type nodules, zonation within the nodules was still observable. In contrast to nodules developed on wild-type roots, nodules collected from *Mtsym20*, *Mtsym19*, NF-FN9363, *Mtdnf4-1* and *Mtdnf7-2* roots exhibited presence of rhizobia only in the infection zone, and interzone when stained for β-galactosidase activity. However, the cells in the nitrogen fixation zone of mutant nodules found to be uninfected suggesting that the defect in these Fix- mutants block the colonization or impeded the persistence of rhizobia (Fig. 7.2c). Additionally, the interzone of *Mtsym20*, *Mtsym19*, NF-FN9363 and *Mtdnf4-1* nodules was noticeably enlarged compared to wild-type nodules. Although the interzone of the nodules developed on *Mtdnf7-2* mutant roots was also exceeding the few cell layers observed in the interzone of wild-type nodules, its extension was less prominent compared to the interzone of the nodules collected from the roots of other tested mutants suggesting that *NCR*169 gene may have an earlier function in the symbiosis compared to the genes responsible from Fix- phenotype of other tested mutants (Fig. 7.2c).

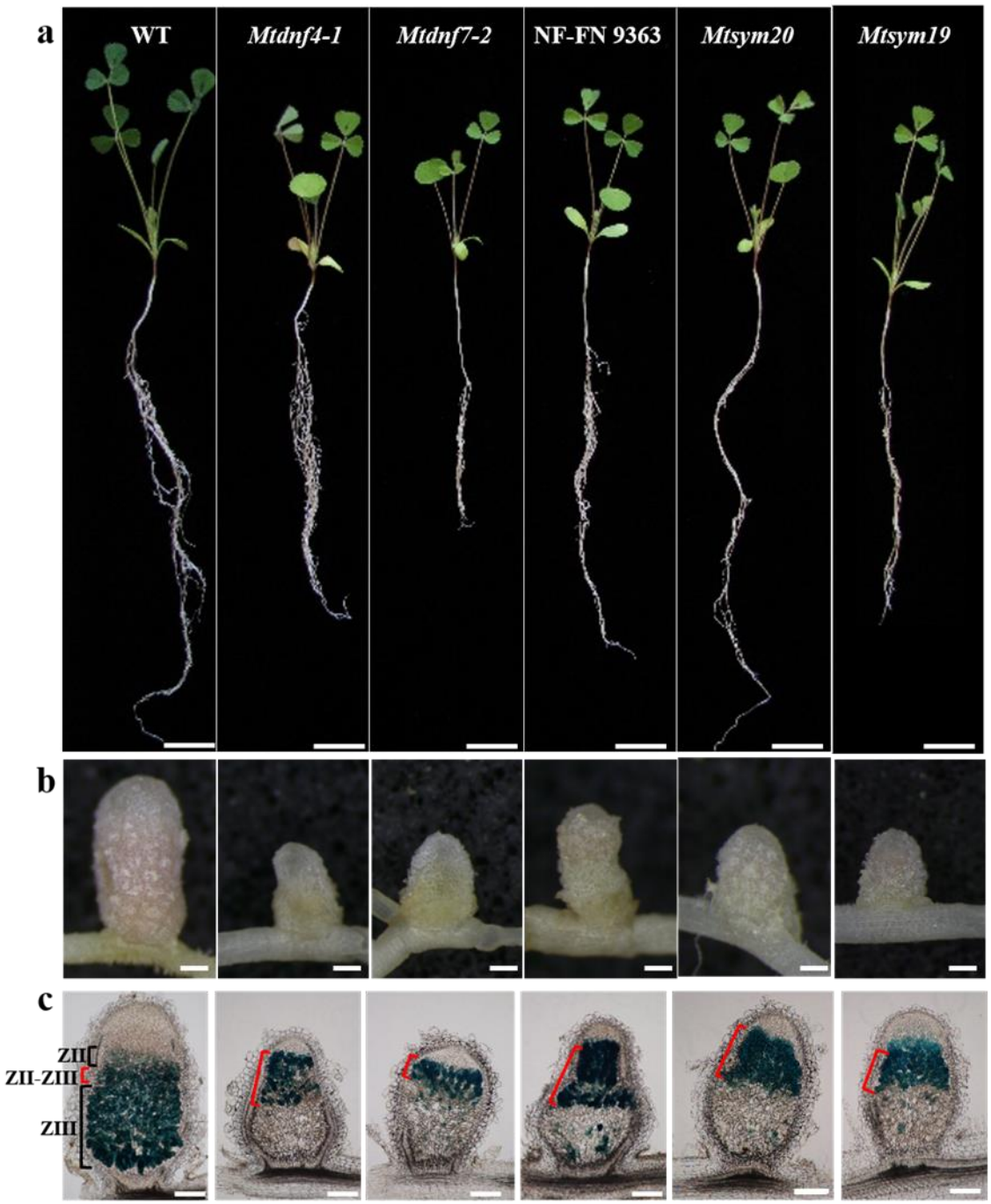


Fig. 7.2 Comparative symbiotic phenotype analysis of nitrogen fixation inefficient mutants

Mtsym20, *Mtsym19 Mtdnf4-1*, *Mtdnf7-2*, NF-FN9363 mutant plants and *M. truncatula* cv. Jemalong wild type plants were evaluated 2 weeks post inoculation with *Sinorhizobium medicae* WSM419 (pXLGD4) (Horvath et al.,2023). Shoot phenotype of wild-type and mutant plants (a). The symbiotic nodule phenotype of wild-type and mutant plants (b). Staining for the β-galactosidase activity in nodule sections of wild-type and mutant plants (c). Extension of each zone in the nodules is indicated by brackets. Infection zone (ZII) and nitrogen fixation zone (ZIII) are marked with black brackets on wild-type nodule whereas interzones (ZII-ZIII) are highlighted in red on the wild-type and mutant nodules (c). Scale Bars: (a) 2 cm; (b,c) 200 µm.

7.3. Infected symbiotic cells do not persist in the region corresponding to the nitrogen fixation zone of *Mtsym20* Nodules

Differentiation of bacteroids in indeterminate nodules is associated with morphological and cytological changes such as elongation of rhizobia and amplification of DNA content (Mergaert et al., 2006). In addition to rhizobia, symbiotic cells also undergo differentiation to facilitate the formation of nitrogen-fixing nodules (Vinardell et al., 2003; Nagymihály et al., 2017). To analyze colonization in details, morphology of bacteroids and symbiotic cells in *Mtsym20* mutant nodules was compared to wild-type and *Mtsym19* (analyzed by Mónika Tünde Tóth), *Mtdnf4-1*, *Mtdnf7-2* and NF-FN9363 (analyzed by Beatrix Horváth) mutant nodules using laser scanning confocal microscopy. SmWSM419 infected nodules were fixed with paraformaldehyde and 70 µm thick sections were prepared at 2 wpi. Sections were stained using the SYTO13 nucleic acid binding dye to visualize rhizobia in green. Plant nuclei also fluoresced in green, while the autofluorescent cell wall and accumulated polyphenolic compounds were observed in the red channel under the confocal microscopy allowing a simultaneous assessment of both bacterial components and plant cellular composition in the analyzed sections (Fig. 7.3).

The examination of nodules on wild-type roots demonstrated rhizobia-colonized cells in the infection (ZII), interzones (ZII-ZIII), and nitrogen fixation zones (ZIII) (Fig. 7.3a1). The invasion of cells in ZII and ZII-III of *Mtsym20*, *Mtsym19*, NF-FN 9363, *Mtdnf4-1* and *Mtdnf7-2* mutants and wild-type nodules were similar. The interzones, where bacteria differentiate into their symbiotic form, were extended in mutant nodules compared to wild-type nodules, consistent with the findings in the β-galactosidase activity experiments. Additionally, in contrast to wild-type nodules, cells in the basal part (ZIII) of mutant nodules were not colonized by rhizobia but exhibited sporadic autofluorescence speckles (Fig. 7.3a1-f1 and a5-f5). Autofluorescence in wild-type nodule tissues can be derived from phenolic compounds that cause lignification in response to pathogens, resulting in intense and extensive defense mechanisms (Van de Velde et al., 2006). In addition to these cases, autofluorescence can be observed in early senescent nodule cells in nitrogen fixation deficient nodule tissues due the accumulation of polyphenolics. Early senescence is induced by the deficiency of bacterial persistence in mutant nodules or the block of the incomplete differentiation of host or rhizobia cells (Van de Velde et al., 2006; Pérez Guerra et al., 2010; Horváth et al., 2015; Kim et al., 2015).

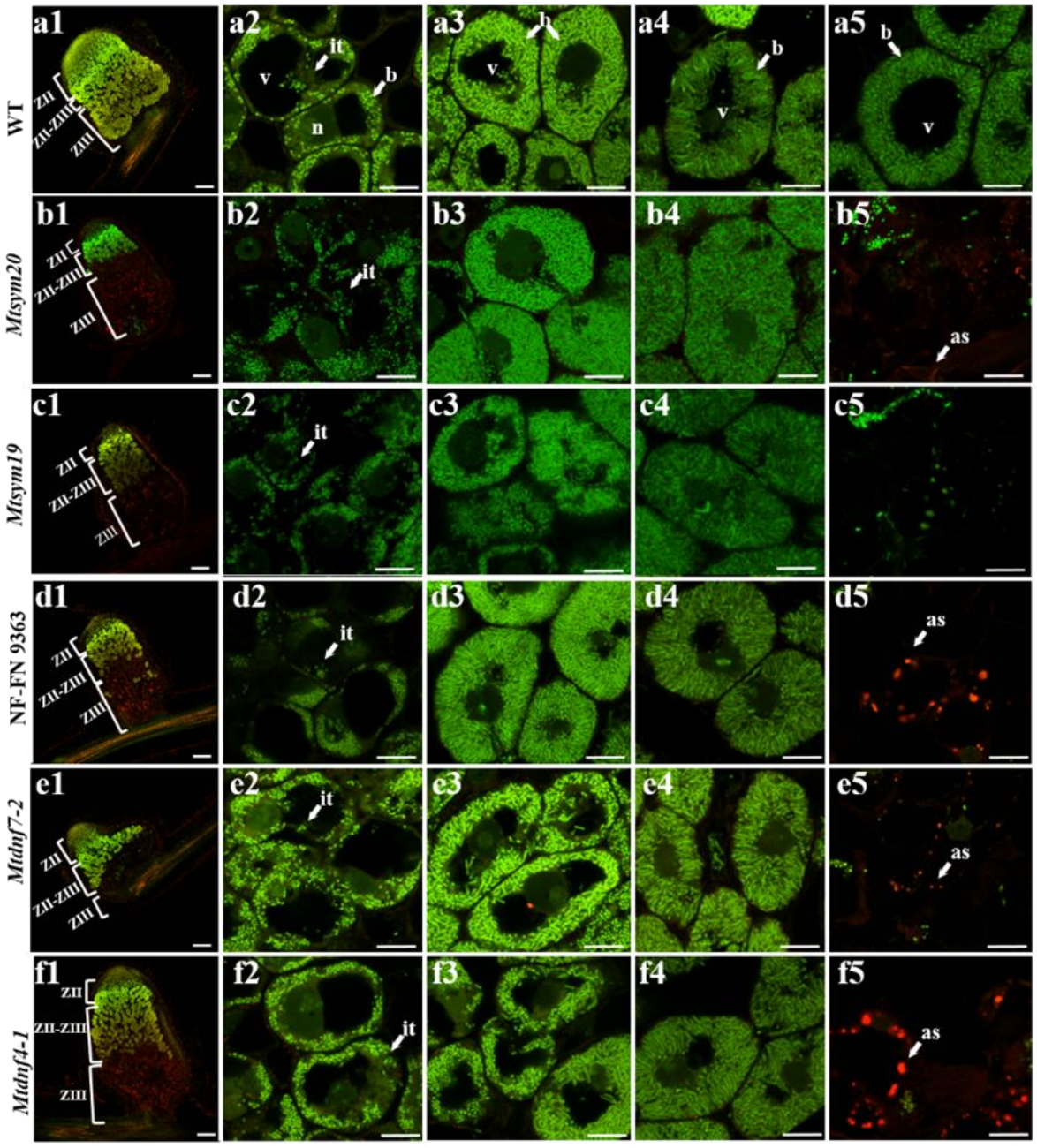


Fig. 7.3 Morphological comparison of symbiotic nodule cells and bacteroids in *Mtsym20*, *Mtsym19*, *NF-FN9363*, *Mtdnf4-1*, and *Mtdnf7-2* mutant and wild-type nodules

SYTO13-stained nodule sections of *Mtsym20*, *Mtsym19*, *Mtdnf4-1*, *Mtdnf7-2* and NF-FN9363 mutant and wild type nodules were observed 2 weeks post-inoculation with *Sinorhizobium medicae* WSM419 (Horvath et al.,2023). Whole nodule sections of wild type and Fix- mutants (a1-f1). Rhizobia invasion in the infection zone (ZII) of wild-type (a2) and mutant nodules (b2-f2). Morphology and organization of bacteroids around vacuoles in the distal part of interzone (ZII-ZIIId) (a3-f3) and in the proximal interzone (ZII-ZIIIp) (a4-f4). Nitrogen fixation zone (ZIII) of wild-type nodule sections, occupied with elongated nitrogen-fixing bacteroids (a5), and ZIII of mutant nodule sections deficient in viable bacteroids (b5-f5). SYTO13 staining of nodule sections visualized bacteria and host cell nuclei in green (a1-f5). Non-viable nitrogen-fixing bacteria in ZIII of mutant nodules observed with autoflourescence red color (b5-f5). It - infection thread v - vacuole; n - nucleus; b - bacteroids; as-autofluorescence. Scale Bars: (a1-f1) 200 µm; (a2-f5) 20 µm.

Higher magnification of nodule cells revealed that bacteria were elongated in the interzone and the proximal part of nitrogen fixation zone of wild-type nodules indicating the complete differentiation of rhizobia to bacteroids (Fig. 7.3a3,a4). The colonized cells in infection and interzone of *Mtsym20* as well as *Mtsym19*, NF-FN 9363, *Mtdnf4-1*, *Mtdnf7-2* mutant nodules showed similar morphology to those in corresponding zones of wild-type nodules. Infected plant cells contained large vacuoles characteristic of the infection zone and these vacuoles were surrounded by rod-shaped and slightly elongated rhizobia indicating that the invasion of rhizobia was completed in infection zone of all observed mutant and wild-type nodules. Infection threads containing rod-shaped bacteria did not show any abnormal morphology in mutant nodules compared to wild-type nodules (Fig. 7.3a2-f2). The terminal differentiation of bacteroids begins in the distal part of the interzone (ZII-ZIIId) in both mutant and wild-type nodules. Rhizobia in these cells were observed to enlarge due to endoreduplication of the bacterial genome. Elongated and well-organized bacteroids, densely located between the cell wall and membrane of the central vacuoles, were found in the cytoplasm both in wild-type and in all Fix- mutant nodules (Fig. 7.3a3-f3). In the proximal part of interzone (ZII-ZIIIp), cells had smaller vacuoles and bacteroids were less dense both in wild-type and mutant nodules. Additionally, shorter and less organized bacteroids were observed in these nodule cells of all ineffective symbiotic mutants compared to wild-type nodules indicating that elongation and differentiation of bacteroids is not consistent in the proximal part of interzone in the mutant nodules (Fig. 7.3.1a4-f4). In the nitrogen fixation zone of wild-type nodules, large plant cells and fully elongated bacteroids were found (Fig. 7.3a5). In contrast, no elongated bacteria were observed in the ZIII of *Mtsym20, Mtsym19*, NF-FN 9363, *Mtdnf4-1*, *Mtdnf7-2* nodules but rod-shaped non-differentiated saprophytic rhizobia were detected sporadically (Fig. 7.3b5-f5). These findings together with the result of β -galactosidase staining, described in sub-chapter 7.2, confirm the dysfunctional state of nitrogen fixation in the *Mtsym20, Mtsym19*, NF-FN 9363, *Mtdnf4-1*, *Mtdnf7-2* mutants.

Due to the result of our parallel genetic mapping experiments, an allelism test was conducted between *Mtsym19* and *Mtsym20* as explained in sub-chapter 7.6. Results indicated that *Mtsym20* is allelic to *Mtsym19* in contrary to the findings in the previous study (Morandi et al., 2005) describing these mutants as distinct alleles. Therefore, *Mtsym19* was excluded from the further phenotypic characterization experiments.

7.4. Mutation in the *Mtsym20* plants causes defects in the late stage of bacteroid differentiation

In *M. truncatula*, the conversion of bacteria into nitrogen fixing bacteroids is associated with the increase in cell volume and genome amplification, leading to the formation of elongated bacteroids known as E-type bacteroids (Mergaert et al., 2006). These E-type bacteroids can undergo substantial elongation, expanding up to 10-fold of their initial size during the differentiation process (Mergaert et al., 2006; Van de Velde et al., 2010; Montiel et al., 2017). Consequently, the examination of bacteroid size in mutant plants compared with free-living bacteria and with rhizobia in wild-type nodules provides information regarding the degree of bacteroid development and provides insights into the specific stages affected by the mutation.

To analyze the differences in bacteroid length and genome amplification level of rhizobia in *Mtsym20*, NF-FN9363, *Mtdnf4-1*, *Mtdnf7-2* and wild-type nodules, bacteroids were isolated from nodules 2 wpi with SmWSM419. The isolated bacteroids were stained with propidium iodide (PI) fluorescent dye and captured using confocal laser scanning microscopy. These images were utilized for size measurements of bacteroids with the ImageJ program (Schneider et al., 2012) (Fig. 7.4.1b). A total of 1000 bacteroids from each of the wild-type A17 and mutant lines were measured and categorized into 15 size groups ranging from 1 µm to 15 µm. The percentage of bacteroids in each category was calculated to illustrate the differences in bacteroid size isolated from wild-type and mutant nodules (Fig. 7.4.1a).

The count of bacteroids in category 1 and 2 exhibited an increase in *Mtsym20*, NF-FN9363, *Mtdnf4-1*, *Mtdnf7-2* mutant nodules compared to A17 nodules, indicating the higher abundance of undifferentiated bacteria in these mutants. These bacteria probably correspond to those found in the infection thread or newly released intracellular bacteria. In categories 4, 5 and 6, the number of bacteroids in mutant nodules slightly surpassed that in A17 nodules, suggesting an accumulation of bacteroids whose elongation was impeded before completing the differentiation process. In contrast, in categories 7, 8 and 9, where bacteroid elongation is completed, mutant nodules contained a reduced percentage of bacteroids compared to wild-type nodules indicating the effect of mutations on implemented bacteroid differentiation (Fig. 7.4.1a).

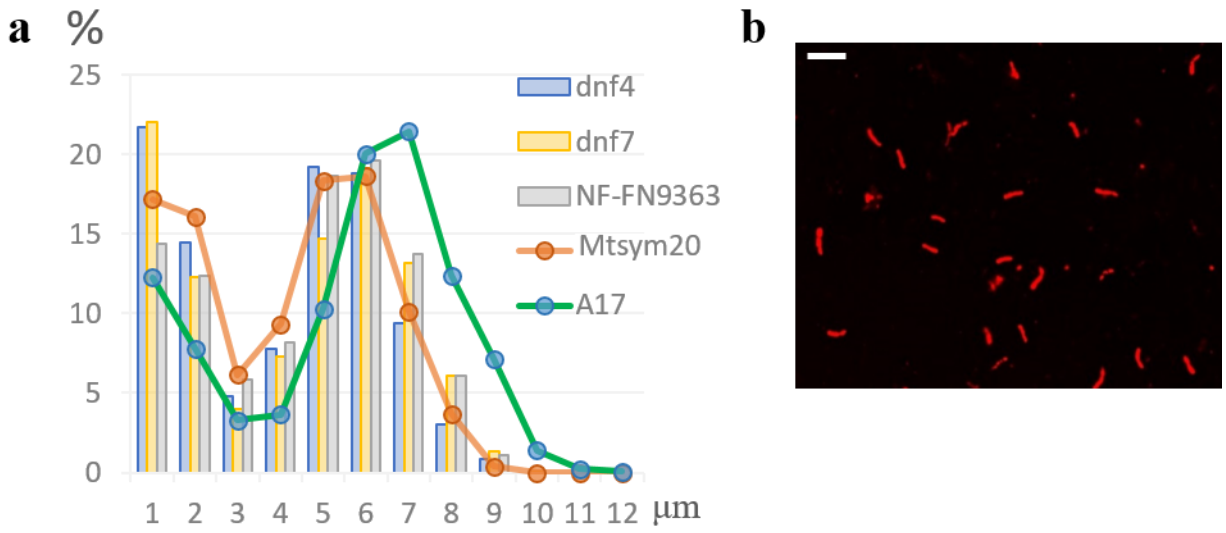


Fig. 7.4.1 Comparative analysis of bacteroid size

Bacteroids were extracted from wild type and Mtsym20, Mtdnf4-1, Mtdnf7-2, and NF-FN9363 mutant nodules two weeks post inoculation with S. medicae WSM419 (pXLGD4) and stained with propidium iodide (PI). Bacteroid lengths were quantified from images obtained from confocal laser scanning microscopy and analyzed with the ImageJ software. A higher proportion of shorter bacterial cells (1-3 µm) was observed in the mutants compared to those from wild type nodules. Conversely, the proportion of longer bacteroids (over 6 µm) was significantly higher in wild type nodules compared to those found in the mutant nodules (a). Image of bacteroids captured under confocal laser scanning microscope (b). Scale Bar: 10 µm

In collaboration with Mickaël Bourge (Institut de Biologie Intégrative de la Cellule, CNRS, Gif-sur-Yvette, France), further analyses were conducted to assess bacteroid size and to evaluate endoreduplication by measuring the DNA content level using flow cytometry (Fig. 7.4.2). The flow cytometry method employs light scatter for morphological features and fluorescence signals for cell properties, where Forward Scatter (FSC) represents cell size, and Side Scatter (SSC) indicates cell internal content. The preference for flow cytometry arises from its ability to rapidly assess numerous events, thereby generating essential statistical data for population evaluation (Narayana et al., 2020).

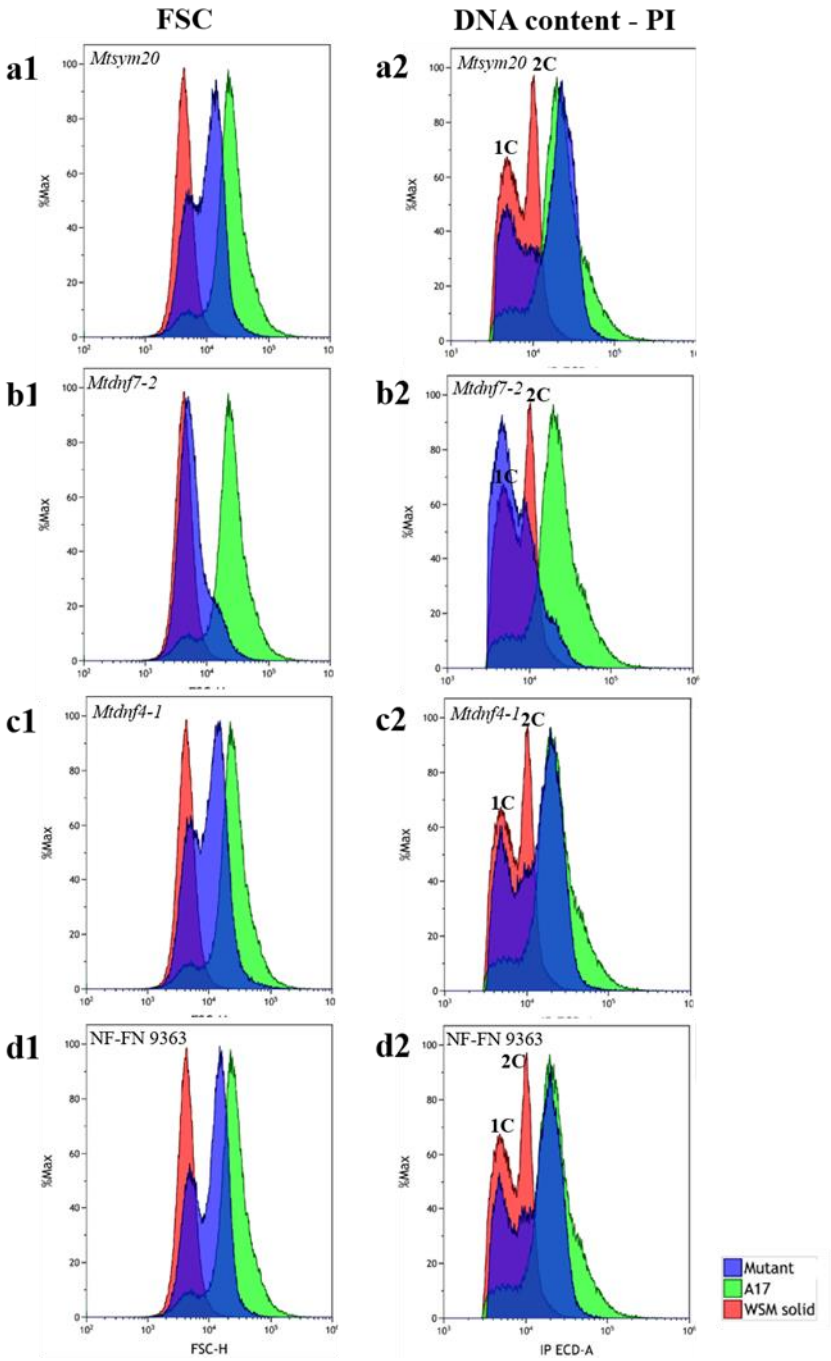


Fig. 7.4.2 Comparative flow cytometry analysis of bacteroid size and DNA content level

S. medicae WSM419 bacteroids were isolated from *Mtsym20*, *Mtdnf4-1*, *Mtdnf7-2*, NF-FN9363 and wild-type A17 nodules at 16 days post-inoculation (dpi). Bacteroid size was inferred from forward light scatter (FSC) signals (a1-d1), while endoreduplication levels were measured by fluorescence intensity (a2-d2). The bacteroid populations from mutant nodules showed a shift towards smaller sizes (a1-d1) and reduced ploidy levels were measured when compared to bacteroids isolated from wild-type nodules (a2-d2). WSM solid means free living SmWSM419 rhizobia grown on plates and washed off for analysis.

The results of flow cytometry analyses revealed that the overlap between *S. medicae* WSM419 free-living bacteria and bacteroids, isolated from *Mtsym20*, NF-FN9363, *Mtdnf4-1* and *Mtdnf7-2* mutants, is more pronounced compared to the overlap of the cultured free living bacteria with A17-derived bacteroids (Fig. 7.4.2a1-d1). This indicates a greater presence of non-elongated bacteroids in mutant nodules compared to A17 nodules. Analyzing the percentage of elongated bacteroids, an initial overlap is observed between mutant and A17-derived bacteroids, however as bacteroid size increases, the percentage of the bacteroids isolated from mutants decrease dramatically. Eventually, only fully elongated bacteroids from wild-type A17 nodules are present, suggesting a reduction in the number of elongated bacteroids in mutant nodules compared to wild-type nodules, consistent with our previous bacteroid size measurements using ImageJ (Fig. 7.4.2a1-d1). Remarkably, the bacteroid population isolated from *Mtdnf7-2* mutant nodules demonstrates a smaller overlap with those isolated from A17 nodules compared to other mutants. This observation is consistent with findings of our colonization phenotype experiments as detailed in sub-chapters 7.2 and 7.3, revealing a less extended interzone in the nodules of *Mtdnf7-2* mutants compared nodules of *Mtsym20*, NF-FN9363, *Mtdnf4-1*. Additionally the result of transcriptomics data of different nodule zones obtained by laser-capture microdissection corroborate these findings (LCM) (Roux et al., 2014) (Fig. 7.4.2b1). Together, these results suggest that *NCR169* gene may have an earlier function in differentiation process comparing to genes responsible for the mutant phenotypes of *Mtsym20*, NF-FN9363 and *Mtdnf4-1*.

During differentiation, bacteroids undergo an endoreduplication process characterized by genome duplications, leading to an increase in bacterial ploidy level (Mergaert et al., 2006). To assess the extent of endoreduplication, the DNA content of bacteroids was measured using flow cytometry. Initially, the fluorescent nucleic acid binding dye propidium iodide (PI) was introduced to the bacteroid suspension. The fluorescent signal emitted by the bacteroids is then quantified using a flow cytometer, with the fluorescence emission level considered indicative of the DNA content in the bacteroids (Institut de Biologie Intégrative de la Cellule, CNRS, Gif-sur-Yvette, France by Mickaël Bourge). The average DNA content of bacteroids obtained from *Mtsym20*, NF-FN9363, *Mtdnf4*-1 mutant nodules exceed the peaks of 1C or 2C ploidy levels observed in cultured *S. medicae* WSM419 cells, implying that bacteroids in mutant nodules were polyploids. The DNA content levels of rhizobia purified from *Mtsym20*, NF-FN9363, *Mtdnf4-1* mutant showed a moderate decrease compared with samples isolated from wild-type nodules, suggesting an advanced genome amplification in bacteroids from mutant nodules (Fig. 7.4.2a2,c2,d2). Exceptionally, DNA content level of bacteroids derived from *Mtdnf7-2* mutant was lower compared to other mutants aligning with the bacteroid size

measurements observed in this mutant. (Fig. 7.4.2b2). The fluorescent range of bacterial cells isolated from *Mtsym20*, NF-FN9363, *Mtdnf4-1* mutant nodules was narrower at higher levels compared to rhizobia from wild-type nodules, indicating a reduced number of bacteroids with highly amplified genomes. Taking together the results of the analysis of bacteroid size and DNA content levels, a slight difference was observed between bacteroids isolated from *Mtsym20* and wild-type A17 plants. Based on these results, we concluded that the mutation in *Mtsym20* results in defects at the late phase of bacteroid differentiation similar to NF-FN9363 and *Mtdnf4-1* mutants lacking essential *NCR* (*Nodule specific Cysteine Rich*) genes.

7.5. Defining the position of *Mtsym20* locus in linkage group 4 of *Medicago truncatula*

To identify the gene responsible for nitrogen fixation deficiency in *Mtsym20* and to study further its function, fine mapping and positional cloning experiments were carried out. In the genetic map co-segregation of genetic markers with the symbiotic phenotype of *Mtsym20* mutant was analyzed to identify the genomic region of the mutant on the chromosome.

Previously, the position of *Mtsym20* was found in a 3 Mbp region in the upper arm of chromosome 4 using the color map method developed by Kiss and co-workers (1998) (Róbert Barassó, Msc Thesis, Szent-István University, Gödöllő, 2017). We continued the genetic mapping using the genotypic and phenotypic data of 593 plants from *Mtsym20*x*A20* F2, F3 and F4 segregation populations. The genetic markers which were previously mapped into the region were screened with PCR amlifications and plants showing recombination events in the genomic region were selected to further narrow down the genomic region. The genotypes of selected plants from segregation populations were determined for these genetic markers designed based on *M. truncatula* A17 genome sequence obtained from Mt4.0 JBrowse (JCVI.4) or on an earlier version of the database (Supplementary Table 1). These markers were positioned on sequences where one parental allele contains insertions or deletions compared to the other parental allele, leading to fragment size differences that reveal polymorphisms which enables identification of homozygous (A20 or *Mtsym20*) or heterozygous alleles (A20 and *Mtsym20*).

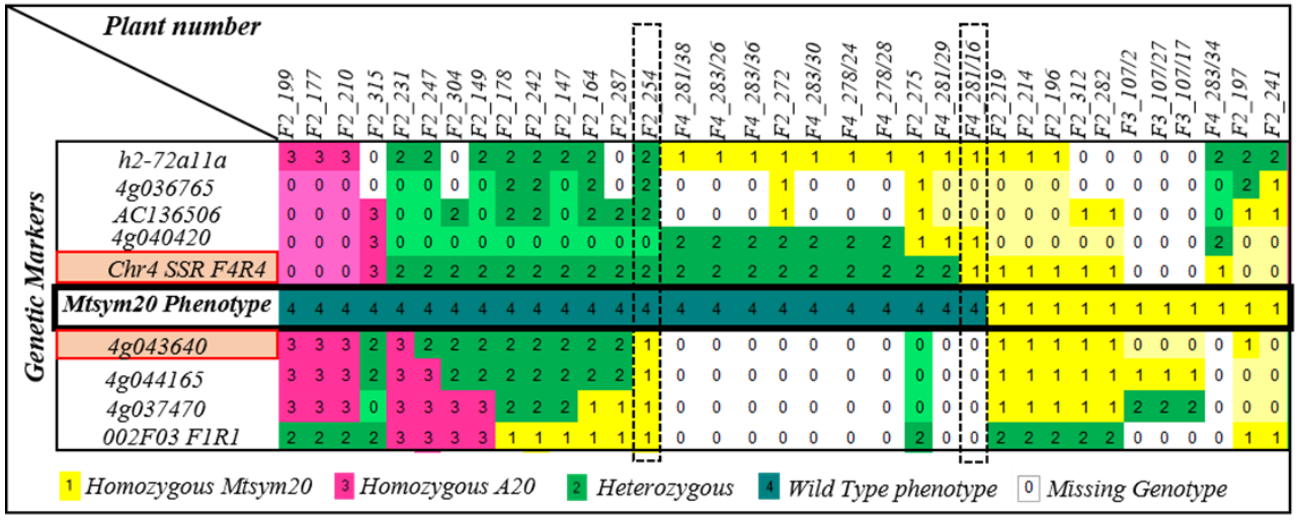


Fig. 7.5.1 The fine mapping of *Mtsym20* mutant

The color map contains phenotype data of individuals showing recombination and their genotype data for markers located on chromosome 4. Genotype and phenotype of each individual is encoded with colors and numbers as indicated below the figure. Border markers of Mtsym20 genomic region are highlighted with orange and outlined in red. Columns of the plants displaying recombination sites are enclosed in black dashed lines, and phenotype row is bordered in solid black. The predicted missing genotypes are filled in with a different shade of the color and non-predictable genotypes were left blank.

The genomic region of the symbiotic locus of *Mtsym20* mutant was successfully decreased by identifying individuals showing recombination. The color map of the region shows the genotype and the phenotype of selected plants from the segregation populations (Fig. 7.5.1) and two plants, F2/254 and F4_281/16 clearly defined the position of the symbiotic locus of *Mtsym20* on chromosome 4. The heterozygous genotype of the plant F2_254 showing wild type symbiotic phenotype, changes to maternal (mutant) homozygous genotype between the genetic markers Chr4 SSR F4R4 and 4g043640, indicating a recombination between these genetic markers and positioning the symbiotic locus of *Mtsym20* downstream of Chr4 SSR F4R4. Similiarly, the plant no. F4_281/16 has a heterozygous genotype for the genetic marker 4g043640, and it showed mutant genotype for the previous marker upwards, Chr4 SSR F4R4, indicating that a recombination occurred between these two markers in this plant. Based on these recombination events, and the symbiotic phenotype of the plants, the *sym20* gene could be located between the genetic markers Chr4 SSR F4R4 and 4g043640 (Fig. 7.5.1). This genomic region covers 89,398 bp (base pairs) in the *M. truncatula* genome, including 12 identified genes and two gaps based on *M. truncatula* genome (Mt4.0) JCVI.4 database. The genes within the genomic region of *Mtsym20* are detailed along with their genomic positions and annotated functions in Table 7.5.1. Based on the predicted function of the genes, we cannot point out a single candidate gene to be definitely responsible for nitrogen fixation deficiency of the mutant. However, gaps

identified in this region could potentially harbor additional genes, indicating that the currently identified list of genes within this region may not be complete and the actual size of the region might be larger than initially estimated.

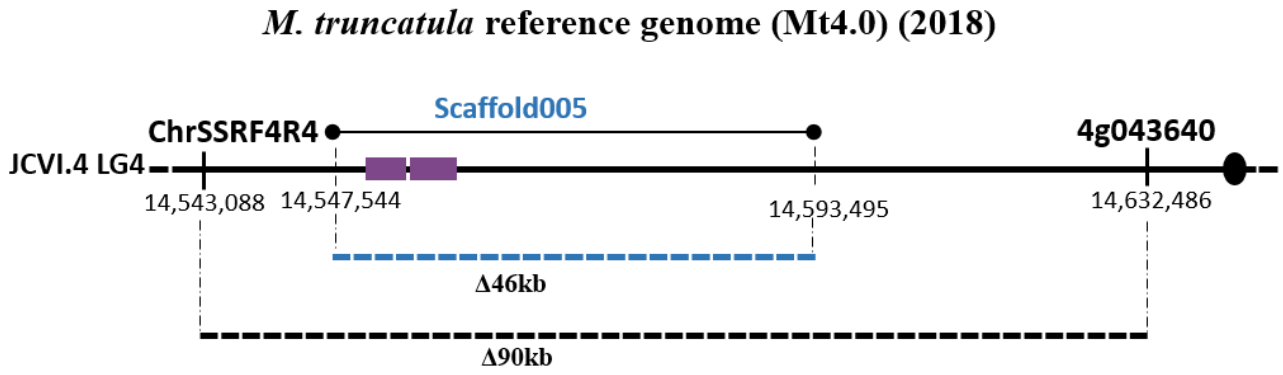


Fig. 7.5.2 Genomic region of *Mtsym20* mutant based on *Medicago truncatula* JCVI.4 genome database

M. truncatula JCVI.4 database indicates the genomic region of Mtsym20 between genetic markers Chr4SSR F4R4 and 4g043640 spreading appr. 90 kb. Scaffold005 aligned to an appr.46 kb long sequence in the genomic region is shown in blue color while gaps in the overlapping region are shown in purple. The genomic positions are indicated below the scale in base pairs.

To further refine the genomic region and clarify the presence of the gaps in the genomic region, new genetic markers were designed based on the available genome sequence located between the flanking markers. The specificity of the designed primer sequences was assessed by screening the *Medicago truncatula* genome (Mt4.0) using the JCVI.4 database. However, the primers designed for the genomic region of *Mtsym20* were not only aligned with the targeted genomic region on chromosome 4 but also showed a high degree of similarity (~99%) to another location in the *M. truncatula* A17 genome sequence on Scaffold005, which was aligning to the genomic region of *Mtsym20* spanning to 45,950 bp. This overlapping genomic sequence also contains two gaps making it challenging to further investigate the size of the genomic region of *Mtsym20* mutant (Fig. 7.5.2 and Table 7.5.1).

Table 7.5.1. The gene content in the genomic region of *Mtsym20* based on the *Medicago truncatula* JCVI.4 genome database

Mtsym20 genomic region flanked by the Chr4SSRF4R4 and 4g043640 genetic markers contains 12 genes, 8 of them overlapping the sequence of Scaffold005 based on the *M. truncatula* JCVI.4 reference genome database. The overlapping gene content on chromosome 4 and Scaffold005 in the genomic region of Mtsym20 is highlighted in blue while sequence gaps are marked in pink.

Genetic Marker	*Gene ID*	*Genomic Position*	*Gene Function*
4g043640	Medtr4g043640.1	14632486 - 14633293	glutathione S-transferase tau, putative
	Medtr4g043650.1	14621396..14625115	glycerol-3-phosphate dehydrogenase
	Medtr4g043660.1	14609657..14612948	RNA-binding domain CCCH-type zinc finger protein
	Medtr4g043715.1	14608363..14609121	hypothetical protein
	Medtr4g043690.1	14583510..14596432	calcium-transporting ATPase 4, plasma membrane-type protein
	Medtr4g043700.1	14584711..14585244	hypothetical protein
	Medtr4g043720.1	14579528..14582035	hypothetical protein
	Medtr4g043730.1	14576803..14577956	hypothetical protein
	Medtr4g043740.1	14570899..14572196	hypothetical protein
	Medtr4g043750.1	14568374..14569676	transmembrane protein, putative
	Medtr4g043760.1	14565252..14567768	mannan endo-1,4-beta-mannosidase-like protein
		GAP 01830	
		GAP 01829	
ChrSSRF4R4	Medtr4g040560.1	14543203..14544607	transmembrane protein, putative

Consequently, considering the possibility that the representation of the genomic region in the JCVI.4 database might be inaccurate or incomplete, a *de novo* assembly strategy was employed as a further approach to reconstruct the SYM20 genomic region in *M. truncatula* A17. Overlapping scaffolds from various *M. truncatula* accessions were identified and assembled using the data available on website of the *M. truncatula* HAPMAP program (www.medicagohapmap.org). The sequence alignments revealed that Scaffold157 within the HM022 accession provides significant coverage of both Scaffold005 and Scaffold003 from *M. truncatula* A17. Consequently, scaffold157 was used as a reference sequence; however, it also contained gaps and did not provide a complete sequence of the region of interest. Therefore gaps were bridged with comprehensive assessment of sequence overlaps between additional accessions for a more complete representation of the targeted region.

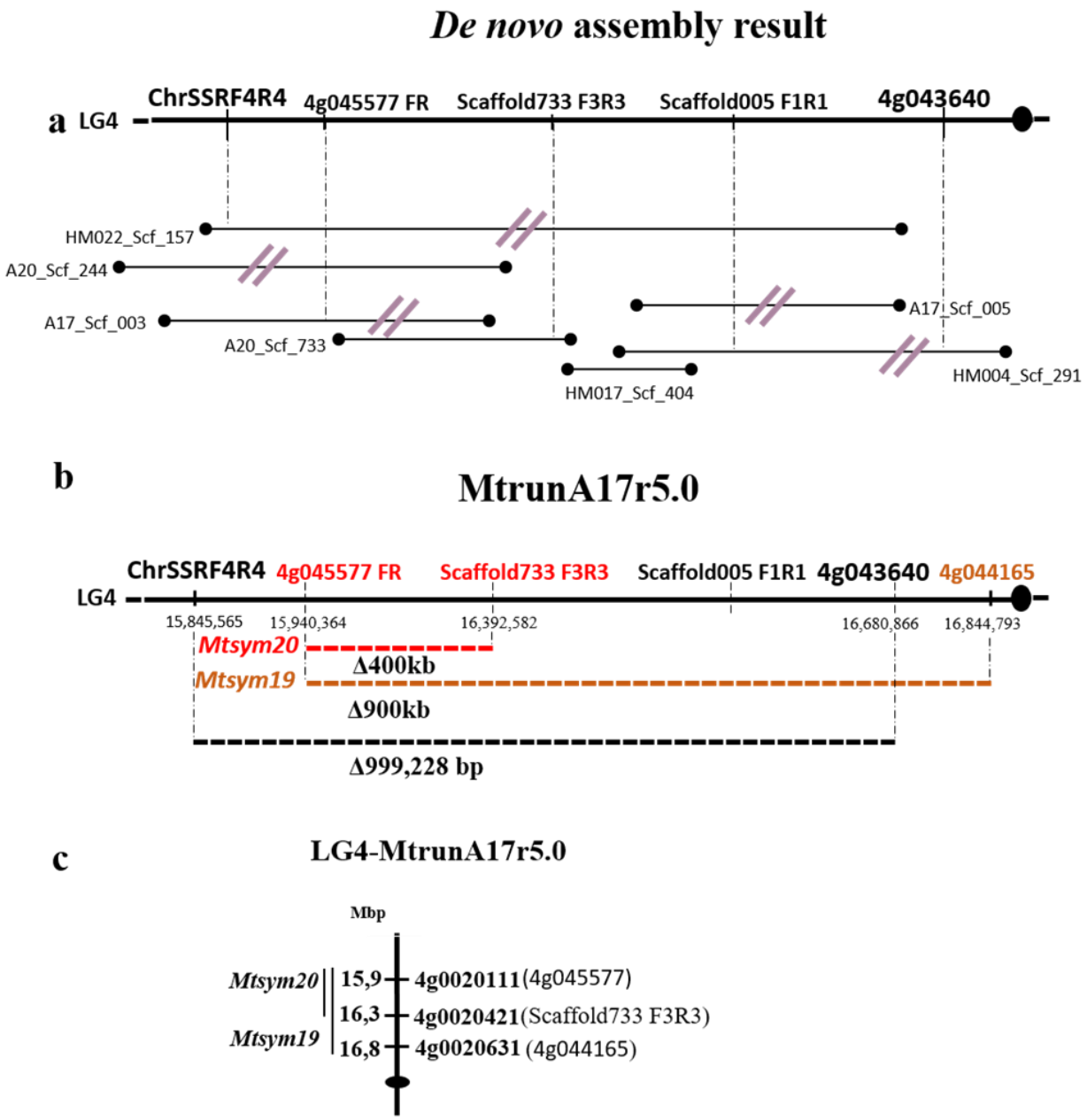


Fig.7.5.3 The genomic region of *Mtsym20* as identified by *De novo* assembly and based on the genome release MtrunA17r5.0

De novo assembly was conducted with scaffold sequences from various Medicago truncatula accessions. These scaffolds were mapped to the genomic region of Mtsym20, with novel genetic markers positioned on chromosome 4. Vertical lines are connecting the novel genetic markers to their corresponding positions on the scaffold bars. Gaps within the scaffold sequences are represented by purple dashes (a). The sequence between the Chr4SSR F4R4 and 4g043640 border markers based on latest revealed *M. truncatula* database (MtrunA17r5.0) confirmed the result of de novo assembly. The identified genomic region of Mtsym20 is spreading to 400 kb as indicated in red and genomic region of Mtsym19 is spreading 900 kb as shown in orange (b). MtrunA17r5.0 annotations of border genetic markers for the Mtsym20 and Mtsym19 genomic regions and their positions on chromosome 4 (c).

The systematic reconstruction of the genomic region was facilitated by assembling Scaffold244 within A20, Scaffold003 within A17, Scaffold733 within A20, Scaffold404 within HM017, Scaffold291 within HM004, and Scaffold005 within A17 cultivars, (Fig. 7.5.3a). The reconstructed region contained 81 additional genes along with the previously identified 12 genes on JCVI.4 database (Supplementary Table 5). After reconstructing the genomic region with *de novo* assembly, new molecular markers localized on the newly assembled sequences were used in genetic mapping to validate the accuracy of the assembly (Fig. 7.5.3a). These PCR-based markers were amplified using the genomic DNA of plants from the genomic region of *Mtsym20* and they could be integrated into the genetic map without any contradiction in genotypes or phenotypes validating the accuracy of our *de novo* assembly (Fig. 7.5.4)

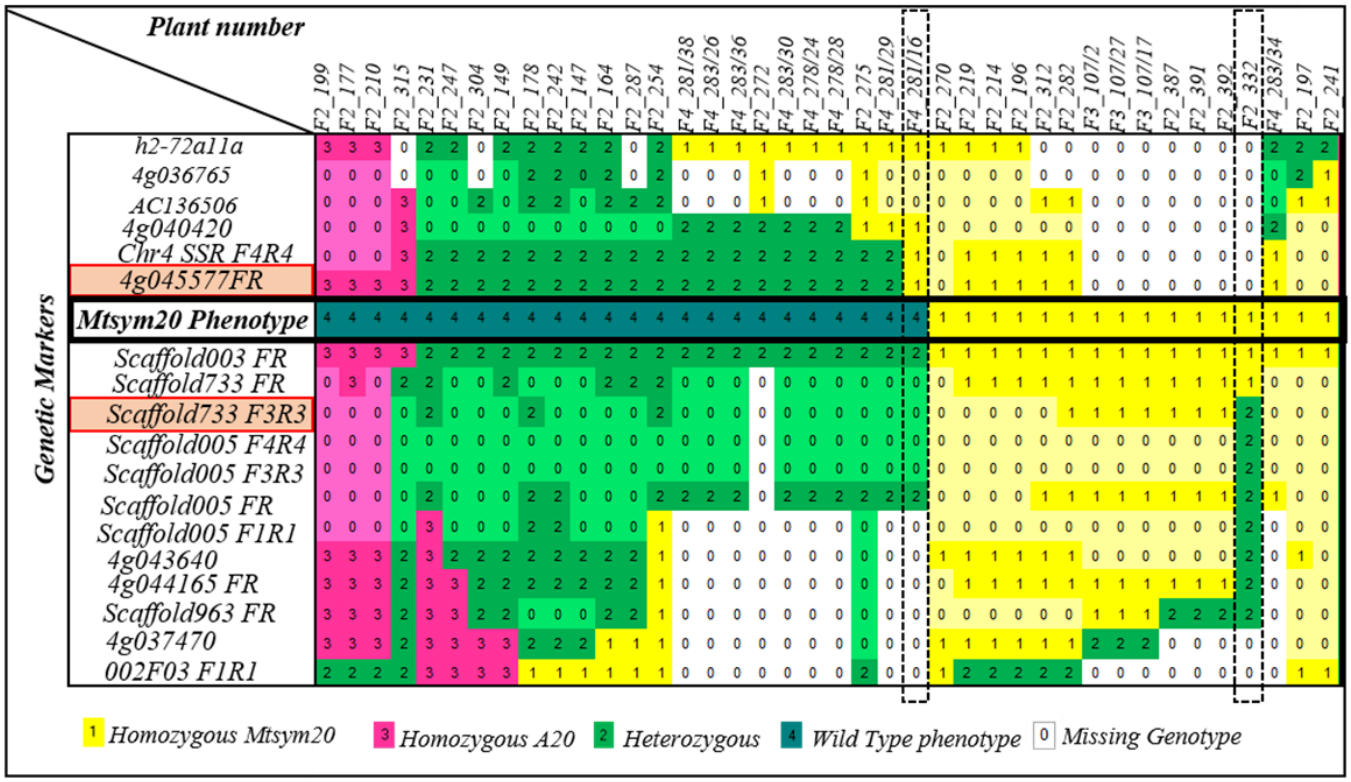


Fig. 7.5.4 Identification of a new *Mtsym20* genomic region using novel genetic markers

The integration of novel genetic markers into the color map confirmed the results of our *de novo* assembly, and the MtrunA17r5.0 database sequence, identifying an extended genomic region for *Mtsym20* between the 4g045577 and Scaffold733 F3R3 genetic markers. The border markers identifying the new *Mtsym20* genomic region are highlighted with orange background and in red rectangles. Black dashed lines surround the plants displaying the recombination sites, and the solid black line is used to highlight the phenotype row. The missing genotypes are highlighted with a different shade of the color used for identified genotypes when they could be predicted based on the neighbouring genotypes.

Based on the corrected sequence of the region, additional genetic markers were developed and introduced into color map and the borders of the *Mtsym20* genomic region was redefined. The upper

border was determined by the flanking marker 4g045577 based on the genotype and the phenotype of plant F4_281/16. This plant displayed a genotype transition from mutant to heterozygous and showed a wild-type symbiotic phenotype which narrowed down the genomic region. The mutant plant no. 332 from the extended F2 segregation population presented a genotype switch from heterozygous to mutant between markers Scaffold733 F3R3 and Scaffold733 FR. Therefore, the lower border of the region was identified at the Scaffold733 F3R3 molecular marker (Fig.7.5.4).

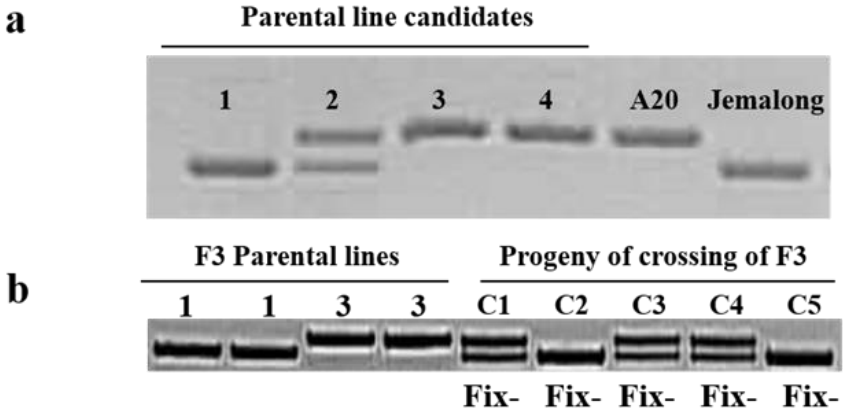
Meanwhile, the symbiotic locus of another mutant *Mtsym19*, which had been isolated in the same mutant screening program (Morandi et al., 2005), was found between genetic markers 4g045577 and 4g044165, in an overlapping genomic region of *Mtsym20*, as identified by our colleague Mónika Tünde Tóth. The updated version of *M. truncatula* database, the MtrunA17r5.0, was revealed at this stage of our genetic mapping. This new version of *M. truncatula* reference genome sequence incorporated new PacBio sequencing data and has substantially improved the genome assembly of *M. truncatula*. Thus, many genes previously defined to be on scaffolds have been integrated into specific chromosomal locations in the genome (Pecrix et al., 2018). Our sequence assembly of the genomic region was confirmed to be correct based on most recent *M. truncatula* database (MtrunA17r5.0). Based on the MtrunA17r5.0 database, the size of the genomic region of *Mtsym20* was identified as about 400 kb and the genomic region of *Mtsym1*9 as 900 kb whereas the region between the genetic markers Chr4SSR F4R4 and 4g043640, which was estimated to be approximately 90 kb on JCVI.4, is 999,228 bp (Fig. 7.5.3a,b). A list of the genes in the genomic regions of *Mtsym20 and Mtsym19* based on *M. truncatula* database (A17.v5) and based on our *de novo* assembled reconstruction of JCVI.4 database is presented together in Supplementary Table5.

The flanking border markers for the *Mtsym20* and *Mtsym19* mutants, 4g045577, Scaffold733 F3R3, and 4g044165, were designed referencing the JCVI (Mt.4.0) database and *de novo* assemblies of different *M. truncatula* accessions. These markers correspond to the genomic locations of genes annotated as 4g002011, 4g0020421, and 4g0020631 on the MtrunA17r5.0 genome database, respectively (Fig. 7.5.3c).

7.6. *Mtsym20* is allelic to *Mtsym19*

Mtsym20 and *Mtsym19* were previously reported to be non-allelic (Morandi et al., 2005). However, the comparative analyses of mutant phenotypes have revealed very similar patterns of colonization, as well as similarities in bacteroid and host cell morphology, raising questions about

genetic relationship of *Mtsym20* and *Mtsym19* (Fig. 7.2 and Fig. 7.3.1). Additionally, our mapping experiments revealed a significant overlap in the genomic positions of *Mtsym20* and Mtsym19 suggesting either defects in two closely situated genes or a single gene malfunction affecting both mutants (sub-chapter 7.5).



c

F3 Parental line candidtates	1	2	3	4	
	Mtsym20 X A20 F3-105	*Mtsym20* X A20 F3-81	*Mtsym19* x A20 F3-201	*Mtsym19* x A20 F3-124	
Progeny of crossings of F3	**C1**	**C2**	**C3**	**C4**	**C5**
	Mtsym20 x A20 F3_105/2 X *Mtsym19* x A20 F3_201/8	*Mtsym20* x A20 F3_105/2 X *Mtsym19* x A20 F3_201/5	*Mtsym20* x A20 F3_105/2 X *Mtsym19* x A20 F3_201/8	*Mtsym20* x A20 F3_105/7 X *Mtsym19* x A20 F3_201/2	*Mtsym20* x A20 F3_105/7 X *Mtsym19* x A20 F3_201/5
Number of Progeny	4	4	1	3	4
Phenotype	Fix-	Fix-	Fix-	Fix-	Fix-

Fig. 7.6. Allelism test between *Mtsym20* and *Mtsym19*

Based on the genotypes for the marker Mtb276, the progenies of the *Mtsym20*xA20-105 and *Mtsym19*xA20-201 with opposite homozygous genotypes were selected for allelism tests (a). The Fix- phenotype observed in the hybrids indicates allelism between *Mtsym20* and *Mtsym19* (b). PCR amplification with the Mtb276 marker is visualized on gel electrophoresis images (a,b). A summary of the F3 parental line candidates, progeny from five crosses (C1-C5) - with the details of the crosses-, number of progenies, and their symbiotic phenotypes 3 weeks post-inoculation with *S. medicae* WSM419 (c).

To clarify the allelic relationship between *Mtsym20* and *Mtsym19* mutants, F1 allelism tests were carried out using individuals from the mapping populations of *Mtsym19* and *Mtsym20*. F3 maternal plants were chosen from the genetic mapping population of *Mtsym20* (*Mtsym20* x A20 F3_105 and *Mtsym20* x A20 F3_81) and *Mtsym19* (*Mtsym19* x A20 F3_201 and *Mtsym19* x A20 F3_124) mutants

among the plants exhibiting mutant genotype. To assure that the progeny of crosses is not derived from self-mating, we genotyped them with a genetic marker outside of the genomic region of these mutants as both mutants are expected to have Jemalong background in the overlapping genomic region. The selected plants were paternal or maternal homozygous for the PCR-based genetic marker Mtb276 from chromosome 6 (LG 6) of *M. truncatula* (Mun et al., 2006). This polymorphism allowed us to verify the hybrids of the crosses between the selected *Mtsym20* and *Mtsym19* plants from the segregating populations (Fig. 7.6a). Progeny of five individual crosses were collected from matured pods and evaluated for their genotypes and symbiotic phenotypes 3 wpi with SmWSM419. Mtb267 marker was subsequently amplified from the DNA isolated from the progeny of the crosses. Based on the genotype analysis presented in Fig. 7.6b, three out of five crosses (C1, C3, and C4) exhibited a heterozygous genotype. However, two of the crosses (C2, C5) showed homozygous genotypes indicating self-pollination of the maternal plant (Fig. 7.6b). Phenotypic characterization of nodules from progenies of successful crosses showed that all hybrid genotypes showed Fix- phenotype similar to mutant phenotype of *Mtsym20* and *Mtsym19* (Fig. 7.6c). This result indicated that *Mtsym20* is allelic to *Mtsym19,* contrary to the result obtained previously (Morandi et al., 2005).

7.7. Identification of the deletion in the symbiotic loci of *Mtsym20*

Fine mapping experiments could restrict the genomic region of *Mtsym20* between genetic markers 4g0020111 (4g045577 on Mt.4.0) and 4g0020421 (scaffold733 F3R3 on Mt.4.0) spanning a ~400 kb region while the symbiotic region of *Mtsym19* was identified between 4g0020111 (4g045577 on Mt.4.0) and 4g0020631 (4g044165 on Mt.4.0) (Fig. 7.5.3b,c). We hypothesized that deletions induced by gamma irradiation in the genomes of *Mtsym20* and *Mtsym19* are responsible for the ineffective symbiotic phenotype of the mutants. Consequently, RNA sequencing (RNAseq) analysis of nodules was conducted and the transcript levels in the mutants were evaluated to identify genes that were either deleted or exhibited significantly reduced expression within the genomic regions of corresponding mutants. The comparative transcriptome analysis involved the *M. truncatula* mutants *Mtsym20*, *Mtsym19*, and *Mtsym18*. The *Mtsym18*, which is described as a mutant in a distinct symbiotic complementation group by Morandi et al. (2005), was used as a reference sequence enabling assessment of transcriptomic variations between *Mtsym* mutants. The qualified reads were aligned to the reference genome of *M. truncatula* (MtrunA17r5.0) and the number of reads present within the specific genomic region between the 4g0020111 and 4g0020631 markers was evaluated using the CLC Genomics Workbench 9.5.3.

Upon comprehensive examination of the RNAseq results, we found that three proximal genes (MtA17_Chr4g0020281, MtA17_Chr4g0020291 and MtA17_Chr4g0020301) had reduced read count levels in the genomic region of *Mtsym20* mutant when compared to the RNA sequence of the *Mtsym18* mutant used for reference. The comparison of read counts between *Mtsym19* and *Mtsym18* revealed that in *Mtsym19* mutant, two additional neighboring genes, MtA17_Chr4g0020311 and MtA17_Chr4g0020321, alongside the three genes identified with reduced number of reads in *Mtsym20*, also exhibited reduced expression. Notably, although MtA17_Chr4g0020281 gene displayed a noticeable quantity of reads both in *Mtsym20* and *Mtsym19* mutants at the first analysis, the detailed evaluations revealed that these reads were accumulated as a result of the nonspecific alignment of a short repeat sequence found within the sequence of the gene (Fig. 7.7a). The comparison of total RNA sequencing read counts suggested potential deletions in the genomes of *Mtsym2*0 and *Mtsym19*.

7.7.1. Localization of the deletion in the genomic region of *Mtsym20*

To investigate the extension of the deletions, PCR markers were designed based on the sequence of MtrunA17r5.0 genome database and the genomic regions in the *Mtsym20* and *Mtsym19* mutants were surveyed for the presence of the genes or other sequences (Supplementary Table 2). Expected fragment sizes were obtained with all the designed primers following PCR amplification on *M. truncatula* J5 genomic DNA, while no fragments could be amplified for genes Chr4g0020281, Chr4g0020291, Chr4g0020301 from *Mtsym20* indicating the deletion of the corresponding genes in the mutant genome (Fig. 7.7c). In case of *Mtsym19*, amplification of fragments failed not only for Chr4g0020281, Chr4g0020291, Chr4g0020301 genes but also for Chr4g0020311 and Chr4g0020321 confirming the deletion of all the five genes in *Mtsym19* genome which corresponded to the transcriptome sequencing data (Fig. 7.7c).

Following the verification of the deletions, additional primers were designed to identify the extension of the deletions in the genome of the corresponding mutants (Supplementary Table 2). However, it was sometimes a challenge to design specific markers due to the presence of repetitive sequences in the genomic regions which made it difficult to identify the deletion boundaries accurately. A repeat region spanning approximately 10 kb was found downstream of the MtrunA17Chr4g0020281 gene sequence. Therefore, unique sequences preceding this large repeat region were targeted for the development of genetic markers to identify approximate size of the deletions. Consequently, one of the markers was placed on a sequence that lacked annotation in the

MtrunA17r5.0 database but was identified as Medtr003te0501 in the JCVI (Mt4.0) database. The Medtr003te0501 marker defined the start of the deletion both in the genome of *Mtsym20* and *Mtsym19* mutants. In *Mtsym20*, the deletion was predicted to extend to the Chr4g0020311 gene, confirming the deletion between Medtr003te0501 and Chr4g0020311 markers, while in *Mtsym19*, the deletion was anticipated to span between Medtr003te0501 and Chr4g0020331 markers (Fig. 7.7b). However, attempts to amplify genomic DNA samples from *Mtsym20* or *Mtsym19*, using the corresponding primer pairs, were unsuccessful, likely due to the extensive repeat sequence situated between the markers. Despite this, primer pairs targeting specific sequences within the Medtr003te0501 and Chr4g0020311 genes for *Mtsym20*, and within the Medtr003te0501 and Chr4g0020331 genes for *Mtsym19*, successfully amplified the expected fragments from the genomic DNA of each mutant, confirming the presence of these gene sequences. Although the exact size of the deletions could not be determined, the successful amplification of fragments at the first possible sequence bordering the deleted regions indicates that the deletion in the genome of *Mtsym19* spans approximately a 45 kb, located between the markers Medtr0003te0510 and Chr4g0020331. Similarly, the deletion in the genome of *Mtsym20* is estimated to be approximately 30 kb, spanning between Medtr0003te0510 and Chr4g0020311 markers. (Fig. 7.7b,c). The identified deleted regions in *Mtsym19* and *Mtsym20* genomes had a continuous sequence and did not include any gaps based on MtrunA17r5.0 genome database. A list of the deleted genes and their predicted gene functions are presented in Table 7.7, along with their positions in the A17 the genome (release MtrunA17r5.0) (Pecrix et al., 2018).

The overlapping deletions in the genomes of *Mtsym20* and *Mtsym19* mutants support the result of the allelism test we previously conducted, confirming that *Mtsym20* and *Mtsym19* are allelic. Thus, the nitrogen-fixing deficiency observed in these mutants was likely caused by the deletion of one of the three shared genes found in the overlapping deletion regions between the *Mtsym20* and *Mtsym19* genomes. This region includes two hypothetical genes (Chr4g0020281 and Chr4g0020301) and one putative late nodulin gene (Chr4g0020291): the *Nodule-specific Cysteine-Rich new35* (*NCR-new35*) based on MtrunA17r5.0 genome database (Pecrix et al., 2018). It is important to note that, *NCR-new35* was annotated as *NCR014* in a previous study (De Bang et al., 2017) before the release of MtrunA17r5.0 and thus most of the literature related to *NCR-new35* was available with the name of *NCR014*.

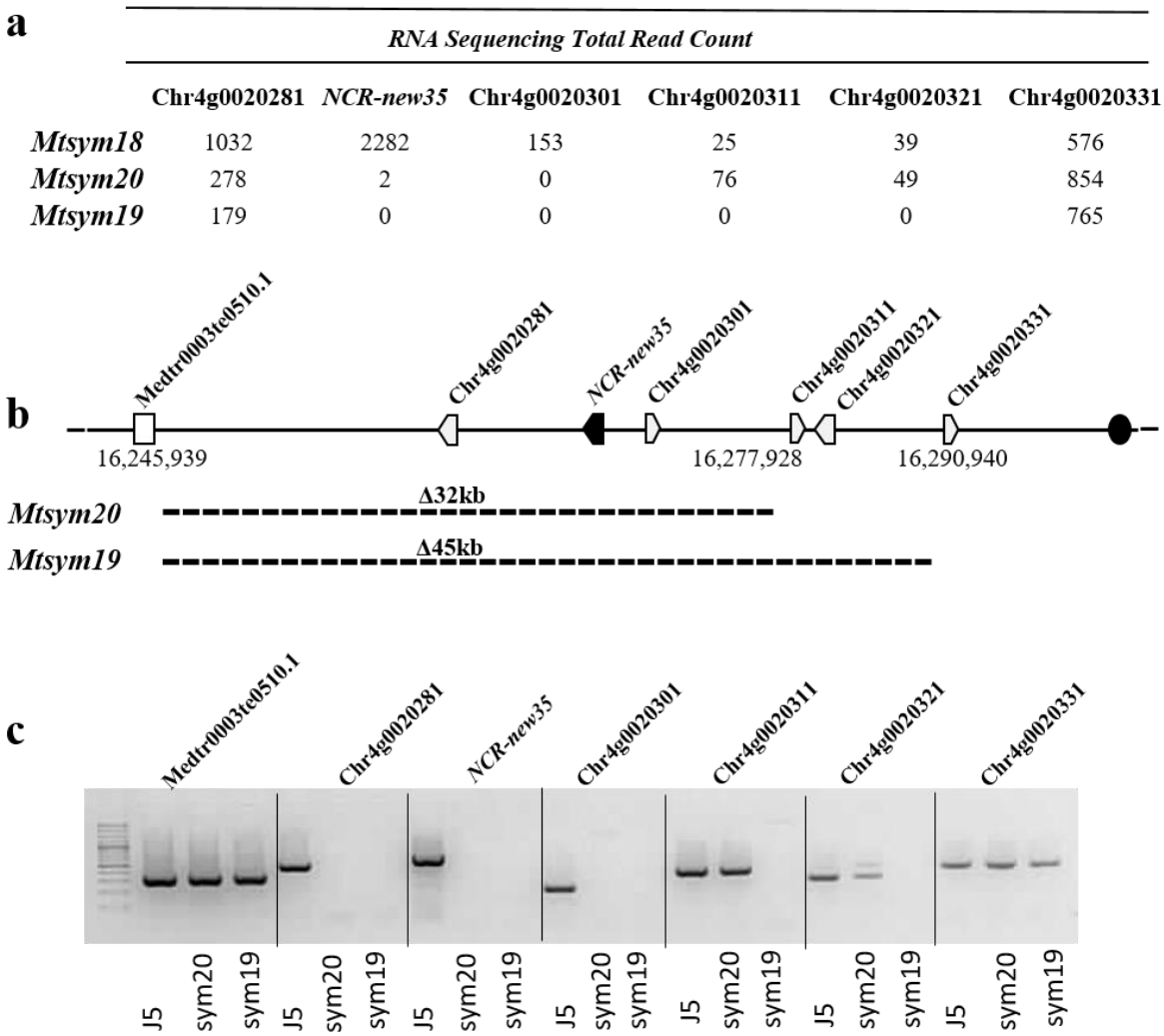


Figure 7.7 Identification of genomic deletions in *Mtsym20* and *Mtsym19*

The RNA sequencing total read count of genes in the deletion regions of *Mtsym20* and *Mtsym19* showing reduced transcription compared to the reference *Mtsym18* (a). The deleted regions on linkage group 4 for both *Mtsym20* and *Mtsym19* are aligned under the solid line representing the genomic region and the estimated extents of the deletions represented by dashed lines below (b). The gel electrophoresis image validating the absence of specific gene fragments, confirming the predicted deletions. Genetic markers used for PCR amplifications are labeled above each gel lane, and the DNA samples are identified below the gel image (c). J5 = wild type *M. truncatula* J5 line, sym20 = *Mtsym20*; sym19 = *Mtsym19*

Table 7.7. Gene content of the deletion in the symbiotic region of the *Mtsym20* and *Mtsym19*

Gene ID	Gene Position	Predicted Gene Function	Mutant Line
MtrunA17Chr4g0020281	16256649..16262089	Putative protein	*Mtsym20 - Mtsym19*
MtrunA17Chr4g0020291	16267079..16268349	Nodule-specific Cysteine-Rich peptide	*Mtsym20 - Mtsym19*
MtrunA17Chr4g0020301	16268538..16269391	Putative protein	*Mtsym20 - Mtsym19*
MtrunA17Chr4g0020311	16277110..16278540	Putative long-chain-alcohol O-fatty-acyltransferase	*Mtsym19*
MtrunA17Chr4g0020321	16279358..16287505	Putative GIY-YIG endonuclease	*Mtsym19*

7.8. NCR-new35 restores the effective symbiosis in *Mtsym20*

To identify the gene responsible for the mutation in *Mtsym20*, genetic complementation experiments were carried out. The candidate genes, MtrunA17Chr4g0020281, MtrunA17Chr4g0020291, MtrunA17Chr4g0020301 within the deleted region of *Mtsym20* were evaluated for their capacity to complement the mutant phenotype of *Mtsym20* based on their expression levels in nodules and their predicted functions. MtrunA17Chr4g0020281 and MtrunA17Chr4g0020301 were predicted to encode hypothetical putative peptides whereas MtrunA17Chr4g0020291 was predicted to code for Nodule-specific Cysteine-Rich (*NCR)* peptide based on MtrunA17r5.0 database (Table 7.7). NCR peptides have been shown to be targeted to the symbiosome in previous studies, which demonstrated that deficiencies in certain *NCR* members, specifically *NCR211* and *NCR169*, can lead to ineffective nitrogen-fixing symbiosis (Horváth et al., 2015; Kim et al., 2015). Furthermore, examination of gene expressions from transcriptome analysis available in the 'Symbimics Gene Atlas' database (https://iant.toulouse.inra.fr/symbimics/; Roux et al., 2014) revealed that the hypothetical genes are hardly expressed in nodules whereas *NCR-new35* shows pronounced nodule-specific expression (Fig. 7.8.1). Comprehensive analyses of both the expression level and predicted function of *NCR-new35* made this gene as the primary candidate responsible for the symbiotic phenotype observed in *Mtsym20*. To confirm gene identity, *NCR-new35* was selected for the subsequent complementation experiments to restore the symbiosis in *Mtsym20*.

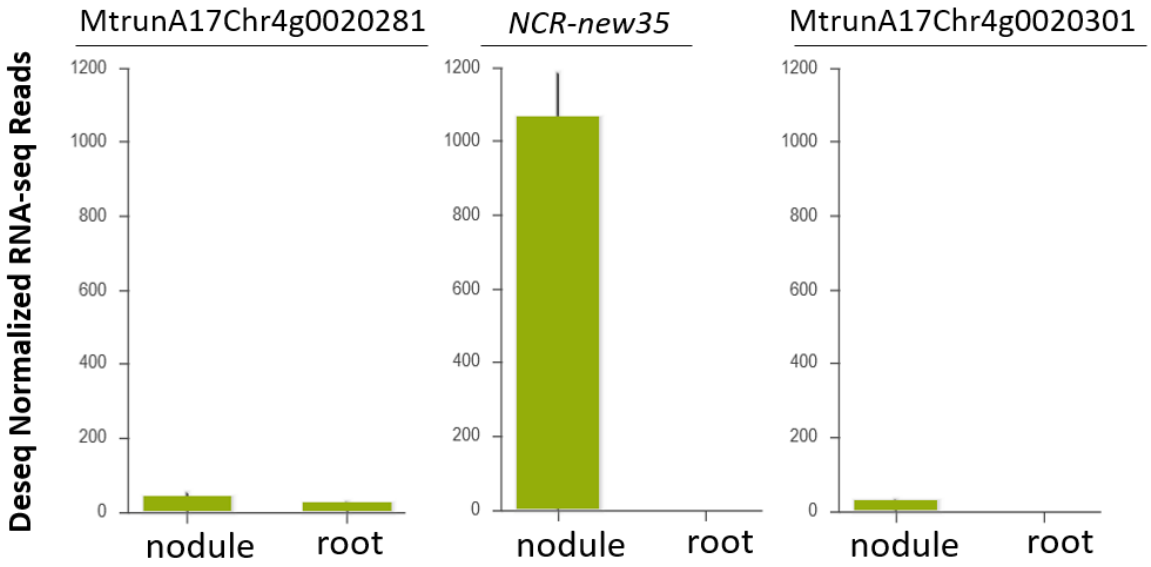


Fig. 7.8.1 *NCR-new35* has high nodule specific expression

Bar graphs indicating deseq normalized RNA sequencing reads obtained from whole organ analysis, demonstrating expression levels of genes MtrunA17Chr4g0020281, *NCR-new35* and MtrunA17Chr4g0020301. Data are obtained from the Symbimics Gene Atlas database (https://iant.toulouse.inra.fr/symbimics/; Roux et al., 2014).

To confirm that the loss of *NCR-new35* gene causes the ineffective symbiotic phenotype of *Mtsym20*, complementation experiments were carried out utilizing the *A. rhizogenes*-mediated hairy root transformation (Boisson-Dernier et al., 2001). In a previous study, it was determined that at least a 400 bp promoter length is essential for driving the expression of *NCR* genes in nodules, but it may not be sufficient for all *NCR* genes (Nallu et al., 2013). In order to assess necessary promoter length for the function of *NCR-new35* gene, we amplified two distinct sequences to use for complementation experiments, one of them with a shorter 516 bp promoter and another with a longer 819 bp promoter. Both of these sequences were fused with the *NCR-new35* gene sequence and a 903 bp 3'UTR sequence originated from the genomic region downstream of *NCR-new35*. Constructs were cloned into pKGWR vector carrying the dsRed transformation marker and then introduced into *Mtsym20* mutant roots with *A. rhizogenes*-mediated hairy root transformation. To ensure optimal symbiotic conditions during the complementation experiment, pKGWR-dsRed empty vector constructs were also introduced into the roots of mutant and wild-type plants. The composite plants, possessing wild-type shoots and transgenic roots, were inoculated with *S. medicae* WSM419 (pXLGD4) and phenotypic assessment was conducted 4 wpi. The red fluorescent signal of dsRed marker was used to identify transformed roots and nodules (Fig. 7.8.2i,j,k,l). Wild-type control plants transformed with the empty vector exhibited typical wild-type characteristics, including well-developed shoots and dark green leaves along with elongated pink nodules (Fig. 7.8.2a,e). In contrast, mutant plants transformed with the empty vector retained their mutant phenotype, characterized by stunted shoots, small yellowish leaves, and slightly elongated white nodules (Fig. 7.8.2d,h). The nitrogen starvation phenotype of the mutant was rescued in *Mtsym20* plants transformed either with 516pr::*NCR-new35*::3'UTR or 819pr::*NCR-new35*::3'UTR constructs. The complemented phenotype of *Mtsym20* mutant resembled the symbiotic phenotype of wild-type plants, characterized by a healthy shoot system and elongated pink nodules (Fig. 7.8.2b,c,f,g). Nodule sections were subsequently stained for β-galactosidase activity and analyzed under the light microscopy. Longitudinal nodule sections prepared from empty vector-transformed wild-type roots displayed typical indeterminate nodule zonation and β-galactosidase activity was observed across all zones (Fig. 7.8.2m). The sections prepared from nodules developed on mutant roots transformed with the empty vector showed β-galactosidase activity in the infection and interzones, while the nitrogen fixation zone was devoid of rhizobia (Fig. 7.8.2p). Histochemical staining of the nodule sections, collected from mutant roots transformed with 516pr:*NCR-new035*:::3'UTR or 819pr::*NCR-new035*::3'UTR complementation constructs, displayed β-

galactosidase activity throughout all nodule zones (except for ZI, as appropriate) similar to zones of wild-type Jemalong nodules (Fig. 7.8.2n,o). The results of complementation experiments demonstrated that the introduction of the *NCR-new35* gene into the genome of the *Mtsym20* mutant rescued the mutant phenotype of *Mtsym20* plants. This finding indicates that the *NCR-new35* gene plays a crucial role in the symbiosis between *M. truncatula* and *S. medicae*.

According to transcriptomic and genome sequencing studies, the putative late nodulin family of *Nodule-Specific Cysteine-Rich* (*NCR*) genes encompasses over 700 genes in *M. truncatula* genome. Although many of these genes have not been yet characterized or experimentally confirmed to function as an active NCR, they are classified in the same gene family based on sequence similarities, computational predictions and evolutionary conservation.

In this study, ineffective mutants *Mtdnf4-1*, *Mtdnf7-2*, and NF-FN9363, each of them deficient for *NCR169*, *NCR211*, and *NCR343*, respectively, were used for the comparative phenotypic examination of *Mtsym20*. Our complementation experiments revealed one more essential *NCR* for effective nitrogen fixation in *M. truncatula*, the *NCR-new35*. On the other hand, some *NCR*s were found to be non-essential, as their absence in mutant genotypes did not hinder the rescue of mutant phenotypes with another *NCR* and their introduction into the genome could not restore the nitrogen-fixing symbiosis. The genome of *Mtdnf4-1* mutant is not only deficient in *NCR211* but also *NCR178* (Kim et al., 2015). Similarly, in NF-FN9363, along with the *NCR343* gene, *NCR341*, *NCR344*, and *NCR345* are also deleted in the symbiotic region (Horváth et al., 2023). Complementation experiments with these *NCR*s failed to restore the mutant phenotypes when *NCR178* was introduced into *Mtdnf4-1* or *NCR341*, *NCR344*, *NCR345* were introduced into NF-FN9363, indicating their non-essential role in nitrogen-fixing symbiosis with the studied rhizobia strains (Kim et al., 2015; Horváth et al., 2023).

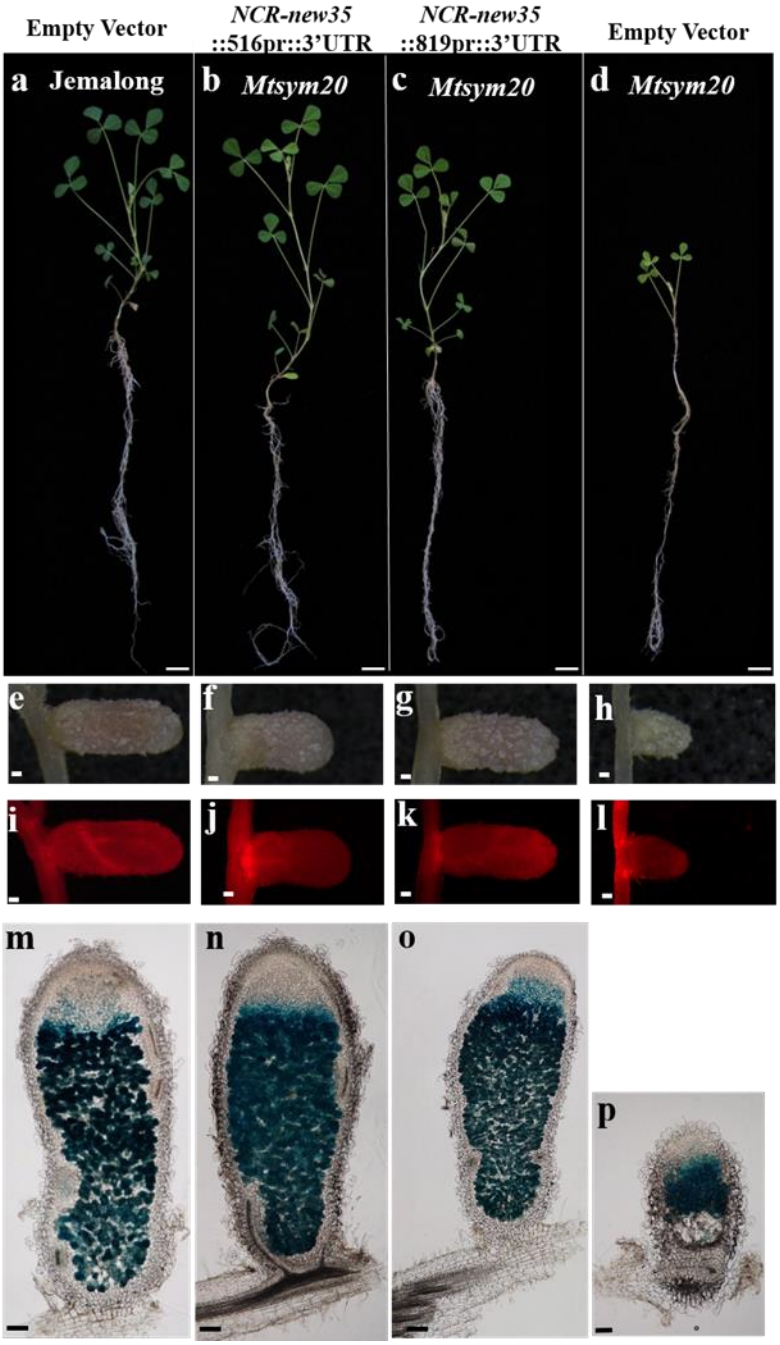


Fig. 7.8.2 *NCR-new35* is responsible for the nitrogen-fixing deficient phenotype of *Mtsym20*

Introduction of *NCR-new35* constructs with the different promoter lengths, 516 bp and 819 bp, rescued the mutant phenotype of the *Mtsym20* mutant 4 weeks post-inoculation with *Sinorhizobium medicae* WSM419. The rescued mutants displayed healthy shoot growth, formed elongated pink nodules, and β-Galactosidase activity was detected in all symbiotic zones of the nodules (b, c, f, g, n, o), similar to wild-type Jemalong plants (a, e, m). In contrast, *Mtsym20* plants introduced with the empty vector exhibited nitrogen starving shoots and leaves and nodules with reduced size, and β-galactosidase activity was absent in the basal part of these nodules (d, h, p). The transformation marker dsRed signal of the nodules was observed under the red fluorescent light (i-l). Scale Bars: (a-d) 2 cm; (e-p) 200 μm

7.9. Comparative genetic analysis of *NCR-new35*

The result of our complementation experiment led to the questions: Why does the *M. truncatula* genome contain so many *NCR* genes? And why are some essential for symbiosis in *M. truncatula* while others are not? Analysis of NCR peptides shows evidence of conservation of certain functions as well as the evolution of new ones, distinguishing them from defensin-like peptides (DEFLs) that typically preserve the gene functions. In the *M. truncatula* genome, *NCR* genes are often observed in small clusters next to each other in the genome, showing a higher percentage of similarity in peptide sequence compared to other NCRs in the genome. This evidence suggests local duplication or multiplication of the genes providing opportunity to develop new subfunctions through evolution (Stage et al., 2016). Beside this, in *M. truncatula* genome, many *NCR* genes are located close to transposons, suggesting the possibility of transposon-mediated long-distance duplications (Young et al., 2011). Consequently, some *NCR*s might be essential for symbiosis through preserved functions or evolving new functions, while some others do not have a critical role, perhaps due to the loss of function or duplications with minor mutations which does not result in a functional gene. However, nonfunctional genes may be activated under certain conditions or can compensate for each other in a collaborative function. To reveal the evolutionary role of NCR peptides in *M. truncatula*, further research would be beneficial.

In our study, to better understand the distinct characteristics of essential *NCR*s and attempt to reveal common features of crucial NCR peptides, we carried out a comparative analysis of *NCR-new35 with NCR169, NCR211, NCR343,* genes and the encoded peptides.

The gene of *NCR-new35* contains two exons; the length of first exon is 295 bp and the second one is 884 bp long, while the intron located in-between is 92 bp long (Fig. 7.9a). When compared to the other essential *NCRs*, the variability in the number and length of exons (*NCR343*: 1 exon, *NCR211*: 3 exons, *NCR169*: 3 exons, *NCR247*: 2 exons) suggests no clear correlation with gene requirement for symbiosis. Similar to other crucial *NCRs* (*NCR343*, *NCR211, NCR169* and *NCR247*), *NCR-new35* also codes for a signal peptide including a highly conserved N-terminal sequence recognized by the secretion system of nodule cells. This signal provides the delivery of mature peptides to the symbiosome by the secretory pathway to initiate bacteroid differentiation (Wang et al., 2010).

The signal peptide of NCRs within the gene family is highly conserved. The amino acid composition of mature peptides varies among NCRs whereas their most characteristic feature is to contain four or six cysteine residues that are regularly spaced (Alunni et al., 2007).

All known essential NCRs, including NCR-new35, contain four cysteine residues at conserved positions (five amino acids located between the first two conserved cysteines and 4 amino acids between the last two conserved cysteines) in the mature peptide (Fig. 7.9a b,c). NCR247 which is recently proved to be essential is the only exception with containing 5 amino acid between the third and fourth cysteine residues in its mature peptide. The lengths of the mature peptides vary between 24 and 47 residues among essential NCRs, and NCR-new35 has the longest mature peptide. Amino acid compositions of mature peptides significantly differ across NCRs suggesting no clear pattern on the functional impacts. However, the presence of a valine, an isoleucine and one or two aspartic acid residues between the first two cysteines is prevalent (Fig. 7.9a,c). The positioning of the cysteine residues is specifically important for the structural stability of the peptide due to their ability to form covalent linkages between the sulfur atoms of two cysteine amino acids, known as disulfide bonds. The effect of disulfide bonds on stabilization of the three-dimensional structure of NCR044 and NCR169 peptides has been recently demonstrated (Velivelli et al., 2020; Isozumi et al., 2021). The *M. truncatula* Thioredoxin s1 protein, known for breaking disulfide bonds, is also targeted the symbiosomes and proposed to influence the function of NCR peptides through adjusting their oxidation and reduction state (Ribeiro et al., 2017). Additionally, previous studies have also shown functional impairment in NCR169 when either cysteine residue has been substituted with a serine residue, indicating the importance of the positions of cysteine residues (Horváth et al., 2015). However, further investigations are required to understand the evolutionary advantage of the conservation of cysteine residues at specific locations in NCR peptides.

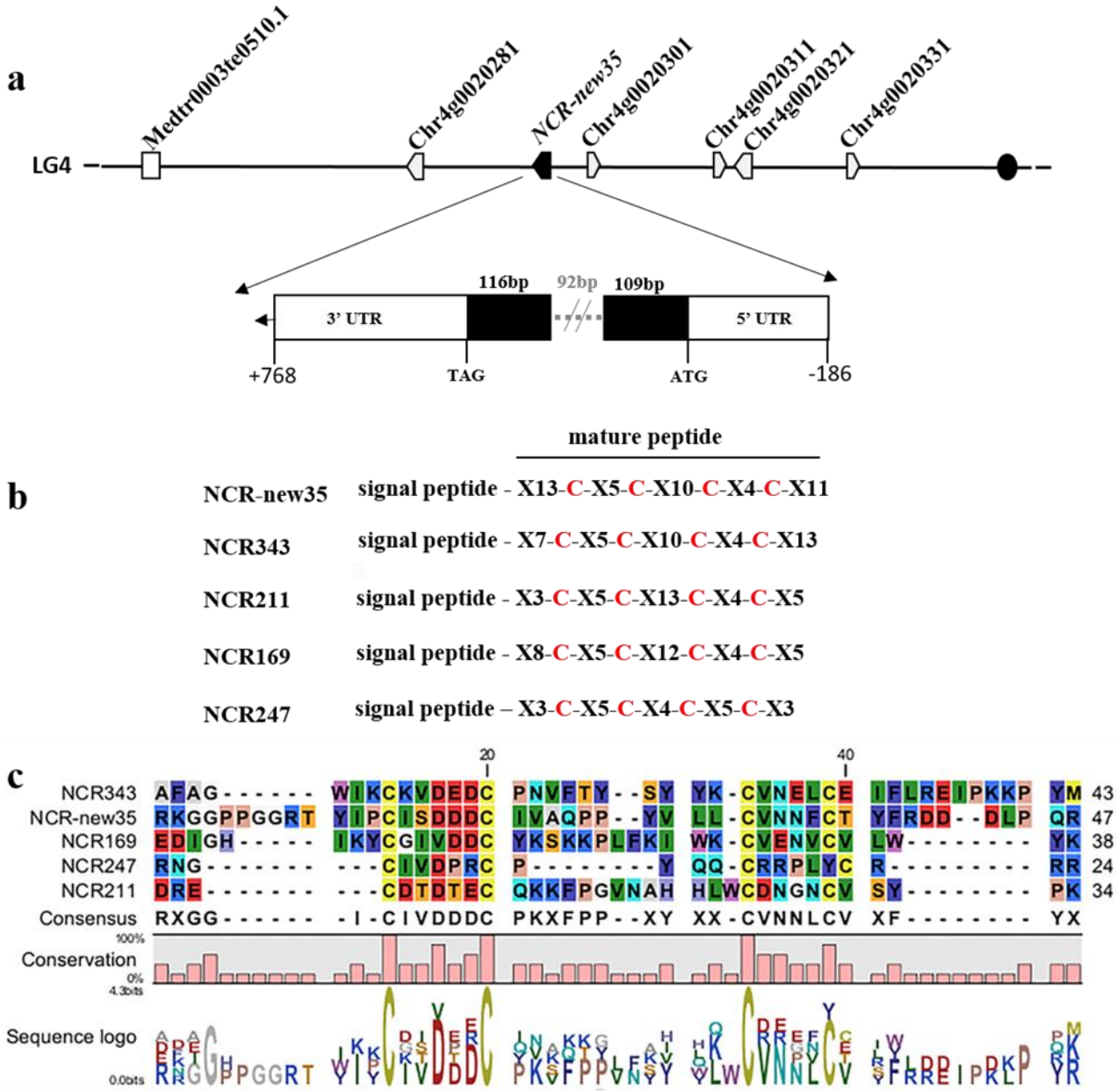


Fig. 7.9 Gene structure of *NCR-new35* and sequence alignment of NCR-new35, NCR343, NCR211, NCR169 and NCR247 peptides

In the gene structure of *NCR-new35*, exons are shown with black boxes and untranslated regions (UTRs) indicated by white boxes. The positions relative to start (ATG) and stop (TAG) codons are indicated with "-","+" symbols respectively. The non-coding region is shown in grey with dashed lines (a). The mature peptide of NCR-new35, NCR343, NCR211, NCR169 and NCR247 represented following their respective signal peptides and conserved cysteine residues are highlighted in red while variable amino acids denoted by X with the number indicating the count of residues between cysteines (b). Multitiple alignment of NCR343, NCR-new35, NCR169, NCR247 and NCR211 peptides with conservation bar below the alignment showing amino acid frequency while size of the letters in sequence logo are proportional to their frequency at that position (c).

NCRs are derived from defensin-like plant peptides that are formed following proteolytic cleavage of the signal sequence (Tavormina et al., 2015). NCR peptides are small (20–50 amino acids long) molecules and defensins range in size from 45-54 amino acids. Essential NCR peptides, unlike from their evolutionary precursors, the plant defensins which typically have eight cysteine residues involved in four disulfide bridges, contain four cysteines which potentially form two disulfide bonds (de Oliveira Carvalho and Gomes, 2009b; Farkas et al., 2018).

The isolectronic points (pI) of known essential NCRs differ in a wide range between 4.78 and 10.15 (*NCR169* pI=8.45, *NCR211* pI=5.38, *NCR343* pI=6.34, *NCR247* pI=10.15 and *NCR-new35*=4.78). This variation in the charge of mature peptides suggests that neutral, anionic or cationic NCR peptides can play a crucial role in nitrogen-fixing symbiosis in *M. truncatula.*

The antimicrobial role of certain cationic NCRs, such as NCR247 and NCR335, demonstrated *in vitro*, supports the suggestion that NCRs may have defensin-like functions (Van de Velde et al., 2010; Farkas et al., 2018). However, *in vitro* studies often use peptide amounts much higher than what plants produce *in vivo*. Moreover, NCRs detected within bacteroids are mostly anionic or neutral (Durgo et al., 2015). A possible hypothesis is that negatively charged (anionic) or neutral NCRs might bind to positively charged (cationic) NCRs to compensate for the toxic effects of their antimicrobial activity (Mergaert, 2018). The varying isoelectric points of essential NCRs also suggest that the function of NCRs may be more diverse *in vivo* than the antimicrobial defensins.

Supporting this hypothesis, a recent study revealed that *NCR247* has a specialized iron uptake function due to the unique features of its sequence, enabling it to bind and sequester heme effectively. However, NCR211, NCR169, and NCR035 did not show the same performance when tested for heme binding using UV-Vis spectrometry (Sankari et al., 2022). These results suggest that NCRs may have distinct or unique functions that potentially diverge from the common features of identified essential NCRs.

7.9.1. *NCR-new35* is expressed low in symbiotic cells compared with *NCR169, NCR211* and *NCR343*

The vast majority of *NCR* genes in *M. truncatula* are exclusively expressed in symbiotic nodule cells infected with rhizobia. These genes undergo activation during the differentiation process of host cells (Guefrachi et al., 2014) Our comparative examination of bacteroid and host cell morphology using confocal and scanning electron microscopy (SEM), suggested that in the absence of essential *NCR*s, initial differentiation patterns can be observed but bacteroids do not persist in basal

part of the nodules, in all tested mutants, including *Mtdnf4-1*, *Mtdnf7-2*, NF-FN9363, and *Mtsym20* (sub-chapter 7.3). Similarly, size and DNA content measurements of bacteroids isolated from mutants lacking essential *NCR*s suggested an impaired differentiation process, highlighting the critical roles of *NCR-new35*, *NCR343*, *NCR211*, and *NCR169* in the complete bacteroid differentiation process (sub-chapter 7.4).

In this study, we utilized the *GUS* promoter reporter system to visualize the expression patterns of *NCR-new35* and undertake a comparative analysis with other identified crucial *NCR*s, *NCR211*, *NCR343* and *NCR169* (Horváth et al., 2015, 2023; Kim et al., 2015). To analyze promoter activity, the promoters of these *NCR* genes were fused with the β-glucuronidase (*GUS*) reporter gene and the constructs were introduced into wild-type *M. truncatula* roots using the *A. rhizogenes*-mediated hairy root transformation. Histochemical staining of nodules sections was carried out by our colleague Beatrix Horváth, 2 and 3 wpi with *S. medicae* WSM419 strain to analyze GUS activity. The promoters of *NCR343* and *NCR211* displayed activity in the interzone of two-weeks-old nodules but exhibit sporadic staining in nitrogen fixation zone (Fig. 7.9.1d,e). However, *GUS* expression of the *NCR-new35* promoter showed a very mild level of activity, mainly limited to the initial cell layers of the interzone. Additionally, sporadic and even weaker activity was noticed in the distant part of the nitrogen fixation zone of these two-weeks old nodules (Fig. 7.9.1c). In contrast *NCR169* promoter displayed pronounced activity not only in interzone but also in nitrogen fixation zone of two weeks old nodules (Fig. 7.9.1f)). Evaluating our results in comparison to the nodule transcriptome data from different zones of *M. truncatula* nodules obtained from LCM (Roux et al., 2014), we found a consistent pattern (Fig. 7.9.1a). Our observations confirm that *NCR-new35*, *NCR343*, *NCR211*, and *NCR169* are predominantly active in the interzone of two-weeks-old nodules and *NCR-new35* has a significantly lower expression level compared to the *NCR343*, *NCR211*, and *NCR169* genes.

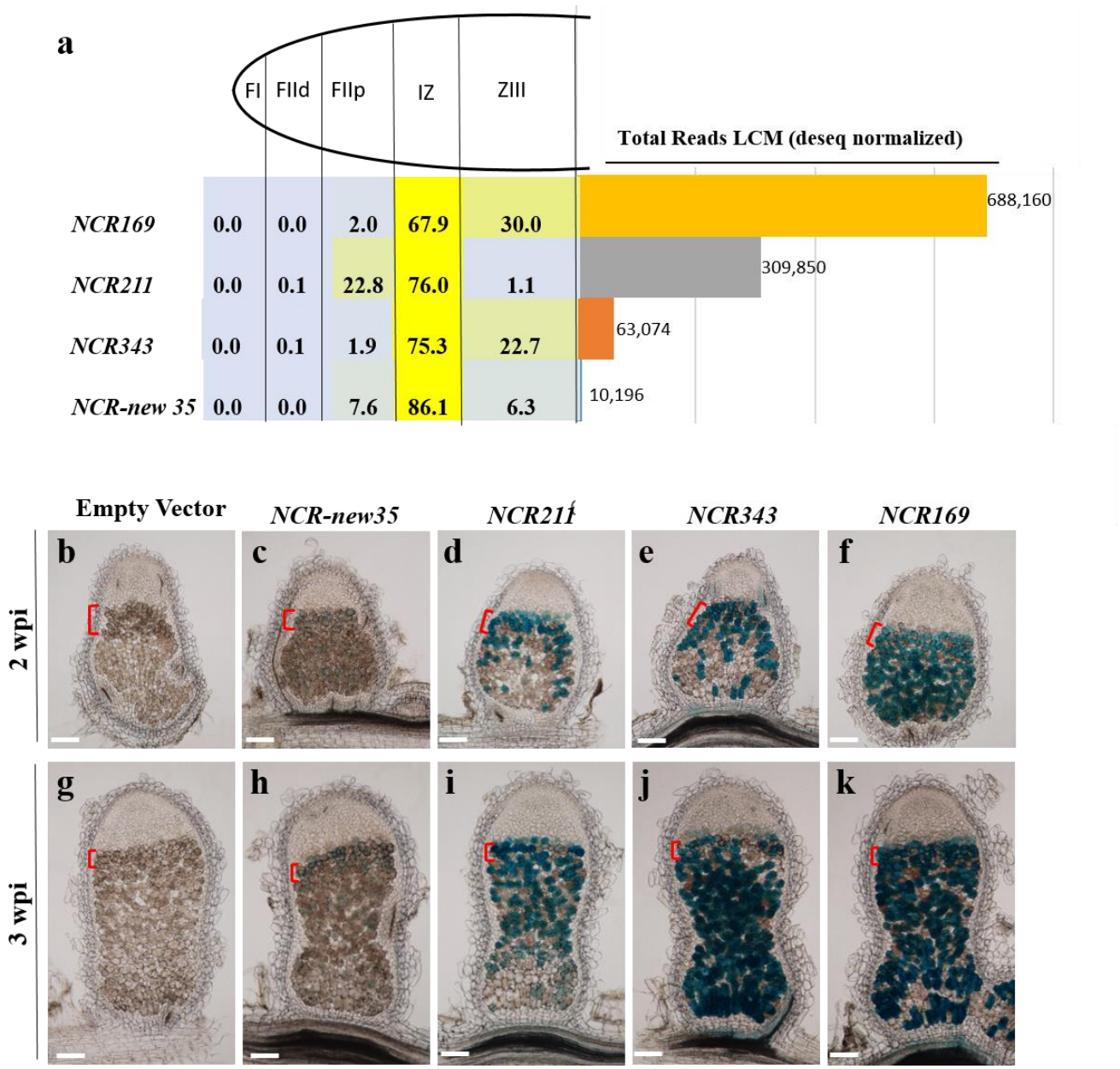


Fig. 7.9.1 Comparative promotor activity assays of *NCR-new35, NCR343, NCR211*, and *NCR169*

Deseq normalized RNA sequence read counts, which are proportional to the expression in different nodule zones, were obtained from laser-capture microdissected nodule zones (Roux et al., 2014). Each column in the diagram represents the percentage of RNA-seq reads in each respective nodule zone, with the total read count for each NCR gene displayed alongside its corresponding row (a). Longitudinal nodule sections from wild-type Jemalong plants introduced with GUS reporter constructs were evaluated using histochemical staining at 2 weeks post-inoculation (wpi) (b-f) and 3 wpi (g-k) with Sinorhizobium medicae (Horvath et al. 2023). Zone specific promoter activity was detected on the nodule sections of plants introduced with the GUS reporter gene fused with promoter sequence of NCR-new35, NCR343, NCR211 or NCR169, respectively (c-f and h-k) while no GUS activity was detected in the nodule sections of plants introduced with the empty vector (b,g). ZII-ZIII is labelled with red in the presented nodule sections (b-k). Scale bars: 200 µm

When the promoter activities of the nodules were examined 3 wpi with *S. medicae* using the *GUS* reporter gene, the promoters of *NCR169* and *NCR343* showed dominant activity both in the interzone and in the nitrogen fixation zone (Fig. 7.9.1j,k). These results suggest that *NCR169*, and *NCR343* may have a key role for the transition of rhizobia from their undifferentiated form to differentiated nitrogen-fixing bacteroids and maintaining this state. While the promoter activity of *NCR211* was also observed in the nitrogen fixation zone, the extent of cells showing activity was less compared to *NCR169* and *NCR343* (Fig. 7.9.1i,j,k). This result may indicate a short-time role of *NCR211* in fine-tuning the differentiation process. However, to confirm that the observed expression patterns emerge indeed from the different functions of *NCR*s, it would be worth to fuse the promoter of *NCR211* with the gene sequence of *NCR169* or *NCR343* and analyze the complementation phenotype of the corresponding mutants *Mtdnf7-2* and NF-FN9363. If the heterologous promoter construct fails to rescue the mutant symbiotic phenotype, we could conclude that *NCR211* indeed has a more specialized function and its limited expression would not be sufficient to rescue the lack of function of *NCR169* or *NCR343*.

The promoter activity of *NCR-new35* was found to be the weakest, being hardly detectable in both the interzone and nitrogen fixation zone of 3 weeks old nodules (Fig. 7.9.1h). The barely detectable expression of *NCR-new35* may indicate that very low levels of expression of the gene might be sufficient to fulfill its essential role in symbiosis. Alternatively, *NCR-new35* may have a specialized function and NCR-new35 might be modulating the activity of other gene products or signaling between the host plant and the bacteroids, as such regulatory roles might not require high levels of expression to be effective.

The expression of *NCR* genes can vary significantly in different periods of nodule development and maturation. Therefore, *NCR-new35* might have higher expression levels at a different developmental stage. It was suggested in an earlier study that some *NCR*s may have a role during nodule senescence which could be another explanation for low expression detected in young, 3wpi nodules (Nallu et al., 2013). However, in another study the expression patterns of *NCR* genes in nodules subjected to senescence through nitrate or herbicide treatments were measured and results indicated that *NCR* genes, including *NCR-new35*; referred as *NCR014* in the corresponding study, are significantly downregulated during the senescence (Guefrachi et al., 2014). Moreover, our phenotypic characterization experiments, demonstrating in the absence of *NCR-new35* nodule development and

bacteroid differentiation is initiated but not completed, suggest that *NCR-new35* rather plays a crucial role during the formation and differentiation of the symbiotic nodules.

The very low expression of *NCR-new35* compared to the other three *NCR* genes may indicate that the abundance of the *NCR* peptide may not necessarily correlate with its expression as monitored by GUS staining. The gene expression might be affected by post-transcriptional regulatory mechanisms that control the stability or translation of its mRNA resulting in higher protein levels. Further complementation experiments fusing the promoter of *NCR-new35* with the gene sequence of other tested essential *NCR*s might provide insights in the reasons of observed low promoter activity.

7.10. NCR-new35 Peptide co-localize with rhizobia in the interzone and nitrogen fixation zones of complemented nodules

The analysis of the sub-cellular localization of NCR peptides within symbiotic cells is fundamental to unravel the complexity of the IRLC legume-rhizobia interactions. The encoded NCR peptides contain conserved signal peptides, which play a crucial role in their traffic to bacteroids (Wang et al., 2010). The understanding of the precise subcellular distribution of these peptides is essential to identify their functional significance in the establishment and maintenance of nitrogen-fixing symbiosis in *M. truncatula* and other IRLC legumes. In this context, the localization of NCR169 and NCR211 peptides to bacteroids within symbiotic cells was demonstrated previously in the interzone and nitrogen fixation zone (Horváth et al., 2015; Kim et al., 2015).

In this study, we employed translational fusions of the *mCherry* reporter with a construct under the control of the native promoter of the *NCR-new35* gene to investigate the intercellular localization of the NCR-new35 peptide. The pK7m34GW-rolD::EGFP-*NCRnew35::mCherry* construct was introduced into *Mtsym20* mutant roots. Additionally, the empty vector (pK7m34GW-rolD::EGFP) was introduced into both wild-type A17 and *Mtsym20* mutant roots as controls. The green signal of GFP transformation marker was visualized under fluorescent light (Fig. 7.10b,d,f). As expected, the wild-type plants introduced with empty vector developed their characteristic nitrogen fixing, pink, elongated nodules whereas empty vector-transformed *Mtsym20* plants did not show the rescue of the symbiotic phenotype and developed white, round shaped nodules 4 wpi with rhizobia. (Fig. 7.10a,e). We found that *Mtsym20* mutant plants transformed with the fluorescent tagged-version of *NCR-new35*, developed pink and elongated wild-type-like nodules 4 wpi with rhizobia strain *Sm* WSM419 indicating that the construct was able to complement the symbiotic defect of *Mtsym20* (Fig. 7.10c).

Functional complementation assays on the mutant nodules validated the preserved activity of the fusion protein and ensured that the fluorescent tag did not compromise the functionality of NCR-new35 protein. However, it should also be noted that the fluorescent tag of the fusion protein can be possibly cleaved and the cleaved NCR-new35 peptide may enter into the bacteroids and act there.

Nodules developed on roots transformed with the *NCRnew35::mCherry* construct (pK7m34GW-rolD::EGFP-*NCRnew35::mCherry*) were sectioned, stained with SYTO13 and analyzed under the confocal microscope. The red fluorescence of mCherry was distinctly observed in the interzone and nitrogen fixation zone of complemented nodules (Fig7.10 g). Further magnification of symbiotic cells revealed the co-localization of mCherry-tagged NCR-new35 peptide with SYTO13-stained rhizobia in both interzone and nitrogen fixation zone (Fig. 7.10h,i,j)

Previously, the presence of NCR peptides were examined in bacteroid preparations and NCR343 and NCR169 peptides were identified within the bacteroids, suggesting their association with these crucial symbiotic structures (Durgo et al., 2015). However, the absence of NCR211 and NCR-new35 in the bacteroids spurred our interest in understanding the localization patterns and potential entry mechanisms of these peptides during the symbiotic relationship. Our results revealed that NCR-new35 similarly to NCR169, exhibit co-localization with bacteroids, forming distinct ring-shaped fluorescent signals (Fig. 7.10). However, mechanism of action of NCR-new35 on bacteroids remains uncertain. It may act on the bacteroid surface, aligning with our *in-situ* localization experiment results. Alternatively, it may enter the bacteroids, potentially following proteolytic cleavage although this hypothesis does not fully correspond with our current findings, but it could potentially explain the absence of *NCR-new35* in the bacteroids. Following proteolytic cleavage, only certain fragments might be translocated into the bacteroids, resulting in different sizes or amounts of fragments in the bacteroids that could not be detected with the conducted proteomic analysis. The interaction of NCR-new35 with bacteroids could also be temporary or occur at a specific time during nodule formation, which may not align with the sampling time of the proteomic study, making its identification challenging.

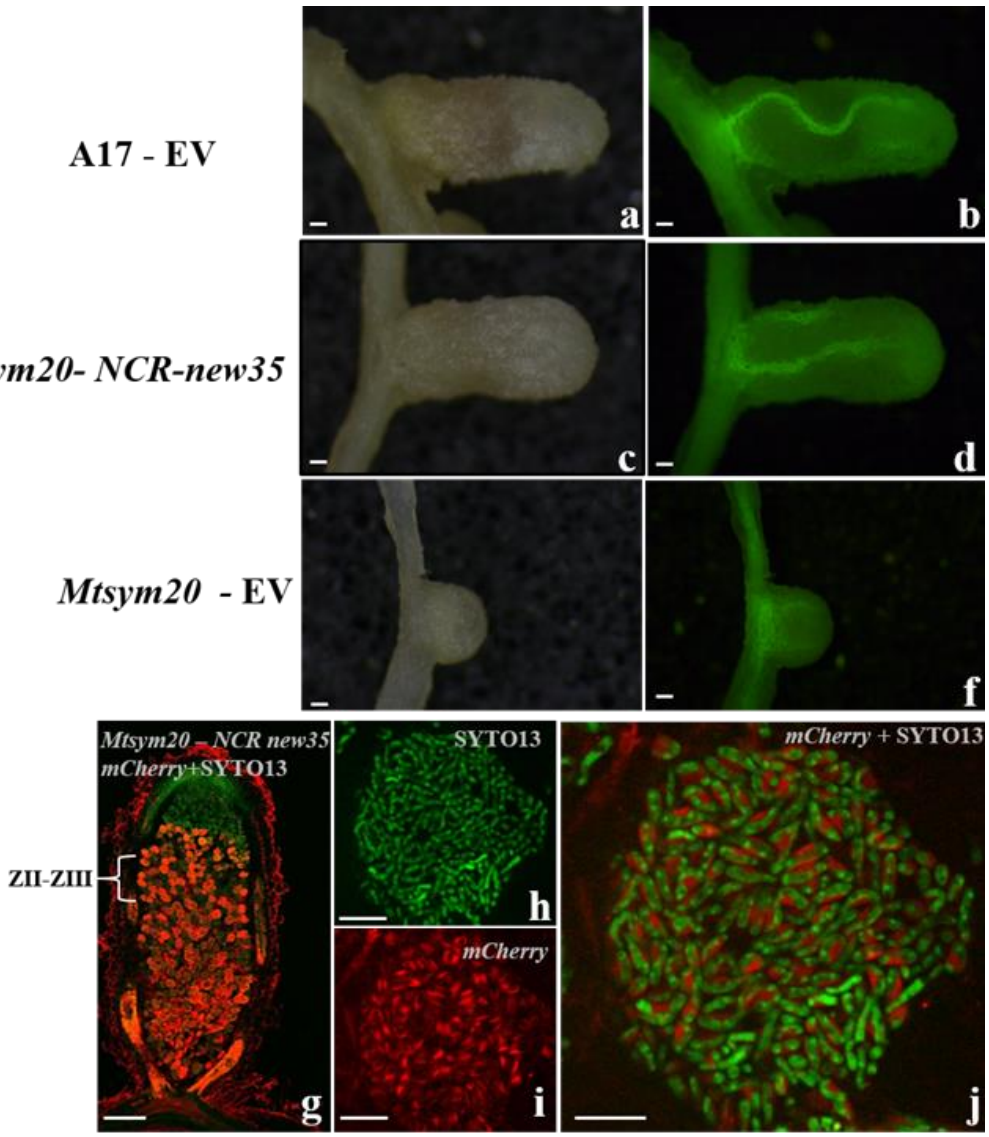


Fig. 7.10 NCR-new35 peptide localize to the symbiosomes

Fusion of NCR-new35 peptide with mCherry fluorescent protein rescued the mutant phenotype of *Mysym20* mutant 4wpi with *Sinorhizobium medicae* WSM419 (c). The transgenic roots were detected with GFP fluorescence (b,d,f) mCherry signal localized the NCR-new35 peptide in the cells of interzone (ZII-ZIII) and nitrogen fixing zone (ZIII) (g) the mCherry signal co-localized with the SYTO13 stained bacteroids (h-j) . Bars: (a-g) 200 μm; (h,i,j) 20 μm.

The localization of the NCR-new35 peptide was assessed in parallel experiments with NCR343 peptide localization studies in the nodule cells (Horvath et al., 2023). Results indicated that despite the barely detectable promoter activity of *NCR-new35* using the GUS reporter and the low level of read numbers found in transcriptomic studies, the peptide encoded by *NCR-new35* was localized in the interzone and nitrogen fixation zone similarly to NCR343 and the previously reported NCR211 and NCR169 peptides (Horvath et al., 2015, Kim et al., 2023, Horvath et al., 2023). This may suggest an enhanced stability of the NCR-new35 peptide, or the translation efficiency of the

NCR-new35 transcript might be particularly high, allowing for the production of peptide from relatively low levels of mRNA. (Fig. 9.9.1c,h and Fig. 7.10g).

Another explanation for the better detection of the mCherry-tagged peptide compared to GUS driven by the same promoter is that, following the translation of the peptide, it is concentrated in specific cell compartments, resulting in a locally accumulated presence of the NCR-new35 peptide. The discrepancy between the relatively strong signal with the mCherry-tagged NCR-new35 peptide and the weak signal with GUS expression driven by the native promoter of the gene could also be due to the presence of a promoter/enhancer element within the intron or the coding sequence of *NCR-new35*, which could enhance its translational efficiency.

7.11. Targeted mutagenesis of the *M. truncatula NCR-new35* gene using the *Agrobacterium rhizogenes*-mediated CRISPR/Cas9 system

The pivotal role of *NCR-new35* in establishing effective nitrogen-fixing symbiosis between *M. truncatula* and its rhizobia partner has been demonstrated in our study. These results are generated based on the forward genetic studies of *Mtsym20*. However, the mutations in *NCR*s and other genes are random in the mutants identified in mutant screens and further investigation of the symbiotic functions of other *NCR* genes in different *M. truncatula* accessions is a challenge. Therefore, the application of reverse genetic methods, such as CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated protein 9), to enable targeted gene editing and the analysis of gene function across different accessions would be useful. Although the CRISPR/Cas9 system is widely applicable in a broad field of biology due to its simplicity and robustness, it is rarely used in studies involving *M. truncatula*, mainly because of the low *in vitro* regeneration capacity of *M. truncatula*. Therefore, we aimed to develop a gene-editing system using the *Agrobacterium rhizogenes*-mediated CRISPR/Cas9 technology. To achieve this, we used two accessions of *M. truncatula*, the Jemalong 2HA, which is closely related to the original background of the *Mtsym20* mutant, and a genetically distinct yet well-characterized line, the *M. truncatula* R108, recently renamed as *Medicago littoralis* R108 (Choi et al., 2022) and targeted the *NCR-new35* gene with single guide RNA (sgRNA) CRISPR/Cas9 constructs. The constructs were delivered into the wild-type roots with the *Agrobacterium rhizogenes* mediated hairy root transformation. Comparing to other genetic transformation techniques, *A. rhizogenes*-mediated hairy root transformation has the advantage that it does not require the complete regeneration of transformed plants. Composite plants have wild-type shoots and transgenic, mutagenized roots generated by transformation and these hairy-roots are

particularly beneficial for investigating traits associated with roots, such as symbiotic nodule development.

CRISPR/Cas9 mutagenesis system requires a specific short RNA molecule for target site recognition, consisting of a 20-nucleotide sequence homologous to a specific genomic locus to be recognized by *Streptococcus pyogenes* Cas9 (SpCas9) endonuclease and a constant region ensuring binding to Cas9. SpCas9 cleaves only those targets where the 20 nucleotides recognized by the sgRNA are immediately followed by a protospacer adjacent motif (PAM) sequence, denoted as NGG.

To design an optimal single-guide RNA which targets the *NCR-new35* gene, the CRISPOR online tool (http://crispor.tefor.net/) referring to the *M. truncatula* A17 genome MtrunA17r5.0 version was utilized. Previous findings demonstrated the preference of the U6 promoter for a guanine nucleotide (G) at the 5' end of the sgRNA, the sgRNA was designed to have a GN_{19} NGG motif ensuring optimal expression and functionality of the sgRNA driven by the U6 promoter within the CRISPR/Cas9 system. Potential sgRNA target sites were predicted on the coding sequence of *NCR-new35* based on *M. truncatula* A17 reference genome. Subsequently, both A17 and R108 genomes were screened with the possible guide RNA sequences in order to reduce the possibility of the sgRNA binding to and cleaving non-target genomic regions, known as off targets. The sgRNA was designed to have mismatches when compared to potential off-target sites identified through the CRISPOR tool. Based on the analysis, only two effective guides could be designed for *NCR-new35*, one of them with a lower off-target chance (5' GGAAGGAAAGGAGGACCACC 3') was chosen to introduce into pKSE466_RR CRISPR vector carrying dsRed fluorescent selective marker. In pKSE466_RR vector, the sgRNA is driven by MtU6.6 and SpCas9 expression is controlled by the SV40 promoter.

The gene editing constructs were introduced into the roots of *Medicago* accessions using *A. rhizogenes*-mediated hairy root transformation and plants were analyzed for their symbiotic nodule phenotype 3-5wpi with rhizobia. Following the phenotypic characterization, the regions surrounding the targeted sequences were amplified by PCR using DNA isolated from transgenic nodule sections. The gene editing events in the targeted region were assessed using the Next Generation Sequencing (NGS) enabling accurate evaluation of genetic variations and their correlation with observed phenotypes. The genotypes identified by NGS were further examined for quantitative allele composition and verification of genome editing mutations using the online tool CRISPResso2.

Based on described methodology, the empty pKSE466_RR vector and the construct of pKSE466_RR CRISPR/Cas9 targeting *NCR-new35* were introduced into *M. truncatula* 2HA and *M. littoralis* R108 roots. Following hairy root transformation, newly emerged roots were screened under a fluorescent stereomicroscope, and the roots that did not show the red fluorescence of the constitutive dsRed marker gene were removed. *M. truncatula* 2HA plants were inoculated with their highly effective rhizobium partner SmWSM419 strain and *M. littoralis* R108 plants were inoculated with *S. meliloti* FSM-MA strain. The symbiotic phenotypes of the plants were evaluated 3-5 wpi. Roots which were transformed with the gene editing construct of *NCR-new35* displayed undeveloped white mutant nodules inefficient in nitrogen fixation on the roots of both *M. truncatula* 2HA and *M. littoralis* R108 plants (Fig. 7.11.1c,e,i,k and Fig. 7.11.2b,c,d,j,k,l). Additionally, some of these plants developed only pink elongated nodules characteristic of wild-type development although they were transformed with the gene editing construct of *NCR-new35* (Fig. 18a,g). The wild-type nodules developed on transgenic roots could be a result of heterozygous alleles carrying both wild-type and mutant genotypes, chimeric tissues containing both transformed and non-transformed cells or the complete lack of gene edition by Cas9. The development of transgenic pink nodules on roots with gene editing constructs could be also due to mutations resulting in a frame deletion of an amino acid which does not necessarily affect the symbiotic function of the gene. Such deletions may not alter the overall structure or activity of the protein, allowing it to retain its function despite the genetic modification. This phenomenon was exemplified in nodules developed on roots introduced with *NCR169* gene editing construct, where the deletion of the first Glu residue in the mature peptide of NCR169 did not impair its symbiotic efficacy (Güngör et al., 2023).

In this study, we only observed wild type nodule samples resulting from partial or unsuccessful gene editing of *NCR-new35* (Fig. 7.11.1a,g). These samples were compared to nodules that have been successfully gene-edited and the high mutation ratios within the targeted regions revealed correlation with ineffective symbiotic phenotype (Fig. 7.11.1c,e,i,k). The transgenic nodules showing nitrogen fixing deficient (Fix-) phenotype, very similar to nodules formed on *Mtsym20* mutant roots and nodules developed on empty vector introduced roots were further analyzed for phenotypic characterization experiments. Transgenic nodules were sectioned, stained with the DNA-binding fluorescent dye SYTO13 and examined under a confocal microscope. Pink, elongated nodules that developed on empty vector-transformed roots of both *M. truncatula* 2HA and *M. littoralis* R108 plants displayed the characteristic zonation of indeterminate nodules showing cells occupied by rhizobia in infection zone, interzone and nitrogen fixation zone (Fig. 7.11.2a,e,i,m). In contrary, SYTO13

fluorescence was detected only at the apical region of white, round shaped nodules formed on roots gene edited for *NCR-new35* whereas only empty or sporadic cells observed in nitrogen fixation zone (Fig. 7.11.2b,c,d,f,g,h,j,k,l,n,o,p).

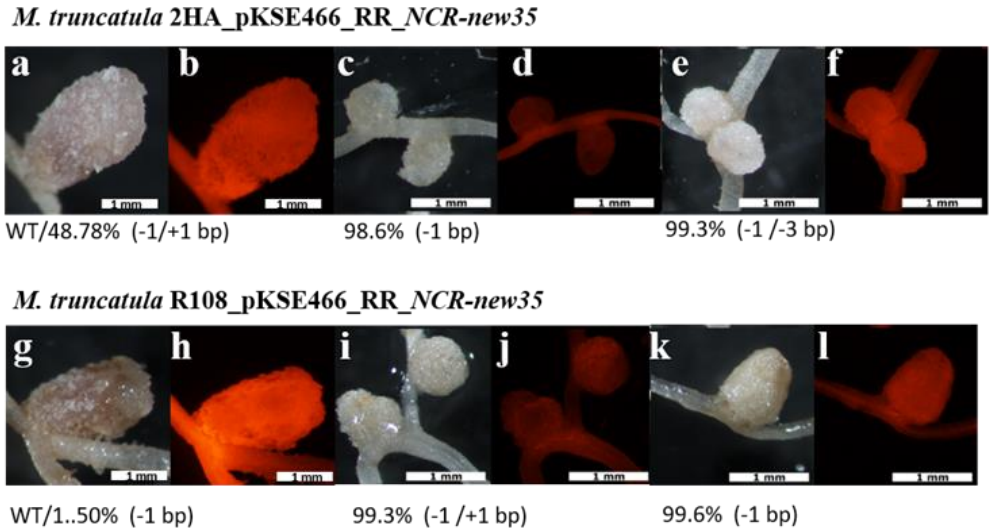


Fig.7.11.1 The nodulation phenotype of plants transformed with the gene editing construct of *NCR-new35*

Transgenic nodules were collected from hairy roots of *M. truncatula* 2HA (a-f) or *M. littoralis* R108 (g-l) transformed with the *NCR-new35* targeting CRISPR/Cas9 construct 3-5-week-post inoculation with *S. medicae* WSM419 (a-f) or *S. meliloti* FSM-MA (g-l), respectively. Transgenic roots were identified by the fluorescence of the DsRed protein (b,d,f,h,j,l). The nodulation phenotype of hairy roots transformed with the construct pKSE466-RR-*NCR-new35* was associated with the relative proportion of wild-type and mutant alleles. The *NCR-new35*-edited white undeveloped nodules carried almost fully mutant homozygous or biallelic mutations of few base pair deletions and/or insertions(c,e,i,k). Contrary, pink nodules carried high percentage of unmodified alleles (>40%) (a,g). The relative percentage and nature of mutant alleles indicated below each corresponding nodule was identified by the sequencing of the targeted region using the Illumina platform.

Following the phenotypic characterization, we analyzed the gene editing events in the *NCR-new35* targeted nodules using NGS. The undeveloped, white nodules displaying defects in colonization by rhizobia in the nitrogen fixation zone similar to *Mtsym20* mutants were selected from the transformed roots of *M. truncatula* 2HA and *M. littoralis* R108 plants had either biallelic or monoallelic homozygous mutations in the amplified regions including the target sequence of *MtNCR-new35* resulting in frame shifts at the end of the sequence coding for conserved signal peptide (Fig. 7.11.2q,r,s).

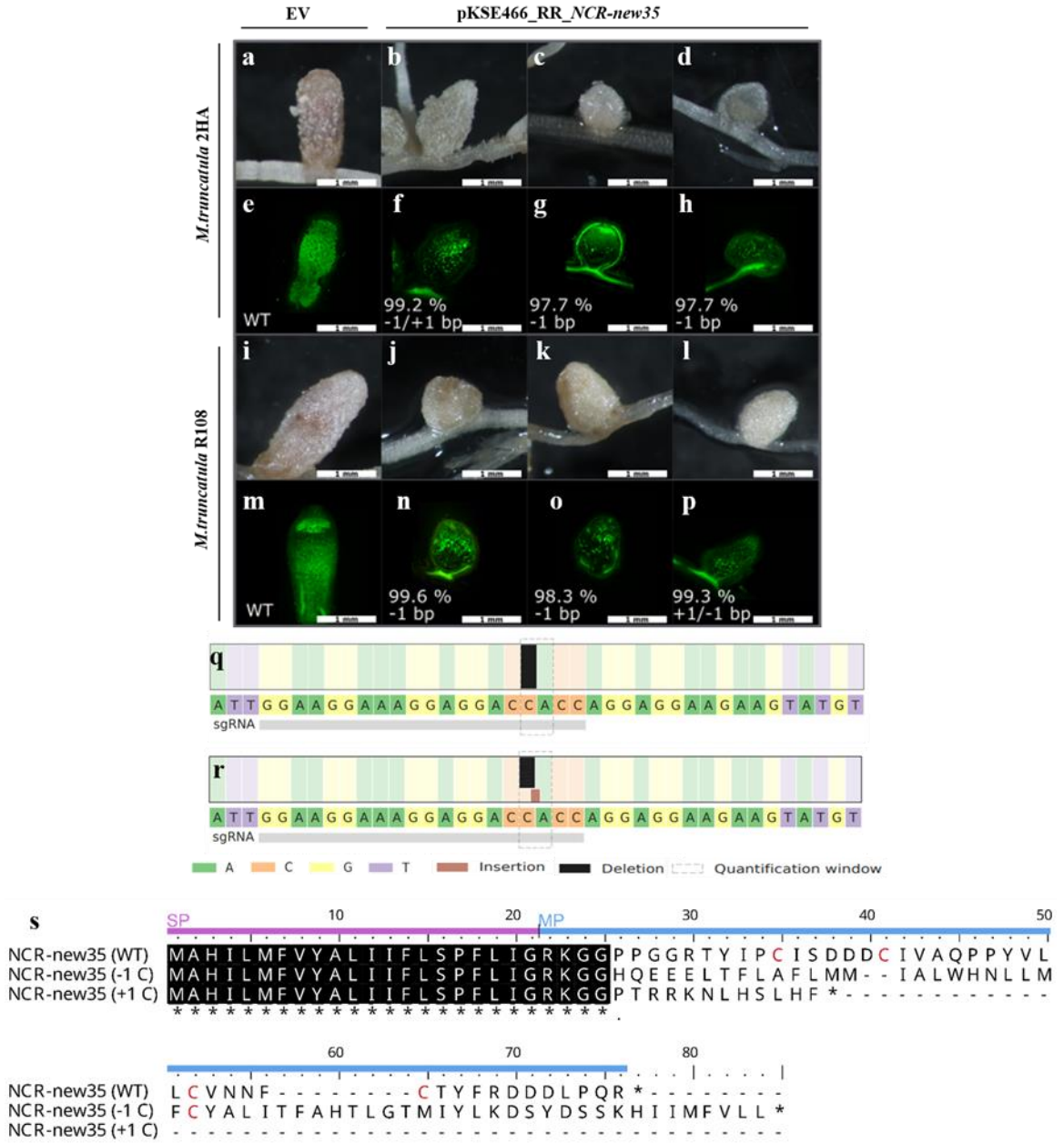


Fig. 7.11.2. The editing of *NCR-new35* induces the formation of nodules impaired in rhizobia colonization of the nitrogen-fixing zone

A. rhizogenes-mediated CRISPR/Cas9 targeting of *NCR-new35* in *M. truncatula* 2HA and *M. litorallis* R108 hairy roots resulted in ineffective nitrogen fixing nodules 3-5 weeks post inoculation with *S. medicae* WSM419 in the case of *2HA* (b,c,d) and with *S. meliloti FSM-MA* in the case of R108 plants (j,k,l). SYTO13 staining of the nodules revealed that the white undeveloped nodules (b,c,d,j,k,l) are hardly colonized at the basal part of the nodules (f,g,h,n,o,p) whereas empty vector (EV) introduced nodules developed elongated pink nodules (a,i) showing characteristic zonation and colonization of indeterminate nodules (e,m). The variety and distribution of alleles of two gene edited nodules were determined by Illumina amplicon sequencing (q,r) Alignment of NCR-new35 petide sequence of wild type plants and *NCR-new35* edited plants. Identical residues are marked with black background and conserved cysteine residues are higligted with red (s).

Alignment of the NCR-new35 peptide sequence of the wild type plants and *NCR-new35* edited plants revealed that all mature peptide sequence was completely different and did not contain cysteine residues that could form disulfide bonds. Therefore, the structure of the gene edited NCR-new35 peptide could not be further investigated for functional analysis Fig. (7.11.2s). A non-significant amount of reads of the wild-type allele detected in these nodules are likely originated from sporadic wild-type cells or a trace amount of wild-type contamination. The observed morphology and bacterial presence in zones of the genetically edited nodules corresponded to the described nodulation traits of *Mtsym20* nodules. These findings verified that *NCR-new35* plays a crucial role in the symbiotic relationships not only between *M. truncatula* 2HA and *S. medicae* WSM419, but also in the interactions between *M. littoralis* R108 and *S. meliloti* FSM-MA, underlying its conserved function of NCR-new35 across different *Medicago* accessions. In addition, our results proved that the CRISPR/Cas9 system is suitable to screen the symbiotic phenotype of genes in different *M. truncatula* lines using hairy root transformation.

8. CONCLUSIONS AND PERSPECTIVES

The *Mtsym20* symbiotic mutant was in the focus of our research, aiming to identify the impaired gene which is responsible for the symbiotic nitrogen fixation deficiency in the *Mtsym20* mutant. Experimental results led to the following conclusions:

Mtsym20 was reported previously as a Fix- mutant and identified as a distinct allele from *Mtsym19* (Morandi et al., 2005). Our phenotypic characterization experiments confirmed that *Mtsym20* mutant showed inefficient nitrogen fixing symbiotic phenotype. However, allelism tests which we conducted because we found that the *Mtsym19* and *Mtsym20* loci mapped to same genomic location, revealed that *Mtsym20* was allelic to *Mtsym19* mutant.

Further investigations of the mutant plants with transcriptome analysis, highlighted *NCR-new35* as a primary candidate for the symbiotic locus of *Mtsym20*. The introduction of the *NCR-new35* gene into the *Mtsym20* mutant roots, rescued the nitrogen starvation and the nodulation phenotype of the mutant confirming that the lack of *NCR-new35* gene is responsible for the nitrogen fixation deficient phenotype of *Mtsym20*. Although the *M. truncatula* genome contains more than 700 *NCR* genes, to date only a few of them including the *NCR169, NCR211, NCR343* and *NCR-new35* (Horváth et al., 2015, 2023; Kim et al., 2015) have been identified as crucial factors for effective nitrogen-fixing symbiosis using forward genetic approaches. Contrary, some other *NCR*s such as *NCR341, NCR344, NCR345* and *NCR178* (Horváth et al., 2015, 2023; Kim et al., 2015) have been identified as non-essential peptides for the studied symbiotic interactions. The sequential induction and overlapping expression of the high number of *NCR* genes in the *M. truncatula* genome suggests a collaborative function of the NCR peptides to enhance the symbiotic interaction between rhizobia and the host plant (Mergaert, 2018). Therefore, in this study, mutants deficient in *NCR* genes were assessed comprehensively to understand their individual and collaborative functions in the symbiosis.

On the roots of *Mtsym20* mutant plants, undeveloped, slightly cylindrical and white nodules were formed similar to those observed on the roots of other mutants defective in essential *NCR* genes, *Mtdnf4-1*, *Mtdnf7-2*, NF-FN9363, following inoculation with *S. medicae*. These nodules were characterized by the recognizable zones of indeterminate nodules but the colonization in these nodules was restricted to the distal zones and no colonized cells were observed in the nitrogen fixation zones. The comprehensive histological analysis of the mutants showed that the cell structure in the distal

zone of mutant nodules did not show any detectable difference comparing to wild-type nodules indicating that the differentiation process is initiated in the symbiotic cells of the mutants. However, an extended transition zone was observed in mutant nodules compared to wild-type nodules, indicating defects in the complete differentiation process. These results are consistent with previous results, which suggest that the lack of essential *NCR*s can lead to either impaired bacteroid differentiation or the death of rhizobia within the host cell (Pan and Wang, 2017; Roy et al., 2020).

Detailed examination of the size and DNA content level of bacteroids revealed that mutants lacking crucial *NCR* genes, *NCR-new35*, *NCR343*, *NCR211*, predominantly accumulate non-elongated bacteria and exhibits a slightly decreased percentage of the elongated and endoreduplicated bacteroids compared to those observed in wild-type nodules. These findings, in line with the results of our microscopic analysis, led to the conclusion that mutations in the *NCR-new35*, *NCR343*, and *NCR211* genes cause defects in the late stage of bacteroid differentiation. In contrast, *Mtdnf7-2* mutant lacking the *NCR169* gene displayed a reduced bacteroid size and endoreduplication level suggesting an earlier function of the *NCR169* compared to *NCR-new35*, *NCR343* and *NCR211* genes.

Expression levels of essential *NCR* genes were compared using a *GUS*-reporter assay. Promoter activity of *NCR-new35* was observed primarily in the interzone, in contrast to the expression of other critical *NCR*s, which was detected in both the interzone and the nitrogen fixation zone. This pattern confirms the RNA sequencing data derived from laser-capture microdissected nodule zones (Roux et al., 2014). This finding suggests that *NCR-new35* may have a specialized role, regulated differently from other members of the family of crucial *NCR*s. While the promoter activity of other essential *NCR*s was clearly visible with GUS staining, that of *NCR-new35* was barely detectable. This can be due to the presence of special regulatory elements that suppress or restrict the expression of *NCR-new35*. It would be an interesting line of research in the future to identify the element responsible for the lower activity of *NCR-new35.*

The efficient symbiotic phenotype of the *Mtsym20* could be restored following the introduction of the *NCR-new35::mCherry* fluorescent tag fusion into the mutant roots. Bacteroids surrounded by the fluorescent signal were observed in the nodules of mutants, indicating the co-localization of the *NCR-new35* peptide with bacteroids. In a previous proteomic study by Durgo et al. (2015), *NCR343* and *NCR169* were identified in bacteroid preps, whereas *NCR211* and *NCR-new35* were not. The absence of NCR-new35 in bacteroids may indicate that at least a subset of the fluorescent-tagged *NCR* fusion proteins—possibly following proteolytic cleavage—enters the bacteroids and could not be

identified due to their size or abundance. Additionally, NCR-new35 may interact with bacteroids transiently, at a specific stage of nodule development or under the specific conditions which could not be captured at the time of sampling of tested nodules.

Despite low promoter activity detected by the GUS reporter, the peptide encoded by *NCR-new35* was localized in the interzone and nitrogen fixation zone similarly to NCR343, NCR211 and NCR169 peptides, suggesting possible enhanced stability or high translation efficiency of the peptide from minimal mRNA. The results alternatively suggest a higher peptide concentration in specific cellular compartments after translation.

The gene edited nodules formed on *M. truncatula* 2HA and *M. littoralis* R108 transformed with *NCR-new35* editing CRISPR/Cas9 construct showed ineffective phenotype and displayed non-colonized cells in the nitrogen fixation zone. The sequence of the DNA isolated from the sections of these nodules were carrying homozygous or biallelic mutations indicating that *NCR-new35* is essential for symbiosis between *M. truncatula* 2HA and *S. medicae* WSM419 as well as between *M. littoralis* R108 and *S. meliloti* FSM-MA indicating conservation of the gene function across various *Medicago* accessions.

Considering large number of *NCR*s in the *M. truncatula* genome, further large-scale forward and reverse genetic studies are necessary to investigate the symbiotic functions of additional NCR peptides in *M. truncatula*. The development of the *A. rhizogenes* hairy root system-based gene editing system in *M. truncatula* provide an excellent tool to study the requirement and function of *NCR*s and other symbiotic genes.

9. NEW SCIENTIFIC FINDINGS

This research reveals novel findings that enhance our understanding of the genetic and molecular mechanisms of nitrogen fixing symbiosis in *Medicago truncatula*, specifically focusing on the genetic, molecular and phenotypic characterization of *NCR-new35* gene and the comparative analysis of *NCR-new35* with other reported crucial *NCR* genes, *NCR169, NCR211* and *NCR343.*

Mtsym20* is an allelic counterpart of *Mtsym19: Our allelism tests demonstrated that *Mtsym20* and *Mtsym19* are allelic to each other contrary to previously reported results (Morandi et al., 2005).

The nitrogen fixation zone of *Mtsym20* mutant nodules is not colonized with rhizobia: The *Mtsym20* mutant, lacking the *NCR-new35* gene, similarly to the *Mtdnf4-1*, *Mtdnf7-2*, NF-FN9363 mutants lacking *NCR169, NCR211* and *NCR343* genes, respectively, showed that the rhizobia colonization is restricted to the distal zones of the symbiotic nodules indicating the defect in the persistence of rhizobia in the nitrogen fixation zone of all tested mutants.

Bacteroid elongation and endoreduplication are impaired in *Mtsym20* nodules: The length and DNA content level of the bacteroids isolated from the nodules of the *Mtsym20* mutant were slightly reduced compared to rhizobia isolated from wild-type *M. truncatula* A17 nodules, suggesting an incomplete differentiation process disrupted by a mutation at a late stage of bacteroid differentiation.

***NCR-new35* is essential for effective nitrogen fixing symbiosis between *Medicago truncatula* and the tested rhizobia**: The forward genetic analysis revealed that *NCR-new35* is responsible for the symbiotic phenotype of *Mtsym20.*

NCR-new35 peptides localize to symbiosomes: The fluorescence of the mCherry-tagged reporter version of NCR-new35 was detected around the bacteroids revealing the co-localization of NCR-new35 s with rhizobia.

The essential role of NCR-new35 in symbiotic nitrogen fixation was demonstrated across different Medicago accessions using the CRISPR/Cas9 gene editing system: *NCR-new35* gene-edited plants of *M. truncatula* 2HA and *M. littoralis* R108 developed ineffective symbiotic nodules similar to *Mtsym20* plants, confirming the conserved essential role of *NCR-new35* in nitrogen fixation symbiosis across different *Medicago* accessions.

10. SUMMARY

Medicago truncatula is one of the model organisms for investigating the genetic and molecular background of the nitrogen fixation mechanism of leguminous plants. In this process, atmospheric nitrogen is efficiently assimilated in the root nodules in a symbiotic interaction with rhizobia.

In recent years, numerous genes necessary for the formation of nitrogen-fixing nodules have been identified in the genome of *M. truncatula*, mainly by applying forward genetic analysis of plants produced by random mutagenesis.

In this study, we focused on *Mtsym20* symbiotic mutant plant line, which forms nodules that is defective in symbiotic nitrogen fixation. We proved that bacteroid development is incomplete in *Mtsym20* mutant, similarly to other Fix- mutants, *dnf7-2*, *dnf4-1* and NF-FN9363, using light microscopy and fluorescent microscopy. Combining the results from our confocal microscopy examinations and flow cytometric data analysis on bacteroid isolates from mutant and wild-type nodules, we concluded that bacteroid elongation is moderately decreased and genome endoreduplication is slightly less advanced in the bacteroids isolated from the nodules of mutant plants compared to those in wild-type nodules.

Using genetic mapping, we identified an approximately 400kb genetic region on upper arm of chromosome 4 in which the gene responsible for the mutation was located. Combining this result with RNA sequencing data obtanied from the mutant nodules, we identifed deletion of three neighbouring genes in an approximately 32 kb deletion region in genomic region of *Mtsym20*. Subsequently, we proved that *NCR-new35* is the gene responsible for the symbiotc defect of *Mtsym20* with genetic complementation experiments.

We analyzed the subcellular localization of NCRnew35 peptide with fluorescently tagged version of the peptided which was able to restore the effective symbiosis in *Mtsym20*. We observed NCR-new35 peptides in the interzone and nitrogen fixing zone of the nodules. The fluorescent-tagged version of NCR-new35 localized together with bacteroids in the symbiosomes.

Using CRISPR/Cas9 based targeted genome editing technology, we proved that *NCR-new 35* is not only essential in the most commonly used *M. truncatula* 2HA cultivar, but also essential for symbiotic nitrogen fixation of *M. littoralis* R108 ecotype.

11. REFERENCES

Alunni, B., and Gourion, B. (2016). Terminal bacteroid differentiation in the legume−rhizobium symbiosis: nodule-specific cysteine-rich peptides and beyond. *New Phytologist* 211, 411–417. doi: 10.1111/nph.14025

Alunni, B., Kevei, Z., Redondo-Nieto, M., Kondorosi, A., Mergaert, P., and Kondorosi, E. (2007). Genomic Organization and Evolutionary Insights on GRP and NCR Genes, Two Large Nodule-Specific Gene Families in Medicago truncatula. *MPMI* 20, 1138–1148. doi: 10.1094/MPMI-20-9-1138

Arrighi, J.-F., Barre, A., Ben Amor, B., Bersoult, A., Soriano, L. C., Mirabella, R., et al. (2006). The Medicago truncatula Lysine Motif-Receptor-Like Kinase Gene Family Includes NFP and New Nodule-Expressed Genes. *Plant Physiology* 142, 265–279. doi: 10.1104/pp.106.084657

Barker, D. G., Pfaff, T., Moreau, D., Groves, E., Ruffel, S., Lepetit, M., et al. (2006). Growing M. truncatula: choice of substrates and growth conditions.

Batut, J., Mergaert, P., and Masson-Boivin, C. (2011). Peptide signalling in the rhizobium–legume symbiosis. *Current Opinion in Microbiology* 14, 181–187. doi: 10.1016/j.mib.2010.12.010

Baudin, M., Laloum, T., Lepage, A., Rípodas, C., Ariel, F., Frances, L., et al. (2015). A Phylogenetically Conserved Group of Nuclear Factor-Y Transcription Factors Interact to Control Nodulation in Legumes. *Plant Physiology* 169, 2761–2773. doi: 10.1104/pp.15.01144

Bell, C. J., Dixon, R. A., Farmer, A. D., Flores, R., Inman, J., Gonzales, R. A., et al. (2001). The Medicago Genome Initiative: a model legume database. *Nucleic Acids Research* 29, 114–117. doi: 10.1093/nar/29.1.114

Benedito, V. A., Torres-Jerez, I., Murray, J. D., Andriankaja, A., Allen, S., Kakar, K., et al. (2008). A gene expression atlas of the model legume Medicago truncatula. *The Plant Journal* 55, 504–513. doi: 10.1111/j.1365-313X.2008.03519.x

Bennett, M. D., and Leitch, I. J. (1995). Nuclear DNA Amounts in Angiosperms. *Annals of Botany* 76, 113–176. doi: 10.1006/anbo.1995.1085

Benson, D. R., and Silvester, W. B. (1993). Biology of Frankia strains, actinomycete symbionts of actinorhizal plants. *Microbiological Reviews* 57, 293–319. doi: 10.1128/mr.57.2.293-319.1993

Beringer, J. E. (1974). R Factor Transfer in Rhizobium leguminosarum. *Microbiology* 84, 188–198. doi: 10.1099/00221287-84-1-188

Bhardwaj, A., and Sinharoy, S. (2022). "Transcriptional Networks in Medicago truncatula: Genomic and Functional Overview During Root Nodule Symbiosis," in *The Medicago truncatula Genome*, eds. S. Sinharoy, Y. Kang, and V. Benedito (Cham: Springer International Publishing), 71–90. doi: 10.1007/978-3-030-90757-0_6

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

Bonaldi, K., Gargani, D., Prin, Y., Fardoux, J., Gully, D., Nouwen, N., 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. *MPMI* 24, 1359–1371. doi: 10.1094/MPMI-04-11-0093

Brassó, R. (2017). Medicago Truncatula TRV43 Mutáns Vonal Ineffektív Szimbiotikus Nitrogénkötő Kapcsolatának Genetikai Vizsgálata. Szent István University.

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

Cebolla, A., María Vinardell, J., Kiss, E., Oláh, B., Roudier, F., Kondorosi, A., et al. (1999). The mitotic inhibitor ccs52 is required for endoreduplication and ploidy-dependent cell enlargement in plants. *The EMBO Journal* 18, 4476–4484. doi: 10.1093/emboj/18.16.4476

Cerri, M. R., Frances, L., Laloum, T., Auriac, M.-C., Niebel, A., Oldroyd, G. E. D., et al. (2012). Medicago truncatula ERN Transcription Factors: Regulatory Interplay with NSP1/NSP2 GRAS Factors and Expression Dynamics throughout Rhizobial Infection. *Plant Physiology* 160, 2155–2172. doi: 10.1104/pp.112.203190

Cheng, Jiujun., Sibley, C. D., Zaheer, Rahat., and Finan, T. M. (2007). A Sinorhizobium meliloti minE mutant has an altered morphology and exhibits defects in legume symbiosis. *Microbiology* 153, 375–387. doi: 10.1099/mic.0.2006/001362-0

Choi, H.-K., Kim, D., Uhm, T., Limpens, E., Lim, H., Mun, J.-H., et al. (2004). A sequence-based genetic map of Medicago truncatula and comparison of marker colinearity with M. sativa. *Genetics* 166, 1463–1502.

Choi, I.-S., Wojciechowski, M. F., Steele, K. P., Hopkins, A., Ruhlman, T. A., and Jansen, R. K. (2022). Plastid phylogenomics uncovers multiple species in Medicago truncatula (Fabaceae) germplasm accessions. *Sci Rep* 12, 21172. doi: 10.1038/s41598-022-25381-1

Cooper, J. E. (2004). "Multiple Responses of Rhizobia to Flavonoids During Legume Root Infection," in *Advances in Botanical Research*, (Academic Press), 1–62. doi: 10.1016/S0065-2296(04)41001-5

Crawford, N. M., and Forde, B. G. (2002). Molecular and Developmental Biology of Inorganic Nitrogen Nutrition. *Arabidopsis Book* 1, e0011. doi: 10.1199/tab.0011

Crespi, M., and Frugier, F. (2008). De Novo Organ Formation from Differentiated Cells: Root Nodule Organogenesis. *Sci. Signal.* 1. doi: 10.1126/scisignal.149re11

Czernic, P., Gully, D., Cartieaux, F., Moulin, L., Guefrachi, I., Patrel, D., et al. (2015). Convergent Evolution of Endosymbiont Differentiation in Dalbergioid and Inverted Repeat-Lacking Clade

Legumes Mediated by Nodule-Specific Cysteine-Rich Peptides. *Plant Physiology* 169, 1254–1265. doi: 10.1104/pp.15.00584

De Bang, T. C., Lundquist, P. K., Dai, X., Boschiero, C., Zhuang, Z., Pant, P., et al. (2017). Genome-Wide Identification of *Medicago* Peptides Involved in Macronutrient Responses and Nodulation. *Plant Physiol.* 175, 1669–1689. doi: 10.1104/pp.17.01096

de Lajudie, P. M., Andrews, M., Ardley, J., Eardly, B., Jumas-Bilak, E., Kuzmanović, N., et al. (2019). Minimal standards for the description of new genera and species of rhizobia and agrobacteria. *International Journal of Systematic and Evolutionary Microbiology* 69, 1852–1863. doi: 10.1099/ijsem.0.003426

de Oliveira Carvalho, A., and Gomes, V. M. (2009a). Plant defensins—prospects for the biological functions and biotechnological properties. *Peptides* 30, 1007–1020.

de Oliveira Carvalho, A., and Gomes, V. M. (2009b). Plant defensins—prospects for the biological functions and biotechnological properties. *Peptides* 30, 1007–1020.

Domonkos, A., Horvath, B., Marsh, J. F., Halasz, G., Ayaydin, F., Oldroyd, G. E., et al. (2013). The identification of novel loci required for appropriate nodule development in Medicago truncatula. *BMC Plant Biol* 13, 157. doi: 10.1186/1471-2229-13-157

Dong, W., Zhu, Y., Chang, H., Wang, C., Yang, J., Shi, J., et al. (2021). An SHR–SCR module specifies legume cortical cell fate to enable nodulation. *Nature* 589, 586–590. doi: 10.1038/s41586-020-3016-z

Downie, J. A., and Kondorosi, E. (2021). Why Should Nodule Cysteine-Rich (NCR) Peptides Be Absent From Nodules of Some Groups of Legumes but Essential for Symbiotic N-Fixation in Others? *Front. Agron.* 3, 654576. doi: 10.3389/fagro.2021.654576

Downie, J. A., and Walker, S. A. (1999). Plant responses to nodulation factors. *Current Opinion in Plant Biology* 2, 483–489. doi: 10.1016/S1369-5266(99)00018-7

Durgo, H., Klement, E., Hunyadi-Gulyas, E., Szucs, A., Kereszt, A., Medzihradszky, K. F., et al. (2015). Identification of nodule-specific cysteine-rich plant peptides in endosymbiotic bacteria. *Proteomics* 15, 2291–2295. doi: 10.1002/pmic.201400385

Farkas, A., Maróti, G., Dürgő, H., Györgypál, Z., Lima, R. M., Medzihradszky, K. F., et al. (2014). *Medicago truncatula* symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5183–5188. doi: 10.1073/pnas.1404169111

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

Farkas, A., Pap, B., Kondorosi, É., and Maróti, G. (2018). Antimicrobial Activity of NCR Plant Peptides Strongly Depends on the Test Assays. *Front. Microbiol.* 9, 2600. doi: 10.3389/fmicb.2018.02600

Felle, H. H., Kondorosi, É., Kondorosi, Á., and Schultze, M. (1998). The role of ion fluxes in Nod factor signalling in *Medicago sativa*. *The Plant Journal* 13, 455–463. doi: 10.1046/j.1365-313X.1998.00041.x

Fonouni-Farde, C., Tan, S., Baudin, M., Brault, M., Wen, J., Mysore, K. S., et al. (2016). DELLA-mediated gibberellin signalling regulates Nod factor signalling and rhizobial infection. *Nat Commun* 7, 12636. doi: 10.1038/ncomms12636

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

Gage, D. J. (2004). Infection and Invasion of Roots by Symbiotic, Nitrogen-Fixing Rhizobia during Nodulation of Temperate Legumes. *Microbiol Mol Biol Rev* 68, 280–300. doi: 10.1128/MMBR.68.2.280-300.2004

Galloway, J. N., Schlesinger, W. H., Levy II, H., Michaels, A., and Schnoor, J. L. (1995). Nitrogen fixation: Anthropogenic enhancement-environmental response. *Global Biogeochemical Cycles* 9, 235–252. doi: 10.1029/95GB00158

Geurts, R., and Bisseling, T. (2002). *Rhizobium* Nod Factor Perception and Signalling. *Plant Cell* 14, S239–S249. doi: 10.1105/tpc.002451

Gibson, A. H., and Nutman, P. S. (1960). Studies on the Physiology of Nodule Formation: With two Figures in the Text: A Reappraisal of the Effect of Preplanting. *Annals of Botany* 24, 420–433. doi: 10.1093/oxfordjournals.aob.a083715

González-Sama, A., de la Peña, T. C., Kevei, Z., Mergaert, P., Lucas, M. M., de Felipe, M. R., et al. (2006). Nuclear DNA Endoreduplication and Expression of the Mitotic Inhibitor Ccs52 Associated to Determinate and Lupinoid Nodule Organogenesis. *MPMI* 19, 173–180. doi: 10.1094/MPMI-19-0173

Gough, C., and Cullimore, J. (2011). Lipo-chitooligosaccharide Signaling in Endosymbiotic Plant-Microbe Interactions. *MPMI* 24, 867–878. doi: 10.1094/MPMI-01-11-0019

Graham, M. A., Silverstein, K. A. T., Cannon, S. B., and VandenBosch, K. A. (2004). Computational Identification and Characterization of Novel Genes from Legumes. *Plant Physiology* 135, 1179–1197. doi: 10.1104/pp.104.037531

Gresshoff, P. M., and Rolfe, B. G. (1978). Viability of Rhizobium bacteroids isolated from soybean nodule protoplasts. *Planta* 142, 329–333. doi: 10.1007/BF00385085

Guefrachi, I., Nagymihaly, M., Pislariu, C. I., Van De Velde, W., Ratet, P., Mars, M., et al. (2014). Extreme specificity of NCR gene expression in Medicago truncatula. *BMC Genomics* 15, 712. doi: 10.1186/1471-2164-15-712

Güngör, B., Biró, J. B., Domonkos, Á., Horváth, B., and Kaló, P. (2023). Targeted mutagenesis of Medicago truncatula Nodule-specific Cysteine-Rich (NCR) genes using the Agrobacterium rhizogenes-mediated CRISPR/Cas9 system. *Sci Rep* 13, 20676. doi: 10.1038/s41598-023-47608-5

Haag, A. F., Arnold, M. F. F., Myka, K. K., Kerscher, B., Dall'Angelo, S., Zanda, M., et al. (2013). Molecular insights into bacteroid development during *Rhizobium*– legume symbiosis. *FEMS Microbiol Rev* 37, 364–383. doi: 10.1111/1574-6976.12003

Haag, A. F., Baloban, M., Sani, M., Kerscher, B., Pierre, O., Farkas, A., et al. (2011). Protection of Sinorhizobium against Host Cysteine-Rich Antimicrobial Peptides Is Critical for Symbiosis. *PLOS Biology* 9, e1001169. doi: 10.1371/journal.pbio.1001169

Haag, A. F., Kerscher, B., Dall'Angelo, S., Sani, M., Longhi, R., Baloban, M., et al. (2012). Role of Cysteine Residues and Disulfide Bonds in the Activity of a Legume Root Nodule-specific, Cysteine-rich Peptide. *Journal of Biological Chemistry* 287, 10791–10798. doi: 10.1074/jbc.M111.311316

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT., in *Nucleic acids symposium series*, (Oxford), 95–98.

Heidstra, R., and Bisseling, T. (1996). Nod factor-induced host responses and mechanisms of Nod factor perception. *New Phytologist* 133, 25–43. doi: 10.1111/j.1469-8137.1996.tb04339.x

Helsel, Z. R. ed. (1987). *Energy in plant nutrition and pest control*. Amsterdam: Elsevier.

Horváth, B., Domonkos, Á., Kereszt, A., Szűcs, A., Ábrahám, E., Ayaydin, F., et al. (2015). Loss of the nodule-specific cysteine rich peptide, NCR169, abolishes symbiotic nitrogen fixation in the *Medicago truncatula dnf7* mutant. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15232–15237. doi: 10.1073/pnas.1500777112

Horváth, B., Güngör, B., Tóth, M., Domonkos, Á., Ayaydin, F., Saifi, F., et al. (2023). The *Medicago truncatula* nodule-specific cysteine-rich peptides, NCR343 and NCR-new35 are required for the maintenance of rhizobia in nitrogen-fixing nodules. *New Phytologist* 239, 1974–1988. doi: 10.1111/nph.19097

Horváth, B., Yeun, L. H., Domonkos, Á., Halász, G., Gobbato, E., Ayaydin, F., et al. (2011). Medicago truncatula IPD3 Is a Member of the Common Symbiotic Signaling Pathway Required for Rhizobial and Mycorrhizal Symbioses. *MPMI* 24, 1345–1358. doi: 10.1094/MPMI-01-11-0015

Howieson, J. G., and Ewing, M. A. (1986). Acid tolerance in the Rhizobium meliloti – Medicago symbiosis. *Aust. J. Agric. Res.* 37, 55–64. doi: 10.1071/ar9860055

Isozumi, N., Masubuchi, Y., Imamura, T., Mori, M., Koga, H., and Ohki, S. (2021). Structure and antimicrobial activity of NCR169, a nodule-specific cysteine-rich peptide of Medicago truncatula. *Sci Rep* 11, 9923. doi: 10.1038/s41598-021-89485-w

Jarzyniak, K., Banasiak, J., Jamruszka, T., Pawela, A., Di Donato, M., Novák, O., et al. (2021). Early stages of legume–rhizobia symbiosis are controlled by ABCG-mediated transport of active cytokinins. *Nat. Plants* 7, 428–436. doi: 10.1038/s41477-021-00873-6

Jin, Y., Liu, H., Luo, D., Yu, N., Dong, W., Wang, C., et al. (2016). DELLA proteins are common components of symbiotic rhizobial and mycorrhizal signalling pathways. *Nat Commun* 7, 12433. doi: 10.1038/ncomms12433

Jones, K. M., Kobayashi, H., Davies, B. W., Taga, M. E., and Walker, G. C. (2007). How rhizobial symbionts invade plants: the Sinorhizobium–Medicago model. *Nat Rev Microbiol* 5, 619–633. doi: 10.1038/nrmicro1705

Kaló, P., Gleason, C., Edwards, A., Marsh, J., Mitra, R. M., Hirsch, S., et al. (2005). Nodulation Signaling in Legumes Requires NSP2, a Member of the GRAS Family of Transcriptional Regulators. *Science* 308, 1786–1789. doi: 10.1126/science.1110951

Kang, Y., Sakiroglu, M., Krom, N., Stanton-Geddes, J., Wang, M., Lee, Y.-C., et al. (2015). Genome-wide association of drought-related and biomass traits with HapMap SNPs in Medicago truncatula. *Plant, Cell & Environment* 38, 1997–2011. doi: 10.1111/pce.12520

Kereszt, A., Mergaert, P., and Kondorosi, E. (2011). Bacteroid Development in Legume Nodules: Evolution of Mutual Benefit or of Sacrificial Victims? *MPMI* 24, 1300–1309. doi: 10.1094/MPMI-06-11-0152

Kereszt, A., Mergaert, P., Montiel, J., Endre, G., and Kondorosi, É. (2018). Impact of Plant Peptides on Symbiotic Nodule Development and Functioning. *Front. Plant Sci.* 9, 1026. doi: 10.3389/fpls.2018.01026

Kevei, Z., Vinardell, J. M., Kiss, G. B., Kondorosi, A., and Kondorosi, E. (2002). Glycine-Rich Proteins Encoded by a Nodule-Specific Gene Family Are Implicated in Different Stages of Symbiotic Nodule Development in Medicago Spp. *MPMI* 15, 922–931. doi: 10.1094/MPMI.2002.15.9.922

Kim, M., Chen, Y., Xi, J., Waters, C., Chen, R., and Wang, D. (2015). An antimicrobial peptide essential for bacterial survival in the nitrogen-fixing symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15238–15243. doi: 10.1073/pnas.1500123112

Kiss, G. B., Kaló, P., Kiss, P., Felföldi, K., Kereszt, A., and Endre, G. (1998a). "Construction of an Improved Genetic Map of Diploid Alfalfa (Medicago Sativa) Using a Novel Linkage Analysis for Chromosomal Regions Exhibiting Extreme Distorted Segregation," in *Biological Nitrogen Fixation for the 21st Century: Proceedings of the 11th International Congress on Nitrogen Fixation, Institut Pasteur, Paris, France, July 20–25, 1997*, eds. C. Elmerich, A. Kondorosi, and W. E. Newton (Dordrecht: Springer Netherlands), 313–313. doi: 10.1007/978-94-011-5159-7_176

Kiss, G. B., Kereszt, A., Kiss, P., and Endre, G. (1998b). Colormapping: a non-mathematical procedure for genetic mapping. *Acta Biologica Hungarica* 49, 125–142.

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

Kondorosi, E., Roudier, F., and Gendreau, E. (2000). Plant cell-size control: growing by ploidy? *Current Opinion in Plant Biology* 3, 488–492. doi: 10.1016/S1369-5266(00)00118-7

Laffont, C., Ivanovici, A., Gautrat, P., Brault, M., Djordjevic, M. A., and Frugier, F. (2020). The NIN transcription factor coordinates CEP and CLE signaling peptides that regulate nodulation antagonistically. *Nat Commun* 11, 3167. doi: 10.1038/s41467-020-16968-1

Laporte, P., Lepage, A., Fournier, J., Catrice, O., Moreau, S., Jardinaud, M.-F., et al. (2014). The CCAAT box-binding transcription factor NF-YA1 controls rhizobial infection. *Journal of Experimental Botany* 65, 481–494. doi: 10.1093/jxb/ert392

Laporte, P., Satiat-Jeunemaître, B., Velasco, I., Csorba, T., Van de Velde, W., Campalans, A., 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.

Latch, J. N., and Margolin, W. (1997). Generation of buds, swellings, and branches instead of filaments after blocking the cell cycle of Rhizobium meliloti. *Journal of Bacteriology* 179, 2373–2381. doi: 10.1128/jb.179.7.2373-2381.1997

Limpens, E., Moling, S., Hooiveld, G., Pereira, P. A., Bisseling, T., Becker, J. D., et al. (2013). Cell- and Tissue-Specific Transcriptome Analyses of Medicago truncatula Root Nodules. *PLOS ONE* 8, e64377. doi: 10.1371/journal.pone.0064377

Liu, C.-W., Breakspear, A., Guan, D., Cerri, M. R., Jackson, K., Jiang, S., et al. (2019). NIN Acts as a Network Hub Controlling a Growth Module Required for Rhizobial Infection. *Plant Physiol.* 179, 1704–1722. doi: 10.1104/pp.18.01572

Maróti, G., Downie, J. A., and 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.

Maróti, G., and Kondorosi, É. (2014). Nitrogen-fixing Rhizobium-legume symbiosis: are polyploidy and host peptide-governed symbiont differentiation general principles of endosymbiosis? *Front. Microbiol.* 5. doi: 10.3389/fmicb.2014.00326

Maunoury, N., Kondorosi, A., Kondorosi, E., and Mergaert, P. (2008). "Cell Biology Of Nodule Infection And Development," in *Nitrogen-fixing Leguminous Symbioses*, eds. M. J. Dilworth, E. K. James, J. I. Sprent, and W. E. Newton (Dordrecht: Springer Netherlands), 153–189. doi: 10.1007/978-1-4020-3548-7_6

Maunoury, N., Redondo-Nieto, M., Bourcy, M., Van De Velde, W., Alunni, B., Laporte, P., et al. (2010). Differentiation of Symbiotic Cells and Endosymbionts in Medicago truncatula Nodulation Are Coupled to Two Transcriptome-Switches. *PLoS ONE* 5, e9519. doi: 10.1371/journal.pone.0009519

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

Mergaert, P., Nikovics, K., Kelemen, Z., Maunoury, N., Vaubert, D., Kondorosi, A., et al. (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. doi: 10.1104/pp.102.018192

Mergaert, P., Uchiumi, T., Alunni, B., Evanno, G., Cheron, A., Catrice, O., et al. (2006). Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium* –legume symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5230–5235. doi: 10.1073/pnas.0600912103

Mikuláss, K. R., Nagy, K., Bogos, B., Szegletes, Z., Kovács, E., Farkas, A., et al. (2016). Antimicrobial nodule-specific cysteine-rich peptides disturb the integrity of bacterial outer and inner membranes and cause loss of membrane potential. *Ann Clin Microbiol Antimicrob* 15, 43. doi: 10.1186/s12941-016-0159-8

Moll, K. M., Zhou, P., Ramaraj, T., Fajardo, D., Devitt, N. P., Sadowsky, M. J., et al. (2017). Strategies for optimizing BioNano and Dovetail explored through a second reference quality assembly for the legume model, Medicago truncatula. *BMC Genomics* 18, 578. doi: 10.1186/s12864-017-3971-4

Montiel, J., Downie, J. A., Farkas, A., Bihari, P., Herczeg, R., Bálint, B., et al. (2017). Morphotype of bacteroids in different legumes correlates with the number and type of symbiotic NCR peptides. *Proc. Natl. Acad. Sci. U.S.A.* 114, 5041–5046. doi: 10.1073/pnas.1704217114

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

Morandi, D., Prado, E., Sagan, M., and Duc, G. (2005). Characterisation of new symbiotic Medicago truncatula (Gaertn.) mutants, and phenotypic or genotypic complementary information on previously described mutants. *Mycorrhiza* 15, 283–289. doi: 10.1007/s00572-004-0331-4

Morgan, T. H., Sturtevant, A. H., Muller, H. J., and Bridges, C. B. (1923). *The mechanism of Mendelian heredity*. H. Holt and Company.

Mousavi, S. A., Li, L., Wei, G., Räsänen, L., and Lindström, K. (2016). Evolution and taxonomy of native mesorhizobia nodulating medicinal *Glycyrrhiza* species in China. *Systematic and Applied Microbiology* 39, 260–265. doi: 10.1016/j.syapm.2016.03.009

Mun, J.-H., Kim, D.-J., Choi, H.-K., Gish, J., Debellé, F., Mudge, J., et al. (2006). Distribution of Microsatellites in the Genome of Medicago truncatula: A Resource of Genetic Markers That Integrate Genetic and Physical Maps. *Genetics* 172, 2541–2555. doi: 10.1534/genetics.105.054791

Murray, J. D. (2011). Invasion by Invitation: Rhizobial Infection in Legumes. *MPMI* 24, 631–639. doi: 10.1094/MPMI-08-10-0181

Nagymihály, M., Veluchamy, A., Györgypál, Z., Ariel, F., Jégu, T., Benhamed, 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, 4543–4548. doi: 10.1073/pnas.1704211114

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

Narayana, S. K., Mallick, S., Siegumfeldt, H., and van den Berg, F. (2020). Bacterial Flow Cytometry and Imaging as Potential Process Monitoring Tools for Industrial Biotechnology. *Fermentation* 6, 10. doi: 10.3390/fermentation6010010

Oldroyd, G. E. D., and Downie, J. A. (2004). Calcium, kinases and nodulation signalling in legumes. *Nat Rev Mol Cell Biol* 5, 566–576. doi: 10.1038/nrm1424

Oldroyd, G. E. D., and Downie, J. A. (2008). Coordinating Nodule Morphogenesis with Rhizobial Infection in Legumes. *Annu. Rev. Plant Biol.* 59, 519–546. doi: 10.1146/annurev.arplant.59.032607.092839

Oldroyd, G. E. D., Murray, J. D., Poole, P. S., and Downie, J. A. (2011). The Rules of Engagement in the Legume-Rhizobial Symbiosis. *Annual Review of Genetics* 45, 119–144. doi: 10.1146/annurev-genet-110410-132549

Ovchinnikova, E., Journet, E.-P., Chabaud, M., Cosson, V., Ratet, P., Duc, G., et al. (2011). IPD3 Controls the Formation of Nitrogen-Fixing Symbiosomes in Pea and Medicago Spp. *MPMI* 24, 1333–1344. doi: 10.1094/MPMI-01-11-0013

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

Pecrix, Y., Staton, S. E., Sallet, E., Lelandais-Brière, C., Moreau, S., Carrère, S., et al. (2018). Whole-genome landscape of Medicago truncatula symbiotic genes. *Nature Plants* 4, 1017–1025. doi: 10.1038/s41477-018-0286-7

Penterman, J., Abo, R. P., De Nisco, N. J., Arnold, M. F. F., Longhi, R., Zanda, M., et al. (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. doi: 10.1073/pnas.1400450111

Pérez Guerra, J. C., Coussens, G., De Keyser, A., De Rycke, R., De Bodt, S., Van De Velde, W., et al. (2010). Comparison of Developmental and Stress-Induced Nodule Senescence in Medicago truncatula. *Plant Physiology* 152, 1574–1584. doi: 10.1104/pp.109.151399

Perret, X., Staehelin, C., and Broughton, W. J. (2000). Molecular Basis of Symbiotic Promiscuity. *Microbiol Mol Biol Rev* 64, 180–201. doi: 10.1128/MMBR.64.1.180-201.2000

Pislariu, C. I., Sinharoy, S., Wen, J., Murray, J. D., Ratet, P., and Udvardi, M. K. (2015). Retrotransposon (Tnt1)-Insertion Mutagenesis in Medicago as a Tool for Genetic Dissection of Symbiosis in Legumes. *Biological Nitrogen Fixation*, 837–854.

Postgate, J. R. (1982). *The fundamentals of nitrogen fixation*. Cambridge: Univ. Pr.

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

Reeve, W., Chain, P., O'Hara, G., Ardley, J., Nandesena, K., Bräu, L., et al. (2010). Complete genome sequence of the Medicago microsymbiont Ensifer (Sinorhizobium) medicae strain WSM419. *Stand in Genomic Sci* 2, 77–86. doi: 10.4056/sigs.43526

Ribeiro, C. W., Baldacci-Cresp, F., Pierre, O., Larousse, M., Benyamina, S., Lambert, A., et al. (2017). Regulation of Differentiation of Nitrogen-Fixing Bacteria by Microsymbiont Targeting of Plant Thioredoxin s1. *Current Biology* 27, 250–256. doi: 10.1016/j.cub.2016.11.013

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

Roux, B., Rodde, N., Jardinaud, M., Timmers, T., Sauviac, L., Cottret, L., 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. doi: 10.1111/tpj.12442

Roy, P., Achom, M., Wilkinson, H., Lagunas, B., and Gifford, M. L. (2020). Symbiotic Outcome Modified by the Diversification from 7 to over 700 Nodule-Specific Cysteine-Rich Peptides. *Genes* 11, 348. doi: 10.3390/genes11040348

Sagan, M., deLarambergue, H., and Morandi, D. (1998). "Genetic Analysis of Symbiosis Mutants in Medicago truncatula," in *Biological Nitrogen Fixation for the 21st Century: Proceedings of the 11th International Congress on Nitrogen Fixation, Institut Pasteur, Paris, France, July 20–25, 1997*, eds. C. Elmerich, A. Kondorosi, and W. E. Newton (Dordrecht: Springer Netherlands), 317–318. doi: 10.1007/978-94-011-5159-7_178

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular cloning: a laboratory manual.* Cold spring harbor laboratory press.

Sambrook, J., and Russell, D. W. (2001). *Molecular cloning: a laboratory manual*., 3rd ed. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* 74, 5463–5467. doi: 10.1073/pnas.74.12.5463

Sankari, S., Babu, V. M. P., Bian, K., Alhhazmi, A., Andorfer, M. C., Avalos, D. M., et al. (2022). A haem-sequestering plant peptide promotes iron uptake in symbiotic bacteria. *Nat Microbiol* 7, 1453–1465. doi: 10.1038/s41564-022-01192-y

Schiessl, K., Lilley, J. L. S., Lee, T., Tamvakis, I., Kohlen, W., Bailey, P. C., et al. (2019). NODULE INCEPTION Recruits the Lateral Root Developmental Program for Symbiotic Nodule Organogenesis in Medicago truncatula. *Current Biology* 29, 3657-3668.e5. doi: 10.1016/j.cub.2019.09.005

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9, 671–675. doi: 10.1038/nmeth.2089

Shabab, M., Arnold, M. F. F., Penterman, J., Wommack, A. J., Bocker, H. T., Price, P. A., et al. (2016). Disulfide cross-linking influences symbiotic activities of nodule peptide NCR247. *Proc. Natl. Acad. Sci. U.S.A.* 113, 10157–10162. doi: 10.1073/pnas.1610724113

Sibley, C. D., MacLellan, S. R., and Finan, T. (2006). The Sinorhizobium meliloti chromosomal origin of replication. *Microbiology* 152, 443–455. doi: 10.1099/mic.0.28455-0

Singh, S., Katzer, K., Lambert, J., Cerri, M., and Parniske, M. (2014). CYCLOPS, A DNA-Binding Transcriptional Activator, Orchestrates Symbiotic Root Nodule Development. *Cell Host & Microbe* 15, 139–152. doi: 10.1016/j.chom.2014.01.011

Smit, P., Limpens, E., Geurts, R., Fedorova, E., Dolgikh, E., Gough, C., et al. (2007). Medicago LYK3, an Entry Receptor in Rhizobial Nodulation Factor Signaling. *Plant Physiology* 145, 183–191. doi: 10.1104/pp.107.100495

Smit, P., Raedts, J., Portyanko, V., Debellé, F., Gough, C., Bisseling, T., et al. (2005). NSP1 of the GRAS Protein Family Is Essential for Rhizobial Nod Factor-Induced Transcription. *Science* 308, 1789–1791. doi: 10.1126/science.1111025

Sprent, J. I. (2008). 60Ma of legume nodulation. What's new? What's changing? *Journal of Experimental Botany* 59, 1081–1084. doi: 10.1093/jxb/erm286

Stacey, G., Libault, M., Brechenmacher, L., Wan, J., and May, G. D. (2006). Genetics and functional genomics of legume nodulation. *Current Opinion in Plant Biology* 9, 110–121. doi: 10.1016/j.pbi.2006.01.005

Starker, C. G., Parra-Colmenares, A. L., Smith, L., Mitra, R. M., and Long, S. R. (2006). Nitrogen Fixation Mutants of Medicago truncatula Fail to Support Plant and Bacterial Symbiotic Gene Expression. *Plant Physiology* 140, 671–680. doi: 10.1104/pp.105.072132

Stonoha-Arther, C., and 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. doi: 10.1016/j.pbi.2018.04.017

Sturtevant, A. H. (1913). The linear arrangement of six sex? linked factors in Drosophila, as shown by their mode of association. *Journal of experimental zoology* 14, 43–59.

Tadege, M., Wen, J., He, J., Tu, H., Kwak, Y., Eschstruth, A., et al. (2008). Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume Medicago truncatula. *The Plant Journal* 54, 335–347. doi: 10.1111/j.1365-313X.2008.03418.x

Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., et al. (2014). An improved genome release (version Mt4.0) for the model legume Medicago truncatula.

Tavormina, P., De Coninck, B., Nikonorova, N., De Smet, I., and Cammue, B. P. A. (2015). The Plant Peptidome: An Expanding Repertoire of Structural Features and Biological Functions. *The Plant Cell* 27, 2095–2118. doi: 10.1105/tpc.15.00440

Timmers, A. C. J., Soupène, E., Auriac, M.-C., de Billy, F., Vasse, J., Boistard, P., et al. (2000). Saprophytic Intracellular Rhizobia in Alfalfa Nodules. *MPMI* 13, 1204–1213. doi: 10.1094/MPMI.2000.13.11.1204

Tirichine, L., de Billy, F., and Huguet, T. (2000). Mtsym6, a Gene ConditioningSinorhizobium Strain-Specific Nitrogen Fixation inMedicago truncatula 1. *Plant Physiology* 123, 845–852. doi: 10.1104/pp.123.3.845

Tiricz, H., Szűcs, A., Farkas, A., Pap, B., Lima, R. M., Maróti, G., et al. (2013). Antimicrobial Nodule-Specific Cysteine-Rich Peptides Induce Membrane Depolarization-Associated Changes in the Transcriptome of Sinorhizobium meliloti. *Appl Environ Microbiol* 79, 6737–6746. doi: 10.1128/AEM.01791-13

Udvardi, M. K., and Day, D. A. (1997). METABOLITE TRANSPORT ACROSS SYMBIOTIC MEMBRANES OF LEGUME NODULES. *Annual Review of Plant Biology* 48, 493–523. doi: 10.1146/annurev.arplant.48.1.493

Vadakattu, G., and Paterson, J. (2006). Free-living bacteria lift soil nitrogen supply. *Farming Ahead* 169, 40.

Van de Velde, W., Guerra, J. C. P., Keyser, A. D., De Rycke, R., Rombauts, S., Maunoury, N., et al. (2006). Aging in Legume Symbiosis. A Molecular View on Nodule Senescence in Medicago truncatula. *Plant Physiology* 141, 711–720. doi: 10.1104/pp.106.078691

Van de Velde, W., Zehirov, G., Szatmari, A., Debreczeny, M., Ishihara, H., Kevei, Z., et al. (2010). Plant Peptides Govern Terminal Differentiation of Bacteria in Symbiosis. *Science* 327, 1122–1126. doi: 10.1126/science.1184057

Vasse, J., De Billy, F., Camut, S., and Truchet, G. (1990). Correlation between ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa nodules. *J Bacteriol* 172, 4295–4306. doi: 10.1128/jb.172.8.4295-4306.1990

Velivelli, S. L. S., Czymmek, K. J., Li, H., Shaw, J. B., Buchko, G. W., and Shah, D. M. (2020). Antifungal symbiotic peptide NCR044 exhibits unique structure and multifaceted mechanisms of action that confer plant protection. *Proceedings of the National Academy of Sciences* 117, 16043–16054. doi: 10.1073/pnas.2003526117

Vernié, T., Kim, J., Frances, L., Ding, Y., Sun, J., Guan, D., et al. (2015). The NIN Transcription Factor Coordinates Diverse Nodulation Programs in Different Tissues of the Medicago truncatula Root. *The Plant Cell* 27, 3410–3424. doi: 10.1105/tpc.15.00461

Vernié, T., Moreau, S., de Billy, F., Plet, J., Combier, J.-P., Rogers, C., et al. (2008). EFD Is an ERF Transcription Factor Involved in the Control of Nodule Number and Differentiation in Medicago truncatula. *The Plant Cell* 20, 2696–2713. doi: 10.1105/tpc.108.059857

Vinardell, J. M., Fedorova, E., Cebolla, A., Kevei, Z., Horvath, G., Kelemen, Z., 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. doi: 10.1105/tpc.014373

Vlassak, K., and Reynders, L. (1979). Factors affecting biological dinitrogen fixation by associate symbiosis in temperate regions.

Wang, D., Griffitts, J., Starker, C., Fedorova, E., Limpens, E., Ivanov, S., et al. (2010). A Nodule-Specific Protein Secretory Pathway Required for Nitrogen-Fixing Symbiosis. *Science* 327, 1126–1129. doi: 10.1126/science.1184096

Wang, Q., Liu, J., Li, H., Yang, S., Körmöczi, P., Kereszt, A., et al. (2018a). Nodule-Specific Cysteine-Rich Peptides Negatively Regulate Nitrogen-Fixing Symbiosis in a Strain-Specific Manner in *Medicago truncatula*. *MPMI* 31, 240–248. doi: 10.1094/MPMI-08-17-0207-R

Wang, Q., Liu, J., and Zhu, H. (2018b). Genetic and Molecular Mechanisms Underlying Symbiotic Specificity in Legume-Rhizobium Interactions. *Front. Plant Sci.* 9, 313. doi: 10.3389/fpls.2018.00313

Weber, M. (1998). Representing genes: classical mapping techniques and the growth of genetical knowledge. *Studies in History and Philosophy of Biological and Biomedical Sciences* 29, 295–315.

Werner, D., Mörschel, E., Kort, R., Mellor, R. B., and Bassarab, S. (1984). Lysis of bacterioids in the vicinity of the host cell nucleus in an ineffective (fix-) root nodule of soybean (Glycine max). *Planta* 162, 8–16. doi: 10.1007/BF00397414

White, J., Prell, J., James, E. K., and Poole, P. (2007). Nutrient Sharing between Symbionts. *Plant Physiology* 144, 604–614. doi: 10.1104/pp.107.097741

Wright, R., Stephens, C., and Shapiro, L. (1997). The CcrM DNA methyltransferase is widespread in the alpha subdivision of proteobacteria, and its essential functions are conserved in Rhizobium meliloti and Caulobacter crescentus. *Journal of bacteriology* 179, 5869–5877.

Xi, J., Chen, Y., Nakashima, J., Wang, S., and Chen, R. (2013). Medicago truncatula esn1 Defines a Genetic Locus Involved in Nodule Senescence and Symbiotic Nitrogen Fixation. *MPMI* 26, 893–902. doi: 10.1094/MPMI-02-13-0043-R

Xiao, T. T., Schilderink, S., Moling, S., Deinum, E. E., Kondorosi, E., Franssen, H., et al. (2014). Fate map of Medicago truncatula root nodules. *Development* 141, 3517–3528. doi: 10.1242/dev.110775

Yang, S., Wang, Q., Fedorova, E., Liu, J., Qin, Q., Zheng, Q., et al. (2017). Microsymbiont discrimination mediated by a host-secreted peptide in *Medicago truncatula*. *Proc. Natl. Acad. Sci. U.S.A.* 114, 6848–6853. doi: 10.1073/pnas.1700460114

Young, N. D., Debellé, F., Oldroyd, G. E. D., Geurts, R., Cannon, S. B., Udvardi, M. K., et al. (2011). The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* 480, 520–524. doi: 10.1038/nature10625

12. APPENDICES

Supplementary Table 1. Genetic markers used for genetic mapping or allelism test

Genetic Marker ID	Orientation	Genetic Marker Sequence
h2-72a11aF	Forward (5' to 3')	CTCCCACCAGGCTCTGATAA
h2-72a11aR	Reverse (5' to 3')	GCCCGCATCTCTAAATCAAA
4g036765F	Forward (5' to 3')	TTTCCAGAACACACAAAACCTTAT
4g036765R	Reverse (5' to 3')	TCTTCACCACTGTAGAAAAATTGC
AC136506F	Forward (5' to 3')	GACTTTTCACCTCCTATTTTAC
AC136506R	Reverse (5' to 3')	ACCTTGTTGTTCTTTCTTTTTA
4g040420F	Forward (5' to 3')	ACCATCCTCTATAGCTTGAATCCT
4g040420R	Reverse (5' to 3')	TGACTTAATTTCATTGACCTTCGG
Chr4SSRF4	Forward (5' to 3')	AGTGATTTGTGACCCATTTTTCCT
Chr4SSRR4	Reverse (5' to 3')	AAGTTGTGAGTTTGTCCAAGTCCA
4g045577F	Forward (5' to 3')	AGATCGACTTGTATCCTCGGAA
4g045577R	Reverse (5' to 3')	TGGTTTAGTATCGGTGATTCAGAG
Scaffol003F	Forward (5' to 3')	GGCAGTCCCGGAACAAATTT
Scaffol003R	Reverse (5' to 3')	TCTGCCAATAACTGTTCAGTGT
Scaffold733F	Forward (5' to 3')	CCGCCTCTCCTACTCTCAAA
Scaffold733R	Reverse (5' to 3')	AGAAAGAGAGGGCTTGAAATGA
Scaffold733F3	Forward (5' to 3')	CAACCATATTGTCACCCCTTT
Scaffold733R3	Reverse (5' to 3')	TTTTACAAGTCTGGGCTCAAATC
Scaffold005F4	Forward (5' to 3')	CAAATCTCCGGACGAAGAAA
Scaffold005R4	Reverse (5' to 3')	TCAAAGTTTCCGACAAGAAGTT
Scaffold005F3	Forward (5' to 3')	ACCACAACACGCCAACATTA
Scaffold005R3	Reverse (5' to 3')	AACTCAGTTGGGTTGGCAAG
Scaffold005F	Forward (5' to 3')	CGATCTCAAATTTCTCACTTCTCC
Scaffold005R	Reverse (5' to 3')	CAAATGGTATAATTGACAAATGGTACT
Scaffold005F1	Forward (5' to 3')	GGCTTGTGCTTATCCACCAT
Scaffold005R1	Reverse (5' to 3')	TCGTATCAGGGCTTCACAGA
4g043640F	Forward (5' to 3')	GATGTCAAGCATTCAAGGTAGGA
4g043640R	Reverse (5' to 3')	AAGCAATGTCAACAAACCCAAAT
4g044165F	Forward (5' to 3')	CCCCAATCCAAATCCTCCTCTT
4g044165R	Reverse (5' to 3')	CACCAACGACCACCTTTCAC
Scaffold963F	Forward (5' to 3')	AACAAACTAATCTAAATCAATGAG
Scaffold963R	Reverse (5' to 3')	TATTCAAATGATAAAGAAAACCTT
4g037470F	Forward (5' to 3')	TGAGTAATTCAGCGTTCAGGA
4g037470R	Reverse (5' to 3')	ACTTAAGGGACCAAAAGAGCA
002F03F	Forward (5' to 3')	TGGCTTTCTTACGATCCACA
002F03R	Reverse (5' to 3')	CAACTTGGTCACTCAGCATT
Mtb267F	Forward (5' to 3')	GCGTTAGCATGGGTTAATGG
Mtb267R	Reverse (5' to 3')	GCAAACAATGGTGTGTCGAG

Supplementary Table.2. Primers used for detection of the deletion borders and deleted genes in the genomic regions of *Mtsym20 and Mtsym19* mutants

Primer	Orientation	Primer Sequence	Gene ID
Medtr0003te0510F	Forward (5' to 3')	GACAGGTGGCTTTGGAGATT	Medtr0003te0510
Medtr0003te0510R	Reverse (5' to 3')	CATTAGGGATTACCGAGCAGAC	
Chr4g0020281F	Forward (5' to 3')	TCCAAACGGAGCCTTAATCT	MtrunA17Chr4g0020281
Chr4g0020281R	Reverse (5' to 3')	GCAAAATAGTGAAAGCCCGA	
Chr4g0020291F	Forward (5' to 3')	CAAATTTACAAGGACGTTTTGCA	MtrunA17Chr4g0020291
Chr4g0020291R	Reverse (5' to 3')	AGGATACACGCATAAGAGACATTTA	
Chr4g0020301F	Forward (5' to 3')	AGGTATCTTTCGCCCAGA	MtrunA17Chr4g0020301
Chr4g0020301R	Reverse (5' to 3')	CGGGGTGTTACATAATACATATAC	
Chr4g0020311F	Forward (5' to 3')	TGGTGACAAGCATCCTAAGAC	MtrunA17Chr4g0020311
Chr4g0020311R	Reverse (5' to 3')	AAGCCCTAACATCAAGCCTAC	
Chr4g0020321F	Forward (5' to 3')	GGCGAGGGAAATTGGTAGTG	MtrunA17Chr4g0020321
Chr4g0020321R	Reverse (5' to 3')	TCAAGAGCTGGTCGGAGAA	
Chr4g0020331F	Forward (5' to 3')	CAACAATGGTATTCACCGATTCC	MtrunA17Chr4g0020331
Chr4g0020331R	Reverse (5' to 3')	CGACCCGACCCTTAACATAAA	

Supplementary Table.3. List of oligonucleotides used for assembly of constructs

Single site gateway (GW) primers used for complementation assays		
NCR-new35-GW	Forward (5' to 3')	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGTTGTAGAAGGAAGAGGAGGAG
	Reverse (5' to 3')	GGGGACCACTTTGTACAAGAAAGCTGGGTCCGTTTTGACAGCTCTAACAGTA
Multisite gateway (MsGW) used for sub cellular localization assays		
pNCR-new35-MsGW	Forward (5' to 3')	GGGGACAACTTTGTATAGAAAAGTTGGGTGTTGTAGAAGGAAGAGGAGGAG
	Reverse (5' to 3')	GGGGACTGCTTTTTTGTACAAACTTGGGTAAAACAATTTTTTTAATCTCTTTATAAGTAAAC
NCR-new35-coding-MsGW	Forward (5' to 3')	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGATGCAAAGGAAGAAAAATATGGC
	Reverse (5' to 3')	GGGGACCACTTTGTACAAGAAAGCTGGGTGAATAATAAACATTTCAGGTTGTCA
NCR-new35-3UTR-MsGW	Forward (5' to 3')	GGGGACAGCTTTCTTGTACAAAGTGGGGATAATAAGTTACTTTAATAAGTAA
	Reverse (5' to 3')	GGGGACAACTTTGTATAATAAAGTTGGCCGTTTTGACAGCTCTAACAGTA
Single site gateway (GW) primers used for promoter activity assays		
pNCR-new35-GUS-GWF	Forward (5' to 3')	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGTTGTAGAAGGAAGAGGAGGAG
pNCR-new35-GUS-GWR	Reverse (5' to 3')	GGGGACCACTTTGTACAAGAAAGCTGGGTGTAAAACAATTTTTTTAATCTCTTTATAAGTAAAC
pNCR343-GUS-GWF	Forward (5' to 3')	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGCAACATTTTTCTCTCTTATGCCTTT
pNCR343-GUS-GWR	Reverse (5' to 3')	GGGGACCACTTTGTACAAGAAAGCTGGGTATTTGTTTTCTTTGTTTTTGTAACAAATAGG
pNCR-169-GUS-GWF	Forward (5' to 3')	GGGGACAAGTTTGTACAAAAAAGCAGGCTAAAACCGCTACAATCAGATCG
pNCR-169-GUS-GWR	Reverse (5' to 3')	GGGGACCACTTTGTACAAGAAAGCTGGGTTGTTTTTCCCCTTTGCGTGAAA
pNCR-211-GUS-GWF	Forward (5' to 3')	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGGAGTGTAGGGGTGATGTTGTT
pNCR-211-GUS-GWR	Reverse (5' to 3')	GGGGACCACTTTGTACAAGAAAGCTGGGTGTCTTTTTTAACTTTGTATATAAC
List of oligonucleotides used in gene targeting experiments to generate gene constructs, and sequence the targeted regions		
New35_gF	Forward (5' to 3')	CTTGGAAGGAAAGGAGGACCACC
New35_gR	Reverse (5' to 3')	AAACGGTGGTCCTCCTTTCCTTC
New35_ngs_F	Forward (5' to 3')	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGtGGCTCACATTCTCATGTTTGT
New35_ngs_R	Reverse (5' to 3')	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGcAAGGAATGTAAGCTGCATTTCA

Supplementary Table 4. The target site predictions performed by the CRISPOR online tool

mismatch Position	mismatchCount	mitOfftargetScore	cfdOfftargetScore	Annotation
99forw	GGAAGGAAAGGAGGACCACCAGG	GGAAGGAAAAGAGGAAAACCTGG	.....*.....**...	intergenic:mRNA:MtrunA17Chr3g0105401-mRNA:MtrunA17Chr3g0105411
99forw	GGAAGGAAAGGAGGACCACCAGG	GGAAGAAAAGTGGGACCATCTGG	.....*....**.....*.	intergenic:mRNA:MtrunA17Chr3g0146071-mRNA:MtrunA17Chr3g0146081
99forw	GGAAGGAAAGGAGGACCACCAGG	GTACGGAAAGGAGAACCATCAGG	.*.*.....*....*.	intergenic:mRNA:MtrunA17Chr4g0069401-mRNA:MtrunA17Chr4g0069411
99forw	GGAAGGAAAGGAGGACCACCAGG	GGAAGGAAAGAAGGACAAATTGG	.....*.....*.**	intron:mRNA:MtrunA17Chr2g0315071
99forw	GGAAGGAAAGGAGGACCACCAGG	GGAAGGGAAGGAGTACCAAAAGG	.....*.....*....**	exon:mRNA:MtrunA17Chr5g0442801
99forw	GGAAGGAAAGGAGGACCACCAGG	GGAAGGGAAGGAGTACCAAAAGG	.....*.....*....**	exon:mRNA:MtrunA17Chr5g0442781
99forw	GGAAGGAAAGGAGGACCACCAGG	GGAAGGAGAGGAGGACATGCAGG	.....*.....***.	intergenic:mRNA:MtrunA17Chr3g0138981-mRNA:MtrunA17Chr3g0138991/mRNA:MtrunA17Chr3g0139001
99forw	GGAAGGAAAGGAGGACCACCAGG	GGAACGAAAGGAGGATCAAGGGG	....*.....*..**	exon:mRNA:MtrunA17Chr8g0353081
99forw	GGAAGGAAAGGAGGACCACCAGG	GGAAGAAAAGGAGGAGGACGCGG	.....*.....**..*	exon:mRNA:MtrunA17Chr7g0217261

Supplementary Table 5. Genes in the genomic region of *Mtsym20* according to *de novo* assembly. Gene IDs used in the genome release of *M.truncatula* JCVI.4. and MtrunA17r5.0 are presented.

Genetic Marker	JCVI.4 Reconstruction	MtrunA17r5.0	Gene Function (MtrunA17r5.0)
4g040420		MtrunA17Chr4g0019791	Putative Cystatin domain-containing protein
		MtrunA17Chr4g0019801	Putative non-specific serine/threonine protein kinase
		MtrunA17Chr4g0019811	Putative histone deacetylase
		MtrunA17Chr4g0019821	Putative nucleic acid-binding protein
		MtrunA17Chr4g0019841	Putative non-specific serine/threonine protein kin
		MtrunA17Chr4g0019851	Putative tetratricopeptide-like helical domain-con
		MtrunA17Chr4g0019861	Putative non-specific serine/threonine protein kin
		MtrunA17Chr4g0019881	hypothetical protein
		MtrunA17Chr4g0019891	hypothetical protein
		MtrunA17Chr4g0019911	Putative Ulp1 protease family catalytic domain-con
		MtrunA17Chr4g0019921	hypothetical protein
		MtrunA17Chr4g0019931	Putative nepenthesin
		MtrunA17Chr4g0019941	hypothetical protein
		MtrunA17Chr4g0019961	Putative F-box domain-containing protein
		MtrunA17Chr4g0019971	Putative Cystatin domain-containing protein
		MtrunA17Chr4g0019981	Putative leucine-rich repeat domain, L domain-cont
		MtrunA17Chr4g0019991	Putative F-box domain-containing protein
		MtrunA17Chr4g0020001	Putative tetratricopeptide-like helical domain, DY
		MtrunA17Chr4g0020011	Putative beta-amyrin synthase
		MtrunA17Chr4g0020021	hypothetical protein
		MtrunA17Chr4g0020031	Putative beta-amyrin synthase
		MtrunA17Chr4g0020041	Putative cytochrome P450
		MtrunA17Chr4g0020051	Putative ribosomal protein L1
		MtrunA17Chr4g0020061	hypothetical protein
		MtrunA17Chr4g0020071	hypothetical protein
		MtrunA17Chr4g0020081	Putative protein kinase RLK-Pelle-DLSV family
Chr4SSRF4R4	Medtr4g040480.1	MtrunA17Chr4g0020081	Putative protein kinase RLK-Pelle-DLSV family
	Medtr4te040490.1	MtrunA17Chr4R0093400	gene
	Medtr4te040500.1		gene

	Medtr4g040530.1		gene
	Medtr4te040540.1		gene
	Medtr4te040550.1	MtrunA17Chr4R0093410	gene
	Medtr4g040560.1		gene
	Medtr4te040590.1		gene
	Medtr4te040600.1	MtrunA17Chr4R0093530	gene
	Medtr4te044650.1	MtrunA17Chr4R0093570	gene
		MtrunA17Chr4g0020091	gene
		MtrunA17Chr4g0020092	gene
	Medtr4g044663.1	MtrunA17Chr4g0020101	Putative protein kinase RLK-Pelle-CR4L family
	Medtr4g044667.1		gene
	Medtr4te044670.1		gene
4g045577	Medtr4g045577.1	MtrunA17Chr4g0020111	Putative transcription factor WD40-like family
	Medtr4te045583.1		gene
		MtrunA17Chr4g0020121	snoR71
	Medtr0816te0010.1		gene
	Medtr4te046610.1	MtrunA17Chr4R0094100	gene
	Medtr4g046607.1	MtrunA17Chr4R0094110	gene
	Medtr4te046620.1	MtrunA17Chr4R0094200	gene
	Medtr4te046623.1		gene
	Medtr4te046630.1	MtrunA17Chr4R0094270	gene
		MtrunA17Chr4g0020131	Putative Cystatin domain-containing protein
		MtrunA17Chr4g0020141	hypothetical protein
	Medtr3g009360.1		gene
		MtrunA17Chr4g0020151	gene
	Medtr0003te0700.1	MtrunA17Chr4R0095040	gene
	Medtr0003s0690.1	MtrunA17Chr4g0020161	Putative RNA polymerase Rpc34
	Medtr0003te0680.1		gene
	Medtr0518te0010.1		gene
	Medtr0518te0020.1	MtrunA17Chr4R0095270	gene
	Medtr0518te0030.1	MtrunA17Chr4R0095290	gene
	Medtr0003s0660.1	MtrunA17Chr4g0020171	Putative S-adenosyl-l-methionine decarboxylase leader peptide
		MtrunA17Chr4g0020181	Putative adenosylmethionine decarboxylase
	Medtr0003s0650.1	MtrunA17Chr4g0020191	hypothetical protein
	Medtr0003te0640.1		gene
	Medtr0003s0630.1	MtrunA17Chr4g0020201	Putative lactoylglutathione lyase
	Medtr0003s0620.1	MtrunA17Chr4g0020211	Putative endothelin-converting enzyme 1
	Medtr0003s0610.1		gene
	Medtr0003te0600.1		gene
	Medtr0003s0590.1	MtrunA17Chr4g0020221	Putative transcription factor MADS-MIKC family

	Medtr0003s0580.1	MtrunA17Chr4g0020231	hypothetical protein
		MtrunA17Chr4g1016181	gene
	Medtr0003s0560.1	MtrunA17Chr4g0020241	Putative carboxypeptidase D
	Medtr0003s0540.1	MtrunA17Chr4g0020251	Putative transferase
Scaffold003FR		MtrunA17Chr4g0020261	tRNA-Pseudo
	Medtr0003s0530.1	MtrunA17Chr4g0020271	Putative strictosidine synthase transcription factor WD40-like f
	Medtr0003te0510.1		gene
	Medtr0512s0010.1	MtrunA17Chr4g0020281	hypothetical protein
	Medtr0512s0030.1	MtrunA17Chr4g0020291	Putative Late nodulin
		MtrunA17Chr4g1016193	gene
		MtrunA17Chr4g0020301	hypothetical protein
	Medtr0003s0500.1	MtrunA17Chr4g0020311	Putative long-chain-alcohol O-fatty-acyltransferase
	Medtr0003s0480.1	MtrunA17Chr4g0020321	Putative GIY-YIG nuclease superfamily
	Medtr0003s0470.1	MtrunA17Chr4g0020331	Putative equilibrative nucleoside transporter
		MtrunA17Chr4g0020341	hypothetical protein
	Medtr0003s0460.1	MtrunA17Chr4g0020351	Putative structural maintenance of chromosomes protein
	Medtr0003te0450.1	MtrunA17Chr4R0096320	gene
	Medtr0077s0060.1	MtrunA17Chr4R0096360	gene
		MtrunA17Chr4g1016196	gene
	Medtr0003te0420.1		gene
	Medtr0077s0060.1	MtrunA17Chr4R0096440	gene
	Medtr0077te0070.1		gene
	Medtr0003s0410.1		gene
	Medtr0077te0050.1	MtrunA17Chr4R0096480	gene
	Medtr0077s0040.1	MtrunA17Chr4R0096560	gene
	Medtr0003s0290.1	MtrunA17Chr4R0096670	gene
	Medtr0003s0280.1		gene
	Medtr0003te0270.1	MtrunA17Chr4R0096700	gene
	Medtr0003s0260.1		gene
	Medtr0003s0250.1		gene
	Medtr0003s0210.1	MtrunA17Chr4g0020361	Putative Late nodulin
Scaffold733FR	Medtr0003s0190.1	MtrunA17Chr4g0020371	hypothetical protein
	Medtr0003s0180.1	MtrunA17Chr4g0020381	Putative Late nodulin
		MtrunA17Chr4g0020391	hypothetical protein
	Medtr0003s0170.1	MtrunA17Chr4g0020401	hypothetical protein
		MtrunA17Chr4g0020411	Putative P-loop containing nucleoside triphosphate hydrolase
	Medtr0003s0160.1	MtrunA17Chr4R0096840	gene
Scaffold733F3R3	Medtr0003s0130.1	MtrunA17Chr4g0020421	Putative Late nodulin
	Medtr0003te0140.1	MtrunA17Chr4R0096880	gene
	Medtr0003te0110.1	MtrunA17Chr4R0096900	gene

Published papers in referred journals

- Horváth, B*., **Güngör, B*.,** Tóth, M., Domonkos, Á., Ayaydin, F., Saifi, F., ... & Kaló, P. (2023). The *Medicago truncatula* nodule-specific cysteine-rich peptides, NCR343 and NCR-new35 are required for the maintenance of rhizobia in nitrogen-fixing nodules. *New Phytologist*, *239*(5), 1974-1988.
  IF: 9.4
- **Güngör, B*.,** Biró, J. B*., Domonkos, Á., Horváth, B., & Kaló, P. (2023). Targeted mutagenesis of *Medicago truncatula Nodule-specific Cysteine-Rich* (*NCR*) genes using the *Agrobacterium rhizogenes*-mediated CRISPR/Cas9 system. *Scientific Reports*, *13*(1), 20676.
  IF: 4.6
- Van Velzen, R., Holmer, R., Bu, F., Rutten, L., van Zeijl, A., Liu, W., Santuari, L., Cao, Q., Sharma, T., Shen, D., Roswanjaya, Y., Wardhani, T.A.K., Seifi Kalhor, M., Jansen, J., Van den Hoogen, J., **Güngör, B.,** Hartog, M., Hontelez,J., Verver, J., Yang, W.C., Schijlen, E., Repin, R., Schilthuizen, M., Schranz, M.E., Heidstra, R., Miyata, K., Fedorova, E., Kohlen, W., Bisseling, T., Smit, S., Geurts, R. (2018). Comparative genomics of the nonlegume Parasponia reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proceedings of the National Academy of Sciences*, *115*(20), E4700-E4709.
  IF: 9.580

Conferences Presentations Related to Thesis

- **Berivan Güngör**, János Barnabás Biró, Ágota Domonkos, Beatrix Horváth, Péter Kaló. Targeted mutagenesis of *Medicago truncatula Nodule-specific Cysteine-Rich (NCR)* genes using the *Agrobacterium rhizogenes*-mediated CRISPR/Cas9 system,
  Institute of Genetics and Biotechnology Days, Gödöllő, Hungary, Nov, 2023

- **B. Güngör**, B. Horvath, M. Toth , A. Domonkos , Z. Szabó et al. Identification of *NCR* genes essential for symbiotic nitrogen fixation in *Medicago truncatula* using forward and reverse genetic approaches, *Internal Scientific Conference of Agricultural Biotechnology Institute, Gödöllő, Hungary, Nov, 2019*
  Special Prize of Greiner Bio-One GmbH

Poster Presentations Related to Thesis

- J. B. Biró*, **B. Güngör***, Á. Domonkos, B. Horváth, P. Kaló. Targeted Mutagenesis of *Medicago truncatula Nodule-specific Cysteine-rich* (*NCR*) Genes Using the *Agrobacterium Rhizogenes*-Mediated CRISPR/Cas9 System**,**
  15th European Nitrogen Fixation Conference, Naples, Italy, Aug 2023
- J. B. Biró, K. Kristóf, Z. Szegletes, **B. Güngör**, P. Kaló, A. Kereszt. A simplified Golden Gate cloning system
  XXI. "Genetics Workshops in Hungary" mini conference, Szeged, Hungary, Sep 2022

- B. Horváth*, **B. Güngör***, M. Tóth , A. Domonkos , Z. Szabó et al. Comparative analysis of *Medicago truncatula* mutants deficient in NCR peptides essential for symbiotic nitrogen fixation,
  9. International Conference on Legume Genetics and Genomics (ICLGG 2019), Let's harness the potential of legumes, Dijon, France, May 2019

13. ACKNOWLEDGMENT

I would like to express my appreciation to Stipendium Hungaricum board members for providing the financial support necessary to complete my PhD studies, and to all members of MATE Institute of Genetics and Biotechnology for supplying the necessary materials and conditions for a conducive work environment.

I would like to express my gratitude for my supervisor Dr. Péter Kaló for providing and supervising my research projects, for his professional advices and for reviewing the dissertation. I owe thanks to Beatrix Horváth for her guidance in the laboratory, Zoltán Szabó for his collaborations regarding data analysis, to Ágota Domonkos for her support regarding RNA-related experiments, to Hajnalka Csákány for her technical assistance, Zoltán Tóth and Ferhan Ayaydın for their support in imaging techniques.

I would also like to say thank you to my colleges Mónika Tünde Tóth and Farheen Saifi for their support and friendship.

Finally, I would like to thank my family for their support despite the distances, and my husband, with whom I have walked this journey hand in hand. I dedicate this dissertation to my son Ármin Robin Biró as an expression of my gratitude for his presence.