Lotus japonicus natural diversity unveils genetic components to host rhizobia inside plant cells

Dissertation zur Erlangung des Doktorgrades der
Naturwissenschaften
Doctor rerum naturalium (Dr. rer. nat.) an der Fakultät für Biologie der
Ludwig-Maximilians-Universität München

vorgelegt von
Oscar Rafael Espejel Venado
München, August 2022

1. Gutachter: PD Dr. Macarena Marín

2. Gutachter: Prof. Dr. Wolfgang Frank

Tag der Abgabe: 13 Juni 2022

Tag der mündlichen Prüfung: 9 August 2022

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

München, den 08.08.2022

Oscar Rafael Espejel Venado

Erklärung

Hiermit erkläre ich, dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist. Ich habe nicht versucht, anderweitig eine Dissertation einzureichen oder mich einer Doktorprüfung zu unterziehen.

München, den 08.08.2022

Oscar Rafael Espejel Venado

Table of Contents

Abbreviations

°C	Degree Celsius
4CL	4-Coumarate-CoA ligase 1 and paralogs
ABCG	ABC transporter G
ABHE	Alpha/beta hydrolase
Ah or *A. hypogaea*	*Arachis hypogaea*
ALA	5-Aaminolevulinic acid
AM	Arbuscular mycorrhizal
ANOVA	Analysis of variance
APN1	Aspartic peptidase nodule-induced 1
ARA	Acetylene reduction assay
ARP3	Actin related protein
ARPC1	Actin-related protein component1
ASL	Asymmetric leaves 2-Like
At	*Arabidopsis thaliana*
BNF	Biological nitrogen fixation
bp	Base pair
C3H1	p-Coumaroyl shikimate 30-hydroxylase
C4H	Cinnamate 4-hydroxylase and paralogs
CA-OMT	Caffeoyl-CoA *O*-methyltransferase
CAD	Cinnamoyl alcohol dehydrogenase
CASP	Casparian strip domain membrane
CASPL	CASP-like
CCaMK	Calcium/calmodium dependent protein kinase
CCR	Cinnamoyl-CoA reductase
CK	Cytokinin
CKX3	Cytokinin oxidase/dehydrogenase3
CLSM	Confocal laser scanning microscopy
CO_2	Carbon dioxide
COMT	Caffeic acid *O*-methyltransferase
CRISPR	Clustered regularly interspaced short palindromic repeats
CS	Casparian strips
cv.	Cultivar
CYP735a	Cytochrome P450 monooxygenase 735a
CYP86A1	Cytochrome 86A1
CYP86B1	Cytochrome 86B1
DE	Differential expression
DEG	Differentially expressed genes
DMI1	Does not make Infection 1
DMI2	Does not make Infection 2
DMI3	Does not make Infection 3
dNRE	Double NRE
dpi	Days post inoculation

1

*Ds*Red	*Discosoma* sp. red fluorescent protein
EMS	Ethyl methanesulfonate
ENOD	Early nodulin
EPS	Exopolysaccharide
ERN1	Ethylene responsive factor required for nodulation 1
ESB1	Enhanced suberin 1
ETI	Effector-triggered immunity
FACT	Fatty alcohol:caffeoyl-CoA caffeoyl transferase
FAD	Fatty acid desaturase
FaFaCuRo	Fabales, Fagales, Cucurbitales, and Rosales
FAR	Fatty acyl-CoA reductase
GC-FID	Gas chromatography – flame ionization detection
gDNA	Genomic DNA
GFP	Green fluorescent protein
Glb	Globulins
GO	Gene ontology
GPAT	Glycerol-3-phosphate acyltransferase
gRNA	Guide RNA
GUS	β-glucuronidase
Hbs	Non-symbiotic leghemoglobins
HCT	Hydroxycinnamoyl-CoA transferase
HMGR	3-hydroxy-3-methyglutaryl-CoA reductase
IPD3	Interacting protein of DMI3
IPT2	Isopentenyl transferase gene 2
IPT4	Isopentenyl transferase gene 4
IT	Infection thread
KCS	ß-ketoacyl-CoA synthase
Km	Michaelis-Menten constant
La	*Lachnospiraceae* sp.
LAC	Laccase
LACS	Long-chain acyl-CoA synthetase
Lb	Leghemoglobins
LB	Luria-Bertani broth
LHK1	Lotus histidine kinase
LIN	Lumpy infection
Lj or *L. japonicus*	*Lotus japonicus*
LOG2	Lonely guy 2
LOG4	Lonely guy 4
LORE1	Lotus retrotransposon 1
LPS	Lipopolysaccharides
LYK3	LysM receptor kinase 3
LysM	Lysin motif
M	Molar
m	Milli

MCA8	Calcium pump ATPases
MED5a	Mediator of RNA polymerase II transcription
min	Minutes
Ml or *M. loti*	*Mesorhizobium loti*
Mt or *M. truncatula*	*Medicago truncatula*
MYB	MYB transcription factor
n	Nano
N_2	Atmospheric nitrogen
N_2O	Nitrous oxide
NAC	NAC transcription factor
Nap1	Nck-associated protein1
nd	Nodule distance
NF	Nodulation factor
NF-Y	Nuclear factor Y
NF-*YA1*	Nuclear factor-Y subunit A-1
NF-YB1	Nuclear factor-Y subunit B-1
NFR1	Nod factor receptor 1
NFR5	Nod factor receptor 5
NGS	Next generation sequencing
NIN	Nodule inception
NLP	NIN-like protein
NPL	Nodulation pectate lyases
NRE	Nitrate-responsive element
NSP1	Nodule signaling pathway 1
NSP2	Nodule signaling pathway 2
NUP	Nucleoporin
O_2	Atmospheric oxygen
OD	Optical density
PAL	Phenylalanine ammonia lyase
PAM	Protospacer-adjacent motif
PCR	Polymerase chain reaction
PI3K	Phosphatidylinositol 3 kinase
Pir1	p53 inducible RNA
PPP	Phenylpropanoid pathway
PR	Permeability ratio
PRX	Peroxidase
Ps	*Pisum sativum*
PUB1	Plant U-box protein 1
RBOH	Respiratory burst oxidase homologs
RIT	Required for infection thread
Rl or *R. leguminosarum*	*Rhizobium leguminosarum*
RNAi	RNA interference
RNS	Root nodule symbiosis
ROS	Reactive oxygen species

3

RPG	Rhizobium-directed polar growth
RT-qPCR	Quantitative reverse transcription polymerase chain reaction
s	Seconds
SCARN	Suppressor of cAMP receptor defect-Nodulation
SNARE	N-ethylmaleimide-sensitive factor attachment protein receptor
SNF	Symbiotic nitrogen fixation
snf	Spontaneous nodule formation
Sr or *S. rostratata*	*Sesbania rostrata*
SST1	Symbiotic sulfate transporter 1
STY	Short internodes/stylish
SYMREM	Symbiotic remorin
SYMRK	Symbiosis receptor-like kinase
SYP71	Syntaxin of plant71
td	Toluidine blue distance
ttCas12a	Temperature-tolerant Cas12a
TY	Tryptone yeast
V	Volt
VAMP	Vesicle-associated membrane protein
WGCNA	Weighted correlation network analysis
WT	Wild-type
YEB	Yeast extract broth
YMB	Yeast mannitol broth
μ	Micro

List of publications

Publication I

Venado RE, Liang J, Marín M (2020). 'Chapter Four - Rhizobia infection, a journey to the inside of plant cells' in Frendo P, Frugier F, Masson-Boivin C (ed.) *Regulation of Nitrogen-Fixing Symbioses in Legumes*, 94, pp. 97-118, Academic Press

Manuscript I

Venado RE, Wange LE, Pinnau F, Shen D, Grube Andersen T, Enard W, Marín M. Tissue-specific regulation of fatty acid synthesis genes controlling oxygen permeation into *Lotus japonicus* nodules. (In revision)

Manuscript II

Venado RE, Akindele O, Marín M. Characterization of Casparian strip domain protein-like family in *Lotus japonicus*. (In preparation)

Summary

Legume plants form a symbiosis with nitrogen-fixing rhizobia bacteria. This symbiosis occurs within cells of specialized root organs called nodules in which a bidirectional nutrient exchange between the symbionts takes place. During this process, legumes obtain reduced nitrogenated compounds whereas rhizobia receive carbon compounds derived from plant photosynthesis. Therefore, there has been great interest in unveiling the genetic architecture of this phenomenon to reduce the use of inorganic nitrogen fertilizers in agriculture.

The study of root nodule symbiosis has determined the function of nearly 200 genes in the last 20 years. This has been possible through forward and reverse genetic screenings and the development of various other genomic tools in model organisms such as *Lotus japonicus* (*L. japonicus*). Gene discovery using conventional screenings has almost reached its limit. Thus, new approaches are needed to uncover new genetic players, particularly genes involved in the plant cell mechanisms required to host bacteria inside the plant cells. Molecular understanding of the tissue and cellular adaptations required to host rhizobia inside nodules remains extremely limited, because of the difficulty in disconnecting nodule formation from infection. These modifications provide an adequate environment for effective nitrogen fixation. A key modification is the development of a cell layer that surrounds the inner cells and reduces the amount of oxygen that enters these cells, thus, protecting the bacterial nitrogenase, which is oxygen-labile. However, the genetic components to form this barrier are still unknown.

Recently, Liang et al. characterized a system that opens up the possibility of studying the mechanism of internalization of the bacteria inside the plant cells and by consequence the bacterial accommodation. This system describes the interaction between a *Lotus* species and a subcompatible rhizobium strain, *Rhizobium leguminosarum* (*R. leguminosarum*) Norway. In this system, the bacteria enter the plant cell without the aid of specialized structures in the root hair called infection threads but rather from an alternative mechanism of infection independent of these threads. In this work, this system was expanded by exploring the natural diversity of different *L. japonicus* accessions in combination with *R. leguminosarum* Norway. By using this approach, the nodule organogenesis and infection programs were uncoupled, as it induces nodules that remain uninfected in some *L. japonicus* accessions. Comparative transcriptomic analysis of infected and uninfected nodules yielded 167 differentially regulated genes. Among these, genes with functions associated with plant barrier formation were specifically upregulated in infected nodules. Among the genes uncovered, two *fatty acyl-CoA reductases* (*FAR*s) genes that are involved

in the production of cuticular waxes, seed coat, bundle sheath, bark tissues and two putative scaffold proteins Casparian Strip Domain-Like Proteins (CASPLs) were studied. It was hypothesized that these genes are involved in the formation of a cellular barrier that controls the delicate oxygen homeostasis in the root nodule.

Spatiotemporal analysis of promoter activity controlling the expression of *FAR*s and *CASPL*s revealed tissue-specific activation in the nodule endodermis and infected cells, respectively. Reverse genetic analysis was performed by investigating two Lotus retrotransposon lines (*LORE1*) in a nodule specific *FAR* and a double mutant in a *CASPL* generated by CRISPR-Cas12a editing. In the first case, mutants compared to wild-type nodules displayed a significant reduction in hydrophobic polyesters in the surrounding cell layer termed nodule endodermis, an increase in their oxygen concentration inside the mutant nodules, and impaired nitrogen fixation activity. This transduced in mutant plants having significantly shorter shoots. These results support a model in which disruption in the composition of the nodule oxygen barrier alters nitrogen fixation. In the second case, infected nodule cells in the double mutant line showed an irregular morphology with an undefined nucleus compared with wild-type cells. This suggests that local cell wall modifications are required to properly accommodate the symbiont. These results pave the way for understanding how plants modify their cell walls locally to host the symbiont.

Introduction

1. The goal of achieving food security and sustainable agriculture

The world population is expected to grow to 9.9 billion over the next 30 years, resulting in the need to increase food production to meet global population demand. This will happen under a scenario where climate change and a reduction in arable land are viewed as major obstacles to reaching this goal (FAO, 2017). Despite these limitations, food production has significantly increased over the last 50 years thanks to the Green Revolution and the adoption of improved crops, irrigation methods, pesticides, and fertilizers. Supplying fields with fertilizers, in particular the nitrogen-based ones, has resulted in a several-fold increase in agricultural production (Bohlool et al., 1992). Plants require reduced nitrogenated compounds as they play important roles in different metabolic processes, including the synthesis of proteins, nucleic acids, vitamins, and chlorophyll (Mcallister et al., 2012). In particular, chlorophyll is essential for photosynthesis as this process provides the energy for plant metabolism, growth and reproduction. Plants are limited by the nitrogen availability in the soil, which can be taken up in the form of nitrate (NO_3^-), nitrite (NO_2^-), and ammonium (NH_4^+). These are easily accessible in inorganic nitrogen fertilizers (Hachiya and Sakakibara, 2017). The standard agricultural practice to use synthetic nitrogen-rich fertilizers has brought about other issues (Rockström et al., 2009; Oldroyd and Dixon, 2014). These includes unequal access to and lack of affordability of nitrogenated fertilizers, and the negative impact on the environment reflected in water and air pollution (Pingali, 2012). The latter problem is attributed to the indiscriminate use of these fertilizers, which are responsible for eutrophication and production of two greenhouse gases: carbon dioxide (CO_2) and nitrous oxide (N_2O). Eutrophication is the overgrowth of plants and algae, a consequence of nutrient enrichment that leads to oxygen depletion in freshwater and marine ecosystems (Smith and Schindler, 2009). Greenhouse gases contribute to the warming effect on the climate. Production of CO_2 is a consequence of burning fossil fuels during the Haber–Bosch process, the industrial method to produce nitrogenated fertilizers (Leigh, 2004), while N_2O is emitted from fertilized soils (Bouwman, 1996). Despite the aforementioned agricultural benefits, this has come at the high cost of polluting air and water, which makes this option unsustainable (Bohlool et al., 1992; Spiertz, 2009; FAO, 2017; Calabi-Floody et al., 2018).

A sustainable alternative is biological nitrogen fixation (BNF), which is the reduction of atmospheric nitrogen to ammonia (NH_3) by free-living or symbiotic diazotrophs (De Bruijn, 2015). BNF offers a sustainable way to integrate nitrogen into soils without causing severe environmental problems. This is possible as BNF systems are capable of supplying soils

9

with a high amount of nitrogenated compounds, the access is affordable and fixed nitrogen is less susceptible to leaching and volatilization (Bohlool et al., 1992; De Bruijn, 2015; Saha et al., 2017; Goyal et al., 2021). Furthermore, access to BNF comes with the economic benefit of reducing the cost of using fertilizers, which is heavily dependent on fossil fuel prices (Bohlool et al., 1992). Plants that belong to the Fabales, Fagales, Cucurbitales, and Rosales (FaFaCuRo) orders have evolved a mutualistic association with diazotrophs to obtain nitrogen (Kistner and Parniske, 2002). By partnering with plants, nitrogen fixing microorganisms obtain in exchange carbon components derived from photosynthesis, a process that is known as symbiotic nitrogen fixation (SNF) (Mus et al., 2016). Understanding this phenomenon is critical if we want to use this system to reduce the damage inflicted by synthetic nitrogen-based fertilizers on the environment and achieve food security for the growing population.

2. Overview of legume root nodule symbiosis

Legumes (Fabaceae or Leguminosae family) are plants with high agricultural relevance due to their protein-rich profile for human and animal nutrition (Gowda et al., 2009; Cernay et al., 2016; Ferreira et al., 2021). The legume family includes almost 20,000 species, with common bean (*Phaseolus vulgaris*), soybean (*Glycine max*), pea (*Pisum sativum*), peanut (*Arachis hypogaea, Ah*), alfalfa (*Medicago sativa*), lentils (*Lens culinaris*), and lupin (*Lupinus spp.*) being the most agronomically relevant species (Stagnari et al., 2017). Agricultural production of these legumes reaches 150 million tons worldwide (https://ourworldindata.org/agricultural-production) and it is estimated that approximately 2.5×10^{11} kg NH_3 are fixed annually (Saha et al., 2017). Their ability to engage with nitrogen-fixing soil bacteria known as rhizobia makes these plants interesting from a sustainability perspective (Goyal et al., 2021; Rogers & Oldroyd, 2014).

Nitrogen-fixing soil rhizobia colonize the plant root of legumes to form specialized organs called nodules, which provide the adequate conditions to host the symbiotic bacteria inside the plant cells (Wagner, 2011).This has been termed as root nodule symbiosis (RNS) (Kouchi et al., 2010; Oldroyd et al., 2011; Oldroyd, 2013; Parniske, 2018). Two legume model organisms, *Lotus japonicus* (*L. japonicus*, *Lj*) and *Medicago truncatula* (*M. truncatula, Mt*), have been intensively studied to reveal genetic players involved in RNS. Use of forward and reverse genetic screening using Ethyl methanesulfonate (EMS) populations, lines obtained from T-DNA insertions, *Tnt1* retrotransposon, and Lotus retrotransposon 1 (*LORE1*), and more recently CRISPR gene tool editing, have revealed genes required for RNS (Penmetsa and Cook, 2000; Perry, 2003; Tadege et al., 2008; Małolepszy et al.,

2016a; Roy et al., 2020). Furthermore, other legumes such as soybean, common bean, pea, peanut, and the semi-aquatic legume *Sesbania rostrata* (*S. rostrata, Sr*) have also helped us to understand specific aspects of RNS.

In order to establish RNS, a complex signaling cascade and downstream morphological changes happen in the plant (Figure 1). Early signaling starts with the release of chemical signals that allow the crosstalk between plants and rhizobia. Plant roots release flavonoids or isoflavonoids, secondary metabolites derived from the phenylpropanoid pathway (PPP) that are perceived by the rhizobia (Liu and Murray, 2016). These molecules induce genes encode proteins involved in the production and secretion of nodulation factors (NFs), which are a complex mix of lipochitooligosaccharides (Cooper, 2004). The chemical decoration of the (iso)flavonoids and the NFs determine the host-range specificity (Liu and Murray, 2016). The NFs are ligands of LysM receptor-like kinase receptors called Nod Factor Receptor 1 (*Lj*NFR1) and Nod Factor Receptor 5 (*Lj*NFR5) in *L. japonicus* and LysM Receptor Kinase 3 (*Mt*LYK3) and Nod Factor Perception (*Mt*NFP) in *M. truncatula*, which are localized at the plasma membrane of root epidermal cells (Madsen et al., 2003; Radutoiu et al., 2003; Limpens et al., 2003; Arrighi, 2006). The perception of NF triggers the first changes in the root tissues, primarily the epidermis, cortex and pericycle. In the epidermis, the induction of specific genes has been observed with root hairs which undergo deformation to entrap the bacteria and initial cell division occurs in the pericycle to form a nodule primordium. Cortical cell reorient for the progression of the bacteria (Geurts and Bisseling, 2002).

Downstream of the NF perception the leucine-rich repeat SYMbiosis Receptor-like Kinase (*Lj*SYMRK) or Does not Make Infection 2 (*Mt*DMI2) interacts with *Lj*NFR1 and *Lj*NFR5 (Stracke et al., 2002; Endre et al., 2002; Ried et al., 2014). The role of SYMRK has not been completely elucidated. *SYMRK* mutants lack IT and bacteria entrapment (Stracke et al., 2002). The kinase domain of SYMRK/DMI2 interacts with the 3-hydroxy-3-methyglutaryl-CoA reductase (*Mt*HMGR), an enzyme that produces the isoprenoid mevalonate (Venkateshwaran et al., 2015). It is suggested that mevalonate acts as a secondary messenger. This molecule transmits the signal activated by the NF to the nucleus and putatively triggers oscillation in the calcium (Ca^{2+}) concentration around and inside the nucleus (Limpens et al., 2005; Sieberer et al., 2009; Venkateshwaran et al., 2015). Ca^{2+} spiking requires multiple components located in the nuclear membrane, which include the potassium channels *Lj*CASTOR and *Lj*POLLUX or Does not Make Infection 1 (*Mt*DMI1) (Ané et al., 2004; Charpentier et al., 2008), a Ca^{2+} pump ATPase (*Mt*MCA8) (Capoen et al., 2011), and the Ca^{2+} channels *Mt*CNGC15a, b and c (Charpentier et al., 2016). Symbiotic-specific nucleoporins are hypothesized to move components to the

11

nuclear membrane required for the spiking. These nucleoporins include the proteins *Lj*NUP133, *Lj*NUP85, and *Lj*NENA (Kanamori et al., 2006; Saito et al., 2007; Groth et al., 2010).

It is generally accepted that Ca^{2+} spiking is decoded by the calcium/calmodium dependent protein kinase (*Lj*CCaMK) or Does not Make Infection 3 (*Mt*DMI3) that has three EF domains for binding Ca^{2+} and calmodulin (Levy et al., 2004; Tirichine et al., 2006). The active version of *Lj*CCaMK phosphorylates a coiled-coil domain containing transcription factor called *Lj*CYCLOPS or *Mt*IPD3 (Interacting Protein of DMI3) (Messinese et al., 2007; Yano et al., 2008), which in turn controls the expression of *Nodule inception* (*NIN*) (Schauser et al., 1999; Singh et al., 2014). *NIN* expression also requires two GRAS-domain-type transcription factors, Nodule Signaling Pathway 1 and 2 (*Lj*NSP1 and *Lj*NSP2), and *Mt*DELLA (Heckmann et al., 2006; Jin et al., 2016) (Figure 1). Regulation of *NIN* and of ERF Required for Nodulation (*Mt*ERN1) is controlled by *Lj*NSP1 and *Lj*NSP2. NIN and ERN1 are transcription factors that control the expression of several genes that lead to the nodule organogenesis and infection (Schauser et al., 1998; Middleton et al., 2007; Liu et al., 2019a).

2.1 Rhizobia infection

Nodule infection is diverse. Three mechanisms of infection have been described: i) root hair, ii) intercellular, and iii) "crack-entry" infection (Gage, 2004; Held et al., 2010; Ibáñez et al., 2016). Independently of the mechanism of infection, rhizobia attach to the root surface, cross the epidermis, colonize cortical cells, and establish inside them. Root hair infection is the most common and well-described mechanism. It takes place in plants belonging to the Mimosoideae-Caesalpinieae-Cassieae clade (Sprent et al., 2017). A tubular structure termed an infection thread (IT) develops inside the root hair. Rhizobia are trapped due to mechanical curling of the root hair cell. Microcolonies are formed after rhizobia divide. The local plant cell wall remodels and the plasma membrane invaginates to allow the bacteria to progress (Fournier et al., 2015; Ibáñez et al., 2016). Infection proceeds through the elongation and branching of the IT in the outer cortical cells (van Brussel et al., 1992). Eventually rhizobia reach the cells of the dividing cortex (Murray, 2011). Around 25% of legumes use different mechanisms of infection that are independent of root hair infection. These include *Mimosa*, *Neptunia*, *Stylosanthes*, *Cytisus*, *Sesbania*, *Arachis*, and *Lupinus,* among others (Sprent, 2007). Intercellular invasion involves the entry of the bacteria through the middle lamella. For "crack entry", rhizobia enter through colonization of natural cracks or fissures that occur at lateral root protrusion sites. Once inside the host cells,

rhizobia develop into bacteroids and engage in a bidirectional nutrient exchange with the host in organelle-like structures called symbiosomes (Oldroyd et al., 2011).

2.1.1 Root hair infection

A number of changes are required to develop an IT. These range from cytoskeleton modification, cell wall degradation, and elongation of the root hair. Mutant screens have identified genes mediating these processes. After bacteria attach to the root surface and release the NFs, one of the first changes is the deformation and curling of the root hair. This encloses the bacteria to form a microcolony in the so-called 'shepherd's crook' (Esseling et al., 2003). Mutants such as *nfr1-1* and *nfr5-1* are unable to perceive the NF, and there is an absence of root hair deformation (Radutoiu et al., 2003; Madsen et al., 2003). In *M. truncatula*, mutation in genes encoding the phosphatidylinositol 3 kinase (*Mt*PI3K) and the Rho family of small GTPases 10 (*Mt*ROP10) decreases the deformation and curling of root hairs (Peleg-Grossman et al., 2007; Lei et al., 2015).

Root hair infection requires the cytoskeleton to rearrange for the redirection and development of the IT. There are different genes responsible for actin cytoskeleton modifications belonging to the SCAR/WAVE-ARP2/3 complex. These include the genes *121F-specific p53 inducible RNA control* (*LjPir1*), *Nck-associated protein1* (*LjNap1*) or *required for infection thread* (*MtRIT*), and the *Actin-related protein component1* (*LjARPC1*). The latter genetic component is responsible for actin polymerization in the root hair. Mutations in these genes developed arrested, deformed, and swollen IT in the epidermis. (Yokota et al., 2009; Miyahara et al., 2010; Hossain et al., 2012). Actin elongation requires the SCAR-Nodulation (*Lj*SCARN). Mutant lines in *LjSCARN* have a similar phenotype as the previously mentioned genes (Qiu et al., 2015).

In order for rhizobia to enter, the plant cell wall must be modified and expanded. Pectin is one of the main components of the plant cell wall (Houston et al., 2016). Local degradation of the cell wall requires the Nodulation Pectate Lyase (*Lj*NPL). In the mutants *Ljnpl-1* and *-2*, bacteria are arrested in the IT (Xie et al., 2012). Elongation of the IT by polar growth follows the migration path of the nucleus. Abnormal and misdirected ITs have been observed when several genes associated with polar growth are mutated. These include the genes *Rhizobium-directed Polar Growth* (*MtRPG*) (Arrighi et al., 2008), the *Cytoplasmic exocyt subunit* (*MtEXO70I*) (Liu et al., 2019b), and the DOCK family guanine nucleotide exchange factor *Spike 1*(*LjSPK1*) and its interactor *Rho-family ROP GTPases* 6 (*LjROP6*) (Liu et al., 2020).

Mutations in other genes arrest the elongation of the IT. Bacteria remain in the microcolony stage without further progress in mutant lines of *LjCERBERUS* or *Lumpy infection (MtLIN) (Kuppusamy et al., 2004; Yano et al., 2009)*, *Vapyrin* (*MtVPY*) (Murray et al., 2011), and *LjCYCLOPS* (Yano et al., 2008). In other mutants, IT can reach the base of the epidermal cell but does not continue into deeper cell layers; such a case is observed in the mutant *crinkle (LjCrinkle)* (Tansengco et al., 2003). Other genes encoding different proteins have an impact on the infection at multiple stages, for example the transcription factor NF-YA1 (*Mt*NF-YA1) (Laporte et al., 2013), the membrane trafficking proteins flotillin-like Flotillin 2 and 4 (*Mt*FLOT2 and *Mt*FLOT4) (Haney and Long, 2010), the Symbiotic remorin1 (*Mt*SYMREM1) (Lefebvre et al., 2010), and the Plant U-box protein 1 (*Mt*PUB1) (Mbengue et al., 2010). In mutant lines of these genes, organogenesis is not impaired, but few nodules are formed and uncolonized.

2.1.2 "Crack-entry"

The semiaquatic legume *S. rostrata* has two types of infection depending on the growth conditions; under dry conditions it is infected via IT, whereas in flooding conditions bacteria penetrate via "crack entry" (Capoen et al., 2010). Under flooding, rhizobia induce cortical cell death in a NF-dependent manner. This requires the production of peroxide, ethylene, and gibberellins that lead to the formation of an infection pocket (d'Haeze et al., 2003; Lievens et al., 2005). From here bacteria are released into the surrounding cortical cells that activate the formation of intracellular infection threads using NF, *A. hypogaea* is infected via "crack-entry" with *Bradyrhizobium spp.* (Sharma et al., 2020). During infection, bacterial exopolysaccharide (EPS) plays an important role in "crack-entry". The formation of nodule-like structures in *A. hypogaea* was observed after inoculating with a mutant strain defective in the production of EPS (Morgante et al., 2007). It is proposed that production of EPS protects the symbiont from plant defenses during entry, unlike in IT infection where bacteria are less exposed to these defenses (Leigh and Coplin, 1992; Morgante et al., 2007). NF perception by *Ah*NFR1 and *Ah*NFP, orthologs of *Lj*NFR1 and *Lj*NFR5, respectively, is indispensable for infection in a partner-specific manner (Noisangiam et al., 2012). Likewise, *Ah*SYMRK, another component of early signaling, is able to complement *Mtsymrk-3 (*null mutant*)*, which suggests that the receptor from *A. hypogaea* has a conserved function. However, the role of these receptors in "crack-entry" still remains undetermined. Downstream of the NF perception, *Ah*CCaMK affects nodulation as silencing of this gene reduces the number of nodules and alters the presence of symbiosomes (Sharma et al., 2020). In addition, when *Ah*CYCLOPS is silenced the expression of

important downstream genes such as *AhNIN*, *AhHK1*, *AhCCaMK*, and *AhENOD40* is affected leading to a delay in nodulation. Establishing the molecular players required for all of the above infection mechanisms is key to understanding the evolution of symbiotic infection.

2.1.3 Intercellular infection

The species *A. hypogaea* and *S. rostrata* are better models to study the other two types of infections. For instance, *A. hypogaea* gets infected intercellularly where the infection site occurs in the middle lamellae (Uheda et al., 2001). In *L. japonicus* genetic mutations in the NF receptor genes (*Lj*NFR1, *Lj*NFR5, *Lj*SYMRK) revealed genetic evidence of an alternative signaling pathway resembling the intercellular infection. This was observed in a *snf1* mutant background (Madsen et al., 2010a). Thus, an alternative mechanism of infection happens in the absence of NF signaling but at a low frequency. More recently, comparative transcriptomic analysis between *L. japonicus - M. loti* R7A infected via IT and *L. japonicus - Rhizobium sp.* IRBG74, which infects *Sesbania* species intercellularly (Cummings et al., 2009; Aguilar et al., 2016), showed that common and distinctive genetic players for both mechanisms of infection are needed. Genes like *LjNFR5*, *LjSYMRK*, *LjCCaMK*, *LjCYCLOPS*, *LjNIN*, *LjNSP1*, and *LjNSP2* are equally important for both mechanisms of infection since nodulation was also affected in mutants of these genes. In the case of the intercellular mechanism of infection, cytokinin (CK) signaling plays an important role (Montiel et al., 2020). CKs are phytohormones that impact RNS either positively or negatively (for a more comprehensive review see Gamas et al., 2017). Genes such as *Lotus Histidine Kinase* (*LjLHK1*) (Murray et al., 2007), *Cytochrome P450 monooxygenase* (*LjCYP735a*), and *Isopentyl transferase* (*LjIPT4*) are necessary for the synthesis of CKs. Mutant analysis of these genes inoculated with *Rhizobium sp.* IRBG74 revealed a reduced number of nodules as an indirect measurement of infection (Montiel et al., 2020). On the other hand, genes related with the formation of the IT are imperative for the intracellular mechanism of infection as demonstrated by the analysis of *rinrk1*, *ern1*, *rbohE*, *rbohG*, *rpg*, *rpg-like*, and *vpy2* mutants. Only in *vpy2* and *rpg* mutants ITs were quantified showing a reduction in this phenotype. (Montiel et al., 2020).

2.2 Nodule organogenesis

Similar to infection, nodule morphology is diverse. Species such as *M. truncatula*, pea, lentil, and fava bean (*Vicia faba*) develop indeterminate nodules that have an elongated shape with a long-lived meristem at the nodule apical end (Guinel, 2009). Indeterminate nodules

have five defined histological zones, each with a specific feature. Zone I is the meristematic zone that generates the majority of the cells that make the mature nodule, Zone II is the early symbiotic zone where ITs penetrate, Zone III is the nitrogen fixation zone and is called the symbiotic zone, Zone IV is the senescent zone where bacteroids degrade, and Zone V is the saprophytic zone (Vasse et al., 1990; Timmers et al., 2000). *L. japonicus*, soybean, and common bean develop a second type of nodules called determinate, which have a spherical form due to a transient meristem. The anatomy of the determinate nodule consists of the epidermis, nodule endodermis, nodule cortex, and vascular bundles (Guinel, 2009). The nodule endodermis separates the nodule cortex into an inner and outer cortex (Frazer, 1942) and it is proposed to act as a gas diffusion barrier (Witty and Minchin, 1990). The vasculature trace is the connection between the nodule and the root (Spratt, 1919). Determinate nodules are considered to be more advanced compared to indeterminate as they have radial symmetry with only the fixation zone where nitrogen fixation happens and eventually senesce (Sprent and Platzmann, 2001; Sprent, 2008; Guinel, 2009). Despite all these differences, the formation of both types of nodules requires a similar set of genes.

NIN controls organogenesis and infection by activating several genes, therefore acting as a master regulator (Liu et al., 2019a; Liu and Bisseling, 2020). NIN controls the infection in root epidermal cells but in the pericycle and cortical cells is responsible for nodule primordium formation (Liu et al., 2019c). The targets of NIN involved in organogenesis are connected with lateral root development and hormonal regulation. It is not clear how exactly the lateral root program and nodule organogenesis are connected, but neo-functionalization of specific genes requires the genetic players to form the lateral roots (Soyano et al., 2021). For instance, in both *M. truncatula* and *L. japonicus* the *Asymmetric Leaves 2-Like/Lateral Organ Boundaries domain 16a (ASL18/LBD16a)* gene regulates nodule primordium formation (Soyano et al., 2019). The *A. thaliana* ortholog *AtASL18/LBD16a* is required for lateral root primordium development (Okushima et al., 2007). The *Ljasl18a-1* has a significantly reduced number of nodules and lateral roots (Soyano et al., 2019). Other targets of NIN are the *Nuclear Factor-Y subunit A-1* and *B-1* (*LjNF-YA1* and *LjNF-YB1*) genes, which activate cortical cell divisions (Soyano et al., 2013). When *LjASL18a* and *LjNF-YA1/B1* are co-expressed, lateral root density increases and bumps form (Soyano et al., 2019).

NF-Ys induce the expression of the transcription factor STY (Short internodes/stylish) that in turn regulates the expression of the *YUCCA* genes required in the biosynthesis of the phytohormone auxin (Shrestha et al., 2021). The exogenous application of auxin leads to nodule primordium development (Allen et al., 1953; Libbenga et al., 1973). Expression of

ASL18/LBD16a is induced by auxin in both legumes and non-legumes (Okushima et al., 2007; Schiessl et al., 2019). In addition, NF-YA1 interacts with ASL18/LBD16a (Soyano et al., 2019). CKs are also positive regulators of nodule organogenesis as application of exogenous CKs in *L. japonicus* and *P. sativum* induces nodule primordium formation (Libbenga et al., 1973; Heckmann et al., 2011). Moreover, the treatment of CKs in a broad range of plants showed that only nodulating species are capable of forming pseudonodules in response to CKs (Gauthier-Coles et al., 2019). Isoprenoid CKs are synthesized by the Isopentenyl transferase 2 (*Lj*IPT2) and Lonely guy 4 (*Lj*LOG4) and regulated by the genes *cytokinin oxidase/dehydrogenase3* (*LjCKX3*) and *Lonely guy 1* (*MtLOG1*) (Chen et al., 2014; Mortier et al., 2014; Reid et al., 2016, 2017). In addition, the CK receptors *Lj*LHK1, *Lj*LHK1A*, and *Lj*LHK3 are also required for nodule formation (Murray et al., 2007; Held et al., 2014). In the mutant *spontaneous nodule formation 2* (*Ljsnf2-1* and *Ljsnf2-2*), nodules develop in the absence of rhizobia (Tirichine et al., 2006). *Ljsnf2* is a gain-of-function allele in which *LjLHK1* is constitutively active (Tirichine et al., 2007). CKs regulate *NIN* expression in cortical cells via the presence of a distal cis-regulatory element called the Cytokinin-responsive element (Liu et al., 2019c). These examples illustrate the key regulatory role of phytohormones such as auxins and CKs during nodule organogenesis.

3. Bacterial uptake and accommodation

Once rhizobia reach the cortical cells, they are taken up inside them. Here the bacteria differentiate into bacteroids and are surrounded by a plant-derived peribacteroid membrane. Different cell modifications are required to host the symbiont, referred as bacterial accommodation. In addition, plant cells need to modify their composition without compromising their integrity (Parniske, 2018). In the following subsections, changes that the plant cell undergoes will be introduced to explain the diversity in these processes.

3.1 Symbiosome

At the cell entry point, the rhizobia become enclosed by a host-derived membrane, which is called the peribacteroid membrane. In the model organism *L. japonicus*, rhizobia divide and differentiate into bacteroids (Oke and Long, 1999; Whitehead and Day, 1997). Encapsulated bacteroid constitutes the symbiosome, which is a special organelle-like structure where nitrogen fixation takes place (Roth and Stacey, 1989). Release of the bacteria from the IT and symbiosome formation require an exocytotic pathway that delivers

membrane vesicles to the plasma membrane (Limpens et al., 2009). SNARE (soluble N-ethylmaleimide-sensitive factor attachment protein receptor) proteins are needed to deliver membrane vesicles during endocytosis (Wickner and Schekman, 2008). Knocking down two SNARE genes, *LjVAMPd* and *LjVAMPe* (Vesicle-associated membrane protein), reduces the number of bacteroids and nodule formation (Sogawa et al., 2019). Mutants in symbiosome development genes go through changes in size. For instance, a mutant in *LjSYP71* (Syntaxin of plant 71) causes enlarged symbiosomes that have impaired nitrogen fixation (Hakoyama et al., 2012b), whereas a knockout version of the *MtARP3* (*Actin Related Protein*) gene produces smaller symbiosomes (Gavrin et al., 2015). In the space between the symbiosome membrane and the bacteroids, a bidirectional nutrient exchange between bacteroids and the host cell takes place (Perret et al., 2000; Haag et al., 2013). Across the symbiosome membrane, exchange of fixed nitrogen, reduced carbon, amino acids, and inorganic cations such as iron, copper, molybdenum, nickel, and cobalt is essential to fuel the activity of the nitrogen-fixing nitrogenase (reviewed in Udvardi and Poole, 2013). There have been several transporters classified based on the substrate transported (Table 1). Mutation in these transporters impairs RNS and in some cases produces irregular-shaped and vacuolated symbiosomes.

Table 1. Transporters required for SNF

Gene	Substrate	Mutant phenotype	Reference
MtMATE67 *LjMATE1*	Citrate	Reduced nitrogen fixation and absence of pink nodules	(Kryvoruchko et al., 2018)
LjAMT1;1	Ammonium	Reduced nitrogen fixation and an increased number of nodules	(Rogato et al., 2008)
*LjNPF8.6/LjNPF2.4/LjNPF3.1/MtNPF7.6**	Nitrate	Reduced nitrogen fixation, increased ROS production and anthocyanin accumulation	(Valkov et al., 2017, 2020; Wang et al., 2020a; Vittozzi et al., 2021)
GmVTL1	Iron	Expression of *GmVTL1a* in *Ljsen1-1* background mutant to evaluate iron transport	(Brear et al., 2020)
LjSEN1	Iron	Reduced nitrogen fixation, enlarged symbiosomes and small bacteroids	(Hakoyama et al., 2012a)
LjSST1	Sulfur	Reduced nitrogen fixation, reduction in leghemoglobin concentration, small and early senescent nodules	(Krusell et al., 2005)
MtTIP1g (Aquaporin)	Water	Reduced nitrogen fixation and premature symbiosome maturation from elongation to nitrogen-fixing stage.	(Gavrin et al., 2014)

* proposed to be in the peribacteroid membrane

3.2 Oxygen homeostasis

Oxygen is the second most abundant gas in the atmosphere and is a vital component for life. Despite its primordial role, oxygen also causes damage to living organisms via the production of reactive oxygen species (ROS). In RNS, the regulation of this molecule is of the utmost priority as oxygen is a denaturing agent of the bacterial nitrogenase complex

(Whiting and Dilworth, 1974). This complex is made up of six protein subunits and different metallo-centers: two iron-sulfur clusters [4Fe-4S] and (Fe_8S_7) and two iron-molybdenum cofactors (Fe_7MoS_9N) (Eady and Postgate, 1974; Downie, 2014). The iron-sulfer cluster is more susceptible than the iron-molybdenum ones (Wang et al., 1985). Therefore, the concentration of oxygen inside nodules needs to be tightly regulated to create a low oxygen environment but at the same time it must be transported and delivered where it is required. The reduction of atmospheric nitrogen requires 16 ATP molecules; therefore, this process consumes a large amount of ATP (Berg et al., 2002). Bacteroids have developed a very efficient respiration. This is possible because bacteroids use a high affinity oxygen cytochrome oxidase (cytochrome cbb_3), the terminal enzyme in the respiration chain, that reduces atmospheric oxygen and releases energy in the form of electrons (Blomberg, 2016). This produces energy to support nitrogen fixation (Marchal and Vanderleyden, 2000). This oxidase has a high affinity for oxygen and is able to consume the delivered oxygen by specific plant proteins called leghemoglobins (Appleby, 1984).

3.2.1 Mechanisms to control oxygen homeostasis

Plant cells use three different mechanisms to create the low oxygen environment: a) formation of an oxygen diffusion barrier, b) clustering of mitochondria in the periphery of infected cells, and c) expression of symbiotic leghemoglobins.

a) Nodule barrier

Plant barriers have evolved as a fundamental mechanism to endure the transition from aquatic to land environments (Pollard et al., 2008). They regulate the uptake of nutrients and gas exchange and protect against different biotic and abiotic stresses (Pollard et al., 2008). The composition of these barriers includes polymers like lignin, cutin, and suberin whose chemical profile includes monolignols for lignin and glycerol, long aliphatic chain fatty acids, long-chain fatty alcohols, and phenolic compounds for cutin and suberin (Fich et al., 2016; Barberon, 2017). The root endodermis is a cell layer that separates the inner vascular tissues from the cortex (Barberon, 2017). The cells in the endodermis contain the lignified Casparian Strips and suberin lamellae and together they regulate water and mineral uptake (Miyashima and Nakajima, 2011).

Inside the nodule, it has been proposed that an oxygen barrier in the periphery of the nodules controls the diffusion of oxygen (Witty and Minchin, 1990). Nodules have a specialized endodermis termed the nodule endodermis, which differs between legume

species (Guinel, 2009). Independently of the legume species, it is believed to restrict gas diffusion and control pathogen entry (Hartmann et al., 2002). More than three decades ago, the presence of a gas diffusion barrier was demonstrated by microelectrode oxygen measurements (Witty and Minchin, 1990). However, some authors propose that an additional layer called the sclerenchyma layer exists (Frazer, 1942; Hirsch, 1992; Guinel, 2009), and thus it has been proposed that a combination of different cell layers are responsible for generating the low oxygen environment (Minchin, 1997; King and Layzell, 1991). The nodule endodermis and the sclerenchyma are made of lignin and suberin with values that are around 27 and 72 μg/mg, respectively in broad bean (*Vicia faba*) nodules (Brown and Walsh, 1994; Hartmann et al., 2002). Although different studies have determined the anatomy and chemical composition of this barrier in a few species, we do not know the genetic determinants or the signals that regulate its formation. Only one study has pointed out some putative genetic components expressed in the parenchyma layer of *G. max*. These genes are the early nodulins *GmENOD2*, *GmENOD13*, and *GmENOD40,* but no further validation has been undertaken (Franssen et al., 1992).

b) Clustering of mitochondria

Soybean nodules have evenly distributed gas-filled intercellular spaces, which serve as a conduit for oxygen to the infected cells (Bergersen and Goodchild, 1973). In infected cells, mitochondria provide the energy required for assimilation of NH_3 and transport of different compounds into the symbiosomes (Day and Copeland, 1991). In soybean, the mitochondria from infected cells showed a higher Michaelis-Menten constant (*Km*) in comparison with root and cotyledon mitochondria, which were also less adapted to low oxygen environments (Millar et al., 1995). It is proposed that the accumulation of mitochondria at the periphery of infected cells occurs to consume oxygen from intercellular spaces (Millar et al., 1995; Bergersen et al., 1995; Bergersen, 1997).

c) Expression of symbiotic leghemoglobins

Hemoglobins are present in both animals and plants and have functions related to carry oxygen transport (Czelusniak et al., 1982; Bogusz et al., 1988; Trevaskis et al., 1997). The most well-characterized hemoglobins in plants are the ones produced in legumes, but other non-legume species also have them (Taylor et al., 1994; Trevaskis et al., 1997). In legumes, leghemoglobins are proteins of 16kDa composed of a protoporphyrin IX (heme moiety) and polypeptide (globin). These proteins give the characteristic pink color to the nodules due to the ferrous state of the iron core (Becana et al., 1995). In the nodule, they serve as oxygen

carriers that deliver it to the bacteroid surface for the respiration needed to perform nitrogen fixation (Appleby, 1984). Robertson proposed that leghemoglobins deliver oxygen to the host's mitochondria (Robertson et al., 1984). However, the ccb_3 has a high affinity for oxygen and works under low levels of oxygen (5 to10 nM) compared with the host's mitochondria. Furthermore, the observation that mitochondria from the host are close to air-filled intercellular spaces eliminates the possibility of oxygen being delivered by leghemoglobins to the host's mitochondria (Appleby, 1984). Early works suggested that the heme group was synthesized by the bacteria and the assembly of the leghemoglobins happened in the plant cell cytoplasm (Cutting and Schulman, 1969; Verma et al., 1979). However, this was later excluded as plants have all the genes required for the synthesis of the heme prosthetic group (review in O'Brian, 1996). Seven reactions are needed to produce the heme group from aminolevulinic acid (ALA). Interestingly, the reactions take place in different parts of the plant where the three last and most critical reactions happen in the plant mitochondria (Dimitrijevic et al., 1989; O'Brian, 1996). Evidence that heme biosynthetic enzymes are expressed during nodulation has been reported in soybean nodules where there is a strong induction of ALA dehydratase, coproporphyrinogen oxidase, and ferrochelatase activity (Dimitrijevic et al., 1989; Madsen et al., 1993; Kaczor et al., 1994).

In *L. japonicus*, three leghemoglobin encoding genes are located in the same chromosome and the encoded proteins share more than 80% sequence identity. Regulation of leghemoglobin gene expression is controlled by the *Mt*NPL2 (NIN-like protein 2); as in the *Mtnlp2-1* mutant, leghemoglobin transcript levels are downregulated. Mutants have small nodules, reduced nitrogen fixation activity, and lower leghemoglobin content. Promoter analysis in several downregulated genes in *Mtnlp2-1* revealed the presence of two regulatory elements: the nitrate-responsive element (NRE) and two partly overlapping NRE designated as 'double NRE' (dNRE), which are required by NPL2 and NIN to regulate the expression of leghemoglobin genes (Jiang et al., 2021). On the other hand, silencing of the three leghemoglobin causes an increase in oxygen level, loss of nitrogenase protein (Ott et al., 2005). Single mutants created by CRISPR/Cas9 editing show a reduction in nitrogenase activity but a more drastic phenotype is observed in double and triple mutant plants (Wang et al., 2019). The presence of multiple leghemoglobin genes in legumes highlights the importance of these proteins in RNS.

Altogether these mechanisms create the perfect environment to protect the oxygen-labile nitrogenase and support nitrogen fixation. However, genetic and biochemical evidence is only available for the leghemoglobins function. What signals are required for the relocation

of mitochondria and the genetic components aiding in the formation of the nodule barrier remain unknown.

3.3 Reactive oxygen species and antioxidants

Although not precisely part of the bacteria accommodation production of ROS, such as superoxide (O_2^-) and peroxide (H_2O_2), accompanies RNS at different stages either during IT formation, maturation, or senescence. This is because ROS are signal molecules that lead to activation of transcription factors and enzymes. In order to avoid the harmful chemical nature of these molecules, plants produce an arsenal of antioxidants (review in Matamoros and Becana, 2020).

ROS are produced in root hairs primarily by RBOHs within seconds of NF perception (Cárdenas et al., 2008). The quick response of the RBOHs suggests that a mechanism independent of gene expression is responsible. Induction in root hairs of expression of several *MtRBOHs* after NF treatment indicates a role in early signaling by expressing genes involved in IT formation (Ramu et al., 2002; Damiani et al., 2016). During infection, ROS work as loosening agents of the cell wall for IT progression; in common bean, *Pv*RBOHA and *Pv*RBOHB are located along the IT (Arthikala et al., 2017). Silencing of those genes results in the abortion of the IT at the base of the root hair but when *PvRBOHB* is overexpressed the number of ITs and of nodules increase (Arthikala et al., 2017).

Production of ROS in nodule primordium has been observed in alfalfa (*Medicago sativa)* and bean. It is suggested that O_2^- produced by the RBOH are required during mitosis for primordium formation (Montiel et al., 2016). In mature nodules, different sources such as the bacteroidal transport chain, hydrogenases, and ferrodixin contribute to the production of ROS. For instance, *MtRBOHA* is more expressed in nodules compared to other *MtRBOHs* and its silencing leads to downregulation of the genes coding for the nitrogenase and a reduction in nitrogen fixation (Marino et al., 2011). Mitochondria of infected cells also contribute to ROS production through respiration during nitrogen fixation (Møller, 2001). Another source of superoxide is generated by the oxidation of leghemoglobins (Becana and Klucas, 1992) and even mutations in leghemoglobin genes produce a high amount of superoxide (Wang et al., 2019). During nodule senescence, the activities of the nitrogenase and the leghemoglobins diminish over time. This makes iron available, which contributes to ROS production as iron is a catalyst to produce hydroxyl radicals (Becana and Klucas, 1992; Puppo et al., 2005).

Plants produce antioxidants to neutralize the damaging effect of ROS. These come as secondary metabolites, degrading enzymes, or chelators. Nodules produce different types of antioxidants (reviewed in Becana et al., 2010; Matamoros and Becana, 2020). Ascorbate and glutathione serve as ROS scavengers. Synthesis of ascorbate requires different enzymes that belong to the Smirnoff–Wheeler pathway (Ishikawa et al., 2006). Transcripts encoding enzymes of the Smirnoff–Wheeler pathway have been achieved in nodules of *L. japonicus* (Matamoros et al., 2006). Enzymatic activity of glutathione biosynthetic enzymes has been detected in soybean nodules (Moran et al., 2000). The nodule defense arsenal includes: i) catalases that are localized in the peroxisome of infected cells in lupin (Lorenzo et al., 1990), ii) superoxide dismutases that are metalloenzymes situated in infected root hair, and infected cells of *L. japonicus*, pea, and alfalfa (Rubio et al., 2004, 2007), and iii) thioredoxins and glutaredoxins, components of the thiol-disulfide redox regulatory network, which are expressed in *L. japonicus* nodules (Tovar-Méndez et al., 2011). Non-symbiotic leghemoglobins or globulins (Glb) are present in legumes and also serve as antioxidants but particularly of reactive nitrogen species, such as nitric oxide (NO). The overexpression of *LjGlb1-1* helps in nitrogen fixation as it delays senescence by scavenging NO (Fukudome et al., 2019); furthermore, *LjGlb2* and *LjGlb3* are highly expressed in nodules compared with roots but with an unknown function (Shimoda et al., 2009).

3.4 Cell biological changes

Infected nodule cells undergo profound cell biological changes to host rhizobia. Host cells go through endoreduplication (Kondorosi and Kondorosi, 2004), their vacuoles contract (Gavrin et al., 2014), and the tubulin cytoskeleton realigns (Kitaeva et al., 2016), which coincides with dramatic cell expansion (Tsyganova et al., 2018). The genetic base of these modifications is very diverse. For instance, the *MtCCS52* (*cell cycle switch*) gene induces endoreduplication in nodules and is expressed in infection zone II (Cebolla et al., 1999). Downregulation of *MtCCS52* leads to the formation of irregular shaped infected cells, lower ploidy, decreased cell size and premature senescence (Vinardell et al., 2003). Vacuole formation and function are altered in infected cells. In *M. truncatula*, when two members of the HOPS vacuole-tethering complex, VPS11 and VPS39, are downregulated, vacuoles contract allowing space for symbiosome expansion in size (Gavrin et al., 2014). Symbiosome distribution around the central vacuole requires microtubules in *M. truncatula*. Different microtubule organizations have been observed in *M. truncatula* and *P. sativum.* In *M. truncatula*, microtubules are positioned parallel to the symbiosomes but in *P. sativum* they are disorganized (Kitaeva et al., 2016). Altogether, these modifications show the striking change in size of host cells to accommodate the symbiont.
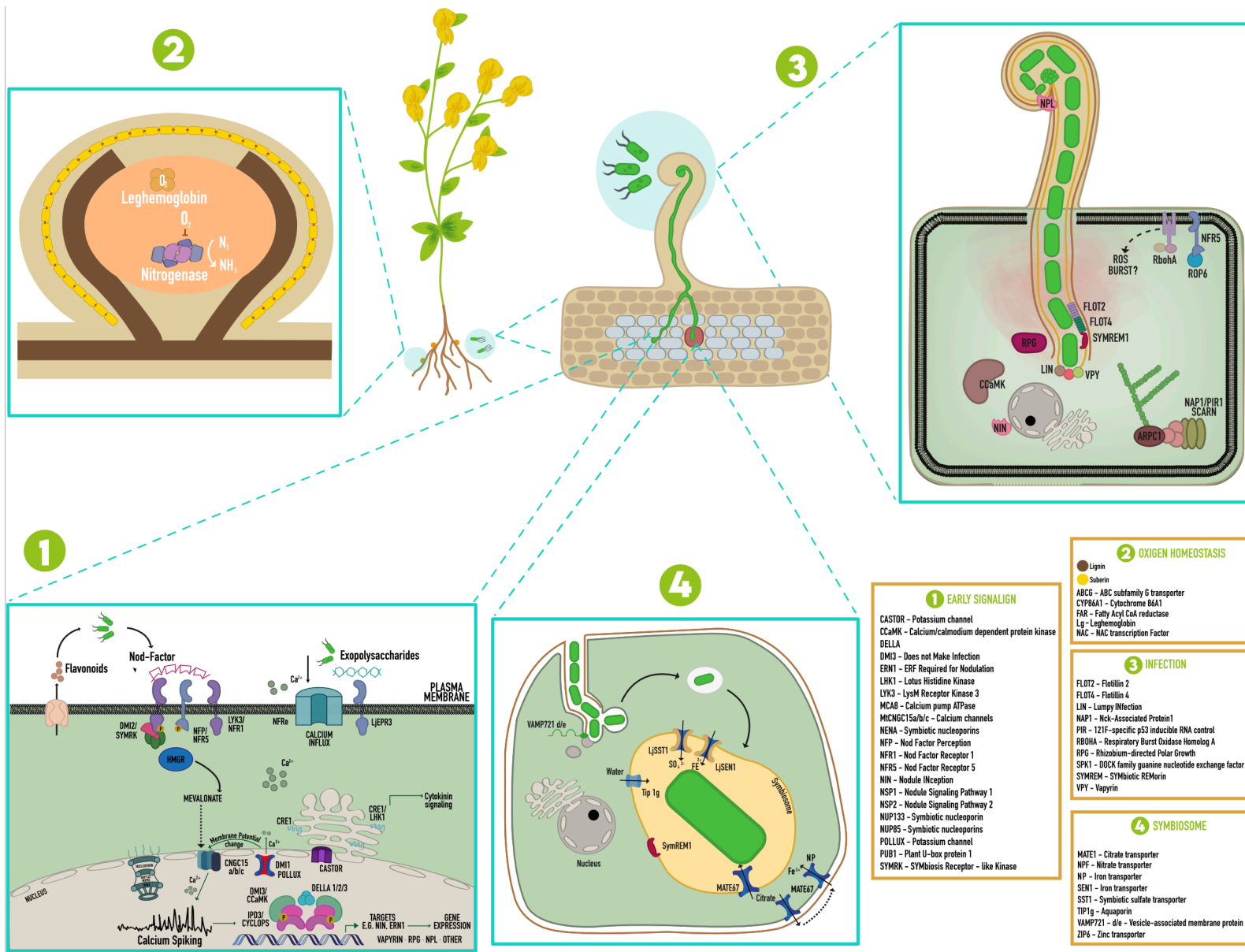
The morphological and physiological changes derived from the colonization of the symbiont have not been studied as intensely as other processes. In general, the complexity to study these changes requires not only conventional forward and reverse genetic screenings but also new approaches. Furthermore, genetic redundancy also complicates the identification and contribution of small effect genes which are likely to be involved in accommodation. Therefore, new strategies are required to reveal genetic determinants of these processes.

4. The *Lotus - Rhizobium leguminosarum* Norway system

Lotus belongs to the Loteae tribe and comprises 150 species. Centers of diversity are located in the Mediterranean region and Western North America. From the economic point of view, different *Lotus* species have been improved by domestication and breeding, including *L. corniculatus*, *L. pedunculatus*, *L. tenuis*, and *L. subbiflorus* (Escaray et al., 2012). *L. japonicus* serves as a model organism for molecular studies of RNS (Márquez et al., 2005). It was originally believed that only two symbionts, *Mesorhizobium* and *Bradyrhizobium*, were compatible partners of *Lotus*. However, an increasing number of reports have revealed that species from the genera *Rhizobium*, *Ensifer*, and *Aminobacter* engage in symbiosis with *Lotus* species (reviewed in Lorite et al., 2018). Thus, a great potential to study host-bacteria compatibility remains in this rhizobia-*Lotus* interaction.

Grossman *et al*. 2012 proposed to screen the *Lotus* natural diversity to identify new phenotypic variation. In their study, they isolated and characterized nodule-associated bacteria from two *Lotus* species. The strain *Rhizobium leguminosarum* (*R. leguminosarum*) Norway was isolated from nodules of *Lotus corniculatus* and exhibited polymorphic symbiotic phenotypes when inoculated in other *Lotus* species, which include contrasting infection and organogenesis. The contrasting phenotypes included the degree of colonization and the presence of well-formed nodules or tumors (Gossmann et al., 2012). When *R. leguminosarum* Norway is inoculated onto *Lotus burttii* nodules develop, whereas in *Lotus filicaulis* and *L. japonicus* ecotypes MG-20, Nepal, and Gifu are absent (Gossmann et al., 2012; Liang et al., 2019). In *L. burttii* nodule organogenesis and infection are uncoupled upon inoculation with *R. leguminosarum* Norway (Liang et al., 2019). No ITs are detected at the epidermal and cortex cells, but rather bacteria penetrate plant cells through 'peg-like' structures. These are tubular invaginations in the cell wall that enclose bacteria but do not transverse cells (Rae et al., 1992). The infection does not require NFs and plant transcriptional response changes when inoculated with *M. loti* MAFF303099 or *R. leguminosarum* Norway (Liang et al., 2019).

Figure 1. A simplified overview of the genetic players during Symbiotic Nitrogen Fixation. 1) Early signaling; bacteria release NFs that are perceived by plant receptors. The recognition activates a signal cascade that travels to the nucleus where it triggers calcium spiking. The spiking is decoded to activate the organogenesis and infection programs. 2) Oxygen homeostasis. 3) Rhizobia infection; the bacteria are entrapped by the root hair and an IT develops along the root hair to guide the rhizobia to the inner cells. 4) Symbiosome formation; rhizobia are delivered to the cell where they are surrounded by the plant membrane to form the symbiosome. Different transporters are required for nutrient exchange that are localized in the symbiosome and the adjacent plant cells. Adapted from (Venado et al., 2019; Roy et al., 2020).

25

②

Leghemoglobin

O_2

O_2

N_2

NH_3

Nitrogenase

③

NPL

ROS BURST?

RbohA

NFR5

ROP6

FLOT2

FLOT4

SYMREM1

RPG

LIN

VPY

CCaMK

NIN

NAP1/PIR1 SCARN

ARPC1

①

Flavonoids

Nod–Factor

Exopolysaccharides

Ca^{2+}

PLASMA MEMBRANE

DMI2/ SYMRK

NFP/ NFR5

LYK3/ NFR1

NFRe

CALCIUM INFLUX

LjEPR3

HMGR

Ca^{2+}

MEVALONATE

Ca^{2+}

Cytokinin signaling

CRE1/ LHK1

Membrane Potential change

Ca^{2+}

CRE1

CNGC15 a/b/c

DMI1 POLLUX

CASTOR

Ca^{2+}

NUCLEUS

DMI3/ CCaMK

DELLA 1/2/3

IPD3/ CYCLOPS

TARGETS E.G. NIN, ERN1

GENE EXPRESSION

Calcium Spiking

VAPYRIN · RPG · NPL · OTHER

④

VAMP721 d/e

Nucleus

LjSST1

Water

Tip 1g

SO_4^{2-}

FE

LjSEN1

Symbiosome

SymREM1

MATE67

Citrate

Fe^{2+}

MATE67

NP

① EARLY SIGNALIGN

CASTOR – Potassium channel
CCaMK – Calcium/calmodium dependent protein kinase
DELLA
DMI3 – Does not Make Infection
ERN1 – ERF Required for Nodulation
LHK1 – Lotus Histidine Kinase
LYK3 – LysM Receptor Kinase 3
MCA8 – Calcium pump ATPase
MtCNGC15a/b/c – Calcium channels
NENA – Symbiotic nucleoporins
NFP – Nod Factor Perception
NFR1 – Nod Factor Receptor 1
NFR5 – Nod Factor Receptor 5
NIN – Nodule INception
NSP1 – Nodule Signaling Pathway 1
NSP2 – Nodule Signaling Pathway 2
NUP133 – Symbiotic nucleoporin
NUP85 – Symbiotic nucleoporins
POLLUX – Potassium channel
PUB1 – Plant U-box protein 1
SYMRK – SYMbiosis Receptor – like Kinase

② OXIGEN HOMEOSTASIS

⬤ Lignin
⬤ Suberin

ABCG – ABC subfamily G transporter
CYP86A1 – Cytochrome 86A1
FAR – Fatty Acyl CoA reductase
Lg – Leghemoglobin
NAC – NAC transcription Factor

③ INFECTION

FLOT2 – Flotilin 2
FLOT4 – Flotilin 4
LIN – Lumpy INfection
NAP1 – Nck–Associated Protein1
PIR – 121F–specific p53 inducible RNA control
RBOHA – Respiratory Burst Oxidase Homolog A
RPG – Rhizobium–directed Polar Growth
SPK1 – DOCK family guanine nucleotide exchange factor
SYMREM – SYMbiotic REMorin
VPY – Vapyrin

④ SYMBIOSOME

MATE1 – Citrate transporter
NPF – Nitrate transporter
NP – Iron transporter
SEN1 – Iron transporter
SST1 – Symbiotic sulfate transporter
TIP1g – Aquaporin
VAMP721 – d/e – Vesicle–associated membrane protein
ZIP6 – Zinc transporter

26

Aim of the thesis

Nitrogen is a limiting factor for plant development as this element is crucial in different metabolic processes. Legumes form a symbiosis with nitrogen-fixing rhizobia bacteria. This symbiosis occurs within cells of specialized root organs called nodules in which a bidirectional nutrient exchange between the symbionts takes place. Several studies have elucidated the molecular signal communication between the symbionts leading to nodule organogenesis.

However, our knowledge of the genetic basis for the tissue and cellular adaptations required to host rhizobia inside nodules remains extremely limited, because of the difficulty in disconnecting nodule formation from infection. Despite the key importance of some of these adaptations for effective nitrogen fixation, genes controlling, for instance, modification in the nodule endodermis formation required to control oxygen diffusion into the nodules are still unknown. This thesis aimed to identify genes involved in the rhizobia accommodation into nodule cells, particularly those associated with cell wall modifications.

Lotus natural diversity of symbiotic phenotypes in response to a subcompatible strain was explored to identify genes associated with bacteria accommodation. The particular objectives were to:

1) Explore the diversity in nodule infection phenotypes elicited by *R. leguminosarum* Norway in different *L. japonicus* accessions.
2) Perform RNA-seq using prime-seq in *L. japonicus* accessions with contrasting infection phenotypes.
3) Identify candidate genes using a differentially expressed and co-expression analysis for further molecular characterization.
4) Characterize selected candidates using reverse genetics and physiological and molecular methods.
5) Generate knowledge that will hopefully aid programs aiming to transfer the symbiotic genetic toolbox into non-legume species.

Materials and methods

1. General plant growth and inoculation conditions

1.1 Bacterial growth conditions

The strains used in this work are listed in Supplemental Table 1. Liquid cultures of the bacterial strains were grown in tryptone yeast extract broth (Beringer, 1974) supplemented with 5 mM $CaCl_2$ and grown for 48 hours at 28°C and 180 rpm before inoculation.

1.2 Plant material, growth conditions, and inoculation

L. japonicus seeds (Supplemental Table 2) were surface sterilized with a sterilizing solution (1.2% NaClO and 1% SDS), soaked with sterile water for 2 h, and germinated in 0.8% agar plates with ½ Gamborg B5 medium (Gamborg et al., 1968). Seeds were incubated at 24°C for 3 days in darkness, followed by 3 days under a long-day photoperiod (16h:8h, light:dark). Ten seedlings per condition were transplanted into sterile Tulip-shaped Weck jars (WECK) containing 300 ml of a sand:vermiculite mixture (1:2) supplemented with 40 ml of FAB medium. Two days after transferring, each seedling was inoculated with 1 ml of a bacteria suspension. Suspensions were generated by washing bacteria grown as described in 1.1 with sterile water and by adjusting the OD_{600} to 0.005 in FAB medium. Plants were grown under a long-day photoperiod at 24°C. Specific details for each experiment will be mentioned in each subsection.

2. Lotus accessions screening

2.1 Specific bacterial and plant growth conditions

Liquid cultures of *R. leguminosarum* Norway-GFP Norway were grown as described in section 1.1. Media were supplemented with antibiotics as follows: tetracycline (Tc, 2 µg ml^{-1}) and streptomycin (Sm, 500 µg ml^{-1}). Forty *L. japonicus* accessions (https://www.legumebase.brc.miyazaki-u.ac.jp/lotus/wildStrainListAction.do and Supplemental Table 2) were grown as described in section 1.2. Ten seedlings of each accession were inoculated with *R. leguminosarum* Norway-GFP (OD_{600} = 0.005).

2.3 Nodule infection screening

Roots and nodules were harvested at 35 days post inoculation (dpi). Samples were immersed in ClearSee solution for 24 h and fixed with 4% paraformaldehyde as described by Kurihara et al. (2015). All accessions were screened qualitatively for the formation of nodules, bumps, tumors, and presence/absence of nodule infection. Ten accessions which displayed round-shaped nodules were quantitatively screened for percentage of infected

nodules. Two groups of accessions were selected to quantify the area colonized by *R. leguminosarum* Norway-GFP. The groups contained three accessions with either the highest or the lowest number of infected nodules. For each accession, ten nodules were embedded in 6% low melting agarose and sectioned with a VT1000S vibratome (Leica Biosystems). The 50-µm-thin sections were visualized with a DM6 B upright microscope (Leica Microsystems). Images were quantified using Fiji ImageJ (Schindelin et al., 2012). The percentage of colonization was calculated as the percentage of GFP area relative to the inner tissue area and using the default function Measure. Each data point is the average of three sections from a single nodule.

3. Prime-seq

3.1 RNA extraction

Six biological replicates were collected at 35 dpi for each infected (MG-70, MG-79, MG-136) and non-infected (MG-9, MG-112, MG-136) accession. Each biological replicate comprised 20 nodules from at least five different plants. All material was frozen in liquid nitrogen and ground with two steel beads in a MM400 tissue lyser (Retsch) until fine powder was obtained. Total RNA was extracted using the Spectrum™ Plant Total RNA kit (Sigma-Aldrich, STRN250-1KT) and treated with DNAse I (Ambion) according to the manufacturer's instructions. RNA integrity was examined with a 1% agarose gel in TAE buffer.

3.2 Library preparation

36 libraries were prepared using the prime-sequencing (prime-seq) method (Janjic et al., 2022). For library preparation, 4 ng of RNA were mixed with 5 µl of reverse transcriptase mix (0.15 µl Maxima H Minus reverse transcriptase, 2 µl Maxima RT 5X buffer, 0.4 µl 25 mM dNTP, 0.1 µl 100 µM TSO, and 2.35 µl UltraPure water) and 1 µl barcoded oligo dT (10 µM). The reaction was incubated at 42°C for 90 min. Cleaning of pooled samples was done with homemade SPRI beads (Sera-Mag SpeedBeads) in 22% PEG. The cDNA was treated with exonuclease I for 20 min at 37°C and 10 min of inactivation at 80°C in a final volume of 20 µl. A second purification step was carried out as described above. The cleaned cDNA was amplified by PCR using 25 µl KAPA HiFi 2X RM, 3 µl 10 µM pre-amp primer and 2 µl UltraPure water. Thermocycler conditions were 98°C 3 min, 10 cycles of 98°C 15 s, 65°C 30 s, and 72°C 4 min, followed by 72°C 10 min for final extension. Quality and quantity were assessed using the Agilent 2100 Bioanalyzer with a High Sensitivity DNA Analysis Kit and with Quant-iTPicoGreen dsDNA, respectively. The library was prepared with the Nextera XT Library Prep Kit. Three replicates with 0.8 ng of cDNA were tagmented following the

manufacturer's protocol. A three prime specific primer was used to amplify the barcode and UMI sequences introduced in the reverse transcription step.

3.3 Sequencing and mapping

Libraries were paired-end sequenced at the LAFUGA Gene Center, Munich, Germany with an Illumina HiSeq 1500. Deep sequencing was between 5 to 10 Mio raw reads per sample. Raw data were processed using the zUMIs pipeline (ver 2.5.4) (Parekh et al., 2018) and mapped with STAR aligner (ver 2.6.0) against the reference *L. japonicus* genomes Gifu v1.2 and MG-20 v3.0 obtained from the Lotus Base (https://lotus.au.dk/).

4. RNA-seq downstream analysis

4.1 Differential Expression (DE) analysis

The DESeq2 package (Love et al., 2014b) within R (R Core Team, 2013) was used for differential expression (DE) analysis. A total of nine pairwise comparisons were performed between the transcriptomes of infected and non-infected accessions. For each individual analysis, a False discovery rate (FDR) ≤ 0.05, $\alpha = 0.01$, and a $\log_2$fold change ≥ 1 were set as a threshold to identify differentially expressed genes (DEGs). The UpSetR package was used to find the shared DEGs across all nine pairwise comparisons (Conway et al., 2017). Volcano plots were obtained with the package EnhancedVolcano (Blighe et al., 2021) with a foldchange cutoff of 2 and a p-value cutoff of 10^{-10}. This enabled a quick visualization of transcripts with a large foldchange in the different pairwise comparisons. A second DE analysis using a relaxed $\alpha = 0.05$ was performed to do a gene ontology analysis.

4.2 Gene ontology (GO) analysis

Enrichment analysis for gene ontology (GO) using the output of the relaxed DE analysis ($\alpha = 0.05$) was conducted with the topGO package (Alexa and Rahnenführer, 2009). This package performed a Fisher's exact test. All GO terms were extracted by matching significant DEGs with the gene identifier of the *Lj* Gifu v1.2 gene ontology annotations file. The molecular function was set as the ontology category and only the first 10 nodes were extracted.

4.3 Gene co-expression analysis

Gene clusters that co-expressed together were identified by a weighted correlation network analysis (WGCNA) (Langfelder and Horvath, 2008). A network dendrogram was created with the shared DEGs (first stringent analysis) across all pairwise combinations and normalized expression data of those genes in seven different *Lotus* tissues and treatments:

leaf, mature flower, seed, root, arbuscular mycorrhizal (AM) symbiosis at 15 days post inoculation (dpi), and nodules at 10 and 21 dpi. Expression values for the different conditions were retrieved from the Lotus Expression atlas tool, specifically the RNA-seq data expression atlas data from *L. japonicus* Gifu v1.2 (https://lotus.au.dk/expat/). Due to the low number of genes, a one-step gene network construction and module detection were used. First, a network topology analysis was done to select proper soft thresholding based on a Pearson correlation. Second, the WGCNA function blockwiseModules was used to detect modules of co-expressed genes. The minimum number of genes detected by a module was set to 30 with a standard merging threshold of 0.25. Results were plotted with the function plotDendroAndColors, also within the WGCNA package, and the heatmap for each module with the function pheatmap (Kolde, 2019).

5. DEGs validation

5.1 Heatmap

The relative gene expression of selected DEGs was depicted using a heatmap in different tissues and conditions. Expression data were retrieved from the Lotus base from *Lj* Gifu v1.2 RNA sequencing for seed, flower, leaf, root inoculated with *M. loti* R7A, root mock treatment, AM 15 dpi, AM mock treatment, root hair 24 h post inoculation (hpi), root hair 72 hpi, root hair mock treatment, nodule primordium at 7 dpi, young nodule at 10 dpi, and mature nodule at 21 dpi. The different Heatmaps were built within R with the package "gplots" and the function heatmap.2 (Warnes et al., 2016).

5.2 Quantitative RT-PCR

Differential expression of candidate genes was validated by RT-qPCR. Whole roots and nodules of *L. japonicus* Gifu inoculated with *M. loti* MAFF303099 or mock-treated roots were collected at 3, 7, and 14 dpi, frozen immediately in liquid nitrogen and total RNA was extracted as described in section 3.1. cDNA synthesis was performed according to the manufacturer's instructions using SuperScript III reverse transcriptase (Invitrogen). RT-qPCR was performed in a Quantstudio5 system (Thermo Fisher) in a final volume of 7 µl with 3.5 µl of 2X SYBR Green master mix (Invitrogen, Thermo Fisher Scientific), 1:10 (v/v) dilution of the cDNA, and 0.3 µM of each primer. The thermal cycler conditions were: 95 °C for 2 min, 40 cycles of 95 °C for 30 s, 56 °C for 30 s, and 72 °C for 20 s. Normalized expression was calculated as $2^{-\Delta CT}$ relative to the housekeeping genes *LjPPA2A* or *LjUbiquitin*. RT-qPCR primers used in this study are listed in Supplemental Table 3.

31

6. Phylogenetic analyses

Accession numbers of all protein sequences used in this study are listed in Supplemental Tables 4 and 5. Phylogenies for the Casparian Strip Membrane Domain Proteins-like (CASPL) and Fatty Acyl-CoA Reductases (FARs) were created by retrieving the protein identifiers of published phylogenies for CASPL (Roppolo et al., 2014) and FARs (Rowland and Domergue, 2012) and by adding protein sequences retrieved from other legume proteins. The non-legume species *Arabidopsis thaliana* [At], *Zea mays* [Zm], *Oryza sativa* [Os], *Solanum lycopersicum* [Sl], and *Parasponia andersonii* [Pa] were included in addition to the legumes *L. japonicus* Gifu [Lj], *Medicago truncatula* [Mt], *Cicer arietinum* [Ca], and *Arachis hypogaea* [Ah]. Protein sequences for the legume species were retrieved from the NCBI (https://www.ncbi.nlm.nih.gov/proteins/) and Lotus Base (https://lotus.au.dk/) by using blastP and the respective *L. japonicus* protein sequence as a query. The protein alignment was done in MAFFT using default settings (Rozewicki et al., 2019). Non-conserved regions were removed by manually trimming gaps in the alignment. Maximum-likelihood phylogeny trees were constructed in CIPRES (www.phylo.org/) using default parameters in the RAxML-HPC BlackBox tool version 8.2.12. Trees were displayed with Interactive Tree Of Life v5 (https://itol.embl.de/).

7. Transient expression experiments

7.1 Golden Gate constructs

Primers and plasmid constructs are listed in Supplemental Tables 3 and 6. All promoters and genes were amplified from *L. japonicus* Gifu genomic DNA and cloned using the Golden Gate toolkit (Binder et al., 2014). Primers were designed with the Design Primers tool from the CLC Main Workbench (ver 7.7.3). *In silico* cloning was done in the same software.

7.2 Promoter cloning

A 3kb promoter region of *LjFAR*, *LjCASPL*, *LjRBOHB*, *LjNACD* genes was amplified with Phusion DNA polymerase (Thermo Fisher Scientific) according to manufacturer's instructions. When amplification was unsuccessful a 2 or 1 kb region was cloned instead. The reaction was carried out in a thermocycler (Applied Biosystem, Thermo Fisher Scientific Inc., USA) with the following conditions: denaturing step 98°C for 2 min, then 35 cycles of 98°C for 30 s, a variable Tm per primer combination for 30 s, and 72°C for 3 min, followed by a final elongation at 72°C for 3 min. Tm was calculated with the Tm Calculator (New England Biolabs, ver 1.13.1) per primer combination from Supplemental Table 3. The products were purified with a GeneJET Gel Extraction Kit (Thermo Fisher Scientific Inc.,

USA) according to manufacturer's instructions. Blunt end cloning with StuI or SmaI was employed to insert the fragments into a level 1 pUC57 backbone by cut-ligation in a reaction volume of 15 μl: 1.5 μl 10X T4 Ligase buffer supplemented with ATP, 0.75 μl of restriction enzyme, 0.75 μl of T4 ligase, double distilled water (ddH$_2$O), vector, and DNA fragments with a ratio insert:vector of 3:1. The reaction was performed in a thermocycler (Applied Biosystem) under the following conditions: 50 cycles of 37°C for 5 min and 16°C for 2 min, and a final ligation step of 16°C overnight. Cut-ligation products were transformed into TOP10 *E. coli* competent cells via heat-shock by adding 5 μl of the reaction product and 15 μl of the bacteria. Transformed bacteria were plated in LB medium supplemented with gentamycin (Gm, 15 μg ml^{-1}) and grown for 16 h at 37 °C. Plasmids were purified with NucleoSpin Plasmid EasyPure (Macherey-Nagel,Germany) following manufacturer's instructions. Quality control included restriction digestion based on each particular promoter region and visualized by 1% TAE gel electrophoresis: 120 V, 35 min. Furthermore, plasmids were sequenced using primers M13F and M13R (Supplemental Table 3) using the Sanger method (Sequencing service, Biocentre LMU Munich).

Promoter regions were subsequently moved into level 3 expression vectors containing a β-Glucuronidase GUS reporter gene (DoGUS) or an *NLS-2xYFP* fluorescent reporter via a cut-ligation reaction with the Esp3I enzyme. The cut-ligation reaction, bacteria transformation, plasmid purification and quality control followed the same conditions as the blunt end cloning with some modifications. Antibiotic selection of the transformed *E. coli* TOP10 was kanamycin (Km, 50 μg ml^{-1}) and only restriction digestion was performed as quality control.

7.3 Gene cloning

For the cloning of *LjCASPL* genes, BsaI and BpiI restriction sites were removed from the gene sequence by mutagenizing those sites via PCR and the stop codon (TAG) was removed to fuse the gene with GFP in a C-terminal position. In some cases, the gene of interest was fused at the N-terminal position. PCR amplification and fragment purification were performed as described in section 7.2. Fragments were introduced into a Ll+Bpi pUC57 vector (BB3) by a cut-ligation reaction using BpiI and conditions described in section 6.1.1. Transformed *E. coli* TOP10 bacteria were selected using Gm, 15 μg ml^{-1}. Restriction digestion was done for quality control based on restriction sites of the gene of interest. All level 1 plasmids were sequenced using the Sanger method (Sequencing service, Biocentre LMU Munich). Level 2 assemblies included the *L. japonicus Ubiquitin1* promoter (*LjUbq1pro*), an N-terminal GFP tag, the genes of interest (i.e. *LjCASPL* genes), the 35S terminator, and two dummy sequences. Elements were introduced into a binary expression

33

vector LIIβ F 3-4 Xpre2-S (BB24) using BsaI and following the same protocol as in 6.1.1. Transformed bacteria were selected with Sm, 100 µg ml^{-1} and restriction digestion was done as quality control based on restriction sites of the LII constructs.

7.4 Hairy root transformation

For hairy root transformation, *Agrobacterium rhizogenes* (*A. rhizogenes*) 1193 cells were transformed by electroporation with the level 3 plasmids carrying the promoter reporter constructs. Between 100-200 ng of the plasmid was mixed with 50 µl of *A. rhizogenes* 1193 in an electroporation cuvette and incubated in ice for 20 min. Electroporation was done with a single pulse from the MicroPlulser™ (BIORAD) with the program Ecl (1.25 kV, 400 Ω). Subsequently, 1 ml of YEB medium was added to the *A. rhizogenes* suspension and incubated for 2 h at 28°C and 180 rpm. Finally, 50 µl of the suspension was plated in YEB agar plates supplemented with kanamycin (Km, 50 µg ml^{-1}), rifampicin (Rf, 50 µg ml^{-1}), and carbenicillin (Cb, 50 µg ml^{-1}) and incubated at 28°C for 2 days. Transformants were evaluated by colony PCR using primers targeting the promoter region (Supplemental Table 3).

Transient root transformation was conducted by the hairy root method (Stougaard et al., 1987). The *A. rhizogenes* 1193 carrying the desired construct were plated 24 h before plant transformation. Roots of *L. japonicus* Gifu seedlings, previously germinated as described in section 1.2, were cut and hypocotyls were submerged in an *A. rhizogenes* 1193 suspension (Supplemental Table 6). Treated hypocotyls were placed on ½ Gamborg's B5 agar (Gamborg et al., 1968) and incubated for 2 days in the dark at room temperature. Plates were then moved into a growth chamber under a long-day photoperiod for 3 days. To remove *A. rhizogenes*, hypocotyls were transferred to Gamborg's B5 agar plates supplemented with cefotaxime (300 µg ml^{-1}). After two weeks post transformation, plants were screened for the presence of an *NLS-2xmCherry* or an *NES-2xmCherry* transformation marker under a M165FC stereo microscope (Leica Biosystem) equipped with a red filter. Three weeks after transformation, plants bearing transformed roots were transferred into sterile Weck jars with 300 ml of a sand:vermiculite mixture (1:2) supplemented with 40 ml of a low nitrogen FAB medium. After 2-3 days, plants were watered with 10 ml of FAB medium containing *M. loti* MAFF303099-GFP (OD$_{600}$ = 0.005) for *NLS-2xYFP (*Binary expression vector for promoter *NLS-2xYFP* fusions with *NLS-* or *NES-2xmCherry* transformation marker, Km^{R}*)* or *M. loti* MAFF303099-*Ds*Red for DoGUS (Binary expression vector for GUS with NLS 2xGFP-*lacZ*dy – DoGUS, Km^{R}) as described in section 1.2.

7.5 Promoter GUS assay

Transgenic roots carrying the promoter-DoGUS constructs (Supplementary Table 6) were harvested at 7 and 21 dpi. Roots were incubated in staining buffer containing 100 mg/ml X-Gluc in DMSO (62.5 μl in 10 ml buffer), and buffer contained 100 mM phosphate buffer (pH = 7.0), 0.5 mM EDTA (pH = 7.0), 0.5 mM $K_3[Fe(CN)_6]$ and 0.5 mM $K_4[Fe(CN)_6]$ at 37 °C for 8 h to 24 h. After staining, plant material was fixed with 2.5 % glutaraldehyde in 0.1 M (pH = 7.0) sodium phosphate buffer with vacuum infiltration (Cerri et al., 2012). Roots were inspected with a VHX-6000 digital microscope (Keyence) and pictures were taken at 20X, 50X, and 200X.

7.6 Confocal microscopy

Transformed roots carrying the *NLS-2xYFP* reporters were harvested at 10 and 21 dpi. Roots were fixed with a 4% paraformaldehyde solution in PIPES buffer (50 mM, pH = 7) and vacuum infiltrated for 1 h at room temperature. Single nodules from transformed roots were embedded in 7% low melting agarose (Roth) and cut with a VT1000S vibratome (Leica Biosystems) in 65-μm-thin sections. A TCS SP5 confocal microscope (Leica Microsystems) equipped with a 20x HCX PL APO water immersion lens was used to look for reporter signal inside the nodules. Secondary cell wall components were excited with a diode lamp and detected at 405-450 nm. GFP, from the tagged *M. loti* and the *NLS-2xYFP* reporter, was excited with an argon laser at 488 and 514 nm, respectively, and 20% power. For GFP and YFP the emission was detected at 493-515 and 520-540 nm, respectively.

7.7 Subcellular localization

Agrobacterium tumefaciens (*A. tumefanciens*) AGL1 was transformed as described in section 6.3 with the level 2 plasmids by electroporation. Bacteria were incubated at 28°C for 2 days and selected with Sm, 50 μg ml^{-1}; Rf, 50 μg ml^{-1}; and Cb, 50 μg ml^{-1}. Transformants were evaluated by colony PCR using the cloning primers targeting one region of the gene (Supplemental Table 3).

Transient expression of protein fusions was performed in epidermal cells of *Nicotiana benthamiana* leaves. Constructs for expression of CASPL proteins fused to GFP in N- or C-terminal position were introduced to *A. tumefaciens* AGL1 by electroporation as described in section 6.1. Three-weeks old plants were infiltrated with a mix of the *Agrobacterium* carrying the desired construct and the post-transcriptional gene silencing inhibitor P19 in equal amounts to reach a final OD_{600} = 0.1. Leaf discs were observed 36 hours post infiltration using confocal microscopy as described in section 7.6. GFP was

excited at 488 nm and detected at 515 nm. Plasmid constructs are listed in Supplemental Table 6.

8. Molecular characterization of the *Ljfar3.2* LORE mutant lines

8.1 DNA extraction and genotyping

Two independent *LORE1* lines with insertions in *LjFAR3.2* (Supplemental Table 7) were ordered from the Lotus Base (https://lotus.au.dk) from a segregating population. To identify mutant lines within a segregating population, primers flanking the insertion of the retrotransposon element were used. For both *LORE1* lines genomic DNA was extracted using a hazard free protocol (Kotchoni and Gachomo, 2009). A single young *Lotus* leaf was cut, frozen in liquid nitrogen and ground for 1 min at 30 Hz in a MM 400 tissue lyser (Retsch). Fine powder was suspended in extraction buffer (1% SDS and 0.5 M NaCl). Subsequently, the mixture was centrifuged for 1 min at 12,000 g and the supernatant was transferred into a new tube where it was precipitated with isopropanol at a 1:1 ratio. A second centrifugation step was performed to obtain the pellet, which was then rinsed with 70% ethanol. Air-dried DNA was suspended in ddH$_2$O. Standard PCR using 0.25 units of GoTaq (Promega) was performed. PCR amplification used the following program: denaturing step at 95°C for 2 min, then 35 cycles of 95°C for 30 s, variable Tm according to the primer combination for 30 s, and 72°C for 5 min, followed by a final extension step at 72°C for 5 min. Tm was calculated as in section 6.1.1. The products were analyzed by gel electrophoresis in 1% agarose gel at 120 V for 40 min. Mutant plants carrying the retrotransposon element as well as the respective wild-type (WT*) plants containing background mutations were selected for greenhouse propagation. All primers were obtained from the Lotus Base (https://lotus.au.dk/) (Supplementary Table 3).

8.2 Growth and inoculation conditions of mutant plants

The *far3.2-1* and *far3.2-2* mutant lines and their respective wild-types were germinated and inoculated with *M. loti* MAFF303099-*Ds*Red as described in section 1. Symbiotic phenotypes were quantified 21 dpi. These included number of nodules, number of ITs, shoot length, root length, and more specific phenotypes described below.

8.3 Nodule permeability assay

Intact nodules were incubated for 30 min in a 0.1% Toluidine blue solution in water. Nodules were then sectioned (100 μm) as previously described in section 6.3. Dye diffusion was observed using a CTR 6000 upright microscope (Leica Microsystems). Permeability to the dye was estimated as a permeability ratio (PR). The PR was defined as PR = td/nd where

nodule distance (nd) indicates the distance from the nodule border to the closest infected cell border and the Toluidine blue distance (td) measures the distance that the dye penetrates. nd and td were measured in Fiji (Schindelin et al., 2012).

8.4 Nodule staining

The staining was adapted from a protocol to detect secondary cell wall modifications in Arabidopsis roots (Ursache et al., 2018). Nodules were submerged overnight in a ClearSee solution supplemented with either 0.2% Basic Fuchsin or 0.05% Nile Red for detection of lignin and suberin, respectively. Nodules were then washed in a ClearSee solution for 1 h with constant shaking. The solution was replaced every 20 min. Fluorol yellow staining to visualized suberin in *L. japonicus* nodules was performed as described in Sexauer, Shen, Schön, Andersen, & Markmann, 2021. Nodules at 21 dpi were fixed and cleared using a ClearSee solution. Subsequently, nodules were embedded in a 6% (w/v) agarose solution (NuSieve™ GTG™), sectioned (100 μm thickness) in a VT1000S vibratome (Leica), and stained with Fluorol yellow 088. Signal intensity in the whole nodule endodermis was quantified using ImageJ as described in (https://theolb.readthedocs.io/en/latest/imaging/measuring-cell-fluorescence-using-imagej.html).

8.5 Oxygen measurements

A pre-calibrated profiling oxygen microsensor PM-Pst7 (PreSens, Regensburg, Germany) was used to measure oxygen concentration inside nodules. Recording was done with the PreSens Measurement Studio2 (PreSens, Regensburg, Germany). Fresh nodules were embedded in 7% low melting agarose (Roth). The microelectrode was positioned perpendicularly at the top of the nodule using a manual micromanipulator (PreSens, Regensburg, Germany). Measurements were taken at the surface (0 µm) and the inner nodule cortex (75 µm). For each point, seven to eight measurements were taken under room temperature conditions. The average of all replicates was plotted.

8.6 Acetylene reduction assay (ARA)

Nitrogenase activity was evaluated by the reduction of acetylene into ethylene and detected using Gas Chromatography - Flame Ionization Detection (GC-FID). Mutant and WT* plants were inoculated with *M. loti* MAFF303099-*Ds*Red, as described in section 1, and harvested at 21 dpi. Five biological replicates were analyzed. A single replicate comprised two nodulated roots in a 25 ml glass tube with 500 µl of FAB medium and sealed with a rubber stopper. Subsequently, 1 ml of air was extracted and replaced with 1 ml of acetylene. Then, 1 ml of the mixture was injected into a GC 2010 Pro (Shimadzu). Five different time points

were measured at 0, 20, 40, 60, and 80 min while keeping the samples at 28 °C in a water bath. Using regression analysis, the area under the curve was converted into nanomoles of ethylene based on an ethylene standard curve. The nanomoles of ethylene per hour were obtained by multiplying the slope of a linear regression model (*Ethylene nanomoles = mt + b,* where m: slope, t: time and b: intercept) by 1 h (60 min). The statistical analysis was performed with R.

9. CRISPR/Cas12a gene editing in *L. japonicus* Gifu

9.1 Guide RNA (gRNA) design

Single and double mutant lines were generated using the CRISPR/Cas12a gene-editing technology. Guide RNAs (gRNAs) targeted the first exon and the 3'UTR region of each gene to generate a gene deletion when possible. The gRNAs were designed using the CRISPOR website (http://crispor.tefor.net/) (Concordet and Haeussler, 2018). The parameters were set as follows: *L. japonicus* as the reference genome and the protospacer adjacent motif (PAM) was 'TTT(A/C/G)-21bp-Cas12a (Cpf1)-21bp guides-recommend by IDT'. The gRNAs were selected based on high efficiency and a low number of off-targets. For cloning BpiI and BsaI, recognition sites were added to the 5'- and 3'-ends.

9.2 Plasmid cloning

Cloning was done using the Golden gate cloning toolbox (Binder et al., 2014). Each gRNA and its reverse complement sequence were ordered from Sigma Aldrich and hybridized at 98°C for 5 min followed by cooling at room temperature. Hybridized gRNAs were cloned into a level 1 backbone LI-BpiI entry plasmid via cut-ligation (section 6.1.1). Restriction digestion was done to confirm the insert of the plasmid based on the restriction sites of each particular construct. Level 2 plasmids were constructed by assembling two *LjU6* promoters driving the expression of each gRNA and "gSNR2" terminator via BsaI cut-ligation into BB24. Level 3 plasmids contained a hygromycin selection marker driven by a *Nos* promoter, the Cas12a coding gene driven by the *LjUbi1* promoter and the level 2 element containing the gRNAs driven by the *LjU6* promoter. All elements were inserted into a LIIIβ fin Xpre2-K (pCAMBIA) (BB52) expression vector via a BpiI cut-ligation reaction. Plasmids were purified and sequenced according to the procedure described in section 7.2.

9.3 Stable transformation

A. tumefaciens AGL1 was transformed with the constructs as described before (section 6.2). *L. japonicus* Gifu hypocotyls were stably transformed with *A. tumefaciens* AGL1 (Handberg et al., 1994 and M. Bircheneder, personal communication). Seedlings were

germinated in ½ B5 plates (section 1.2). *A. tumefaciens* AGL1 cells carrying the desired constructs were grown on LB plates with appropriate antibiotics (Km, 50 μg ml^{-1}; Rf, 50 μg ml^{-1}; and Cb, 50 μg ml^{-1}) for 24 h at 28 °C before transformation. Bacteria were collected from the plate and resuspended in 4 ml sterile YMB liquid medium and 40 μl of phosphate-buffer (0.3 M, pH 6.8). A sterilized blotting paper (Grade.: BF4, AHLSTROM MUNKSJÖ) was saturated with the bacterial suspension. The seedlings were placed on top of the paper and their hypocotyls were cut into 3 mm fragments. Cut hypocotyls were incubated on the paper for 10 min after which they were transferred onto a new plate with 5 blotting papers soaked in co-cultivation medium (Supplemental Table 8). The plate was sealed with parafilm and placed for 6 days in dark conditions at 20°C.

After 6 days, the hypocotyls were moved to a new plate containing callus induction medium (Supplemental Table 8) for callus formation and selection. The plates were placed in a phytochamber (MLR-352H-PE, Panasonic) with a 16h light/8h darkness photocycle. Transformed calli displaying a green color were transferred to fresh callus medium once a week for 6 to 8 weeks. Non-transgenic white or brown calli were discarded.

Shoots were induced by moving the green transgenic calli onto 12-well plates (one or two calli per well) with shoot induction medium (Supplemental Table 8). Calli were maintained in plates for 4 to 8 weeks until leaf-like structures developed. Then they were moved to 12-well plates (one callus-shoot per well) containing shoot elongation medium. Calli were kept on the plates for 3 to 6 weeks until the shoots were circa 0.5 cm.

The shoots were cut and transferred to 12-well plates (one shoot per well) containing root induction medium (Supplemental Table 8). Plantlets were incubated for 10 days until white swellings on the base of the shoot developed. Calli with root primordia were transferred into Magenta Boxes (Magenta GA-7 Plant Culture Box, PlantMedia) containing 50 ml root elongation medium (Supplemental Table 8). The plantlets were incubated in the semisolid root elongation medium for 3 to 5 weeks until the roots were 2 to 3 cm. Regenerated plants were transferred to pots with Stender soil substrate (A210, Stender GmbH) fertilized with Osmocote Exact Standard fertilizer (3 g fertilizer per litter substrate). They were kept in a growth chamber under a long-day photoperiod for 2 to 4 weeks before being transferred to the green house for seed production.

9.4 Plant growth

Plants in the greenhouse were grown in Stender soil substrate. Light intensity was 400W with additional light from 6 to 10 am and 3 to 10 pm. The day and night temperatures ranged from 18 to 24°C. Plants were grown for 5 to 8 months for seed production.

9.5 Genotyping

DNA was extracted from transformed plants as described before (section 7.1). Flanking primers were designed to amplify a 300 to 700 nt region surrounding the gRNA targeting site and the genome editing was supposed to take place. PCR amplification used the same conditions as in section 7.1. Tm was calculated as in section 7.2 and based on primers listed in Supplemental Table 3. PCR product was purified using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) according to manufacturer's instructions and sequenced with the Sanger method (Sequencing service, Biocentre LMU Munich). All primers were designed with CLC Main Workbench v7.7.3 (Supplemental Table 3).

9.6 Phenotyping of *Ljcasp4.1 caspl4.2* mutant line

A heterozygous double mutant *Ljcasp4.1 caspl4.2* line was germinated in ½ B5 medium and grown and inoculated with *M. loti* MAFF303099-*Ds*Red as described in section 1. Recording of the phenotype included shoot length, root length, and number of nodules. In addition, nodule sections were performed as described in section 2.3. Genotyping was performed for individual plants as described in section 8.1.

10. Statistical analysis

All statistical analyses were conducted in R (RDevelopment, 2012). The following packages were used: Tukey's HSD and ANOVA tests were performed with the package agricolae (De Mendiburu and Simon, 2015). The Wilcox-test and Krustal-test were performed with the function compare_mean from the package ggpubr (Kassambara and Kassambara, 2020).

Results

1. *Lotus japonicus* accessions displayed contrasting nodule infection phenotypes

In order to identify *Lotus japonicus* accessions in which nodule organogenesis and nodule infection were uncoupled, 40 accessions were qualitatively and quantitatively phenotyped after inoculation with *Rhizobium leguminosarum* Norway (Table 2). Plants were harvested 35 dpi as *R. legumminosarum* Norway takes at least 3 weeks to induce sizable nodules on *Lotus* (Liang et al., 2019). The presence of nodule primordia, nodules, and tumors was assessed. Irregular or multilobular organs were considered as tumors, whereas organs with a well-defined round shape were classified as nodules (Table 2). As described for *Lotus burttii* (Liang et al., 2019), *R. leguminosarum* Norway did not induce epidermal infection threads in any of the *L. japonicus* accessions. To facilitate the visualization of nodule infection by *R. leguminosarum* Norway-GFP, nodulated roots were immersed in a ClearSee solution that reduces plant tissue autofluorescence (Kurihara et al., 2015). Although *R. leguminosarum* Norway induced nodule organogenesis in all accessions, nodule infection varied among them, ranging from highly infected to almost complete absence of bacteria (Table 2).

Table 2. Qualitative screening of *Lotus japonicus* accessions. All accessions were inoculated with *Rhizobium leguminosarum* Norway–GFP. After 5 weeks nodule organogenesis (primordia, nodules, and tumors) and infection phenotypes were scored. To facilitate the visualization, nodules were treated with the ClearSee method (Kurihara et al. 2015) before observation under a fluorescence microscope. Black and white boxes indicate presence and absence of the phenotype, respectively. *Accessions selected for prime-seq.

Ecotype	Nodule	Bumps	Tumor	Infected	Ecotype	Nodule	Bumps	Tumor	Infected
MG-7	■	■	■	■	MG-106	■	■	■	■
MG-9*	■	□	■	■	MG-110	■	□	■	■
MG-25	■	■	■	□	MG-111	■	□	■	■
MG-28	■	■	□	■	MG-112*	■	□	□	□
MG-29	■	■	■	■	MG-113	□	□	■	■
MG-50	■	■	■	■	MG-115*	■	■	□	□
MG-52	■	■	■	□	MG-119	■	□	□	■
MG-55	■	■	□	■	MG-123	■	□	□	■
MG-60	■	■	□	■	MG-128	■	□	□	■
MG-64	■	■	■	□	MG-129	■	■	■	□
MG-70*	■	□	■	■	MG-133	■	■	■	■
MG-78	■	□	■	■	MG-135	■	■	■	■
MG-79*	■	□	■	■	MG-136*	■	■	■	■
MG-81	■	■	■	■	MG-137	■	■	■	■
MG-83	■	□	□	■	MG-139	■	□	■	■
MG-85	■	■	□	■	MG-140	■	■	■	■
MG-86	□	■	□	■	MG-143	□	■	■	■
MG-102	■	■	■	■	MG-142	■	■	■	■
MG-103	■	■	■	■	MG-144	■	■	□	■
MG-104	■	■	■	■	MG-146	■	■	■	■

To quantify nodule infection, nine accessions that harbored nodules with a regular shape and had qualitative differences in the nodule infection phenotype were selected. First, the number of infected nodules was quantified. Some accessions were more susceptible to infection by *R. leguminosarum* Norway-GFP than others (Figure 2A and 2B), as supported by a Tukey's honestly significant difference (HSD) test, which revealed five different significance groups (Figure 2A). As the percentage of nodules infected does not necessarily reflect the quantity of bacteria inside nodule cells, nodules were sectioned and the nodule area colonized by *R. leguminosarum* Norway-GFP was quantified. Based on the HSD test, two groups of samples were selected: i) the infected group that included accessions MG-70, MG-79, and MG-136, which had the highest percentage of infected nodules, and ii) the non-infected group that contained accessions MG-9, MG-112, and MG-115 with the fewest infected nodules. In addition, the area of *L. japonicus* Gifu nodules infected by *M. loti* MAFF303099-*Ds*Red, a compatible symbiont (Kaneko et al., 2000), was also quantified. Nodules infected with *M. loti* MAFF303099-*Ds*Red showed colonization with a mean value of around 60%. In comparison, the first group of accessions was significantly more colonized by *R. leguminosarum* Norway-GFP (mean = 20%) than the second group (mean = 5%) (Figure 2C and 2D). Altogether these results indicate that upon inoculation with *R. leguminosarum* Norway, *L. japonicus* accessions can display similar nodulation phenotypes, but significantly differ in the degree of infection. This suggests that the organogenesis and infection programs can be uncoupled under specific conditions.
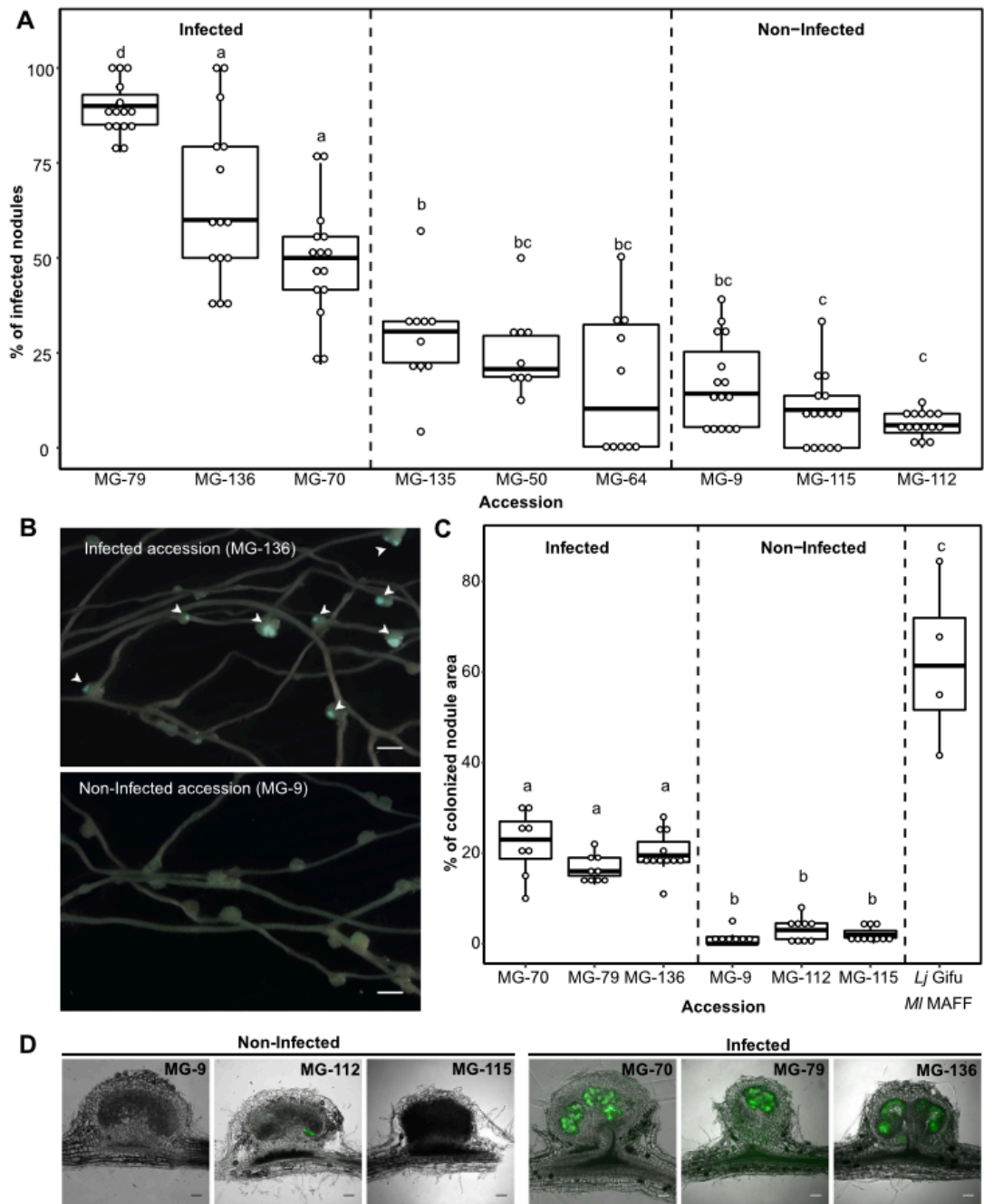
Figure 2. Quantification of the nodule infection phenotype of different *Lotus japonicus* accessions inoculated with *Rhizobium leguminosarum* Norway. A) Representative images of nodules for one infected (MG-136) and one not infected accession (MG-9). Plants were harvested 5 weeks after inoculation with *R. leguminosarum* Norway-GFP. Scale bar = 7 mm. Arrowheads indicate infected nodules. B) The percentage of *R. leguminosarum* Norway-infected nodules of 10 to 15 plants of selected *L. japonicus* accessions with well-defined nodules was quantified and displayed in a boxplot. Results from Tukey's honestly significant difference (HSD) test are displayed as lowercase letters. C) Quantification of nodule colonization of *L. japonicus* (*Lj*) accessions with *R. leguminosarum* Norway-GFP and *M. loti* MAFF 303099-*Ds*Red (*Ml* MAFF). Nodule colonization was quantified 5 weeks post inoculation by measuring the infected area in the nodule inner tissue (n=8 to 12). A Tukey's HSD test (p = 0.01) was applied and the differences are displayed as lowercase letters. D) Representative images of nodule sections for the infected (MG-79, MG-70, MG-136) and non-infected (MG-9, MG-112, MG-115) accessions. Plants were harvested 5 weeks after inoculation with *R. leguminosarum* Norway-GFP. Scale bar = 100 µm

43

2. Natural diversity of *L. japonicus* in response to *R. leguminosarum* Norway as an approach to identify new players during bacterial accommodation

The infection phenotype triggered in *L. japonicus* accession by *R. leguminosarum* Norway revealed accessions that were either highly or poorly colonized (Figure 2). It was hypothesized that the expression of several genes differs across these accessions and therefore it could be used to identify genes associated with the accommodation of rhizobia inside nodules. Based on the results of the *L. japonicus* accessions screening, an experiment was designed to look at the transcriptome of nodules with contrasting infection phenotypes (Figure 3). At 35 dpi all *L. japonicus* accessions had a mix of primordium and fully developed nodules. Only nodules were collected for RNA-seq and their transcriptomes were explored using a method called prime-seq (Janjic et al., 2022). Different bioinformatic tools and methodologies were used to identify and validate the candidate genes.
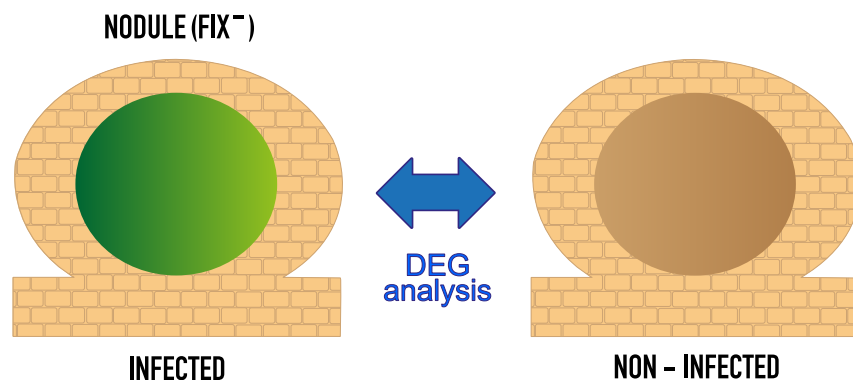

Figure 3. Experimental setting for the identification of genes required during bacteria uptake and accommodation. Scheme of the experimental design for this project. At 35 dpi *R. leguminosarum* Norway has completely colonized the nodule and it is at this stage when genes required for the accommodation can be identified. Six biological replicates were obtained for each accession. Differentially expressed genes (DEG) analysis was used to compare the transcriptomes.

3. prime-seq as a tool to obtain the transcriptome of different *Lj* accessions

prime-seq is a sensitive bulk RNA sequencing method based on the molecular crowding single-cell RNA barcoding and sequencing (mcSCRB-seq) protocol, which uses oligodT priming, early barcoding, and unique molecular identifiers (UMIs) to efficiently generate 3' tagged RNA-seq libraries (Bagnoli et al., 2018; Janjic et al., 2022). This method was used to sequence the nodule transcriptomes of the top infected (MG-70, MG-79, and MG-136) and non-infected (MG-9, MG-112, and MG-115) accessions (Figure 2D).

Nodules infected with *R. leguminosarum* Norway-GFP were selected for the infected group, whereas nodules completely lacking infection were sampled from the non-infected group.

Six biological replicates were taken from each accession comprising a total of 36 samples. A single biological replicate pooled 20 nodules from five plants at 35 dpi. Total RNA was subjected to prime-seq library preparation (Janjic et al., 2022) and the resulting average 4.2 million 3' cDNA reads/sample were mapped to the reference genome of *L. japonicus* Gifu v1.2 (Kamal et al., 2020) using the *zUMIs* pipeline (Parekh et al., 2018) with the splice-aware aligner STAR. The *L. japonicus* MG-20 v3.0 reference genome (Sato et al., 2008) was used as a comparison yielding similar results as with *L. japonicus* Gifu (Figure 4).
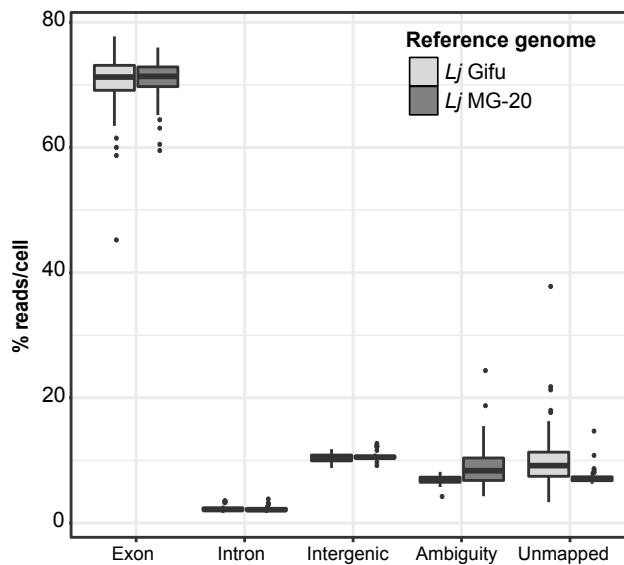

Figure 4. Mapping comparison between the *Lotus japonicus* Gifu and MG-20 genomes. Percentage of reads per cell was plotted for five categories of detected reads: Exon, Intron, Intergenic, Ambiguity and Unmapped. *Lj*, *Lotus japonicus*.

Transcript reads were assigned to different mapping categories. Reads were mapped to exonic (~ 65 %), intronic (~ 4 %), and intergenic (~ 10 %) regions similarly across samples and accessions (Figure 5). Using reads mapping to exonic and intronic regions we detected on average 571,467 UMIs and 19,547 genes per sample and 29,670 genes in total, which is over 98% of the genes annotated in *L. japonicus* Gifu v1.2 (Kamal et al., 2020). Thus, the sequencing approach captured the majority of the genomic regions of all the accessions by employing *L. japonicus* Gifu as a reference genome.
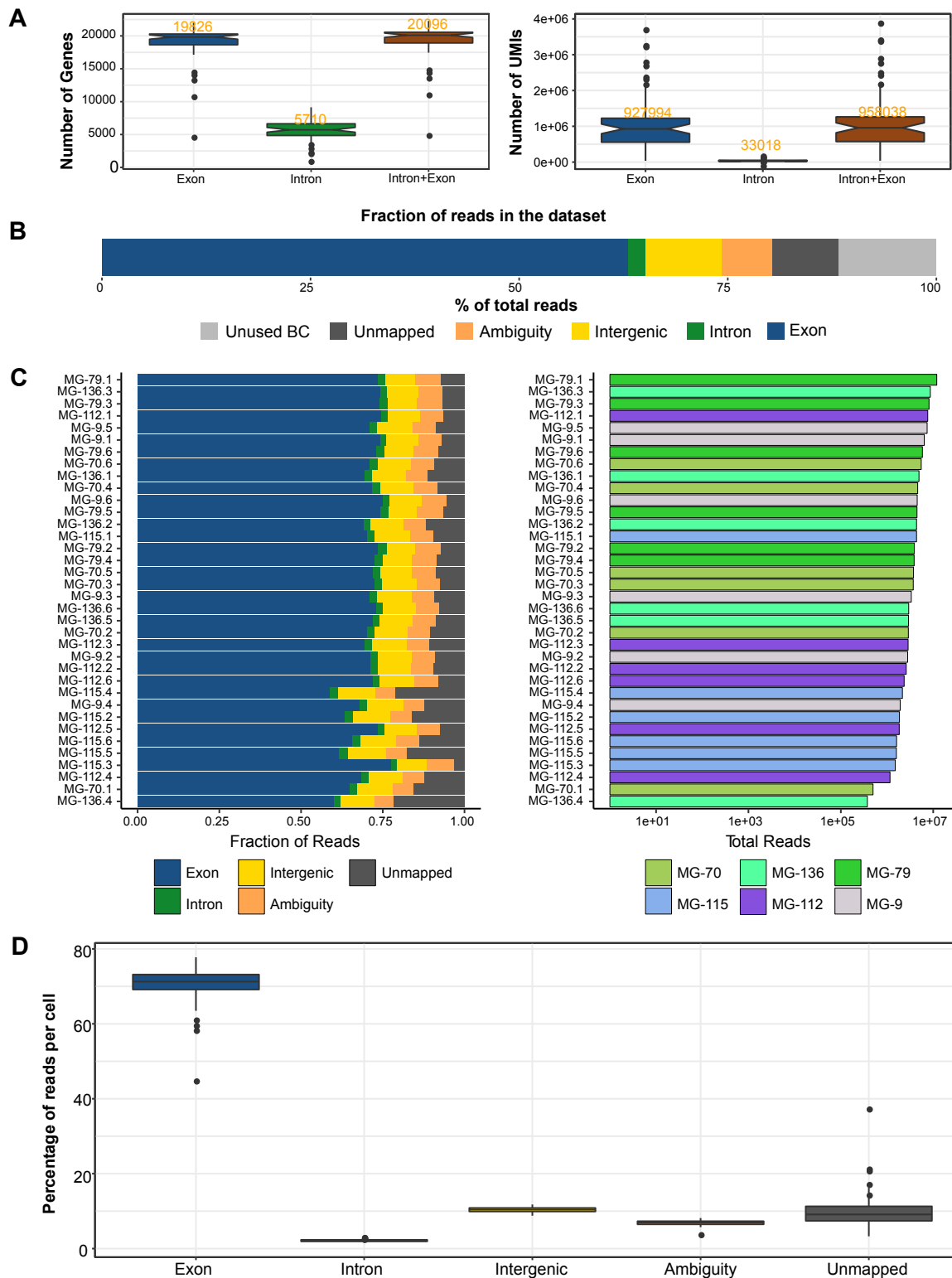
Figure 5. Mapping statistics for prime-seq. A) Number of detected genes and unique molecular identifiers (UMIs) and the read distribution in different genomic regions. These regions include exonic, intronic, and intergenic regions as well as ambiguous and unmapped ones. B) Percentage of the reads per barcoding cell in different genomic regions. C) Distribution of reads for each accession including six biological replicates. D) Percentage of reads per cell. All results were obtained directly as the output of the zUMI pipeline (Parekh et al., 2018).

4. The transcriptomes of infected and non-infected nodules are significantly different

To visualize the variation and data distribution, we plotted the first and the second principal components, which explained more than 60% of the variation in the dataset (Figure 6A). With the exception of one outlier, two groups clustered mainly by the infection status of a sample, suggesting that this is the main factor influencing the variation in data. To identify differentially expressed genes (DEGs), nine pairwise comparisons between infected and non-infected accessions were performed and visualized using an upset plot (Figure 6B) and volcano plots (Figure 7).
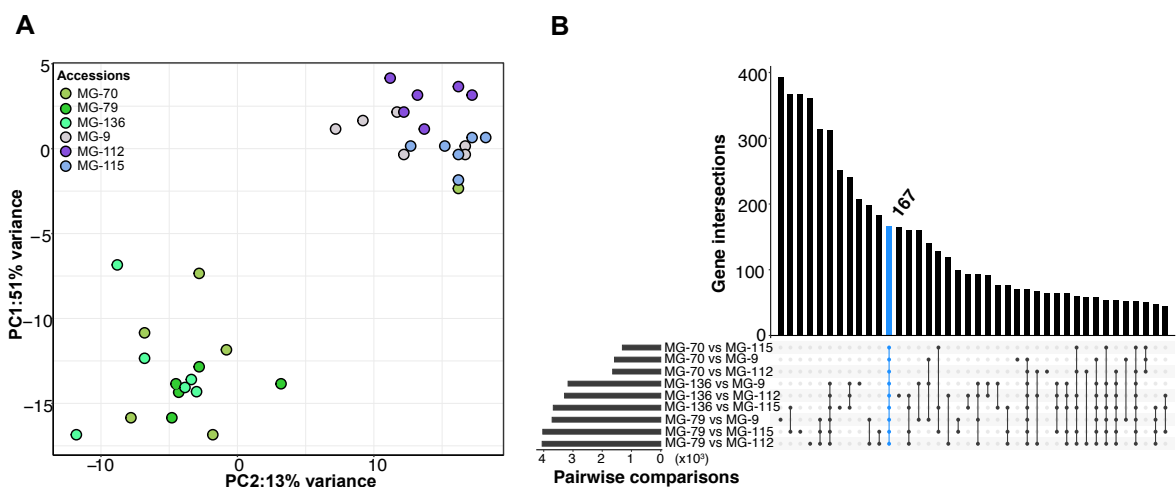

Figure 6. PCA and identification of common DEGs in two contrasting infection phenotypes. A) PCA showing the variation between the transcriptomes of infected and non-infected groups by plotting the first two principal components using the package DESeq2 (Love et al., 2014a). The infected accessions are depicted in shades of green whereas the non-infected ones are depicted in shades of purple. B) Upset plot for the DEGs shared among nine pairwise comparisons. The number of DEGs in each pairwise comparison is observed at the bottom left side of the plot. The blue bar indicates the intersection for all 9 pairwise comparisons. Plot was created with the function "upset" from the UpSetR package (Conway et al., 2017).
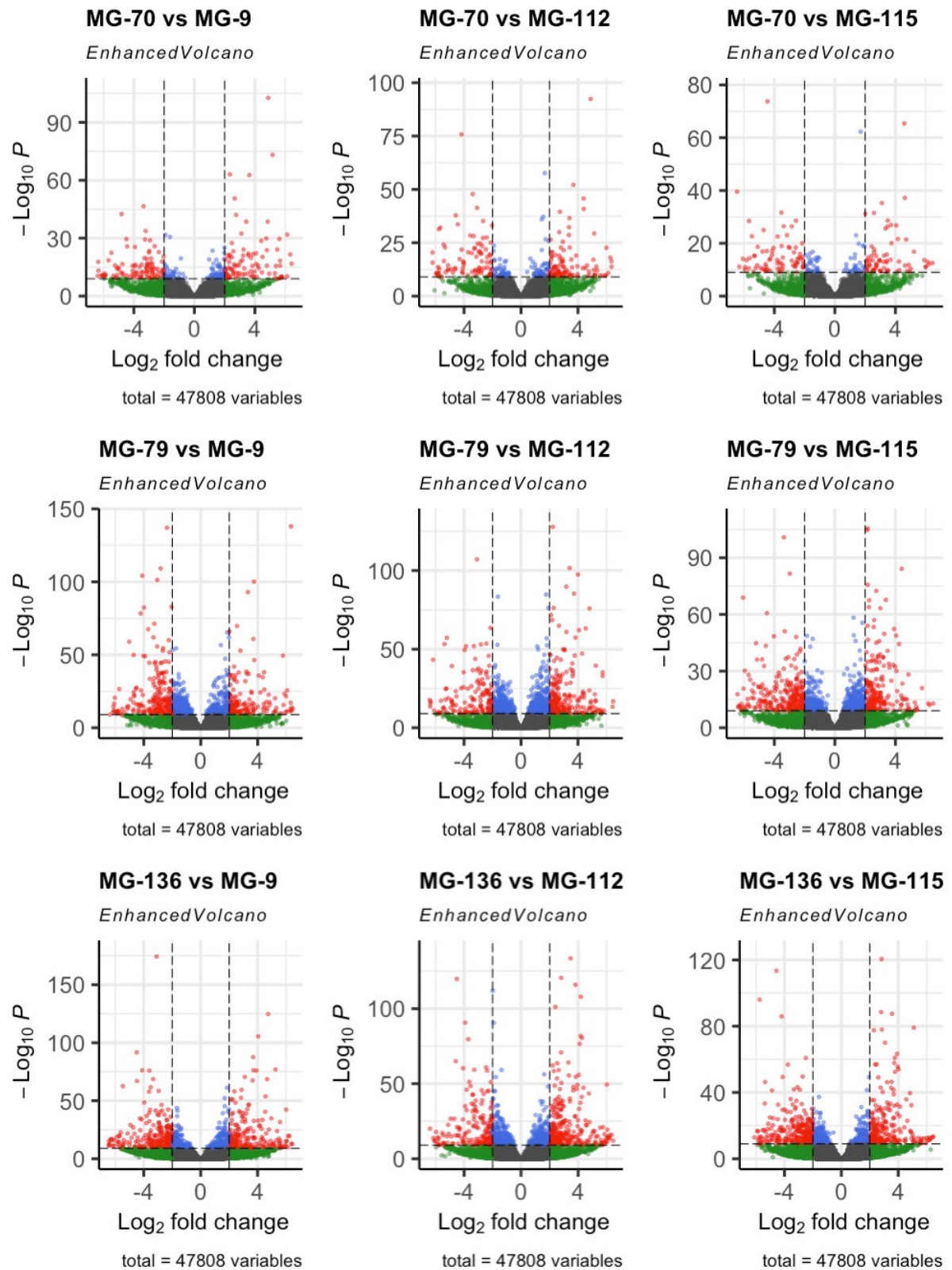
Figure 7. Volcano plots for the nine pairwise comparisons. Scatterplot of the $Log_2FoldChange$ (x-axis) against the $-log10$ p-value (y-axis). Genes with a cutoff below or above 2 for $Log_2FoldChange$ are indicated as grey and green dots, respectively. Genes with a cutoff higher than 10^{-10} for p-value with a $Log_2FoldChange$ below or above 2 are indicated as blue and red dots, respectively.

Comparisons generated between 1,310 and 4,053 DEGs. Common DEGs were defined as the intersection of DEGs with at least a 2-fold change (Supplemental Table 9). This indicates that 167 genes were shared among the 9 pairwise comparisons (Figure 6B). These common DEGs showed contrasting expression patterns between the two infection phenotypes and contained 53 down-regulated and 114 up-regulated genes with reference to the infected group (Figure 8). Several genes encoding proteins with already reported functions in RNS were identified, including the receptor kinase SYMRK (Stracke et al., 2002), the GRAS protein NSP2 (Kaló et al., 2005), three symbiotic leghemoglobins (Ott et al., 2005), and the symbiotic sulfate transporter SST1 (Krusell, 2005). Hence, the identification of already reported genes served as a proof of concept for the approach taken here.
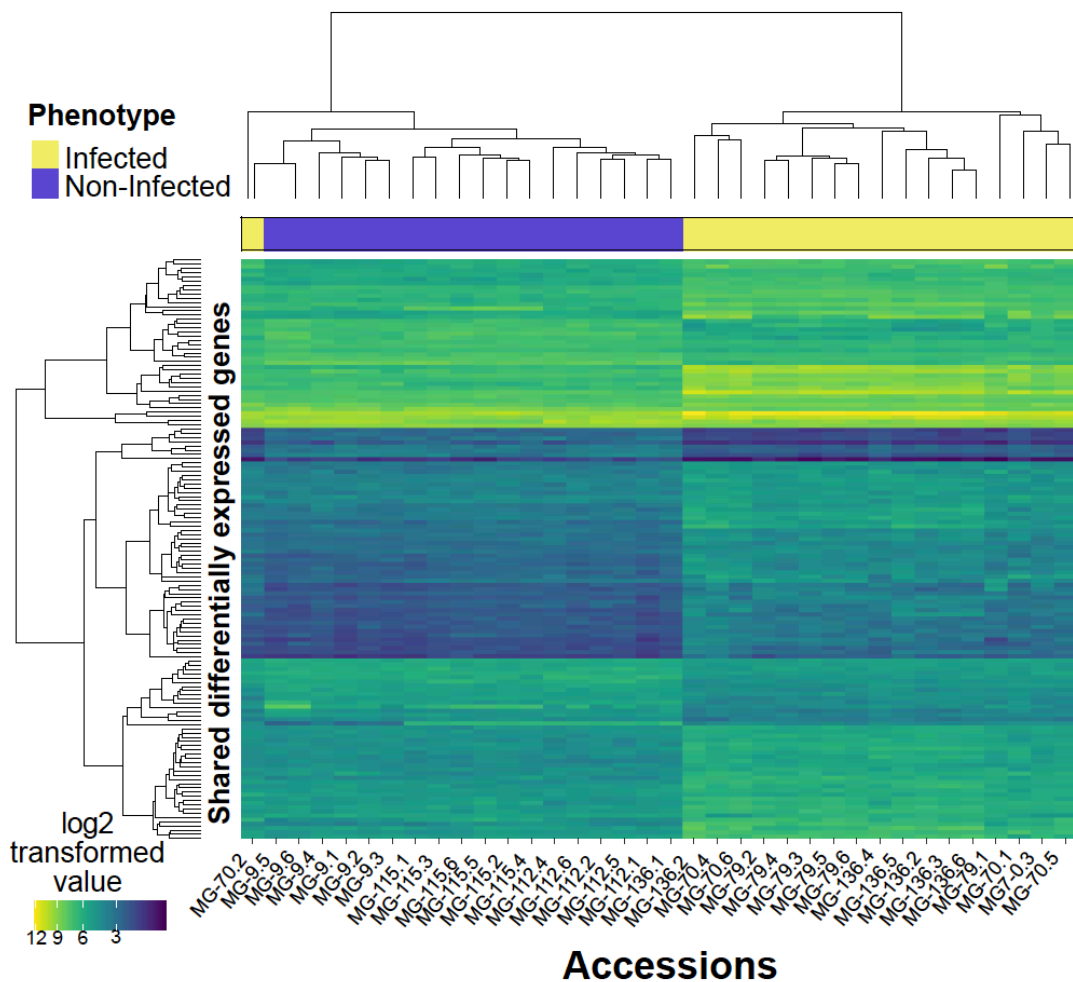

Figure 8. Transcriptomic variation in nodules with contrasting infection phenotypes. Heatmap for the 167 DEGs (p-adjust < 0.01) in different biological replicates from six different accessions. The hierarchical clustering dendrograms for the accessions and the DEGs are displayed at the top and left side, respectively.

To detect functional gene categories a Gene Ontology (GO) analysis was performed. A less stringent DE analysis (p-adjusted < 0.05) was conducted and 1,774 DEGs were used to run

the GO analysis of the dataset. Molecular function was set as the term for the analysis. Among the represented GO categories, there were processes associated with oxidation-reduction reactions (GO:0016684, GO:0004601, GO:0016491), transcription factors (GO:0003700, GO:0140110), and regulation and binding to coordinate compounds (GO:0046906, GO:0020037, GO:0048037) (Figure 9). Altogether these results revealed that genes with oxido-reductase functions were abundant among differentially regulated genes when comparing transcriptomes of infected and non-infected nodules.
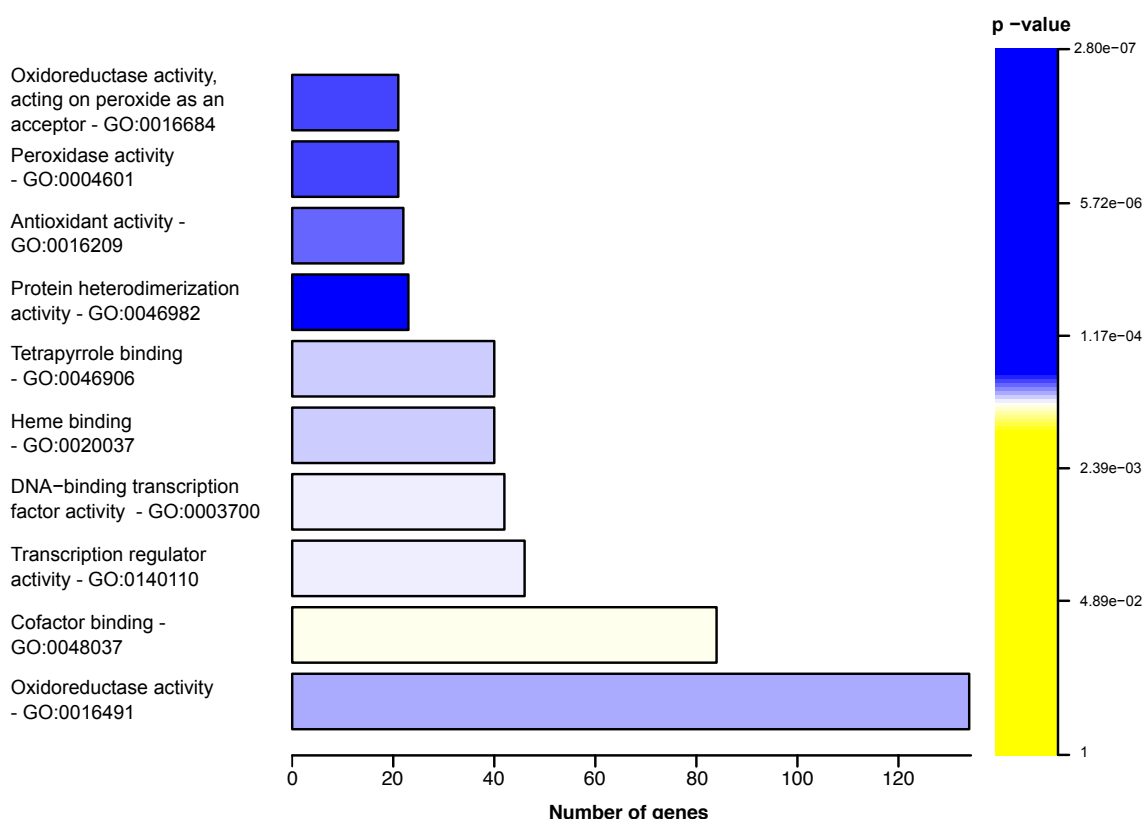

Figure 9. Gene ontology analysis. The analysis used DEGs from a less stringent analysis (p-adjust < 0.05). The top ten nodes are displayed in the bar plot. The p-value scale bar indicates the statistical significance of each node. The analysis was done with the package TopGo (Alexa and Rahnenführer, 2010) and the gene ontology (GO) terms were extracted from the gaf file and the *Lotus japonicus* Gifu v1.2 GO annotation file (https://lotus.au.dk/data/).

5. DEGs cluster in three co-expression modules

To identify associations between the DEGs as clusters of co-expressed genes, a Weighted Gene Co-Expression Network Analysis (WGCNA) was performed, as genes involved in the same processes are often co-expressed (Wolfe et al., 2005). This analysis separated genes into co-expression modules using network topology (Zhang and Horvath, 2005). Available transcriptome data from the Lotus base (https://lotus.au.dk/) were extracted from four different tissues (leaf, mature flower, seed, and root) and three inoculation conditions (roots

15 dpi with arbuscular mycorrhiza (AM), and nodules at 10 and 21 dpi with *M. loti* R7A) for the 167 DEGs. Three modules were generated with module 1 containing 92 genes, and the second and third modules having 28 and 47 genes, respectively (Figure 10A). Genes in module 1 clustered separately as they had a weak correlation with genes in the other two modules, whereas modules 2 and 3 grouped genes with a moderate Pearson correlation ($R^2 = 0.66$). Within the three modules, module 1 had the majority of the downregulated genes from the 167 DEGs and for the other two modules most of the genes were upregulated. Module 2 included genes that have their highest expression in nodules at 10 dpi. In contrast, module 3 harbored genes that are slightly expressed at 10 dpi, but are strongly upregulated at 21 dpi (Figure 10B). As genes in these two modules are specifically expressed in nodules, they might play roles during RNS.
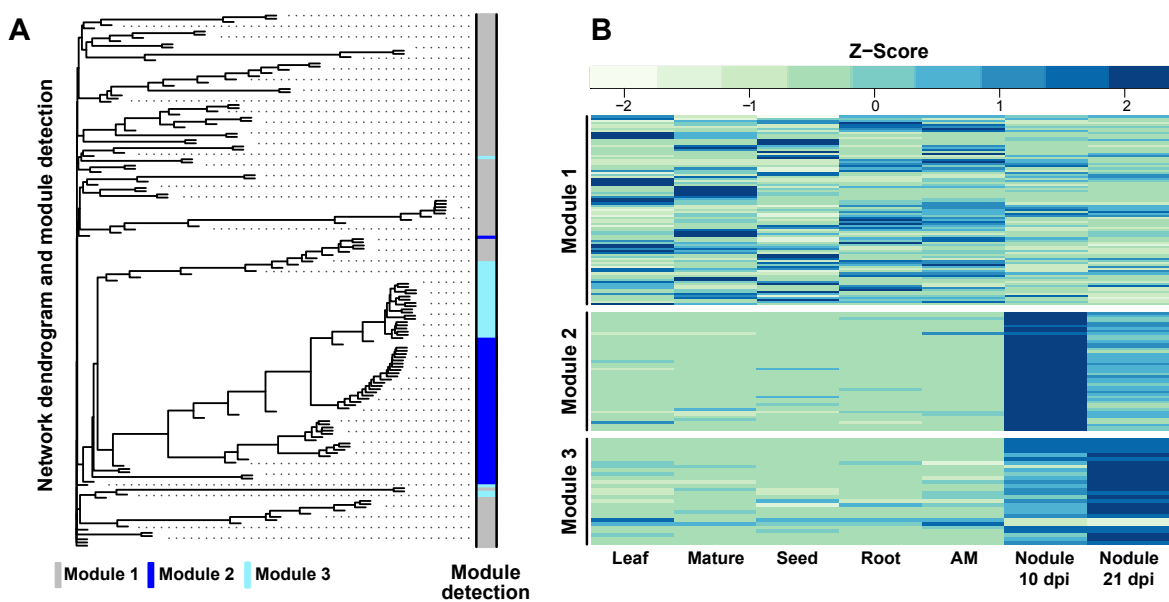

Figure 10. Gene co-expression analysis and transcriptomic variation per module. A) Cluster dendrogram and module detection assignment for the shared 167 DEGs. Three modules were detected using the cutreeDynamic function from a Topological Overlap Matrix (TOM) within the WGCNA package (Langfelder and Horvath, 2008). B) Relative expression pattern of the genes located in each module across seven different tissues and treatments. High expression levels are depicted in darker blue, while light green is for low expression levels. The heatmap was created from the genes in each module and the expression values for all conditions were calculated with the function heatmap.2 from the package gplots (Warnes et al., 2016).

From the starting 167 DEGs, the co-expression analysis helped to reduce the number of candidate genes to 75. In module 1 genes related with ROS production and leghemoglobin synthesis were identified whereas in module 2 genes related with antioxidant activity were identified. Interestingly, in both modules there were genes that had putative functions related to cell wall modifications, such as localized lignin polymerization and suberin deposition. It was hypothesized that a subset of genes from module 2 and 3 take part in the formation of the nodule endodermis as different genes within these modules have putative

function in plant barrier formations. The nodule endodermis is composed of cells that have a suberized and lignified cell wall (Hartmann et al., 2002). The chemical composition of those molecules includes products derived from the phenylpropanoid pathway (PPP). These products are aliphatic long-chain fatty acids or alcohols for suberin and monolignols for lignin. In addition, glycerol is required for suberin but do not belong to the PPP pathway (Brown and Walsh, 1994; Hartmann et al., 2002; Pollard et al., 2008). Genes associated with suberin biosynthesis and lignin formation were characterized in the context of the RNS (section 6 and 7).

6. Putative suberin biosynthesis genes are expressed during RNS

In module 2, four genes had functions associated with the synthesis of suberin, a cell wall biopolymer that is deposited in root endodermal cells (Graça, 2015). These included a Cytochrome 86A1 encoding gene (*LjCYP86A1*, LotjaGi6g1v0111000), which is a *Lotus* homolog of *HORST* in *Arabidopsis*, and a gene encoding a fatty acyl-CoA reductase (*LjFAR3.2*, LotjaGi3g1v0478900). Both genes are involved in the synthesis of aliphatic long-chain alcohols or fatty acids, which are suberin precursors (Höfer et al., 2008; Domergue et al., 2010). In addition, a gene encoding a transporter of the ABC-G family (*LjABC-G1*, LotjaGi5g1v0696000) was identified, a family whose members have been associated with the transport of suberin monomers from the cytoplasm to the cell wall (Kreszies et al., 2018); and a gene encoding a NAC transcription factor (*LjNAC*, LotjaGi1g1v0398100) homologous to *A. thaliana AtNAC38*, which has a putative role in suberin biosynthesis (Lashbrooke et al., 2016). When including putative genes from modules 1 and 3, the list expanded to contain genes encoding a second fatty acyl-CoA reductase (*LjFAR3.1*, LotjaGi3g1v0175200), a fatty acid desaturase (*LjFAD*, LotjaGi1g1v0726200), a second ABC-G transporter (*LjABC-G2,* LotjaGi5g1v0359700), a GDSL esterase/lipase (*LjGDSL*, LotjaGi1g1v0221300), and a LOB domain (LotjaGi1g1v0550200) transcription factor. Genes belonging to the families of this expanded list have been identified in a suberin study in poplar and also some of them have been included in comprehensive reviews about suberin (Rains et al., 2018; Serra and Geldner, 2022). Suberin in plant cells has a heterogeneous composition and is associated with lignin deposition (Zeier et al., 1999). Interlinked macromolecules of suberin and lignin act as cellular fences with diverse roles that include abiotic and biotic protection, solute diffusion limitation, and structural support (Graça, 2015; Meents et al., 2018; Zhong et al., 2019). Thus, it was hypothesized that the products of these genes are involved in the suberization of the nodule endodermis. It has been proposed that this limits the diffusion of oxygen into the nodule protecting the bacterial nitrogenase (Becana Ausejo et al., 1995).

To investigate the regulation of genes associated with the biosynthesis of suberin, e.g. genes encoding enzymes from the phenylpropanoid pathway (PPP) and fatty acid synthesis, *Lotus japonicus* orthologs from genes with reported functions in these processes were extracted from the Lotus base from five different tissues: root, root hair, nodule primordium, and nodule at 10 dpi and 21 dpi. In addition, genes involved in lignin biosynthesis were included, as it shares precursors from the PPP pathway and together with suberin confer plant barrier stability (Vogt, 2010). A list of consensus genes was obtained from reports in Arabidopsis, poplar (*Populus*), and Cork oak (*Quercus suber*) (Pollard et al., 2008; Graça, 2015; Zhong and Ye, 2015; Rains et al., 2018; Zhong et al., 2019) (Figure 11A). Genes associated with suberin biosynthesis were in general upregulated in nodules compared to roots and root hairs, while lignin and PPP genes were preferentially expressed in roots (Figure 11B).
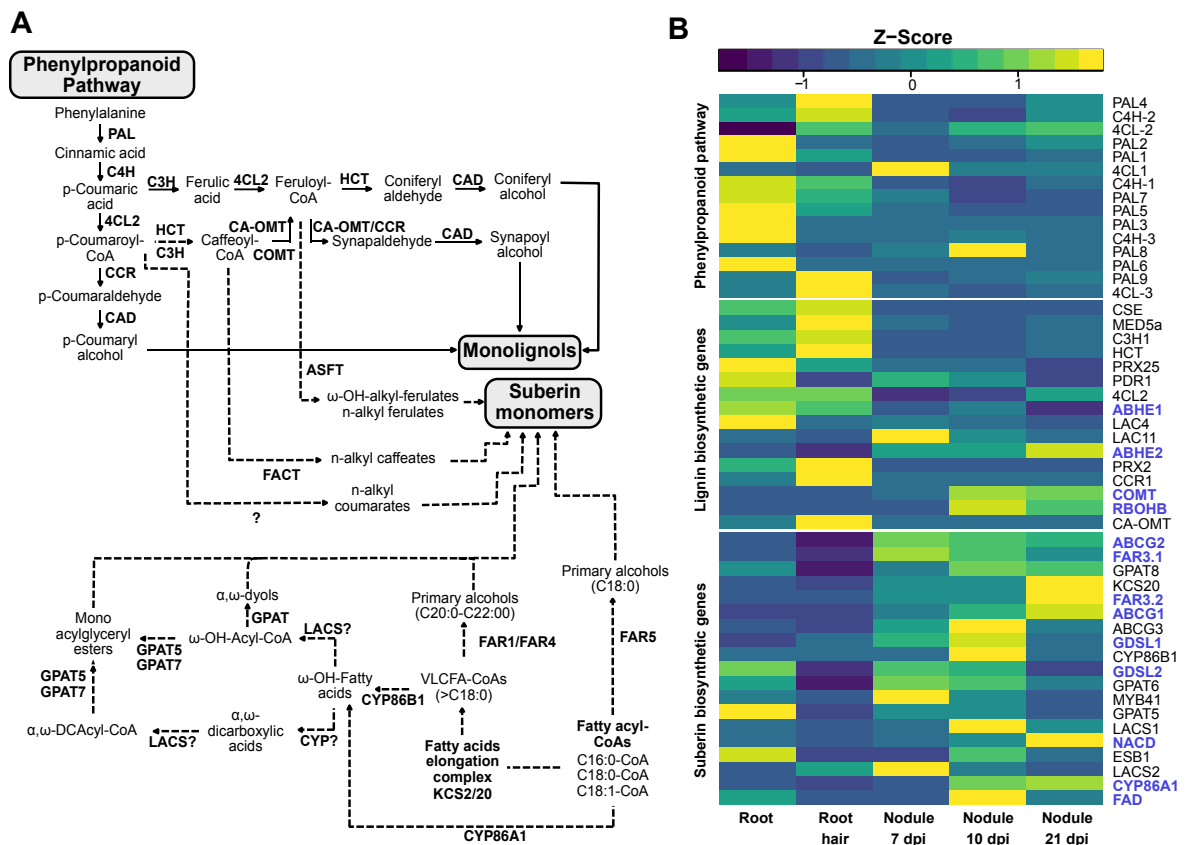

Figure 11. Expression pattern of the phenylpropanoid, lignin and suberin biosynthetic genes. A) Schematic representation of Lignin and Suberin biosynthetic pathways. B) Expression values from different *L. japonicus* genes with predicted functions in phenylpropanoid, lignin, and suberin biosynthesis in different tissues. The heatmap includes detected genes from the DE analysis (blue color) and a consensus of genes with putative functions in secondary cell wall formation in Arabidopsis and poplar. The hierarchical clustering dendrograms were built with the stat package within R (RDevelopment, 2012). ABC transporter G (ABCG), Alpha/beta hydrolase (ABHE), Caffeoyl CoA *O*-methyltransferase (CA-OMT), Caffeic acid *O*-methyltransferase (COMT), Cinnamoyl CoA reductase (CCR), Cinnamoyl alcohol dehydrogenase (CAD), Cinnamate 4-

hydroxylase and paralogs (C4H), 4-coumarate-CoA ligase 1 and paralogs (4CL), p-coumaroyl shikimate 30-hydroxylase (C3H1), Cytochrome 86A1 (CYP86A1), Cytochrome 86B1 (CYP86B1), Dirigent protein (ESB1), Fatty acyl-CoA reductase (FAR), Fatty alcohol:Caffeoyl-CoA Caffeoyl Transferase (FACT), Fatty acid desaturase (FAD), Glycerol-3-phosphate acyltransferase (GPAT), Hydroxycinnamoyl CoA transferase (HCT), ß-ketoacyl-CoA synthase (KCS), Laccase and paralogs (LAC), long-chain acyl-CoA synthetase (LACS), Peroxidase (PRX), MYB transcription factor (MYB41), Mediator of RNA polymerase II transcription (MED5a), NAC transcription factor (NAC), Phenylalanine ammonia lyase and paralogs (PAL), Peroxidase (PRX), Phenylalanine ammonia lyase (PAL), Respiratory burst oxidase homolog (RBOHB).

In addition, a similar analysis was conducted to observe if the same genes were expressed in the transcriptomic data from infected and non-infected nodules (Figure 12).

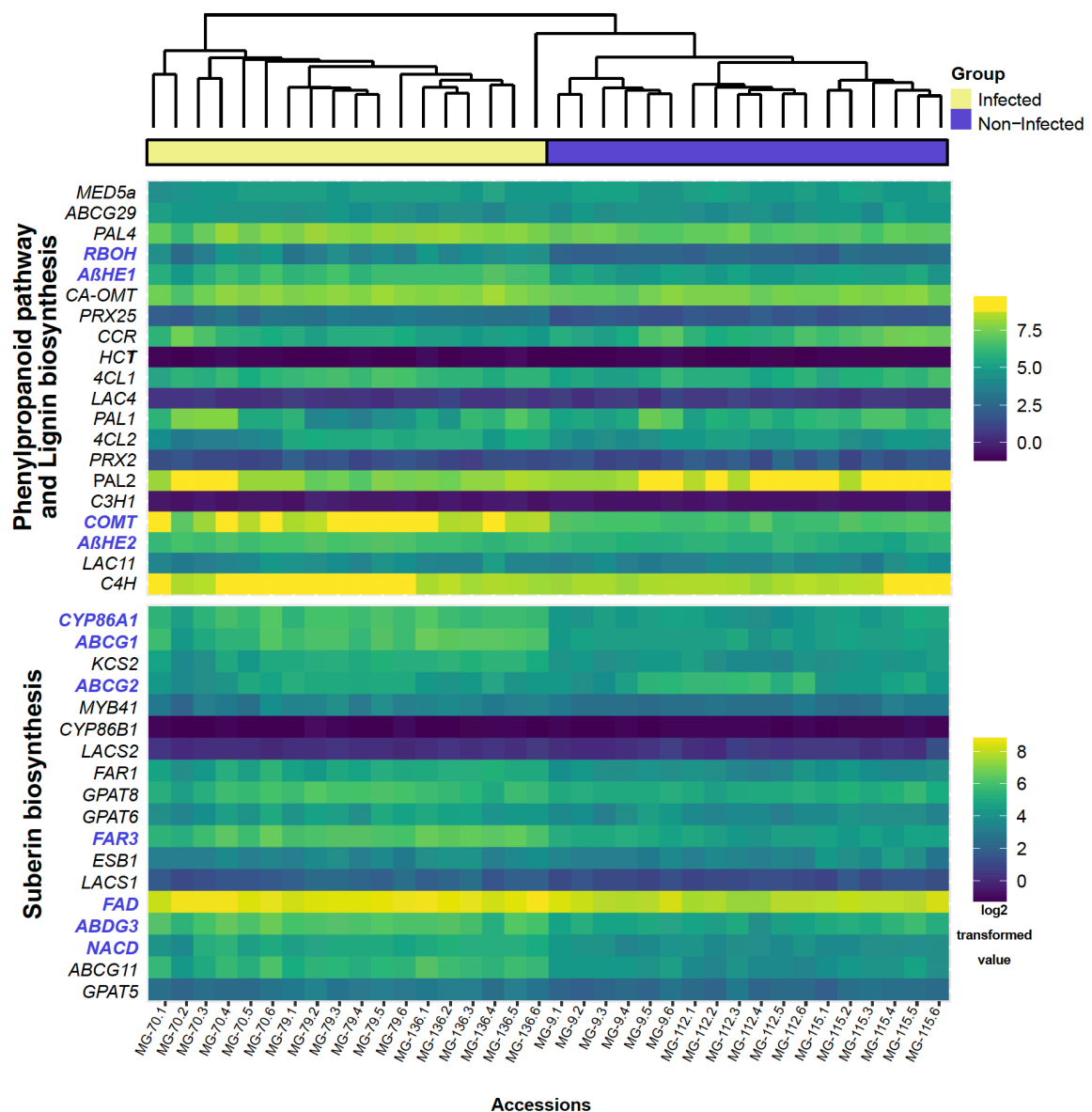**Figure 12. Transcriptomic variation in genes associated with secondary cell wall modifications in infected and non-infected nodules.** Heatmap of genes involved in the phenylpropanoid pathway, and in lignin and suberin biosynthesis in infected and non-infected *Lotus japonicus* accessions. The heatmap included genes detected in the DE analysis (blue color) and a consensus of genes with putative functions in secondary cell wall

54

formation described in Arabidopsis and poplar. The hierarchical clustering dendrogram was built with the stat package within R (RDevelopment, 2012).

To independently validate these results, RT-qPCR was conducted on *L. japonicus* Gifu roots inoculated with *M. loti* MAFF303099. This system is more robust compared to the sub-compatible interaction with *R. leguminosarum* Norway, as *M. loti* is the natural symbiont of *Lotus spp.* (Rodpothong et al., 2009). The analysis included four genes with a putative function in suberin biosynthesis (*LjCYP86A1, LjFAR3.1*, *LjFAR3.2,* and *LjNAC*) and two additional genes with putative cell wall associated functions (*LjRBOHB* and *LjCOMT*). All genes were specifically upregulated in rhizobia inoculated compared to mock-treated roots (Figure 13). The putative suberin-related genes and *LjCOMT* reached peak expression at 14 dpi, whereas *LjRBOHB* was expressed higher at 7 dpi, which coincides with the presence of young nodules. This independent validation indicated that genes with putative functions in suberin synthesis and regulation are induced during RNS.
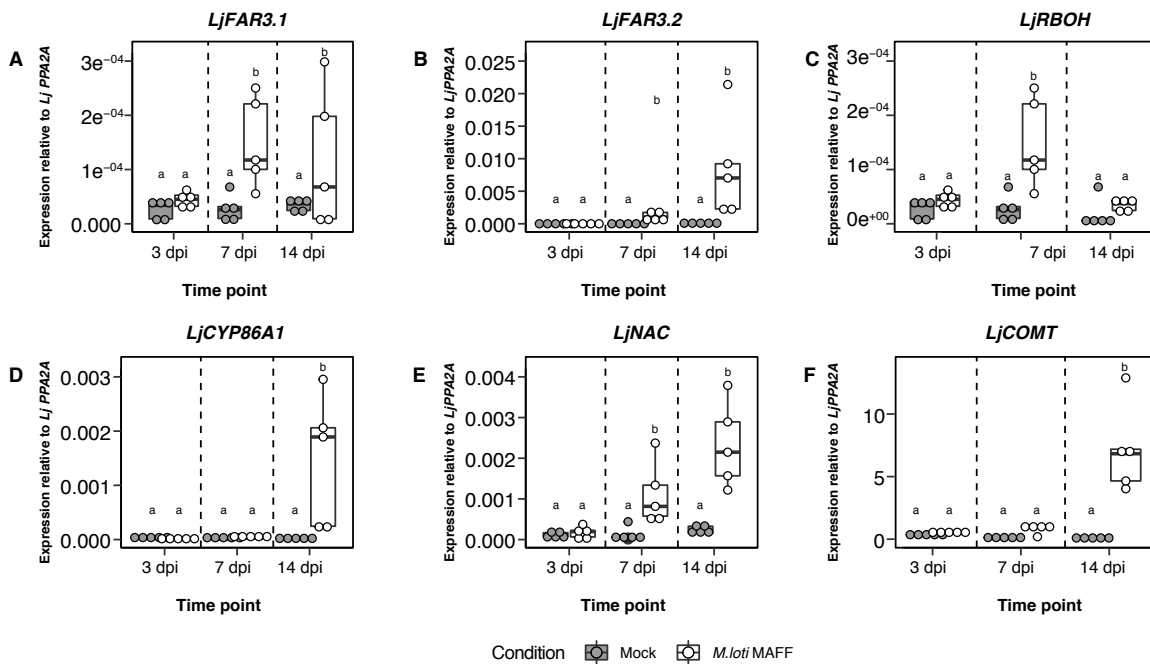

Figure 13. Gene expression analysis of putative secondary cell wall genes in *Lotus japonicus* roots upon inoculation with rhizobia. Quantification of transcript abundance by RT-qPCR of *LjFAR3.1* (A), *LjFAR3.2* (B), *LjRBOHB* (C), *LjCYP86A1* (D), *LjNAC* (E), and *LjCOMT* (F). Total RNA was extracted from *L. japonicus* Gifu whole roots three, seven, and fourteen days after inoculation with *M. loti* MAFF 303099. Relative expression levels were normalized against the *LjPPA2A* housekeeping gene. Each dot represents one independent biological replicate. The bold black line and the box represent the median and the interquartile range, respectively. The statistical analysis was performed using R. Lowercase letters indicate significance groups within each time point.

6.1 *FAR3* genes belong to a multigene family

The distinct expression pattern of *LjFAR3.1* and *LjFAR3.2* suggested neofunctionalization. To investigate possible functional diversification, a phylogenetic analysis was conducted. A previously reported phylogeny of FAR proteins (Rowland and Domergue, 2012) was expanded by including homologs of legume species. A maximum-likelihood tree was generated with homologs of *Arabidopsis thaliana*, *Solanum lycopersicum*, *Parasponia andersonii*, *M. truncatula*, *Cicer arietinum*, *L. japonicus* Gifu, *Oryza sativa*, and *Zea mays*, and outgroups reported on that phylogeny (Supplemental Table 5). The well-supported subfamily 3 branch included the two FARs encoded by genes identified in the DE analysis (*Lj*FAR3.1 and *Lj*FAR3.2). Legumes possessed a larger number of FARs inside this subfamily compared to non-legumes (Figure 14).
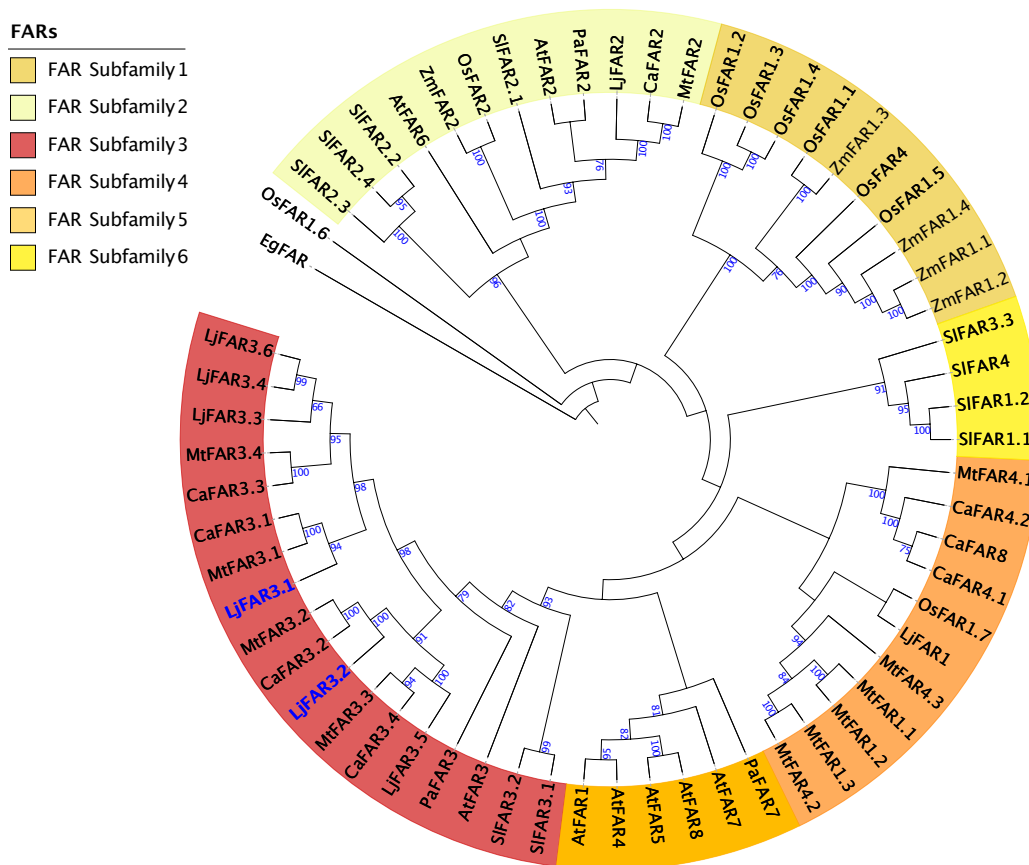

Figure 14. Maximum-likelihood phylogenetic tree of *Lj*FARs protein family. The dendrogram was created using sequences from an already reported phylogeny from Rowland et al. 2012 with the addition of sequences belonging to legume species. The FARs detected with the DEG analysis are highlighted in blue. Abbreviations are as follows: Fatty Acyl-CoA Reductase (FAR), *Arabidopsis thaliana* (*At*), *Arachis hypogaea* (*Ah*), *Cicer arietinum* (*Ca*), *Lotus japonicus* (*Lj*), *Medicago truncatula* (*Mt*), *Oryza sativa* (*Os*), *Parasponia andersonii* (*Pa*), and *Zea mays* (*Zm*). Protein alignment was done in MAFFT (Rozewicki et al., 2019) and maximum-likelihood tree was built using the RAxML-HPC BlackBox tool (version 8.2.12) in CIPRES (www.phylo.org), which uses a rapid bootstrap algorithm (Stamatakis, 2014). Trees were displayed with Interactive Tree Of Life v5 (https://itol.embl.de/). Bootstrap values over 60 are indicated in the nodes.

In *Lotus*, from the five paralogs in this subfamily, four of them (*LjFAR3.1*, *LjFAR3.2*, *LjFAR3.4*, and *LjFAR3.6*) are induced during nodulation (Figure 15). Interestingly, *FAR3.1* is also expressed in roots. In Arabidopsis, the closest homologs with a known function are *AtFAR1*, *AtFAR4*, and *AtFAR5* (Figure 14), which are involved in suberin synthesis in the root endodermis (Domergue et al., 2010), while *AtFAR3* is associated with cuticular wax synthesis (Rowland et al., 2006). These results suggest that *LjFAR3.1* might have retained a function in root endodermis suberization, while *LjFAR3.2*, *LjFAR3.4*, and *LjFAR3.6* might have evolved to fulfill nodule-specific functions.
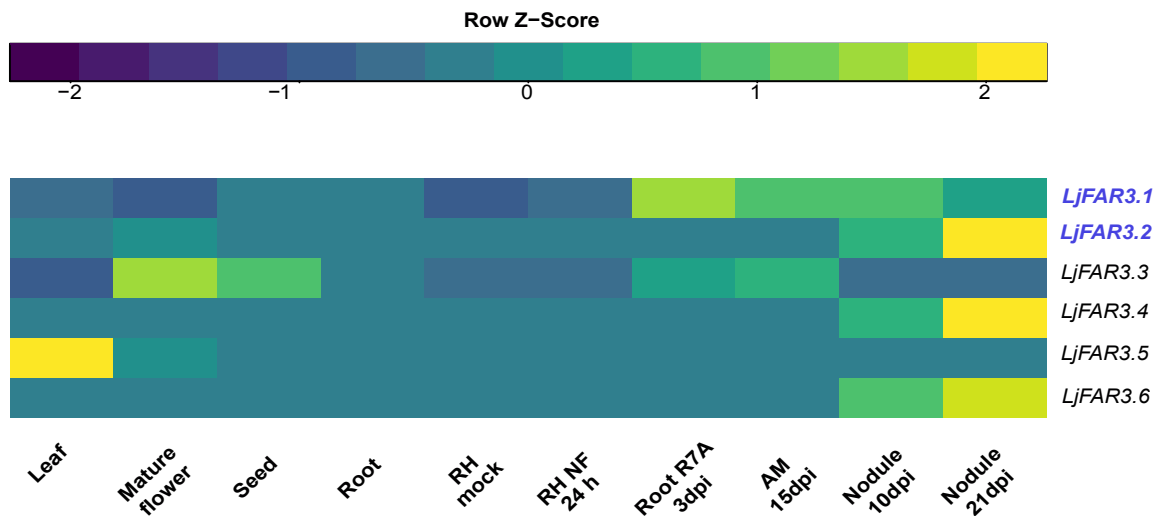
Figure 15. Heatmap of *Lotus japonicus* paralogs from subfamily 3. Heatmap illustrating the expression levels of the six *Lotus japonicus* paralogs present in subfamily 3 of the FAR phylogeny. Expression data were extracted for different tissues: leaf, mature flower, seed, root, root hair (RH), AM at 15 dpi, and nodules at 10 and 21 dpi.

Determinate nodules, in addition to the nodule vascular bundles, possess a suberized cell layer that divide the outer and inner cortex (Hartmann et al., 2002; Guinel, 2009). This cell layer is known as the nodule endodermis (Figure 16) and has been hypothesized to aid in the formation of the oxygen diffusion barrier (Hartmann et al., 2002), which is essential for efficient nitrogen fixation (Becana Ausejo et al., 1995).

Figure 16. Nodule suberized tissues. Modified secondary cell walls in *Lotus japonicus* nodule stained with Nile red. White arrows indicate the nodule endodermis En(n), the root endodermis En(r), and the nodule vascular endodermis En(nv). Infected cells, ic. Scale bar = 100 μm.
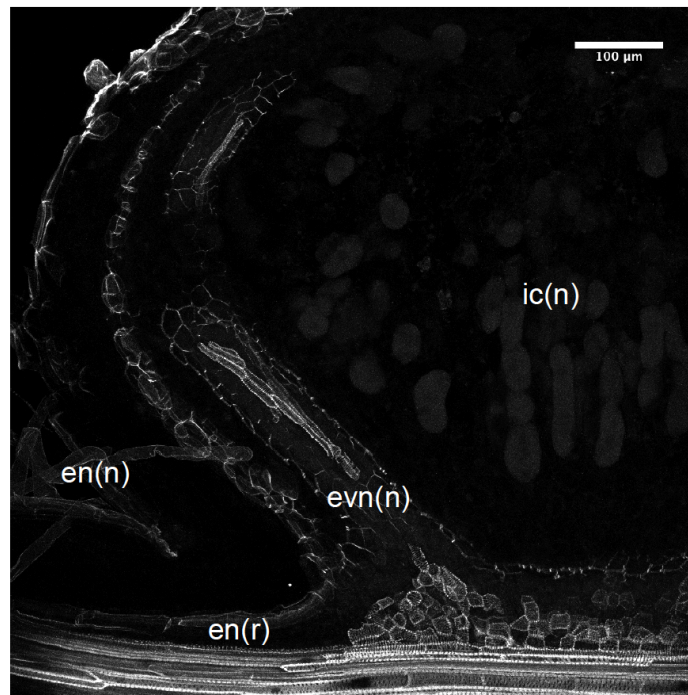
6.2 Promoters of suberin biosynthesis genes are active in the nodule endodermis

To determine the temporal and tissue-specific activation pattern from promoters driving the expression of genes, which were hypothesized to be involved in the deposition of suberin in the nodule endodermis, 3kb upstream region from the start codon were cloned and fused either to the *β-glucuronidase* (*GUS*) gene or a two-times nuclear localized YFP (*NLS-2xYFP)*. The promoters of *LjFAR3.1*, *LjFAR3.2*, and *LjNAC* were selected based on the function of the controlled genes (biosynthesis, transcriptional regulation). Constructs were introduced by hairy root transformation into *Lj* Gifu and roots were then inoculated with *M. loti* MAFF303099-GFP. All promoters were active in nodule primordia and mature nodules at 7 and 21 dpi, respectively (Figure 17). Strong and more localized activation appeared for the *FAR3.2pro:DoGUS* in mature nodules (Figure 17F). Both *FAR3.1pro:DoGUS* and *NACpro:DoGUS* had an nonspecific pattern including roots, nodule primordia, and nodules (Figure 17A and C).

To investigate if the activity of the promoters coincided with the nodule endodermis, autofluorescence of suberin and lignin was detected by UV-excitation (405 nm) and confocal microscopy. These polymers exhibit autofluorescence due to their chemical composition (Donaldson, 2020). The activities of the *FAR3.1pro:NLS-2xYFP* and *FAR3.2pro:NLS-2xYFP* reporters were located to a single cell layer in the outer nodule tissue

that coincided with secondary cell wall autofluorescence (Figure 17H and I). In addition *FAR3.1pro:NLS-2xYFP* was also active in the vasculature and in the root endodermis. The *NACpro:NLS-2xYFP* reporter was also active in the vasculature, but to a lesser degree (Figure 17J). The specific activity of these promoters in the nodule endodermis suggests that these genes are associated with suberization of this cell layer.
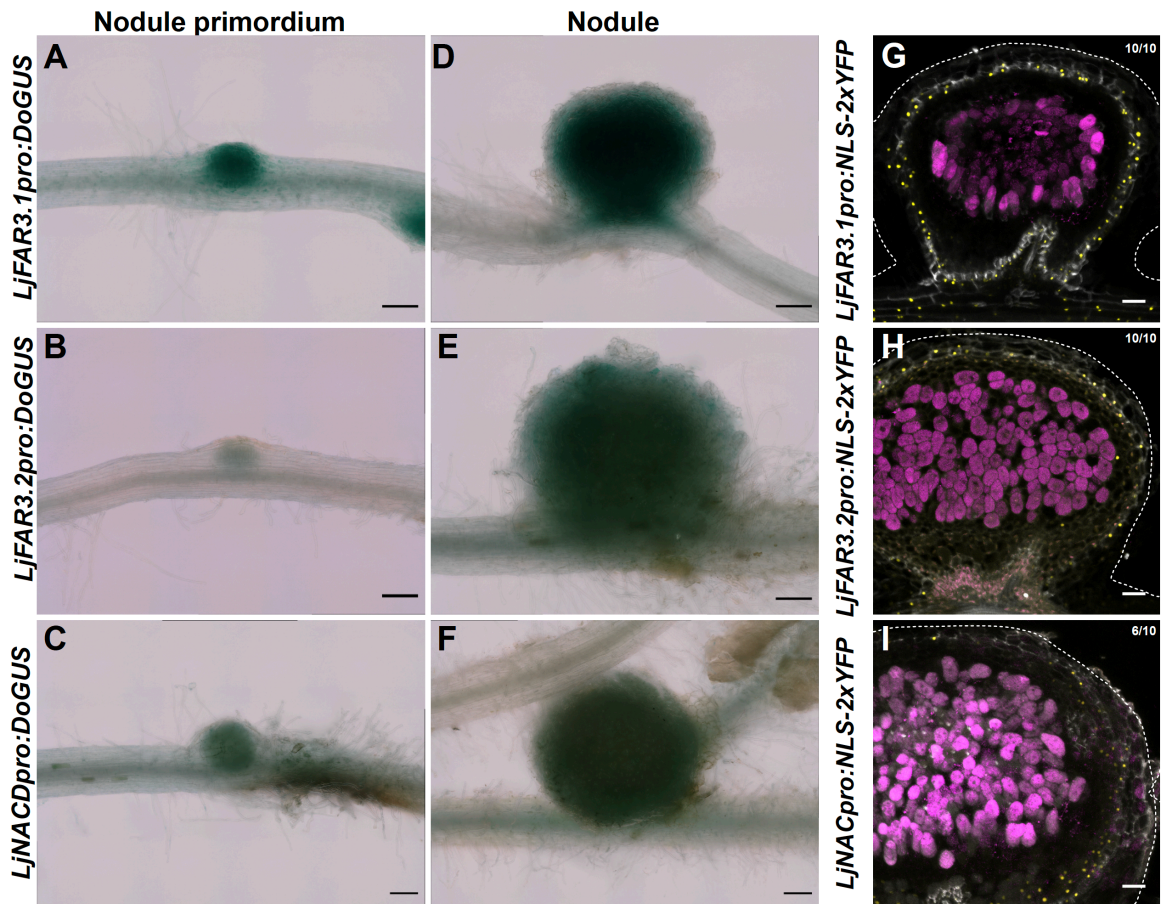
Figure 17. Promoter activity driving the expression of suberin-related genes at different stages of root nodule development. Representative images of DoGUS histochemical staining of roots and nodules indicating the activity of the *LjFAR3.1pro, LjFAR3.2pro,* and *LjNACpro* promoters. The panels (A-C) and (D-F) display the staining in nodule primordia and mature nodules, respectively. Scale bar = 100 μm. (G-I) Representative images of 65 μm-thick nodule sections panels displaying fluorescent reporter (NLS-2xYFP, yellow), infected cells (*M. loti* MAFF 303099-GFP, magenta) and the auto-fluorescent lignin and suberin (grey). The border of the nodule is marked with dashed lines. Scale bar = 50 μm.

6.3 Mutations in *LjFAR3.2* impair endodermis suberization and nitrogen fixation

The specific pattern of the *LjFAR3.2* promoters suggested that *LjFAR3.2* function in the nodule endodermis. To determine if disruption of *LjFAR3.2* alters nodule endodermis permeability and nodule function, two independent *LOTUS RETROTRANSPOSON 1*

59

(*LORE1*) mutant lines (Małolepszy et al., 2016b) in this gene were characterized (L30127714 and L30165196). The two *LORE1* insertions were located in the last exon of the *LjFAR3.2* gene (Figure 18A). Both lines carried additional insertions in exonic and intronic regions of other genes. Therefore, homozygous plants for the wild-type *FAR3.2* allele from the respective segregating populations carrying the background mutations were used for comparison (referred to here as WT*). Suberin lamellae staining of nodules was performed using Fluorol yellow (Naseer et al., 2012; Andersen et al., 2021). WT* nodules showed significantly higher Fluorol yellow signal in the endodermis than *far3.2-2* and *far3.2-1* nodules (Figure 18B and C).
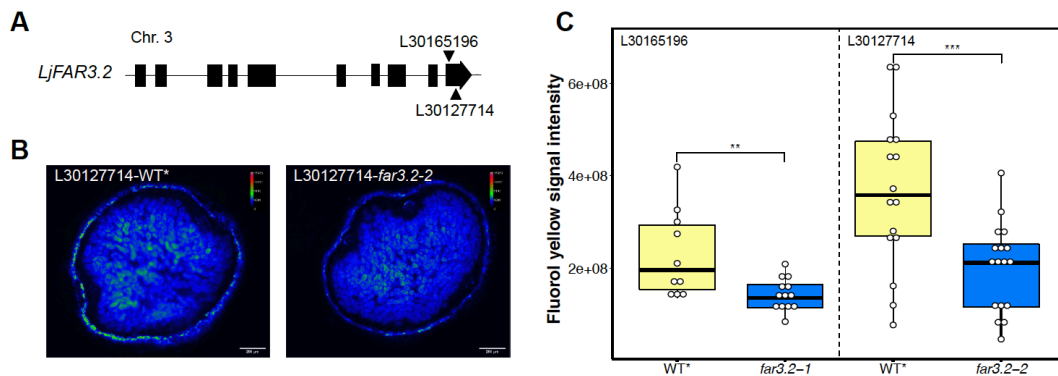
Figure 18. Fluorol yellow staining in *far3.2* mutant lines. A) Gene structure for the *LjFAR3.2* gene displaying the intron/exon structure and the *LORE1* retrotransposon insertion sites. B) Representative images of nodules stained with Fluorol yellow and signal quantification for the *far3.2-2* and its WT* plants. Scale bar = 200 μm. The significant levels of the p-values were ** ≤ 0.01 and *** ≤ 0.001 based on the Welch's t-test.

Toluidine blue staining has been used to evaluate root permeability in *A. thaliana* (Andersen et al., 2021). Mutant and WT* nodules of both lines were immersed in Toluidine blue to evaluate if interruption of this gene increased the permeability of the nodule endodermis to the dye. As nodules vary in size, a permeability ratio (PR) was calculated. This ratio was defined as PR = td/nd where nodule distance (nd) measures the length from the nodule border to the closer infected cell border and the Toluidine blue distance (td) measures the distance that the dye penetrates in the nodule (Figure 19A). The dye permeated more in the mutant nodules compared to the WT*, as indicated by a significantly larger PR in both *far3.2* lines (Figure 19B and C).

Figure 19. Toluidine permeability in *far3.2* mutant lines. A) Graphical depiction of the permeability ratio (PR) where Toluidine distance (td) and nodule distance (nd) were measur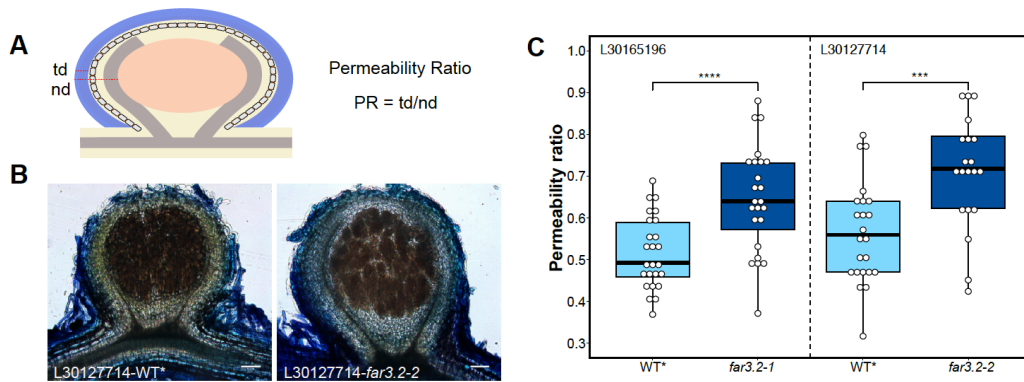ed in µm. B) Representative images of 100 µm-thick nodule sections of WT* and *far3.2-2* mutant plants. Scale bar = 100 µm. C) Box plot displaying the PR quantification for the *far3.2-1* and *far3.2-2* mutants and the respective WT* plants. The significant levels of the p-values were *** ≤ 0.001 and *** ≤ 0.0001 based on the Welch's t-test.

To evaluate the permeability of the nodule endodermis to oxygen, a microelectrode was used to measure the oxygen concentration in the nodule surface and in the inner cortex. We inserted a needle-type microelectrode perpendicularly to the top of the nodule (Figure 20A). Both *far3.2* mutant nodules had significantly higher oxygen concentrations compared with the respective wild-type nodules (Figure 20B and C). This supports the concept that deposition of lipid long aliphatic fatty acids in the nodule endodermis restricts oxygen diffusion.
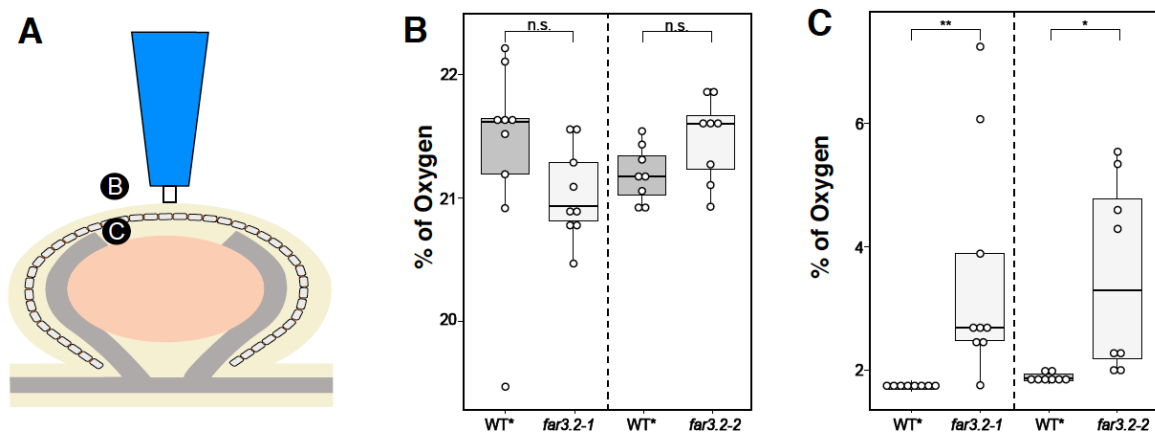
Figure 20. Oxygen measurement in the nodule. A) Schematic representation of oxygen measurement setup. B) Measurement on the nodule surface (10 µm). C) Measurement in the nodule inner cortex (75 µm). The p-values were * ≤ 0.05 and ** ≤ 0.01 based on the Wilcoxon test.

Oxygen is a potent inhibitor of the bacterial nitrogenase (Gallon, 1981). Thus, nitrogen fixation activity was quantified using an acetylene reduction assay (ARA). Mutant lines showed a decline in the ARA activity compared with their respective WT* (Figure 21A). To

assess the effect of the mutations on nodule functionality, plant growth was quantified (Figure 21B). Shoot length and root length were significantly different for both lines (Figure 21C and D). To determine if *LjFAR3.2* affects organogenesis or infection, the number of nodules and ITs were quantified. Only *far3.2-2* had significantly less nodules, whereas *far3.2-1* had significantly less infection threads (Figure 21E and 1F).
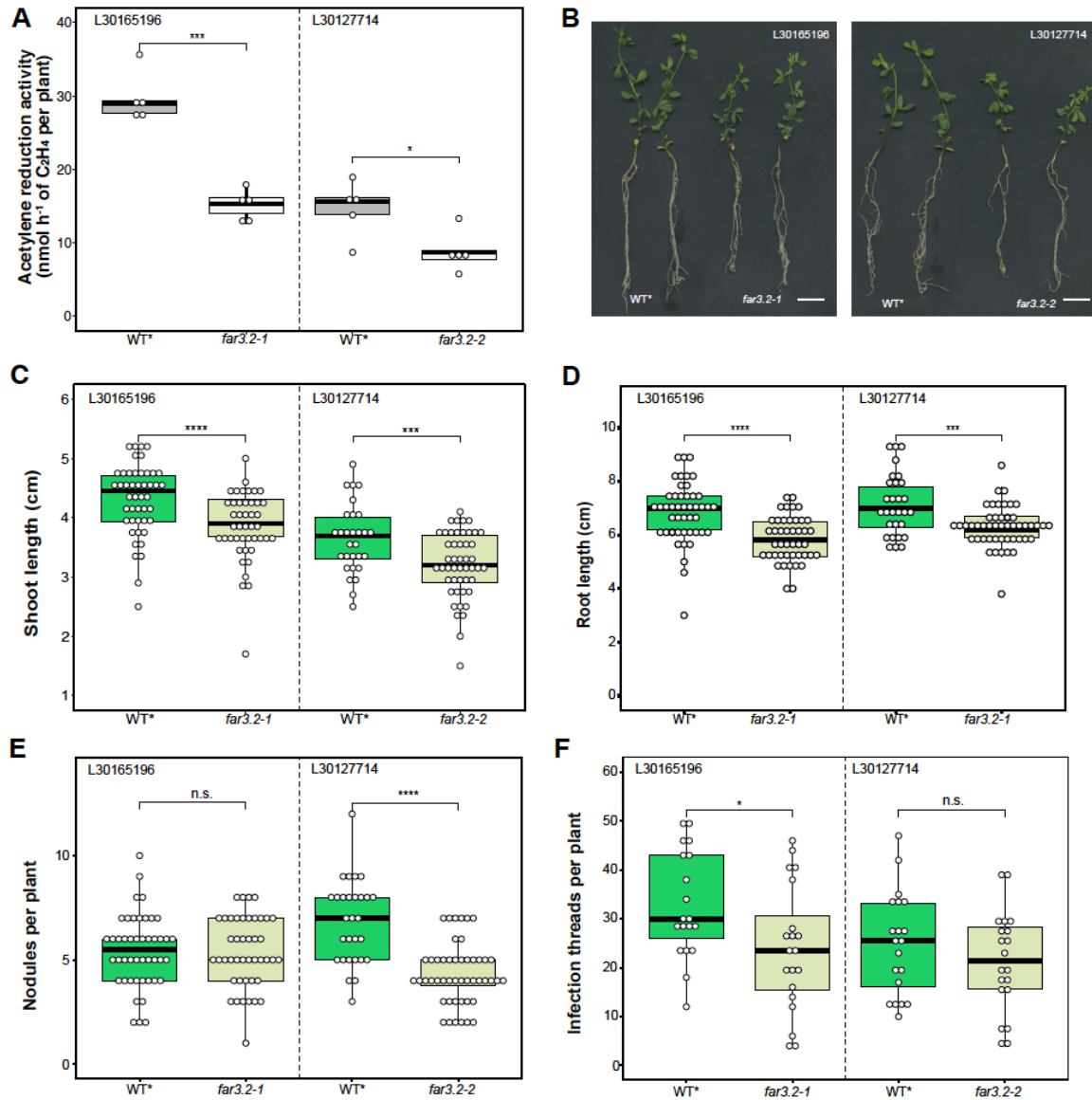
Figure 21. Acetylene reduction assay and plant growth characterization of *far3.2* mutant lines. A) Box plot displaying the acetylene reduction results where each point is the mean of 5 biological replicates in the *far3.2-1* and *far3.2-2* mutant lines. B) Box plot displaying the shoot phenotype for the *far3.2-1* and *far3.2-2* mutants and the respective WT* plants. C) Representative image for the shoot length in the *far3.2-1* and *far3.2-2* mutant lines. Scale bar = 1 cm. Box plot displaying the D) number of nodules/plant, E) the root length, and F) number of infection threads/plant for the *far3.2-1* and *far3.2-2* mutants and the respective WT* plants. The significant levels of the p-values were n.s > 0.05, * ≤ 0.05, *** ≤ 0.001 and *** ≤ 0.0001 based on the Welch's t-test.

7. A subset of the *CASPL* genes is induced during RNS in a tissue-specific manner

The root endodermis acts as a barrier that selectively allows gas and nutrient permeation. As stated before this endodermis is composed of cells that have a suberized and lignified cell wall (Hartmann et al., 2002). Lignin is required for the formation of the Casparian Strip, a ring-like structure in the root endodermis (Geldner, 2013). A family of proteins called Casparian strip membrane domain proteins (CASPs) is required for the formation of the Casparian strip (Roppolo et al., 2011). *CASP* and *CASPL* (CASP-like) genes are part of a multigene family. CASPs recruit cell wall enzymes for lignin formation, e.g. peroxidases, RBOHs, and laccases. CASPLs have been hypothesized to perform similar functions but at localized levels and in different tissues (Roppolo et al., 2014). In module 2, two *CASPL* genes and one *RBOHB* gene were identified and a third CASPL was also identified in module 1. All these genes were upregulated in infected nodules compared with non-infected ones.

A phylogenetic analysis was conducted to investigate possible genetic redundancy and relationships between the CASPLs from *L. japonicus* and from other plant species, particularly legumes. This analysis was done in a similar manner as the FAR phylogeny by including proteins from legume species described in a previously reported phylogeny. The maximum-likelihood tree of CASPs and CASPLs grouped proteins into six subfamilies (Figure 22). The proteins encoded by the three *CASPL* genes identified in the DE analysis were placed in subfamily 4. *L. japonicus* had the highest number of homologs (11 proteins) among legume species within subfamily 4. However, these genes are not unique to legumes, as orthologs from the non-legume species *A. thaliana*, *O. sativa*, *Z. mays*, and *S. lycopersicum* were found in this subfamily.
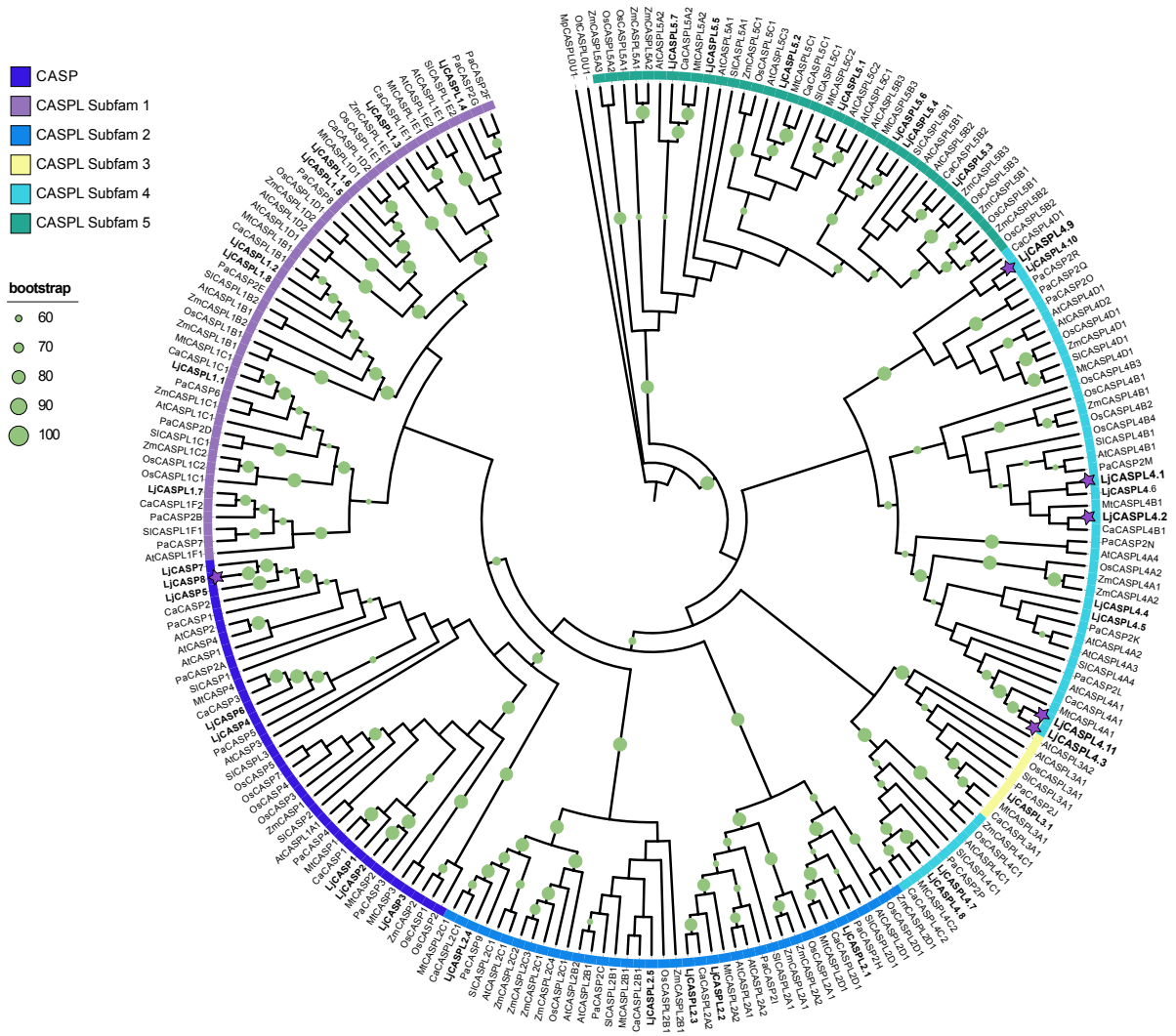
Figure 22. Maximum-likelihood phylogenetic tree of *Lj*CASP protein family. The dendrogram was created using sequences from a phylogeny already reported by Roppolo et al. 2014 with the addition of sequences belonging to legume species. *Lj*CASPLs are highlighted in bold. Abbreviations are as follows: Casparian strip membrane domain protein (CASP) and Casparian strip membrane domain protein-like (CASPL), *Arabidopsis thaliana* (*At*), *Arachis hypogaea* (*Ah*), *Cicer arietinum* (*Ca*), *Lotus japonicus* (*Lj*), *Medicago truncatula* (*Mt*), *Oryza sativa* (*Os*), *Parasponia andersonii* (*Pa*), and *Zea mays* (*Zm*). Protein alignment was created in MAFFT (Rozewicki et al., 2019) and maximum-likelihood tree was built using the RAxML-HPC BlackBox tool (version 8.2.12), which uses a rapid bootstrap algorithm (Stamatakis, 2014) in CIPRES (www.phylo.org). Trees were displayed with Interactive Tree Of Life v5 (https://itol.embl.de/). Bootstrap values over 60 are indicated in the nodes as green dots. Purple stars indicate *Lotus* CASPs and CASPLs relevant for this study.

L. japonicus has the highest number of homologs in subfamily 4 compared with other subfamilies. Proteins subfamilies have similar sequences and related functions (Nei and Rooney, 2005). Thus, it was hypothesized that *L. japonicus* subfamily 4 members could have related functions, specifically with RNS as *LjCASPL4.1*, *4.2*, and *4.3* were detected in the DE analysis. To investigate the regulation of these genes during RNS, *L. japonicus CASP* and *CASPL* expression values across different treatments were extracted from the lotus base (https://lotus.au.dk/). The selected tissues and conditions included: seed, flower,

leaf, root inoculated with *M. loti* R7A, arbuscular mycorrhiza symbiosis (AMS), root hairs treated and 24 h and 72 h post inoculation, and nodules at 7, 10, and 21 dpi (Figure 23). Members of subfamily 4 are induced under symbiotic conditions, e.g. *LjCASPL4.6*, *4.10*, and *4.11* are expressed in root hairs upon rhizobia inoculation whereas *LjCASPL4.1*, *4.2*, *4.3*, *4.5*, *4.7*, and *4.9* are induced at different stages of nodule development. In the case of *LjCASPL4.4*, expression only occurs during AMS. Other CASPLs outside the subfamily 4 are induced during symbiotic conditions. Therefore, several members of the *LjCASPs* and *LjCASPL*s respond to symbiosis either RNS or AMS with subfamily 4 being predominantly expressed during symbiosis (Figure 23).
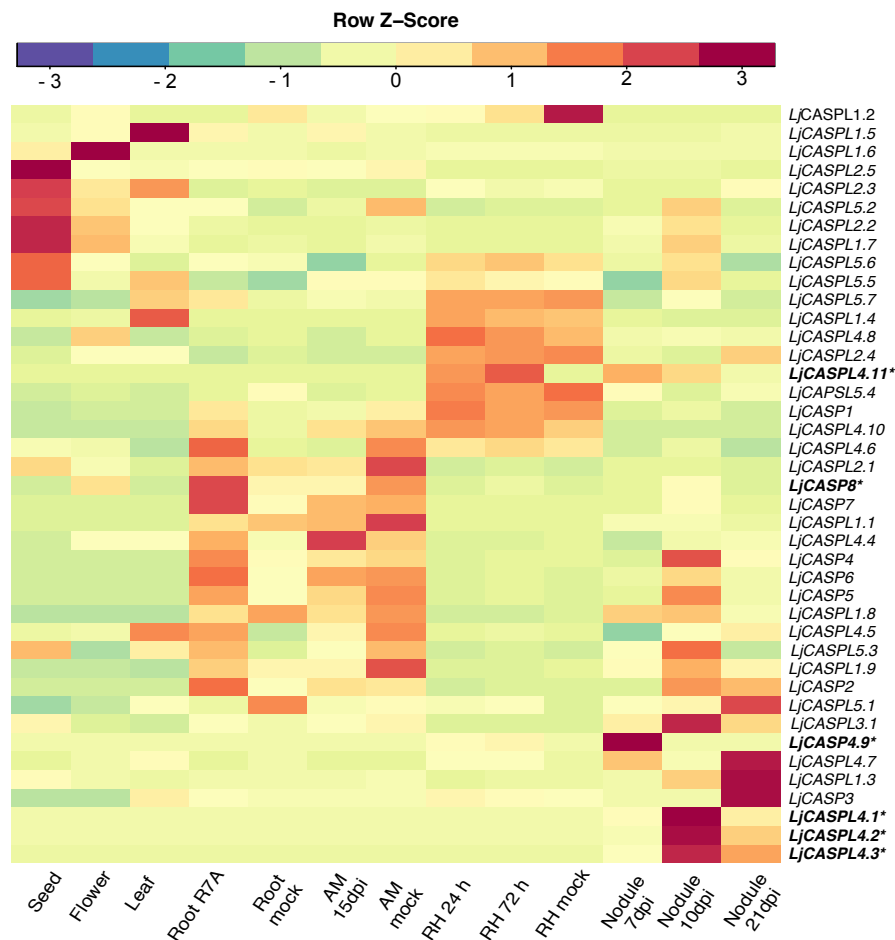
Figure 23. Heatmap illustrating the expression levels of *LjCASP*s and *LjCASPL*s. The expression levels of all *Lotus japonicus CASPs* and *CASPL* genes in different tissues and treatments from Lotus Base. *CASPs* and *CASPL*s relevant for this study.

To validate the expression data, a RT-qPCR was performed. Nodulated roots and mock roots as described in section 6 were used (Figure 13). For this analysis, genes representative of different RNS conditions were selected based on the heatmap (Figure 23). *LjCASPL4.1* and *LjCASPL4.2* were detected in the DE analysis and have the highest expression in nodules at 10 dpi. *LjCASPL4.3* was also included as it is a close paralog of

65

LjCASPL4.1 and *LjCASPL4.2*. *LjCASPL4*.9 is expressed primarily in primordia whereas *LjCASPL4.11* is induced in inoculated root hairs and nodules at 7 and 10 dpi. *LjCASP8* was included as it does not express in the nodule or the root hair but is induced in the root. All subfamily 4 *CASPL*s genes were induced upon inoculation with *M. loti* compared with mock-treated roots and partially agreed with the heatmap results (Figure 24). All *LjCASP*s and *LjCASPL*s were significantly induced at least at one time point. *LjCASPL4.1*, *4.3*, and *4.11* were expressed as displayed by the heatmap (Figure 24A, C, and E). The RT-qPCR results differ between *LjCASPL4.2* and *LjCASPL4.1* despite similar expression patterns in the heatmap (Figure 23A and B). *LjCASPL4.9* was induced at all time points as opposed to being only expressed in primordia according to the heatmap (Figure 24D). Finally, *LjCASPL8* should be expressed only in root tissue. However, at 3 dpi a significant difference can be observed (Figure 24F). Combined results from the heatmap and the RT-qPCR indicate that *LjCASP* and *LjCASPL* genes are expressed during RNS.
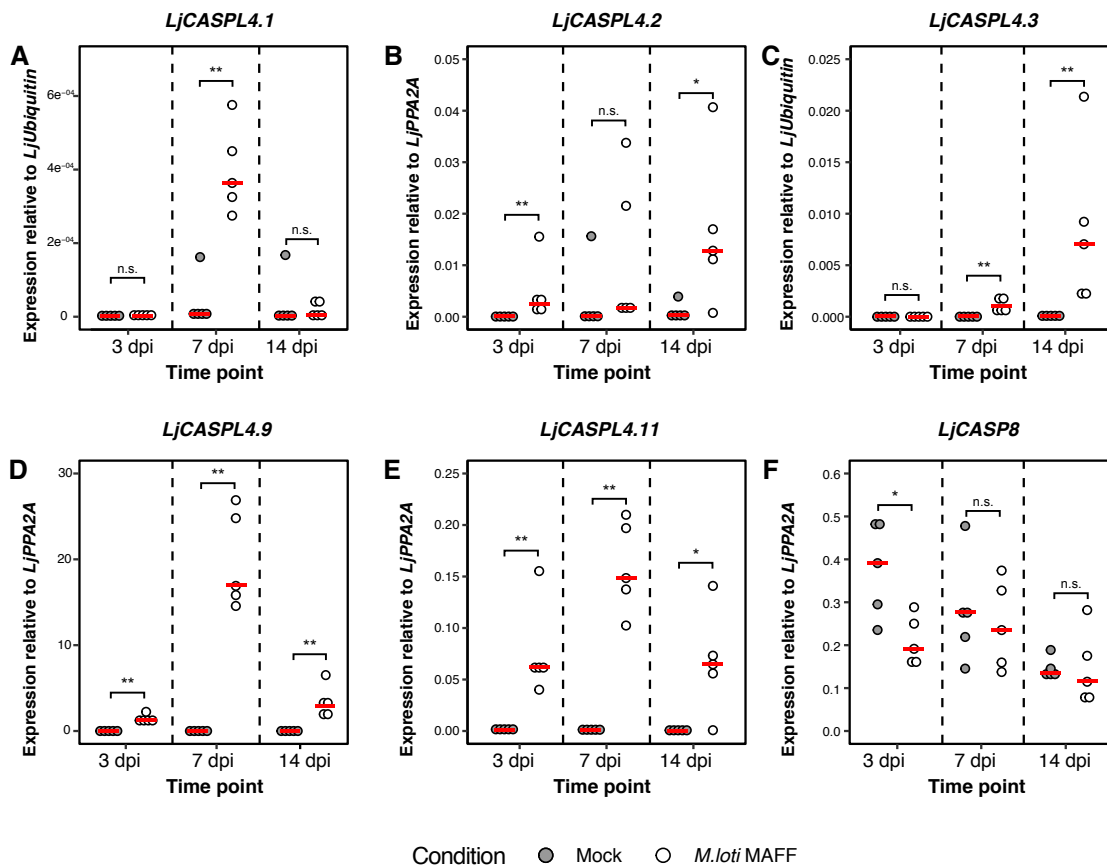
Figure 24. Gene expression analysis of a set of *LjCASPL* upon inoculation with rhizobia. Transcript abundance was quantified by RT-qPCR for *LjCASPL4.1* (A), *LjCASPL4.2* (B), *LjCASPL4.3* (C), *LjCASPL4.9* (D), *LjCASPL4.11* (E), and *LjCASPL8* (F). Total RNA was extracted from *Lotus japonicus* Gifu whole roots three, seven, and fourteen days after inoculation with *M. loti* MAFF 303099. Relative expression levels were normalized against the *LjPPA2A* or *LjUbiquitin* housekeeping genes. Each dot represents one independent biological replicate. The bold red line represents the median. The statistical analysis was the Wilcox test and it was performed using R. The p-values were * ≤ 0.05, ** ≤ 0.01 and n.s. for no significant difference.

7.1 Promoters of *LjCASPL*s are active at different stages during RNS

The specific induction of subfamily 4 *CASPL* genes in symbiotic tissues prompted us to explore the spatio-temporal activity of selected *LjCASPL* promoters. The promoters of *LjCASPL4.1*, *LjCASPL4.9* and *LjCASPL4.11* were selected as these genes had a distinctive pattern of expression based on the heatmap and RT-qPCR results (Figure 23 and 24).

When possible a 3kb upstream region from the start codon was cloned and fused to a two-times nuclear localized YFP (*NLS-2xYFP)* otherwise shorter fragments of at least 1kb in length were cloned. Constructs were introduced by hairy root transformation into *L. japonicus* Gifu and transformed plants were inoculated with *M. loti* MAFF303099-GFP. The *LjCASPL4.11* promoter was not active in any tissue or symbiotic stage at 10 or 21 dpi (Figure 25). The *LjCASPL4.9pro* was active in the root epidermis, IT containing cells, primordia, and young nodules (Figure 25). The *LjCASPL4.1pro* was active in infected cells in primordia and nodules at 10 dpi and 21 dpi (Figure 25). Interestingly, *LjCASPL4.1pro* and *LjCASPL4.9pro* were active in nodules at 10 dpi but *LjCASPL4.1pro* activity was restricted to infected cells whereas *LjCASPL4.9pro* is also active in the inner cortex. This indicates that certain *Lj*CASPLs subfamily 4 genes are active during different stages of the symbiosis.
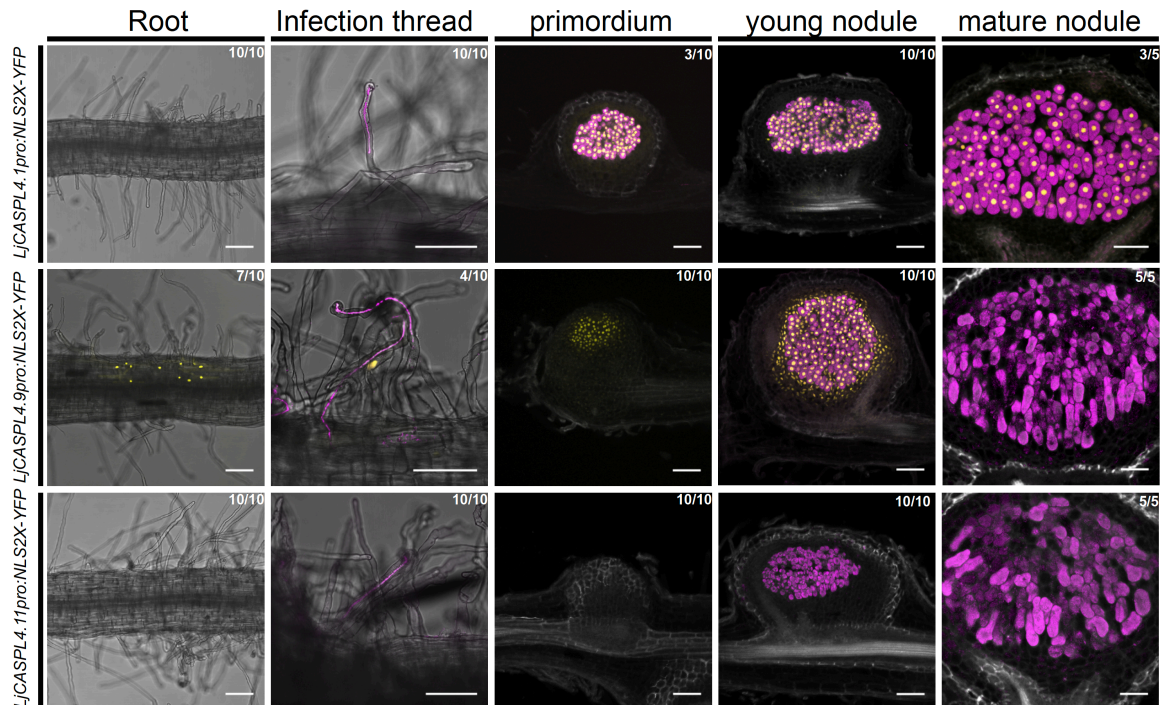
Figure 25. Activity of promoters driving the expression of different *CASPL* genes in *Lotus japonicus*. Five different stages are depicted: root, infection thread, primordium, young nodule (10 dpi) and mature nodule (21 dpi), for the *LjCASPL4.1_{pro}, LjCASPL4.9_{pro}* and *LjCASPL4.11_{pro}*. Representative images of 65 μm-thick nodule sections displaying the fluorescent reporter (NLS2x-YFP, yellow), infected cells (*M. loti* MAFF 303099-GFP, magenta) and the auto-fluorescent cell wall components for contrast (grey). Scale bar = 100 μm.

In *Arabidopsis*, *At*CASP1 recruits RBOHF for Casparian strip formation. This is possible as RBOHF has a specific N-terminal domain and can enter the Casparian strip domain (Roppolo et al., 2011; Lee et al., 2013; Fujita et al., 2020). In the same co-expression module as the *LjCASPL4.1* and *LjCASPL4.2*, a *LjRBOHB* was identified. Therefore, the *LjRBOHB* promoter was studied to see if there was an overlap in activity between the promoters of both genes. In a similar fashion, a 3 kb promoter was cloned and fused with the *β-glucoronidase* (*DoGUS*) and the *NLS-2xYFP* reporters. The constructs were introduced by hairy root transformation into *L. japonicus* Gifu and roots were then inoculated with *M. loti* MAFF303099-GFP. The signal for the *RBOHB$_{pro}$:DoGUS* was detected in both young and mature nodules and the *RBOHB$_{pro}$:NLS-2xYFP* reporter was detected exclusively in infected cells (Figure 26). This hints at the possibility that CASPL and RBOHB might interact as their promoter activity happens in the same tissue and they are co-expressed together.
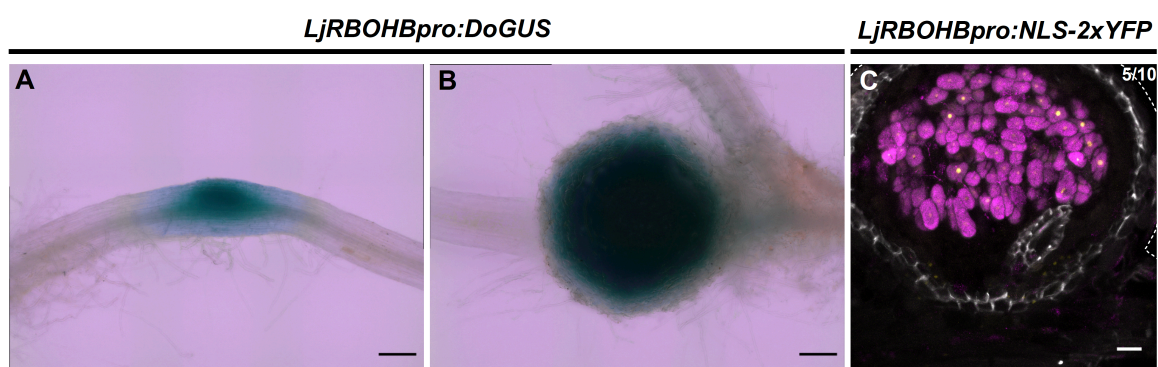
Figure 26. Activity of promoters driving the expression *LjRBOHB* in *Lotus japonicus*. A) Representative images of DoGUS histochemical staining of nodule primordia at 7dpi (A) and nodules at 21 dpi (B). Scale bar = 100 μm. Representative images of 65 μm-thick nodule sections of *LjRBOHBpro:NLS*-2xYFP in yellow (C), infected cells (*M. loti* MAFF 303099-GFP, magenta) and the auto-fluorescent lignin and suberin (grey). The border of the nodule is marked with dashed lines. The number in the upper corner indicates the cut sections. Scale bar = 50 μm.

7.2 *Lj*CASPL subfamily 4 members have similar domains as *At*CASPLs

To investigate the protein structure and function, a domain analysis was performed in *Lj*CASPLs. Although not in all cases, members of LjCASPL subfamily 4 appear to have a greater role during symbiosis than other CASPLs, therefore the analysis only included this subfamily. Protein sequences for all eleven members were extracted and domains were annotated using InterProScan. CASPs have a signature region called the Casparian strip membrane domain (CSD) that comprises four transmembrane regions and variable N- and

C- terminal regions (Roppolo et al., 2014). As CSD is constant in different orthologs from different species, the domain analysis only considered the CSD in *Lj*CASPLs of subfamily 4. The four transmembrane domains (TMs) were present in all subfamily 4 *Lj*CASPLs with the exception of *Lj*CASPL4.4, which had only two predicted domains. In a protein, specific residues are required for proper folding, localization, and function (Haspel and Jagodzinski, 2017). Similar to *Arabidopsis* CASPs, the amino acids arginine (R, position 6) and asparagine (D, position 55) were conserved in the TM1 and TM3, respectively (Roppolo et al., 2014). However, hydrophobic residues such as leucine in TM1 (L, positions 7 and 18), tyrosine (Y, position 73) in TM2, phenylalanine (F, position 102) in TM3, and phenylalanine (position 158) in TM4 appear to be unique to the *Lj*CASPLs subfamily 4 (Figure 27A). Additionally, the intracellular loops had variable length as depicted in the number of gaps in the alignment. This has also been observed in CASPL proteins in *Arabidopsis* (Roppolo et al., 2014). Protein function is determined by the subcellular localization (Pan et al., 2021). To explore the *Lj*CASPLs subcellular localization the genomic sequence of *CASPL4.1*, *4.2*, and *4.3* and *CASPL8* were cloned and tagged with GFP. In the DE analysis the *CASPL4.1*, *4.2*, and *4.3* were upregulated in infected compared with non-infected nodules. These genes were tagged in the N-terminal position. *Lj*CASP8 was selected as this protein is an homolog to *At*CASP1. It was tagged at the C-terminus due to the presence of a predicted signal peptide at the N-terminus. Transient expression in *Nicotiana benthamiana* leaves revealed that *Lj*CASP8, *Lj*CASPL4.1 and *Lj*CASPL4.3 localized to the plasma membrane, whereas *Lj*CASPL4.2 and *Lj*CASPL4.3 displayed a cytosolic signal (Figure 27B). *Lj*CASPL4.2 has a methionine in position 115 in the TM3, which differs from the leucine in the other members of CASPL subfamily 4. In general, *Lj*CASPLs have the same protein structure and localization as the *At*CASPLs, which suggests that these proteins might have similar biological functions.
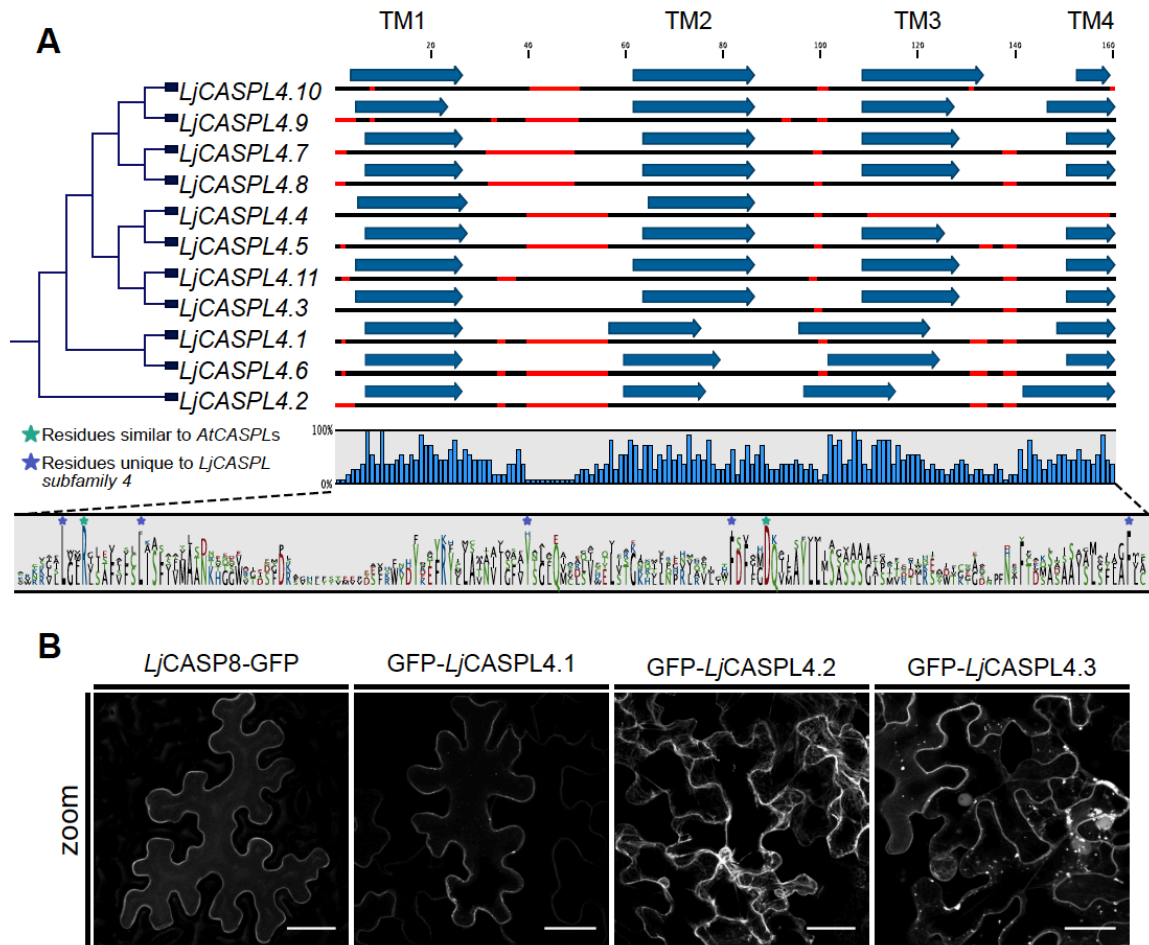
Figure 27. *Lj*CASPL domain analysis and subcellular localization. A) The Casparian strip membrane domain, including N- and C-terminal region, from the eleven protein sequences of *Lj*CASPL subfamily4 was aligned with CLC main workbench (Ver. 7.7.3) and annotated with InterProScan. Transmembrane domains are depicted with a blue arrow. Gaps in the alignment are shown in red and the consensus sequence is depicted at the bottom. Green stars indicate conserved residues in *Arabidopsis* whereas purple stars highlight residues unique to *Lotus*. B) Subcellular localization of selected *Lj*CASP and *Lj*CASPL proteins tagged with GFP. Scale bar = 50 μm.

7.3 Infected nodule cells in *Ljcaspl4.1 caspl4.2* double mutant line have an irregular morphology

To provide insights into the role of CASPLs during SNF, mutant lines were generated using CRISPR/Cas gene editing. Genetic redundancy could prevent observing striking phenotypes in multigene families. In addition, *LORE1* lines were not available. Double mutant lines were generated by individually targeting the first exons of *LjCASPL4.1* and *LjCASPL4.2*. Constructs were generated via Golden Gate cloning (Binder et al., 2014). Constructs carried two guide RNAs (gRNAs) targeting the first exon of each gene, the temperature tolerant and intronic version of the *Lachnospiraceae* sp. Cas gene (*LaCas12a*) and the hygromycin resistance gene for selection (Figure 28A). Stable lines were produced

by inserting these constructs into *L. japonicus* hypocotyls. Genotyping was performed to identify the mutations in *Ljcaspl4.1 caspl4.2* T0 lines. This revealed that T0 plants were heterozygous and therefore *Ljcaspl4.1 caspl4.2* T1 were also genotyped to identify homozygous mutant plants. Two lines were identified: i) *caspl4.1 caspl4.2-1* plants have an 8 nucleotide deletion in *LjCASPL4.1* and a 6 nucleotide deletion in *LjCASPL4.2* (Figure 28B, Supplemental table 10) and ii) *caspl4.1 caspl4.2-2* plants have a 10 nucleotide deletion in *LjCASPL4.1* and 6 nucleotides deletion in *LjCASPL4.2*.
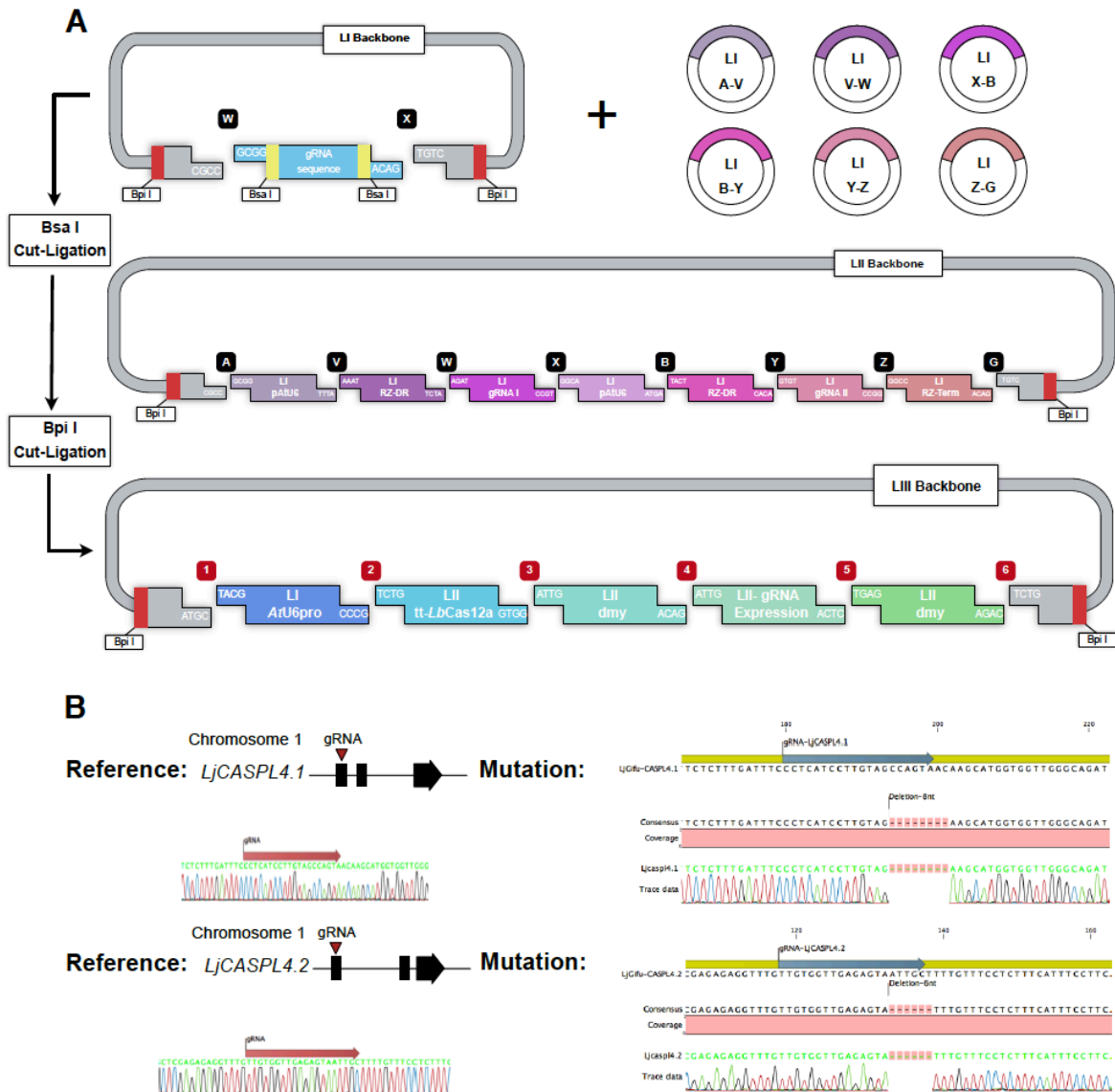
Figure 28. Gene editing using the CRISPR/Cas12a system for *Ljcaspl4.1 caspl4.2*. A) The guide RNAs were cloned into a LI backbone. The binary vector *LIIβF 4-5* contained the *Lotus U6* promoter (*LjU6pro*), the two guide RNAs (gRNAs), two direct repeat sequences (RZ-DR), and "gSNR2" terminator (Term). The vector *LIIIβF 3-4 – BB24* harbored the Hygromycin resistance gene driven by the *Nos_pro* for selection, the temperature tolerant and intronic version of *LbCas12a* driven by the *LjUbq1_pro* for gene editing and the binary vector *LIIβF 4-5* with the gRNAs. B) *LjCASPL4.1* and *LjCASPL4.2* have three exons. Two different gRNAs were designed to target the first exon of each gene indicated by the marron arrows. DNA from T0 plants were extracted and PCR was performed to identify mutations via Sanger sequencing.

Phenotyping was conducted in *caspl4.1 caspl4.2-1* T1 segregating and wild-type *L. japonicus* Gifu (WT) plants, plants with a heterozygous allele in both *caspl4.1-1 and caspl4.2-1* genes (named HET), and plants carrying a homozygous allele for *caspl4.1-1* but heterozygous for *caspl4.2-1* (referred to *caspl4.1*). A significant reduction in shoot length and number of nodules was observed between the WT and HET plants. Homozygous *caspl4.1* plants did not show any significant phenotype when compared with WT plants (Figure 27). The *CASPL4.1pro* was active in infected cells, which suggested a function inside these cells. Thus, infected cells were examined in WT and homozygous double mutant *caspl4.1 caspl4.2-1* plants by fluorescent microscopy. Nodules of wild-type plants contained fully colonized cells with a visible nucleus and with define appearance (Figure 29D and E). In contrast, double mutant *caspl4.1 caspl4.2-1* nodules had an undefined or absent nucleus and granular-like appearance in infected cells (Figure 29F and G).
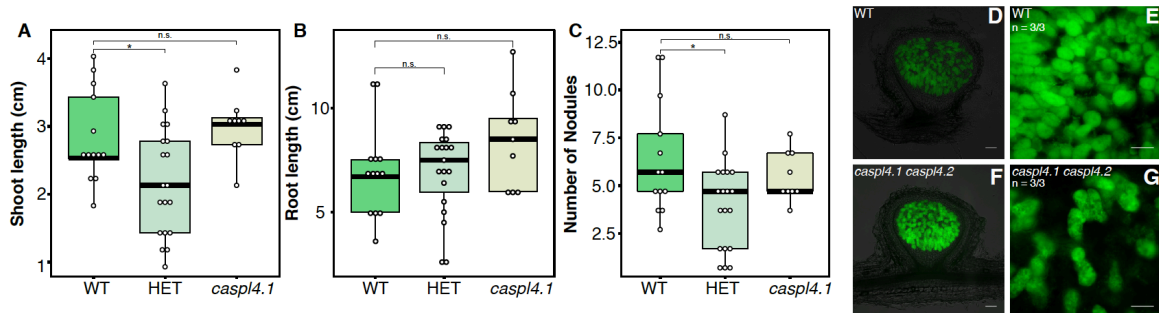
Figure 29. Phenotypic characterization of the CRISPR/Cas line *caspl4.1 caspl4.2-1.* A) shoot length, B) root length, and C) number of nodules were measured in *Lotus japonicus* Gifu plants, heterozygous plants for both *CASPL* alleles and *caspl4.1* (heterozygous for *caspl4.2-1*) mutant plants. Representative nodule sections of *L. japonicus* Gifu (D and E) and homozygous *caspl4.1 caspl4.2-1* double mutant (F and G). Scale bar 100 μm.

In addition, other CRISPR/Cas12a lines were generated targeting *LjCASPL4.3 and LjRBOHB.* A second double mutant line *Ljcaspl4.1 caspl4.3* was generated. The first gRNA targeted the exon 1 of *LjCASPL4.1* and the second gRNA targeted the 3'UTR region. Two independent mutant lines were generated for *LjCASPL4.1* where the gRNAs targeted the first and the third exon. Finally, for *LjRBOHB*, five independent lines were generated. The gRNA targeted the first and the eleventh exon. Four independent lines were generated. In the majority of the cases, heterozygous lines were generated (Supplemental Table 10)

Discussion

1. Natural diversity as a tool to dissect the genetic landscape of root nodule symbiosis

One of the long-standing goals in the RNS field has been to transfer nitrogen-fixation to non-legume crops. This is a monumental task, as this symbiosis is a complex trait regulated by hundreds of genes (reviewed in Mus et al., 2016; Pankievicz et al., 2019). For this endeavor to be successful, it is essential to understand the developmental programs that control nodule organogenesis, the cell biology that underlies the hosting of bacteria and the metabolic adaptations that fuel nitrogen-fixation. However, the high interconnectivity between different pathways makes it difficult to genetically dissect the contribution of candidate genes to each process.

Exploration of the natural diversity of RNS harbors great potential to study specific traits. It has been argued that natural diversity can be used to investigate context-dependent interactions, such as epistasis or developmental dependency (Eguchi et al., 2019). Moreover, the avenue of modern sequencing technologies makes the study of multiple non-model species or accessions possible. Exploration of natural diversity by comparative RNA-seq has been successfully used to identify genes in different plant species such as *Arabidopsis*, soybean (*Glycine max),* cotton (*Gossypium spp.*), and maize (*Zea mays*) (Kusunoki et al., 2017; Du et al., 2019; Xu et al., 2020; Kost et al., 2020). Here, the nodule transcriptomes of six *L. japonicus* accessions were sequenced to bypass the epistatic effect of nodule organogenesis over nodule infection and reveal genes whose expression was specifically associated with infected nodules.

The phenotypic diversity of *L. japonicus* accessions in response to *R. leguminosarum* Norway was used to identify combinations that disconnected nodule formation and nodule cell infection. In contrast to other well-characterized interactions (e.g. *L. japonicus* and *M. loti; M. truncatula* and *S. meliloti*), in this system, the formation of nodules and their infection do not co-occur simultaneously (Liang et al., 2019). Consequently, in some accessions, the nodules remained uninfected, while in others they were highly colonized (Table 2 and Figure 2). In addition, the nodules induced by *R. leguminosarum* Norway did not fix nitrogen. Nodule formation and diversity have been studied in the context of symbiont compatibility where rhizobia trigger different nodule outcomes (Walker et al., 2020). Compatibility depends on both the host (NF recognition) and the symbiont (*nod* genes) (Jiao et al., 2015). One example that resembles the results presented here is the legume *Sophora flavescens* inoculated with a broad spectrum of rhizobia (Jiao et al., 2015). In those interactions, the

symbiosomes in infected cells display a range of phenotypes among the combinations. These include size, number of bacteroids, and structural differences in the peribacteroid membrane, among others (Jiao et al., 2015). Thus, these interactions and the study presented here offer the opportunity to do comparative transcriptomic analysis.

To examine the transcriptomic response of nodules in accessions with infected and non-infected phenotypes, the highly sensitive and efficient prime-seq method was adapted (Janjic et al., 2022). Currently, there is a trend to sequence at the level of single-cell level; however, bulk RNA-seq methods still offer more experimental flexibility and complement single-cell studies (Janjic et al., 2022). When comparing prime-seq with the most common standard protocols like Truseq (Illumina) and NEBNext (New England Biolabs), three unique characteristics make the prime-seq approach powerful: i) adding UMIs reduce the amplification noise derived from the PCR amplification and offer a more accurate estimation of the gene expression in the samples, ii) the addition of a barcode allows high-throughput processing reducing costs, and iii) by sequencing from the 3' end any bias originated from longer or shorter transcripts is minimized (Janjic et al., 2022). Unfortunately, this method is not suitable for studying gene isoforms, and it does not capture mRNA from bacteria and/or organelles (Janjic et al., 2022). Yet, the combination of the distinct phenotypes with sensitive transcriptomics provided a unique opportunity to identify genes directly connected to nodule infection, independent of genes associated with nodule organogenesis or nitrogen fixation. As stated before, the system reported here disconnects nodule formation from infection (Figure 2) and *R. leguminosarum* Norway does not fix nitrogen. We found a discrete number of genes encoding transporters, proteins involved in oxygen homeostasis, and enzymes involved in redox reactions and secondary cell wall modifications among the candidates identified. These constitute a valuable source to explore the genetic base of the adaptations required to host bacteria inside of nodules.

2. A discrete set of genes is differentially regulated in infected nodules

Transcriptome studies complement genetic screenings in RNS. Few studies resemble the approach taken here as they compare specific tissues of infected cells against non-infected cells, in the model organisms *L. japonicus* and *M. truncatula*. The most noteworthy examples are: i) a cell- and tissue-specific transcriptome of *M. truncatula* nodules using laser-capture microdissection (Limpens et al., 2013), ii) a transcriptome in different tissues of *M. truncatula* nodules using laser-capture microdissection and high-depth RNA-seq (Jardinaud et al., 2016), iii) a study comparing the transcriptomes of *L. japonicus* upon inoculation with rhizobia that colonize either inter- or intracellularly (Montiel et al., 2020),

and iv) a single cell-type transcriptome between infected and non-infected cells in *L. japonicus* (Wang et al., 2022). Unlike some of the other studies, total RNA was extracted from whole nodules reducing the complexity of sample handling that is required in laser-capture microdissection. Furthermore, using entire nodules allowed flexibility in the experimental design and cost reduction. The approach taken here serves as a method to identify new genetic players in RNS and to complement previous studies.

Despite the differences between this and previous studies, genes within similar functional categories were identified. Previous studies identified genes related to nodule cell differentiation, cell wall modifications, ROS production, and metabolite transport (Figure 10 and 11). Co-expression analysis from transcriptome studies can identify genes with related functions. In this study, three clusters of genes were identified suggesting similar functions. A similar approach was taken by Poehlman et al. where using co-expression and link community network analysis revealed highly interconnected subnetworks in the transcriptome of the maturation zone from *M. truncatula* nodules. The major findings included carbohydrate and CKs production (Poehlman et al., 2019). Identification of differently regulated linked networks in combination with biological testing offers a framework for new hypotheses and experiments. Therefore, the power of this study also resides in the identification of co-expressed modules during symbiosis.

Among the 167 DEGs, we detected genes from the common symbiotic pathway, such as the *Symbiosis receptor kinase* (*SYMRK*) and *Nodulation Signaling Pathway 2* (*NSP2*). SYMRK is essential for calcium spiking and infection thread formation (Stracke et al., 2002; Miwa et al., 2006). In addition, it has been proposed that SYMRK plays a role during bacteria release in nodules (Kosuta et al., 2011; Saha et al., 2016). In *Sesbania rostrata,* a legume that is infected via crack entry, RNAi of *SYMRK* impairs bacteria release into nodule cells. In addition, upon infection high expression of *SrSYMRK* occurs in epidermal cells but diminishes in mature nodules (Capoen et al., 2005). In our data, *SYMRK* is downregulated in infected nodules (Supplemental Table 9), indicating that higher expression is no longer required once cells have been infected. However, in *M. truncatula* the expression of *Does not Make Infection* 2 (*DMI2*) the ortholog of *LjSYMRK* is required for cell division in cortical cells and symbiosome formation (Catoira et al., 2000; Limpens et al., 2005). The transcription factor NSP2*,* along with NSP1 and DELLA forms a complex which activates the expression of different symbiotic genes (Hirsch et al., 2009; Jin et al., 2016). In our analyses, *NSP2* was specifically upregulated in nodules of infected accessions, suggesting a role in later steps of the symbiosis. It is possible that NSP2 activates the expression of other unknown genes to promote the infection in *Lotus* accessions with an infected

phenotype, but further studies are needed. These examples illustrate the sensitivity of our approach and open up the possibility that already discovered genes have additional functions at later stages of the symbiosis. Other well characterized genes included leghemoglobins, which highlight the importance of oxygen homeostasis. Tight control of oxygen levels is required for the nitrogenase activity and therefore efficient nitrogen fixation (review in Rutten and Poole, 2019). Other findings, common among all the studies previously mentioned, were the genes encoding the symbiotic sulfate transporter SST1, members of the families of calmodium-like proteins, BAM3-like receptors, expansin-like genes, and peptide transporters. As many of these genes are common across several studies but have no reported functions, they are strong candidates for genetical characterization. One such candidate is the gene *MtENOD8.1*, which is the ortholog of the *GDSL lipase esterase* gene identified in this study. This gene is a common findings between this study and the one reported by Limpens et al. Interestingly, *MtENOD8.1pro* was active in the nodule endodermis/parenchyma (Limpens et al., 2013). GDSL lipases have putative functions in suberin biosynthesis and have been identified in suberin-enriched tissues in poplar (Rains et al., 2018). A combination of all these studies could identify specific candidate genes for RNS. This is important as RNA-seq studies end up with a large number of transcripts that are difficult to validate.

The last category includes genes associated with ROS production. ROS have a harmful chemical nature, but evidence indicates that they serve as secondary messengers in plant cells under specific conditions (Choudhury et al., 2013). There is evidence that these compounds are required for both infection and organogenesis during RNS (Mandon et al., 2009). The generation of ROS is linked to NF perception (Ramu et al., 2002) and to nodule metabolic activity (Dalton et al., 1991). Furthermore, different transcriptomic studies have identified enrichment of genes involved in redox reactions and ROS production during RNS at both intracellular and intercellular infection points (Høgslund et al., 2009; Roux et al., 2014; Montiel et al., 2021). Thus, nodules require mechanisms to cope with the effect of oxidative stress. The transcripts of two genes encoding key enzymes in the synthesis of L-ascorbate, a major antioxidant, have been detected in *Lotus* nodules (Matamoros et al., 2006). It is suggested that this compound along with others protects the nitrogenase, the leghemoglobins, and other proteins, which are prone to oxidation (Dalton et al., 1986). This study revealed the presence of genes that take part in the synthesis of L-ascorbate. For example, a gene encoding a GDP-mannose 3,5-epimerase (reviewed in Akram et al., 2017) was upregulated in infected nodules. In addition, transcripts encoding two glutathione S-transferases were also identified. These are part of a ubiquitous gene family that attenuates oxidative stress (Gullner et al., 2018).

3. Symbiotic transporters are induced during RNS

One category enriched in the data produced in this work was transporters. During symbiosis, numerous bacteria and plant transporters are required for the exchange of metabolites and minerals between the symbionts (Udvardi and Poole, 2013). These include a broad range of compounds containing elements such as carbon, nitrogen, phosphorus, sulfur, peptides, iron, copper, and molybdenum, in addition to peptides. Among the identified genes, eleven encoded transporters were induced in infected compared to non-infected nodules (Supplemental Table 9). From these transporters, only *Lj*SST1 has been characterized during RNS. *Lj*SST1 locates in the symbiosome membrane and transports sulfate to the rhizobia. It has been speculated that sulfate is required for the synthesis of the nitrogenase and other proteins (Krusell, 2005; Schneider et al., 2019).

Metal ions are important as they serve as cofactors for different proteins; for instance, iron is a cofactor of plant leghemoglobins and bacterial nitrogenase cores (Brear et al., 2013). On the other hand, copper forms cupro-proteins, which include cytochromes, superoxide dismutase, and laccases for cell wall remodeling (Senovilla et al., 2018). In general, iron and copper transporters are the means by which plants provide components with these elements to the bacteroids (Johnston et al., 2001). A search for these transporters has been done in the *M. truncatula* genome, which revealed the genes *MtNRAMP1* and *MtCOPT1* that are iron and copper transporters, respectively (Tejada-Jiménez et al., 2015; Senovilla et al., 2018). In this study, an iron transporter and a copper transporter were upregulated in infected nodules. The iron transporter was co-expressed with the leghemoglobin genes, which perhaps provides iron for the synthesis of the leghemoglobin core. As for the copper trasporter, it was not expressed in module 2 or 3 but cannot be ruled out as having a function for RNS.

Nitrate and ammonium are sources of nitrogen for plant growth and development. Assimilation, sensing, and distribution of these compounds in the plant require nitrate transporters, which are specialized in the uptake of nitrogen in the form of peptides and a wide variety of nitrogenated compounds (Léran et al., 2014; Valkov et al., 2020). *L. japonicus* has 86 nitrate transporters (Sol et al., 2019; Valkov et al., 2020) divided into four subfamilies (Criscuolo et al., 2012). So far four different symbiotic nitrate transporters have been characterized in both *L. japonicus* and *M. truncatula* (Table 1). Mutations in those transporter genes lead to impaired nitrogen fixation and starvation (Valkov et al., 2017, 2020; Wang et al., 2020a; Vittozzi et al., 2021). Single mutants produce strong phenotypes

despite their coding proteins being part of a multigene family, thus the nitrate transporter gene identified in this study (*LjNRT1.1*, LotjaGi4g1v0207100) could be a potential target to characterize. Other potential substrates for this class of transporters are plant peptides affecting nodulation, such as CLV3/ESR-related (CLE) peptides, which negatively regulate nodulation, and C-terminally encoded peptides (CEP) that positively control nodulation under low nitrogen (review in Kereszt et al., 2018). Nitrate transporters have a promiscuous nature as they can transport other nitrogenated compounds. Although it is possible that the identified *Lj*NRT1.1 transports peptides, a transcript encoding a peptide transporter was also identified in infected nodules (gene ID LotjaGi1g1v0309300).

Carbon supply to the bacteria is of utmost importance during the symbiosis, e.g. malate is used by rhizobia. Moreover, polyols are specific metabolites that serve as a carbon and energy source (Noiraud et al., 2001). The *Lj*SWEET3 and *Mt*SWEET11 transporters are expressed in nodules and they do not have any symbiotic phenotype (Kryvoruchko et al., 2016; Sugiyama et al., 2017). It has been argued that other carbon transporters could compensate for single mutant phenotypes because they have similar biochemical activities such as *Lj*SUT4 and/or *Lj*ALMT4 in *L. japonicus* (Flemetakis et al., 2003; Takanashi et al., 2016). In this study, a polyol transporter (LotjaGi2g1v0391600) and a sugar transporter (LotjaGi4g1v0116000) were identified in co-expressed modules that are specific for nodule tissue.

In general, the majority of the identified transporter candidates have not been described in the context of RNS. Thus, they constitute a valuable resource to investigate new transporters while considering possible genetic redundancy. It is also tempting to speculate that some of these candidates localize in the symbiosome membrane and mediate the transport of compounds required to sustain bacteroid maintenance.

4. Nodule-induced suberin biosynthesis genes that play a role in the suberization of the nodule endodermis

In plants, the root endodermis serves as a point for nutrient uptake, pathogen-induced defense, abiotic stress protection, and gas exchange (review in Barberon, 2017). The endodermis formation requires different levels of modifications: i) Casparian strip, a ring-like structure in the center of the endodermis, ii) suberin lamellae, a secondary cell wall that covers endodermal cells, and iii) in some plants a tertiary cell wall thickening (Krömer, 1903; Layers, 2013). Cell wall modifications are known to be important for the successful establishment of RNS (Brewin, 2004), secondary cell walls have been poorly investigated

in the context of the symbiosis despite their potential roles. In the present study, the differentially expressed analysis revealed a large number of genes associated with secondary cell wall biosynthesis.

Determinate nodules have an endodermis or sclerenchyma layer (nodule endodermis) that it is primarily composed of suberin and lignin (Frazer, 1942; Guinel, 2009). It has been proposed that this cell layer reduces oxygen diffusion (Dakora and Atkins, 1989; Łotocka, 2007). Witty and Minchin first described a drop of oxygen to nanomolar concentration in the inner cortex outside the bacteroids (Witty et al., 1987; King et al., 1988; Kuzma et al., 1993). Thus, the authors propose a diffusion barrier in the nodule cortex of soybean (*G. max*) (Witty and Minchin, 1990). In soybean, the outer and inner cortex are divided by a cell layer that holds secondary cell wall modifications. In broad bean, this layer is composed of cells with suberized cell walls that have similar chemical composition as root endodermal walls (Hartmann et al., 2002). Despite the morphological and chemical composition evidence, we still do not know the genetic components responsible for the formation of this barrier. In this study, we identified genes with functions associated with suberin regulation, biosynthesis, and export that are highly expressed in mature nodules (Figure 12 and 17). The expression of four of these genes was confirmed with RT-qPCR in nodulated roots inoculated with *M. loti* MAFF 303099 (Figure 13). In addition, the *LjFAR3.1*, *LjFAR3.2*, and *LjNAC* promoters were specifically active in the nodule endodermis (Figure 17). This provided evidence that aliphatic long-chain fatty acid synthesis genes are expressed in the nodule endodermis. However, as the constitution of the nodule endodermis is diverse, it still remains unclear if other types of components, e.g. monolignols or polyesters, are required for this cell layer. To provide some evidence of the role of long-chain fatty acid genes in the nodule endodermis, two independent *LORE1* lines in the *LjFAR3.2* gene were characterized (Figure 18). In the two *Ljfar3.2* mutant alleles, disruption in the nodule endodermis was shown by a reduction in the staining with fluorol yellow which stains suberin (Figure 18) (Sexauer et al., 2021). In addition, there was an increase in gas permeability in the two *far3.2* mutant lines compared with their segregating WT* (Figures 20). These lines also had a significant reduction in nitrogen-fixing activity and shoot length (Figures 21). Based on previous data and the results in this study, it can be postulated that alteration in the composition of the nodule endodermis affects nodule function.

Genetic redundancy by paralogs could partially compensate for function and therefore prevent more striking growth phenotypes. *FARs* belong to a multigene family, which makes their study technically challenging. Phylogeny of the FAR protein family revealed that *Lj*FAR3.1 and *Lj*FAR3.2 cluster together with *At*FAR3 and another four potential *L.*

79

japonicus paralogs (Figure 14). *At*FAR3 is associated with cuticular wax synthesis and expressed in leaves, stems, flowers, siliques, and roots (Rowland et al., 2006). In *Arabidopsis FAR1*, *FAR4,* and *FAR5* are responsible for suberin biosynthesis in roots and are particularly expressed at suberin deposition sites (Domergue et al., 2010). On the other hand, *FAR3* orthologs are expressed in bark tissue in poplar (Rains et al., 2018). This suggests that *FAR3* genes have a different role in other plant species. This role could be associated with suberin deposition. Due to the high number of FAR3 homologs in *L. japonicus* and other legumes, it is tempting to speculate on the functional diversification of this family to fulfill nodule-specific functions.

Transcriptional regulation by NAC or other transcription factors plays a role in secondary cell wall modification by regulating the expression of biosynthetic and transport genes in a tissue-specific manner (Mitsuda et al., 2007; Zhong et al., 2007; Mahmood et al., 2019). The specific activation of the *LjNAC* promoter in the nodule endodermis suggests a function in this tissue. Furthermore, *Lj*NAC is the ortholog of *At*NAC38, a transcription factor co-expressed with suberin biosynthesis genes in *Arabidopsis* (Lashbrooke et al., 2016). Thus, *Lj*NAC might either directly activate the expression of suberin biosynthetic genes in a tissue-specific manner or activate the expression of other transcription factors controlling these genes. In addition, it cannot be ruled out that *LjNAC* is activated by an upstream transcriptions factor. Activation of genes encoding proteins that generate long-chain alcohols, such as FARs and CYP86A1, is likely to lead to the production of suberin monomers that later can be exported by transporters of the ABC-G family (Yadav et al., 2014). Lignin and suberin often have interlinked functions in the formation of plant barriers (Graça, 2015; Meents et al., 2018; Zhong et al., 2019). In the transcriptome of this study, lignin biosynthetic genes are less expressed in nodule tissues compared to roots (Figure 11 and 12), so it is still unknown if the list of lignin biosynthetic genes explored here are likely to be involved in the formation of the nodule endodermis. In *Arabidopsis*, it is suggested that lignin metabolites produced in distal cells can be transported from other cells/tissues where they are needed (Andersen et al., 2021); therefore a similar situation could occur in the nodules.

5. Oxygen homeostasis in infected nodules

Oxygen homeostasis is key for optimal nodule functioning. Oxygen is both a potent denaturing agent of the nitrogenase enzyme complex and essential for bacteroid respiration (Gallon, 1981). To reconcile this apparent contradiction, the host creates the oxygen

diffusion barrier in the outer nodule tissue and expresses high quantities of leghemoglobin proteins (reviewed in Rutten and Poole, 2019). Leghemoglobins have a high oxygen-binding capacity via their heme group (Kundu et al., 2003). Nevertheless, they serve as oxygen donors to the cbb_3-type terminal oxidase in the bacteroid respiratory chain encoded by the *fixNOPQ* operon (reviewed in Rutten and Poole, 2019). Furthermore, this oxidase has a high affinity for oxygen, which helps to have an efficient respiration (Preisig et al., 1996). Three leghemoglobin paralogs are specifically induced in *Lotus* nodules, and a triple mutant line interrupted in these genes cannot fix nitrogen (Ott et al., 2005; Wang et al., 2019). These genes were upregulated in infected compared to non-infected nodules (Figure 8). In addition, we identified a gene that encodes a coproporphyrinogen III oxidase (*Lj*COP). An ortholog of this gene in soybean participates in the synthesis of the heme moiety (Madsen et al., 1993), which is a cofactor of hemoproteins, such as the leghemoglobins (Singh and Varma, 2017). Induction of *COP* was detected by comparing soybean and pea nodules against uninfected roots (Santana et al., 1998). Our DE analysis detected the transcript of a *heme oxidase* (*LjHO*) in *Lotus*. This gene encodes a rate-limiting enzyme in the catabolism of the heme moiety and is present in eukaryotes and prokaryotes (reviewed in Lyles and Eichenbaum, 2018). HO has been detected in plants like *Arabidopsis*, wheat (*Triticum aestivum*), rice (*Oryza sativa*), maize, and soybean (He and He, 2014). Production of this enzyme in soybean nodules has been reported and its role as an antioxidant has been suggested in nodules and roots (Balestrasse et al., 2008).

Based on all these results it is proposed that the formation of a molecular fence composed of suberin and lignin in the nodule endodermis prevents the diffusion of oxygen into inner nodule tissues to protect the nitrogenase. Suberin biosynthetic genes such as *FAR*, *CYP86A1* are required to produce both aliphatic and aromatic constituents of suberin. The monomeric units of suberin are transported to the cell wall through ABC transporters where they polymerize. It is suggested here that the NAC transcription factor regulates the expression of these genes. Lignin is also required, but just how this secondary cell wall component is deposited remains unknown. In addition, leghemoglobins, using their prosthetic heme group, bind oxygen to transfer to the bacteroid. In the bacteroid, the highly efficient respiration fuels the nitrogenase. These mechanisms are part of the oxygen homeostasis in the root nodule (Figure 30).
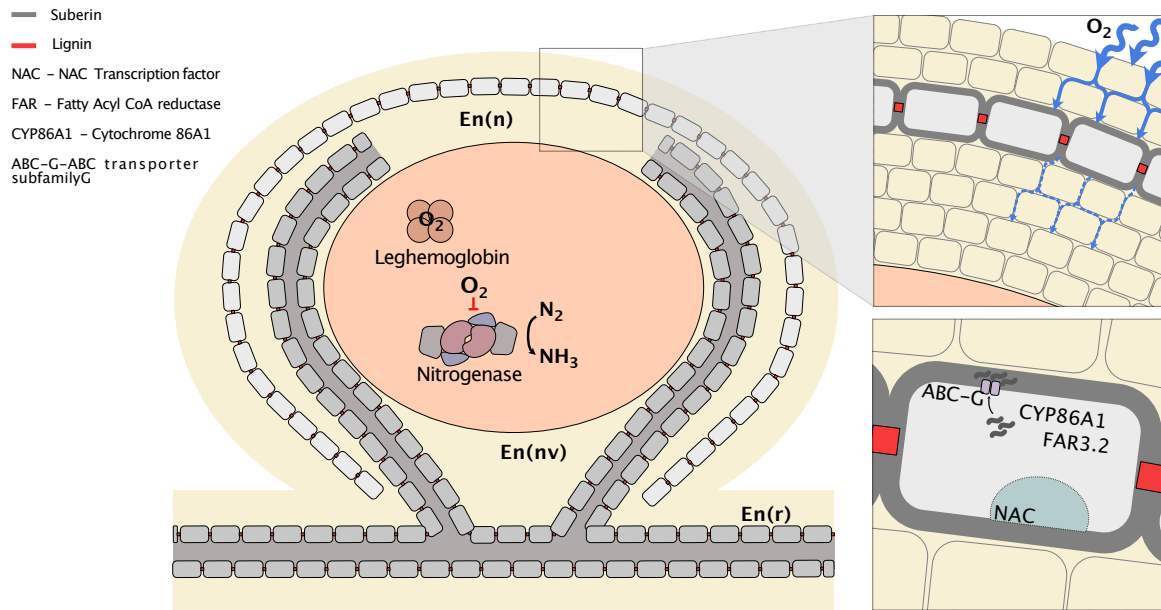
Figure 30. Model of suberization of nodule endodermis. The nodule endodermis located in the periphery of the nodule acts as an oxygen diffusion barrier. Genes encoding putative suberin-related functions are specifically regulated in infected nodules. Interlinked suberin and lignin coat endodermal cells limiting oxygen diffusion, thereby protecting the oxygen-sensitive nitrogenase enzyme. Oxygen is bound by nodule leghemoglobins, ensuring its supply for respiration. NAC: NAC transcription factor, FAR: fatty acyl CoA reductase, CYP86A1: cytochrome 86A1, ABC-G: ABC transporter subfamily G, En: endodermis, r: root, nv: nodule vasculature, n: nodule.

6. A putative role for nodule-induced *Lj*CASPLs during RNS

Casparian strips are ring-like cell wall modifications in the root endodermis of vascular plants and were first described by Robert Caspary (Caspary, 1864). More than a century later their chemical composition was determined to be primarily monolignols which polymerize to form lignin (Schreiber et al., 1999). In the last decade, different genetic players have been identified. Roppolo et al. described the first genetic components to form the Casparian strip, which were named CASPARIAN STRIP MEMBRANE DOMAIN PROTEINS (CASPs) (Roppolo et al., 2011). These proteins localize at what is called the Casparian membrane domain where they recruit lignifying enzymes (Roppolo et al., 2011). In addition, a complex regulatory network is required for Casparian Strip formation. This involves the action of a master transcriptional regulator of the MYB family (AtMYB36) (Liberman et al., 2015). This transcription factor controls the expression of *AtCASP*s, *PEROXIDASE 64* which encodes an enzyme responsible for polymerizing monolignols using ROS (*At*PER64), and the *ENHANCED SUBERIN 1* (*At*ESB1) that is also required for proper formation of the Casparian strip (Roppolo et al., 2011; Hosmani et al., 2013; Lee et al., 2013). CASP-like (CASPL) proteins belong to the same family but unlike CASP they are expressed in different tissues from floral organs to root tips (Roppolo et al., 2014). CASPs

and CASPL proteins have been identified and characterized in different organisms, including cotton (*Gossypium spp.*), watermelon (*Citrullus lanatus, Cl*), tomato (*Solanum lycopersicum*), and maize (Yang et al., 2015; Li et al., 2018; Wang et al., 2020b; Pan et al., 2020). In this work we investigated the function of CASPLs in *Lotus* and their putative role in RNS.

Arabidopsis contains 33 CASPs and CASPLs clustered into five subfamilies (Roppolo et al., 2014). This work revealed that *L. japonicus* has the highest number of orthologs among the investigated species within subfamily 4. This suggests an important role of this subfamily for *L. japonicus;* however, it could not be completely associated with RNS as the other legumes did not have the same number of orthologs (Figure 22). The majority of the *LjCASPL* subfamily 4 members are induced in symbiotic conditions (Figure 23). This provides more evidence of a role in RNS as it is during this process that roots undergo different changes to form the nodule. In watermelon, the *ClCASPL* (Cla004012), the closest ortholog to *AtCASPL4C1,* negatively regulates growth and cold tolerance (Yang et al., 2015). In this case, the function of *ClCASPL* is related to abiotic stress and development rather than Casparian strip formation. Based on this evidence the CASPLs subfamily 4 has a diverse range of biological functions.

The domain topology of *L. japonicus* CASPLs resembled the results reported in *Arabidopsis*, cotton and maize where the Casparian domain has four transmembrane regions, two extracellular loops, an intracellular loop, and cytosolic amino and carboxyl termini (Roppolo et al., 2014; Wang et al., 2020b; Pan et al., 2020). Specific residues were also conserved such as arginine (position R10) and aspartic acid (position D88) in the first and third transmembrane domains. In the *At*CASPL subfamily 4 absence of two cysteine residues in the second extracellular loop is conserved (Roppolo et al., 2014). This was also observed in the *Lj*CASPLs (Figure 27A). Proteins with a similar biochemical function have conserved residues (Mirny and Gelfand, 2002). *Lj*CASPL subfamily 4 has some unique residues in transmembrane regions 1, 3, and 4 (Figure 27A). These conserved positions within a subfamily are called 'specificity-determining positions' and normally are involved in controlling protein function (Benítez-Páez et al., 2012; Chagoyen et al., 2016; Pitarch et al., 2021). Therefore, it is possible that these residues in the *Lj*CASPL proteins confer specificity. In addition, the presence of the hydrophobic residues in the transmembrane regions, leucine (TM1: position L7 and L18) and phenylalanine (TM3: position F102 and TM4 position F158), could be important for proper localization in the plasma membrane as their chemical nature favors interactions with lipids (Donev, 2014). In *Arabidopsis*, CASP1 is likely to missfold when the conserved aspartate (TM3: position D134) in transmembrane

domain 3 is mutated to histidine (Roppolo et al., 2014). Mutating specific residues of *Lj*CASPL could reveal whether these residues are dispensable for proper localization in *Lotus*. Subcellular localization in tobacco epidermal leaves showed that only *Lj*CASPL8 and *Lj*CASPL4.1 were confined to the plasma membrane whereas *Lj*CASPL4.2 and *Lj*CASPL4.3 displayed cytosolic threads (Figure 27B). Roppolo et al. revealed that the absence of specific residues prevents plasma membrane localization, e.g. *At*CASPL3A1, *At*CASPL5A2 and *At*CASPL4D1 lack tryptophan (EL2: position W164) and did not locate at the plasma membrane. In the case *of Lj*CASPL4.2, the presence of methionine (TM3: position M114) that differs from the consensus sequence and occurs in a critical transmembrane domain could modify the localization in the plasma membrane but further studies are needed (Figure 25B).

β-glucuronidase staining in different tissues of species including *Arabidopsis*, tomato, cotton and watermelon revealed that *CASPL*s are expressed in a tissue-specific manner (Roppolo et al., 2014; Yang et al., 2015; Li et al., 2018; Wang et al., 2020b). The *LjCASPL* promoters revealed specific activity patterns in the nodule. For instance, the promoters of *LjCASPL4.1* and *LjCASPL4.9* are active at different stages of the symbiosis (Figure 25). It has been proposed that the diversity and specific expression in different tissues suggest a specialized function. This function could be to generate membrane scaffolds and/or recruit cell-wall modifying enzymes in a tissue-specific manner (Roppolo et al., 2014). Cell wall composition and modifications vary between species and tissues during plant growth development (Montes et al., 2008). During nodule formation, epidermal cells and cells in nodule primordia undergo profound cell wall remodeling (Brewin, 2004; Guinel, 2009). For instance, when IT develops the NPL enzyme is required for localized degradation of the root hair cell wall (Xie et al., 2012) and in soybean the β-expansin gene *GmEXPB2* is proposed to be involved in cell wall modification and expansion during nodule formation (Li et al., 2015). *Lj*CASPL subfamily 4 proteins could be required for local cell wall modification in a tissue-specific manner during RNS. The expression of *LjCASPL*s at different stages of symbiosis provides a valuable tool to study different phases of rhizobia colonization. The promoter activity and expression pattern in different symbiotic conditions could work as specific switches that show the progress in the development of the RNS. In the interaction between the Rosales species *Discaria trinervis* and its symbiont the filamentous nitrogen-fixing bacteria *Frankia* BCU110501, the gene *Dt12* has been used as a marker of epidermal penetration sites (Fournier et al., 2018). In this context, *LjCASPLs* could be used as markers to perform *in vivo* microscopy of different stages of the symbiosis. These markers could be used to study different time points of either IT-dependent or IT-independent ('crack-entry' or intercellular) infection.

Currently, there are few examples outside Arabidopsis where *CASPLs* have been studied to understand their function. In cotton, silencing of *GaCASPL27* increases the number of lateral roots, thus indicating a negative regulation of lateral root development (Wang et al., 2020b). A CRISPR mutant in maize *SEMI-ROLLED LEAF 5* (*SRL5*), a homolog of *AtCASPL2B2*, leads to a disorganized cuticular wax and a semi-rolled leaf phenotype (Pan et al., 2020). In these examples, single mutants showed a strong phenotype, but in other cases, only double or higher order mutants provide a phenotype. In *Arabidopsis*, single mutants in *AtCASPs* did not reveal any phenotype, only when a double mutant was generated (*Atcasp1-1 casp3-1*) disorganized Casparian strip were observed (Roppolo et al., 2011). Roppolo et al. proposed that absence of a phenotype in a single mutant is due to similarities in sequence, expression pattern, and localization among *At*CASPs (Roppolo et al., 2011). Therefore, in this study double mutant lines were developed using CRISPR Cas12a, a gene editing method, which has the capacity to perform multiple editing events and has a superior efficiency compared with other methods such as meganucleases, ZFN (zinc-finger nucleases), and TALEN (transcription activator-like effector nucleases). Furthermore, the CRISPR/Cas method only requires a guide RNA and Cas enzyme which is easily engineered as opposed to complex proteins like TAL effectors (Zhu et al., 2020). In the segregating double mutant *caspl4.1 caspl4.2-1* population, only heterozygous and not *caspl4.1* homozygous mutants showed a significant reduction in shoot length and number of nodules. As *CASPL4.1*, *4.2* and *4.3* have similar expression patterns (Figure 23 and 24), it is likely that they have redundant functions. Nodule sections in *caspl4.1 caspl4.2-1* revealed a granular-like appearance in infected cells (Figure 29G). This resembled the phenotype of *Ljsen1-1*, *Ljsst1-1* and *Ljsym105* mutants. These mutants have infected nodule cells that are highly vacuolated, with irregularly-shaped symbiosomes, and less electron dense material. Nodules in these mutant lines senesced prematurely (Krusell, 2005; Hossain et al., 2006; Hakoyama et al., 2012a). Interestingly, these mutants have different times and degrees of senescence, suggesting that premature senescence is associated with the function of the defective gene. Transmission electron microscopy will be required to confirm the resemblance between the *caspl* mutant and the aforementioned mutants. Based on the putative role of CASPLs and the resemblance of the *caspl4.1 caspl4.2-1* mutant with the other *Lotus* symbiosome mutants, I propose the following hypothesis: local cell wall remodeling takes place at different stages of the symbiosis and they are required to properly allocate the rhizobia in the infected cells. In this case, *Lj*CASPL4.1, *Lj*CASPL4.2, and even *Lj*CASPL4.3 could recruit unknown cell wall modifiers to allow proper development during the intracellular bacterial accommodation.

Conclusion

The interaction between *L. japonicus – R. leguminosarum* Norway can be used to identify genetic determinants of how bacteria enter the plant cell (Liang et al. 2020). In this work, *L. japonicus* natural diversity in combination with prime-seq was used to identify genes related to the intracellular accommodation of bacteria. We identified 167 genes from infected or non-infected nodules. Within the differentially expressed genes a group of genes required in cell wall modifications was identified. Validation and characterization of the *LjFAR3.2* gene indicate that long-chain fatty acids are required in the nodule endodermis to generate a low oxygen environment for the oxygen-labile nitrogenase (Figure 30). In *Ljfar3.2* mutants, the reduction in the content of long-chain fatty acids alters the nodule endodermis and thus alters the oxygen homeostasis.

*LjCASPL*s are other types of genes that highlight the modifications that plant cells undergo to host rhizobia at different stages of the symbiosis. LjCASPLs belong to a multigene family with a putative role in local cell wall modification in different plant tissues. The development of a double mutant using CRISPR/Cas gene editing revealed an interesting phenotype with a granular-like appearance in the infected cells. Further characterization of this new phenotype will aid to understand the role of these genes in rhizobia accommodation. Altogether, this study provides new evidence of the spatiotemporal control of different genes, *FAR*s and *CASPL*s, in the nodule to host the symbiont and exemplifies the importance of the bacteria accommodation for proper nitrogen fixation. Finally, these new players will be essential if in the future we intend to engineer efficient nitrogen fixation in non-legumes.

References

Aguilar, A., Peralta, H., Mora, Y., Diaz, R., Vargas-Lagunas, C., Girard, L., and Mora, J. (2016). Genomic comparison of *Agrobacterium* pusense strains isolated from bean nodules. Front. Microbiol. **7**: 1720.

Akram, N.A., Shafiq, F., and Ashraf, M. (2017). Ascorbic acid-a potential oxidant scavenger and its role in plant development and abiotic stress tolerance. Front. Plant Sci. **8**: 613.

Alexa, A. and Rahnenfuhrer, J. (2010). topGO: enrichment analysis for gene ontology. R Packag. version **2**: 2010.

Alexa, A. and Rahnenführer, J. (2009). Gene set enrichment analysis with topGO. Bioconductor Improv **27**.

Allen, E.K., Allen, O.N., and Newman, A.S. (1953). Pseudonodulation of leguminous plants induced by 2-bromo-3, 5-dichlorobenzoic acid. Am. J. Bot.: 429–435.

Andersen, T.G., Molina, D., Kilian, J., Franke, R.B., Ragni, L., and Geldner, N. (2021). Tissue-autonomous phenylpropanoid production is essential for establishment of root barriers. Curr. Biol.

Ané, J.-M., Kiss, G.B., Riely, B.K., Penmetsa, R.V., Oldroyd, G.E.D., Ayax, C., Lévy, J., Debellé, F., Baek, J.-M., and Kalo, P. (2004). *Medicago truncatula* DMI1 required for bacterial and fungal symbioses in legumes. Science (80). **303**: 1364–1367.

Antolín-Llovera, M., Ried, M.K., and Parniske, M. (2014). Cleavage of the SYMBIOSIS RECEPTOR-LIKE KINASE Ectodomain Promotes Complex Formation with Nod Factor Receptor 5. Curr. Biol. **4**: 422–427.

Appleby, C.A. (1984). Leghemoglobin and Rhizobium respiration. Annu. Rev. Plant Physiol. **35**: 443–478.

Arrighi, J.-F. (2006). The *Medicago truncatula* Lysine Motif-Receptor-Like Kinase gene family includes NFP and new nodule-expressed genes. Plant Physiol. **142**: 265–279.

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

Arthikala, M.-K., Montiel, J., Sánchez-López, R., Nava, N., Cárdenas, L., and Quinto, C. (2017). Respiratory burst oxidase homolog gene a is crucial for rhizobium infection and nodule maturation and function in common bean. Front. Plant Sci. **8**: 2003.

Bagnoli, J.W., Ziegenhain, C., Janjic, A., Wange, L.E., Vieth, B., Parekh, S., Geuder, J., Hellmann, I., and Enard, W. (2018). Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq. Nat. Commun. **9**: 1–8.

Balestrasse, K.B., Yannarelli, G.G., Noriega, G.O., Batlle, A., and Tomaro, M.L. (2008). Heme oxygenase and catalase gene expression in nodules and roots of soybean plants subjected to cadmium stress. Biometals **21**: 433–441.

Barberon, M. (2017). The endodermis as a checkpoint for nutrients. New Phytol. **213**: 1604–1610.

Becana Ausejo, M., Morán, J.F., Iturbe-Ormaetxe, I., Gogorcena Aoiz, Y., and Escuredo, P.R. (1995). Structure and function of leghemoglobins.

Becana, M. and Klucas, R. V (1992). Oxidation and reduction of leghemoglobin in root nodules of leguminous plants. Plant Physiol. **98**: 1217–1221.

Becana, M., Matamoros, M.A., Udvardi, M., and Dalton, D.A. (2010). Recent insights into antioxidant defenses of legume root nodules. New Phytol. **188**: 960–976.

Becana, M., Moran, J.F., and Iturbe-Ormaetxe, I. (1995). Structure and function of leghemoglobins.

Benítez-Páez, A., Cárdenas-Brito, S., and Gutiérrez, A.J. (2012). A practical guide for the computational selection of residues to be experimentally characterized in protein families. Brief. Bioinform. **13**: 329–336.

Berg, J.M., Tymoczko, J.L., and Stryer, L. (2002). Section 24.1, Nitrogen Fixation: Microorganisms Use ATP and a Powerful Reductant to Reduce Atmospheric Nitrogen to Ammonia. Biochemistry.

Bergersen, F.J. (1997). Regulation of nitrogen fixation in infected cells of leguminous root nodules in relation to O_2 supply. Plant Soil **191**: 189–203.

Bergersen, F.J., Gibson, A.H., and Licis, I. (1995). Growth and N2-fixation of soybeans inoculated with strains of *Bradyrhizobium japonicum* differing in energetic efficiency and PHB utilization. Soil Biol. Biochem. **27**: 611–616.

Bergersen, F.J. and Goodchild, D.J. (1973). Aeration pathways in soybean root nodules. Aust. J. Biol. Sci. **26**: 729–740.

Beringer, J.E. (1974). R factor transfer in *Rhizobium leguminosarum*. Microbiology **84**: 188–198.

Binder, A., Lambert, J., Morbitzer, R., Popp, C., Ott, T., Lahaye, T., and Parniske, M. (2014). A modular plasmid assembly kit for multigene expression, gene silencing and silencing rescue in plants. PLoS One **9**: e88218.

Blighe, K., Rana, S., and Lewis, M. (2021). EnhancedVolcano: publication-ready volcano plots with enhanced colouring and labeling. 2020. R package version 1.8. 0.

Blomberg, M.R.A. (2016). Mechanism of oxygen reduction in cytochrome c oxidase and the role of the active site tyrosine. Biochemistry **55**: 489–500.

Bogusz, D., Appleby, C.A., Landsmann, J., Dennis, E.S., Trinick, M.J., and Peacock, W.J. (1988). Functioning haemoglobin genes in non-nodulating plants. Nature **331**:

178–180.

Bohlool, B.B., Ladha, J.K., Garrity, D.P., and George, T. (1992). Biological nitrogen fixation for sustainable agriculture: A perspective. Plant Soil **141**: 1–11.

Bouwman, A.F. (1996). Direct emission of nitrous oxide from agricultural soils. Nutr. Cycl. agroecosystems **46**: 53–70.

Brear, E.M., Bedon, F., Gavrin, A., Kryvoruchko, I.S., Torres-Jerez, I., Udvardi, M.K., Day, D.A., and Smith, P.M.C. (2020). *Gm*VTL1a is an iron transporter on the symbiosome membrane of soybean with an important role in nitrogen fixation. New Phytol. **228**: 667–681.

Brear, E.M., Day, D.A., and Smith, P.M.C. (2013). Iron: an essential micronutrient for the legume-rhizobium symbiosis. Front. Plant Sci. **4**: 359.

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

Brown, S.M. and Walsh, K.B. (1994). Anatomy of the legume nodule cortex with respect to nodule permeability. Funct. Plant Biol. **21**: 49–68.

De Bruijn, F.J. (2015). Biological nitrogen fixation. In Principles of plant-microbe interactions (Springer), pp. 215–224.

van Brussel, A.A.N., Bakhuizen, R., van Spronsen, P.C., Spaink, H.P., Tak, T., Lugtenberg, B.J.J., and Kijne, J.W. (1992). Induction of pre-infection thread structures in the leguminous host plant by mitogenic lipo-oligosaccharides of Rhizobium. Science (80-.). **257**: 70–72.

Calabi-Floody, M., Medina, J., Rumpel, C., Condron, L.M., Hernandez, M., Dumont, M., and de la Luz Mora, M. (2018). Smart fertilizers as a strategy for sustainable agriculture. Adv. Agron. **147**: 119–157.

Capoen, W., Goormachtig, S., De Rycke, R., Schroeyers, K., and Holsters, M. (2005). *Sr*SymRK, a plant receptor essential for symbiosome formation. Proc. Natl. Acad. Sci. **102**: 10369–10374.

Capoen, W., Oldroyd, G., Goormachtig, S., and Holsters, M. (2010). Sesbania rostrata: a case study of natural variation in legume nodulation. New Phytol. **186**: 340–345.

Capoen, W., Sun, J., Wysham, D., Otegui, M.S., Venkateshwaran, M., Hirsch, S., Miwa, H., Downie, J.A., Morris, R.J., and Ané, J.-M. (2011). Nuclear membranes control symbiotic calcium signaling of legumes. Proc. Natl. Acad. Sci. **108**: 14348–14353.

Cárdenas, L., Martínez, A., Sánchez, F., and Quinto, C. (2008). Fast, transient and specific intracellular ROS changes in living root hair cells responding to Nod factors (NFs). Plant J. **56**: 802–813.

Caspary, R. (1864). Bemerkungen über die Schutzscheide und die Bildung des Stammes

und der Wurzel.

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

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

Cernay, C., Pelzer, E., and Makowski, D. (2016). A global experimental dataset for assessing grain legume production. Sci. data **3**: 1–20.

Cerri, M.R., Frances, L., Laloum, T., Auriac, M.-C., Niebel, A., Oldroyd, G.E.D., Barker, D.G., Fournier, J., and de Carvalho-Niebel, F. (2012). *Medicago truncatula* ERN transcription factors: regulatory interplay with NSP1/NSP2 GRAS factors and expression dynamics throughout rhizobial infection. Plant Physiol. **160**: 2155–2172.

Chagoyen, M., García-Martín, J.A., and Pazos, F. (2016). Practical analysis of specificity-determining residues in protein families. Brief. Bioinform. **17**: 255–261.

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

Charpentier, M., Sun, J., Martins, T.V., Radhakrishnan, G. V, Findlay, K., Soumpourou, E., Thouin, J., Véry, A.-A., Sanders, D., and Morris, R.J. (2016). Nuclear-localized cyclic nucleotide–gated channels mediate symbiotic calcium oscillations. Science (80-.). **352**: 1102–1105.

Chen, Y., Chen, W., Li, X., Jiang, H., Wu, P., Xia, K., Yang, Y., and Wu, G. (2014). Knockdown of LjIPT3 influences nodule development in *Lotus japonicus*. Plant Cell Physiol. **55**: 183–193.

Choudhury, S., Panda, P., Sahoo, L., and Panda, S.K. (2013). Reactive oxygen species signaling in plants under abiotic stress. Plant Signal. Behav. **8**: e23681.

Concordet, J.-P. and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res. **46**: W242–W245.

Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics **33**: 2938–2940.

Cooper, J.E. (2004). Multiple responses of rhizobia to flavonoids during legume root infection. In Advances in Botanical Research (Elsevier), pp. 1–62.

Criscuolo, G., Valkov, V.T., Parlati, A., Alves, L.M., and Chiurazzi, M. (2012). Molecular characterization of the *Lotus japonicus* NRT1 (PTR) and NRT2 families.

Plant. Cell Environ. **35**: 1567–1581.

Cummings, S.P., Gyaneshwar, P., Vinuesa, P., Farruggia, F.T., Andrews, M., Humphry, D., Elliott, G.N., Nelson, A., Orr, C., and Pettitt, D. (2009). Nodulation of Sesbania species by Rhizobium (*Agrobacterium*) strain IRBG74 and other rhizobia. Environ. Microbiol. **11**: 2510–2525.

Cutting, J.A. and Schulman, H.M. (1969). The site of heme synthesis in soybean root nodules. Biochim. Biophys. Acta (BBA)-General Subj. **192**: 486–493.

Czelusniak, J., Goodman, M., Hewett-Emmett, D., Weiss, M.L., Venta, P.J., and Tashian, R.E. (1982). Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. Nature **298**: 297–300.

d'Haeze, W., De Rycke, R., Mathis, R., Goormachtig, S., Pagnotta, S., Verplancke, C., Capoen, W., and Holsters, M. (2003). Reactive oxygen species and ethylene play a positive role in lateral root base nodulation of a semiaquatic legume. Proc. Natl. Acad. Sci. **100**: 11789–11794.

Dakora, F.D. and Atkins, C.A. (1989). Diffusion of oxygen in relation to structure and function in legume root nodules. Funct. Plant Biol. **16**: 131–140.

Dalton, D.A., Post, C.J., and Langeberg, L. (1991). Effects of ambient oxygen and of fixed nitrogen on concentrations of glutathione, ascrobate, and associated enzymes in soybean root nodules. Plant Physiol. **96**: 812–818.

Dalton, D.A., Russell, S.A., Hanus, F.J., Pascoe, G.A., and Evans, H.J. (1986). Enzymatic reactions of ascorbate and glutathione that prevent peroxide damage in soybean root nodules. Proc. Natl. Acad. Sci. **83**: 3811–3815.

Damiani, I., Pauly, N., Puppo, A., Brouquisse, R., and Boscari, A. (2016). Reactive oxygen species and nitric oxide control early steps of the legume–rhizobium symbiotic interaction. Front. Plant Sci. **7**: 454.

Day, D.A. and Copeland, L. (1991). Carbon metabolism and compartmentation in nitrogen-fixing legume nodules [symbiotic nitrogen-fixation, dicarboxylic acids, oxygen-limitation, nodule structure]. Plant Physiol. Biochem.

Dimitrijevic, L., Puppo, A., Trinchant, J.-C., and Rigaud, J. (1989). Ferrochelatase activities and heme contents in purified mitochondria from soybean roots and root nodules. J. Plant Physiol. **134**: 642–644.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29**: 15–21.

Domergue, F., Vishwanath, S.J., Joubès, J., Ono, J., Lee, J.A., Bourdon, M., Alhattab, R., Lowe, C., Pascal, S., and Lessire, R. (2010). Three Arabidopsis fatty acyl-coenzyme A reductases, FAR1, FAR4, and FAR5, generate primary fatty

alcohols associated with suberin deposition. Plant Physiol. **153**: 1539–1554.

Donaldson, L. (2020). Autofluorescence in plants. Molecules **25**: 2393.

Donev, R. (2014). Advances in Protein Chemistry and Structural Biology (Academic Press).

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

Eady, R.R. and Postgate, J.R. (1974). Nitrogenase. Nature **249**: 805–810.

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

Escaray, F.J., Menendez, A.B., Gárriz, A., Pieckenstain, F.L., Estrella, M.J., Castagno, L.N., Carrasco, P., Sanjuán, J., and Ruiz, O.A. (2012). Ecological and agronomic importance of the plant genus Lotus. Its application in grassland sustainability and the amelioration of constrained and contaminated soils. Plant Sci. **182**: 121–133.

Esseling, J.J., Lhuissier, F.G.P., and Emons, A.M.C. (2003). Nod factor-induced root hair curling: continuous polar growth towards the point of nod factor application. Plant Physiol. **132**: 1982–1988.

FAO, F. (2017). The future of food and agriculture–Trends and challenges. Annu. Rep. **296**.

Ferreira, H., Pinto, E., and Vasconcelos, M.W. (2021). Legumes as a cornerstone of the transition toward more sustainable agri-food systems and diets in Europe. Front. Sustain. Food Syst. **5**.

Fich, E.A., Segerson, N.A., and Rose, J.K.C. (2016). The plant polyester cutin: biosynthesis, structure, and biological roles. Annu. Rev. Plant Biol. **67**: 207–233.

Flemetakis, E., Dimou, M., Cotzur, D., Efrose, R.C., Aivalakis, G., Colebatch, G., Udvardi, M., and Katinakis, P. (2003). A sucrose transporter, Lj SUT4, is up-regulated during *Lotus japonicus* nodule development. J. Exp. Bot. **54**: 1789–1791.

Fournier, J., Imanishi, L., Chabaud, M., Abdou-Pavy, I., Genre, A., Brichet, L., Lascano, H.R., Muñoz, N., Vayssières, A., and Pirolles, E. (2018). Cell remodeling and subtilase gene expression in the actinorhizal plant Discaria trinervis highlight host orchestration of intercellular *Frankia* colonization. New Phytol. **219**: 1018–1030.

Fournier, J., Teillet, A., Chabaud, M., Ivanov, S., Genre, A., Limpens, E., de Carvalho-Niebel, F., and Barker, D.G. (2015). Remodeling of the infection chamber before infection thread formation reveals a two-step mechanism for rhizobial entry into the host legume root hair. Plant Physiol. **167**: 1233–1242.

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

Frazer, H.L. (1942). XXIV.—The Occurrence of Endodermis in Leguminous Root Nodules

and Its Effect upon Nodule Function. Proc. R. Soc. Edinburgh, Sect. B Biol. Sci. **61**:
328–343.

**Fujita, S., De Bellis, D., Edel, K.H., Köster, P., Andersen, T.G., Schmid-Siegert, E.,
Dénervaud Tendon, V., Pfister, A., Marhavý, P., and Ursache, R.** (2020).
SCHENGEN receptor module drives localized ROS production and lignification in
plant roots. EMBO J. **39**: e103894.

**Fukudome, M., Watanabe, E., Osuki, K.-I., Imaizumi, R., Aoki, T., Becana, M., and
Uchiumi, T.** (2019). Stably transformed *Lotus japonicus* plants overexpressing
phytoglobin LjGlb1-1 show decreased nitric oxide levels in roots and nodules as well
as delayed nodule senescence. Plant Cell Physiol. **60**: 816–825.

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.

Gallon, J.R. (1981). The oxygen sensitivity of nitrogenase: a problem for biochemists and
micro-organisms. Trends Biochem. Sci. **6**: 19–23.

Gamas, P., Brault, M., Jardinaud, M.-F., and Frugier, F. (2017). Cytokinins in symbiotic
nodulation: when, where, what for? Trends Plant Sci. **22**: 792–802.

Gamborg, O.L., Miller, R., and Ojima, K. (1968). Nutrient requirements of suspension
cultures of soybean root cells. Exp. Cell Res. **50**: 151–158.

Gauthier-Coles, C., White, R.G., and Mathesius, U. (2019). Nodulating legumes are
distinguished by a sensitivity to cytokinin in the root cortex leading to pseudonodule
development. Front. Plant Sci. **9**: 1901.

Gavrin, A., Jansen, V., Ivanov, S., Bisseling, T., and Fedorova, E. (2015). ARP2/3-
mediated actin nucleation associated with symbiosome membrane is essential for
the development of symbiosomes in infected cells of *Medicago truncatula* root
nodules. Mol. Plant-Microbe Interact. **28**: 605–614.

**Gavrin, A., Kaiser, B.N., Geiger, D., Tyerman, S.D., Wen, Z., Bisseling, T., and
Fedorova, E.E.** (2014). Adjustment of host cells for accommodation of symbiotic
bacteria: vacuole defunctionalization, HOPS suppression, and TIP1g retargeting in
Medicago. Plant Cell **26**: 3809–3822.

Geldner, N. (2013). The endodermis. Annu. Rev. Plant Biol. **64**: 531–558.

Geurts, R. and Bisseling, T. (2002). Rhizobium Nod factor perception and signalling.
Plant Cell **14**: S239–S249.

Gong, X., Bucerius, S., Jensen, E., and Parniske, M. (2021). A CCaMK/Cyclops
response element in the promoter of *L. japonicus* Calcium-Binding Protein 1 (CBP1)
mediates transcriptional activation in root symbioses. bioRxiv.

Gossmann, J.A., Markmann, K., Brachmann, A., Rose, L.E., and Parniske, M. (2012).
Polymorphic infection and organogenesis patterns induced by a *Rhizobium*

leguminosarum isolate from Lotus root nodules are determined by the host genotype. New Phytol. **196**: 561–573.

Gowda, C.L.L., Rao, P.P., and Bhagvatulu, S. (2009). Global Trends in Production and Trade of Major Grain Legumes, paper presented in International Conference on Grain Legumes: Quality Improvement, Value Addition and Trade, February 14-16, 2009, Indian Society of Pulses Research and Development, Indian Inst.

Goyal, R.K., Mattoo, A.K., and Schmidt, M.A. (2021). Rhizobial–Host Interactions and Symbiotic Nitrogen Fixation in Legume Crops Toward Agriculture Sustainability. Front. Microbiol. **12**: 1290.

Graça, J. (2015). Suberin: the biopolyester at the frontier of plants. Front. Chem. **3**: 62.

Groth, M., Takeda, N., Perry, J., Uchida, H., Dräxl, S., Brachmann, A., Sato, S., Tabata, S., Kawaguchi, M., and Wang, T.L. (2010). NENA, a *Lotus japonicus* homolog of Sec13, is required for rhizodermal infection by arbuscular mycorrhiza fungi and rhizobia but dispensable for cortical endosymbiotic development. Plant Cell: tpc-109.

Guinel, F.C. (2009). Getting around the legume nodule: I. The structure of the peripheral zone in four nodule types. Botany **87**: 1117–1138.

Gullner, G., Komives, T., Király, L., and Schröder, P. (2018). Glutathione S-transferase enzymes in plant-pathogen interactions. Front. Plant Sci. **9**: 1836.

Haag, A.F., Arnold, M.F.F., Myka, K.K., Kerscher, B., Dall'Angelo, S., Zanda, M., Mergaert, P., and Ferguson, G.P. (2013). Molecular insights into bacteroid development during Rhizobium–legume symbiosis. FEMS Microbiol. Rev. **37**: 364–383.

Hachiya, T. and Sakakibara, H. (2017). Interactions between nitrate and ammonium in their uptake, allocation, assimilation, and signaling in plants. J. Exp. Bot. **68**: 2501–2512.

Hakoyama, T., Niimi, K., Yamamoto, T., Isobe, S., Sato, S., Nakamura, Y., Tabata, S., Kumagai, H., Umehara, Y., and Brossuleit, K. (2012a). The integral membrane protein SEN1 is required for symbiotic nitrogen fixation in *Lotus japonicus* nodules. Plant Cell Physiol. **53**: 225–236.

Hakoyama, T., Oi, R., Hazuma, K., Suga, E., Adachi, Y., Kobayashi, M., Akai, R., Sato, S., Fukai, E., and Tabata, S. (2012b). The SNARE protein SYP71 expressed in vascular tissues is involved in symbiotic nitrogen fixation in *Lotus japonicus* nodules. Plant Physiol. **160**: 897–905.

Handberg, K., Stiller, J., Thykjær, T., and Stougaard, J. (1994). Transgenic plants: Agrobacterium-mediated transformation of the diploid legume Lotus japonicus. In Cell Biology (Elsevier), pp. 119–127.

Haney, C.H. and Long, S.R. (2010). Plant flotillins are required for infection by nitrogen-fixing bacteria. Proc. Natl. Acad. Sci. **107**: 478–483.

Hartmann, K., Peiter, E., Koch, K., Schubert, S., and Schreiber, L. (2002). Chemical composition and ultrastructure of broad bean (*Vicia faba* L.) nodule endodermis in comparison to the root endodermis. Planta **215**: 14–25.

Haspel, N. and Jagodzinski, F. (2017). Methods for Detecting Critical Residues in Proteins. In *In Vitro* Mutagenesis (Springer), pp. 227–242.

He, H. and He, L. (2014). Heme oxygenase 1 and abiotic stresses in plants. Acta Physiol. Plant. **36**: 581–588.

Heckmann, A.B., Lombardo, F., Miwa, H., Perry, J.A., Bunnewell, S., Parniske, M., Wang, T.L., and Downie, J.A. (2006). *Lotus japonicus* nodulation requires two GRAS domain regulators, one of which is functionally conserved in a non-legume. Plant Physiol. **142**: 1739–1750.

Heckmann, A.B., Sandal, N., Bek, A.S., Madsen, L.H., Jurkiewicz, A., Nielsen, M.W., Tirichine, L., and Stougaard, J. (2011). Cytokinin induction of root nodule primordia in *Lotus japonicus* is regulated by a mechanism operating in the root cortex. Mol. plant-microbe Interact. **24**: 1385–1395.

Held, M., Hossain, M.S., Yokota, K., Bonfante, P., Stougaard, J., and Szczyglowski, K. (2010). Common and not so common symbiotic entry. Trends Plant Sci. **15**: 540–545.

Held, M., Hou, H., Miri, M., Huynh, C., Ross, L., Hossain, M.S., Sato, S., Tabata, S., Perry, J., and Wang, T.L. (2014). *Lotus japonicus* cytokinin receptors work partially redundantly to mediate nodule formation. Plant Cell **26**: 678–694.

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

Hirsch, S., Kim, J., Muñoz, A., Heckmann, A.B., Downie, J.A., and Oldroyd, G.E.D. (2009). GRAS proteins form a DNA binding complex to induce gene expression during nodulation signaling in *Medicago truncatula*. Plant Cell **21**: 545–557.

Höfer, R., Briesen, I., Beck, M., Pinot, F., Schreiber, L., and Franke, R. (2008). The Arabidopsis cytochrome P450 CYP86A1 encodes a fatty acid ω-hydroxylase involved in suberin monomer biosynthesis. J. Exp. Bot. **59**: 2347–2360.

Høgslund, N., Radutoiu, S., Krusell, L., Voroshilova, V., Hannah, M.A., Goffard, N., Sanchez, D.H., Lippold, F., Ott, T., and Sato, S. (2009). Dissection of symbiosis and organ development by integrated transcriptome analysis of *Lotus japonicus* mutant and wild-type plants. PLoS One **4**: e6556.

Hosmani, P.S., Kamiya, T., Danku, J., Naseer, S., Geldner, N., Guerinot, M. Lou, and Salt, D.E. (2013). Dirigent domain-containing protein is part of the machinery

required for formation of the lignin-based Casparian strip in the root. Proc. Natl. Acad. Sci. **110**: 14498–14503.

Hossain, M.S., Liao, J., James, E.K., Sato, S., Tabata, S., Jurkiewicz, A., Madsen, L.H., Stougaard, J., Ross, L., and Szczyglowski, K. (2012). Lotus japonicus ARPC1 is required for rhizobial infection. Plant Physiol. **160**: 917–928.

Hossain, M.S., Umehara, Y., and Kouchi, H. (2006). A novel fix symbiotic mutant of *Lotus japonicus*, Ljsym105, shows impaired development and premature deterioration of nodule infected cells and symbiosomes. Mol. plant-microbe Interact. **19**: 780–788.

Houston, K., Tucker, M.R., Chowdhury, J., Shirley, N., and Little, A. (2016). The plant cell wall: a complex and dynamic structure as revealed by the responses of genes under stress conditions. Front. Plant Sci. **7**: 984.

Ibáñez, F., Wall, L., and Fabra, A. (2016). Starting points in plant-bacteria nitrogen-fixing symbioses: intercellular invasion of the roots. J. Exp. Bot. **68**: 1905–1918.

Ishikawa, T., Dowdle, J., and Smirnoff, N. (2006). Progress in manipulating ascorbic acid biosynthesis and accumulation in plants. Physiol. Plant. **126**: 343–355.

Janjic, A., Wange, L.E., Bagnoli, J.W., Geuder, J., Nguyen, P., Richter, D., Vieth, B., Vick, B., Jeremias, I., and Ziegenhain, C. (2022). Prime-seq, efficient and powerful bulk RNA sequencing. Genome Biol. **23**: 1–27.

Jardinaud, M.-F., Boivin, S., Rodde, N., Catrice, O., Kisiala, A., Lepage, A., Moreau, S., Roux, B., Cottret, L., and Sallet, E. (2016). A laser dissection-RNAseq analysis highlights the activation of cytokinin pathways by Nod factors in the *Medicago truncatula* root epidermis. Plant Physiol. **171**: 2256–2276.

Jiang, S., Jardinaud, M.-F., Gao, J., Pecrix, Y., Wen, J., Mysore, K., Xu, P., Sanchez-Canizares, C., Ruan, Y., and Li, Q. (2021). NIN-like protein transcription factors regulate leghemoglobin genes in legume nodules. Science (80-.). **374**: 625–628.

Jiao, Y.S., Liu, Y.H., Yan, H., Wang, E.T., Tian, C.F., Chen, W.X., Guo, B.L., and Chen, W.F. (2015). Rhizobial diversity and nodulation characteristics of the extremely promiscuous legume *Sophora flavescens*. Mol. Plant-Microbe Interact. **28**: 1338–1352.

Jin, Y., Liu, H., Luo, D., Yu, N., Dong, W., Wang, C., Zhang, X., Dai, H., Yang, J., and Wang, E. (2016). DELLA proteins are common components of symbiotic rhizobial and mycorrhizal signalling pathways. Nat. Commun. **7**: 12433.

Johnston, A.W.B., Yeoman, K.H., and Wexler, M. (2001). Metals and the rhizobial-legume symbiosis—uptake, utilization and signalling.

Kaczor, C.M., Smith, M.W., Sangwan, I., and O'Brian, M.R. (1994). Plant [delta]-Aminolevulinic Acid Dehydratase (Expression in Soybean Root Nodules and

Evidence for a Bacterial Lineage of the Alad Gene). Plant Physiol. **104**: 1411–1417.

Kaló, P., Gleason, C., Edwards, A., Marsh, J., Mitra, R.M., Hirsch, S., Jakab, J., Sims, S., Long, S.R., and Rogers, J. (2005). Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. Science (80-.). **308**: 1786–1789.

Kamal, N., Mun, T., Reid, D., Lin, J., Akyol, T.Y., Sandal, N., Asp, T., Hirakawa, H., Stougaard, J., and Mayer, K.F.X. (2020). Insights into the evolution of symbiosis gene copy number and distribution from a chromosome-scale *Lotus japonicus* Gifu genome sequence. bioRxiv.

Kanamori, N., Madsen, L.H., Radutoiu, S., Frantescu, M., Quistgaard, E.M.H., Miwa, H., Downie, J.A., James, E.K., Felle, H.H., and Haaning, L.L. (2006). A nucleoporin is required for induction of Ca2+ spiking in legume nodule development and essential for rhizobial and fungal symbiosis. Proc. Natl. Acad. Sci. **103**: 359–364.

Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idesawa, K., Ishikawa, A., and Kawashima, K. (2000). Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. DNA Res. **7**: 331–338.

Kassambara, A. and Kassambara, M.A. (2020). Package 'ggpubr.' R Packag. version 0.1.6.

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.

King, B.J., Hunt, S., Weagle, G.E., Walsh, K.B., Pottier, R.H., Canvin, D.T., and Layzell, D.B. (1988). Regulation of O_2 concentration in soybean nodules observed by *in situ* spectroscopic measurement of leghemoglobin oxygenation. Plant Physiol. **87**: 296–299.

King, B.J. and Layzell, D.B. (1991). Effect of increases in oxygen concentration during the argon-induced decline in nitrogenase activity in root nodules of soybean. Plant Physiol. **96**: 376–381.

Kistner, C. and Parniske, M. (2002). Evolution of signal transduction. **7**: 511–518.

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

Kolde, R. (2019). Pheatmap: Pretty heatmaps (R package version, 2012).

Kondorosi, E. and Kondorosi, A. (2004). Endoreduplication and activation of the

anaphase-promoting complex during symbiotic cell development. FEBS Lett. **567**: 152–157.

Kosuta, S., Held, M., Hossain, M.S., Morieri, G., MacGillivary, A., Johansen, C., Antolín-Llovera, M., Parniske, M., Oldroyd, G.E.D., and Downie, A.J. (2011). *Lotus japonicus* symRK-14 uncouples the cortical and epidermal symbiotic program. Plant J. **67**: 929–940.

Kotchoni, S.O. and Gachomo, E.W. (2009). A rapid and hazardous reagent free protocol for genomic DNA extraction suitable for genetic studies in plants. Mol. Biol. Rep. **36**: 1633–1636.

Kouchi, H., Imaizumi-Anraku, H., Hayashi, M., Hakoyama, T., Nakagawa, T., Umehara, Y., Suganuma, N., and Kawaguchi, M. (2010). How many peas in a pod? Legume genes responsible for mutualistic symbioses underground. Plant Cell Physiol. **51**: 1381–1397.

Kreszies, T., Schreiber, L., and Ranathunge, K. (2018). Suberized transport barriers in *Arabidopsis*, barley and rice roots: From the model plant to crop species. J. Plant Physiol. **227**: 75–83.

Krömer, K. (1903). Wurzelhaut, Hypodermis und Endodermis der Angiospermenwurzel: Arbeit aus d. Bot. Inst. d. Univ. Marburg; Mit 6 Taf.

Krusell, L. (2005). The Sulfate Transporter SST1 Is Crucial for Symbiotic Nitrogen Fixation in *Lotus japonicus* Root Nodules. Plant Cell Online **17**: 1625–1636.

Krusell, L., Krause, K., Ott, T., Desbrosses, G., Krämer, U., Sato, S., Nakamura, Y., Tabata, S., James, E.K., and Sandal, N. (2005). The sulfate transporter SST1 is crucial for symbiotic nitrogen fixation in *Lotus japonicus* root nodules. Plant Cell **17**: 1625–1636.

Kryvoruchko, I.S., Routray, P., Sinharoy, S., Torres-Jerez, I., Tejada-Jiménez, M., Finney, L.A., Nakashima, J., Pislariu, C.I., Benedito, V.A., and González-Guerrero, M. (2018). An iron-activated citrate transporter, MtMATE67, is required for symbiotic nitrogen fixation. Plant Physiol. **176**: 2315–2329.

Kryvoruchko, I.S., Sinharoy, S., Torres-Jerez, I., Sosso, D., Pislariu, C.I., Guan, D., Murray, J., Benedito, V.A., Frommer, W.B., and Udvardi, M.K. (2016). MtSWEET11, a nodule-specific sucrose transporter of *Medicago truncatula*. Plant Physiol. **171**: 554–565.

Kuppusamy, K.T., Endre, G., Prabhu, R., Penmetsa, R.V., Veereshlingam, H., Cook, D.R., Dickstein, R., and VandenBosch, K.A. (2004). LIN, a *Medicago truncatula* gene required for nodule differentiation and persistence of rhizobial infections. Plant Physiol. **136**: 3682–3691.

Kurihara, D., Mizuta, Y., Sato, Y., and Higashiyama, T. (2015). ClearSee: a rapid

optical clearing reagent for whole-plant fluorescence imaging. Development **142**: 4168–4179.

Kuzma, M.M., Hunt, S., and Layzell, D.B. (1993). Role of oxygen in the limitation and inhibition of nitrogenase activity and respiration rate in individual soybean nodules. Plant Physiol. **101**: 161–169.

Kuzma, M.M., Winter, H., Storer, P., Oresnik, I., Atkins, C.A., and Layzell, D.B. (1999). The site of oxygen limitation in soybean nodules. Plant Physiol. **119**: 399–408.

Langfelder, P. and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics **9**: 559.

Laporte, P., Lepage, A., Fournier, J., Catrice, O., Moreau, S., Jardinaud, M.-F., Mun, J.-H., Larrainzar, E., Cook, D.R., and Gamas, P. (2013). The CCAAT box-binding transcription factor NF-YA1 controls rhizobial infection. J. Exp. Bot. **65**: 481–494.

Lashbrooke, J., Cohen, H., Levy-Samocha, D., Tzfadia, O., Panizel, I., Zeisler, V., Massalha, H., Stern, A., Trainotti, L., and Schreiber, L. (2016). MYB107 and MYB9 homologs regulate suberin deposition in angiosperms. Plant Cell **28**: 2097–2116.

Layers, T.S.C. (2013). Structure and function of three suberized cell layers: epidermis, exodermis, and endodermis. Plant roots hidden half.

Lazo, G.R., Stein, P.A., and Ludwig, R.A. (1991). A DNA transformation–competent *Arabidopsis* genomic library in Agrobacterium. Bio/technology **9**: 963–967.

Lee, Y., Rubio, M.C., Alassimone, J., and Geldner, N. (2013). A mechanism for localized lignin deposition in the endodermis. Cell **153**: 402–412.

Lefebvre, B., Timmers, T., Mbengue, M., Moreau, S., Hervé, C., Tóth, K., Bittencourt-Silvestre, J., Klaus, D., Deslandes, L., and Godiard, L. (2010). A remorin protein interacts with symbiotic receptors and regulates bacterial infection. Proc. Natl. Acad. Sci. **107**: 2343–2348.

Lei, M.-J., Wang, Q., Li, X., Chen, A., Luo, L., Xie, Y., Li, G., Luo, D., Mysore, K.S., and Wen, J. (2015). The small GTPase ROP10 of *Medicago truncatula* is required for both tip growth of root hairs and nod factor-induced root hair deformation. Plant Cell **27**: 806–822.

Leigh, G.J. (2004). Haber-bosch and other industrial processes. In Catalysts for nitrogen fixation (Springer), pp. 33–54.

Leigh, J.A. and Coplin, D.L. (1992). Exopolysaccharides in plant-bacterial interactions. Annu. Rev. Microbiol. **46**: 307–346.

Léran, S., Varala, K., Boyer, J.-C., Chiurazzi, M., Crawford, N., Daniel-Vedele, F., David, L., Dickstein, R., Fernandez, E., and Forde, B. (2014). A unified nomenclature of NITRATE TRANSPORTER 1/PEPTIDE TRANSPORTER family

members in plants. Trends Plant Sci. **19**: 5–9.

Levy, J., Bres, C., Geurts, R., Chalhoub, B., Kulikova, O., Journet, E., Ane, J.-M., Lauber, E., Bisseling, T., Denarie, J., Rosenberg, C., and Debelle, F. (2004). A Putative Ca 2 and Calmodulin- Dependent Protein Kinase Required. Science (80-.). **303**: 1361–1364.

Li, P., Yang, M., Chang, J., Wu, J., Zhong, F., Rahman, A., Qin, H., and Wu, S. (2018). Spatial expression and functional analysis of Casparian strip regulatory genes in endodermis reveals the conserved mechanism in tomato. Front. Plant Sci. **9**: 832.

Li, X., Zhao, J., Tan, Z., Zeng, R., and Liao, H. (2015). GmEXPB2, a cell wall β-expansin, affects soybean nodulation through modifying root architecture and promoting nodule formation and development. Plant Physiol. **169**: 2640–2653.

Liang, J., Klingl, A., Lin, Y.-Y., Boul, E., Thomas-Oates, J., and Marín, M. (2019). A subcompatible rhizobium strain reveals infection duality in Lotus. J. Exp. Bot. **70**: 1903–1913.

Libbenga, K.R., Van Iren, F., Bogers, R.J., and Schraag-Lamers, M.F. (1973). The role of hormones and gradients in the initiation of cortex proliferation and nodule formation in *Pisum sativum* L. Planta **114**: 29–39.

Liberman, L.M., Sparks, E.E., Moreno-Risueno, M.A., Petricka, J.J., and Benfey, P.N. (2015). MYB36 regulates the transition from proliferation to differentiation in the Arabidopsis root. Proc. Natl. Acad. Sci. **112**: 12099–12104.

Lievens, S., Goormachtig, S., Den Herder, J., Capoen, W., Mathis, R., Hedden, P., and Holsters, M. (2005). Gibberellins are involved in nodulation of *Sesbania rostrata*. Plant Physiol. **139**: 1366–1379.

Limpens, E., Franken, C., Smit, P., Willemse, J., Bisseling, T., and Geurts, R. (2003). LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. Science. **302**: 630–633.

Limpens, E., Ivanov, S., van Esse, W., Voets, G., Fedorova, E., and Bisseling, T. (2009). *Medicago* N$_2$-fixing symbiosomes acquire the endocytic identity marker Rab7 but delay the acquisition of vacuolar identity. Plant Cell **21**: 2811–2828.

Limpens, E., Mirabella, R., Fedorova, E., Franken, C., Franssen, H., Bisseling, T., and Geurts, R. (2005). Formation of organelle-like N2-fixing symbiosomes in legume root nodules is controlled by DMI2. Proc. Natl. Acad. Sci. **102**: 10375–10380.

Limpens, E., Moling, S., Hooiveld, G., Pereira, P.A., Bisseling, T., Becker, J.D., and Küster, H. (2013). Cell-and tissue-specific transcriptome analyses of *Medicago truncatula* root nodules. PLoS One **8**: e64377.

Liu, C.-W., Breakspear, A., Guan, D., Cerri, M.R., Abbs, K., Jiang, S., Robson, F.C., Radhakrishnan, G., Roy, S., and Bone, C. (2019a). NIN acts as a Network Hub

Controlling a Growth Module Required for Rhizobial Infection. Plant Physiol.: pp-01572.

Liu, C.-W., Breakspear, A., Stacey, N., Findlay, K., Nakashima, J., Ramakrishnan, K., Liu, M., Xie, F., Endre, G., and de Carvalho-Niebel, F. (2019b). A protein complex required for polar growth of rhizobial infection threads. Nat. Commun. **10**: 1–17.

Liu, C.-W. and Murray, J.D. (2016). The role of flavonoids in nodulation host-range specificity: an update. Plants **5**: 33.

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

Liu, J., Liu, M.X., Qiu, L.P., and Xie, F. (2020). SPIKE1 activates the GTPase ROP6 to guide the polarized growth of infection threads in *Lotus japonicus*. Plant Cell **32**: 3774–3791.

Liu, J., Rutten, L., Limpens, E., Van Der Molen, T., Van Velzen, R., Chen, R., Chen, Y., Geurts, R., Kohlen, W., and Kulikova, O. (2019c). A remote cis-regulatory region is required for NIN expression in the pericycle to initiate nodule primordium formation in *Medicago truncatula*. Plant Cell **31**: 68–83.

Lorenzo, C. de, Lucas, M.M., Vivo, A., and De Felipe, M.R. (1990). Effect of nitrate on peroxisome ultrastructure and catalase activity in nodules of *Lupinus albus* L. cv. Multolupa. J. Exp. Bot. **41**: 1573–1578.

Lorite, M.J., Estrella, M.J., Escaray, F.J., Sannazzaro, A., Videira e Castro, I.M., Monza, J., Sanjuán, J., and León-Barrios, M. (2018). The Rhizobia-Lotus symbioses: deeply specific and widely diverse. Front. Microbiol.: 2055.

Łotocka, B. (2007). Vascular endodermis in root nodules of *Lupinus luteus* L.(Fabaceae). Acta Biol. Cracoviensia Ser. Bot. **49**: 73–80.

Love, M., Anders, S., and Huber, W. (2014a). Differential analysis of count data–the DESeq2 package. Genome Biol **15**: 10–1186.

Love, M.I., Huber, W., and Anders, S. (2014b). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. **15**: 550.

Lyles, K. V and Eichenbaum, Z. (2018). From host heme to iron: the expanding spectrum of heme degrading enzymes used by pathogenic bacteria. Front. Cell. Infect. Microbiol. **8**: 198.

Madsen, E.B., Madsen, L.H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., Sato, S., Kaneko, T., Tabata, S., and Sandal, N. (2003). A receptor kinase gene of the LysM type is involved in legumeperception of rhizobial signals. Nature **425**: 637.

Madsen, L.H., Tirichine, L., Jurkiewicz, A., Sullivan, J.T., Heckmann, A.B., Bek, A.S., Ronson, C.W., James, E.K., and Stougaard, J. (2010a). The molecular network

governing nodule organogenesis and infection in the model legume *Lotus japonicus*. Nat. Commun. **1**.

Madsen, L.H., Tirichine, L., Jurkiewicz, A., Sullivan, J.T., Heckmann, A.B., Bek, A.S., Ronson, C.W., James, E.K., and Stougaard, J. (2010b). The molecular network governing nodule organogenesis and infection in the model legume *Lotus japonicus*. Nat. Commun. **1**: 10.

Madsen, O., Sandal, L., Sandal, N.N., and Marcker, K.A. (1993). A soybean coproporphyrinogen oxidase gene is highly expressed in root nodules. Plant Mol. Biol. **23**: 35–43.

Mahmood, K., Zeisler-Diehl, V.V., Schreiber, L., Bi, Y.-M., Rothstein, S.J., and Ranathunge, K. (2019). Montes et al. Int. J. Mol. Sci. **20**: 6117.

Małolepszy, A. et al. (2016a). The *LORE1* insertion mutant resource. Plant J. **88**: 306–317.

Małolepszy, A., Mun, T., Sandal, N., Gupta, V., Dubin, M., Urbański, D., Shah, N., Bachmann, A., Fukai, E., and Hirakawa, H. (2016b). The *LORE1* insertion mutant resource. Plant J. **88**: 306–317.

Mandon, K., Pauly, N., Boscari, A., Brouquisse, R., Frendo, P., Demple, B., and Puppo, A. (2009). ROS in the legume-Rhizobium symbiosis. React. Oxyg. species plant Signal.: 135–147.

Marchal, K. and Vanderleyden, J. (2000). The" oxygen paradox" of dinitrogen-fixing bacteria. Biol. Fertil. soils **30**: 363–373.

Marino, D., Andrio, E., Danchin, E.G.J., Oger, E., Gucciardo, S., Lambert, A., Puppo, A., and Pauly, N. (2011). A *Medicago truncatula* NADPH oxidase is involved in symbiotic nodule functioning. New Phytol. **189**: 580–592.

Márquez, A.J., Stougaard, J., Udvardi, M., Parniske, M., Spaink, H., Saalbach, G., Webb, J., and Chiurazzi, M. (2005). *Lotus japonicus* handbook (Springer).

Matamoros, M.A. and Becana, M. (2020). Redox control of the legume-Rhizobium symbiosis. Adv. Bot. Res. **94**: 67–96.

Matamoros, M.A., Loscos, J., Coronado, M.J., Ramos, J., Sato, S., Testillano, P.S., Tabata, S., and Becana, M. (2006). Biosynthesis of ascorbic acid in legume root nodules. Plant Physiol. **141**: 1068–1077.

Mbengue, M., Camut, S., de Carvalho-Niebel, F., Deslandes, L., Froidure, S., Klaus-Heisen, D., Moreau, S., Rivas, S., Timmers, T., and Hervé, C. (2010). The *Medicago truncatula* E3 ubiquitin ligase PUB1 interacts with the LYK3 symbiotic receptor and negatively regulates infection and nodulation. Plant Cell: tpc-110.

Mcallister, C.H., Beatty, P.H., and Good, A.G. (2012). Engineering nitrogen use efficient crop plants: The current status. Plant Biotechnol. J. **10**: 1011–1025.

Meents, M.J., Watanabe, Y., and Samuels, A.L. (2018). The cell biology of secondary cell wall biosynthesis. Ann. Bot. **121**: 1107–1125.

De Mendiburu, F. and Simon, R. (2015). Agricolae-Ten years of an open source statistical tool for experiments in breeding, agriculture and biology (PeerJ PrePrints).

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

Middleton, P.H., Jakab, J., Penmetsa, R.V., Starker, C.G., Doll, J., Kaló, P., Prabhu, R., Marsh, J.F., Mitra, R.M., and Kereszt, A. (2007). An ERF transcription factor in *Medicago truncatula* that is essential for Nod factor signal transduction. Plant Cell **19**: 1221–1234.

Millar, A.H., Day, D.A., and Bergersen, F.J. (1995). Microaerobic respiration and oxidative phosphorylation by soybean nodule mitochondria: implications for nitrogen fixation. Plant. Cell Environ. **18**: 715–726.

Minchin, F.R. (1997). Regulation of oxygen diffusion in legume nodules. Soil Biol. Biochem. **29**: 881–888.

Mirny, L.A. and Gelfand, M.S. (2002). Using orthologous and paralogous proteins to identify specificity determining residues. Genome Biol. **3**: 1–20.

Mitsuda, N., Iwase, A., Yamamoto, H., Yoshida, M., Seki, M., Shinozaki, K., and Ohme-Takagi, M. (2007). NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of *Arabidopsis*. Plant Cell **19**: 270–280.

Miwa, H., Sun, J., Oldroyd, G.E.D., and Downie, J.A. (2006). Analysis of Nod-factor-induced calcium signaling in root hairs of symbiotically defective mutants of *Lotus japonicus*. Mol. Plant-Microbe Interact. **19**: 914–923.

Miyahara, A., Richens, J., Starker, C., Morieri, G., Smith, L., Long, S., Downie, J.A., and Oldroyd, G.E.D. (2010). Conservation in function of a SCAR/WAVE component during infection thread and root hair growth in *Medicago truncatula*. Mol. plant-microbe Interact. **23**: 1553–1562.

Miyashima, S. and Nakajima, K. (2011). The root endodermis: a hub of developmental signals and nutrient flow. Plant Signal. Behav. **6**: 1954–1958.

Møller, I.M. (2001). Plant mitochondria and oxidative stress: electron transport, NADPH turnover, and metabolism of reactive oxygen species. Annu. Rev. Plant Biol. **52**: 561–591.

Montes, R.A.C., Ranocha, P., Martinez, Y., Minic, Z., Jouanin, L., Marquis, M., Saulnier, L., Fulton, L.M., Cobbett, C.S., and Bitton, F. (2008). Cell Wall

Modifications in *Arabidopsis* Plants with Altered a-L-Arabinofuranosidase Activity [C][W]. Plant Physiol. **147**: 63–77.

Montiel, J., Arthikala, M.-K., Cárdenas, L., and Quinto, C. (2016). Legume NADPH oxidases have crucial roles at different stages of nodulation. Int. J. Mol. Sci. **17**: 680.

Montiel, J., Reid, D., Grønbæk, T.H., Benfeldt, C.M., James, E.K., Ott, T., Ditengou, F.A., Nadzieja, M., Kelly, S., and Stougaard, J. (2021). Distinct signaling routes mediates intercellular and intracellular rhizobial infection in *Lotus japonicus*. Plant Physiol. **185**: 1131–1147.

Montiel, J., Reid, D.E., Grønbæk, T.H., Benfeldt, C.M., James, E.K., Ott, T., Ditengou, F.A., Nadzieja, M., Kelly, S., and Stougaard, J. (2020). Distinct signalling routes mediates intercellular and intracellular rhizobial infection in *Lotus japonicus*. bioRxiv.

Moran, J.F., Iturbe-Ormaetxe, I., Matamoros, M.A., Rubio, M.C., Clemente, M.R., Brewin, N.J., and Becana, M. (2000). Glutathione and homoglutathione synthetases of legume nodules. Cloning, expression, and subcellular localization. Plant Physiol. **124**: 1381–1392.

Morgante, C., Castro, S., and Fabra, A. (2007). Role of rhizobial EPS in the evasion of peanut defense response during the crack-entry infection process. Soil Biol. Biochem. **39**: 1222–1225.

Mortier, V., Wasson, A., Jaworek, P., De Keyser, A., Decroos, M., Holsters, M., Tarkowski, P., Mathesius, U., and Goormachtig, S. (2014). Role of LONELY GUY genes in indeterminate nodulation on *Medicago truncatula*. New Phytol. **202**: 582–593.

Murray, J.D. (2011). Invasion by invitation: rhizobial infection in legumes. Mol. Plant-Microbe Interact. **24**: 631–639.

Murray, J.D., Karas, B.J., Sato, S., Tabata, S., Amyot, L., and Szczyglowski, K. (2007). A cytokinin perception mutant colonized by Rhizobium in the absence of nodule organogenesis. Science (80-.). **315**: 101–104.

Murray, J.D., Muni, R.R.D., Torres-Jerez, I., Tang, Y., Allen, S., Andriankaja, M., Li, G., Laxmi, A., Cheng, X., and Wen, J. (2011). Vapyrin, a gene essential for intracellular progression of arbuscular mycorrhizal symbiosis, is also essential for infection by rhizobia in the nodule symbiosis of *Medicago truncatula*. Plant J. **65**: 244–252.

Mus, F., Crook, M.B., Garcia, K., Costas, A.G., Geddes, B.A., Kouri, E.D., Paramasivan, P., Ryu, M.-H., Oldroyd, G.E.D., and Poole, P.S. (2016). Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. Appl. Environ. Microbiol. **82**: 3698–3710.

Naseer, S., Lee, Y., Lapierre, C., Franke, R., Nawrath, C., and Geldner, N. (2012).

Casparian strip diffusion barrier in *Arabidopsis* is made of a lignin polymer without suberin. Proc. Natl. Acad. Sci. **109**: 10101–10106.

Nei, M. and Rooney, A.P. (2005). Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. **39**: 121–152.

Noiraud, N., Maurousset, L., and Lemoine, R. (2001). Transport of polyols in higher plants. Plant Physiol. Biochem. **39**: 717–728.

Noisangiam, R., Teamtisong, K., Tittabutr, P., Boonkerd, N., Toshiki, U., Minamisawa, K., and Teaumroong, N. (2012). Genetic diversity, symbiotic evolution, and proposed infection process of *Bradyrhizobium* strains isolated from root nodules of Aeschynomene americana L. in Thailand. Appl. Environ. Microbiol. **78**: 6236–6250.

O'Brian, M.R. (1996). Heme synthesis in the rhizobium-legume symbiosis: a palette for bacterial and eukaryotic pigments. J. Bacteriol. **178**: 2471–2478.

Oke, V. and Long, S.R. (1999). Bacteroid formation in the Rhizobium–legume symbiosis. Curr. Opin. Microbiol. **2**: 641–646.

Okushima, Y., Fukaki, H., Onoda, M., Theologis, A., and Tasaka, M. (2007). ARF7 and ARF19 regulate lateral root formation via direct activation of LBD/ASL genes in *Arabidopsis*. Plant Cell **19**: 118–130.

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

Oldroyd, G.E.D. and Dixon, R. (2014). Biotechnological solutions to the nitrogen problem. Curr. Opin. Biotechnol. **26**: 19–24.

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

Ott, T., van Dongen, J.T., Gu, C., Krusell, L., Desbrosses, G., Vigeolas, H., Bock, V., Czechowski, T., Geigenberger, P., and Udvardi, M.K. (2005). Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. Curr. Biol. **15**: 531–535.

Pan, X., Li, H., Zeng, T., Li, Z., Chen, L., Huang, T., and Cai, Y.-D. (2021). Identification of protein subcellular localization with network and functional embeddings. Front. Genet.: 1800.

Pan, Z., Liu, M., Zhao, H., Tan, Z., Liang, K., Sun, Q., Gong, D., He, H., Zhou, W., and Qiu, F. (2020). *Zm*SRL5 is involved in drought tolerance by maintaining cuticular wax structure in maize. J. Integr. Plant Biol. **62**: 1895–1909.

Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., and Hellmann, I. (2018). zUMIs-a fast and flexible pipeline to process RNA sequencing data with UMIs. Gigascience **7**: giy059.

Parniske, M. (2018). Uptake of bacteria into living plant cells, the unifying and distinct feature of the nitrogen-fixing root nodule symbiosis. Curr. Opin. Plant Biol. **44**: 164–174.

Peleg-Grossman, S., Volpin, H., and Levine, A. (2007). Root hair curling and Rhizobium infection in *Medicago truncatula* are mediated by phosphatidylinositide-regulated endocytosis and reactive oxygen species. J. Exp. Bot. **58**: 1637–1649.

Penmetsa, R.V. and Cook, D.R. (2000). Production and characterization of diverse developmental mutants of *Medicago truncatula*. Plant Physiol. **123**: 1387–1398.

Perret, X., Staehelin, C., and Broughton, W.J. (2000). Molecular basis of symbiotic promiscuity. Microbiol. Mol. Biol. Rev. **64**: 180–201.

Perry, J.A. (2003). A TILLING Reverse Genetics Tool and a Web-Accessible Collection of Mutants of the Legume *Lotus japonicus*. Plant Physiol. **131**: 866–871.

Pingali, P.L. (2012). Green revolution: impacts, limits, and the path ahead. Proc. Natl. Acad. Sci. **109**: 12302–12308.

Pitarch, B., Ranea, J.A.G., and Pazos, F. (2021). Protein residues determining interaction specificity in paralogous families. Bioinformatics **37**: 1076–1082.

Poehlman, W.L., Schnabel, E.L., Chavan, S.A., Frugoli, J.A., and Feltus, F.A. (2019). Identifying temporally regulated root nodulation biomarkers using time series gene co-expression network analysis. Front. Plant Sci.: 1409.

Pollard, M., Beisson, F., Li, Y., and Ohlrogge, J.B. (2008). Building lipid barriers: biosynthesis of cutin and suberin. Trends Plant Sci. **13**: 236–246.

Preisig, O., Zufferey, R., Thöny-Meyer, L., Appleby, C.A., and Hennecke, H. (1996). A high-affinity cbb3-type cytochrome oxidase terminates the symbiosis-specific respiratory chain of *Bradyrhizobium japonicum*. J. Bacteriol. **178**: 1532–1538.

Puppo, A., Groten, K., Bastian, F., Carzaniga, R., Soussi, M., Lucas, M.M., De Felipe, M.R., Harrison, J., Vanacker, H., and Foyer, C.H. (2005). Legume nodule senescence: roles for redox and hormone signalling in the orchestration of the natural aging process. New Phytol. **165**: 683–701.

Qiu, L., Lin, J., Xu, J., Sato, S., Parniske, M., Wang, T.L., Downie, J.A., and Xie, F. (2015). SCARN a novel class of SCAR protein that is required for root-hair infection during legume nodulation. PLoS Genet. **11**: e1005623.

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

Rae, A.L., Bonfante-Fasolo, P., and Brewin, N.J. (1992). Structure and growth of infection threads in the legume symbiosis with *Rhizobium leguminosarum*. Plant J. **2**: 385–395.

Rains, M.K., Gardiyehewa de Silva, N.D., and Molina, I. (2018). Reconstructing the suberin pathway in poplar by chemical and transcriptomic analysis of bark tissues. Tree Physiol. **38**: 340–361.

Ramu, S.K., Peng, H.-M., and Cook, D.R. (2002). Nod factor induction of reactive oxygen species production is correlated with expression of the early nodulin gene rip1 in *Medicago truncatula*. Mol. plant-microbe Interact. **15**: 522–528.

RDevelopment, C. (2012). TEAM 2009: R: A language and environment for statistical computing. Vienna, Austria. Internet http//www. R-project. org.

Reid, D., Nadzieja, M., Novák, O., Heckmann, A.B., Sandal, N., and Stougaard, J. (2017). Cytokinin biosynthesis promotes cortical cell responses during nodule development. Plant Physiol. **175**: 361–375.

Reid, D.E., Heckmann, A.B., Novák, O., Kelly, S., and Stougaard, J. (2016). CYTOKININ OXIDASE/DEHYDROGENASE3 maintains cytokinin homeostasis during root and nodule development in *Lotus japonicus*. Plant Physiol. **170**: 1060–1074.

Ried, M.K., Antolín-Llovera, M., and Parniske, M. (2014). Spontaneous symbiotic reprogramming of plant roots triggered by receptor-like kinases. Elife **3**: e03891.

Robertson, J.G., Lyttleton, P., and Tapper, B.A. (1984). The role of peribacteroid membrane in legume root nodules. In Advances in nitrogen fixation research (Springer), pp. 475–481.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F.S., Lambin, E.F., Lenton, T.M., Scheffer, M., Folke, C., and Schellnhuber, H.J. (2009). A safe operating space for humanity. Nature **461**: 472–475.

Rodpothong, P., Sullivan, J.T., Songsrirote, K., Sumpton, D., Cheung, K.W.-T., Thomas-Oates, J., Radutoiu, S., Stougaard, J., and Ronson, C.W. (2009). Nodulation Gene Mutants of *Mesorhizobium loti* R7A—nodZ and nolL Mutants Have Host-Specific Phenotypes on Lotus spp. Mol. plant-microbe Interact. **22**: 1546–1554.

Rogato, A., D'Apuzzo, E., Barbulova, A., Omrane, S., Stedel, C., Simon-Rosin, U., Katinakis, P., Flemetakis, M., Udvardi, M., and Chiurazzi, M. (2008). Tissue-specific down-regulation of *Lj*AMT1; 1 compromises nodule function and enhances nodulation in *Lotus japonicus*. Plant Mol. Biol. **68**: 585–595.

Rogers, C. and Oldroyd, G.E.D. (2014). Synthetic biology approaches to engineering the nitrogen symbiosis in cereals. J. Exp. Bot. **65**: 1939–1946.

Roppolo, D., Boeckmann, B., Pfister, A., Boutet, E., Rubio, M.C., Dénervaud-Tendon, V., Vermeer, J.E.M., Gheyselinck, J., Xenarios, I., and Geldner, N. (2014). Functional and evolutionary analysis of the CASPARIAN STRIP MEMBRANE DOMAIN PROTEIN family. Plant Physiol. **165**: 1709–1722.

Roppolo, D., De Rybel, B., Tendon, V.D., Pfister, A., Alassimone, J., Vermeer, J.E.M., Yamazaki, M., Stierhof, Y.-D., Beeckman, T., and Geldner, N. (2011). A novel protein family mediates Casparian strip formation in the endodermis. Nature **473**: 380–383.

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

Roux, B., Rodde, N., Jardinaud, M., Timmers, T., Sauviac, L., Cottret, L., Carrère, S., Sallet, E., Courcelle, E., and Moreau, S. (2014). An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. Plant J. **77**: 817–837.

Rowland, O. and Domergue, F. (2012). Plant fatty acyl reductases: enzymes generating fatty alcohols for protective layers with potential for industrial applications. Plant Sci. **193**: 28–38.

Rowland, O., Zheng, H., Hepworth, S.R., Lam, P., Jetter, R., and Kunst, L. (2006). CER4 encodes an alcohol-forming fatty acyl-coenzyme A reductase involved in cuticular wax production in Arabidopsis. Plant Physiol. **142**: 866–877.

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

Rozewicki, J., Li, S., Amada, K.M., Standley, D.M., and Katoh, K. (2019). MAFFT-DASH: integrated protein sequence and structural alignment. Nucleic Acids Res. **47**: W5–W10.

Rubio, M.C., Becana, M., Sato, S., James, E.K., Tabata, S., and Spaink, H.P. (2007). Characterization of genomic clones and expression analysis of the three types of superoxide dismutases during nodule development in *Lotus japonicus*. Mol. plant-microbe Interact. **20**: 262–275.

Rubio, M.C., James, E.K., Clemente, M.R., Bucciarelli, B., Fedorova, M., Vance, C.P., and Becana, M. (2004). Localization of superoxide dismutases and hydrogen peroxide in legume root nodules. Mol. Plant-Microbe Interact. **17**: 1294–1305.

Rutten, P.J. and Poole, P.S. (2019). Oxygen regulatory mechanisms of nitrogen fixation in rhizobia. In Advances in microbial physiology (Elsevier), pp. 325–389.

Saha, B., Saha, S., Das, A., Bhattacharyya, P.K., Basak, N., Sinha, A.K., and Poddar, P. (2017). Biological nitrogen fixation for sustainable agriculture. In Agriculturally important microbes for sustainable agriculture (Springer), pp. 81–128.

Saha, S., Paul, A., Herring, L., Dutta, A., Bhattacharya, A., Samaddar, S., Goshe,

M.B., and DasGupta, M. (2016). Gatekeeper tyrosine phosphorylation of SYMRK is essential for synchronizing the epidermal and cortical responses in root nodule symbiosis. Plant Physiol. **171**: 71–81.

Saito, K., Yoshikawa, M., Yano, K., Miwa, H., Uchida, H., Asamizu, E., Sato, S., Tabata, S., Imaizumi-Anraku, H., and Umehara, Y. (2007). NUCLEOPORIN85 is required for calcium spiking, fungal and bacterial symbioses, and seed production in *Lotus japonicus*. Plant Cell **19**: 610–624.

Santana, M.A., Pihakaski-Maunsbach, K., Sandal, N., Marcker, K.A., and Smith, A.G. (1998). Evidence that the plant host synthesizes the heme moiety of leghemoglobin in root nodules. Plant Physiol. **116**: 1259–1269.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Nakao, M., Sasamoto, S., Watanabe, A., Ono, A., and Kawashima, K. (2008). Genome structure of the legume, *Lotus japonicus*. DNA Res. **15**: 227–239.

Schauser, L., Handberg, K., Sandal, N., Stiller, J., Thykjær, T., Pajuelo, E., Nielsen, A., and Stougaard, J. (1998). Symbiotic mutants deficient in nodule establishment identified after T-DNA transformation of *Lotus japonicus*. Mol. Gen. Genet. **259**: 414–423.

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

Schiessl, K., Lilley, J.L.S., Lee, T., Tamvakis, I., Kohlen, W., Bailey, P.C., Thomas, A., Luptak, J., Ramakrishnan, K., and Carpenter, M.D. (2019). NODULE INCEPTION recruits the lateral root developmental program for symbiotic nodule organogenesis in *Medicago truncatula*. Curr. Biol. **29**: 3657–3668.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., and Schmid, B. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods **9**: 676–682.

Schneider, S., Schintlmeister, A., Becana, M., Wagner, M., Woebken, D., and Wienkoop, S. (2019). Sulfate is transported at significant rates through the symbiosome membrane and is crucial for nitrogenase biosynthesis. Plant. Cell Environ. **42**: 1180–1189.

Schreiber, L., Hartmann, K., Skrabs, M., and Zeier, J. (1999). Apoplastic barriers in roots: chemical composition of endodermal and hypodermal cell walls. J. Exp. Bot. **50**: 1267–1280.

Senovilla, M., Castro-Rodríguez, R., Abreu, I., Escudero, V., Kryvoruchko, I., Udvardi, M.K., Imperial, J., and González-Guerrero, M. (2018). *Medicago truncatula* copper transporter 1 (Mt COPT 1) delivers copper for symbiotic nitrogen fixation. New Phytol. **218**: 696–709.

109

Serra, O. and Geldner, N. (2022). The making of suberin. New Phytol.

Sexauer, M., Shen, D., Schön, M., Andersen, T.G., and Markmann, K. (2021). Visualizing polymeric components that define distinct root barriers across plant lineages. Development **148**: dev199820.

Sharma, V., Bhattacharyya, S., Kumar, R., Kumar, A., Ibañez, F., Wang, J., Guo, B., Sudini, H.K., Gopalakrishnan, S., and DasGupta, M. (2020). Molecular basis of root nodule symbiosis between *Bradyrhizobium* and 'crack-entry' legume groundnut (*Arachis hypogaea* L.). Plants **9**: 276.

Shimoda, Y., Shimoda-Sasakura, F., Kucho, K., Kanamori, N., Nagata, M., Suzuki, A., Abe, M., Higashi, S., and Uchiumi, T. (2009). Overexpression of class 1 plant hemoglobin genes enhances symbiotic nitrogen fixation activity between *Mesorhizobium loti* and *Lotus japonicus*. Plant J. **57**: 254–263.

Shrestha, A., Zhong, S., Therrien, J., Huebert, T., Sato, S., Mun, T., Andersen, S.U., Stougaard, J., Lepage, A., and Niebel, A. (2021). *Lotus japonicus* Nuclear Factor YA1, a nodule emergence stage-specific regulator of auxin signalling. New Phytol. **229**: 1535–1552.

Sieberer, B.J., Chabaud, M., Timmers, A.C., Monin, A., Fournier, J., and Barker, D.G. (2009). A nuclear-targeted cameleon demonstrates intranuclear Ca2+ spiking in *Medicago truncatula* root hairs in response to rhizobial nodulation factors. Plant Physiol. **151**: 1197–1206.

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.

Singh, S. and Varma, A. (2017). Structure, function, and estimation of leghemoglobin. In Rhizobium Biology and Biotechnology (Springer), pp. 309–330.

Smith, V.H. and Schindler, D.W. (2009). Eutrophication science: where do we go from here? Trends Ecol. Evol. **24**: 201–207.

Sogawa, A., Yamazaki, A., Yamasaki, H., Komi, M., Manabe, T., Tajima, S., Hayashi, M., and Nomura, M. (2019). SNARE proteins *Lj*VAMP72a and *Lj*VAMP72b are required for root symbiosis and root hair formation in *Lotus japonicus*. Front. Plant Sci. **9**: 1992.

Sol, S., Valkov, V.T., Rogato, A., Noguero, M., Gargiulo, L., Mele, G., Lacombe, B., and Chiurazzi, M. (2019). Disruption of the *Lotus japonicus* transporter LjNPF2. 9 increases shoot biomass and nitrate content without affecting symbiotic performances. BMC Plant Biol. **19**: 1–14.

Soyano, T., Kouchi, H., Hirota, A., and Hayashi, M. (2013). Nodule inception directly targets NF-Y subunit genes to regulate essential processes of root nodule

development in *Lotus japonicus*. PLoS Genet. **9**: e1003352.

Soyano, T., Liu, M., Kawaguchi, M., and Hayashi, M. (2021). Leguminous nodule symbiosis involves recruitment of factors contributing to lateral root development. Curr. Opin. Plant Biol. **59**: 102000.

Soyano, T., Shimoda, Y., Kawaguchi, M., and Hayashi, M. (2019). A shared gene drives lateral root development and root nodule symbiosis pathways in *Lotus*. Science (80-.). **366**: 1021–1023.

Spiertz, J.H.J. (2009). Nitrogen, sustainable agriculture and food security: a review. Sustain. Agric.: 635–651.

SPRATT, E.R. (1919). A comparative account of the root-nodules of the Leguminosae. Ann. Bot. **33**: 189–199.

Sprent, J.I. (2008). Evolution and diversity of legume symbiosis. In Nitrogen-fixing leguminous symbioses (Springer), pp. 1–21.

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

Sprent, J.I., Ardley, J., and James, E.K. (2017). Biogeography of nodulated legumes and their nitrogen-fixing symbionts. New Phytol. **215**: 40–56.

Sprent, J.I. and Platzmann, J. (2001). Nodulation in legumes (Royal Botanic Gardens Kew).

Stagnari, F., Maggio, A., Galieni, A., and Pisante, M. (2017). Multiple benefits of legumes for agriculture sustainability: an overview. Chem. Biol. Technol. Agric. **4**: 1–13.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30**: 1312–1313.

Stougaard, J., Abildsten, D., and Marcker, K.A. (1987). The *Agrobacterium* rhizogenes pRi TL-DNA segment as a gene vector system for transformation of plants. Mol. Gen. Genet. MGG **207**: 251–255.

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

Sugiyama, A., Saida, Y., Yoshimizu, M., Takanashi, K., Sosso, D., Frommer, W.B., and Yazaki, K. (2017). Molecular characterization of *Lj*SWEET3, a sugar transporter in nodules of *Lotus japonicus*. Plant Cell Physiol. **58**: 298–306.

Tadege, M., Wen, J., He, J., Tu, H., Kwak, Y., Eschstruth, A., Cayrel, A., Endre, G., Zhao, P.X., and Chabaud, M. (2008). Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume *Medicago truncatula*. Plant J. **54**: 335–347.

Takanashi, K., Sasaki, T., Kan, T., Saida, Y., Sugiyama, A., Yamamoto, Y., and Yazaki, K. (2016). A dicarboxylate transporter, *Lj*ALMT4, mainly expressed in nodules of *Lotus japonicus*. Mol. Plant-Microbe Interact. **29**: 584–592.

Tansengco, M.L., Hayashi, M., Kawaguchi, M., Imaizumi-Anraku, H., and Murooka, Y. (2003). crinkle, a novel symbiotic mutant that affects the infection thread growth and alters the root hair, trichome, and seed development in *Lotus japonicus*. Plant Physiol. **131**: 1054–1063.

Taylor, E.R., Nie, X.Z., MacGregor, A.W., and Hill, R.D. (1994). A cereal haemoglobin gene is expressed in seed and root tissues under anaerobic conditions. Plant Mol. Biol. **24**: 853–862.

Team, R.C. (2013). R: A language and environment for statistical computing.

Tejada-Jiménez, M., Castro-Rodríguez, R., Kryvoruchko, I., Lucas, M.M., Udvardi, M., Imperial, J., and González-Guerrero, M. (2015). *Medicago truncatula* natural resistance-associated macrophage protein1 is required for iron uptake by rhizobia-infected nodule cells. Plant Physiol. **168**: 258–272.

Timmers, A.C.J., Soupène, E., Auriac, M.-C., de Billy, F., Vasse, J., Boistard, P., and Truchet, G. (2000). Saprophytic intracellular rhizobia in alfalfa nodules. Mol. Plant-Microbe Interact. **13**: 1204–1213.

Tirichine, L., Imaizumi-Anraku, H., Yoshida, S., Murakami, Y., Madsen, L.H., Miwa, H., Nakagawa, T., Sandal, N., Albrektsen, A.S., and Kawaguchi, M. (2006). Deregulation of a Ca 2+/calmodulin-dependent kinase leads to spontaneous nodule development. Nature **441**: 1153.

Tirichine, L., James, E.K., Sandal, N., and Stougaard, J. (2007). Spontaneous Root-Nodule Formation in the Model Legume *Lotus japonicus* : A Novel Class of Mutants Nodulates in the Absence of Rhizobia . Mol. Plant-Microbe Interact. **19**: 373–382.

Tovar-Méndez, A., Matamoros, M.A., Bustos-Sanmamed, P., Dietz, K.-J., Cejudo, F.J., Rouhier, N., Sato, S., Tabata, S., and Becana, M. (2011). Peroxiredoxins and NADPH-dependent thioredoxin systems in the model legume *Lotus japonicus*. Plant Physiol. **156**: 1535–1547.

Trevaskis, B., Watts, R.A., Andersson, C.R., Llewellyn, D.J., Hargrove, M.S., Olson, J.S., Dennis, E.S., and Peacock, W.J. (1997). Two hemoglobin genes in *Arabidopsis thaliana*: the evolutionary origins of leghemoglobins. Proc. Natl. Acad. Sci. **94**: 12230–12234.

Tsyganova, A. V, Kitaeva, A.B., and Tsyganov, V.E. (2018). Cell differentiation in nitrogen-fixing nodules hosting symbiosomes. Funct. Plant Biol. **45**: 47–57.

Udvardi, M. and Poole, P.S. (2013). Transport and metabolism in legume-rhizobia symbioses. Annu. Rev. Plant Biol. **64**: 781–805.

Uheda, E., Daimon, H., and Yoshizako, F. (2001). Colonization and invasion of peanut (*Arachis hypogaea* L.) roots by gusA-marked *Bradyrhizobium* sp. Can. J. Bot. **79**: 733–738.

Ursache, R., Andersen, T.G., Marhavý, P., and Geldner, N. (2018). A protocol for combining fluorescent proteins with histological stains for diverse cell wall components. Plant J. **93**: 399–412.

Valkov, V.T., Rogato, A., Alves, L.M., Sol, S., Noguero, M., Léran, S., Lacombe, B., and Chiurazzi, M. (2017). The nitrate transporter family protein *Lj*NPF8. 6 controls the N-fixing nodule activity. Plant Physiol. **175**: 1269–1282.

Valkov, V.T., Sol, S., Rogato, A., and Chiurazzi, M. (2020). The functional characterization of *Lj*NRT2. 4 indicates a novel, positive role of nitrate for an efficient nodule N_2-fixation activity. New Phytol. **228**: 682–696.

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.

Venado, R.E., Liang, J., and Marín, M. (2019). Rhizobia infection, a journey to the inside of plant cells.

Venkateshwaran, M., Jayaraman, D., Chabaud, M., Genre, A., Balloon, A.J., Maeda, J., Forshey, K., den Os, D., Kwiecien, N.W., and Coon, J.J. (2015). A role for the mevalonate pathway in early plant symbiotic signaling. Proc. Natl. Acad. Sci. **112**: 9781–9786.

Verma, D.P.S., Ball, S., Guerin, C., and Wanamaker, L. (1979). Leghemoglobin biosynthesis in soybean root nodules. Characterization of the nascent and released peptides and the relative rate of synthesis of the major leghemoglobins. Biochemistry **18**: 476–483.

Vinardell, J.M., Fedorova, E., Cebolla, A., Kevei, Z., Horvath, G., Kelemen, Z., Tarayre, S., Roudier, F., Mergaert, P., and Kondorosi, A. (2003). Endoreduplication mediated by the anaphase-promoting complex activator CCS52A is required for symbiotic cell differentiation in *Medicago truncatula* nodules. Plant Cell **15**: 2093–2105.

Vittozzi, Y., Nadzieja, M., Rogato, A., Radutoiu, S., Valkov, V.T., and Chiurazzi, M. (2021). The *Lotus japonicus* NPF3. 1 Is a Nodule-Induced Gene That Plays a Positive Role in Nodule Functioning. Front. Plant Sci. **12**: 1110.

Vogt, T. (2010). Phenylpropanoid biosynthesis. Mol. Plant **3**: 2–20.

Wagner, S.C. (2011). Biological nitrogen fixation. Nat. Educ. Knowl. **3**: 15.

Walker, L., Lagunas, B., and Gifford, M.L. (2020). Determinants of host range specificity in legume-rhizobia symbiosis. Front. Microbiol.: 3028.

Wang, L., Rubio, M.C., Xin, X., Zhang, B., Fan, Q., Wang, Q., Ning, G., Becana, M., and Duanmu, D. (2019). CRISPR/Cas9 knockout of leghemoglobin genes in Lotus japonicus uncovers their synergistic roles in symbiotic nitrogen fixation. New Phytol. **224**: 818–832.

Wang, L., Zhou, Y., Li, R., Liang, J., Tian, T., Ji, J., Chen, R., Zhou, Y., Fan, Q., and Ning, G. (2022). Single cell-type transcriptome profiling reveals genes that promote nitrogen fixation in the infected and uninfected cells of legume nodules. Plant Biotechnol. J.

Wang, Q., Huang, Y., Ren, Z., Zhang, X., Ren, J., Su, J., Zhang, C., Tian, J., Yu, Y., and Gao, G.F. (2020a). Transfer cells mediate nitrate uptake to control root nodule symbiosis. Nat. Plants **6**: 800–808.

Wang, X., Zhang, Y., Wang, L., Pan, Z., He, S., Gao, Q., Chen, B., Gong, W., and Du, X. (2020b). Casparian strip membrane domain proteins in *Gossypium arboreum*: genome-wide identification and negative regulation of lateral root growth. BMC Genomics **21**: 1–16.

Wang, Z.C., Burns, A., and Watt, G.D. (1985). Complex formation and oxygen sensitivity of Azotobacter vinelandii nitrogenase and its component proteins. Biochemistry **24**: 214–221.

Warnes, M.G.R., Bolker, B., Bonebakker, L., Gentleman, R., and Huber, W. (2016). Package 'gplots.' Var. R Program. tools plotting data.

Whitehead, L.F. and Day, D.A. (1997). The peribacteroid membrane. Physiol. Plant. **100**: 30–44.

Whiting, M.J. and Dilworth, M.J. (1974). Legume root nodule nitrogenase: Purification, properties, and studies on its genetic control. Biochim. Biophys. Acta (BBA)-Protein Struct. **371**: 337–351.

Wickner, W. and Schekman, R. (2008). Membrane fusion. Nat. Struct. Mol. Biol. **15**: 658–664.

Witty, J., Skøt, L., and Revsbech, N.P. (1987). Direct evidence for changes in the resistance of legume root nodules to O_2 diffusion. J. Exp. Bot. **38**: 1129–1140.

Witty, J.F. and Minchin, F.R. (1990). Oxygen diffusion in the legume root nodule. In Nitrogen Fixation (Springer), pp. 285–292.

Wolfe, C.J., Kohane, I.S., and Butte, A.J. (2005). Systematic survey reveals general applicability of guilt-by-association within gene coexpression networks. BMC Bioinformatics **6**: 1–10.

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

Yadav, V., Molina, I., Ranathunge, K., Castillo, I.Q., Rothstein, S.J., and Reed, J.W. (2014). ABCG transporters are required for suberin and pollen wall extracellular barriers in Arabidopsis. Plant Cell **26**: 3569–3588.

Yang, J., Ding, C., Xu, B., Chen, C., Narsai, R., Whelan, J., Hu, Z., and Zhang, M. (2015). A Casparian strip domain-like gene, CASPL, negatively alters growth and cold tolerance. Sci. Rep. **5**: 1–11.

Yano, K., Shibata, S., Chen, W., Sato, S., Kaneko, T., Jurkiewicz, A., Sandal, N., Banba, M., Imaizumi-Anraku, H., and Kojima, T. (2009). CERBERUS, a novel U-box protein containing WD-40 repeats, is required for formation of the infection thread and nodule development in the legume–*Rhizobium* symbiosis. Plant J. **60**: 168–180.

Yano, K., Yoshida, S., Müller, J., Singh, S., Banba, M., Vickers, K., Markmann, K., White, C., Schuller, B., and Sato, S. (2008). CYCLOPS, a mediator of symbiotic intracellular accommodation. Proc. Natl. Acad. Sci. **105**: 20540–20545.

Yokota, K., Fukai, E., Madsen, L.H., Jurkiewicz, A., Rueda, P., Radutoiu, S., Held, M., Hossain, M.S., Szczyglowski, K., and Morieri, G. (2009). Rearrangement of actin cytoskeleton mediates invasion of *Lotus japonicus* roots by *Mesorhizobium loti*. Plant Cell **21**: 267–284.

Zeier, J., Goll, A., Yokoyama, M., Karahara, I., and Schreiber, L. (1999). Structure and chemical composition of endodermal and rhizodermal/hypodermal walls of several species. Plant. Cell Environ. **22**: 271–279.

Zhang, B. and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. **4**.

Zhong, R., Cui, D., and Ye, Z. (2019). Secondary cell wall biosynthesis. New Phytol. **221**: 1703–1723.

Zhong, R., Richardson, E.A., and Ye, Z.-H. (2007). Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of Arabidopsis. Planta **225**: 1603–1611.

Zhong, R. and Ye, Z.-H. (2015). Secondary cell walls: biosynthesis, patterned deposition and transcriptional regulation. Plant Cell Physiol. **56**: 195–214.

Zhu, H., Li, C., and Gao, C. (2020). Applications of CRISPR–Cas in agriculture and plant biotechnology. Nat. Rev. Mol. Cell Biol. **21**: 661–677.

Supplemental Tables

Supplemental Table 1. Strains used for this work.

Strain	Information	Source
Agrobacterium rhizogenes AR1193	pRi1193 carrying pBR322 in the TL segment, Rf^R, Cm^R	Stougaard et al., 1987
Agrobacterium tumefaciens AGL1	pTiBo542 ΔT, RfR, CbR	Lazo et al., 1991
Escherichia coli TOP10	F$^-$ *mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80*lacZ*ΔM15 Δ *lacX74 recA1 araD139* Δ(*araleu*)*7697 galU galK rpsL endA1 nupG*, Sm^R	Invitrogen
Mesorhizobium loti MAFF 303099 *Ds*Red	*M. loti* MAFF 303099 expressing *Ds*Red, Gm^R	MaekawaYoshikawa et al., 2009
Mesorhizobium loti MAFF 303099-GFP	*M. loti* MAFF 303099 containing the pFAJ-GFP plasmid, Fm^R, Tc^R	Liang et al. 2019
Rhizobium leguminosarum Norway-GFP	*R. leguminosarum* Norway containing the pHC60 plasmid, IncP, Sm^R, Tc^R	Liang et al., 2019

Supplemental Table 2. List of *Lotus japonicus* accessions used in this work.

Accession	Seed bag Number	Tribal name	Location
Gifu B-129	110902	Gifu	Gifu, Japan
Gifu B-129	110903	Gifu	Gifu, Japan
Gifu B-129	110927	Gifu	Gifu, Japan
MG-7	20016498	Not available	Not available
MG-9*	20016173	Oga	Oga Peninsula
MG-20	110934	Miyakojima	Miyakojima Island
MG-25	20016350	Ambo	Ambo Haruta
MG-28	20015841	Nagasakibana	Nagasakibana
MG-29	20015834	Tassobe	Tassobe
MG-50	20016643	Tohaku	Kaseichi River
MG-52	20016032	Hikawa	Hii River
MG-55	20016347	Not available	Not available
MG-60	20015856	Makurazaki	Sea side Cliff
MG-64	20015656	Miyadai1	Univ. of Miyazaki Coop
MG-70*	20015696	Inami	Minamihata
MG-78	20016099	Kuzuryuko	The mouth of Lake Kuzuryu
MG-79*	20016172	Mikuni	The mouth of Kuzuryuu River
MG-81	20015968	Omi	The mouth of Hime River
MG-83	20016181	Sugahira	Sugahira Mountain trail
MG-85	20016496	Ishikawa	Abukuma River
MG-86	20016467	Naruko	Arao River
MG-102	20016362	Not available	Not available
MG-103	20016238	Not available	Not available
MG-104	20015838	Not available	Not available
MG-106	20015676	Not available	Not available
MG-110	20016735	Takamori1	Murayama
MG-111	20016341	Choyo	Choyo
MG-112*	20016564	Namino	Namino
MG-113	20016566	Takamori2	Kamishikimi
MG-115*	20016112	Ibigawa	Ibi River
MG-119	20016356	Shinanogawa	Shinanogawa
MG-123	20016143	Chino	Tateshina
MG-128	20016026	Hokigawa	Hoki River
MG-129	20016576	Nishisenboku	Kariwano
MG-133	20015819	Not available	Not available
MG-135	20016324	Not available	Not available
MG-136*	20016500	Not available	Not available
MG-137	20015988	Not available	Not available
MG-139	20015897	Not available	Not available
MG-140	20015398	Not available	Not available
MG-143	20016151	Not available	Not available
MG-142	20016101	Not available	Not available
MG-144	20016126	Not available	Not available
MG-146	20016077	Not available	Not available

Lotus accessions used for prime-seq.

Supplemental Table 3. List of PCR, sequencing and RT-qPCR primers used in this work. Oligos for CRISPR/Cas12a were included. BpiI, BsaI, and Esp3I recognition sites are underlined.

Primer	Sequence (5' – 3')
LORE1	
far3.2-1-F	TCAATTCATGTCTCAAGGACGGAACCA
far3.2-1-R	CCTGCCATCAAGAAACAAATGCAGA
far3.2-2-F	TTTTATCTCGGGGCGACTGGCTGC
far3.2-2-R	CCTGCCATCAAGAAACAAATGCAGA
LORE1 5'LTR rv[1]	CCATGGCGGTTCCGTGAATCTTAGG
gRNA for CRISPR/Cas12a	
gRNA-I-CASPL4.1	CCTCATCCTTGTAGCCAGTA
gRNA-II-CASPL4.1	TGTCCTGATATAATGGCTGA
gRNA-II-CASPL4.2	TTGTGGTTGAGAGTAATTGC
gRNA-I-CASPL4.3	TGTACACCTTCACCAACCCA
gRNA-II-CASPL4.3	CCTACCTCAACACTAACAGG
gRNA-I-CYP86A1	ACACCGCCACTGAAACCACC
gRNA-II-CYP86A1	CTCATGAACCGCGATAAAAG
gRNA-I-RBOHB	AGGGTCCTGACAAAAAACGT
gRNA-II-RBOHB	ATCACAATGCTTCAATCCCT
CRISPR/Cas12a sequencing primers	
I-CASPL4.1-F[2]	AT<u>GAAGAC</u>ATTACG<u>GGTCTC</u>ACACCATGTTGGATTCTTCTAACAACTCA
I-CASPL4.1-R[2]	AT<u>GAAGAC</u>ATGAGACTTCCACGGCTACTAGA
II-CASPL4.1-F[3]	AT<u>GAAGAC</u>ATATCAGCGATTCCAATAACAGA
II-CASPL4.1-R[3]	AT<u>GAAGAC</u>ATCAGA<u>GGTCTC</u>ACCTTGATACAAGTTTGAGCTGATAGTTT
II-CASPL4.2-F	GCAGTTTCAAAGTCAGA
II-CASPL4.2-R	CGAAATCTTTCCAACCGC
I-CASPL4.3-F[4]	AT<u>GAAGAC</u>ATTACG<u>GGTCTC</u>ACACCATGAGTAATAATGATCAAAG
I-CASPL4.3-R[4]	<u>GAAGAC</u>TAAGAGATTAATCACTATTAACAA
II-CASPL4.3-F	CACACATACACACATTCCC
II-CASPL4.3-R	CTCCACCAGCTATTCCTT
I-II-CYP86A1-F	AAAGCCGCAAAGGAGAAA
I-II-CYP86A1-R	GCCAGAAAAACCAGCTCA
I-RBOHB-F	TGTGACAAGCTCAAGGTC
I-RBOHB-R	CAGGTCGTAAAAAGCAAGAA
II-RBOHB-F	AAGGAAGGAGTGATTGAG
II-RBOHB-R	TCAGGGTGTTTGAGAGAA
Promoters	
promLjCASPL4.1 F	AT<u>CGTCTC</u>AGCGGCGTAGGCGCAACATGAAA
promLjCASPL4.1 R	AT<u>CGTCTC</u>ACAGACTCTATCTGCTTTTTCTTCC
promLjCASPL4.3 F	AT<u>CGTCTC</u>AGCGGGCAGCGTACTTGGTCTTA
promLjCASPL4.3 R	AT<u>CGTCTC</u>ACAGAATTCAAATTTTTTGGAACTAG
promLjFAR3.1 F	AT<u>CGTCTC</u>TGCGGAGAGCAGTAAAAGGATGAAG
promLjFAR3.1 R	AT<u>CGTCTC</u>TCAGATGGAGAGAAAATTGAATTAT
promLjFAR3.2 F	AT<u>CGTCTC</u>AGCGGCAGACTGCCACATAGGAT
promLjFAR3.2 R	AT<u>CGTCTC</u>ACAGATGTGGAAAAAATAAAGCAA
promLjNACD F	AT<u>CGTCTC</u>AGCGGTTCTGGAATTGGTCAGGG
promLjNACD R	AT<u>CGTCTC</u>ACAGAGGGGAGACAAACTACTCTT
promLjRBOHB F	AT<u>CGTCTC</u>ACTTTCATAAAAAGGGGGACACAG
promLjRBOHB R	AT<u>CGTCTC</u>ACAGATTTGCAAGCCTTTAGTAG
Sequencing[5]	
GGP1 M13 F	TGTAAAACGACGGCCAGT
GGP1 M13 FR	GGAAACAGCTATGACCAT
GGP5 35Sterm R	GCTCAACACATGAGCGAAACC
GGP6 35Sterm F	GGTTTCGCTCATGTGTTGAGC
GGP7 Nosterm R	CCCATCTCATAAATAACGTCATGC
GGP8 Nosterm F	TTGAATCCTGTTGCCGGTCTTG

GGP12 LjUbq Pro F	CGATGGCTTGCGATGTAGATCT
GGP19 HSPterm F	CTGCAGCATATAACTACTGTATG
GGP20 HSPterm R	CATACAGTAGTTATATGCTGCAG
RT-qPCR	
qPCR-CASPL4.1-F	GAAAATGCTTCAGCCAAG
qPCR-CASPL4.1-R	TCCTCCTTCTCTTATGGT
qPCR-CASPL4.11-F	ACACAAGTAGAAGAAGGCA
qPCR-CASPL4.11-R	TCAGGATCTGTCACCATT
qPCR-CASPL4.2-F	AGTAGTGGTGCTCCTAAG
qPCR-CASPL4.2-R	TGCCATGGTGATGAAGGA
qPCR-CASPL4.3-F	TCAAAGCCTAGCTCAAGC
qPCR-CASPL4.3-R	AGAAACCCGATACACTGC
qPCR-CASPL4.9-F	AACACAGCCCTCATCAAA
qPCR-CASPL4.9-R	AGGCAAACGAGCTGAAAA
qPCR-CASPL8-F	GCTCACAACGGCAATCAA
qPCR-CASPL8-R	CAACAACACAAAGACAACCA
qPCR-COMT-F	TACCAACCCAACACCACC
qPCR-COMT-R	ACCCCATACACTCTCACC
qPCR-CYP86A1-F	CTCTTCTTCACCATAGCA
qPCR-CYP86A1-R	GCCCACGAAAGGATACAC
qPCR-FAR3.1-F	AGGGCTGATAGTAGAGGAT
qPCR-FAR3.1-R	ATTTTGTGGCTCCCTCTT
qPCR-FAR3.2-F	AATGTAGAGAGGTTGCGA
qPCR-FAR3.2-R	ACTTCACAATACCAGGGA
qPCR-NAC-F	ACTTCCAACCATACAAGG
qPCR-NAC-R	ATAGAGGAAATGAGCAGCA
qPCR-RBOHB-F	TGTGACAAGCTCAAGGTC
qPCR-RBOHB-R	AGTTTTCCCTCAATTGCC
Genes	
LjCASPL4.1 Fg1 F	ATGAAGACATTACGGGTCTCACACCATGTTGGATTCTTCTAACAACTCA
LjCASPL4.1 Fg1 R	ATGAAGACATGAGACTTCCACGGCTACTAGA
LjCASPL4.1 Fg2 F	ATGAAGACATTCTCAAGTTTTTTCCTTATGCAG
LjCASPL4.1 Fg2 R	ATGAAGACATTGATGACGCTGATGATATCAAGAG
LjCASPL4.1 Fg3 F	ATGAAGACATATCAGCGATTCCAATAACAGA
LjCASPL4.1 Fg3 R	ATGAAGACATCAGAGGTCTCACCTTGATACAAGTTTGAGCTGATAGTTT
LjCASPL4.2 Fg1 F	ATGAAGACATTACGGGTCTCACACCATGGCAGTTTCAAAGTCAGA
LjCASPL4.2 Fg1 R	ATGAAGACATGTCCTCTACACTCCTCATCC
LjCASPL4.2 Fg2 F	ATGAAGACATGGACAATATATTCACAGATTCTTTAGC
LjCASPL4.2 Fg2 R	ATGAAGACATCAGAGGTCTCACCTTTTAGCTATAAATTTGAG
LjCASPL4.3 Fg1 F	ATGAAGACATTACGGGTCTCACACCATGAGTAATAATGATCAAAG
LjCASPL4.3 Fg1 R	ATGAAGACTAAGAGATTAATCACTATTAACAA
LjCASPL4.3 Fg2 F	ATGAAGACATCTCTTCAGTTTTGAAGTTAA
LjCASPL4.3 Fg2 R	ATGAAGACATCAGAGGTCTCACCTTGTTGAACCTAAGCATTAAAC

[1] Małolepszy et al., 2016
[2] same forward and reverse primers for fragment 1 gCASPL4.1
[3] same forward and reverse primers for fragment 3 gCASPL4.1
[4] same forward and reverse primers for fragment 1 gCASPL4.3
[5] Binder et al., 2014

Supplemental Table 4. List of FAR sequences used in phylogenetic analyses.

Subfamily	Nomenclature	Species	UniProt or gene ID or accession number
Outgroup	EgFAR		A0A6G0XCH3
1	OsFAR1.2	*Oryza sativa subsp. japonicus*	B8B671
	OsFAR1.1	*Oryza sativa subsp. japonicus*	B7F9S3
	OsFAR1.2	*Oryza sativa subsp. japonicus*	A0A0N7KNG3
	OsFAR1.3	*Oryza sativa subsp. japonicus*	Q7XS02
	OsFAR1.4	*Oryza sativa subsp. japonicus*	Q7XRZ6
	OsFAR1.5	*Oryza sativa subsp. japonicus*	Q6ZJ06
	OsFAR2	*Oryza sativa subsp. japonicus*	Q8S7T9
	ZmFAR1.1	*Zea mays*	A0A1D6KMQ9
	ZmFAR1.2	*Zea mays*	A0A1D6KMQ9
	ZmFAR1.3	*Zea mays*	NP_001357621.1
	ZmFAR1.4	*Zea mays*	ACG42983.1
2	AtFAR2	*Arabidopsis thaliana*	Q08891
	AtFAR6	*Arabidopsis thaliana*	B9TSP7
	CaFAR2	*Cicer arietinum*	A0A1S2Z323
	LjFAR2	*Lotus japonicus* Gifu	LotjaGi3g1v0479100.3
	MtFAR2	*Medicago truncatula*	A0A072VJJ5
	OsFAR2	*Oryza sativa subsp. japonicus*	Q8S7T9
	PaFAR2	*Parasponia andersonii*	A0A2P5BCV3
	SlFAR2.1	*Solanum lycopersicum*	A0A3Q7G705
	SlFAR2.2	*Solanum lycopersicum*	A0A3Q7HZ68
	SlFAR2.3	*Solanum lycopersicum*	A0A3Q7HXS9
	SlFAR2.4	*Solanum lycopersicum*	A0A3Q7HXG0
	ZmFAR2	*Zea mays*	A0A1D6PJD1
3	AtFAR3	*Arabidopsis thaliana*	Q93ZB9
	CaFAR3.1	*Cicer arietinum*	A0A1S2XD65
	CaFAR3.2	*Cicer arietinum*	A0A1S2YKG6
	CaFAR3.3	*Cicer arietinum*	A0A1S2XG61
	CaFAR3.4	*Cicer arietinum*	A0A1S2YKS5
	LjFAR3.1*	***Lotus japonicus* Gifu**	LotjaGi3g1v0175200.1
	LjFAR3.2*	***Lotus japonicus* Gifu**	LotjaGi3g1v0478900.1
	LjFAR3.3	*Lotus japonicus* Gifu	LotjaGi3g1v0479000.1
	LjFAR3.4	*Lotus japonicus* Gifu	LotjaGi3g1v0479300.2
	LjFAR3.5	*Lotus japonicus* Gifu	LotjaGi3g1v0175300.1
	LjFAR3.6	*Lotus japonicus* Gifu	LotjaGi3g1v0479100.1
	MtFAR3.1	*Medicago truncatula*	G7IRD7
	MtFAR3.2	*Medicago truncatula*	LOC11446356
	MtFAR3.3	*Medicago truncatula*	LOC11425913
	MtFAR3.4	*Medicago truncatula*	XP_003606163.2
	PaFAR3	*Parasponia andersonii*	A0A2P5DU66
	SlFAR3.1	*Solanum lycopersicum*	A0A3Q7HW29
	SlFAR3.2	*Solanum lycopersicum*	A0A3Q7HW29
4	CaFAR4.1	*Cicer arietinum*	XP_004490308.1
	CaFAR4.2	*Cicer arietinum*	XP_027187874.1
	CaFAR8	*Cicer arietinum*	A0A1S2XL45
	LjFAR1	*Lotus japonicus* Gifu	LotjaGi4g1v0129900.1
	MtFAR1.1	*Medicago truncatula*	G7JJS1
	MtFAR1.2	*Medicago truncatula*	G7JJP4
	MtFAR1.3	*Medicago truncatula*	AES88362.2
	MtFAR4.1	*Medicago truncatula*	G7JJM2
	MtFAR4.2	*Medicago truncatula*	XP_039690485.1
	MtFAR4.3	*Medicago truncatula*	XP_039690758.1
	OsFAR1.7	*Oryza sativa subsp. japonicus*	A0A0E0HAC6

5	AtFAR1	*Arabidopsis thaliana*	Q39152
	AtFAR4	*Arabidopsis thaliana*	Q9LXN3
	AtFAR5	*Arabidopsis thaliana*	Q0WRB0
	AtFAR7	*Arabidopsis thaliana*	Q9FMQ9
	AtFAR8	*Arabidopsis thaliana*	Q1PEI6
	PaFAR7	*Parasponia andersonii*	A0A2P5DU79
6	SlFAR1.1	*Solanum lycopersicum*	XP_010313341.1
	SlFAR1.2	*Solanum lycopersicum*	XP_004251041.1
	SlFAR3.2	*Solanum lycopersicum*	A0A3Q7FAJ0
	SlFAR4	*Solanum lycopersicum*	XP_004241312.1

*Candidate genes detected by DE analysis

Supplemental Table 5. List of CASP and CASPLs sequences used in phylogenetic analyses.

Subfamily	Nomenclature	Species	UniProt ID or Gene ID
Outgroup	OtCASPL0U1	*Ostreococcus tauri*	Ot16g01510
	MpCASPL0U1	*Micromonas pusilla*	C1N652
1	PaCASP1	*Parasponia andersonii*	PON78046
	PaCASP2A	*Parasponia andersonii*	PON74504
	PaCASP3	*Parasponia andersonii*	PON65358
	PaCASP4	*Parasponia andersonii*	PON54141
	PaCASP5	*Parasponia andersonii*	PON65410
1A	AtCASP1	*Arabidopsis thaliana*	Q9SIH4
	AtCASP2	*Arabidopsis thaliana*	Q9CAX3
	AtCASP3	*Arabidopsis thaliana*	Q9ZQI2
	AtCASP4	*Arabidopsis thaliana*	Q9FFZ7
	AtCASP5	*Arabidopsis thaliana*	Q9LXF3
	AtCASPL1A1	*Arabidopsis thaliana*	Q9XI72
	CaCASP1	*Cicer arietinum*	A0A1S2XLI9
	CaCASP2	*Cicer arietinum*	A0A1S2Z791
	CaCASP3	*Cicer arietinum*	A0A1S2XD38
	LjCASP1	*Lotus japonicus* Gifu	P0DKC2
	LjCASP1	*Lotus japonicus* Gifu	P0DI58
	LjCASP1	*Lotus japonicus* Gifu	P0DI56
	LjCASP1	*Lotus japonicus* Gifu	P0DI55
	LjCASP1	*Lotus japonicus* Gifu	P0DI57
	MtCASP1	*Medicago truncatula*	G7KGQ4
	MtCASP2	*Medicago truncatula*	G7JG80
	MtCASP3	*Medicago truncatula*	G7L218
	MtCASP4	*Medicago truncatula*	G7IHF9
	OsCASP1	*Oryza sativa subsp. japonicus*	Q7XPU9
	OsCASP2	*Oryza sativa subsp. japonicus*	Q6Z1Y7
	OsCASP3	*Oryza sativa subsp. japonicus*	Q67X40
	OsCASP4	*Oryza sativa subsp. japonicus*	Q6Z2U5
	OsCASP5	*Oryza sativa subsp. japonicus*	Q7XUV7
	OsCASP7	*Oryza sativa subsp. japonicus*	Q6EP58
	SlCASP1	*Solanum lycopersicum*	A0A3Q7ILZ7
	SlCASP2	*Solanum lycopersicum*	A0A3Q7GAI4
	SlCASPL3	*Solanum lycopersicum*	A0A3Q7H363
	ZmCASP1	*Zea mays*	B6T959
	ZmCASP2	*Zea mays*	B6U045
1B	AtCASPL1B1	*Arabidopsis thaliana*	Q9FI10
	AtCASPL1B2	*Arabidopsis thaliana*	Q9SUP0
	CaCASPL1B1	*Cicer arietinum*	A0A1S2XGX0
	LjCASP4	*Lotus japonicus subsp.* Gifu	LotjaGi6g1v0214000.1
	MtCASPL1B1	*Medicago truncatula*	A0A072TGS8
	OsCASPL1B1	*Oryza sativa subsp. japonicus*	Q0ILZ7
	SlCASPL1B2	*Solanum lycopersicum*	A0A3Q7F1I4
	ZmCASPL1B1	*Zea mays*	B6TUH4
	ZmCASPL1B2	*Zea mays*	A7QC16
1C	AtCASPL1C1	*Arabidopsis thaliana*	Q9ZT81
	AtCASPL1C2	*Arabidopsis thaliana*	D7KCH2
	CaCASPL1C1	*Cicer arietinum*	A0A1S2YLL3
	LjCASP1.1	*Lotus japonicus* Gifu	LotjaGi3g1v0454600.1
	MtCASPL1C1	*Medicago truncatula*	A0A072TJ77
	OsCASPL1C1	*Oryza sativa subsp. japonicus*	A2WMK7
	OsCASPL1C2	*Oryza sativa subsp. japonicus*	Q84UT5
	SlCASPL1C1	*Solanum lycopersicum*	A0A3Q7F4H0
	ZmCASPL1C1	*Zea mays*	A7PJ32

	ZmCASPL1C2	*Zea mays*	B6SZU6
	AtCASPL1D1	*Arabidopsis thaliana*	Q9FE29
	AtCASPL1D2	*Arabidopsis thaliana*	D7L5G6
	CaCASPL1D2	*Cicer arietinum*	A0A1S2YQE8
1D	LjCASP1.3	*Lotus japonicus* Gifu	LotjaGi6g1v0214000.1
	LjCASP1.4	*Lotus japonicus* Gifu	LotjaGi4g1v0346000.1
	MtCASPL1D1	*Medicago truncatula*	G7JJ29
	OsCASPL1D1	*Oryza sativa subsp. japonicus*	Q6YT98
	ZmCASPL1D2	*Zea mays*	B6U361
	AtCASPL1E1	*Arabidopsis thaliana*	Q8L8Z1
	AtCASPL1E2	*Arabidopsis thaliana*	O23413
	CaCASPL1E1	*Cicer arietinum*	A0A1S2XTP9
1E	LjCASP1.3	*Lotus japonicus* Gifu	LotjaGi6g1v0214000
	MtCASPL1E1	*Medicago truncatula*	G7KQZ0
	OsCASPL1E1	*Oryza sativa subsp. japonicus*	Q9ARX2
	SlCASPL1E1	*Solanum lycopersicum*	M1C2K8
	ZmCASPL1E1	*Zea mays*	B4FAP1
	AtCASPL1F1	*Arabidopsis thaliana*	Q9M0L3
1F	CaCASPL1F2	*Cicer arietinum*	A0A1S2XRW1
	LjCASP1.6	*Lotus japonicus* Gifu	LotjaGi4g1v0346100
	SlCASPL1F1	*Solanum lycopersicum*	A0A3Q7GYS9
	PaCASP2B	*Parasponia andersonii*	PON61236
	PaCASP2D	*Parasponia andersonii*	PON58427
	PaCASP2E	*Parasponia andersonii*	PON63623
1-Pa**	PaCASP2F	*Parasponia andersonii*	PON80263
	PaCASP2G	*Parasponia andersonii*	PON80265
	PaCASP6	*Parasponia andersonii*	PON39284
	PaCASP7	*Parasponia andersonii*	PON77208
	PaCASP8	*Parasponia andersonii*	PON80266
	AtCASPL2A1	*Arabidopsis thaliana*	Q8VZQ3
	AtCASPL2A2	*Arabidopsis thaliana*	Q9LUL1
	CaCASPL2A1	*Cicer arietinum*	A0A1S2YD42
	LjCASP2.2	*Lotus japonicus* Gifu	LotjaGi6g1v0050500
	LjCASP2.3	*Lotus japonicus* Gifu	LotjaGi1g1v0101000
2A	MtCASPL2A1	*Medicago truncatula*	G7J7X5
	OsCASPL2A1	*Oryza sativa subsp. japonicus*	Q0JEF7
	PaCASP2	*Parasponia andersonii*	PON74504
	SlCASPL2A1	*Solanum lycopersicum*	A0A3Q7HDI7
	ZmCASPL2A1	*Zea mays*	B4FBQ7
	ZmCASPL2A2	*Zea mays*	B6SR79
	AtCASPL2B1	*Arabidopsis thaliana*	Q8L9B5
	AtCASPL2B2	*Arabidopsis thaliana*	Q8L924
	CaCASPL2B1	*Cicer arietinum*	A0A1S2XCV6
2B	LjCASP2.5	*Lotus japonicus* Gifu	LotjaGi6g1v0284400
	MtCASPL2B1	*Medicago truncatula*	G7IPB7
	OsCASPL2B1	*Oryza sativa subsp. japonicus*	Q0IN16
	SlCASPL2B1	*Solanum lycopersicum*	A0A3Q7IS57
	ZmCASPL2B1	*Zea mays*	K7TYK1
	AtCASPL2C1	*Arabidopsis thaliana*	Q8L8U9
	CaCASPL2C1	*Cicer arietinum*	A0A1S2YPE5
	LjCASP2.4	*Lotus japonicus* Gifu	LotjaGi4g1v0299300
	MtCASPL2C1	*Medicago truncatula*	A0A072TPZ4
2C	OsCASPL2C1	*Oryza sativa subsp. japonicus*	Q67W83
	SlCASPL2C1	*Solanum lycopersicum*	M1BM50
	ZmCASPL2C1	*Zea mays*	B6U769
	ZmCASPL2C2	*Zea mays*	B6TUB4
	ZmCASPL2C3	*Zea mays*	B6SZA7

123

	ZmCASPL2C4	*Zea mays*	B6U8R7
	AtCASPL2D1	*Arabidopsis thaliana*	Q9FFT2
	CaCASPL2D1	*Cicer arietinum*	A0A1S3E243
	LjCASP2.1	*Lotus japonicus* Gifu	LotjaGi1g1v0601200
2D	MtCASPL2D1	*Medicago truncatula*	I3S6M6
	OsCASPL2D1	*Oryza sativa subsp. japonicus*	A2X2I0
	SlCASPL2D1	*Solanum lycopersicum*	A0A3Q7J1J2
	ZmCASPL2D1	*Zea mays*	B6TUW9
	PaCASP2C	*Parasponia andersonii*	PON34013
2-Pa**	PaCASP2H	*Parasponia andersonii*	PON51483
	PaCASP2I	*Parasponia andersonii*	PON68600
	PaCASP9	*Parasponia andersonii*	PON56854
	AtCASPL3A1	*Arabidopsis thaliana*	Q3EB59
	AtCASPL3A2	*Arabidopsis thaliana*	Q1PFB8
	CaCASPL3A1	*Cicer arietinum*	A0A1S2YUU3
3A	LjCASP3.1	*Lotus japonicus* Gifu	LotjaGi3g1v0368900.1
	MtCASPL3A1	*Medicago truncatula*	G7LFX8
	OsCASPL3A1	*Oryza sativa subsp. japonicus*	Q5JM57
	PaCASP2J	*Parasponia andersonii*	PON54327
	SlCASPL3A1	*Solanum lycopersicum*	A0A3Q7GKQ1
	AtCASPL4A1	*Arabidopsis thaliana*	Q9FNE8
	AtCASPL4A2	*Arabidopsis thaliana*	Q501G6
	AtCASPL4A3	*Arabidopsis thaliana*	Q84WP5
	AtCASPL4A4	*Arabidopsis thaliana*	Q3EA54
	CaCASPL4A1	*Cicer arietinum*	A0A1S2Z5F1
	LjCASP4.4	*Lotus japonicus* Gifu	LotjaGi1g1v0181300.1
	LjCASP4.5	*Lotus japonicus* Gifu	LotjaGi4g1v0235000.1
4A	**LjCASPL4.3***	***Lotus japonicus* Gifu**	LotjaGi2g1v0246900.1
	MtCASPL4A1	*Medicago truncatula*	O24088
	OsCASPL4A1	*Oryza sativa japonicus*	A2X0L7
	OsCASPL4A2	*Oryza sativa japonicus*	A3A2W2
	SlCASPL4A1	*Solanum lycopersicum*	A0A3Q7FAD4
	ZmCASPL4A1	*Zea mays*	B6TWJ1
	ZmCASPL4A2	*Zea mays*	C4JAF2
	AtCASPL4B1	*Arabidopsis thaliana*	Q8LE26
	CaCASPL4B1	*Cicer arietinum*	A0A1S2XRL8
	LjCASP4.1*	***Lotus japonicus* Gifu**	LotjaGi3g1v0201700.1
	LjCASP4.2*	***Lotus japonicus* Gifu**	LotjaGi3g1v0201500.1
4B	MtCASPL4B1	*Medicago truncatula*	G7KV13
	OsCASPL4B1	*Oryza sativa subsp. japonicus*	Q84NQ7
	OsCASPL4B2	*Oryza sativa subsp. japonicus*	Q10MR5
	OsCASPL4B3	*Oryza sativa subsp. japonicus*	B9F6Z0
	SlCASPL4B1	*Solanum lycopersicum*	A0A3Q7JDY9
	ZmCASPL4B1	*Zea mays*	B6SM80
	AtCASPL4C1	*Arabidopsis thaliana*	Q9M2U0
	CaCASPL4C1	*Cicer arietinum*	A0A1S2Y265
	LjCASP4.6	*Lotus japonicus* Gifu	LotjaGi1g1v0449300
	LjCASP4.7	*Lotus japonicus* Gifu	LotjaGi1g1v0075300
4C	LjCASP4.8	*Lotus japonicus* Gifu	LotjaGi5g1v0355600
	MtCASPL4C1	*Medicago truncatula*	A0A072VR44
	OsCASPL4C1	*Oryza sativa subsp. japonicus*	Q2QNE3
	SlCASPL4C1	*Solanum lycopersicum*	A0A3Q7GQM5
	ZmCASPL4C1	*Zea mays*	A0A1D6MWW7
	AtCASPL4D1	*Arabidopsis thaliana*	Q8GWD5
4D	AtCASPL4D2	*Arabidopsis thaliana*	Q56X75
	CaCASPL4D1	*Cicer arietinum*	A0A3Q7XXQ2
	LjCASP4.10	*Lotus japonicus* Gifu	LotjaGi1g1v0539300

	LjCASP4.11*	*Lotus japonicus* Gifu	LotjaGi2g1v0247000
	LjCASP4.9*	*Lotus japonicus* Gifu	LotjaGi1g1v0539200
	MtCASPL4D1	*Medicago truncatula*	G7KIJ0
	OsCASPL4D1	*Oryza sativa subsp. japonicus*	Q2R2T4
	SlCASPL4D1	*Solanum lycopersicum*	A0A3Q7FN59
	ZmCASPL4D1	*Zea mays*	A0A1D6F6J3
4-Pa**	PaCASP2K	*Parasponia andersonii*	PON61109
	PaCASP2L	*Parasponia andersonii*	PON36660
	PaCASP2M	*Parasponia andersonii*	PON53595
	PaCASP2N	*Parasponia andersonii*	PON58340
	PaCASP2O	*Parasponia andersonii*	PON51062
	PaCASP2P	*Parasponia andersonii*	PON56589
	PaCASP2Q	*Parasponia andersonii*	PON33834
	PaCASP2R	*Parasponia andersonii*	PON35555
5A	AtCASPL5A1	*Arabidopsis thaliana*	Q6NPF8
	AtCASPL5A2	*Arabidopsis thaliana*	Q9SKN3
	CaCASPL5C1	*Cicer arietinum*	A0A1S2YKI3
	LjCASP5.5	*Lotus japonicus* Gifu	LotjaGi2g1v0120700.1
	MtCASPL5C1	*Medicago truncatula*	I3SCI9
	OsCASPL5A1	*Oryza sativa subsp. japonicus*	Q10Q78
	OsCASPL5A2	*Oryza sativa subsp. japonicus*	Q339M6
	SlCASPL5C1	*Solanum lycopersicum*	A0A3Q7JH14
	ZmCASPL5A1	*Zea mays*	P0DI67
	ZmCASPL5A2	*Zea mays*	P0DI66
	ZmCASPL5A3	*Zea mays*	B4FNS3
5B	AtCASPL5B1	*Arabidopsis thaliana*	Q9LZM5
	AtCASPL5B2	*Arabidopsis thaliana*	Q945M8
	AtCASPL5B3	*Arabidopsis thaliana*	NA
	CaCASPL5B1	*Cicer arietinum*	A0A1S2YLI8
	LjCASP5.3	*Lotus japonicus* Gifu	LotjaGi3g1v0228900.1
	LjCASP5.4	*Lotus japonicus* Gifu	LotjaGi5g1v0183900.1
	MtCASPL5B1	*Medicago truncatula*	A0A072VX83
	OsCASPL5B1	*Oryza sativa subsp. japonicus*	Q5N794
	OsCASPL5B2	*Oryza sativa subsp. japonicus*	Q0DHM7
	SlCASPL5B1	*Solanum lycopersicum*	A0A3Q7GMZ1
	ZmCASPL5B1	*Zea mays*	B6TM88
	ZmCASPL5B2	*Zea mays*	B6TAX2
	ZmCASPL5B3	*Zea mays*	B6T990
5C	AtCASPL5C1	*Arabidopsis thaliana*	Q66GI1
	AtCASPL5C2	*Arabidopsis thaliana*	P0CB17
	AtCASPL5C3	*Arabidopsis thaliana*	Q3ECT8
	CaCASPL5C1	*Cicer arietinum*	A0A1S2YND9
	LjCASP5.1	*Lotus japonicus* Gifu	LotjaGi4g1v0113000.1
	LjCASP5.2	*Lotus japonicus* Gifu	LotjaGi3g1v0503900.5
	MtCASPL5C1	*Medicago truncatula*	I3TAN6
	OsCASPL5C1	*Oryza sativa subsp. japonicus*	Q10EJ2
	SlCASPL5C1	*Solanum lycopersicum*	A0A3Q7FXT7
	ZmCASPL5C1	*Zea mays*	B6U300

*Candidate genes detected by DE analysis. In bold all the *Lj*CASPL Subfamily 4 proteins
** Parasponia accessions

Supplemental Table 6. List of plasmids used in this work and cloning backbones (BB).

Plasmid	Description	Reference
LI – pUC57-BB2	Vector for blunt end cloning, Gm^R	Binder et al. 2014
LI – Bpi – BB3	LI for subcloning and mutagenesis, Gm^R	Binder et al. 2014
LIIβF – 3-4 – BB24	Binary expression vector, Sm^R	Binder et al. 2014
LIIIβfin-BB52	LIIIIβ expression vector, Km^R	Binder et al. 2014
Promoters		
LI – pUC57-CASPL4.1pro	LI pUC57 carrying a 3 kb fragment upstream the *L. japonicus* Gifu *LjCASPL4.1* gene, Gm^R	This study
LI – pUC57-CASPL4.9pro	LI pUC57 carrying a 1.5 kb fragment upstream the *L. japonicus* Gifu *LjCASPL4.9* gene, Gm^R	Koyinde Akindele Master thesis 2022
LI – pUC57-CASPL4.11pro	LI pUC57 carrying a 1 kb fragment upstream the *L. japonicus* Gifu *LjCASPL4.11* gene, Gm^R	Koyinde Akindele Master thesis 2022
LI – pUC57-FAR3.1pro	LI pUC57 carrying a 3 kb fragment upstream the *L. japonicus* Gifu *LjFAR3.1* gene, Gm^R	This study
LI – pUC57-FAR3.2pro	LI pUC57 carrying a 3 kb fragment upstream the *L. japonicus* Gifu *LjFAR3.2* gene, Gm^R	This study
LI – pUC57-NACDpro	LI pUC57 carrying a 3 kb fragment upstream the *L. japonicus* Gifu *LjNAC* gene, Gm^R	This study
LI – pUC57-RBOHBpro	LI pUC57 carrying a 3 kb fragment upstream the *L. japonicus* Gifu *LjRBOHB* gene, Gm^R	This study
GUS promoters		
LIII – pCASPL4.1pro:DoGUS	LIII XGp179a expressing DoGUS reporter under the control of the 3kb *CASPL4.1* promoter	This study
LIII – pCASPL4.3pro:DoGUS	LIII XGp179a expressing DoGUS reporter under the control of the 3kb *CASPL4.3* promoter	This study
LIII – pFAR3.1pro:DoGUS	LIII XGp179a expressing DoGUS reporter under the control of the 3kb *FAR3.1* promoter	This study
LIII – pFAR3.2pro:DoGUS	LIII XGp179a expressing DoGUS reporter under the control of the 3kb FAR3.*2* promoter	This study
LIII – pNACDpro:DoGUS	LIII XGp179a expressing DoGUS reporter under the control of the 3kb *NACD promoter*	This study
LIII – pRBOHBpro:DoGUS	LIII XGp179a expressing DoGUS reporter under the control of the 3kb *RBOHB promoter*	This study
LIII – XGp179a pUbi10 DoGUS	Binary expression vector for GUS with NLS 2xGFP-*lacZ*dy – GUS, Km^R	(Gong et al., 2021)
NLS-2xYFP promoters		
LIII – TM- lacZdy NLS-2xYFP	Binary expression vector for promoter *NLS-2xYFP* fusions with *NLS-2xmChe*rry transformation marker, Km^R	This study
LIII – pCASPL4.1pro:NLS-2xYFP	LIII vector expressing the *NLS-2xYFP* reporter under the control of the 3kb *CASPL4.1* promoter	This study
LIII – pCASPL4.9pro:NLS-2xYFP	LIII vector expressing the *NES-2xYF*P reporter under the control of the 1.5kb *CASPL4.9* promoter	Koyinde Akindele Master thesis 2022
LIII – pCASPL4.11pro:NLS-2xYFP	LIII vector expressing the *NES-2xYFP* reporter under the control of the 1 kb *CASPL4.11* promoter	Koyinde Akindele Master thesis 2022
LIII – pFAR3.1pro:NLS-2xYFP	LIII vector expressing the *NLS-2xYFP* reporter under the control of the 3kb *FAR3.1* promoter	This study
LIII – pFAR3.2pro:NLS-2xYFP	LIII vector expressing the *NLS-2xYFP* reporter under the control of the 3kb *FAR3.2* promoter	This study

LIII – pRBOHBpro:NLS-2xYFP	LIII vector expressing the *NLS-2xYFP* reporter under the control of the 3kb *RBOHB* promoter	This study
LIII – pNACpro:NLS-2xYFP	LIII vector expressing the *NLS-2xYFP* reporter under the control of the 3kb *NAC* promoter	This study

Genes

LI – LjCASP8	*LI+Bpi – BB3* carrying the genomic sequence for *LjCAPSL8*	Koyinde Akindele Master thesis 2022
LI – LjCASPL4.1	*LI+Bpi – BB3* carrying the genomic sequence for *LjCAPSL4.1*	This study
LI – LjCASPL4.2	*LI+Bpi – BB3* carrying the genomic sequence for *LjCAPSL4.2*	This study
LI – LjCASPL4.3	*LI+Bpi – BB3* carrying the genomic sequence for *LjCAPSL4.3*	This study
LII – GFP- LjCASPL4.1	Binary vector for expression of *LjCASPL4.1* fused internally to GFP. Expression assembled via *BsaI* into *LIIβF 3-4 – BB24*	This study
LII – GFP- LjCASPL4.2	Binary vector for expression of *LjCASPL4.2* fused internally to GFP. Expression assembled via *BsaI* into *LIIβF 3-4 – BB24*	This study
LII – GFP- LjCASPL4.3	Binary vector for expression of *LjCASPL4.3* fused internally to GFP. Expression assembled via *BsaI* into *LIIβF 3-4 – BB24*	This study
LII – GFP- LjCASPL8	Binary vector for expression of *LjCASPL8* fused internally to GFP. Expression assembled via *BsaI* into *LIIβF 3-4 – BB24*	Koyinde Akindele Master thesis 2022

CRISPR/Cas12a

LI 1-2 *LjU6pro*	LI plasmid containing the *LjU6* promoter, GmR	Martin Bircheneder personal communication
LI 1-2 *LjU6*-T	LI plasmid containing the *LjU6* terminator, GmR	M. Bircheneder personal communication
LII 1-2 *Nospro:Hygromycin*	LII hygromycine resistance gene under the control of *Nos* promoter for stable transformation and plant selection	M. Bircheneder personal communication
LII 2-3 *LjUb10pro:ttCas12a*	LII *Cas12a* under the control of *LjUb10* promoter for CRISPR gene editing	Martin Bircheneder personal communication
LIII- HygR-Cas12a-LjU6pro:CASPL4.1	LIII vector expressing *Nospro:Hygromycin* for selection and *LjUb10pro:ttCas12a* for gene editing the *LjCASPL4.1.* Expression assembled via *BpiI* into *LIIIβF 3-4 – BB24*	This study
LIII- HygR-Cas12a-LjU6pro:CASPL4.3	LIII vector expressing *Nospro:Hygromycin* for selection and *LjUb10pro:ttCas12a* for gene editing the *LjCASPL4.3*	This study
LIII- HygR-Cas12a-LjU6pro:CASPL4.1 and CASPL4.2-dm	LIII vector expressing *Nospro:Hygromycin* for selection and *LjUb10pro:ttCas12a* for gene editing the *LjCASPL4.1* and *LjCASPL4.2* - double mutant	This study
LIII- HygR-Cas12a-LjU6pro:CASPL4.1 and CASPL4.3-dm	LIII vector expressing *Nospro:Hygromycin* for selection and *LjUb10pro:ttCas12a* for gene editing the *LjCASPL4.1* and *LjCASPL4.3* - double mutant	This study
LIII- HygR-Cas12a-LjU6pro:RBOHB	LIII vector expressing *Nospro:Hygromycin* for selection and *LjUb10pro:ttCas12a* for gene editing the *LjRBOHB*	This study

Supplemental Table 7. List of *Lotus japonicus LORE1* lines used in this work.

Accession number	Position	Allele name	Off-targets			
			Exonic	Intronic	Intergenic	Total
30165196	Chr3. 37238378	*far3.2-1*	14	6	4	24
30127714	Chr3. 37238398	*far3.2-2*	4	5	3	12

Supplemental Table 8. Composition for different medias used during stable line generation.

Components	Co-Cultivation	*Callus Induction	*Shoot Induction	*Shoot Elongation	*Root Induction	*Root elongation
Water	100 ml	100 ml	100 ml	100 ml	100 ml	100 ml
Gamborg B5 basal salt	0.33 g	0.33 g	0.33 g	0.33 g	0.165 g	0.165 g
Sucrose	-	2 g	2 g	2 g	1 g	1 g
Gelrite	-	0.5 g	0.5 g	0.5 g	0.5 g	0.5 g
$(NH_4)_2SO_4$	-	125 µl	125 µl	-	-	-
Gamborg B5 vitamins (1000X)	10 µl	100 µl	100 µl	100 µl	50 µl	50 µl
NAA (1 mg ml^{-1})	5 µl	5 µl	5 µl	-	50 µl	-
BAP (1 mg ml^{-1})	50 µl	20 µl	20 µl	20 µl	-	-
1 M MES (pH = 5.2)	500 µl	-	-	-	-	-
Acetosyringone	100 µl	-	-	-	-	-
Cefotaxime (300 mg ml^{-1})	-	120 µl	-	-	-	120 µl
Hygromicin (50 mg ml^{-1})	-	80 µl	-	-	-	-

* Autoclave and adjust pH = 5.5

Supplemental Table 9. Shared DEG for all nine pairwise comparisons.

Gene ID	Description	Regulation
LotjaGi0g1v0000800	Agenet domain protein	Up
LotjaGi1g1v0003900	Oxygen-dependent coproporphyrinogen-III oxidase	Up
LotjaGi1g1v0035700	Small nuclear ribonucleoprotein-associated protein	Down
LotjaGi1g1v0048000	Non-specific serine/threonine protein kinase	Up
LotjaGi1g1v0059700	Aldehyde oxidase	Down
LotjaGi1g1v0150600	transcription factor-like protein	Up
LotjaGi1g1v0199700	Heme oxygenase 1	Up
LotjaGi1g1v0217000	Retrovirus-related Pol polyprotein from transposon TNT 1-94	Up
LotjaGi1g1v0221300	GDSL esterase/lipase	Up
LotjaGi1g1v0292700	Cytochrome B5 isoform E	Up
LotjaGi1g1v0293900	Homeobox protein	Up
LotjaGi1g1v0309300	Peptide transporter	Up
LotjaGi1g1v0340900	Basic helix-loop-helix transcription factor	Up
LotjaGi1g1v0353600	alpha/beta-Hydrolases	Up
LotjaGi1g1v0387300	EG45-like domain containing protein	Up
LotjaGi1g1v0397700	plant/protein (Protein of unknown function)	Up
LotjaGi1g1v0398100	NAC domain-containing protein	Up
LotjaGi1g1v0412200	Cysteine protease	Up
LotjaGi1g1v0420000	DNA polymerase epsilon catalytic subunit A	Up
LotjaGi1g1v0440900	Vacuolar cation/proton exchanger, putative	Up
LotjaGi1g1v0444900	GRAS family transcription factor	Up
LotjaGi1g1v0449100	P-loop containing nucleoside triphosphate hydrolases	Down
LotjaGi1g1v0451500	LURP-one-like protein	Up
LotjaGi1g1v0491200	Glycosyltransferase	Up
LotjaGi1g1v0508300	O-methyltransferase	Up
LotjaGi1g1v0517100	B12D-like protein	Up
LotjaGi1g1v0550200	LOB domain-containing protein	Up
LotjaGi1g1v0562700	Arginine decarboxylase	Down
LotjaGi1g1v0563500	Serpin-like protein	Down
LotjaGi1g1v0577100	Glutathione S-transferase	Down
LotjaGi1g1v0578900	Zinc finger protein	Up
LotjaGi1g1v0591900	GTP cyclohydrolase 1	Up
LotjaGi1g1v0603000	Cytochrome P450 83B1	Up
LotjaGi1g1v0613400	2-isopropylmalate synthase	Up
LotjaGi1g1v0614400	Elongation factor 1-alpha	Down
LotjaGi1g1v0627500	Bifunctional uridylyltransferase/uridylyl-removing enzyme	Down
LotjaGi1g1v0630600	Enolase	Up
LotjaGi1g1v0630800	Enolase	Up
LotjaGi1g1v0639600	Ricin B lectin domain-containing protein	Down
LotjaGi1g1v0686100	demeter-like protein 3	Up
LotjaGi1g1v0696000	ABC transporter G family member	Up
LotjaGi1g1v0720200	Alkaline alpha-galactosidase seed imbibition protein	Up

129

LotjaGi1g1v0726200	Fatty acid desaturase	Up
LotjaGi1g1v0728600	Protein trichome birefringence	Down
LotjaGi1g1v0733300	Dirigent protein	Down
LotjaGi1g1v0750700	Pyridoxine/pyridoxamine 5'-phosphate oxidase	Down
LotjaGi1g1v0753100	Cathepsin B-like cysteine protease	Up
LotjaGi1g1v0754800	ABC transporter B family protein	Up
LotjaGi1g1v0758000	RNA binding protein, putative	Down
LotjaGi1g1v0758200	ATP-dependent zinc metalloprotease FTSH protein	Down
LotjaGi2g1v0069900	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase	Up
LotjaGi2g1v0167800	Glutathione S-transferase T3	Down
LotjaGi2g1v0187400	Chalcone reductase	Up
LotjaGi2g1v0209000	LOB domain-containing protein, putative	Up
LotjaGi2g1v0213700	Sulfate transporter	Up
LotjaGi2g1v0257400	2-aminoethanethiol dioxygenase-like protein	Up
LotjaGi2g1v0261000	Methionine S-methyltransferase	Down
LotjaGi2g1v0275400	Leguminosin group485 secreted peptide / IgA FC receptor	Up
LotjaGi2g1v0308400	Hypersensitive-induced response protein 1	Up
LotjaGi2g1v0330500	Receptor kinase (*LjSYMRK*)	Down
LotjaGi2g1v0349200	Glycine-rich protein	Down
LotjaGi2g1v0372000	70 kDa heat shock protein	Up
LotjaGi2g1v0391600	Polyol transporter	Up
LotjaGi2g1v0394000	Geranylgeranyl pyrophosphate synthase	Down
LotjaGi2g1v0405300	Cytochrome P450 81E8	Up
LotjaGi2g1v0453000	12-oxophytodienoate reductase-like protein	Down
LotjaGi3g1v0021400	Histidine phosphotransfer protein	Down
LotjaGi3g1v0033700	Calmodulin, putative	Down
LotjaGi3g1v0061100	BTB/POZ domain-containing family protein	Down
LotjaGi3g1v0079300	Wound-responsive family protein	Up
LotjaGi3g1v0134800	Carbohydrate-binding X8 domain superfamily protein	Up
LotjaGi3g1v0148700	BTB/POZ domain-containing family protein	Down
LotjaGi3g1v0148900	Transposon Ty3-G Gag-Pol polyprotein	Down
LotjaGi3g1v0149500	Thiamine thiazole synthase	Up
LotjaGi3g1v0175200	Fatty acyl-CoA reductase	Up
LotjaGi3g1v0183600	Protein sawadee homeodomain-like 2	Down
LotjaGi3g1v0201500	CASP-like protein	Up
LotjaGi3g1v0201700	CASP-like protein	Up
LotjaGi3g1v0249300	Cysteine protease, putative	Up
LotjaGi3g1v0254300	Cysteine protease	Up
LotjaGi3g1v0263000	Receptor-like protein kinase	Down
LotjaGi3g1v0312900	WAT1-related protein	Up
LotjaGi3g1v0326800	LINE-1 reverse transcriptase like	Down
LotjaGi3g1v0335400	Type IV secretion system protein PtlG	Down
LotjaGi3g1v0351700	Ethylene-responsive transcription factor	Up
LotjaGi3g1v0419900	Unknown	Up

LotjaGi3g1v0432200	transcription regulatory protein SNF2	Down
LotjaGi3g1v0439600	Wound-responsive family protein	Up
LotjaGi3g1v0439700	Wound-responsive family protein	Up
LotjaGi3g1v0445300	14 kDa proline-rich protein DC2.15	Up
LotjaGi3g1v0476100	serine/arginine repetitive matrix protein	Down
LotjaGi3g1v0478900	Fatty acyl-CoA reductase	Up
LotjaGi3g1v0489400	Cysteine protease, putative	Up
LotjaGi3g1v0491700	Copper transporter family protein	Up
LotjaGi3g1v0504600	Non-symbiotic hemoglobin 1	Down
LotjaGi3g1v0545600	Purple acid phosphatase	Up
LotjaGi4g1v0022600	Gibberellin 2-oxidase	Up
LotjaGi4g1v0031000	Vacuolar iron transporter-like protein	Up
LotjaGi4g1v0059500	Trigger factor	Up
LotjaGi4g1v0088600	Thioredoxin-like protein 4B	Down
LotjaGi4g1v0097700	Subtilisin-like protease	Up
LotjaGi4g1v0108600	Kinase family protein	Up
LotjaGi4g1v0114400	Calcium-binding protein	Up
LotjaGi4g1v0116000	Sugar transporter, putative	Up
LotjaGi4g1v0117800	Alpha/beta fold hydrolase	Up
LotjaGi4g1v0173200	1-deoxy-D-xylulose 5-phosphate synthase	Down
LotjaGi4g1v0207100	Nitrate transporter 1.1	Up
LotjaGi4g1v0210200	Plant invertase/pectin methylesterase inhibitor	Up
LotjaGi4g1v0212500	Senescence-associated protein	Up
LotjaGi4g1v0228700	Octicosapeptide/Phox/Bem1p domain-containing protein kinase	Down
LotjaGi4g1v0258100	Ypt/Rab-GAP domain of gyp1p superfamily protein	Up
LotjaGi4g1v0279900	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase	Up
LotjaGi4g1v0296300	Receptor-like kinase	Down
LotjaGi4g1v0317900	Jasmonate ZIM domain-containing protein	Down
LotjaGi4g1v0339000	senescence-associated family protein	Up
LotjaGi4g1v0381200	Unknown	Up
LotjaGi4g1v0390100	CASP-like protein	Up
LotjaGi4g1v0396000	Cytochrome P450 71D8	Down
LotjaGi4g1v0401300	Ring finger protein	Up
LotjaGi4g1v0428200	M1 family aminopeptidase	Down
LotjaGi4g1v0429900	Monodehydroascorbate reductase	Down
LotjaGi4g1v0466300	Tetratricopeptide repeat	Down
LotjaGi5g1v0014600	DNA polymerase epsilon catalytic subunit A, putative	Up
LotjaGi5g1v0024700	Non-symbiotic hemoglobin 1	Up
LotjaGi5g1v0024900	Non-symbiotic hemoglobin 1	Up
LotjaGi5g1v0046500	Non-symbiotic hemoglobin	Up
LotjaGi5g1v0051600	Chorismate mutase	Up
LotjaGi5g1v0089600	Transcription factor, putative	Up
LotjaGi5g1v0095500	Carbonic anhydrase family protein	Up
LotjaGi5g1v0144700	Cytochrome P450 93A1	Up

131

LotjaGi5g1v0169600	Lipid transfer protein	Up
LotjaGi5g1v0178400	Zinc finger family protein	Up
LotjaGi5g1v0184800	Repetitive proline-rich cell wall protein 2	Up
LotjaGi5g1v0187200	Zinc finger family protein	Up
LotjaGi5g1v0192400	Auxin-responsive protein	Down
LotjaGi5g1v0211100	Auxin-induced in root cultures protein 12	Down
LotjaGi5g1v0213200	GDP-mannose 3,5-epimerase	Up
LotjaGi5g1v0216300	Kinesin, putative	Up
LotjaGi5g1v0216400	WD repeat-containing protein-like protein	Up
LotjaGi5g1v0224200	Respiratory burst oxidase-like protein	Up
LotjaGi5g1v0232400	Hexosyltransferase	Down
LotjaGi5g1v0253300	Myosin heavy chain-like protein, putative	Up
LotjaGi5g1v0269800	Extensin	Up
LotjaGi5g1v0287700	GIGANTEA	Up
LotjaGi5g1v0297000	Ferric reductase oxidase	Down
LotjaGi5g1v0329100	Kynurenine formamidase	Up
LotjaGi5g1v0335200	---	Up
LotjaGi5g1v0359700	ABC transporter G family member;	Up
LotjaGi6g1v0031500	En/Spm-like transposon protein	Down
LotjaGi6g1v0051900	Myb transcription factor	Down
LotjaGi6g1v0094800	Transposon protein, putative,	Down
LotjaGi6g1v0103900	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase	Down
LotjaGi6g1v0105200	Transcription factor bHLH93-like protein	Up
LotjaGi6g1v0111000	Cytochrome P450 86A1	Up
LotjaGi6g1v0145100	Serine acetyltransferase	Up
LotjaGi6g1v0152200	Transposon protein, putative	Up
LotjaGi6g1v0152300	7,8-dihydroneopterin aldolase	Up
LotjaGi6g1v0152500	7,8-dihydroneopterin aldolase	Up
LotjaGi6g1v0193000	Disease resistance protein	Up
LotjaGi6g1v0195800	Homoserine dehydrogenase	Up
LotjaGi6g1v0199600	Heptahelical transmembrane protein 2	Up
LotjaGi6g1v0202200	Ribosomal RNA small subunit methyltransferase G	Down
LotjaGi6g1v0224100	LURP-one-like protein	Up
LotjaGi6g1v0246500	Galactan beta-1,4-galactosyltransferase GALS1	Down
LotjaGi6g1v0260900	Pectinesterase inhibitor	Up
LotjaGi6g1v0321700	Receptor kinase, putative	Down
LotjaGi6g1v0358300	Protein SUPPRESSOR OF GENE SILENCING 3	Down

Supplemental Table 10. List of mutant lines generated via CRISPR/Cas12a with their target and genotype.

Mutant	1st gRNA position	2nd gRNA position	Lines+	gRNA 1 mutation	gRNA 2 mutation
Ljcaspl4.1	Exon 1 *LjCASPL4.1*	Exon 3 *LjCASPL4.1*	1	Heterozygous	Unknown
			2	Heterozygous	Unknown
Ljcaspl4.1 caspl4.2	Exon 1 *LjCASPL4.1*	Exon 1 *LjCASPL4.2*	1	10 bp deletion*	6 bp deletion*
			2	8 bp deletion*	6 bp deletion*
Ljcaspl4.1 casp4.3	Exon 1 *LjCASPL4.1*	3'-UTR *LjCASPL4.3*	1	8 bp deletion*	Heterozygous
			2	Heterozygous	Heterozygous
			3	7 bp deletion*	Heterozygous
			4	Heterozygous	Heterozygous
Ljrbohb	Exon 1 *LjRBOHB*	Exon 11 *LjRBOHB*	1	Heterozygous	Heterozygous
			2	Heterozygous	Heterozygous
			3	Heterozygous	Heterozygous
			4	Heterozygous	Heterozygous
			5	Heterozygous	Heterozygous

*indicate homozygosity.
+ All lines are coming from independent calli

List of tables

List of figures

135

Declaration of contribution

Rafael Venado

- All figures were edited and modified by Rafael Venado
- Figure 1 and 3: Schemes were drawn by Rafael Venado
- Figure 2, 6-16, 19-29: Experiments were performed and analyzed by Rafael Venado
- Figure 4-5, 14,17-19, 21, 25, 27 and 29: Experiments were performed by collaborators or by supervised students. Exact contributions are listed below

Lucas Wange

- He made libraries, sent for sequencing and mapped the raw data
- Figure 4: He performed mapping to both *L. japonicus* genomes (Gifu and MG-20)
- Figure 6: He generated mapping statistics from the zUMI pipeline

Fabienne Pinnau

- Figure 14: She obtained FAR orthologs, performed alignment and made the dendrogram
- Figure 17 and 26: She helped with GUS staining and microscopy for the promoters
- Figure 19: She helped with the toluidine staining assay
- Figure 21: She recorded the phenotype of *far3.2* mutant lines and their WT* (panels C to E) and took pictures (panel B)

Defeng Shen

- Figure 18: He performed fluorol yellow staining and quantification of the signal in the *far3.2* mutant lines and their WT*

Koyinde Akindele

- Figure 25: She cloned the promoters *CASPL4.9* and *CASPL4.11* and helped with microscopy
- Figure 27: She annotated the protein domains and cloned the CASP8. In addition, transient expression in *N. benthamiana* was done by her.
- Figure 29: She helped with phenotyping and genotyping of the *caspl4.1 caspl4.2* segregation population.

Acknowledgements

So, this is it! The highest academic degree one can get and now I want to self-reflect and thank everyone that helped and supported me during this great adventure.

To my parents and relatives:

Maria Elena Venado (my mom) and Jaime Espejel (my dad), who love me no matter what, with or without a Ph.D. degree. My sister Alejandra Espejel Venado and my three nephews Arturo, Alex, and Angel. To all members of my two big families: the Espejels and Venados. Many relatives and friends of my parents have prayed for me. Thanks to all of you. Two people deserve an honorable mention my uncle Russ Paradise for all these years of constant help and second my grandmother Gudelia Soriano who passed away but before I came here believed I could achieve this goal. You will always be in my thoughts.

To the Marín Lab:

I believe we all need a "master", someone who sees our potential and guides us to become a better person. I consider Dr. Macarena Marín as my "master", a person who made me a scientist and supported me all these years during this journey. Her maternal Latin instinct pampered us and for all that, I sincerely thank you. I will always be your lifelong student Rafa.

My lab friends, current and previous grad students: Juan Liang, Yen-Yu Lin, Duncan Crosbie, and Sara Masachis. You made lab life way better.

To my thesis committee:

Thanks for spending time evaluating and reading my thesis. Thanks to Prof. Dr. Korbinian Schneerger, Prof. Dr. Thorben Cordes, Prof. Dr. Wolfgang Enard, and a special thank you to Prof. Dr. Silke Werth and Prof. Dr. Wolfgang Frank for all their ideas to improve this project.

To my friends:

First off, all members of "La bolita retejaladora": Areli Gutiérrez, Diana Jeronimo, Eduardo Jiménez, Humberto Herrera, Luis Dávila, Rafael León and Raquel Díaz. Second off, to Abby Pepin, thanks to you I am a much better person. You are like my second family; I am so blessed to have you in my life. I can only wish you all the best in your life.

Finally, I want to conclude with a brief reflection and an oath. When I was 15 years old, I had the opportunity to live in the poor conditions of farmers. That made me realize that I want to contribute positively to reducing poverty through innovation in agriculture. I have forged a path to prepare myself for this moment. Therefore, I am committed to working hard and improving agriculture to support farmers and achieve food security.

Curriculum vitae

Oscar Rafael Espejel Venado

About me

Currently, I have 10 years of experience in the fields of molecular biology, next generation sequencing, plant breeding, capacity development and preparation of workshops to support farmers, among others. I am looking for the opportunity to positively contribute to society through research and student training.

Work Experience

2017-2018

International Maize and Wheat Improvement Center (CIMMYT). Consultant
Working for MasAgro (Sustainable Modernization of Traditional Agriculture):
1. Collaborate with activities related to research partnership
2. Handle and process samples with the objective of obtaining genotypic data (using DArTseq) and their subsequent analysis and use in research.
3. Assist with planning and development of MasAgro training workshops
4. Collaborate in the preparation of reports to donors and research partners

2014-2016

Purdue University. Department of Agronomy. Graduate Research Assistant
Interdisciplinary work with people from other departments:
1. Field responsibilities: field designs and field phenotyping
2. Lab responsibilities: carotenoid extraction and quantification with high performance liquid chromatography (HPLC)
3. Data analysis: Analysis of whole genome sequence (WGS) and detection of quantitative trait locus (QTL)

2011-2014

International Maize and Wheat Improvement Center (CIMMYT). Lab Assistant
Participated in the following projects:
1. Implementing KASP genotyping for proVitamin A in tropical maize lines
2. Development of doubled haploids in tropical maize lines with molecular markers and the detection of the QTL-*hir1*
3. Automatization of DNA extraction with a "Beckman-Coulter: Biomek FX"
4. Validation and design of molecular markers associated with the allele *R1-nj* for detecting doubled haploid maize lines

Education

2018-present

Ludwig Maximilian University of Munich, Germany
Doctoral candidate Plant molecular biology in the field of Root Nodule Symbiosis. Advisor: Dr. Macarena Marín

2014-2016

Purdue University, United States of America
Master of Science in Agronomy, specializing in Plant Breeding and Genetics.
Thesis Title: Genetic Study of Carotenoids in Maize (*Zea mays* L) Grain.
Advisor: Dr. Torbert Rocheford

2006-2010

National Autonomous University of Mexico, Mexico
Bachelor of Science, majoring in Pharmaceutical Chemist Biologist
Thesis Tittle: Synthesis of 2-imino-3,4-dihydro- (1H) -pyrimidine through a specific cleavage of heterocyclic N-cyanoguanidine. Advisor: Dr. Hulme Ríos
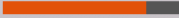
Publications and Awards

* Genetic analysis of provitamin A carotenoid β- cryptoxanthin concentration and relationship with other carotenoids in maize grain (Zea mays L.). Molecular Breeding. Sep 22, 2017

* Rhizobia infection, a journey to the inside of plant cells. Advances in Botanical Research. Feb 3, 2020

* Ph.D. scholarship awarded by the DAAD (Deutscher Akademischer Austauschdienst) - German Academic Exchange Service. Aug 2018

* Scholarship for graduate studies abroad by Mexico's National Council on Science and Technology (CONACyT). Sep 2015

Skills

Languages

Spanish
English
German

Softwares

R Core Team
ImageJ
AffinityDesigner
Linux

Teaching and other activities

* Supervision of 2 bachelor thesis and 2 masters students lab rotation under the scope of my project

* Teaching the module Fiji ImageJ for the Computational Biology course. Since October 2019

* Flash talk at the 14th European Nitrogen Fixation Conference 2021. Tittle "Comparative transcriptomics reveals tissue-specific regulation of suberin genes during root nodule endosymbiosis". October 2021

* Participation and certification at the online Good Manufacturing Practice. August 2020

* ECSEQ Bioinformatics. Workshop "RNA-seq Data Analysis". LMU Biocenter. April 2019

* University and Industry Consortium (UIC 2016). Dow Agroscience, Indianapolis, Indiana. October 2016. Poster title: "Initial evaluation of candidate genes sequence variation in a population selected for orange color"

* Corn Showcase Poster Session. Purdue University. West Lafayette, Indiana. July 2016. Poster title:"Genetic analysis of carotenoids and color in a biparental maize population"

* Plant Science Networking Gathering. Purdue University. West Lafayette, Indiana. October 2015. Poster title: "Genetic Analysis of provitamin A and total Carotenoids in Maize Grain"

* Participation at the "21st International Triticeae Mapping Initiative". Mexico City. September 2011

* Research internship at the International Maize and Wheat Improvement Center (CIMMYT) on "Validation of molecular markers associated with wheat (Triticum aestivum) quality". El Batán. State of Mexico. August 2010.

* Participation at the "XLI Annual Theoretical-Practice Course in Human Genetics". Faculty of Chemistry - UNAM. Mexico City. June 2009.

* Research internship at the National "Ignacio Chavez" Cardiology Institute on "Developing of a new method for cholesterol quantification in the high-density subunits of lipoproteins (HDL)". Mexico City. January 2009.

* Research internship at UNAM's Faculty of Chemistry on "Synthesis of organic heterocyclic compounds". Mexico City. June 2008.

* Research internship at UNAM's Nuclear Science Centre on "Synthesis of smart polymers through gamma radiation". Mexico City. January 2008.