Cyclops response elements in the evolution and function of root endosymbioses in *Lotus japonicus*

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

---

Xiaoyun Gong

München, September 2021

1. Gutachter: Prof. Dr. Martin Parniske
2. Gutachter: Prof. Dr. Nicolas Gompel

Tag der Abgabe: 02. September 2021
Tag der mündlichen Prüfung: 10 Dezember 2021

Eidesstattliche Erklärung

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

München, den Jan 20, 2022 Xiaoyun Gong

(Unterschrift)

Erklärung

Hiermit erkläre ich, *

☒ dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist.

☒ dass ich mich anderweitig einer Doktorprüfung ohne Erfolg **nicht** unterzogen habe.

☐ dass ich mich mit Erfolg der Doktorprüfung im Hauptfach .....

und in den Nebenfächern .....

bei der Fakultät für ..... der .....

(Hochschule/Universität)

unterzogen habe.

☐ dass ich ohne Erfolg versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.

München, den Jan 20, 2022 Xiaoyun Gong

(Unterschrift)

*) Nichtzutreffendes streichen

Summary

Plant growth is dependent on sufficient supply of nitrogen, an essential component for important macromolecules such as proteins and nucleic acids. A small group of plants belonging to four orders - Fabales, Fagales, Cucurbitales and Rosales (the FaFaCuRo clade) - have evolved the ability to engage in a mutually beneficial interaction, namely the nitrogen-fixing root nodule symbiosis (RNS). Fixed nitrogen is supplied by the bacterial symbiont to the plant host, hence helping the host to overcome nitrogen limitation. RNS development is a complex procedure that involves massive transcriptional reprogramming carried out by a cohort of *cis*- and *trans*-acting regulators. The evolutionary steps leading to the emergence of RNS have been at the center of interest for decades. Understanding the main genetic differences between plants that can form RNS and those that cannot is assumed to provide the key for installing RNS in important crop plants that are currently unable to engage in RNS.

Evolution of *cis*- and *trans*-acting elements has played key roles in the evolution of novel biological traits. This work focussed on two *cis*-regulatory elements that have played distinct roles in the evolution and maintenance of RNS. One *cis*-element, *PACE* (*Predisposition Associated Cis-regulatory Element*) was identified *via* a phylogenomic approach. *PACE* was discovered to be exclusively present in species within the FaFaCuRo clade in the promoter of the *Nodule Inception* (*NIN*) gene that encodes a master transcription factor (TF) positioned at the top of the transcriptional regulatory hierarchy specific for RNS. *PACE* confers responsiveness to bacterial signals and dictates gene expression in cortical cells forming infection threads (ITs), a tube-like plant-derived structure through which bacteria enter the root. *PACE* is essential for restoring IT formation in the *Lotus japonicus nin-15* mutant even when engineered into the *NIN* promoter of tomato, a species outside of the FaFaCuRo clade. *PACE* contains the binding site of a TF Cyclops that is indispensable for transcriptional rewiring during RNS as well as for the evolutionarily older arbuscular mycorrhizal symbiosis. *PACE* confers transactivation mediated by Cyclops in combination with the Calcium and Calmodulin-dependent protein kinase (CCaMK). These results suggested that *PACE* allows the induction of *NIN via* the symbiosis-induced signalling cascade common for RNS and AM. The phylogenetic restriction of *PACE* is congruent with that of RNS and consistent with an emergence in the last common ancestor of the FaFaCuRo clade.

A related, yet functionally distinct *cis*-element was identified in the promoter of the *Calcium Binding Protein 1 (CBP1)* gene utilising the transgenic line T90 that originated from a promoter tagging program of *L. japonicus*. T90 carries a promoterless *GUS* gene that is specifically induced during RNS and AM. Dissection of the regulatory region of the T90 *GUS* gene led to the identification of one *cis*-regulatory element required for reporter expression in the epidermis and a second element, *CYC-RE*$_{CBP1}$ (*Cyclops response element in the CBP1 promoter*), necessary and sufficient for transactivation mediated by CCaMK/Cyclops and driving gene expression during both AM and RNS. The lack of *GUS* expression in three T90 *white* mutants that were identified from an ethyl methanesulfonate-mutagenised T90 population could be traced to DNA hypermethylation detected in and around *CYC-RE*$_{CBP1}$. Two additional regulatory regions also impact *CBP1* expression. This work showcases Cyclops response elements as an essential building block for engineering RNS in crops, with a long-term goal of reducing agricultural fertiliser application.

List of abbreviations

aa	Amino acid
AM	Arbuscular mycorrhiza
A. arabidopsis & *At*	*Arabidopsis thaliana*
A. glutinosa	*Alnus glutinosa*
A. rhizogenes	*Agrobacterium rhizogenes*
AMF	Arbuscular mycorrhizal fungus
ASL18a/LBD16	Asymmetric leaves 2-like 18a/Lob-domain protein 16
BF	brightfield
BLAST	Basic local alignment search tool
bp	Base pair
C-terminus	Carboxyl-terminus
C. arietinum	*Cicer arietinum*
C. glauca & Cg	*Casuarina glauca*
Ca^{2+}-ATPase	Ca^{2+}-dependent adenosine triphosphatase
CaM	Calmodulin
CBP1	Calcium binding protein 1
CCaMK	Calcium-Calmodulin dependent kinase
CE	*Cytokinin element*
CEP	C-terminally encoded peptides
ChIP-seq	Chromatin immune-precipitation sequencing
CLE	Clavata3/embryo-surrounding region related
CLE-RS2	CLE-Root signal 2
CLSM	Confocal laser scanning microscopy
CNGC15a,b,c	Cyclic nucleotide-gated channels 15 a , b & c
CRE1	Cytokinin response 1
CYC-RE	Cyclops response element
D. drummondii & *Dd*	*Dryas drummondii*
D. glomerata & *Dg*	*Datisca glomerata*
D. trinervis	*Discaria trinervis*
DNA	Deoxyribonucleic acid
*Ds*Red	*Discosoma spp.* red fluorescent protein
EMS	Ethyl methanesulfonate
ENOD11	Early nodulin 11
ERN1	ERF required for nodulation 1
EPS	Exopolysaccharides
EPR3	EPS reporter 3
FaFaCuRo	Fagales, Fabales, Cucurbitales and Rosales
GFP	Green fluorescent protein
GUS	β-Glucuronidase
IPN2	Interacting protein of NSP2
IT	Infection thread

J. regia & *Jr*	*Juglans regia*
kb	Kilobase
L. japonicus & *Lj*	*Lotus japonicus*
LB	Luria-Bertani broth
M. loti	*Mesorhizobium loti*
M. truncatula & *Mt*	*Medicago truncatula*
NOOT1	Nodule root 1
Mya	Million years ago
N-terminus	Amino-terminus
N. benthamiana	*Nicotiana benthamiana*
NCBI	National center for biotechnology information
NF	Nod factor
NFR1	Nod-factor receptor 1
NFR5	Nod-factor receptor 5
NFRe	Epidermal Nod-factor receptor
NiCK4	NFR5-interacting cytoplasmic kinase 4
NIN	Nodule inception
NLP	NIN-like protein
NPL	Nodulation pectate lyase
NRD	Nitrate-responsive domain
NRSYM1	Nitrate unresponsive symbiosis 1
NSP1	Nodulation signalling pathway 1
NSP2	Nodulation signalling pathway 2
O. sativa & *Os*	*Oryza sativa*
P. andersonii	*Parasponia andersonii*
P. patens	*Physcomitrella patens*
P. persica & *Pp*	*Prunus persica*
PACE	Predisposition associated *cis*-regulatory element
PB1	Phox and Bem1 domain
PCR	Polymerase chain reaction
PIT	Pre-infection thread
PPA	Pre-penetration apparatus
R. irregularis	*Rhizophagus irregularis*
RAM1	Reduced arbuscular mycorrhization 1
rcf	Relative centrifugal force
RINRK	Rhizobial infection receptor-like kinase1
RNA	Ribonucleic acid
RNS	Nitrogen-fixing root nodule symbiosis
RPG	Rizhobia-directed polar growth
rpm	Revolutions per minute
RT	Room temperature
S. lycopersicum & *Sl*	*Solanum lycopersicum*
SCARN	Suppressor of cAMP receptor defect
SCR	Scarecrow
SHR	Shoot root

SIP1	SymRK-interacting protein 1
SNARE	Soluble N-ethyl maleimide sensitive factor attachment protein receptor
spp.	Species
STARR-seq	Self-transcribing active regulatory region sequencing
SymRK	Symbiosis receptor-like kinase
TF	Transcription factor
tRNA	Transfer RNA
TY	Tryptone yeast extract
UTR	Untranslated region
VAMP	Vesicle-associated membrane proteins
YFP	Yellow fluorescent protein
Z. jujuba & *Zj*	*Ziziphus jujuba*

List of publications and manuscripts

The following publication containing data presented in this work is in preparation (1):

Cathebras, C*, **Gong, X.***, Andrade, R.A., Vondenhoff, K., Keller, J., Delaux, P.M., Griesmann, M. & Parniske, M. Acquisition of a *cis*-regulatory element in the *NIN* promoter enabled the emergence of the nitrogen-fixing root nodule symbiosis.

*these authors contributed equally to the work

The relevant data is found in section 3.1. The Data presented in Figures 8 - 11 was a result of collaborative efforts between Rosa Elena Andrade (RA), Chloé Cathebras (CC) and Xiaoyun Gong (XG). CC, RA and XG collected the data displayed in Fig. 8. CC and RA collected the data displayed in Fig. 9 & 10a. CC and XG collected the data displayed in Fig. 10b and 11. CC, RA and XG all participated in experimental design and data interpretation. CC performed the analysis that is summarised in a graphic illustration in Fig. 6g. XG prepared all of the figures. Contributors are listed in each figure legend.

The following publication containing data presented in this work has been accepted for publication by *New Phytologist* (2):

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

The relevant data can be found in section 3.2. The data displayed in Figure 14 was generated with assistance from Sarah Zeitlmayr.

Publications from which data is not presented in this work

Gong, X.*, Bräcker*, L., Bölke, N., Plata, C., Zeitlmayr, S., Metzler, D., et al. (2016). Strawberry accessions with reduced *Drosophila suzukii* emergence from fruits. *Frontiers in Plant Science* 7, 1880. doi:10.3389/fpls.2016.01880.

*these authors contributed equally to the work

Bräcker, L. B., **Gong, X.**, Schmid, C., Dawid, C., Ulrich, D., Phung, T., et al. (2020). A strawberry accession with elevated methyl anthranilate fruit concentration is naturally resistant to the pest fly *Drosophila suzukii*. *PloS One* 15, e0234040. doi:10.1371/journal.pone.0234040.

Shen, D., Xiao, T. T., van Velzen, R., Kulikova, O., **Gong, X**., Geurts, R., et al. (2020). A homeotic mutation changes legume nodule ontogeny into actinorhizal-type ontogeny. *Plant Cell* 32, 1868-1885. doi:10.1105/tpc.19.00739.

Table of content

List of figures

1. INTRODUCTION

1.1 Plant endosymbiosis

Mutually-beneficial endosymbiotic interactions usually bring together a prokaryote or fungus and an eukaryote; the eukaryote gains access to biochemical processes that can only be executed by prokaryotes such as photosynthesis, chemosynthesis and nitrogen fixation, while the prokaryote gains a protected ecological niche (Dubilier et al., 2008; Wernegreen, 2012). In the case of fungal microsymbionts, the plant profits from their higher ability for soil exploration and more complex secondary metabolism. Endosymbiotic associations therefore allow the eukaryotic hosts to thrive in normally inaccessible habitats and are profound shapers of the ecosystem as well as critical components of the global nutrient cycling. Amongst these remarkably diverse interactions, endosymbioses that enhance essential nutrient supply to plant hosts are “self-regulating, customised and optimised fertiliser factories for plants” that can readily be taken advantage of in agriculture. An enhanced nutrient supply *via* endosymbiotic interactions benefits plant fitness and growth, and therefore directly contributes to yield.

Two types of plant root endosymbiosis, nitrogen-fixing root nodule symbiosis (RNS) and arbuscular mycorrhizal symbiosis (AM), are perhaps the most relevant for agricultural applications. RNS is formed with nitrogen-fixing soil bacteria and the association allows plants to grow in nitrogen-limited soils due to the supply of ammonium derived from air dinitrogen by the bacterial symbiont (Udvardi and Day, 1997). AM on the other hand relies on the vast fungal hyphal network to deliver nutrients such as phosphate, nitrogen and water from a larger soil area than accessible by plant roots (Garcia et al., 2016). In exchange, the microsymbionts receive photosynthetically fixed carbon from host plants and additionally lipids for the AM fungi (AMF) (Pimprikar and Gutjahr, 2018). The two types of endosymbiosis greatly benefit the plant hosts because phosphate and nitrogen are essential for plant growth as basic building components for macromolecules such as proteins and lipids (Bowler et al., 2010). Yet both nitrogen and phosphate are often the most limiting nutrients, due to a limited uptake capability of plants and different sources of loss from soil. As a consequence, nitrogen and phosphate are regularly applied in substantial amounts in the form of chemical fertiliser to support plant growth and ensure a good agricultural yield; and the fertiliser use is predicted to continuously increase in the future (Bouwman et al., 2013; Fowler et al., 2013).

The problems brought by the fertiliser application in current amounts are multi-fold and long-term (San Martín, 2020). First, industrial fertiliser production relies on non-renewable resources: one main ingredient, reduced nitrogen, is generated by the Haber-Bosch process that currently requires extensive use of fossil fuels (Erisman et al., 2008); and the phosphorous ingredients come from mining of rock phosphate reserves (Tiessen, 1995). Second, a large portion of applied fertiliser can not be utilised by crops and is therefore lost. Only 35% to 75% of the applied amount is actually taken up by crops (50% for major cereal crops including rice, maize and wheat; Ladha et al., 2016). The unused fertiliser is lost predominately by leaching into surface- and groundwater pools, denitrification in the soil (nitrogen) or formation of insoluble complexes (phosphate). The adverse effect of the excessive fertiliser run-off is evident in aquatic ecosystems, where eutrophication and hypoxia consequently occur (Diaz

and Rosenberg, 2008). Third, fertiliser loading negatively impacts local natural biological nitrogen fixation (Tamagno et al., 2018). This practice antagonises the efforts to utilise biological nitrogen fixation as a sustainable source of nitrogen. Today's agriculture simultaneously faces two pressing challenges: 1) meeting the need of a growing population that requires a stable supply of food, the production of which is largely dependent on fertiliser application in agriculture and 2) achieving this goal while limiting carbon emissions and leaking of excessive amounts of nitrogenous and phosphorous fertiliser into ground water and water bodies (San Martín, 2020). In the context of imbalanced nitrogen and phosphate cycles on a global scale and the ever-growing need of food supply, the potential of employment of RNS and AM as alternative sustainable fertiliser sources has attracted increasing interests.

Means to use plant endosymbioses to improve yield have been investigated. RNS and AM have been traditionally used "as they are" in agriculture to improve soil quality and crop health. Common practices include rotational cultivation of leguminous crops to improve soil fertility and intercropping of leguminous plants (e.g., clover or alfalfa) with other crops that are unable to establish RNS (e.g., triticale; Zhao et al., 2020). To this end, the natural diversity within RNS and AM is under investigation because the symbiotic outcome of enhanced plant growth varies greatly in different microbe-host and host-environment combinations (Lanfranco et al., 2018). The development of optimised AM and rhizobial inocula for diverse types of agricultural crops that are also suitable for local growth conditions is crucial to maximise efficiency. For instance, the "N2Africa" project presents an example of scientific efforts to put symbiotic nitrogen fixation to work, specifically for legume growers in Africa (https://www.n2africa.org/home).

Since the ability to engage in RNS is limited to a small group of plant species (Fig. **1**; 1.2.1), there has long been a key interest to transfer this ability to form nitrogen-fixing nodules to important agricultural crops. Currently, amongst RNS-forming species, only leguminous plants are suited for application of RNS in agriculture. This is in stark contrast to the widespread AM formed by the majority of land plants, including the important cereal crops. To engineer RNS in a non-host is undoubtedly challenging given that RNS is a genetically complex trait. The establishment and management of RNS by the host is achieved by the actions of a cohort of regulators and underpinned by drastic transcriptional changes leading to the formation of a novel root organ (1.2.2; 1.3). Understandings of the genetic make-up and the evolutionary origin of RNS are fundamental to achieve the long-term goal of engineering RNS in important non-host crops. Up to date, the roles of almost 200 RNS-related genes have been characterised thanks to the available mutant populations (Roy et al., 2020). Advanced Omics techniques have enabled studies of genetic, proteomic and metabolic changes in response to RNS on a large scale (e.g., Demina et al., 2013; Zhang et al., 2019). The availability of a growing number of high-quality plant genomes has fuelled phylogenomic studies to investigate genome-wide changes through the evolutionary history of RNS (e.g., Delaux et al., 2015; Griesmann et al., 2018). "Are we there yet?" (Pankievicz et al., 2019). Not yet, but we are getting closer with an increasing pace.

1.2 Nitrogen-fixing root nodule symbiosis

1.2.1 Evolutionary origin of RNS

Two types of nutrient-providing root endosymbiosis promote the growth of plant hosts in nutrient limiting conditions. Interestingly, the range of plants capable of engaging in the two types of endosymbioses differs drastically (Fig. **1**): the nitrogen-fixing root nodule symbiosis (RNS or nodulation; hereafter used interchangeably) shows a much more restricted phylogenetic distribution than the arbuscular mycorrhizal symbiosis (AM). An estimate of 70 - 90% of land plant species, spanning from bryophytes to angiosperms, establish symbiotic interactions with arbuscular mycorrhizal fungi (AMF), making it perhaps the most widespread endosymbiosis on land (Parniske, 2008). Fossil records indicate that AM has existed for at least 400 million years, possibly accompanying the first plants that colonised land (Remy et al., 1994; Krings et al., 2007). On the contrary, plants engaging in RNS are restricted to a monophyletic clade that encompasses four orders, the Fabales, Fagales, Cucurbitales and Rosales, collectively referred to as the FaFaCuRo clade (alternatively referred to as nitrogen-fixation or -fixing clade in the literature; Fig. **1**; Soltis et al., 1995). Hundreds of millions of years younger than AM, the oldest fossil record of root nodules dates back to 84 million years ago (Mya; based solely on appearance of nodule-like structures; Herendeen et al., 1999), while the last putative RNS-forming common ancestor was dated back to 92 - 110 Mya (Wang et al., 2009; Bell et al., 2010). Interestingly even within the FaFaCuRo clade, occurrence of RNS is still sparse: it is only present in some genera in 10 out of the 28 families (Doyle, 2011).

This restricted and scattered distribution of RNS is in itself an intriguing phenomenon. What is so special about the members of the FaFaCuRo clade that only they evolved RNS? What caused the scattered distribution of RNS: losses of nodulation in plant lineages after its emergence, or multiple parallel gains? The evolutionary origin of RNS and the genetic mechanism leading to its emergence have been a research focus in the last decades and sparked long-lasting discussions. A number of reviews have speculated about the origin of RNS from various perspectives, for example, just to list a few, based on the phylogenetic signatures of RNS-forming plants, morphological features of nodules, conservation of gene functions and neofunctionalisation of duplicated genes (Soltis et al., 1995; Doyle, 2011, 2016; Soyano and Hayashi, 2014; Geurts et al., 2016; van Velzen et al., 2019).

The discovery that RNS-forming plants belong to a monophyletic clade elicited the notion of a predisposition event that took place in the last common ancestor of the FaFaCuRo clade (Soltis et al., 1995). The prevailing evolutionary model for RNS, **the predisposition model**, predicts a two-step process: a not-yet-characterised propensity was acquired in the last common ancestor of the FaFaCuRo clade, and subsequently independent acquisition or convergent evolution of nodulation arose in lineages scattered in the four plant orders. Hence, the predisposition model treats nodulation as a nonhomologous trait and predicts approximately eight independent parallel gains of RNS (Werner et al., 2014). The gap time

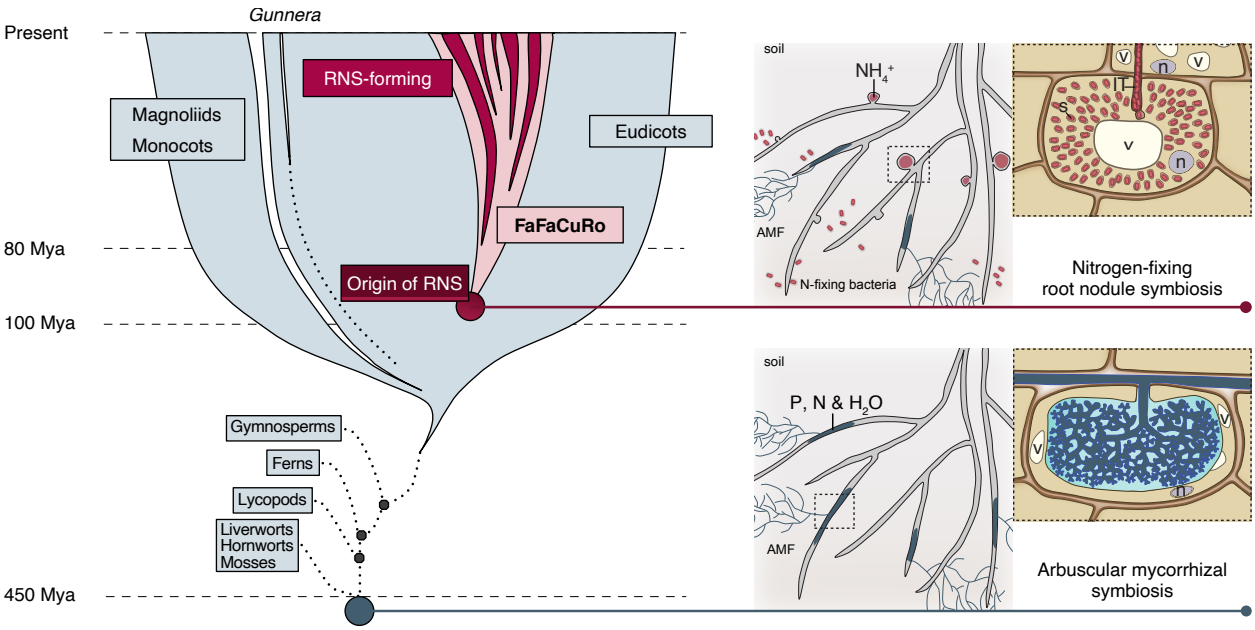


Figure 1. **Evolution of plant root endosymbioses**. Symbiotic associations with nutrient-delivering arbuscular mycorrhizal fungus (AMF) evolved approximately 450 million years ago (Mya; dark blue circle) and are present in major land plants (light blue shade). AMF hyphae invades the root cells of host plants and develop into highly branched tree-like structures called arbuscules that are enclosed in plant membranes. Nutrients, predominantly phosphate (P), but also nitrogen (N) and water (H_2O), are delivered to host plants. In contrast, nitrogen-fixing root nodule symbiosis formed with nitrogen-fixing (N-fixing) soil bacteria occurs in a restricted group of plants comprised of four plant orders, the Fagales, Fabales, Cucurbitales and Rosales (FaFaCuRo; light pink shade), that evolved ca. 100 mya (dark red circle). Within this clade, the occurrence of RNS is scattered (dark red shade). N-fixing bacteria are taken up into nodule cells *via* a tubular plant-derived structure, namely the infection thread (IT) and are either released to form symbiosomes (s) or maintained in fixation thread. Bacterial symbionts provide fixed atmospheric nitrogen (NH_4^+; ammonium) to host plants. Intracellular uptake of bacteria is only observed in the FaFaCuRo clade and in *Gunnera* spp. that host *Nostoc* cyanobacteria in stem glands. n: nucleus. v: vacuole. The phylogenetic tree is adapted from Kistner and Parniske (2002). Fixation threads are not illustrated for simplicity reason (for a comprehensive review on infection by N-fixing bacteria, Parniske 2018). The graphic illustration of infection structures of N-fixing bacteria and AMF is based on those observed in leguminous species.

between the last common ancestor of the FaFaCuRo clade and the oldest fossil records of nodules can be interpreted to mean that this important propensity (the genetic change and associated plant traits) was maintained, setting a foundation for RNS to evolve.

The strongest evidence to support this predisposition hypothesis is the vast diversity in the two essential components of RNS - infection mechanisms through which bacterial symbionts reach and are accommodated in nodule primordial cells; and nodule organogenesis that leads to the formation of root nodules to stably accommodate the microsymbionts. Phenotypic variations in infection modes and developmental and anatomical features of

nodules have been reported across the RNS-forming species. To begin with, two groups of bacteria engage in RNS: plants belonging to the order Fabales and the genus *Parasponia* (Rosales) partner with Gram-negative proteobacteria called rhizobia, whereas plants from eight families of the Fagales, Rosales and Cucurbitales host Gram-positive Actinobacteria of the genus *Frankia,* and are thus collectively referred to as the actinorhizal plants (Sprent, 2007; Santi et al., 2013). The infection mechanisms differ between plant species. The intracellular infection mode refers to a guided bacterial invasion across cells using a tubular-like plant-derived structure called infection thread (IT) within root hairs and cortical cells. This mode has been reported for plants belonging to the orders Fagales (e.g., *Casuarina glauca*) and Fabales (represented by the leguminous plants). Alternatively, bacterial partners can enter the roots through intercellular spaces between cells. For example, cluster III *Frankia* enters between root epidermal cells of actinorhizal members of the Rosales (e.g., *Discaria trinervis;* Fournier et al. 2018; Valverde and Wall, 1999). Differently, the infection of *Bradyrhizobium* on *Aeschynomene evenia* (Fabales) roots takes place at the lateral root emergence sites (Okubo et al., 2012).

Nodule morphologies vary greatly across the FaFaCuRo clade and within a plant order. Nodules can be indeterminate, such as actinorhizal nodules and those formed on *Medicago spp.* (Fabales), or determinate such as those formed on *Lotus japonicus* (Fabales). The major difference lies in the presence of a persistent apical meristematic region after the formation of a nodule primordium in the indeterminate type of nodules. Within the indeterminate nodules, the nodule vasculature can be placed in the central location as those formed by *Parasponia andersonii* (Rosales) and actinorhizal plants or in the periphery as those formed by legume plants. The evolution of peripheral vasculature is thought to be important for a symbiotic relationship with rhizobia, since they need a plant-derived protection mechanism against oxygen, the excess of which irreversibly damages the nitrogenase enzyme complex (Fay 1992). Furthermore, symbiont accommodation strategies also differ. Bacterial symbionts are either maintained in fixation threads that branch off from ITs (Behm et al., 2014) or formed after uptake of bacteria from the apoplast and typically occupy the whole cell or released as symbiosomes that structurally resemble organelles (de La Peña et al., 2018). The accommodation mechanism is dependent both on the bacterial symbiont and host species, e.g., fixation threads are formed in actinorhizal nodules that host *Frankia* as well as those formed on roots of *P. andersonii* and several species within the genus *Chamaerista* that host rhizobia. These vast phenotypic variations across plant orders and types of symbionts support the argument of the employment of distinct genetic mechanisms for independent gains of nodulation.

Recently, an opposing model, **a single origin model**, was proposed. The single origin model hypothesises that nodulation evolved only once in an ancestor of the FaFaCuRo clade and was subsequently lost in descendants, leading to non-RNS-forming FaFaCuRo lineages (van Velzen et al., 2019). In this model, the ancestral package of nodulation-essential genes evolved altogether only once in the ancestor host, making RNS a homologous trait among the descendant plant lineages. Nodulation therefore is postulated to have evolved quickly within a few million years after the divergence of the FaFaCuRo clade from its sister clade ca. 110 Mya,

since the divergence of the four orders of the FaFaCuRo clade took place ca. 104 Mya (Wang et al., 2009; Bell et al., 2010).

The hypothesis of a single origin of RNS is mainly argued from the commonality perspective. The first line of argument focuses on the similarities with regard to developmental and structural features of nodules. In *Parasponia*-rhizobia symbiosis and actinorhizal plant-*Frankia* RNS systems, symbionts are hosted in fixation threads, which are considered the "ancestral" mode of infection (Behm et al., 2014). Van Velzen and colleagues (2019) proposed that the presence of the "ancestral" form of intracellular uptake in distant RNS-forming lineages suggests a common origin. Moreover, although leguminous and actinorhizal nodules were previously thought to differ significantly from each other in terms of developmental and anatomical features, recently analyses of detailed fate maps of nodules formed on *Alnus glutinosa* (Fagales), *M. truncatula* (Fables) and *P. andersonii* (Rosales) roots revealed that the main difference between these nodules is the type of founder cells for the nodule vasculature (Shen et al., 2020). Together with the observation that a mutation in the *Nodule Root 1* (*MtNOOT1*) gene in legume *M. truncatula* partially suffices for the formation of actinorhizal-like nodules, it is proposed that actinorhizal nodules represent the ancestral type and nodules from different plant orders share more similarities than previously thought.

The second line of argument deals with commonalities in functions of symbiosis genes. The strongest evidence is perhaps the requirement of a common set of functionally conserved genes, the common symbiosis genes, in legumes, *Parasponia* and actinorhizal plants to form nodules (Gherbi et al., 2008; Markmann et al., 2008; Svistoonoff et al., 2013; Clavijo et al., 2015; Granqvist et al., 2015). The third line of argument is related to the coincidence of loss of RNS and varying degrees of mutations in one or a few RNS-essential genes in non-RNS-forming FaFaCuRo species (Griesmann et al., 2018; van Velzen et al., 2018). This discovery of the massive loss of RNS agrees with the general understanding that RNS is a genetically complex trait, hence gaining is more difficult than loss. Following this argument, parallel gains of nodulations after the acquisition of a yet-not-identified propensity (as suggested by the predisposition model) seems less probable than loss of nodulation after a single gain event.

Both evolutionary models currently lack experimental evidence that could pinpoint to the plausible genetic components leading to either the emergence of the proposed predisposition event or an ancestral nodule in the FaFaCuRo common ancestor. Nevertheless, what is clear is that the genetic changes in the last common ancestor must have enabled one or a few novel traits (e.g., an infection thread-based symbiosis without organ formation; Parniske 2018) that provided evolutionary advantages or at least neutrality. This genetic novelties and associated traits must have been maintained by descendants of the common ancestor for ca. 20 - 30 million years and are still shared by all modern RNS-forming species (Fig. **1**). A comprehensive understanding of the function of RNS-relevant genetic components and their phylogenetic distribution is necessary to uncover the potential candidates that contribute to the emergence of RNS and further discuss these two hypotheses.

1.2.2 Establishment of RNS

The establishment of nodulation can be subdivided into three major processes: pre-infection communication, bacterial infection and nodule organogenesis. The pre-infection communication assists the bidirectional selection of symbiotic partners, the success of which prompts plants to take up bacterial symbionts into their roots. The process through which bacteria invades the roots, namely bacterial infection, and the organogenesis event that leads to the formation of a nodule, namely nodule organogenesis, are temporally and spatially coordinated. This concerted action ensures formation of nodules that stably host bacterial symbionts in their intracellular space. All three processes are essential for successful formation of nitrogen-fixing nodules. Largely thanks to the discovery of plant mutants that are impaired in one or more of processes, nearly 200 genes have been identified to be involved in the modulation of cellular responses underpinning the establishment of RNS (for a list of RNS-relevant genes, see Roy et al., 2020). Functional characterisations of these genes have allowed a reconstruction of the RNS-related genetic program, which continues to evolve as the list of relevant genetic components expands.

1.2.2.1 Overview

RNS is initiated by the pre-infection chemical cross-talk between the symbiotic partners to identify a compatible pair of bacterial symbiont and plant host. Successful perception of the symbiont by the host is followed by uptake of bacteria into host roots. Amongst the diverse infection modes observed in different RNS systems, the best characterised one is the intracellular infection through root hairs in legumes. During this process, deformation and curling of root hairs to trap rhizobia are induced first, leading to the formation of a closed infection chamber. An IT is initiated from this infection chamber and elongates inwards *via* polar growth to reach the root cortex (Fournier et al., 2008, 2015). The initiation and elongation of IT are regulated by a number of plant proteins that remodel cell walls and guide the growth of the IT (Fournier et al., 2015; Liu et al., 2020a, 2020b; Roy et al., 2020 and references within; Su et al., 2020). Bacterial symbionts therefore follow the path laid by the growing IT to enter the root cortex.

Simultaneous with the elongation of the IT, the onset of organogenesis is induced, a process which in legumes is underpinned by a complex interplay of phytohormones including the core regulator cytokinin (Gamas et al., 2017; Reid et al., 2017; Tan et al., 2020). Cell divisions are induced in the root cortex and pericycle, giving rise to a nodule primordium. Some of these newly divided cortical cells are competent for stable intracellular accommodation of bacteria, and later become the central nodule tissue colonised by bacteria (van Spronsen et al., 2001; Xiao et al., 2014; Shen et al., 2020). Following the arrival of ITs in the nodule primordium, ITs ramify in the newly divided cortical cells (van Spronsen et al., 2001). Bacteria are subsequently either released from these cortical ITs leading to formation of symbiosomes, organelle-like structures in which bacteria are enclosed in plant-derived peribacterioid membrane (de La Peña et al., 2018) or maintained in fixation threads that branch off from ITs

(in several tropical legume trees, Parasponia and actinorhizal plants; Pawlowski and Bisseling 1996). After bacterial symbionts develop into the nitrogen-fixing forms, bacteroids in case of rhizobia and vesicles in most actinorhizal symbioses, biological nitrogen fixation is executed by the intracellularly-hosted bacterial symbionts (Vasse et al., 1990; Pawlowski and Demchenko 2012; Behm et al., 2014). Bidirectional exchange of nutrients occurs within the infected cells of nodules that sustain both partners: most plant hosts receive reduced nitrogen in the form of ammonium supplied by the bacterial symbionts and provide primarily dicarboxylates in return (for a detailed review see Liu et al., 2018). To balance the cost and benefits of nodulation, the number of nodules in legumes is systemically regulated through a process called autoregulation of nodulations, controlled by short peptides and associated receptors, plant hormone as well as micro RNAs (Soyano et al., 2014; Tsikou et al., 2018; Laffont et al., 2019, 2020).

1.2.2.2 Bacterial symbiont-induced signalling

The establishment process of RNS is enabled by a large network of genetic regulators that induce transcriptional reprogramming in host roots. The starting point of this transcriptional programming is the successful recognition of a compatible bacterial symbiont by the plant host. The genetic program enabling the bacterial symbiont-induced signalling is currently better understood in model legumes (Fabales) than in plants from the other three orders. In legumes, flavonoid compounds exuded by the host roots are perceived by rhizobia in the rhizosphere (Liu and Murray, 2016). In return, lipo-chitooligosaccharides, collectively referred to as Nodulation factors (NFs), are produced by the rhizobia (Gough and Cullimore, 2011) and perceived by the host roots *via* plasma membrane localised dimers of LysM receptors and LysM receptor-like kinases, NFR1, NFR5 and NFRe (Madsen et al., 2003; Radutoiu et al., 2003, 2007; Murakami et al., 2018; in this thesis, the names of the *L. japonicus* genes will be used). Additionally, bacterial exopolysaccharides (EPS) play a role in microbe recognition in the subsequent infection process (Kawaharada et al., 2015). NF-induced signalling is mediated by the Symbiosis Receptor-like Kinase (SymRK) that interacts with NFR1 and NFR5 (Endre et al., 2002; Stracke et al., 2002; Chen et al., 2012; Yan et al., 2020), as well as several other proteins (see a summary in Roy et al., 2020) to transmit microsymbiont perception to downstream responses in the nucleus (Fig. **2**, step 1). The transmission of symbiont perception at the plasma membrane to the nucleus likely involves two proteins SymRK-Interacting Protein 1 (SIP1) and NFR5-interacting Cytoplasmic Kinase 4 (NiCK4), that locate to the nucleus and nuclear envelope, respectively (Zhu et al., 2008; Wong et al., 2019).

A hallmark downstream of the NF-induced signalling is the generation of nuclear and perinuclear calcium oscillations or “spiking” (Sieberer et al., 2009). Generation of calcium spiking is facilitated by a number of nucleoporin complexes and ion channels localised in the nuclear envelope - Castor, Pollux, nucleoporins NUP85 and NUP133, NENA and Cyclic Nucleotide-Gated Channels 15a, b & c (CNGC15a,b,c) - and a nuclear envelope and endoplasmic reticulum localised Ca^{2+}-dependent adenosine triphosphatase (Ca^{2+}-ATPase) MCA8 (Fig **2**, step 2) (Ané et al., 2004; Imaizumi-Anraku et al., 2005; Kanamori et al., 2006; Saito et al., 2007; Charpentier et al., 2008, 2016; Capoen et al., 2011; see a review Kim et al.,

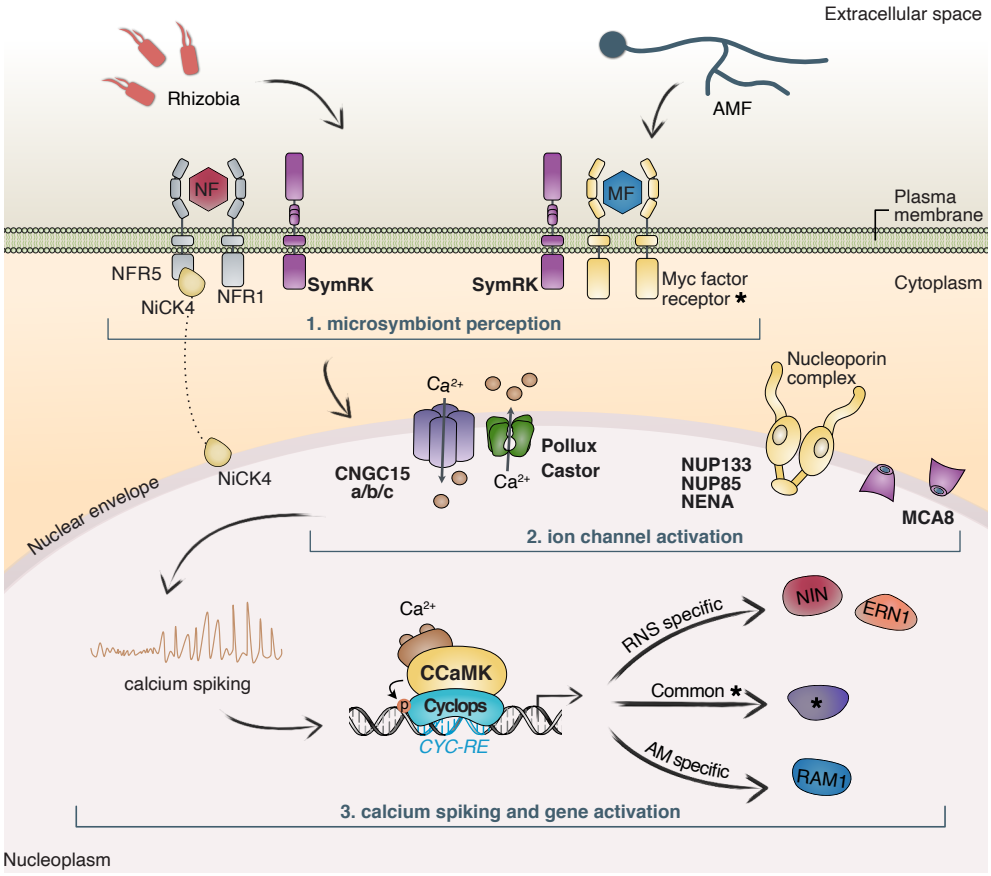


Figure 2. Overview of the microsymbiont-induced signalling and consequential transcriptional activation. The establishment of RNS and AM requires a set of common genes (encoded protein in bold). Signalling molecules exuded by rhizobia or AMF are recognised by respective plant receptors (NFR1, NFR5 or the hypothetical Myc factor receptors) at the plasma membrane and a receptor-like kinase is activated (SymRK) (1). Calcium spiking is generated as a core downstream response facilitated by ion channels (CNGC15a,b,c, Pollux, Castor) and other proteins (nucleoporin complex including NUP133, NUP85 and NENA, and MCA8) at the nuclear envelope (2). CCaMK presumably decodes the calcium signature upon activation by Ca^{2+}-binding calmodulin (CaM). Activated CCaMK interacts with and phosphorylates a transcription factor Cyclops (3). Cyclops recognises the response elements (*CYC-RE*) in the promoters of its target genes that are specifically required for RNS (*NIN* and *ERN1*) or AM (*RAM1*), or hypothetically common responses (asterisk) for both types of endosymbioses. p, phosphate group. This figure is simplified to highlight the members of common symbiosis genes and the biological process they are involved in. Proteins participating for only one type of symbiosis and some additional interacting proteins of the components illustrated in the figure are not included (refer to a timely review Roy et al., 2020 for the full list of relevant proteins).

2019). Nuclear calcium spiking is presumably decoded by a Calcium-Calmodulin dependent kinase (CCaMK), that is postulated to be activated by a Ca^{2+}-binding calmodulin (Lévy et al., 2004; Tirichine et al., 2006; Hayashi et al., 2010). Activated CCaMK interacts and subsequently phosphorylates Cyclops, a DNA-binding transcription factor (TF).

Ultimately, the NF-induced signalling leads to expression of downstream RNS-related genes to facilitate different developmental stages of RNS. Cyclops is an important player in this process as it induces the expression of two RNS-essential TFs in a phosphorylation-dependent manner (Fig. **2**, step 3): *Nodule Inception* (*NIN*) (Singh et al., 2014) and an APETALA2/ Ethylene Responsive Factor (AP2/ERF) *ERF Required for Nodulation 1* (*ERN1*) (Cerri et al., 2017). Both TFs are necessary for bacterial infection; and *NIN* also regulates nodule organogenesis.

1.2.2.3 *Nodule Inception:* a central regulatory hub

The *NIN* gene was first cloned from a transposon-tagged *L. japonicus* mutant *nin-1* that displays severe defects in nodulation (Schauser et al., 1999) and has been more and more recognised as a central hub of RNS. *NIN* encodes a TF protein belonging to a group of plant-specific RWP-RK domain containing proteins that also encompass the NIN-like proteins (NLPs) (Chardin et al., 2014). Phylogenetic analysis supports *NLP1* as the closest relative of *NIN,* resulting from a gene duplication event that occurred in the ancestor ca. 110 Mya before the divergence of Fabidae (including the FaFaCuRo clade) and Malvidae (including the Brassicales) (Schauser et al., 2005; Wang et al., 2009; Bell et al., 2010; Liu and Bisseling, 2020). The molecular structure of NIN proteins resembles a typical NLP protein that consists of three domains: the nitrate-responsive domain (NRD), DNA-binding RWP-RK domain and Phox and Bem1 domain (PB1) mediating protein-protein interaction (Schauser et al., 2005; Konishi and Yanagisawa, 2013, 2019; Lin et al., 2018). Among these three domains, functions of RWP-RK and PB1 domains for both NIN and NLPs in and outside of the FaFaCuRo clade are conserved. The NRD domain of NIN*,* despite the high sequence similarity to that of NLPs, is incompletely conserved and carries deletions of variable length. A key feature that distinguishes NIN from NLP proteins is their differential response to nitrate sensing. When sensing nitrate, the N-terminus of NLPs mediates shuffling of NLPs form the cytosol to the nucleus and post-translational activation of nitrate-induced genes. This mechanism is observed for NLPs outside of the FaFaCuRo clade, such as *At*NLP6 & 7, *Zm*NLP6 & 8 and *Os*NLP4 (Alfatih et al.; Konishi and Yanagisawa, 2013; Cao et al., 2017; Wu et al., 2020), as well as inside such as *Lj*NLP4/NRSYM1 and *Mt*NLP1 (Lin et al., 2018; Nishida et al., 2018). On the contrary, nuclear localisation of NIN and the induction of nitrate-induced genes by NIN is independent of nitrate concentration, a fact likely caused by deletions in NRD domain of NIN (Suzuki et al., 2013). This loss of the response to nitrate might have been a key step during the exaptation of NIN function for RNS (Suzuki et al., 2013; Soyano and Hayashi, 2014).

NIN functions downstream of NF-signalling and is specifically required for RNS but not AM (Schauser et al., 1999; Kumar et al., 2020). The *nin* mutant displays normal morphological features and can successfully establish an AM symbiosis. When inoculated with rhizobia, the *nin* mutant exhibits excessive root hair curling/deformation in an expanded susceptible zone of the root, and formations of ITs or nodule primordia are not observed (Schauser et al., 1999; Marsh et al., 2007; Clavijo et al., 2015). *NIN* is amongst the earliest induced genes during rhizobial infection and directly activates expression of genes required for IT formation: *Nodulation Pectate Lyase* (*NPL*) encoding a pectate lyase*, Rizhobia-directed Polar Growth*

(*RPG*), *Nuclear Factor-YA1* (*MtNF-YA1*), *EPS Reporter 3* (*EPR3*) encoding a LysM type polysaccharide receptor, *Rhizobial Infection Receptor-like Kinase 1* (*RINRK1*) encoding an atypical leucine-rich repeat receptor-like kinase as well as *SCAR-Nodulation* (*SCARN*) encoding a Suppressor of cAMP Receptor defect (SCAR) protein (Arrighi et al., 2008; Xie et al., 2012; Laporte et al., 2014; Kawaharada et al., 2015; Qiu et al., 2015; Li et al., 2019). Loss of function of these genes results in similar defects of RNS: failure in initiating IT growth despite the formation of infection loci and/or arrested IT growth within root hairs; as well as abnormal progression of ITs in root hairs or in the occasionally-formed nodules (only in some mutants). Additionally, the pectate lyase NPL is secreted into the infection chamber and tightly accompanies the growing IT (Liu et al., 2019a), presumably engaging in cell wall modification. A recent study transcriptionally profiled the *M. truncatula nin-1* mutant at the onset of IT development and identified downregulated expression of 63 genes presumably involved in cell-wall modification, including *NPL* (Liu et al., 2019a). Moreover, *NIN* expression is co-ordinately mediated by *Cyclops* and *ERN1*, another TF essential for IT formation (Liu et al., 2019d).

Another role of *NIN* is to regulate nodule organogenesis. Cytokinin signalling is necessary and sufficient for the organogenesis program; exogenous application of a synthetic cytokinin (6-Benzylaminopurine) is sufficient to induce *NIN* expression (Murray et al. 2007; Tirichine et al. 2007; Heckmann et al. 2011). The cytokinin-mediated activation of *NIN* is modulated in part *via* activation of a GRAS domain TF *Nodulation Signalling Pathway 2* (*NSP2*) that in turn induces *NIN* expression in a heterodimer with *NSP1* (Hirsch et al., 2009; Ariel et al., 2012). *NIN* directly binds to the promoter of *Cytokinin Response 1* (*CRE1*) that encodes a cytokinin receptor, and transcriptionally activates its transcription (Vernié et al. 2015), creating a positive feedback loop for cytokinin induced signalling. In the legume *M. truncatula*, a distant regulatory region containing several putative cytokinin response elements, the *Cytokinin Element* (*CE*), was recently reported to be essential for *NIN* to regulate nodule organogenesis, although whether *CE* directly mediated cytokinin-induced *NIN* expression is yet to be clarified (Liu et al., 2019c). Both *CE* and the 5 kb proximal promoter of the *NIN* gene are required for successful restoration of infection and nodule organogenesis in the *Mtnin-1* mutant. *LjNF-YA1* and *LjNF-YB1* as well as *Asymmetric Leaves 2-like 18a/Lob-Domain protein 16* (*ASL18a/LBD16*) are direct transcriptional targets of *NIN,* that interact with each other and control cell division *via* regulation of auxin biosynthesis genes (Soyano et al., 2013, 2019; Schiessl et al., 2019). Ectopic expression of ASL18a/LBD16 and NF-Ys is sufficient to induce spontaneous nodules in *L. japonicus* wild-type and *nin-9* mutant; a similar phenomenon is observed when *NIN* is ectopically expressed (Soyano et al. 2013 & 2019). Additionally, a SHORT ROOT-SCARECROW (SHR-SCR) TF pair is reported to modulate cell division and nodule organogenesis in *M. truncatula*. Accumulation of SHR protein is likely dependent on *NIN* as a reduced protein accumulation was observed in the *nin-1* mutant (Dong et al., 2020). The induction of *LBD16* during nodulation is dependent on *SHR* and *SCR*, and that of *SCR* is dependent on *LBD16.* Therefore, *NIN* likely oversees the two interdependent modules for nodule organogenesis.

Moreover, *NIN* plays dual antagonistic roles in the regulation of nodule numbers. On one hand, *NIN* is involved in the program of autoregulation of nodulation to restrict nodule numbers.

CLAVATA3/embryo-surrounding region related (CLE) peptides are reported to function as the mobile root-to-shoot signal to initiate the negative regulation on nodulation (for a detailed review, see Nishida and Suzaki, 2018). *NIN* induces the expression of the CLE-peptide encoding genes as early as during the initiation and formation of ITs in root hairs in *L. japonicus* (Soyano et al. 2014; Laffont et al., 2020). NIN interacts with a NIN-like protein SYMRM/NLP4 that senses nitrate concentrations and regulates the expression of *CLE-Root Signal 2* (*CLE-RS2*) to control nodule numbers (Lin et al., 2018). On the other hand, the production of C-terminally encoded peptides (CEPs) which promotes nodulation is also under control of *NIN* (Laffont et al., 2020). Together, these two *NIN*-dependent antagonistic processes maintain the homeostasis of RNS.

The role of *NIN* in nodulation likely extends beyond the previously discussed processes. It has been reported that *NIN* controls symbiosome formation process based on the observations that two weak *M. truncatula nin* mutant alleles are able to form rather normal infection some in the nodules, however these nodules experience premature senescence and arrested symbiosome development (Liu et al., 2021). Transcriptomic data of *Mtnin-1* suggests potential involvement of *NIN* in plant defence responses and nutrient transport, based on the observation of downregulation of the genes in the mutant (Liu et al., 2019a). The sucrose transporter SWEET11 is active in not-yet-infected and symbiosome-containing cells in the nodules formed on *M. truncatula* roots and its expression depends on *NIN* (Kryvoruchko et al. 2016). Moreover, although *NIN* does not directly activate genes encoding nitrate transporters, ectopic expression of *NIN* antagonises the nitrate-dependent induction of these genes (Soyano et al., 2015). Taken together, *NIN* is a central regulation hub, i.e., it coordinates several aspects of RNS including but not restricted to bacterial infection, nodule primordia formation, regulation of nodule numbers and symbiosome formation.

1.3 Transcriptional rewiring for the evolution of root nodule symbiosis

Indubitably, the establishment of RNS is a complex process that requires a large network of regulators (1.2.2). How these regulators attained a function in RNS is a fascinating and important research subject. For instance, the question arises whether these regulators did evolve specifically for the emergence of RNS or were “adopted” from other existing processes and repurposed for RNS? What are the elemental genetic and molecular qualities of these regulators to enable RNS, regardless whether these qualities evolved directly or were adapted? Was is it a novel invention or adoption of one regulator that contributes to a large extent to the emergence of RNS or was is it an accumulative effect of multiple regulators, each of which contributed one specific aspect? Studying how each individual regulator and their entire community were recruited for RNS simultaneously provides insights into the evolutionary pathway leading to the emergence of RNS and the mechanistic requirements for their RNS-related regulation and function.

A recent study searched for gene gains specific to the FaFaCuRo clade *via* genome-wide comparative analysis of 37 plant species that covered a diversity of infection modes and symbiont types (Griesmann et al., 2018). It was hypothesised that if one or several novel genes (“a genetic toolbox for RNS”) acquired in the last common ancestor were essential for

emergence of nodulation, these genes must have been maintained and should be exclusively present in RNS-forming FaFaCuRo members. Surprisingly, no genes matching the aforementioned evolutionary pattern were discovered, indicating that gain of gene(s) might not have been associated with the origin of root nodule symbiosis. Instead, a gene-loss pattern was observed: variable degrees of fragmentations and loss of RNS-essential genes, namely *NIN* and *RPG*, correlate with loss of nodulation. A similar gene-loss pattern was reported from another study comparing genomes of three *Parasponia* species and five *Trema* species that are close relatives of *Parasponia* but unable to engage in RNS (van Velzen et al., 2018). That no such genetic toolbox could be detected brought up the question that had been already asked by Markmann and Parniske (2009) over 10 years ago: “how novel are nodules?”

Indeed, several genetic components play dual roles in establishment of RNS and other biological processes such as establishment of AM, lateral root development, microbe recognition, plant defence and nutrient transport. These observations have led to the hypothesis that repurposing of already functional regulatory networks *via* transcriptional rewiring was an important driver for the emergence of RNS. Transcriptional rewiring of pre-existing machineries contributes significantly to phenotypic novelties and variations and is remarkably dynamic. Examples of evolutionary studies in other organisms have revealed that rewiring of regulatory networks could be achieved by changes in *cis*-acting elements such as enhancers (Jeong et al., 2006; Rebeiz and Williams, 2017); or *trans*-acting elements such as TFs (Lynch and Wagner, 2008; Wagner and Lynch, 2010); or concerted changes in both (Britton et al., 2020). Compared to the knowledge of the function of the *cis*- and *trans*-acting elements, relatively little is known about the evolutionary mechanisms through which they were rewired for the evolution of RNS. It is possible to peek at their evolutionary history of functional adaptations for RNS by studying their phylogenetic distribution, roles in biological processes and associated regulatory networks. For the sake of discussion, these regulatory elements are sorted into three main realms/processes that are fundamental to forming a nitrogen-fixing nodule: microsymbiont induced signalling, intracellular uptake of symbiont (bacterial infection) and nodule organogenesis. The latter two often act in concert, but are genetically separable (Madsen et al., 2010), hence are treated here as two independent topics.

1.3.1 Microsymbiont-induced signalling: trigger downstream transcriptional changes

An important discovery supporting the hypothesis of transcriptional rewiring is that a common set of genes induced by microsymbiont-exuded signals orchestrates infection and intracellular accommodation of both AMF and nitrogen-fixing bacteria (Kistner and Parniske, 2002; Kistner et al., 2005). This set of genes is therefore referred to as “the common symbiosis genes” (Fig. **2**). Members of the common symbiosis genes mediate distinct biological processes including: (1) perception of the microsymbionts (*SymRK*); (2) generation of calcium oscillation as the messenger to induce downstream response (*Castor*, *Pollux*, *NUP85*, *NUP133* and *CNGC15a,b,c*); and (3) transcriptional activation of symbiosis-related genes to stably accommodate the microsymbionts in living plant cells (*CCaMK* and *Cyclops*).

SymRK from different dicot species and from the monocot *Oryza sativa* (rice; Poales) can restore AM in a *L. japonicus* (Fabales) *symrk* mutant (Markmann et al., 2008). Likewise, *Castor, Cyclops* and *CCaMK* are indispensable for AM in rice and can rescue both RNS and AM in respective mutants of legume plants (Chen et al., 2007, 2008; Banba et al., 2008; Markmann et al., 2008; Yano et al., 2008). Remarkably, *CCaMK* and *Cyclops* from the AM-forming liverwort *Marchantia paleacea* can successfully restore AM in the corresponding *M. truncatula* mutants, although two plant species are separated by more than 450 million years of diversification (Radhakrishnan et al., 2020). Furthermore, the function of these common symbiosis genes is conserved across the FaFaCuRo clade in different RNS systems. *SymRK* from the actinorhizal plants *C. glauca* (Fagales) and *D. glomerate* (Cucurbitales) can restore both RNS and AM in a *L. japonicus symrk* mutant (Fabales) (Gherbi et al., 2008; Markmann, 2008). Similarly, *CCaMK* from the actinorhizal plants *Discaria trinervis* (Rosales) and *C. glauca* can rescue nodulation in a *M. truncatula ccamk* mutant (Svistoonoff et al., 2013; Radhakrishnan et al., 2020).

Although the roles of actinorhizal homologs of several common symbiosis genes have not yet been demonstrated in detail, the functional equivalence of the common symbiosis genes across two types of endosymbiosis that evolved approximately 350 million years apart, distant plant lineages and different RNS systems strongly supports the rewiring of an evolutionary older mechanism (functional in AM) for a new purpose (RNS).

1.3.2 Intracellular uptake of bacterial symbiont: a central feature of RNS

In spite of the great diversity in infection mechanisms and nodule organogenesis and morphology, one and perhaps the only shared feature emerges among the RNS-forming lineages: the uptake of bacteria into living plant cells (Parniske, 2018). Diazotrophic microbes commonly coexist with plants as endophytic or associative organisms, which promote plant health by performing nitrogen fixation and modulating plant growth (Carvalho et al., 2014). Comparatively, the more intimate association - hosting the microbes in the intracellular space - provides a more efficient exchange of nutrients and controlled selection of symbionts (Gage, 2004; Carvalho et al., 2014). Nevertheless, intracellular accommodation of bacteria in living cells is a rare phenomenon among land plants, observed only in the FaFaCuRo clade and one exception outside of this clade, *Gunnera* spp. (Gunnerales) that host diazotrophic cyanobacteria *Nostoc* in stem glands (Johansson and Bergman, 1992). It is hence argued by Parniske (2018) that given the uniqueness and rareness of intracellular uptake of bacteria into living cells, this trait must have been acquired by the last common ancestor of the FaFaCuRo clade and thereby been maintained in the descendant lineages.

Curiously, an analogy can be drawn between endosymbiotic interactions with two distinct types of symbionts, rhizobia and AMF, in that the infectious unit is always surrounded by a plant-derived peri-microbial membrane separating the microbe from the cytoplasm of the plant cell during the course of infection and stable accommodation (Parniske, 2000). It was therefore postulated by Parniske (2000) that an “intracellular accommodation program” is in place to mediate this process. Since then, several lines of evidences were found that support this hypothesis. First, early cellular responses including relocation and movement of nuclei in cells

anticipating and undergoing infection has been observed during the infection process by AMF and rhizobia (Fåhraeus, 1957; Timmers et al., 1999; Sieberer and Emons, 2000; Gage, 2004; Genre et al., 2005; Fournier et al., 2018). Second, formation the membrane-matrix structure composed of cytoskeletal and cytoplasmic aggregations in cells anticipating infection, namely the pre-infection thread (PIT) structures for RNS and the pre-penetration apparatus (PPA) for AM, is suggested to be associated with cell-cycle and cell division related mechanisms in both endosymbioses (van Brussel et al. 1992; Yang et al., 1994; Breakspear et al., 2014; Smertenko et al., 2017; Russo et al., 2019). Third, in later developmental stages when the peri-microbial membrane is produced to surround to symbiont, specific proteins functioning in membrane trafficking pathways are essential for establishment and maintenance of the host-microbe interface for both types of endosymbiosis. These members include VAPYRIN (Murray et al., 2011), SNAREs (Huisman et al., 2016; Sogawa et al., 2019) and VAMP72s (Ivanov et al., 2012). How these common components were incorporated to RNS by plants of the FaFaCuRo clade is not yet known. It is possible that an upstream central regulator governing the expression of these components was recruited, or alternatively independent recruitment of them occurred in the FaFaCuRo clade.

Furthermore, it is perhaps suitable in this context to mention that the common symbiosis gene *Cyclops*, plays a crucial role in intracellular accommodation of microsymbionts during both AM and RNS (Yano et al., 2008). Upon infection by its compatible symbiont *Mesorhizobium loti*, the *L. japonicus cyclops* mutant fails in the initiation of ITs although bacteria are entrapped by curled root hairs. Although nodule organogenesis is induced in this mutant, the nodules are devoid of bacterial infection. This mutant also fails to intracellularly host AMF in its root cortex. Its importance for intracellular accommodation of symbionts is indirectly supported by the finding of a positive correlation between the presence of *Cyclops* (and its interacting and phosphorylating partner *CCaMK*) and the ability to intracellular accommodate symbionts across a wide range of land plants (Radhakrishnan et al., 2020). It therefore seems reasonable that the co-option of *Cyclops* might have enabled the FaFaCuRo ancestor to extend its “residents” from AMF to nitrogen-fixing bacteria.

1.3.3 Nodule organogenesis: production of a novel organ

Nodule organogenesis is a coordinated consequence involving two main processes: 1) cell dedifferentiation, division and proliferation and 2) the ultimate step leading to the formation of a novel organ. However, both processes are not unique to RNS. In fact, organogenesis occurs frequently in the above- and belowground parts of plants as response to prevailing and changing environmental conditions, such as modification of the root architecture in response to nutrient availability in soil (Luo et al., 2020) and formation of the invasion structure, haustoria, during infection by biotrophic fungi or oomycetes (Ichihashi et al., 2020).

On roots, lateral root formation is a frequently occurring organogenesis event to modify root architecture. Interestingly, actinorhizal and *Parasponia* nodules bear high resemblance to lateral roots, possessing a pericycle-derived central vasculature adjacent to cortex-derived tissue and an apical meristematic region (Pawlowski and Demchenko, 2012; Shen et al., 2020).

Leguminous nodules are in comparison less similar to later roots, displaying a peripheral vasculature and in some species lacking a persistent apical meristem (e.g., in *L. japonicus* nodule). The discovery of several shared components functioning in both lateral root development and RNS suggests the repurposing of the lateral root program for nodulation. First, cytokinin is a core phytohormone promoting nodule organogenesis (Gamas et al., 2017); it also regulates lateral root development in a complex interplay with other hormones such as auxin, ethylene and gibberellin (Qin and Huang, 2018; Jing and Strader, 2019). Downregulation of *CRE1* in *M. truncatula* rendered the roots insensitive to cytokinin and significantly increased lateral root density, while negatively impacting nodulation (Gonzalez-Rizzo et al., 2006). Second, a *NIN*-dependent *ASL18a/LBD16*-mediated pathway was recently found to control both nodule organogenesis and lateral root development. Ectopic expression of *ASL18a/LBD16* alone in *L. japonicus* roots induces ectopic formation of lateral root primordia, and together with its interacting partners, NF-Ys, leads to spontaneous nodule formation (Soyano et al., 2019). These reports highlight the overlaps in the genetic control of the two different organogenetic processes. Moreover, the role of theses shared regulators in two organogenetic processes further implies the evolution of additional control of organ identity. In *M. truncatula,* the *NOOT* gene maintains the identity of the indeterminate nodule (Magne et al., 2018). The *Mtnoot1noot2* mutant displays a dramatic phenotype where a nodule transitions into a lateral root that originates from the apical nodule meristem (Shen et al., 2020).

It is important to note that the NIN-*ASL18a/LBD16* connection appears to be specific for legumes. Together with the observation that the response to cytokinin treatment in actinorhizal plants differs from that of legumes (Gauthier-Coles et al., 2019), this indicates that nodule organogenesis is regulated through distinct mechanisms in legumes and actinorhizal plants. A gene that appears to regulate nodule organogenesis for both leguminous and actinorhizal plants is *CCaMK*. The deregulated *C. glauca* CCaMK protein that lacks the auto-inhibitory domain triggers the organogenesis program resulting in spontaneous nodule formation on *D. trinervis* and *C. glauca* roots in the absence of the symbiont, a phenomenon that have been also observed in leguminous plants (Gleason et al., 2006; Hayashi et al., 2010; Svistoonoff et al., 2013). Activation of organogenesis can be independent from Cyclops, as deregulated CCaMK is capable of inducing spontaneous nodules in a *cyclops* mutant (Yano et al., 2008). Altogether, these findings suggest a close evolutionary relationship between lateral root development and nodules, and CCaMK appears to present as a central regulator of nodule organogenesis across the FaFaCuRo clade, although the recruitment mechanism for CCaMK is not yet understood.

1.4 *Cis*-regulatory elements in root nodule symbiosis

1.4.1 *Cis*-elements for the establishment of RNS

As previously introduced, it is evident that the establishment and maintenance of RNS is accompanied by dramatic local and systemic transcriptional reprogramming, involving a large network of regulatory proteins that modulate a wide variety of biological processes. A transcriptional change is the result of the combined activity of *trans*-acting regulatory elements

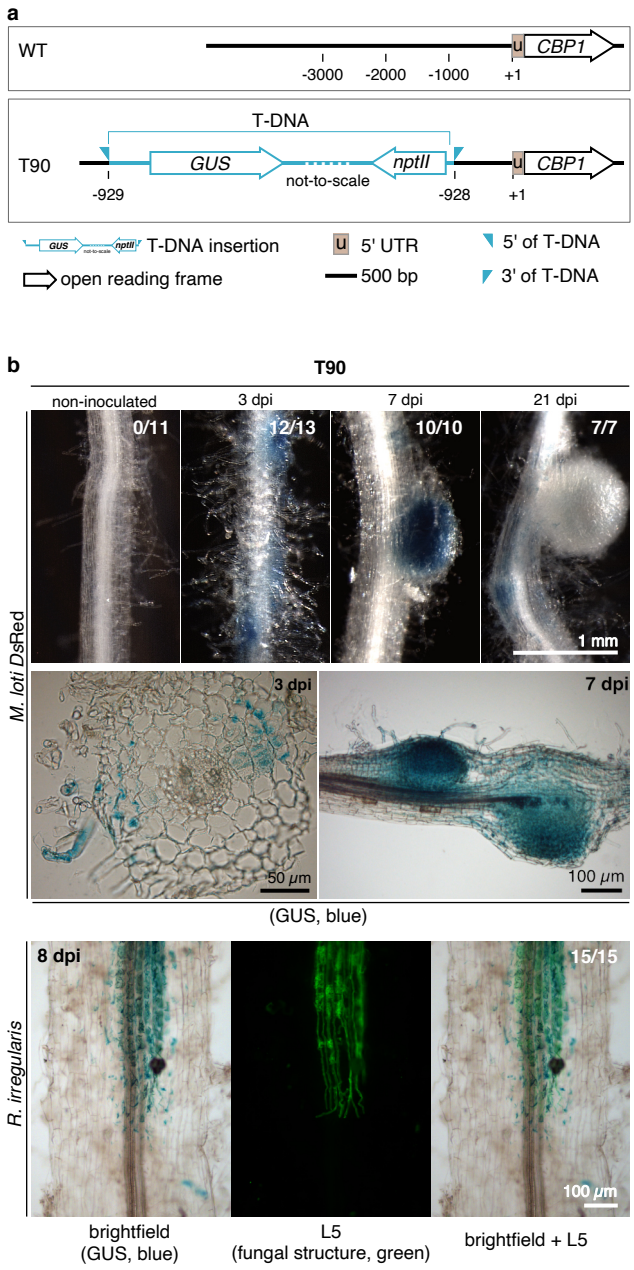


Fig. 3 *GUS* expression pattern in T90 roots. a, Position of the T-DNA insertion within the promoter of *CBP1* gene in the transgenic line T90. The respective region in *L. japonicus* ecotype Gifu is shown as a reference. **b**, T90 roots were stained with X-Gluc to reveal a blue coloration generated by GUS enzyme activity at indicated days post inoculation (dpi) with *M. loti Ds*Red or AMF *Rhizophagus irregularis*. Note the blue staining at 3 dpi in patches in root hair cells, the presence and absence of blue staining in the central nodule tissue at 7 dpi and 21 dpi, respectively. At 21 dpi, blue staining could be seen in the inner tissue of developing nodules, in contrast to mature nodules that remained white. Green, Alexa Fluor-488 WGA-stained *R. irregularis* visualised with a Leica Filter Cube L5 next to a brightfield image of the same root segment. #/# in **b**: number of plants displaying GUS activity / total number of plants analysed.

including regulatory proteins and non-coding regulatory RNAs, and *cis*-acting regulatory elements, such as enhancer, promoter and untranslated regions of mRNA (Ong and Corces, 2011; Spitz and Furlong, 2012; Kim and Shiekhattar, 2015). Together, *cis*- and *trans*-elements define the landscape of the regulatory networks (e.g., see more about gene regulatory network Peter & Davidson, 2017). Each type of these components has been demonstrated as important regulatory units for RNS. Amongst the *trans*-acting elements, TFs are key regulatory proteins that govern the expression of a set of RNS-related genes for distinct developmental processes and make up a considerable portion of currently known RNS-related players (Roy et al., 2020). *Trans*-acting micro RNAs produced by plant hosts and rhizobial tRNA-derived small RNAs that act post-transcriptionally to modulate gene expression level have also been demonstrated as important players in RNS (for timely reviews, see Ren et al., 2019; Valdés-López et al., 2019). For example, miR2111 post-transcriptionally controls the expression level of the *Too Much Love* gene that encodes a repressive F-box protein to balance infection and nodule formation (Tsikou et al., 2018). The advent of Omics technologies, either alone such as transcriptomics or in combinations of other Omics, such as proteomics and transcriptomics, have greatly

accelerated the understanding of the regulatory circuitry mediated by these *trans*-acting elements (for a summary, see Mergaert et al., 2020).

Alongside the identification of a large number of *trans*-acting TFs and their target genes, the list of RNS-related *cis*-acting regulatory elements (hereafter used to refer to motifs in promoters and enhancers collectively; Table S1) has expanded. The majority of these *cis*-elements were identified *via* classical promoter deletion experiments, while in some studies, genome-wide identification of *cis*-elements was employed. A promoter tagging program using a promoterless reporter gene has been performed to guide identification of RNS-related regulatory regions and genes. This approach relies on the generation of transgenic lines whose reporter activity is detected, or in an ideal scenario, specifically induced during establishment of RNS. It theoretically allows the genome-wide mapping of regulatory regions active during RNS by locating the insertion sites of the reporter gene in transgenic lines of interest and has the advantage of providing stable transgenic lines as final products that can be used readily. The promoter tagging approach has led to the identification of a *L. japonicus* transgenic line, T90 (Webb et al., 2000) and several others (Buzas et al., 2005). T90 has served as a useful marker line for the study of plant symbiotic signal transduction over the last two decades (Kistner et al., 2005; Gossmann et al., 2012; Ried et al., 2014; Banhara et al., 2015). T90 carries a single copy of a transfer DNA (T-DNA), containing a promoterless *GUS* gene, which is inserted in the promoter region of the *Calcium Binding Protein 1* gene (*CBP1*; gene ID Lj3g3v0381710; Fig. **3a**). After staining with 5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid (X-Gluc), T90 roots display blue coloration indicative of *GUS* gene expression. The T90 *GUS* gene expression was so far exclusively observed in plant roots inoculated with AMF (Kistner et al., 2005) or rhizobia (Fig. 3**b**), including *M. loti* strain R7A in an NF-dependent manner (Webb et al., 2000) and treatment with *M. loti* strain R7A Nod factor (Gossmann et al., 2012; Webb et al., 2000) but in no other tissues or treatments tested (Kistner et al., 2005; Gossmann et al., 2012). For example, T90 *GUS* expression was neither detected in T90 shoots or leaves (Webb et al., 2000) nor inducible by synthetic hormones 1-Naphthaleneacetic acid or 6-Benzylaminopurine (Tuck, 2006). It was also not induced upon inoculation with the growth-promoting fungus *Serendipita indica* (previously known as *Piriformospora indica*) (Banhara et al., 2015). The T90 GUS phenotype hence indicates presence of a regulatory region responsible for symbiosis-related expression that was not yet identified. In addition to T90, four lines with *GUS*-trapped promoters were identified in an independent promoter-tagging mutagenesis study on *L. japonicus*, showing GUS activity covering a range of tissues and organ types including nodules (Buzas et al., 2005). Nevertheless, Buzas and colleagues reported that ca. 2% to 5% of the generated transgenic lines exhibiting GUS activity. Out of the 284 promoter tagging lines that were initially screened by Webb and colleagues, T90 is the only one that is still used for the genetic dissection of RNS. The low yield speaks against a broader use of promoter tagging approaches.

More recent techniques enable the discovery of *cis*-regulatory elements associated with individual TF (TF binding sites) in a high-throughput manner. One example is the employment of Chromatin immunoprecipitation followed by sequencing (ChIP-seq) to identify and verify NIN binding sites in target genes (Soyano et al., 2013; Laffont et al., 2020). This method was able to

reveal the difference in binding sequences by NIN and NLPs that are from the same protein family, assisting the study of target gene specificity of these TFs (Nishida et al., 2021).

The common feature of RNS-related *cis*-elements is their ability to define gene expression patterns in appropriate cell and tissue types based on binding and/or transcriptional activation by regulatory proteins. For instance, two regions of the promoter of a well-characterised symbiosis marker gene *Early Nodulin 11* (*ENOD11*) mediate gene expression in different tissues: one region is recognised by ERN1 in response to NF signalling while the other (within -257 bp 5' of the translational start site) is recognised by the NSP1/NSP2 protein complex in response to bacterial infection, leading to expression of *ENOD11* in the respective tissue types (Boisson-Dernier et al., 2005; Andriankaja et al., 2007; Cerri et al., 2012). The *cis*-elements situated between 280 bp and 257 bp as well as within 257 bp 5' of the translational start site of the *Epr3* gene are bound by ERN1 and NIN, and required for gene expression in the epidermis and nodule primordia, respectively (Kawaharada et al., 2017). NIN induces the expression of *ASL18a/LBD16* via one or two putative NIN-binding sites in the first intron of this gene, which was sufficient to confer expression in the nodule primordia. This expression domain of *ASL18a/LBD16* is consistent with its role in nodule primordia formation (Soyano et al., 2019). The distant *Cytokinin Element* (*CE*) of *M. truncatula NIN* is sufficient to mediate *NIN* expression in the pericycle in the *M. truncatula daphne-like* mutant. Additionally, another *cis*-element presumably facilitating binding by Cyclops within the proximal 5 kb of the promoter is required for the epidermal *NIN* expression as deletion of this *cis*-element rendered the 5 kb promoter incapable of rescuing IT formation in root hairs in a *M. truncatula nin-1* mutant (Liu et al., 2019c). Furthermore, distinct regulatory regions of the *NIN* promoter are directly bound by protein regulators to induce its expression - Cyclops, NSP1 (interacting with NSP2), a MYB coiled-coil type TF Interacting Protein of NSP2 (IPN2; interacting with NSP2) or an ARID (AT-rich Interaction Domain)-containing protein SIP1 (Zhu et al., 2008; Hirsch et al., 2009; Kang et al., 2014; Xiao et al., 2020). A similar phenomenon has been described for *ERN1* that is transcriptionally activated by Cyclops, IPN2 and NSP1 which recognise and bind to the promoter of *ERN1* at different regions (Hirsch et al., 2009; Cerri et al., 2017; Xiao et al., 2020). These findings support that expression of a RNS-related gene is regulated by multiple *cis*-elements during transcriptional reprogramming for RNS. The *cis*-elements therefore represent an important group of regulators for RNS.

1.4.2 Contribution of *cis*-elements to evolution of RNS

In an evolutionary context, RNS likely evolved by co-opting genes from existing machineries (1.3), implying the emergence of *cis*-elements in these co-opted target promoters that confer regulation of gene expression specific for RNS. Currently known examples of *cis*-elements that likely contributed to the evolution of RNS appear to be specific for the legume branch of the FaFaCuRo clade. These elements include the regulatory region containing putative cytokinin response elements, *CE*, that mediates *NIN* expression (Liu and Bisseling, vv2020), the *NIN*-binding sites in the intron of *ASL18a/LBD16* gene (Soyano et al., 2019), a conserved region in the promoter of *IPN2* (Xiao et al., 2020) and two elements in the promoter of *SCR* (Dong et al., 2020). The restricted presence of these *cis*-elements indicates recruitment

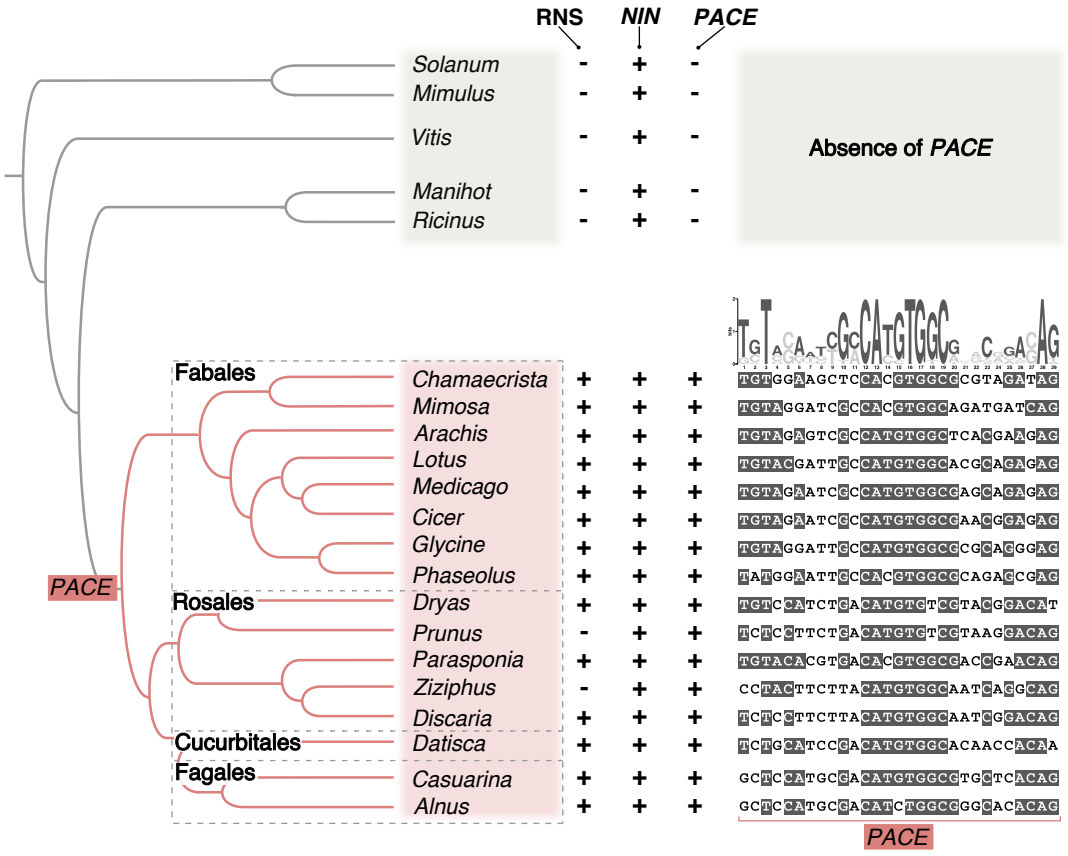


Figure 4. ***PACE* is exclusively present in the FaFaCuRo clade.** Left: Schematic illustration of phylogenetic relationships between species inside (light red shade) and outside (light grey shade) the FaFaCuRo clade and presence (+) and absence (-) pattern of RNS, *NIN* and *PACE*. Right: *PACE* sequence alignment of the displayed species in which grey shadings indicate more than 50% sequence identity. On top of the alignment the *PACE* consensus sequence depicted as Position Weight Matrix calculated from the displayed RNS-competent species. MEME analyses of *NIN* promoters and the identification of *PACE* were performed by Maximilian Griesmann. The phylogenetic tree was drafted by Ksenia Vondenhoff.

of these target genes for nodulation in leguminous plants, distant from the assumed genetic changes that enabled evolution of RNS at the base of the FaFaCuRo clade.

A potential evolutionary event that took place at the base of the FaFaCuRo clade is the co-option of the common symbiosis genes; the members coordinate two central responses including microsymbiont-induced signalling and intracellular uptake of symbiont (1.3) already for the ancient AM that evolved earlier than RNS. How the co-option process of this set of genes for RNS took place still remains elusive. According to the predisposition model (1.2.1), these common symbiosis genes could have been independently recruited several times in different plant lineages, or alternatively recruited by the last common ancestor as part of, or representing the entirety of, the predisposition event. In the single origin model, the last common ancestor would have had simultaneously recruited all common symbiosis genes. Soyano and Hayashi (2014) brought forward an inspiring hypothesis that co-option of the

common symbiosis genes to RNS was achieved through recruiting *NIN* as its downstream target. The hypothesis is supported by findings that *NIN* is transcriptionally activated by Cyclops (a TF encoded by a common symbiosis gene; Singh et al., 2014) and the first induced regulator downstream of common symbiosis genes to specifically trigger RNS-related responses. *NIN* is a central regulator, as it modulates all aspects of the establishment of RNS from bacterial infection, nodule organogenesis to balance of cost and benefits of nodulation *via* regulation of nodule numbers (1.2.2.3). It was hence proposed that connecting the common symbiosis genes and *NIN* through the gain of *cis*-regulatory elements in the *NIN* promoter was an important step during the evolution of nodulation (Soyano and Hayashi, 2014).

Following the logic of this hypothesis, a *cis*-regulatory element that recruited *NIN* should be: 1) exclusively present in the *NIN* promoters from FaFaCuRo members and absent outside of this clade and 2) maintained at least in the RNS-forming species (if assuming mutations of the element accumulated after the loss of RNS). A search for novel *cis*-regulatory elements in regulatory regions 5' of the open reading frame of the *NIN* gene from 37 plant species inside and outside of the FaFaCuRo clade was carried out (analysis performed by Maximilian Griesmann, unpublished). Only one motif fulfilling the aforementioned criteria was identified, namely the *Predisposition Associated cis-regulatory Element* (*PACE;* Fig. **4**). The 29-nucleotide long *PACE* is present at varying locations in the regulatory regions of the *NIN* genes of tested RNS-forming FaFaCuRo species and two FaFaCuRo species that have lost RNS but maintained the *NIN* gene (Fig. **4**). Additionally, a *PACE*-like motif was detected in the promoter of *Juglans regia NLP1b* gene and no *PACE* was identified, although this species has also maintained the *NIN* gene despite the loss of RNS. Importantly, no motif displaying a phylogenetic distribution like *PACE* was discovered in the promoters of *NLP* genes, which belong to the same gene family encompassing the *NIN* gene (analysis performed by Maximilian Griesmann, unpublished). The phylogenetic signature of *PACE* strongly suggests that its emergence occurred in the last common ancestor of the FaFaCuRo clade. The assumed time point of *PACE* emergence coincides with the origin of RNS (Fig. **1**), suggesting that it may have contributed to the emergence of RNS. Whether *PACE* serves a function in RNS and if so, which function, is therefore worth further investigation.

2. AIM OF THE THESIS

This study followed up two approaches that aimed to identify *cis*-regulatory elements that could have played a role in the establishment and evolution of nodulation (Fig. **3** & **4**). One was the phylogenomic approach that led to the identification of *PACE*, a *cis*-element exclusively present in the *NIN* promoters from FaFaCuRo members. *NIN* has been proposed to be a target of a co-option process through emergence of novel *cis*-elements in its promoter, that rewired it to function in RNS; however the responsible *cis*-elements remain unknown. Given the phylogenetic position of *PACE*, it is highly likely that *PACE* is an important contributor to the rewiring of *NIN* for the evolution of RNS. Using *L. japonicus* and its compatible bacterial symbiont *M. loti* as an experimental system, this work aimed to study **the relevance and importance of *PACE* in the evolution of RNS** by: 1) investigating the impact of *PACE* on the expression of *NIN* gene utilising promoter:reporter fusion constructs; 2) investigating the role of *PACE* in connecting *NIN* to Cyclops, a transcription factor participating in the ancient AM-induced signalling cascade, using transient expression assays; 3) functionally characterising the role of *PACE* in the establishment of RNS in a transgenic complementation approach using a *nin* mutant allele; and 4) examining the functional conservation of *PACE* sequence variants identified from FaFaCuRo members with regard to their ability in determining expression domains and function in RNS.

The other approach focused on the transgenic line T90, in the roots of which the *GUS* reporter gene was induced in response to colonisation by rhizobia or an AMF. The previous work suggested that the transcriptional activation of the T90 *GUS* gene is positioned downstream of common symbiosis genes *NFR1*, *SymRK*, *Pollux* and *CCaMK* because F2 plants from the following crosses, *ccamk-2* x T90, *nfr1-1* x T90 and T90 x *symrk-10* (Gossmann et al., 2012) did not respond with *GUS* expression after *M. loti* inoculation. Additionally, screenings of an ethyl methanesulfonate (EMS)-mutagenised T90 population resulted in three mutants that did not respond with *GUS* expression after *M. loti* inoculation (T90 *white;* Fig. **13a**). In line with these observations, the T90 *GUS* gene expression was observed to be activated by transgenically-expressed autoactive CCaMK^{T265D} in T90, but not in T90 *white* mutants' hairy roots. The second aim was therefore built on these observations to **resolve the relationship of the T90 *GUS* gene activation to CCaMK and its interacting partner Cyclops** and **identify symbiosis-related regulatory regions responsible for the T90 *GUS* gene expression**. For this purpose, this work 1) investigated whether the putative regulatory region of the *GUS* gene in T90 could drive gene expression in RNS and AM using promoter:reporter fusion constructs; 2) whether the so-identified regulatory region was responsive to CCaMK/Cyclops-mediated transcriptional activation in transient expression assay; 3) aimed to identify the symbiosis-relevant regulatory region(s) utilising 5' deletion series generated in the context of promoter:reporter fusion constructs; 4) investigated the reason for the inability to induce *GUS* gene in response to symbiosis in the T90 *white* mutants.

3. RESULTS

3.1 *PACE* is associated with origin of root nodule symbiosis

3.1.1 *PACE* connects *NIN* to Cyclops

Lotus japonicus PACE is located -965 to -936 bp from the translational start site of *NIN* gene, encompassing the 12 nt-long *CYC-box* that is recognised and directly bound by Cyclops to facilitate transcriptional activation of the *NIN* gene (Singh et al., 2014). We tested whether *PACE* could achieve Cyclops-mediated gene activation in transient expression assays in *Nicotiana benthamiana* leaves. *L. japonicus* Cyclops transcriptionally activated the expression of a *GUS* reporter gene driven by the 3 kb *LjNIN* promoter (NIN_{pro}) in the presence of $CCaMK^{1-314}$ (the kinase domain of *Lj*CCaMK; Hayashi et al., 2010), and this activation was almost completely demolished when *PACE* was mutated or deleted (NIN_{pro}*::mPACE* and NIN_{pro}*::ΔPACE,* respectively; Fig. **5d**). Moreover, *PACE* fused to the minimal promoter of the *NIN* gene ($NINmin_{pro}$*;* Singh et al., 2014) was sufficient for Cyclops-mediated transcriptional activation (Fig. **5c**). By contrast, no reporter gene expression was observed when $NINmin_{pro}$ was tested.

PACE was detected by MEME searches as a conserved motif within *NIN* promoters of the FaFaCuRo clade. However, the individual *PACE* sequences from different species differed from each other (Fig. **4**). The consensus sequence of *PACE* variants identified from 14 RNS-competent FaFaCuRo members - 8 species from the order Fagales and 6 species from actinorhizal members - revealed 100% conservation of nucleotides at only 7 positions (3, 12, 13, 16, 17, 19 and 28; Fig. **4**). To investigate whether these bioinformatically identified *PACE* variants could also confer transcriptional activation mediated by Cyclops, *PACE* sequence variants from RNS-forming FaFaCuRo species that belong to different plant orders, *Casuarina glauca* (*CgPACE;* Fagales), *Dryas drummondii* (*DdPACE;* Rosales) and *Datisca glomerata* (*DgPACE;* Cucurbitales) were tested (Fig. **5a**). *PACE* variants alone (*Species abbreviation PACE:*$NINmin_{pro}$) or in the context of the *LjNIN* promoter (replacing *LjPACE;* NIN_{pro}*::Species abbreviationPACE*) achieved $CCaMK^{1-314}$/Cyclops-mediated gene activation (Fig. **5b**), indicating that *PACE* variants are functionally equivalent for this purpose. These observations altogether suggested the necessity and sufficiency of *PACE* for CCaMK/Cyclops-mediated transcriptional activation.

3.1.2 *PACE* drives gene expression in nodules

Cis-elements often determine tissue-specific gene expression (spatial pattern) for the respective developmental stage of the relevant biological process (temporal pattern). To determine the *PACE*-mediated spatio-temporal expression domain, we introduced a series of promoter:GUS fusion constructs individually into the model legume *L. japonicus* Gifu wild-type roots *via Agrobacterium rhizogenes*-mediated hairy root transformation (Fig. **6**). Roots were subsequently inoculated with its compatible nitrogen-fixing microsymbiont *Mesorhizobium loti* MAFF 303099 expressing *Ds*Red (*M. loti Ds*Red) that facilitated detection of the bacteria

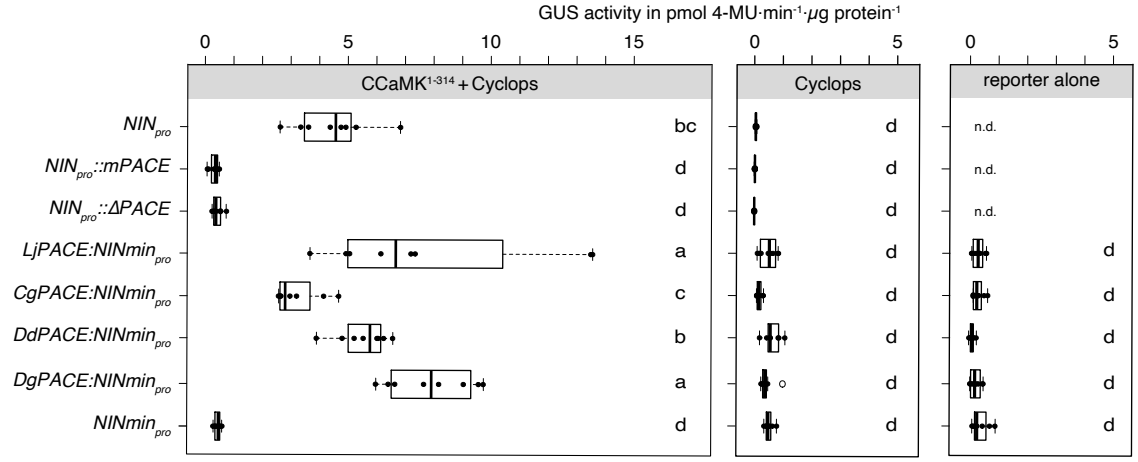
a

Abbre-viation	Species	Plant Order	RNS	FaFaCuRo member	*PACE*
Lj	*Lotus japonicus*	Fabales	Yes	Yes	Yes
Cg	*Casuarina glauca*	Fagales	Yes	Yes	Yes
Jr	*Juglans regia*	Fagales	No	Yes	No
Dd	*Dryas drummondii*	Rosales	Yes	Yes	Yes
Zj	*Ziziphus jujuba*	Rosales	No	Yes	Yes
Pp	*Prunus persica*	Rosales	No	Yes	Yes
Dg	*Datisca glomerata*	Cucurbitales	Yes	Yes	Yes
Sl	*Solanum lycopersicum* (tomato)	Solanales	No	No	No

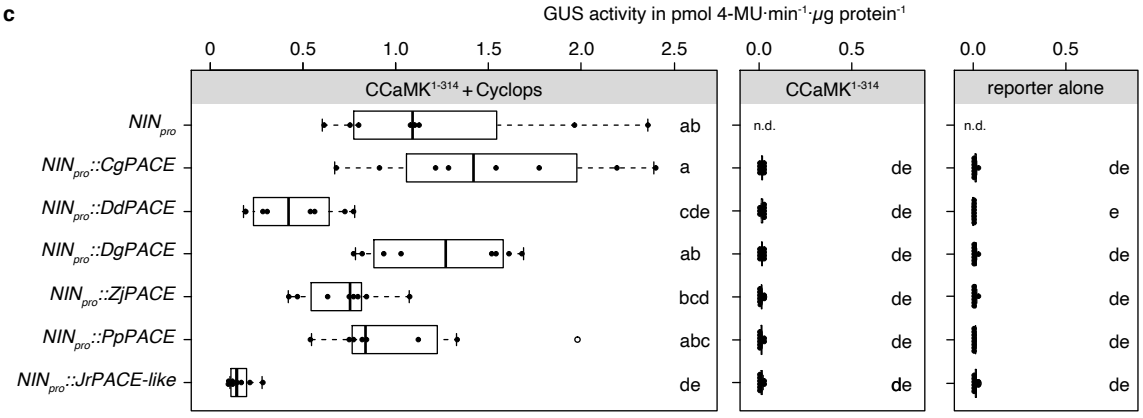
b

LjPACE TGTACGATTGCCATGTGGCACGCAGAGAG

mPACE ***GTGGATCCGTATCGTCTTAGATAGTCTGT***



c



d

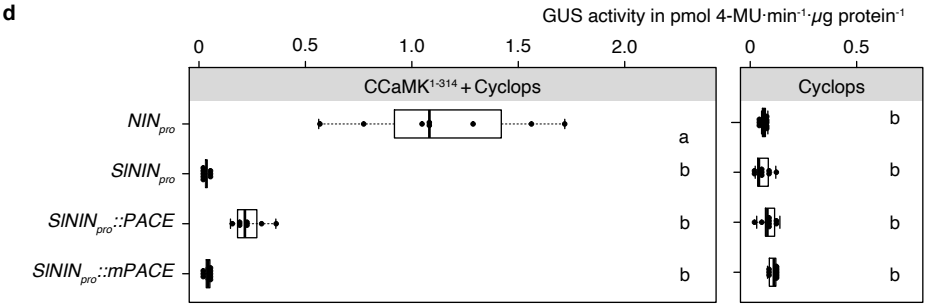


Figure 5 (legend on the next page)

Figure 5. *PACE* sequence variants from species across the FaFaCuRo clade were able to functionally replace *L. japonicus PACE* in a *LjNIN$_{pro}$:GUS* reporter fusion. a, List of species within the FaFaCuRo clade (light red shade) and outside (light blue shade) and abbreviations. **b-d**, *Nicotiana benthamiana* leaf cells were transformed with T-DNAs carrying a *GUS* reporter gene driven by either of the indicated promoters: **b**, The 3 kb *L. japonicus NIN* promoter (*NIN$_{pro}$*), the *LjNIN* promoter with *PACE* mutated or deleted (*NIN$_{pro}$::mPACE* and *NIN$_{pro}$::ΔPACE*, respectively), or *PACE* sequence variants from the nodulating FaFaCuRo species fused to the *LjNIN* minimal promoter (*NINmin$_{pro}$*); **c**, chimeric promoters where *LjPACE* in the *LjNIN* promoter was replaced with either one of the PACE variants from species tested in **b** or from non-nodulating FaFaCuRo species including the *Juglans regia PACE*-like motif (*JrPACE-like*); **d**, the *S. lycopersicum NIN* promoter (*SlNIN$_{pro}$*) and the *SlNIN* promoter with *LjPACE* (*SlNIN$_{pro}$::PACE*) or *mPACE* (*SlNIN$_{pro}$::mPACE*) inserted. Note in **b** that the deletion or mutation of *PACE* in *LjNIN* promoter resulted in a drastic reduction in reporter gene expression and in **d** insertion of *LjPACE* but not *mPACE* into the *SlNIN$_{pro}$* confers transactivation by CCaMK$^{1-314}$/Cyclops. The applied statistical method was ANOVA with *post hoc* Tukey: **b**, $F_{20,144} = 51.38$, $p < 2x10^{-16}$; **c**, $F_{18,166} = 149.1$, $p < 2x10^{-16}$; **d**, $F_{7,62} = 30.5$, $p = 7.02x10^{-7}$. Different small letters indicate significant difference. n.d., not determined.

through their fluorescence signal in root hairs and nodules 10 to 14 days post inoculation (dpi). The process by which *M. loti* is taken up by *L. japonicus* roots can be subdivided into successive stages: (1) entrapment of bacteria in a pocket formed by a curled root hair (Perrine-Walker et al., 2014), (2) development of an IT within that root hair (Perrine-Walker et al., 2014), (3) IT progression into and through the outer cortical cell layers (van Spronsen et al., 2001), (4) IT branching and extension within the nodule primordium (Yoon et al., 2014) (5) release of bacteria from ITs into symbiosomes (Yoon et al., 2014) leading to (6) mature nodules characterised by infected cells densely packed with symbiosomes and the pink colour of leghemoglobin (Ott et al., 2005).

The infection process was therefore categorised into four stages (**I** - **IV**) based on the microscopic detection method used (exemplified in Fig. **6**): (**I**) the root hair IT could be visualised as a continuous red thread on top of a developing nodule primordium; (**II**) root hair ITs progressed into the nodule primordium, but the central tissue of the primordium was not yet fully infected. At this stage, the *Ds*Red signal was not evenly distributed in the central tissue, but rather accumulated at certain sites; (**III**) the entire central tissue of the nodule primordium became colonised by bacteria, due to ramification of cortical ITs and release of bacteria into cells to form symbiosomes. The *Ds*Red signal could be detected in the entire central tissue of the nodule. The nodule typically had a defined round shape and displayed a white colour when illuminated with white light; (**IV**) the cortical cells were filled with symbiosomes in a mature nodule. In addition to the *Ds*Red signal in their entire central tissue, nodules displayed a characteristic pink colour when illuminated with white light. Transformed root pieces corresponding to these four stages were excised and stained to detect reporter activity.

The *NIN* minimal promoter did not mediate reporter gene expression at any stage of bacterial infection (Fig. **6e**). The *L. japonicus* 3 kb promoter drove reporter expression throughout the four stages (stage **I** - **IV**): from root hair bearing ITs to nodules (Fig. **6a**).

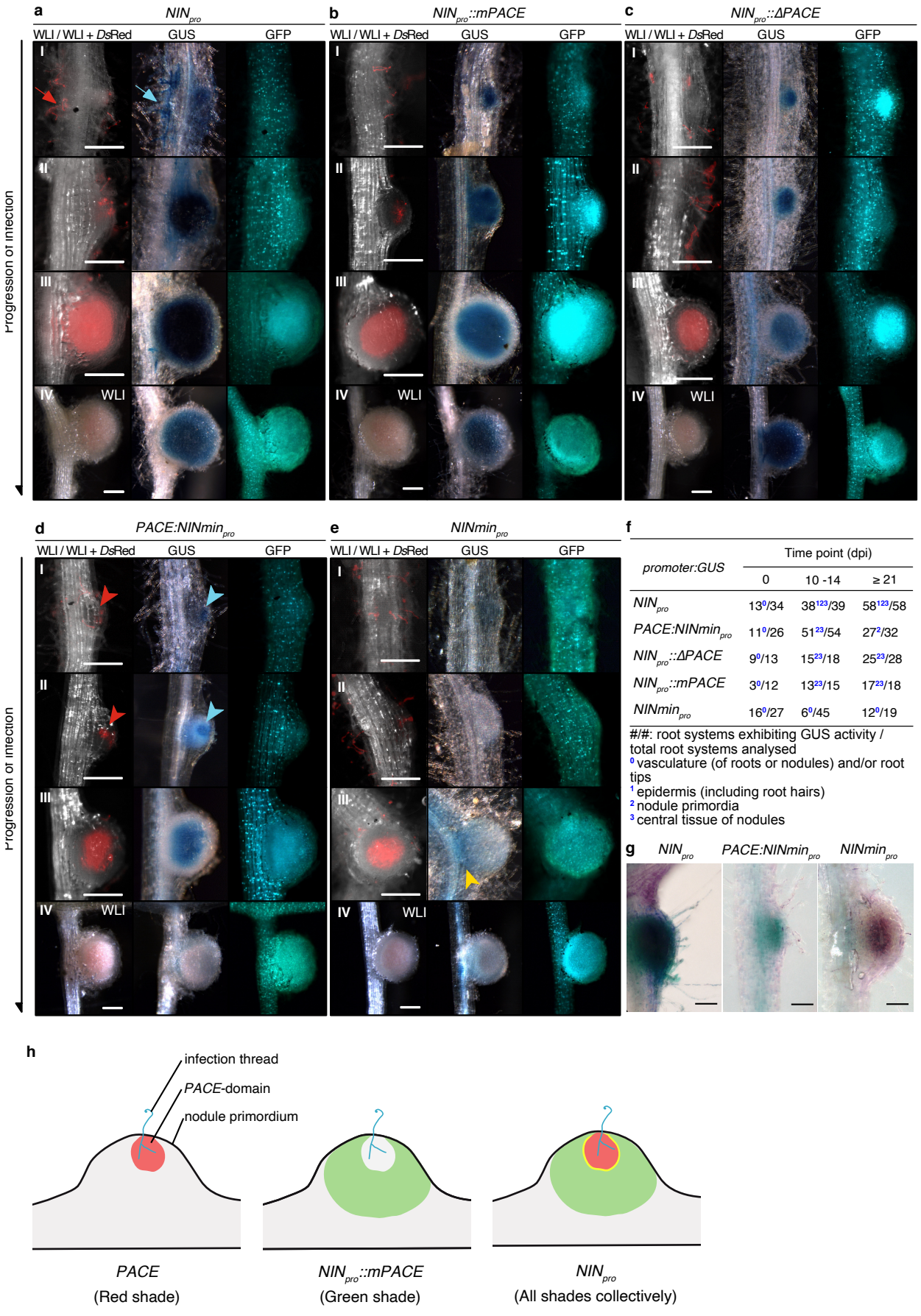
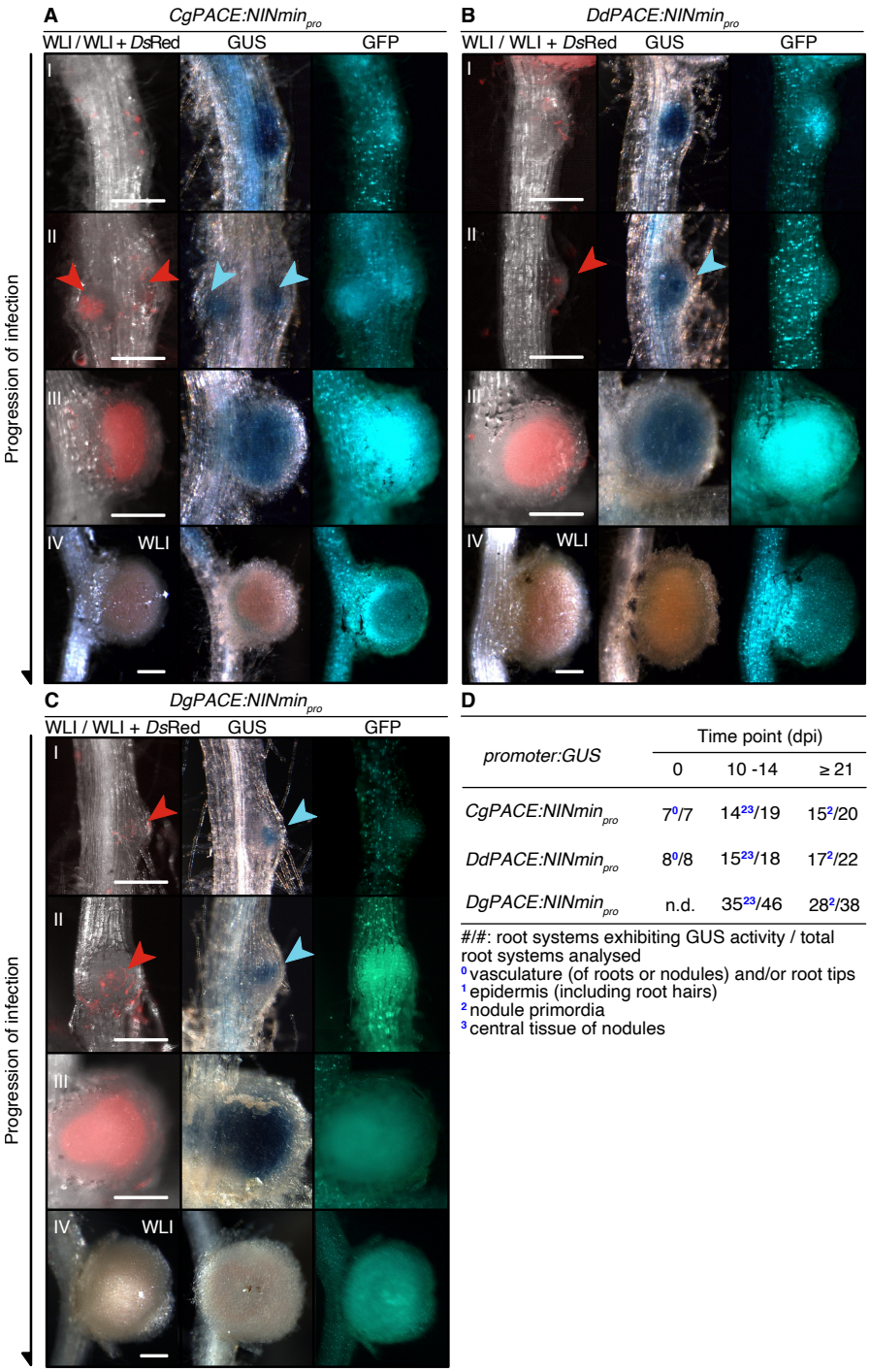


Figure 6 (legend on the next page)

Figure 6. Spatio-temporal *GUS* expression driven by *PACE* and the *NIN* promoter in *L. japonicus* roots during the bacterial infection process. a - **f**, *L. japonicus* Gifu wild-type hairy roots were transformed with T-DNAs carrying a Ubq_{10}*:NLS-GFP* transformation marker together with a *GUS* reporter gene driven by either of the indicated promoters: **a**, the 3 kb *LjNIN* promoter (NIN_{pro}); NIN_{pro} with *PACE* **b**, mutated ($LjNIN_{pro}$*::mPACE*) or **c**, deleted (NIN_{pro}*::ΔPACE*); **d**, *PACE* fused to the *LjNIN* minimal promoter (*PACE:*$NINmin_{pro}$); or **e**, *LjNIN* minimal promoter ($NINmin_{pro}$). The progression of bacterial infection was determined by the *Ds*Red signal 10 - 14 dpi with *M. loti Ds*Red. Nodules undergoing different stages of infection (panels I to IV) were stained with X-Gluc. Note the overlapping bacterial invasion zone and *PACE:*$NINmin_{pro}$*:GUS* expression in early infection stages (red and blue arrowheads in **d**) as well as the differences between *PACE:*$NINmin_{pro}$*:GUS* and the much broader NIN_{pro}*:GUS* expression at that stage (red and blue arrows in **a**). Red arrow and arrowheads: *M. loti Ds*Red. Blue arrow and arrowheads: GUS activity in root hairs bearing ITs and nodule primordia, respectively. The $NINmin_{pro}$*:GUS* fusion gave only rarely detectable signal, and if so in the vasculature (yellow arrowhead in **e**). Only white light illumination (WLI) pictures are displayed for nodules in panel VI to reveal the pink colour of leghemoglobin, characteristic for mature and fully infected nodules. Note that *PACE:*$NINmin_{pro}$*:GUS* expression was absent at this stage, whereas the $NIN_{pro:}$*GUS* resulted in strong blue staining in the nodule regardless of the presence of *PACE* (compare panel IV in **d** and **a** - **c**). **f**, Quantification of transgenic root systems exhibiting *GUS* expression in different cell types and tissues exemplarily displayed in **a** - **e**. **g**, *PACE* drove *GUS* reporter gene expression in the central tissue of primordia and nodules, but was not sufficient for expression in root hairs. Transgenic *L. japonicus* WT hairy roots carrying promoter:GUS fusions same as in **a**, **d** & **e** were inoculated with *M. loti lacZ* and dual-stained with X-Gluc and Magenta-Gal. Purple: *M. loti lacZ*. Blue, GUS activity. Note the co-existence of blue and purple staining in root hairs on roots transformed by NIN_{pro}*:GUS*, but not that transformed by *PACE:*$NINmin_{pro}$*:GUS*. Bars, 250 µm. **h**, Graphic summary of expression domain achieved by *PACE,* NIN_{pro} and NIN_{pro}*::mPACE* determined by fluorescent reporters (analysis performed by Chloé Cathebras). Yellow shade: overlapping domain.

Mutation or deletion of *PACE* in the context of the *NIN* promoter (NIN_{pro}*::mPACE* and NIN_{pro}*::ΔPACE,* respectively) resulted in a loss of *GUS* expression in the root hairs bearing ITs, but did not affect that in nodule primordia and nodules (stage **II** - **IV**; Fig. **6c - d**). Intriguingly, the earliest detectable GUS activity mediated by *PACE:*$NINmin_{pro}$*:GUS* was clearly restricted to a zone in the nodule primordia (panel **I** - **II** in Fig. **6d**) that roughly correlated with the site of bacterial infection (indicated by a local accumulation of *Ds*Red signal) and later expanded to the entire central tissue of the nodule (panel **III** in Fig. **6d**). *PACE*-driven reporter expression was neither detected in root hairs bearing ITs (Fig. **6g**) nor in nodules in which cortical cells were filled with symbiosomes (panel **IV** in Fig. **6d**). *PACE*-driven expression was distinct from that mediated by NIN_{pro} or the NIN_{pro} with *PACE* mutated or deleted that conferred reporter expression across the central tissue of the nodule (panels **II** - **IV** in Fig. **6a-c**).

Moreover, to confirm the absence of reporter expression in the root hairs, a *M. loti* MAFF 303099 *lacZ* (*M. loti lacZ*) strain was used to inoculate transgenic roots systems transformed with NIN_{pro}*:GUS*, *PACE:*$NINmin_{pro}$*:GUS* or NIN_{pro}*::mPACE:GUS* (Fig. **6g**). Dual staining for lacZ and GUS activity allowed simultaneous visualisation of bacteria and reporter gene expression



D

promoter:GUS	Time point (dpi)		
	0	10 -14	≥ 21
CgPACE:NINmin pro	7[0]/7	14[23]/19	15[2]/20
DdPACE:NINmin pro	8[0]/8	15[23]/18	17[2]/22
DgPACE:NINmin pro	n.d.	35[23]/46	28[2]/38

#/#: root systems exhibiting GUS activity / total root systems analysed
[0] vasculature (of roots or nodules) and/or root tips
[1] epidermis (including root hairs)
[2] nodule primordia
[3] central tissue of nodules

Figure 7. Spatio-temporal *GUS* expression driven by *PACE* variants in *L. japonicus* roots during the bacterial infection process. *L. japonicus* Gifu wild-type roots were transformed with T-DNAs carrying a *Ubq10*$_{pro}$*:NLS-GFP* transformation marker together with a *GUS* reporter gene driven by either of the *PACE* variants from nodulating FaFaCuRo species fused to the *LjNIN* minimal promoter (*NINmin*$_{pro}$). For species abbreviations and experimental details see Fig. S5a and Fig. 6, respectively. Note the overlapping bacterial invasion zone and *PACE:NINmin*$_{pro}$*:GUS* expression in early infection stages (red and blue arrowheads in **a** - **c**). Red arrowheads: *M. loti* *Ds*Red. Blue arrowheads: GUS activity in nodule primordia. Note that like *LjPACE*, the *PACE* variants-driven *GUS* expressions were absent at this stage (panel IV in **a** - **c** and panel IV in Fig. 7D). **d**, Quantification of transgenic root systems exhibiting *GUS* expression in different cell types and tissues exemplarily displayed in **a** - **c**. n.d., not determined. Bars, 250 µm.

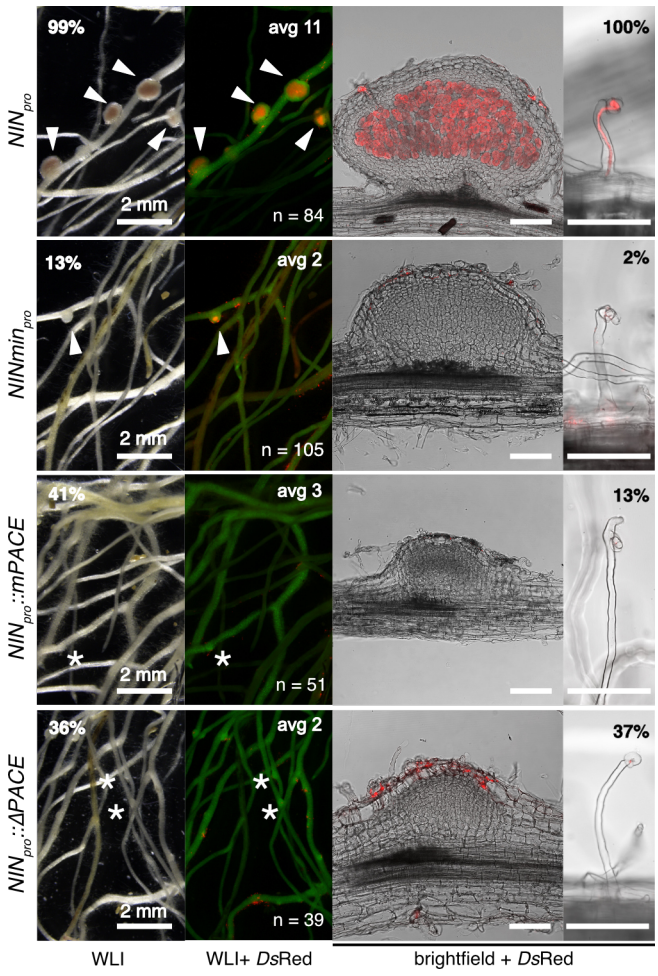


Figure 8. *PACE* is required for the restoration of the bacterial infection in the *L. japonicus nin-15* mutant. *L. japonicus nin-15* hairy roots were transformed with T-DNAs carrying *Ubq10*$_{pro}$*:NLS-GFP* as a transformation marker in tandem with the *LjNIN* gene driven by either of the following promoters: the 3 kb *LjNIN* promoter (*NIN*$_{pro}$), *LjNIN* minimal promoter (*NINmin*$_{pro}$), and the *LjNIN* promoter with *PACE* mutated (*NIN*$_{pro}$*::mPACE*) or deleted (*NIN*$_{pro}$*::ΔPACE*) and analysed 21 dpi vwith *M. loti Ds*Red. Images of sections of representative nodules and those of an IT in a root hair or bacterial entrapment are displayed. Note the drastic reduction of restoration of infection in nodules and root hairs associated with the mutation or deletion of *PACE.* Quantification of root hair ITs and infection nodules of these transgenic systems is included in Fig. 10. % in images: percentage of transgenic root systems bearing at least one infected nodule or root hair IT. avg: average number of infected nodules on plants bearing infected nodules. n: number of transgenic root systems analysed. White arrowheads and asterisks: infected and non-infected nodules, respectively. WLI: white light illumination. Bars, 100 μm unless labeled. Characterisation of the *nin-15* mutant was performed by Rosa Elena Andrade. The data presented in this figure was generated by Rosa Elena Andrade, Chloé Cathebras and Xiaoyun Gong.

(purple and blue colour, respectively). Co-localised blue and purple colour was observed in root hairs of *NIN*$_{pro}$*:GUS*-transformed roots, indicating co-localization of bacteria and GUS activity. In comparison, while purple root hairs were visible on *PACE:NINmin*$_{pro}$*:GUS*-transformed roots, GUS activity was restricted to nodule primordia. *NINmin*$_{pro}$*:GUS*-transformed roots displayed purple colour in root hairs and nodule primordia, however GUS activity was not detected. Based on these observations, we concluded that the *PACE*-driven expression domain is temporally and spatially restricted and possibly accompanies the development of bacterial accommodation structures in the nodule.

The observation of localised GUS activity in *PACE:NINmin*$_{pro}$*:GUS*-transformed nodule primordia (stage **I - II**) suggested that *PACE*-driven expression was likely regulated in a cell-specific manner in response to bacterial infection in the nodule primordia. To further resolve this relationship between *PACE*-driven gene expression and bacterial accommodation at the

cellular level, an independent analysis using a red and a yellow fluorescent proteins (mCherry and YFP, respectively) was performed by Chloé Cathebras (LMU Biocenter, Munich). Using a fusion construct that contained *PACE:NINmin$_{pro}$:NLS-mCherry* and *NIN$_{pro}$::mPACE:NLS-YFP* placed in tandem in the same T-DNA region, the expression domain of *PACE* could be directly compared to that of *NIN$_{pro}$::mPACE* (analysis performed by Chloé Cathebras; unpublished data; graphic summary for this result included in Fig. **6h**). In sections of developing nodules, in which infection had progressed to stage **III** or **IV**, *PACE*-mediated mCherry was expressed specifically in a - hereafter called "infection zone" - comprising cortical cells that carried ITs and in some, but not all, directly adjacent cells. Intriguingly, the expression domains marked by mCherry and YFP fluorescence were distinct from each other: while the *PACE*-driven mCherry signal was consistently marking the infection zone (red in Fig. **6h**), the *NIN$_{pro}$::mPACE*-driven YFP signal was observed in cortical cells surrounding this zone (green in Fig. **6h**). The thin (approx. 1-2 cells thick) border between the two domains was characterised by nuclei emitting both YFP and mCherry signals (yellow in Fig. **6h**). In so-marked cells, ITs were typically not detected. The expression pattern mediated by the *NIN* promoter (containing *PACE*) was congruent with the sum of both promoter fragments.

These observations suggested that *PACE* directs *NIN* expression to a specific infection zone and that the *NIN* promoter comprises *cis*-regulatory elements that drive expression outside the *PACE* territory i.e., in root hairs (together with *PACE*), non-infected cortical cells and cells filled with symbiosomes.

3.1.3 *PACE*-driven expression pattern is conserved across the FaFaCuRo clade

The expression pattern mediated by *PACE* variants from RNS-forming species belonging to the other three clades, *C. glauca* (*CgPACE;* Fagales), *D. drummondii* (*DdPACE;* Rosales) and *D. glomerata* (*DgPACE;* Cucurbitales) were investigated as described in **3.1.2** for *LjPACE* to test whether the *PACE*-driven expression domain was conserved (Fig. **7**). Similar to *LjPACE, CgPACE, DdPACE* and *DgPACE*, individually was sufficient to drive gene expression in nodule primordia and nodules (stage **II** - **III**), but neither in root hairs bearing ITs nor mature nodules (stage **I** & **IV**, respectively). For all *PACE* variants analysed, GUS activity was observed in nodule primordia at stage **II** (yellow arrowheads), resembling what was observed when *LjPACE* was tested (Fig. **6e**). These results suggested that the expression pattern conferred by *PACE* variant is conserved across the FaFaCuRo clade.

3.1.4 *PACE* enables formation of infection threads in nodules

The fact that *PACE*-mediated expression marks an infection zone indicates that *PACE*-mediated *NIN* expression is likely to be associated with bacterial infection. To investigate the relevance and specific role of *PACE* in relation to bacterial infection, we utilised a *L. japonicus nin-15* mutant carrying a *Lotus Retrotransposon 1* insertion in the *NIN* promoter 143 bp 3' of *PACE* (*LORE1* line 30003529; Małolepszy et al., 2016). *nin-15* is impaired in IT formation but retains the capacity to form nodules. Most of these nodules were uninfected (92% and 86% plants carrying no root hair ITs and no infected nodules, respectively) and cortical cells filled

with symbiosomes were never observed (analysis performed by Rosa Elena Andrade and Chloé Cathebras; unpublished data). This mutant therefore provided an ideal background to enable a focussed analysis of *PACE* in cortical IT formation, circumventing the negative epistatic effect of the inability of nin loss-of-function mutants to initiate cell divisions. The *L. japonicus NIN* gene driven by a series of promoters (*Promoter:NIN*) individually was introduced into *nin-15* hairy roots. We examined the restoration of bacterial infection 21 dpi with *M. loti Ds*Red by quantifying the number of root hairs harbouring ITs and the number of infected nodules (Fig. **8 - 11**; Material and Method).

The success of restoration of bacterial infection was described in two respects: (1) infection frequency: percentage of transgenic root systems bearing at least one roots hair IT or infected nodule and (2) the absolute count: average number of infected nodules in transgenic root systems bearing bearing nodules. Transformation with the *LjNIN* gene driven by the *NIN* minimal promoter (*NINmin$_{pro}$:NIN*) did not alter the symbiotic phenotype of *nin-15* roots: only 2% and 13% of the transgenic root systems bore root hair ITs and infected nodules, respectively, with an average of less than 2 infected nodules (Fig. **8**). In contrast, bacterial infection in root hairs and nodules of *nin-15* could be fully restored by introducing the *LjNIN* gene driven by its 3 kb promoter (*NIN$_{pro}$:NIN*): the majority of the transgenic root systems (92%) bore an average of 11.2 infected nodules while root hair ITs were observed in 100% of the root systems analysed (Fig. **8**). Mutation or deletion of *PACE* in the *NIN* promoter (*NIN$_{pro}$::mPACE* and *NIN$_{pro}$::ΔPACE,* respectively) resulted in drastically reduced restoration of infection both in root hairs and nodules (Fig. **8**). These results indicated that *PACE* is involved in bacterial infection in both root hairs and nodules.

We tested whether *PACE* fused to the *NINmin$_{pro}$* (*PACE:NINmin$_{pro}$:NIN*) could rescue bacterial infection of nodules formed on *nin-15* roots, considering that *PACE:NINmin$_{pro}$* was sufficient to drive gene expression in the nodules (Fig. 9). In line with its expression pattern, *PACE*-mediated *NIN* expression failed to restore infection in root hairs: only 8% of transgenic root systems bore root hair ITs. Transgenic root systems bearing infected nodules increased from 17% of *NINmin$_{pro}$*:*NIN*-transformed root systems to 49% of those transformed by *PACE:NINmin$_{pro}$:NIN*, with an increase of the average number of infected nodules from 1.7 to 2.8 (Fig. **9a-b**). Importantly, sectioning of the infected nodules formed on *PACE:NINmin$_{pro}$:NIN*-transformed roots revealed that the vast majority of these nodules carried ITs in the outer cortex, originating from a hyperaccumulation of *M. loti Ds*Red that seemed to be locally constricted by root cell wall boundaries (25 out of 28 nodule sections inspected). This phenomenon was not observed in most of the rarely occurring infected nodules formed on *NINmin$_{pro}$*:*NIN*-transformed roots (11 out of 16 nodule sections inspected). Taken together, these results suggested that *PACE* promotes IT development in cortex cells but not within root hairs.

3.1.5 *PACE* is functionally conserved across the FaFaCuRo clade

We tested whether and to what extend the sequence variation of *PACE* would affect its function. We exchanged *LjPACE* (Fabales) with either of the *PACE* variants (listed in Fig. **5a**)

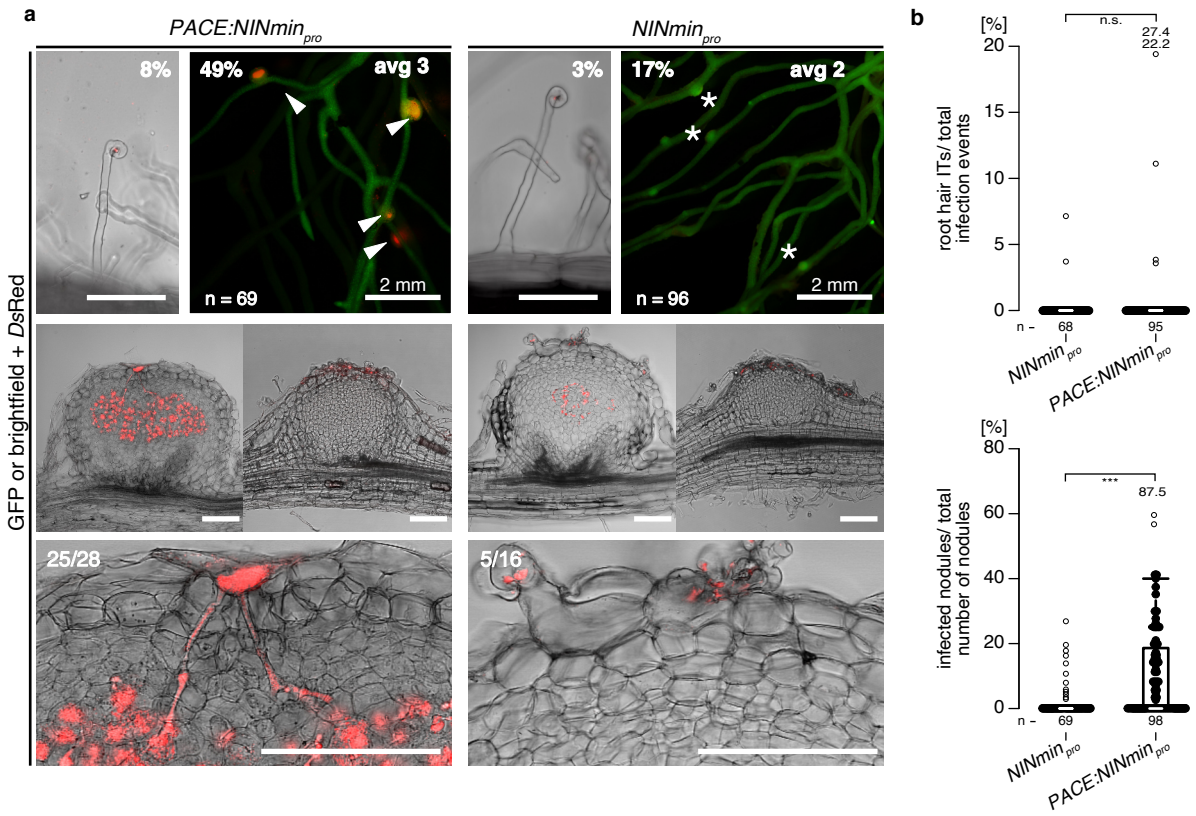
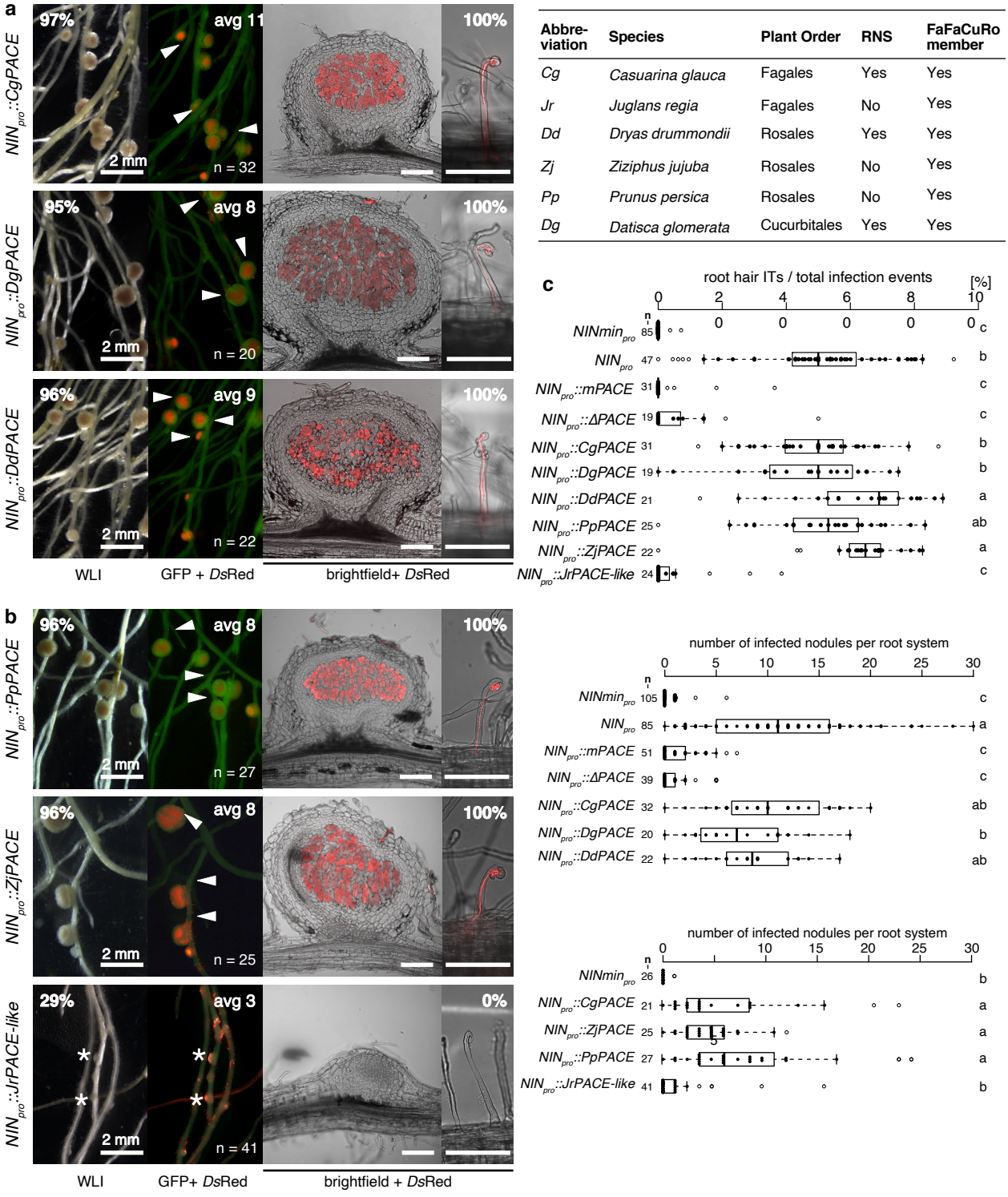


Figure 9. *PACE* enables IT formation in the cortex in *L. japonicus nin-15* mutant. *L. japonicus nin-15* hairy roots were transformed with T-DNAs carrying *Ubq10$_{pro}$:NLS-GFP* as a transformation marker together with the *LjNIN* gene driven by either of the following promoters: the *LjNIN* minimal promoter (*NINmin$_{pro}$*) or *PACE* fused to the *LjNIN* minimal promoter (*PACE:NINmin$_{pro}$*) and analysed 21 dpi with *M. loti Ds*Red. **a**, images (*Ds*Red and GFP images merged) of root systems, root hair ITs and nodule sections. White arrowheads and asterisks: infected and non-infected nodules, respectively. Note the long cortical infection threads on the section of *PACE:NINmin$_{pro}$:NIN*-transformed nodules. % in images: percentage of transgenic root systems bearing at least one infected nodule or root hair IT. n: number of transgenic root systems analysed. #/# on nodule section picture: number of nodule sections bearing long cortical ITs / total number of nodule sections inspected. Bars, 100 *µ*m unless labeled. **b**, Boxplots displaying the percentage of root hair ITs among total infection events (sum of bacterial entrapments and ITs) or the percentage of infected nodules of total number of nodules. Each dot represents one *nin-15* transgenic root system or root piece. n: number of transgenic root systems or root pieces analysed. Numbers above the boxplots: the value of individual data points outside of the plotting area. The applied statistical method was Fisher's exact test: *$p < 0.05$; ***$p < 0.001$; n.s.: not significant. The data presented in this figure was generated by Rosa Elena Andrade and Chloé Cathebras.

from the RNS-forming species tested in Fig. **5a**: *DgPACE* (Cucurbitales), *DdPACE* (Rosales), and *CgPACE* (Fagales) in the context of the *L. japonicus NIN* 3 kb promoter (*NIN$_{pro}$::Species abbreviation PACE:NIN*), and tested the ability of these chimeric promoters to restore bacterial infection in *nin-15* roots (as in **3.1.4**; Fig. **10**). The *LjNIN* gene driven by either of these chimeric restored the complete infection process in *nin-15* to a similar level as *NIN$_{pro}$:NIN*, indicating the



Abbre-viation	Species	Plant Order	RNS	FaFaCuRo member
Cg	*Casuarina glauca*	Fagales	Yes	Yes
Jr	*Juglans regia*	Fagales	No	Yes
Dd	*Dryas drummondii*	Rosales	Yes	Yes
Zj	*Ziziphus jujuba*	Rosales	No	Yes
Pp	*Prunus persica*	Rosales	No	Yes
Dg	*Datisca glomerata*	Cucurbitales	Yes	Yes

Figure 10. *PACEs* from FaFaCuRo species are functionally equivalent in restoring bacterial infection in the *L. japonicus nin-15* mutant. **a** - **b**, *L. japonicus nin-15* roots were transformed with T-DNAs carrying *Ubq10*$_{pro}$*:NLS-GFP* as a transformation marker together with the *LjNIN* gene driven by either of the promoters described in Fig. 5. and identified by GFP fluorescence emanating nuclei 21 dpi with *M. loti Ds*Red. Images of root systems, nodule sections and root hair Its or bacteria entrapments are displayed. n, number of transgenic root systems analysed. Note the drastic reduction of restoration of infection in nodules and root hairs when *LjPACE* was replaced with *JrPACE-like* in the context of *L. japonicus NIN* promoter. White arrowheads and

asterisk: infected and non-infected nodules, respectively. % in images: percentage of transgenic root systems bearing at least one infected nodule or root hair IT. **c**, Boxplots displaying the percentage of root hair ITs among total infection events (sum of bacterial entrapments and Its; top) and the number of infected nodules per *nin-15* transgenic root system (middle and bottom). Middle and bottom plots display data collected from two independent experiments and each dot represents one *nin-15* transgenic root system. Top plots display merged data from all experiments as the percentage represents a normalised value calculated for each root piece. n: number of transgenic root systems or root pieces analysed. The applied statistical method was ANOVA with *post hoc* Tukey: **c**, top to bottom, $F_{9,313} = 106.7$, $p < 2x10^{-16}$; $F_{6,346} = 82.89$, $p < 2x10^{-16}$; $F_{4,135} = 20.18$, $p = 4.76x10^{-13}$, respectively. Different small letters indicate significant differences. WLI: white light illumination. Bars, 100 μm unless labeled. The data presented in **a**-**b** was generated by Rosa Elena Andrade, Chloé Cathebras and Xiaoyun Gong.

functional conservation of *PACE* from nodulating species across the entire FaFaCuRo clade (Fig. **10a&c**). *PACE* variants were also identified in two species that are unable to establish RNS but maintain the *NIN* gene: *Prunus persica* (*PpPACE;* Rosales) and *Ziziphus jujuba* (*ZjPACE;* Rosales). The 7 conserved nucleotides among *PACE* variants from RNS-forming members (Fig. **4**) are also conserved in *PpPACE* and *ZjPACE.* In the context of the *LjNIN* promoter, *PpPACE* or *ZjPACE* in the context of $LjNIN_{pro}$ were able to fully restore bacterial infection of *nin-15* roots: 96% and 96% transgenic root systems bore an average of 8.3 and 7.6 infected nodules, respectively; and 95% and 96% transgenic root systems bore root hair ITs, respectively.

PACE was not detected in the promoters of *NLP* genes (analysis by Maximilian Griesmann; **1.4.2**). Curiously it was also absent from the promoter of the so-annotated *NIN* gene of *Juglans regia* (Fagales). However, a *PACE*-like motif was identified in the promoter of the closest gene family member, *NIN-like protein 1 JrNLP1b* (*JrPACE-like*). The *LjNIN* promoter with *LjPACE* exchanged with this *PACE*-like element behaved similarly to the *NIN* promoter where *PACE* was mutated or deleted (Fig. **10b**): only 29% and 29% transgenic roots systems bore an average of 3.3 infected nodules and root hairs with ITs, respectively. This observation indicated that this *PACE*-like element could not restore bacterial infection in *nin-15.*

3.1.6 *PACE* insertion into the tomato *NIN* promoter confers RNS capability

To artificially recapitulate the functional consequence of *PACE* acquisition into a non-FaFaCuRo *NIN* promoter, we chose tomato (*Solanum lycopersicum*) which belongs to the Solanaceae, a family phylogenetically distant from the FaFaCuRo clade. *S lycopersicum* was chosen has the common symbiosis genes and is able to establish AM, but not RNS. The *SlNLP2* gene (gene ID Solyc01g112190.2.1) was identified based on a previously published phylogenetic tree as the closest homolog of *LjNIN* (Griesmann et al., 2018), and is hereafter referred to as the *SlNIN* gene. A 3 kb region of the *SlNIN* promoter fused to the endogenous 238 bp *SlNIN* 5'UTR was cloned from *S. lycopersicum* cv. "Moneymaker" and *PACE* or *mPACE* (Fig. **5b**) was inserted 184 bp upstream of the *SlNIN* 5'UTR. Consistent with the absence of *PACE*, a GUS reporter gene driven by the tomato *NIN* promoter ($SlNIN_{pro}$) was not transactivated by Cyclops in *N. benthamiana* leaf cells, while the insertion of the *L. japonicus*

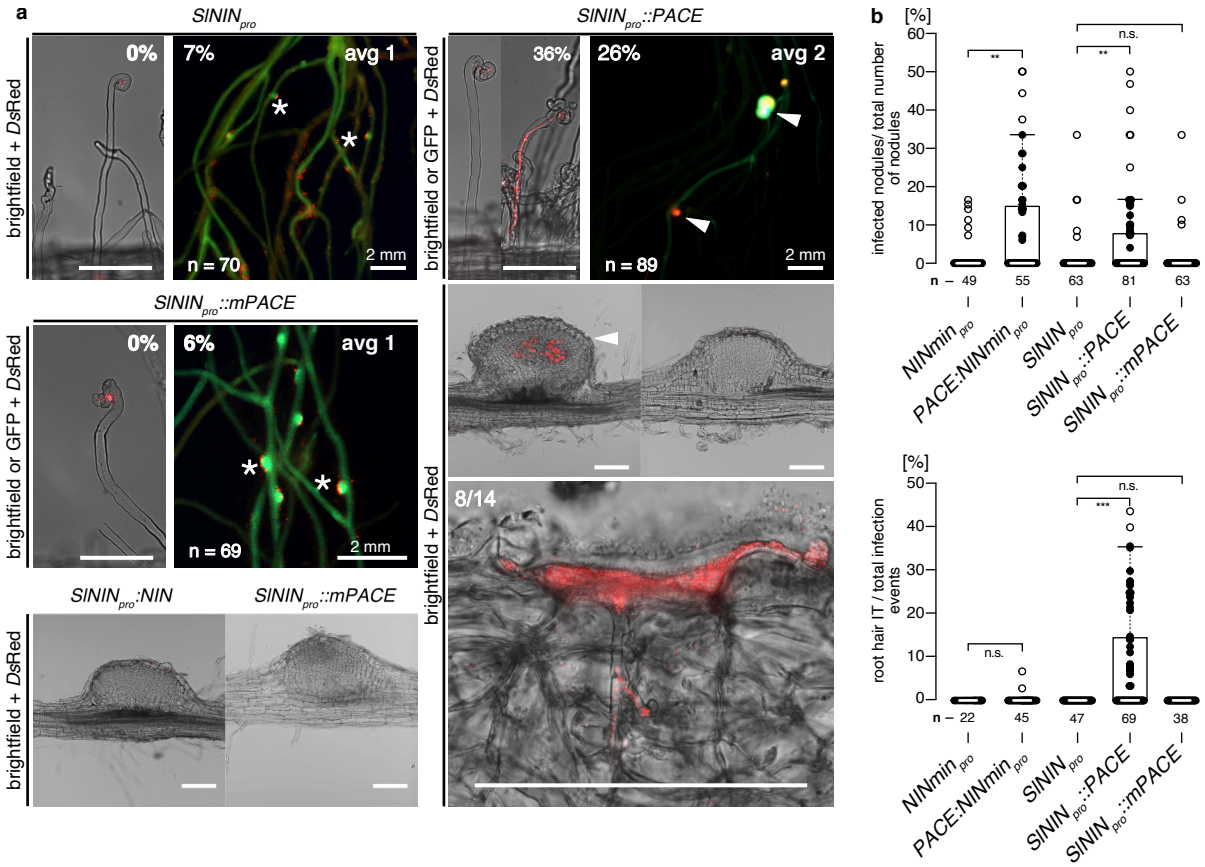


Figure 11. Insertion of *LjPACE* into the *NIN* promoter from a non-FaFaCuRo species *S. lycopersicum* (tomato) could restore bacterial infection in roots hairs and nodules in the *L. japonicus nin-15* mutant. *nin-15* hairy roots were transformed with T-DNAs carrying *Ubq10*$_{pro}$*:NLS-GFP* as a transformation marker together with the *LjNIN* gene driven by either of the following promoters: the *S. lycopersicum NIN* promoter (*SlNIN*$_{pro}$), *SlNIN*$_{pro}$ with *L. japonicus PACE* (*SlNIN*$_{pro}$*::PACE*) or *mPACE* inserted (*SlNIN*$_{pro}$*::mPACE*) and analysed 21 dpi with *M. loti Ds*Red.. White arrowheads and asterisks: infected and non-infected nodules, respectively. **a**, Images of root systems (merge of GFP and *Ds*Red images), a root hair IT or bacterial entrapments and sections of infected (when applied) or non-infected nodules are displayed. % in images: percentage of transgenic root systems bearing infected nodules or root hair ITs. avg: average number of infected nodules on plants bearing infected nodules. n: number of transgenic root systems analysed. #/# on section images: number of nodule sections bearing long cortical ITs / total number of nodule sections inspected. Bars, 100 μm unless labeled. **b**, Boxplots displaying the percentage of root hair ITs among total infection events (sum of bacterial entrapments and ITs) or the percentage of infected nodules of total number of nodules. Each dot represents one *nin-15* transgenic root system or root piece. n: number of transgenic root systems or root pieces analysed. The applied statistical method was Fisher's exact test: **p<0.01; ***p<0.001; n.s., not significant. Inspection of nodule sections, and imaging of nodule sections and root hair ITs were performed by Chloé Cathebras.

PACE (*SlNIN*$_{pro}$*::PACE*), but not of a mutated *PACE* (*SlNIN*$_{pro}$*::mPACE*) resulted in a significant transactivation by Cyclops (Fig. **5c**).

We tested the ability of the *LjNIN* expressed under the control of these synthetic promoters to restore the bacterial infection process in *nin-1*5 (Fig. **11**). Similar to *NINmin*$_{pro}$*:NIN*-transformed *nin-15* roots, *SlNIN*$_{pro}$*:NIN* did not restore bacterial infection (0% and 7% of transgenic root systems bore root hair ITs and infected nodules, respectively; Fig. **11a-b**). In contrast, *nin-15* roots transformed with *SlNIN*$_{pro}$*::PACE:NIN* restored the formation of root hair ITs and infected nodules on 36% and 26% of transgenic root systems, respectively (Fig. **11a-b**). This increase in infection success was not observed on *SlNIN*$_{pro}$*::mPACE:NIN*-transformed roots. ITs in the outer cortex that originated from a focal accumulation of bacteria were also observed in the *SlNIN*$_{pro}$*::PACE:NIN*-transformed *nin-15* nodules (8 out of 14 nodules inspected; Fig. **11a**) resembling those in the *PACE:NINmin*$_{pro}$*:NIN*-transformed nin-15 nodules (Fig. **8**). The gained ability of the *SlNIN::PACE* promoter to restore root hair ITs suggested that additional *cis*-regulatory elements within the *SlNIN* promoter function together with *PACE* for root hair IT formation.

All together, these findings obtained with the tomato *NIN* promoter carrying an artificially inserted *PACE* agree with the hypothesis that the acquisition of *PACE* by a non-FaFaCuRo *NIN* promoter enabled its regulation via Cyclops and laid the foundation for IT formation in cortical cells.

3.2 *CYC-RE*$_{CBP1}$ confers gene expression in endosymbioses

3.2.1 GUS activity is absent in the T90 *white* mutants during RNS and AM

To identify the regulators of the T90 *GUS* gene, two independent genetic screens were performed using an ethyl methanesulfonate (EMS)-mutagenised T90 population. The rationale was as follows: mutants with altered GUS activity (and/or impaired symbiotic behaviour) likely possess defects in the upstream regulatory machinery that directly or indirectly regulates the transcription of the *GUS* gene. Since transcriptional activation of the *GUS* gene in T90 occurs in response to symbiotic interactions, these impaired machineries potentially play a role in RNS and/or AM. In brief, an EMS-mutagenised T90 M_2 population was generated by separately harvesting seeds of 1342 M_1 plants labelled T0001-T1342 (screening performed by Elaine Jesen née Tuck and Simone Bucerius). Two independent screens were conducted at the seedling stage utilising individual M_2 families to identify (Fig. **12a**): (A) individual M_2 plants displaying spontaneous activation of the *GUS* gene in the absence of symbionts or (B) individual M_2 plants with altered GUS activity in presence of *M. loti* (for further details see Tuck 2006). For this purpose, root pieces were removed and stained with X-Gluc whereas the rest of the seedling was maintained to allow for seed production and analysis of heritability. Screen A of 519 M_2 and 203 M_3 lines resulted in 84 plants from 55 lines, which showed spontaneous *GUS* expression in the roots, however no progenies from them inherited this phenotype.

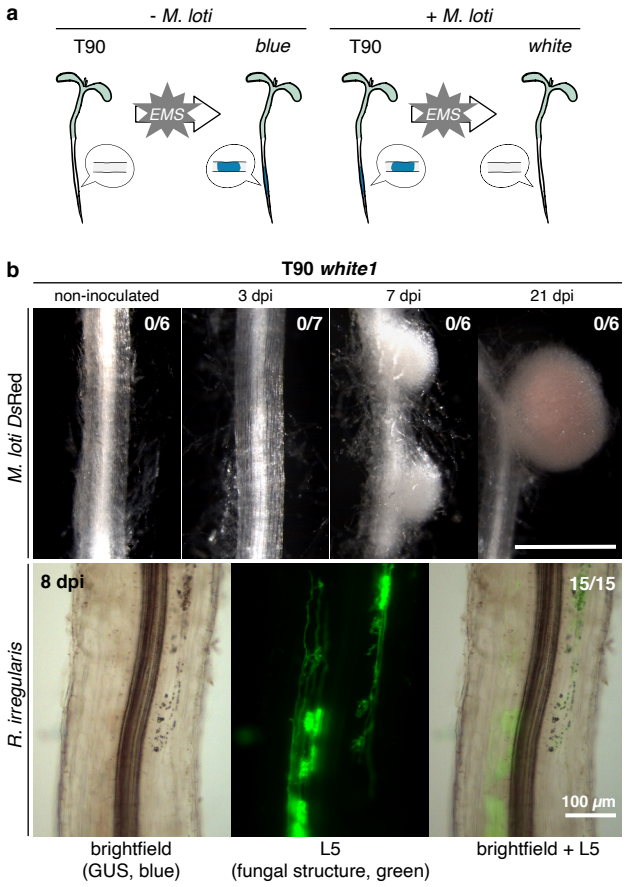


Figure 12. Absence of GUS activity in T90 *white* mutant roots during AM or RNS. a, Schematics for the two screens of EMS-induced mutant populations for M_2 seedlings with altered GUS activity: spontaneous activation of the *GUS* gene in the absence of symbionts (left) or undetectable GUS activity in the presence of symbionts (right; resulting mutants are referred to as T90 *white* mutants). **b**, T90 *white1* roots were stained with X-Gluc at indicated dpi with *M. loti* *Ds*Red or *R. irregularis*. Note the total absence of GUS activity in T90 *white* roots, compared to those of T90 upon inoculation with microsymbionts (tested side-by-side in the same experiment; see Fig. 1a; Fig. S1). Pictures of T90 *white1* root systems and analysis of T90 *white3* are included in Fig. S2c-d. Green: Alexa Fluor-488 WGA-stained *R. irregularis* visualised with a Leica Filter Cube L5 next to a brightfield image of the same root segment.. #/#: number of plants displaying GUS activity / total number of plants analysed.

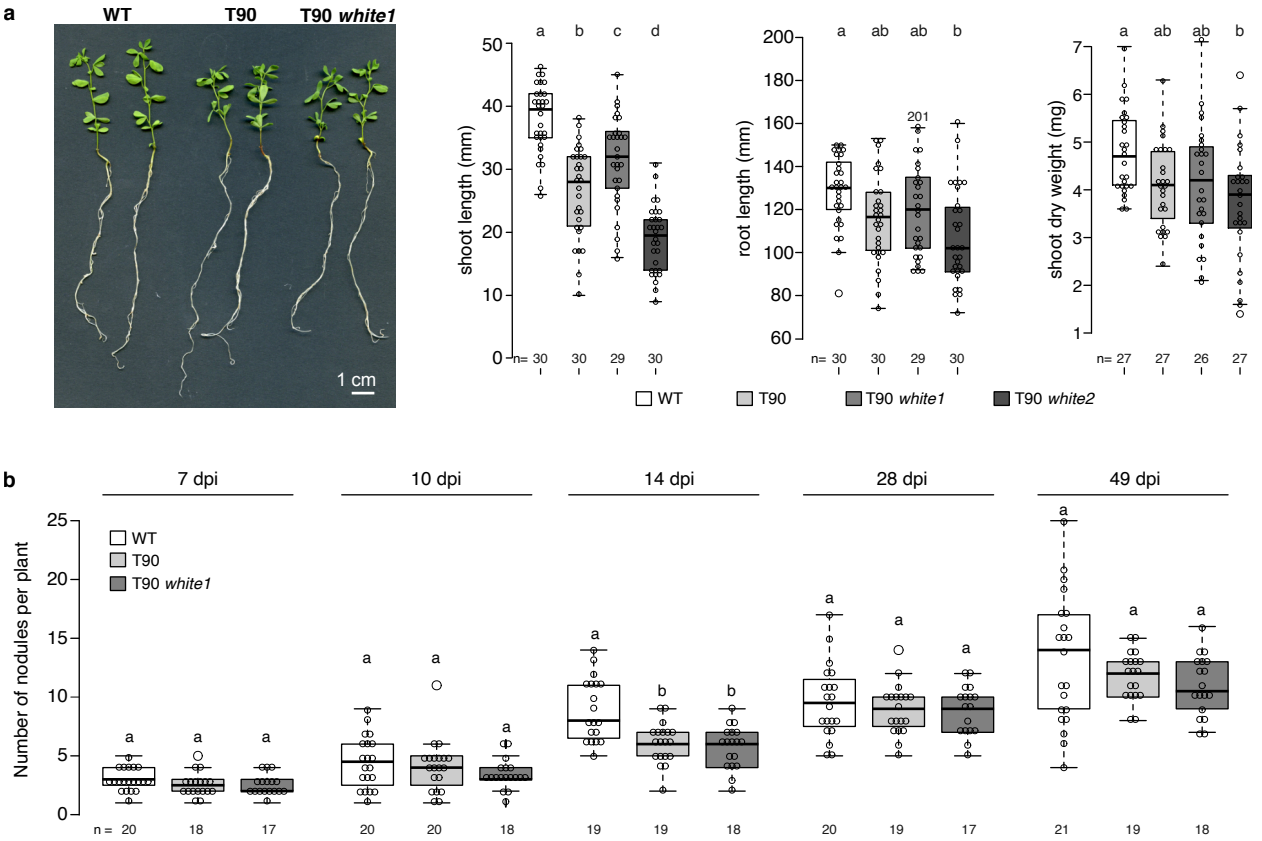


Figure 13. T90 *white* mutants retain symbiosis competence. *L. japonicus* ecotype Gifu, T90 and T90 *white* mutants (T90 *white1* and/or *white2*) were grown in **a,** a nitrogen-rich medium (15 mM KNO_3) in the absence of symbiont or **b,** a nitrogen-poor medium (100 μM KNO_3) and inoculated with *M. loti Ds*Red. Boxplots display **a**, the shoot length, root length or shoot dry weight measured 24 days post transfer from the germination medium to the nitrogen-rich medium; or **b**, number of nodules quantified at indicated dpi. n: number of plants analysed. Bars, 1 mm unless labeled. Statistical method was ANOVA with *post hoc* Tukey: **a**, boxplots from left to right, $F_{3,115} = 48.08$, $p < 2\text{x}10^{-16}$; $F_{3,115} = 5.29$, $p = 0.002$; $F_{3,103} = 4.881$, $p = 0.003$; **b**, boxplots from left to right $F_{2,52} = 1.418$, $p = 0.251$; $F_{2,55} = 1.514$, $p = 0.229$; $F_{2,53} = 12.49$, $p = 3.6\text{x}10^{-5}$; $F_{2,53} = 0.527$, $p = 0.593$; $F_{2,55} = 1.728$, $p = 0.187$. Different small letters above the boxplots indicate significant difference.

Screen B of 709 M_2 lines for loss of *M. loti*-induced GUS activity resulted in three lines that exhibited heritable aberrant GUS phenotypes (unpublished; Simone Bucerius). In detail, three M_2 plants (L8668 and L8686-8687, progeny from M1 plant T614 and T1305, respectively), were identified that did not exhibit blue staining after incubation with *M. loti*. Based on the white colour of their roots after GUS staining, these three plants were renamed T90 *white* mutants (L8668 *white1*, L8686 *white2* and L8687 *white3*). The progeny of all three T90 *white* plants displayed normal shoot and root morphology and could successfully establish AM and RNS similarly as T90 and *L. japonicus* Gifu (Fig. **13**), however GUS activity could not be detected in their roots during both symbioses (Fig. **12b; S1a-b**). T90 *white2* was less healthy than *white3*

and produced limited seeds at the time of this study; therefore, only the progeny of T90 *white3* was included in some subsequent experiments.

3.2.2 Transgenic insertion of a T90 promoter:GUS fusion in the T90 *white* mutant background restored symbiosis-inducible *GUS* expression

Based on the observed dependency of T90 *GUS* expression on genes involved in early symbiotic signalling, together with the success of T90 *white* to form both RNS and AM, we hypothesised that these symbiosis genes are likely functional in T90 *white*. We consequently directed our focus onto the regulatory region of the T90 *GUS* gene. To this end, we cloned a chimeric region of 2530 bp directly 5' of the *GUS* gene in T90, hereafter called "T90 promoter". This region comprised a 1942 bp fragment positioned between -2870 bp to -929 bp relative to the transcriptional start site of *CBP1*, followed at the 3' end by 588 bp of the T-DNA sequence 5' of the ATG of the *GUS* gene (Fig. **14a**; **3a**). This fusion is identical to the original T90 fusion and contains all elements necessary for the transcription of the *GUS* gene, such as a minimal promoter and a transcriptional start site (Jefferson et al., 1987; Topping et al., 1991). We transformed *L. japonicus* Gifu hairy roots with T-DNAs containing a *GUS* reporter gene driven by the T90 promoter (*T90$_{pro}$:GUS*) and analysed the *GUS* expression in the transgenic roots followed by inoculation with *M. loti Ds*Red. Upon exposure of roots to X-Gluc, blue staining indicative of *GUS* expression was detected exclusively in root hairs (Fig. **15b**) and nodules (Fig. **14b**). By contrast, *T90$_{pro}$:GUS*-transformed roots grown in the absence of microsymbionts or roots transformed with an identical construct in which the T90 promoter was replaced by a 4 bp spacer sequence did not exhibit any blue staining (Fig. **14b**). The staining pattern achieved by

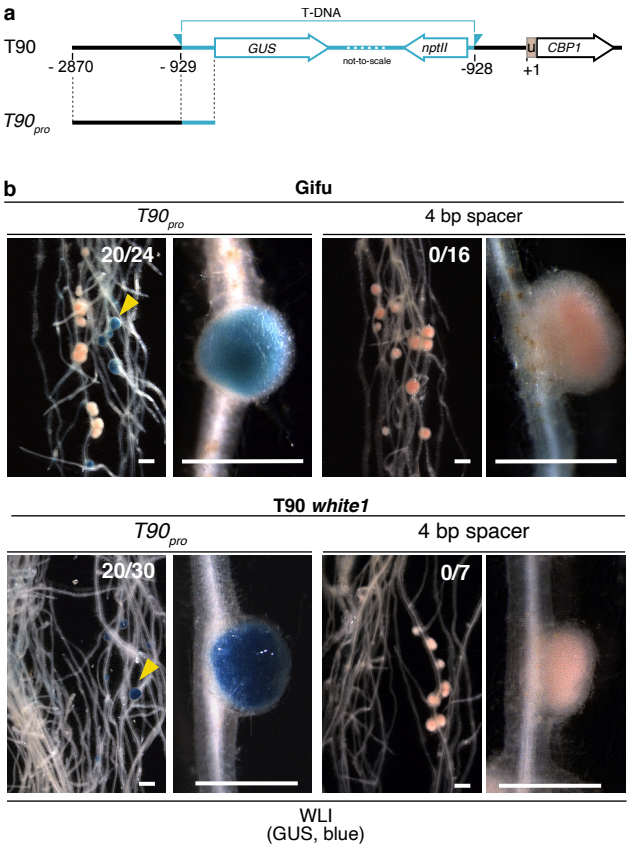


Figure 14. Absence of GUS activity in T90 *white* mutants can be restored by transgenic *T90$_{pro}$:GUS* fusions. a, T-DNA insertion in the T90 line as in Fig. 3, here including the *CBP1* promoter reference coordinates used for definition of the length of the individual fragments tested in the T90 promoter deletion series (for symbols refer to Fig. 1b). **b**, *L. japonicus* Gifu or **c**, T90 *white1* hairy roots transformed with T-DNAs carrying a *Ubq10$_{pro}$:NLS-GFP* transformation marker together with a *GUS* reporter gene driven by the full length T90 promoter (-2870) or a 4 bp spacer sequence were analysed at 14 dpi with *M. loti Ds*Red. yellow arrows: nodules displaying GUS activity . #/#: number of plants showing GUS activity in nodules / total number of transgenic root systems analysed. Bars, 1 mm unless labeled. WLI: white light illumination.

the *T90$_{pro}$:GUS* fusion in transgenic roots resembled strongly the pattern observed in T90 completely matching each other in all key aspects investigated: expression in response to *M. loti* inoculation exclusively in root hairs in early infection stage and later in nodules, which eventually disappeared in mature nodules.

We sequenced the corresponding T90 promoter region of the three T90 *white* mutants and could not detect any sequence alteration (data not shown). We consequently hypothesised that these mutants may suffer from an epigenetic change that renders its corresponding T90 promoter region non-functional. To test this hypothesis, T90 *white1* and *white3* hairy roots were

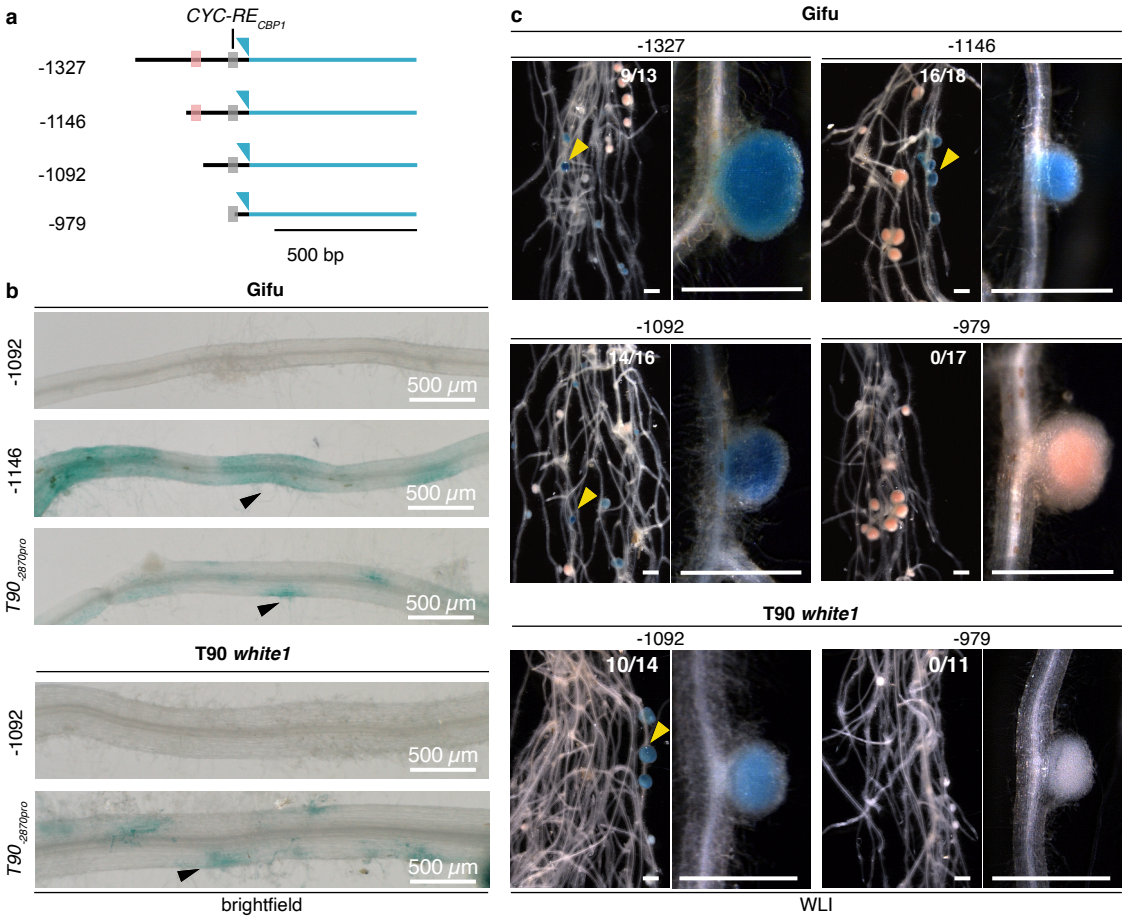


Figure 15. Two regions of the T90 promoter confer tissue specific gene expression during RNS. a, graphic illustrations of the promoter regions driving the *GUS* reporter gene. blue half triangle, the T-DNA insertion site projected onto the *CBP1* promoter; grey and pink boxes, *CYC-RE$_{CBP1}$* and *EPRE$_{CBP1}$*. **b** - **c**, *L. japonicus* ecotype Gifu or T90 *white* mutant hairy roots transformed with either of the deletion series (starting at -1327, -1146, -1092 and -979 of the *CBP1* promoter) that was generated based on the *T90$_{pro}$:GUS* reporter fusion. Note the inability to drive reporter expression in root hairs when the region between -1146 and -1092 bp was deleted in **b**. GUS activity was lost when more than half of *CYC-RE$_{CBP1}$* was deleted from the promoter region (see -979 bp region in **c**). Black and yellow arrowheads: GUS activity in root hairs and nodules, respectively. #/#, number of plants showing GUS activity in nodules / total number of transgenic root systems analysed. WLI: white light illumination. Bars, 1 mm unless labeled.

transformed with the *T90$_{pro}$:GUS* reporter fusion construct which gave nodulation-specific *GUS* expression in *L. japonicus* Gifu hairy roots and analysed the *GUS* expression after inoculation with *M. loti Ds*Red. In the absence of microsymbionts, no blue staining was detected. After inoculation with *M. loti D*sRed, blue staining could be observed in both areas characteristic for T90: patches of root epidermal cells (Fig. **15b**) and in the inner tissue of nodules (Fig. **15c**). Hairy root transformation did not result in a revival of the T90 endogenous *GUS* gene, as T90 *white*1 hairy roots transformed with a *GUS* reporter gene driven by a 4 bp spacer sequence did not exhibit any blue staining (Fig. **14b**). Because the endogenous *GUS* genes in the T90 *white* mutants were not induced during nodulation, any detected blue staining could only result from expression of the introduced reporter gene. When these transgenic roots were grown in the absence of microsymbionts, no blue staining was observed. These observations suggested that the machinery targeting the T90 promoter to induce gene expression during nodulation is intact in T90 *white* mutants and supported the hypothesis that epigenetic changes block the expression of the endogenous *GUS* genes in T90 *white* mutants.

3.2.3 A 54 bp and a 113 bp region in the T90 promoter are required for tissue specific expression

To further dissect the promoter and identify relevant regions and *cis*-elements, we generated 5' deletion series of the T90 promoter in the context of the *T90$_{pro}$:GUS* reporter fusion (starting at -2870 bp relative to the *CBP1* transcriptional start site), without modifications of the rest of the promoter or reporter. The resulting individual constructs started at -1327, -1146, -1092 and -979 bp (Fig. **15a**) and were introduced individually into *L. japonicus* Gifu or T90 *white1* hairy roots. We observed the characteristic blue patches of root epidermal cells on Gifu and T90 *white* roots transformed with the *GUS* reporter gene driven by *T90$_{pro}$* or a shorter promoter -1146 bp (Fig. **14c**). The epidermal blue staining pattern was no longer detected when a region -1092 bp was tested (Fig. **15b**). This 54 bp region (-1146 to -1092 bp) was therefore called the "*Epidermal Patch Response Element in the CBP1 promoter*" (*EPRE$_{CBP1}$*). Blue staining was observed in nodules transformed with either of the T90 promoter:GUS fusions starting at -1327, -1146 or -1092 bp (Fig. **15b-c**). Further deletion to -979 bp, as well as promoter replacement with a 4 bp spacer sequence eliminated blue staining in nodules (Fig. **15c**). An identical pattern was observed in hairy roots of T90 *white1* where a -1092 bp region could achieve GUS activity in transgenic nodules, but not a -979 bp region (Fig. **15b**). We concluded based on these findings that a region of 113 bp (-1092 to -979 bp) and a stretch of 54 bp located directly 5' (-1146 to -1092 bp; *EPRE$_{CBP1}$*) was necessary for gene expression in nodules and the root epidermis, respectively.

3.2.4 T90 promoter hypermethylation was detected in three T90 *white* mutants

DNA methylation is an important and frequently occurring driver of epigenetic changes, which can attenuate binding by the transcription regulatory proteins, thereby inhibiting the activation of target genes (Medvedeva et al., 2014; Yang et al., 2020). We investigated whether an epigenetic event interfering with the endogenous *GUS* expression in T90 *white* mutants could be related to DNA methylation. To detect differences in the methylation pattern between

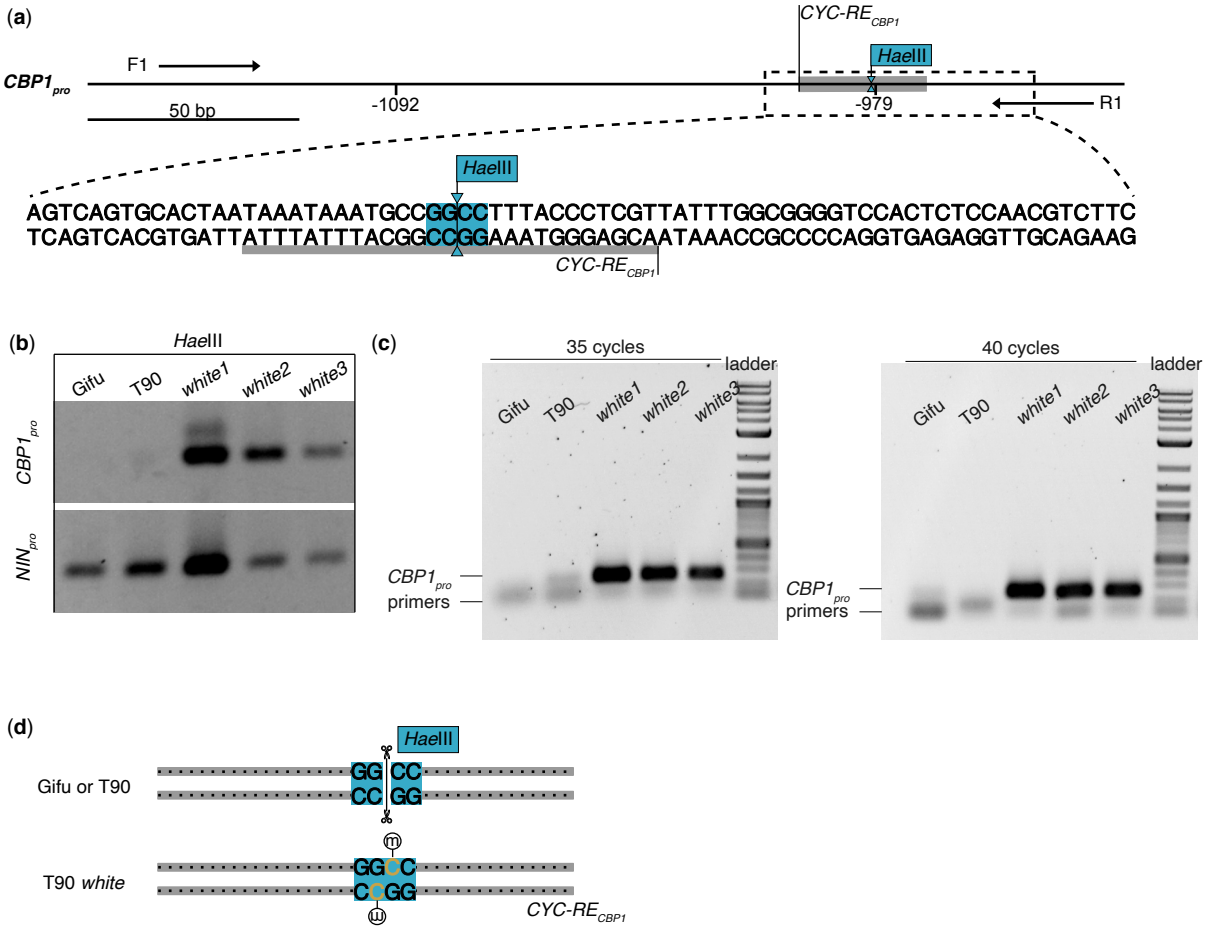


Figure 16. Cytosine methylation within a 113 bp T90 promoter region of T90 *white* mutants but not those of T90 or *L. japonicus* Gifu. a, Genomic DNA (gDNA) from *L. japonicus* Gifu, T90 or T90 *white* mutants (*white1*, *white2* or *white3*) was digested with the methylation-sensitive restriction enzyme *Hae*III. Blue shade on DNA sequence, recognition site of *Hae*III. Grey shade in the restriction map or grey underline of promoter sequence, *CYC-RE*$_{CBP1}$. Arrowheads and lines, endonucleolytic cleavage site and outline of the restriction digestion products. **b**, Analysis of the success of restriction digestion by PCR. Digested gDNA from the indicated genotype was used as a template for PCR amplification with primers F1 and R1 flanking a 218 bp promoter region (top panel) and primers flanking a 198 bp stretch of the *L. japonicus NIN* promoter as control (bottom panel). This region in the *NIN* promoter does not contain the restriction site for *Hae*III. Note that PCR products could be obtained using digested gDNA from the T90 white mutants but not from Gifu or T90 as the amplification template. **c**, the same PCR as in **b** was repeated with 35 or 40 cycles. **d**, Graphic summary of the results in **b** projected onto the promoter region together with the recognition sites of *Hae*III. m in an open circle, methyl groups. Scissor cartoon: successful endonucleolytic cleavage.

T90 and T90 *white* mutants, we took advantage of the DNA restriction endonuclease *Hae*III whose activity is impaired by cytosine methylation in its GGCC recognition site (rebase.neb.com). Its recognition site is present in the short 113 bp region, deletion of which led to a complete absence of *GUS* gene expression during nodulation (Fig. **15c**). To detect

differential methylation at this site, we performed restriction digestion of genomic DNA (gDNA) extracted from roots of *L. japonicus* Gifu, T90 and T90 *white* mutants grown in the absence of microsymbionts, followed by Polymerase Chain Reaction (PCR). PCR based amplification of a gDNA region containing a *Hae*III recognition site and digested with *Hae*III is only successful when its recognition site is methylated and thus protected from cleavage. A promoter region of the *NIN* gene that did not contain any recognition site could be successfully amplified with digested gDNA from all genotypes, demonstrating that the gDNA quality was suitable for PCR after restriction digestion (Fig. **16b**; bottom panel). In contrast, a PCR product using the primers covering the *CBP1* promoter region displayed in Fig 16a was only obtained using digested gDNA from T90 *white* mutants as the amplification template, but not with that from Gifu or T90 (Fig. **16b**; top panel). Amplicons for the latter two were also not detected when increasing the PCR amplification cycle to 35 or 40 (Fig. **16c**), indicating a complete *Hae*III digestion of the gDNA. Embedded into the context of *CYC-RE*$_{CBP1}$, the critical C for *Hae*III's methylation sensitivity falls into the CHH sequence pattern of cytosine methylation sites in plants, frequently associated with epigenetic transcriptional gene silencing (Iwasaki & Paszkowski, 2014). Taken together, the so discovered differential methylation within the *CYC-RE*$_{CBP1}$ is likely the cause for the loss of *GUS* expression in the T90 *white* mutants (Fig. **16d**).

3.2.5 *CBP1* is regulated by the CCaMK/Cyclops complex via a *cis*-element

The observations that the T90 *GUS* gene expression can be induced by exposure to rhizobia and an AMF as well as spontaneously by autoactive CCaMK suggested that the T90 promoter is likely subject to CCaMK/Cyclops regulation. We used transient expression assays in *N. benthamiana* leaves to test whether the CCaMK/Cyclops protein complex could transcriptionally induce expression of a *GUS* reporter gene under the control of the CBP1 promoter or the T90 promoter (Fig. **17; S2**). A 2870 bp region 5' of the transcriptional start site of CBP1 was cloned together with the 177 bp 5' untranslated region (UTR) of *CBP1* from *L. japonicus* Gifu (*CBP1*$_{-2870pro}$). The expression of the reporter gene driven by *T90*$_{pro}$ or *CBP1*$_{-2870pro}$ was induced in the presence of Cyclops and the autoactive CCaMK$^{1-314}$ (CCaMK$^{1-314}$/Cyclops; Fig. **17a; S2b**). In addition, *T90*$_{pro}$ achieved transcriptional activation mediated by CCaMKT265D/Cyclops (Fig. **S2a**). A 928 bp stretch of *CBP1* promoter corresponding to the region 3' to the T90 T-DNA insertion site (*CBP1*$_{-928pro}$; fused to the *CBP1* 5' UTR in the reporter fusion) did not achieve reporter gene induction by CCaMK$^{1-314}$/Cyclops (Fig. **17a; S2b**). These observations together indicated the presence of putative *cis*-regulatory elements responsive to CCaMK/Cyclops-mediated transactivation between -2870 and -928 bp of the *CBP1* promoter.

To identify the CCaMK/Cyclops-responsive *cis*-regulatory element, we generated a promoter 5' deletion series and investigated reporter gene activation by CCaMK$^{1-314}$/Cyclops-mediated transactivation in transient expression assays in *N. benthamiana* leaves (Fig. **17b-d; S2c-e**). The 5' deletion series was built on the basis of *T90*$_{pro}$*:GUS* (constructed in the same way as those tested in Fig. **15a**). Each construct comprises a *CBP1* promoter stretch of variable length. The nucleotide position at the 5' end of the deletions is based on the coordinates of the *CBP1* promoter (Fig. **3a**). An initial comparison of the transactivation strength across the deletion series revealed a reduction to approximately 50% when comparing

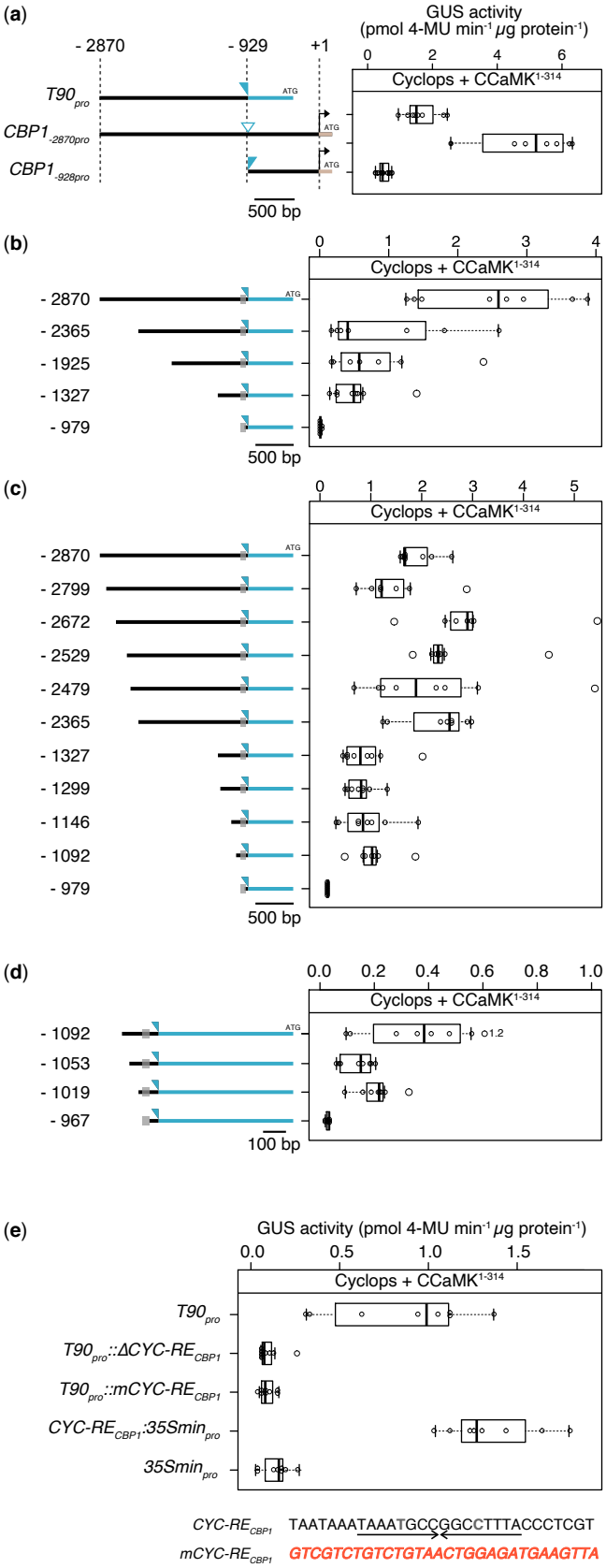


Figure 17. *A cis*-element in the promoter of *CBP1* is necessary and sufficient for the CCaMK$^{1-314}$/Cyclops-mediated transactivation of the reporter gene in *N. benthamiana* leaf cells. *N. benthamiana* leaf cells were transformed with T-DNAs carrying a *GUS* reporter gene driven by either of the indicated promoters: **a**, the T90 promoter (labeled as $T90_{pro}$ in a&d or the simplified version -2870 in b-c; see Fig. 15 legend); one of the two *CBP1* promoter regions ($CBP1_{-2870pro}$ or $CBP1_{-928pro}$); **b-d**, promoter deletion series generated in the context of $T90_{pro}$ including promoter regions that were **b**, ca. 300 - 500 bp different in length; **c**, ca. 50 to 100 bp different in length within -2870 to -2365 bp or -1327 to -979 bp; **d**, ca. 35 - 50 bp different in length within -1092 to -967 bp; **e**, $T90_{pro}$, $T90_{pro}$ with the 30 nt long *cis*-element ($CYC\text{-}RE_{CBP1}$) mutated or deleted ($T90_{pro}$::$mCYC\text{-}RE_{CBP1}$ or $T90_{pro}$::$\Delta CYC\text{-}RE_{CBP1}$, respectively), a 35S minimal promoter ($35Smin_{pro}$), or $CYC\text{-}RE_{CBP1}$ fused to $35Smin_{pro}$ ($CYC\text{-}RE_{CBP1}$:$35Smin_{pro}$). The numbers in **a-d** indicating length of promoter were based on *CBP1* promoter taking its transcriptional start site as +1. Left of the boxplots in **a-d** are graphic illustrations of the promoter regions driving the *GUS* reporter gene with the open triangle and grey boxes illustrating the T-DNA insertion site projected onto the *CBP1* promoter and of $CYC\text{-}RE_{CBP1}$, respectively. Blue and black line indicates sequence originating from *L. japonicus* wild-type genomic sequence and that from the T-DNA sequence in T90, respectively. The larger experimental set-up boxplots including the results of negative controls is presented Fig. S2 with statistical tests.

-2780 and -2365 bp and a complete loss of activity when comparing -1327 to -979 bp (Fig. **17b**), indicating that these two regions might contain the responsible *cis*-elements. Testing further deletions that were ca. 100 bp different in length within these two regions revealed that the series between -2870 and -2365 exhibited large variations in responsiveness between replicates and was therefore not investigated further. In the -1327 to -979 series, fragments equal to or longer than -1092 resulted in similar reporter gene activation mediated by

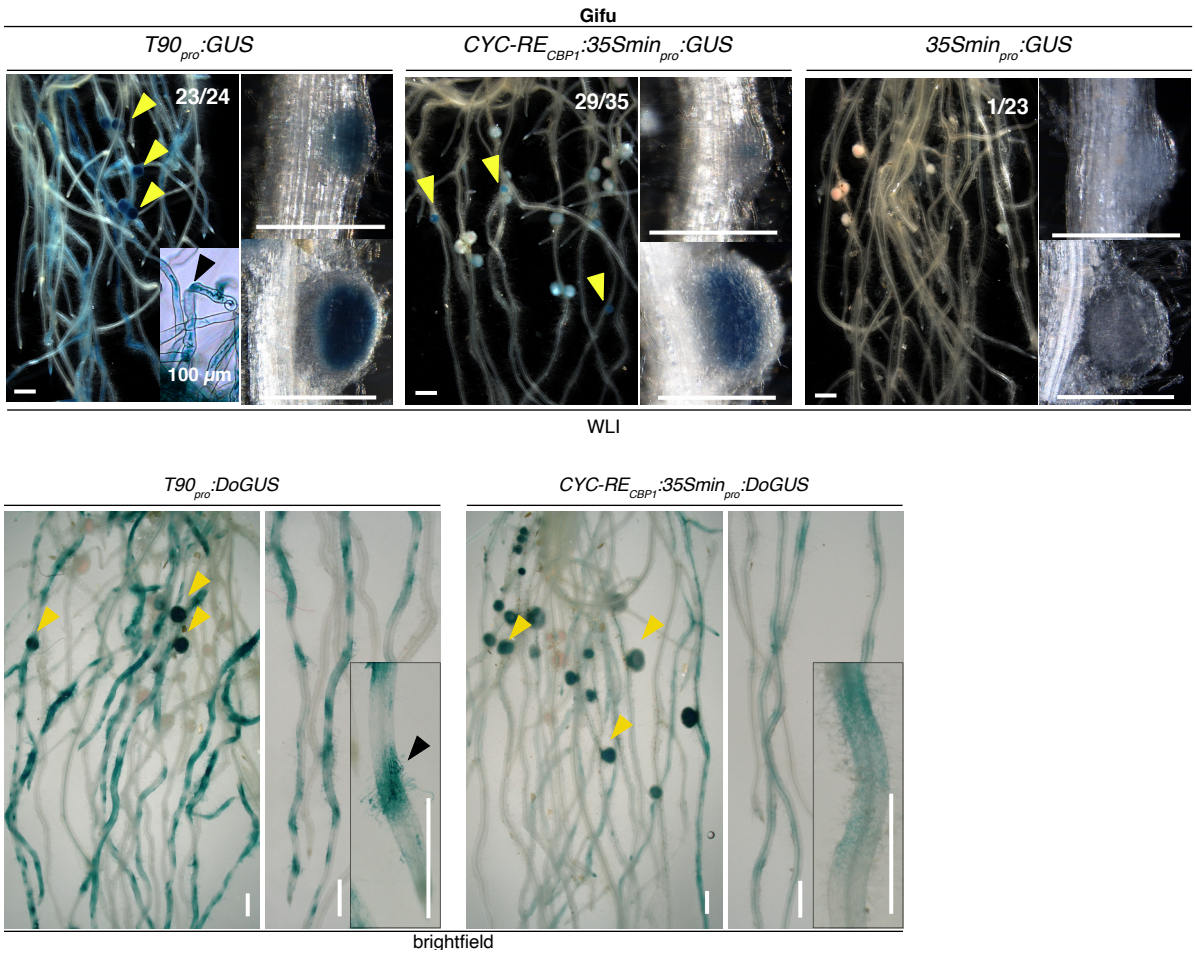


Figure 18. *CYC-RE$_{CBP1}$* drives gene expression in *L. japonicus* hairy roots during nodulation. *L. japonicus* Gifu hairy roots transformed with T-DNAs carrying a *Ubq10$_{pro}$:NLS-GFP* transformation marker together with a *GUS* or *DoGUS* gene, driven by the T90 promoter (*T90$_{pro}$*); a 35S minimal promoter (*35Smin$_{pro}$*) or *CYC-RE$_{CBP1}$* fused to a 35S minimal promoter (*CYC-RE$_{CBP1}$:35Smin$_{pro}$*) were stained with X-Gluc 10 - 14 dpi with *M. loti Ds*Red. Note that the T90 promoter could drive *GUS* expression in root hairs (black arrowheads) but *CYC-RE$_{CBP1}$:35Smin$_{pro}$* could not. Overall the T90 promoter gave stronger GUS activity (darker and more widespread) in nodules than *CYC-RE$_{CBP1}$:35Smin$_{pro}$* (yellow arrowheads; compare the overview images of root systems). Note that roots transformed with *35Smin$_{pro}$:GUS* did not show exhibit GUS activity during nodulation except for GUS activity in vasculature in rare cases. Black and yellow arrowheads: GUS activity in root hairs and nodules, respectively. #/#, number of plants showing GUS activity in nodules out of total number of chimeric root systems analysed. WLI: white light illumination. Bars, 1 mm unless labeled.

CCaMK$^{1-314}$/Cyclops, while construct -979 was inactive, suggesting the presence of relevant *cis*-element(s) between -1092 and -979 bp (Fig. **17c; S2c**). By testing a higher resolution series with 35 bp to 50 bp length difference, we narrowed down the responsible cis-element to the 30 nucleotides between -997 and -967 bp that contained an almost perfect (only two non-matching basepairs; Fig. 5e) palindromic sequence of 16 bp. We called this element "*Cyclops-response element within the CBP1 promoter*" (*CYC-RE$_{CBP1}$*) because a loss of reporter gene induction by CCaMK$^{1-314}$/Cyclops was observed when this element was deleted (Fig. **17d; S2d**). To test the relevance of *CYC-RE$_{CBP1}$* in the context of the T90 promoter, we mutated or deleted *CYC-RE$_{CBP1}$* (*T90$_{pro}$::mCYC-RE$_{CBP1}$* and *T90$_{pro}$::ΔCYC-RE$_{CBP1}$*, respectively). Both resulted in an almost complete loss of CCaMK$^{1-314}$/Cyclops-mediated transcriptional activation, indicating that *CYC-RE$_{CBP1}$* was essential for this transcriptional activation (Fig. **17e; S2e**). Moreover, *CYC-RE$_{CBP1}$* fused to a 35S minimal promoter (*CYC-RE$_{CBP1}$:35Smin$_{pro}$*) was sufficient for the activation of reporter gene (Fig. **17e; S2e**). These results together indicated that *CBP1$_{pro}$* (and *T90$_{pro}$* in the context of T90 genome) is regulated by the CCaMK/Cyclops complex through a cis-regulatory element, *CYC-RE$_{CBP1}$*.

3.2.6 *CYC-RE$_{CBP1}$* drives gene expression during RNS and AM

CYC-RE$_{CBP1}$ is located only 39 bp 5' of the T-DNA insertion site in T90 and we noticed that *CYC-RE$_{CBP1}$* sits within the hypermethylated region in the T90 white mutants (Fig. **17**). Given the necessity and sufficiency of *CYC-RE$_{CBP1}$* for CCaMK/Cyclops-mediated transcriptional activation (Fig. **17**) as well as the common requirement of this protein complex in AM and RNS, we hypothesised that this *cis*-element might be responsible for the symbioses-specific *GUS* expression in T90. To test this, a *GUS* or *DoGUS* (a variant of *GUS*) gene driven by *CYC-RE$_{CBP1}$* fused to a 35S minimal promoter (*CYC-RE$_{CBP1}$:35Smin$_{pro}$*) or the T90 promoter (*T90$_{pro}$*) was introduced into *L. japonicus* Gifu hairy roots, followed by inoculation with *M. loti Ds*Red or the AMF *R. irregularis* (Fig. **18**). During nodulation, GUS activity in roots transformed with *CYC-RE$_{CBP1}$:35Smin$_{pr}$:GUS* exhibited blue staining specifically in nodule primordia and nodules, but not root hairs (Fig. **18**). In contrast, T90pro:GUS- transformed roots displayed a much broader GUS activity in epidermis including root hairs, in addition to that observed in nodule primordia and nodules (Fig. **18**, bottom panel). The same promoter:reporter fusions constructed with *DoGUS* instead of *GUS* led to similar results (Fig. **18**). During mycorrhization, blue staining were detected in segments in roots transformed with *T90$_{pro}$:DoGUS* and correlated strongly with the presence of *R. irregularis*, at the entry site of fungal hyphae crossing the epidermis and in cortical cells containing arbuscles (Fig. **19**). Blue staining in roots transformed with *CYC-RE$_{CBP1}$:35Smin$_{pro}$:DoGUS* could be specifically detected in the cortex in segments of roots, where cells were infected by *R. irregularis* (Fig. **19**). In both cases, GUS activity was visibly stronger in cells that were just invaded or had developing arbuscles, compared to those that the arbuscles almost occupying the entire cells. By contrast, roots transformed with GUS or *DoGUS* driven by the 35S minimal promoter did not display GUS activity during RNS or AM. Roots transformed with either one of the mentioned fusion constructs grown in the absence of microsymbionts exhibited only rarely blue staining, and if so, in vasculature or root tips

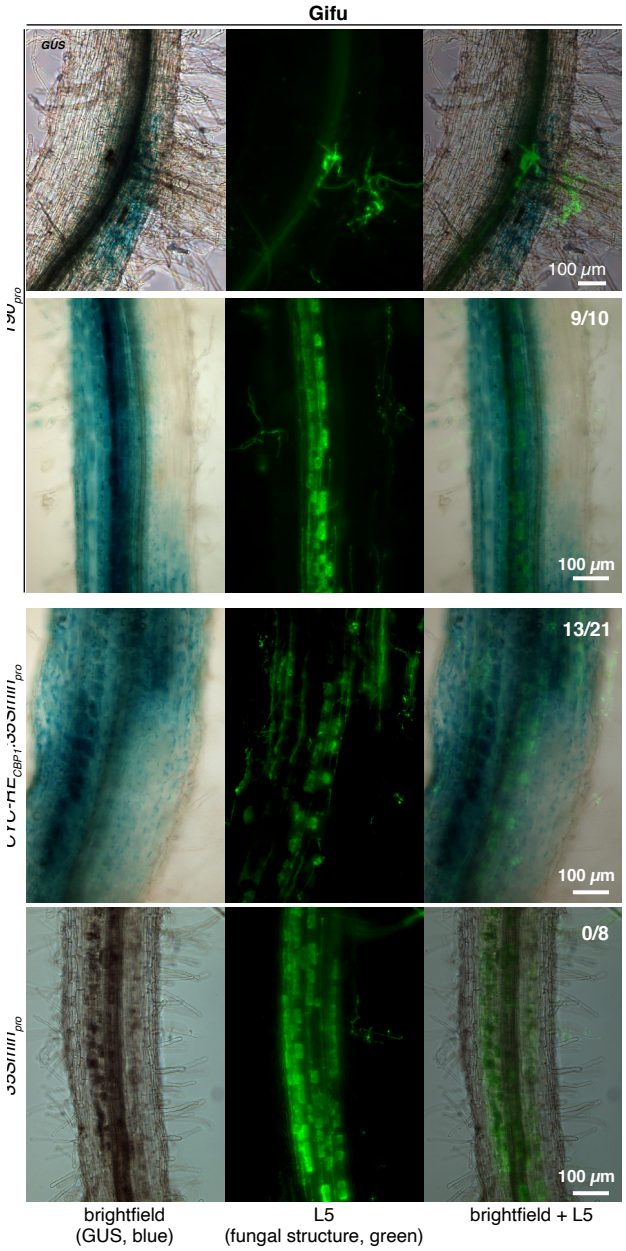


Figure 19. *CYC-RE*$_{CBP1}$ drives gene expression in *L. japonicus* hairy roots during AM. *L. japonicus* Gifu hairy roots transformed with T-DNAs carrying a *Ubq10*$_{pro}$*:NLS-GFP* transformation marker together with a *DoGUS* gene driven by promoters indicated in Fig. 19. Transformed roots stained with X-Gluc 12 dpi with AMF *R. irregularis*. Note the overlapping GUS activity and fungal infection structures in root cortex transformed with *DoGUS* equipped with *T90*$_{pro}$ or *CYC-RE*$_{CBP1}$*:35Smin*$_{pro}$. Roots transformed with *35Smin*$_{pro}$*:DoGUS* did not exhibit any GUS activity during mycorrhization except for GUS activity in vasculature in rare cases. Green: Alexa Fluor-488 WGA-stained *R. irregularis* visualised with a Leica Filter Cube L5. #/#, number of root systems showing GUS activity in root cortex out of total number of chimeric root systems analysed.

regardless of the reporter fusion. We concluded that *CYC-RE*$_{CBP1}$ confers AM- and RNS- related gene expression specifically in the fungal-colonised root cortical cells and in nodules, respectively.

3.2.7 *CBP1* promoter drives reporter expression during nodulation

We observed that *CYC-RE*$_{CBP1}$ in the context of the T90 promoter mediated responsiveness to transactivation by CCaMK/Cyclops and conferred gene expression during symbioses. The T-DNA insertion in T90 physically separated the promoter of *CBP1* into two regions: one containing *CYC-RE*$_{CBP1}$ located 5' of the insertion (5' region) and the other 3' of the insertion (3' region). It has been hypothesised previously that the 5' region enhances *CBP1* expression during symbiosis while the 3' region was responsible for its basal expression (Tuck, 2006). To investigate the role of the 3' region in more detail, we generated *L. japonicus* Gifu

hairy roots transformed with a *GUS* reporter gene driven by $CBP1_{-2870pro}$ that represents the "native" full length promoter comprising both the 5' and the 3' region or $CBP1_{-928pro}$ consists of only the 3' region (Fig. **3a; 17a**). Transgenic roots were analysed 14 or 21 dpi with *M. loti Ds*Red for *GUS* expression (Fig. **20**). $CBP1_{-2870pro}$*:GUS*-transformed roots exhibited strong blue staining in nodules, vasculature tissue, lateral root primordia and root tips (93% of transgenic root systems displaying blue staining in nodules). In comparison, blue staining in $CBP1_{-928pro}$*:GUS*-transformed roots was observed in the same tissue and organ types, however at a lower efficiency (ca. 50% of transgenic root systems displaying blue staining in nodules), and the blue staining was overall visibly weaker in nodules. These observations were consistent with the hypothesis that the 5' region enhances *CBP1* expression during nodulation.

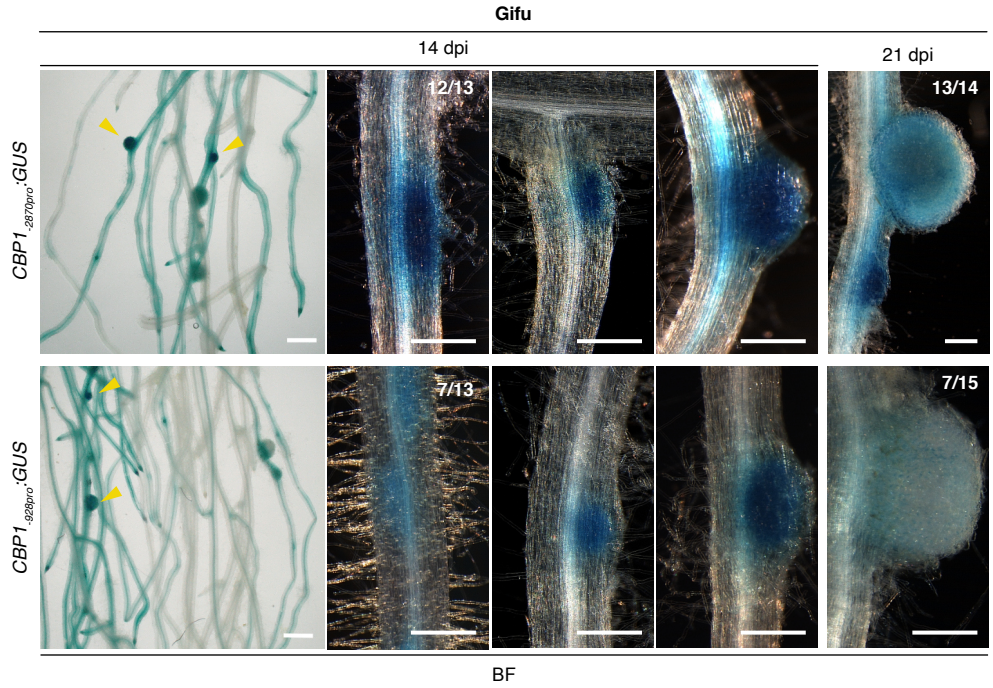


Figure 20. *CBP1* promoter-driven reporter gene expression during nodulation in *L. japonicus* roots. *L. japonicus* ecotype Gifu hairy roots transformed with T-DNAs carrying a $Ubq10_{pro}$*:NLS-GFP* transformation marker together a *GUS* reporter gene driven by either of the two *CBP1* promoter regions: a 2870 bp region containing $CYC\text{-}RE_{CBP1}$ ($CBP1_{-2870pro}$) or a 928 bp region that did not contain $CYC\text{-}RE_{CBP1}$ ($CBP1_{-928pro}$), were stained with X-Gluc at indicated dpi with *M. loti Ds*Red. Note that only ca. 50% roots transformed with $CBP1_{-928pro}$*:GUS* had blue staining in nodules compared to over 93% of those transformed with $CBP1_{-2870pro}$*:GUS*. #/#, number of plants showing GUS activity in nodules / total number of transgenic root systems analysed. Bars, 1 mm. WLI: white light illumination.

4. DISCUSSION

4.1 Acquisition of *cis*-regulatory elements for the evolution of nodulation

Combinatory efforts of *cis*-elements (such as enhancers and core promoters) and associated transcription regulatory proteins (such as transcription factors, TFs) in a transcriptional network determine the spatiotemporal gene expression. Evolutionary studies on *Drosophila* and fungal lineages demonstrate that alterations or emergence of both *cis*-regulatory elements as well as protein-protein interactions between transcriptional regulators contribute significantly to phenotypic novelties and variations (Wagner and Lynch, 2010). Evolution of *cis*-regulatory elements, especially transcriptional enhancers, is often stressed as a source of novel traits (e.g., pelvic loss in threespine stickleback fish occurred through mutations in enhancer regions; Chan et al., 2010). This is due to modularity and high-turnover rates of enhancers that potentially allow gain of gene expression in discrete spatiotemporal domains with little risk of pleiotropy. RNS is postulated to co-opt genes that are functional in other processes such as the most wide-spread plant root endosymbiosis, the AM, as well as in lateral root development. However, little is known about the involvement of *cis*-elements in these co-option events. This study provides evidence that the acquisition of key *cis*-elements contributes to evolution of RNS by two case studies of RNS-related *cis*-elements. The findings of this study also demonstrate that the acquisition may or may not be crucial for the establishment of RNS; the outcome is likely dependent on the role of the target genes and their relevant cellular responses.

4.1.1 *PACE*, the *cis*-element linking microsymbiont signalling to infection

With the growing number of available plant genomes, phylogenomic approaches offer the opportunity to identify *cis*-regulatory elements that might play a crucial role in the evolution of RNS. *PACE* was identified *via* a phylogenomic approach covering a range of RNS-forming and non-RNS-forming plant species. *PACE* turned out to be the only element in the tested *NIN* promoters that is exclusively present in those from members of the FaFaCuRo clade, sharing the same phylogenetic distribution as RNS (1.4.2). *Cis*-elements that were restricted to the FaFaCuRo clade could not be detected in the promoters of other members of the same gene family as *NIN*, indicating that the acquisition of *PACE* might have taken place later than the duplication event that created *NIN*. The strong phylogenetic signature of *PACE,* combined with the fact that NIN is the first RNS-specific TF activated in response to bacterial signalling, underlines the importance of *PACE* for RNS and its potential association with the origin of RNS.

It is important to point out that current evidence can not rule out the presence of other *cis*-elements exhibiting a *PACE*-like distribution. Altogether, the targeted phylogenetic approach was limited in two respects: (1) the search was limited to promoter regions 5' upstream of the coding sequence. The choice is justified since most reported symbiosis-relevant regulatory elements are situated in close vicinity 5' of the translational start sites of genes (within a few kb distance; Table S1). Genome-wide analyses of plant enhancers in other plant species also demonstrate occurrence of the majority of enhancers in the proximal region of genes (Yan et

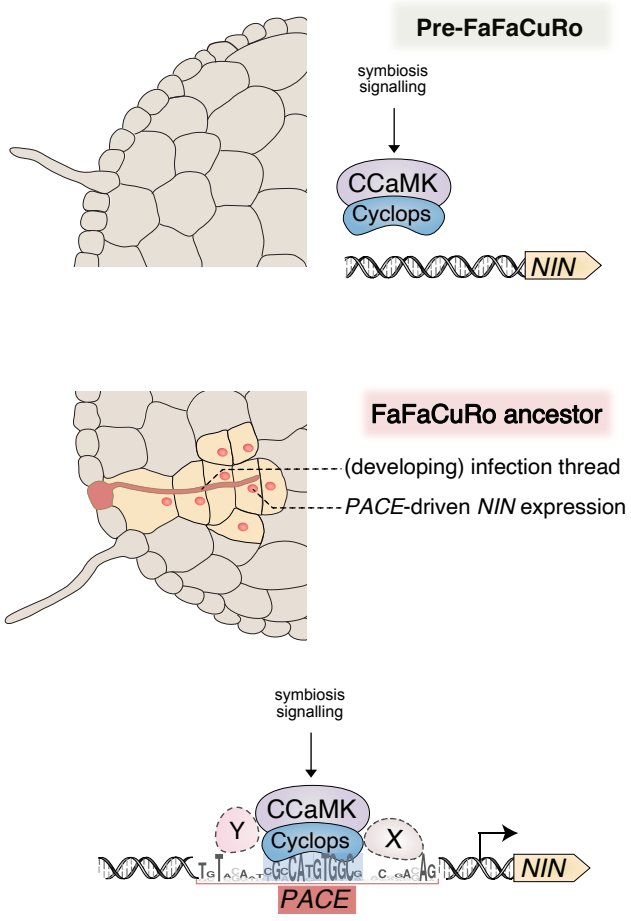
al., 2019). Nevertheless, distant regulatory regions and untranslated regions that are not covered in this search have also been reported functional in RNS (Liu et al., 2019c; Soyano et al., 2019). (2) The search was only conducted for the promoters of a handful of known symbiosis genes, although the rationale of choosing these candidate genes is well justified. To complete the search, proximal and distal regulatory regions of all symbiosis-related genes known thus far should be included in the analysis. Yet, such an analysis would entail technical challenges regarding the identification of orthologous and equivalent regulatory regions in plants from different orders. If a genome-wide list of RNS-relevant enhancers was available, it could be used for phylogenomic approaches for targeted search of *cis*-elements contributing to the evolution of nodulation.

PACE contains a Cyclops binding site (Singh et al., 2014) and we observed that *PACE* is essential for Cyclops-mediated transcriptional activation of the *NIN* promoters across the FaFaCuRo clade and sufficient for this purpose (Fig. **5**) in transactivation assays in *N. benthamiana* leaves. Additionally, despite the considerable degree of sequence variation (only 7 out 29 nucleotides are completely conserved), *PACE* appeared to be functionally conserved across the FaFaCuRo clade when tested in the context of the *L. japonicus NIN* promoter. Congruent the phylogenomic distribution of *PACE*, Cyclops-mediated transcriptional activation could be conferred by the *NIN* promoters from RNS-forming members of the FaFaCuRo clade. The *NIN* promoters from two species outside of this clade, *Solanum lycopersicum* (tomato) and *Vitis vinifera* (grape), could not mediate transcriptional activation by Cyclops (unpublished data included in publication 1). Two motifs were also identified in the promoters of *Reduced Arbuscular Mycorrhiza 1* (*RAM1*) and *ERF Required for Nodulation1* (*ERN1*) that encompass or partially overlap with previously identified Cyclops response elements, respectively (Pimprikar et al., 2016; Cerri et al., 2017); and their distributions extend beyond the FaFaCuRo clade (unpublished data included in publication 1). In line with these data, the ability of CCaMK/Cyclops to transcriptionally induce the reporter gene expression mediated by the *RAM1* promoter is not restricted to those from the FaFaCuRo clade, i.e., both *S. lycopersicum* and *L. japonicus RAM1* promoter could mediate transcriptional activation by *Sl*Cyclops or *Lj*Cyclops (unpublished data included in publication 1). Together with *PACE*'s phylogenetic signature, these findings strongly support the hypothesis that emergence of *PACE* connected the *NIN* gene to an ancient signalling transduction pathway that had been already functional in the AM, *via* the CCaMK/Cyclops protein complex.

The consequence of *PACE* acquisition appears to be directing *NIN* expression in a discrete infection zone in nodule primordia. Importantly, while *PACE* and the rest of the *NIN* promoter (exemplified by *NIN::mPACE* and *NIN::ΔPACE*; 3.1.2) both drive *NIN* expression in the nodule primordia, their respective domains are not identical. The rest of the *NIN* promoter directs expression in the surrounding cortical cells including those bordering on the *PACE* core territory (unpublished data included in publication 1). The nearly non-overlapping expression domains suggest distinct transcriptional regulation of these two elements, and a unique role of *PACE* that is not fulfilled by the rest of the *NIN* promoter. This discrete expression domain conferred by *PACE* is consistent with the sufficiency of *PACE* to restore infection thread (IT) formation in the cortex of nodule primordia formed on roots of the *nin-15* mutant (3.1.4). The

appearance of long cortical IT originating from an accumulation of bacteria fits the expectation of a primitive entry mode (Sprent, 2007; Madsen et al., 2010), which is also observed in a variety of legumes including for example *Sesbania* and *Mimosa* as well as a *L. japonicus* root hair-less mutant (Karas et al., 2005). Combined with the discovery that cortical ITs might have been invented by the last common ancestor of the FaFaCuRo clade, these findings support the hypothesis that the acquisition of *PACE* likely enabled the common ancestor to form cortical ITs. Furthermore, *PACE* and the rest of the *NIN* promoter, respectively, are both necessary but insufficient to mediate gene expression in root hairs bearing ITs and to rescue IT formation in root hairs of *nin-15* roots. This observation suggests that collaborative efforts of multiple *cis*-regulatory elements are required for *NIN* expression in root hairs.

The inability to restore bacterial infection in root hairs and nodules of *nin-15* associated with deletion or mutation of *PACE* supported its essential role in infection during RNS. The fact that *PACE* sequence variants are functionally equivalent for this purpose further indicates the evolutionary importance and maintenance of *PACE* function. It is intriguing that two *PACE* variants identified from non-RNS-forming FaFaCuRo members are also functionally equivalent to those from RNS-forming species while the *PACE*-like motif from *J. regia* is not. The point mutation of one of these 7 conserved nucleotides in *JrPACE*-like motif likely explains its inability to restore infection in the *nin-15* mutant (Fig. **10b**). The loss or mutation pattern of *PACE* needs further analysis as it is possible that changes in *PACE* also contribute to the loss of RNS.



Interestingly, the artificial insertion of an exotic *PACE* sequence is able to render a *NIN* promoter from tomato (*Solanum lycopersicum*) - a plant that is outside of the FaFaCuRo clade and whose endogenous *NIN* promoter does not contain *PACE* - to be responsive to Cyclops-mediated transcriptional activation (Fig. **5c**). This "evolved" version of the tomato *NIN* promoter is capable of restoring infection of *nin-15* mutant to some degree both in root hairs and nodules, as well as restoring

Figure 21. Graphical illustration of how *PACE* connected *NIN* to symbiotic transcriptional regulation by CCaMK/Cyclops, enabling IT development in the root cortex. This acquisition coincided with the predisposition event. X and Y: hypothetical proteins binding to PACE outside the Cyclops binding site.

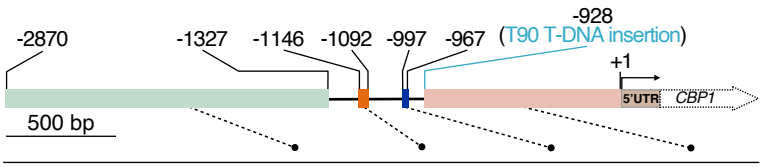
formation of symbiosomes. Combined with the observations that not-yet-identified epidermal elements and *PACE* are necessary for *NIN* expression in root hairs or in cells packed with symbiosomes, this result implies that the tomato *NIN* promoter already contains *cis*-elements that can partially substitute for the elements in *L. japonicus NIN* promoter. Further evolution of these epidermal and symbiosome-related *cis*-elements may have taken place in the FaFaCuRo members that efficiently take up bacterial symbiont through root hairs or forms bacterial symbiosome, or both, respectively.

The strong phylogenetic signature of *PACE,* its discrete expression domain and its role in IT formation as well as connection to Cyclops strongly suggest that acquisition of *PACE* was a key evolutionary event in the last common ancestor of the FaFaCuRo clade. Based on the findings presented in this work, the following model for the evolution of nodulation could be envisioned (Fig. **21**): acquisition of *PACE* resulted in a novel *NIN* expression domain due to activation by CCaMK/Cyclops and enabled cortical infection thread formation in the last common ancestor of the FaFaCuRo clade. This event consequently brought nitrogen-fixing bacteria in close contact to living plant cells. *PACE* acquisition was then followed by parallel gains of nodule organogenesis, evolution of diverse entry modes and other necessary refinements to give rise to functional nodules in the descendant lineages. Fixation threads in nodules formed on roots of actinorhizal plants, *Parasponia* and several tropical trees may be considered as an actualisation of this ancestral state. This model also implies that the *NIN*-regulon at the time of *PACE* acquisition could at least partially function in infection thread formation.

This *PACE* model may be viewed as an analogy of the emergence of *cis*-elements in *Drosophila* and fungi that expanded the regulation territory of existing transcriptional factors, leading to phenotypic novelties (Gompel et al., 2005; Jeong et al., 2006; Sorrells et al., 2018; Britton et al., 2020). The contribution of *PACE* to the evolution of RNS is not matched by the regulatory region of *NIN* promoter (*CE* element) that modulates nodule organogenesis, considering the occurrence of *CE* is restricted to only one plant order (Fabales) of the FaFaCuRo clade. The *PACE* model clearly agrees with the hypothesis proposed by Soyano and Hayashi (2014) in which the gain of novel *cis*-elements in the *NIN* promoter was predicted to be the first step in the evolution of RNS. *PACE* and *PACE*-enabled traits are good candidates for the long-sought-after features associated with the predisposition event (Soltis et al., 1995). This model is intrinsically not in conflict with the single origin model (1.2.1) because Cyclops-*PACE* connection is an important adaptation of transcriptional regulation of the *NIN* gene and must have been present when RNS evolved. It is noteworthy that a gap of at least 20 million years exists between the acquisition of *PACE* and the oldest fossil record of a nodule (Fig. **1**). Million years of maintenance implies that an evolutionary advantage or at least evolutionary neutrality was provided by the putative ancestral cortical ITs. A plausible candidate for this evolutionary advantage may be the fixed nitrogen provided by symbionts, considering that biologically fixed nitrogen can be transferred to plants that associate with endophytic or associative nitrogen-fixing bacteria to support plant growth (Carvalho et al., 2014).

4.1.2 *CYC-RE*$_{CBP1}$, the *cis*-element tagged in the T90 line

With the goal to uncover regulatory elements involved in symbiosis, the regulatory circuits underlying the symbiosis-specific *GUS* expression of the *L. japonicus* promoter tagging line T90 was investigated. We observed that the T90 *GUS* expression can be largely recapitulated in hairy roots transformed with a GUS reporter gene driven by a region between -2870 and -967 bp of the *CBP1* promoter fused to the same T-DNA border found in the genomic arrangement of the T90 line. The expression pattern achieved by this region matches all key aspects of that of the T90 line: in root hairs and nodules in presence of *M. loti*; as well as root epidermis and cortical cells when roots were colonised by an AMF; and in both symbioses, absent from other tissues such as root vasculature and root tips. We therefore used this transgenic setting as the starting point to dissect the promoter function using a classical promoter deletion series. This analysis revealed at least four regions/elements with significant impact on *CBP1* expression (Fig. **22**): **(1)** a **30 bp *CYC-RE*$_{CBP1}$** is essential for gene expression in nodules and root cortex. A 30-nucleotide long element named*CYC-RE*$_{CBP1}$ was identified within the region -997 and -967 bp which is only 39 bp 5' of the T-DNA insertion in T90. This element, when equipped with a minimal promoter, was able to confer gene expression during both RNS and AM (Fig. **18; 19**), specifically in nodules and infected cortical cells, respectively. The features of *CYC-RE*$_{CBP1}$ provide a plausible explanation for the common and symbioses-specific GUS activity in T90: as a result of the T-DNA insertion in T90, the promoterless *GUS* gene was coincidently brought in proximity at the 3' of *CYC-RE*$_{CBP1}$, a *cis*-element that drives gene expression during colonisation by rhizobia and AM fungi. In the presence of microsymbionts, the GUS gene was consequently activated generating a symbioses-specific expression pattern. These results provided evidence of the involvement of CCaMK/Cyclops in mediating activation of the *CBP1* gene encoding a putative calcium-binding protein for both RNS and AM, through a common cis-element *CYC-RE*$_{CBP1}$. In *L. japonicus*, the expression of the endogenous *CBP1* during nodulation in roots is likely enhanced by the presence of C*CYC-RE*$_{CBP1}$ in its promoter (Fig. **20**). This conclusion is based on the fact that *CBP1* is expressed at a reduced level in T90 roots in which the T-DNA insertion presumably reduces or entirely blocks the activity of this cis-element (Webb et al., 2000); **(2) a 54 bp region 5' of *CYC-RE*$_{CBP1}$ (*EPRE*$_{CBP1}$)** is essential for gene expression in root hairs. A region between -1146 and -1092 bp was necessary for *GUS* expression in patches of root epidermal cells in proximity to or undergoing IT formation (Fig. **15b**). Expression in root epidermal cells could not be achieved when the region was deleted from the T90 promoter or when *CYC-RE*$_{CBP1}$ was tested on its own. Whether this region is required for epidermal expression achieved by T90 promoter during AM development (Fig. **19**) requires a refined analysis of early infection stages. Moreover, *in silico* analysis predicts a number of transcription factors that potentially bind to *EPRE*$_{CBP1}$ (Table **S5**), the involvement of which in *CBP1* regulation remains to be investigated; **(3) the region 3' of the *CYC-RE*$_{CBP1}$** is boosting CCaMK/Cyclops-mediated expression. The region 3' of *CYC-RE*$_{CBP1}$, between -928 and -1 (region 2 in Fig. **22**), had on its own very little or no responsiveness to CCaMK/Cyclops mediated gene activation in *N. benthamiana* leaf cells, but



	region 1	*EPRE*$_{CBP1}$	*CYC-RE*$_{CBP1}$	region 2
required for gene expression in / in response to:				
CCaMK/Cyclops mediated activation	E	-	+,+	-
root hair (RNS)	n.d.	+	-*	n.d.
nodules (RNS)	E	n.d.	+,+	+
root cortical cells (AM)	E$^{\psi}$	n.d.	+,+	n.d.
root apical meristem (RNS)	-	-	-	+
root vasculature (RNS)	-	-	-	+

* insufficient; necessity not determined
ψ enhanced compared to *CYC-RE*$_{CBP1}$

Figure 22. Four regulatory regions of the *CBP1* promoter and their impact on gene expression. *EPRE*$_{CBP1}$ and *CYC-RE*$_{CBP1}$ confer tissue specificity while region 1 and region 2 contribute to the expression strength of the *CBP1* gene. +: necessary; - : not necessary; E: enhancing expression strength; +,+: necessary and sufficient; n.d. : not determined. Note that the quality attribute "necessary" is based on results obtained in different promoter contexts. For *EPRE*$_{CBP1}$ and *CYC-RE*$_{CBP1}$, this annotation is based on the T90 promoter context and region 2 may contain at least partially redundant functions. The entire promoter and 5' UTR are drawn to scale, while *CBP1* (dotted arrow) it is not.

such responsiveness was significantly boosted when this region was combined with its 5' region containing *CYC-RE*$_{CBP1}$ (Fig. **17a**). Interestingly, region 2, despite being devoid of a CCaMK/Cyclops response in in *N. benthamiana* leaf cells, conferred gene expression in early nodule development, root vasculature and root tips in *L. japonicus* hairy roots infected by *M. loti.* **(4) the region 5' of *EPRE*$_{CBP1}$** contains probably multiple regulatory elements. A region between -2870 and -1327 bp (region 1 in Fig. **22**) significantly enhances gene activation mediated by CCaMK/Cyclops in *N. benthamiana* leaf cells. Interestingly, the inclusion of this region in *N. benthamiana* transient assays resulted in much larger variation between different leaf discs and unusually strong inter-experimental variation. We interpreted this variation as a sensitivity of the underlying regulatory machinery to subtle diurnal, developmental or environmental differences of the leaf tissue that are not observed in other promoter fusions. Deletion of this region did not result in loss of reporter expression in nodules, rather seemed to affect the expression strength of the reporter gene (Fig. **17b**). Further investigation is needed for the possible quantitative contribution of the -2780 and -1327 bp region.

In summary, the *CBP1* promoter contains at least 4 distinct regulatory regions that contribute to the expression strength or the tissue specificity or the stimulus specificity of the *CBP1* expression (Fig. **22**).

CYC-RE$_{CBP1}$ equipped with a minimal promoter was sufficient for CCaMK/Cyclops-mediated transcriptional activation (Fig. **17e**). Three *CYC-REs* identified earlier come from the promoters of two RNS-induced genes, *NIN* and *ERN1* (*CRE* or *PACE* and *CYC-RE*$_{ERN1}$,

respectively; Singh et al., 2014; Cerri et al., 2017) and an AM-induced gene, *RAM1* (*AMCYC-RE*; Pimprikar et al., 2016). *ERN1* encodes an AP2/ERF transcription factor that is essential for bacterial infection during RNS (Middleton et al., 2007; Cerri et al., 2016, 2017). *RAM1* encodes the first transcription factor activated *via* symbiotic signalling specific to AM, the AM equivalent of *NIN* in RNS (Gobbato et al., 2012; Pimprikar et al., 2016). The core of *CYC-RE*$_{CBP1}$ shares a high sequence similarity with *AMCYC-RE* and to a lesser extent with *CRE* and *CYC-RE*$_{ERN1}$. Similar to the situation of these three *CYC-RE*s, deletion or mutation of *CYC-RE*$_{CBP1}$ drastically impaired Cyclops-mediated transcriptional activation (Fig. **17e**). These observations together suggest that the induction of T90 *GUS* gene is at least in the nodules, achieved by *CYC-RE*$_{CBP1}$ via CCaMK/Cyclops-mediated activation and that it contributes essentially, but can not on its own, mediate expression in root hairs (Fig. **15**). The binding of Cyclops to *CYC-RE*$_{CBP1}$ could be further verified in, for example, an *in vitro* electrophoretic mobility shift assay or a ChIP-seq assay.

The initial forward genetic approach to screen an EMS-mutagenised T90 population led to the identification of the T90 *white* mutants (Fig. **12**). Based on an analysis using cytosine methylation-sensitive restriction enzymes, we detected hypermethylation specifically in the T90 *white* mutants within *CYC-RE*$_{CBP1}$ (Fig. **16**). This region contains the Cyclops target *cis*-element, *CYC-RE*$_{CBP1}$, which is capable on its own, to drive gene expression during AM and RNS (Fig. **18; 19**). Cytosine methylation is a well-studied phenomenon and frequently associated with heterochromatin-based gene inactivation (Iwasaki & Paszkowski, 2014). The element's ability to achieve gene expression during both symbioses and its hypermethylation in the T90 white mutants are in line with the T90 and T90 white phenotype. However, hypermethylation is typically not restricted to single bases but generally affects longer DNA stretches that undergo heterochromatin formation. Indeed, we have obtained preliminary evidence for additional methylated cytosines within the 113 bp element depicted in Fig. **16** (data not shown). It is therefore likely that the T90 *white* mutants suffered from a broader hypermethylation within the T90 promoter and that the silencing of T90 cannot be attributed solely to the single cytosine methylation within *CYC-RE*$_{CBP1}$. Independent of the extend of the hypermethylation in T90 *white* mutants, our observation revealed that Cyclops activity can be severely impeded by DNA methylation of its target promoters. As DNA methylation is overall dynamically altered during nodulation (Satgé, et al., 2016), studying the methylation status of Cyclops target promoters may reveal another layer of transcriptional regulation during symbiotic development.

For most genes that are activated in both RNS and AM, the *cis*-elements responsible for the common induction are not yet known. Only one other element, an AT-rich motif identified in the promoter of *M. truncatula ENOD11* gene, was reported important for high-level gene expression during both RNS and AM (Boisson-Dernier et al., 2005). *ENOD11* is one of the earliest marker genes induced by rhizobia as well as an AMF (Chabaud et al., 2002; Journet et al., 2001). The discovery of *CYC-RE*$_{CBP1}$ supported a possible scenario that at least a subset of the genes induced during both RNS and AM development could be regulated by the CCaMK/Cyclops complex. This hypothesis is in line with the observations that *CBP1* expression is specifically induced in roots upon Nod factor treatment or inoculation with *M. loti* (Cathrine Kistner, unpublished data) as well as with the AMF *R. irregularis*; and the upregulation of *CBP1*

during nodulation is dependent on functional *Cyclops*, *NFR*1 and *NFR5* genes (data retrieved from LotusBase, Webb et al., 2000; Mun et al., 2016). The observation that four regions of the *CBP1* promoter impact its gene expression suggest that additional pathways are in play to achieve tissue-specific expression pattern and enhance gene expression.

In *L. japonicus*, the expression of *CBP1* during nodulation in roots is likely enhanced by the presence of *CYC-RE*$_{CBP1}$ in its promoter (Fig. **20**). This conclusion is consistent with the fact that *CBP1* is expressed at a reduced level in T90 roots in which the T-DNA insertion presumably reduces or entirely blocks the activity of this cis-element (Webb et al., 2000). Nevertheless, current evidence seems to suggest that the Cyclops-regulated enhancement of *CBP1* expression is dispensable for symbiosis because T90 and T90 *white* mutants are neither impaired in RNS nor in AM. It is likely that the residual expression of *CBP1* mediated by the 928 bp regulatory region 3' of the T-DNA insertion site in T90 may be adequate for endosymbiosis; or that the impact of the reduced expression levels could not be detected by phenotyping the number of nodules and shoot dry weight of plants. Alternatively, although it appears that there is only one copy of *CBP1* gene present in the *L. japonicus* genome, other calcium-binding proteins could function redundantly with CBP1. A study of genome-wide identification of Ca^{2+} binding-protein in *L. japonicus* has reported 47 proteins containing an identifiable CaM-domain, amongst which several are encoded by genes highly expressed in nodules (Liao et al., 2017). Given the conserved function of the CaM domain to interact with Ca^{2+} ions, these proteins might carry the potential to function redundantly with CBP1 during nodulation.

The CBP1 protein is comprised of 230 amino acids (aa). Bioinformatic analysis of CBP1 predicts an N-terminal cleavable signal sequence (1 to 29 aa) and a C-terminal classic calmodulin (CaM) domain consisting of a double pair of EF hand Ca^{2+}-binding motifs (Cal-EF-Afi & SignalP-5.0). Considering its potential to bind to calcium due to its CaM domain and its participation in both types of endosymbiosis, it is tempting to speculate that CBP1 plays a role in calcium signalling. Calcium homeostasis is realised by at least four processes: entry of Ca^{2+} from the apoplast into the cytosol, intracellular compartments involved in Ca^{2+} release and uptake, Ca^{2+} buffers in the cytoplasm and extrusion mechanisms (Schwaller et al., 2002). Up to date, apart from a few players enabling symbiotic Ca^{2+} spiking in the nucleus (1.3.1; Fig. 2), other mediators of Ca^{2+} homeostasis during RNS are unknown. Proteins capable of binding to Ca^{2+} ions such as CaM, calmodulin-like proteins (CML) and calcineurin B-like proteins (CBL) are attractive candidates, due to their concerted ability to interact with Ca^{2+} and regulatory proteins such as Ca^{2+}-dependent protein kinases to exert various biological functions (Kudla et al., 2018). A *CML* gene, *MtCaML1*, is highly expressed in nodules and the encoded protein localises to bacterial symbiosomes in *M. truncatula* and *M. sativa* nodules (Liu et al., 2006); however its specific role in Ca^{2+} homeostasis has not been extensively characterised, largely because the attempt to knock down the six identified *CML* copies simultaneously did not yield reproducible phenotypes. The high sequence similarity of the N-terminus of CBP1 to that of *At*CML4 and *At*CML5 suggests that CBP1 might localise to specific intracellular compartments. Interestingly, *At*CML5 is a single-membrane anchored protein with its CaM domain exposed to the cytosol, enabling it to sense cytosolic Ca^{2+} changes (Ruge et al., 2016). A possible scenario would be that CBP1 is anchored to specific subcellular structures and senses regional Ca^{2+}

concentration changes in the cytosol, as for instance a Ca^{2+} buffer. Examples of Ca^{2+} buffers are members of the calbindin-D_{28K} family (Ca^{2+}-binding proteins containing four EF hands) that principally localise to the cytosol and act as slow or fast Ca^{2+} buffers to maintain calcium homeostasis and prevent pro-apoptosis pathways in mammals (Schwaller et al., 2002; Kook et al., 2014). This hypothesis requires information on the subcellular localisation of CBP1 in terms of tissue type and subcellular space during RNS and AM, respectively, for further clarification. Previous studies have demonstrated that CBP1 protein exhibits a mobility shift in a sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) experiment when protein was extracted from *Escherichia coli* cells in the presence of an excess amount of $CaCl_2$ (Tuck, 2006). The Ca^{2+}-binding ability and kinetics of CBP1 are worth re-evaluating in depth, for instance, using flash photolysis analysis, a technique that has been successfully applied to determine binding kinetics of three types of Ca^{2+}-binding proteins (calbindin, calretinin and calmodulin) and provided insights in their specific roles in calcium signalling (Faas and Mody, 2012).

4.2 Similarities and differences between *PACE* and *CYC-RE$_{CBP1}$*

The *cis*-elements, *PACE* and *CYC-RE$_{CBP1}$*, identified in promoters of two genes of distinct functions, bear notable resemblances. First, both *cis*-elements confer tissue specific gene expression in the cortex of nodules in response to microsymbiont-induced signalling. It is not known yet whether the expression domains of *PACE* and *CYC-RE$_{CBP1}$* overlap, and if so, how much. A focal accumulation of reporter activity driven by *CYC-RE$_{CBP1}$* was observed during the progression of bacterial infection into nodule primordia (Fig. **18b**), similar to that driven by *PACE* (Fig. **6e**). A detailed analysis at the cellular level is necessary to trace the physical relation of bacterial infection to the *CYC-RE$_{CBP1}$* domain at different stages of bacterial infection. For this purpose, *Medicago truncatula* might be a better experimental system than *L. japonicus*. Cells in the indeterminate nodules formed on *M. truncatula* roots are arranged in zones representing successive developmental stages ranging from the anticipation of infection threads to hosting fully differentiated bacterial symbiosomes (Roux et al., 2014; Xiao et al., 2014). A single *M. truncatula* nodule can therefore provide a high-resolution snapshot of *CYC-RE$_{CBP1}$* activity covering the entire process of bacterial infection. This approach also eliminates the laborious microscopic screening to locate *L. japonicus* nodules at specific developmental stages and the technical difficulty to determine the infection status of individual cells in the central tissue of nodules.

Moreover, neither *PACE* nor *CYC-RE$_{CBP1}$* is sufficient for gene expression in the root epidermis. It appears as a recurrent phenomenon that different regulatory regions are required for epidermal and cortical expression, as it has been observed for the promoter of *NIN* (this study and Liu et al. 2019c), *CBP1* (this study) and *EPR3* which encodes a receptor for the perception of rhizobial exopolysaccharides during infection (Kawaharada et al., 2017). This phenomenon is not unexpected considering that a community of TFs in addition to Cyclops are critical for IT formation in root hairs. These TFs include the previously described ERN1 and NIN as well as not-yet-identified TFs inducing the production of necessary protein components to facilitate initiation and elongation of ITs, which involves the rearrangements of actin filaments

(Yokota et al., 2009; Hossain et al., 2012) and microtubules (Su et al., 2020), cell wall remodelling (Xie et al., 2012) and several others including RPG (Arrighi et al., 2008), *Lj*SPK1 & *Lj*ROP6 (Liu et al., 2020a), CBS1 (Sinharoy et al., 2016), VAPYRIN, LIN and EXO70H4 (Murray et al., 2011; Liu et al., 2019b, 2020b). Logically, regulatory elements associated with the transcription of these IT-essential proteins should mediate gene expression in root hairs. In the case of *NIN* promoter, direct binding by the regulatory proteins IPN2, NSP1 or SIP1 in addition to Cyclops has been demonstrated although the function of the responsible *cis*-element has not yet been characterised in similar depth as for *PACE*.

Second, both *PACE* and *CYC-RE*$_{CBP1}$ confer Cyclops-mediated transcriptional activation. Cyclops transcriptionally induce expressions of differential sets of genes in response to infection by AMF or rhizobia (Fig. **2**), e.g., *RAM1* for arbuscule formation in AM (Pimprikar et al., 2016); and *NIN* and *ERN1* for RNS controlling both infection and organogenesis (Singh et al., 2014; Cerri et al., 2017) *via* binding to its response elements. So far, it remains elusive how Cyclops, which is essential for both AM and RNS, discerns between two different microsymbiont signals and induces the respective transcriptional changes. Current findings suggest that the transcriptional specificity by Cyclops likely depends on the context. Cyclops interacts with CCaMK and this complex recruits a GRAS-TF DELLA, the repressor of gibberellin signalling, during AM to regulate *RAM1* expression (Pimprikar et al., 2016). The resulting tripartite protein complex may further associate with more members based on the findings of additional proteins directly interacting with CCaMK (Kang et al., 2011) or DELLA (Yu et al., 2014; Heck et al., 2016). Indeed, recruitment of additional regulatory proteins is a commonly used strategy, for instance, to determine binding specificity for TF members belonging to the same family and sharing very similar DNA binding domains (Slattery et al., 2011; Merabet and Mann, 2016). Studies from other organisms also suggest that community composition of regulatory proteins and the connections amongst them as well as with their associated *cis*-regulatory elements are more determinant than individual players in terms of tissue-specific expression of target genes (Vaquerizas et al., 2009; Sonawane et al., 2017). Moreover, biochemical analyses indicate that calmodulin binding is differentially needed for CCaMK activation during mycorrhization and nodulation, respectively, suggesting that two activities of CCaMK can be induced in response to different microbial signals (Shimoda et al., 2012; Routray et al., 2013). Therefore, the protein-protein interaction and differential activation of CCaMK may provide crucial contextual information for Cyclops activity.

An aspect so far neither investigated nor discussed, is the influence of *CYC-REs* on the activity of Cyclops, considering that the differential activities of the three *CYC-REs* analysed suggest distinct regulatory mechanisms. Assembly of regulatory protein complexes is known to depend on their associated *cis*-regulatory elements; the sequence and corresponding DNA ultrastructure of *cis*-elements influence the composition of associated protein complexes (Kribelbauer et al., 2020). The detectable variation in nucleotide composition, particularly those flanking the Cyclops binding site, among *PACEs* variants form different plant orders indirectly suggests that distinct regulatory mechanisms other than Cyclops are in play. The *CYC-REs* (including *PACE* variants) may be used to uncover additional layers of regulation by Cyclops,

including the dissection of protein complex assembly involving CCaMK, Cyclops and other proteins.

Last but not the least, the Cyclops-*PACE* and Cyclops-*CYC-RE*$_{CBP1}$ connections render *NIN* and *CBP1,* respectively, responsive to an ancient signal transduction pathway induced by signalling molecules exuded by microsymbionts. The exclusive and restricted presence of *PACE* suggests that recruitment of *NIN via PACE* occurred in the last common ancestor of the FaFaCuRo clade. In contrast with *PACE,* the phylogenetic distribution of *CYC-RE*$_{CBP1}$ and phenotypic consequence of its acquisition has not yet been investigated. It is helpful to identify the orthologues of the CBP1 proteins and enable analysis of their promoters to trace the the evolutionary pathway of *CYC-RE*$_{CBP1}$.

4.3 Evolution of NIN and CBP1 proteins

The transcriptional rewiring for the evolution of RNS is certainly a combinatory effort of evolution of *cis*- and *trans*-acting elements (1.3). What is beyond the scope of this study is the contribution of regulatory protein evolution. Given its importance, we consider it necessary to discuss the potential adaptations of protein function during the evolution of RNS. In terms of CBP1, a functional divergence of the protein appears less likely considering the conserved function of EF-hands containing CaM motif to bind to Ca^{2+}. Nevertheless, it is certainly interesting that the CBP1 N-terminal extension shares high sequence homology to the N-terminal signal of *At*CML4 and *At*CML5 from *A. thaliana* (Ruge et al., 2016), that anchors them to endomembrane systems and expose CaM domain to the cytosol. A similar case is described for two EF-hand containing small Ca^{2+} binding proteins, CaBP7 and CaBP8, that are unique amongst human Ca^{2+} binding proteins due to the presence of a C-terminal anchoring signal. CaBP7 and CaBP8 localise to trans-Golgi network with its CaM domain facing the cytosol (McCue et al., 2009, 2011). It is suggested that sorting these Ca^{2+} sensors to specific subcellular compartments might be essential to sustain distinct roles of Ca^{2+}-binding proteins, likely refines the Ca^{2+} signalling on top of a global Ca^{2+} response. It is well probable that the adoption of (or alternatively loss it in some species) the N-terminal extension of CBP1 might be an important evolutionary adaptation, and likely related to refined response towards Ca^{2+} signalling during symbiosis.

The case of NIN differs from CBP1 in that NIN is a transcription factor, the change of its expression domain potentially affecting the expression of its whole regulon. In fact, NIN has probably experienced functional divergence over evolutionary time, based on the observations that NIN has lost the nitrate-mediated post-transcriptional regulation that has been reported for its close relative *NLPs* (Schauser et al., 2005; Konishi and Yanagisawa, 2013; Suzuki et al., 2013) and that *A. thaliana* NIN ortholog can not rescue *M. truncatula nin-1* mutant (Liu and Bisseling, 2020). The pre-FaFaCuRo NIN regulon therefore is very likely to differ from its modern regulon. However given the timing of the duplication event giving rise to *NLPs* and *NIN* roughly coincides (if not earlier) with that of birth of the FaFaCuRo clade, this pre-FaFaCuRo NIN regulon may contain the overlapping regulatory modules between the modern NIN and NLPs regulons (Liu and Bisseling, 2020). The overlapping modules include *NF-Y* and *LBD* members that are directly regulated by *NIN* in RNS to induce nodule organogenesis (Soyano et

al., 2013, 2019; Schiessl et al., 2019). The *NF-YA1* orthologs of *A. thaliana,* a plant that does not engage in RNS or AM, modulate root architecture through regulating lateral root development and root growth, which is consistent with the idea of an origin of NIN as a regulator of lateral root development. The expression pattern of *NF-YA1* orthologs are altered in the *A. thaliana nlp7* mutant (Sorin et al., 2014; Zhao et al., 2018; Alvarez et al., 2020). Furthermore, the *LBD16* ortholog playing a role in *A. thaliana* root development is a direct target of *At*NLP7 (Alvarez et al., 2020). A comparison of binding-sites of *L. japonicus* NIN and NLP4 revealed that they recognise highly similar *cis*-elements and that NIN and NLP4 bind to the same regulatory regions of *CLE-RS2, LBD38* and *Nitrate Reductase1* (*NIR1*) genes (Nishida et al., 2021). These observations strongly supports that NIN and NLP4 share binding sites and downstream target genes. Intriguingly, analysis on NIN-binding sites from 17 genes that are up-regulated by rhizobial inoculation in a NIN-dependent manner and down-regulated by nitrate in a NLP4-dependent manner suggested that NIN-specific binding sites appear less palindromic than those of NLP4. This sequence signature appears to help distinguish the binding protein because replacing part of a NIN-specific binding site in the *NF-YB* promoter with the counterpart from that of NLP4-specific binding site in *LjNIR1* promoter leads to gain of binding of this chimeric element by NLP4 (Nishida et al., 2021). The similarity and specificity seen in the NIN and NLP4 binding sites might be a result of target gene adaptation during the evolution of NIN function.

The *PACE* model suggests that the Cyclops-*PACE* connection recruited the ancestral *NIN* into a novel process and a novel expression domain, probably leading to the formation of cortical ITs in the root cortex, one of the essential building blocks of RNS. It is not yet understood which parts of pre-FaFaCuRo NIN regulon are involved in this process. One potential candidate could be *NIN*-regulated *NF-Y* members that are proposed to have an ancestral function in IT formation (Bu et al., 2020). Moreover, a SHR-SCR module involved in cell division during nodulation depends on *NIN* in *M. truncatula* (Dong et al., 2020) and the orthologous module is functional for root development in *A. thaliana* (outside of the FaFaCuRo clade). Although whether *NIN* directly regulates the SHR-SCR module and if so, when the recruitment of this module into NIN regulon occurred is not yet clear, this finding together with the *LBD* module regulated by modern *NLPs* and *NIN* maybe interpreted to mean that the pre-FaFaCuRo NIN regulon had the capacity to induce cell division. Combined with the necessity of newly divided cells for IT formation, cell division in the *PACE*-domain might have already taken place in the last common ancestor to produce competent cells for bacterial infection (Fig. **21**).

Another possibility, though currently lacking experimental evidence, is that a *NIN*-independent pathway in the common ancestor of the FaFaCuRo clade could activate cell division in the *PACE*-domain. The pre-FaFaCuRo ancestor presumably engaged in symbiosis with AMF, a symbiosis that does not involve the induction of novel organs in the vast majority of modern plant species. Nevertheless, several lines of evidences - the observation of sparse cell division induced by AMF in legume roots (Russo et al., 2019), fossil records of short arbusculated cells (Strullu-Derrien et al., 2018) and specialised mycorrhizal nodules on some gymnosperm and angiosperm plant roots (Duhoux et al., 2001; Schwendemann et al., 2011;

Harper et al., 2015; Nunes et al., 2020) - might be interpreted to mean that cell division in response to AMF infection was once more frequent. The intriguing observation that AM-induced ectopic cell divisions may be dependent on *CCaMK* deserves detailed examination (Russo et al., 2019), considering that *CCaMK* is fundamental for the establishment of both types of endosymbiosis and a key regulator for nodule organogenesis. Unlike the *cyclops* mutant, the *ccamk* mutant is unable to initiate cell division upon rhizobial infection. The deregulated variants of CCaMK induce spontaneous nodules in both wild-type and interestingly, in the *cyclops* mutant (Yano et al., 2008; Hayashi et al., 2010; Shimoda et al., 2012; Takeda et al., 2012). Whether a hypothetical AMF-induced cell division program exists and whether CCaMK plays a role in this process requires further functional dissection of the role of CCaMK in AM-induced ectopic cell divisions and formation of mycorrhizal nodules.

5. CONCLUDING REMARKS

The functional and phylogenomic analysis of the two *cis*-regulatory elements, two Cyclops response elements, in this study highlights their importance in establishment and evolution of the nitrogen-fixing root nodule symbiosis (RNS). In particular, *PACE* may stand at the basis of the evolution of RNS, serving as a pre-requisite for the emergence of functional nodules. The evolution of *cis*-regulatory elements is dynamic and multi-functional, that may alone suffice for phenotypic novelty (Gompel et al., 2005) or function as an initial permissive event before another novel factor causes an evolutionary shift in phenotypes (Sorrells et al., 2015). Fine tuning of gene expression for a particular biological process often involves several *cis*-elements acting *via* diverse mechanisms (Baudouin-Gonzalez et al., 2017). Therefore, it is almost certain that more *cis*-elements involved in the evolution of RNS trait modulation await to be discovered. In addition to *PACE* that directs *NIN* expression in a discrete infection zone, the findings presented in this study indicate the presence of *cis*-elements that confer *NIN* expression in root hairs, cortical dividing cells and cells filled with symbiosomes. Yet undiscovered repressive *cis*-elements might have also emerged to turn a ubiquitously active *NLP* promoter into a *NIN* promoter that is only induced by bacterial infection. Additionally, *cis*-elements encompassing binding sites in the *NIN* promoter for TFs that induce *NIN* expression (e.g., IPN2 & NSP2) have to be analysed to understand the full spectrum transcriptional control of *NIN.* Changes in protein coding sequence, that are not addressed in this study, are another profound contributor of evolutionary novelties, often leading to altered protein-protein interactions and post-translational modifications. In this context, the N-terminal domain adjacent to the conserved EF hands Ca^{2+}-binding motif of CBP1 may represent an important evolutionary adaptation. A similar suggestion has been made for the human calcium binding proteins *Hs*CaBP7 and *Hs*CaBP8 in human (Haynes et al., 2012). Altogether, this work proposes that the reconstruction of RNS in crops should consider evolutionary adaptations of essential transcriptional regulators and crucial *cis*-elements, that ultimately made up the full set of the RNS genetic toolbox.

Recent technical advances based on next-generation sequencing technologies and computational prediction of *cis*-elements enable genome-wide identification of enhancers in a high-throughput format. This type of genome-wide approach has been successfully implemented for humans (Andersson et al., 2014), *Drosophila* (Arnold et al., 2013), zebra fish (Taminato et al., 2016) and more recently for plants (Ricci et al., 2019; Sun et al., 2019; Yan et al., 2019; Jores et al., 2020) independent from prior identification of relevant transcription regulators. In an ideal scenario, a genome-wide search for symbiosis-relevant enhancers could be conducted for different type of endosymbioses, at different developmental stages and in different cell and tissue types. Coupled with Omics data such as that generated from transcriptomic analysis, the genome-wide list of RNS- or AM-relevant enhancers carries the potential to uncover symbiosis-essential genes with high efficiency, bypassing the classical mutant screening process.

6. MATERIAL AND METHODS

Plant, bacterial and fungal material

Lotus japonicus genotypes used in this thesis are: Gifu (wild-type, accession B-129, Handberg & Stougaard, 1992); T90 (Webb et al., 2000); T90 *white 1* (original seeds harvested from L8668); T90 *white 2* (original seeds harvested from L8686); T90 *white 3* (original seeds harvested from L8687); and *nin-15* (*LORE1* line 30003529). The *LORE1* line was genotyped to select plant homozygous for *LORE1* insertions as instructed (Małolepszy et al., 2016) by Rosa Elena Andrade. Seed bag numbers are listed in Supplementary Table 2.

Mesorhizobium loti MAFF 303099 constitutively expressing *Discosoma sp.* red fluorescent protein (*Ds*Red) (*M. loti Ds*Red) or lacZ (*M. loti lacZ*, provided by David Chiasson, Saint Mary's University, Canada) were used to inoculate *L. japonicus* plants. *M. loti* was grown in Tryptone yeast extract (TY) liquid medium (Beringer, 1974) supplied with the appropriate antibiotics: 25 μg/ml gentamicin and 2 μl/ml tetracycline for *M. loti Ds*Red and *M. loti lacZ,* respectively. Bacterial liquid cultures were incubated at 180 rpm at 28 ^{o}C overnight and bacteria were collected by centrifugation at 3400 rcf (Multifuge 3L-R; Thermo Electron Corporation) for 10 min at room temperature (RT). *M. loti* was washed twice with nitrogen-reduced FAB medium containing the following components: 500 μM $MgSO_4 \cdot 7H_2O$, 250 μM KH_2PO_4, 250 μM KCl, 250 μM $CaCl_2 \cdot 2H_2O$, 100 μM KNO_3, 25 μm Fe-EDDHA (catalog no. F0527.0250, Duchefa Biochemie), 50 μM H_3BO_3, 25 μM $MnSO_4 \cdot H_2O$, 10 μM $ZnSO_4 \cdot 7H_2O$, 0.5 μM $Na_2MoO_4 \cdot 2H_2O$, 0.2 μM $CuSO_4 \cdot 5H_2O$ and 0.2 μM $CoCl_2 \cdot 6H_2O$ (pH 5.7) and resuspended in the same medium to reach the desired final optical density (determined at 600 nm, OD_{600}).

The arbuscular mycorrhizal fungus (AMF) *Rhizophagus irregularis* was used to inoculate *L. japonicus* roots in a chive nurse plant system (based on Wegel et al., 1998). To prepare the nurse plants, *Rhizophagus irregularis* spores (DAOM197198; Connectis, Agronutrition) were used to inoculate chive seedlings (ca. 200 spores for 40 plants). *R. irregularis* spores were collected by centrifugation at 805 rcf for 10 min at 4 ^{o}C and resuspended in 10 ml of ¼ strength modified Hoagland's solution (based on the nitrogen-free medium described in Hoagland and Arnon (1938) with the following modifications: 1 mM KNO_3, 100 μM KH_2PO_4 and replacing half of the chelated iron stock solution with 12.5 μm Fe-EDDHA). Chive seeds were briefly sterilised with 1.2% NaClO and 0.1% SDS for 1 - 2 min and thoroughly washed with sterile distilled water. Pots used to grow chive plants were washed and sterilised with 70% ethanol before use. Sterilised chive seeds were placed on the surface of a sterile sand-vermiculite mixture (2:1) in a pot and watered with 35 ml of modified 1/4 Hoagland's solution and spore solution. Chive pots were kept in a growth chamber (24 ^{o}C, 16 h light /8 h dark; light intensity of 180 μmol/m^2·s) and were covered with a plastic lid for the first 3 days. Chive plants were watered with 1/12 modified Hoagland's solution three times a week (20 ml solution for each pot). Six weeks post inoculation, chive roots were stained with black ink to monitor the colonisation by AMF. After verification of AM colonisation, two chive plants were transferred to a new pot containing a sterile sand-vermiculite mixture and 40 ml of 1/12 modified Hoagland's solution and allowed to grow for another 4 - 6 weeks before being used as nurse plants. The shoot systems of chive

nurse plants were cut off for AM inoculation experiments, leaving only colonised roots in the growth substrate.

Agrobacterium tumefaciens strains AGL1 and GV3101 were used for the transient expression assays in *Nicotiana benthamiana* leaves. *Agrobacterium rhizogenes* AR1193 (Stougaard et al., 1987a) was used for hairy root transformation to generate chimeric transgenic root systems. *A. tumefaciens* and *A. rhizogenes* carrying plasmids of interest were grown in 2 ml Lysogeny broth (LB) medium for two days and then transferred to fresh LB medium of an appropriate volume or plated on LB plates (300 µl for each plate), respectively, one day before the respective experiment. LB medium and plates were supplied with the appropriate antibiotics; relevant information can be found in Supplementary Table 3.

Plant growth conditions

L. japonicus seeds were scarified and surface sterilised as previously described (Groth et al. 2010) and plated on ½ Gamborg's B5 (catalog no. G0209.0050; Duchefa Bichemie) with 0.8% BD Bacto™ agar (BD 214010; Becton, Dickinson and Company) plates. Seeds were kept in a Panasonic growth chamber (24 °C; 16 h light/8 h dark; MLR-352H-PE) first in dark for three days and then in light condition for days as stated. Seedlings were either directly transferred to Weck jars (SKU743 or SKU745; J.WECK GmbH u. Co. KG) for phenotypic analysis or subjected to *A. rhizogenes*-mediated hairy root transformation based on the protocol described (Charpentier et al., 2008) with the following modifications: 1) roots of seedlings were cut while seedlings were immersed in *A. rhizogenes* re-suspended in sterile water; 2) after removal of roots, shoots of the seedlings were transferred to plates containing Gamborg's B5 medium (without sucrose) and 0.8% BD Bacto™ agar. Shoots were kept on this sucrose-absent B5 medium for a week to avoid excessive growth of *A. rhizogenes.*

For phenotypic analysis under symbiotic condition (Fig. 3, 12), ten-day-old seedlings were transferred to chive nurse pots (5 plants in each pot) watered with ¼ modified Hoagland's solution (20 ml for each pot; supplied with 9 mM KNO_3) three times a week or Weck jars (SKU745) containing 300 ml of sterile sand-vermiculite mixture (2:1) and 25 ml Fahraeus medium (Fåhraeus, 1957) containing *M. loti Ds*Red with an OD_{600} of 0.03. The same amount of nitrogen-reduced version of FAB medium without *M. loti Ds*Red was used for the non-inoculated control. Plants were kept in a growth chamber (24 °C, 16 h light /8 h dark; light intensity of 180 µmol/m^2s) and roots were harvested at days post inoculation as stated in each figure legend. For phenotypic analysis under nitrogen-sufficient condition (Fig. 13), seven-day-old seedlings were transferred to Weck jars (SKU745) containing 300 ml of a seramis-vermiculite mixture (4:1; seramis from Seramis Zimmerpflanzen, Westland Deutschland GmbH) mixed well with 30 ml of ¼ Hoagland's solution (15mM KNO_3). Plants were analysed at 26 days post transfer. Roots of T90 *white* mutants (*w2* & *w3*) grown under these conditions were harvested for extraction of genomic DNA (used in the analysis illustrated in Fig. 16). To grow plants for genomic DNA extraction (Gifu, T90 and T90 *w1*; Fig.16), seven-day-old seedlings were transferred to Weck jars (SKU745) containing 300 ml of a sand-vermiculite mixture (2:1) mixed well with 60 ml of the nitrogen-sufficient version of FAB medium (other components

remain the same except for KNO_3 adjusted to 1.5 mM). Roots were harvested 25 days post transfer.

Plants with chimeric transgenic root systems generated *via* hairy root transformation were transferred to: (a) Weck jars (SKU745 & SKU743) containing 300 ml of sterile sand-vermiculite mixture mixed well with 60 ml nitrogen-reduced FAB medium containing *M. loti Ds*Red or *M. loti* lacZ with an OD_{600} of 0.01 (Fig. 6 - 7, 14, 15c & 18); (b) Weck jars (SKU745 & SKU743) containing 300 ml of a sterile sand-vermiculite mixture and 30 ml nitrogen-reduced FAB medium containing *M. loti Ds*Red with an OD_{600} of 0.05 (Fig. 8 - 11); (c) Weck jars (SKU745) containing 300 ml of sterile sand-vermiculite mixture and 25 ml Fahraeus medium containing *M. loti Ds*Red with an OD_{600} of 0.03 (Fig. 15b, S1c); (d) chive nursing pots (5 plants in each pot) and watered with ¼ modified Hoagland's solution (20 ml for each pot; supplied with KNO_3 to a final concentration of 9 mM) three times a week (Fig. 19). Plants were grown in the same growth chamber as the chive plants.

Staining method for arbuscular mycorrhizal fungal visualisation

To determine AMF colonisation in chive nurse plants, chive roots were first cleared with 10% KOH, followed by two times rinsing with distilled water. Cleared roots were acidified by rinsing roots with 10% acetic acid and then stained with 5% black ink in 5% acetic acid at 95 °C for 5 minutes. Roots were kept in 5% acetic acid for detaining for at least 20 min at RT before microscopic evaluation. Plants co-cultivated with mycorrhized chive plants were harvested at indicated days post transfer to the chive pots (referred as dpi; Table S2). Roots or transgenic root systems were subjected to GUS staining for 14 to 16 h at 37 °C, followed by cleaning with 50% ethanol overnight at RT and then cleared with 10% KOH. Roots were thoroughly rinsed with distilled water and incubated in 0.1 M HCl for 2 hours at RT before overnight staining with 1 µg/ml WGA Alexa Fluor 488 (catalog no. W11261;Thermo Fisher Scientific) dissolved in PBS buffer (140 mM NaCl; 2 mM KCl; 10 mM Na_2HPO_4; 2 mM KH_2PO_4; pH 7.4). Roots were kept in staining buffer at 4 °C in the dark until microscopic analysis.

Equipment and settings

For Fig. 3b, 6, 7, 12b, 14, 15c, 18, 20 & S1, roots were examined using a Leica MZ16 FA stereomicroscope equipped with a DFC 300FX digital camera (Leica Microsystems) using the GFP and *Ds*Red filters. The white light source was Schott KL 1500-Z. For Fig. 8 - 11, roots were evaluated using a Leica M165FC stereomicroscope equipped with a DFC 450C camera and a Leica DM6B-Z light microscope equipped with a DFC 9000GT camera. Sections of nodule primordia and nodules were imaged with a TCS SP5 confocal laser-scanning microscope (CLSM) equipped with a HCX PL APO CS 20x/0.7 IMM CORR CS objective lens (Leica Microsystems). The argon laser band of 514 nm was used to excite *Ds*Red, with the emission window set between 559 to 633 nm. The images displayed are maximal projections of selected planes of a Z-stack. To detect AMF in Fig. 3, 12b, 19 & S1, roots were examined with a Leica DM6B-Z light microscope equipped with a DFC 9000GT camera. Presence of WGA-stained AMF was detected using the L5 filter cube (512nm - 542nm; size K, catalogue no. 11513880; Leica Microsystems).

For Fig. 15b & 18, roots were imaged with a Keyence VHX-6000 digital microscope with full ring light setting (KEYENCE Deutschland GmbH). For Fig. 3b, images of root and nodule sections were obtained using a Leica DM6B-Z light microscope equipped with a DFC320R2 camera. For Fig. 13a, plants were scanned at 600 dots per inch with an Epson V700 scanner. All microscopic images were processed using the Fiji ImageJ v2.0.0 rc 44/1.50e software and the vector progressing software Affinity designer (Serif (Europe) Ltd).

Cloning and DNA constructs

The *NIN* gene used in this study consists of the *NIN* genomic sequence without the 5' and 3' UTRs. For all the versions of the *L. japonicus* NIN promoter tested, the *LjNIN* 5'UTR was fused to 3' end of the promoter. The *Solanum lycopersicum NLP2* gene (gene ID Solyc01g112190.2.1) was identified as the closest homolog of *LjNIN* based on a previously published phylogenetic analysis (Griesmann et al., 2018) and was referred to as the *SlNIN* gene. A 3 kb region of the *SlNIN* promoter fused to the endogenous 238 bp *SlNIN* 5'UTR was cloned from *S. lycopersicum* cv. "Moneymaker" (seeds obtained from Arne Weiberg and propagated in a greenhouse located the at LMU biocenter). *PACE* or *mPACE* (Fig. 5a) was inserted 184 bp upstream of the *SlNIN* 5'UTR. Promoter fragments used to generate the synthetic promoter with *PACE* or *mPACE* inserted at in the *SlNIN* promoter were obtained by Polymerase Chain Reaction (PCR) or directly synthesised (eurofins Genomics). To clone *cis*-regulatory elements, the two complementary strands of the *cis*-element with the appropriate flacking sequence required for subsequent cloning steps were synthesised as single-stranded DNA fragment (100 μM; Merck KGaA). One microliter of each strand was added to annealing buffer (10 mM Tris-HCl, 0.1 mM Na_2EDTA, 50 mM NaCl) to make up a total volume of 20 μl, which was incubated at 95 ^{o}C for 5 min and allowed to cool for 45 min to 60 min at RT. One microliter of the annealing mix was used in subsequent cloning experiments.

A variant of the *GUS* gene, *DoGUS* (from plasmid C204, DNA Cloning Service), used for cloning was kindly provided by David Chiasson (SMU, Halifax, Canada). Constructs were generated with the Golden Gate cloning system (Binder et al., 2014). A detailed description of the constructs used in this study is provided in Supplementary Table 3 and primers used for cloning are included in Supplementary Table 4.

Phenotypic analysis

For promoter activity analysis with the *GUS* reporter gene, GUS staining was performed as previously described with the incubation time adjusted (Groth et al., 2010). Roots or transgenic root systems were harvested as indicated time point and subjected to GUS staining for 6 h at 37 ^{o}C before imaging with white light illumination. To investigate GUS activity in nodules at different development stages, the analysis was performed 10 to 14 dpi with *M. loti Ds*Red as follows (Fig. 6 - 7): (1) transformed nodule primordia and nodules were identified by the presence of GFP fluoresence emitting nuclei (*$Ubi10_{pro}$:NLS-2xGFP;* GFP transformation marker encoded on the T-DNA) with a GFP filter. Bacterial infection of each nodule primordium and nodule was determined by detecting the *Ds*Red signal indicative of *M. loti Ds*Red with a *Ds*Red filter. Nodule primordia and nodules were imaged with the GFP and *Ds*Red filters; (2) transformed nodule primordia and nodules of interest were excised, placed in 24-well plates

(one nodule in each well containing 500 µl GUS staining buffer; catalog no. 83.3922, Sarstedt Ltd.) and incubated for 3 h at 37 °C. Meanwhile the individual transgenic root systems from which the nodule primordia and nodules of interests had been excised were also subjected to GUS staining; (3) images of nodule primordia and nodules were obtained with the brightfield after staining.

For complementation experiments of *nin-15* (Fig. 8 - 11), the terms 'infected' and 'non-infected' nodules refer to nodules with *Ds*Red signal detected or not detected inside nodules, respectively. Presence or absence of bacteria was later confirmed by examination of sections of representative nodules. Quantifications and sectioning were performed 21dpi with *M. loti* *Ds*Red as follows: (1) transgenic roots were identified by GFP fluorescence emitting nuclei with a GFP filter; (2) infected nodules were identified and quantified with a *Ds*Red filter; (3) the total number of nodules (including infected and non-infected ones) was then determined under WLI; (4) the number of non-infected nodules was calculated by subtracting the number of infected nodules from the total number of nodules. To quantify the infection events in root hairs, the number of bacterial entrapment and ITs in root hairs were scored with a *Ds*Red filter on a 0.5 cm root piece for each transgenic root system, excised from a region where bacterial accumulation was visible. Sectioning was performed on non-infected and infected nodules and the presence/absence of ITs and symbiosomes in cortical cells was examined. Nodule primordia and nodules were embedded in 6% low-melting agarose and sliced into 40 - 50 µm thick sections using a vibrating-blade microtome (Leica VT1000 S).

For Fig. 13, shoot height was evaluated as the distance between shoot apical meristem to the lower end of the hypocotyl. Root length referred to as the length of whole root system. Shoot dry weight was measured after drying the shoots at 60 °C for 1 h in an incubator (Memmert GmbH & Co.KG.).

Genomic DNA extraction and investigation of promoter methylation pattern

Roots (ca. 100 mg) from plants grown in the absence of symbiont of each genotype were harvested, frozen and ground in liquid nitrogen with mortar and pestle. Genomic DNA (gDNA) was extracted as previously described (Lueders et al. 2004). Concentration of gDNA was determined with a Nanodrop photometer (DS-11; DeNovix Inc.). In total, 25 ng gDNA was subjected to restriction digestion by *Hae*III (New England Biolabs GmbH) in a 10 µl reaction that contained 1 µl appreciate NEB Cutsmart restriction buffer (supplied with the enzyme), 1 µl gDNA, 1 µl (ca. 10 units) of enzyme and 7 µl MilliQ water for 18 h at 37 °C. PCR was performed with 1 µl of digestion mix as template and the primer pair 5'-AATAGTGGCATATGAAAATGTTGG-3' (F1) and 5'-AATTATAGGAAGACGTTGGAGAGT-3' (R1; Fig. 16) to amplify a 220 bp region in the T90 promoter containing a single recognition site of each of the enzymes, or the primer pair 5'-TTTCGCCGATATCGTAGAC-3' and 5'-GCAACACCGGCTATATAATAGTG-3' to amplify a 199 bp region of the NIN promoter that does not contain recognition sites for any of the enzymes, as a control for the quality of digested gDNA. The PCR was performed with Taq DNA polymerase (catalog no. M7848 Promega) as follows: initial denaturation at 95 °C for 3 min, 30, 35 or 40 cycles of amplification (95 °C for 20 s, then 53 °C for 30 s, followed by 72 °C for 25 s); and a final extension step at 72 °C for 5 min.

PCR products were detected using agarose gel electrophoresis (2.0 % agarose gel with 9.5V/cm for 40 min for Fig **16b**; 1.0% agarose gel with 8v/cm for 35 min for Fig. **16c**).

Transient expression in *Nicotiana benthamiana* leaves

Nicotiana benthamiana plants were grown as previously described (Cerri et al., 2017), and infiltration of *N. benthamiana* leaves with *A. tumefaciens* (Cerri et al., 2012) with the acetosyringone concentration in the infiltration buffer changed to 150 µM. *A. tumefaciens* carrying promoter:GUS fusion constructs of interest (strain AGL1) were co-infiltrated with *A. tumefaciens* containing plasmid *$35S_{pro}$:3xHA-Cyclops* (strain AGL1; Singh et al., 2014), *$35S_{pro}$:$CCaMK^{1-314}$-NLS-mOrange* (strain GV3101; Takeda et al., 2012) or *$35S_{pro}$:$CCaMK^{T265D}$-3xHA* (strain GV3101; Yano et al. 2008) as indicated in Fig. 5 & 18. An AGL1 strain carrying a K9 plasmid constitutively expressing *Ds*Red was used as needed to equalise the amount of *A. tumefaciens* infiltrated per leaf, together with an *A. tumefaciens* strain carrying a plasmid for the expression of the viral P19 silencing suppressor (Voinnet et al., 2003). *N. benthamiana* leaf discs with a diameter of 0.5 cm were harvested 60 hours post infiltration and used for either quantitative fluorometric GUS assays.

Quantitative fluorometric GUS assay and analysis

The quantitative fluorometric GUS assay was adapted (Jefferson 1987) to be suitable for 96-well plate format. *N. benthamiana* leaf discs were harvested into a 96-deep well plate (one leaf disc for each well; catalog no. 260252; Thermo Fisher Scientific) in liquid nitrogen with two glass beads for each well (2.7 mm; catalog no. N032.1; Carl Roth GmbH & Co. KG) and closed tightly with a silicon lid (catalog no. 276000; Thermo Fisher Scientific). Leaf discs were ground into fine powder with a Retsch mill (MM400; Retsch) for 1 min at a frequency of 30/sec. The plate was then briefly centrifuged to collect and remove leaf powder from the lid at 4 °C, kept on ice while 350 µl of extraction buffer (50 mM sodium-phosphate buffer, pH 7.0; 10 mM Na_2EDTA; pH 8.0; 10 mM beta-Mercaptoethanol; 0.1% Triton X-100; 0.1% N-Laurylsarcosine) was added to each well. The fine leaf powder was allowed to thaw on ice and completely homogenised with the extraction buffer by gentle inversions, followed by a centrifugation at 3220 rcf for 30 min at 4 °C. Supernatant (100 µl for each well) was transferred to a 96-well PCR plate (catalog no. I1402-9700; Starlab GmbH) and kept on ice, referred to as the protein extract. A “GUS reaction plate” was prepared by adding 100 µl of extraction buffer supplied with 1 mM 4-Methylumbelliferyl-ß-D-glucuronide hydrate (4-MUG) for each well of a 96-well PCR plate and preheated to 37 °C for 5 min in a Thermomixer (Eppendorf Thermomixer® C equipped with a 96-well Eppendorf SmartBlock™; Eppendorf Deutschland). Ten microliter of protein extract were added into each well of the GUS reaction plate that was kept at 37 °C. After 20 min of incubation, 10 µl of the reaction mix were pipetted into a black fluorescence measuring plate (Costar 96-well Flat Black, Merck KGaA) containing 100 µl of 0.2 M Na_2CO_3 in each well. The fluorescent signal of 4-Methylumbelliferone (4-MU; released product from 4-MUG substrate by GUS) was measured with a TECAN Infinite® 200 PRO (Tecan Trading AG) under the following conditions: excitation at 360 nm; emission at 465 nm; 15 flashes; manual gain at 50. One microliter of the protein extract per well was added into a transparent protein measurement plate (Costar 96 flat transparent or Greiner 96 flat transparent; Greiner Bio-One

International GmbH) containing 100 μl Bradford solution (1:5 dilution in MilliQ water) and the amount of total protein was measured with TECAN Infinite® 200 PRO (absorbance at 595 nm; 15 flashes). Two standard curves were calculated with different concentrations of BSA protein (3.0, 2.5, 2.0, 1.5, 1.0, 0.5, 0.25 and 0 μg/μl) and 4-MU (500, 250, 100, 10, 1, 0.5, 0.25 and 0.125 pmol/μl), to assist calculating the total amount of protein and 4-MU, respectively. The amount of GUS protein in each leaf disc was indicated by the amount of 4-MU, normalised to reaction time and total amount of protein. A total number of four to eight leaf discs per indicated combinations was analysed in at least two independent experiments performed in different weeks.

Prediction of transcription factor binding sites

The binding sites were predicted by Tomtom (version 5.4.1; accessed on https://meme-suite.org/meme/tools/tomtom; accessed Nov 8th, 2021). The $EPRE_{CBP1}$ was used as input query motifs against the JASPAR (non-redundant) core plant database (2018). The motif column comparison function was Pearson correlation coefficient and a significance threshold of E-value <10 was used. The identified putative motif sequences and corresponding known motif were summarised in Table S5.

Data visualisation and statistical analysis

Statistical analyses and data visualisation were performed with RStudio 1.1. 383 (RStudio Inc.). Boxplots were used to display data in Fig 5, 9 -11, 13 (Wickham & Stryjewski, 2011). Individual data points were added to boxplots using R package “Beeswarm” with the method “center” and spacing of 0.1 or 0.3 (https://github.com/aroneklund/beeswarm). R package “agricolae” was used to perform ANOVA statistical analysis with post hoc Tukey (Mendiburu 2018). Statistical results were presented in small letters where different letters indicate statistical significance, while identical letters indicate no significant statistical difference.

7. REFERENCE

Alfatih, A., Wu, J., Zhang, Z.-S., Xia, J.-Q., Ullah Jan, S., Yu, L.-H., et al. (2020). Rice NIN-LIKE PROTEIN 1 rapidly responds to nitrogen deficiency and improves yield and nitrogen use efficiency. *Journal of Experimental Botany* 71, 6032-6042. doi:10.1093/jxb/eraa292.

Alvarez, J. M., Schinke, A. L., Brooks, M. D., Pasquino, A., Leonelli, L., Varala, K., et al. (2020). Transient genome-wide interactions of the master transcription factor NLP7 initiate a rapid nitrogen-response cascade. *Nature Communications* 11, 1157. doi:10.1038/s41467-020-14979-6.

Akamatsu, A., Nagae, M., Nishimura, Y., Romero Montero, D., Ninomiya, S., Kojima, M., et al. (2020). Endogenous gibberellins affect root nodule symbiosis via transcriptional regulation of *NODULE INCEPTION* in *Lotus japonicus*. *The Plant Journal*, tpj.15128. doi:10.1111/tpj.15128.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., et al. (2014). An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455 - 461. doi:10.1038/nature12787.

Andriankaja, A., Boisson-Dernier, A., Frances, L., Sauviac, L., Jauneau, A., Barker, D. G., et al. (2007). AP2-ERF transcription factors mediate nod factor-dependent *MtENOD11* activation in root hairs via a novel *cis*-regulatory motif. *Plant Cell* 19, 2866-2885. doi:10.1105/tpc.107.052944.

Ané, J. M., Kiss, G. B., Riely, B. K., Penmetsa, R. V., Oldroyd, G. E. D., Ayax, C., et al. (2004). *Medicago truncatula DMI1* required for bacterial and fungal symbioses in legumes. *Science* 303, 1364-1367. doi:10.1126/science.1092986.

Ariel, F., Brault-Hernandez, M., Laffont, C., Huault, E., Brault, M., Plet, J., et al. (2012). Two direct targets of cytokinin signaling regulate symbiotic nodulation in *Medicago truncatula*. *Plant Cell* 24, 3838-3852. doi:10.1105/tpc.112.103267.

Arnold, C. D., Gerlach, D., Stelzer, C., Boryń, Ł. M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* 339, 1074-1077. doi:10.1126/science.1232542.

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. *Proceedings of the National Academy of Sciences of the United States of America* 105, 9817-9822. doi:10.1073/pnas.0710273105.

Banba, M., Gutjahr, C., Miyao, A., Hirochika, H., Paszkowski, U., Kouchi, H., et al. (2008). Divergence of evolutionary ways among common sym genes: CASTOR and CCaMK show functional conservation between two symbiosis systems and constitute the root of a common signaling pathway. *Plant and Cell Physiology* 49, 1659-1671. doi:10.1093/pcp/pcn153.

Banhara, A., Ding, Y., Kühner, R., Zuccaro, A., and Parniske, M. (2015). Colonization of root cells and plant growth promotion by *Piriformospora indica* occurs independently of plant common symbiosis genes. *Frontiers in Plant Science* 6, 667. doi:10.3389/fpls.2015.00667.

Baudouin-Gonzalez, L., Santos, M. A., Tempesta, C., Sucena, E., Roch, F., and Tanaka, K. (2017). Diverse *cis*-regulatory mechanisms contribute to expression evolution of tandem gene duplicates. *Molecular Biology and Evolution* 34, 3132-3147. doi:10.1093/molbev/msx237.

Behm, J. E., Geurts, R., and Kiers, E. T. (2014). *Parasponia*: A novel system for studying mutualism stability. *Trends in Plant Science* 19, 757-763. doi:10.1016/j.tplants.2014.08.007.

Bell, C. D., Soltis, D. E., and Soltis, P. S. (2010). The age and diversification of the angiosperms re-revisited. *American Journal of Botany* 97, 1296-1303. doi:10.3732/ajb.0900346.

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

Beringer, J. E. (1974). R factor transfer in *Rhizobium leguminosarum*. *Journal of General Microbiology* 84, 188-198. doi:10.1099/00221287-84-1-188.

Binder, A., Lambert, J., Morbitzer, R., Popp, C., Ott, T., Lahaye, T., et al. (2014). A modular plasmid assembly kit for multigene expression, gene silencing and silencing rescue in plants. *PLoS One* 9, e88218. doi:10.1371/journal.pone.0088218.

Boisson-Dernier, A., Andriankaja, A., Chabaud, M., Niebel, A., Journet, E. P., Barker, D. G., et al. (2005). MtENOD11 gene activation during rhizobial infection and mycorrhizal arbuscule development requires a common AT-rich-containing regulatory sequence. *Molecular Plant-Microbe Interactions* 18, 1269-1276. doi:10.1094/MPMI-18-1269.

Bouwman, A. F., Beusen, A. H. W., Griffioen, J., van Groenigen, J. W., Hefting, M. M., Oenema, O., et al. (2013). Global trends and uncertainties in terrestrial denitrification and N_2O emissions. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368, 20130112. doi:10.1098/rstb.2013.0112.

Bowler, M. W., Cliff, M. J., Waltho, J. P., Blackburn, G. M. (2010). Why did Nature select phosphate for its dominant roles in biology? *New Journal of Chemistry* 34: 784–794. doi.org/10.1039/B9NJ00718K.

Breakspear, A., Liu, C., Roy, S., Stacey, N., Rogers, C., Trick, M., et al. (2014). The root hair "infectome" of *Medicago truncatula* uncovers changes in cell cycle genes and reveals a requirement for Auxin signaling in rhizobial infection. *Plant Cell* 26, 4680-4701. doi:10.1105/tpc.114.133496.

Britton, C. S., Sorrells, T. R., and Johnson, A. D. (2020). Protein-coding changes preceded *cis*-regulatory gains in a newly evolved transcription circuit. *Science* 367, 96-100. doi:10.1126/science.aax5217.

Bu, F., Rutten, L., Roswanjaya, Y. P., Kulikova, O., Rodriguez-Franco, M., Ott, T., et al. (2020). Mutant analysis in the nonlegume *Parasponia andersonii* identifies NIN and NF-YA1 transcription factors as a core genetic network in nitrogen-fixing nodule symbioses. *New Phytologist* 226, 541-554. doi:10.1111/nph.16386.

Buzas, D. M., Lohar, D., Sato, S., Nakamura, Y., Tabata, S., Vickers, C. E., et al. (2005). Promoter trapping in *Lotus japonicus* reveals novel root and nodule GUS expression domains. *Plant and Cell Physiology* 46, 1202-1212. doi:10.1093/pcp/pci129.

Cao, H., Qi, S., Sun, M., Li, Z., Yang, Y., Crawford, N. M., et al. (2017). Overexpression of the maize *ZmNLP6* and *ZmNLP8* can complement the *Arabidopsis* nitrate regulatory mutant *nlp7* by restoring nitrate signaling and assimilation. *Frontiers in Plant Science* 8, 1703. doi:10.3389/fpls.2017.01703.

Capoen, W., Sun, J., Wysham, D., Otegui, M. S., Venkateshwaran, M., Hirsch, S., et al. (2011). Nuclear membranes control symbiotic calcium signaling of legumes. *Proceedings of the National Academy of Sciences of the United States of America* 108, 14348-14353. doi:10.1073/pnas.1107912108.

Carvalho, T. L. G., Balsemão-Pires, E., Saraiva, R. M., Ferreira, P. C. G., and Hemerly, A. S. (2014). Nitrogen signalling in plant interactions with associative and endophytic diazotrophic bacteria. *Journal of Experimental Botany* 65, 5631-5642. doi:10.1093/jxb/eru319.

Cerri, M. R., Frances, L., Kelner, A., Fournier, J., Middleton, P. H., Auriac, M. C., et al. (2016). The Symbiosis-Related ERN transcription factors act in concert to coordinate rhizobial host root infection. *Plant Physiol* 171, 1037-1054. doi:10.1104/pp.16.00230.

Cerri, M. R., Frances, L., Laloum, T., Auriac, M. C., Niebel, A., Oldroyd, G. E. D., et al. (2012). *Medicago truncatula* ERN transcription factors: regulatory interplay with NSP1/NSP2 GRAS factors and expression dynamics throughout rhizobial infection. *Plant Physiology* 160, 2155-2172. doi:10.1104/pp.112.203190.

Cerri, M. R., Wang, Q., Stolz, P., Folgmann, J., Frances, L., Katzer, K., et al. (2017). The *ERN1* transcription factor gene is a target of the CCaMK/CYCLOPS complex and controls rhizobial infection in *Lotus japonicus*. *New Phytologist* 215, 323-337. doi:10.1111/nph.14547.

Chan Y. F., Marks M. E., Jones F. C., Villarreal G. Jr, Shapiro M. D., Brady S. D., et al. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* 327: 302-305. doi: 10.1126/science.1182213.

Chardin, C., Girin, T., Roudier, F., Meyer, C., and Krapp, A. (2014). The plant RWP-RK transcription factors: Key regulators of nitrogen responses and of gametophyte development. *Journal of Experimental Botany* 65, 5577-5587. doi:10.1093/jxb/eru261.

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. doi:10.1105/tpc.108.063255.

Charpentier, M., Sun, J., Martins, T. V., Radhakrishnan, G. v., Findlay, K., Soumpourou, E., et al. (2016). Nuclear-localized cyclic nucleotide-gated channels mediate symbiotic calcium oscillations. *Science* 352, 1102-1105. doi:10.1126/science.aae0109.

Chen, C., Ané, J. M., and Zhu, H. (2008). OsIPD3, an ortholog of the *Medicago truncatula* DMI3 interacting protein IPD3, is required for mycorrhizal symbiosis in rice. *New Phytologist* 180. doi:10.1111/j.1469-8137.2008.02612.x.

Chen, C., Gao, M., Liu, J., and Zhu, H. (2007). Fungal symbiosis in rice requires an ortholog of a legume common symbiosis gene encoding a Ca^{2+}/calmodulin-dependent protein kinase. *Plant Physiology* 145, 1619-1628. doi:10.1104/pp.107.109876.

Chen, T., Zhu, H., Ke, D., Cai, K., Wang, C., Gou, H., et al. (2012). A MAP kinase kinase interacts with SymRK and regulates nodule organogenesis in *Lotus japonicus*. *Plant Cell* 24, 823-838. doi:10.1105/tpc.112.095984.

Clavijo, F., Diedhiou, I., Vaissayre, V., Brottier, L., Acolatse, J., Moukouanga, D., et al. (2015). The *Casuarina NIN* gene is transcriptionally activated throughout Frankia root infection as well as in response to bacterial diffusible signals. *New Phytologist* 208, 887-903. doi:10.1111/nph.13506.

De Faria, S. M., Hay, G. T., Sprent, J. I. (1988) Entry of rhizobia into roots of *Mimosa scabrella* Bentham occurs between epidermal cells. *Microbiology* 134, 2291–2296. doi:10.1099/00221287-134-8-2291.

de La Peña, T. C., Fedorova, E., Pueyo, J. J., and Mercedes Lucas, M. (2018). The symbiosome: Legume and rhizobia co-evolution toward a nitrogen-fixing organelle? *Frontiers in Plant Science* 8, 2229. doi:10.3389/fpls.2017.02229.

Delaux, P. M., Radhakrishnan, G. v, Jayaraman, D., Cheema, J., Malbreil, M., Volkening, J. D., et al. (2015). Algal ancestor of land plants was preadapted for symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* 112, 13390-13395. doi:10.1073/pnas.1515426112.

Demina, I. v., Persson, T., Santos, P., Plaszczyca, M., and Pawlowski, K. (2013). Comparison of the nodule vs. root transcriptome of the actinorhizal plant *Datisca glomerata*: actinorhizal nodules contain a apecific class of defensins. *PLoS One* 8, e72442. doi:10.1371/journal.pone.0072442.

D' Haeze, W., Gao, M., De Rycke, R., Van Montagu, M., Engler, G., Holsters, M. (1998). Roles for azorhizobial nod factors and surface polysaccharides in intercellular invasion and nodule penetration, respectively. *Molecular Plant-Microbe Interaction* 11, 999–1008. doi:10.1094/MPMI.1998.11.10.999.

Dong, W., Zhu, Y., Chang, H., Wang, C., Yang, J., Shi, J., et al. (2020). An SHR-SCR module specifies legume cortical cell fate to enable nodulation. *Nature,* 589, 586-590. doi:10.1038/s41586-020-3016-z.

Doyle, J. J. (2011). Phylogenetic perspectives on the origins of nodulation. *Molecular Plant-Microbe Interactions* 24, 1289-1295. doi:10.1094/MPMI-05-11-0114.

Doyle, J. J. (2016). Chasing unicorns: Nodulation origins and the paradox of novelty. *American Journal of Botany* 103, 1865-1868. doi:10.3732/ajb.1600260.

Dubilier, N., Bergin, C., and Lott, C. (2008). Symbiotic diversity in marine animals: The art of harnessing chemosynthesis. *Nature Reviews Microbiology* 6, 725-740. doi:10.1038/nrmicro1992.

Duhoux, E., Rinaudo, G., Diem, H. G., Auguy, F., Fernandez, D., Bogusz, D., et al. (2001). Angiosperm *Gymnostoma* trees produce root nodules colonized by arbuscular mycorrhizal fungi related to *Glomus*. *New Phytologist* 149, 115-125. doi:10.1046/j.1469-8137.2001.00005.x.

Endre, G., Kereszt, A., Kevei, Z., Mihacea, S., Kaló, P., and Kiss, G. B. (2002). A receptor kinase gene regulating symbiotic nodule development. *Nature* 417, 962-966. doi:10.1038/nature00841.

Erisman, J. W., Sutton, M. A., Galloway, J., Klimont, Z., and Winiwarter, W. (2008). How a century of ammonia synthesis changed the world. *Nature Geoscience* 1, 636-639. doi:10.1038/ngeo325.

Faas, G. C., and Mody, I. (2012). Measuring the kinetics of calcium binding proteins with flash photolysis. *Biochimica et Biophysica Acta - General Subjects* 1820, 1195-1204. doi:10.1016/j.bbagen.2011.09.012.

Fay, P. (1992). Oxygen relations of nitrogen fixation in cyanobacteria. *Microbiological Reviews* 56, 340-373. doi:10.1128/mmbr.56.2.340-373.1992.

Fåhraeus, G. (1957). The infection of clover root hairs by nodule bacteria studied by a simple glass slide technique. *Journal of general microbiology* 16, 374-381. doi:10.1099/00221287-16-2-374.

Fehlberg, V., Vieweg, M. F., Dohmann, E. M. N., Hohnjec, N., Pühler, A., Perlick, A. M., et al. (2005). The promoter of the leghaemoglobin gene *VfLb29*: Functional analysis and identification of modules necessary for its activation in the infected cells of root nodules and in the arbuscule-containing cells of mycorrhizal roots. *Journal of Experimental Botany* 56, 62-69. doi:10.1093/jxb/eri074.

Fournier, J., Imanishi, L., Chabaud, M., Abdou-Pavy, I., Genre, A., Brichet, L., et al. (2018). Cell remodeling and subtilase gene expression in the actinorhizal plant *Discaria trinervis* highlight host orchestration of intercellular *Frankia* colonization. *New Phytologist* 219, 1018-1030. doi:10.1111/nph.15216.

Fournier, J., Teillet, A., Chabaud, M., Ivanov, S., Genre, A., Limpens, E., et al. (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. doi:10.1104/pp.114.253302.

Fournier, J., Timmers, A. C. J., Sieberer, B. J., Jauneau, A., Chabaud, M., and Barker, D. G. (2008). Mechanism of infection thread elongation in root hairs of *Medicago truncatula* and dynamic interplay with associated rhizobial colonization. *Plant Physiology* 148, 1985-1995. doi:10.1104/pp.108.125674.

Fowler, D., Coyle, M., Skiba, U., Sutton, M. A., Cape, J. N., Reis, S., et al. (2013). The global nitrogen cycle in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368, 20130164. doi:10.1098/rstb.2013.0164.

Gage, D. J. (2004). Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiology and Molecular Biology Reviews* 68, 280-300. doi:10.1128/mmbr.68.2.280-300.2004.

Gamas, P., Brault, M., Jardinaud, M. F., and Frugier, F. (2017). Cytokinins in symbiotic nodulation: when, where, what for? *Trends in Plant Science* 22, 792-802. doi:10.1016/j.tplants.2017.06.012.

Garcia, K., Doidy, J., Zimmermann, S. D., Wipf, D., and Courty, P. E. (2016). Take a Trip through the plant and fungal transportome of Mycorrhiza. *Trends in Plant Science* 21, 937-950. doi:10.1016/j.tplants.2016.07.010.

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. *Frontiers in Plant Science* 9, 1901. doi:10.3389/fpls.2018.01901.

Genre, A., Chabaud, M., Timmers, T., Bonfante, P., and Barker, D. G. (2005). Arbuscular mycorrhizal fungi elicit a novel intracellular apparatus in *Medicago truncatula* root epidermal cells before infection. *Plant Cell* 17, 3489-3499. doi:10.1105/tpc.105.035410.

Geurts, R., Xiao, T. T., and Reinhold-Hurek, B. (2016). What does it take to evolve a nitrogen-fixing endosymbiosis? *Trends in Plant Science* 21, 199-208. doi:10.1016/j.tplants.2016.01.012.

Gherbi, H., Markmann, K., Svistoonoff, S., Estevan, J., Autran, D., Giczey, G., et al. (2008). SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and *Frankia* bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 105, 4928-4932. doi:10.1073/pnas.0710618105.

Gleason, C., Chaudhuri, S., Yang, T., Muñoz, A., Poovaiah, B. W., and Oldroyd, G. E. D. (2006). Nodulation independent of rhizobia induced by a calcium-activated kinase lacking autoinhibition. *Nature* 441, 1149-1152. doi:10.1038/nature04812.

Gobbato, E., Marsh, J. F., Vernié, T., Wang, E., Maillet, F., Kim, J., et al. (2012). A GRAS-type transcription factor with a specific function in mycorrhizal signaling. *Current Biology* 22, 2236-2241. doi:10.1016/j.cub.2012.09.044.

Gompel, N., Prud'Homme, B., Wittkopp, P. J., Kassner, V. A., and Carroll, S. B. (2005). Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in Drosophila. *Nature* 433, 481-487. doi:10.1038/nature03235.

Gonzalez-Rizzo, S., Crespi, M., and Frugier, F. (2006). The *Medicago truncatula* CRE1 cytokinin receptor regulates lateral root development and early symbiotic interaction with Sinorhizobium meliloti. *Plant Cell* 18, 2680-2693. doi:10.1105/tpc.106.043778.

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. doi:10.1111/j.1469-8137.2012.04281.x.

Gough, C., and Cullimore, J. (2011). Lipo-chitooligosaccharide signaling in endosymbiotic plant-microbe interactions. *Molecular Plant-Microbe Interactions* 24, 867-878. doi:10.1094/MPMI-01-11-0019.

Granqvist, E., Sun, J., Op den Camp, R., Pujic, P., Hill, L., Normand, P., et al. (2015). Bacterial-induced calcium oscillations are common to nitrogen-fixing associations of nodulating legumes and nonlegumes. *New Phytologist* 207, 551-558. doi:10.1111/nph.13464.

Griesmann, M., Chang, Y., Liu, X., Song, Y., Haberer, G., Crook, M. B., et al. (2018). Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* 361, eaat1743. doi:10.1126/science.aat1743.

Groth, M., Takeda, N., Perry, J., Uchida, H., Dräxl, S., Brachmann, A., et al. (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. *The Plant cell* 22, 2509-2526. doi:10.1105/tpc.109.069807.

Handberg, K., and Stougaard, J. (1992). *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. *The Plant Journal* 2, 487-496. doi:10.1111/j.1365-313X.1992.00487.x.

Harper, C. J., Taylor, T. N., Krings, M., and Taylor, E. L. (2015). Arbuscular mycorrhizal fungi in a voltzialean conifer from the Triassic of Antarctica. *Review of Palaeobotany and Palynology* 215, 76-84. doi:10.1016/j.revpalbo.2015.01.005.

Hayashi, T., Banba, M., Shimoda, Y., Kouchi, H., Hayashi, M., and Imaizumi-Anraku, H. (2010). A dominant function of CCaMK in intracellular accommodation of bacterial and fungal endosymbionts. *Plant Journal* 63, 141-154. doi:10.1111/j.1365-313X.2010.04228.x.

Haynes, L. P., McCue, H. v., and Burgoyne, R. D. (2012). Evolution and functional diversity of the Calcium Binding Proteins (CaBPs). *Frontiers in Molecular Neuroscience* 5, 9. doi:10.3389/fnmol.2012.00009.

He, J., Benedito, V. A., Wang, M., Murray, J. D., Zhao, P. X., Tang, Y., et al. (2009). The *Medicago truncatula* gene expression atlas web server. *BMC Bioinformatics* 10, 441. doi:10.1186/1471-2105-10-441.

Heck, C., Kuhn, H., Heidt, S., Walter, S., Rieger, N., and Requena, N. (2016). Symbiotic fungi control plant root cortex development through the novel GRAS transcription factor MIG1. *Current Biology* 26, 2770-2778. doi:10.1016/j.cub.2016.07.059.

Herendeen, P. S., Magallon-Puebla, S., Lupia, R., Crane, P. R., and Kobylinska, J. (1999). A preliminary conspectus of the allon flora from the late cretaceous (Late Santonian) of central Georgia, U.S.A. *Annals of the Missouri Botanical Garden* 86, 407-471. doi:10.2307/2666182.

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. doi:10.1105/tpc.108.064501.

Hoagland, D. R., and Arnon, D. I. (1938). *The water-culture method for growing plants without soil*. Berkeley: University of California. Revised by Arnon 1950.

Hossain, M. S., Liao, J., James, E. K., Sato, S., Tabata, S., Jurkiewicz, A., et al. (2012). *Lotus japonicus ARPC1* is required for rhizobial infection. *Plant Physiology* 160, 917-928. doi:10.1104/pp.112.202572.

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

Ichihashi, Y., Hakoyama, T., Iwase, A., Shirasu, K., Sugimoto, K., and Hayashi, M. (2020). Common mechanisms of developmental reprogramming in plants-lessons from regeneration, symbiosis, and parasitism. *Frontiers in Plant Science* 11, 1084. doi:10.3389/fpls.2020.01084.

Imaizumi-Anraku, H., Takeda, N., Charpentier, M., Perry, J., Miwa, H., Umehara, Y., et al. (2005). Plastid proteins crucial for symbiotic fungal and bacterial entry into plant roots. *Nature* 433, 527-531. doi:10.1038/nature03237.

Ivanov, S., Fedorova, E. E., Limpens, E., de Mita, S., Genre, A., Bonfante, P., et al. (2012). Rhizobium-legume symbiosis shares an exocytotic pathway required for arbuscule formation. *Proceedings of the National Academy of Sciences of the United States of America* 109, 8316-8321. doi:10.1073/pnas.1200407109.

Iwasaki, M. and Paszkowski, J.. 2014. Epigenetic memory in plants. *The EMBO Journal* 33. 1987–1998. doi: 10.15252/embj.201488883.

Jefferson, R. A. 1987. Assaying chimeric genes in plants: the GUS gene fusion system. *Plant Molecular Biology Reporter* 5:387-405. doi:10.1007/BF02667740.

Jefferson, R. A., Kavanagh, T. A., Bevan, M. W. 1987. GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *The EMBO journal* 6. 3901–3907. doi: 10.1002/j.1460-2075.1987.tb02730.x

Jeong, S., Rokas, A., and Carroll, S. B. (2006). Regulation of body pigmentation by the Abdominal-B Hox protein and its gain and loss in *Drosophila* evolution. *Cell* 125, 1387-1399. doi:10.1016/j.cell.2006.04.043.

Jing, H., and Strader, L. C. (2019). Interplay of auxin and cytokinin in lateral root development. *International Journal of Molecular Sciences* 20, 486. doi:10.3390/ijms20030486.

Johansson, C., and Bergman, B. (1992). Early events during the establishment of the Gunnera/Nostoc symbiosis. *Planta* 188, 403-413. doi:10.1007/BF00192808.

Jores, T., Tonnies, J., Dorrity, M. W., Cuperus, J. T., Fields, S., and Queitsch, C. (2020). Identification of plant enhancers and their constituent elements by STARR-seq in tobacco leaves. *Plant Cell* 32, 2120-2131. doi:10.1105/tpc.20.00155.

Jørgensen, J.E., Stougaard, J., Marcker, A., and Marcker, K. A. (1988) Root nodule specific gene regulation: analysis of the soybean nodulin N23 gene promoter in heterologous symbiotic systems. *Nucleic Acids Research* 16, 39-50. doi: 10.1093/nar/16.1.39.

Kanamori, N., Madsen, L. H., Radutoiu, S., Frantescu, M., Quistgaard, E. M., Miwa, H., et al. (2006). A nucleoporin is required for induction of Ca^{2+} spiking in legume nodule development and essential for rhizobial and fungal symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* 103, 359-364. doi:10.1073/pnas.0508883103.

Kang, H., Zhu, H., Chu, X., Yang, Z., Yuan, S., Yu, D., et al. (2011). A novel interaction between CCaMK and a protein containing the scythe_N ubiquitin-like domain in *Lotus japonicus*. *Plant Physiology* 155, 1312-1324. doi:10.1104/pp.110.167965.

Kang, H., Chu, X., Wang, C., Xiao, A., Zhu, H., Yuan, S., et al. (2014). A MYB coiled-coil transcription factor interacts with NSP2 and is involved in nodulation in *Lotus japonicus*. *New Phytologist* 201, 837-849. doi:10.1111/nph.12593.

Karas, B., Murray, J., Gorzelak, M., Smith, A., Sato, S., Tabata, S., et al. (2005). Invasion of *Lotus japonicus root hairless 1* by *Mesorhizobium loti* involves the nodulation factor-dependent induction of root hairs. *Plant Physiology* 137, 1331-1344. doi:10.1104/pp.104.057513.

Katoh, K., Rozewicki, J., and Yamada, K. D. (2018). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* 20, 1160-1166. doi:10.1093/bib/bbx108.

Kawaharada, Y., Kelly, S., Nielsen, M. W., Hjuler, C. T., Gysel, K., Muszyński, A., et al. (2015). Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* 523, 308-312. doi:10.1038/nature14611.

Kawaharada, Y., Nielsen, M. W., Kelly, S., James, E. K., Andersen, K. R., Rasmussen, S. R., et al. (2017). Differential regulation of the *Epr3* receptor coordinates membrane-restricted rhizobial colonization of root nodule primordia. *Nature Communications* 8, 14534. doi:10.1038/ncomms14534.

Kim, S., Zeng, W., Bernard, S., Liao, J., Venkateshwaran, M., Ane, J. M., et al. (2019). Ca^{2+}-regulated Ca^{2+} channels with an RCK gating ring control plant symbiotic associations. *Nature Communications* 10, 3703. doi:10.1038/s41467-019-11698-5.

Kim, T. K. and Shiekhattar, R. (2015). Architectural and functional commonalities between enhancers and promoters. *Cell* 162, 948-59. doi:10.1016/j.cell.2015.08.008.

Kistner, C. and Parniske, M. (2002). Evolution of signal transduction in intracellular symbiosis. *Trends in Plant Science* 7, 511-518. doi:10.1016/S1360-1385(02)02356-7.

Kistner, C., Winzer, T., Pitzschke, A., Mulder, L., Sato, S., Kaneko, T., et al. (2005). Seven *Lotus japonicus* genes required for transcriptional reprogramming of the root during fungal and bacterial symbiosis. *The Plant cell* 17, 2217-2229. doi:10.1105/tpc.105.032714.

Konishi, M. and Yanagisawa, S. (2013). *Arabidopsis* NIN-like transcription factors have a central role in nitrate signalling. *Nature Communications* 4, 1617. doi:10.1038/ncomms2621.

Konishi, M. and Yanagisawa, S. (2019). The role of protein-protein interactions mediated by the PB1 domain of NLP transcription factors in nitrate-inducible gene expression. *BMC Plant Biology* 19, 90. doi:10.1186/s12870-019-1692-3.

Kook, S. Y., Jeong, H., Kang, M. J., Park, R., Shin, H. J., Han, S. H., et al. (2014). Crucial role of calbindin-D 28k in the pathogenesis of Alzheimer's disease mouse model. *Cell Death and Differentiation* 21, 1575-1587. doi:10.1038/cdd.2014.67.

Kribelbauer, J. F., Loker, R. E., Feng, S., Rastogi, C., Abe, N., Rube, H. T., et al. (2020). Context-dependent gene regulation by homeodomain transcription factor complexes revealed by shape-readout deficient proteins. *Molecular Cell* 78, 152-167.e11. doi:10.1016/j.molcel.2020.01.027.

Krings, M., Taylor, T. N., Hass, H., Kerp, H., Dotzler, N., and Hermsen, E. J. (2007). Fungal endophytes in a 400-million-yr-old land plant: Infection pathways, spatial distribution, and host responses. *New Phytologist* 174, 648-657. doi:10.1111/j.1469-8137.2007.02008.x.

Kudla, J., Becker, D., Grill, E., Hedrich, R., Hippler, M., Kummer, U., et al. (2018). Advances and current challenges in calcium signaling. *New Phytologist* 218, 414-431. doi:10.1111/nph.14966.

Kumar, A., Cousins, D. R., Liu, C. W., Xu, P., and Murray, J. D. (2020). *Nodule inception* is not required for arbuscular mycorrhizal colonization of *Medicago truncatula*. *Plants* 9, 71. doi:10.3390/plants9010071.

Ladha, J. K., Tirol-Padre, A., Reddy, C. K., Cassman, K. G., Verma, S., Powlson, D. S., et al. (2016). Global nitrogen budgets in cereals: A 50-year assessment for maize, rice, and wheat production systems. *Scientific Reports* 6, 19355. doi:10.1038/srep19355.

Laffont, C., Huault, E., Gautrat, P., Frugier, F., Endre, G., Kalo, P., et al. (2019). Independent Regulation of symbiotic nodulation by the SUNN Negative and CRA2 positive systemic pathways. *Plant Physiology* 180, 559-570. doi:10.1104/pp.18.01588.

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

Lanfranco, L., Fiorilli, V., and Gutjahr, C. (2018). Partner communication and role of nutrients in the arbuscular mycorrhizal symbiosis. *New Phytologist* 220, 1031-1046. doi:10.1111/nph.15230.

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

Letunic, I., and Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research* 47, W256-W259. doi:10.1093/nar/gkz239.

Lévy, J., Bres, C., Geurts, R., Chalhoub, B., Kulikova, O., Duc, G., et al. (2004). A Putative Ca^{2+} and Calmodulin-dependent protein kinase required for bacterial and fungal symbioses. *Science* 303, 1361-1364. doi:10.1126/science.1093038.

Li, X., Zheng, Z., Kong, X., Xu, J., Qiu, L., Sun, J., et al. (2019). Atypical receptor kinase RINRK1 required for rhizobial infection but not nodule development in *Lotus japonicus*. *Plant Physiology* 181, 804-816. doi:10.1104/pp.19.00509.

Liao, J., Deng, J., Qin, Z., Tang, J., Shu, M., Ding, C., et al. (2017). Genome-wide identification and analyses of calmodulins and calmodulin-Like proteins in *Lotus japonicus*. *Frontiers in Plant Science* 8, 482. doi:10.3389/fpls.2017.00482.

Lin, J. shun, Li, X., Luo, Z. L., Mysore, K. S., Wen, J., and Xie, F. (2018). NIN interacts with NLPs to mediate nitrate inhibition of nodulation in *Medicago truncatula*. *Nature Plants* 4, 942-952. doi:10.1038/s41477-018-0261-3.

Liu, A., Contador, C. A., Fan, K., and Lam, H. M. (2018). Interaction and regulation of carbon, nitrogen, and phosphorus metabolisms in root nodules of legumes. *Frontiers in Plant Science* 871, 1860. doi:10.3389/fpls.2018.01860.

Liu, C. W., Breakspear, A., Guan, D., Cerri, M. R., Jackson, K., Jiang, S., et al. (2019a). NIN acts as a network hub controlling a growth module required for rhizobial infection. *Plant Physiology* 179, 1704-1722. doi:10.1104/pp.18.01572.

Liu, C. W., Breakspear, A., Stacey, N., Findlay, K., Nakashima, J., Ramakrishnan, K., et al. (2019b). A protein complex required for polar growth of rhizobial infection threads. *Nature Communications* 10, 2848. doi:10.1038/s41467-019-10029-y.

Liu, C. W., and Murray, J. D. (2016). The role of flavonoids in nodulation host-range specificity: An update. *Plants* 5, 33. doi:10.3390/plants5030033.

Liu, J., and Bisseling, T. (2020). Evolution of *NIN* and *NIN-like* genes in relation to nodule symbiosis. *Genes* 11, 777. doi:10.3390/genes11070777.

Liu, J., Liu, M., Qiu, L., and Xie, F. (2020a). SPIKE1 activates the GTPase ROP6 to guide the polarized growth of infection threads in *Lotus japonicus*. *The Plant Cell*, 32, 3774-3791. doi:10.1105/tpc.20.00109.

Liu, J., Miller, S. S., Graham, M., Bucciarelli, B., Catalano, C. M., Sherrier, D. J., et al. (2006). Recruitment of novel calcium-binding proteins for root nodule symbiosis in *Medicago truncatula*. *Plant Physiol* 141, 167-177. doi:10.1104/pp.106.076711.

Liu, J., Rutten, L., Limpens, E., van der Molen, T., van Velzen, R., Chen, R., et al. (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. doi:10.1105/tpc.18.00478.

Liu, J., Rasing, M., Zeng, T., Klein, J., Kulikova, O., Bisseling, T. (2021) NIN is essential for development of symbiosomes, suppression of defence and premature senescence in *Medicago truncatula* nodules. *New Phytologist,* 230: 290-303 doi: 10.1111/nph.17215.

Liu, M., Jia, N., Li, X., Liu, R., Xie, Q., Murray, J. D., et al. (2020b). CERBERUS is critical for stabilization of VAPYRIN during rhizobial infection in *Lotus japonicus*. *New Phytologist,* 229, 1684-1700. doi:10.1111/nph.16973.

Liu, M., Soyano, T., Yano, K., Hayashi, M. and Kawaguchi, M. (2019d). ERN1 and CYCLOPS coordinately activate NIN signaling to promote infection thread formation in *Lotus japonicus*. *Journal of Plant Research* 132, 641-653. doi:10.1007/s10265-019-01122-w.

Luo, L., Zhang, Y. and Xu, G. (2020). How does nitrogen shape plant architecture? *Journal of Experimental Botany* 71, 4415-4427. doi:10.1093/jxb/eraa187.

Lynch, V. J. and Wagner, G. P. (2008). Resurrecting the role of transcription factor change in developmental evolution. *Evolution* 62, 2131-2154. doi:10.1111/j.1558-5646.2008.00440.x.

Madsen, E. B., Madsen, L. H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., et al. (2003). A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. *Nature* 425, 637-640. doi:10.1038/nature02045.

Madsen, L. H., Tirichine, L., Jurkiewicz, A., Sullivan, J. T., Heckmann, A. B., Bek, A. S., et al. (2010). The molecular network governing nodule organogenesis and infection in the model legume *Lotus japonicus*. *Nature Communications* 1, 10. doi:10.1038/ncomms1009.

Magne, K., Couzigou, J. M., Schiessl, K., Liu, S., George, J., Zhukov, V., et al. (2018). *MtNODULE ROOT1* and *MtNODULE ROOT2* are essential for indeterminate nodule identity. *Plant Physiology* 178, 295-316. doi:10.1104/pp.18.00610.

Małolepszy, A., Mun, T., Sandal, N., Gupta, V., Dubin, M., Urbański, D., et al. (2016). The *LORE1* insertion mutant resource. *Plant Journal* 88, 306-317. doi:10.1111/tpj.13243.

Markmann, K. (2008). Functional adaptation of the plant receptor-kinase gene *SYMRK* paved the way for the evolution of root endosymbioses with bacteria. Doctoral thesis. University of Munich, Munich.

Markmann, K., Giczey, G., and Parniske, M. (2008). Functional adaptation of a plant receptor-kinase paved the way for the evolution of intracellular root symbioses with bacteria. *PLoS Biology* 6, e68. doi:10.1371/journal.pbio.0060068.

Markmann, K., and Parniske, M. (2009). Evolution of root endosymbiosis with bacteria: how novel are nodules? *Trends in Plant Science* 14, 77-86. doi:10.1016/j.tplants.2008.11.009.

Marsh, J. F., Rakocevic, A., Mitra, R. M., Brocard, L., Sun, J., Eschstruth, A., et al. (2007). *Medicago truncatula NIN* is essential for rhizobial-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase. *Plant Physiology* 144, 324-335. doi:10.1104/pp.106.093021.

McCue, H. v., Burgoyne, R. D., and Haynes, L. P. (2009). Membrane targeting of the EF-hand containing calcium-sensing proteins CaBP7 and CaBP8. *Biochemical and Biophysical Research Communications* 380, 825-831. doi:10.1016/j.bbrc.2009.01.177.

McCue, H. v., Burgoyne, R. D., and Haynes, L. P. (2011). Determination of the membrane topology of the small EF-hand Ca^{2+}-sensing proteins CaBP7 and CaBP8. *PLoS ONE* 6, e17853. doi:10.1371/journal.pone.0017853.

Medvedeva, Y. A., Khamis, A. M., Kulakovskiy, I. v., Ba-Alawi, W., Bhuyan, M. S. I., Kawaji, H., et al. (2014). Effects of cytosine methylation on transcription factor binding sites. *BMC Genomics* 15, 199. doi:10.1186/1471-2164-15-119.

Merabet, S., and Mann, R. S. (2016). To be specific or not: the critical relationship between Hox and TALE proteins. *Trends in Genetics* 32, 334-347. doi:10.1016/j.tig.2016.03.004.

Mendiburu, F. 2018. Package 'agricolae'. [WWW document] URL https://tarwi.lamolina.edu.pe/~fmendiburu/ [accessed 9 August 2021].

Middleton, P. H., Jakab, J., Penmetsa, R. V., Starker, C. G., Doll, J., Kaló, P., et al. (2007). An ERF transcription factor in *Medicago truncatula* that is essential for nod factor signal transduction. *Plant Cell* 19, 1221-1234. doi:10.1105/tpc.106.048264.

Mun, T., Bachmann, A., Gupta, V., Stougaard, J., and Andersen, S. U. (2016). Lotus Base: An integrated information portal for the model legume *Lotus japonicus*. *Scientific Reports* 6, 39447. doi:10.1038/srep39447.

Murakami, E., Cheng, J., Gysel, K., Bozsoki, Z., Kawaharada, Y., Toftegaard Hjuler, C., et al. (2018). Epidermal LysM receptor ensures robust symbiotic signalling in *Lotus japonicus*. *eLIFE* 7, e33506. doi:10.7554/eLife.33506.001.

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* 315, 101-104. doi:10.1126/science.1132514.

Murray, J. D., Muni, R. R. D., Torres-Jerez, I., Tang, Y., Allen, S., Andriankaja, M., et al. (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 Journal* 65, 244-252. doi:10.1111/j.1365-313X.2010.04415.x.

Nishida, H., and Suzaki, T. (2018). Two negative regulatory systems of root nodule symbiosis: how are symbiotic benefits and costs balanced? *Plant and Cell Physiology* 59, 1733-1738. doi:10.1093/pcp/pcy102.

Nishida, H., Tanaka, S., Handa, Y., Ito, M., Sakamoto, Y., Matsunaga, S., et al. (2018). A NIN-LIKE PROTEIN mediates nitrate-induced control of root nodule symbiosis in *Lotus japonicus*. *Nature Communications* 9, 499. doi:10.1038/s41467-018-02831-x.

Nishida, H., Nosaki, S., Suzuki, T., Ito, M., Miyakawa, T., Nomoto, M., et al. (2021). Different DNA-binding specificities of NLP and NIN transcription factors underlie nitrate-induced control of root nodulation. *The Plant Cell* 33, 2340–2359. doi:10.1093/plcell/koab103.

Nunes, C. I., Massini, J. L. G., Escapa, I. H., Guido, D. M., and Campbell, K. (2020). Conifer root nodules colonized by arbuscular mycorrhizal fungi in jurassic geothermal settings from Patagonia, Argentina. *International Journal of Plant Sciences* 181, 196-209. doi:10.1086/706857.

Okubo, T., Fukushima, S., and Minamisawa, K. (2012). Evolution of *Bradyrhizobium-Aeschynomene* mutualism: living testimony of the ancient world or highly evolved state? *Plant and Cell Physiology* 53, 2000-2007. doi:10.1093/pcp/pcs150.

Ong, C. T., and Corces, V. G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Reviews Genetics* 12, 283-293. doi:10.1038/nrg2957.

Ott, T., van Dongen, J. T., Günther, C., Krusell, L., Desbrosses, G., Vigeolas, H., et al. (2005). Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. *Current Biology* 15, 531-535. doi:10.1016/j.cub.2005.01.042.

Pankievicz, V. C. S., Irving, T. B., Maia, L. G. S., and Ané, J. M. (2019). Are we there yet? The long walk towards the development of efficient symbiotic associations between nitrogen-fixing bacteria and non-leguminous crops. *BMC Biology* 17, 283-293. doi:10.1186/s12915-019-0710-0.

Parniske, M. (2000). Intracellular accommodation of microbes by plants: a common developmental program for symbiosis and disease? *Current Opinion in Plant Biology* 3, 320–328. doi: 10.1016/s1369-5266(00)00088-1.

Parniske, M. (2008). Arbuscular mycorrhiza: The mother of plant root endosymbioses. *Nature Reviews Microbiology* 6, 763-775. doi:10.1038/nrmicro1987.

Parniske, M. (2018). Uptake of bacteria into living plant cells, the unifying and distinct feature of the nitrogen-fixing root nodule symbiosis. *Current Opinion in Plant Biology* 44, 164-174. doi:10.1016/j.pbi.2018.05.016.

Perrine-Walker, F. M., Lartaud, M., Kouchi, H. Ridge, R. W. (2014). Microtubule array formation during root hair infection thread initiation and elongation in the *Mesorhizobium-Lotus* symbiosis. *Protoplasma* 251, 1099–1111. doi:10.1007/s00709-014-0618-z.

Peter I. S. and Davidson E. H. (2017). Assessing regulatory information in developmental gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America* 114, 5862-5869. doi: 10.1073/pnas.1610616114.

Pawlowski, K., and Bisseling, T. (1996). Rhizobial and actinorhizal symbioses: what are the shared features? *Plant Cell* 8, 1899-1913. doi:10.1105/tpc.8.10.1899.

Pawlowski, K., and Demchenko, K. N. (2012). The diversity of actinorhizal symbiosis. *Protoplasma* 249, 967-979. doi:10.1007/s00709-012-0388-4.

Pimprikar, P., Carbonnel, S., Paries, M., Katzer, K., Klingl, V., Bohmer, M.J., et al. (2016). A CCaMK-CYCLOPS-DELLA complex activates transcription of *RAM1* to regulate arbuscule branching. *Current Biology* 26, 987-998. doi:10.1016/j.cub.2016.01.069.

Pimprikar, P., and Gutjahr, C. (2018). Transcriptional regulation of arbuscular mycorrhiza development. *Plant and Cell Physiology* 59, 673-690. doi:10.1093/pcp/pcy024.

Qin, H., and Huang, R. (2018). Auxin controlled by ethylene steers root development. *International Journal of Molecular Sciences* 19, 3656. doi:10.3390/ijms19113656.

Qiu, L., Lin, J. S., Xu, J., Sato, S., Parniske, M., Wang, T. L., et al. (2015). SCARN a novel class of SCAR protein that is required for root-hair infection during legume nodulation. *PLoS Genetics* 11, e1005623. doi:10.1371/journal.pgen.1005623.

Radhakrishnan, G. v., Keller, J., Rich, M. K., Vernié, T., Mbadinga Mbadinga, D. L., Vigneron, N., et al. (2020). An ancestral signalling pathway is conserved in intracellular symbioses-forming plant lineages. *Nature Plants* 6, 280-289. doi:10.1038/s41477-020-0613-7.

Radutoiu, S., Madsen, L. H., Madsen, E. B., Felle, H. H., Umehara, Y., Gronlund, M., et al. (2003). Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* 425, 585-592. doi:10.1038/nature02039.

Radutoiu, S., Madsen, L. H., Madsen, E. B., Jurkiewicz, A., Fukai, E., Quistgaard, E. M. H., et al. (2007). LysM domains mediate lipochitin-oligosaccharide recognition and *Nfr* genes extend the symbiotic host range. *EMBO Journal* 26, 3923-3935. doi:10.1038/sj.emboj.7601826.

Rebeiz, M., and Williams, T. M. (2017). Using *Drosophila* pigmentation traits to study the mechanisms of *cis*-regulatory evolution. *Current Opinion in Insect Science* 19, 1-7. doi:10.1016/j.cois.2016.10.002.

Reid, D., Nadzieja, M., Heckmann, A., Sandal, N., Stougaard, J., and Novák, O. (2017). Cytokinin biosynthesis promotes cortical cell responses during nodule development. *Plant Physiology* 175, 361-375. doi:10.1104/pp.17.00832.

Remy, W., Taylor, T. N., Hass, H., and Kerp, H. (1994). Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proceedings of the National Academy of Sciences of the United States of America* 91, 11841-11843. doi:10.1073/pnas.91.25.11841.

Ren, B., Wang, X., Duan, J., and Ma, J. (2019). Rhizobial tRNA-derived small RNAs are signal molecules regulating plant nodulation. *Science* 365, 919-922. doi:10.1126/science.aav8907.

Ricci, W. A., Lu, Z., Ji, L., Marand, A. P., Ethridge, C. L., Murphy, N. G., et al. (2019). Widespread long-range *cis*-regulatory elements in the maize genome. *Nature Plants* 5, 1237-1249. doi:10.1038/s41477-019-0547-0.

Ried, M. K., Antolín-Llovera, M., and Parniske, M. (2014). Spontaneous symbiotic reprogramming of plant roots triggered by receptor-like kinases. *eLife* 3, e03891. doi:10.7554/eLife.03891.

Routray, P., Miller, J. B., Du, L., Oldroyd, G., and Poovaiah, B. W. (2013). Phosphorylation of S344 in the calmodulin-binding domain negatively affects CCaMK function during bacterial and fungal symbioses. *Plant Journal* 76, 287-296. doi:10.1111/tpj.12288.

Roux, B., Rodde, N., Jardinaud, M. F., Timmers, T., Sauviac, L., Cottret, L., et al. (2014). An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant Journal* 77, 817-837. doi:10.1111/tpj.12442.

Roy, S., Ernst, J., Kharchenko, P. v., Kheradpour, P., Negre, N., Eaton, M. L., et al. (2010). Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science* 330, 1787-1792. doi:10.1126/science.1198374.

Roy, S., Liu, W., Nandety, R. S., Crook, A., Mysore, K. S., Pislariu, C. I., et al. (2020). Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. *The Plant cell* 32, 15-41. doi:10.1105/tpc.19.00279.

Roy, S., Robson, F., Lilley, J., Liu, C. W., Cheng, X., Wen, J., et al. (2017). MtLAX2, a functional homologue of the *Arabidopsis* auxin influx transporter AUX1, is required for nodule organogenesis. *Plant Physiology* 174, 326-338. doi:10.1104/pp.16.01473.

Ruge, H., Flosdorff, S., Ebersberger, I., Chigri, F., and Vothknecht, U. C. (2016). The calmodulin-like proteins AtCML4 and AtCML5 are single-pass membrane proteins targeted to the endomembrane system by an N-terminal signal anchor sequence. *Journal of Experimental Botany* 67, 3985-3996. doi:10.1093/jxb/erw101.

Russo, G., Carotenuto, G., Fiorilli, V., Volpe, V., Chiapello, M., van Damme, D., et al. (2019). Ectopic activation of cortical cell division during the accommodation of arbuscular mycorrhizal fungi. *New Phytologist* 221, 1036-1048. doi:10.1111/nph.15398.

Saito, K., Yoshikawa, M., Yano, K., Miwa, H., Uchida, H., Asamizu, E., et al. (2007). NUCLEOPORIN85 is required for calcium spiking, fungal and bacterial symbioses, and seed production in *Lotus japonicus*. *Plant Cell* 19, 610-624. doi:10.1105/tpc.106.046938.

San Martín, W. (2020). “Global nitrogen in sustainable development: four challenges at the interface of science and policy,” doi:10.1007/978-3-319-71065-5_114-2.

Santi, C., Bogusz, D., Franche, C., Perpignan, D., Domitia, V., Alduy, A. P., et al. (2013). Biological nitrogen fixation in non-legume plants. *Annuals of Botany* 111, 743-767. doi:10.1093/aob/mct048.

Satgé, C., Moreau, S., Sallet, E., Lefort, G., Auriac, MC., Remblière, C., et al. (2016). Reprogramming of DNA methylation is critical for nodule development in *Medicago truncatula*. Nature Plants 2: 16166. doi.org/10.1038/nplants.2016.166

Schauser, L., Roussis, A., Stiller, J., and Stougaard, J. (1999). A plant regulator controlling development of symbiotic root nodules. *Nature* 402, 191-195. doi:10.1038/46058.

Schauser, L., Wieloch, W., and Stougaard, J. (2005). Evolution of NIN-like proteins in *Arabidopsis*, rice, and *Lotus japonicus*. *Journal of Molecular Evolution* 60, 229-237. doi:10.1007/s00239-004-0144-2.

Schiessl, K., Lilley, J. L. S., Lee, T., Tamvakis, I., Kohlen, W., Bailey, P. C., et al. (2019). NODULE INCEPTION recruits the lateral root developmental program for symbiotic nodule organogenesis in *Medicago truncatula*. *Current Biology* 29, 3657-3668.e5. doi:10.1016/j.cub.2019.09.005.

Schwaller, B., Meyer, M., and Schiffmann, S. (2002). “New” functions for “old” proteins: The role of the calcium-binding proteins calbindin D-28k, calretinin and parvalbumin, in cerebellar physiology. Studies with knockout mice. *Cerebellum* 1, 241-258. doi:10.1080/147342202320883551.

Schwendemann, A. B., Decombeix, A. L., Taylor, T. N., Taylor, E. L., and Krings, M. (2011). Morphological and functional stasis in mycorrhizal root nodules as exhibited by a Triassic conifer. *Proceedings of the National Academy of Sciences of the United States of America* 108, 13630-13634. doi:10.1073/pnas.1110677108.

Shen, D., Xiao, T. T., van Velzen, R., Kulikova, O., Gong, X., Geurts, R., et al. (2020). A homeotic mutation changes legume nodule ontogeny into actinorhizal-type ontogeny. *Plant Cell* 32, 1868-1885. doi:10.1105/tpc.19.00739.

Shimoda, Y., Han, L., Yamazaki, T., Suzuki, R., Hayashi, M., and Imaizumi-Anraku, H. (2012). Rhizobial and fungal symbioses show different requirements for calmodulin binding to calcium calmodulin-dependent protein kinase in *Lotus japonicus*. *Plant Cell* 24, 304-321. doi:10.1105/tpc.111.092197.

Sieberer, B.J., Chabaud, M., Timmers, A.C., Monin, A., Fournier, J., Barker, D.G. (2009). A nuclear-targeted cameleon demonstrates intranuclear Ca^{2+} spiking in *Medicago truncatula* root hairs in response to rhizobial nodulation factors. Plant Physiology 151: 1197-1206. doi: 10.1104/pp.109.142851.

Sieberer, B., and Emons, A. M. C. (2000). Cytoarchitecture and pattern of cytoplasmic streaming in root hairs of *Medicago truncatula* during development and deformation by nodulation factors. *Protoplasma* 214, 118-127. doi:10.1007/BF02524268.

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 and Microbe* 15, 139-152. doi:10.1016/j.chom.2014.01.011.

Sinharoy, S., Liu, C., Breakspear, A., Guan, D., Shailes, S., Nakashima, J., et al. (2016). A *Medicago truncatula* cystathionine-β-synthase-like domain-containing protein is required for rhizobial infection and symbiotic nitrogen fixation. *Plant Physiology* 170, 2204-2217. doi:10.1104/pp.15.01853.

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., et al. (2011). Cofactor binding evokes latent differences in DNA binding specificity between hox proteins. *Cell* 147, 1270-1282. doi:10.1016/j.cell.2011.10.053.

Smertenko, A., Assaad, F., Baluška, F., Bezanilla, M., Buschmann, H., Drakakaki, G., et al. (2017). Plant cytokinesis: terminology for structures and processes. *Trends in Cell Biology* 27, 885-894. doi:10.1016/j.tcb.2017.08.008.

Sogawa, A., Yamazaki, A., Yamasaki, H., Komi, M., Manabe, T., Tajima, S., et al. (2019). SNARE proteins LjVAMP72a and LjVAMP72b are required for root symbiosis and root hair formation in *Lotus japonicus*. *Frontiers in Plant Science* 9, 1992. doi:10.3389/fpls.2018.01992.

Soltis, D. E., Soltis, P. S., Morgant, D. R., Swensent, S. M., Mullin, B. C., Dowdi, J. M., et al. (1995). Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proceedings of the National Academy of Sciences of the United States of America* 92, 2647-2651. doi:10.1073/pnas.92.7.2647.

Sonawane, A. R., Platig, J., Fagny, M., Chen, C. Y., Paulson, J. N., Lopes-Ramos, C. M., et al. (2017). Understanding tissue-specific gene regulation. *Cell Reports* 21, 1077-1088. doi:10.1016/j.celrep.2017.10.001.

Sorin, C., Declerck, M., Christ, A., Blein, T., Ma, L., Lelandais-Brière, C., et al. (2014). A miR169 isoform regulates specific NF-YA targets and root architecture in *Arabidopsis*. *New Phytologist* 202, 1197-1211. doi:10.1111/nph.12735.

Sorrells, T. R., Booth, L. N., Tuch, B. B., and Johnson, A. D. (2015). Intersecting transcription networks constrain gene regulatory evolution. *Nature* 523, 361-365. doi:10.1038/nature14613.

Sorrells, T. R., Johnson, A. N., Howard, C. J., Britton, C. S., Fowler, K. R., Feigerle, J. T., et al. (2018). Intrinsic cooperativity potentiates parallel *cis*-regulatory evolution. *eLife* 7, e37563. doi:10.7554/eLife.37563.

Soyano, T., and Hayashi, M. (2014). Transcriptional networks leading to symbiotic nodule organogenesis. *Current Opinion in Plant Biology* 20, 146-154. doi:10.1016/j.pbi.2014.07.010.

Soyano, T., Hirakawa, H., Sato, S., Hayashi, M., and Kawaguchi, M. (2014). NODULE INCEPTION creates a long-distance negative feedback loop involved in homeostatic regulation of nodule organ production. *Proceedings of the National Academy of Sciences of the United States of America* 111, 14607-14612. doi:10.1073/pnas.1412716111.

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 Genetics* 9, e1003352. doi:10.1371/journal.pgen.1003352.

Soyano, T., Shimoda, Y., and Hayashi, M. (2015). NODULE INCEPTION antagonistically regulates gene expression with nitrate in *Lotus japonicus*. *Plant and Cell Physiology* 56, 368-376. doi:10.1093/pcp/pcu168.

Soyano, T., Shimoda, Y., Kawaguchi, M., and Hayashi, M. (2019). A shared gene drives lateral root development and root nodule symbiosis pathways in *Lotus*. *Science* 366, 1021-1023. doi:10.1126/science.aax2153.

Spitz, F., and Furlong, E. E. M. (2012). Transcription factors: From enhancer binding to developmental control. *Nature Reviews Genetics* 13, 613-626. doi:10.1038/nrg3207.

Sprent, J. I. (2007). Evolving ideas of legume evolution and diversity: A taxonomic perspective on the occurrence of nodulation: Tansley review. *New Phytologist* 174, 11-25. doi:10.1111/j.1469-8137.2007.02015.x.

Stougaard, J., Abildsten, D., and Marcker, K. A. (1987a). The *Agrobacterium rhizogenes* pRi TL-DNA segment as a gene vector system for transformation of plants. *Molecular & General Genetics* 20, 251-255. doi:10.1007/BF00331586.

Stougaard, J., Sandal, N. N., Grøn, A., Kühle, A., and Marcker, K. A. (1987b). 5′ Analysis of the soybean leghaemoglobin *lbc 3* gene: regulatory elements required for promoter activity and organ specificity. *The EMBO Journal* 6, 3565-3569. doi:10.1002/j.1460-2075.1987.tb02686.x.

Stracke, S., Kistner, C., Yoshida, S., Mulder, L., Sato, S., Kaneko, T., et al. (2002). A plant receptor-like kinase required for both bacterial and fungal symbiosis. *Nature* 417, 959-962. doi:10.1038/nature00841.

Strullu-Derrien, C., Selosse, M. A., Kenrick, P., and Martin, F. M. (2018). The origin and evolution of mycorrhizal symbioses: from palaeomycology to phylogenomics. *New Phytologist* 220, 1012-1030. doi:10.1111/nph.15076.

Su, C., Klein, M. L., Hernández-Reyes, C., Batzenschlager, M., Ditengou, F. A., Lace, B., et al. (2020). The *Medicago truncatula* DREPP protein triggers microtubule fragmentation in membrane nanodomains during symbiotic infections. *Plant Cell* 32, 1689-1702. doi:10.1105/tpc.19.00777.

Sun, J., He, N., Niu, L., Huang, Y., Shen, W., Zhang, Y., et al. (2019). Global quantitative mapping of enhancers in rice by STARR-seq. *Genomics, Proteomics and Bioinformatics* 17, 140-153. doi:10.1016/j.gpb.2018.11.003.

Suzuki, W., Konishi, M., and Yanagisawa, S. (2013). The evolutionary events necessary for the emergence of symbiotic nitrogen fixation in legumes may involve a loss of nitrate responsiveness of the NIN transcription factor. *Plant Signaling and Behavior* 8, e25975. doi:10.4161/jrn.25975.

Svistoonoff, S., Benabdoun, F. M., Nambiar-Veetil, M., Imanishi, L., Vaissayre, V., Cesari, S., et al. (2013). The independent acquisition of plant root nitrogen-fixing symbiosis in Fabids recruited the same genetic pathway for nodule organogenesis. *PLoS ONE* 8, e64515. doi:10.1371/journal.pone.0064515.

Takeda, N., Maekawa, T., and Hayashi, M. (2012). Nuclear-localized and deregulated calcium- and calmodulin-dependent protein kinase activates rhizobial and mycorrhizal responses in *Lotus japonicus*. *Plant Cell* 24, 810-822. doi:10.1105/tpc.111.091827.

Tamagno, S., Sadras, V. O., Haegele, J. W., Armstrong, P. R., and Ciampitti, I. A. (2018). Interplay between nitrogen fertilizer and biological nitrogen fixation in soybean: implications on seed yield and biomass allocation. *Scientific Reports* 8, 17502. doi:10.1038/s41598-018-35672-1.

Taminato, T., Yokota, D., Araki, S., Ovara, H., Yamasu, K., and Kawamura, A. (2016). Enhancer activity-based identification of functional enhancers using zebrafish embryos. *Genomics* 108, 102-107. doi:10.1016/j.ygeno.2016.05.005.

Tan, S., Sanchez, M., Laffont, C., Boivin, S., le Signor, C., Thompson, R., et al. (2020). A cytokinin signaling Type-B response regulator transcription factor acting in early nodulation. *Plant physiology* 183, 1319-1330. doi:10.1104/pp.19.01383.

Tiessen, H. (1995). Phosphorus in the global environment: transfers, cycles and management. *Phosphorus in the global environment: transfers, cycles and management*. doi:10.1016/s0169-555x(96)00035-9.

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

Tirichine, L., Imaizumi-Anraku, H., Yoshida, S., Murakami, Y., Madsen, L. H., Miwa, H., et al. (2006). Deregulation of a Ca^{2+}/calmodulin-dependent kinase leads to spontaneous nodule development. *Nature* 441, 1153-1156. doi:10.1038/nature04862.

Topping, J. F., Wei, W., Lindsey, K. 1991. Functional tagging of regulatory elements in the plant genome. *Development* 112: 1009–1019. doi.org/10.1242/dev.112.4.1009

Tsikou, D., Yan, Z., Holt, D. B., Abel, N. B., Reid, D. E., Madsen, L. H., et al. (2018). Systemic control of legume susceptibility to rhizobial infection by a mobile microRNA. *Science* 362, 233-236. doi:10.1126/science.aat6907.

Tuck, E. (2006). Molecular and genetic analysis of a symbiosis-specific locus of the *Lotus japonicus* genome and its potential as a tool for dissecting the symbiotic signal transduction pathway. Doctoral thesis. accessed from https://pure.aber.ac.uk/portal/files/49036338/Jensen_Elaine.pdf.

Udvardi, M. K., and Day, D. A. (1997). Metabolite transport across symbiotic membranes of legume nodules. *Annual Review of Plant Biology* 48, 493-523. doi:10.1146/annurev.arplant.48.1.493.

Valdés-López, O., Formey, D., Isidra-Arellano, M. C., Reyero-Saavedra, M. del R., Fernandez-Göbel, T. F., and Sánchez-Correa, M. del S. (2019). Argonaute proteins: why are they so important for the legume-rhizobia symbiosis? *Frontiers in Plant Science* 10, 1177. doi:10.3389/fpls.2019.01177.

van Spronsen, P. C., Grønlund, M., Bras, C. P., Spaink, H. P., and Kijne, J. W. (2001). Cell biological changes of outer cortical root cells in early determinate nodulation. *Molecular Plant-Microbe Interactions* 14, 839-847. doi:10.1094/MPMI.2001.14.7.839.

van Velzen, R., Doyle, J. J., and Geurts, R. (2019). A resurrected scenario: single gain and massive loss of nitrogen-fixing nodulation. *Trends in Plant Science* 24, 49-57. doi:10.1016/j.tplants.2018.10.005.

van Velzen, R., Holmer, R., Bu, F., Rutten, L., van Zeijl, A., Liu, W., et al. (2018). Comparative genomics of the nonlegume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proceedings of the National Academy of Sciences of the United States of America* 115, E4700-E4709. doi:10.1073/pnas.1721395115.

Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., and Luscombe, N. M. (2009). A census of human transcription factors: Function, expression and evolution. *Nature Reviews Genetics* 10, 252-263. doi:10.1038/nrg2538.

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

Vernié, T., Kim, J., Frances, L., Ding, Y., Sun, J., Guan, D., et al. (2015). The NIN transcription factor coordinates diverse nodulation programs in different tissues of the *Medicago truncatula* root. *The Plant Cell* 27, 3410-3424. doi:10.1105/tpc.15.00461.

Voinnet, O., Rivas, S., Mestre, P., Baulcombe, D. 2003. An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant Journal* 33: 949-56. doi: 10.1046/j.1365-313x.2003.01676.x.

Wagner, G. P., and Lynch, V. J. (2010). Evolutionary novelties. *Current Biology* 20. doi:10.1016/j.cub.2009.11.010.

Wang, H., Moore, M. J., Soltis, P. S., Bell, C. D., Brockington, S. F., Alexandre, R., et al. (2009). Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proceedings of the National Academy of Sciences of the United States of America* 106, 3853-3858. doi:10.1073/pnas.0813376106.

Webb, K. J., Skøt, L., Nicholson, M. N., Jørgensen, B., and Mizen, S. (2000). *Mesorhizobium loti i*ncreases root-specific expression of a calcium-binding protein homologue identified by promoter tagging in *Lotus japonicus*. *Molecular Plant-Microbe Interactions* 13, 606-616. doi:10.1094/MPMI.2000.13.6.606.

Wegel, E., Schauser, L., Sandal, N., Stougaard, J., Parniske, M. (1998). Mycorrhiza mutants of *Lotus japonicus* define genetically independent steps during symbiotic infection. *Molecular Plant-Microbe Interactions* 11: 933–936.doi.org/10.1094/MPMI.1998.11.9.933.

Wernegreen, J. J. (2012). Endosymbiosis. *Current Biology* 22, R555-R561. doi:10.1016/j.cub.2012.06.010.

Werner, G. D. A., Cornwell, W. K., Sprent, J. I., Kattge, J., and Kiers, E. T. (2014). A single evolutionary innovation drives the deep evolution of symbiotic N_2-fixation in angiosperms. *Nature Communications* 5, 4087. doi:10.1038/ncomms5087.

Wickham, H., & Stryjewski, L. 2011. 40 years of boxplots. [WWW document] URL https://vita.had.co.nz/papers/boxplots.pdf [accessed 9 August 2021]

Wong, J. E. M. M., Nadzieja, M., Madsen, L. H., Bücherl, C. A., Dam, S., Sandal, N. N., et al. (2019). A *Lotus japonicus* cytoplasmic kinase connects Nod factor perception by the NFR5 LysM receptor to nodulation. *Proceedings of the National Academy of Sciences of the United States of America* 116, 14339-14348. doi:10.1073/pnas.1815425116.

Wu, J., Zhang, Z., Xia, J., Alfatih, A., Song, Y., Huang, Y., et al. (2020). Rice NIN-LIKE PROTEIN 4 is a master regulator of nitrogen use efficiency. *Plant Biotechnology Journal*. doi:10.1111/pbi.13475.

Valverde, C., and Wall, L. G. (1999). Time course of nodule development in the *Discaria trinervis* (Rhamnaceae) - *Frankia* symbiosis. *New Phytologist* 141, 345-354. doi:10.1046/j.1469-8137.1999.00345.x.

Xiao, A., Yu, H., Fan, Y., Kang, H., Ren, Y., Huang, X., et al. (2020). Transcriptional regulation of *NIN* expression by IPN2 is required for root nodule symbiosis in *Lotus japonicus*. *New Phytologist* 227, 513-528. doi:10.1111/nph.16553.

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

Xie, F., Murray, J. D., Kim, J., Heckmann, A. B., Edwards, A., Oldroyd, G. E. D., et al. (2012). Legume pectate lyase required for root infection by rhizobia. *Proceedings of the National Academy of Sciences of the United States of America* 109, 633-638. doi:10.1073/pnas.1113992109.

Yan, W., Chen, D., Schumacher, J., Durantini, D., Engelhorn, J., Chen, M., et al. (2019). Dynamic control of enhancer activity drives stage-specific gene expression during flower morphogenesis. *Nature Communications* 10, 1-16. doi:10.1038/s41467-019-09513-2.

Yan, Z., Cao, J., Fan, Q., Chao, H., Guan, X., Zhang, Z., et al. (2020). Dephosphorylation of LjMPK6 by phosphatase LjPP2C is involved in regulating nodule organogenesis in *Lotus japonicus*. *International Journal of Molecular Sciences* 21, 1-12. doi:10.3390/ijms21155565.

Yang, wei cai, de Blank, C., Meskiene, I., Hirt, H., Bakker, J., van Kammen, A., et al. (1994). Rhizobium Nod factors reactivate the cell cycle during infection and nodule primordium formation, but the cycle is only completed in primordium formation. *Plant Cell* 6, 1415-1426. doi:10.1105/tpc.6.10.1415.

Yang, L., Chen, Z., Stout, E. S., Delerue, F., Ittner, L. M., Wilkins, M. R., et al. (2020). Methylation of a CGATA element inhibits binding and regulation by GATA-1. *Nature Communications* 11, 2560. doi:10.1038/s41467-020-16388-1.

Yano, K., Yoshida, S., Müller, J., Singh, S., Banba, M., Vickers, K., et al. (2008). CYCLOPS, a mediator of symbiotic intracellular accommodation. *Proceedings of the National Academy of Sciences of the United States of America* 105, 20540-20545. doi:10.1073/pnas.0806858105.

Yokota, K., Fukai, E., Madsen, L. H., Jurkiewicz, A., Rueda, P., Radutoiu, S., et al. (2009). Rearrangement of actin cytoskeleton mediates invasion of *Lotus japonicus* roots by Mesorhizobium loti. *Plant Cell* 21, 267-284. doi:10.1105/tpc.108.063693.

Yoon, H.J., Hossain, M.S., Held, M., Hou, H., Kehl, M., Tromas, A., et al. (2014). *Lotus japonicus SUNERGOS1* encodes a predicted subunit A of a DNA topoisomerase VI that is required for nodule differentiation and accommodation of rhizobial infection. *Plant Journal* 78, 811-21. doi:10.1111/tpj.12520.

Yu, N., Luo, D., Zhang, X., Liu, J., Wang, W., Jin, Y., et al. (2014). A DELLA protein complex controls the arbuscular mycorrhizal symbiosis in plants. *Cell research* 24, 130-133. doi:10.1038/cr.2013.167.

Zhang, Z., Ke, D., Hu, M., Zhang, C., Deng, L., Li, Y., et al. (2019). Quantitative phosphoproteomic analyses provide evidence for extensive phosphorylation of regulatory proteins in the rhizobia-legume symbiosis. *Plant Molecular Biology* 100, 265-283. doi:10.1007/s11103-019-00857-3.

Zhao, L., Zhang, W., Yang, Y., Li, Z., Li, N., Qi, S., et al. (2018). The *Arabidopsis NLP7* gene regulates nitrate signaling via NRT1.1-dependent pathway in the presence of ammonium. *Scientific Reports* 8, 1487. doi:10.1038/s41598-018-20038-4.

Zhu, H., Chen, T., Zhu, M., Fang, Q., Kang, H., Hong, Z., et al. (2008). A novel ARID DNA-binding protein interacts with SymRK and is expressed during early nodule development in *Lotus japonicus*. *Plant Physiology* 148, 337-347. doi:10.1104/pp.108.119164.

8. ACKNOWLEDGEMENT

My most special thanks and most gratitudes go to my supervisor Prof. Dr. Martin Parniske for supporting me throughout my doctoral study. I enjoyed the inspiring discussions about science and life with him and (most of the time) his jokes. I am especially grateful for him always believing in me and motivating me to accomplish some seemingly impossible tasks. I would like to thank my “second supervisor” Prof. Dr. Katharina Pawlowski. She supported me during my master study as a supervisor and has since has always been there whenever I needed advice or help. I want to thank her for giving advice for experiments and valuable feedback of my thesis.

I am also thankful to all members of AG Parniske for being a supportive group and the great working atmosphere that they created. Everyone was kind and showed zero hesitation to lend a hand when I sought for help. I want to thank the PACE making team, Rosa, Chloe, Max and Ksenia, for their insights and work, which together we brought a challenging project forward. I want to also thank Phillip for his insights on Cyclops function and working with me on setting up a new assay. Alongside, I would also like to thank colleagues who have provided valuable knowledge on the T90 project. Amongst them are Simone Bucerius, Elaine Jesen and Marion Cerri.

The strawberry project was an important part of my doctoral study although not discussed in this thesis. It could only be possible thanks to the collaboration with *Drosophila* expert Prof. Dr. Nicolas Gompel and Dr. Lasse Bräcker, strawberry master Prof. Dr. Klaus Olbricht and his co-workers, as well as Sarah Zeiltmayr who was a bachelor and master student under my supervision for this project. I would like to thank them for their support, patience and understanding.

I had the greatest pleasure of meeting some of my best friends during this time, Juan, Fang-yu, Yen-yu, Anne, Chloe and Jessica who were also great colleagues. We shared knowledge, laughter, tears and lots of fun. I am very grateful for having them along my side during the doctoral study. They have always offered me help and support.

My warmest thanks go to a very important person in my life, Thomas Krieger, and my Chinese as well as German family members for always being there for me in the high and low times.

It was a journey that is only possible because of all of these people. Thank you all very much from the bottom of my heart.

9. SUPPLEMENTARY INFORMATION

9.1 Supplementary Figures and Tables

List of Supplementary Figures and Tables

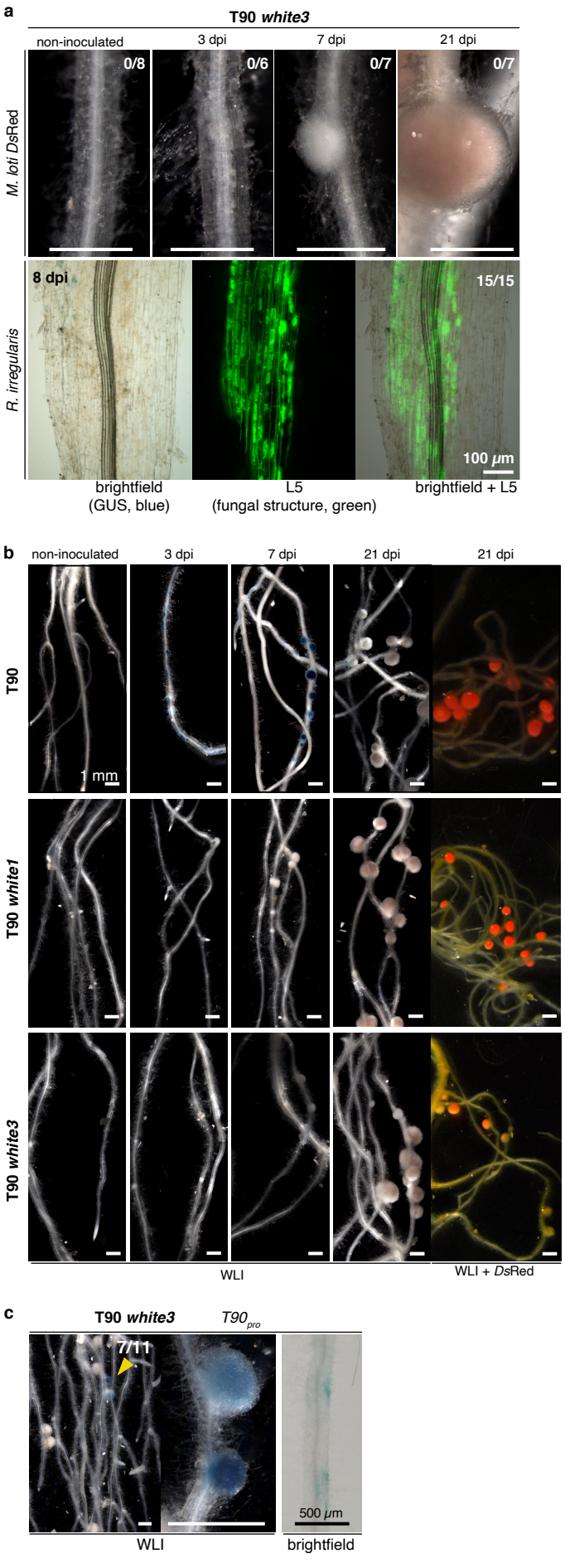


Figure S1. Absence of GUS activity in T90 *white* mutant roots during AM or RNS. **a** - **b**, T90, T90 *white1* or T90 *white3* roots were stained with X-Gluc at indicated dpi with *M. loti* *Ds*Red or *R. irregularis*. Note the total absence of GUS activity in T90 *white* roots, compared to those of T90 upon inoculation with microsymbionts (see also Fig. 3 & 13). Green: Alexa Fluor-488 WGA-stained *R. irregularis* visualised with a Leica Filter Cube L5. #/# top right corner of images: number of plants displaying GUS activity / total number of plants analysed. WLI: white light illumination. **c**, T90 *white3* hairy roots transformed with T-DNAs carrying a $Ubq10_{pro}$*:NLS-GFP* transformation marker together with a *GUS* reporter gene driven by the the T90 promoter ($T90_{pro}$) were analysed 21 dpi with *M. loti* *Ds*Red. Chimeric root systems (including transformed and non-transformed roots) were stained with X-Gluc. #/#: number of plants showing GUS activity in nodules / total number of chimeric root systems analysed. Yellow arrowhead: nodules showing *GUS* expression. Bars, 1 mm unless labeled.

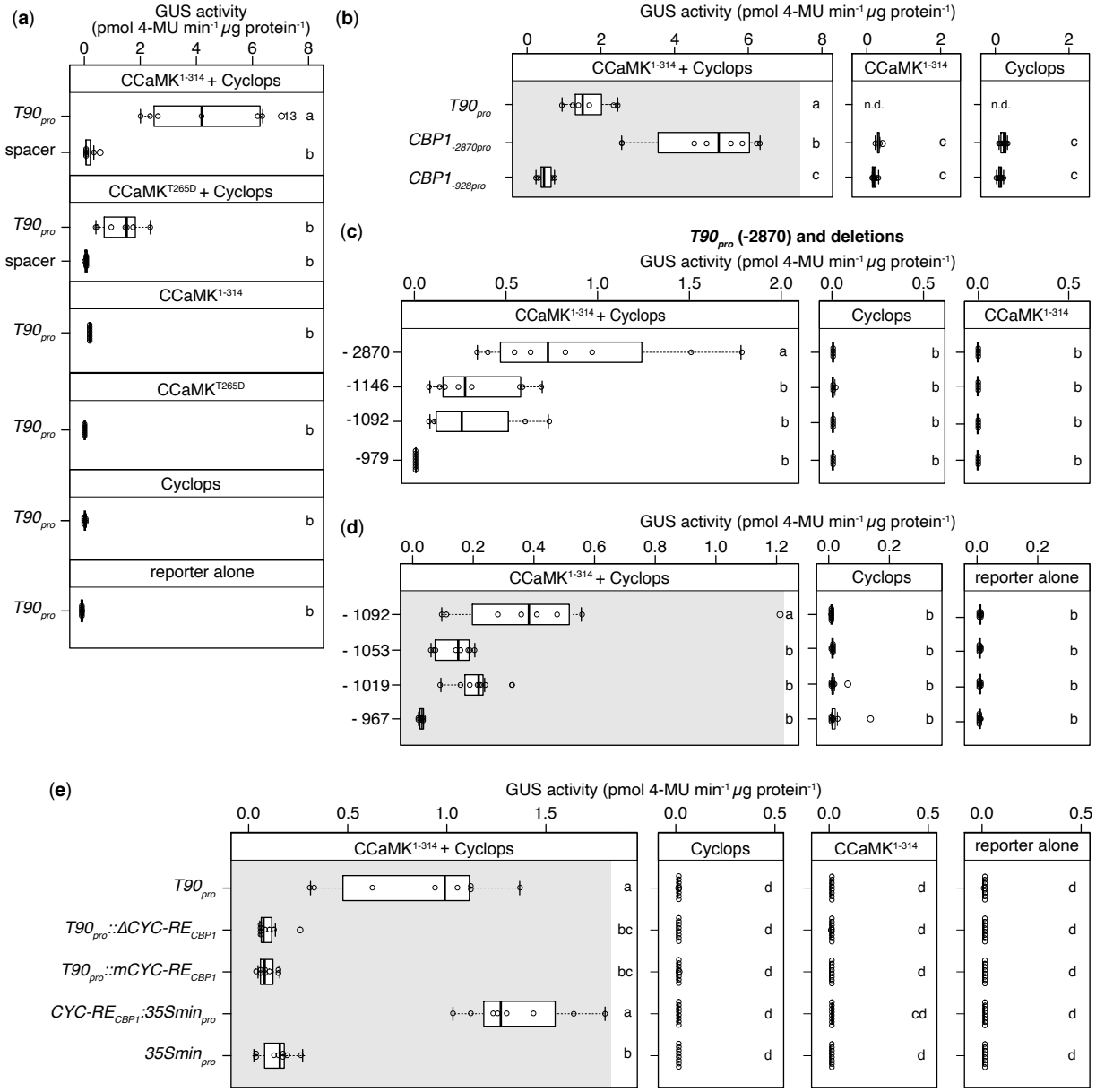


Figure. S2 A *cis*-element in the *CBP1* promoter is necessary and sufficient for CCaMK$^{1-314}$/ Cyclops -mediated transactivation. *N. benthamiana* leaf cells were transformed with T-DNAs carrying a *GUS* reporter gene driven by either of the indicated promoters: **a**, the T90 promoter (labeled as $T90_{pro}$ in a-b & e or -2870 in **c**); or a 4 bp spacer sequence; **b**, $T90_{pro}$; either one of the two *CBP1* promoters of varying length ($CBP1_{-2870pro}$ or $CBP1_{-928pro}$); **c-d**, promoter deletion series generated in the context of $T90_{pro}$ (see Fig. 18a-b); (e) same constructs as in Fig. 18e. (b-e) include controls for data depicted in Fig. 18a, d & e (grey shaded areas) and the indicated promoter regions in Fig. 18c. The applied statistical method was ANOVA with *post hoc* Tukey: (a), $F_{9,67} = 13.91$, $p = 2.59\text{x}10^{-12}$; (b), $F_{6,47} = 59.55$, $p < 2\text{x}10^{-16}$; (c), $F_{11,52} = 9.558$, $p = 4.05\text{x}10^{-9}$; (d), $F_{11,94} = 40.27$, $p = 7.71\text{x}10^{-9}$; (e), $F_{23,190} = 75.81$, $p = 1.5\text{x}10^{-15}$. Different small letters on the right side of the boxplots indicate significant difference. n.d.: not determined.

Supplementary Table 1. *Cis*-regulatory sequences identified in the promoters of RNS-related genes (continued to page 98).

Identified in promoter	Associated regulatory proteins	Plant species	*cis*-element sequence or regulatory region	Reference
NIN	Cyclops	*L. japonicus*	*CRE* (CGATTGCCATGTGGCACGCAGAGAGGAGCC); located between 961 bp and 931 bp 5' of start codon	Singh et al., 2014
NIN	n.a.	*M. truncatula*	*CE* located between 20 kb and 5 kb 5' of start codon; infection related region: within 5 kb 5' of start codon	Liu et al., 2019c
NIN	IPN2	*L. japonicus*	*IPN-RE* (AAAGAATATTTATTTATATGTTATATGTATT); located between 192 bp and 162 bp 5' of start codon	Xiao et al., 2020
ERN1	IPN2	*L. japonicus*	located between 524 bp to 76 bp 5' of start codon	Xiao et al., 2020
NPL	IPN2	*L. japonicus*	located between 914 bp to 401 bp 5' of start codon	Xiao et al., 2020
NF-Y	NIN	*L. japonicus*	NIN binding sites in promtoers of NF-YB1a (TGTATCTTTAGAGCTTTCCAAGGGATATTT); NF-YB1b (ACTCTTTTGAGCGGCCTCAAAGGCCTTCTC); NF-YA1 (GGCCCCCTTCTTATCTGGACAGGCATTTTC)	Soyano et al., 2013
*ASL18a**	NIN	*L. japonicus*	putative *NBS-S1* (TCGGCCTCTTTAATGCTTCAAGGGCCAGT); putative *NBS-S2* (TTGGCTCTTCGAATTTTCTAACCTTTTCC)	Soyano et al., 2019
EPR3	ERN1	*L. japonicus*	putative ERN1 binding site (TGTCATTGTTATAGCCGTCTGAGATCCCAC); located within 265 bp 5' of start codon	Kawaharada et al., 2017
EPR3	NIN	*L. japonicus*	putative NIN binding site (ATACCCTGCCCTTACAGACAAGTATAAGAGGCACAAAA)	Kawaharada et al., 2017
NIN	SIP1	*L. japonicus*	putative SIP1 binding site (TCAATTCAATTAAT)	Zhu et al., 2008

Identified in promoter	Associated regulatory proteins	Plant species	*cis*-element sequence or regulatory region	Reference
NIN	SIP1	*L. japonicus*	putative SIP1 binding site (ACAATTAAAGA)	Zhu et al., 2008
NIN	NSP1	*L. japonicus*	located between -892 bp and -13 bp	Hirsch et al., 2009
ERN1	NSP1	*L. japonicus*	located between -862 bp and -29 bp	Hirsch et al., 2009
CRE1	NIN	*M. truncatula*	located between -1487 bp to -971 bp; multiple putative sites	Vernie et al., 2015
ENOD11	NIN	*M. truncatula*	located between -1046 to +3, that contains the NF-box)	Vernie et al., 2015
CLE-RS1	NIN	*L. japonicus*	two regions: located between –5193 to –4631 bp and –2368 to –2187 bp	Soyano et al., 2014
CLE-RS2	NIN	*L. japonicus*	two regions: located between –5193 to –4631 bp and –2368 to –2187 bp	Soyano et al., 2014
MtCLE13	NIN	*M. truncatula*	located within ca. 2000 bp 5' of predicted start codon, at ca. -700 bp	Laffont et al., 2020
MtCEP7	NIN	*M. truncatula*	located within ca. 2400 bp 5' of predicted start codon; at ca. -250 bp	Laffont et al., 2020
RR4	RR1	*M. truncatula*	*RRBS* 12 bp consensus {AAT(G/A)AGA(C/T)TAGT}; box2 (AATGAGACATAT)	Ariel et al., 2012
NSP2	RR1	*M. truncatula*	*RRBS* 12 bp consensus {AAT(G/A)AGA(C/T)TAG}T; box5 (AATGAGATTAGC); box6 (GATGAGACTTAA)	Ariel et al., 2012
MIR171§	RR2	*M. truncatula*	box7 (AATAAGATTCAC); *RRBS*-like (AATAAGATCTCT)	Ariel et al., 2012
MtLAX2§	n.a.	*M. truncatula*	three TGTCTC auxin responsive elements and two truncated TGTCT motifs located within ca. 2000 bp 5' of predicted start codon	Roy et al., 2017

Identified in promoter	Associated regulatory proteins	Plant species	*cis*-element sequence or regulatory region	Reference
ENOD11 *	n.a.	*M. truncatula*	AT-rich motif {TTATT(N)7-12AATAA}	Boisson-Dernier et al., 2005
ENOD11	ERN1	*M. truncatula*	NF box (TAATAACATAAAATAATTGCAGGCCTAAAGCT); located between 390 bp and 358 bp 5' of start codon	Andriankaja et al., 2007
ENOD11	NSP1 & NSP2	*M. truncatula*	within 257 bp 5' of start codon	Cerri et al., 2012; Boisson-Dernier et al., 2005
ERN1	Cyclops	*L. japonicus*	*CYC-RE*$_{ERN1}$ (TTTGGAGCCTCCATGTGGCAGTCGTTCATG)	Cerri et al., 2017
MtNSP2	MtRRB3	*M. truncatula*	located at ca. 850 bp 5' of start codon	Tan et al., 2020
MtCCS52A	MtRRB3	*M. truncatula*	located at ca. 1673 bp 5' of start codon	Tan et al., 2020
MtSCR	n.a.	*M. truncatula*	AT1-box (AATATTTTTTTT) located between 1604 and 1615 bp presumbaly 5' of start codon	Dong et al. 2020
MtSCR	n.a.	*M. truncatula*	En (GTAATTC) located between 1632 and 1638 bp presumbaly 5' of start codon	Dong et al. 2020
N23	n.a.	*G. max*	located between 344 bp and 293bp as well as between 247 bp and 165 bp	Jorgensen et al., 1988
lbc3	n.a.	*G. max*	*OSE* (GTTTTTGAAAAGATGATTGTCTCTTCACCATACCAAT)	Stougaard et al., 1987b
Lb29	n.a.	*V. faba*	located between 325 bp and 175 bp 5' of start codon	Fehlberg et al., 2005

Identified in promoter	Associated regulatory proteins	Plant species	*cis*-element sequence or regulatory region	Reference
NIN	n.a.	*L. japonicus*	Gibberellic acid responsive region located between -1244 bp to -1169 bp	Akamatsu et al., 2021
NIN	n.a.	*L. japonicus*	Gibberellic acid responsive region located between -1245 bp to -1086 bp	Akamatsu et al., 2021

n.a.: not applicable or not available

*identified in the intron of the gene

§ *in silico* or mannual prediction

Supplementary Table 2. Plant and bacterial material used in this work.

Figure	Plant material		Symbiont	Time point
	Plant genotype	Seedbag no.		
Fig. 3	T90	91664	*M. loti Ds*Red	3, 7 or 21 dpi
Fig. 3	T90	92731	*R. irregularis*	8 dpi
Fig. 6	Gifu WT	92673	*M. loti Ds*Red	10 to 14 dpi
Fig. 7a	Gifu WT	111268	*M. loti lacZ*	10 to 14 dpi
Fig. 7b	*lhk1-1*	92056	*M. loti Ds*Red	7, 8 or 21 dpi
Fig. 8	*nin-15*	111284 111281 111292	*M. loti Ds*Red	21 dpi
Fig. 9	*nin-15* *nin-15* *nin-15* *nin-15* *nin-15* *nin-15* *nin-15* *nin-15*	111278 111638 111281 111286 111293 111292 111285 111636	*M. loti Ds*Red	21 dpi
Fig. 10	*nin-15*	same as Fig. 9	*M. loti Ds*Red	21 dpi
Fig. 11	*nin-15*	same as Fig. 10	*M. loti Ds*Red	21 dpi
Fig. 12c	T90 *w1*	92822; 92823; 93824	*M. loti Ds*Red	14 dpi
Fig. 12c	T90 *w3*	88256	*M. loti Ds*Red	21 dpi
Fig. 12b	T90 *w1*	88502	*M. loti Ds*Red	3, 7 or 21 dpi
Fig. 12b	T90 *w1*	110519	*R. irregularis*	8 dpi
Fig. 13a	Gifu WT T90 T90 *w1* T90 *w2*	111218 92730 110518 113530	n.a.	25 dpt
Fig. 13b	Gifu WT T90 T90 *w1*	92673 92813 110519	*M. loti Ds*Red	7, 10, 14, 28 or 49 dpi
Fig. 14	Gifu WT T90 *w1* T90 *w3*	92665 92822; 92823; 93824 88256	*M. loti Ds*Red	15 dpi 14 dpi 21 dpi

Supplementary Table 2. Continued.

Figure	Plant material		Symbiont	Time point
	Plant genotype	Seedbag no.		
Fig. 15	Gifu WT	92665	*M. loti Ds*Red	25 dpt
	T90	92809		25 dpt
	T90 *w1*	92822		25 dpt
	T90 *w2*	113530		26 dpt
	T90 *w3*	110511		26 dpt
Fig. 17	Gifu WT	111220	*R. irregularis*	12 dpi
Fig. 17	Gifu WT	111218	*M. loti Ds*Red	10 to 14 dpi
Fig. 18	Gifu WT	111218	*M. loti Ds*Red	11 to 14 dpi
Fig. 19	Gifu WT	111220	*R. irregularis*	12 or 13 dpi
Fig. 20	Gifu WT	111268	*M. loti Ds*Red	14 or 21 dpi
Supplementary Fig. 1a	T90 *w3*	110512	*R. irregularis*	8 dpi
	T90 *w3*	110512	*M. loti Ds*Red	3, 7 or 21 dpi
Supplementary Fig. 1b	T90	91664	*M. loti Ds*Red	3, 7 or 21 dpi
	T90 *w1*	88502		
	T90 *w3*	88257		
Supplementary Fig. 1c	T90 *w3*	88256	*M. loti Ds*Red	21 dpi

Gifu*: Lotus japonicus* accession B-129
nin-15: Lotus japonicus LORE1 line 30003529
*M. loti Ds*Red*: Mesorhizobium loti* MAFF 303099 *Ds* Red
M. loti lacZ: Mesorhizobium loti MAFF 303099 *lacZ*
R. irregularis: Rhizophagus irregularis inoculated in a chive nursing pot
dpi and dpt: days post inoculation and days post transfer, respectively
n.a.: not applicable

Supplementary Table 3. Plasmids used in this study (continued to page 106). Plasmids (LIII) were used for experiments presented in this week. LIII plasmids were generated with standard Golden Gate modules (Binder et al., 2014) and customized L0, LI and LII modules. XGp8 and XGp179a are master backbone only used to generate other fusion constructs. Ref no.: reference number. Plasmids with a reference number containing XG or RA were generated by Xiaoyun Gong and Rosa Elena Andrade, respectively.

LIII Ref no.	Name	LIII Ref no.	Name
XGp8	LIIIβ *lacZ:GUSi* (master backbone)	**XGp190**	LIIIβ*CYC-RE$_{CBP1}$:35Smin$_{pro}$:DoGUS*
XGp179a	LIIIβ *lacZ:DoGUS* (master backbone)	**XGp192**	LIIIβ *35Smin$_{pro}$:DoGUS*
XGp9	LIIIβ *T90$_{pro}$:GUSi*	**XGp75**	LIIIβ *LjNIN$_{pro}$:GUSi*
XGp10	LIIIβ *T90$_{pro}$* (-2365 bp):*GUSi*	**pRA121**	LIIIβ *LjNIN$_{pro}$::ΔPACE:GUSi*
XGp11	LIIIβ *T90$_{pro}$* (-1925 bp):*GUSi*	**pRA122**	LIIIβ *LjNIN$_{pro}$::mPACE:GUSi*
XGp12	LIIIβ *T90$_{pro}$* (-1327 bp):*GUSi*	**XGp74**	LIIIβ *LjNINmin$_{pro}$:GUSi*
XGp13	LIIIβ *T90$_{pro}$* (-979 bp):*GUSi*	**XGp71**	LIIIβ *LjPACE:NINmin$_{pro}$:GUSi*
XGp14	LIIIβ *dy*:*GUSi*	**XGp97**	LIIIβ *CgPACE:NINmin$_{pro}$:GUSi*
XGp15	LIIIβ *T90$_{pro}$*(-1299 bp):*GUSi*	**XGp98**	LIIIβ *DdPACE:NINmin$_{pro}$:GUSi*
XGp16	LIIIβ *T90$_{pro}$* (-1146 bp):*GUSi*	**XGp99**	LIIIβ *Dg1PACE:NINmin$_{pro}$:GUSi*
XGp17	LIIIβ *T90$_{pro}$* (-1092 bp):*GUSi*	**XGp100**	LIIIβ *Dg2PACE:NINmin$_{pro}$:GUSi*
XGp18	LIIIβ *T90$_{pro}$*(-2799 bp):*GUSi*	**pRA66**	LIIIβ *lacZ:NINmin$_{pro}$:NIN*
XGp19	LIIIβ *T90$_{pro}$* (-2672 bp):*GUSi*	**pRA146**	LIIIβ *LjPACE:NINmin$_{pro}$:NIN*
XGp20	LIIIβ *T90$_{pro}$* (-2329 bp):*GUSi*	**pRA20**	LIIIβ *lacZ:NIN*
XGp21	LIIIβ *T90$_{pro}$* (-2479 bp):*GUSi*	**pRA21**	LIIIβ *NIN$_{pro}$:NIN*
XGp126	LIIIβ *T90$_{pro}$* (-1053 bp):*GUSi*	**pRA119**	LIIIβ *NIN$_{pro}$::ΔPACE:NIN*
XGp127	LIIIβ *T90$_{pro}$* (-1019 bp):*GUSi*	**pRA120**	LIIIβ *NIN$_{pro}$::mPACE:NIN*
XGp128	LIIIβ *T90$_{pro}$* (-967 bp):*GUSi*	**pRA135**	LIIIβ *LjNIN$_{pro}$::CgPACE:NIN*
XGp125	LIIIβ *CYC-RE$_{CBP1}$:35Smin$_{pro}$:GUSi*	**pRA136**	LIIIβ *LjNIN$_{pro}$::DdPACE:NIN*
XGp41	LIIIβ *35Smin$_{pro}$:GUSi*	**pRA137**	LIIIβ *LjNIN$_{pro}$::Dg2PACE:NIN*
XGp136	LIIIβ *T90$_{pro}$::ΔCYC-RE$_{CBP1}$:GUSi*	**pRA138**	LIIIβ *LjNIN$_{pro}$::Dg1PACE:NIN*
XGp137	LIIIβ *T90$_{pro}$::mCYC-RE$_{CBP1}$:GUSi*	**XGp142**	LIIIβ *LjNIN$_{pro}$::ZjPACE:NIN*
XGp65	LIIIβ *CBP1$_{-2870pro}$:GUSi*	**XGp143**	LIIIβ *LjNIN$_{pro}$::PpPACE:NIN*
XGp66	LIIIβ *CBP1$_{-928pro}$:GUSi*	**XGp144**	LIIIβ *LjNIN$_{pro}$::JrPACE-like:NIN*
XGp187	LIIIβ *T90$_{pro}$:DoGUS*	**XGp110**	LIIIβ *SlNIN$_{pro}$::184LjPACE:NIN*
XGp115	LIIIβ *SlNIN$_{pro}$:NIN*	**XGp111**	LIIIβ *SlNIN$_{pro}$::184mLjPACE:NIN*
XGp215	LIIIβ F *Ub10$_{pro}$:NLS-2xGFP - LjUbi$_{pro}$:CBP1-mCherry*		

LII Ref no.	Name
XGp3	LII F 3-4 *lacZ:GUSi*
CCp5	LIIc F 1-2 *Ub10$_{pro}$:NLS-2xGFP*
XGp134	LIIc F 4-5 *T90$_{pro}$::ΔCYC-RE$_{CBP1}$*
XGp135	LIIc F 4-5 *T90$_{pro}$::mCYC-RE$_{CBP1}$*
XGp172	LII F 3-4 *lacZ:DoGUS*
XGp122	LIIβ F 3-4 *LjUbi$_{pro}$:CBP1-mCherry*
pRA65	LIIβ F 3-4 *lacZ:NINmin$_{pro}$:NIN*
pRA19	LIIβ F 3-4 *lacZ:NIN*

LI Ref no.	Name
XGp5	L0 *T90$_{pro}$*
XGp91	L0 *LjNINmin$_{pro}$* (96 bp)
XGp92	L0 *35Smin$_{pro}$*
XGp68	LI A-B *LjPACE:NINmin$_{pro}$*
XGp93	LI A-B *CgPACE:NINmin$_{pro}$*
XGp94	LI A-B *DdPACE:NINmin$_{pro}$*
XGp95	LI A-B *Dg1PACE:NINmin$_{pro}$*
XGp96	LI A-B *Dg2PACE:NINmin$_{pro}$*
XGp62	L0 *CBP1$_{-2870pro}$*
XGp63	L0 *CBP1$_{-928pro}$*
XGp119	L0 *CBP1* without stop codon
pRA130	LI *LjNIN$_{pro}$::CgPACE* (2 kb)
pRA131	LI *LjNIN$_{pro}$::DdPACE* (2 kb)
pRA133	LI *LjNIN$_{pro}$::DgPACE* (2 kb)
XGp129	L0 *T90$_{pro}$* fragment 1

LI Ref no.	Name
XGp130	L0 *T90$_{pro}$::ΔCYC-RE$_{CBP1}$* Frag 2.1
XGp131	L0 *T90$_{pro}$::ΔCYC-RE$_{CBP1}$* Frag 2.2
XGp132	L0 *T90$_{pro}$::mCYC-RE$_{CBP1}$* Frag 2.1
XGp133	L0 *T90$_{pro}$::mCYC-RE$_{CBP1}$* Frag 2.2
XGp138	LI *LjNIN$_{pro}$::ZjPACE* (2 kb)
XGp139	LI *LjNIN$_{pro}$::PpPACE* (2 kb)
XGp140	LI *LjNIN$_{pro}$::JrPACE-like* (2 kb)
pRA116	LI *LjNIN$_{pro}$::mPACE* (2 kb)
pRA117	LI *LjNIN$_{pro}$::ΔPACE* (2 kb)
pRA64	LI B-C *LjNINmin$_{pro}$* (96 bp)
pRA1	LI C-D *LjNIN*
pRA145	L0 A-B *LjPACE*
XGp106	L0 *pSly184PACE-1*
XGp107	L0 *pSly184PACE-2*
XGp108	L0 *pSly184mPACE-1*
XGp109	L0 *pSly184mPACE-2*
XGp124	LI *CYC-RE$_{CBP1}$:35Smin$_{pro}$*
KPp10	LI A-B *SlNIN$_{pro}$* (3kb)
KPp13	LI A-B *LjNIN$_{pro}$* (1 kb)
KPp14	LI A-B *LjNIN$_{pro}$* (2 kb)

Plasmid construction							
LIII Ref no.	**Restriction enzyme for cut-ligation reaction**	**Golden Gate modules used for assembly**					**Backbone**
XGp8	BpiI	LIIc F 1-2 *Ub10$_{pro}$:NLS-2xGFP* (CCp5)	LII 2-3 ins (BB43)	LII 3-4 *lacZ:GUSi* (XGp3)	LII 4-5 dummy (BB40)	LII 5-6 dy (BB65)	LIIIβ fin BB52
XGp9	Esp3I	L0 *T90$_{pro}$* (XGp5)					LIIIβ *lacZ:GUSi* (XGp8)
XGp10	Esp3I	PCR fragment amplifed with primers XG8 and XG13 using XGp5 as template					LIIIβ *lacZ:GUSi* (XGp8)
XGp11	Esp3I	PCR fragment amplifed with primers XG8 and XG14 using XGp5 as template					LIIIβ *lacZ:GUSi* (XGp8)
XGp12	Esp3I	PCR fragment amplifed with primers XG8 and XG15 using XGp5 as template					LIIIβ *lacZ:GUSi* (XGp8)
XGp13	Esp3I	PCR fragment amplifed with primers XG8 and XG5 using XGp5 as template					LIIIβ *lacZ:GUSi* (XGp8)
XGp14	Esp3I	LI A-B BB05					LIIIβ *lacZ:GUSi* (XGp8)
XGp15	Esp3I	PCR fragment amplifed with primers XG8 and XG16 using XGp5 as template					LIIIβ *lacZ:GUSi* (XGp8)
XGp16	Esp3I	PCR fragment amplifed with primers XG8 and XG17 using XGp5 as template					LIIIβ *lacZ:GUSi* (XGp8)
XGp17	Esp3I	PCR fragment amplifed with primers XG8 and XG18 using XGp5 as template					LIIIβ *lacZ:GUSi* (XGp8)
XGp18	Esp3I	PCR fragment amplifed with primers XG8 and XG19 using XGp5 as template					LIIIβ *lacZ:GUSi* (XGp8)
XGp19	Esp3I	PCR fragment amplifed with primers XG8 and XG20 using XGp5 as template					LIIIβ *lacZ:GUSi* (XGp8)
XGp20	Esp3I	PCR fragment amplifed with primers XG8 and XG21 using XGp5 as template					LIIIβ *lacZ:GUSi* (XGp8)
XGp21	Esp3I	PCR fragment amplifed with primers XG8 and XG22 using XGp5 as template					LIIIβ *lacZ:GUSi* (XGp8)
XGp126	Esp3I	PCR fragment amplifed with primers XG8 and XG91 using XGp5 as template					LIIIβ *lacZ:GUSi* (XGp8)
XGp127	Esp3I	PCR fragment amplifed with primers XG8 and XG92 using XGp5 as template					LIIIβ *lacZ:GUSi* (XGp8)
XGp128	Esp3I	PCR fragment amplifed with primers XG8 and XG93 using XGp5 as template					LIIIβ *lacZ:GUSi* (XGp8)
XGp125	Esp3I	LI *CYC-RE$_{T90}$:35Smin$_{pro}$* (XGp124)					LIIIβ *lacZ:GUSi* (XGp8)
XGp41	Esp3I	PCR fragment amplifed with primers XG39 and XG40 using LI A-B 35S (G005) as template					LIIIβ *lacZ:GUSi* (XGp8)
XGp136	Esp3I	L0 *T90$_{pro}$* fragment 1	LIIc F 4-5 *T90$_{pro}$::ΔCYC-RE$_{T90}$*				LIIIβ *lacZ:GUSi* (XGp8)
XGp137	Esp3I	L0 *T90$_{pro}$* fragment 1	LIIc F 4-5 *T90$_{pro}$::mCYC-RE$_{T90}$*				LIIIβ *lacZ:GUSi* (XGp8)
XGp65	Esp3I	L0 *CBP1-F$_{pro}$* (XGp62)					LIIIβ *lacZ:GUSi* (XGp8)
XGp66	Esp3I	L0 *CBP1-d$_{pro}$* (XGp63)					LIIIβ *lacZ:GUSi* (XGp8)
XGp179a	Esp3I	LIIc F 1-2 *Ub10$_{pro}$:NLS-2xGFP* (CCp5)	LII 2-3 ins (BB43)	LII F 3-4 *lacZ:DoGUS* (XGp172)	LII 4-5 dummy (BB40)	LII 5-6 dy (BB65)	LIIIβ fin BB52
XGp187	Esp3I	L0 *T90$_{pro}$* (XGp5)					LIIIβ *lacZ:DoGUSi* (XGp179a)
XGp190	Esp3I	LI *CYC-RE$_{T90}$:35Smin$_{pro}$* (XGp124)					LIIIβ *lacZ:DoGUSi* (XGp179a)
XGp192	Esp3I	PCR fragment amplifed with primers XG39 and XG40 using LI A-B 35S (G005) as template					LIIIβ *lacZ:DoGUSi* (XGp179a)
XGp75	Esp3I	LI *LjNIN$_{pro}$* (1kb) (KPp13)	LI *LjNIN$_{pro}$* (2kb) (KPp14)				LIIIβ *lacZ:GUSi* (XGp8)

Plasmid construction

LIII Ref no.	Restriction enzyme for cut-ligation reaction	Golden Gate modules used for assembly					Backbone
pRA121	Esp3I	LI *LjNIN$_{pro}$* (1kb) (KPp13)	LI *LjNINΔPACE$_{pro}$* (2kb) (pRA117)				LIIIβ *lacZ:GUSi* (XGp8)
pRA122	Esp3I	LI *LjNIN$_{pro}$* (1kb) (KPp13)	LI *LjNINmPACE$_{pro}$* (2kb) (pRA116)				LIIIβ *lacZ:GUSi* (XGp8)
XGp74	Esp3I	*LjNINmin$_{pro}$* amplified with primers XG24 and XG70 using KPp14 as template					LIIIβ *lacZ:GUSi* (XGp8)
XGp71	Esp3I	LI A-B *LjPACE:NINmin$_{pro}$* (XGp68)					LIIIβ *lacZ:GUSi* (XGp8)
XGp97	Esp3I	LI A-B *CgPACE:NINmin$_{pro}$* (XGP93)					LIIIβ *lacZ:GUSi* (XGp8)
XGp98	Esp3I	LI A-B *DdPACE:NINmin$_{pro}$* (XGP94)					LIIIβ *lacZ:GUSi* (XGp8)
XGp99	Esp3I	LI A-B *Dg1PACE:NINmin$_{pro}$* (XGp95)					LIIIβ *lacZ:GUSi* (XGp8)
XGp100	Esp3I	LI A-B *Dg2PACE:NINmin$_{pro}$* (XGp96)					LIIIβ *lacZ:GUSi* (XGp8)
pRA66	Esp3I	LIIc F 1-2 *UBI10$_{pro}$:NLS-2xGFP* (CCp4)	LII 2-3 dy (BB39)	LIIβ F 3-4 *lacZ:NINmin$_{pro}$:NIN* (pRA65)	II 4-6 dy (BB41)		LIIIβ F A-B BB53
pRA146	Esp3I	L0 A-B *LjPACE* (pRA145)					LIIIβ *lacZ:NINmin$_{pro}$:NIN* (pRA66)
pRA20	BpiI	LIIc F 1-2 *AtUBI10$_{pro}$:NLS-2xGFP* (CCp4)	LII 2-3 dy (BB39)	LIIβ F 3-4 *lacZ:NIN* (pRA19)	LII 4-6 dy (BB41)		LIIIβ F A-B BB53
pRA21	Esp3I	LI *LjNIN$_{pro}$* (1kb) (KPp13)	LI *LjNIN$_{pro}$* (2kb) (KPp14)				LIIIβ lacZ:NIN (pRA20)
pRA119	Esp3I	LI *LjNIN$_{pro}$* (1kb) (KPp13)	LI *LjNIN$_{pro}$::ΔPACE$_{pro}$* (2 kb) (pRA117)				LIIIβ lacZ:NIN (pRA20)
pRA120	Esp3I	LI *LjNIN$_{pro}$* (1kb) (KPp13)	LI *LjNIN$_{pro}$::mPACE* (2 kb) (pRA116)				LIIIβ lacZ:NIN (pRA20)
pRA135	Esp3I	LI *LjNIN$_{pro}$* (1kb) (KPp13)	LI *LjNIN$_{pro}$::CgPACE* (2 kb) (pRA130)				LIIIβ lacZ:NIN (pRA20)
pRA136	Esp3I	LI *LjNIN$_{pro}$* (1kb) (KPp13)	LI *LjNIN$_{pro}$::DdPACE* (2 kb) (pRA131)				LIIIβ lacZ:NIN (pRA20)
pRA137	Esp3I	LI *LjNIN$_{pro}$* (1kb) (KPp13)	LI *LjNIN$_{pro}$::Dg1PACE* (2 kb) (pRA132)				LIIIβ lacZ:NIN (pRA20)
pRA138	Esp3I	LI *LjNIN$_{pro}$* (1kb) (KPp13)	LI *LjNIN$_{pro}$::Dg2PACE* (2 kb) (pRA133)				LIIIβ lacZ:NIN (pRA20)
XGp142	Esp3I	LI *LjNIN$_{pro}$* (1kb) (KPp13)	LI *LjNIN$_{pro}$::ZjPACE* (2 kb) (XGp138)				LIIIβ lacZ:NIN (pRA20)
XGp143	Esp3I	LI *LjNIN$_{pro}$* (1kb) (KPp13)	LI *LjNIN$_{pro}$::PpPACE* (2 kb) (XGp139)				LIIIβ lacZ:NIN (pRA20)
XGp144	Esp3I	LI *LjNIN$_{pro}$* (1kb) (KPp13)	LI *LjNIN$_{pro}$::JrPACE-like* (2 kb) (XGp140)				LIIIβ lacZ:NIN (pRA20)
XGp110	Esp3I	Three PCR fragment amplifed with primers KP50 and KP60, KP61 and KP59, or KP58 and KP63 using *S. lycopersicum* (cultivarmoneymaker) genomic DNA as template			L0 *pSly184PACE-1* (XGp106)	*L0 pSly184PACE-2* (XGp107)	LIIIβ lacZ:NIN (pRA20)
XGp111	Esp3I	Three PCR fragment amplifed with primers KP50 and KP60, KP61 and KP59, or KP58 and KP64 using S. lycopersicum (cultivar moneymaker) genomic DNA as template			L0 *pSly184mPACE-1* (XGp108)	*L0 pSly184mPACE-2* (XGp109)	LIIIβ lacZ:NIN (pRA20)
XGp115	Esp3I	LI A-B *SlNIN$_{pro}$* (3kb) (KPp10)					LIIIβ lacZ:NIN (pRA20)
XGp215	BpiI	LIIc F 1-2 *Ub10$_{pro}$:NLS-2xGFP* (CCp5)	LII 2-3 ins (BB43)	LIIβ F 3-4 *LjUbi$_{pro}$:CBP1-mCherry* (XGp122)	LII 4-5 dummy (BB40)	LII 5-6 dy (BB65)	LIIIβ fin BB52

Plasmid construction								
LII Ref no.	**Restriction enzyme for cut-ligation reaction**	**Golden Gate modules used for assembly**						**Backbone**
XGp3	BsaI	LI A-B lacZ (G082)	LI B-E *GUSi* (G079)	LI B-E *GUSi* (G079)	LI B-E GUSi (G079)	LI E-F Nos-T (G006)	LI F-G dy (BB09)	LII F 3-4 (BB33)
CCp5	BsaI	LI A-B $AtUBI10_{pro}$	LI B-C NLS (BB060)	LI C-D GFP (G053)	LI D-E GFP (G011)	LI E-F 35S-T (G059)	LI F-G dy (BB09)	LII F 1-2 (BB30)
XGp134	BsaI	L0 $T90_{pro}$::$\Delta CYC\text{-}RE_{CBP1}$ Frag 2.1 (XGp130)	L0 $T90_{pro}$::$\Delta CYC\text{-}RE_{T90}$ Frag 2.2 (XGp131)					LIIc F 4-5 (BB35)
XGp135	BsaI	L0 $T90_{pro}$::$mCYC\text{-}RE_{CBP1}$ Frag 2.1 (XGp132)	L0 $T90_{pro}$::$mCYC\text{-}RE_{T90}$ Frag 2.2 (XGp133)					LIIc F 4-5 (BB35)
XGp172	BsaI	LI A-B lacZ (G082)	LI B-C dy (BB06)	LI C-D *DoGUS* (provided by David Chiasson)	LI D-E dy (BB08)	LI E-F Nos-T (G006)	LI F-G dy (BB09)	LII F 3-4 (BB33) LII F 1-2 (BB30)
XGp122	BsaI	LI A-B $LjUbi1_{pro}$ (G007)	LI B-C dy (BB06)	L0 *CBP1* without stop codon (XGp119)	LI D-E mCherry (G025)	LI E-F nos-T (G006)	LI F-G dy (BB09)	LIIβ F 3-4 (BB24) LIIβ F 3-4 (BB24)
pRA65	BsaI	LI A-B lacZ (G082)	LI B-C $LjNINmin_{pro}$ (pRA64)	LI C-D *LjNIN*	LI D-E dy (BB08)	LI E-F nos-T (G006)	LI F-G dy (BB09)	LIIβ F 3-4 (BB24)
pRA19	BsaI	LI A-B lacZ (G082)	LI B-C dy (BB06)	LI C-D *LjNIN*	LI D-E dy (BB08)	LI E-F 35S-T (G059)	LI F-G dy (BB09)	LIIβ F 3-4 (BB24)

Plasmid construction

LI Ref no.	Restriction enzyme for cut-ligation reaction	Inserts used for Golden Gate module assembly		Backbone	Notes
XGp5	StuI	PCR fragment amplified with primer XG8 and XG9		L0 (BB02)	PCR fragments were amplified using T90 genomic DNA as template
XGp91	StuI	*NINminpro* amplified with primers XG23 & XG24		L0 (BB01)	PCR fragment was amplified using $LjNIN_{pro}$ (2 kb)(KPp14l as template
XGp92	StuI	*35minpro* amplified with primers XG38 & XG39		L0 (BB01)	PCR fragment was amplified using 35S promoter as template
XGp68	BpiI	*LjPACE* annealed with oligos XG32 & XG33	$NINmin_{pro}$ (XGp91)	L0 (BB03)	
XGp93	BpiI	*CgPACE* annealed with oligos XG79 & XG80	$NINmin_{pro}$ (XGp91)	L0 (BB03)	
XGp94	BpiI	*DdPACE* annealed with oligos XG81 & XG82	$NINmin_{pro}$ (XGp91)	L0 (BB03)	
XGp95	BpiI	*Dg1PACE* annealed with oligos XG83 & XG84	$NINmin_{pro}$ (XGp91)	L0 (BB03)	
XGp96	BpiI	*Dg2PACE* annealed with oligos XG85 & XG86	$NINmin_{pro}$ (XGp91)	L0 (BB03)	
XGp62	StuI	PCR fragment amplified with primer XG9 and XG53		L0 (BB01)	PCR fragments were amplified using Gifu wild-type genomic DNA as template
XGp63	StuI	PCR fragment amplified with primer XG52 and XG53		L0 (BB01)	PCR fragments were amplified using Gifu wild-type genomic DNA as template
XGp119	StuI	PCR fragment amplified with primer XG89 and XG90		L0 (BB01)	PCR fragment was amplified using XGp67 (Ubi_{10}:NLS:2xGFP-$CBP1_{pro}$:gCBP1) as template. XGp67 not included in this table
pRA130	BpiI	Fragment 1 amplified with primers RA114 and RA124	Fragment 2 amplified with primers RA125 and RA117	L0 (BB03)	two PCR fragments were amplified using LI $LjNIN_{pro}$ (2 kb)(KPp14) as template
pRA131	BpiI	Fragment 1 amplified with primers RA114 and RA126	Fragment 2 amplified with primers RA127 and RA117	L0 (BB03)	two PCR fragments were amplified using LI $LjNIN_{pro}$ (2 kb)(KPp14) as template
pRA132	BpiI	Fragment 1 amplified with primers RA114 and RA128	Fragment 2 amplified with primers RA129 and RA117	L0 (BB03)	two PCR fragments were amplified using LI $LjNIN_{pro}$ (2 kb)(KPp14) as template
pRA133	BpiI	Fragment 1 amplified with primers RA114 and RA130	Fragment 2 amplified with primers RA131 and RA117	L0 (BB03)	two PCR fragments were amplified using LI $LjNIN_{pro}$ (2 kb)(KPp14) as template
XGp129	StuI	PCR fragment amplified with primer XG96 and XG9		L0 (BB02)	PCR fragment was amplified using XGp5 as template
XGp130	StuI	PCR fragment amplified with primer XG97 and XG98		L0 (BB02)	PCR fragment was amplified using XGp5 as template
XGp131	StuI	PCR fragment amplified with primer XG99 and XG100		L0 (BB02)	PCR fragment was amplified using XGp5 as template
XGp132	StuI	PCR fragment amplified with primer XG97 and XG101		L0 (BB02)	PCR fragment was amplified using XGp5 as template
XGp133	StuI	PCR fragment amplified with primer XG100 and XG102		L0 (BB02)	PCR fragment was amplified using XGp5 as template
XGp138	BpiI	Fragment 1 amplified with primers RA114 and XG105	Fragment 2 amplified with primers XG106 and RA117	L0 (BB03)	two PCR fragments were amplified using LI $LjNIN_{pro}$ (2 kb)(KPp14) as template
XGp139	BpiI	Fragment 1 amplified with primers RA114 and XG107	Fragment 2 amplified with primers XG108 and RA117	L0 (BB03)	two PCR fragments were amplified using LI $LjNIN_{pro}$ (2 kb)(KPp14) as template
XGp140	BpiI	Fragment 1 amplified with primers RA114 and XG109	Fragment 2 amplified with primers XG110 and RA117	L0 (BB03)	two PCR fragments were amplified using LI $LjNIN_{pro}$ (2 kb)(KPp14) as template
pRA116	BpiI	Fragment 1 amplified with primers RA114 and RA118	Fragment 2 amplified with primers RA119 and RA117	L0 (BB03)	two PCR fragments were amplified using LI $LjNIN_{pro}$ (2 kb)(KPp14) as template
pRA117	BpiI	Fragment 1 amplified with primers RA114 and RA115	Fragment 2 amplified with primers RA116 and RA117	L0 (BB03)	two PCR fragments were amplified using LI $LjNIN_{pro}$ (2 kb)(KPp14) as template
pRA64	BpiI	PCR fragment amplified with primers RA87 and RA88		L0 (BB03)	PCR fragments were amplified using LI $LjNIN_{pro}$ (2 kb)(KPp14) as template
pRA1	BpiI	Three fragments primers JL65 and JL66, JL67 and JL68, or JL69 and JL70		L0 (BB03)	PCR fragment amplified using *L. japonicus* gDNA as template; Mutagenized for removing BpiI/BsaI sites for Golden Gate cloning
pRA145	StuI	*LjPACE* annealed with oligos XG71 & XG72		L0 (BB02)	
XGp106	StuI	PCR fragment amplified with primers KP62 and SZ08		L0 (BB02)	PCR fragment amplified using *S. lycopersicum* gDNA as template
XGp107	StuI	PCR fragment amplified with primers SZ02 and SZ07		L0 (BB02)	PCR fragment amplified using *S. lycopersicum* gDNA as template
XGp108	StuI	PCR fragment amplified with primers KP62 and SZ10		L0 (BB02)	PCR fragment amplified using *S. lycopersicum* gDNA as template
XGp109	StuI	PCR fragment amplified with primers SZ02 and SZ09		L0 (BB02)	PCR fragment amplified using *S. lycopersicum* gDNA as template
XGp124	BpiI	$CYC\text{-}RE_{T90}$ annealed with oligos XG94 and XG95	$35Smin_{pro}$ (XGp92)		
KPp10	StuI	$SlNIN_{pro}$ (3kb) ampflied with primer KP50 & KP52		L0 (BB01)	PCR fragment was amplified using *S. lycopersicum* genomic DNA as template
KPp13	StuI	$LjNIN_{pro}$ (1 kb) amplified with primers KP91 & KP92		L0 (BB01)	PCR fragment was amplified using *L. japonicus* genomic DNA as template
KPp14	StuI	$LjNIN_{pro}$ (2 kb) amplified with primers KP93 & KP84		L0 (BB01)	PCR fragment was amplified using *L. japonicus* genomic DNA as template

Supplementary Table 4. Primers used in this study. Primer names beging with JL, KP, RA and XG (and SZ) were generated by Jane Lambert, Ksenia Vondenhoff, Rosa Elena Andrade and Xiaoyun Gong, respectively.

Primer name	Sequence 5' - 3'
JL65	ATGAAGACTTTACGGGTCTCTCACCATGGAATATGGTTCATTACTAGTGC
JL66	TTGAAGACTTGAACACAGGAAGGGCTAAAGA
JL67	ATGAAGACTTGTTCGAAAGAGGCACCGG
JL68	TTGAAGACTTAGATTCTTGATCCCCACCCTC
JL70	TTGAAGACTTCAGAGGTCTCTCCTTAGATGGGCTGCTATTGC
KP50	ATCGTCTCAGCGGCTGCTTTGGACTATATTTTCTTGAG
KP52	TACGTCTCTCAGAGCTGCTTCCTTCTTACCTC
KP58	ATCGTCTC ATACAACATGACACACGGATG
KP59	TACGTCTC TTGTAAAAATGGTATGTCCTAGACAAAC
KP60	TACGTCTC TGGAGACCAGCTACACTCAAATG
KP61	ATCGTCTC ACTCCAATCGAAGGTATATTAGTATGGATC
KP62	TACGTCTCTGCAGGTTGTGGGTTCCATTATTG
KP63	ATCGTCTCACTGCCGCTAATCAGGAAGCA
KP84	TACGTCTCACAGAGCTAGCTGATCCAATTAAGTACCTG
KP91	ATCGTCTCAGCGGGCTCCGTTTGGTCAACAGAC
KP92	ATCGTCTCTAAGCTAATTTGCAGCGACTTTTTCC
KP93	TACGTCTCAGCTTATATCGCAGCGACCAG
RA114	ACGAAGACAATACGTACGTCTCAGCTTATATCGCA
RA115	AAGAAGACAAGCTCAAATTTGTGTACCTAAAAATGC
RA116	TCAGAAGACAAGAGCCCACAAGAGGC
RA117	AAGAAGACAACAGATACGTCTCACAGAGCTAGCTG
RA118	AAGAAGACAAAGACGATACGGATCCACAAATTTGTGTACCTAAAAATGCAA
RA119	TCAGAAGACAAGTCTTAGATAGTCTGTGAGCCCACAAGAGGC
RA124	AAGAAGACAAACATGTCGCATGGAGCAAATTTGTGTACCTAAAAATGC
RA125	TCAGAAGACAAATGTGGCGTGCTCACAGGAGCCCACAAGAGGC
RA126	AAGAAGACAAACATGTCAGATGGACAAAATTTGTGTACCTAAAAATGC
RA127	TCAGAAGACAAATGTGTCGTACGGACATGAGCCCACAAGAGGC
RA128	AAGAAGACAACATATGTCACCGCAAGCAAATTTGTGTACCTAAAAATGC

Supplementary Table 4. Primers used in this study (continued).

Primer name	Sequence 5'-3'
RA129	TCAGAAGACAAATGCGCACAGGCACGAGCCCACAAGAGGC
RA130	AAGAAGACAAACATGTCGGATGCAGAAAATTTGTGTACCTAAAAATGC
RA131	TCAGAAGACAAATGTGGCACAACCACAAGAGCCCACAAGAGGC
RA87	ATGAAGACTTTACGGGTCTCATCTGTGCTTACACTTGTGGGTC
RA88	ATGAAGACTTCAGAGGTCTCAGGTGCTAGCTGATCCAATTAAGTACCT
SZ02	ATCGTCTCCCAGAATTATTACTGATAAAAAATCAAATGTTGC
SZ07	ATCGTCTCGCATGTGGCACGCAGAGAGCAGATTTTAAGGTTCACTACTCTATTTCT
SZ08	TACGTCTCACATGGCAATCGTACAAAATCCAGAGTGGTGAGGAT
SZ09	ATCGTCTCGATCGTCTTAGATAGTCTGTCAGATTTTAAGGTTCACTACTCTATTTCT
SZ10	TACGTCTCACGATACGGATCCACAAATCCAGAGTGGTGAGGAT
XG100	TAGGTCTCTGACATATCGTCTCACAGAAAGGGACTGACCACCCGGGAATTCACTGG
XG101	TAGGTCTCTTTACAGACAGACGACGTGCACTGACTTTAGTCACTTGA
XG102	ATGGTCTCAGTAACTGGAGATGAAGTTATATTTGGCGGGGTCCACT
XG105	TAGAAGACTAATGTAAGAAGTAGGAAATTTGTGTACCTAAAAATGCAA
XG106	TAGAAGACTAACATGTGGCAATCAGGCAGGAGCCCACAAGAGGCGAGA
XG107	TAGAAGACTAGACACATGTCAGAAGGAGAAAATTTGTGTACCTAAAAATGCAA
XG108	TAGAAGACTATGTCGTAAGGACAGGAGCCCACAAGAGGCGAGA
XG109	TAGAAGACTACTGATGTTGGTTGGATCAAATTTGTGTACCTAAAAATGCAA
XG110	TAGAAGACTATCAGGCGTAGCGTCAGGAGCCCACAAGAGGCGAGA
XG13	TAACGTCTCAGCGGTGCATCTAAACGGGTATTTGAG
XG14	TAACGTCTCAGCGGAGGATAGTGAGTTGTGCTCAC
XG15	TAACGTCTCAGCGGTCAGTTAATTGGTGTGTAATTCTCC
XG16	CTTCGTCTCAGCGGGATCCTTTGGCGCTAGCTCTTAC
XG17	TAACGTCTCAGCGGAATAGTGGCATATGAAAATGTTGG
XG18	TAACGTCTCAGCGGAATTGAAGATCCCAACAAGAT
XG19	TAACGTCTCAGCGGGCTAGTACTCCCTCCGTTCCAATA
XG20	TAACGTCTCAGCGGTGGTGGTAGGAAACTTGAAAAG
XG21	TAACGTCTCAGCGGGGAACGGATGGAGTATTACATAAA

Supplementary Table 4. Primers used in this study (continued).

Primer name	Sequence 5' - 3'
XG22	TAACGTCTCAGCGGTAAGTAATCCTATGAATGTAGGCTTT
XG23	TAGAAGACTACACCTGCTTACACTTGTGGGTCCTA
XG24	TAGAAGACTACAGATATCGTCTCACAGAGCTAGCTGATCCAATTAAGTAC
XG32	TACGTAACGTCTCAGCGGTGTACGATTGCCATGTGGCACGCAGAGAG
XG33	GGTGCTCTCTGCGTGCCACATGGCAATCGTACACCGCTGAGACGTTA
XG38	TAGAAGACTACACCCAAGACCCTTCCTCTATATAAGGA
XG39	TAGAAGACTACAGATATCGTCTCACAGATGTAATTGTAAATAGTAATTGTAATGTTGTTTG
XG40	TAACGTCTCAGCGGCAAGACCCTTCCTCTATATAAGGA
XG5	TAACGTCTCAGCGGCTTTACCCTCGTTATTTGGCG
XG52	TAACGTCTCAGCGGTACTTTAGATCCGTCATCAACGG
XG53	TATCGTCTCACAGATGGAGAGATTTTGGTGGTGTC
XG70	TAACGTCTCAGCGGTGCTTACACTTGTGGGTCCTA
XG71	ATCGTCTCAGCGGTGTACGATTGCCATGTGGCACGCAGAGAGTCTGTGAGACGAT
XG72	ATCGTCTCACAGACTCTCTGCGTGCCACATGGCAATCGTACACCGCTGAGACGAT
XG79	CTGAAGACTATACGTAACGTCTCAGCGGGCTCCATGCGACATGTGGCGTGCTCACAGCACCAAGTCTTCTT
XG8	TATCGTCTCACAGAAAGGGACTGACCACCCG
XG80	AAGAAGACTTGGTGCTGTGAGCACGCCACATGTCGCATGGAGCCCGCTGAGACGTTACGTATAGTCTTCAG
XG81	CTGAAGACTATACGTAACGTCTCAGCGGTGTCCATCTGACATGTGTCGTACGGACATCACCAAGTCTTCTT
XG82	AAGAAGACTTGGTGATGTCCGTACGACACATGTCAGATGGACACCGCTGAGACGTTACGTATAGTCTTCAG
XG83	CTGAAGACTATACGTAACGTCTCAGCGGGCTTGCGGTGACATATGGCGCACAGGCACCACCAAGTCTTCTT
XG84	AAGAAGACTTGGTGGTGCCTGTGCGCCATATGTCACCGCAAGCCCGCTGAGACGTTACGTATAGTCTTCAG
XG85	CTGAAGACTATACGTAACGTCTCAGCGGTCTGCATCCGACATGTGGCACAACCACAACACCAAGTCTTCTT
XG86	AAGAAGACTTGGTGTTGTGGTTGTGCCACATGTCGGATGCAGACCGCTGAGACGTTACGTATAGTCTTCAG
XG89	ATGGTCTCACACCATGCCAACTATTTTGCATAGGA
XG9	TAACGTCTCAGCGGTTAACACCATATATAATTGTCCCTTGC
XG90	TAGGTCTCTCCTTAGTGAGAGCGCTAAAGCCA
XG91	TAACGTCTCAGCGGGAGTGAGCACGCGCAAATA
XG92	TAACGTCTCAGCGGCAAGTGACTAAAGTCAGTGCACT

Supplementary Table 4. Primers used in this study (continued).

Primer name	Sequence 5' - 3'
XG93	TAACGTCTCAGCGGTATTTGGCGGGGTCCACT
XG94	TAGAAGACTATACGTAACGTCTCAGCGGTAATAAATAAATGCCGGCCTTTACCCTCGTCACCAAGTCTTCTA
XG95	TAGAAGACTTGGTGACGAGGGTAAAGGCCGGCATTTATTTATTACCGCTGAGACGTTACGTATAGTCTTCTA
XG96	TATCGTCTCAGTTCACGTTGTGAAACAAGACTA
XG97	ATGGTCTCAGCGGTAACGTCTCAGAACACAAACTTACCAACTTACCTT
XG98	TAGGTCTCTAATAGTGCACTGACTTTAGTCACTTGA
XG99	ATGGTCTCATATTTGGCGGGGTCCACT
JL69	ATGAAGACTTATCTTCTTATACCTTTGGAAGCCGCCGCTCtTCTTCTGGTGGAAGAAAGTCAGGCGAGAAAAGACGAACCAAGGCAGAAAAGACTATCAG

Supplementary Table 5. Transcription factors predicted to bind to *EPRE*$_{CBP1}$

Query_ID	Target_ID	Transcription factor class	Optimal_of	p-value	E-value	q-value	Overlap	Predicted seuqences in RHE	Target_consensus	Orientation	JASPAR link
EPRE_CBP1	MA0952.1 (ATHB-51)	Homeo domain factors	-29	0.00051478	0.251727	0.381388	8	CAATAATT	CAATAATT	-	https://jaspar2018.genereg.net/matrix/MA0952.1/
EPRE_CBP1	MA0953.1 (ATHB-6)	Homeo domain factors	-28	0.000779934	0.381388	0.381388	9	GCAATAATT	GCAATAATT	+	https://jaspar2018.genereg.net/matrix/MA0953.1/
EPRE_CBP1	MA1168.1 (AT3G04030)	other (G2like)	-30	0.00270646	1.32346	0.882306	14	AATAATTTTCTCTT	AAGAAGATTCTTTT	-	https://jaspar2018.genereg.net/matrix/MA1168.1/
EPRE_CBP1	MA1163.1 (AT5G45580)	other (G2like)	-30	0.0059978	2.93293	0.988309	14	AATAATTTTCTCTT	AAGAATATTCTTTT	-	https://jaspar2018.genereg.net/matrix/MA1163.1/
EPRE_CBP1	MA0951.1 (ATHB-16)	Homeo domain factors	-29	0.00606324	2.96493	0.988309	8	CAATAATT	TAATAATT	+	https://jaspar2018.genereg.net/matrix/MA0951.1/
EPRE_CBP1	MA1009.1 (ARF3)	B3 domain	-18	0.00606324	2.96493	0.988309	8	TGTTGGAA	TGTCGGAA	+	https://jaspar2018.genereg.net/matrix/MA1009.1/
EPRE_CBP1	MA0950.1 (ATHB-12)	Homeo domain factors	-29	0.0106738	5.21949	1	8	CAATAATT	CAATCATT	-	https://jaspar2018.genereg.net/matrix/MA0950.1/
EPRE_CBP1	MA1385.1 (AT2G40260)	other (G2like)	-29	0.0106829	5.22393	1	15	CAATAATTTTCTCTT	TTAAAATATTCTTTT	+	https://jaspar2018.genereg.net/matrix/MA1385.1/
EPRE_CBP1	MA1280.1 (OBP4)	C2H2 zinc finger factors	-13	0.0126487	6.1852	1	14	GAAAATGTTGGAAA	CAAAAAGTTAAAAA	+	https://jaspar2018.genereg.net/matrix/MA1280.1/
EPRE_CBP1	MA1278.1 (OBP1)	C2H2 zinc finger factors	-14	0.0127921	6.25534	1	21	AAAATGTTGGAAATGCAATAA	AAAAAGTAAAAAAAAAAAAAA	-	https://jaspar2018.genereg.net/matrix/MA1278.1/
EPRE_CBP1	MA0008.2 (HAT5)	Homeo domain factors	-28	0.0153804	7.52104	1	12	GCAATAATTTTC	GCAATAATTGAA	-	https://jaspar2018.genereg.net/matrix/MA0008.2/
EPRE_CBP1	MA1026.2 (ATHB15)	Homeo domain factors	-24	0.0157594	7.70634	1	15	AAATGCAATAATTTT	AAAAGTAATGATTAC	+	https://jaspar2018.genereg.net/matrix/MA0008.2/
EPRE_CBP1	MA1277.1 (Adof1)	C2H2 zinc finger factors	-8	0.0193449	9.45966	1	21	CATATGAAAATGTTGGAAATG	AAAAAGAAAAAGTAAAAAAAA	+	https://jaspar2018.genereg.net/matrix/MA1277.1/