Parasponia;

a missing piece of the evolutionary puzzle of nitrogen–fixing nodule symbiosis

Fengjiao Bu

Parasponia; a missing piece of the evolutionary puzzle of nitrogen-fixing nodule symbiosis

Fengjiao Bu

Thesis committee

Promotor

Prof. Dr T. Bisseling
Professor of Molecular Biology
Wageningen University & Research

Co-promotor

Dr R. Geurts
Associate professor, Laboratory of Molecular Biology
Wageningen University & Research

Other members

Prof. Dr M. G. M. Aarts, Wageningen University & Research
Prof. Dr K. Pawlowski, Stockholm University, Sweden
Dr. J van Heerwaarden, Wageningen University & Research
Dr. A. Niebel, Director de recherche CNRS, Auzeville, France

This research was conducted under the auspices of the Graduate School
Experimental Plant Sciences

Parasponia; a missing piece of the evolutionary puzzle of nitrogen-fixing nodule symbiosis

Fengjiao Bu

Thesis

submitted in fulfilment of the requirements for the degree of doctor

at Wageningen University

by the authority of the Rector Magnificus

Prof. Dr A. P. J. Mol,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Tuesday March 31 2020

at 1:30 p.m. in the Aula.

Fengjiao Bu
Parasponia; a missing piece of the evolutionary puzzle of nitrogen-fixing nodule symbiosis

PhD thesis, Wageningen University, Wageningen, NL (2020)
With references, with summary in English
ISBN 978-94-6395-327-6
DOI: https://doi.org/10.18174/516173

Dedicated to my beloved parents

献给我敬爱的爸妈

Table of content

Chapter 1

General Introduction

Introduction

Nitrogen (N) is a primary macronutrient for plants. It is a critical element in amino acids, nucleotides and other compounds. Plants can only take up fixed-forms of nitrogen, such as ammonium (NH_4^+), nitrate (NO_3^-) and urea, which usually are present in soils, albeit at different levels. However, most nitrogen on earth is present in the atmosphere as the kinetically stable form of dinitrogen (N_2). Limitations of bioavailable nitrogen and its importance on crop production invoke a massive usage of chemical nitrogen fertilizer. But breaking the triple bond of the di-nitrogen (N_2) during nitrogen fertilizer production requires high pressure and temperature thus is environmentally and economically costly (Worrell *et al.*, 2000; Sutton *et al.*, 2011). Furthermore, inefficient use of chemical nitrogen fertilizer causes contamination of soil and ground-water (European Fertilizer Manufacturers Association, 2000).

In contrast to plants, some prokaryotes contain genes encoding a protein complex called nitrogenase, which can convert di-nitrogen gas (N_2) into ammonia (NH_3). This process is called biological nitrogen fixation. Prokaryotes capable of catalyzing this reduction of N_2 are called diazotrophs. Diazotrophs form a paraphyletic group of prokaryotic species and display different lifestyles ranging from free-living marine and soil microbes to others which live in a symbiotic relationship with fungi or plants (Santi *et al.*, 2013; de Bruijn, 2015; Provorov & Onishchuk, 2018; Zhou *et al.*, 2019). Free-living soil diazotrophs like *Azotobacter vinelandii* (Pseudomonadaceae, Pseudomonadales) only fix relative small amounts of nitrogen but are essential for sustainable ecosystems and often used as plant growth-promoting rhizobacteria.

Symbiotic associations can occur at different levels of intimacy (Coba de la Peña *et al.*, 2018). Diazotrophic cyanobacteria can extracellularly associate with a wide range of plants from nonvascular plants such as mosses, liverworts and hornworts (Bryophyta) to vascular plants from the genus of *Azolla* (Pteridophytes) and Cycadaceae (Gymnosperms) and intracellularly associate with *Gunnera* (Gunneraceae, order Gunnerales, Angiosperms) (Adams *et al.*, 2006; Bergman *et al.*, 2007). In the latter case, cyanobacteria of the genus *Nostoc* can establish an endosymbiosis in stem glands of *Gunnera* species, which evolved more than 100 million years ago (Johansson & Bergman, 1992; Warshan *et al.*, 2018). The diazotrophic *Frankia* and rhizobia can also establish intracellular endosymbioses with specific host plants. In these cases, novel lateral organs are formed known as nodules. Nodule cells can accommodate the diazotrophic microsymbiont intracellularly. This nodule endosymbiosis occurs in plant species belonging to four related orders from the Fabid clade: the Fagales, Fabales, Cucurbitales and

Rosales, which diverged more than 100 million years ago (Soltis *et al.*, 1995; Wang *et al.*, 2009). Together these orders are known as the nitrogen-fixing clade, even though many lineages within this clade are non-nodulating plants (Soltis *et al.*, 1995; Wang *et al.*, 2009; Doyle, 2011; Werner *et al.*, 2014).

Legumes (Fabaceae, order Fabales) represent the most prominent family of nodulating species as it includes important crops such as soybean (*Glycine max*), common bean (*Phaseolus vulgaris*), chickpea (*Cicer arietinum*), lens (*Lens culinaris*) and pea (*Pisum sativum*). Due to the high nitrogen-fixing efficiency of these legume crops (e.g. up to 100-300 kg/hectare of fixed nitrogen annually for alfalfa, red clover, pea, soybean, cowpea, and vetch) (Wani *et al.*, 1995), they do not require chemical nitrogen fertilizer. This phenomenal nitrogen-fixing capacity has raised two long-standing objectives to 1), uncover the evolution and molecular mechanisms leading to the formation of nitrogen-fixing nodules, and 2) extend the nitrogen-fixing nodule symbiosis to other crops outside the nitrogen-fixing clade; e.g. rice (*Oryza sativa*), wheat (*Triticum aestivum*) and maize (*Zea mays*) (Markmann & Parniske, 2009; Charpentier & Oldroyd, 2010; Oldroyd & Dixon, 2014; Stokstad, 2016). In this chapter, I will summarize knowledge at the start of this thesis and address critical questions related to evolution and core genetic basis of nodulation, and report new strategies applied in this thesis.

Taxonomic relations of nodulating plants and diazotrophic partners

The nitrogen-fixing clade contains 28 families, of which only 10 families represent species that can form nitrogen-fixing root nodules with either diazotrophic rhizobia or *Frankia* (**Fig. 1**). Thus symbiotic plant species are largely scattered by non-nodulating species (Werner *et al.*, 2014). Among the nodulating lineages, legumes are one of the world's largest families, representing close to 20,000 species belonging to 750 genera of which many can establish nitrogen-fixing symbiosis with rhizobia. Unlike legumes, the other 9 families are much smaller. About 220 species in eight families can form nodules with filamentous *Frankia* spp. bacteria, including the Rhamnaceae, Elaeagnaceae and Rosaceae (Rosales); Casuarinaceae, Betulaceae and Myricaceae (Fagales); Datiscaceae and Coriariaceae (Cucurbitales) (Roy & Bousquet, 1996; Schwencke & Carú, 2001; Svistoonoff *et al.*, 2014; Li *et al.*, 2015). These nodulating species are collectively named actinorhizal plants. Besides legumes (Fabaceae, Fabales), *Parasponia* (Cannabaceae, Rosales) comprising 5 spec*ies -P. andersonii, P. melastomatifolia, P. parviflora, P. rigida and P. rugosa-* is the only non-legume lineage that can form nitrogen-fixing root nodules with diazotrophic rhizobia.

Phylogenetically, the *Frankia* genus (Frankiaceae, Frankiales) can be divided into four taxonomic clusters, three of which are symbiotic and show a certain level of specificity to their actinorhizal host plants (**Fig. 1**) (Pawlowski & Demchenko, 2012). Cluster II *Frankia* spp. are more basal compared to other clusters and these species can interact with plants from the Rosaceae family and the genus *Ceanothus* from Rhamnaceae family (Rosales), as well as Datiscaceae and Corariaceae families (Cucurbitales). Cluster III *Frankia* spp. also can interact with plant species from Rhamnaceae and Elaeagnaceae family (Rosales), and also the genus *Gymnostoma* from Casuarinaceae family and the genus Myrica from Myricaceae family (Fagales). Unlike clade II and III, cluster I *Frankia* species only interact with plant species from Fagales.

Diazotrophic rhizobia represent over 100 different bacterial species divided over 14 genera representing eight largely unrelated families of **α**-proteobacteria (Rhizobiaceae, Phyllobacteriaceae, Methylobacteriaceae, Brucellaceae, Hyphomicrobiaceae and Bradyrhizobiaceae, belong to order Rhizobiales), **β**-proteobacteria (Burkholderiaceae, order Burkholderiales) and **γ**-proteobacteria (Pseudomonaceae, Pseudomonadales) (Remigi et al. 2016; Limpens et al. 2015; Berrada and Benbrahim, 2014). These microbes have obtained the nitrogen fixation (*fix* & *nif*) gene clusters most probably by horizontal gene transfer (Raymond *et al.*, 2004; Dos Santos *et al.*, 2012; Poole *et al.*, 2018).

Although legumes and *Parasponia* species can establish nitrogen-fixing nodulation with rhizobia, the way how rhizobia infect and colonize nodule cells is different (**Fig. 1**). Rhizobia infect *Parasponia* nodule cells via intercellular crack entry, while for most legumes, such as the model species *Medicago truncatula* and *Lotus japonicus*, infection occurs intracellularly via curled root hairs that capture a single bacterium. From the micro-colony in the root hair curl, the plant cell wall is degraded, followed by invagination of the plasma membrane and the formation of a tubular infection thread. This infection thread will grow towards the newly formed nodule primordium.

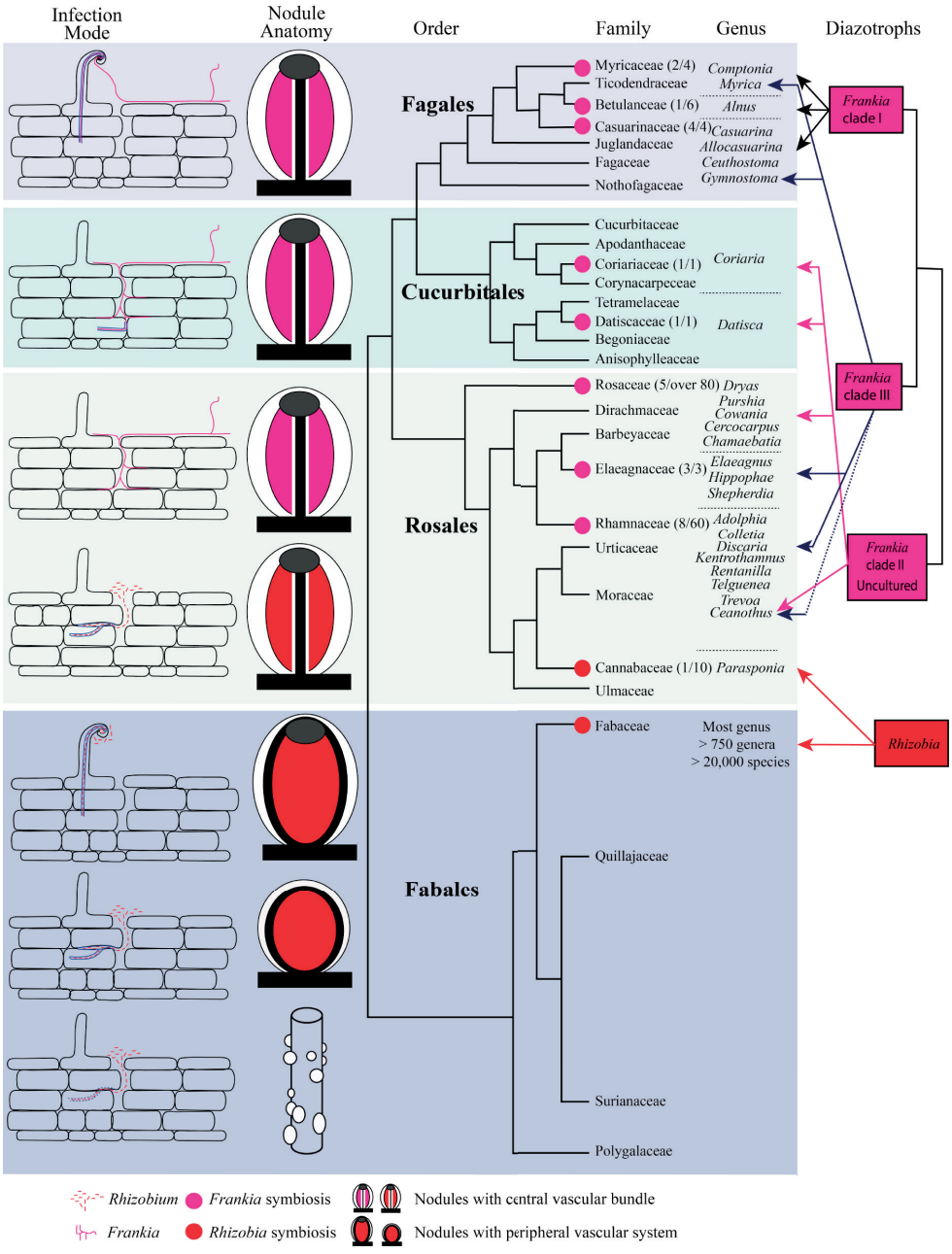

Infection Mode
Nodule Anatomy
Order
Family
Genus
Diazotrophs
Fagales
Myricaceae (2/4)
Comptonia
Myrica
Ticodendraceae
Betulanceae (1/6)
Alnus
Casuarinaceae (4/4)
Casuarina
Allocasuarina
Juglandaceae
Fagaceae
Ceuthostoma
Nothofagaceae
Gymnostoma
Frankia clade I
Cucurbitales
Cucurbitaceae
Apodanthaceae
Coriariaceae (1/1)
Coriaria
Corynacarpeceae
Tetramelaceae
Datiscaceae (1/1)
Datisca
Begoniaceae
Anisophylleaceae
Frankia clade III
Rosales
Rosaceae (5/over 80)
Dryas
Purshia
Dirachmaceae
Cowania
Barbeyaceae
Cercocarpus
Chamaebatia
Elaeagnaceae (3/3)
Elaeagnus
Hippophae
Shepherdia
Rhamnaceae (8/60)
Adolphia
Colletia
Urticaceae
Discaria
Kentrothamnus
Rentanilla
Moraceae
Telguenea
Trevoa
Ceanothus
Frankia clade II Uncultured
Cannabaceae (1/10)
Parasponia
Ulmaceae
Fabales
Fabaceae
Most genus
> 750 genera
> 20,000 species
Rhizobia
Quillajaceae
Surianaceae
Polygalaceae
Rhizobium
Frankia symbiosis
Frankia
Rhizobia symbiosis
Nodules with central vascular bundle
Nodules with peripheral vascular system

Fig. 1 Phylogenetic distribution and diversity of nitrogen-fixing root nodule symbiosis. Phylogenetic distribution of plant family (Family) able to establish nitrogen-fixing root nodule symbiosis with the NFC (Order). Circles represent actinorhizal (pink) or rhizobial (red) nodulation. Nodulation genus for each nodulating family is listed (Genus). Arrows directly pointed to certain genus show exceptional associations, arrows did not point to certain genus mean broad association between plant family and Frankia clade/rhizobium (Diazotrophs). Dashed arrows indicate that members of this cluster have been isolated from or detected in an effective or ineffective nodule of a member of the plant group at least once. Numbers in parenthesis indicate the total number of genera and the number of genera known to be nodulated. Schematic views of infection mechanisms (Infection mode) and the anatomy of nodules (Nodule anatomy) are shown for each group. When present, infection threads are shown in blue. Nodules formed on legumes have a peripheral vascular system and a large central zone of infected cells, while nodules formed on *Parasponia* and actinorhizal plants with a central vascular bundle and infected cells in the periphery. Nodule apical meristems are in grey and tissues colonised by Frankia or rhizobia are in pink and red respectively. Note in some Aeschynomene species, nodules are formed on stems instead of roots (shown in white). Figure is modified based on Svistoonoff et al. (2014) with permission from the publisher. Host specificity in actinorhizal symbioses is based on Pawlowski and Demchenko (2012).

Once infection thread reaches primordium cells, rhizobium will be released into host cells and developed into transient organelle-like structures called symbiosomes. In contrast, inside cells of *Parasponia* nodules, rhizobium is hosted in so-called fixation threads without being released. Actinorhizal nodule cells also host *Frankia* spp. in the form of fixation threads. Fixation threads are considered more basal when compared to symbiosomes. It is hypothesized that fixation threads mimic the structure of arbuscules formed by endomycorrhizal (AM) fungi.

A second divergence between legumes and other nodulating plants is the ontogeny of the nodule. Legume nodules show a 'stem-like' ontogeny with a peripheral vascular system and a large central zone of infected cells, while nodules formed on *Parasponia* and actinorhizal plants have a 'lateral root-like' ontogeny with a central vascular bundle and infected cells in the periphery.

The above-mentioned divergence in infection mode, nodule ontology, dispersed phylogenetic distribution of nodulating lineages, as well as the involvement of two different classes of diazotrophic microsymbionts invoked hypotheses regarding the evolution of nitrogen-fixing symbiosis (Soltis *et al.*, 1995; Swensen, 1996). Prior to my thesis the most widely accepted hypothesis was that nodulation originated independently multiple times. This convergent evolution was predicted to be preceded by a 'predisposition' event that has occurred at the root of the nitrogen-fixing clade, which lifted plant species in a precursor state for nodulation (Swensen, 1996; Doyle, 1998, 2011, 2016; Werner *et al.*, 2014; Li *et al.*, 2015; Martin *et al.*, 2017). A second hypothesis that proposed was a single origin of nodulation in the nitrogen-

fixation clade followed by multiple parallel losses (Soltis *et al.*, 1995; Swensen, 1996). This second hypothesis was largely dismissed because it is not parsimonious, nor it could explain why plants would lose a favourable trait. No matter which hypothesis is correct, it remains elusive which genetic adaptations were essential to allow the evolution of the nitrogen-fixing root nodule symbiosis trait, and to what extent a conserved genetic network may exist and is shared by all nodulating plants.

Symbiotic LCO signalling

During the past three decades, studies on the two legume models *M. truncatula* and *L. japonicus* provided in-depth knowledge on the molecular and genetic signalling pathways that control root nodule formation and bacterial infection. Upon sensing of root secreted (iso-)flavonoids, rhizobium produces symbiotic signalling molecules named Nod(ulation) factors. Nod factors of legume micro-symbionts were first characterized in the early 1990s and showed to be lipo-chitooligosaccharides (LCOs) composed of three to five β-1,4-linked N-acetyl-D-glucosamine with an N-acyl group at the non-reducing terminal residue (Lerouge *et al.*, 1990; Dénarié *et al.*, 1996). LCOs produced by different rhizobia vary in length of chitin backbone, length and saturation of the acyl group, and the presence of one or more of the following substitutions at both ends of the chitin backbone: methylation, acetylation, arabinosylation, carbamoylation, fucosylation, glycerolation, mannosylation, and/or sulfation. Expression of rhizobial LCO biosynthesis genes is controlled by LysR-family transcriptional regulator NodD (and in some species also homologous proteins named NolR and NrcR commit a similar function) upon direct binding of (iso-)flavonoids (del Cerro *et al.*, 2015; Del Cerro *et al.*, 2016; Peters *et al.*, 1986; Redmond *et al.*, 1986; Mulligan & Long, 1989). Core genes in LCO biosynthesis are the *nodC* encoded N-acetyl glucosamine transferase, the *nodB* encoded Deacetylase, and the *nodA* encoded Acyltransferase. The complexity of LCOs synthesized by certain rhizobium determines - at least in part - whether it can nodulate only a few or a broad range of host plants. For example, *Sinorhizobium fredii* NGR234 and *Rhizobium tropici* CIAT899 can interact with a broad range of hosts -at least partially- due to that a large set of structurally different LCO molecules they produce (Price *et al.*, 1992, 1996; Folch-Mallol *et al.*, 1996; Pueppke & Broughton, 1999; Morón *et al.*, 2005; Estévez *et al.*, 2009). Rhizobia nodulating the non-legume *Parasponia* are also dependent on LCO-induced signalling pathway (Marvel, 1987; Op den Camp *et al.*, 2011a). And also theses plants have a relatively broad host range (Op den Camp et al. 2012).

In a few legumes, like some Aeschynomene species, and LCOs receptor knockout mutants of soybean, LCO perception can be bypassed by rhizobium secreted effector-like molecules (Okazaki *et al.*, 2013, 2016; Teulet *et al.*, 2019). In line with this is the finding that nodulation of actinorhizal plants associated with *Frankia* species of cluster I or cluster III is also independent of LCO signalling (Normand *et al.*, 2007). Moreover, research shows that the *Frankia* signal molecules are chemically distinct from rhizobium LCOs since they are not chitinase degradable, like LCOs. Yet they are able to trigger expression of symbiosis genes in *Casuarina glauca* (Casuarinaceae, Fagales) that are homologous to genes induced in legumes by rhizobium LCOs (Chabaud *et al.*, 2016). In contrast to cluster I and III *Frankia*, *Frankia* strains belonging to the basal cluster II contain homologs of rhizobium LCO biosynthesis genes *nodABC*, and some strains even contain homologs of rhizobium sulfotransferase gene *nodH* (Normand et al., 1996; Sen et al., 2014; Persson et al., 2015; Van Nguyen et al., 2016). Candidatus *Frankia* strain Dg1 (cluster II) possesses two *nodABC* operons and these genes are expressed in *Datisca glomerata* nodules based on quantitative RT-PCR (Persson *et al.*, 2015). Also, it was shown that a *nodC* homologous gene from this *Frankia* strain can functionally complement a *Rhizobium leguminosarum nodC* mutant (Persson *et al.*, 2015). This indicates that the *Frankia* Dg1 *nodC* homologous gene can function as an N-acetyl glucosamine transferase. Despite this, so far no LCO-like molecules from *Frankia* cluster II strains have been structurally characterized. This, mainly because of the technical difficulties of culturing such strains (Persson *et al.*, 2011).

In the course of evolution, diazotrophic bacteria most probably acquire LCO biosynthesis genes from AM fungi. AM fungi can also secrete diffusible signal molecules, called 'Myc factors'. Myc factors are composed of a mixture of sulphate and non-sulphate LCOs, which are structurally reminiscent of rhizobium secreted LCOs (Maillet *et al.*, 2011). AM symbiosis can occur with 71% of vascular plants and is considered to have co-evolved with land plants (Wang & Qiu, 2006; Delaux *et al.*, 2014). AM fungi colonize the roots via fungal hyphae and form specialized nutrient exchange structures in root cortical cells called arbuscules. Unlike AM symbiosis, root nodule symbiosis involves two parallel biological processes that are initiated simultaneously: bacterial infection and nodule organogenesis. In most legumes, both processes are dependent on the LCO-induced signalling pathway that is shared with AM fungi. This pathway is named common symbiosis signalling pathway (CSSP) as AM symbioses are widely spread among land plants. As all plants that can establish an AM symbiosis possess this CSSP, the question that remains is to what extent this pathway has been adapted to allow nodulation?

Common symbiosis signalling pathway

In legumes, rhizobium LCOs are perceived by a LysM receptor-like kinase heterodimer named *LjNFR1/LjNFR5* in *L. japonicus* and *MtLYK3/MtNFP* in *M. truncatula* (Limpens *et al.*, 2003; Madsen *et al.*, 2003; Radutoiu *et al.*, 2003; Arrighi *et al.*, 2006; Moling *et al.*, 2014). Downstream of LCO perception, legumes co-opted the CSSP from arbuscular mycorrhizal symbiosis. The common symbiosis signalling pathway starts from the LRR-type receptor LjSYMRK/MtDMI2, which interacts with the kinase domain of LjNFR5/MtNFP to form a receptor heteromeric complex (Endre *et al.*, 2002; Stracke *et al.*, 2002). LjSYMRK/MtDMI2 also interacts with MtHMGR1, an enzyme involved in mevalonate biosynthesis (Kevei *et al.*, 2007; Venkateshwaran *et al.*, 2015). Activation of MtHMGR1 triggers nuclear calcium spiking, a response that is dependent on a nuclear-associated machinery including the nuclear pore complex subunits LjNUP85, LjNUP133 and LjNENA, the nuclear-localized calcium-dependent adenosine triphosphatase MtMCA8, the potassium channels LjCASTOR/MtDMI1 and LjPOLLUX, and the calcium channels MtCNGC15a,b,c (Kanamori *et al.*, 2006; Charpentier *et al.*, 2008, 2016; Chen *et al.*, 2009; Capoen *et al.*, 2011). Decoding of calcium spiking by the calcium- and calmodulin-dependent kinase LjCCaMK/MtDMI3 will phosphorylate the transcription factor LjCYCLOPS/MtIPD3 upon direct interaction (Yano *et al.*, 2008; Miller *et al.*, 2013). LjCYCLOPS/MtIPD3 represents the last shared component between rhizobium and AM fungi in the CSSP. Downstream of the CSSP, the signalling subsequently diverges and leads to very different transcriptional reprogramming for both symbioses. For rhizobium, it involves the activation of mitotic processes and rewiring of phytohormone pathways. This involves genes such as the transcription factors *NIN*, *NF-YA1* and *ERN1/ERN2*, which are transcriptionally activated upon rhizobium induced LCO signalling. These genes either do not have, or only play a very minor role in AM symbiosis, and their expression is also not induced upon AM induced signalling.

Several lines of evidence suggest that the CSSP is also recruited to function in actinorhizal and *Parasponia* nodule formation. Homologous CSSP genes are expressed in young nodules of the actinorhizal plant species *Datisca glomerata* and *Casuarina glauca* (Hocher et al. 2011; Tromas et al. 2012; Svistoonoff et al. 2014; Gherbi et al. 2008; Granqvist et al. 2015; Svistoonoff et al. 2013; Markmann et al. 2008), whereas RNA interference (RNAi) mediated knockdown of the LRR-type receptor SYMRK in these species abolishes nodule formation (Gherbi *et al.*, 2008; Markmann *et al.*, 2008; Fabre *et al.*, 2015). Also, it was found that calcium spiking can be induced in *P. andersonii* by rhizobium LCOs and in two actinorhizal plants

Alnus glutinosa and *C. glauca* upon application of *Frankia* exudates (Granqvist *et al.*, 2015; Chabaud *et al.*, 2016). In consistence with that, knockdown of the calcium spiking signalling decoding gene *CCaMK* in *C. glauca* reduces nodule formation (Svistoonoff *et al.*, 2013), whereas ectopic expression of an autoactive *CCaMK* allele in *C. glauca*, *D. trinervis* and *P. andersonii* induces spontaneous nodules (Svistoonoff *et al.*, 2013; Op den Camp et al. 2011), similar as observed in *L. japonicus* and *M. truncatula* (Gleason *et al.*, 2006; Tirichine *et al.*, 2006; Ovchinnikova *et al.*, 2011; Hocher *et al.*, 2011; Ried *et al.*, 2014; Singh *et al.*, 2014; Saha *et al.*, 2016). This underlines that not only in legumes but also in non-legumes the CSSP from AM fungi symbiosis is recruited for nitrogen-fixing nodule symbiosis.

The question remains how a nodulating plant can discriminate between both micro-symbiotic partners. The calcium spiking profiles triggered by both rhizobia and arbuscular mycorrhiza are remarkably similar (Sieberer *et al.*, 2012). Though rhizobium and AM fungi induced signalling shows different requirements regarding the binding of CALMODULIN (CaM) to CCaMK. Whereas this is essential for rhizobium-induced nodulation, it is not in the case of AM symbiosis (Shimoda *et al.*, 2012).

Evolution of nitrogen-fixing symbiosis

As stated above, it is generally anticipated that nodulation evolved several times in parallel, and was preceded by a predisposition event in the root of the nitrogen-fixing clade (Li et al. 2015; Swensen 1996; Werner et al. 2014; Martin et al. 2017; Doyle 1998; Doyle 2011; Doyle 2016). However, molecular support for this hypothesis is lacking.

Trans-complementation studies in legume nodulation mutant using rice (*Oryza sativa*) putative orthologs of *OsSYMRK, OsPOLLUX, OsCASTOR, OsCCaMK, OsCYCLOPS, OsNSP1* and *OsNSP2* showed that the encoded proteins possess -at least in part- the functionality to support rhizobium induced nodule formation (Godfroy *et al.*, 2006; Banba *et al.*, 2008; Chen *et al.*, 2009). Interestingly, *OsCCaMK* and *OsPOLLUX* can only restore nodulation, not infection, suggesting that there might be specific adaptations in these proteins in legumes. It is likely that such specific adaptations also resides in proteins that are not part of the CSSP. In legumes, genes in the orthogroup of *LjNFR1/MtLYK3* experienced several duplications, which has driven functional specificity of this rhizobium LCO receptor (De Mita *et al.*, 2014; Bozsoki *et al.*, 2017). For example in rice, the orthologous gene *OsCERK1* commits a dual function. It is essential for chitin-triggered innate immune responses as well as for the establishment of the AM symbiosis. Domain swapping experiment shows that the kinase domain of rice OsCERK1

can also functionally replace the kinase domain of LjNFR1 (Miyata et al. 2014). Therefore, specific adaptations in the LjNFR1/MtLYK3 receptor are most probably not sufficient to explain the evolution of nitrogen-fixing root nodule symbiosis.

In most legumes, rhizobium intracellular infection happens in epidermal root hairs by the formation of infection threads. Nodule organogenesis is initiated from the inner root layers which are physically not linked to the epidermis. In order to coordinate these two parallel biological processes, there must exist a tightly regulated 'signal dialogue'. *NODULE INCEPTION* (*NIN*), a key transcription factor, which among the first genes transcriptionally induced downstream of the CSSP, is likely to act in this signal dialogue (Schauser *et al.*, 1999; Marsh, 2007; Vernié *et al.*, 2015). In *M. truncatula*, *NIN* is first induced in the epidermis upon rhizobium infection (Yoro *et al.*, 2014). It has been demonstrated that *NIN* induction in the cortical is essential for nodule organogenesis and this induction is dependent on cytokinin (Yoro *et al.*, 2014) and requires a *cis*-regulatory element within the *NIN* promoter (Liu *et al.*, 2019). In legumes, NIN-mediated nodule organogenesis and rhizobium intracellular infection are - at least partially- dependent on *NF-YA1*, a member of *NUCLEAR FACTOR Y* gene family (Combier *et al.*, 2006a; Soyano *et al.*, 2013a; Rípodas *et al.*, 2014). Knockdown or knockout *NF-YA1* blocks nodule development in early stages in *L. japonicus*, and disturbs the formation and functional maintenance of the nodule apical meristem in *M. truncatula*, resulting in nodules of variable size, but all smaller than wild type (Combier et al. 2006; Laporte et al. 2014; Laloum et al. 2014; Soyano et al. 2013; Hossain et al. 2016; Xiao et al. 2014). NF-YA1 also functions in regulating rhizobium intracellular infection, which has been shown in multiple legume species, and this function is probably redundant with other genes from the *NF-YA* family (Soyano *et al.*, 2013a; Laporte *et al.*, 2014; Battaglia *et al.*, 2014; Laloum *et al.*, 2014b; Xiao *et al.*, 2014; Hossain *et al.*, 2016a; Rípodas *et al.*, 2019). As for non-legumes, *NIN* and *NF-YA1* are highly upregulated in nodules of *C. glauca* and *Alnus glutinosa*, suggesting *NIN* and *NF-YA1* might be essential also for actinorhizal nodule organogenesis (Diédhiou *et al.*, 2014). Indeed, it has been demonstrated that RNA interference (RNAi)-mediated knockdown of *NIN* in *C. glauca* reduces nodule formation (Clavijo *et al.*, 2015). However, it remains unclear whether non-legumes have recruited also *NF-YA1* to function in nodule formation.

Thesis outline

The aim of this thesis is to gain insights into the evolution and to identify the core genetic basis of the nitrogen-fixing nodulation trait. To do so, I adopted *Parasponia* as a comparative system

and applied different strategies, including comparative transcriptomics, comparative structural genomics and reverse genetics approaches.

Parasponia is the only non-legume plant which can establish nitrogen-fixing endosymbiosis with rhizobium, and it is the only nodulating plant within the Cannabaceae. Based on its primitive nodulation trait it is hypothesized that *Parasponia* has deployed rhizobium nitrogen-fixing symbiosis more recent compared to legumes (Op den Camp *et al.*, 2011b, 2012; Geurts *et al.*, 2012; Werner *et al.*, 2014). The establishment of a highly efficient stable transformation and CRISPR-CAS9 mediated platform for *P. andersonii* will be helpful in testing the role of identified target genes (van Zeijl *et al.*, 2018). Therefore, *Parasponia* forms a unique complementary system to legumes to understand the evolution of nodulation.

In **Chapter 2**, we aimed to identify the genetic basis that confers the nodulation capacity of *Parasponia.* To do so, *Parasponia* and its closely related non-nodulating sister species *Trema* is used in a comparative analysis. In this chapter, we first conducted forward genetics by creating intergeneric hybrids between *Parasponia* and *Trema,* aiming to obtain a segregating population that would have allowed us a QTL analysis. With an infertile F1 hybrid resulting from a cross between a diploid *Parasponia andersonii* and a tetraploid *Trema tomentosa*, we found that the nodule formation and rhizobium intracellular infection can be genetically separated. Further, comparative transcriptomics between *P. andersonii* and *M. truncatula* nodules reveal that these two remotely related plant species share a large set of symbiosis genes. These results indicate that nodulation in *Parasponia* shares the same genetic basis as legumesSubsequent comparative genomics showed that *Trema* and other plant species within the Rosales order have lost orthologs of key symbiosis genes, among which is *NFP* and *NIN*. This strongly suggests parallel loss of the nitrogen-fixing nodulation trait in these species.

P. andersonii x *T. tomentosa* F1 hybrid plants contain a diploid *T. tomentosa* with a haploid genome of *P. andersonii* complement introduced. Despite efficient nodulation, rhizobium is unable to establish intracellular infections within hybrid nodules. As the nodulation trait of the *P. andersonii* x *T. tomentosa* hybrid may reflect a future engineering result, hybrid plants represent a valuable experimental tool to study the mechanism controlling intracellular rhizobium infection. In **Chapter 3**, our focus is to understand the nodulation features of hybrid plants. We showed that the block in intracellular infection within hybrid nodules is consistent for all tested rhizobial strains. This block of intracellular infection cannot be overcome by increased LCO biosynthesis nor by mutating the type III or IV secretion systems of the

nodulating strains. Besides, we also identified that the host range of hybrid plants is more narrow when compared to *P. andersonii.* Block of intracellular infection within hybrid nodules is likely specific to nitrogen-fixing root nodule symbiosis since hybrid plants can establish arbuscular mycorrhization effectively. Taken together, this indicates a yet unknown mechanism leading to an impaired host range and block of intracellular infection of hybrid plants.

In **Chapter 4**, we focus on the key symbiotic transcriptional module NIN - NF-YA1 that act downstream of the CSSP. By reverse genetics, we showed that *NIN* and *NF-YA1* are essential for nitrogen-fixing nodule symbiosis in *P. andersonii*. *Parasponia* NIN is essential for nodule initiation, whereas *NF-YA1* is essential for intracellular infection. This provides further evidence that nodulation in legumes and *Parasponia* is founded on a conserved genetic network.

In **Chapter 5**, I investigated the specificity of *P. andersonii* towards LCOs. *Parasponia* species are known to be promiscuous. However, I noted that several rhizobial strains are unable to trigger root nodule formation on this species. By comparing the gene repertoire of nodulating and non-nodulating strains, we found that compatibility with *P. andersonii* correlates with the presence of a functional *nodS* gene. *nodS* encodes an N-methyltransferase, which controls N-methylation decoration in its LCOs to form functional nodules on *P. andersonii*. Rhizobium strains that lack the *nodS* gene or produce LCOs lacking N-methylation either cannot induce nodules or only induce limited nodule-like structures without intracellular infection. Low *NF-YA1* induction in response to incompatible strains correlates with the finding described in **Chapter 4** that *PanNF-YA1* is essential for intracellular infection. This suggests that *P. andersonii* intracellular infection requires a high stringency towards the structure of LCOs, similar as observed in legumes.

In **Chapter 6,** I discussed the implications of the results presented in this thesis and provided a broader perspective on the evolution of the nitrogen-fixing nodulation trait. Furthermore, I also discussed how the studies described in my thesis can support engineering of the nitrogen-fixing nodulation trait in non-legume crop plants.

References

Adams DG, Bergman B, Nierzwicki-Bauer SA, Rai AN, Schüßler A. **2006**. Cyanobacterial-Plant Symbioses. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, eds. The Prokaryotes: Volume 1: Symbiotic associations, Biotechnology, Applied Microbiology. New York, NY: Springer New York, 331–363.

Arrighi JF, Barre A, Ben Amor B, Bersoult A, Soriano LC, Mirabella R, de Carvalho-Niebel F, Journet EP, Gherardi M, Huguet T, *et al.* 2006. The *Medicago truncatula* lysine motif-receptor-like kinase gene family includes NFPand new nodule-expressed genes. *Plant physiology* **142**: 265–279.

Banba M, Gutjahr C, Miyao A, Hirochika H, Paszkowski U, Kouchi H, Imaizumi-Anraku H. **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 & cell physiology* **49**: 1659–1671.

Battaglia M, Rípodas C, Clúa J, Baudin M, Aguilar OM, Niebel A, Zanetti ME, Blanco FA. **2014**. A nuclear factor Y interacting protein of the GRAS family is required for nodule organogenesis, infection thread progression, and lateral root growth. *Plant physiology* **164**: 1430–1442.

Bergman B, Rai AN, Rasmussen U. **2007**. Cyanobacterial Associations. *Associative and Endophytic Nitrogen-fixing Bacteria and Cyanobacterial Associations*: 257–301.

Berrada H, Benbrahim KF. **2014**. Taxonomy of the Rhizobia: current perspectives. *British Microbiology Research Journal* **4**: 616–639.

Bozsoki Z, Cheng J, Feng F, Gysel K, Vinther M, Andersen KR, Oldroyd G, Blaise M, Radutoiu S, Stougaard J. **2017**. Receptor-mediated chitin perception in legume roots is functionally separable from Nod factor perception. *Proceedings of the National Academy of Sciences of the United States of America* **114**: E8118–E8127.

de Bruijn FJ. **2015**. Biological Nitrogen Fixation. In: Lugtenberg B, ed. Principles of Plant-Microbe Interactions: Microbes for Sustainable Agriculture. Cham: Springer International Publishing, 215–224.

Capoen W, Sun J, Wysham D, Otegui MS, Venkateshwaran M, Hirsch S, Miwa H, Downie JA, Morris RJ, Ané J-M, *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.

del Cerro P, Rolla-Santos AAP, Gomes DF, Marks BB, Pérez-Montaño F, Rodríguez-Carvajal MÁ, Nakatani AS, Gil-Serrano A, Megías M, Ollero FJ, *et al.* 2015. Regulatory nodD1 and nodD2 genes of Rhizobium tropici strain CIAT 899 and their roles in the early stages of molecular signaling and host-legume nodulation. *BMC genomics* **16**: 251.

Chabaud M, Gherbi H, Pirolles E, Vaissayre V, Fournier J, Moukouanga D, Franche C, Bogusz D, Tisa LS, Barker DG, *et al.* 2016. Chitinase-resistant hydrophilic symbiotic factors secreted by Frankia activate both Ca(2+) spiking and NIN gene expression in the actinorhizal plant *Casuarina glauca*. *The New phytologist* **209**: 86–93.

Charpentier M, Bredemeier R, Wanner G, Takeda N, Schleiff E, Parniske M. **2008**. *Lotus*

japonicus CASTOR and POLLUX are ion channels essential for perinuclear calcium spiking in legume root endosymbiosis. *The Plant cell* **20**: 3467–3479.

Charpentier M, Oldroyd G. **2010**. How close are we to nitrogen-fixing cereals? *Current opinion in plant biology* **13**: 556–564.

Charpentier M, Sun J, Martins TV, Radhakrishnan GV, Findlay K, Soumpourou E, Thouin J, Véry A-A, Sanders D, Morris RJ, *et al.* 2016. Nuclear-localized cyclic nucleotide–gated channels mediate symbiotic calcium oscillations. *Science* **352**: 1102–1105.

Chen C, Fan C, Gao M, Zhu H. **2009**. Antiquity and function of CASTOR and POLLUX, the twin ion channel-encoding genes key to the evolution of root symbioses in plants. *Plant physiology* **149**: 306–317.

Clavijo F, Diedhiou I, Vaissayre V, Brottier L, Acolatse J, Moukouanga D, Crabos A, Auguy F, Franche C, Gherbi H, *et al.* 2015. The Casuarina NIN gene is transcriptionally activated throughout Frankia root infection as well as in response to bacterial diffusible signals. *The New phytologist* **208**: 887–903.

Coba de la Peña T, Fedorova E, Pueyo JJ, Lucas MM. **2018**. The Symbiosome: Legume and Rhizobia Co-evolution toward a Nitrogen-Fixing Organelle? *Frontiers in plant science* **8**: 2229.

Combier J-PP, Frugier F, Billy FD, Boualem A, El-yahyaoui F, Moreau S, Vernié T, Ott T, Gamas P, Crespi M, *et al.* 2006. MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. *Genes & development* **20**: 3084–3088.

Delaux P-M, Varala K, Edger PP, Coruzzi GM, Pires JC, Ané J-M. **2014**. Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. *PLoS genetics* **10**: e1004487.

Del Cerro P, Rolla-Santos AAP, Valderrama-Fernández R, Gil-Serrano A, Bellogín RA, Gomes DF, Pérez-Montaño F, Megías M, Hungría M, Ollero FJ. **2016**. NrcR, a New Transcriptional Regulator of *Rhizobium tropici* CIAT 899 Involved in the Legume Root-Nodule Symbiosis. *PloS one* **11**: e0154029.

De Mita S, Streng A, Bisseling T, Geurts R. **2014**. Evolution of a symbiotic receptor through gene duplications in the legume--rhizobium mutualism. *The New phytologist* **201**: 961–972.

Dénarié J, Debellé F, Promé J-C. **1996**. Rhizobium lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. *Annual review of biochemistry* **65**: 503–535.

Diédhiou I, Tromas A, Cissoko M, Gray K, Parizot B, Crabos A, Alloisio N, Fournier P, Carro L, Svistoonoff S, *et al.* 2014. Identification of potential transcriptional regulators of actinorhizal symbioses in *Casuarina glauca* and *Alnus glutinosa*. *BMC plant biology* **14**: 1–13.

Dos Santos PC, Fang Z, Mason SW, Setubal JC, Dixon R. **2012**. Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC genomics* **13**: 162.

Doyle JJ. **1998**. Phylogenetic perspectives on nodulation: evolving views of plants and symbiotic bacteria. *Trends in plant science* **3**: 473–478.

Doyle JJ. **2011**. Phylogenetic perspectives on the origins of nodulation. *Molecular plant-microbe interactions: MPMI* **24**: 1289–1295.

Doyle JJ. **2016**. Chasing unicorns: Nodulation origins and the paradox of novelty. *American journal of botany* **103**: 1–4.

Endre G, Kereszt A, Mihaceae S, Kalo P, Kiss GB. **2002**. A receptor kinase gene regulating symbiotic nodule development. *Nature* **417**: 962–966.

Estévez J, Soria-Díaz ME, De Córdoba FF, Morón B, Manyani H, Gil A, Thomas-Oates J, Van Brussel AAN, Dardanelli MS, Sousa C, *et al.* 2009. Different and new Nod factors produced by *Rhizobium tropici* CIAT899 following Na+ stress. *FEMS microbiology letters* **293**: 220–231.

European Fertilizer Manufacturers Association. **2000**. Best available techniques for pollution prevention and control in the European fertilizer industry. *Production of Urea and Urea Ammonium Nitrate*.

Fabre S, Gully D, Poitout A, Patrel D, Arrighi J-F, Giraud E, Czernic P, Cartieaux F. **2015**. The Nod factor-independent nodulation in *Aeschynomene evenia* required the common plant-microbe symbiotic 'toolkit'. *Plant physiology* **169**: 2654–2664.

Folch-Mallol JL, Marroqui S, Sousa C, Manyani H, López-Lara IM, van der Drift KM, Haverkamp J, Quinto C, Gil-Serrano A, Thomas-Oates J, *et al.* 1996. Characterization of *Rhizobium tropici* CIAT899 nodulation factors: the role of nodH and nodPQ genes in their sulfation. *Molecular plant-microbe interactions* **9**: 151–163.

Geurts R, Lillo A, Bisseling T. **2012**. Exploiting an ancient signalling machinery to enjoy a nitrogen fixing symbiosis. *Current opinion in plant biology* **15**: 438–443.

Gherbi H, Markmann K, Svistoonoff S, Estevan J, Autran D, Giczey G, Auguy F, Péret B, Laplaze L, Franche C, *et al.* 2008. SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and Frankiabacteria. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 4928–4932.

Gleason C, Chaudhuri S, Yang T, Muñoz A, Poovaiah BW, Oldroyd GED. **2006**. Nodulation independent of rhizobia induced by a calcium-activated kinase lacking autoinhibition. *Nature* **441**: 1149–1152.

Godfroy O, Debellé F, Timmers T, Rosenberg C. **2006**. A rice calcium- and calmodulin-dependent protein kinase restores nodulation to a legume mutant. *Molecular plant-microbe interactions: MPMI* **19**: 495–501.

Granqvist E, Sun J, Op Den Camp R, Pujic P, Hill L, Normand P, Morris RJ, Downie JA, Geurts R, Oldroyd GED, *et al.* 2015. Bacterial-induced calcium oscillations are common to nitrogen-fixing associations of nodulating legumes and nonlegumes. *The New phytologist* **207**: 551–558.

Hocher V, Alloisio N, Auguy F, Fournier P, Doumas P, Pujic P, Gherbi H, Queiroux C, Da Silva C, Wincker P, *et al.* 2011. Transcriptomics of actinorhizal symbioses reveals homologs of the whole common symbiotic signaling cascade. *Plant physiology* **156**: 700–711.

Hossain MS, Shrestha A, Zhong S, Miri M, Austin RS, Sato S, Ross L, Huebert T, Tromas A, Torres-Jerez I, *et al.* 2016. *Lotus japonicus* NF-YA1 Plays an Essential Role During

Nodule Differentiation and Targets Members of the SHI/STY Gene Family. *Molecular plant-microbe interactions: MPMI* **29**: 950–964.

Johansson C, Bergman B. **1992**. Early events during the establishment of the Gunnera/Nostoc symbiosis. *Planta* **188**: 403–413.

Kanamori N, Madsen LH, Radutoiu S, Frantescu M, Quistgaard EMH, Miwa H, Downie JA, James EK, Felle HH, Haaning LL, *et al.* 2006. A nucleoporin is required for induction of Ca2+ 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* **545**: 359–364.

Kevei Z, Lougnon G, Mergaert P, Horvath GV, Kereszt A, Jayaraman D, Zaman N, Marcel F, Regulski K, Kiss GB, *et al.* 2007. 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase1 Interacts with NORK and Is Crucial for Nodulation in *Medicago truncatula*. *The Plant cell* **19**: 3974–3989.

Laloum T, Baudin M, Frances L, Lepage A, Billault-Penneteau B, Cerri MR, Ariel F, Jardinaud M-F, Gamas P, de Carvalho-Niebel F, *et al.* 2014a. Two CCAAT-box-binding transcription factors redundantly regulate early steps of the legume-rhizobia endosymbiosis. *The Plant journal* **79**: 757–768.

Laporte P, Lepage A, Fournier J, Catrice O, Moreau S, Jardinaud M-F, Mun J-H, Larrainzar E, Cook DR, Gamas P, *et al.* 2014. The CCAAT box-binding transcription factor NF-YA1 controls rhizobial infection. *Journal of experimental botany* **65**: 481–494.

Lerouge P, Roche P, Faucher C, Maillet F, Truchet G, Promé JC, Dénarié J. **1990**. Symbiotic host-specificity of Rhizobium melilotiis determined by a sulphated and acylated glucosamine oligosaccharide signal. *Nature* **344**: 781–784.

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

Limpens E, van Zeijl A, Geurts R. **2015**. Lipochitooligosaccharides modulate plant host immunity to enable endosymbioses. *Annual review of phytopathology* **53**: 311–334.

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

Li H-L, Wang W, Mortimer PE, Li R-Q, Li D-Z, Hyde KD, Xu J-C, Soltis DE, Chen Z-D. **2015**. Large-scale phylogenetic analyses reveal multiple gains of actinorhizal nitrogen-fixing symbioses in angiosperms associated with climate change. *Scientific reports* **5**: 14023.

Madsen EB, Madsen LH, Radutoiu S, Sato S, Kaneko T, Tabata S, Sandal N. **2003**. A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. *Nature* **425**: 637–640.

Maillet F, Poinsot V, André O, Puech-Pagès V, Haouy A, Gueunier M, Cromer L, Giraudet D, Formey D, Niebel A, *et al.* 2011. Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature* **469**: 58–63.

Markmann K, Giczey G, 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.

Markmann K, Parniske M. **2009**. Evolution of root endosymbiosis with bacteria: how novel are nodules? *Trends in plant science* **14**: 77–86.

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

Martin FM, Uroz S, Barker DG. **2017**. Ancestral alliances: Plant mutualistic symbioses with fungi and bacteria. *Science* **356**: eaad4501.

Marvel DJ. **1987**. STUDIES OF NODULATION GENES OF A RHIZOBIUM STRAIN THAT NODULATES A NON-LEGUME, PARASPONIA RIGIDA (SYMBIOSIS, NITROGEN FIXATION). https://elibrary.ru/item.asp?id=7479807

Miller JB, Pratap A, Miyahara A, Zhou L, Bornemann S, Morris RJ, Oldroyd GED. **2013**. Calcium/Calmodulin-Dependent Protein Kinase Is Negatively and Positively Regulated by Calcium, Providing a Mechanism for Decoding Calcium Responses during Symbiosis Signaling. *The Plant cell* **25**: 5053–5066.

Moling S, Pietraszewska-Bogiel A, Postma M, Fedorova E, Hink MA, Limpens E, Gadella TWJ, Bisseling T. **2014**. Nod factor receptors form heteromeric complexes and are essential for intracellular infection in medicago nodules. *The Plant cell* **26**: 4188–4199.

Morón B, Soria-Díaz ME, Ault J, Verroios G, Noreen S, Rodríguez-Navarro DN, Gil-Serrano A, Thomas-Oates J, Megías M, Sousa C. **2005**. Low pH changes the profile of nodulation factors produced by *Rhizobium tropici* CIAT899. *Chemistry and Biology* **12**: 1029–1040.

Mulligan JT, Long SR. **1989**. A family of activator genes regulates expression of Rhizobium meliloti nodulation genes. *Genetics* **122**: 7–18.

Miyata K, Kozaki T, Kouzai Y, Ozawa K, Ishii K, Asamizu E, Okabe Y, Umehara Y, Miyamoto A, Kobae Y, *et al.* 2014. Bifunctional plant receptor, OsCERK1, regulates both chitin-triggered immunity and arbuscular mycorrhizal symbiosis in rice. *Plant & cell physiology* 55: 1864–1872.

Normand P, Orso S, Cournoyer B, Jeannin P, Chapelon C, Dawson J, Evtushenko L, Misra AK. **1996**. Molecular phylogeny of the genus Frankia and related genera and emendation of the family Frankiaceae. *International journal of systematic bacteriology* **46**: 1–9.

Normand P, Queiroux C, Tisa LS, Benson DR, Rouy Z, Cruveiller S, Médigue C. **2007**. Exploring the genomes of Frankia. *Physiologia plantarum* **130**: 331–343.

Okazaki S, Kaneko T, Sato S, Saeki K. **2013**. Hijacking of leguminous nodulation signaling by the rhizobial type III secretion system. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 17131–17136.

Okazaki S, Tittabutr P, Teulet A, Thouin J, Fardoux J, Chaintreuil C, Gully D, Arrighi J-F, Furuta N, Miwa H, *et al.* 2016. Rhizobium–legume symbiosis in the absence of Nod factors: two possible scenarios with or without the T3SS. *The ISME journal* **10**: 64–74.

Oldroyd GED, Dixon R. **2014**. Biotechnological solutions to the nitrogen problem. *Current opinion in biotechnology* **26**: 19–24.

Op den Camp RHM, Polone E, Fedorova E, Roelofsen W, Squartini A, Op den Camp HJM, Bisseling T, Geurts R. **2012**. Nonlegume Parasponia andersonii deploys a broad rhizobium host range strategy resulting in largely variable symbiotic effectiveness. *Molecular plant-microbe interactions: MPMI* **25**: 954–963.

Op den Camp R, Streng A, De Mita S, Cao Q, Polone E, Liu W, Ammiraju JSS, Kudrna D, Wing R, Untergasser A, *et al.* 2011a. LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume Parasponia. *Science* **331**: 909–912.

Op den Camp R, Streng A, De Mita S, Cao Q, Polone E, Liu W, Ammiraju JSS, Kudrna D, Wing R, Untergasser A, *et al.* 2011b. LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume Parasponia. *Science* **331**: 909–912.

Ovchinnikova E, Journet E-P, Chabaud M, Cosson V, Ratet P, Duc G, Fedorova E, Liu W, den Camp RO, Zhukov V, *et al.* 2011. IPD3 controls the formation of nitrogen-fixing symbiosomes in pea and *Medicago* spp. *Molecular plant-microbe interactions: MPMI* **24**: 1333–1344.

Pawlowski K, Demchenko KN. **2012**. The diversity of actinorhizal symbiosis. *Protoplasma* **249**: 967–979.

Persson T, Battenberg K, Demina IV, Vigil-Stenman T, Heuvel BV, Pujic P, Facciotti MT, Wilbanks EG, O'Brien A, Fournier P, *et al.* 2015. *Candidatus* Frankia datiscae Dg1, the actinobacterial microsymbiont of *Datisca glomerata*, expresses the canonical nod genes nodABC in symbiosis with its host plant. *PloS one* **10**: e0127630.

Persson T, Benson DR, Normand P, Vanden Heuvel B, Pujic P, Chertkov O, Teshima H, Bruce DC, Detter C, Tapia R, *et al.* 2011. Genome Sequence of '*Candidatus* Frankia datiscae' Dg1, the Uncultured Microsymbiont from Nitrogen-Fixing Root Nodules of the Dicot *Datisca glomerata*. *Journal of bacteriology* **193**: 7017–7018.

Peters NK, Frost JW, Long SR. **1986**. A plant flavone, luteolin, induces expression of *Rhizobium meliloti* nodulation genes. *Science* **233**: 977–980.

Poole P, Ramachandran V, Terpolilli J. **2018**. Rhizobia: from saprophytes to endosymbionts. *Nature reviews. Microbiology* **16**: 291–303.

Price NPJ, RelicA B, Talmont F, Lewin A, Promé D, Pueppke SG, Maillet F, Dénarié J, Promé J-C, Broughton WJ. **1992**. Broad-host-range *Rhizobium* species strain NGR234 secretes a family of carbamoylated, and fucosylated, nodulation signals that are O-acetylated or sulphated. *Molecular microbiology* **6**: 3575–3584.

Price NP, Talmont F, Wieruszeski JM, Promé D, Promé JC. **1996**. Structural determination of symbiotic nodulation factors from the broad host-range *Rhizobium* species NGR234. *Carbohydrate research* **289**: 115–136.

Provorov NA, Onishchuk OP. **2018**. Microbial Symbionts of Insects: Genetic Organization, Adaptive Role, and Evolution. *Microbiology* **87**: 151–163.

Pueppke SG, Broughton WJ. **1999**. *Rhizobium* sp. strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges. *Molecular plant-microbe interactions: MPMI*

12: 293–318.

Radutoiu S, Madsen LH, Madsen EB, Felle HH, Umehara Y, Grønlund M, Sato S, Nakamura Y, Tabata S, Sandal N, *et al.* 2003. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* **425**: 585–592.

Raymond J, Siefert JL, Staples CR, Blankenship RE. **2004**. The natural history of nitrogen fixation. *Molecular biology and evolution* **21**: 541–554.

Redmond JW, Batley M, Djordjevic MA, Innes RW, Kuempel PL, Rolfe BG. **1986**. Flavones induce expression of nodulation genes in Rhizobium. *Nature* **323**: 632–635.

Remigi P, Zhu J, Young JPW, Masson-Boivin C. **2016**. Symbiosis within Symbiosis: Evolving Nitrogen-Fixing Legume Symbionts. *Trends in microbiology* **24**: 63–75.

Ried MK, Antolín-Llovera M, Parniske M. **2014**. Spontaneous symbiotic reprogramming of plant roots triggered by receptor-like kinases. *eLife* **3**: 1–17.

Rípodas C, Castaingts M, Clúa J, Villafañe J, Blanco FA, Zanetti ME. **2019**. The PvNF-YA1 and PvNF-YB7 Subunits of the Heterotrimeric NF-Y Transcription Factor Influence Strain Preference in the *Phaseolus vulgaris-Rhizobium etli* Symbiosis. *Frontiers in plant science* **10**: 221.

Rípodas C, Clúa J, Battaglia M, Baudin M, Niebel A, Zanetti ME, Blanco F. **2014**. Transcriptional regulators of legume-rhizobia symbiosis: nuclear factors Ys and GRAS are two for tango. *Plant signaling & behavior* **9**: e28847.

Roy A, Bousquet J. **1996**. The evolution of the actinorhizal symbiosis through phylogenetic analysis of host plants. *Acta botanica Gallica: bulletin de la Societe botanique de France* **143**: 635–650.

Saha S, Paul A, Herring L, Dutta A, Bhattacharya A, Samaddar S, Goshe MB, DasGupta M. **2016**. Gatekeeper Tyrosine phosphorylation of SYMRK is essential for synchronising the epidermal and cortical responses in root nodule symbiosis. *Plant physiology* **171**: 01962.2015.

Santi C, Bogusz D, Franche C. **2013**. Biological nitrogen fixation in non-legume plants. *Annals of botany* **111**: 743–767.

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

Schwencke J, Carú M. **2001**. Advances in Actinorhizal Symbiosis: Host Plant- Frankia Interactions, Biology, and Applications in Arid Land Reclamation. A Review. *Arid Land Research and Management* **15**: 285–327.

Sen A, Daubin V, Abrouk D, Gifford I, Berry AM, Normand P. **2014**. Phylogeny of the class Actinobacteria revisited in the light of complete genomes. The orders 'Frankiales' and Micrococcales should be split into coherent entities: proposal of Frankiales ord. nov., Geodermatophilales ord. nov., Acidothermales ord. nov. and Nakamurellales ord. nov. *International journal of systematic and evolutionary microbiology* **64**: 3821–3832.

Shimoda Y, Han L, Yamazaki T, Suzuki R, Hayashi M, Imaizumi-Anraku H. **2012**. Rhizobial and fungal symbioses show different requirements for calmodulin binding to calcium calmodulin--dependent protein kinase in *Lotus japonicus*. *The Plant cell* **24**: 304–321.

Sieberer BJ, Chabaud M, Fournier J, Timmers ACJ, Barker DG. **2012**. A switch in Ca 2+ spiking signature is concomitant with endosymbiotic microbe entry into cortical root cells of *Medicago truncatula*. *The Plant journal: for cell and molecular biology* **69**: 822–830.

Singh S, Katzer K, Lambert J, Cerri M, Parniske M. **2014**. CYCLOPS, A DNA-binding transcriptional activator, orchestrates symbiotic root nodule development. *Cell host & microbe* **15**: 139–152.

Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JM, Martin PG. **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* **92**: 2647–2651.

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

Soyano T, Kouchi H, Hirota A, Hayashi M. **2013b**. Nodule inception directly targets NF-Y subunit genes to regulate essential processes of root nodule development in *Lotus japonicus*. *PLoS genetics* **9**: e1003352.

Stokstad E. **2016**. The nitrogen fix. *Science* **353**: 1225–1227.

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

Sutton MA, Oenema O, Erisman JW, Leip A, van Grinsven H, Winiwarter W. **2011**. Too much of a good thing. *Nature* **472**: 159–161.

Svistoonoff S, Benabdoun FM, Nambiar-veetil M, Imanishi L, Bonneau J, Wall L, Ykhlef N, Rosenberg C, Bogusz D, Franche C. **2013**. The independent acquisition of plant root nitrogen-fixing symbiosis in fabids recruited the same genetic pathway for nodule organogenesis. *PloS one* **8**: e64515.

Svistoonoff S, Hocher V, Gherbi H. **2014**. Actinorhizal root nodule symbioses: what is signalling telling on the origins of nodulation? *Current opinion in plant biology* **20**: 11–18.

Swensen SM. **1996**. The evolution of actinorhizal symbioses: Evidence for multiple origins of the symbiotic association. *American journal of botany* **83**: 1503–1512.

Teulet A, Busset N, Fardoux J, Gully D, Chaintreuil C, Cartieaux F, Jauneau A, Comorge V, Okazaki S, Kaneko T, *et al.* 2019. The rhizobial type III effector ErnA confers the ability to form nodules in legumes. *Proceedings of the National Academy of Sciences of the United States of America* **116**: 21758–21768.

Tirichine L, Imaizumi-Anraku H, Yoshida S, Murakami Y, Madsen LH, Miwa H, Nakagawa T, Sandal N, Albrektsen AS, Kawaguchi M. **2006**. Deregulation of a Ca2+/calmodulin-dependent kinase leads to spontaneous nodule development. *Nature* **441**: 1153–1156.

Tromas A, Parizot B, Diagne N, Champion A, Hocher V, Cissoko M, Crabos A, Prodjinoto H, Lahouze B, Bogusz D, *et al.* 2012. Heart of endosymbioses: transcriptomics reveals a conserved genetic program among arbuscular mycorrhizal, actinorhizal and legume-

rhizobial symbioses. *PloS one* **7**: e44742.

Van Nguyen T, Wibberg D, Battenberg K, Blom J, Vanden Heuvel B, Berry AM, Kalinowski J, Pawlowski K. **2016**. An assemblage of FrankiaCluster II strains from California contains the canonical nodgenes and also the sulfotransferase gene nodH. *BMC genomics* **17**: 796.

Venkateshwaran M, Jayaraman D, Chabaud M, Genre A, Balloon AJ, Maeda J, Forshey K, den Os D, Kwiecien NW, Coon JJ, *et al.* 2015. A role for the mevalonate pathway in early plant symbiotic signaling. *Proceedings of the National Academy of Sciences* **112**: 9781–9786.

Vernié T, Kim J, Frances L, Ding Y, Sun J, Guan D, Niebel A, Gifford ML, de Carvalho-Niebel F, Oldroyd GED. **2015**. The NIN Transcription Factor Coordinates Diverse Nodulation Programs in Different Tissues of the Medicago truncatula Root. *The Plant cell*: tpc.15.00461.

Wang H, Moore MJ, Soltis PS, Bell CD, Brockington SF, Alexandre R, Davis CC, Latvis M, Manchester SR, Soltis DE. **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.

Wang B, Qiu Y-L. **2006**. Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* **16**: 299–363.

Wani SP, Rupela OP, Lee KK. **1995**. Sustainable agriculture in the semi-arid tropics through biological nitrogen fixation in grain legumes. Symposium on Biological Nitrogen Fixation for Sustainable Agriculture at the 15th Congress of Soil Science, Acapulco, Mexico, 1994. Dordrecht: Springer Netherlands, 29–49.

Warshan D, Liaimer A, Pederson E, Kim S-Y, Shapiro N, Woyke T, Altermark B, Pawlowski K, Weyman PD, Dupont CL, *et al.* 2018. Genomic Changes Associated with the Evolutionary Transitions of Nostoc to a Plant Symbiont. *Molecular biology and evolution* **35**: 1160–1175.

Werner GDA, Cornwell WK, Sprent JI, Kattge J, Kiers ET. **2014**. A single evolutionary innovation drives the deep evolution of symbiotic N2-fixation in angiosperms. *Nature communications* **5**: 4087.

Worrell E, Phylipsen D, Einstein D, Martin N. **2000**. Energy use and energy intensity of the U.S. chemical industry.

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

Yano K, Yoshida S, Müller J, Singh S, Banba M, Vickers K, Markmann K, White C, Schuller B, Sato S, *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.

Yoro E, Suzaki T, Toyokura K, Miyazawa H, Fukaki H, Kawaguchi M. **2014**. A Positive Regulator of Nodule Organogenesis, NODULE INCEPTION, Acts as a Negative Regulator of Rhizobial Infection in *Lotus japonicus*. *Plant physiology* **165**: 747–758.

van Zeijl A, Wardhani TAK, Seifi Kalhor M, Rutten L, Bu F, Hartog M, Linders S, Fedorova EE, Bisseling T, Kohlen W, *et al.* 2018. CRISPR/Cas9-Mediated Mutagenesis of Four Putative Symbiosis Genes of the Tropical Tree *Parasponia andersonii* Reveals Novel Phenotypes. *Frontiers in plant science* **9**: 284.

Zhou J, Duan J, Gao M, Wang Y, Wang X, Zhao K. **2019**. Diversity, Roles, and Biotechnological Applications of Symbiotic Microorganisms in the Gut of Termite. *Current microbiology* **76**: 755–761.

Chapter 2

Comparative genomics of the non-legume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobium symbioses

Robin van Velzen[*], Rens Holmer[*], Fengjiao Bu[$], Luuk Rutten[$], Arjan van Zeijl, Wei Liu, Luca Santuari, Qingqin Cao, Trupti Sharma[1], Defeng Shen, Yuda Roswanjaya, Titis A.K. Wardhani, Maryam Seifi Kalhor, Joelle Jansen, Johan van den Hoogen, Berivan Güngör, Marijke Hartog, Jan Hontelez, Jan Verver, Wei-Cai Yang, Elio Schijlen, Rimi Repin, Menno Schilthuizen, M. Eric Schranz, Renze Heidstra, Kana Miyata, Elena Fedorova, Wouter Kohlen, Ton Bisseling, Sandra Smit, & Rene Geurts
(2018) Comparative genomics of the nonlegume Parasponia reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proc Natl Acad Sci USA* 115:E4700-E4709.

*: contributed equally

$: contributed equally

Abstract

Nodules harboring nitrogen-fixing rhizobia are a well-known trait of legumes, but nodules also occur in other plant lineages either with rhizobia or the actinomycete *Frankia* as microsymbiont. It is generally assumed that nodulation evolved independently multiple times. However, molecular genetic support for this hypothesis is lacking, as the genetic changes underlying nodule evolution remain elusive. We conducted genetic and comparative genomics studies using *Parasponia* species (Cannabaceae), the only non-legumes that can establish nitrogen-fixing nodules with rhizobium. Intergeneric crosses between *Parasponia andersonii* and its non-nodulating relative *Trema tomentosa* demonstrated that nodule organogenesis, but not intracellular infection, is a dominant genetic trait. Comparative transcriptomics of *P. andersonii* and the legume *Medicago truncatula* revealed utilization of at least 290 orthologous symbiosis genes in nodules. Among these are key genes that in legumes are essential for nodulation, including *NODULE INCEPTION* (*NIN*) and *RHIZOBIUM-DIRECTED POLAR GROWTH (RPG)*. Comparative analysis of genomes from three *Parasponia* species and related non-nodulating plant species show evidence of parallel loss in non-nodulating species of putative orthologs of *NIN*, *RPG*, and *NOD FACTOR PERCEPTION*. Parallel loss of these symbiosis genes indicates that these non-nodulating lineages lost the potential to nodulate. Taken together, our results challenge the view that nodulation evolved in parallel and raises the possibility that nodulation originated ~100 million years ago in a common ancestor of all nodulating plant species, but was subsequently lost in many descendant lineages. This will have profound implications for translational approaches aimed at engineering nitrogen-fixing nodules in crop plants.

Introduction

Nitrogen sources such as nitrate or ammonia are key nutrients for plant growth, but their availability is frequently limited. Some plant species in the related orders Fabales, Fagales, Rosales, and Cucurbitales -collectively known as the nitrogen-fixing clade- can overcome this limitation by establishing a nitrogen-fixing endosymbiosis with either *Frankia* or rhizobium bacteria (Soltis *et al.*, 1995). These symbioses require specialized root organs, known as nodules, that provide optimal physiological conditions for nitrogen fixation (Udvardi & Poole, 2013). For example, nodules of legumes (Fabaceae, order Fabales) contain a high concentration of hemoglobin that is essential to control oxygen homeostasis and protect the rhizobial nitrogenase enzyme complex from oxidation (Ott *et al.*, 2005; Udvardi & Poole, 2013). Legumes, such as soybean (*Glycine max*), common bean (*Phaseolus vulgaris*), and peanut (*Arachis hypogaea*) represent the only crops that possess nitrogen-fixing nodules, and engineering this trait in other crop plants is a long-term vision in sustainable agriculture (Burrill & Hansen, 1917; Stokstad, 2016).

Nodulating plants represent ~10 related clades that diverged >100 million years ago, supporting a shared evolutionary origin of the underlying capacity for this trait (Soltis *et al.*, 1995). Nevertheless, these nodulating clades are interspersed with many non-nodulating lineages. This has led to two hypotheses explaining the evolution of nodulation (Soltis *et al.*, 1995). (i) Nodulation has a single origin in the root of the nitrogen-fixation clade, followed by multiple independent losses. (ii) Nodulation originated independently multiple times, preceded by a single hypothetical predisposition event in a common ancestor of the nitrogen-fixing fixation clade. The latter of these hypotheses is most widely accepted (Swensen, 1996; Doyle, 1998, 2011, 2016; Werner *et al.*, 2014; Li *et al.*, 2015; Martin *et al.*, 2017).

Genetic dissection of rhizobium symbiosis in two legume models -*Medicago truncatula* (medicago) and *Lotus japonicus* (lotus)- has uncovered symbiosis genes that are essential for nodule organogenesis, bacterial infection, and nitrogen fixation (Dataset S1). These include genes encoding LysM-type receptors that perceive rhizobial lipo-chitooligosaccharides (LCOs, also known as Nod factors) and transcriptionally activate the *NODULE INCEPTION* (*NIN*) transcription factor (Limpens *et al.*, 2003; Madsen *et al.*, 2003; Radutoiu *et al.*, 2003; Arrighi *et al.*, 2006; Marsh *et al.*, 2007; Broghammer *et al.*, 2012). Expression of *NIN* is essential and sufficient to set in motion nodule organogenesis (Schauser *et al.*, 1999; Marsh *et al.*, 2007; Soyano *et al.*, 2013; Vernié *et al.*, 2015). Some symbiosis genes have been co-opted from the

more ancient and widespread arbuscular mycorrhizal symbiosis (Parniske, 2008; Oldroyd, 2013). However, causal genetic differences between nodulating and non-nodulating species have not been identified (Geurts *et al.*, 2016).

To obtain insight in the molecular genetic changes underlying evolution of nitrogen-fixing root nodules we conducted comparative studies using *Parasponia* (Cannabaceae, order Rosales). The genus *Parasponia* is the only lineage outside the legume family establishing a nodule symbiosis with rhizobium (Clason, 1936; Trinick, 1973; Akkermans *et al.*, 1978; Becking, 1992). Similar as shown for legumes, nodule formation in *Parasponia* is initiated by rhizobium-secreted LCOs (Marvel *et al.*, 1987; Op den Camp *et al.*, 2011; Granqvist *et al.*, 2015). This suggests that *Parasponia* and legumes utilize a similar set of genes to control nodulation, but the extent of common gene utilization between distantly related nodulating species remains unknown. The genus *Parasponia* represents a clade of five species that is phylogenetically embedded in the closely related *Trema* genus (Yang *et al.*, 2013). Like *Parasponia* and most other land plants, *Trema* species can establish an arbuscular mycorrhizal symbiosis (SI Appendix, Fig. S1). However, they are non-responsive to rhizobium LCOs and do not form nodules (Becking, 1992; Granqvist *et al.*, 2015). Taken together, *Parasponia* is an excellent system for comparative studies with legumes and non-nodulating *Trema* species to provide insights into the molecular genetic changes underlying evolution of nitrogen-fixing root nodules.

Results

Nodule organogenesis is a genetically dominant trait

First, we took a genetics approach for understanding the rhizobium symbiosis trait of *Parasponia* by making intergeneric crosses (SI Appendix, Table S1). Viable F_1 hybrid plants were obtained only from the cross *Parasponia andersonii* (2n=20) x *Trema tomentosa* (2n=4x=40) (Fig. 1A, SI Appendix, Fig. S2). These triploid hybrids (2n=3x=30) were infertile, but could be propagated clonally. We noted that F_1 hybrid plants formed root nodules when grown in potting soil, similar as earlier observations for *P. andersonii* (Op den Camp *et al.*, 2012). To further investigate the nodulation phenotype of these hybrid plants, clonally propagated plants were inoculated with two different strains; *Bradyrhizobium elkanii* strain WUR3 (Op den Camp *et al.*, 2012) or *Mesorhizobium plurifarium* strain BOR2. The latter strain was isolated from the rhizosphere of *Trema orientalis* in Malaysian Borneo and showed to be an effective nodulator of *P. andersonii* (SI Appendix, Fig. S3). Both strains induced nodules on F_1 hybrid plants (Fig. 1B,D,E; SI Appendix, Fig. S4) but, as expected, not on *T. tomentosa*, nor on any other *Trema* species investigated. Using an acetylene reduction assay we noted that, in contrast to *P. andersonii* nodules, in F_1 hybrid nodules of plant H9 infected with *M. plurifarium* BOR2 there is no nitrogenase activity (Fig. 1C). To further examine this discrepancy, we studied the cytoarchitecture of these nodules. In *P. andersonii* nodules, apoplastic *M. plurifarium* BOR2 colonies infect cells to form so-called fixation threads (Fig. 1F,H-J), whereas in F_1 hybrid nodules these colonies remain apoplastic, and fail to establish intracellular infections (Fig. 1G,K). To exclude the possibility that the lack of intracellular infection is caused by heterozygosity of *P. andersonii* where only a nonfunctional allele was transmitted to the F_1 hybrid genotype, or by the particular rhizobium strain used for this experiment, we examined five independent F_1 hybrid plants either inoculated with *M. plurifarium* BOR2 or *B. elkanii* WUR3. This revealed a lack of intracellular infection structures in nodules of all F_1 hybrid plants tested, irrespective which of both rhizobium strains was used (Fig. 1G,K, SI Appendix, Fig. S4), confirming that heterozygosity of *P. andersonii* does not play a role in the F_1 hybrid infection phenotype. These results suggest, at least partly, independent genetic control of nodule organogenesis and rhizobium infection. Because F_1 hybrids are nodulated with similar efficiency as *P. andersonii* (Fig. 1B), we conclude that the network controlling nodule organogenesis is genetically dominant.

Parasponia and *Trema* genomes are highly similar

Based on preliminary genome size estimates using FACS measurements, three *Parasponia* and five *Trema* species were selected for comparative genome analysis (SI Appendix, Table S2). K-mer analysis of medium-coverage genome sequence data (~30x) revealed that all genomes had low levels of heterozygosity, except those of *Trema levigata* and *T. orientalis* accession RG16 (SI Appendix, Fig. S5). Based on these k-mer data we also generated more accurate

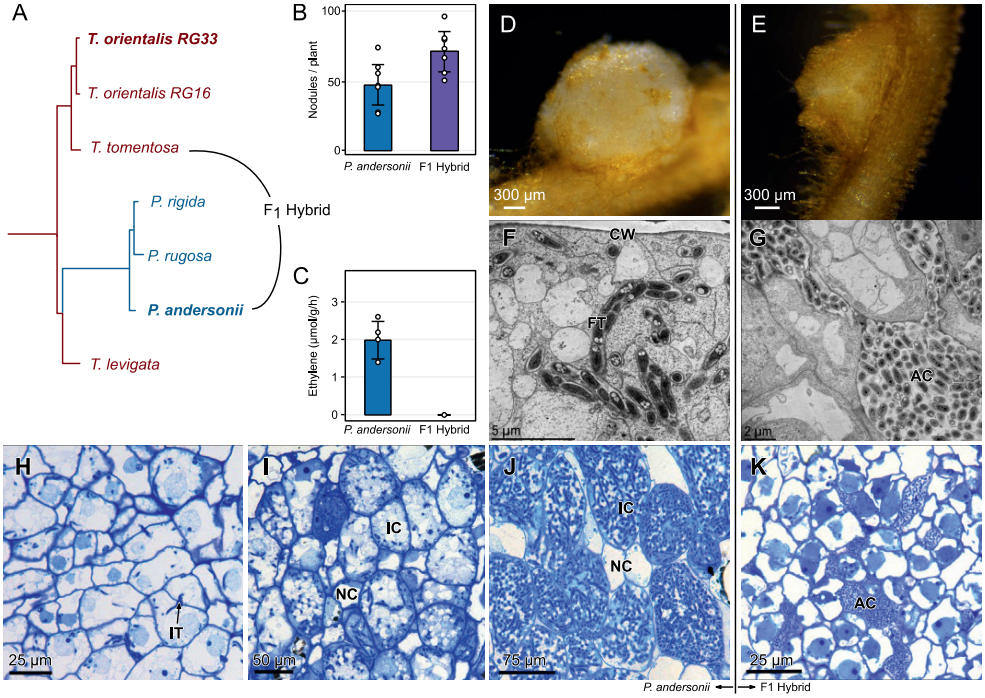


Fig.1 Nodulation phenotype of *Parasponia andersonii* and interspecific *P. andersonii* x *Trema tomentosa* F_1 hybrid plants. A Phylogenetic reconstruction based on whole chloroplast of *Parasponia* and *Trema*. The *Parasponia* lineage (blue) is embedded in the *Trema* genus (red). Species selected for interspecific crosses are indicated, species used for reference genome assembly are in bold. All nodes had a posterior probability of 1. **B** Mean number of nodules on roots of *P. andersonii* and F_1 hybrid plants (n=7). **C** Mean nitrogenase activity in acetylene reductase assay of *P. andersonii* and F_1 hybrid nodules (n=4). Barplot error bars indicate standard deviations; dots represent individual measurements **D** *P. andersonii* nodule. **E** F_1 hybrid nodule. **F,G** Ultrastructure of nodule tissue of *P. andersonii* **F** and F_1 hybrid **G**. Note the intracellular fixation thread (FT) in the cell of *P. andersonii* in comparison with the extracellular, apoplastic colonies of rhizobia (AC) in the F_1 hybrid nodule. **H-J** Light microscopy images of *P. andersonii* nodules in three subsequent developmental stages. **H** Stage 1: initial infection threads (IT) enter the host cells. **I** Stage 2: progression of rhizobium infection in nodule host cell, **J** Stage 3: nodule cells completely filled with fixation threads. Note difference in size between the infected (IC) and non-infected cells (NC). **K** Light microscopy image of F1 hybrid nodule cells. Note rhizobium colonies in apoplast, surrounding the host cells (AC). Nodules have been analysed 6 weeks post inoculation with *Mesorhizobium plurifarium* BOR2. Abbreviations: FT: fixation thread, CW: cell wall, AC: apoplastic colony of rhizobia, IT: infection threads, IC: infected cell, NC: non-infected cell.

estimates of genome sizes. Additionally, we used these data to assemble chloroplast genomes based on which we obtained additional phylogenetic evidence that *T. levigata* is sister to *Parasponia* (Fig. 1A, SI Appendix, Fig. S6-8). Graph-based clustering of repetitive elements in the genomes (calibrated with the genome size estimates based on k-mers) revealed that all selected species contain roughly 300 Mb of non-repetitive sequence, and a variable repeat content that correlates with the estimated genome size that ranges from 375 to 625 Mb (SI Appendix, Fig. S9, Table S3). Notably, we found a *Parasponia*-specific expansion of ogre/tat LTR retrotransposons comprising 65 to 85 Mb (SI Appendix, Fig. S9b). We then generated annotated reference genomes using high-coverage (~125X) sequencing of *P. andersonii* accession WU1 (Op den Camp *et al.*, 2011) and *T. orientalis* accession RG33 (SI Appendix, Tables S4-5). These species were selected based on their low heterozygosity levels in combination with relatively small genomes. *T. tomentosa* was not used for a high-quality genome assembly because it is an allotetraploid (SI Appendix, Fig. S5, Tables S2-3).

We generated orthogroups for *P. andersonii* and *T. orientalis* genes and six other Eurosid species, including arabidopsis (*Arabidopsis thaliana*) and the legumes medicago and soybean. From both *P. andersonii* and *T. orientalis* approximately 35,000 genes could be clustered into >20,000 orthogroups (Dataset S2, SI Appendix, Table S6, note that there can be multiple orthologous gene pairs per orthogroup). Within these orthogroups we identified 25,605 *P. andersonii* - *T. orientalis* orthologous gene pairs based on phylogenetic analysis as well as whole genome alignments (SI Appendix, Table S6). These orthologous gene pairs had a median percentage nucleotide identity of 97% for coding regions (SI Appendix, Fig. S10-11). This further supports the recent divergence of the two species and facilitates their genomic comparison.

Common utilization of symbiosis genes in *Parasponia* and medicago

To assess commonalities in the utilization of symbiosis genes in *Parasponia* species and legumes we employed two strategies. First, we performed phylogenetic analyses of close homologs of genes that were characterized to function in legume-rhizobium symbiosis. This revealed that *P. andersonii* contains putative orthologs of the vast majority of these legume symbiosis genes (96 out of 126; Dataset S1, S3). Second, we compared the sets of genes with enhanced expression in nodules of *P.andersonii* and medicago. RNA sequencing of *P. andersonii* nodules revealed 1,719 genes that are functionally annotated and have a significantly enhanced expression level (fold change > 2, $p < 0.05$, DESeq2 Wald test) in any

of three nodule developmental stages compared with uninoculated roots (SI Appendix, Fig. S12, Dataset S4). For medicago, we generated a comparable data set of 2,753 nodule-enhanced genes based on published RNA sequencing data (Roux *et al.*, 2014). We then determined the overlap of these two gene sets based on orthogroup membership and found that 382 orthogroups comprise both *P. andersonii* and medicago nodule-enhanced genes. This number is significantly larger than is to be expected by chance (permutation test, $p < 0.00001$) (Dataset S5, SI Appendix, Fig. S13). Based on phylogenetic analysis of these orthogroups we found that in 290 cases putative orthologs have been utilized in both *P. andersonii* and medicago root nodules (Dataset S5, S6). Among these 290 commonly utilized genes are 26 putative orthologs of legume symbiosis genes; e.g. the LCO-responsive transcription factor *NIN* and its downstream target *NUCLEAR TRANSCRIPTION FACTOR-YA1* (*NFYA1*) that are essential for nodule organogenesis (Schauser *et al.*, 1999; Combier *et al.*, 2006; Soyano *et al.*, 2013; Baudin *et al.*, 2015), and *RHIZOBIUM DIRECTED POLAR GROWTH* (*RPG*) involved in intracellular infection (Arrighi *et al.*, 2008). Of these 26, five are known to function also in arbuscular mycorrhizal symbiosis (namely *VAPYRIN*, *SYMBIOTIC REMORIN,* the transcription factors *CYCLOPS* and *SAT1*, and a cysteine proteinase gene) (Kistner *et al.*, 2005; Deguchi *et al.*, 2007; Yano *et al.*, 2008; Pumplin *et al.*, 2010; Murray *et al.*, 2011; Horváth *et al.*, 2011; Tóth *et al.*, 2012; Chiasson *et al.*, 2014). To further assess whether commonly utilized genes may be co-opted from the ancient and widespread arbuscular mycorrhizal symbiosis we determined which fraction is also induced upon mycorrhization in medicago based on published RNA sequencing data (Afkhami & Stinchcombe, 2016). This revealed that only 8% of the commonly utilized genes have such induction in both symbioses (Dataset S5).

By exploiting the insight that nodule organogenesis and rhizobial infection can be genetically dissected using hybrid plants we classified these commonly utilized genes into two categories based on their expression profiles in roots and nodules of both *P. andersonii* and F1 hybrids (Fig. 2). The first category comprises 126 genes that are upregulated in both *P. andersonii* and hybrid nodules and that we associate with nodule organogenesis. The second category comprises 164 genes that are only upregulated in the *P. andersonii* nodule and that we therefore associate with infection and/or fixation (Dataset S5). Based on these results we conclude that

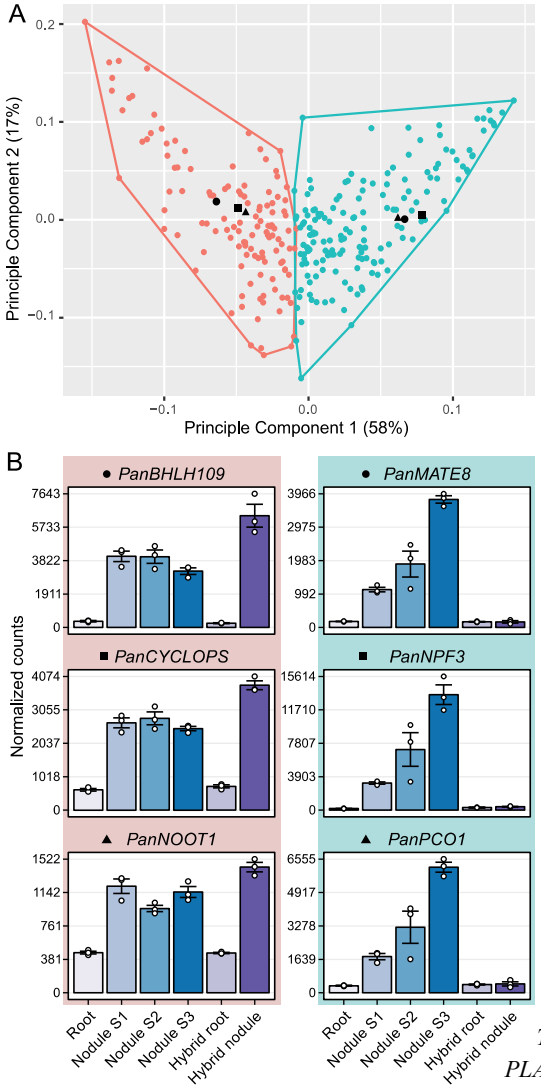


Fig. 2 Clustering of commonly utilized symbiosis genes based on expression profile in *Parasponia andersonii*. **A** Principal component analysis plot of the expression profile of 290 commonly utilized symbiosis genes in 18 transcriptome samples: *P. andersonii* roots and nodules (stage 1-3), hybrid roots and nodules (line H9). All samples have three biological replicates. First two components are shown, representing 75% of the variation in all samples. Colors indicate clusters (K-means clustering using Pearson correlation as distance measure, k=2) of genes with similar expression patterns. The three genes with the highest pearson correlation to the cluster centroids are indicated as black dots, triangles, and squares, and their expression profiles are given in panel **B**. Cluster 1 (pink) represents genes related to nodule organogenesis: these genes are upregulated in both *P. andersonii* and hybrid nodules. Cluster 2 (green) represents genes related to infection and fixation: these genes are highly upregulated in *P. andersonii* nodules but do not respond in the hybrid nodule. *PanBHLH109: BASIC HELIX-LOOP-HELIX DOMAIN CONTAINING PROTEIN 109*; *PanNOOT1: NODULE ROOT 1*; *PanMATE8: MULTI ANTIMICROBIAL EXTRUSION PROTEIN 8*; *PanNPF3*: *NITRATE/PEPTIDE TRANSPORTER FAMILY 3*; *PanPCO1: PLANT CYSTEINE OXIDASE 1.*

Parasponia and medicago utilize orthologous genes that commit various functions in at least two different developmental stages of the root nodule.

Lineage-specific adaptation in *Parasponia* HEMOGLOBIN 1

A notable exception to the pattern of common utilization in root nodules are the oxygen-binding hemoglobins. Earlier studies showed that *Parasponia* and legumes have recruited different hemoglobin genes (Sturms *et al.*, 2010). Whereas legumes use class II LEGHEMOGLOBIN to control oxygen homeostasis, *Parasponia* recruited the paralogous class I HEMOGLOBIN 1

(*HB1*) for this function (Fig. 3A,B). Biochemical studies have revealed that *P. andersonii* PanHB1 has oxygen affinities and kinetics that are adapted to their symbiotic function, whereas this is not the case for *T. tomentosa* TtoHB1 (Sturms *et al.*, 2010; Kakar *et al.*, 2011). We therefore examined HB1 from *Parasponia* species, *Trema* species, and other non-symbiotic Rosales species to see if these differences are due to a gain of function in *Parasponia* or a loss of function in the non-symbiotic species. Based on protein alignment we identified *Parasponia*-specific adaptations in 7 amino acids (Fig. 3C,D). Among these is Ile(101) for which it is speculated to be causal for a functional change in *P. andersonii* HB1 (Kakar *et al.*, 2011). Hemoglobin-controlled oxygen homeostasis is crucial to protect the rhizobial nitrogen-fixing enzyme complex Nitrogenase in legume rhizobium-infected nodule cells (Ott *et al.*, 2005; Udvardi & Poole, 2013). Therefore, *Parasponia*-specific gain of function adaptations in HB1 may have comprised an essential evolutionary step towards functional nitrogen-fixing root nodules with rhizobium endosymbionts.

Parallel loss of symbiosis genes in *Trema* and other relatives of *Parasponia*

Evolution of complex genetic traits is often associated with gene copy number variations (CNVs) (Żmieńko *et al.*, 2014). To test if CNVs were associated with the generally assumed independent evolution of nodulation in *Parasponia*, we focused on two gene sets: (i) close homologs and putative orthologs of the genes that were characterized to function in legume-rhizobium symbiosis, and (ii) genes with a nodule-enhanced expression and functional annotation in *P. andersonii* (these sets partially overlap and together comprise 1,813 genes; SI Appendix, Fig. S14). We discarded *Trema*-specific duplications as we considered them irrelevant for the nodulation phenotype. To ensure that our findings are consistent between the *Parasponia* and *Trema* genera and not due to species-specific events, we analyzed the additional draft genome assemblies of two *Parasponia* and two *Trema* species (SI Appendix Table S5). As these additional draft genomes were relatively fragmented, we sought additional support for presence and absence of genes by mapping sequence reads to the *P. andersonii* and *T. orientalis* reference genomes and by genomic alignments. This procedure revealed only 11 consistent CNVs in the 1,813 symbiosis genes examined, further supporting the recent divergence between *Parasponia* and *Trema* (SI Appendix, Fig. S15). Due to the dominant inheritance of nodule organogenesis in F_1 hybrid plants, we anticipated finding *Parasponia*-

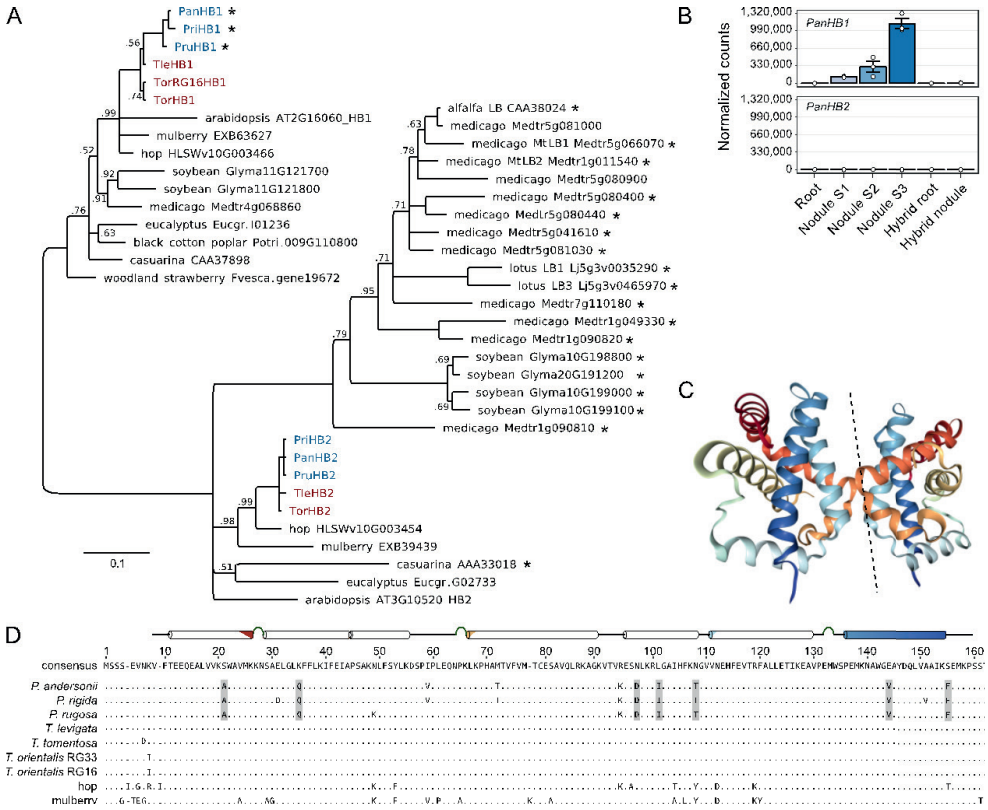


Fig. 3 *Parasponia*-specific adaptations in class 1 hemoglobin protein HB1. A Phylogenetic reconstruction of class 1 (OG0010523) and class 2 hemoglobins (OG0002188). Symbiotic hemoglobins are marked with an asterisk; legumes and the actinorhizal plant casuarina have recruited class 2 hemoglobins for balancing oxygen levels in their nodules. Conversely, *Parasponia* has recruited a class 1 hemoglobin *PanHB1* confirming parallel evolution of symbiotic oxygen transport in this lineage. *Medicago truncatula* (Medtr); *Glycine max* (Glyma), *Populus trichocarpa* (Potri); *Fragaria vesca* (Fvesca); *Eucalyptus grandis* (Eugr); *Arabidopsis thaliana* (AT). Node values indicate posterior probabilities below 1; Scale bar represents substitutions per site. *Parasponia* marked in blue, *Trema* in red. **B** Expression profile of *PanHB1* and *PanHB2* in *P. andersonii* roots, stage 1-3 nodules, and in *P. andersonii* x *T. tomentosa* F_1 hybrid roots and nodules (line H9). Expression is given in DESeq2 normalized read counts, error bars represent standard error of three biological replicates, dots represent individual expression levels. **C** Crystal structure of the asymmetric dimer of PanHB1 as deduced by Kakar *et al.* 2011(Kakar *et al.*, 2011). Dashed line separates the two units. **D** Protein sequence alignment of class 1 hemoglobins from *Parasponia* spp., *Trema* spp., hop (*Humulus lupulus*), and mulberry (*Morus notabilis*). Only amino acids that differ from the consensus are drawn. A linear model of the crystal structure showing alpha helices and turns is depicted above the consensus sequence. There are seven amino acids (marked grey) that consistently differ between all *Parasponia* and all other sampled species: Ala(21), Gln(35), Asp(97), Ile(101), Thr(108), Val(144), and Phe(155). These differences therefore correlate with the functional divergence between *P. andersonii* PanHB1 and *T. tomentosa* TtoHB1 (Sturms *et al.*, 2010; Kakar *et al.*, 2011).

specific gene duplications that could be uniquely associated with nodulation. Surprisingly, we found only one consistent *Parasponia*-specific duplication in symbiosis genes; namely for a *HYDROXYCINNAMOYL-COA SHIKIMATE TRANSFERASE* (*HCT*) (SI Appendix, Fig. S16-17). This gene has been investigated in the legume forage crop alfalfa (*Medicago sativa)*, where it was shown that *HCT* expression correlates negatively with nodule organogenesis (Shadle *et al.*, 2007; Gallego-Giraldo *et al.*, 2014). Therefore, we do not consider this duplication relevant for the nodulation capacity of *Parasponia*. Additionally, we identified three consistent gene losses in *Parasponia* among which is the ortholog of *EXOPOLYSACCHARIDE RECEPTOR 3* that in lotus inhibits infection of rhizobia with incompatible exopolysaccharides (Kawaharada *et al.*, 2015, 2017) (SI Appendix, Fig. S18-20, Table S7). Such gene losses may have contributed to effective rhizobium infection in *Parasponia* and their presence in *T. tomentosa* could explain the lack of intracellular infection in the F1 hybrid nodules. However, they cannot explain the dominance of nodule organogenesis in the F1 hybrid.

Contrary to our initial expectations, we discovered consistent loss or pseudogenization of seven symbiosis genes in *Trema* (SI Appendix, Fig. S21-23, Table S7). Based on our current sampling, these genes have a nodule-specific expression profile in *P. andersonii*, suggesting that they function exclusively in symbiosis (Fig. 4). Three of these are orthologs of genes that are essential for establishment of nitrogen-fixing nodules in legumes: *NIN, RPG,* and the LysM-type LCO receptor *NFP*/*NFR5*. In the case of *NFP/NFR5*, we found two close homologs of this gene, *NFP1* and *NFP2*, a duplication that predates the divergence of legumes and *Parasponia* (Fig. 5). In contrast to *NFP1*, *NFP2* is consistently pseudogenized in *Trema* species (Fig. 5; SI Appendix, Fig. S22-23). In an earlier study we used RNA interference (RNAi) to target *PanNFP1* (previously named *PaNFP*), which led to reduced nodule numbers and a block of intracellular infection by rhizobia as well as arbuscular mycorrhiza (Op den Camp *et al.*, 2011). However, we cannot rule out that the RNAi construct unintentionally also targeted *PanNFP2*, as both genes are ~70% identical in the 422 bp RNAi target region. Therefore, the precise functioning of both receptors in rhizobium and mycorrhizal symbiosis remains to be elucidated. Based on phylogenetic analysis the newly discovered *PanNFP2* is the ortholog of the legume *MtNFP/LjNFR5* genes encoding rhizobium LCO receptors required for nodulation, while *PanNFP1* is most likely a paralog (Fig. 5). Also, *PanNFP2* is significantly higher expressed in nodules than *PanNFP1* (SI Appendix, Fig. S25). Taken together, this indicates that PanNFP2 may represent a key LCO receptor required for nodulation in *Parasponia*.

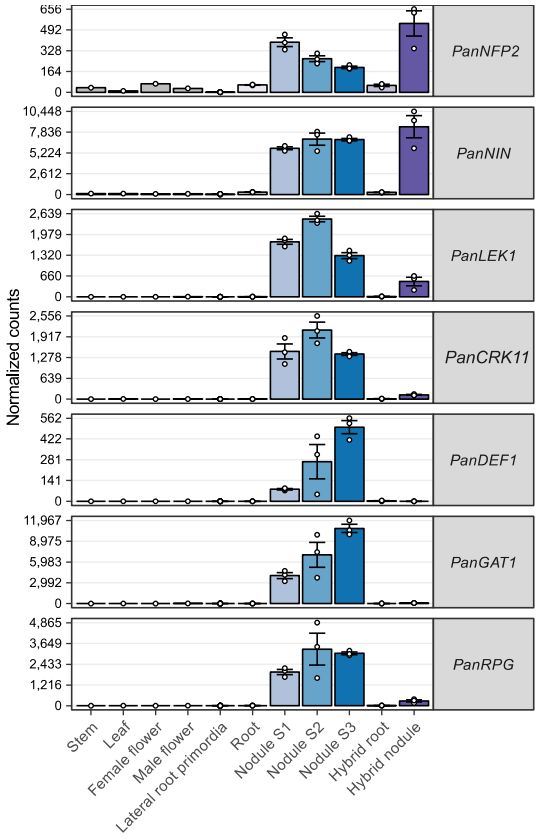
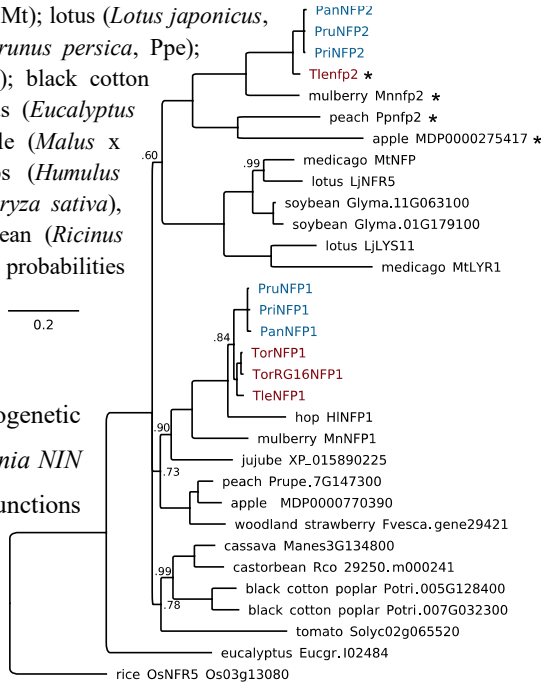


Fig. 4 Expression profile of *Parasponia andersonii* symbiosis genes that are lost in *Trema* species. Expression of symbiosis genes in *P. andersonii* stem, leaf, female and male flowers, lateral root primordia, roots and 3 nodule stages (S1-3), and in F_1 hybrid roots and nodules (line H9). Expression is given in DESeq2 normalized read counts, error bars represent standard error of three biological replicates for lateral root primordia, root, and nodule samples. Dots represent individual expression levels. *PanNFP2: NOD FACTOR PERCEPTION 2*; *PanNIN: NODULE INCEPTION*; *PanLEK1: LECTIN RECEPTOR KINASE 1*; *PanCRK11: CYSTEINE-RICH RECEPTOR KINASE 11*; *PanDEF1: DEFENSIN 1*; *PanRPG: RHIZOBIUM DIRECTED POLAR GROWTH.*

Contrary to our initial expectations, we discovered consistent loss or pseudogenization of seven symbiosis genes in *Trema* (SI Appendix, Fig. S21-23, Table S7). Based on our current sampling, these genes have a nodule-specific expression profile in *P. andersonii*, suggesting that they function exclusively in symbiosis (Fig. 4). Three of these are orthologs of genes that are essential for establishment of nitrogen-fixing nodules in legumes: *NIN, RPG,* and the LysM-type LCO receptor *NFP*/*NFR5*. In the case of *NFP/NFR5*, we found two close homologs of this gene, *NFP1* and *NFP2*, a duplication that predates the divergence of legumes and *Parasponia* (Fig. 5). In contrast to *NFP1*, *NFP2* is consistently pseudogenized in *Trema* species (Fig. 5; SI Appendix, Fig. S22-23). In an earlier study we used RNA interference (RNAi) to target *PanNFP1* (previously named *PaNFP*), which led to reduced nodule numbers and a block of intracellular infection by rhizobia as well as arbuscular mycorrhiza (Op den Camp *et al.*, 2011). However, we cannot rule out that the RNAi construct unintentionally also targeted *PanNFP2*, as both genes are ~70% identical in the 422 bp RNAi target region. Therefore, the precise functioning of both receptors in rhizobium and mycorrhizal symbiosis remains to be elucidated. Based on phylogenetic analysis the newly discovered *PanNFP2* is the ortholog of

the legume *MtNFP/LjNFR5* genes encoding rhizobium LCO receptors required for nodulation, while *PanNFP1* is most likely a paralog (Fig. 5). Also, *PanNFP2* is significantly higher expressed in nodules than *PanNFP1* (SI Appendix, Fig. S25). Taken together, this indicates that PanNFP2 may represent a key LCO receptor required for nodulation in *Parasponia*.

Fig. 5 *Parasponia NFP2* are putative orthologs of legume LCO receptors MtNFP/LjNFR5. Phylogenetic reconstruction of the NFP/NFR5 orthogroup based on kinase domain. Protein sequences deduced from pseudogenes are marked with an asterisk. Included species: *Parasponia andersonii* (Pan); *P. rigida* (Pri); *P. rugosa* (Pru); *Trema orientalis* RG33 (Tor); *T. orientalis* RG16 (TorRG16); *T. levigata* (Tle); medicago (*Medicago truncatula*, Mt); lotus (*Lotus japonicus*, Lj); soybean (*Glycine max*, Glyma); peach (*Prunus persica*, Ppe); woodland strawberry (*Fragaria vesca*, Fvesca); black cotton poplar (*Populus trichocarpa*, Potri); eucalyptus (*Eucalyptus grandis*, Eugr); jujube (*Ziziphus jujube*), apple (*Malus* x *domestica*), mulberry (*Morus notabilis*), hops (*Humulus lupulus*), cassave (*Manihot esculenta*), rice (*Oryza sativa*), tomato (*Solanum lycopersicum*), and castor bean (*Ricinus communis*). Node numbers indicate posterior probabilities below 1, scale bar represents substitutions per site. *Parasponia* proteins are marked in blue, *Trema* in red.



Based on expression profiles and phylogenetic relationships we postulate that also *Parasponia NIN* and *RPG* commit essential symbiotic functions similar as in other nodulating species (Fig. 3; SI Appendix, Fig. S25-28) (Schauser *et al.*, 1999; Borisov *et al.*, 2003; Marsh *et al.*, 2007; Arrighi *et al.*, 2008; Clavijo *et al.*, 2015). Compared with uninoculated roots, expression of *PanRPG* is >300 fold higher in *P. andersonii* nodules that become intracellularly infected (nodule stage 2), whereas in F_1 hybrid nodules, which are devoid of intracellular rhizobium infection- *PanRPG* this difference is less than 20-fold (Fig. 4). This suggests that PanRPG commits a function in rhizobium infection, similar as found in medicago (Arrighi *et al.*, 2008). The transcription factor NIN has been studied in several legume species as well as in the actinorhizal plant casuarina (*Casuarina glauca*) and in all cases shown to be essential for nodule organogenesis (Schauser *et al.*, 1999; Borisov *et al.*, 2003; Marsh *et al.*, 2007; Clavijo *et al.*, 2015). Loss of *NIN* and possibly *NFP2* in *Trema* species can explain the genetic dominance of nodule organogenesis in the *Parasponia* x *Trema* F1 hybrid plants.

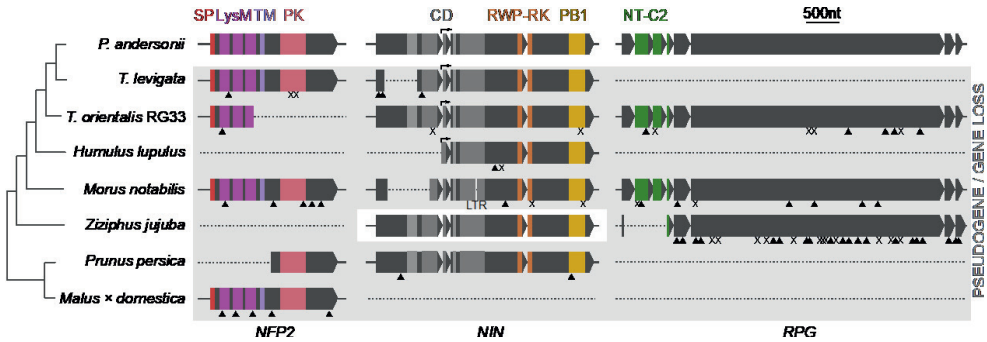


Fig. 6 Parallel loss of symbiosis genes in non-nodulating Rosales species. Pseudogenization or loss of *NFP2*, *NIN,* and *RPG* in two phylogenetically independent *Trema* lineages, *Humulus lupulus* (hop), *Morus notabilis* (mulberry), *Ziziphus jujuba* (jujube)*, Prunus persica* (peach)*,* and *Malus* x *domestica* (apple). In *H. lupulus NIN* is pseudogenized, whereas *NFP2* and *RPG* were not found (this may due to the low N50 of the publicly available assembly). In *Z. jujuba NFP2* is lost and *RPG* is pseudogenized, but *NIN* is intact. In *Fragaria vesca* all three genes are lost (not shown). Introns are indicated but not scaled. Triangles indicate frame-shifts; X indicates premature stop codons; LTR indicates long terminal repeat retrotransposon insertion (not scaled); arrows indicate alternative transcriptional start site in *NIN*. SP = signal peptide (red); LysM: 3 Lysin Motif domains (magenta); TM = transmembrane domain (lilac); PK = protein kinase (pink); CD = 4 conserved domains (grey); RWP-RK: conserved amino acid domain (orange); PB1 = Phox and Bem1 domain (yellow); NT-C2 = N-terminal C2 domain (green).

Next, we assessed whether loss of these symbiosis genes also occurred in more distant relatives of *Parasponia*. We analysed non-nodulating species representing 6 additional lineages of the Rosales clade; namely hop (*Humulus lupulus,* Cannabaceae) (Natsume *et al.*, 2015), mulberry (*Morus notabilis,* Moraceae) (He *et al.*, 2013), jujube (*Ziziphus jujuba*, Rhamnaceae) (Huang *et al.*, 2016), peach (*Prunus persica,* Rosaceae) (Verde *et al.*, 2013), woodland strawberry (*Fragaria vesca*, Rosaceae) (Shulaev *et al.*, 2011), and apple (*Malus x domestica,* Rosaceae) (Velasco *et al.*, 2010). This revealed a consistent pattern of pseudogenization or loss of *NFP2, NIN* and *RPG* orthologs; the intact jujube *ZjNIN* being the only exception (Fig. 6). We note that for peach *NIN* was previously annotated as protein-coding gene (Verde *et al.*, 2013). However, based on comparative analysis of conserved exon structures we found two out-of-frame mutations (SI Appendix, Fig. S28). We therefore conclude that also in peach the *NIN* gene is pseudogenized. Because the pseudogenized symbiosis genes are largely intact in most of these species and differ in their deleterious mutations, the loss of function of these essential symbiosis genes should have occurred relatively recently and in parallel in at least seven Rosales lineages.

Discussion

Here we present the nodulating non-legume *Parasponia* as a comparative system to obtain insights in molecular genetic changes underlying evolution of nitrogen-fixing root nodules. We show that nodulation is a genetically dominant trait and that *P. andersonii* and the legume medicago share a set of 290 genes that have a nodule enhanced expression profile. Among these are *NIN* and *RPG,* two genes that in legumes are essential for nitrogen-fixing root nodulation (Schauser *et al.*, 1999; Borisov *et al.*, 2003; Marsh *et al.*, 2007; Arrighi *et al.*, 2008). Both these genes as well as a putative ortholog of the NFP/NFR5-type LysM receptor for rhizobium LCO signal molecules -named *NFP2* in *Parasponia*- are consistently pseudogenized or lost in *Trema* and other non-nodulating species of the Rosales order. This challenges the current view on the evolution of nitrogen-fixing plant-microbe symbioses.

Evolution of nodulation is generally viewed as a two-step process: first an unspecified predisposition event in the ancestor of all nodulating species, bringing species in the nitrogen-fixing clade to a precursor state for nodulation. Subsequently, nodulation originated in parallel; eight times with *Frankia* and twice with rhizobium (Soltis *et al.*, 1995; Swensen, 1996; Doyle, 1998, 2011, 2016; Werner *et al.*, 2014; Li *et al.*, 2015; Martin *et al.*, 2017). This hypothesis is most parsimonious and suggests a minimum number of independent gains and losses of symbiosis. Based on this hypothesis it is currently assumed that non-host relatives of nodulating species are generally in a precursor state for nodulation (Werner *et al.*, 2014).

Our results are difficult to explain under the hypothesis of parallel origins of nodulation. The functions of *NFP2, NIN,* and *RPG* currently cannot be linked to any non-symbiotic processes. Therefore it remains obscure why these symbiosis genes were maintained over an extended period of time in non-nodulating plant species, and were subsequently independently lost. Additionally, the hypothesis of parallel origins of nodulation would imply convergent recruitment of at least 290 genes to commit symbiotic functions in *Parasponia* and legumes. Because these 290 genes encode proteins with various predicted functions (e.g. from extracellular signaling receptors to sugar transporters; Dataset S5), as well as comprise at least two different developmental expression patterns (nodule organogenesis and intracellular infection and/or fixation; Fig. 2, Dataset S5), this would imply parallel evolution of a genetically complex trait.

Alternatively, the parallel loss of symbiosis genes in non-nodulating plants can be interpreted as parallel loss of nodulation (Soltis *et al.*, 1995). Under this hypothesis nodulation possibly

evolved only once in an ancestor of the nitrogen-fixing clade. Subsequently, nodulation was lost in most descendant lineages. This single gain-massive loss hypothesis fits our data better in two ways. First, a single gain explains the origin of the conserved set of at least 290 symbiosis genes utilized by both *Parasponia* and medicago, because they then result from the same ancestral recruitment event. Second, it more convincingly explains the parallel loss of symbiosis genes in non-nodulating plants, because then gene loss correlates directly with loss of nodulation. Additionally, the single gain-massive loss model eliminates the predisposition event, a theoretical concept that currently cannot be addressed experimentally. We therefore favor this alternative hypothesis over the currently most widely held assumption of parallel origins of nodulation.

Loss of nodulation is not controversial, as it is generally considered to have occurred at least 20 times in the legume family (Werner *et al.*, 2014; Li *et al.*, 2015). Nevertheless, the single gain-massive loss hypothesis implies many more evolutionary events than the current hypothesis of parallel gains. On the other hand, it is conceptually easier to lose a complex trait, such as nodulation, rather than to gain it (Doyle, 2016). Genetic studies in legumes demonstrated that nitrogen-fixing symbioses can be abolished by a single knockout mutation in tens of different genes, among which are *NFP/NFR5, NIN*, and *RPG* (Dataset S1). Because parsimony implies equal weights for gains and losses, it may therefore not be the best way to model the evolution of nodulation.

Preliminary support for the single gain-massive loss hypothesis can be found in fossil records. Putative root nodule fossils have been discovered from the late Cretaceous (approximately 84 million years ago), which corroborates our hypothesis that nodulation is much older than is generally assumed (Herendeen *et al.*, 1999). Legumes are the oldest and most diverse nodulating lineage, however the earliest fossils that can be definitively assigned to the legume family appeared in the late Palaeocene (approximately 65 million years ago) (Bruneau *et al.*, 2008). Notably, the age of the nodule fossils coincides with the early diversification of the nitrogen-fixing clade that has given rise to the 4 orders Fabales, Rosales, Cucurbitales, and Fagales (Li *et al.*, 2015). As it is generally agreed that individual fossil ages provide minimum bounds for dates of origins it is therefore not unlikely that the last common ancestor of the nitrogen-fixing clade was a nodulator.

Clearly, the single gain-massive loss hypothesis that is supported by our comparative studies using *Parasponia* requires further substantiation. First, the hypothesis implies that many

ancestral species in the nitrogen-fixing clade were able to nodulate. This should be further supported by fossil evidence. Second, the hypothesis implies that actinorhizal plant species maintained *NIN, RPG,* and possibly *NFP2* (the latter only in case LCOs are used as symbiotic signal) (Nguyen *et al.*, 2016). Third, these genes should be essential for nodulation in these actinorhizal plants as well as in *Parasponia*. This can be shown experimentally, as was done for *NIN* in casuarina (Clavijo *et al.*, 2015).

Loss of symbiosis genes in non-nodulating plant species is not absolute, as we observed a functional copy of *NIN* in jujube. This pattern is similar to the pattern of gene loss in species that lost endomycorrhizal symbiosis where occasionally endomycorrhizal symbiosis genes have been maintained in non-mycorrhizal plants (Delaux *et al.*, 2015; Kamel *et al.*, 2017). Conservation of *NIN* in jujube suggests that this gene has a non-symbiotic function. Contrary to *NFP2*, which is the result of a gene duplication near the origin of the nitrogen-fixing clade, functional copies of *NIN* are also present in species outside the nitrogen-fixing clade (SI Appendix, Fig. S26). This suggests that these genes may have retained -at least in part- an unknown ancestral non-symbiotic function in some lineages within the nitrogen-fixing clade. Alternatively, *NIN* may have acquired a new non-symbiotic function within some lineages in the nitrogen-fixing clade.

As hemoglobin is crucial for rhizobium symbiosis in legumes (Ott *et al.*, 2005), it is striking that *Parasponia* and legumes do not use orthologous copies of hemoglobin genes in their nodules (Sturms *et al.*, 2010). Superficially, this seems inconsistent with a single gain of nodulation. However, hemoglobin is not crucial for all nitrogen-fixing nodule symbioses because several *Frankia* microsymbionts possess intrinsic physical characteristics to protect the Nitrogenase enzyme for oxidation (Winship *et al.*, 1987; Silvester *et al.*, 1990, 2007; Silvester & Winship, 1990). In line with this, *Ceanothus* spp. (Rhamnaceae, Rosales) - which represent actinorhizal nodulating relatives of *Parasponia* - do not express a hemoglobin gene in their *Frankia*-infected nodules (Silvester *et al.*, 1990, 2007; Silvester & Winship, 1990). Consequently, hemoglobins may have been recruited in parallel after the initial gain of nodulation as parallel adaptations to rhizobium microsymbionts. Based on the fact that *Parasponia* acquired lineage-specific adaptations in HB1 that are considered to be essential for controlling oxygen homeostasis in rhizobium root nodules (Sturms *et al.*, 2010; Kakar *et al.*, 2011), a symbiont switch from *Frankia* to rhizobium may have occurred recently in an ancestor of the *Parasponia* lineage.

Our study provides novel leads for attempts to engineer nitrogen-fixing root nodules in agricultural crop plants. Such a translational approach is anticipated to be challenging (Rogers & Oldroyd, 2014), and the only published attempt so far, describing transfer of 8 LCO signaling genes, was unsuccessful (Untergasser *et al.*, 2012). Our results suggest that transfer of symbiosis genes may not be sufficient to obtain functional nodules. Even though F1 hybrid plants contain a full haploid genome complement of *P. andersonii* they lack intracellular infection. This may be due to haploinsufficiency of *P. andersoni*i genes in the F1 hybrid, or due to an inhibitory factor in *T. tomentosa*. For example, inhibition of intracellular infection may be the result of a dominant negative factor, or the result of heterozygosity negatively affecting the formation of e.g. LysM receptor complexes required for appropriate perception of microsymbionts. Such factors may also be present in other non-host species. Consequently, engineering nitrogen-fixing nodules may require gene knockouts in non-nodulating plants to overcome inhibition of intracellular infection. *Trema* may be the best candidate species for such a (re)engineering approach, due to its high genetic similarity with *Parasponi*a and the availability of transformation protocols (Cao *et al.*, 2012). Therefore, the *Parasponia-Trema* comparative system may not only be suited for evolutionary studies, but also can form an experimental platform to obtain essential insights for engineering nitrogen-fixing root nodules.

Materials and Methods

Parasponia - *Trema* intergeneric crossing and hybrid genotyping

Parasponia and *Trema* are wind-pollinated species. A female-flowering *P. andersonii* individual WU1.14 was placed in a plastic shed together with a flowering *T. tomentosa* WU10 plant. Putative F_1 hybrid seeds were germinated (see SI Appendix, Supplementary Methods) and transferred to potting soil. To confirm the hybrid genotype a PCR marker was used that visualizes a length difference in the promoter region of *LIKE-AUXIN 1* (*LAX1*) (primers: LAX1-f: ACATGATAATTTGGGCATGCAACA, LAX1-r: TCCCGAATTTTCTACGAATTGAAA, amplicon size *P. andersonii*: 974 bp; *T. tomentosa*: 483 bp). Hybrid plant H9 was propagated *in vitro* (Davey *et al.*, 1993; Op den Camp *et al.*, 2011). The karyotype of the selected plants was determined according to Geurts and de Jong (Geurts & de Jong, 2013).

Assembly of reference genomes

Cleaned DNA sequencing reads were *de novo* assembled using ALLPATHS-LG (release 48961) (Gnerre *et al.*, 2011) After filtering of any remaining adapters and contamination, contigs were scaffolded with two rounds of SSPACE-standard (v3.0) (Boetzer *et al.*, 2011) with the mate-pair libraries using default settings. We used the output of the second run of SSPACE scaffolding as the final assembly (See SI Appendix, Supplementary Methods for full details and parameter choices). Validation of the final assemblies showed that 90-100% of the genomic reads mapped back to the assemblies (SI Appendix, Table S4), and 94-98% of CEGMA (Parra *et al.*, 2007) and BUSCO (Simão *et al.*, 2015) genes were detected (SI Appendix, Table S5).

Annotation of reference genomes

Repetitive elements were identified following the standard Maker-P recipe (http://weatherby.genetics.utah.edu/MAKER/wiki/index.php/Repeat_Library_Construction-Advanced accessed October 2015) as described on the GMOD site: (i) RepeatModeler with Repeatscout v1.0.5, Recon v1.08, RepeatMasker version open4.0.5, using RepBase version 20140131 (Bao *et al.*, 2015) and TandemRepeatFinder; (ii) GenomeTools: LTRharvest and LTRdigest (Gremme *et al.*, 2013); (iii) MITEhunter with default parameters (Han & Wessler, 2010). We generated species-specific repeat libraries for both *P. andersonii* and *T. orientalis*

separately and combined these into a single repeat library, filtering out sequences that are >98% similar. We masked both genomes using RepeatMasker with this shared repeat library.

To aid the structural annotation we used 11 *P. andersonii* and 6 *T. orientalis* RNA sequencing datasets (SI Appendix, Table S8). All RNA-seq samples were assembled *de novo* using genome-guided Trinity (Grabherr *et al.*, 2011), resulting in one combined transcriptome assembly per species. In addition all samples were mapped to their respective reference genomes using BWA-MEM and processed into putative transcripts using cufflinks (Trapnell *et al.*, 2010) and transdecoder (Haas *et al.*, 2013). As protein homology evidence, only Swiss-Prot (UniProt Consortium, 2015) entries filtered for plant proteins were used. This way we only included manually verified protein sequences and prevented the incorporation of erroneous predictions. Finally, four gene-predictor tracks were used: 1) SNAP (Korf, 2004), trained on *P. andersonii* transdecoder transcript annotations; 2) SNAP, trained on *T. orientalis* transdecoder transcript annotations; 3) Augustus (Stanke *et al.*, 2008) as used in the BRAKER pipeline, trained on RNA-seq alignments (Hoff *et al.*, 2016); 4) GeneMark-ET as used in the BRAKER pipeline, trained on RNA-seq alignments (Lomsadze *et al.*, 2014).

First, all evidence tracks were processed by Maker-P (Campbell *et al.*, 2014). The results were refined with EVidenceModeler (EVM) (Haas *et al.*, 2008), which was used with all the same tracks as Maker-P, except for the Maker-P blast tracks and with the addition of the Maker-P consensus track as additional evidence. Ultimately, EVM gene models were preferred over Maker-P gene models, except when there was no overlapping EVM gene model. Where possible, evidence of both species was used to annotate each genome (i.e. *de novo* RNA-seq assemblies of both species were aligned to both genomes).

To take maximum advantage of annotating two highly similar genomes simultaneously we developed a custom reconciliation procedure involving whole genome alignments. The consensus annotations from merging the EVM and Maker-P annotations were transferred to their respective partner genome using nucmer (Kurtz *et al.*, 2004) and RATT revision 18 (Otto *et al.*, 2011) (i.e. the *P. andersonii* annotation was transferred to *T. orientalis* and *vice versa*), based on nucmer whole genome alignments (SI Appendix, Fig. S10). Through this reciprocal transfer, both genomes had two candidate annotation tracks. This allowed for validation of annotation differences between *P. andersonii* and *T. orientalis*, reduced technical variation, and consequently improved all downstream analyses. After automatic annotation and reconciliation 1,693 *P. andersonii* genes and 1,788 *T. orientalis* genes were manually curated.

These were mainly homologs of legume symbiosis genes and genes that were selected based on initial data exploration.

To assign putative product names to the predicted genes we combined BLAST results against Swiss-Prot, TrEMBL, and nr with InterProScan results (custom script). To annotate Gene ontology (GO) terms and kyoto encyclopedia of Genes and Genomes (KEGG) enzyme codes we used Blast2GO based on the nr BLAST results and InterProScan results. Finally, we filtered all gene models with hits to InterPro domains that are specific to repetitive elements.

Orthogroup inference

To determine relationships between *P. andersonii* and *T. orientalis* genes, as well as with other plant species we inferred orthogroups with OrthoFinder version 0.4.0 (Emms & Kelly, 2015). Since orthogroups are defined as the set of genes that are descended from a single gene in the last common ancestor of all the species being considered, they can comprise orthologous as well as paralogous genes. Our analysis included proteomes of selected species from the Eurosid clade: *Arabidopsis thaliana* TAIR10 (Brassicaceae, Brassicales) (Swarbreck *et al.*, 2008) and *Eucalyptus grandis* v2.0 (Myrtaceae, Myrtales) from the Malvid clade (Myburg *et al.*, 2014); *Populus trichocarpa* v3.0 (Salicaeae, Malpighiales) (Tuskan *et al.*, 2006), legumes *Medicago truncatula* Mt4.0v1 (Young *et al.*, 2011) and *Glycine max* Wm82.a2.v1 (Fabaceae, Fabales) (Schmutz *et al.*, 2010), *Fragaria vesca* v1.1 (Rosaceae, Rosales) (Shulaev *et al.*, 2011), *P. andersonii* and *T. orientalis* (Cannabaceae, Rosales) from the Fabid clade (Dataset S2). Sequences were retrieved from phytozome (www.phytozome.net).

Gene copy number variant detection

To assess orthologous and paralogous relationships between *Parasponia* and *Trema* genes, we inferred phylogenetic gene trees for all 21,959 orthogroups comprising *Parasponia* and/or *Trema* genes using the neighbor-joining clustering algorithm (Saitou & Nei, 1987). Based on these gene trees, for each *Parasponia* gene its relationship to other *Parasponia* and *Trema* genes was defined as follows. 1) orthologous pair: the sister lineage is a single gene from the *Trema* genome suggesting that they are the result of a speciation event; 2) inparalog: the sister lineage is a gene from the *Parasponia* genome, suggesting that they are the result of a gene duplication event; 3) singleton: the sister lineage is a gene from a species other than *Trema*, suggesting that the *Trema* gene was lost; 4) multi-ortholog: the sister lineage comprises multiple genes from the *Trema* genome, suggesting that the latter are inparalogs. For each

Trema gene, relationship was defined in the same way but with respect to the *Parasponia* genome (SI Appendix, Table S6). Because phylogenetic analysis relies on homology we assessed the level of conservation in the multiple-sequence alignments by calculating the trident score using MstatX (https://github.com/gcollet/MstatX) (Valdar, 2002). Orthogroups with a score below 0.1 were excluded from the analysis. Examination of orthogroups comprising >20 inparalogs revealed that some represented repetitive elements; these were also excluded. Finally, orthologous pairs were validated based on the whole-genome alignments used in the annotation reconciliation.

Nodule-enhanced genes

To assess gene expression in *Parasponia* nodules, RNA was sequenced from the three nodule stages described above as well as uninoculated roots (SI Appendix, Table S8). RNA-seq reads were mapped to the *Parasponia* reference genome with HISAT2 version 2.02 (Kim *et al.*, 2015) using an index that includes exon and splice site information in the RNA-seq alignments. Mapped reads were assigned to transcripts with featureCounts version 1.5.0 (Liao *et al.*, 2014). Normalization and differential gene expression were performed with DESeq2. Nodule enhanced genes were selected based on >2.0 fold-change and p<=0.05 in any nodule stage compared with uninoculated root controls. Genes without functional annotation or orthogroup membership or from orthogroups with low alignment scores (<0.1 trident score, see above) or representing repetitive elements were excluded from further analysis. To assess expression of *Parasponia* genes in the hybrid nodules, RNA was sequenced from nodules and uninoculated roots. Here, RNA-seq reads were mapped to a combined reference comprising two parent genomes from *P. andersonii* and *T. tomentosa*. To assess which genes are nodule-enhanced in medicago we re-analyzed published RNA-seq read data from Roux *et al.* [archived at the National Center for Biotechnology Information (NCBI) under sequence read archive (SRA) study ID code SRP028599] (Roux *et al.*, 2014). To assess which of these genes may be co-opted from the ancient and widespread arbuscular mycorrhizal symbiosis we generated a set of 575 medicago genes induced upon mycorrhization in medicago by re-analyzing published RNA-seq read data from Afkhami and Stinchcombe (archived at NCBI under SRA study SRP078249) (Afkhami & Stinchcombe, 2016). Both medicago data sets were analysed as described above for *Parasponia* but using the medicago genome and annotation version 4.0v2 as reference (Young *et al.*, 2011).

To assess common recruitment of genes in nodules from *Parasponia* and medicago we counted orthogroups comprising both *P. andersonii* and medicago nodule-enhanced genes. To assess whether this number is higher than expected by chance we performed the hypergeometric test as well as three different permutation tests where we randomized either the *Parasponia* gene set, the medicago gene set, or both sets with 10,000 permutations. We then determined putative orthology between the *Parasponia* and medicago genes within the common orthogroups based on phylogenetic analysis. *Parasponia* and medicago genes were considered putative orthogroups if they occur in the same subclade with more than 50% bootstrap support; otherwise they were considered close homologs.

Availability of data and materials

The data reported in this study are tabulated in Datasets S1-S7 and SI Appendix; sequence data are archived at NCBI (https://www.ncbi.nlm.nih.gov) under BioProject numbers PRJNA272473 and PRJNA272482; draft genome assemblies, phylogenetic datasets, and orthogroup data are archived at the Dryad Digital Repository: https://doi.org/10.5061/dryad.fq7gv88. All analyzed data can also be browsed or downloaded through a web portal on www.parasponia.org. All custom scripts and code are available on https://github.com/holmrenser/parasponia_code.

Supplemental data

Supplemental data belonging to this chapter are available at *Proc Natl Acad Sci* USA online (https://doi.org/10.1073/pnas.1721395115).

Acknowledgments

We thank Shelley James, Thomas Marler, Giles Oldroyd, and Johan van Valkenburg for providing germplasm and Ries de Visser (IsoLife) for supporting acetylene reduction assays.

References

Afkhami ME, Stinchcombe JR. **2016**. Multiple mutualist effects on genomewide expression in the tripartite association between *Medicago truncatula*, nitrogen-fixing bacteria and mycorrhizal fungi. *Molecular ecology* **25**: 4946–4962.

Akkermans ADL, Abdulkadir S, Trinick MJ. **1978**. Nitrogen-fixing root nodules in Ulmaceae. *Nature* **274**: 190–190.

Arrighi JF, Barre A, Ben Amor B, Bersoult A, Soriano LC, Mirabella R, de Carvalho-Niebel F, Journet EP, Gherardi M, Huguet T, *et al.* 2006. The *Medicago truncatula* lysine motif-receptor-like kinase gene family includes *NFP* and new nodule-expressed genes. *Plant physiology* **142**: 265–279.

Arrighi J-F, Godfroy O, de Billy F, Saurat O, Jauneau A, 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.

Bao W, Kojima KK, Kohany O. **2015**. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**: 11.

Baudin M, Laloum T, Lepage A, Rípodas C, Ariel F, Frances L, Crespi M, Gamas P, Blanco FA, Zanetti ME, *et al.* 2015. A Phylogenetically conserved group of Nuclear Factor-Y transcription factors interact to control nodulation in legumes. *Plant physiology* **169**: 2761–2773.

Becking JH. **1992**. The rhizobium symbiosis of the nonlegume *Parasponia*. In: Stacey G, Burris RH, Evans HJ, eds. Biological nitrogen fixation. New York: Routledge, Chapman and Hall, 497–559.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. **2011**. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**: 578–579.

Borisov AY, Madsen LH, Tsyganov VE, Umehara Y, Voroshilova VA, Batagov AO, Sandal N, Mortensen A, Schauser L, Ellis N, *et al.* 2003. The *Sym35* gene required for root nodule development in pea is an ortholog of *Nin* from *Lotus japonicus*. *Plant physiology* **131**: 1009–1017.

Broghammer A, Krusell L, Blaise M, Sauer J, Sullivan JT, Maolanon N, Vinther M, Lorentzen A, Madsen EB, Jensen KJ, *et al.* 2012. Legume receptors perceive the rhizobial lipochitin oligosaccharide signal molecules by direct binding. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 13859–13864.

Bruneau A, Mercure M, Lewis GP, Herendeen PS. **2008**. Phylogenetic patterns and diversification in the caesalpinioid legumes. *Botany* **86**: 697–718.

Burrill TJ, Hansen R. **1917**. Is symbiosis possible between legume bacteria and non-legume plants? *Bulletin - University of Illinois Agricultural Experiment Station* **202**: 115–181.

Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, Lei J, Achawanantakun R, Jiao D, Lawrence CJ, *et al.* 2014. MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant physiology* **164**: 513–524.

Cao Q, Op den Camp R, Kalhor MS, Bisseling T, Geurts R. **2012**. Efficiency of *Agrobacterium rhizogenes*–mediated root transformation of *Parasponia* and *Trema* is temperature dependent. *Plant growth regulation* **68**: 459–465.

Chiasson DM, Loughlin PC, Mazurkiewicz D, Mohammadidehcheshmeh M, Fedorova EE, Okamoto M, McLean E, Glass ADM, Smith SE, Bisseling T, *et al.* 2014. Soybean SAT1 (Symbiotic Ammonium Transporter 1) encodes a bHLH transcription factor involved in nodule growth and NH4+ transport. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 4814–4819.

Clason EW. **1936**. The vegetation of the upper-Badak region of mount Kelut (East Java). *Bulletin du Jardin botanique de Buitenzorg, sér. 3* **13**: 509–518.

Clavijo F, Diedhiou I, Vaissayre V, Brottier L, Acolatse J, Moukouanga D, Crabos A, Auguy F, Franche C, Gherbi H, *et al.* 2015. The *Casuarina NIN* gene is transcriptionally activated throughout *Frankia* root infection as well as in response to bacterial diffusible signals. *The New phytologist* **208**: 887–903.

Combier J-PP, Frugier F, Billy FD, Boualem A, El-yahyaoui F, Moreau S, Vernié T, Ott T, Gamas P, Crespi M, *et al.* 2006. MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. *Genes & development* **20**: 3084–3088.

Davey MR, Webster G, Manders G, Ringrose FL, Power JB, Cocking EC. **1993**. Effective nodulation of micro-propagated shoots of the non-legume *Parasponia andersonii* by *Bradyrhizobium*. *Journal of experimental botany* **44**: 863–867.

Deguchi Y, Banba M, Shimoda Y, Chechetka SA, Suzuri R, Okusako Y, Ooki Y, Toyokura K, Suzuki A, Uchiumi T, *et al.* 2007. Transcriptome profiling of *Lotus japonicus* roots during arbuscular mycorrhiza development and comparison with that of nodulation. *DNA research: an international journal for rapid publication of reports on genes and genomes* **14**: 117–133.

Delaux P-M, Radhakrishnan GV, Jayaraman D, Cheema J, Malbreil M, Volkening JD, Sekimoto H, Nishiyama T, Melkonian M, Pokorny L, *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.

Doyle JJ. **1998**. Phylogenetic perspectives on nodulation: evolving views of plants and symbiotic bacteria. *Trends in plant science* **3**: 473–478.

Doyle JJ. **2011**. Phylogenetic perspectives on the origins of nodulation. *Molecular plant-microbe interactions: MPMI* **24**: 1289–1295.

Doyle JJ. **2016**. Chasing unicorns: Nodulation origins and the paradox of novelty. *American journal of botany* **103**: 1865–1868.

Emms DM, Kelly S. **2015**. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome biology* **16**: 157.

Gallego-Giraldo L, Bhattarai K, Pislariu CI, Nakashima J, Jikumaru Y, Kamiya Y, Udvardi MK, Monteros MJ, Dixon RA. **2014**. Lignin modification leads to increased nodule numbers in alfalfa. *Plant physiology* **164**: 1139–1150.

Geurts R, de Jong H. **2013**. Fluorescent In Situ Hybridization (FISH) on pachytene chromosomes as a tool for genome characterization. *Methods in molecular biology* **1069**: 15–24.

Geurts R, Xiao TT, Reinhold-Hurek B. **2016**. What does it take to evolve a nitrogen-fixing endosymbiosis? *Trends in plant science* **21**: 199–208.

Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, *et al.* 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 1513–1518.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* **29**: 644–652.

Granqvist E, Sun J, Op den Camp R, Pujic P, Hill L, Normand P, Morris RJ, Downie JA, Geurts R, Oldroyd GED. **2015**. Bacterial-induced calcium oscillations are common to nitrogen-fixing associations of nodulating legumes and non-legumes. *The New phytologist* **207**: 551–558.

Gremme G, Steinbiss S, Kurtz S. **2013**. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM* **10**: 645–656.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, *et al.* 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* **8**: 1494–1512.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. **2008**. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome biology* **9**: R7.

Han Y, Wessler SR. **2010**. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic acids research* **38**: e199.

Herendeen PS, Magallon-Puebla S, Lupia R, Crane PR, Kobylinska J. **1999**. A preliminary conspectus of the Allon flora from the Late Cretaceous (late Santonian) of central Georgia, USA. *Annals of the Missouri Botanical Garden. Missouri Botanical Garden*: 407–471.

He N, Zhang C, Qi X, Zhao S, Tao Y, Yang G, Lee T-H, Wang X, Cai Q, Li D, *et al.* 2013. Draft genome sequence of the mulberry tree *Morus notabilis*. *Nature communications* **4**: ncomms3445.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. **2016**. BRAKER1: Unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**: 767–769.

Horváth B, Yeun LH, Domonkos Á, Halász G, Gobbato E, Ayaydin F, Miró K, Hirsch S, Sun J, Tadege M, *et al.* 2011. Medicago truncatula IPD3 is a member of the common symbiotic signaling pathway required for rhizobial and mycorrhizal symbioses. *Molecular plant-microbe interactions: MPMI* **24**: 1345–1358.

Huang J, Zhang C, Zhao X, Fei Z, Wan K, Zhang Z, Pang X, Yin X, Bai Y, Sun X, *et al.* 2016. The jujube genome provides insights into genome evolution and the domestication of sweetness/acidity taste in fruit trees. *PLoS genetics* **12**: e1006433.

Kakar S, Sturms R, Tiffany A, Nix JC, Dispirito AA, Hargrove MS. **2011**. Crystal structures of *Parasponia* and *Trema* hemoglobins: Differential heme coordination is linked to quaternary structure. *Biochemistry* **50**: 4273–4280.

Kamel L, Keller-Pearson M, Roux C, Ané J-M. **2017**. Biology and evolution of arbuscular mycorrhizal symbiosis in the light of genomics. *The New phytologist* **213**: 531–536.

Kawaharada Y, Kelly S, Nielsen MW, Hjuler CT, Gysel K, Muszyński A, Carlson RW, Thygesen MB, Sandal N, Asmussen MH, *et al.* 2015. Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* **523**: 308–312.

Kawaharada Y, Nielsen MW, Kelly S, James EK, Andersen KR, Rasmussen SR, Füchtbauer W, Madsen LH, Heckmann AB, Radutoiu S, *et al.* 2017. Differential regulation of the Epr3 receptor coordinates membrane-restricted rhizobial colonization of root nodule primordia. *Nature communications* **8**: 14534.

Kim D, Langmead B, Salzberg SL. **2015**. HISAT: a fast spliced aligner with low memory requirements. *Nature methods* **12**: 357–360.

Kistner C, Winzer T, Pitzschke A, Mulder L, Sato S, Kaneko T, Tabata S, Sandal N, Stougaard J, Webb KJ, *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.

Korf I. **2004**. Gene finding in novel genomes. *BMC bioinformatics* **5**: 59.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. **2004**. Versatile and open software for comparing large genomes. *Genome biology* **5**: R12.

Liao Y, Smyth GK, Shi W. **2014**. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930.

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

Li H-L, Wang W, Mortimer PE, Li R-Q, Li D-Z, Hyde KD, Xu J-C, Soltis DE, Chen Z-D. **2015**. Large-scale phylogenetic analyses reveal multiple gains of actinorhizal nitrogen-fixing symbioses in angiosperms associated with climate change. *Scientific reports* **5**: 14023.

Lomsadze A, Burns PD, Borodovsky M. **2014**. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic acids research* **42**: e119.

Madsen EB, Madsen LH, Radutoiu S, Sato S, Kaneko T, Tabata S, Sandal N. **2003**. A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. *Nature* **425**: 637–640.

Marsh JF, Rakocevic A, Mitra RM, Brocard L, Sun J, Eschstruth A, Long SR, Schultze M, Ratet P, Oldroyd GEDD. **2007**. *Medicago truncatula* NIN is essential for rhizobial-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase. *Plant physiology* **144**: 324–335.

Martin FM, Uroz S, Barker DG. **2017**. Ancestral alliances: Plant mutualistic symbioses with fungi and bacteria. *Science* **356**.

Marvel DJ, Torrey JG, Ausubel FM. **1987**. *Rhizobium* symbiotic genes required for nodulation of legume and nonlegume hosts. *Proceedings of the National Academy of Sciences of the United States of America* **84**: 1319–1323.

Murray JD, Muni RRD, Torres-Jerez I, Tang Y, Allen S, Andriankaja M, Li G, Laxmi A, Cheng X, Wen J, *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. *The Plant journal: for cell and molecular biology* **65**: 244–252.

Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D, *et al.* 2014. The genome of *Eucalyptus grandis*. *Nature* **510**: 356–362.

Natsume S, Takagi H, Shiraishi A, Murata J, Toyonaga H, Patzak J, Takagi M, Yaegashi H, Uemura A, Mitsuoka C, *et al.* 2015. The draft genome of hop (*Humulus lupulus*), an essence for brewing. *Plant & cell physiology* **56**: 428–441.

Nguyen TV, Wibberg D, Battenberg K, Blom J, Vanden Heuvel B, Berry AM, Kalinowski J, Pawlowski K. **2016**. An assemblage of *Frankia* Cluster II strains from California contains the canonical nod genes and also the sulfotransferase gene nodH. *BMC genomics* **17**: 796.

Oldroyd GED. **2013**. Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nature reviews. Microbiology* **11**: 252–263.

Op den Camp RHM, Polone E, Fedorova E, Roelofsen W, Squartini A, Op den Camp HJM, Bisseling T, Geurts R. **2012**. Nonlegume *Parasponia andersonii* deploys a broad rhizobium host range strategy resulting in largely variable symbiotic effectiveness. *Molecular plant-microbe interactions: MPMI* **25**: 954–963.

Op den Camp R, Streng A, De Mita S, Cao Q, Polone E, Liu W, Ammiraju JSS, Kudrna D, Wing R, Untergasser A, *et al.* 2011. LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume *Parasponia*. *Science* **331**: 909–912.

Ott T, van Dongen JT, Günther C, Krusell L, Desbrosses G, Vigeolas H, Bock V, Czechowski T, Geigenberger P, Udvardi MK. **2005**. Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. *Current biology: CB* **15**: 531–535.

Otto TD, Dillon GP, Degrave WS, Berriman M. **2011**. RATT: Rapid annotation transfer tool. *Nucleic acids research* **39**: e57.

Parniske M. **2008**. Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nature reviews. Microbiology* **6**: 763–775.

Parra G, Bradnam K, Korf I. **2007**. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**: 1061–1067.

Pumplin N, Mondo SJ, Topp S, Starker CG, Gantt JS, Harrison MJ. **2010**. Medicago truncatula Vapyrin is a novel protein required for arbuscular mycorrhizal symbiosis. *The Plant journal: for cell and molecular biology* **61**: 482–494.

Radutoiu S, Madsen LH, Madsen EB, Felle HH, Umehara Y, Gronlund M, Sato S, Nakamura Y, Tabata S, Sandal N, *et al.* 2003. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* **425**: 585–592.

Rogers C, Oldroyd GED. 2014. Synthetic biology approaches to engineering the nitrogen symbiosis in cereals. *Journal of experimental botany* **65**: 1939–1946.

Roux B, Rodde N, Jardinaud M-F, Timmers T, Sauviac L, Cottret L, Carrère S, Sallet E, Courcelle E, Moreau S, *et al.* 2014a. An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *The Plant journal: for cell and molecular biology* **77**: 817–837.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* **4**: 406–425.

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

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, *et al.* 2010. Genome sequence of the palaeopolyploid soybean. *Nature* **463**: 178–183.

Shadle G, Chen F, Srinivasa Reddy MS, Jackson L, Nakashima J, Dixon RA. 2007. Down-regulation of hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase in transgenic alfalfa affects lignification, development and forage quality. *Phytochemistry* **68**: 1521–1529.

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, *et al.* 2011. The genome of woodland strawberry (*Fragaria vesca*). *Nature genetics* **43**: 109–116.

Silvester WB, Berg RH, Schwintzer CR, Tjepkema JD. 2007. Oxygen responses, hemoglobin, and the structure and function of vesicles. In: Pawlowski K, Newton WE, eds. Nitrogen Fixation: Origins, Applications, and Research Progress. Nitrogen-fixing actinorhizal symbioses. Springer Netherlands, 105–146.

Silvester WB, Harris SL, Tjepkema JD. 1990. Oxygen regulation and hemoglobin. In: Schwintzer CR, Tjepkema JD, eds. The biology of *Frankia* and actinorhizal plants. New York: Academic Press, 157–176.

Silvester WB, Winship LJ. 1990. Transient responses of nitrogenase to acetylene and oxygen in actinorhizal nodules and cultured *Frankia*. *Plant physiology* **92**: 480–486.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212.

Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JM, Martin PG. 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.

Soyano T, Kouchi H, Hirota A, 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.

Stanke M, Diekhans M, Baertsch R, Haussler D. **2008**. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**: 637–644.

Stokstad E. **2016**. The nitrogen fix. *Science* **353**: 1225–1227.

Sturms R, Kakar S, Trent J, Hargrove MS. **2010**. *Trema* and *Parasponia* hemoglobins reveal convergent evolution of oxygen transport in plants. *Biochemistry* **49**: 4085–4093.

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, *et al.* 2008. The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucleic acids research* **36**: D1009–14.

Swensen SM. **1996**. The Evolution of Actinorhizal Symbioses: Evidence for Multiple Origins of the Symbiotic Association. *American journal of botany* **83**: 1503–1512.

Tóth K, Stratil TF, Madsen EB, Ye J, Popp C, Antolín-Llovera M, Grossmann C, Jensen ON, Schüßler A, Parniske M, *et al.* 2012. Functional domain analysis of the remorin protein LjSYMREM1 in lotus japonicus. *PloS one* **7**.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. **2010**. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* **28**: 511–515.

Trinick MJ. **1973**. Symbiosis between *Rhizobium* and the non-legume, *Trema aspera*. *Nature* **244**: 459–460.

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, *et al.* 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596–1604.

Udvardi M, Poole PS. **2013**. Transport and metabolism in legume-rhizobia symbioses. *Annual review of plant biology* **64**: 781–805.

UniProt Consortium. **2015**. UniProt: a hub for protein information. *Nucleic acids research* **43**: D204–12.

Untergasser A, Bijl GJM, Liu W, Bisseling T, Schaart JG, Geurts R. **2012**. One-step *Agrobacterium* mediated transformation of eight genes essential for rhizobium symbiotic signaling using the novel binary vector system pHUGE. *PloS one* **7**: e47885.

Valdar WSJ. **2002**. Scoring residue conservation. *Proteins* **48**: 227–241.

Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, *et al.* 2010. The genome of the domesticated apple (*Malus* × *domestica* Borkh.). *Nature genetics* **42**: 833–839.

Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, Cattonaro F, *et al.* 2013. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature genetics* **45**: 487–494.

Vernié T, Kim J, Frances L, Ding Y, Sun J, Guan D, Niebel A, Gifford ML, de Carvalho-Niebel F, Oldroyd GED. **2015**. The NIN transcription factor coordinates diverse nodulation programs in different tissues of the *Medicago truncatula* root. *The Plant cell* **27**: 3410–3424.

Werner GDA, Cornwell WK, Sprent JI, Kattge J, Kiers ET. **2014**. A single evolutionary innovation drives the deep evolution of symbiotic N_2-fixation in angiosperms. *Nature communications* **5**: 4087.

Winship LJ, Martin KJ, Sellstedt A. **1987**. The acetylene reduction assay inactivates root nodule uptake hydrogenase in some actinorhizal plants. *Physiologia plantarum* **70**: 361–366.

Yang M-Q, Van Velzen R, Bakker FT, Sattarian A, Li D-Z, Yi T-S. **2013**. Molecular phylogenetics and character evolution of Cannabaceae. *Taxon* **62**: 473–485.

Yano K, Yoshida S, Müller J, Singh S, Banba M, Vickers K, Markmann K, White C, Schuller B, Sato S, *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.

Young ND, Debellé F, Oldroyd GED, Geurts R, Cannon SB, Udvardi MK, Benedito V a., Mayer KFX, Gouzy J, Schoof H, *et al.* 2011. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**: 520–524.

Żmieńko A, Samelak A, Kozłowski P, Figlerowicz M. **2014**. Copy number polymorphism in plant genomes. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **127**: 1–18.

Chapter 3

Characterization of the nodulation phenotype of *Parasponia andersonii x Trema tomentosa* F1 hybrid plants

Fengjiao Bu, Yuda Roswanjaya, Elena Fedorova, Ton Bisseling, Rene Geurts

Laboratory of Molecular Biology, Department of Plant Science, Wageningen University, 6708PB Wageningen, The Netherlands.

Abstract

Nitrogen-fixing root nodule symbiosis occurs in ten taxonomic lineages from four related orders -Fagales, Fabales, Rosales and Cucurbitales- that together are called the nitrogen-fixing clade. A long-standing aim is to engineer this nitrogen-fixing nodulation trait into non-legume crops. Two comparative studies have been conducted to identify the core genes underlying this trait. Both reveal multiple losses of essential symbiotic genes in non-nodulating species of the nitrogen-fixing clade, including the LCO receptor encoding gene *NOD FACTOR PERCEPTION* (*NFP*), the LCO-responsive transcription factor encoding gene *NODULE INCEPTION* (*NIN*), and the coil-coiled protein-encoding gene *RHIZOBIUM DIRECTED POLAR GROWTH* (*RPG*). This suggests that these genes represent essential targets in future engineering approaches. Previously we obtained an F1 hybrid of the cross of diploid *Parasponia andersonii* and tetraploid *Trema tomentosa.* This F1 hybrid can form nodules, whereas it is devoid of intracellular infection when inoculated with either *Mesorhizobium plurifarium* BOR2 or *Bradyrhizobium elkanii* WUR3. Based on its genetic composition and symbiotic phenotype, we argue that the F1 hybrid may mimic future engineer results. Here we aimed to obtain a better understanding of the deviation in nodulation phenotype of wild type *P. andersonii* and F1 hybrid plants. To achieve this, we compared nodulation efficiencies and intracellular infection within nodule cells upon inoculation with a range of rhizobium strains on these plants. This revealed that the host range of hybrid plants is more narrower when compared to *P. andersonii*. Furthermore, we showed that the block in intracellular infection within hybrid nodules is consistent for all nodulating strains identified, and cannot be overcome by increased LCO biosynthesis nor by mutating the type III or IV secretion systems of nodulating strains. Finally, we found that hybrid plants can establish arbuscular mycorrhization effectively, suggesting that the block of intracellular infection is rhizobium specific. Taken together, this indicate the occurrence of a yet unknown mechanism leading to an impaired host range and block of intracellular infection of hybrid plants.

Introduction

Fixed nitrogen is an indispensable nutrient for plant growth but often limited in soils. To achieve a higher yield, farmers apply chemical nitrogen fertilizers. The downside of this practice is often eutrophication of the environment (Worrell et al. 2000; Sutton et al. 2011). Some plant species of the so-called nitrogen-fixing clade (the orders Fagales, Fabales, Rosales and Cucurbitales) can establish a nitrogen-fixing nodule symbiosis with either diazotrophic *Frankia* or rhizobium bacteria. For example, the actinorhizal plant species *Ceanothus thyrsiflorus* (Rhamnaceae, Rosales) and *Datisca glomerata* (Datiscaceae, Cucurbitales) that nodulate with *Frankia*, whereas *Parasponia* species (Cannabaceae, Rosales) and legume crops (Fabaceae, Fabales) such as soybean (*Glycine max*) nodulate with rhizobia. Such plants do not rely on exogenous fixed nitrogen sources, but rather obtain it from their nitrogen-fixing microsymbiont. These nitrogen-fixing plants form specialized organs called nodules to host the diazotrophic bacteria intracellularly. Inside nodule cells, bacteria are converting atmospheric di-nitrogen into ammonium driven by a protein complex named nitrogenase (Eady and Postgate 1974). The produced ammonium is assimilated into glutamine (or glutamate) that can be directly used by the plants (Groat and Vance 1981). In return, the bacteria receive photosynthetic carbohydrates from its host to fuel the nitrogen-fixing process.

A long-standing aim is to engineer this symbiotic nitrogen-fixing trait into non-legume crops (Markmann and Parniske 2009; Oldroyd and Dixon 2014; Stokstad 2016; Charpentier and Oldroyd 2010). Over the past two decades, crucial genes within the signal pathways underlying this symbiotic trait have been uncovered through studies in legumes and other nodulating plant species (Oldroyd 2013; Roy et al. 2019; Geurts et al. 2016; Mergaert et al. 2019). The general consensus now is that the capacity to establish nitrogen-fixing root nodule symbiosis partially recruited from the symbiotic signalling pathway that is used by the more ancient arbuscular mycorrhiza (AM) symbiosis, which is widespread among higher plants (Oldroyd 2013; Oldroyd et al. 2011; Geurts et al. 2016). Similar to AM fungi, some rhizobia and possibly clade II actinorhizal *Frankia* bacteria produce lipo-chitooligosaccharide (LCO)-type signal molecules that are perceived by specific LysM-type receptor kinases (Limpens et al. 2003; Madsen et al. 2003; Radutoiu et al. 2003; Maillet et al. 2011; Genre et al. 2013; Op den Camp et al. 2011; van Velzen et al. 2018; Van Nguyen et al. 2016; Persson et al. 2015). In the case of nodulating plants, perception of LCOs sets in motion nodule organogenesis. Also, LCO signalling is essential to allow bacterial intracellular infection (Moling et al. 2014).

Some micro-symbiotic partners can bypass LCO perception but still rely on the same common symbiosis signalling pathway (Fabre et al. 2015; Hocher et al. 2011; Tromas et al. 2012; Svistoonoff et al. 2014; Gherbi et al. 2008; Chabaud et al. 2016; Svistoonoff et al. 2013). So far the sequenced *Frankia* strains of the taxonomic clusters I and III do not contain the LCO biosynthesis genes encoded in rhizobium by the canonical *nodABC* operon (reviewed by Normand et al. 2007). It was shown that the signal molecules of unknown nature secreted by cluster I *Frankia* sp. strain CcI3 can trigger symbiotic gene expression in *Casuarina glauca* (Casuarinaceae, Fagales) and *Alnus glutinosa* as well as nuclear calcium spiking. The latter is considered as a hallmark response in symbiotic signalling (Chabaud et al. 2016; Granqvist et al. 2015). Likewise, LCO-independent nodulation occurs also in legumes. For example, the soybean LCO receptor mutant *nfr1* can nodulate by a few wild type *Bradyrhizobium elkanii* strains, e.g. USDA61 (Okazaki et al. 2013), whereas basal legume species such as *Aeschynomene indica* can establish nitrogen fixing nodules with photosynthetic *Bradyrhizobia* lacking the *nodABC*-encoded LCO biosynthesis genes (Okazaki et al. 2016). Recently, it was shown that ernA - a *Bradyrhizobium* specific effector-like protein - is essential for LCO-independent nodulation on *A. indica* (Teulet et al. 2019). Rhizobium deploys two specialized secretion systems called type III (T3SS) or type IV (T4SS) secretion systems to trigger nodule formation to secrete effector-like molecules (Nelson and Sadowsky 2015; Masson-Boivin et al. 2009; Marie et al. 2001; Fauvart and Michiels 2008; Deakin and Broughton 2009). Rhizobium genes encoding T3SS or T4SS are induced by flavonoids, similar as observed for LCO biosynthesis genes (Viprey et al. 1998). Nevertheless, having a T3SS also may affect the microbial host range negatively. For example, T3SS harbouring *Bradyrhizobium* spp. will be ineffective to nodulate soybean genotypes that harbour the *Rj4* locus, a thaumatin-like protein (TLPs) encoding gene, which function as a host restriction protein (Tsurumaru et al. 2015; Faruque et al. 2015; Vest and Caldwell 1972; Hayashi et al. 2014; Sadowsky and Cregan 1992). Another locus that controls host specificity in soybean is *Rj2*/*Rfg1*. This gene encodes a putative plant resistance protein homologous to a Toll-interleukin/nucleotide-binding site/leucine-rich repeat (TIR-NBS-LRR) receptor (Yang et al. 2010).

To identify symbiosis genes that form essential targets in an engineering approach, comparative phylogenomics studies have been conducted (Griessmann et al., 2019; Van Velzen 2018 Chapter 2). This uncovered three symbiosis genes, the LCO receptor encoding gene *NOD FACTOR PERCEPTION* (*NFP*), the LCO-responsive transcription factor encoding gene *NODULE INCEPTION* (*NIN*), and the coil-coiled protein-encoding gene *RHIZOBIUM*

DIRECTED POLAR GROWTH (*RPG*), of which loss or pseudogenization correlates with absence of nodulation in species within the nitrogen-fixing clade ((van Velzen et al. 2018; Griesmann et al. 2018; **Chapter 2**). This favours the hypothesis that nodulation evolved only once at the root of the nitrogen-fixing clade, and was subsequently lost multiple times (van Velzen et al. 2019; Soltis et al. 1995; Swensen 1996). As *NFP*, *NIN* and *RPG* have a nodulation specific expression profile and have shown to be essential for nodulation in legumes and non-legumes, these three genes form essential targets in engineering approaches.

The *Parasponia* lineage represents five species and phylogenetic analysis shows that this lineage is embedded within the *Trema* clade (van Velzen et al. 2018; Yang et al. 2013). As *Parasponia* and *Trema* are closely related, F1 hybrids could be created by crossing of the diploid *Parasponia andersonii* (2n=20) and the allotetraploid *Trema tomentosa* (2n=4X=40). Conceptually, *P. andersonii x T. tomentosa* F1 hybrid plants reflects a diploid *T. tomentosa* with a haploid genome of *P. andersonii* introduced. As can be anticipated, such hybrids can form nodules (van Velzen et al. 2018; **Chapter 2**). However, despite efficient nodulation, rhizobium is unable to establish intracellular infections within hybrid nodules. As the nodulation trait of the *P. andersonii x T. tomentosa* hybrid may reflect a future engineering result, hybrid plants represent a valuable experimental tool to study the mechanism controlling intracellular rhizobium infection. In this chapter, we conducted studies aiming to understand the importance of LCO signalling and putative effector signalling in nodulation and intracellular infection of *P. andersonii x T. tomentosa* hybrid plants.

Results

P. andersonii x T. tomentosa F1 hybrid plants have a more narrow host range

Previously we have reported that nodules formed on *P. andersonii* x *T. tomentosa* F1 hybrids are unable to establish intracellular infection, irrespective whether inoculated with *Mesorhizobium plurifarium* BOR2 or *Bradyrhizobium elkanii* WUR3 (**Chapter 2**). To test whether this is a generic characteristic of *P. andersonii x T. tomentosa* F1 hybrid plants that result from the hybridization of the two parental genomes, we tested different rhizobial species (**Table 2**). *P. andersonii* is a promiscuous host (Op den Camp et al. 2012; van Velzen et al. 2018), and it engaged with all the five strains tested albeit with different nodulation efficiencies (**Fig. 1A-E**). Analyses of nodule cytoarchitecture showed that *P. andersonii* nodules induced by all tested strains contained intracellular infections (**Fig. 2**). Also hybrid line H9 plants nodulated upon inoculation with *Rhizobium tropici* CIAT899 and the β-proteobacteria

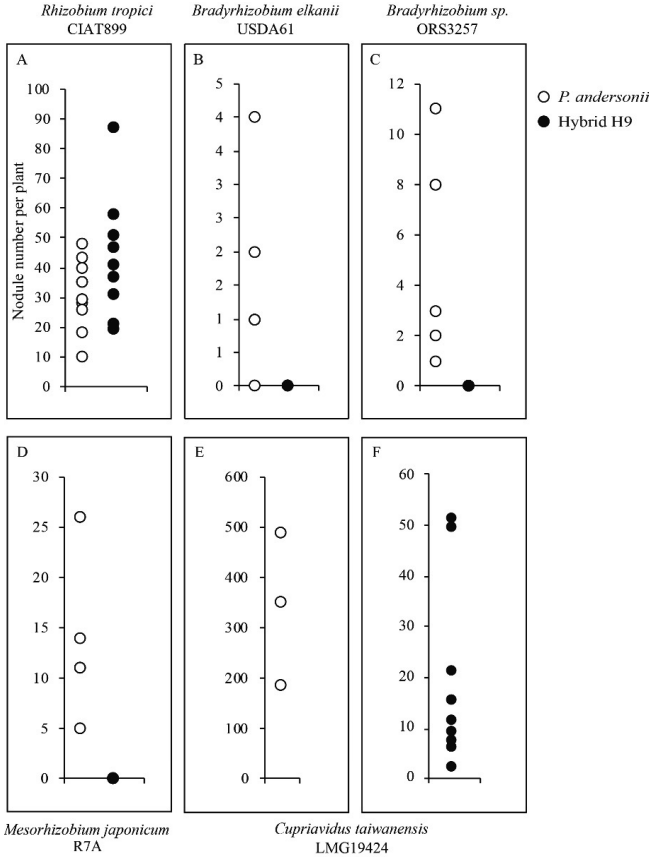


Fig. 1 Nodulation efficiency of *P. andersonii* and *P. andersonii* x *T. tomentosa* hybrid line H9 inoculated with different rhizobium strains. *Rhizobium tropici* CIAT899 (**A**), *Bradyrhizobium elkanii* USDA61 (**B**), *Bradyrhizobium* sp. ORS3257 (**C**), *Mesorhizobium loti* R7A (Martínez-Hidalgo et al. 2016) (**D**) and *Cupriavidus taiwanensis LMG19424* (**E, F**). Open dots represent nodule number on *P. andersonii,* filled dots represent nodule number on hybrid line H9. Each dot represents nodule number formed on an individual plant. Nodulation were scored after 4 weeks post inoculation at an OD_{600} = 0.03 for A-D and F (van Zeijl et al., 2018; van Velzen et al., 2018; Wardhani et al., 2019). Note a different nodulation system was used for E (see Material and Methods).

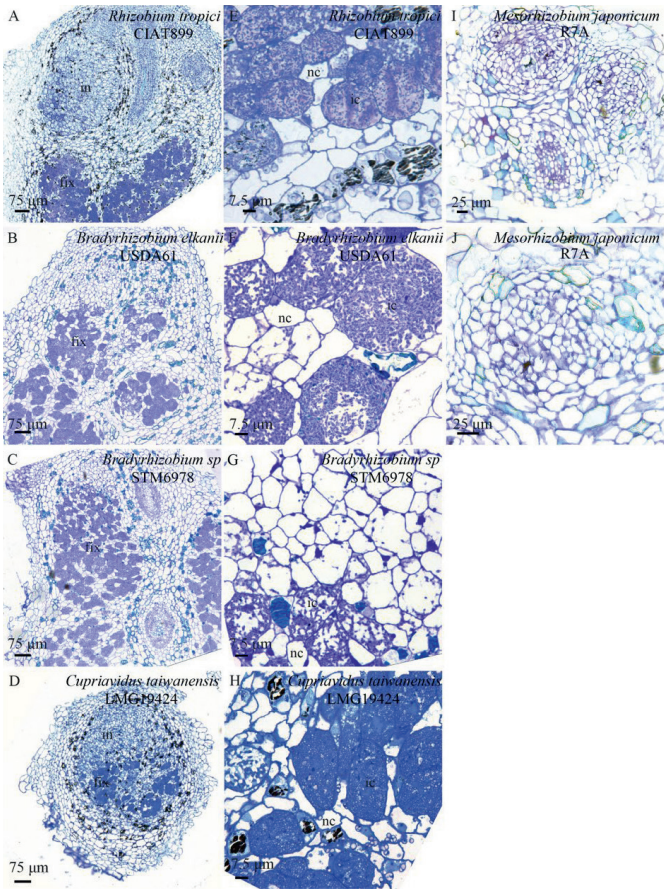


Fig. 2 *P. andersonii* nodules induced by five different rhizobium species form intracellular infection threads. (**A-D**, **I**) Sections of a mature *P. andersonii* nodule containing infection zone (in) and fixation zone (fix). (**E-H**, **J**) *P. andersonii* nodule cells are filled with intracellular infection threads. (**A**, **E**) *P. andersonii* nodule induced by *Rhizobium tropici* CIAT899; (B, F) *P. andersonii* nodule induced by *Bradyrhizobium elkanii* USDA61; (**C**, **G**) *P. andersonii* nodule induced by *Bradyrhizobium* sp. ORS3257; (**D**, **H**) *P. andersonii* nodule induced by *Cupriavidus taiwanensis* LMG19424 and (I, **J**) *P. andersonii* nodule induced by *Mesorhizobium loti* R7A. Sections were stained using Toluidine Blue. in: infection zone; fix: fixation zone; ic: infected cells; nc: non-infected cells. Nodules were isolated 4 weeks post-inoculation for **A-C**, and **I**, 8 weeks post-inoculation for **D**.

Cupriavidus taiwanensis LMG19424 (**Fig. 1A, E-F**). Similar to what we found previously when inoculated with *M. plurifarium* BOR2 or *B. elkanii* WUR3, hybrid nodules induced by *R. tropici* CIAT899 and *C. taiwanensis* LMG19424 were devoid of intracellular infection threads (**Fig. 3A-B**, **4A-B**). Based on the consistency in lack of intracellular infections of hybrid nodules induced by a wide range of rhizobium species, we hypothesize that this phenotype is unlikely caused by a particular rhizobium strain or genus, but rather by a general characteristic of the *P. andersonii* x *T. tomentosa* F1 hybrid plants. Regarding *M. loti* R7A, *B. elkanii* USDA61 and *Bradyrhizobium sp.* ORS3257, no nodules were found on hybrid H9 plants (**Fig. 1B-D**). This shows that a *P. andersonii x T. tomentosa* F1 hybrid has a narrower host range when compared to *P. andersonii*.

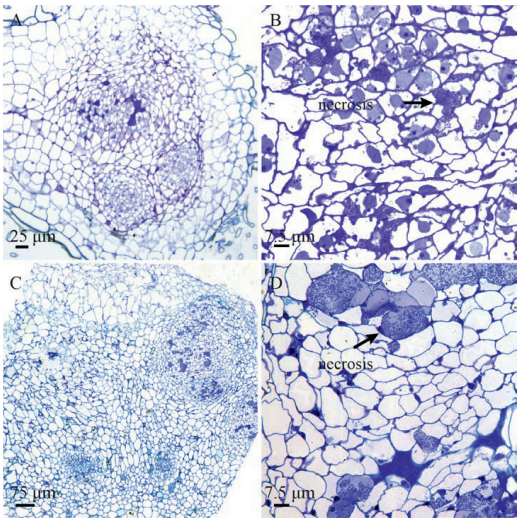


Fig. 3 Cytoarchitecture of *P. andersonii* x *T. tomentosa* hybrid nodules induced by *Rhizobium tropici* CIAT899 wild type and its LCO overproducing *Rhizobium tropici* CIAT899.pMP604. Hybrid line H9 nodules induced by *Rhizobium tropici* CIAT899 (**A, B**) and *R. tropici* CIAT899.pMP604 (**C, D**). (**A**) Sections of hybrid H9 nodules induced by wild type *R. tropici* CIAT899 lacks an infection and fixation zone (**Fig. 2A**). (B) Hybrid H9 nodules induced by *R. tropici* CIAT899 lack intracellular infection threads. Instead, cell necrosis can be detected (arrow). (**C**) Sections of hybrid H9 nodules induced by *R. tropici* CIAT899.pMP604. (**D**) Hybrid H9 nodules induced by *R. tropici* CIAT899.pMP604 lack intracellular infection threads. Instead, cell necrosis can be detected (arrow). Note cell necrosis is stronger in nodules induced by *Rhizobium tropici* CIAT899.pMP604 (D). Nodules were isolated 4 weeks post-inoculation.

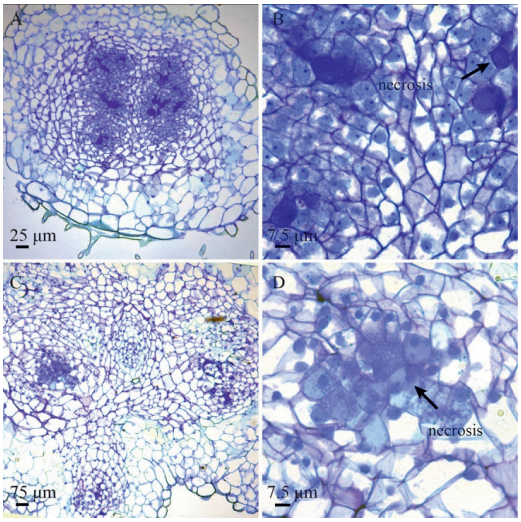


Fig. 4 Cytoarchitecture of *P. andersonii* x *T. tomentosa* hybrid (line H9) nodules induced by *Cupriavidus taiwanensis* LMG19424 (CBM777) and its type III secretion system (T3SS) mutant strain *C. taiwanensis* LMG19424. Hybrid H9 nodules induced by wild type *C. taiwanensis* LMG19424 (**A, B**) and T3SS mutant strain (**C, D**). (**A**) Sections of hybrid (line H9) nodules induced by *C. taiwanensis* LMG19424 lacks infection and fixation zone (see also **Fig. 2A**). (**B**) Hybrid H9 nodules induced by *C. taiwanensis* LMG19424 lack intracellular infection threads. Instead, cell necrosis can be detected (arrow). (**C**) Sections of hybrid H9 nodules induced by T3SS mutant strain *C. taiwanensis* LMG19424. (**D**) Hybrid H9 nodules induced by the mutant strain of *C. taiwanensis* LMG19424 lack intracellular infection threads. Instead, cell necrosis can be detected (arrow). Nodules were isolated 4 weeks post-inoculation.

Block of intracellular infection in hybrid plants does not affect mycorrhization

To test whether the block of intracellular infection in hybrid plants is specific to nitrogen-fixing root nodule symbiosis, we determined whether arbuscular mycorrhiza symbiosis could be established. We tested the mycorrhization of hybrid line H9 by inoculating with spores of *Rhizophagus irregularis* DOAM197198. Both parental lines, *P. andersonii* and *T. tomentosa*,

can establish an arbuscular mycorrhizal symbiosis (van Velzen et al. 2018**; Fig. 5A-B**). Analysis of hybrid roots at 6 weeks post inoculation showed that well-developed arbuscules were present (**Fig. 5C**). This shows that the lack of intracellular infection is not a generic phenotype of hybrid plants, but most probably specific to nitrogen-fixing root nodule symbiosis.

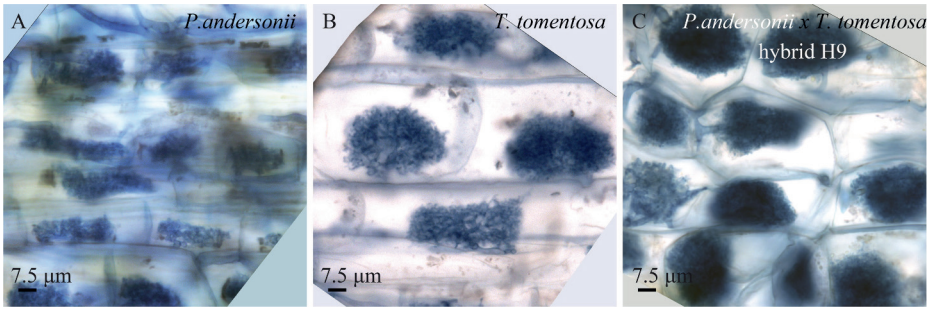


Fig. 5 Mycorrhization phenotype of *Parasponia andersonii*, *Trema tomentosa* and their intergeneric F1 hybrid line H9. Arbuscules in root cells of *P. andersonii* (**A**), *T. tomentosa* (**B**) and intergeneric F1 hybrid H9 inoculated with *Rhizophagus irregularis* strain DOAM197198. Shown are representative root segments stained with 0.05% (w/v) trypan blue harvested at 6 weeks post-inoculation.

Block of intracellular rhizobium infection in hybrid plants can't be rescued by enhanced LCO biosynthesis

Previous work showed that the nodulation of *P. andersonii* is dependent on LCO signalling (Op den Camp et al. 2011; van Zeijl et al. 2018; **Chapter 2**; **Chapter 4**; Luuk et al., unpublished). As hybrid plants have only a single allele of all *P. andersonii* genes essential for nodulation and intracellular infection, we questioned whether the absence of deficiency of intracellular infection can be overcome by increased LCO biosynthesis of rhizobium. To answer this question, we made use of rhizobium strains that constitutively produce LCOs. Such strains were obtained by transforming plasmid pMP604, which harbors a gene encoding a flavonoid-independent NodD transcription factor (Spaink et al. 1989). We transformed pMP604 into *R. tropici* CIAT899 and *M. loti* R7A. qRT-PCR analysis confirmed increased expression of the LCO biosynthesis genes encoded by *nodABC* operon, irrespective of the presence or absence of the inducer flavonoid naringenin. LCO biosynthesis genes were up to 20-fold higher expressed under non-inductive conditions (**Fig. 6** or **Chapter 5, Fig. S3**). Nodulation assays using *R. tropici* CIAT899.pMP604 and *M. loti* R7A.pMP604 on *P. andersonii* showed that both strains induce significantly more nodules when compared to the wild type counterparts (**Fig. 7A-B**). This indicates that the enhanced nodulation efficiency of

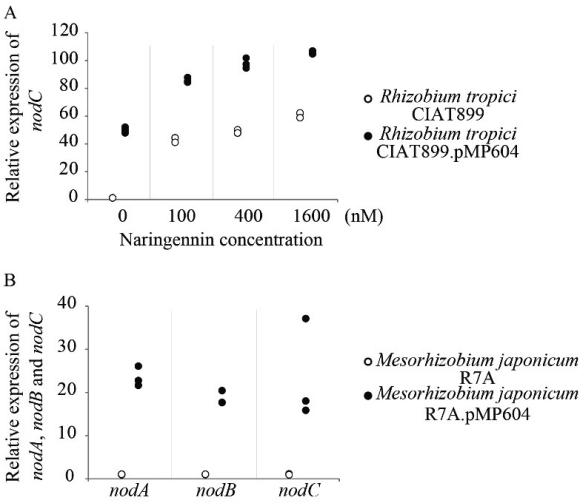


Fig. 6 *Rhizobium tropici* CIAT899.pMP604 and *Mesorhizobium loti* R7A.pMP604 constitutively expresses LCO biosynthesis genes. The pMP604 plasmid encodes a constitutively expressed variant of nodD, a transcription factor that regulates LCO biosynthesis genes. **A**, Shown is the relative expression of *nodC* in wild-type *Rhizobium tropici* CIAT899 (white dots) and *Rhizobium tropici* CIAT899.pMP604 (black dots) in the presence and absence of the flavonoid naringenin (Han et al. 2009). **B**, Shown is the relative expression of *nodABC* in wild-type *M. loti* R7A (white dots) and *M. loti* R7A.pMP604 (black dots) in the absence of (iso)flavonoid (Han et al. 2009). Dots represent technical repeats.

R. tropici CIAT899. pMP604 also induces significantly higher amount of nodules on the hybrid line H9 compared to the wild type strain (**Fig. 7C**). However, these nodules remain devoid of intracellular infections (**Fig. 3C-D**). This indicates that the intracellular infection phenotype of hybrid plants can't be rescued by enhanced LCO biosynthesis of the rhizobia. Furthermore, we found that *M. loti* R7A.pMP604 remains unable to induce any nodules on hybrid plants (**Fig. 7D**).

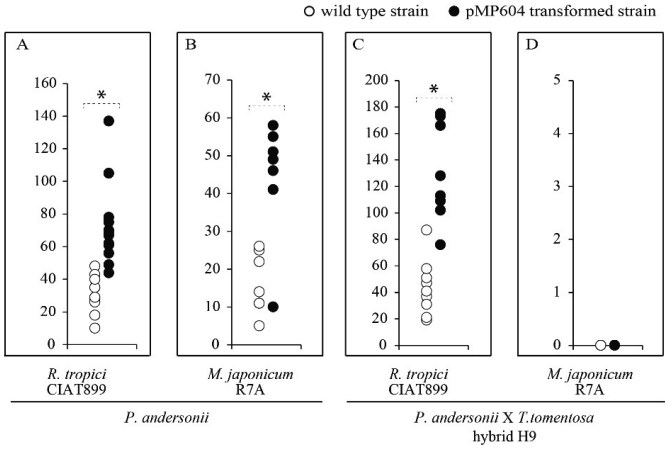


Fig. 7 Nodulation efficiency of *P. andersonii* and *P. andersonii* x *T. tomentosa* hybrid inoculated with *Rhizobium tropici* CIAT899 and *Mesorhizobium loti* R7A and the LCO-overproducing strains carrying pMP604. Nodulation efficiency induced by wild type and pMP604 transformed strains of *R. tropici* CIAT899 and *M. loti* R7A on *P. andersonii* (**A, B**) and intergeneric F1 hybrid line H9 (**C, D**). Open dots represent nodule number induced by wild type strain, filled dots represent nodule number induced by pMP604 transformed strains. Each dot represents nodule number formed on an individual plant. Nodulation were scored after 4 weeks post inoculation with an OD_{600} = 0.03.

This indicates that the impaired host range of *P. andersonii* x *T. tomentosa* F1 hybrid plants is caused by a yet unknown mechanism locate downstream of, or parallel to LCO signalling.

Mutations in rhizobium type III/IV secretion systems are unable to restore intracellular infection

It has been reported that rhizobium T3SS and T4SS can negatively affect host range and intracellular infection of some soybean cultivars (Tsurumaru et al. 2015; Faruque et al. 2015; Vest and Caldwell 1972; Hayashi et al. 2014; Sadowsky and Cregan 1992; Yang et al. 2010). We questioned whether the presence of either of these secretion systems has an effect on nodulation as well as intracellular infection capacity of the rhizobium in *P. andersonii* and/or hybrid nodules. *R. tropici* CIAT899 and *C. taiwanensis* LMG19424, the two strains which can nodulate the hybrid, contain an operon encoding proteins constituting a T3SS (Ormeño-Orrillo et al. 2012; Amadou et al. 2008; Saad et al. 2012). This suggests a possible role of rhizobium T3SS in nodulation of *P. andersonii* and/or hybrid. To seek support for this, we first sequenced the genome of *M. plurifarium* BOR2 and re-analyzed the genome of *B. elkanii* WUR3, the two strains which also can nodulate hybrid plants. This revealed the presence of genes encoding T4SS system in *M. plurifarium* BOR2 and genes encoding T3SS within *B.elkanii* WUR3 genome (**Table 3, Table 4**). Next, we tested whether a T4SS mutant of *M. loti* R7A (AH34) (Hubber et al. 2004), T3SS mutants of *B. elkanii* USDA61 (△T3SS) (Okazaki et al. 2013), *Bradyrhizobium* sp. ORS3257 (△T3SS) (Okazaki et al. 2016) and *C. taiwanensis* LMG19424 (CBM312) (Saad et al. 2012) are affected in the nodulation efficiency on *P.andersonii*. This revealed a range of effects. The T3SS mutants of *B. elkanii* USDA61 (△T3SS) and *Bradyrhizobium* sp. ORS3257 (△T3SS) showed an increased nodulation efficiency on *P. andersonii* (**Fig. 8A-B**), while a T4SS mutation in *M. loti* R7A (AH34) reduced its nodulation efficiency **(Fig. 8C**). Worth noting is that both *B. elkanii* USDA61 and *Bradyrhizobium* sp. ORS3257 induced a large amount of small nodule-like structures on *P. andersonii* roots **(Fig. 9A-B**, **D-E**), which was not observed in case of the T3SS mutant counterparts (**Fig. 9C**, **F**). Cytological analysis of these nodule-like structures showed that they lacked rhizobium intracellular infection structures (**Fig 9G-H**). This shows that T3SS of *B. elkanii* USDA61 and *Bradyrhizobium* sp. ORS3257 do affect the nodulation efficiency as well the capability of the strain to establish intracellular infections in *P. andersonii* nodules via a yet unknown mechanism.

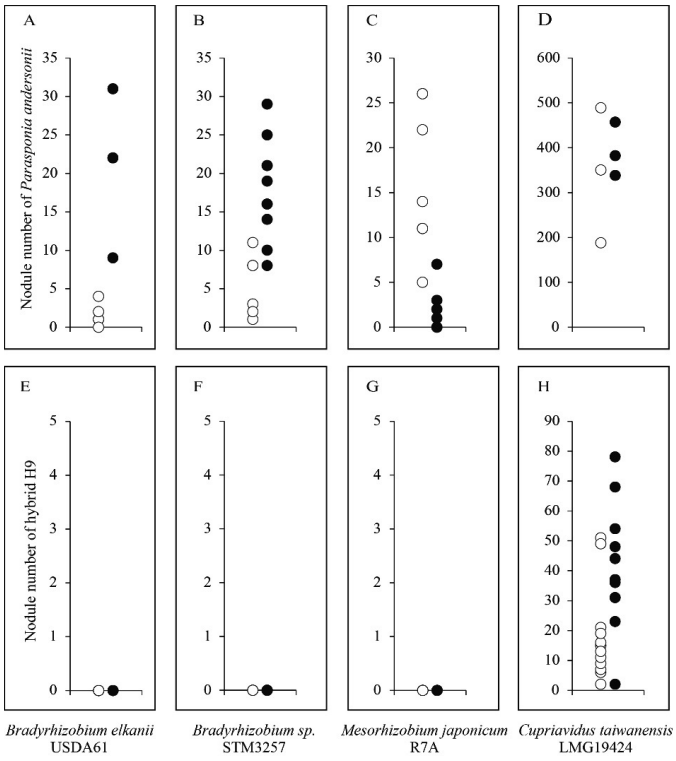


Fig. 8 Nodulation efficiency of *P. andersonii* and *P. andersonii* x *T. tomentosa* hybrid inoculated wild type strains and their type III or type IV secretion system mutants. Nodulation efficiency of *P. andersonii* (**A-D**) and hybrid line H9 (**E-H**) inoculated with wild type rhizobium (open dots) and their type III or type IV secretion system mutants (filled dots). Each circle represents nodule number formed on an individual parasponia or hybrid plant. Nodulation were scored after 4 weeks post inoculation with an OD_{600} = 0.03 for **A** to **C**, and **E** to **H**. Note a different nodulation system was used for **D** (see Material and Methods).

Further, we tested whether the T4SS mutant of *M. loti* R7A (AH34), the T3SS mutants of *B. elkanii* USDA61 (△T3SS), *Bradyrhizobium* sp. ORS3257 (△T3SS), and *C. taiwanensis* LMG19424 (CBM312) gain an increased capacity to nodulate hybrid plants, and if so, whether intracellular infections could be established. Like their wild type counterparts, the *M. loti* R7A T4SS mutant AH34, and the *B. elkanii* USDA61 T3SS mutant (△T3SS) and *Bradyrhizobium* sp. ORS3257 T3SS mutant (△T3SS) were unable to induce nodule formation on hybrid plants. Nodules on hybrid roots were only induced by the *C. taiwanensis* LMG19424 T3SS mutant strain CBM312, of which the wild type counterpart was also capable to do so **(Fig. 8H)**. Analysis of the nodule cytoarchitecture showed that hybrid H9 nodules induced by CBM312 did not establish any intracellular infection (**Fig. 4C-D**). Thus, we conclude that T3SS/T4SS of the rhizobium strains we have tested is neither responsible for the intracellular infection phenotype, nor for the host range of *P. andersonii* x *T. tomentosa* hybrid plants.

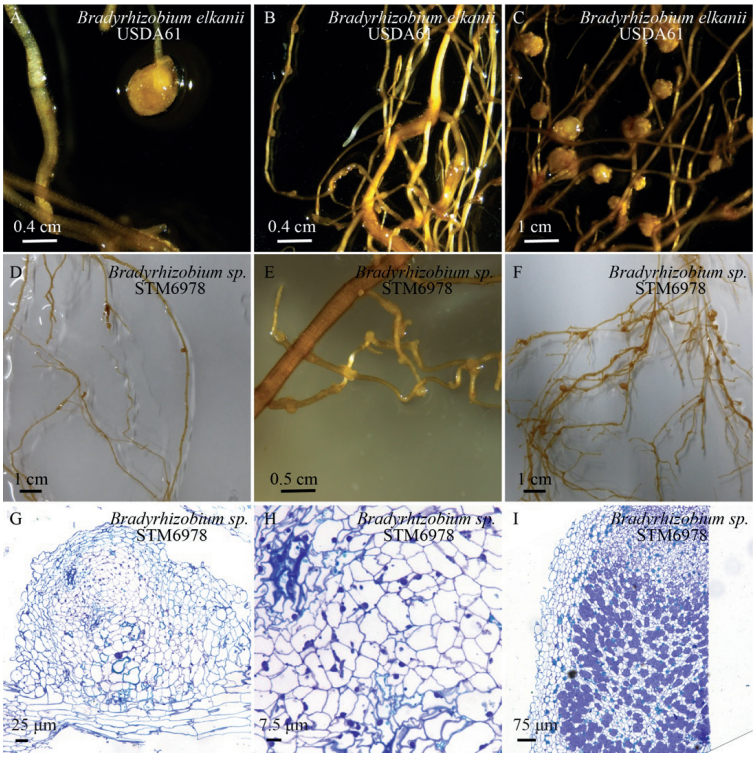


Fig. 9 *P. andersonii* nodulation efficiency upon inoculation with *Bradyrhizobium elkanii* USDA61, *Bradyrhizobium sp.* ORS3257 and their subsequent T3SS mutants. (**A-B**) *P. andersonii* roots inoculated with *B. elkanii* USDA61 forms nodules (**A**) and small nodule-like structures (**B**). (**C**) *P. andersonii* roots inoculated with *B. elkanii* USDA61 T3SS mutant. Note the absence of nodule-like structures. (**D-E**) *P. andersonii* roots inoculated with *Bradyrhizobium* sp. ORS3257 form nodules (**D**) and nodule-like structures (**E**). (**F**) *P. andersonii* roots inoculated with *Bradyrhizobium sp.* ORS3257 T3SS mutant. Note the absence of nodule-like structures. (**G-H**) Cytoarchitecture of nodule-like structures induced by *Bradyrhizobium* sp. ORS3257 on *P. andersonii* roots. Note absence of intracellular infection, and apoplastic infections were rarely observed. (**I**) Section of *P. andersonii* nodule induced by *Bradyrhizobium sp.* ORS3257 T3SS mutant shows a cytoarchitecture similar as nodules induced by wild type *Bradyrhizobium* sp. ORS3257.

Increased defence responses within Hybrid H9 nodules

In legumes, rhizobial mutants showing a deficiency during the development of nitrogen fixation nodules can trigger defence responses. Extensive formation of callose, which is a cell wall component, is a hallmark of such a response (Gaudioso-Pedraza et al. 2018). Since nodulation of hybrid plants is often accompanied by apoplastic rhizobium colonization and death of hybrid nodules cells, we studied whether callose deposition in hybrid nodules is more severe when compared to wild type nodules. Comparison of the amount of callose in *P. andersonii* and hybrid H9 nodules suggests a stronger callose deposition in hybrid nodules (**Fig. 10**). This indicates an increased defence response in hybrid H9 nodules when compared to *P. andersonii* nodules.

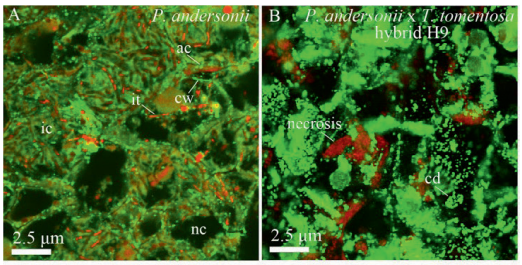


Fig. 10 Nodules formed on *P. andersonii* x *T. tomentosa* intergeneric F1 hybrid H9 show callose deposition. Fluorescence microscopy detection of immunolocalization of callose within *P. andersonii* (**A**) and hybrid line H9 nodule cells (**B**) induced by *Mesorhizobium plurifarium* BOR2. Red fluorescence represents rhizobium. Green fluorescence signal represents callose. Note nodule cells shown in **A** are infected by rhizobium intracellular infection or fixation threads. ac: apoplast colony; it: infection threads; cw: cell wall; ic: infected cells; nc: non-infected cells; necrosis: dead cells colonized by necrosis rhizobium; cd: callose deposition.

Discussion

In this study we compared the nodulation capacity of *P. andersonii* to an intergeneric hybrid *P. andersonii x T. tomentosa*, which represents the first hybrid of a cross between an nodulating and a non-noduling species. This hybrid is capable to establish the arbuscular mycorrhizal symbiosis as well as form nodules with a variety of rhizobium bacteria. Though nodules induced by all the four nodulating rhizobia strains, *M. plurifarium* BOR2 (**Chapter 2**), *B. elkanii* WUR3 (**Chapter 2**), as well as *R. tropici* CIAT899 and β-proteobacteria *C. taiwanensis* LMG19424 (this Chapter), showed to be not intracellularly infected. This is in contrast to *P. andersonii.* This indicates that, like in legumes, nodule organogenesis and rhizobium intracellular infection can be genetically uncoupled in Cannabaceae species. The strong defence response within hybrid nodules as visualized by increased callose deposition, is similar to responses provoked in legume nodules when the symbiosis is aborted (Mithöfer 2002). Therefore defence response is unlikely a causal but rather a general outcome of the interaction of nodulating rhizobia and *P. andersonii x T. tomentosa* hybrid plants. However, the molecular nature of that causes this block of intracellular infection in hybrid plants remains elusive.

P. andersonii is known as highly promiscuous towards nitrogen-fixing rhizobial micorsymbionts (Op den Camp et al. 2012). In contrast, the *P. andersonii x T. tomentosa* hybrid showed to be more restrictive. Whereas *P. andersonii* can nodulate with *B. elkanii* USDA61, *Bradyrhizobium* sp. ORS3257 and *M. loti* R7A, hybrid plants can not. We found that this incompatibility could not be rescued by increasing LCO production. These findings revealed that nodulation efficiency on *P. andersonii* as well as the hybrid increases when inoculated with a compatible strain that produces LCOs in excess and independent of plant secreted flavonoids. However, LCO-overproduction could not overcome the intracellular infection phenotype of hybrid plants e.g. as observed with *R. tropici CIAT899.pMP604*, nor the incompatibility to induce nodules as observed when hybrid plants are inoculated with *M. loti* R7A.pMP604. This suggests that the nodule phenotype of *P. andersonii* x *T. tomentosa* hybrid plants is not related to the dose-responsiveness towards rhizobium secreted LCOs.

Several studies in legumes demonstrated that T3SS or T4SS secretion systems of the rhizobia can affect the nodulation efficiency, most probably due to the secretion of putative effector proteins (Saad et al. 2012; Teulet et al. 2019). We demonstrated that also on *P. andersonii* the nodulation efficiency is affected by these secretion systems. For example, T3SS mutants of *B. elkanii* USDA61 and *Bradyrhizobium* sp. ORS3257 showed an increased nodulation and

infection efficiency, suggesting that both wild type strains may secrete effector-like molecules that hamper nodulation in *P. andersonii*. In contrast, the T4SS mutation in *M. loti* R7A reduces its nodulation efficiency on *P. andersonii*, suggesting an occurrence of effectors that promote nodulation in *P. andersonii*. However, studies on hybrid plants could not reveal a role of the T3SS and/or T4SS in the observed nodulation and/or infection phenotypes.

Interspecific and intergeneric hybridization of plants usually causes pleiotropic effects and gene misexpression, resulting in physiological disorders like growth deficiency and infertility (Barr and Fishman 2011; MacNair and Christie 1983). Similarly, the *P. andersonii* x *T. tomentosa* F1 hybrid also shows growth distortions (**Chapter 2**). However, we do not anticipate that this generic pleiotropic effect is the causal of the blocked intracellular infection. This, because hybrid plants can still form arbuscules upon inoculation with the AM fungus *Rhizophagus irregularis*. Comparative transcriptome studies in *P. andersonii* and hybrid nodules have uncovered dozens of genes that do not show a nodule-enhanced expression profile in hybrid nodules. One such gene is *RPG*, which in the legume Medicago truncatula controls intracellular infection (Arrighi et al. 2008). While *RPG* expression is 300 fold upregulated in *P. andersonii* nodules compared to uninoculated roots in hybrid nodules this gene only upregulated 20 fold, 100 times less compared to that in *P. andersonii* (van Velzen et al. 2018). Interestingly, *NUCLEAR FACTOR YA1* (NF-YA1), which is known to control intracellular infection in *P. andersonii* is not affected in its expression in hybrid nodules (Bu et al., 2020 in press; Van Velzen et al., 2018). This suggests that hybrid plants are affected specifically in a symbiotic response upstream of *RPG* transcriptional regulation.

Besides gene loss in nonsymbiotic *Trema* species, comparative genomics also revealed that *Parasponia* species consistently lost a few genes, one of which is the symbiotic related gene *EPR3* (*EXOPOLYSACCHARIDE RECEPTOR 3*). In the legume *Lotus japonicus* the LysM receptor kinase *LjEPR3* is functioning in recognition of surface exopolysaccharides of compatible bacteria. *LjEPR3* functions in facilitating infection thread progression in root cortex as well as in nodule primordium of compatible strains (Kawaharada et al. 2017; Kawaharada et al. 2015). Interestingly, though at low level, the *T. tomentosa EPR3* allele is expressed in hybrid nodules. Considering the two main symbiotic phenotype the intergeneric hybrid shows, namely (i) a more narrow host range, and (ii) a block in intracellular infection, expression of the *T. tomentosa* alleles of the *EPR3* gene in hybrid nodule could be causal for these phenotypes. The reverse genetic tools available for *P. andersonii* will allow to test this hypothesis (van Zeijl et al. 2018).

Material and Methods

Plant Materials and Growth Conditions

P. andersonii WU1.14 and F1 hybrid plants line H9 were maintained as described previously (van Zeijl *et al.*, 2018; Wardhani *et al.*, 2019). Young plantlets for nodulation assays were vegetatively propagated *in vitro* and rooted (van Zeijl *et al.*, 2018; Wardhani *et al.*, 2019).

Rhizobium transformation

Constructs pMP604 was transformed into *R.tropici* CIAT899 and *M. japonica* R7A (Martínez-Hidalgo et al. 2016) using electroporation. Electrocompetent cells of rhizobium *R.tropici* CIAT899 and *M. japonica* R7A (Martínez-Hidalgo et al. 2016) were prepared according to the published methods used for preparing *E.coli* electrocompetent cells with the exception of using low salt LB broth (Sharma and Schimke 1996). Transformants were obtained by tetracycline antibiotics selection as well as PCR amplification using specific primer pairs (**Table 1**) to detect the presence of pMP604 plasmid. Colony PCR was conducted using DreamTaq DNA Polymerases (Thermo Fisher Scientific) according to the manufacturer's protocol.

LCO biosynthesis gene expression in rhizobium strains transformed with pMP604

Flavonoid naringenin was added to rhizobium preculture in a range of 0 nM to 1600 nM and incubate at 28 °C for 5 hours. Bacteria were collected by centrifuge at 4000 rpm at 4°C. The pellet was carried on with RNA isolation conducted according to the protocol provided by the manufacturer (RNeasy Mini Kit, Qiagen, Germany). cDNA library was synthesis using 500 ng for each sample using i-script cDNA synthesis kit (Bio-Rad, United States). qRT-PCR conducted was set up as described in (van Zeijl et al. 2015). Normalization was performed based on the expression of rhizobium 16S *rRNA* gene. Primer pairs used in this study are listed in **Table 1**.

Rhizobium and type III/IV secretion system

Rhizobium strains used in this study were listed in **Table 2**. Composition of Type III/IV secretion system in the tested rhizobium strains and corresponding mutants used in this study were listed in **Table 3**. Proteins from *Mesorhizobium japonica* MAFF303099 (Okazaki et al. 2010) were used as query to detect genes encoding the Type III secretion system (T3SS) in *Bradyrhizobium elkanii* WUR3, and proteins from *Mesorhizobium japonica* R7A (Hubber et

al. 2004) were used as query to detect genes encoding the Type IV secretion system (T4SS) in *Mesorhizobium plurifarium* BOR2, respectively (**Table** 4).

Mycorrhization assay

Mycorrhization of *P. andersonii*, *T. tomentosa* and F1 hybrid line H9 were carried out according to Van Velzen et al. (2018) and Wardhani et al. (2019).

Nodulation assay

Nodulation assay was carried out according to the previous report unless stated otherwise (van Velzen *et al.*, 2018; Wardhani *et al.*, 2019). Nodulation assay for Parasponia inoculated with *Cupriavidus taiwanensis LMG19424* were carried out in greenhouse of Wageningen in winter, 2016. Rooted *P. andersonii* plantlets were grown on an autoclaved fine sand and watered every second week with EKM medium (with 0.0375mM NH_4NO_3) and nodulation efficiency were scored after 3 month post inoculation. Nodulation efficiencies were scored for individual plants.

Microtome Sectioning and Microscopy

Nodules were fixed in 4% paraformaldehyde (w/v), 5% glutaraldehyde (v/v) in 50 mM phosphate buffer (pH = 7.2) at 4°C for 24 hours for plastic sectioning. Subsequently, the samples were dehydrated using an ethanol series and embedded in Technovit 7100 (Heraeus Kulzer, Germany) according to the manufacturer's instructions. Semi-thin sections were cut using a Leica Ultracut microtome (Leica Microsystems, Germany) to 4 µm thickness (7 µm thickness in the case for GUS stained samples). Sections were stained with 0.05% Toluidine Blue or 0.1% Ruthenium Red. Images were photographed using a Leica DM5500B microscope equipped with a DFC425C camera (Leica Microsystems, Germany).

References

Amadou C, Pascal G, Mangenot S, Glew M, Bontemps C, Capela D, Carrère S, Cruveiller S, Dossat C, Lajus A, *et al.* 2008. Genome sequence of the β-rhizobium *Cupriavidus taiwanensis* and comparative genomics of rhizobia. *Genome research* **18**: 1472–1483.

Arrighi J-F, Godfroy O, de Billy F, Saurat O, Jauneau A, Gough C. 2008. The RPGgene 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.

Barr CM, Fishman L. 2011. Cytoplasmic male sterility in *Mimulus* hybrids has pleiotropic effects on corolla and pistil traits. *Heredity* **106**: 886–893.

Fengjiao Bu, Luuk Rutten, Yuda Purwana Roswanjaya, Olga Kulikova, Marta Rodriguez-Franco, Thomas Ott, Ton Bisseling, Arjan van Zeijl & Rene Geurts. 2020. Mutant analysis in the non-legume *Parasponia andersonii* identifies NIN and NF-YA1 transcription factors as a core genetic network in nitrogen-fixing nodule symbioses. *New Phytologist* (In press).

Chabaud M, Gherbi H, Pirolles E, Vaissayre V, Fournier J, Moukouanga D, Franche C, Bogusz D, Tisa LS, Barker DG, *et al.* 2016. Chitinase-resistant hydrophilic symbiotic factors secreted by Frankia activate both Ca(2+) spiking and NIN gene expression in the actinorhizal plant *Casuarina glauca*. *The New phytologist* **209**: 86–93.

Charpentier M, Oldroyd G. 2010. How close are we to nitrogen-fixing cereals? *Current opinion in plant biology* **13**: 556–564.

Deakin WJ, Broughton WJ. 2009. Symbiotic use of pathogenic strategies: rhizobial protein secretion systems. *Nature reviews. Microbiology* **7**: 312–320.

Fabre S, Gully D, Poitout A, Patrel D, Arrighi J-F, Giraud E, Czernic P, Cartieaux F. 2015. Nod Factor-Independent Nodulation in *Aeschynomene evenia* Required the Common Plant-Microbe Symbiotic Toolkit. *Plant physiology* **169**: 2654–2664.

Faruque OM, Miwa H, Yasuda M, Fujii Y, Kaneko T, Sato S, Okazaki S. 2015. Identification of *Bradyrhizobium elkanii* Genes Involved in Incompatibility with Soybean Plants Carrying the *Rj4* Allele. *Applied and Environmental Microbiology* **81**: 6710–6717.

Fauvart M, Michiels J. 2008. Rhizobial secreted proteins as determinants of host specificity in the rhizobium–legume symbiosis. *FEMS microbiology letters* **285**: 1–9.

Gaudioso-Pedraza R, Beck M, Frances L, Kirk P, Ripodas C, Niebel A, Oldroyd GED, Benitez-Alfonso Y, de Carvalho-Niebel F. 2018. Callose-Regulated Symplastic Communication Coordinates Symbiotic Root Nodule Development. *Current biology: CB* **28**: 3562–3577.e6.

Genre A, Chabaud M, Balzergue C, Puech-Pagès V, Novero M, Rey T, Fournier J, Rochange S, Bécard G, Bonfante P, *et al.* 2013. Short-chain chitin oligomers from arbuscular mycorrhizal fungi trigger nuclear C a2+ spiking in *Medicago truncatula* roots and their production is enhanced by strigolactone. *The New phytologist* **198**: 190–202.

Geurts R, Xiao TT, Reinhold-Hurek B. 2016. What Does It Take to Evolve A Nitrogen-Fixing Endosymbiosis? *Trends in plant science* **21**: 199–208.

Gherbi H, Markmann K, Svistoonoff S, Estevan J, Autran D, Giczey G, Auguy F, Péret B, Laplaze L, Franche C, *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.

Granqvist E, Sun J, Op Den Camp R, Pujic P, Hill L, Normand P, Morris RJ, Downie JA, Geurts R, Oldroyd GED, *et al.* 2015. Bacterial-induced calcium oscillations are common to nitrogen-fixing associations of nodulating legumes and nonlegumes. *The New phytologist* **207**: 551–558.

Hayashi M, Shiro S, Kanamori H, Mori-Hosokawa S, Sasaki-Yamagata H, Sayama T, Nishioka M, Takahashi M, Ishimoto M, Katayose Y, *et al.* 2014. A thaumatin-like protein, *Rj4*, controls nodule symbiotic specificity in soybean. *Plant & cell physiology* **55**: 1679–1689.

Hocher V, Alloisio N, Auguy F, Fournier P, Doumas P, Pujic P, Gherbi H, Queiroux C, Da Silva C, Wincker P, *et al.* 2011. Transcriptomics of actinorhizal symbioses reveals homologs of the whole common symbiotic signaling cascade. *Plant physiology* **156**: 700–711.

Hubber A, Vergunst AC, Sullivan JT, Hooykaas PJJ, Ronson CW. 2004. Symbiotic phenotypes and translocated effector proteins of the *Mesorhizobium loti* strain R7A VirB/D4 type IV secretion system. *Molecular microbiology* **54**: 561–574.

Kawaharada Y, Kelly S, Nielsen MW, Hjuler CT, Gysel K, Muszyński A, Carlson RW, Thygesen MB, Sandal N, Asmussen MH, *et al.* 2015. Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* **523**: 308–312.

Kawaharada Y, Nielsen MW, Kelly S, James EK, Andersen KR, Rasmussen SR, Füchtbauer W, Madsen LH, Heckmann AB, Radutoiu S, *et al.* 2017. Differential regulation of the Epr3 receptor coordinates membrane-restricted rhizobial colonization of root nodule primordia. *Nature communications* **8**: 14534.

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

MacNair MR, Christie P. 1983. Reproductive isolation as a pleiotropic effect of copper tolerance in *Mimulus guttatus*? *Heredity* **50**: 295–302.

Madsen EB, Madsen LH, Radutoiu S, Olbryt M, Rakwalska M, Szczyglowski K, Sato S, Kaneko T, Tabata S, Sandal N, *et al.* 2003. A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. *Nature* **425**: 637–640.

Maillet F, Poinsot V, André O, Puech-Pagès V, Haouy A, Gueunier M, Cromer L, Giraudet D, Formey D, Niebel A, *et al.* 2011. Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature* **469**: 58–63.

Marie C, Broughton WJ, Deakin WJ. 2001. Rhizobium type III secretion systems: legume charmers or alarmers? *Current opinion in plant biology* **4**: 336–342.

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

Moling S, Pietraszewska-Bogiel A, Postma M, Fedorova E, Hink MA, Limpens E, Gadella TWJ, Bisseling T. 2014. Nod factor receptors form heteromeric complexes and are essential for intracellular infection in medicago nodules. *The Plant cell* **26**: 4188–4199.

Mithöfer A. **2002**. Suppression of plant defence in rhizobia–legume symbiosis. *Trends in plant science* **7**: 440–444.

Martínez-Hidalgo P, Ramírez-Bahena MH, Flores-Félix JD, Igual JM, Sanjuán J, León-Barrios M, Peix A, Velázquez E. **2016**. Reclassification of strains MAFF 303099T and R7A into Mesorhizobium japonicum sp. nov. *International journal of systematic and evolutionary microbiology* **66**: 4936–4941.

Nelson MS, Sadowsky MJ. **2015**. Secretion systems and signal exchange between nitrogen-fixing rhizobia and legumes. *Frontiers in plant science* **6**: 491.

Normand P, Queiroux C, Tisa LS, Benson DR, Rouy Z, Cruveiller S, Médigue C. **2007**. Exploring the genomes of Frankia. *Physiologia plantarum* **130**: 331–343.

Okazaki S, Kaneko T, Sato S, Saeki K. **2013**. Hijacking of leguminous nodulation signaling by the rhizobial type III secretion system. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 17131–17136.

Okazaki S, Tittabutr P, Teulet A, Thouin J, Fardoux J, Chaintreuil C, Gully D, Arrighi J-F, Furuta N, Miwa H, *et al.* 2016. Rhizobium–legume symbiosis in the absence of Nod factors: two possible scenarios with or without the T3SS. *The ISME journal* **10**: 64–74.

Okazaki S, Okabe S, Higashi M, Shimoda Y, Sato S, Tabata S, Hashiguchi M, Akashi R, Göttfert M, Saeki K. **2010**. Identification and functional analysis of type III effector proteins in *Mesorhizobium loti*. *Molecular plant-microbe interactions: MPMI* **23**: 223–234.

Oldroyd GED. **2013**. Speak , friend , and enter : signalling systems that promote beneficial symbiotic associations in plants. *Nature Publishing Group* **11**: 252–263.

Oldroyd GED, Dixon R. **2014**. Biotechnological solutions to the nitrogen problem. *Current opinion in biotechnology* **26**: 19–24.

Oldroyd GED, Murray JD, Poole PS, Downie JA. **2011**. The Rules of Engagement in the Legume-Rhizobial Symbiosis. *Annual review of genetics* **45**: 119–144.

Op den Camp RHM, Polone E, Fedorova E, Roelofsen W, Squartini A, Op den Camp HJM, Bisseling T, Geurts R. **2012**. Nonlegume *Parasponia andersonii* deploys a broad rhizobium host range strategy resulting in largely variable symbiotic effectiveness. *Molecular plant-microbe interactions: MPMI* **25**: 954–963.

Op den Camp R, Streng A, De Mita S, Cao Q, Polone E, Liu W, Ammiraju JSS, Kudrna D, Wing R, Untergasser A, *et al.* 2011a. LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume *Parasponia*. *Science* **331**: 909–912.

Ormeño-Orrillo E, Menna P, Almeida LGP, Ollero FJ, Nicolás MF, Rodrigues EP, Nakatani AS, Batista JSS, Chueire LMO, Souza RC, *et al.* 2012. Genomic basis of broad host range and environmental adaptability of *Rhizobium tropici* CIAT 899 and *Rhizobium sp.* PRF 81 which are used in inoculants for common bean (*Phaseolus vulgaris L.*). *BMC Genomics* **13**: 735.

Persson T, Battenberg K, Demina IV, Vigil-Stenman T, Heuvel BV, Pujic P, Facciotti MT, Wilbanks EG, O'Brien A, Fournier P, *et al.* 2015. Candidatus *Frankia datiscae* Dg1, the actinobacterial microsymbiont of *Datisca glomerata*, expresses the canonical nod genes

nodABC in symbiosis with its host plant. *PloS one* **10**: e0127630.

Radutoiu S, Madsen LH, Madsen EB, Felle HH, Umehara Y, Grønlund M, Sato S, Nakamura Y, Tabata S, Sandal N, *et al.* 2003. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* **425**: 585–592.

Roy S, Liu W, Nandety RS, Crook AD, Mysore KS, Pislariu CI, Frugoli JA, Dickstein R, Udvardi MK. 2019. Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. *The Plant cell*.

Saad MM, Crèvecoeur M, Masson-Boivin C, Perret X. 2012. The type 3 protein secretion system of *Cupriavidus taiwanensis* strain LMG19424 compromises symbiosis with *Leucaena leucocephala*. *Applied and environmental microbiology* **78**: 7476–7479.

Sadowsky MJ, Cregan PB. 1992. The Soybean Rj4 Allele Restricts Nodulation by *Bradyrhizobium japonicum* Serogroup 123 Strains. *Applied and environmental microbiology* **58**: 720–723.

Sharma RC, Schimke RT. 1996. Preparation of Electro-Competent E. coli Using Salt-Free Growth Medium. *BioTechniques* **20**: 42–44.

Sullivan JT, Trzebiatowski JR, Cruickshank RW, Gouzy J, Brown SD, Elliot RM, Fleetwood DJ, McCallum NG, Rossbach U, Stuart GS, *et al.* 2002. Comparative sequence analysis of the symbiosis island of Mesorhizobium loti strain R7A. *Journal of bacteriology* **184**: 3086–3095.

Spaink HP, Okker RJ, Wijffelman C a., Tak T, Goosen-de Roo L, Pees E, van Brussel A a., Lugtenberg BJ. 1989. Symbiotic properties of rhizobia containing a flavonoid-independent hybrid nodD product. *Journal of bacteriology* **171**: 4045–4053.

Sutton MA, Oenema O, Erisman JW, Leip A, van Grinsven H, Winiwarter W. 2011. Too much of a good thing. *Nature* **472**: 159–161.

Svistoonoff S, Benabdoun FM, Nambiar-Veetil M, Imanishi L, Vaissayre V, Cesari S, Diagne N, Hocher V, de Billy F, Bonneau J, *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.

Svistoonoff S, Hocher V, Gherbi H. 2014. Actinorhizal root nodule symbioses: what is signalling telling on the origins of nodulation? *Current opinion in plant biology* **20**: 11–18.

Teulet A, Busset N, Fardoux J, Gully D, Chaintreuil C, Cartieaux F, Jauneau A, Comorge V, Okazaki S, Kaneko T, *et al.* 2019. The rhizobial type III effector ErnA confers the ability to form nodules in legumes. *Proceedings of the National Academy of Sciences of the United States of America* **116**: 21758–21768.

Tromas A, Parizot B, Diagne N, Champion A, Hocher V, Cissoko M, Crabos A, Prodjinoto H, Lahouze B, Bogusz D, *et al.* 2012. Heart of endosymbioses: transcriptomics reveals a conserved genetic program among arbuscular mycorrhizal, actinorhizal and legume-rhizobial symbioses. *PloS one* **7**: e44742.

Tsurumaru H, Hashimoto S, Okizaki K, Kanesaki Y, Yoshikawa H, Yamakawa T. 2015. A Putative Type III Secretion System Effector Encoded by theMA20_12780Gene in

Bradyrhizobium japonicum Is-34 Causes Incompatibility with Rj4 Genotype Soybeans. *Applied and Environmental Microbiology* **81**: 5812–5819.

Van Nguyen T, Wibberg D, Battenberg K, Blom J, Vanden Heuvel B, Berry AM, Kalinowski J, Pawlowski K. **2016**. An assemblage of FrankiaCluster II strains from California contains the canonical nodgenes and also the sulfotransferase gene nodH. *BMC genomics* **17**: 796.

van Velzen R, Doyle JJ, Geurts R. **2019**. A Resurrected Scenario: Single Gain and Massive Loss of Nitrogen-Fixing Nodulation. *Trends in plant science* **24**: 49–57.

van Velzen R, Holmer R, Bu F, Rutten L, van Zeijl A, Liu W, Santuari L, Cao Q, Sharma T, Shen D, *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.

Vest G, Caldwell BE. **1972**. Rj4 — A Gene Conditioning Ineffective Nodulation in Soybean1. *Crop Science* **12**: 692.

Viprey V, Del Greco A, Golinowski W, Broughton WJ, Perret X. **1998**. Symbiotic implications of type III protein secretion machinery in Rhizobium. *Molecular microbiology* **28**: 1381–1389.

Wardhani TAK, Roswanjaya YP, Dupin S, Li H, Linders S, Hartog M, Geurts R, Van Zeijl A. **2019**. Transforming, genome editing and phenotyping the nitrogen-fixing tropical Cannabaceae tree *Parasponia andersonii*. *Journal of visualized experiments: JoVE*.

Worrell E, Phylipsen D, Einstein D, Martin N. **2000**. Energy use and energy intensity of the U.S. chemical industry.

Yang S, Tang F, Gao M, Krishnan HB, Zhu H. **2010**. R gene-controlled host specificity in the legume–rhizobia symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 18735–18740.

Yang M-Q, van Velzen R, Bakker FT, Sattarian A, Li D-Z, Yi T-S. **2013**. Molecular phylogenetics and character evolution of Cannabaceae. *Taxon* **62**: 473–485.

van Zeijl A, Wardhani TAK, Seifi Kalhor M, Rutten L, Bu F, Hartog M, Linders S, Fedorova EE, Bisseling T, Kohlen W, *et al.* 2018. CRISPR/Cas9-Mediated Mutagenesis of Four Putative Symbiosis Genes of the Tropical Tree *Parasponia andersonii* Reveals Novel Phenotypes. *Frontiers in plant science* **9**: 284.

van Zeijl A, Op den Camp RHM, Deinum EE, Charnikhova T, Franssen H, Op den Camp HJM, Bouwmeester H, Kohlen W, Bisseling T, Geurts R. **2015**. Rhizobium Lipo-chitooligosaccharide Signaling Triggers Accumulation of Cytokinins in *Medicago truncatula* Roots. *Molecular plant* **8**: 1213–1226.

Table S1: Primers used in this work.

pMP604 PCR_F	GACCAGAATTAGGCCGCTCT	genotyping of pMP604 transformants
pMP604 PCR_R	TTCCCAGCATAGCTTCCACT	
16S rRNA qPCR_F	AAGGCCCTAGGGTTGTAAAGC	qRT-PCR quantification of 16S *rRNA*
16S rRNA qPCR_R	AATTCCGAACAACGCTAGCC	
CIAT899 nodC_F	AATGTTGGAAAGCGCAAGGC	qRT-PCR quantification of *nodC* gene of *R. tropici* CIAT899
CIAT899 nodC_R	AGTGCAAGTTTCACGACGAC	
R7A nodA_F	ACCGCCTTCGAATTGCTTTC	RT-PCR quantification of *nodA* gene of *M. japonica* R7A
R7A nodA_R	GGAAAATGAGTTGCAGCTTCCC	
R7A nodB_F	AGACAAGGTTAAGCGTGCAG	T-PCR quantification of *nodB* gene of *M. japonica* R7A
R7A nodB_R	TTCGTCATTGGTGCTTACGC	
R7A nodC_F	TTGGGAGCGCAATGAAGTTG	T-PCR quantification of nodC gene of *M. japonica* R7A
R7A nodC_R	TGGCTTCCATTGCAAGTCAG	

Table S2: List of strains tested in this study.

	Genus	Species	Strain	Origin of strain	References	Host plant
α proteobacteria	*Mesorhizobium*	*plurifarium*	BOR2	Saba, Malaysia	1	*Trema orientalis*
	Rhizobium	*tropici*	CIAT899	Columbia	2	*Phaseolus vulgaris*
	Mesorhizobium	*japonica*	R7A[#]	New Zealand	2,3	*Lotus sp.*
	Bradyrhizobium	*elkanii*	WUR3	*potting soil*	4	*Chamaecrista fasiculata*
	Bradyrhizobium	*elkanii*	USDA61	USA	5	*Glycine max*
	Bradyrhizobium	*sp.*	ORS3257[*]	unknown	6	*Aeschynomene indica*
β proteobacteria	*Cupriviadus*	*tainwanensis*	LMG19424/CBM777	Taiwan, China	7	*Mimosa. pudica/ M. diplotricha*

#, *Mesorhizobium japonica* R7A was previously named *Mesorhizobium loti* R7A (Martínez-Hidalgo et al. 2016).

* *Bradyrhizobium sp.* ORS3257 was previously named *Bradyrhizobium sp.* STM6978 (Okazaki et al. 2016).

1, (van Velzen et al. 2018); 2, (Spaink et al. 1989); 3, (Sullivan et al. 2002); 4, (Op den Camp et al. 2012); 5, Okazaki et al. 2013; 6, Okazaki et al. 2016; 7, Saad et al. 2012.

Table S3: Presenting of type III/IV secretion system in rhizobium strains and corresponding mutants used in this study.

	Genus	Species	Strain	Type III/IV secretion system mutant	Resources
Type IV secretion system	*Mesorhizobium*	*japonica*	R7A AH34	trbE::tn5	1
	Mesorhizobium	*plurifarium*	BOR2	not available	this study
Type III secretion system	*Cupriviadus*	*tainwanensis*	LMG19424	CBM312	2
	Rhizobium	*tropici*	CIAT899	not tested	3
	Bradyrhizobium	*elkanii*	WUR3	not available	this study
	Bradyrhizobium	*elkanii*	USDA61	△T3SS	4
	Bradyrhizobium	*sp.*	ORS3257	△T3SS	5

1, Hubber et al. 2004; 2, Saad et al. 2012; 3, Ormeño-Orrillo et al. 2012; 4, Okazaki et al. 2013; 5, Okazaki et al. 2016.

Table S4: Presenting of genes encoding the T3SS or T4SS secretion system within *Bradyrhizobium elkanii* WUR3 and *Mesorhizobium plurifarium*.

	Gene name	*Bradyrhizobium elkanii* WUR3	*Mesorhizobium plurifarium* BOR2
Type III Secretion System	*rchU*	YES	NO
	rchT	YES	NO
	rchS	YES	NO
	rchR	YES	NO
	rchQ	YES	NO
	rchN	YES	NO
	rchV	YES	NO
	rchJ	YES	NO
	nolU	YES	NO
	rhcC2	YES	NO
	ttsI	YES	NO
Type IV Secretion System	*trbI*	NO	YES
	trbG	NO	YES
	trbF	NO	YES
	trbL	NO	YES
	trbJ	NO	YES
	trbE	NO	YES
	trbB	NO	YES
	CopG	NO	YES
	traG	NO	YES
	traI	NO	YES

Chapter 4

Mutant analysis in the non-legume *Parasponia andersonii* identifies NIN and NF-YA1 transcription factors as a core genetic network in nitrogen-fixing nodule symbioses

Fengjiao Bu, Luuk Rutten, Yuda Purwana Roswanjaya, Olga Kulikova, Marta Rodriguez-Franco, Thomas Ott, Ton Bisseling, Arjan van Zeijl & Rene Geurts, *New Phytologist*. Accepted

Abstract

Nitrogen-fixing nodulation occurs in ten taxonomic lineages, either with rhizobia or *Frankia* bacteria. To establish such an endosymbiosis, two processes are essential: nodule organogenesis and intracellular bacterial infection. In the legume-rhizobium endosymbiosis, both processes are guarded by the transcription factor NODULE INCEPTION (NIN) and its downstream target genes of the NUCLEAR FACTOR Y (NF-Y) complex.It is hypothesized that nodulation has a single evolutionary origin ~110 million years ago, followed by many independent losses. Despite a significant body of knowledge of the legume-rhizobium symbiosis, it remains elusive which signalling modules are shared between nodulating species in different taxonomic clades. We used *Parasponia andersonii* to investigate the role of *NIN* and *NF-YA* genes in rhizobium nodulation in a non-legume system.Consistent with legumes, *P. andersonii PanNIN* and *PanNF-YA1* are co-expressed in nodules. By analyzing single, double and higher-order CRISPR-Cas9 knockout mutants, we show that nodule organogenesis and early symbiotic expression of *PanNF-YA1* are PanNIN-dependent and that PanNF-YA1 is specifically required for intracellular rhizobium infection.This demonstrates that *NIN* and *NF-YA1* commit conserved symbiotic functions. As *Parasponia* and legumes diverged soon after the birth of the nodulation trait, we argue that NIN and NF-YA1 represent core transcriptional regulators in this symbiosis.

Key words: nodulation, evolution, intracellular infection, NIN, NF-YA1, Parasponia, rhizobium

Introduction

Nitrogen is an essential element for plant growth. To cope with nitrogen limitation, some plant species engage with nitrogen-fixing rhizobium or *Frankia* bacteria. These bacteria colonize cells of specialized root organs, called nodules. Inside nodule cells, the bacteria convert atmospheric nitrogen into ammonium that can be exploited by the plant. Plant species capable of forming nitrogen-fixing nodules all belong to one of the four orders Fabales, Fagales, Cucurbitales and Rosales that together form the so-called nitrogen-fixing clade (Soltis *et al.*, 1995; Doyle, 2011). Within this clade, nodulation is limited to ten lineages, of which eight nodulate with *Frankia* and two with rhizobia (Geurts *et al.*, 2012). The nodulating lineages within the nitrogen-fixing clade are interspersed among tens of non-nodulating lineages. The current hypothesis is that this scattered distribution originates from a single evolutionary gain of nodulation in the ancestor to the nitrogen-fixing clade, and subsequent loss of this trait in many descending species (van Velzen *et al.*, 2018, 2019; Griesmann *et al.*, 2018). Such a scenario implies that the nodulation trait in all ten lineages is based on conserved genetic networks.

Rhizobium-induced nodulation occurs in two lineages; *Parasponia* (Cannabaceae, Rosales) and legumes (Fabaceae, Fabales). These lineages diverged >100 million years ago and even though the capacity to live in endosymbiosis with diazotrophic bacteria may have been the result of a shared evolutionary event, *Parasponia* and legumes likely acquired rhizobium as a microsymbiont in parallel (van Velzen *et al.*, 2018, 2019). The molecular and genetic aspects of rhizobium-induced nodulation have been extensively studied in a number of legume species; e.g. pea (*Pisum sativum*), *Medicago truncatula* and *Lotus japonicus*, whereas some data are also available for *Parasponia*. To initiate symbiosis, most rhizobium bacteria excrete lipo-chitooligosaccharide (LCO) signals that are perceived by plant LysM-type receptor kinases (Lerouge *et al.*, 1990; Dénarié *et al.*, 1996; Madsen *et al.*, 2003; Radutoiu *et al.*, 2003; Limpens *et al.*, 2003; Op den Camp *et al.*, 2011). LCO perception activates the so-called 'common symbiosis signalling pathway', which is co-opted from arbuscular mycorrhizal symbiosis (Oldroyd, 2013). Downstream of the common symbiosis signalling pathway, it culminates in the activation of a suite of transcriptional regulators (Soyano & Hayashi, 2014). Among these are NODULE INCEPTION (NIN) and its downstream targets of the NUCLEAR FACTOR Y (NF-Y) complex that are essential for nodule organogenesis and rhizobium infection and among the first genes

transcriptionally induced (Schauser *et al.*, 1999; Combier *et al.*, 2006; Marsh *et al.*, 2007; Soyano *et al.*, 2013; Rípodas *et al.*, 2014; Vernié *et al.*, 2015).

NF-Y complexes are heterotrimeric transcription factors composed of the NF-YC, NF-YB and NF-YA subunits, of which the latter determines the DNA-binding specificity (Baudin *et al.*, 2015; Myers & Holt, 2018). In plants, each of these subunits is encoded by a small family and in legumes several NF-Y-encoding genes display a nodule-enhanced expression profile (Laloum *et al.*, 2013; Baudin *et al.*, 2015). Mutant analysis in *L. japonicus* and *M. truncatula* revealed that *NF-YA1* is required for nodule development (Combier *et al.*, 2006; Soyano *et al.*, 2013; Laporte *et al.*, 2014; Laloum *et al.*, 2014; Hossain *et al.*, 2016). In *L. japonicus nf-ya1* mutants, most nodules do not progress beyond the primordial stage, whereas *M. truncatula nf-ya1* mutants develop nodules of variable size, but all remain substantially smaller than wild-type nodules (Combier *et al.*, 2006; Hossain *et al.*, 2016). The latter is most probably due to disturbed formation of the nodule apical meristem (Combier *et al.*, 2006; Laporte *et al.*, 2014; Laloum *et al.*, 2014; Xiao *et al.*, 2014). Besides problems in nodule organogenesis, *M. truncatula nf-ya1* mutants are also affected in the formation of intracellular infection threads (Laporte *et al.*, 2014). These infection threads initiate at the tip of a root hair and function to guide rhizobium bacteria to the underlying nodule primordium, which is formed in the root cortex. In *M. truncatula nf-ya1* mutants, infection thread progression is hampered and infection thread growth is frequently arrested in the epidermal layer (Laporte *et al.*, 2014). In *L. japonicus*, *Ljnf-ya1* knockdown lines display only a very weak infection phenotype (Soyano *et al.*, 2013; Hossain *et al.*, 2016). Taken-together, this shows that in legumes *NF-YA* genes function during rhizobia infection and nodule organogenesis.

In legumes, *NIN* is among the first genes transcriptionally activated upon rhizobium LCO signalling, which is acting downstream of the common symbiosis signalling pathway, and is essential as well as sufficient to initiate nodule organogenesis (Schauser *et al.*, 1999; Borisov *et al.*, 2003; Marsh *et al.*, 2007; Soyano *et al.*, 2013). NIN belongs to a small family of NIN-Like Proteins (NLPs), of which in *Arabidopsis thaliana* several members are involved in nitrate signalling (Schauser *et al.*, 2005; Castaings *et al.*, 2009; Konishi & Yanagisawa, 2013). Orthologues of *NIN* are found across eudicots, but within the nitrogen-fixation clade functional copies of this gene have been repeatedly lost from the genomes of non-nodulating species (van Velzen *et al.*, 2018; Griesmann *et al.*, 2018). This suggests that within the nitrogen-fixation clade

NIN predominantly performs a nodulation-specific function. The first indication that this is indeed the case is obtained from *Agrobacterium tumefaciens*-mediated stable transformation knockdown studies in *Casuarina glauca*, which resulted in a reduced nodulation efficiency when inoculated with *Frankia* (Clavijo *et al.*, 2015). However, such functional studies to prove that NIN -and its subsequent *NF-YA* target genes- commits key symbiotic roles in nodulating lineages other than legumes remain scarce.

We aimed to use *Parasponia* to investigate to what extent NIN and NF-YA transcriptional regulators commit conserved functions in root nodule formation. Previous studies showed that *NIN* and *NF-YA1* are transcriptionally induced in *Parasponia andersonii* nodules (van Velzen *et al.*, 2018). By creating a series of CRISPR-Cas9 knockout mutants, we provide evidence that *PanNIN* is essential for nodule initiation in the non-legume *P. andersonii*. Furthermore, we show that *PanNF-YA1* is specifically required for intracellular rhizobium infection, whereas nodule organogenesis is controlled by a genetically redundant network of *NF-YA* genes. Taken together, this suggests that *NIN* and *NF-YA1* are part of a core genetic network essential for rhizobium symbiosis in legumes and non-legume species.

Results

P. andersonii NIN and *NF-YA1* are co-expressed during nodule formation

Previously conducted transcriptome studies revealed that *PanNIN* and *PanNF-YA1* have a nodule-enhanced expression profile in *P. andersonii* (van Velzen *et al.*, 2018). To get a first insight into the spatiotemporal expression pattern of both genes, we conducted promoter reporter and/or *in situ* hybridization experiments. To this end, a 3.8-kb sequence upstream of the translational start site of *PanNF-YA1*, containing the putative promoter sequence and the 5'-UTR that includes the first intron, was fused to a β-glucuronidase (GUS)-encoding sequence. The resulting construct was introduced into the *P. andersonii* genome using *Agrobacterium tumefaciens*-mediated stable transformation (van Zeijl *et al.*, 2018). Five lines were selected, for which we compared the GUS reporter activity under symbiotic and non-symbiotic conditions. Four of these lines yielded comparable results, therefore one of these lines (line 1E5) was selected for detailed characterization.

Under sterile conditions, activity of the *PanNF-YA1*$_{pro}$:*GUS* was observed around the vasculature of differentiated root tissue (Fig. S1a,b). Root sections revealed that GUS staining is restricted to the pericycle cells opposite to the protoxylem, but absent from lateral root primordia (Fig. S1b-d). In *M. truncatula*, similar promoter-GUS studies using a 2.2 kb-upstream region revealed that *MtNF-YA1* is induced in root hairs of the pre-infection zone and in the root pericycle upon rhizobium inoculation (Laporte *et al.*, 2014; Liu *et al.*, 2019). We questioned whether this is also the case for *P. andersonii*. To determine this, transgenic plantlets expressing the *PanNF-YA1*$_{pro}$:*GUS* reporter were grown *in vitro* on nitrogen-poor medium (0.375 mM NH_4NO_3) and inoculated with *Mesorhizobium plurifarium* BOR2. In contrast to legumes like *M. truncatula* and *L. japonicus*, *Parasponia* species are not infected via curled root hairs. Instead rhizobia enter apoplastically via cracks that are formed upon cell divisions in the epidermis and outer cortex and only infect intracellularly when a nodule primordium is formed (Lancelle & Torrey, 1984, 1985). At 2 days post inoculation (dpi), *PanNF-YA1*$_{pro}$:*GUS* activity was observed in epidermal and cortical cells located just above the root elongation zone (Fig. S1e). *PanNF-YA1*$_{pro}$*:GUS* is active in clumps of multicellular root hairs and adjacent cortical cells as well as dividing pericycle-derived cells (Fig. 1a; Fig. S1f,g). The formation of multicellular root hairs is one of the earliest responses associated with nodule initiation in *Parasponia* species, and is not observed in non-

inoculated roots (Lancelle & Torrey, 1984, 1985). In young nodule primordia that are visible as small bumps on the root (5 dpi), the *PanNF-YA1*$_{pro}$:*GUS* reporter was highly active in clusters of dividing cells (Fig. 1b). Additionally, activity was observed in dividing pericycle cells that flank the developing nodule vascular bundle (Fig. 1b). In young nodules, *PanNF-YA1*$_{pro}$:*GUS* activity is observed in the central region of the nodule lobes, where intracellular infection by rhizobium will occur (Fig. 1c). In mature nodules, *PanNF-YA1*$_{pro}$:*GUS* activity was mostly confined to the infection zone (Fig. 1d). Additionally, weaker activity is observed in the cell layers surrounding the nodule vascular bundle (Fig. 1c,d). Taken-together, the expression pattern of the *PanNF-YA1*$_{pro}$:*GUS* reporter suggests a symbiotic role of *PanNF-YA1*.

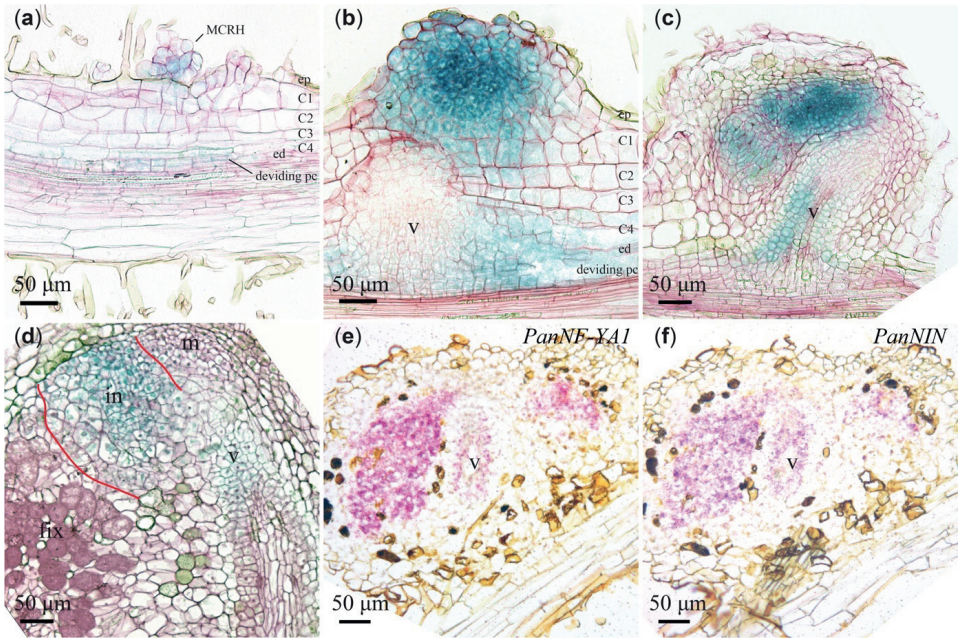


Fig. 1 Spatiotemporal expression pattern of *PanNF-YA1* and *PanNIN* in developing *Parasponia andersonii* root nodules. (**a-d**) Spatiotemporal expression pattern of *PanNF-YA1*pro:GUS in nodules of different developmental stages. (**e, f**) Spatiotemporal expression pattern of *PanNF-YA1* and *PanNIN* visualized by *in situ* hybridization on consecutive sections of a young *P. andersonii* nodule primordium. (**a**) *PanNF-YA1*pro:GUS activity in clustered root hairs that are associated with dividing epidermal, outer cortical and pericycle cells. (**b**) P*anNFYA1*pro:GUS activity in a young but not yet intracellularly infected nodule and in the pericycle derived cells flanking the developing nodule vasculature. (**c**) *PanNF-YA1*pro:GUS activity in the infection zone of young nodules, and in the basal part of the nodule vasculature. (**d**) *PanNF-YA1*pro:GUS activity in a mature nodule is restricted to the infection zone (marked with red lines) and nodule vasculature. *PanNF-YA1* (**e**) and *PanNIN* (**f**) transcripts are detected in the infection zone

and nodule vasculature by *in situ* hybridization on consecutive sections. MCRH: multicellular root hairs. ep: epidermis, C1-C4: 1st to 4th cortical cell layer; ed: endodermis; pc: pericycle; m: nodule meristem; in: infection zone; fix: fixation zone; v: nodule vasculature. Sections (7 µm) were counterstained with Ruthenium Red for a-d. Nodules were isolated 4 weeks post-inoculation with *Mesorhizobium plurifarium* BOR2.

Next, we determined whether *PanNF-YA1* is co-expressed with *PanNIN* in *P. andersonii* nodules. As regulation of *NIN* in legumes showed to be highly complex and determined by distant *cis*-regulatory elements (Heckmann *et al.*, 2011; Popp & Ott, 2011; Kosuta *et al.*, 2011; Yoro *et al.*, 2014; Soyano *et al.*, 2014; Liu *et al.*, 2019), we decided to use RNA *in situ* hybridization. This method showed the accumulation of the *PanNIN* transcripts in the central region of the lobes where rhizobium infection will take place and in the pericycle/endodermis of the vasculature of young nodules (Fig. 1f). *In situ* hybridization on a consecutive section of the same nodule showed that the *PanNF-YA1* transcripts are present in the same cells as *PanNIN* (Fig. 1e,f), and that transcript accumulation is consistent with the activity of the *PanNF-YA1*$_{pro}$*:GUS* reporter in a nodule of a similar developmental stage (Fig. 1c). Therefore, we conclude that *PanNIN* and *PanNF-YA1* are co-expressed in young nodules.

PanNIN is essential for nodule formation and symbiotic expression of _PanNF-YA1_

To determine whether *PanNIN* is essential for nodule formation in *P. andersonii*, we created *Pannin* knockout mutants using CRISPR/Cas9-mediated mutagenesis. The *NIN* gene in *Parasponia* species produces two alternative transcript variants; (i) using a transcriptional initiation site at the 5'-end of the gene (*PanNIN.1*) and (ii) an alternative transcriptional initiation site for *PanNIN.2* located in the second intron of the gene (Fig. S2a) (van Velzen *et al.*, 2018). Quantification of RNAseq reads revealed that both transcripts are expressed in roots, whereas only expression of the long transcript (*PanNIN.1*) encoding a canonical NIN protein is enhanced in nodules (Fig. S2b). Therefore, we decided to create CRISPR-Cas9 mutants exclusively mutated in the long *NIN* transcript (*PanNIN.1*). Two knockout mutant lines (named B1 and B3) were obtained by targeting the first coding-exon using three single guide RNAs (sgRNAs) (Fig. S2c). These mutants contain premature stop codons at amino acid positions 90 (line B1) and 70 (line B3), respectively (Fig. S2d). Inoculation with *M. plurifarium* BOR2 showed that both lines are unable to form root nodules or even nodule primordia (Fig. 2c-d), whereas a transgenic control line

(CTR44) was well nodulated (Fig. 2a-b). This demonstrates that the *PanNIN.1* transcript is essential for nodule organogenesis in *P. andersonii*.

To determine whether rhizobium-induced *PanNF-YA1* expression is dependent on a functional PanNIN.1 protein, we conducted qRT-PCR experiments. Root RNA was isolated from a ~0.5 cm region encompassing part of the root elongation and differentiation zone at 1 dpi with a

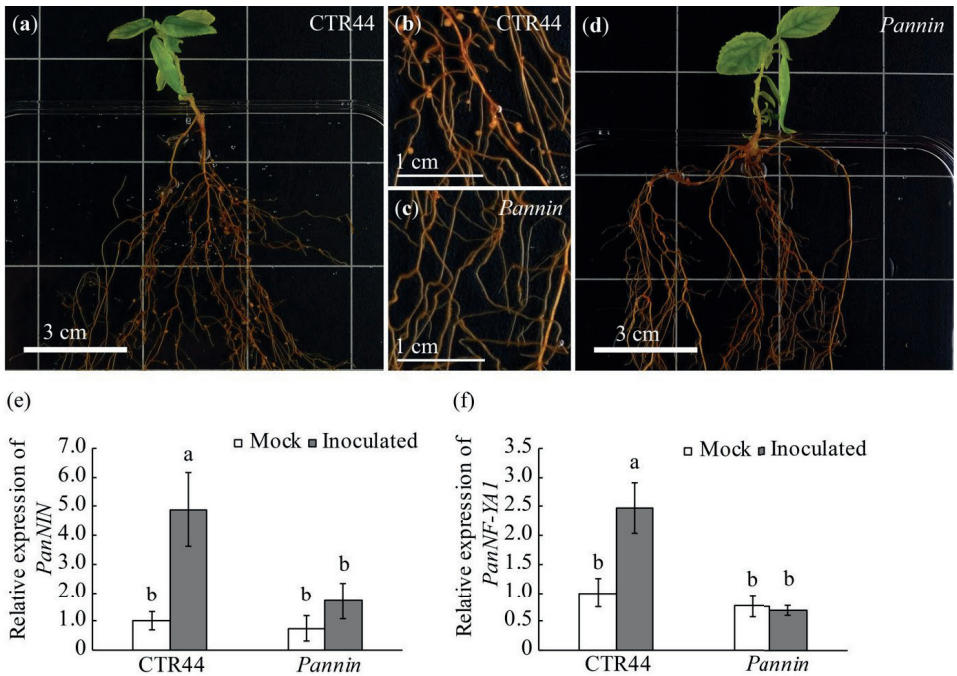


Fig. 2 Symbiotic phenotype of the *P. andersonii nin* mutant. (**a-d**) The symbiotic phenotype of the *Pannin* (line B3) knockout mutant. Shown are (**a, b**) a transgenic control (CTR44) and (**c, d**) a *Pannin* knockout mutant (line B3) at 4 weeks post-inoculation with *M. plurifarium* BOR2. Note that nodules are present on roots of the control (**a, b**), but not on *Pannin* mutant roots (n=50) (**c, d**). These images are representative results obtained from 3 independent experiments, combined >20 plants for each line. (**e**) Relative expression of *PanNIN* in non-inoculated and inoculated transgenic control (CTR44) and *Pannin* mutant (line B3) roots. (**f**) Relative expression of *PanNF-YA1* in non-inoculated and inoculated transgenic control (CTR44) and *Pannin* mutant (line B3) roots. RNA was isolated from root segments encompassing the elongation and part of the differentiation zone at 1 dpi with *Rhizobium tropici* CIAT899 pMP604. Data represent means of 2 independent experiments with a total of 5 biological replicates each ± SE. Data were normalized against the mock-treated CTR44 sample. Different letters indicate statistical significance (Student's t-test, $p < 0.05$).

compatible rhizobium strain that harbors a dominant active NodD protein that transcriptionally activates LCO biosynthesis genes (*Rhizobium tropici* CIAT899 pMP604) (Spaink *et al.*, 1989; Op den Camp *et al.*, 2012) (Fig. S3). In roots of transgenic control line CTR44, expression of *PanNIN.1* and *PanNF-YA1* was induced 5- and 2.5-fold following inoculation, respectively (Fig. 2e-f). In contrast, such induction of *PanNF-YA1* is not observed in *Pannin* mutant roots (Fig. 2e-f). This indicates that the early symbiotic induction of *PanNF-YA1* is downstream of PanNIN.1.

***PanNF-YA1* is essential for rhizobium intracellular infection**

To determine the symbiotic role of *PanNF-YA1*, we mutated this gene using CRISPR/Cas9. To this end, sgRNAs were designed that target the first coding-exon of *PanNF-YA1* (Table S1, Fig. S4a). This allowed the isolation of *Pannf-ya1* knockout mutant line (Fig. S4b).

We noted that *Pannf-ya1* mutant shoots were somewhat more difficult to root (Fig. S5a-b), a phenotype we didn't observe with transgenic control nor *Pannin* mutant shoots. As it was reported previously that *NF-YA1* orthologous genes may function in root growth and lateral root formation (Soyano *et al.*, 2013; Sorin *et al.*, 2014), we quantified root development in the *Pannf-ya1-1* mutant line and transgenic control. This revealed that the *Pannf-ya1-1* mutant formed less lateral roots when compared to transgenic controls (Fig. S5c-f).

To determine the nodulation phenotype, the *Pannf-ya1-1* mutant line plants were grown in perlite and inoculated with *M. plurifarium* BOR2. This showed that *Pannf-ya1-1* can be nodulated at least as efficient as control plants (Fig. S6a). However, quantification of nitrogenase activity using the acetylene reductase assay (ARA) indicated that *Pannf-ya1-1* nodules are unable to fix nitrogen (Fig. S6b).

Next, we studied the cytoarchitecture of *Pannf-ya1-1* nodules using light microscopy as well as transmission electron microscopy. In wild-type *Parasponia*, rhizobium bacteria first colonize the apoplast of the nodule infection zone, after which they enter nearby cells through infection threads (Lancelle & Torrey, 1984) (Fig. 3a,b). *P. andersonii nf-ya1-1* mutant nodules display a wild-type cytology, but cells in the infection zone are devoid of intracellular infection threads (Fig. 3d,e). Instead, large apoplastic colonies of rhizobium can be seen that occasionally occupy dead host cells (Fig. 3e). Transmission electron microscopy showed that apoplastic rhizobia in wild-type nodules are embedded in a thin layer of secreted matrix material from where intracellular infection

can occur (Fig. 3c) (Trinick, 1979). In contrast, no such intracellular infections were observed in *Pannf-ya1-1* mutant nodules. Instead, rhizobium formed large apoplastic colonies embedded in a secreted matrix (Fig. 3f). This infection phenotype was confirmed in two additional *Pannf-ya1* mutant lines (Fig. S4c-d, Fig. S6c-d). Based on these results, we conclude that *PanNF-YA1* commits an essential role in intracellular infection thread formation in *Parasponia* nodules.

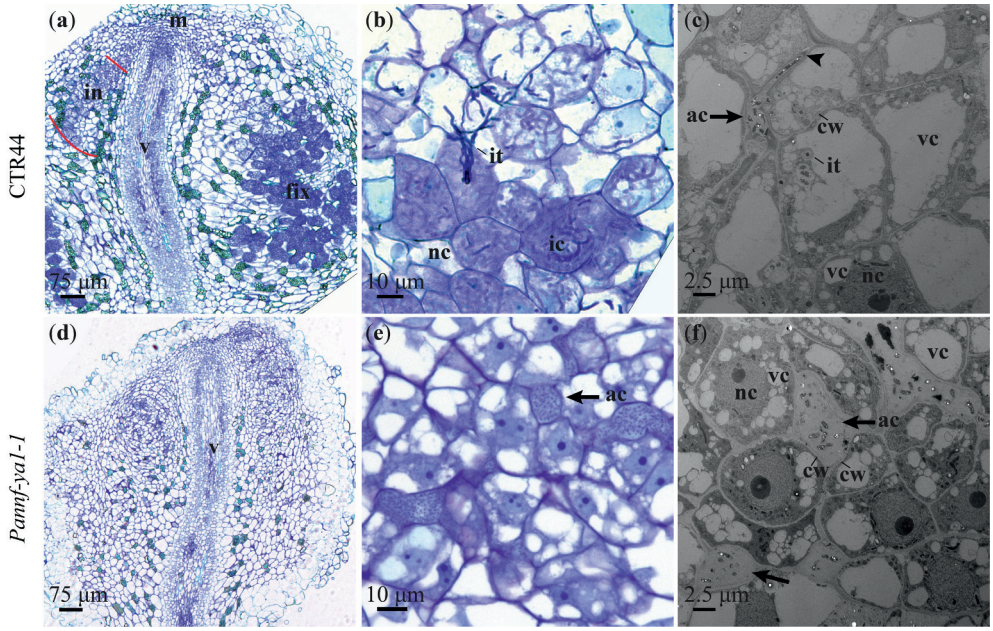


Fig. 3 *PanNF-YA1* is essential for intracellular rhizobium infection. (**a-b**) Nodule cytoarchitecture of *P. andersonii* transgenic control (CTR44) plants studied by light microscopy. (**a**) Sections of a mature transgenic control nodule. The infection zone (in) in one lobe is marked with red lines. (**b**) Formation of intracellular infection threads (arrowhead). Shown is a zoom in on the infection zone of a mature nodule. (**c**) Transmission electron microscopy image of apoplastic rhizobium infection (arrow) and initiation of intracellular infection (arrowheads) in a transgenic control nodule. (**d-e**) Cytoarchitecture of a Pannf-ya1 nodule studied by light microscopy. (**d**) *Pannf-ya1* mutant nodules lack intracellular infection threads. (**e**) In mature *Pannf-ya1-1* nodules, apoplastic colonies of rhizobium can be detected (arrow). (**f**) Transmission electron microscopy image of large apoplastic rhizobium colonies (arrows) in a *Pannf-ya1* mutant nodule. Plastic sections (a, b, d and e) were stained using Toluidine Blue. m: nodule meristem; in: infection zone; fix: fixation zone; v: nodule vasculature; it: intracellular infection thread; ic: infected cells; nc: non-infected cells; ac: apoplastic colonies of rhizobia; cw: cell wall; nc: nucleus; vc: vacuoles. Nodules were isolated 4 weeks post-inoculation with *M. plurifarium* BOR2.

PanNF-YA3 and *PanNF-YA6* are expressed during nodule formation

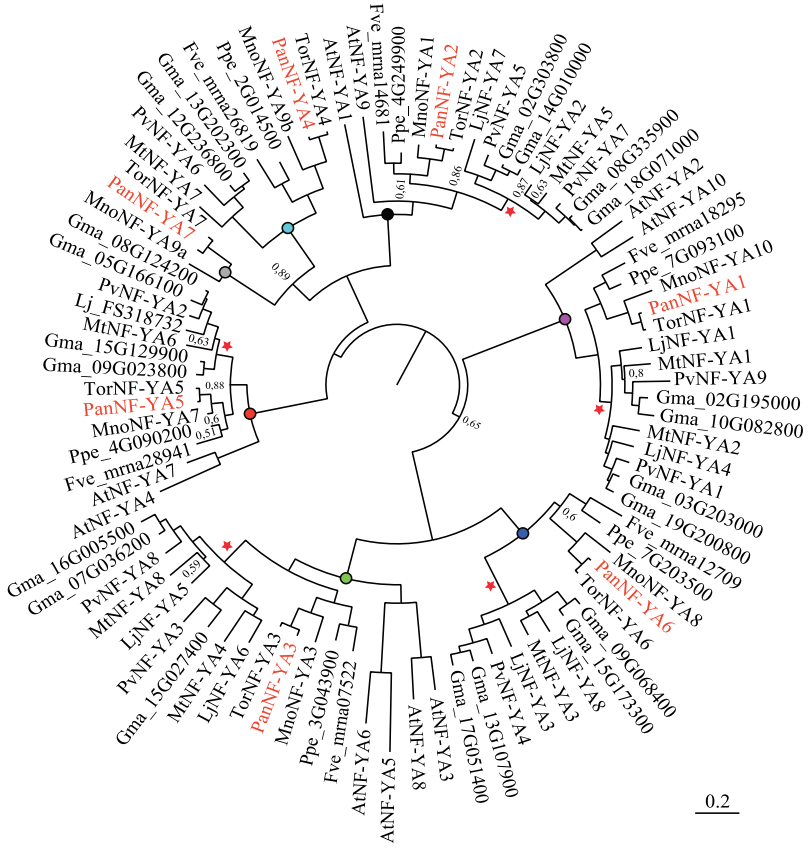
The *nf-ya1* mutants in *M. truncatula* and *L. japonicus* are clearly affected in nodule development (Combier *et al.*, 2006; Soyano *et al.*, 2013; Laporte *et al.*, 2014; Xiao *et al.*, 2014). In contrast, no such phenotype was observed in *P. andersonii nf-ya1-1* mutants. Therefore, we questioned whether additional *NF-YA*-encoding genes perform a symbiotic function in *Parasponia*.

To determine whether close paralogs of *PanNF-YA1* exist in *P. andersonii*, as has been reported for the model legume *M. truncatula* and *L. japonicus* (Laloum *et al.*, 2013; Soyano *et al.*, 2013), we reconstructed the phylogeny of the NF-YA clade. This revealed that *P. andersonii* possesses seven *NF-YA* genes that are divided over seven orthogroups (Fig. 4a, Fig. S7). We noted that legumes experienced gene duplication events in five orthogroups, including the *NF-YA1* lineage (Fig. 4a). In line with this, we conclude that *PanNF-YA1* is the sole orthologue of two legume genes represented by *MtNF-YA1*/*LjNF-YA1* and *MtNF-YA2*/*LjNF-YA4* in *M. truncatula* and *L.japonicus*. Additionally, we noted that legumes have genes only in six orthogroups, lacking an ortholog of *PanNF-YA7*. To determine whether gene duplications are specific to legumes we reconstructed the phylogeny also including *NF-YA* protein family of the actinorhizal plant species *Casuarina glauca* (Fagales) and *Datisca glomerata* (Cucurbitales), and the legume *Arachis duranensis*. This showed that *C. glauca* and *D. glomerata* possess generally a single gene in each of the seven orthogroups, similar as observed for *P. andersonii* supporting the conclusion that duplication of *NF-YA* genes in legumes is the result of a lineage-specific event (Fig. S7).

To study whether other *PanNF-YA* genes might function in rhizobium symbiosis, we determined their expression in nodules using published transcriptome data (van Velzen *et al.*, 2018). This revealed six *PanNF-YA* genes are expressed in nodules (transcripts per million (tpm) > 10), three of which show a nodule enhanced expression profile; namely *PanNF-YA1, PanNF-YA3,* and *PanNF-YA6*, respectively (Fig. 4b). To study the symbiotic expression of *PanNF-YA3* and *PanNF-YA6* in more detail, we created promoter-reporter GUS constructs for both genes. These constructs, contain 3.5 kb and 4.9 kb upstream of the translational start sites of *PanNF-YA3* and *PanNF-YA6*, respectively.

Transgenic *P. andersonii* lines harbouring these constructs revealed that the *PanNF-YA3*$_{pro}$*:GUS* construct is active in the root apical meristem (Fig. S8a). In case of *PanNF-YA6*, the promoter-reporter construct is expressed in young parts of the roots, including the meristem (Fig. S8e). Next, we studied their expression patterns following inoculation with rhizobium. In nodule primordia,

(a)


Fve_mrna18948
Ppe_4G249900
MnoNF-YA1
PanNF-YA2
TorNF-YA2
LjNF-YA7
PvNF-YA5
Gma_02G303800
LjNF-YA2
MtNF-YA5
PvNF-YA7
Gma_14G010000
Gma_08G335900
Gma_18G071000
AtNF-YA2
AtNF-YA10
AtNF-YA9
AtNF-YA1
TorNF-YA4
PanNF-YA4
MnoNF-YA9b
Ppe_2G014500
Fve_mrna26819
Gma_13G202300
Gma_12G256800
PvNF-YA6
MtNF-YA7
TorNF-YA7
PanNF-YA7
MnoNF-YA9a
Gma_08G124200
Gma_05G166100
PvNF-YA2
Lj_FS318732
MtNF-YA6
Gma_15G129900
Gma_09G023800
TorNF-YA5
PanNF-YA5
MnoNF-YA7
Ppe_4G090200
Fve_mrna28941
AtNF-YA7
AtNF-YA4
Fve_mrna12295
Fve_7G093100
Ppe_7G093100
MnoNF-YA10
PanNF-YA1
TorNF-YA1
LjNF-YA1
MtNF-YA1
PvNF-YA9
Gma_02G195000
Gma_10G082800
MtNF-YA2
LjNF-YA4
PvNF-YA1
Gma_03G203000
Gma_19G200800
Fve_mrna12709
Ppe_7G203500
MnoNF-YA8
PanNF-YA6
TorNF-YA6
Gma_09G068400
Gma_15G173300
LjNF-YA8
MtNF-YA3
LjNF-YA3
PvNF-YA4
Gma_13G107900
Gma_17G051400
AtNF-YA3
AtNF-YA8
AtNF-YA5
AtNF-YA6
Fve_mrna07522
Ppe_3G043900
MnoNF-YA3
PanNF-YA3
TorNF-YA3
LjNF-YA6
MtNF-YA4
Gma_15G027400
PvNF-YA3
LjNF-YA5
MtNF-YA5
PvNF-YA8
Gma_07G036200
Gma_16G005500
AtNF-YA8
0.2

(b)

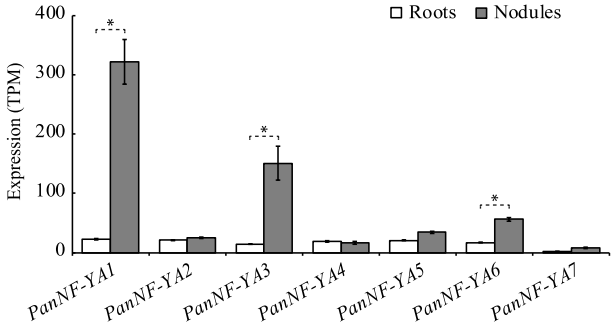

Roots
Nodules
Expression (TPM)
400
300
200
100
0
PanNF-YA1
PanNF-YA2
PanNF-YA3
PanNF-YA4
PanNF-YA5
PanNF-YA6
PanNF-YA7

Fig. 4 Phylogenetic relation and symbiotic expression of *Parasponia andersonii NF-YA* genes. (**a**) Bayesian phylogeny of NF-YA proteins reconstructed based on an alignment of protein sequences from the following species: *Parasponia andersonii* (Pan), *Trema orientalis* (Tor), *Arabidopsis thaliana* (At), *Medicago truncatula* (Mt), *Lotus japonicus* (Lj), *Glycine max* (Glyma), *Phaseolus vulgaris* (Pv), *Morus notabilis* (Mno), *Prunus persica* (Ppe), *Fragaria vesca* (Fve). *P. andersonii* NF-YA proteins are marked in red. Red pentagrams mark duplication events within the legume family. Orthogroups are indicated by a coloured circle. Node labels indicate posterior probability, Node labels with a value above 0.9 are not shown. (**b**) The expression level of *PanNF-YA* genes in roots and mature nodules. Expression was determined by quantification of RNAseq reads. Data represent average expression in transcripts per million (TPM) (n = 3) ± SD, which were obtained from van Velzen et al., 2018. Nodules were isolated 4 weeks post-inoculation with *M. plurifarium* BOR2. Asterisk: p-value < 0.01 adjusted for multiple testing based on false discovery rate estimated for 2 fold-change in mature nodule versus root sample as described by (van Velzen *et al.*, 2018).

we studied their expression patterns following inoculation with rhizobium. In nodule primordia, *PanNF-YA3*$_{pro}$:*GUS* is active in the dividing epidermal, cortical and pericycle cells, mimicking activity of the *PanNF-YA1*$_{pro}$:*GUS* reporter (Fig. 5a; Fig. S8b,c). In young nodules, *PanNF-YA3*$_{pro}$:*GUS* is expressed in the central region of the nodule lobes where rhizobium infection occurs, and in the vascular bundle (Fig. 5b). In mature nodules, *PanNF-YA3*$_{pro}$:*GUS* activity is observed in the infection zone and nodule vasculature (Fig. 5c; Fig. S8d). Activity of the *PanNF-YA6*$_{pro}$:*GUS* reporter is restricted to the nodule vascular meristem (Fig. 5d; Fig. S8f). *In situ* hybridization confirmed the expression patterns of *PanNF-YA3* and *PanNF-YA6* in young nodules (Fig. 5e,f). Additionally, it showed that *PanNF-YA3* is co-expressed with *PanNIN* in the lobes of young nodules (Fig. 5e; Fig. 1f). Therefore, we questioned whether symbiotic *PanNF-YA3* and/or *PanNF-YA6* expression requires a functional *PanNIN* gene. qRT-PCR experiments on the same samples used for studying *PanNF-YA1* expression revealed that neither *PanNF-YA3* nor *PanNF-YA6* expression is enhanced within 24 hr after inoculation (Fig. S8g,h).

Taken-together, these data suggest a possible symbiotic role for *PanNF-YA3* and to a lesser extent *PanNF-YA6,* though in roots both genes are not responsive to rhizobium inoculation (1 dpi).

PanNF-YA1, PanNF-YA3 and *PanNF-YA6* act redundantly in nodule development

To determine the role of *PanNF-YA3* and *PanNF-YA6* during *Parasponia* nodule formation, we created CRISPR/Cas9 mutants for both genes. *Pannf-ya3* and *Pannf-ya6* knockout mutant lines were created using three sgRNAs targeting the second and third exon, respectively (Fig. S9a-d). Inoculation with *M. plurifarium* BOR2 showed that *Pannf-ya3* and *Pannf-ya6* mutants developed

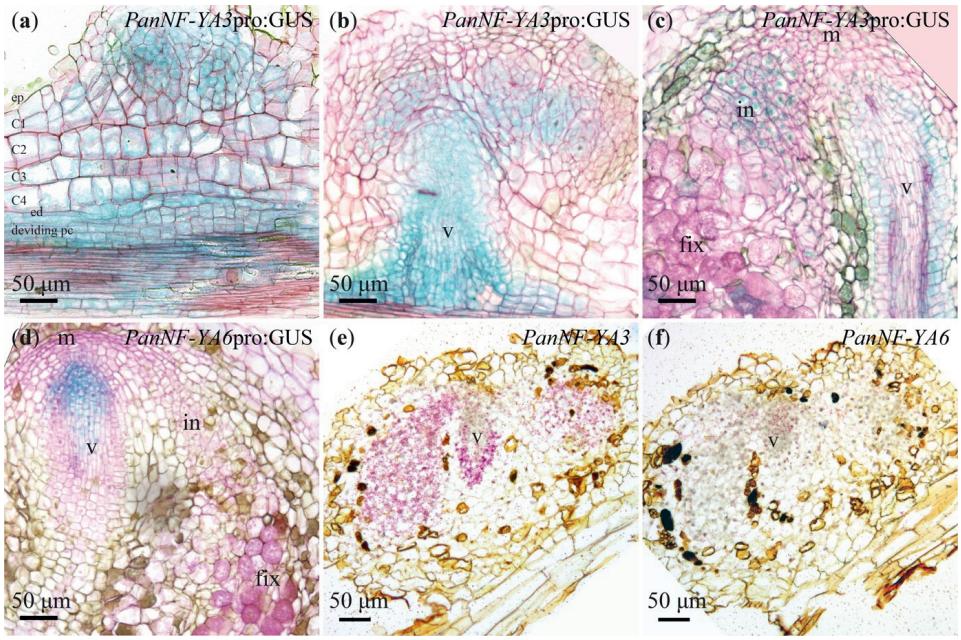


Fig. 5 Spatiotemporal expression pattern of *PanNF-YA3* and *PanNF-YA6* in *Parasponia andersonii* root nodules. (**a-c**) Spatiotemporal expression pattern of *PanNF-YA3*pro:GUS in nodules of different developmental stages. (**a**) *PanNF-YA3*pro:GUS activity is observed in dividing epidermal, cortical, endodermal cells of a nodule primordium as well as the root vasculature. (**b**) In a young nodule, *PanNF-YA3*pro:GUS activity is confined to the nodule lobes that will become intracellularly infected and the nodule vasculature. (**c**) In a mature nodule *PanNF-YA3*pro:GUS is active in the infection zone and the nodule vasculature (v). (**d**) *PanNFYA6*pro:GUS is active at the nodule vascular meristem. (**e, f**) Spatiotemporal expression pattern of *PanNFYA3* and *PanNF-YA6* visualized by *in situ* hybridization on consecutive sections of a young *P. andersonii* nodule primordium. ep: epidermis, C1-C4: 1st to 4th cortical cell layer; ed: endodermis; pc: pericycle; m: nodule meristem; in: infection zone; fix: fixation zone; v: nodule vasculature. Sections (7 µm) were counterstained with Ruthenium Red for a to d. Nodules were isolated 4 weeks post-inoculation with *M. plurifarium* BOR2.

a similar number of nodules as transgenic control plants (Fig. S9e). These mutant nodules were able to fix nitrogen, as determined by ARA (Fig. S9f), and display a wild-type cytoarchitecture (Fig. S9g,h). This indicates that neither *PanNF-YA3* nor *PanNF-YA6* is essential for *Parasponia* nodule formation.

As we cannot rule out that *PanNF-YA1, PanNF-YA3* and/or *PanNF-YA6* function redundantly in nodule organogenesis, we decided to create three double mutants (*Pannf-ya1;Pannf-ya3*, *Pannf-ya1;Pannf-ya6* and *Pannf-ya3;Pannf-ya6*), and higher order triple mutants (*Pannf-ya1;Pannf-*

ya3;Pannf-ya6) (Fig. S10). When inoculated with *M. plurifarium* BOR2, all three double mutant combinations formed nodules (Fig. S11a). Consistent with the phenotype of *Pannf-ya1* single mutant nodules, *Pannf-ya1;Pannf-ya3-1* and *Pannf-ya1;Pannf-ya6-6* double mutant nodules are devoid of intracellular infection structures (Fig. S12a,b,d,e). Intracellular infection in *Pannf-ya3;Pannf-ya6-5* double mutant nodules was not affected (Fig. S12c,f), indicating that intracellular rhizobium infection of *P. andersonii* nodules is specifically controlled by *PanNF-YA1*.

Next, we analysed the nodulation phenotype of three independent *Pannf-ya1;Pannf-ya3;Pannf-ya6* triple mutant lines. All three lines showed initiation of nodule organogenesis upon rhizobium inoculation with similar efficiency when compared to the transgenic control (Fig. S11a). However, *Pannf-ya1;Pannf-ya3;Pannf-ya6* triple mutants nodules were irregular in shape and remain substantially smaller than nodules formed on control (Fig. 6a-c; Fig. S11b). Approximately half of the *Pannf-ya1;Pannf-ya3;Pannf-ya6* triple mutant nodules do not develop beyond the primordial stage (Fig. 6a). These nodule-like structures originated from multiple rounds of cell divisions in the epidermis and outer cortex, but did not develop a vascular bundle (Fig. 6d). In contrast, the somewhat larger nodules formed on the *Pannf-ya1;Pannf-ya3;Pannf-ya6* triple mutant developed a nodule vascular bundle, but were disturbed in growth (Fig. 6b,e; Fig. S11b). In *M. truncatula*, it was shown that the casparian strip was absent from the nodule endodermis in the region close to meristem (Xiao *et al.*, 2014). We used this criterion to determine whether or not the nodule meristem of the *P. andersonii Pannf-ya1;Pannf-ya3;Pannf-ya6* triple mutants remained active. Nodule sections were examined under UV light to detect light emitted by the casparian strips. This showed that the meristematic region in *P. andersonii* triple mutant nodules is fully surrounded by casparian strips, which was not observed in wild type nodules of similar age (Fig. S13). This result indicates that meristematic activity ceased early in the development of *Pannf-ya1;Pannf-ya3;Pannf-ya6* triple mutant nodules. Like *Pannf-ya1* single mutant nodules, *Pannf-ya1;Pannf-ya3;Pannf-ya6* triple mutant nodules contain large apoplastic colonies of rhizobium, but are devoid of intracellular infection structures (Fig. 6f). Taken-together, these data demonstrate that rhizobium intracellular infection is specifically controlled by *PanNF-YA1*, and that *PanNF-YA1*, *PanNF-YA3* and *PanNF-YA6* function redundantly to control nodule growth and development.

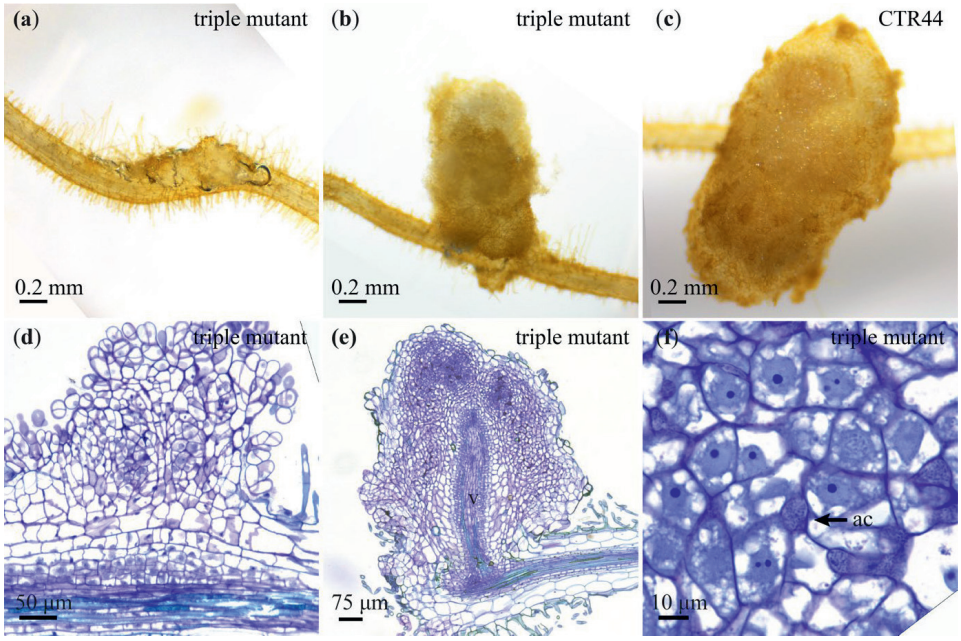


Fig. 6 The *Pannf-ya1;Pannf-ya3;Pannf-ya6* triple mutant is affected in nodule development. (**a**, **b**). Nodule-like structures formed on a *Pannf-ya1;Pannf-ya3;Pannf-ya6* mutant. (**c**) Nodule formed on a transgenic control line (CTR44). (**d, e**) Sections of the nodule-like structure shown in (**a**) and (**b**). (**f**) Apoplastic rhizobia (arrow) in a *Pannf-ya1;Pannf-ya3;Pannf-ya6* mutant nodule, whereas intracellular infection is absent. v indicates nodule vasculature; ac: apoplastic colonies of rhizobia. Sections were stained using Toluidine Blue. Nodules were isolated 4 weeks post-inoculation with *M. plurifarium* BOR2.

As *P. andersonii nf-ya1* mutant nodules are devoid of intracellular infection, we questioned whether this is specific for rhizobium, or alternatively, whether *NF-YA* genes may function also in intracellular colonization by arbuscular mycorrhizal fungi. To test this, control plants, the *Pannf-ya1*, *Pannf-ya3* and *Pannf-ya6* single mutants, and the *Pannf-ya1;Pannf-ya3;Pannf-ya6* triple mutant were grown under phosphate poor conditions and inoculated with 250 spores of *Rhizophagus irregularis* strain DOAM197198. The average colonization and arbuscule formation frequency were scored six weeks post inoculation. This showed that all mutants were equally well mycorrhized when compared to control plants (Fig. S14). Therefore, we conclude that *PanNF-YA1* commits a specific role in rhizobium intracellular infection.

Discussion

The transcription factors *NIN* and *NF-YA1* are essential components in a transcriptional network controlling rhizobium-induced nodule formation in legumes (Soyano & Hayashi, 2014). Here, we showed that the orthologous genes -*PanNIN* and *PanNF-YA1*- are essential for the formation of functional root nodules in the non-legume *P. andersonii*. Earlier studies, using transient RNA interference-mediated knockdown indicated that also *CgNIN* commits a symbiotic function in the nodulating actinorhizal species *Casuarina glauca* (Clavijo *et al.*, 2015). The *Parasponia* (Rosales), *Casuarina* (Fagales) and legume (Fabales) lineages diverged ~110 million years ago, soon after an assumed shared evolutionary event that gave birth to the nodulation trait (Soltis *et al.*, 1995; Wang *et al.*, 2009; van Velzen *et al.*, 2019). As *NIN* and *NF-YA1* are indispensable for the formation of functional nitrogen-fixing nodules in distinct taxonomic lineages, we conclude that these transcription factors represent core genes in the nodulation trait. Furthermore, we hypothesize that this recruitment into the nodulation trait has occurred in a species ancestral to the Fabales, Fagales, Cucurbitales and Rosales split.

In *L. japonicus*, *LjNF-YA1* is a direct transcriptional target of LjNIN (Soyano *et al.*, 2013, 2015). Direct evidence of a similar relationship has not been provided in any other species. Experiments presented here showed that in *P. andersonii* rhizobium-induced *PanNF-YA1* expression is PanNIN-dependent and that both genes are co-expressed in nodule primordia. In line with the hypothesis that both genes have been recruited in nodulation in a common ancestor of legumes and *Parasponia*, it is likely that the direct transcriptional regulation of the *NF-YA1* gene by NIN is conserved in nodulating species. This hypothesis is supported by the occurrence of putative NIN binding sites in the promoter region of *PanNF-YA1* (Fig. S15). In case these bindings sites find experimental support, the question remains whether the NIN-NF-YA1 transcription factor module is ancestral to the nitrogen-fixing clade, or whether it has evolved in concurrence with the nodulation trait.

P. andersonii NF-YA1 controls intracellular rhizobium infection, and knockout mutants of this gene are specifically blocked in infection thread formation. This mutant phenotype is different from the phenotypes reported for legume *nf-ya1* knockout and/or knockdown lines. In *L. japonicus* and *M. truncatula*, *nf-ya1* mutants and RNAi knockdown lines form smaller nodules (Combier *et al.*, 2006; Soyano *et al.*, 2013; Laporte *et al.*, 2014; Hossain *et al.*, 2016). In *M. truncatula*, this

developmental phenotype is due to absence or reduced activity of the nodule meristem (Xiao *et al.*, 2014), whereas in *L. japonicus LjNF-YA1* is indispensable for nodule differentiation, including vascular bundle formation (Hossain *et al.*, 2016). Absence of a functional *Mtnf-ya1* gene in *M. truncatula* also affects rhizobium infection, resulting in an increased number of infection threads that are arrested in the epidermis, and often have a swollen more bulbous appearance (Laporte *et al.*, 2014; Laloum *et al.*, 2014). In *P. andersonii nf-ya1* knockout mutants are not affected in nodule development. This divergence in phenotype between *P. andersonii* and legumes most probably is the result of adaptive evolution and subsequent divergence of the nodulation trait in both lineages. For example, intracellular rhizobium infection in *M. truncatula* and *L. japonicus* is initiated in curled root hairs, whereas in *P. andersonii* only nodule cells become invaded. Consequently, infection phenotypes may be observed in different cell types. Papilionoideae legumes -e.g. *L. japonicus*, *M. truncatula,* soybean (*Glycine max*), and common bean (*Phaseolus vulgaris*)- experienced gene duplication events in five *NF-YA* orthogroups, including *NF-YA1*, which most probably is the result of a whole genome duplication in a common ancestor (Cannon *et al.*, 2006; Young *et al.*, 2011). Subsequent gene redundancy may have allowed sub-neofunctionalization of *NF-YA1* and its closest paralog in legumes. Phenotypic analyses of mutant plants where both *NF-YA1* paralogs are targeted simultaneously, support the idea that controlling rhizobium intracellular infection is an ancestral symbiotic function of NF-YA1 and its closest paralog. For example, by committing *MtNF-YA2* RNAi in a *M. truncatula nf-ya1* mutant background, a more severe rhizobium infection phenotype can be observed (Laloum *et al.*, 2014). Also, in common bean a strong infection phenotype is observed after silencing of both *PvNF-YA9* and *PvNF-YA1* in *A. rhizogenes*-transformed roots (Rípodas *et al.*, 2019). However, in this study, single gene targets have not been analysed. Such gene duplications, which are common in Papilionoideae legumes, complicate reverse genetic studies. *P. andersonii* did not experience any duplication events in any of the seven *NF-YA* orthogroups (Fig. 4a). In line with this, we argue that this species may be more suited to uncover the functioning of *NF-YA* genes by reverse genetics.

We also studied the function of two additional *NF-YA* genes -*PanNF-YA3* and *PanNF-YA6*- in *P. andersonii*, as both these genes have a nodule-enhanced expression profile (Fig. 4b). Such a nodule-enhanced expression profile has been reported also for the *M. truncatula* orthologs *MtNF-YA8* (orthologous to *PanNF-YA3*) and *MtNF-YA3* (orthologous to *PanNF-YA6*) (Baudin *et al.*, 2015). However, no apparent nodulation phenotype could be observed in *P. andersonii* single and

double mutants. Only upon creating a higher order *Pannf-ya1;Pannf-ya3;Pannf-ya6* mutant an effect on nodule organogenesis was observed. This suggests that all three *PanNF-YA* genes act redundantly in controlling nodule development.

Recent phylogenomic analyses revealed that within the nitrogen-fixing clade absence of the nodulation trait is associated with pseudogenization of the *NIN* gene (van Velzen *et al.*, 2018; Griesmann *et al.*, 2018). This shows that within the nitrogen-fixing clade the functioning of this gene correlates with the nodulation trait. In contrast to *NIN*, no such correlation has been reported between the presence of *NF-YA1* orthologs and the nodulation trait (van Velzen *et al.*, 2018; Griesmann *et al.*, 2018), suggesting that these genes commit also non-symbiotic functions. *A. thaliana* has two orthologs of *LjNF-YA1, MtNF-YA1* and *PanNF-YA1*, named *AtNF-YA2* and *AtNF-YA10* (Fig. 4a). Mutant analysis of these genes has been hampered by the sterility phenotype of *Atnf-ya2* insertion and RNAi lines (Pagnussat *et al.*, 2005; Sorin *et al.*, 2014). Mis-expression studies of either gene revealed a function in leaf and root growth and lateral root initiation as well as increased tolerance to several types of abiotic stresses (Leyva-González et al. 2012; Sorin et al. 2014; Zhang et al. 2017; Soyano et al. 2019). Furthermore, it was shown that in *L. japonicus* ectopic expression of *LjNF-YA1* results in lateral roots with malformed tips (Soyano *et al.*, 2013; Sorin *et al.*, 2014) We observed a mild though consistent decrease in lateral roots formed in plantlets containing a mutation in *Pannf-ya1*. This supports the findings that *NF-YA1* orthologous genes commit a non-symbiotic function in root development, and may explain why *NF-YA1* is not pseudogenized in species that have lost the nodulation trait (Soyano *et al.*, 2013; van Velzen *et al.*, 2018; Griesmann *et al.*, 2018). As the *P. andersonii nf-ya1* knockout mutants are not affected in the symbiosis with arbuscular mycorrhiza, it suggests that NF-YA1 symbiotic functioning is exclusively required to allow entry of symbiotic bacteria. As the bacterial infectability of cells is a key characteristic of the nodulation trait, it will be an important future scientific objective to determine the core transcriptional network regulated by *NF-YA1* and its interacting partners. Having a *P. andersonii nf-ya1* mutant available with a strict infection phenotype as a comparative system to legumes where infection and organogenesis phenotypes are intertwined, will be instrumental to achieve this objective.

Materials and Methods

Plant Materials and Growth Conditions

All experiments were done using *P. andersonii* WU1.14 (van Velzen *et al.*, 2018; Wardhani *et al.*, 2019). Plants were maintained as described previously (van Zeijl *et al.*, 2018; Wardhani *et al.*, 2019). Young plantlets for nodulation assays were vegetatively propagated *in vitro*, rooted, and inoculated with *Mesorhizobium plurifarium* BOR2 at an OD_{600} = 0.03 (van Zeijl *et al.*, 2018; van Velzen *et al.*, 2018; Wardhani *et al.*, 2019). For early induction of symbiotic genes, we made use of *Rhizobium tropici* CIAT899 transformed with pMP604 (OD_{600} = 0.03-0.05) (Martínez *et al.*, 1985; Spaink *et al.*, 1989). Nodulation efficiencies were calculated by determining the average nodule number per plant. Nodule size estimates were determined by measuring the 2D nodule surface area using ImageJ (Abràmoff *et al.*, 2004). Comparisons were made based on the average nodule size per plant using at least four replicate plants. Acetylene reductase assays (ARA) were conducted as described previously (van Velzen *et al.*, 2018). Mycorrhization experiments were conducted using 250 spores of *Rhizophagus irregularis* strain DOAM197198 as described previously (van Velzen *et al.*, 2018; Wardhani *et al.*, 2019).

Lateral Root growth assay

Similar sized rooted plantlets were grown on EKM-plates (1% Daishin agar) (Duchefa, Netherlands) in between two cellophane layers cut to 12x8cm (Sigma Aldrich, Gel drying frames) (van Velzen *et al.*, 2018; Wardhani *et al.*, 2019). Plants were grown vertically at a 60 degree angle for 20 days at 28 °C, 16/8h day-night regime. Main roots were determined as all roots directly attached to the shoot present at the start of the experiment. Per plantlet, root length and lateral root number per root was determined. Total “main” root length per shoot and lateral root density in lateral roots per mm root was plotted per plant. Statistical testing was based on a Mann-Whitney U-test with a significance level $P<0.05$.

Vectors and Constructs

Single-guide RNAs (sgRNAs) were designed using the ‘Find CRISPR Targets’ function implemented in Geneious 9.1.5 (Biomatters, New Zealand) and subsequently checked against the *P. andersonii* genome for high identity off-targets. To mutate genes up to three sgRNAs were used

to target either the first or the second coding exon (Table S1). Selected sgRNAs were amplified using sequence-specific forward primers and a universal reverse primer (Table S2), using Addgene plasmid #46966 as template (Nekrasov *et al.*, 2013). Constructs for CRISPR/Cas9-mediated mutagenesis were assembled as described previously (van Zeijl *et al.*, 2018; Wardhani *et al.*, 2019). To allow golden gate cloning of GUS reporter constructs, the BpiI and BsaI restriction sites in putative promoter sequences of *PanNF-YA1, PanNF-YA3* and *PanNF-YA6* were mutated by introducing single nucleotide substitutions (Engler *et al.*, 2014). The putative promoter sequences are provided in Table S3.

Plant Transformation

A. tumefaciens-mediated transformation and genotyping was done as published previously (van Zeijl *et al.*, 2018; Wardhani *et al.*, 2019). Primers used for genotyping are listed in Table S2. For promoter-GUS reporter studies, we investigated 5 independent lines for each construct.

Histochemical Analysis, Microtome Sectioning and Microscopy

Root and nodule samples of the *PanNF-YA$_{pro}$:GUS* lines were incubated in GUS buffer (3% [w/v] sucrose, 10 mM EDTA, 2 mM k-ferrocyanide, 2 mM k-ferricyanide, and 0.5 mg/mL 5-bromo-4-chloro-3-indolyl-beta-D- glucuronic acid, cyclohexylammonium salt [X-Gluc] in 0.1 M phosphate buffer [pH = 7.2]) at 37°C for 2 and 5 hours, respectively. For whole mount sections, GUS-stained samples were embedded in 6% low melting point agarose (in PBS). 70 µm thick sections were made using a vibratome, and were imaged using nomarski microscopy. For plastic sections, root segments and nodules were fixed in 4% paraformaldehyde (w/v), 5% glutaraldehyde (v/v) in 50 mM phosphate buffer (pH = 7.2) at 4°C for 24 hours. Subsequently, the samples were dehydrated using an ethanol series and embedded in Technovit 7100 (Heraeus Kulzer, Germany) according to the manufacturer's instructions. Semi-thin sections were cut using a Leica Ultracut microtome (Leica Microsystems, Germany) to 4 µm thickness for nodules formed on CRISPR mutant lines and 7 µm thickness for GUS stained samples. Sections were stained with 0.05% Toluidine Blue or 0.1% Rethudium Red. Images were photographed using a Leica DM5500B microscope equipped with a DFC425C camera (Leica Microsystems, Germany). Samples for electron microscopy were fixed in MTSB buffer (Pasternak *et al.*, 2015) containing 2.5% glutaraldehyde, post-fixed in aqueous 1% OsO_4 solution, and stained *in bloc* with 1% uranyl acetate. After dehydration in

increasing EtOH concentrations, samples were embedded in Epoxy resin.Ultrathin (70 nm) sections were post-stained with 2% uranyl acetate and observed in a Philips CM-10 TEM. Images were taken using a Gatan BioScan 792 camera.

In situ Hybridization

P. andersonii nodules were fixed with 4% paraformaldehyde, 3% glutaraldehyde in 50 mM phosphate buffer (pH = 7.4) and embedded in paraffin (Paraplast X-tra, McCormick Scientific, United States). Root sections of 7 µm were prepared using an RJ2035 microtome (Leica Microsystems, Germany). RNA *in situ* hybridization (ISH) was conducted using Invitrogen™ ViewRNA™ ISH Tissue 1- Plex Assay kits (ThermoFisher Scientific, United States) according to a protocol previously developed for *M. truncatula* (Kulikova *et al.*, 2018). In short; mRNA detection is based on branched (b)DNA signal amplification technology. A mRNA probe set contains ~20 synthetic adjacent oligonucleotide pairs. Each oligonucleotide is composed of a 20 bp primary sequence to target the sequence of interest and a secondary extended sequence serving as a template for hybridization of a preamplifier oligonucleotide. The preamplifier can hybridize to two adjacent probes. An additional sequence of the preamplifier is designed to hybridize to multiple bDNA amplifier molecules that create a branched structure. Finally, alkaline phosphatase (AP)-labeled oligonucleotides, which are complementary to bDNA amplifier sequences, bind to the bDNA molecule by hybridization. By adding Fast Red substrate (ThermoFisher Scientific, United States), red punctuated precipitates are formed that can be detected by light microscopy. RNA ISH probe sets were designed and synthesized at request by ThermoFisher Scientific. Catalog numbers of probes are VF1-6000380 for *PanNIN*, VF1-6000400 for *PanNF-YA1*, VF1-6000767 for *PanNF-YA3* and VF-6000766 for *PanNF-YA6*. Images were taken with an DM5500B microscope equipped with a DFC425C camera (Leica Microsystems, Germany).

Phylogenetic Reconstruction

Protein sequences of *L. japonicus* (Lj3.0, *Lotus* Base (REF: doi:10.1038/srep39447)) (Sato *et al.*, 2008), *Glycine max* (Wm82.a2.v1) (Sato *et al.*, 2008; Schmutz *et al.*, 2010), *Phaseolus vulgaris* (Pvulgaris V2.1) (Schmutz *et al.*, 2014), *Morus notabilis* (https://morus.swu.edu.cn/morusdb/datasets) (He *et al.*, 2013), *Prunus persica* (Ppersica v2.1) (International Peach Genome Initiative *et al.*, 2013) *Fragaria vesca* (Fvesca v1.1) (Shulaev *et al.*,

2011) were retrieved from Phytozome 12 (http://phytozome.jgi.doe.gov/), unless stated otherwise. *Casuarina glauca* and *Datisca glomerata* assemblies were downloaded and setup as custom blast database in Geneious 8.1.9 (Biomatters, New Zealand)(van Velzen *et al.*, 2018; Griesmann *et al.*, 2018). Sequences from diploid Peanut *Arachis duranensis* were retrieved from NCBI (Bertioli *et al.*, 2016). Protein sequences of *P. andersonii* (PanWU01x14) and *Trema orientalis* (TorRG33x02) were obtained from www.parasponia.org (van Velzen *et al.*, 2018; Holmer *et al.*, 2019). These sequences were mined using sequences from *A. thaliana* (TAIR10) (Lamesch *et al.*, 2012) and *M. truncatula* (Mt4.0v1) (Young *et al.*, 2011; Tang *et al.*, 2014). Protein sequences were aligned using MAFFT v7.017 (Table S4; parameter settings: algorithm, auto; scoring matrix, Blosum62; gap open penalty, 1.53; offset value, 0.123) (Katoh *et al.*, 2002; Katoh & Standley, 2013) implemented in Geneious 8.1.9. (Biomatters, New Zealand). Bayesian phylogeny was reconstructed using MrBayes 3.2.6. (Ronquist & Huelsenbeck, 2003) implemented in Geneious 8.1.9. (parameter settings: rate matrix, poisson; rate variation, gamma; gamma categories, 4; chain length, 5100000; heated chains, 4; heated chain temp, 0.2; subsampling freq, 1000; burn-in length, 100000; random seed, 8681). Mid-point rooting was applied for better tree visualization using FigTree v1.4.2. (http://tree.bio.ed.ac.uk/software/figtree).

RNA Isolation and qRT-PCR Analysis

RNA was isolated from snap-frozen root segments of about 0.5 cm, which includes the elongation zone and the newly formed differentiation zone. cDNA was prepared from 1 µg of total RNA using the i-script cDNA synthesis kit (Bio-Rad, United States), following the manufacturer's instructions. Ten µl qRT-PCR reactions were set up using 2x iQ SYBR Green Supermix (Bio-Rad, United States) and 5 ng template DNA. Quantification was performed using a CFX Connect optical cycler, according to the manufacturer's protocol (Bio-Rad, United States). Normalization was performed based on the stably expressed reference gene *ELONGATION FACTOR 1α* (*PanEF1α*) (van Zeijl *et al.*, 2018). Primers used for qPR-PCR analysis are listed in Table S2.

ACKNOWLEDGEMENTS

This work was supported by an NWO-VICI grant (865.13.001) to RG, the ENSA project funded by the Bill & Melinda Gates Foundation to the University of Cambridge to RG and TO, a CSC Scholarship (201303250067) to FB, and Ministry of Research, Technology and Higher Education of the Republic of Indonesia (RISET-PRO grant 8245-ID) to YPR.

AUTHOR CONTRIBUTIONS

FB, LR and RG planned and designed the research; FB, LR, MR, OK and YPR performed the experiments; FB, AvZ, LR, RG, TB, TO and YPR analyzed the data; and FB, AvZ and RG wrote the manuscript.

References

Abràmoff MD, Magalhães PJ, Ram SJ. **2004**. Image processing with ImageJ. *Biophotonics international* **11**: 36–42.

Baudin M, Laloum T, Lepage A, Ripodas C, Ariel F, Frances L, Crespi M, Gamas PC, Blanco FA, Zanetti ME, *et al.* 2015. A phylogenetically conserved group of NF-Y transcription factors interact to control nodulation in legumes. *Plant physiology*: 01144.2015.

Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EKS, Liu X, Gao D, Clevenger J, Dash S, *et al.* 2016. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nature genetics* **48**: 438–446.

Borisov AY, Madsen LH, Tsyganov VE, Umehara Y, Voroshilova VA, Batagov AO, Sandal N, Mortensen A, Schauser L, Ellis N, *et al.* 2003. The Sym35 gene required for root nodule development in pea is an ortholog of Nin from *Lotus japonicus*. *Plant physiology* **131**: 1009–1017.

Cannon SB, Sterck L, Rombauts S, Sato S, Cheung F, Gouzy J, Wang X, Mudge J, Vasdewani J, Schiex T, *et al.* 2006. Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 14959–14964.

Castaings L, Camargo A, Pocholle D, Gaudon V, Texier Y, Boutet-Mercey S, Taconnat L, Renou J-P, Daniel-Vedele F, Fernandez E, *et al.* 2009. The nodule inception-like protein 7 modulates nitrate sensing and metabolism in Arabidopsis. *The Plant journal* **57**: 426–435.

Clavijo F, Diedhiou I, Vaissayre V, Brottier L, Acolatse J, Moukouanga D, Crabos A, Auguy F, Franche C, Gherbi H, *et al.* 2015. The *Casuarina NIN* gene is transcriptionally activated throughout *Frankia* root infection as well as in response to bacterial diffusible signals. *The New phytologist* **208**: 887–903.

Combier J-P, Frugier F, de Billy F, Boualem A, El-Yahyaoui F, Moreau S, Vernié T, Ott T, Gamas P, Crespi M, *et al.* 2006. MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. *Genes & development* **20**: 3084–3088.

Dénarié J, Debellé F, Promé JC. **1996**. Rhizobium lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. *Annual review of biochemistry* **65**: 503–535.

Doyle JJ. **2011**. Phylogenetic perspectives on the origins of nodulation. *Molecular plant-microbe interactions* **24**: 1289–1295.

Engler C, Youles M, Gruetzner R, Ehnert T-M, Werner S, Jones JDG, Patron NJ, Marillonnet S. **2014**. A golden gate modular cloning toolbox for plants. *ACS synthetic biology* **3**: 839–843.

Geurts R, Lillo A, Bisseling T. **2012**. Exploiting an ancient signalling machinery to enjoy a nitrogen fixing symbiosis. *Current opinion in plant biology* **15**: 438–443.

Griesmann M, Chang Y, Liu X, Song Y, Haberer G, Crook MB, Billault-Penneteau B, Lauressergues D, Keller J, Imanishi L, *et al.* 2018. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* **361**: eaat1743.

Heckmann AB, Sandal N, Bek AS, Madsen LH, Jurkiewicz A, Nielsen MW, Tirichine L, Stougaard J. 2011. Cytokinin Induction of Root Nodule Primordia in *Lotus japonicus* Is Regulated by a Mechanism Operating in the Root Cortex. *Molecular plant-microbe interactions* **24**: 1385–1395.

He N, Zhang C, Qi X, Zhao S, Tao Y, Yang G, Lee T-H, Wang X, Cai Q, Li D, *et al.* 2013. Draft genome sequence of the mulberry tree *Morus notabilis*. *Nature communications* **4**: 2445.

Holmer R, van Velzen R, Geurts R, Bisseling T, de Ridder D, Smit S. 2019. GeneNoteBook, a collaborative notebook for comparative genomics. *Bioinformatics*. doi: 10.1093/bioinformatics/btz491

Hossain MS, Shrestha A, Zhong S, Miri M, Austin RS, Sato S, Ross L, Huebert T, Tromas A, Torres-Jerez I, *et al.* 2016. *Lotus japonicus* NF-YA1 Plays an Essential Role During Nodule Differentiation and Targets Members of the SHI/STY Gene Family. *Molecular plant-microbe interactions* **29**: 950–964.

International Peach Genome Initiative, Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, *et al.* 2013. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature genetics* **45**: 487–494.

Katoh K, Misawa K, Kuma K-I, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* **30**: 3059–3066.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30**: 772–780.

Konishi M, Yanagisawa S. 2013. Arabidopsis NIN-like transcription factors have a central role in nitrate signalling. *Nature communications* **4**: 1617.

Kosuta S, Held M, Hossain MS, Morieri G, Macgillivary A, Johansen C, Antolín-Llovera M, Parniske M, Oldroyd GED, Downie AJ, *et al.* 2011. *Lotus japonicus* symRK-14 uncouples the cortical and epidermal symbiotic program. *The Plant journal* **67**: 929–940.

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

Laloum T, Baudin M, Frances L, Lepage A, Billault-Penneteau B, Cerri MR, Ariel F, Jardinaud M-F, Gamas P, de Carvalho-Niebel F, *et al.* 2014. Two CCAAT-box-binding transcription factors redundantly regulate early steps of the legume-rhizobia endosymbiosis. *The Plant journal* **79**: 757–768.

Laloum T, De Mita S, Gamas P, Baudin M, Niebel A. 2013. CCAAT-box binding transcription factors in plants: Y so many? *Trends in plant science* **18**: 157–166.

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, *et al.* 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic acids research* **40**: D1202–10.

Lancelle SA, Torrey JG. **1984**. Early development of Rhizobium-induced root nodules of *Parasponia rigida*. I. Infection and early nodule initiation. *Protoplasma* **123**: 26–37.

Lancelle SA, Torrey JG. **1985**. Early development of Rhizobium-induced root nodules of *Parasponia rigida*. II. Nodule morphogenesis and symbiotic development. *Canadian journal of botany. Journal canadien de botanique* **63**: 25–35.

Laporte P, Lepage A, Fournier J, Catrice O, Moreau S, Jardinaud M-F, Mun J-H, Larrainzar E, Cook DR, Gamas P, *et al.* 2014. The CCAAT box-binding transcription factor NF-YA1 controls rhizobial infection. *Journal of experimental botany* **65**: 481–494.

Lerouge P, Roche P, Faucher C, Maillet F, Truchet G, Promé JC, Dénarié J. **1990**. Symbiotic host-specificity of Rhizobium meliloti is determined by a sulphated and acylated glucosamine oligosaccharide signal. *Nature* **344**: 781–784.

Leyva-González MA, Ibarra-Laclette E, Cruz-Ramírez A, Herrera-Estrella L. **2012**. Functional and transcriptome analysis reveals an acclimatization strategy for abiotic stress tolerance mediated by *Arabidopsis* NF-YA family members. *PloS one* **7**: e48138.

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

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

Madsen EB, Madsen LH, Radutoiu S, Olbryt M, Rakwalska M, Szczyglowski K, Sato S, Kaneko T, Tabata S, Sandal N, *et al.* 2003. A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. *Nature* **425**: 637–640.

Marsh JF, Rakocevic A, Mitra RM, Brocard L, Sun J, Eschstruth A, Long SR, Schultze M, Ratet P, Oldroyd GED. **2007**. *Medicago truncatula NIN* is essential for rhizobial-independent nodule organogenesis induced by autoactive Calcium/Calmodulin-Dependent Protein Kinase. *Plant physiology* **144**: 324–335.

Martínez E, Pardo MA, Palacios R, Miguel AC. **1985**. Reiteration of Nitrogen Fixation Gene Sequences and Specificity of Rhizobium in Nodulation and Nitrogen Fixation in *Phaseolus vulgaris*. *Microbiology* **131**: 1779–1786.

Myers ZA, Holt BF 3rd. **2018**. NUCLEAR FACTOR-Y: still complex after all these years? *Current opinion in plant biology* **45**: 96–102.

Nekrasov V, Staskawicz B, Weigel D, Jones JDG, Kamoun S. **2013**. Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nature biotechnology* **31**: 691–693.

Oldroyd GED. **2013**. Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nature reviews. Microbiology* **11**: 252–263.

Op den Camp RHM, Polone E, Fedorova E, Roelofsen W, Squartini A, Op den Camp HJM, Bisseling T, Geurts R. **2012**. Nonlegume *Parasponia andersonii* deploys a broad rhizobium host range strategy resulting in largely variable symbiotic effectiveness. *Molecular plant-microbe interactions* **25**: 954–963.

Op den Camp R, Streng A, De Mita S, Cao Q, Polone E, Liu W, Ammiraju JSS, Kudrna D, Wing R, Untergasser A, *et al.* 2011. LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume *Parasponia*. *Science* **331**: 909–912.

Pagnussat GC, Yu H-J, Ngo QA, Rajani S, Mayalagu S, Johnson CS, Capron A, Xie L-F, Ye D, Sundaresan V. **2005**. Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis*. *Development* **132**: 603–614.

Pasternak T, Tietz O, Rapp K, Begheldo M, Nitschke R, Ruperti B, Palme K. **2015**. Protocol: an improved and universal procedure for whole-mount immunolocalization in plants. *Plant methods* **11**: 50.

Popp C, Ott T. **2011**. Regulation of signal transduction and bacterial infection during root nodule symbiosis. *Current opinion in plant biology* **14**: 458–467.

Radutoiu S, Madsen LH, Madsen EB, Felle HH, Umehara Y, Grønlund M, Sato S, Nakamura Y, Tabata S, Sandal N, *et al.* 2003. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* **425**: 585–592.

Rípodas C, Castaingts M, Clúa J, Villafañe J, Blanco FA, Zanetti ME. **2019**. The PvNF-YA1 and PvNF-YB7 Subunits of the Heterotrimeric NF-Y Transcription Factor Influence Strain Preference in the *Phaseolus vulgaris–Rhizobium etli* Symbiosis. *Frontiers in plant science* **10**: 221.

Rípodas C, Clúa J, Battaglia M, Baudin M, Niebel A, Zanetti ME, Blanco F. **2014**. Transcriptional regulators of legume-rhizobia symbiosis: nuclear factors Ys and GRAS are two for tango. *Plant signaling & behavior* **9**: e28847.

Ronquist F, Huelsenbeck JP. **2003**. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K, *et al.* 2008. Genome structure of the legume, *Lotus japonicus. DNA research: an international journal for rapid publication of reports on genes and genomes* **15**: 227–239.

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

Schauser L, Wieloch W, Stougaard J. **2005**. Evolution of NIN-like proteins in *Arabidopsis*, rice, and *Lotus japonicus*. *Journal of molecular evolution* **60**: 229–237.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, *et al.* 2010. Genome sequence of the palaeopolyploid soybean. *Nature* **463**: 178–183.

Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C, *et al.* 2014. A reference genome for common bean and genome-wide analysis of dual domestications. *Nature genetics* **46**: 707–713.

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, *et al.* 2011. The genome of woodland strawberry (*Fragaria vesca*). *Nature genetics* **43**: 109–116.

Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JM, Martin PG. 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.

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

Soyano T, Hayashi M. 2014. Transcriptional networks leading to symbiotic nodule organogenesis. *Current opinion in plant biology* **20**: 146–154.

Soyano T, Hirakawa H, Sato S, Hayashi M, 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.

Soyano T, Kouchi H, Hirota A, 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.

Soyano T, Shimoda Y, Hayashi M. 2015. NODULE INCEPTION antagonistically regulates gene expression with nitrate in *Lotus japonicus*. *Plant & cell physiology* **56**: 368–376.

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

Spaink HP, Okker RJ, Wijffelman CA, Tak T, Goosen-de Roo L, Pees E, van Brussel AA, Lugtenberg BJ. 1989. Symbiotic properties of rhizobia containing a flavonoid-independent hybrid nodD product. *Journal of bacteriology* **171**: 4045–4053.

Tang H, Krishnakumar V, Bidwell S, Rosen B, Chan A, Zhou S, Gentzbittel L, Childs KL, Yandell M, Gundlach H, *et al.* 2014. An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. *BMC genomics* **15**: 312.

Trinick MJ. 1979. Structure of nitrogen-fixing nodules formed by Rhizobium on roots of *Parasponia andersonii* Planch. *Canadian journal of microbiology* **25**: 565–578.

van Velzen R, Doyle JJ, Geurts R. **2019**. A Resurrected Scenario: Single Gain and Massive Loss of Nitrogen-Fixing Nodulation. *Trends in plant science* **24**: 49–57.

van Velzen R, Holmer R, Bu F, Rutten L, van Zeijl A, Liu W, Santuari L, Cao Q, Sharma T, Shen D, *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.

Vernié T, Kim J, Frances L, Ding Y, Sun J, Guan D, Niebel A, Gifford ML, de Carvalho-Niebel F, Oldroyd GED. **2015**. The NIN transcription factor coordinates diverse nodulation programs in different tissues of the *Medicago truncatula* root. *The Plant cell* **27**: tpc.15.00461.

Wang H, Moore MJ, Soltis PS, Bell CD, Brockington SF, Alexandre R, Davis CC, Latvis M, Manchester SR, Soltis DE. **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.

Wardhani TAK, Roswanjaya YP, Dupin S, Li H, Linders S, Hartog M, Geurts R, Van Zeijl A. **2019**. Transforming, genome editing and phenotyping the nitrogen-fixing tropical Cannabaceae tree *Parasponia andersonii*. *Journal of visualized experiments: JoVE*. doi:10.3791/59971

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

Yoro E, Suzaki T, Toyokura K, Miyazawa H, Fukaki H, Kawaguchi M. **2014**. A positive regulator of nodule organogenesis, NODULE INCEPTION, acts as a negative regulator of rhizobial infection in *Lotus japonicus*. *Plant physiology* **165**: 113.233379–.

Young ND, Debellé F, Oldroyd GED, Geurts R, Cannon SB, Udvardi MK, Benedito V a., Mayer KFX, Gouzy J, Schoof H, *et al.* 2011. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**: 520–524.

van Zeijl A, Wardhani TAK, Seifi Kalhor M, Rutten L, Bu F, Hartog M, Linders S, Fedorova EE, Bisseling T, Kohlen W, *et al.* 2018. CRISPR/Cas9-Mediated Mutagenesis of Four Putative Symbiosis Genes of the Tropical Tree *Parasponia andersonii* Reveals Novel Phenotypes. *Frontiers in plant science* **9**: 284.

Zhang M, Hu X, Zhu M, Xu M, Wang L. **2017**. Transcription factors NF-YA2 and NF-YA10 regulate leaf growth via auxin signaling in *Arabidopsis*. *Scientific reports* **7**: 1395.

Supporting information

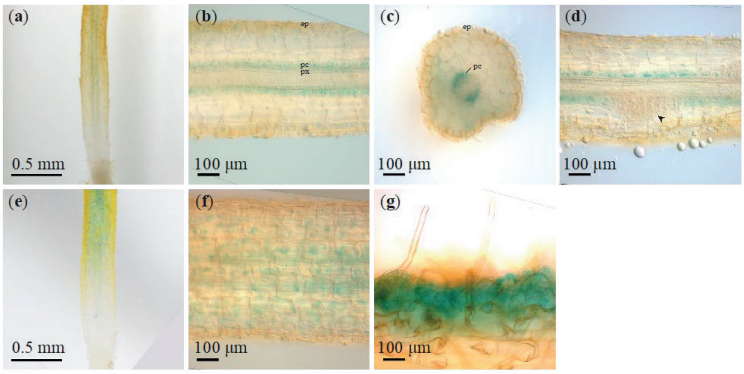


Fig. S1 Spatiotemporal expression pattern of *PanNF-YA1*$_{pro}$:GUS in *Parasponia andersonii* roots. (**a-d**) GUS-stained non-inoculated root segments. (e-g) GUS-stained inoculated root segments. (**a**) Faint *PanNF-YA1*$_{pro}$:GUS activity was observed around the vasculature in the differentiated zone of a young root. (**b**) *PanNF-YA1*$_{pro}$:GUS activity observed in the pericycle cells (70 µm-thick longitudinal root section). (**c**) *PanNF-YA1*$_{pro}$:GUS activity observed in pericycle cells opposite protoxylem poles (70 µm-thick cross-section of the root). (**d**) Spatiotemporal expression of *PanNF-YA1*$_{pro}$:GUS during lateral root initiation (70 µm-thick longitudinal root-section). (**e, f**) *PanNF-YA1*$_{pro}$:GUS activity was induced in epidermal cells at the elongation and differentiation zone of a root at 2 dpi with *M. plurifarium* strain BOR2. (**g**) *PanNF-YA1*$_{pro}$:GUS activity detected in the root epidermis upon rhizobium inoculation at a similar developing stage as shown in Figure 1A. Plants were grown *in vitro* (a-f), or in a perlite potting system (**g**). Data shown are obtained using transgenic *PanNF-YA1*$_{pro}$:GUS line 1E5. ep: epidermis; pc: pericycle; px: protoxylem; arrowhead indicates lateral root primordia.

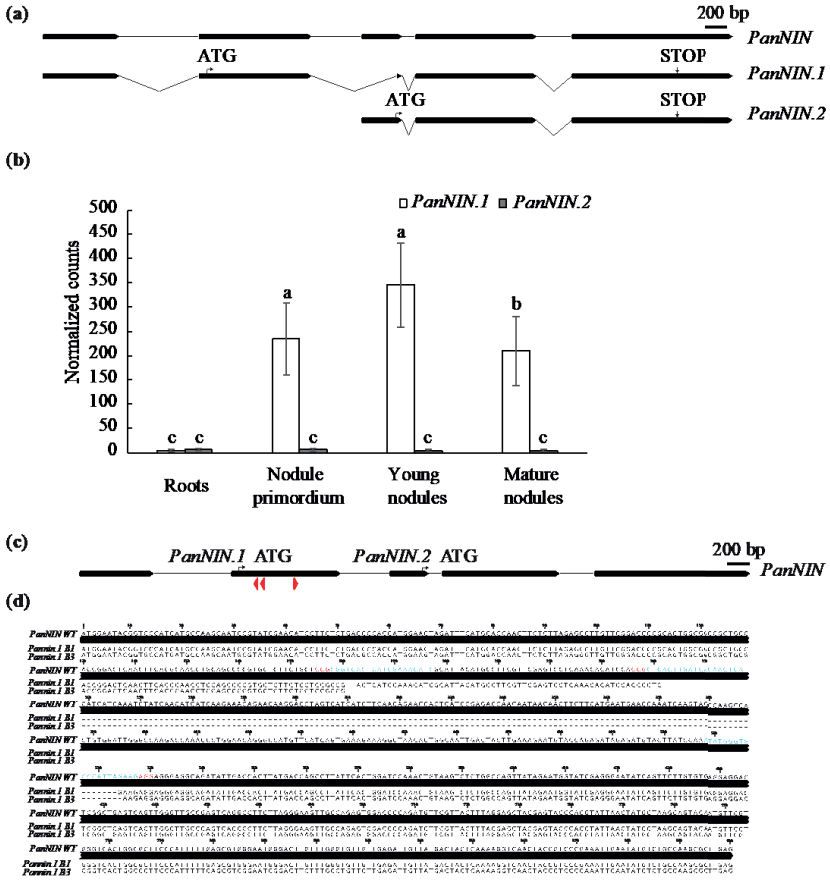


Fig. S2 Structure and expression of the *P. andersonii NIN* gene and the genotype of CRISPR-Cas9 *Pannin* mutants. (**a**) Schematic representation of the *PanNIN* gene model. Indicated are two *PanNIN* transcripts: *PanNIN.1* and *PanNIN.2*. Translational initiation (ATG) and termination (STOP) sites are indicated by arrows. (**b**) Expression of *PanNIN.1* and *PanNIN.2* in roots and nodules. Expression was determined by quantification of RNAseq reads. Data represent DE-seq2-normalized read counts (n = 3) ± SD, which were obtained from van Velzen et al., 2018. Different letters indicate statistical significance (Student's t-test, $p < 0.05$). (**c**) Schematic representation of the *PanNIN* gene model. Indicated are the locations of 3 sgRNA target sites (red triangles) and the translational initiation sites present in *PanNIN.1* and *PanNIN.2*. Note that the sgRNAs target the first coding-exon that is specific for *PanNIN.1*. (**d**) Alignment of the sequence of the first *PanNIN* coding-exon in wild type and *Pannin* mutant lines B1 and B3. sgRNA target sites are marked in blue, PAM sequences are marked in red. Note that in both lines mutations are homozygous.

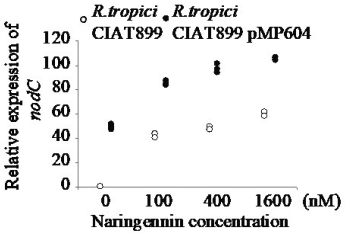


Fig. S3 *Rhizobium tropici* CIAT899.pMP604 constitutively expresses the LCO biosynthesis gene *nodC*. The pMP604 plasmid encodes an autoactive variant of *nodD*, a transcription factor that regulates LCO biosynthesis genes. Shown are the relative expression of *nodC* in wild-type *R. tropici* CIAT899 (white dots) and *R. tropici* strain CIAT899 transformed with pMP604 (black dots) in the absence or presence of increasing concentrations of naringenin (nM). Dots represent technical repeats.

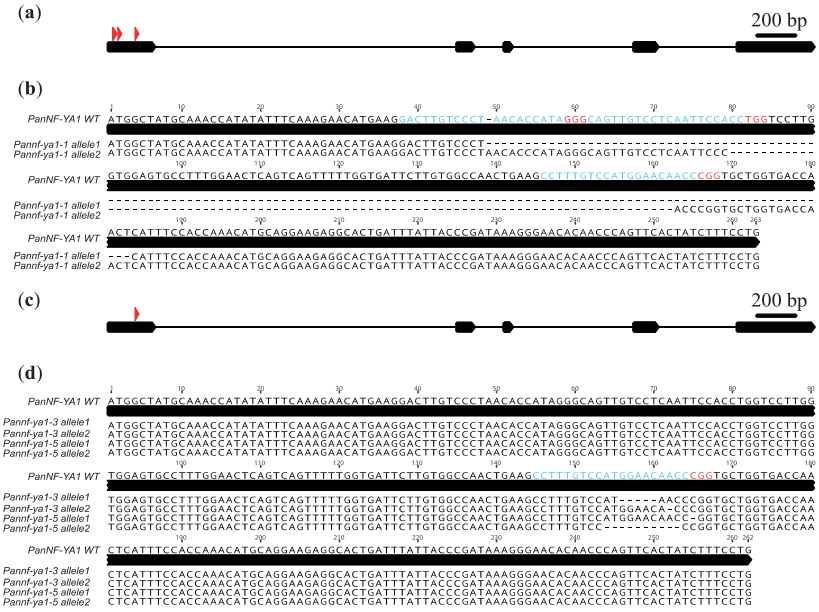


Fig. S4 Structure of the *P. andersonii NF-YA1* gene and genotype of CRISPR-Cas9 *Panf-ya1* mutants. (**a**) Schematic representation of the *PanNF-YA1* gene model. Indicated are the locations of 3 sgRNAs used for mutagenesis (red triangles). (**b**) Genotype of the bi-allelic *Pannf-ya1-1* mutant line. (**c**) Schematic representation of the *PanNF-YA1* gene model. Indicated (red triangles) are the location of sgRNA used in an independent transformation to creating *Pannf-ya1* CRISPR-Cas9 mutants. (**d**) The genotype of the *Pannf-ya1* mutant lines m3 and m5. Shown is the sequence of the first coding exon.

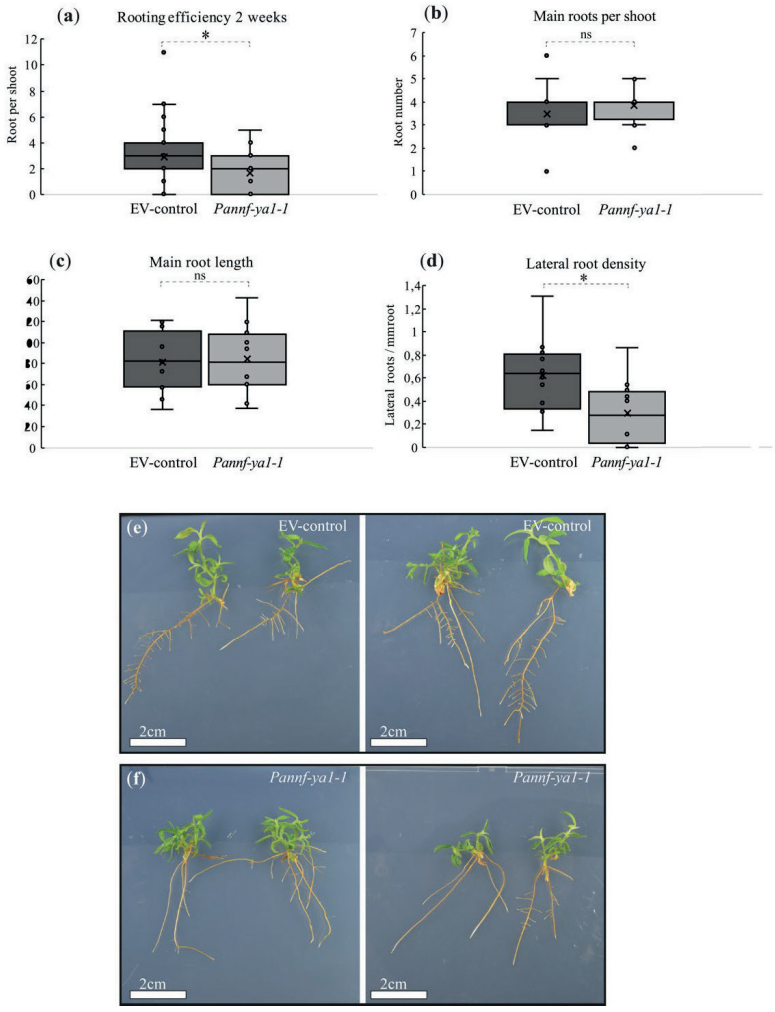


Fig. S5 Lateral root formation is affected in the *P. andersonii nf-ya1* mutant. (a) Number of roots formed on plantlets incubated on rooting medium for 14 days is lower in *Pannf-ya1-1* mutants (n=67 for EV-control n=66 for *Pannf-ya1-1*). Statistical significance based on a student's T-test $p<0.05$. (**b-d**) Plantlets in a comparable developmental stage were selected to be used for the root formation assay, transferred to EKM-agar plates lined with cellophane and scored after 20 days. EV-control n=12, *Pannf-ya1-1* n=12. **(b)** Number of growing main roots on EKM medium is not different between EV-control and *Pannf-ya1-1*. Main roots are characterized as all roots directly attached to the shoot. Given the nature of rooted plantlets there are usually multiple roots. **(c)** Total summed main root length per shoot is not different between EV-control and *Pannf-ya1-1*. **(d)** Number of lateral roots formed on main roots is reduced in *Pannf-ya1-1* mutants. Statistical significance for (b,c,d) based on Mann Whitney U-test $p<0.05$.

(e) Representative examples of EV-control plantlets grown 20 days on EKM-agar plates. **(f)** Representative examples of *Pannf-ya1-1* plantlets grown for 20 days on EKM-agar plates.

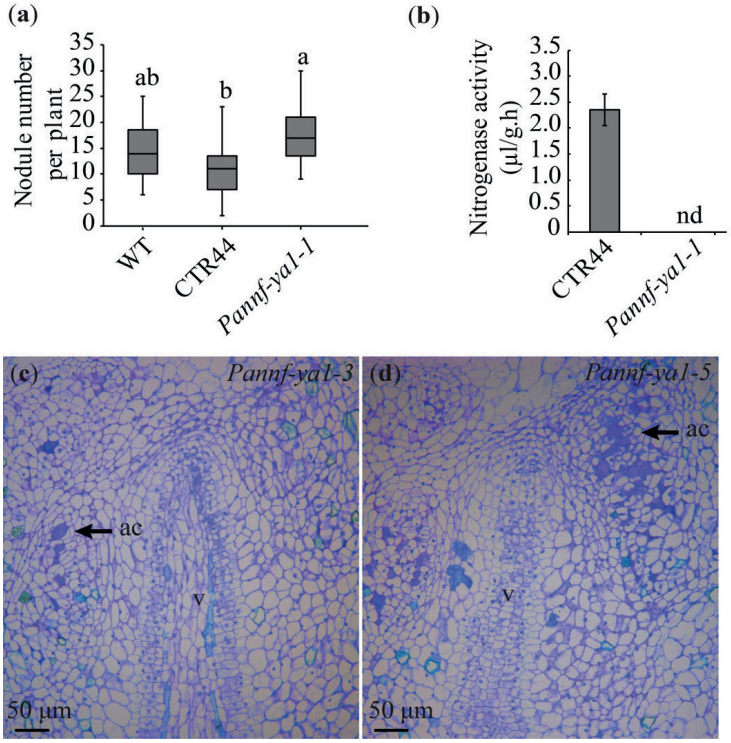


Fig. S6 Phenotyping of *P. andersonii nf-ya1* knockout mutants. (**a**) Average number of nodules formed on wild type (WT), transgenic control line CTR44, and the CRISPR-Cas9 knockout mutant *Pannf-ya1* (line *Pannf-ya1-1*), 5.5 weeks post-inoculation with *M. plurifarium* BOR2. Different letters indicate statistical significance (Student's t-test, $p < 0.05$). (**b**) Nitrogenase activity measured by an acetylene reduction assay (ARA) on nodules formed on transgenic control line CTR44 and *Pannf-ya1* (*Pannf-ya1-1*). Data represent means (n = 15 in panel a, n = 5 in panel b) ± SD. nd.: not detected. (**c**) Cytoarchitecture of nodules formed on *Pannf-ya1-3* and (d) *Pannf-ya1-5*, 4 weeks post inoculation with *M. plurifarium* BOR2. v: nodule vasculature; ac: apoplastic colonies of rhizobia.

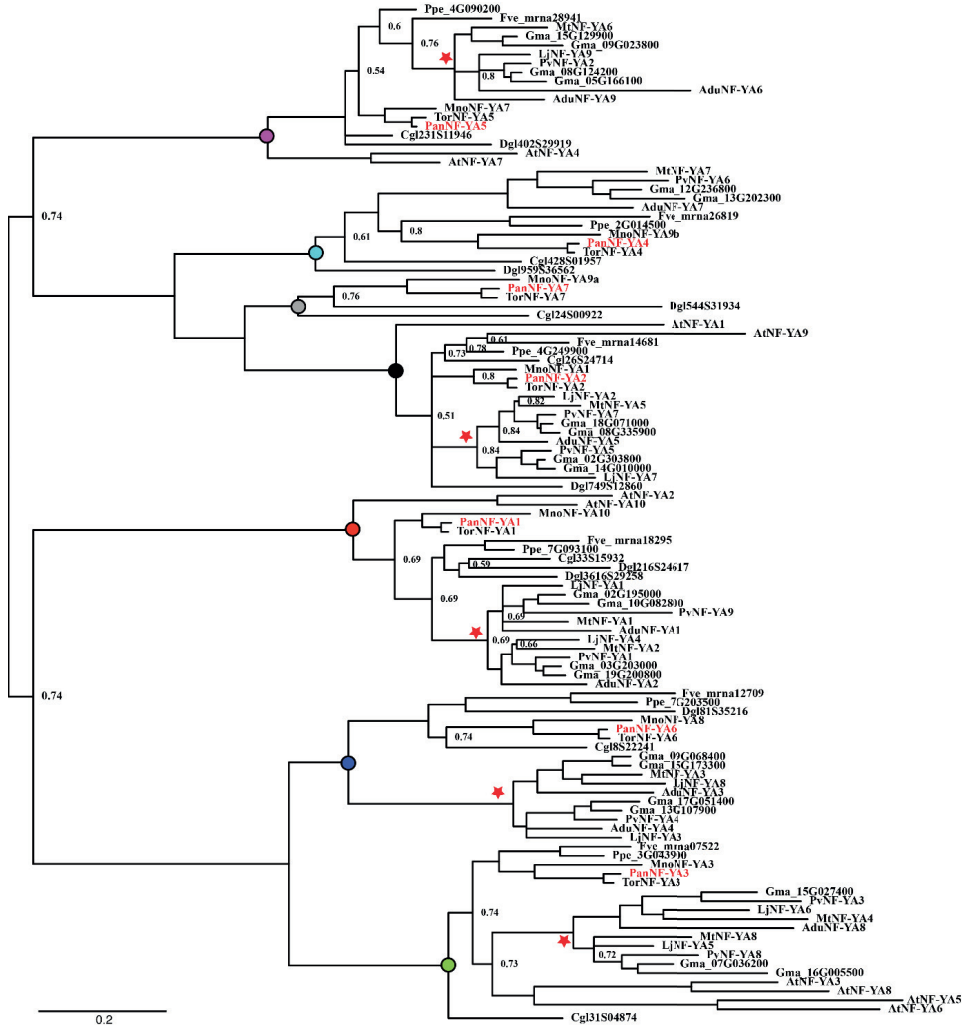


Fig. S7 Phylogenetic analysis of NF-YA in the nitrogen-fixing clade. Bayesian phylogeny of NF-YA proteins reconstructed based on an alignment of protein sequences from the following species: *Parasponia andersonii* (Pan), *Trema orientalis* (Tor), *Arabidopsis thaliana* (At), *Medicago truncatula* (Mt), *Lotus japonicus* (Lj), *Glycine max* (Glyma), *Arachis duranensis* (Adu), *Phaseolus vulgaris* (Pv), *Casuarina glauca* (Cgl), *Datisca glomerata* (Dgl), *Morus notabilis* (Mno), *Prunus persica* (Ppe), and *Fragaria vesca* (Fve). *P. andersonii* NF-YA proteins are marked in red. Red pentagrams mark duplication events within the legume family. Orthogroups are indicated by coloured circles. Node labels indicate posterior probability, Node labels with a value above 0.9 are not shown.

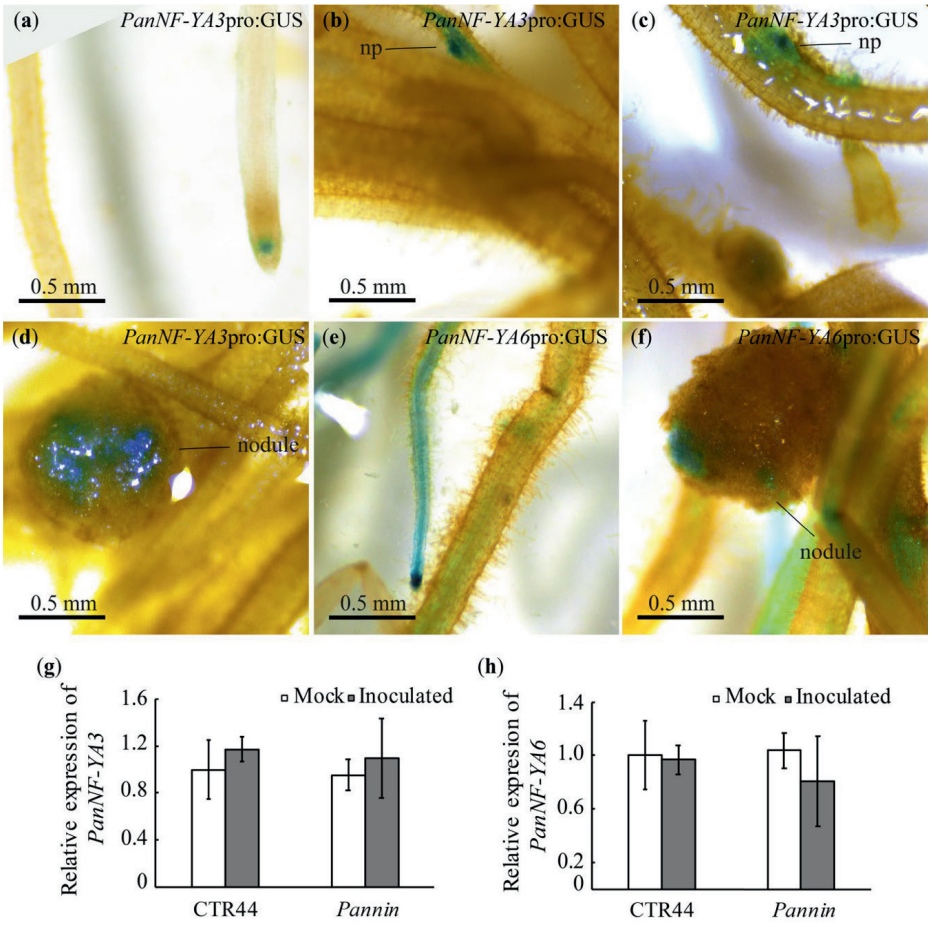


Fig. S8 Expression of *PanNF-YA3* and *PanNF-YA6* in *P. andersonii* roots and nodules.

(**a**) Expression of *PanNF-YA3*$_{pro}$:GUS in uninoculated roots. (b-d) Expression of PanNF-YA3$_{pro}$:GUS following inoculation with *M. plurifarium* strain BOR2. (**e, f**) *PanNF-YA6*$_{pro}$:GUS in uninocualted roots (e) and following inoculation with *M. plurifarium* strain BOR2 (f). (**a**) *PanNF-YA3*$_{pro}$:GUS is expressed in the root meristem, (**b, c**) in discrete spots along the root (**d**) and in mature nodules. (**e**) *PanNF-YA6*$_{pro}$:GUS is expressed at the root meristem and root vasculature, and (**f**) the tip of the nodule. (**g**) Relative expression of *PanNF-YA3* and (**h**) *PanNF-YA6* in non-inoculated and *R. tropici* CIAT899.pMP604 inoculated (1 DPI) transgenic control (CTR44) and *Pannin* mutant (line B3) roots detected by qRT-PCR. Data were generated from the RNA samples used in Figure 2. RNA was isolated from root segments encompassing the elongation and part of the differentiation zone at 1 DPI with *R. tropici* CIAT899.pMP604. Data represent means of 2 independent experiments with a total of 5 biological replicates each ± SE. Data were normalized against the mock-treated CTR44 sample. np: nodule primordium; nodule: mature nodule.

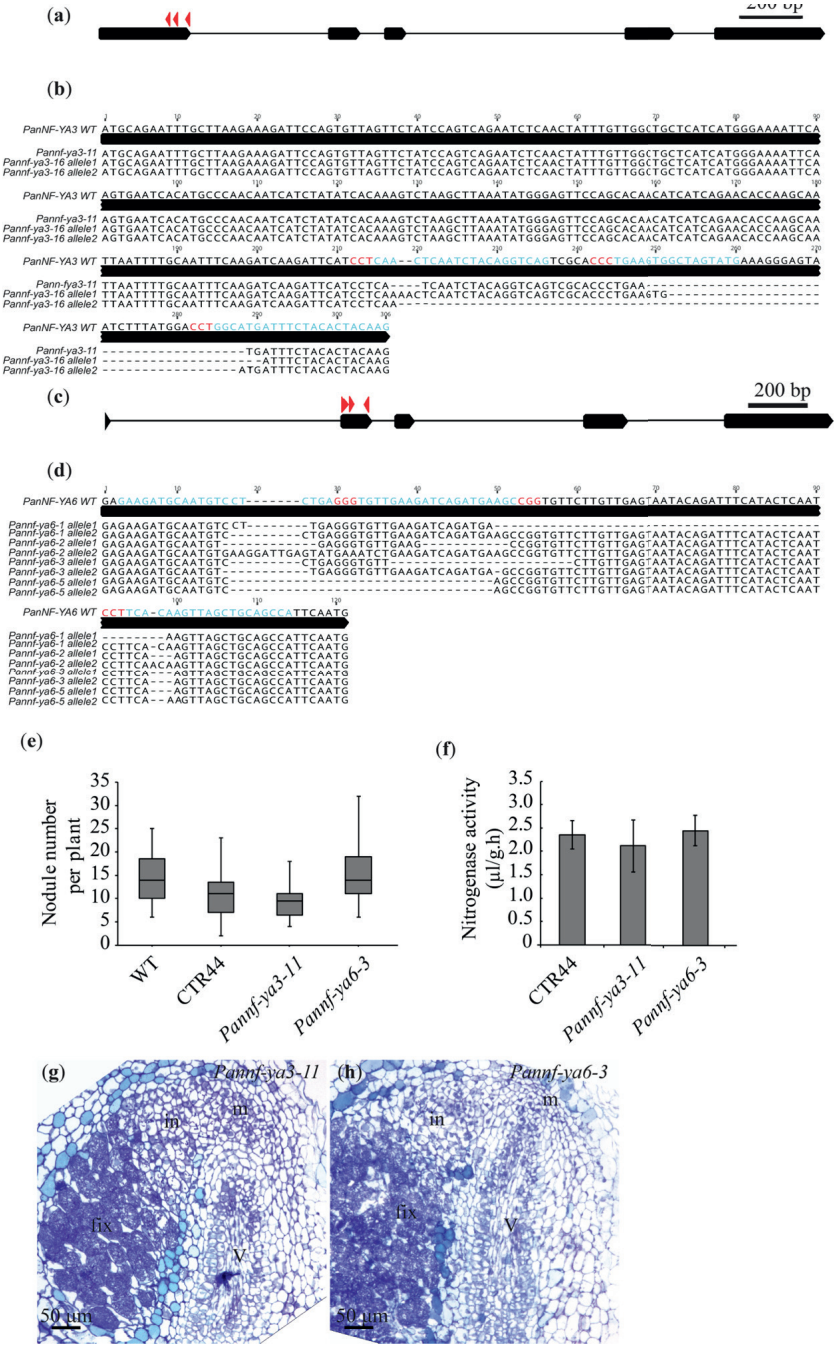
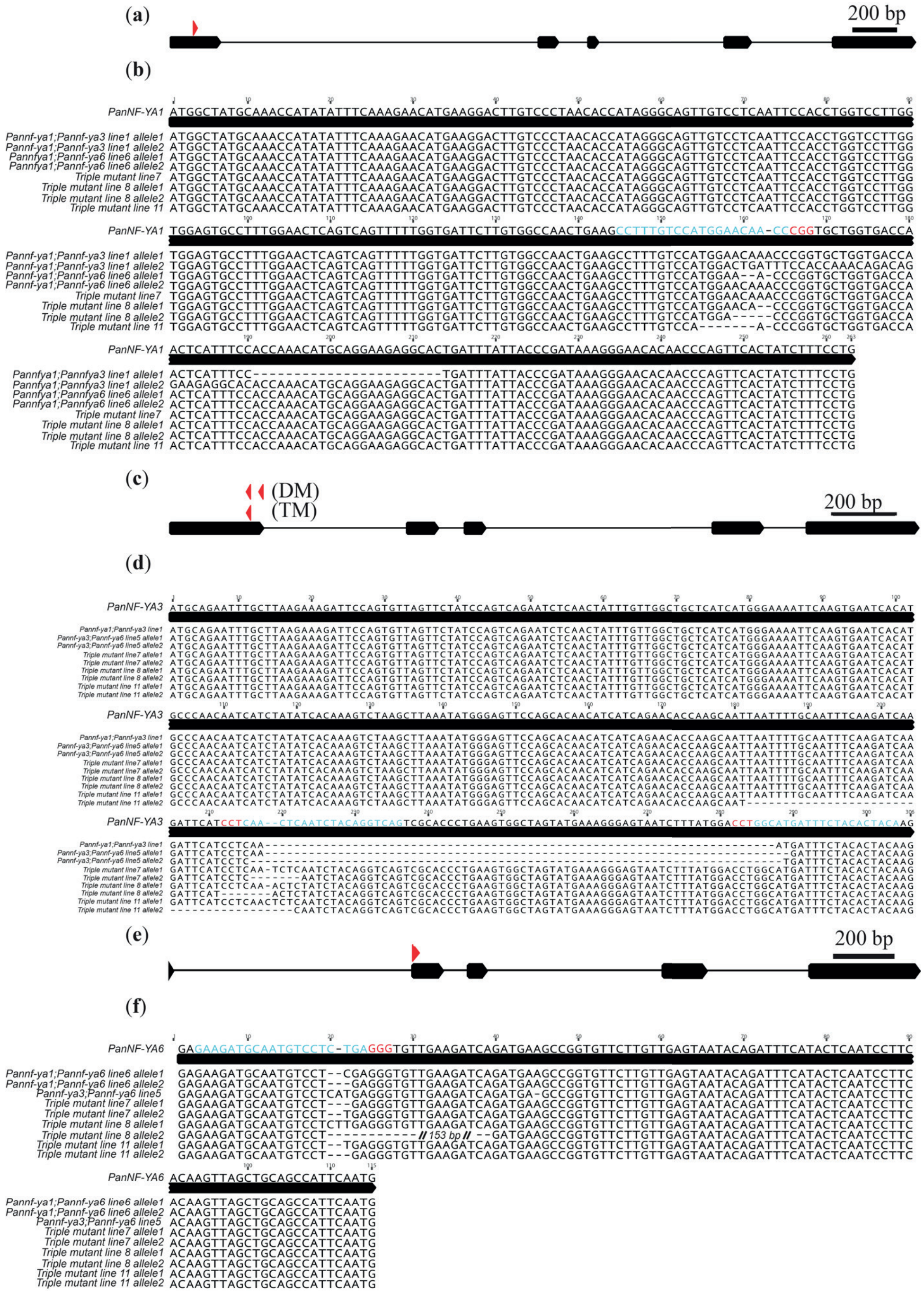

(a)
200 bp
(b)
PanNF-YA3 WT
Pannf-ya3-11
Pannf-ya3-16 allele1
Pannf-ya3-16 allele2
(c)
200 bp
(d)
PanNF-YA6 WT
Pannf-ya6-1 allele1
Pannf-ya6-1 allele2
Pannf-ya6-2 allele1
Pannf-ya6-2 allele2
Pannf-ya6-3 allele1
Pannf-ya6-3 allele2
Pannf-ya6-5 allele1
Pannf-ya6-5 allele2
(e)
Nodule number per plant
WT
CTR44
Pannf-ya3-11
Pannf-ya6-3
(f)
Nitrogenase activity (μl/g.h)
CTR44
Pannf-ya3-11
Pannf-ya6-3
(g)
Pannf-ya5-11
m
Iix
V
50 μm
(h)
Pannf-ya6-3
m
Iin
Iix
V
50 μm

Fig. S9 Gene structure of *P. andersonii NF-YA3* and *NF-YA6,* genotype of CRISPR-Cas9 mutants, and nodulation phenotypes. (**a**) Schematic representation of the *PanNF-YA3* gene model. Indicated are the positions of 3 sgRNAs used for mutagenesis (red triangles). (**b**) The genotype of the *Pannf-ya3* mutant lines 11 and 16. Shown is the sequence of the first coding exon. (**c**) Schematic representation of the *PanNF-YA6* gene model. Indicated are the positions of 3 sgRNAs used for mutagenesis (red triangles). (**d**) The genotype of 4 *Pannf-ya6* mutant lines. Shown is the sequence of the second coding exon. (**e**) Averaged number of nodules formed on wild type (WT), transgenic control line CTR44, and the *Pannf-ya3* (line 11) and *Pannf-ya6* (line 3) CRISPR-Cas9 knockout mutants at 5.5 weeks post-inoculation with *M. plurifarium* BOR2. (**f**) Nitrogenase activity measured by an acetylene reduction assay (ARA) on nodules formed on transgenic control line CTR44 and *Pannf-ya3* (line 11) and *Pannf-ya6* (line 3). Data represent means (n = 15 in E, n = 5 in F) ± SD. (**g, h**) Cytoarchitecture of *Pannf-ya3* (line 11) (**g**) and *Pannf-ya6* (line 3) (**h**) mutant nodules induced by *M. plurifarium* BOR2 (4 weeks post-inoculation). In both cases, nodules are indistinguishable from wild-type. sgRNAs are marked in blue, PAM sequences in red. m: nodule meristem; in: infection zone; fix: fixation zone; v indicates nodule vasculature. Scale bars are equal to 50 μm.


(a)
200 bp
(b)
PanNF-YA1
(c)
(DM)
(TM)
200 bp
(d)
PanNF-YA3
(e)
200 bp
(f)
PanNF-YA6
153 bp

4

Fig. S10 Genotypes of *Pannf-ya1, Pannf-ya3* and *Pannf-ya6* CRISPR-Cas9 double and triple mutants. (**a**) Schematic representation and (**b**) genotype of *PanNF-YA1* for double knockout mutant *Pannf-ya1;Pannf-ya3* (line m1)*, Pannf-ya1;Pannf-ya6* (line m6) and *Pannf-ya1;Pannf-ya3;Pannf-ya6* triple mutants (line m7, m8 and m11). (**c**) Schematic representation and (**d**) genotype of *PanNF-YA3* for double and triple knockout mutants. Notice that a different number of sgRNAs targeting *PanNF-YA3* were used to create double mutants (2 sgRNAs; DM) and triple mutants (1 sgRNA; TM). (**e**) Schematic representation and (**f**) genotype of *PanNF-YA6* for double and triple knockout mutants. Red triangles indicate the positions of sgRNAs used for mutagenesis. Shown in b and d are the sequences of the first coding exons, shown in f is the second coding exon. sgRNAs target sites are marked in blue, PAM sequences in red.

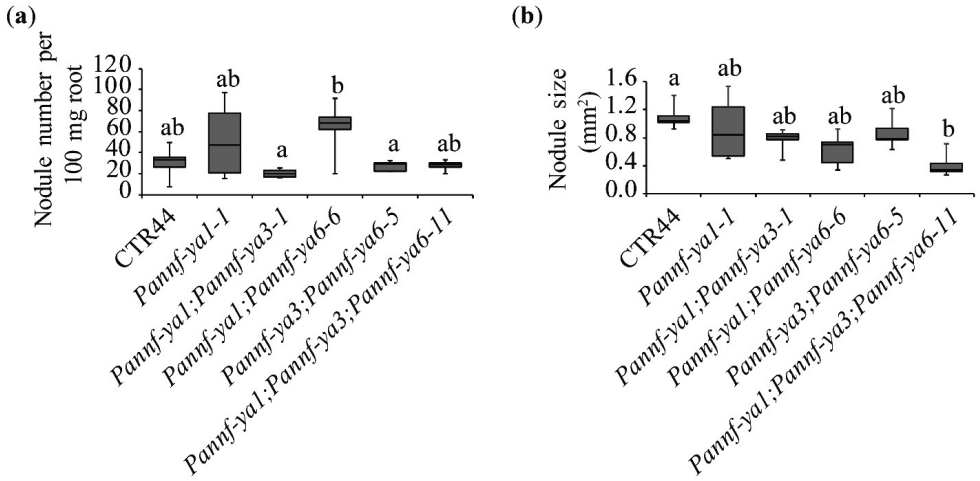


Fig. S11 Nodulation efficiency and nodule size of *Parasponia andersonii nf-ya1* single, double and triple knockout mutants. (**a**) Average nodule number per 100 mg root fresh weight, formed on transgenic control line CTR44, and CRISPR-Cas9 knockout mutants *Pannf-ya1* (line *Pannf-ya1-1*), *Pannnf-ya1;Pannf-ya3* (line 1), *Pannf-ya1;Pannf-ya6* (line 6), *Pannf-ya3;Pannf-ya6* (line 5) and the *Pannf-ya1;Pannf-ya3;Pannf-ya6* triple mutant (line 11), 5.5 weeks post-inoculation with *M. plurifarium* BOR2. (**b**) Nodule size presents averaged nodule size measured by nodule area (2D). In the case of the *Pannf-ya1;Pannf-ya3;Pannf-ya6* triple mutant, only nodule structures as shown in Figure 6B were included in this analysis. Data represent means (n = 4-5 plants) ± SE. Different letters indicate statistical significance (Student's t-test, $p < 0.05$).

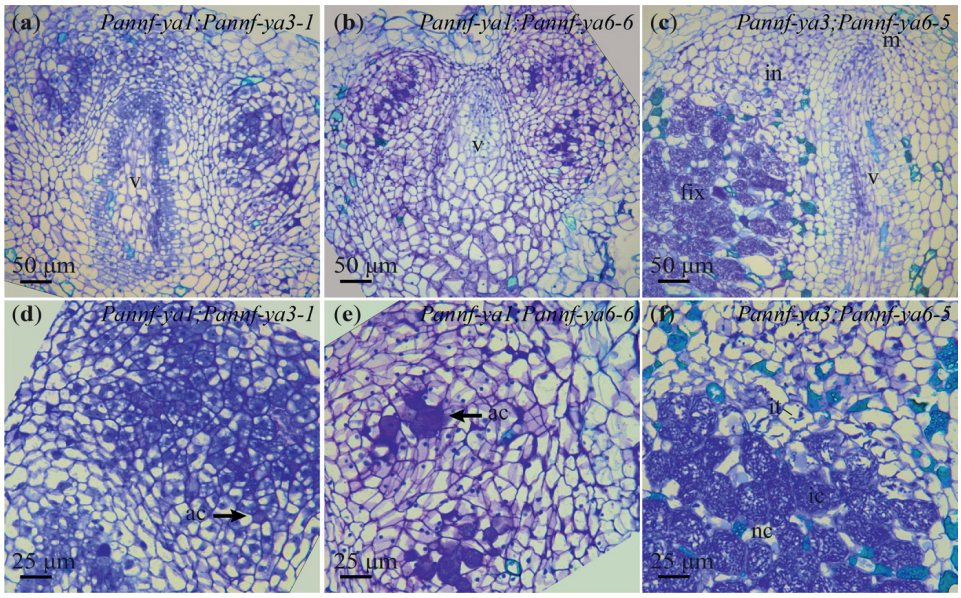


Fig. S12 Nodule cytoarchitecture of *Parasponia andersonii nf-ya* double knockout mutants. (**a, d**) Sections of nodules formed at 5.5 weeks post-inoculation on *Pannf-ya1;Pannf-ya3,* (**b, e**) *Pannf-ya1;Pannf-ya6* and (**c, f**) *Pannf-ya3;Pannf-ya6* mutant plants. (**d, e)** A zoom-in of nodule shown in a and b to visualize the absence of rhizobium intracellular infection threads. (**f**) A zoom-in of the nodule shown in c to visualize normal rhizobium intracellular infection. in: infection zone; fix: fixation zone; v: nodule vasculature; it: intracellular infection thread; ic: infected cells; nc: non-infected cells; ac: apoplastic colonies of rhizobia.

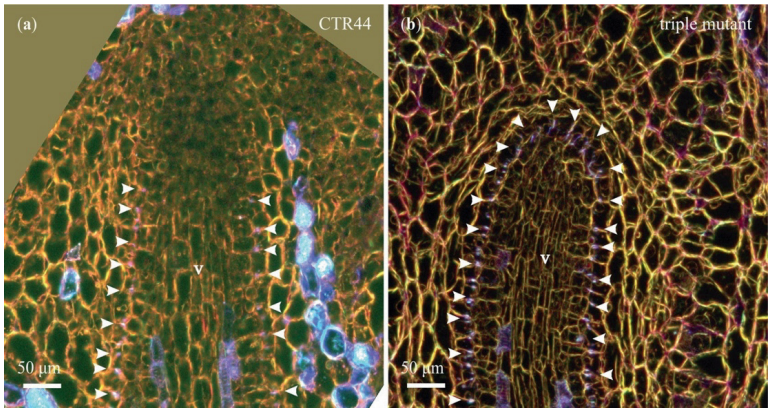


Fig. S13 Casparian strips in the vascular endodermis next to the nodule meristem in *Pannf-ya1;Pannf-ya3;Pannf-ya6* mutant. (**a, b**) Visualization of Casperian strips in nodule sections of transgenic control (CRT44) (**a**) and *Pannf-ya1;Pannf-ya3;Pannf-ya6* triple mutant plants (**b**). Note that casparian strips (white arrows) are not present in the nodule vascular tip in control (**a**), but present in the vascular tip of the mutant nodule (**b**). Casparian strips are detected as auto-fluorescence under UV light. v: nodule vasculature. n = 5 nodules per lines.

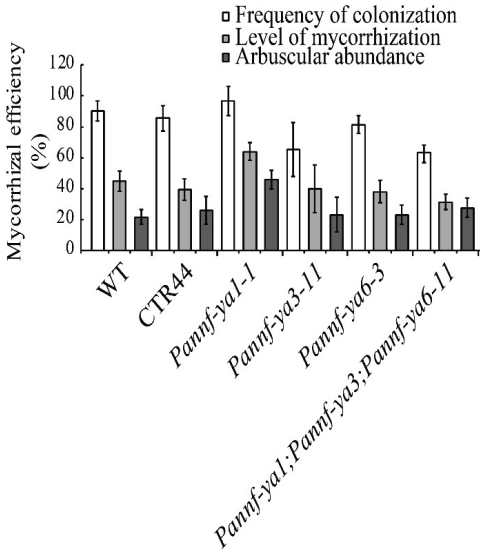


Fig. S14 *Parasponia andersonii nf-ya1, nf-ya3, nf-ya6* and *Pannf-ya1;Pannf-ya3;Pannf-ya6* mutants can form arbuscular mycorrhiza. Mycorrhization efficiency of *P. andersonii* wild type (WT), transgenic control line CTR44, and CRISPR-Cas9 knockout mutants *Pannf-ya1* (line 1), *Pannf-ya3* (line 11), *Pannf-*

ya6 (line 3) and *Pannf-ya1;Pannf-ya3;Pannf-ya6* triple mutant (line 11), 6 weeks post-inoculation with *Rhizophagus irregularis* DOAM197198 (n > 5). Error bars denote standard errors.

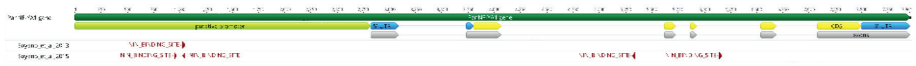


Fig. S15 Putative NIN Binding sites in the *PanNF-YA1* promoter region. Shown is the annotated *PanNF-YA1* gene, including the putative promoter region (2,823 bp) and 5'UTR (977 bp containing an intron of 643 bp) that was used for *PanNF-YA1*$_{pro}$:GUS reporter studies. A single putative NIN binding site with sequence TTAACTTTTAGATGCTAGGTAGGGTTAATT 1,785 bp upstream of the transcriptional start site- is predicted based as the distinct consensus sequences as defined by Soyano et al. (2013) and Soyano et al. (2015).

Table S1: Sequences of sgRNAs used for creating single, double and triple knockout mutants.

sgRNA name	Single mutant	Double mutant	Triple mutant	Location	Genomic target sequence
PanNIN.1 sgRNA1	Y	-	-	exon 1	GATGTTTCGATGAGTGACCG
PanNIN.1 sgRNA2	Y	-	-	exon 1	ATGAGTTGTGATCAAGTGAG
PanNIN.1 sgRNA3	Y	-	-	exon 1	**A**TATGGGTGCCCATTAGAAG
PanNF-YA1 sgRNA1	Y	-	-	exon 1	CAGTTGTCCTCAATTCCACC
PanNF-YA1 sgRNA2	Y	Y	Y	exon 1	CCTTTGTCCATGGAACAACC
PanNF-YA1 sgRNA3	Y	-	-	exon 1	GACTTGTCCCTAACACCATA
PanNF-YA3 sgRNA1	Y	-	-	exon 1	TGTAGTGTAGAAATCATGCC
PanNF-YA3 sgRNA2	Y	Y	Y	exon 1	CTGACCTGTAGATTGAGTTG
PanNF-YA3 sgRNA3	Y	Y	-	exon 1	CTTTCATACTAGCCACTTCA
PanNF-YA6 sgRNA1	Y	-	-	exon 2	TGTTGAAGATCAGATGAAGC
PanNF-YA6 sgRNA2	Y	Y	Y	exon 2	GAAGATGCAATGTCCTCTGA
PanNF-YA6 sgRNA3	Y	-	-	exon 2	TGGCTGCAGCTAACTTGTGA

Table S2: Primers used in this work.

Primer name	Primer sequence	Purpose
sgRNA universal Reverse primer	TGTGGTCTCAAGCGTAATGCCAACTTTGTACGT TTTAGAGCTAGAAATAGCAAG	Cloning of sgRNAs
PanNIN.1 sgRNA1 F	TGTGGTCTCAATTGATGTTTCGATGAGTGACCG GTTTTAGAGCTAGAAATAGCAAG	
PanNIN.1 sgRNA2 F	TGTGGTCTCAATTGATGAGTTGTGATCAAGTGA GGTTTTAGAGCTAGAAATAGCAAG	
PanNIN.1 sgRNA3 F	TGTGGTCTCAATTGATATGGGTGCCCATTAGAA GGTTTTAGAGCTAGAAATAGCAAG	
PanNF-YA1 sgRNA1 F	TGTGGTCTCAATTGCAGTTGTCCTCAATTCCACC GTTTTAGAGCTAGAAATAGCAAG	
PanNF-YA1 sgRNA2 F	TGTGGTCTCAATTGCCTTTGTCCATGGAACAAC CGTTTTAGAGCTAGAAATAGCAAG	
PanNF-YA1 sgRNA3 F	TGTGGTCTCAATTGACTTGTCCCTAACACCATA GTTTTAGAGCTAGAAATAGCAAG	
PanNF-YA3 sgRNA1 F	TGTGGTCTCAATTGTGTAGTGTAGAAATCATGC CGTTTTAGAGCTAGAAATAGCAAG	
PanNF-YA3 sgRNA2 F	TGTGGTCTCAATTGCTGACCTGTAGATTGAGTT GGTTTTAGAGCTAGAAATAGCAAG	
PanNF-YA3 sgRNA3 F	TGTGGTCTCAATTGCTTTCATACTAGCCACTTCA GTTTTAGAGCTAGAAATAGCAAG	
PanNF-YA6 sgRNA1 F	TGTGGTCTCAATTGTGTTGAAGATCAGATGAAG CGTTTTAGAGCTAGAAATAGCAAG	
PanNF-YA6 sgRNA2 F	TGTGGTCTCAATTGAAGATGCAATGTCCTCTGA GTTTTAGAGCTAGAAATAGCAAG	
PanNF-YA6 sgRNA3 F	TGTGGTCTCAATTGTGGCTGCAGCTAACTTGTG AGTTTTAGAGCTAGAAATAGCAAG	
PanNIN.1 geno F	CTCAACTTCACGCAACCTGC	Genotyping of CRISPR mutant lines
PanNIN.1 geno R	TCCCACGCTCAAAAATGGGA	
PanNF-YA1 geno F	TCCCCCTATTTTGGTCTTAGTCT	
PanNF-YA1 geno R	TGCAAACAACAGAGTTATAGGCC	
PanNF-YA3 geno F	CCAGCACAACATCATCAGAACA	
PanNF-YA3 geno R	TTGTTAAGCAACGTAGGGAACT	
PanNF-YA6 geno F	ATCTGGGTGGACAGGCAATG	
PanNF-YA6 geno R	CTTACAAGAGCCCGTGGTCC	
PanEF1a qF	AGACAAGGTTAAGCGTGCAG	qRT-PCR
PanEF1a qR	TGCAACTGGGCAACAAACTC	
PanNIN qF	TGGGAATGGGACTTGTTTGG	
PanNIN qR	GGGAGGGCTGAAGTTTTGAA	
PanNF-YA1 qF	CAGTCATCCCTGCCAGAATATC	
PanNF-YA1 qR	TGCAGTCAAGTTCAGCGG	
PanNF-YA3 qF	TCCCGCTATGATTCACCATTCC	
PanNF-YA3 qR	ATTGCACGGTACTGCTTTGC	
PanNF-YA6 qF	CCATTCAATGGCTGGTGTGC	
PanNF-YA6 qR	TCCAAAGGCAATGGAACTCG	

Table S3: Putative promoter sequences used for promoter-reporter GUS assays.

In red: nucleotide mutations to remove BsaI and/or BpiI restriction sites, which is essential for GoldenGate cloning; in blue: putative NIN binding site; underlined: 5'UTR; yellow highlighting: exon sequences in 5'UTR, in small letters: intron sequences; in bold: translational start site.

>*PanNF-YA1*$_{pro}$ (PanWU01x14_284830)

ATCGAAGCCTCCAAAAGGGGGGCAGAGTTATTAAATGATGAAGAGAATTTTTAGGTCACCTT
AAGATGTGGAAATTAAAGGTTCGATGTAGCACAATGTAAGCATACATTATGTTAGTCATGAT
GTTAGTCACTAATTACAAATAGTTGTACTTGTATTGGTCATATCAGTCGATCTATGCTATTAC
ATTGAACATTTGACATAAAATCCCAATATAATCCATATATTTAATGCTAAATTGACAAAATT
ATTACCAATATATATTCCTACTCTAATAGGTATTTTTTTTTTACCAATTTATGTGTATTCTTTTT
AACCATTGGATCATTATCTAAATATTATTATAATTTGGATAAGTGGAAGAAAAACAAAAATA
AATTAGGTAAAACAATACGTATAAATAGATAAAAAAAAATGAGAAATTTAGGTGGATGACTA
ATATTTTATATTTGATTAAGTAGATATCTATTTTTTAAATTTTATAGCTAAATATCTAGGTATA
TATATTTTTATACTTAGTAGATACATTAATTTACCAATTTTAGCTACTGTACTGAGATGTACT
GGATATAAAATATTTTATATTTAGTATATAAAATTTTACACTTGATATATAAGTATATATACC
GGTTTATTATAATATTTGATAAATAGATATTTACTTAAATTAGGCTTAAAATTTGATCCTGTT
TCTATAAGCCCCAAAAAAATACCTATTGAATCAGCGTCCATATATTTATATATATACAAATG
ACCAAAAAAAAAAAAAAACAATTAGAACATGAAACTACTAATGTTTATTGGCCAGAACCAAG
AAAGTTGCAACACTTTTGTATAGGAATAATCTGGTTACAAGGTCGTCAGATTCCTATTTACA
TACTTTAATAAATGCATTTCCACCAAAAAAAATAAGAAAACAAAAAGCTCATTCCAATCTCA
GCTAGTCTCTCTGCCCTTCTCAAATGACTAAGGGTCAGGACACTTTCACGAAATATAAGCTTT
AGTTTTGGGGTTTACTTCATATAAAGTTAACTTTTAGATGCTAGGTAGGGTTAATTACAAAG
ATTAAAAAAAATGACAATTATAAGTGGAAATGACCGTGACAGTGATAAATATTCAACACCAT
TAATTATATATACAATTGCTAAGCTAATCAGAACTTTTCACGCATATAAAATACAATATTCTC
ATCCCAAATTAAGATTTTTATATTTGAAATGTGACATAATAGTAACATGTTTTTTTTGGTCGT
ATACATTTATATATATGGTTCCTAGTAGGGATTTTTTTTAACTTATCTATAACTATTTTTATTT
TTAGTCATTAGATGATGATCTAACGATTATTATAATTTGAGTAGTTAAAAAAAATAATATAA
ATGAATACTCATTATTCAATTCTTTTCTCTTTCTATCTTTAAAACATGTGTGATGATTTTTAAT
TAGGAAAAATAATACTTATGTATAGGTAAAAAAAAATACCTACTAAATATATATATATATAT
TTGAACACTAATCCAGTAGGTTTTTTTTTTTTTACCTATTTATAGGTATTCTTTTGTCCAATAA
AAATTCATCACACGTATTTTTTATGATAAAAAAAAAAGAGAGAAAAAATTATTAAATAATA
AGTTTTTATTTTTTTATTTTTTTTTTATTTTTGATTGATAACCTATTTATATATTTTCCCACTACG
GCCTGCTAATTAGCAACAGCCATAATCTTGACTGTCATTAAGTATCTTTAGCCATTAGAATTA
TGGTCATATATATCTTCAAGCTGTTAATTTGCACCACCGAATGAATATTATAGTAATATCATT
TAAGCTGAAAAATATTACTATACTATGTTGGAGCATTTGCCAACTCAATTTTTTCCCACTACTT
CACTTGAGTGAAAGGGTAAATTTTTAAGTGATACTGATCCGGCCTCTATGAATATGTTAGCT
AATTATACTAAGGTAATAATAGGAACTAAAGATTGATATCAATGGTTATAAATATTGTTTGG
AATGTTTATGTACACATCACTGAAAAAATTTTCTGAAATTGTTTTTCTAACACGTTTGTGTTC
ATGAGGAATAAAAAAAAATTAACATTAACAAAAAAAAGGACACTAAGCTGTTTTAATTTTT
AAAAACCGGAAAAAATAAAATTATGGTATATATATACAACCTAATCCCACATGCCCTGACA
CTCCTTTTTATTTATTTATTTTTTCAAGAACAGAAAAAAGAATTTTGAGTAGTAGGAGAAAA

GTCTATTTTCTGACAGTTATATAATGTTTAAAATTAAACAATACTCAAGTATATAAAAATAA
CAAAGCAGTTTAAGGATTATAAACTATTTATATAAATAGTATTAAAAAATGGAAATGAAAA
CTATGACATCCCATCGGACCATGGACATGCCCTGGAATTAGAGAACCAAAGGCAAAAACCA
TGCACTCCACAACTCCTAAGCAAAATTTTTAACCTTATCCGCATTATTTTCTTAAACTTATCT
GGAACCCACTTGTATTATATACATATAATAAACAATACGAATTTTACATTACATACATATAT
ATATTATATATATATATATATGGCAGGCTTTTGTATATTATATATAGTAATGAATAATCCAGT
TTTTTTTTTCCCACTGTGACATTCCAACACCCCAAGTTGTGAAGCTTAAGAAAATATTAAAGA
GAAAGGGGTACCTCAACTTTAGCTTAATTTGATAAAAATAAACATAAATAAGAATTGATCAA
AATATGTTGGAGAAAAACGGTGAACCCGAGAGAGAAAGGGAGCACTAAATTCTTTTGTTTT
ATTTATTTTTTTAAAAAAATATATATATACATATTATATAATTTGATCTGTAACACTATACAT
ATTGGCACATACAATACAAAGAGGGGTACTTGCATGAAAGGCAAAGCTTATTGGCCATTGG
CGTCTAGTTCCATTAGAATTGTACGCATTTATCATTAGGGGGTCGTCTCTCTCTCTCTCTCC
TTCCCCATCTCTCTCTCTCTCTCTGTTCTTGAGCGTTGGACACGAAGCAAACTGTTGCTGG
GTACTGCTATTCAGCTAGTTTTCTCTCATATgtgagctctcaaataccaaacttttttctattcttagccttgttgctccaacatagt
tggagggtcttgtcaaagtttctttttttctttttaaagcttgaatttttttctcccttctcatttttgatttctgggttttttttttgttctgagttgatgcttttgttctttttttttt
tcccttctttttttaaatgtatctgtcagttaaagaagatgtaatttcagtgatcaaacagacagaacagaattgtttgtaaccaatatttgatcaaaataaattttga
gagctcattcacaattttcatttttttttttcttcaatgaaacaccgtggttttcatctttttattgtttcatttatatcgacaatttatttttttttcataatctcgaatcaaatca
attttgatcacttattatccttctgtatgagtctgctaatcagaccagaatttttttttttttgggccaaatttttagttcaatattgaaactcaatgaaatttatttccccct
attttggtcttagtctttgtcaactatagtatttttgtcatatttttcaatctcgtgtttacatagaatgttgttagatatagtcttatagactcataattaaggttggtgg
gttttcatatttgtagGCGTTAGATTCAACAATTGGTTTGAAGCTAGCTAGGTACTGTGGTTTTGATCAT
TTTGCAAGTTAGCT**ATG**

>*PanNF-YA3$_{pro}$* (PanWU01x14_246880)

CATAACGAAGAAACTCTTGGAGTAATTAGAAGGTTGGTGGTGGATGATGGGTTGCTTTCAAA
TTTTGTATTTTTATTTTTTAAATGCGACTTTCAAGGTCATTTTATATTGACAGCAGATAAATG
ACATTGATTATATAAAGTAAGCTAGCTAATTCAGGGCTCTTTTGTAATAAATGCTGTGTAGA
GGGTCATTAATTTAAAAGCTGCCATATTATCTTCCTATTAATTGCATTAATAAGTTTTATTTA
ATTAATTGAAATCCATGACAAAGATGGATATCTAGCTACCTTCACTCTCTTGGAGCGCGTGT
GGAAAGTCGTAAGAGAATATTAGAATTAGATTTAGAAGTACATGAGAATTCTTGGAAGCTC
CTCATCCTCATTCCAGAGTAATGGGGACATGATTCAACTTCTTACATGTTATCCAAAAGCATC
ACACAAAAAAAATTCAGTTGGTAAATTTAATTTTTTAAATTTGATCATGAGTTTATTCTAAAAA
CAAATTAATGCAAAAAAAAAAAATTCTTATGTCTAGATTTTCTGTAATTTAAATATTAAATTT
TGGAATCTGTCAAAAAAATATATTTGTATTTTAAAATTTTAAACATTTGTATCCTCTATTCTT
ATTTTGTTAGTTCAACAATATTGTTTTTTTGTAAATATTTTAATTTTTTAAAAATAGAAATAAC
ATAGTTTGGATTTGTAATTTTTTGAGAAATAAATTTATCAAAATATGGGCATAGGATACAAA
TTTTTTTTACGAATTTCAAAATGTAGACATTTTTTTTGGGTAGAAATTTAGGATTTAAACCGC
ATAAGGACTACTAGATTTCTTGGAATTTTTTGCAGTTAACCCAAAAAAAATGACCATGAGTT
GATTACTAATTTAGACCACTAATATTTCTTTTTCTTAAGATAATAGTTGCATTTAGGGTAAAA
TCTTGTATTTGAAATTGAATGCATTTTTTTTGTTGTTAAGAAATAGACTAAGATAATATGTGG
TGTTGAATACCCTATAATTATATGGTCTTTAATATTATTCTTTTGTTAATTGCATTCATTTCGC
AAGCTGTAAAATGCTTTCTACCTATATCAAAAAAATTAATAAGAGATAATTTATAACCTTTA
ATTTATTTTCATGTAAGCGATTAAATTTCTCTTAGCCTTATCATTTCTGAATTATTCAAATGAA
TCTAATTAAACTATAGGTCTTATTATTAGAAAAAGTTTATATAATAGAAAAAACATAAAGGG
TTTTTACTAAACGTAAAGTTAAAATTAATAACTTCAAAAAAAAAAAATTTACCTTAAAACAA

ATTAAAAAAGAAAAAGAAAAAAAAGAAAACAGGGGAGATAGTTGATTGATTTCAATGGGC
CCTTGTAGCGCGTGGGTAGAACACTGGCCGTAATATCAAGTGTGGTGACAACTTAGTATTTA
TTGTACCAGAAACCAGAAAGCGTGGGAAATTCATATTGTCGATTTACAGCAAAATTTATAAT
CTCTCGTACTTAAAAATATATATAAAAAATAAATAATAAAGAATAGAATTTTATAATCATTT
GGTGCTACACGACAAACACAAACTACAATTTAATATTATGTTATTAGAATAATACGTATTAT
AATATACATATGTATATGTATACATCTAGTAAATGGGAAAAATGTAAGATTGGATTGATTGA
TATAAATTTTAATTAAGAATAGCAATAATAATAATAGAAGAAGAAGAATAAAAAAAAAATAA
AAACTCGGAGAAGGTGGGAAGGTGGTGGAGATGCTGGTGAAGTCCCTAGTGAGTTTAATGC
AATAAGAATTGTTTAAGGAGAGATTCAGTCTCAGAATTGACACCAAAAGCCAAGTGTGCGA
GAGAGAGAGAGAGAGAGAGAGAGAGCTTTGTGGGGTTAAACAATAAGAGAAGAAGAGGAA
GAGGACATGCACAGTTCCAAGTCTACAACGATTCCCATTTCTTGTTTTTCCACACTTTACTAA
ATCCAAAAAAGGTCCTTCCTTTACCCAAAGgtttatctttttctctctactttcttgccttttcttcaactgggatttcttcttcttcttctt
gttctctctcctccttttttttttgtgtccttatttttttgtttttttaaactgcccatttttatgtctgctactttttttgttccctcgcggatatcgttgtgtggcagtgaatgtt
acatgagattgcatgtggagacttaagatggtaaagctgatttttagtgctattggggttgtaagttatatgggtttttggttttttgtgagggttatatatgggatg
ctgagttataggttgttcttatgggttcatttgatcattcatttgtctttgtatctattggttccattccaatcatgtgcttttctgattttgaaggtggacttacttgaca
cagtttattcagccttgacttctctctttttgctccttatttgatacttttttcctctttcatagaaagaatagtgtttttaatgaatggttctgttctttgggttcttgacact
taacctcactgctgatttggtttctatgttctgaccatgatatttgcttgttcccatgtcatctttattcttgcaatcattttcttaactgggaattgctcacgtcctctg
catatttagttattggataggtgctgccttttgactactttgctcaactatttttgttctcattcttttcaatcaagtgtgttcatgtacattaaaacgtgctgaattgatc
gtgttcatatgaaagtgaaatttaggattggaggaaagtggaaatggaagtttgtgctatgatttggaactaccgccgtagcgctgacgtatctttgatatttg
cttggttattggttcttattctcacttttgtttatgatttttgggtctgggctaccgaggcttagtggatagagatgataaactagggtcacctagtggtggaggta
caatctgcatctgaattctgtgtactaggacagatgtctgtatgtatatatatatatatatatatatatattcttttctcatatccaagacaactttttgttccatacttt
tgctcaagtccattgaaggcctctctatggagactctttattcttttctatcccggagataggggtggagtggcattttaaattggtttccttgctgcctacctttt
gttagcttacctttttgtggattctgagagtcatgatactttttttagttacgattctttttttaaccaccggtgcattaatttaacctacaaattgagatttttatcttaaaa
tgcattttcagtgctttggtaatataagatagtatcattttcgttctgttgcttagaactgacgtattgtacattgtggcagTTTATGCTTTGTTAGG
GTTCTTGGTTGGCAATGAGTGCACAGCCTAAATATAAGgtcagattctcctgcctttacatatatgaagatctaggaag
tttttcatcataatttctagatgctttgagtcgcagCTTAAAGAGA**ATG**

>*PanNF-YA6*$_{pro}$ (PanWU01x14_192330)

ATATGTCACTAGTCGCGTGCATGAGACTTTTTACAAAGACATTTTAAGCATGTTGCAAGTTA
AAAAAAAAAGGAAGAGACAATAATAATTTTATTTCATTTAAAAAGATAAATAAAAACAAAT
TTTATTTTAGTACAAAAGTTACTAATCTTTTTTGTGGTCATGATTTATGTGAATTTCTATAATA
AACTCAACTATGATTAGGATATATTGTTGTCTTTAATTTTTAAATCAATTCTTAGGAGAATTT
ATACAAAATAACTCAAAATTTTCGTCTTTCTTTTATTTATGTTTTATTTTTCCCAAAAATTTTA
CATTGCACAAGTTTTTGTTTTCTTTTCTCTTTTTTATAGTTTATTACGTTATTACTATATAGTAC
TCAAATATTCAAAAAGTATATATATATGGGTACCAATCCAGTAAGTATTTATTTTTACTTATA
CACAAGTATCATTTTGACCAATAAAATTTTTTTTTTATATTAAAAAGTAAAGGAGAGAAATA
ATATTAATGATGCATGATTTTTCATGCACTTCATTATATAGTAGATTTTTTTTTTAACCTATGTG
TAGGTATTTTTTTTAGCCGTTGGATCATCATCTAATGGTGATCTAATAGTTATAATGAAATGC
ATAAAAAAGACATGCATCATTAATATTATTTCTCTCTTTTATTTTTTAATATAAAAAATATTAT
ATTAAAAAGTAAGAGAAAGAAGTAAAAATGTTATATATAAAACTTATATATTTATATATTCG
CTTATATAAGTTATCAATCCAGTAGATATATTTTTTCCCTATGTGTAAGTATTTTTTTAACCGT
TGGATCATCATTTAATGGTGATCTAATGGCTATAATGAAGTACATGAAAATACGTACATTAT
TAATGTTATTTCTCTTTCTTACTTTTAATATAAAAAAATGTGCAAGAAAAATTTTTATTGGCC
AAAATGATACCTATATATAGATAAAAAAAATACCTACTAGATTGGTGCCTATATATATATAT

ATATAATACATATCCCTTATTTCAAATAGAAAAATTGCCAATTTATTTGATAAAAGCAAATT
GCCCAAAATTTTTTTTCAAATCTACTTCAGTGTTACATACTCATTTCTAACTATTCCTAACAAT
AAAGAAATGATAAGGTATTCACTCAGAAAAAAAAGAAAAGAAATGATAAGGTATAAAATG
AGATACTCTCATAAATCGAGTGCCAAGACAGAAATATTATAAAATTTTAGTCTTATGTATAA
TTCAAGAGCTAAGATAAGGATGTTTTTTTATTTGTTTGTAATTGAAGTGTCACACAAGGAGTT
AAATAAGATGGAAGAAAAACATCTGCGTGCCAAAATGAGGACATTTTTGAAGTTTAGTTGTC
AAAATAGGACTTGAATAATAATCCGGGATCTAAATATGGGATTAACCCCATATTCAGCTATT
TTATATCTTAATTTTTCATCCATTTATTTATTTATTATCTGTTCCCCTTTGCCAGCTGCTGAGA
GGACTTTCCTCTGACTGGGTCTGGTTGGGAAAGGAAGATAAGGAAATATGTGAAAAATCTTT
GGAGTTGGATATGTTCTAAGCAACTGAGCAGATCATTTCTTGGCCATGCCACGTGTATGGCA
TAGTTGCCCTGCGCTAACCAAATATATTAAACGAGTCTCTCAAGTCATCCTACTACCCGACA
ACACAGGACGATAACGACAAATACCTTTTCTTTTCTTTTCTTGTCTTTTCTTTTCCCTTTTTTTC
TTTTCGTTCTCTTATTATTACCTACATCCTATTCATACCAATTTGTCTATTTACTCAAGAGCCC
CACATTGCATCATACATTTACAACCACTGTAATACGTCACCAAAAAAAAAGGAAAACAATTT
GTCTATATTACTGTTGTGCGATCTTGATATTGAGTCTACTTTATTAGTTAACATCGCATATTA
ACATTTTAGCCATAGATGCTACCTGATCCGAATAAAAAATATATGAACGGGCGATAAAAGA
AAATTAGATGAAAATTTTAAATGGATACTTGTGAAGAAAAAGAAACAAATTTATTATAAAG
ATCAAATAATTTATTTATTAGTTGTTATTTAATTTTATATTGATATTTAAAATTATATTTTCTG
ATTATGTAATTTTGTATTATTATTCAAAATTATATCTTCTAATTACTTTCTTGTTTTCGTAATA
AAAAATATTGAAGTACCCTCGACTATATTCTTACATTTTTATTTTGTTTTATTATTAAATTCAT
TTTCACACACAAATTTAATCTAATACATCATCACATTTTTCTCTCCTCTCTTTTGATTTATTGA
GAAAGAAAAAGAGCTAATTATAAATCTGATATACTTCGTGGGGAAATATACTGCAAAATCA
ATTGAATCGAAACCTTCCAATGAAAAAGGTTAAAAATAAAGGTATGACTTCATAAGAAATT
CTTTTTCTGTAGTAACCTTTCTTTAATTTCTTTTTTTTTAGTTATGTCATGCAAATTACTCGTGG
TCCAAATGAACTCCAAGATTATACGTCGTTTTTCTCTAATTATTGTCGGTGAGTTAGTGCTGG
GTATTTCCGAAAATAAACTAACAATTCCGGGGCCAATACTAGTATATACGCGACGATAAGTT
CAGGCAGCATTCTAAGTCACCAGAGAGGAGAGTGACGAACACCGCCAATCAGAAAACACGA
AGGCCTCGGAATTTGGATCGAAAGTGATTGGTCGGAGGGATCGGTTGGGAGTGAAGCGTCA
GGATCCGAGCGTGGGCCGAGAAGTGTTGACACAGTGAAGGGCGGCTGACAACAGAGAGAC
CGAATGAAGCCAAAAGGGTCTTAGCTCTTAAGGCCGTGTTTGGACGTGACTCGTGGGCCCCA
CCCAGAGAAAGACGTTAAATCATAAAATATAATTTGATTTTTTAAAAAACAAAAGCCAGTA
GTGAATCACGACAGACAGAGAGAGAGAGAGAGAGAGAGTCGGTAAAAGTGGTGGTGGGCA
GTGGGTGTTTTTATTGGGGAATTGAATTAAAATGGTTTTTTTTTTTTTGTATTATTATTTACAG
CAAACCCAAGAAGAAGATGATGAGAAGAGAAGGGAGAGAGAGAGAGTAAAATATAGATAG
AGAAAGAGAAGGAAGAGGAGACATGTCCAAGTTCATCCCAACAGCTCTGTAGCTTGCTCCT
TCATAGCGTATCAGATTCTATTCCAAAGgtcgctctttctcgcttctttttttcttttttccttttatttatttgttattattctttctaatcccgt
ccatctccactagtttcccacctttccccctttttttctttttaatgagtggatttctcgtggggttcatttattggtgtagaatataagtgggttttgttctatttgctgttt
agcgagtccatatatacagagtctgagatcattgtggtgtttgttgaggaagggaaagtggtcctggtaaagtctttttatggatgatgttggttttctcattctg
gccttatccctgttttggatcttctgctttttcagttcttggttttgggtttatcatatctttgccatttcatctttatctttttaggtttatgggtatttttcatttttttttttac
tgcatttggttgaaaataaatggatgtgtttatatttgtataaagtttggggcttaaggctttaattagtatgttaaattgttaatatgggtggtttgtgcttatttattt
gtgggtttcatagttgatttcttaagcgatctgctagtttgagctttttctgttgatttggacatatactgcatgtaatgttatctttcttgcccactgaaccataggtc
tttttacttgtaatctgggttttgtctaggtacttttttgagaaaataatcaacaattttggaaaaaatttttgtacccttaagtttacatcttgatgtgtgcctttaagt
agtagctttgtgatgtttgtatgttacatgtggttgtttggttatgtcactgaattttgtacgcagcaagttttggagtcaattattgtgcttgtgggatttcttgtttag
tctgctttcttgctggactagtgcttatgaaacaatggcacgttacctaatctgttattgaaagagctagaatagggaaatgtttaaaggaaacaaaaaggaga

attatgaaaaggaatgcttgatatctattttagctgtttattaatacttattatgtgtaggactcacataatctctttgctaaatttgtaagtaaggacattaatgaata
aaggaaggaaatactccctttcattctctttccatttattacacggaaacatgttatagatatgatagatttttttttaaatcaattttaactctcaattgggacatgc
ctgtactctgtgcatggtagtacgcgtatgcttacatacatacatacatgcatacatacatactatcttggttggcaactataaaaatacattctctcttaagatat
ggagacagaaatagctcattttatgttatcttccaattttggacttttattatattcctgttatttttatcttggggattgattgtttatattgactagattactctttgctat
tgaaacttgcccaacaaaagtgtcattccgagataattaagttagcacctttttaccattaatttccctcaaatttcaGCAAATTCCAGACTAA
ATGTAAGCTTCATAGTGGAAGTTCAACTGGATCTTCATCAACGCGTTACTAAGCATTTAGGG
TTTCAAATTTGGAGTACTTTAGAAGGAAACATCTACAACACGAGCAATTGGTCAATCTCAAG
AAAATTTGTGGTTTGGGAGGGACCAATTAAAAAGATCC**ATG**

Table S4: Gene identifiers for NF-YA proteins used to build the phylogenetic tree depicted in Figure 4 and Figure S7.

Name	Gene Identifier
AtNF-YA1	AT5G12840
AtNF-YA2	AT3G05690
AtNF-YA3	AT1G72830
AtNF-YA4	AT2G34720
AtNF-YA5	AT1G54160
AtNF-YA6	AT3G14020
AtNF-YA7	AT1G30500
AtNF-YA8	AT1G17590
AtNF-YA9	AT3G20910
AtNF-YA10	AT5G06510
Fve_mrna07522	mrna07522.1-v1.0-hybrid
Fve_mrna12709	mrna12709.1-v1.0-hybrid
Fve_mrna14681	mrna14681.1-v1.0-hybrid
Fve_mrna18295	mrna18295.1-v1.0-hybrid
Fve_mrna26819	mrna26819.1-v1.0-hybrid
Fve_mrna28941	mrna28941.1-v1.0-hybrid
Gma_02G195000	Glyma.02G195000
Gma_02G303800	Glyma.02G303800
Gma_03G203000	Glyma.03G203000
Gma_05G166100	Glyma.05G166100
Gma_07G036200	Glyma.07G036200
Gma_08G124200	Glyma.08G124200
Gma_08G335900	Glyma.08G335900
Gma_09G023800	Glyma.09G023800
Gma_09G068400	Glyma.09G068400
Gma_10G082800	Glyma.10G082800
Gma_12G236800	Glyma.12G236800
Gma_13G107900	Glyma.13G107900
Gma_13G202300	Glyma.13G202300
Gma_14G010000	Glyma.14G010000
Gma_15G027400	Glyma.15G027400
Gma_15G129900	Glyma.15G129900
Gma_15G173300	Glyma.15G173300
Gma_16G005500	Glyma.16G005500
Gma_17G051400	Glyma.17G051400
Gma_18G071000	Glyma.18G071000
Gma_19G200800	Glyma.19G200800
Lj_FS318732	FS318732.1
LjNF-YA1	Lj5g3v0841080
LjNF-YA2	Lj6g3v0647470
LjNF-YA3	Lj4g3v2179250

Table S4 Continued

Name	Gene Identifier
LjNF-YA4	Lj1g3v4752710
LjNF-YA5	Lj3g3v2657800
LjNF-YA6	Lj3g3v0338970
LjNF-YA7	Lj2g3v3336090
LjNF-YA8	Lj0g3v0252369
MnoNF-YA1	XP_010102352.1
MnoNF-YA3	XP_010087689.1
MnoNF-YA7	XP_010098569.1
MnoNF-YA8	XP_010090113.1
MnoNF-YA9a	XP_010105454.1
MnoNF-YA10	XP_010106984.1
MnoNF-YA9b	XP_010088228.1
MtNF-YA1	Medtr1g056530
MtNF-YA2	Medtr7g106450
MtNF-YA3	Medtr2g041090
MtNF-YA4	Medtr2g099490
MtNF-YA5	Medtr3g061510
MtNF-YA6	Medtr2g030170
MtNF-YA7	Medtr8g037270
MtNF-YA8	Medtr8g019540
PanNF-YA1	PanWU01x14_284830
PanNF-YA2	PanWU01x14_161830
PanNF-YA3	PanWU01x14_246880
PanNF-YA4	PanWU01x14_050420
PanNF-YA5	PanWU01x14_192760
PanNF-YA6	PanWU01x14_192330
PanNF-YA7	PanWU01x14_231390
Ppe_2G014500	Prupe.2G014500
Ppe_3G043900	Prupe.3G043900
Ppe_4G090200	Prupe.4G090200
Ppe_4G249900	Prupe.4G249900
Ppe_7G093100	Prupe.7G093100
Ppe_7G203500	Prupe.7G203500
PvNF-YA1	Phvul.001G196800
PvNF-YA2	Phvul.002G246600
PvNF-YA3	Phvul.005G156100
PvNF-YA4	Phvul.003G133100
PvNF-YA5	Phvul.008G283100
PvNF-YA6	Phvul.011G211300
PvNF-YA7	Phvul.006G062200
PvNF-YA8	Phvul.010G133300
PvNF-YA9	Phvul.007G267100

Table S4 Continued

Name	Gene Identifier
TorNF-YA1	TorRG33x02_341480
TorNF-YA2	TorRG33x02_150260
TorNF-YA3	TorRG33x02_081410
TorNF-YA4	TorRG33x02_339730
TorNF-YA5	TorRG33x02_031550
TorNF-YA6	TorRG33x02_321500
TorNF-YA7	TorRG33x02_125390
AduNF-YA1	XP_015951283
AduNF-YA2	XP_015946278
AduNF-YA3	XP_015954718.1
AduNF-YA4	Aradu.67X2R
AduNF-YA5	XP_015972671.1
AduNF-YA6	XP_015956129.1
AduNF-YA7	XP_015937254.1
AduNF-YA8	XP_015935517.1
AduNF-YA9	XP_015959290.1
Dgl3616S29258	Dgl3616S29258
Dgl216S24617	Dgl216S24617
Dgl81S35216	Dgl81S35216
Dgl959S36562	Dgl959S36562
Dgl402S29919	Dgl402S29919
Dgl544S31934	Dgl544S31934
Dgl749S12860	Dgl749S12860
Cgl33S15932	Cgl33S15932
Cgl26S24714	Cgl26S24714
Cgl31S04874	Cgl31S04874
Cgl428S01957	Cgl428S01957
Cgl231S11946	Cgl231S11946
Cgl8S22241	Cgl8S22241
Cgl24S00922	Cgl24S00922

Chapter 5

Rhizobium NodS-mediated N-methylation of lipo-chitooligosaccharide signal molecules is essential for functional nodule formation on *Parasponia andersonii*

Fengjiao Bu, Elena Fedorova, Arjan van Zeijl, Rene Geurts

Laboratory of Molecular Biology, Department of Plant Science, Wageningen University, 6708PB Wageningen, The Netherlands.

Abstract

Within the nitrogen-fixing clade, *Parasponia* species are the only non-legume plants that can establish nitrogen-fixing nodule symbiosis with diazotrophic rhizobia. Like legumes, *Parasponia* nodulation is also dependent on lipo-chitooligosaccharide (LCO)-based signalling. Whereas most legumes -including *Medicago truncatula* and *Lotus japonicus* - associate with specific rhizobium species, *Parasponia* has been reported to be promiscuous, as it can interact with a wide range of rhizobial species. However, some LCO-producing rhizobia cannot nodulate *Parasponia*. The molecular basis of this variation remains elusive.We studied the symbiotic interaction between *Parasponia andersonii* and a diverse range of rhizobial species of which the structure of the lipo-chitooligosaccharide (LCO) signal molecules have been elucidated. It is noticed that nodule formation and intracellular infection in *P. andersonii* correlates with the presence of N-methylation conferred by the bacterial *nodS* gene, which encodes an N-methyl transferase that methylates non-reducing terminal residue of LCOs. The importance of *nodS* is shown by demonstrating that LCO signalling induced by *Rhizobium tropici* CIAT899 is *nodS* dependent. We conclude that the N-methyl decoration of the non-reducing terminal residue of LCOs is essential for establishing successful nitrogen-fixing nodule symbiosis between rhizobium and *P. andersonii.*

Introduction

The rhizobium root nodule symbiosis is a trait that is dominantly present in the Fabaceae (order Fabales), comprising more than 20,000 species, but also occurs in five tropical tree species of the genus *Parasponia* (Cannabaceae, Rosales). This endosymbiosis involves the formation of a genuine organ: the root nodule. Root nodules are induced by rhizobia when plants grow under low nitrogen conditions. Individual nodule cells can host hundreds of rhizobium bacteria that find proper environmental conditions to convert atmospheric di-nitrogen (N_2) into ammonia. This enzymatic reaction is fueled by carbohydrates from the host plant, in exchange for fixed nitrogen.

In most cases, the formation of root nodules is triggered by rhizobium secreted signal molecules called nodulation (Nod) factors. Nod factors were first characterized from *Sinorhizobium meliloti* RCR2011 (Lerouge *et al.*, 1990). The major Nod factor of *S. meliloti* RCR2011 consists of a tetrameric chitin chain (four β-1,4-linked N-acetyl-D-glucosamine (GlcNAc)) with an N-acyl group and acetyl group at the non-reducing terminal GlcNAc residue, and a sulphate group at the reducing GlcNac residue. Thus Nod factors are lipo-chitooligosaccharides (LCOs) (Dénarié *et al.*, 1996). Characterization of LCOs of other rhizobium species revealed variation in the number of GlcNAc residues, the length and ratio/location of saturation of the fatty acyl group, and type and position of substitutions including acetyl, arabinosyl, carbamoyl, fucosyl, glycerol, mannosyl, sulfate and N-methylation (D'Haeze & Holsters, 2002).

LCOs determine the rhizobial host range at two different levels. First, specific flavonoids and isoflavonoids secreted by plants under nitrogen starvation can activate rhizobial NodD, NolR and/or NrcR proteins, which are transcriptional regulators of LCO biosynthesis genes (del Cerro *et al.*, 2015; Del Cerro *et al.*, 2016; Peters *et al.*, 1986; Redmond *et al.*, 1986; Mulligan & Long, 1989). NodD-type transcription factors of different rhizobia display variation in specificity towards (iso)flavonoids, whereas variation also occurs in secreted (iso)flavonoids between plants (Zaat *et al.*, 1989; Peck *et al.*, 2006). Partly this determines whether or not a rhizobium will engage with a potential host plant. The second level of host range regulation is determined by the structural variation of the produced LCOs.

The structural complexity of LCOs produced by a certain rhizobium species or strain is determined by its repertoire of nodulation (*nod, noe* and *nol*) genes. The *NodABC* gene operon is known to be

responsible for the biosynthesis of the chitin backbone. *NodC* encodes N-acetylglucosaminyltransferase and catalyzes elongation of chitin backbone by adding β-1,4 glycosidic linked GlcNAc residues. Chitin N-deacetylase (encoded by *nodB*) specifically removes the acetyl group located at the N-atom of the non-reducing terminal GlcNAc so that the *nodA* encoded N-acyltransferase can transfer an acyl chain to it from an acyl carrier protein. The *nodA, nodB* and *nodC* genes are likely present in all LCOs producing rhizobia (Dénarié *et al.*, 1992; Roche *et al.*, 1996). Other nodulation genes can encode enzymes that confer different substitutions on the chitin backbone at both the non-reducing and reducing terminal N-acetyl-D-glucosamine (GlcNAc) residues. For instance, *nodS,* which encodes an N-methyltransferase, is responsible for attaching a methyl group to the same N atom at the non-reducing N-acetylglucosamine where the acyl chain is attached (Geelen *et al.*, 1993; Jabbouri *et al.*, 1995). The *nodS* gene is present in the genome of some rhizobia, for instance, *Rhizobium tropici* CIAT899, while other rhizobia, e.g. *S. meliloti* RCR2011, do not contain this gene. *R. tropici* CIAT899 can nodulate both *Phaseolus vulgaris* and *Leucaena Leucocephala,* which requires LCOs with *nodS* mediated N-methylation substitution (Krishnan *et al.*, 1992) (Waelkens *et al.*, 1995). Similarly, NodE and NodF controlled modifications of the acyl chain and/or the NodL-controlled O-acetyl addition to *S. meliloti* LCOs are essential for successful infection of *Medicago* species *(Ardourel et al., 1994)*, whereas *Mesorhizobium loti* R7A requires NodL-NodZ controlled acetyl fucosylation of LCOs to infect *Lotus* species (Rodpothong *et al.*, 2009). Taken together, this shows that structural variation in LCOs determines -at least in part- the host range of rhizobia.

The five *Parasponia* species -*P. andersonii, P. melastomatifolia, P. parviflora, P. rigida* and *P. rugosa*- are the only non-legume plants that can establish nitrogen-fixing nodule symbiosis with rhizobium (Trinick, 1973; Akkermans *et al.*, 1978; Becking, 1992). Grown in native ecological niches, *Parasponia* species have been found to nodulate mainly with *Bradyrhizobium* sp. (**Table 1**), though lab experiments found that *Parasponia* sp. are generally promiscuous towards rhizobia (**Table 2**) (Trinick, 1980; Marvel *et al.*, 1987; Trinick & Hadobas, 1989a, 1990a; Webster *et al.*, 1995; Op den Camp *et al.*, 2012). However, not all rhizobium species can establish a successful symbiosis with *Parasponia* spp. Especially fast-growing rhizobia like *Rhizobium phaseoli*, *Rhizobium leguminosarum* and *S. meliloti* showed to be hampered in the formation of nitrogen-fixing nodules (Trinick & Galbraith, 1980; Trinick *et al.*, 1989). The molecular basis of this variation in susceptibility to rhizobium of *Parasponia* ssp. remains elusive.

Comparative phylogenomic studies demonstrated that the nodulation trait in legumes and *Parasponia* share a single evolutionary origin, about 110 million years ago (Op den Camp *et al.*, 2011; Griesmann *et al.*, 2018; van Zeijl *et al.*, 2018; van Velzen *et al.*, 2018). It was shown that *P. andersonii* deploys the same LCO signalling cascade to control the formation of nitrogen-fixing nodules, as has been identified in legumes (Op den Camp *et al.*, 2011; Griesmann *et al.*, 2018; van Zeijl *et al.*, 2018; van Velzen *et al.*, 2018; Bu et al., 2020 in press). This suggests that like in legumes, also *Parasponia* spp. may recognize a limited number of LCO variants that are required for establishing nitrogen-fixing symbiosis. To obtain insights in these requirements we quantified the nodulation efficiency of a range of rhizobium strains of which the structure of the main LCOs has been determined. This revealed distinct nodulation phenotypes of rhizobium species that belong to the same genera. Comparing the structural differences of the produced LCOs, we found the nodulation and infection ability on *P. andersonii* correlates with the presence of NodS mediated N-methylation. Therefore, we conclude that N-methylation of the non-reducing N-acetyl-D-glucosamine (GlcNAc) of rhizobium LCOs is essential for nodulation of *P. andersonii.*

Results

Rhizobium species of the same genus show differences in symbiotic compatibility on *Parasponia andersonii*

To obtain insights in the LCO specificity of *P. andersonii* we conducted nodulation experiments with a diverse range of rhizobium strains for which the main LCO structure is elucidated. This revealed three different levels of interaction, namely (i) no initiation of nodule formation; (ii) formation of nodule-like structures without intracellular infection, and (iii) nitrogen-fixing nodules (**Fig. 1, Table 3**). In the latter group of compatible strains, *Bradyrhizobium elkanii* USDA61 and *R. tropici* CIAT899 induced nodules that had a typical cytoarchitecture with a central vascular bundle and infected cells in the periphery (6 wpi) (**Fig. 2A,D**). It should be noted that *R. tropici* CIAT899-infected nodules contained only a few cell layers containing fixation threads, whereas multiple layers of infection cells were dead and fully colonized by bacteria (**Fig. 2A**). A similar phenotype has been reported for *P. andersonii* nodules infected by *Rhizobium tropici* WUR1 (Op den Camp *et al.*, 2012). *P. andersonii* plants inoculated with *S. meliloti* RCR2011 or *R. leguminosarum* bv. *trifolii* ANU843 did not harbour mature nodules, but close inspection revealed the presence of small bumps (**Fig. 1**). No such bumps were found upon inoculation with *B. diazoefficiens* USDA110 (**Fig. 1**). Sectioning showed that these bumps induced by *S. meliloti* RCR2011 and *R. leguminosarum* bv. *trifolii* ANU843 represent nodule-like structures that were devoid of intracellular infecting rhizobia (**Fig. 2B-C, E-F**). This suggests that *S. meliloti* RCR2011 and *R. leguminosarum* bv. *trifolii* ANU843 can trigger nodule initiation, but that further nodule development and intracellular infection is impaired.

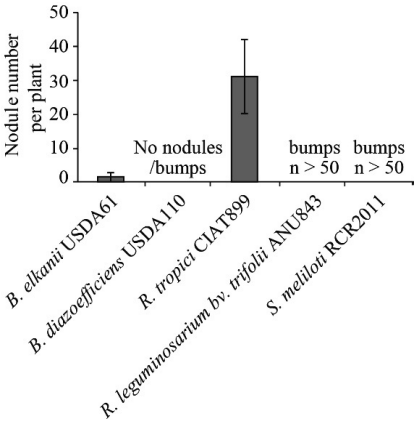


Fig. 1 *Parasponia andersonii* displays rhizobium host strain specificity. Shown are the nodulation efficiencies upon inoculation with *B. elkanii* USDA61, *B. diazoefficiens* USDA110, *R. tropici* CIAT899, *R. leguminosarum* bv. *trifolii* ANU843 and *S. meliloti* RCR2011. Rooted tissue culture plantlets of *P. andersonii* wild type were inoculated and scored for nodulation after 6 weeks. Note that only number of mature nodules are shown in this figure.

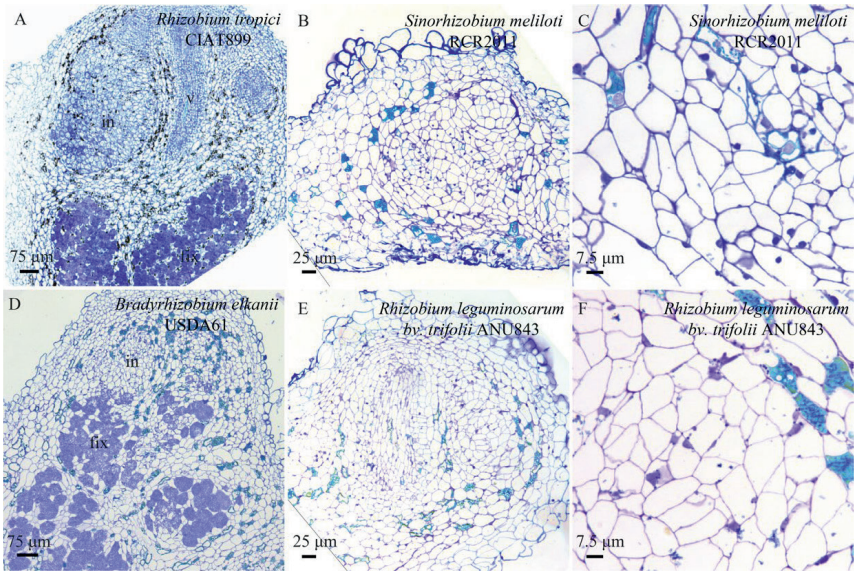


Fig. 2 Cytoarchitecture of nodules formed on *P. andersonii* roots induced by different rhizobium strains. Nodules induced by *R.tropici* CIAT899 (**A**) and *B.elkanii* USDA61 (**D**) display a normal cytoarchitecture including infected cells. Nodule-like structures induced by *S. meliloti* RCR2011 (**B**) and *R. leguminosarum bv. trifolii* ANU843 (**E**) do not contain infected cells (**C**, **F**).

NodS-mediated N-methylation of LCOs is essential for the formation of functional nodules on *P. andersonii*

R. tropici CIAT899 and *R. leguminosarum* bv. *trifolii* ANU843 are two species from the same genus, but both strains display a distinct symbiotic interaction on *P. andersonii.* Similarly, this is the case for *S. fredii* NGR234 (Op den Camp *et al.*, 2012) and *S. meliloti* RCR2011, and *B. elkanii* USDA61 and *B. diazoefficiens* USDA110. To investigate whether this difference in nodulation capacity is related to the difference in LCO repertoire produced by these strains, we compared the structure of LCOs of compatible and incompatible strains based on literature review (**Table 3**). We noticed that nodulation compatibility correlates with the occurrence of a NodS-controlled N-methylation of the non-reducible terminal GlcNAc. To test whether the presence of this methyl group and formation of functional nodules on *P. andersonii* are causally linked, we determined the nodulation ability of an *R. tropici* CIAT899 *nodS* mutant (Waelkens *et al.*, 1995; Geelen *et al.*, 1993; Jabbouri *et al.*, 1995). Whereas wild-type *R. tropici* CIAT899 efficiently induces nodules on *P. andersonii*, the *nodS* mutant is unable to induce nodule formation (**Fig. 1, Fig. 3A**). This

demonstrates a requirement for NodS-mediated N-methylation of *R. tropici* CIAT899 LCOs for the formation of functional nodules on *P. andersonii* roots.

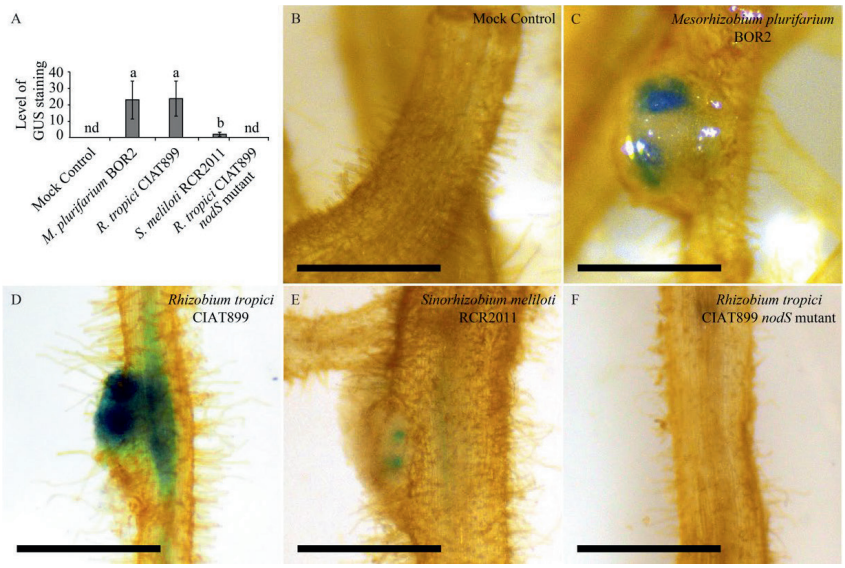


Fig. 3 *PanNFYA1*$_{pro}$*::GUS* expression induced by different rhizobium strains. (**A**) Quantification of 5-h stained GUS signals in non-nodulated control (Mock Control), and inoculated with *M. plurifarium* BOR2, *R. tropici* CIAT899, *S. meliloti* RCR2011 and *R. tropici* CIAT899 *nodS* mutant. nd stands for not detected. Data represent average GUS signal per plant (n > 5) ± SD. GUS signals were scored at 7 dpi. (**B**) Representative roots of Mock Control after Gus staining at 7 dpi. Note no GUS signal was observed. (**C**) A representative image of GUS signal on roots inoculated with *M. plurifarium* BOR2 at 7 dpi. (**D**) A representative image of GUS signal on roots inoculated with *R. tropici* CIAT899 at 7 dpi. (**E**) A representative image of GUS signal on roots inoculated with *S. meliloti* RCR2011 at 7 dpi. (**F**) A representative image of roots inoculated with *R. tropici nodS* mutant at 7 dpi. Note no GUS signal was observed.

It has been reported that *S. fredii USDA257* can nodulate *P. andersonii*, however, based on LCO analysis, it doesn't produce N-methylated LCOs due to absence of a functional *nodS* gene (Bec-Ferté et al., 1994; Pueppke & Broughton, 1999; Schuldes et al., 2012). As this would be an outlier in the strict correlation between nitrogen-fixing nodulation potential on *P. andersonii* and the presence of N-methyl group in the non-reducing terminal of LCOs, we re-investigated nodulation potential of this strain on *P. andersonii.* This reveals that *S. fredii USDA257* does not form functional nodules on *P. andersonii* at 8wpi, but only non-infected nodule-like structures (**Fig. 4**). Taken together, these results show a strict correlation between the capability of formation of

functional nodules on *P. andersonii* that harbor infected cells and the presence of an N-methyl group at the non-reducing terminal GlcNac of LCOs.

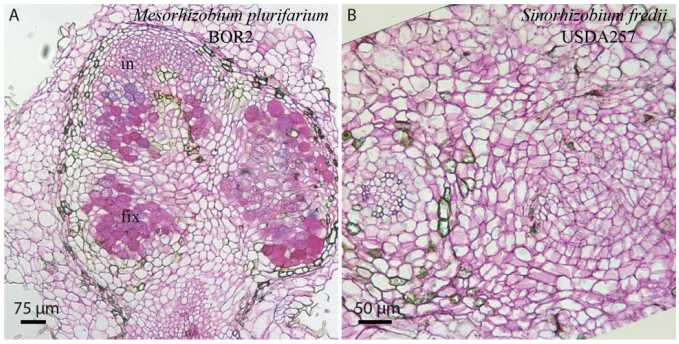


Fig. 4 Section of a *P. andersonii* nodule induced by *M. plurifarium* BOR2 (A) and *S. fredii* USDA257 (B). The nodule induced by *M. plurifarium* BOR2 contains infected cells, which are stained by rethidium red in dark red. Notice the lack of intracellular infection in nodules induced by *S. fredii* USDA257. Sectioned are nodules formed on rooted tissue culture plantlets of *P. andersonii* at 8 weeks post-inoculation.

Rhizobium species that lack NodS are hampered in symbiotic *PanNF-YA1* expression

Recently we demonstrated that the *P. andersonii NUCLEAR FACTOR YA1* (*PanNF-YA1*) is essential for rhizobium intracellular infection (**Chapter 4;** Bu et al., 2020 in press). Expression of this gene as shown with *PanNF-YA1*$_{pro}$:GUS reporter construct is associated with rhizobium infection. We asked the question to what extent incompatible - *nodS* lacking - rhizobium strains are able to induce early symbiotic event in *P. andersonii.* To do so, we used the *P. andersonii PanNF-YA1*$_{pro}$:GUS reporter line and studied GUS activity 7 days post-inoculation. The compatible strain *M. plurifarium* BOR2 induces *PanNF-YA1*$_{pro}$:GUS activity in young nodule primordia (**Fig. 3A, C**). Similarly, GUS expression in nodule primordia was induced by *R. topicii* CIAT899 (**Fig. 3A, D**). In contrast, no such *PanNFYA1*$_{pro}$*:GUS* induction was found upon inoculation with the *R. tropici* CIAT899 *nodS* mutant (**Fig. 3A, B, F**). This shows a correlation of the induction level of *PanNFYA1*$_{pro}$*:GUS* and LCOs with *nodS* mediated N-methyl decoration. We then tested the effect upon inoculation with rhizobium strain *S. meliloti* RCR2011 that form nodule-like structures on *P. andersonii* roots. In these roots, occasionally low GUS activity was observed. The intensity of the GUS signal was significantly less when compared to the signals induced by the compatible strains *M. plurifarium* BOR2 or *R. tropici* CIAT899 (**Fig. 3A, E**). This indicates that *PanNFYA1*$_{pro}$*:GUS* induction by rhizobium is a proxy for nodule formation and infection.

Discussion

Here we showed that N-methylated LCOs are essential for the formation of nitrogen-fixing nodules on *P. andersonii*. *R. tropici* CIAT899 that forms nitrogen-fixing nodules is unable to do so when the *nodS* gene required for LCO N-methylation is non-functional. This indicates that although *P. andersonii* is a promiscuous host, it deploys host specificity by discriminating N-methylated LCOs producing rhizobium from non-N-methylated LCOs producing rhizobium. Uninfected nodule-like structures are induced by rhizobium lacking N-methylated LCOs, suggesting that *P. andersonii* also deploys a higher stringent level of LCO signalling to form intracellular infection than that for nodule formation, similar as reported for legumes (Ardourel *et al.*, 1994).

Although the NodS-mediated N-methyl decoration is not present in LCOs produced by *R. tropici* CIAT899 *nodS* mutant and *S. meliloti* RCR2011, they show a phenotypic difference when used as inoculum on *P. andersonii. S. meliloti* RCR2011 still can trigger weak expression of *PanNF-YA1*$_{pro}$:GUS, this is not the case for the *R. tropici* CIAT899 *nodS* mutant. The reason for this difference could be that other LCOs produced by *S. meliloti* RCR2011 might in part compensate for the N-methyl decoration. One such decoration could be the acetyl group at the non-reducing GlcNac of LCOs of *S. meliloti* RCR2011, which does not exist in *R. tropici*. This acetylation is controlled by the *nodL* encoded N-actyl transferase.

The NodS-controlled N-methylation of LCOs may affect the 3D structure of the Nod factor. A 3D structure model of pentameric Nod factor produced by *Azotobacter caulinodans* ORS571 suggests an important effect of the N-methyl group on the 3D-positioning of the acyl chain. In this model, removing the N-methyl group causes re-orientation of the lipid chain, from almost perpendicularly to almost in parallel with the chitooligosaccharide backbone (D'Haeze & Holsters, 2002). LCO receptor complexes are membrane proteins that reside in lipid raffles, might only have a certain level of flexibility (Moling *et al.*, 2014). Thus changes in the 3D structure could reduce the affinity of LCOs to LCO receptor complexes. To test this, our future work will focus on testing whether the introduction of a functional *nodS* in rhizobium strains that lack *nodS* -like *Sinorhizobium meliloti* RCR2011- will allow them to nodulate *P. andersonii*. Furthermore, experiments will be needed to determine whether the functioning of NodS and NodL are interfering with each other (López-Lara *et al.*, 2001). Thus, *P. andersonii* is an excellent system to conduct those studies

aiming at understanding how certain LCO decorations affect LCO signalling induction and nitrogen-fixing root nodule development.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

All experiments were done using *P. andersonii* WU1.14(van Velzen *et al.*, 2018; Wardhani *et al.*, 2019). Plants were maintained as described previously(van Zeijl *et al.*, 2018; Wardhani *et al.*, 2019). Young plantlets for nodulation assays were vegetatively propagated *in vitro* and rooted (van Zeijl *et al.*, 2018; van Velzen *et al.*, 2018; Wardhani *et al.*, 2019).

Nodulation assay, GUS assay

Nodulation assay was carried out according to the previous report unless stated otherwise (van Velzen *et al.*, 2018; Wardhani *et al.*, 2019). Nodualtion was checked after 4 to 6 weeks post-inoculation. Nodulation efficiencies were estimated by determining the average nodule number per plant. In the case of *S. meliloti* RCR2011 and *R. leguminosarum* bv. *trifolii* ANU843, the nodule like structures were scored as bumps and estimation of the total event were given. For GUS assay, growth systems were sterilized at 120 °C for 21 mins before use to exclude the potential contamination. Plants were transferred to the growth system and grown for 2 weeks before inoculation. The transfer of plants into growth system as well as rhizobium inoculation were carried out in laminar-flow. GUS signal was scored at 7 dpi (OD600 = 0.03). Number of GUS signal were quantified after stain in GUS staining buffer (3% [w/v] sucrose, 10 mM EDTA, 2 mM k-ferrocyanide, 2 mM k-ferricyanide, and 0.5 mg/mL 5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid, cyclohexylammonium salt [X-Gluc] in 0.1 M phosphate buffer [pH = 7.2]) at 37°C for 5 hours. For re-investigation of *S. fredii* USDA257, the same growth conditions were used as for GUS assay. Nodules were checked after 8 weeks post-inoculation at an OD_{600} = 0.03.

Histochemical Analysis, Microtome Sectioning and Microscopy

The same *PanNF-YA$_{pro}$:GUS* transgenic plant (line 1E5) was used as described in Chapter 4. Nodule samples formed on roots of the *PanNF-YA$_{pro}$:GUS* transgenic plant (line 1E5) were incubated in GUS buffer. Then nodules were fixed in 4% paraformaldehyde (w/v), 5% glutaraldehyde (v/v) in 50 mM phosphate buffer (pH = 7.2) at 4°C for 24 hours for plastic sectioning. Subsequently, the samples were dehydrated using an ethanol series and embedded in Technovit 7100 (Heraeus Kulzer, Germany) according to the manufacturer's instructions. Semi-

thin sections were cut using a Leica Ultracut microtome (Leica Microsystems, Germany) to 4 µm thickness (7 µm thickness in the case for GUS stained samples). Sections were stained with 0.05% Toluidine Blue or 0.1% Ruthenium Red. Images were photographed using a Leica DM5500B microscope equipped with a DFC425C camera (Leica Microsystems, Germany).

References

Akkermans ADL, Abdulkadir S, Trinick MJ. **1978**. N2-fixing root nodules on Ulmaceae: Parasponia or (and) Trema spp.? *Plant and soil* **49**: 711–715.

Ardourel M, Demont N, Debellb b. F, Maillet F, de Billy F, Promb J-C, D6narié b. J, Truchet G. **1994**. Rhízobíum meliloti Lipooligosaccharide Nodulation Factors: Different Structural Requirements for Bacterial Entry into Target Root Hair Cells and lnduction of Plant Symbiotic Developmental Responses. *The Plant cell* **6**: 1357–1374.

Appleby CA, Tjepkema JD, Trinick MJ. **1983**. Hemoglobin in a nonleguminous plant, parasponia: possible genetic origin and function in nitrogen fixation. *Science* **220**: 951–953.

Bec-Ferté MP, Krishnan HB, Promé D, Savagnac A, Pueppke SG, Promé JC. **1994**. Structures of nodulation factors from the nitrogen-fixing soybean symbiont *Rhizobium fredii* USDA257. *Biochemistry* **33**: 11782–11788.

Becking JH. **1992**. The Rhizobium symbiosis of the nonlegume Parasponia. In: Stacey G, Burris RH, Evans HJ, eds. Biological nitrogen fixation. New York: Routledge, Chapman and Hall, 497–559.

Becking JH. **1983**. The *Parasponia parviflora—Rhizobium* symbiosis. Host specificity, growth and nitrogen fixation under various conditions. *Plant and soil* **75**: 309–342.

Bender GL, Goydych W, Rolfe BG, Nayudu M. **1987**. The role of Rbizobium conserved and host specific nodulation genes in the infection of the non-legume Pnrasponia andersonii. *Molecular & general genetics: MGG* **210**: 299–306.

Cen Y, Bender GL, Trinick MJ. **1982**. Transposon mutagenesis in rhizobia which can nodulate both legumes and the nonlegume Parasponia. *Applied and Environmental Microbiology* **43**: 233–236.

Fengjiao Bu, Luuk Rutten, Yuda Purwana Roswanjaya, Olga Kulikova, Marta Rodriguez-Franco, Thomas Ott, Ton Bisseling, Arjan van Zeijl & Rene Geurts. 2020. Mutant analysis in the non-legume *Parasponia andersonii* identifies NIN and NF-YA1 transcription factors as a core genetic network in nitrogen-fixing nodule symbioses. *New Phytologist* (In press).

del Cerro P, Rolla-Santos AAP, Gomes DF, Marks BB, Pérez-Montaño F, Rodríguez-Carvajal MÁ, Nakatani AS, Gil-Serrano A, Megías M, Ollero FJ, *et al.* 2015. Regulatory nodD1 and nodD2 genes of *Rhizobium tropici* strain CIAT 899 and their roles in the early stages of molecular signaling and host-legume nodulation. *BMC genomics* **16**: 251.

Del Cerro P, Rolla-Santos AAP, Valderrama-Fernández R, Gil-Serrano A, Bellogín RA, Gomes DF, Pérez-Montaño F, Megías M, Hungría M, Ollero FJ. **2016**. NrcR, a New Transcriptional Regulator of *Rhizobium tropici* CIAT 899 Involved in the Legume Root-Nodule Symbiosis. *PloS one* **11**: e0154029.

Dénarié J, Debellé F, Promé J-C. **1996**. Rhizobium lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. *Annual review of biochemistry* **65**:

503–535.

Dénarié J, Debellé F, Rosenberg C. **1992**. Signaling and host range variation in nodulation. *Annual review of microbiology* **46**: 497–531.

D'Haeze W, Holsters M. **2002**. Nod factor structures, responses, and perception during initiation of nodule development. *Glycobiology* **12**: 79R–105R.

Geelen D, Mergaert P, Geremia RA, Goormachtig S, Van Montagu M, Holsters M. **1993**. Identification of nodSUIJ genes in Nod locus 1 of *Azorhizobium caulinodan*s: evidence that nodS encodes a methyltransferase involved in Nod factor modification. *Molecular microbiology* **9**: 145–154.

Geurts R, Heidstra R, Hadri A-E, Allan Downie J, Franssen H, van Kammen A, Bisseling T. **1997**. Sym2 of Pea 1s lnvolved in a Nodulation Factor-Perception Mechanism That Controls the lnfection Process in the Epidermis. *Plant physiology* **115**: 351–359.

Griesmann M, Chang Y, Liu X, Song Y, Haberer G, Crook MB, Billault-Penneteau B, Lauressergues D, Keller J, Imanishi L, *et al.* 2018. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* **1743**: eaat1743.

Jabbouri S, Fellay R, Talmont F, Kamalaprija P, Burger U, Relić B, Promé JC, Broughton WJ. **1995**. Involvement of nodS in N-methylation and nodU in 6-O-carbamoylation of *Rhizobium* sp. NGR234 nod factors. *The Journal of biological chemistry* **270**: 22968–22973.

Krishnan HB, Lewin A, Fellay R, Broughton WJ, Pueppke SG. **1992**. Differential expression of nodS accounts for the varied abilities of *Rhizobium fredii* USDA257 and *Rhizobium sp.* strain NGR234 to nodulate Leucaena spp. *Molecular microbiology* **6**: 3321–3330.

Lerouge P, Roche P, Faucher C, Maillet F, Truchet G, Promé JC, Dénarié J. **1990**. Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. *Nature* **344**: 781–784.

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

López-Lara IM, Kafetzopoulos D, Spaink HP, Thomas-Oates JE. **2001**. Rhizobial NodL O-acetyl transferase and nodS N-methyl transferase functionally interfere in production of modified nod factors. *Journal of bacteriology* **183**: 3408–3416.

Marvel DJ, Torrey JG, Ausubel FM. **1987**. Rhizobium symbiotic genes required for nodulation of legume and nonlegume hosts. *Proceedings of the National Academy of Sciences of the United States of America* **84**: 1319–1323.

Moling S, Pietraszewska-Bogiel A, Postma M, Fedorova E, Hink MA, Limpens E, Gadella TWJ, Bisseling T. **2014**. Nod factor receptors form heteromeric complexes and are essential for intracellular infection in medicago nodules. *The Plant cell* **26**: 4188–4199.

Mulligan JT, Long SR. **1989**. A family of activator genes regulates expression of *Rhizobium meliloti* nodulation genes. *Genetics* **122**: 7–18.

McDonagh J. **1992**. Nitrogen Fixation: Achievements and Objectives: Proceedings of the 8th International Congress on Nitrogen Fixation, eds P. M. Greshoff, L. E. Roth, G. Stacey & W. E. Newton, xv 869 pp. New York and London: Chapman and Hall (1990). £35.95 (hardback). ISBN 0 412 02591 4. *The Journal of Agricultural Science* **118**: 136–136.

Op den Camp RHM, Polone E, Fedorova E, Roelofsen W, Squartini A, Op den Camp HJM, Bisseling T, Geurts R. **2012**. Nonlegume *Parasponia andersonii* deploys a broad rhizobium host range strategy resulting in largely variable symbiotic effectiveness. *Molecular plant-microbe interactions: MPMI* **25**: 954–963.

Op den Camp R, Streng A, De Mita S, Cao Q, Polone E, Liu W, Ammiraju JSS, Kudrna D, Wing R, Untergasser A, *et al.* 2011. LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume *Parasponia*. *Science* **331**: 909–912.

Peck MC, Fisher RF, Long SR. **2006**. Diverse flavonoids stimulate NodD1 binding to nod gene promoters in *Sinorhizobium meliloti*. *Journal of bacteriology* **188**: 5417–5427.

Peters NK, Frost JW, Long SR. **1986**. A plant flavone, luteolin, induces expression of *Rhizobium meliloti* nodulation genes. *Science* **233**: 977–980.

Pueppke SG, Broughton WJ. **1999**. *Rhizobium sp.* strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges. *Molecular plant-microbe interactions: MPMI* **12**: 293–318.

Redmond JW, Batley M, Djordjevic MA, Innes RW, Kuempel PL, Rolfe BG. **1986**. Flavones induce expression of nodulation genes in Rhizobium. *Nature* **323**: 632–635.

Roche P, Maillet F, Plazanet C, Debellé F, Ferro M, Truchet G, Promé JC, Dénarié J. **1996**. The common nodABC genes of *Rhizobium meliloti* are host-range determinants. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 15305–15310.

Rodpothong P, Sullivan JT, Songsrirote K, Sumpton D, Cheung KWJ-T, Thomas-Oates J, Radutoiu S, Stougaard J, Ronson CW. **2009**. Nodulation gene mutants of *Mesorhizobium loti* R7A-nodZ and nolL mutants have host-specific phenotypes on Lotus spp. *Molecular plant-microbe interactions: MPMI* **22**: 1546–1554.

Schuldes J, Rodriguez Orbegoso M, Schmeisser C, Krishnan HB, Daniel R, Streit WR. **2012**. Complete genome sequence of the broad-host-range strain *Sinorhizobium fredii* USDA257. *Journal of bacteriology* **194**: 4483.

Smit P, Limpens E, Geurts R, Fedorova E, Dolgikh E, Gough C, Bisseling T. **2007**. Medicago LYK3, an entry receptor in rhizobial nodulation factor signaling. *Plant physiology* **145**: 183–191.

Scott KF, Rolfe BG, Shine J. **1983**. Nitrogenase structural genes are unlinked in the nonlegume symbiont Parasponia rhizobium. *DNA* **2**: 141–148.

Trinick MJ. **1973**. Symbiosis between Rhizobium and the non-legume, *Trema aspera*. *Nature* **244**: 459–460.

Trinick MJ, Galbraith J. **1976**. Structure of root nodules formed by Rhizobium on the non-

legume Trema cannabina var. scabra. *Archives of microbiology* **108**: 159–166.

Trinick MJ. **1979**. Structure of nitrogen-fixing nodules formed by Rhizobium on roots of Parasponia andersonii Planch. *Canadian journal of microbiology* **25**: 565–578.

Trinick MJ. **1980**. Growth of Parasponia in agar tube culture and symbiotic effectiveness of isolates from *Parasponia spp*. *The New phytologist* **85**: 37–45.

Trinick MJ, Galbraith J. **1980**. The rhizobium requirements of the non-legume parasponia in relationship to the cross-inoculation group concept of legumes. *The New phytologist* **86**: 17–26.

Trinick MJ, Goodchild DJ, Miller C. **1989**. Localization of Bacteria and Hemoglobin in Root Nodules of *Parasponia andersonii* Containing Both Bradyrhizobium Strains and *Rhizobium leguminosarum* biovar trifolii. *Applied and environmental microbiology* **55**: 2046–2055.

Trinick MJ, Hadobas PA. **1989a**. Competition by Bradyrhizobium Strains for Nodulation of the Nonlegume *Parasponia andersonii*. *Applied and environmental microbiology* **55**: 1242–1248.

Trinick MJ, Hadobas PA. **1989b**. Effectiveness and competition for nodulation of Vigna unguiculata and Macroptilium atropurpureum with Bradyrhizobium from Parasponia. *Canadian journal of microbiology* **35**: 1156–1163.

Trinick MJ, Hadobas PA. **1990a**. Nodulation of Trifofium repens with modified Bradyrhizobium and the nodulation of *Parasponia* with *Rhizobium leguminosarum* biovar trifolii. *Plant and soil* **125**: 49–61.

Trinick MJ, Hadobas PA. **1990b**. Symbiotic effectiveness of Bradyrhizobium strains isolated from Parasponia and tropical legumes on Parasponia host species. *Plant and soil* **124**: 117–126.

Tjepkema JD, Cartica RJ. **1982**. Diffusion Limitation of Oxygen Uptake and Nitrogenase Activity in the Root Nodules of Parasponia rigida Merr. and Perry. *Plant physiology* **69**: 728–733.

van Velzen R, Holmer R, Bu F, Rutten L, van Zeijl A, Liu W, Santuari L, Cao Q, Sharma T, Shen D, *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.

Waelkens F, Voets T, Vlassak K, Vanderleyden J, van Rhijn P. **1995**. The nodS gene of *Rhizobium tropici* strain CIAT899 is necessary for nodulation on Phaseolus vulgaris and on Leucaena leucocephala. *Molecular plant-microbe interactions : MPMI* **8**: 147–154.

Walker SA, Downie JA. **2000**. Entry of *Rhizobium leguminosarum* bv.viciae into root hairs requires minimal Nod factor specificity, but subsequent infection thread growth requires nodO or nodE. *Molecular plant-microbe interactions: MPMI* **13**: 754–762.

Wardhani TAK, Roswanjaya YP, Dupin S, Li H, Linders S, Hartog M, Geurts R, Van Zeijl A. **2019**. Transforming, genome editing and phenotyping the nitrogen-fixing tropical Cannabaceae tree *Parasponia andersonii*. *Journal of visualized experiments: JoVE*.

Webster G, Poulton PR, Cocking EC, Davey MR. **1995**. The nodulation of micro-propagated

plants of *Parasponia andersonii* by tropical legume rhizobia. *Journal of experimental botany* **46**: 1131–1137.

Wittenberg JB, Wittenberg BA, Gibson QH, Trinick MJ, Appleby CA. **1986**. The kinetics of the reactions of Parasponia andersonii hemoglobin with oxygen, carbon monoxide, and nitric oxide. *The Journal of biological chemistry* **261**: 13624–13631.

Zaat SA, Schripsema J, Wijffelman CA, van Brussel AA, Lugtenberg BJ. **1989**. Analysis of the major inducers of the Rhizobium nodA promoter from *Vicia sativa* root exudate and their activity with different nodD genes. *Plant molecular biology* **13**: 175–188.

van Zeijl A, Wardhani TAK, Seifi Kalhor M, Rutten L, Bu F, Hartog M, Linders S, Fedorova EE, Bisseling T, Kohlen W, *et al.* 2018. CRISPR/Cas9-Mediated Mutagenesis of Four Putative Symbiosis Genes of the Tropical Tree *Parasponia andersonii* Reveals Novel Phenotypes. *Frontiers in plant science* **9**: 284.

Table 1 Native *Bradyrhizobium* microsymbiont of *Parasponia* species and their collection sites.

strain	origin	host	reference
NGR231	Pangia, P.N.G.	*P. rugosa*	(*1–7*)
CP241	Lae-Bulolo, P.N.G.	*P. rugosa*	(*4*, *5*)
CP272	Panguna, Bougainville, P.N.G	*P. andersonii*	(*4*, *6*, *8*)
CP273	Panguna, Bougainville, P.N.G	*P. andersonii*	(*4*, *6–9*)
CP274	Panguna, Bougainville, P.N.G	*P. andersonii*	(*4*, *8*)
CP275	Panguna, Bougainville, P.N.G	*P. andersonii*	(*4*, *7*, *8*)
CP276	Panguna, Bougainville, P.N.G	*P. andersonii*	(*4*)
CP277	Panguna, Bougainville, P.N.G	*P. andersonii*	(*6–8*)
CP278	Panguna, Bougainville, P.N.G	*P. andersonii*	(*4*, *8*)
CP279	Panguna, Bougainville, P.N.G	*P. andersonii*	(*4*, *7*, *8*)
CP280	Panguna, Bougainville, P.N.G	*P. andersonii*	(*4*, *8*)
CP281	Panguna, Bougainville, P.N.G	*P. andersonii*	(*6*, *8*)
CP282	Panguna, Bougainville, P.N.G	*P. andersonii*	(*4*)
CP283/ANU298	Panguna, Bougainville, P.N.G	*P. andersonii*	(*4*, *6*, *7*, *10–13*)
CP284	Panguna, Bougainville, P.N.G	*P. andersonii*	(*4*, *6*)
CP285	Panguna, Bougainville, P.N.G	*P. andersonii*	(*8*)
CP286	Panguna, Bougainville, P.N.G	*P. andersonii*	(*8*)
CP287	Panguna, Bougainville, P.N.G	*P. andersonii*	(*8*)
CP288	Panguna, Bougainville, P.N.G	*P. andersonii*	(*7*, *8*)
CP289	Panguna, Bougainville, P.N.G	*P. andersonii*	(*7*, *8*)
CP290	Panguna, Bougainville, P.N.G	*P. andersonii*	(*7*, *8*)
CP291	Panguna, Bougainville, P.N.G	*P. andersonii*	(*7*, *8*)
CP292	Panguna, Bougainville, P.N.G	*P. andersonii*	(*7*, *8*)
CP294	Patep Village, P.N.G.	*P. rigida*	(*8*)
CP295	Patep Village, P.N.G.	*P. rigida*	(*8*)
CP296	Patep Village, P.N.G.	*P. rigida*	(*8*)
CP296	Patep Village, P.N.G.	*P. rigida*	(*8*)
CP297	Patep Village, P.N.G.	*P. rigida*	(*7*, *8*)
CP298	Patep Village, P.N.G.	*P. rigida*	(*7*, *8*)
CP299	Patep Village, P.N.G.	*P. rigida*	(*6–9*)
CP300	Patep Village, P.N.G.	*P. rigida*	(*8*)
CP301	Patep Village, P.N.G.	*P. rigida*	(*8*)
CP300	Patep Village, P.N.G.	*P. rigida*	(*8*)
CP301	Patep Village, P.N.G.	*P. rigida*	(*8*)
CP302	Patep Village, P.N.G.	*P. rigida*	(*8*)
CP303	Patep Village, P.N.G.	*P. rigida*	(*7*, *8*)
CP304	Patep Village, P.N.G.	*P. rigida*	(*8*)
CP305	Patep Village, P.N.G.	*P. rigida*	(*7*, *8*)
CP306	Patep Village, P.N.G.	*P. rigida*	(*8*)

Continued on Table 1 Native *Bradyrhizobium* microsymbiont of *Parasponia* species and their collection sites.

strain	origin	host	reference
CP307	Patep Village, P.N.G.	*P. rigida*	(*8*)
CP308	Patep Village, P.N.G.	*P. rigida*	(*8*)
CP309	Patep Village, P.N.G.	*P. rigida*	(*7*, *8*)
CP310	Patep Village, P.N.G.	*P. rigida*	(*8*)
CP311	Patep Village, P.N.G.	*P. rigida*	(*8*)
CP312	Patep Village, P.N.G.	*P. rigida*	(*7*, *8*)
CP313	Patep Village, P.N.G.	*P. rigida*	(*8*)
CP314	Patep Village, P.N.G.	*P. rigida*	(7, 8)
CP315	Patep Village, P.N.G.	*P. rigida*	(*8*)
CP316	Patep Village, P.N.G.	*P. rigida*	(*8*)
CP317	Patep Village, P.N.G.	*P. rigida*	(*6–8*)
RP501	Indonesia	*P. parviflora*	(*14–16*)
Pp1A	Western Java*, Indonesia	*P. parviflora*	(*17*)
Pp1B	Western Java*, Indonesia	*P. parviflora*	(*17*)
Pp2	Western Java*, Indonesia	*P. parviflora*	(*17*)
Pp4A	Western Java*, Indonesia	*P. parviflora*	(*17*)
Pp4B	Western Java*, Indonesia	*P. parviflora*	(*17*)
Pp6	Western Java*, Indonesia	*P. parviflora*	(*17*)
Pp7A	Western Java*, Indonesia	*P. parviflora*	(*17*)
Pp8	Western Java*, Indonesia	*P. parviflora*	(*17*)
Pp10	Western Java*, Indonesia	*P. parviflora*	(*17*)
Pp25	Western Java*, Indonesia	*P. parviflora*	(*17*)
Pp226	Western Java*, Indonesia	*P. parviflora*	(*17*)
Pp31	Western Java*, Indonesia	*P. parviflora*	(*17*)
Pp50	Western Java*, Indonesia	*P. parviflora*	(*17*)
Pp51	Western Java*, Indonesia	*P. parviflora*	(*17*)
Pp52	Western Java*, Indonesia	*P. parviflora*	(*17*)
Pp53	Western Java*, Indonesia	*P. parviflora*	(*17*)
Pp54	Western Java*, Indonesia	*P. parviflora*	(*17*)

*Strains have been samples on Mt Pangrango and Mt alak, Western Java, Inodonesia (*17*).

P.N.G.: Papua New-Guinea.

1, (Trinick 1973)*; 2,* (Trinick and Galbraith 1976)*; 3,* (Trinick 1979); *4,* (Trinick 1980); *5,* (Trinick and Galbraith 1980); *6,* (Trinick and Hadobas 1989); *7,* (Trinick and Hadobas 1990b); *8,* (Trinick and Hadobas 1989); *9,* (Trinick et al. 1989); *10,*(Cen et al. 1982); *11,* (Scott et al. 1983); *12,* (Appleby et al. 1983);13*,* (Wittenberg et al. 1986); *14,* (Tjepkema and Cartica 1982); *15,*(Marvel et al. 1987); 1*6,* (McDonagh 1992); *17,* (Becking 1983).

Table 2 Nodulation of *Parasponia* sp. with non-native rhizobia.

Genera	Species	Strain	Origin	Host plant	Nodulation on *Parasponia*	References
Aeschnomene	*sp.*	ORS302	West Africa	*Aeschynomene pfundii*	Nod+/Fix+	(1)
Rhizobium	*leguminosarum bv trifolii*	CP/WU strains	Australia	*Trifolium*	**Nod+/Fix$^-$**	(2)
	leguminosarum biovar trifolii	NGR66	Papua New Guinea	Trifolium	**Nod+/IF$^-$**	(3)
	tropici	CIAT899	Columbia	*Phaseolus vulgaris*	Nod+/Fix+	**Chapter 3**
	tropici	WUR1	-	*potting soil*	Nod+/Fix+	(4)
	sullae	IS123T	Southern Spain	*Hedysarum coronarium*	Nod+/Fix+**?**	(4)
Mesorhizobium	*plurifarium*	WUR2	potting soil	*Parasponia andersonii*	Nod+/Fix+	(4)
	plurifarium	BOR2	Saba, Malaysia	*Trema orientalis rhizosphere*	Nod+/Fix+	(5)
	loti	R7A	New Zealand	*Lotus sp.*	Nod+/Fix+	**Chapter 3**
Bradyrhizobium	*elkanii*	WUR3	potting soil	*Chamaecrista fasiculata*	Nod+/Fix+	(4)
	sp.	ORS302	West Africa	*Aeschynomene pfundii*	Nod+/Fix+	(1)
	elkanii	USDA61	USA	*Glycine max*	Nod+/IF+	**Chapter 3**
	sp.	ORS3257	unknown	Aeschynomene indica	Nod+/IF+	**Chapter 3**
Sinorhizobium	*fredii*	NGR234	P.N.G.	*Lablab purpureus*	Nod+/Fix+	(4,6)
	fredii	USDA257	P.N.G.	*Glycine soja*	**Nod+/?**	(6)
Cupriviadus	*tainwanensis*	LMG19424	Taiwan, China	Momosa. pudica/diplotricha	Nod+/IF+	**Chapter 3**

1,(Webster et al. 1995); *2,*(Trinick and Hadobas 1990a); *3,*(Trinick et al. 1989); *4,* (Op den Camp et al. 2012); *5,*(van Velzen et al. 2018); *6,*(Bender et al. 1987); *7,*(Pueppke and Broughton 1999);

Table 3 Comparison of Nod factor structures produced by rhizobium strains showing different nodulation phenotype on parasponia. Summarization is based on (D'Haeze and Holsters 2002).

	Sinorhizobium sp. NGR234*	*B. elkanii* USDA61	*R. tropici* CIAT899	*S.meliloti* RCR2011	*B. diazoefficiens* USDA110	*R. leguminosarum bv. trifolii* ANU843
Nodulation phenotype	Nod^+/Fix^+	Nod^+/Fix^+	Nod^+/Fix^+	Nod^+/Fix^-	Nod^-	Nod^+/Fix^-
n	5	4,5	4,5	4,5	5	4,5
R1-lipid	C16:0,C16:1,C18:0,C18:1,C18;2	C18:1, C16:0	C16:0,C16:1,C18:0, C18:1,C20;0,C20:1	C16:0,C16:1,C16:3,C18-C2(w-1)-OH	C18:1	C16:0,C16:1,C18:0,C18:1,3OH-C16:0,3OH-C14:0,C18:2,3OH-C18:0
R2 (N-methyl)	Me	Me, H	Me, H	H	H	H
R3	Cb, H	Cb,Acr, H	H	H	H	Ac, H
R4	Cb, H	Cb,Ac,H	H	H	H	Ac, H
R5	Cb, H	Cb,Ac, H	H	Ac,H	H	Ac, H
R6	3-O-S-2-O-MeFuc,3-/4-O-Ac-2-O-MeFuc, 2-O-MeFuc	2-O-MeFuc, Fuc	S,H	S	2-O-MeFuc	H
R7	H	Gro,H	Man,H	H	H	H
R8	Me	Me	Me	Me	Me	Me
R9	H	H	H	H	H	H
R10	H	H	H	H	H	H

* was included here to compare LCO structure

n stands for the number of N-acetyl-D-glucosamine (GlcNAc).

R1 to R10 refer to identity of substitutions on the chitin backbone (D'Haeze and Holsters 2002).

Chapter 6

General discussion

Introduction

In nature, some plants can associate with nitrogen-fixing microbes at different levels to enhance nitrogen nutrient availability (Ormeño-Orrillo *et al.*, 2013). Among these, the interaction leading to the formation of nitrogen-fixing nodules is most prominent due to its high efficiency of nitrogen fixation. Plants that can establish a nitrogen-fixing nodule symbiosis are only found in four related taxonomic orders that form a monophyletic lineage: the so-called nitrogen-fixing clade (NFC) representing Fagales, Fabales, Cucurbitales and Rosales (Soltis *et al.*, 1995). Within the NFC, 10 out of 28 plant families contain plant species that can establish a nitrogen-fixing nodule symbiosis (Soltis *et al.*, 1995). Most families contain only a few nodulating plant species. One good example is the Cannabis family (Cannabaceae, order Rosales), of which *Parasponia* is the only genus that can establish root nodule symbiosis, whereas species of the remaining genera are unable to do so. An exception is the legume family (Fabaceae, order Fabales). This family comprises over 20,000 species divided over 750 genera of which most posses the nitrogen-fixing nodule symbiosis trait. Taken together, in the NFC lineages of nodulating plants are dispersed as this clade also represents many lineages of non-nodulating species.

To establish a nitrogen-fixing nodule symbiosis, plants of the NFC associate with one of the two different types of diazotrophic microsymbionts. Legumes interact with a group of gram-negative bacteria collectively known as rhizobia. Also, *Parasponia* species establish a nitrogen-fixing symbiosis with rhizobia. The remaining nodulating plants associate with gram-positive filamentous *Frankia* species and therefore are collectively called actinorhizal plants. Intriguingly, *Parasponia* and legumes that both interact with rhizobia do not represent a monophyletic group, but diverged >100 million years ago and are interspersed with lineages that nodulate with *Frankia*. Also, there is significant phenotypic variation in legume, *Parasponia* and actinorhizal nodules, especially in nodule ontogeny, infection mode, and the way micro-symbionts are hosted. This led to speculations whether nitrogen-fixing nodule symbiosis evolved multiple times independently in a convergent manner, preceded by a predisposition in the last common ancestor of the NFC (Swensen, 1996; Doyle, 1998, 2011, 2016; Werner *et al.*, 2014; Li *et al.*, 2015; Martin *et al.*, 2017). Alternatively, it was suggested that nodulation evolved only once in the root of the NFC followed by massive losses (Soltis *et al.*, 1995; Swensen, 1996). The latter hypothesis was generally refuted, in favour of the hypothesis of the parallel evolution of the nodulation trait. This, because

convergent evolution requires less evolutionary events and therefore is parsimonious of a single gain - massive loss hypothesis (Jeong *et al.*, 1999; Werner *et al.*, 2014; Li *et al.*, 2015). In this chapter, I will discuss the results described in this thesis in which I deploy *Parasponia* as a comparative model system, and what these findings imply concerning the evolution of nodulation.

Comparative genomics studies to explain the current phylogeny distribution of nodulation

Commonly recruited symbiosis genes in legumes, *Parasponia* and actinorhizal plants

Several transcriptomics studies have been conducted on legumes as well as on several non-legumes to understand the transcriptional changes associated with nodulation. Most studies focused on identification of nodule enhanced genes and each study led to the identification of hundreds, if not thousands of such genes (Mergaert *et al.*, 2019). Together with gene functional analysis, one major finding from these studies is that the convergence of the common symbiosis signalling pathway (CSSP) exploited by the more ancient arbuscular mycorrhizal (AM) symbiosis (Gherbi *et al.*, 2008; Markmann *et al.*, 2008; Hocher *et al.*, 2011; Tromas *et al.*, 2012; Svistoonoff *et al.*, 2013, 2014; Granqvist *et al.*, 2015; Fabre et al. 2015; Chabaud et al. 2016; Op den Camp et al. 2011). The CSSP is likely being used also by those plants that can be engaged with ectomycorrhizal fungi, suggesting the recruitment of this signalling pathway for other symbioses than AM might predate the origin of the NFC (Cope *et al.*, 2019). Besides, the tight correlation between presence of key symbiotic genes -including SYMRK, CCaMK and CYCLOPS from the CSSP- in plant species that can establish intracellular endosymbioses such as AM and nitrogen-fixing nodule symbiosis, and loss of these genes in plants which do not engage in any type of those intracellular infection symbioses, defines the CSSP a universal signalling pathway for intracellular mutualistic symbioses in plants (Radhakrishnan *et al.*, 2019).

In legumes, so far 126 symbiotic genes (including genes from CSSP) have been identified through mutant analysis (**Chapter 2**; Roy *et al.*, 2019). These genes cover diverse programs from symbiotic signalling, nodule organogenesis and autoregulation of nodulation, rhizobium infection to symbiosome formation, maturation and senescence. To determine whether the *Parasponia*-rhizobium symbiosis requires the same genetic signalling pathway as legumes, we assessed the commonalities of the genetic basis of nodulation between the model legume *Medicago truncatula*

and *Parasponia andersonii* using orthology assessment (**Chapter 2**). In total, we identified 1,719 *P. andersonii* genes that have a nodule enhanced expression. The *M. truncatula* orthologs of 290 of these genes are also showing a nodule enhanced expression. Interestingly, within this 290 commonly-recruited gene set, only 26 have previously been identified in legumes as symbiotic genes, indicating that a large part of this core genetic basis of nodulation remains to be uncovered.

Two similar studies identified commonly recruited symbiosis genes in legumes and actinorhizal plant species. One study identified 51 of such genes by comparing *M. truncatula* and the actinorhizal plant species *Ceanothus. thyrsiflorus* (Rhamnanceae, Rosales) and *Datisca glomerata* (Datiscaceae, Cucurbitales). The second study compared four legumes - *Glycine max*, *Lotus japonicus*, *Phaseolus vulgaris* and *M. truncatula* - and identified only 10 commonly recruited nodule enhanced genes (Battenberg *et al.*, 2018; Wu *et al.*, 2019). The limited overlap in conserved nodule-ennhanced genes identified suggests that such studies are not yet saturated. This can be due to the genetic variation in the species used for comparisons, and/or due to technical limitations, e.g. synchronization of nodule development in the different species. This urges for further investments in generating transcriptome datasets with higher cellular resolution, e.g. by using laser-capture microdissection (LCM) analysis, cell type specific or single-cell transcriptome technology. Such strategies will be helpful to limit physiological differences and associated identified genes with certain developmental stage or process. Additionally, a systematic meta-analysis can be applied including representative species from different clades which cover differences in engaged symbiont, infection mode and nodule ontology. Such analysis may lead to the identification of core genes recruited at an earlier evolutionary point of the nitrogen-fixing nodule symbiosis.

Massive loss of key symbiotic genes within non-nodulating species

The *Parasponia* lineages comprise five species that are phylogenetically embedded within the non-symbiotic *Trema* genus (Yang *et al.*, 2013; van Velzen *et al.*, 2018), indicating a closely related relationship. This close relationship allowed me to make intergeneric F1 hybrid plant by crossing the diploid *P. andersonii* and tetraploid *Trema tomentosa* (**Chapter 2 & 3**). Although these hybrid plants did not produce a viable F2 generation, it allowed studying the symbiotic phenotype of these plants. Interestingly, *P. andersonii* x *Trema tomentosa* hybrids can be effectively nodulated, although with a more narrow host range when compared to *P. andersonii* (**Chapter 2 & 3**).

However, hybrids plants showed to be unable to host rhizobia inside nodule cells. This indicates that nodule organogenesis and intracellular infection -at least in part- have different genetic requirements in *P. andersonii*. Alternatively, this phenotype can be the result of *T. tomentosa* specific genes that are expressed in root nodules that interfere with nodulation.

Comparative genome analysis revealed that *Parasponia* and *Trema* species are highly similar in genetic makeup (**Chapter 2**). All sequenced species contain around 300 Mbps of non-repetitive sequence and a variable amount of repetitive DNA content. A medium percentage nucleotide identity of 97% for coding regions of 25,605 orthologous gene pairs supports *Parasponia* and *Trema* species diverged only relatively recent (approximately ~17 million years ago) (Li *et al.*, 2015; van Velzen *et al.*, 2019). To obtain insight into the molecular-genetic changes underlying the evolution of nitrogen-fixing nodule symbiosis, we conducted comparative genomics studies on three *Parasponia* and non-nodulating species Rosales species, including *Trema* (**Chapter 2**). This uncovered the pseudogenization -or even loss- of orthologs of essential symbiotic genes in *Trema* and other non-nodulating Rosales species. Among the loss genes are *NOD FACTOR PERCEPTION* (*NFP*) encoding a LysM-type receptor involved in recognizing rhizobial lipo-chitooligosaccharide (LCO) signal molecules, the *NODULE INCEPTION* (*NIN*) encoding a LCO-responsive transcription factor that is essential for nodule organogenesis and bacterial infection, and the *RHIZOBIUM DIRECTED POLAR GROWTH* (*RPG*) encoding a coil-coiled protein that functions in rhizobium intracellular infection threads elongation. These findings were supported by a similar study comparing in total 37 nodulating and non-nodulating species covering the NFC (Griesmann *et al.*, 2018). As *NFP*, *NIN* and *RPG* are only expressed in a symbiotic context and commit specific functions in nodulation, loss of these genes in non-nodulating species of the NFC suggests that these species have lost the capacity to form nitrogen-fixing nodules. These findings are in line with the hypothesis that the nitrogen-fixing nodulation trait evolved only once in the root of the NFC, followed by massive parallel loss of the trait (van Velzen *et al.*, 2019).

NIN and NF-YA1 transcription factors are a core genetic network of the nitrogen-fixing nodule symbiosis

NIN gained nodulation-related functional adaptation

NIN is among the first genes that is transcriptionally activated by the CSSP upon the perception of rhizobium LCO signals. Also, it functions specifically in nodulation and has no -or only very minor- role in AM symbiosis (Guillotin *et al.*, 2016). *NIN* showed to be essential for nodulation in multiple legume plants (Fabaceae, Fabales) and the actinorhizal species *C. glauca* (Casuarinaceae, Fagales) as well as *P. andersonii* (Cannabaceae, Rosales) (**Chapter 4**; Schauser et al., 1999; Borisov et al., 2003; Marsh et al., 2007; Vernié et al., 2015). These studies in species representing three taxonomic orders; Fabales, Fagales and Rosales support that NIN is a transcriptional master regulator of nodulation. As the orders Fabales, Fagales and Rosales diverged shortly after the birth of the NFC, about 100 million years ago, it suggests that *NIN* was among the first genes recruited in nodulation.

The transcription factor NIN is essential for a diverse range of cellular processes and physiological responses required for nodule organogenesis and rhizobium infection. NIN is part of a small gene family of so-called NIN-LIKE PROTEINs (NLPs) that function in regulating nitrogen homeostasis (Schauser *et al.*, 1999; Riechmann, 2002). Nodulation usually is triggered under low nitrate conditions while higher nitrate can inhibit nodulation and nitrogen fixation (Streeter, 1988; Coronado *et al.*, 1995; Matamoros *et al.*, 1999; Carroll & Mathews, 2018). Recent studies in *L. japonicus* showed that several NLP proteins are involved in the inhibition of rhizobium infection and nodule organogenesis under higher nitrate conditions via physical interaction with NIN. In this way, NIN activity is repressed (Lin *et al.*, 2018). Similar to NLPs, NIN can also bind to the nitrate responsive elements (NREs) present in nitrite inducible genes like *NIR1*, *NRT2.1* and *NRT2.2* (Soyano *et al.*, 2013, 2015). In nodulating plants, NIN also transcriptionally activates the symbiosis related targets of the NUCLEAR FACTOR Y (NF-Y) complex *NF-YA1* and *NF-YB1*, which possess a similar NRE motif in their promoter (Soyano *et al.*, 2013, 2015). Interestingly, rhizobium infection does not activate the expression of nitrate responsive genes in *L. japonicus*. Thus, NIN has likely acquired some unknown adaption which enables it to respond differently upon changing of nitrate homeostasis as well as other physiological changes.

The function of NIN in nodule organogenesis and intracellular infection is -at least partially- dependent on genes encoding the heterotrimeric NF-Y transcription factor complex (Soyano *et al.*, 2013, 2015). NF-Y heterotrimeric transcription factors are composed of NF-YA, NF-YB and NF-YC subunits, of which the DNA-binding specificity is determined by NF-YA (Baudin *et al.*, 2015;

Myers & Holt, 2018). In plants, each subunit is encoded by several genes, which play conserved roles in a broad range of processes like plant-microbe interactions, root development and adaptation to abiotic stresses such as drought and nutrient limitation (Leyva-González *et al.*, 2012; Soyano *et al.*, 2013; Sorin *et al.*, 2014; Zanetti *et al.*, 2017). Mutant analysis in *M. truncatula*, *L. japonicus* and *P. vulgaris* showed that several *NF-Y* genes are involved in multiple steps of nodulation, including nodule organogenesis and bacterial infection. Recent studies in *L. japonicus* and *M. truncatula* showed that lateral root development in part mediated by *NF-YA1* and *NF-YB1*, suggesting this is an ancestral function of both genes prior to recruitment into nodulation (Schiessl *et al.*, 2019; Soyano *et al.*, 2019). Ectopic expression of *LjNF-YA1* and *LjNF-YB1* together with the transcription factor *ASYMMETRIC LEAVES2-LIKE* 18/*LATERAL ORGAN BOUNDARIES DOMAIN* 16a (*ASL*18/*LBD*16a) that is known to control lateral root formation induces root cortical cell divisions in the *L. japonicus nin* mutant (Soyano *et al.*, 2019). Such cortical cell divisions are considered a hallmark of nodule organogenesis. These findings strongly suggest an overlap in the nodule and lateral root developmental programs.

Studies in legumes showed that cytokinin signalling is an integral part of rhizobium LCO signalling (Tirichine *et al.*, 2007; van Zeijl *et al.*, 2015). Also, *NIN* expression can be induced by exogenous cytokinin application (van Zeijl et al. 2015). Recently, cytokinin responsive *cis*-regulatory elements (CE) have been found in the *NIN* promoter of legumes (Liu *et al.*, 2019). This CE element is essential for the primary activation of *NIN* expression in inner root layers to activate nodule organogenesis (Heckmann *et al.*, 2011; Liu *et al.*, 2019). Interestingly, a conserved CE *cis*-regulatory element has not been found in the *Parasponia NIN* promoter, nor in *NIN* promoters of actinorhizal plants. This indicates that *NIN* in nodulating plants might have experienced lineage-specific adaptation.

NF-YA1 plays an essential role in intracellular infection during nodulation

Phylogenetic reconstruction of the *NF-YA* gene family revealed seven orthogroups, and legumes experienced duplication events in all except one orthogroup (**Chapter 4**). In *M. truncatula*, the two paralogous genes *MtNF-AY1* and *MtNF-YA2* display distinct expression patterns (Laloum et al. 2014). Whereas *MtNF-YA1* is specifically induced in epidermal cells and pericycle cells of susceptible zone upon rhizobium infection, *MtNF-YA2* has a basal expression level in root cells under symbiotic and non-symbiotic conditions. This indicates that in *M. truncatula MtNF-YA1* and

MtNF-YA2 experienced subfunctionalization after the duplication event. *P. andersonii* has only a single *NF-YA1* gene, of which the expression is responsive to rhizobium-induced signalling. *PanNF-YA1*$_{pro}$:GUS reporter studies revealed specific expression in dividing epidermal, cortical and pericycle cells that are associated with nodule organogenesis. Also, it was found that *PanNF-YA1* is expressed in pericycle cells located opposite the protoxylem cells under non-symbiotic conditions (**Chapter 4**). This suggests that *PanNF-YA1* commits symbiotic as well as non-symbiotic functions.

In *L. japonicus* and *M. truncatula NF-YA1* functions in rhizobium infection and nodule organogenesis. However, in such legume *nf-ya1* knockout mutants nodulation is not fully blocked and nitrogen-fixing nodules are still formed albeit smaller in size and with lower fixation efficiencies (Combier *et al.*, 2006; Soyano *et al.*, 2013; Laporte *et al.*, 2014; Laloum *et al.*, 2014; Xiao *et al.*, 2014; Hossain *et al.*, 2016). This is probably due to the functional redundancy of other *NF-YA* genes that are expressed in nodules (Laloum *et al.*, 2014; Baudin *et al.*, 2015; Rípodas *et al.*, 2019). Also in *P. andersonii* we identified two additional *NF-YA* genes that are induced transcriptionally in nodules; *PanNF-YA3* and *PanNF-YA6*. Through analysis of single, double and higher-order CRISPR-Cas9 knockout mutants of these three nodule-enhanced *NF-YA* genes, we found *PanNF-YA1* plays an exclusive role in intracellular infection thread formation, whereas only a *Pannf-ya1;Pannf-ya3;Pannf-ya6* triple mutant is also affected in nodule organogenesis. This indicates that *PanNF-YA1*, *PanNF-YA3* and *PanNF-YA6* function redundantly in nodule development in *Parasponia* (**Chapter 4).**

NIN and NF-YA1 transcription factors are core genetic network in nitrogen-fixing nodule symbiosis

In. *L. japonicus* it was shown that *LjNF-YA1* is a direct transcriptional target of LjNIN (Soyano et al. 2013; Soyano et al. 2015). We showed that *PanNIN* and *PanNF-YA1* are coexpressed in *P. andersonii* nodules and qRT-PCR analyses revealed that symbiotic induction of *PanNF-YA1* is PanNIN dependent. The presence of NIN binding sites in the putative promoter of *PanNF-YA1*, further supports the notion that *PanNF-YA1* might be a direct target of PanNIN (**Chapter 4**). Besides in *Parasponia*, *NF-YA1* was also found to be induced in nodules of actinorhizal species, such as *C. glauca* and *Alnus glutinosa* (Diédhiou *et al.*, 2014). This suggests a possible role for *NF-YA* genes in nitrogen-fixing nodule symbiosis in these plant species as well. Since *NIN* has

been recruited most probably at the root of the NFC, we hypothesize that *NIN* and *NF-YA1* represent a core transcription network controlling nodulation in the NFC.

When the NIN and NF-YA1 network was recruited for nodulation?

I argue that *NIN* and *NF-YA1* are part of the core genetic basis for nitrogen-fixing nodule symbiosis, thus the question remains when this transcriptional module has been recruited? Studies in *Parasponia* have revealed that NF-YA1 is essential for intracellular infection (**Chapter 4**). Two hypotheses can explain the evolution of intracellular infection. It evolved gradually from bacterial microbes that co-colonize plant roots together with AM fungi (**Fig. 1A-B**). In such a scenario intracellular infection was initially *NF-YA1* independent, as the bacteria joined the AM fungus to enter cells. *NF-YA1* could have been recruited as a NIN target in the early evolutionary trajectory of nodulation to allow AM independent infection of nitrogen-fixing bacteria. Alternatively, early nitrogen-fixing micro-symbionts can't hijack AM infection of cortical cells, but require *de novo* induced cell divisions that are associated with expression of *NIN* (**Fig. 1C-D**). In such a scenario it is most probable that together with *NIN, NF-YA1* was recruited simultaneously. This would imply that the NIN - NF-YA1 transcriptional module predates the NFC. Upon recruitment of *NIN*, e.g. by requiring novel *cis*-regulatory elements that enabled NIN-driven cortical cell divisions, these cells also express *NF-YA1*.

To discriminate both hypotheses, it will be essential to study whether the NIN - NF-YA1 transcriptional module exists also outside of the NFC (**Fig. 1F-G**). Such studies could be done in *Arabidopsis thaliana*, as this model plant possess orthologs of both genes. Alternatively, it could be investigated whether this module also controls a more distinct nitrogen-fixing endosymbiosis that has evolved in parallel to the nitrogen-fixing clade; e.g. the symbiosis between *Gunnera* spp. and nitrogen-fixing cyanobacterial species of the genus *Nostoc*. *Nostoc* spp. are also hosted intracellularly, and infect meristem derived cells of glands (Johansson & Bergman, 1992; Osborne & Bergman, 2008; Geurts *et al.*, 2016). It would be interesting to investigate whether the same transcriptional modules are recruited to support this interaction. The first step to answering such a question would be to conduct comparative transcriptomic and genomics analysis of symbiotic tissue of *Gunnera*, to determine whether *NIN* - *NF-YA1* also plays a symbiotic role in this symbiosis.

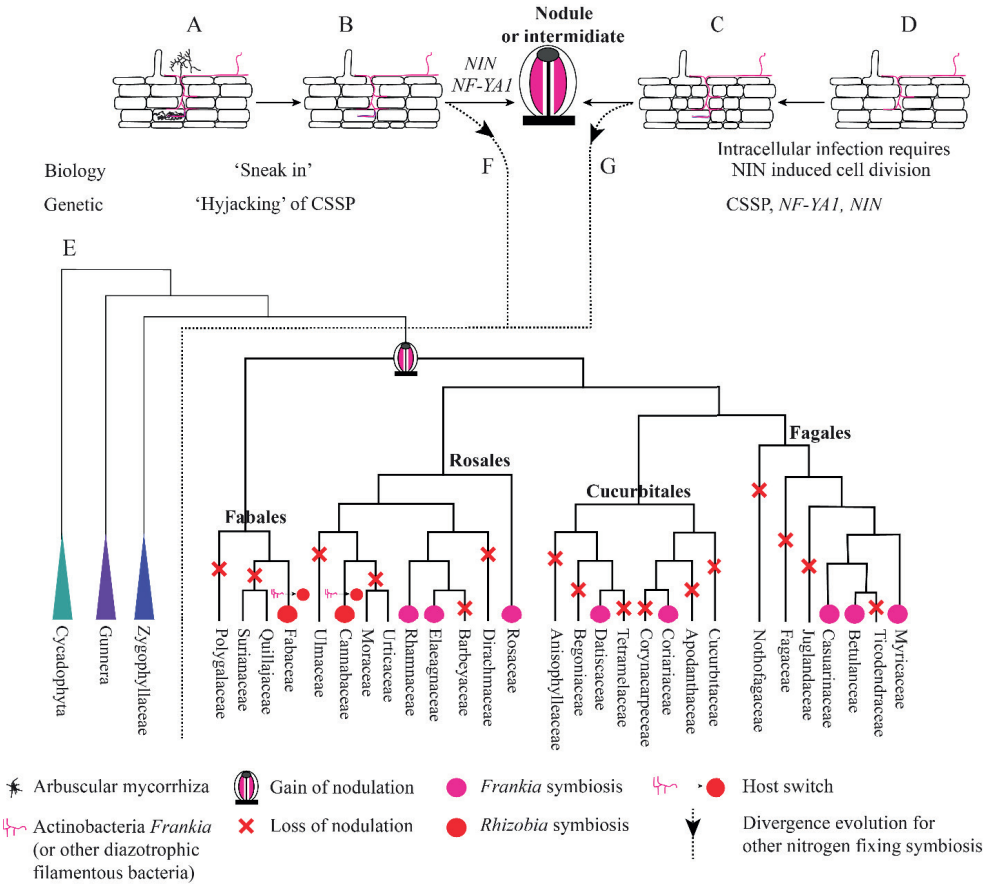


Fig. 1 Evolution of nitrogen-fixing symbiosis in the NFC and recruitment of *NIN* and *NF-YA1*. A -B, Intracellular infection of nitrogen-fixing bacteria evolved by hijacking AM symbiosis, which did not require cortical cell divisions nor *NF-YA1* expression ('sneak in' hypothesis). **A**, Diazotrophic microbes co-colonize plant roots associated with AM fungi. **B,** Intracellular accommodation of root cells by diazotrophic bacteria independently from AM fungi. **C-D** Intracellular accommodation of root cells by diazotrophic bacteria independently from AM fungi, but requiring NIN induced cell division and *NF-YA1* expression. **E**, Current phylogenetic distribution of nitrogen-fixing nodule symbiosis is likely being shaped by a single gain in the common ancestor of the NFC, coupled with multiple losses and two microsymbiont switches that happened in the Fabaceae (Fabales) and *Parasponia* (Cannabaceae, Rosales). **F-G**, Diverged evolution of nitrogen-fixing symbiosis in Cycadophyta, Gunnera and Zygophyllaceae from NFC. Figure is modified based on Mergaert et al. (2019) with permission from the publisher.

Parasponia; a missing piece of the evolutionary puzzle of nitrogen-fixing nodule symbiosis

The current phylogenetic distribution of nodulation species is likely shaped by a single gain and multiple losses (**Figure 1E**). This hypothesis is supported by the finding of loss or pseudogenization of genes (*NFP*, *NIN* and *RPG*) that in nodulating plants confer essential functions in establishing a nitrogen-fixing symbiosis. These genes have a symbiosis-specific expression profile, and from mutant analysis, no function other than nodulation has been identified. Nevertheless, Mergaert *et al.* suggests that an independent evolution of the trait cannot be excluded (Mergaert *et al.*, 2019). Our comparative study on *Parasponia* and *Trema* may shed light on the likeliness of this viewpoint. *Parasponia* represents a lineage within the *Trema* genus and diverged less than 17 million years ago (Li *et al.*, 2015; van Velzen *et al.*, 2019). *Parasponia* has *Trema* sister species, e.g. *Trema levigata*, and a *Trema* outroup, e.g. *Trema orientalis* (Van Velzen et al., 2018). Furthermore, any other Cannabaceae (e.g. *Humulus lupulus*), Moraceae (*Morus notabilis*), non-nodulating Ramanaceae (e.g. Rhamnaceae) and Rosaceae (e.g. *Fragaria vesca, Malus x domestica* or *Prunus persica*) species can be considered as an outgroup. From the point of view of Mergaert *et al.* independent loss of the three key symbiotic genes *NFP, NIN* and *RPG* in non-nodulating *T. levigata, T. orientalis, H. humulus, M. notabilis, F. vesca, Malus x domestica* and *P. persica* does not necessarily mean loss of the nodulation trait in these species. In other words, it means that *Parasponia* is the only lineage that maintained these genes to use them in nodulation, only after the split of *T. levigata*. Theoretically, such a scenario can not be excluded. However, it leaves essential questions unanswered. Why independent gain only happened in *Parasponia* while not in other Cannabaceae species? Why these genes were maintained only in *Parasponia* for a long period of time before it independently evolves nodulation? Therefore, I consider this hypothesis very unlikely to be correct. Instead, it is more likely that *Trema* and all other mentioned outgroups have lost the nodulation trait (**Fig. 1E**). This loss of a trait gradually led to the loss of key symbiotic genes which is called co-elimination (Force *et al.*, 1999; Albalat & Cañestro, 2016). Such co-elimination happened also widely within the plant kingdom when plants lost intracellular AM symbiosis (Radhakrishnan *et al.*, 2019). *Parasponia* is likely the only genus within the Cannabaceae that remained the nitrogen-fixing nodule symbiosis trait, whereas the remaining

species within this family lost the trait by parallel events. In line with this, *Parasponia* can be considered as a living nitrogen-fixing fossil.

Though a single origin coupled with massive loss can fit well in the current phylogenetic distribution of nodulation within the NFC, it does not explain the occurrence of classes of distinct micro symbionts -*Frankia* and rhizobia- that are able to trigger nitrogen-fixing nodules. Most strikingly, the plant families which are able to establish nitrogen fixing symbiosis with the same diazotrophic bacteria do not form a single phylogenetic group. Instead, *Parasponia* represents the only lineage outside the legumes that can form nodules with rhizobia, whereas other nodulating Rosales species represent actinorhizal plants (e.g. *Dryas*, *Discaria*, and *Ceanothus*). This suggests that at least two switches in microsymbiont must have occurred. Based on independent evolution of hemoglobins in *Parasponia*, legumes and the actinorhizal plant *Casuarina glauca* (Casuarinaceae, Fagales), it was hypothesized that *Parasponia* experienced a switch from *Frankia* to rhizobium as microbial host (van Velzen *et al.*, 2019). Hemoglobin is highly expressed in *Parasponia* root nodules and can provide and optimal oxygen homeostasis within the nodule, which is essential to protect the rhizobial nitrogenase enzyme complex. *Frankia* not necessarily relies on plant encoded hemoglobin, as it can maintain its oxygen homeostasis by the formation of vesicles with protective laminar lipid layers. *Parasponia* deploys class I hemoglobin in root nodules, which is different from the one used by legumes (class II). Also, *Parasponia* hemoglobin class I experienced recent adaptations that lead to lower oxygen affinity allowing the protein to function as an oxygen transporter rather than a scavenger (**Chapter** 2). Such adaptations I consider as essential to allow a symbiont switch from a self-supporting *Frankia* to rhizobium that needs protection against oxygen when fixing nitrogen.

As I hypothesize that *Parasponia* only recently experienced a microbial host switch, I questioned whether *Frankia* may still be able to trigger symbiotic responses on this species. To get first insights whether this might be the case, I inoculated the transgenic *P. andersonii PanNF-YA1*$_{pro}$:GUS reporter line with the cluster II *Frankia* strain DG2. This strain contains the canonical LCO biosynthesis *nodABC* genes (Van Nguyen *et al.*, 2016). *PanNF-YA1*$_{pro}$:GUS is detected in epidermal cells at 48 hours post-inoculation with rhizobium (**Chapter 4**). Due to the fact that cluster II *Frankia* strains are unculturable, crushed fresh nodules from *Frankia* DG2 inoculated *D. glomerata* plants were prepared after surface sterilization and used as inoculum. GUS signal was

scored 11 weeks post inoculation. In 2 out of 10 *PanNF-YA1*$_{pro}$:GUS plants, a GUS signal was detected in nodule-like outgrowth (**Fig. 2A-B**). Such signal is not found on roots of uninoculated control plants. Cytoarchitecture analysis of the nodule-like structures revealed that these originate from dividing epidermal and cortical cells. Also, the presence of unknown microorganisms was observed in the apoplastic region as well as GUS signal in the neighbouring cells (**Fig. 2C-D**). However, due to the lack of information about the nature of the observed microbes it yet can not be concluded that *Frankia* sp. DG2 can trigger nodule structures on *P. andersonni* roots. To come to such a conclusion, metagenomic analysis of the initial inoculum prepared from nodules formed on *D. glomerata* plants, as well as the microbiome composition of the rhizosphere samples and nodule-like structures will be required. Besides, I have shown in **Chapter 5** that without functional *nodS*, rhizobium is only able to trigger nodule-like outgrowth instead of functional nodules. In *Frankia* spp. DG2, no such gene with high similarity to rhizobium *nodS* has been identified.

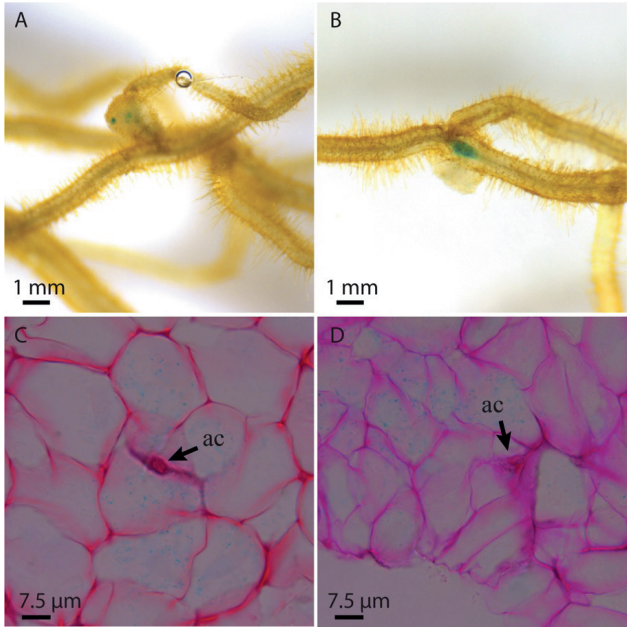


Fig. 2 Nodule-like structures formed on roots of transgenic *Parasponia* plants (*PanNF-YA1*$_{pro}$:GUS) inoculated with *Frankia* sp. DG2. **A-B**, Nodule-like structures as observed 11 weeks post-inoculation on *PanNF-YA1*$_{pro}$:GUS transgenic line 1E5 inoculated with *Frankia* spp. DG2 inoculum prepared from surface-sterilised *Datisca glomerata* nodules. GUS signal was detected after incubating in GUS buffer (3% [w/v] sucrose, 10 mM EDTA, 2 mM k-ferrocyanide, 2 mM k-ferricyanide, and 0.5 mg/mL 5-bromo-4-chloro-3-indolyl-beta-D- glucuronic acid, cyclohexylammonium salt [X-Gluc] in 0.1 M phosphate buffer [pH = 7.2]) at 37°C for 5 hours. **C-D**, Section of nodule-like structures showing apoplastic colonization of non-rhizobial microbes based on its diameter. GUS signalling is detected in the surrounding cells. ac: apoplast infection of non-rhizobial microbes. Sections (7 µm) were counterstained with Ruthenium Red for **C-D**.

Conclusion

In this thesis, I used comparative genomic and transcriptomic analysis to show that key symbiotic genes have been continuously lost in non-nodulating species within the nitrogen-fixing clade. By CRISPR-Cas9 mediated mutagenesis on the non-legume *P. andersonii*, I proved that the transcriptional module NIN - NF-YA1 plays a conserved role in nitrogen-fixing nodule symbioses. This has led to a paradigm shift in the hypothesis concerning the evolution of the nitrogen-fixing nodule symbiosis trait. The parallel evolution hypothesis which was advocated for more than two decades finds no support by current comparative analysis. Instead, a single gain of the nodulation trait followed by many parallel losses is more likely. Although the comparative analysis identified three key symbiosis genes, *NFP, NIN* and *RPG*, the minimal number of genes needed to establish a symbiosis remains elusive. Furthermore, the characterization of the *Parasponia x Trema* hybrid learned that genes preventing interaction between nitrogen-fixing bacteria and host plant are new players in the field that need equal attention to understand the evolution of nodulation.

References

Albalat R, Cañestro C. **2016**. Evolution by gene loss. *Nature reviews. Genetics* **17**: 379–391.

Battenberg K, Potter D, Tabuloc CA, Chiu JC, Berry AM. **2018**. Comparative Transcriptomic Analysis of Two Actinorhizal Plants and the Legume Medicagotruncatula Supports the Homology of Root Nodule Symbioses and Is Congruent With a Two-Step Process of Evolution in the Nitrogen-Fixing Clade of Angiosperms. *Frontiers in Plant Science* **9**.

Baudin M, Laloum T, Lepage A, Ripodas C, Ariel F, Frances L, Crespi MD, Gamas PC, Blanco FA, Zanetti ME, *et al.* 2015. A phylogenetically conserved group of NF-Y transcription factors interact to control nodulation in legumes. *Plant physiology* **169**: 01144.2015.

Borisov AY, Madsen LH, Tsyganov VE, Umehara Y, Voroshilova VA, Batagov AO, Sandal N, Mortensen A, Schauser L, Ellis N, *et al.* 2003. The Sym35 gene required for root nodule development in pea is an ortholog of Nin from Lotus japonicus. *Plant physiology* **131**: 1009–1017.

Carroll BJ, Mathews A. **2018**. Nitrate inhibition of nodulation in legumes. In: Molecular biology of symbiotic nitrogen fixation. CRC Press, 159–180.

Chabaud M, Gherbi H, Pirolles E, Vaissayre V, Fournier J, Moukouanga D, Franche C, Bogusz D, Tisa LS, Barker DG, *et al.* 2016. Chitinase-resistant hydrophilic symbiotic factors secreted by Frankia activate both Ca(2+) spiking and NIN gene expression in the actinorhizal plant Casuarina glauca. *The New phytologist* **209**: 86–93.

Combier J-P, Frugier F, de Billy F, Boualem A, El-Yahyaoui F, Moreau S, Vernié T, Ott T, Gamas P, Crespi M, *et al.* 2006. MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. *Genes & development* **20**: 3084–3088.

Cope KR, Bascaules A, Irving TB, Venkateshwaran M, Maeda J, Garcia K, Rush TA, Ma C, Labbé J, Jawdy S, *et al.* 2019. The Ectomycorrhizal Fungus Laccaria bicolor Produces Lipochitooligosaccharides and Uses the Common Symbiosis Pathway to Colonize Populus Roots. *The Plant cell* **31**: 2386–2410.

Coronado C, Zuanazzi J, Sallaud C, Quirion JC, Esnault R, Husson HP, Kondorosi A, Ratet P. **1995**. Alfalfa Root Flavonoid Production Is Nitrogen Regulated. *Plant physiology* **108**: 533–542.

Diédhiou I, Tromas A, Cissoko M, Gray K, Parizot B, Crabos A, Alloisio N, Fournier P, Carro L, Svistoonoff S, *et al.* 2014. Identification of potential transcriptional regulators of actinorhizal symbioses in Casuarina glauca and Alnus glutinosa. *BMC plant biology* **14**: 342.

Doyle JJ. **1998**. Phylogenetic perspectives on nodulation: evolving views of plants and symbiotic bacteria. *Trends in plant science* **3**: 473–478.

Doyle JJ. **2011**. Phylogenetic perspectives on the origins of nodulation. *Molecular plant-microbe interactions: MPMI* **24**: 1289–1295.

Doyle JJ. **2016**. Chasing unicorns: Nodulation origins and the paradox of novelty. *American*

journal of botany **103**: 1–4.

Fabre S, Gully D, Poitout A, Patrel D, Arrighi J-F, Giraud E, Czernic P, Cartieaux F. 2015. The Nod factor-independent nodulation in Aeschynomene eveniarequired the common plant-microbe symbiotic 'toolkit'. *Plant physiology* **169**: 2654–2664.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531–1545.

Geurts R, Xiao TT, Reinhold-Hurek B. 2016. What Does It Take to Evolve A Nitrogen-Fixing Endosymbiosis? *Trends in plant science* **21**: 199–208.

Gherbi H, Markmann K, Svistoonoff S, Estevan J, Autran D, Giczey G, Auguy F, Péret B, Laplaze L, Franche C, *et al.* 2008. SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and Frankiabacteria. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 4928–4932.

Granqvist E, Sun J, Op Den Camp R, Pujic P, Hill L, Normand P, Morris RJ, Downie JA, Geurts R, Oldroyd GED, *et al.* 2015. Bacterial-induced calcium oscillations are common to nitrogen-fixing associations of nodulating legumes and nonlegumes. *The New phytologist* **207**: 551–558.

Griesmann M, Chang Y, Liu X, Song Y, Haberer G, Crook MB, Billault-Penneteau B, Lauressergues D, Keller J, Imanishi L, *et al.* 2018. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* **361**.

Guillotin B, Couzigou J-M, Combier J-P. 2016. NIN Is Involved in the Regulation of Arbuscular Mycorrhizal Symbiosis. *Frontiers in plant science* **7**: 1704.

Heckmann AB, Sandal N, Bek AS, Madsen LH, Jurkiewicz A, Nielsen MW, Tirichine L, Stougaard J. 2011. Cytokinin induction of root nodule primordia in Lotus japonicusis regulated by a mechanism operating in the root cortex. *Molecular plant-microbe interactions: MPMI* **24**: 1385–1395.

Hocher V, Alloisio N, Auguy F, Fournier P, Doumas P, Pujic P, Gherbi H, Queiroux C, Da Silva C, Wincker P, *et al.* 2011. Transcriptomics of actinorhizal symbioses reveals homologs of the whole common symbiotic signaling cascade. *Plant physiology* **156**: 700–711.

Hossain MS, Shrestha A, Zhong S, Miri M, Austin RS, Sato S, Ross L, Huebert T, Tromas A, Torres-Jerez I, *et al.* 2016. Lotus japonicus NF-YA1 Plays an Essential Role During Nodule Differentiation and Targets Members of the SHI/STY Gene Family. *Molecular plant-microbe interactions* **29**: 950–964.

Jeong SC, Ritchie NJ, Myrold DD. 1999. Molecular phylogenies of plants and Frankia support multiple origins of actinorhizal symbioses. *Molecular phylogenetics and evolution* **13**: 493–503.

Johansson C, Bergman B. 1992. Early events during the establishment of the Gunnera/Nostoc symbiosis. *Planta* **188**: 403–413.

Laloum T, Baudin M, Frances L, Lepage A, Billault-Penneteau B, Cerri MR, Ariel F,

Jardinaud M-F, Gamas P, de Carvalho-Niebel F, *et al.* 2014. Two CCAAT-box-binding transcription factors redundantly regulate early steps of the legume-rhizobia endosymbiosis. *The Plant journal* **79**: 757–768.

Laporte P, Lepage A, Fournier J, Catrice O, Moreau S, Jardinaud M-F, Mun J-H, Larrainzar E, Cook DR, Gamas P, *et al.* 2014. The CCAAT box-binding transcription factor NF-YA1 controls rhizobial infection. *Journal of experimental botany* **65**: 481–494.

Leyva-González MA, Ibarra-Laclette E, Cruz-Ramírez A, Herrera-Estrella L. 2012. Functional and transcriptome analysis reveals an acclimatization strategy for abiotic stress tolerance mediated by Arabidopsis NF-YA family members. *PloS one* **7**: e48138.

Lin J-S, Li X, Luo Z, Mysore KS, Wen J, Xie F. 2018. NIN interacts with NLPs to mediate nitrate inhibition of nodulation in Medicago truncatula. *Nature plants* **4**: 942–952.

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

Li H-L, Wang W, Mortimer PE, Li R-Q, Li D-Z, Hyde KD, Xu J-C, Soltis DE, Chen Z-D. 2015. Large-scale phylogenetic analyses reveal multiple gains of actinorhizal nitrogen-fixing symbioses in angiosperms associated with climate change. *Scientific reports* **5**: 14023.

Markmann K, Giczey G, 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.

Marsh JF, Rakocevic A, Mitra RM, Brocard L, Sun J, Eschstruth A, Long SR, Schultze M, Ratet P, Oldroyd GED. 2007. *Medicago truncatula NIN* is essential for rhizobial-independent nodule organogenesis induced by autoactive Calcium/Calmodulin-Dependent Protein Kinase. *Plant physiology* **144**: 324–335.

Martin FM, Uroz S, Barker DG. 2017. Ancestral alliances: Plant mutualistic symbioses with fungi and bacteria. *Science* **356**: eaad4501.

Matamoros MA, Baird LM, Escuredo PR, Dalton DA, Minchin FR, Iturbe-Ormaetxe I, Rubio MC, Moran JF, Gordon AJ, Becana M. 1999. Stress-induced legume root nodule senescence. Physiological, biochemical, and structural alterations. *Plant physiology* **121**: 97–112.

Mergaert P, Kereszt A, Kondorosi E. 2019. Gene Expression in Nitrogen-Fixing Symbiotic Nodule Cells in Medicago truncatula and Other Nodulating Plants. *The Plant cell*.

Myers ZA, Holt BF 3rd. 2018. NUCLEAR FACTOR-Y: still complex after all these years? *Current opinion in plant biology* **45**: 96–102.

Ormeño-Orrillo E, Hungria M, Martinez-Romero E. 2013. Dinitrogen-Fixing Prokaryotes. *The Prokaryotes*: 427–451.

Osborne B, Bergman B. 2008. Why Does Gunnera Do It and Other Angiosperms Don't? An Evolutionary Perspective on the Gunnera--Nostoc Symbiosis. In: Prokaryotic symbionts in plants.

Springer, 207–224.

Op den Camp R, Streng A, De Mita S, Cao Q, Polone E, Liu W, Ammiraju JSS, Kudrna D, Wing R, Untergasser A, *et al.* 2011. LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume *Parasponia*. *Science* **331**: 909–912.

Radhakrishnan GV, Keller J, Rich MK, Vernié T, Mbadinga Mbaginda DL, Vigneron N, Cottret L, Clemente HS, Libourel C, Cheema J, *et al.* 2019. An ancestral signalling pathway is conserved in plant lineages forming intracellular symbioses. *bioRxiv*: 804591.

Riechmann JL. **2002**. Transcriptional regulation: a genomic overview. *The Arabidopsis book / American Society of Plant Biologists* **1**: e0085.

Rípodas C, Castaingts M, Clúa J, Villafañe J, Blanco FA, Zanetti ME. **2019**. The PvNF-YA1 and PvNF-YB7 Subunits of the Heterotrimeric NF-Y Transcription Factor Influence Strain Preference in the Phaseolus vulgaris-Rhizobium etli Symbiosis. *Frontiers in plant science* **10**: 221.

Roy S, Liu W, Nandety RS, Crook AD, Mysore KS, Pislariu CI, Frugoli JA, Dickstein R, Udvardi MK. **2019**. Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. *The Plant cell*.

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

Schiessl K, Lilley JLS, Lee T, Tamvakis I, Kohlen W, Bailey PC, Thomas A, Luptak J, Ramakrishnan K, Carpenter MD, *et al.* 2019. NODULE INCEPTION Recruits the Lateral Root Developmental Program for Symbiotic Nodule Organogenesis in Medicago truncatula. *Current biology: CB* **29**: 3657–3668.e5.

Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JM, Martin PG. **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* **92**: 2647–2651.

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

Soyano T, Kouchi H, Hirota A, 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.

Soyano T, Shimoda Y, Hayashi M. **2015**. NODULE INCEPTION antagonistically regulates gene expression with nitrate in Lotus japonicus. *Plant & cell physiology* **56**: 368–376.

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

Streeter J. **1988**. Inhibition of legume nodule formation and N2 fixation by nitrate. *Critical reviews in plant sciences* **7**: 1–23.

Svistoonoff S, Benabdoun FM, Nambiar-veetil M, Imanishi L, Bonneau J, Wall L, Ykhlef N, Rosenberg C, Bogusz D, Franche C. 2013. The independent acquisition of plant root nitrogen-fixing symbiosis in fabids recruited the same genetic pathway for nodule organogenesis. *PloS one* **8**: e64515.

Svistoonoff S, Hocher V, Gherbi H. 2014. Actinorhizal root nodule symbioses: what is signalling telling on the origins of nodulation? *Current opinion in plant biology* **20**: 11–18.

Swensen SM. 1996. The evolution of actinorhizal symbioses: Evidence for multiple origins of the symbiotic association. *American journal of botany* **83**: 1503–1512.

Tirichine L, Sandal N, Madsen LH, Radutoiu S, Albrektsen AS, Sato S, Asamizu E, Tabata S, Stougaard J. 2007. A gain-of-function mutation in a cytokinin receptor triggers spontaneous root nodule organogenesis. *Science* **315**: 104–107.

Tromas A, Parizot B, Diagne N, Champion A, Hocher V, Cissoko M, Crabos A, Prodjinoto H, Lahouze B, Bogusz D, *et al.* 2012. Heart of endosymbioses: transcriptomics reveals a conserved genetic program among arbuscular mycorrhizal, actinorhizal and legume-rhizobial symbioses. *PloS one* **7**: e44742.

Van Nguyen T, Wibberg D, Battenberg K, Blom J, Vanden Heuvel B, Berry AM, Kalinowski J, Pawlowski K. 2016. An assemblage of Frankia Cluster II strains from California contains the canonical nod genes and also the sulfotransferase gene nodH. *BMC genomics* **17**: 796.

van Velzen R, Doyle JJ, Geurts R. 2019. A Resurrected Scenario: Single Gain and Massive Loss of Nitrogen-Fixing Nodulation. *Trends in plant science* **24**: 49–57.

van Velzen R, Holmer R, Bu F, Rutten L, van Zeijl A, Liu W, Santuari L, Cao Q, Sharma T, Shen D, *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.

Vernié T, Kim J, Frances L, Ding Y, Sun J, Guan D, Niebel A, Gifford ML, de Carvalho-Niebel F, Oldroyd GED. 2015. The NIN transcription factor coordinates diverse nodulation programs in different tissues of the *Medicago truncatula* root. *The Plant cell* **27**: tpc.15.00461.

Werner GDA, Cornwell WK, Sprent JI, Kattge J, Kiers ET. 2014. A single evolutionary innovation drives the deep evolution of symbiotic N2-fixation in angiosperms. *Nature communications* **5**: 4087.

Wu Z, Wang M, Yang S, Chen S, Chen X, Liu C, Wang S, Wang H, Zhang B, Liu H, *et al.* 2019. A global coexpression network of soybean genes gives insight into the evolution of nodulation in non-legumes and legumes. *The New phytologist*.

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

Yang M-Q, van Velzen R, Bakker FT, Sattarian A, Li D-Z, Yi T-S. 2013. Molecular phylogenetics and character evolution of Cannabaceae. *Taxon* **62**: 473–485.

Zanetti ME, Rípodas C, Niebel A. **2017**. Plant NF-Y transcription factors: Key players in plant-microbe interactions, root development and adaptation to stress. *Biochimica et Biophysica Acta, Gene Regulatory Mechanisms* **1860**: 645–654.

van Zeijl A, Op den Camp RHM, Deinum EE, Charnikhova T, Franssen H, Op den Camp HJM, Bouwmeester H, Kohlen W, Bisseling T, Geurts R. **2015**. Rhizobium Lipo-chitooligosaccharide Signaling Triggers Accumulation of Cytokinins in Medicago truncatula Roots. *Molecular plant* **8**: 1213–1226.

Summary

Nitrogen-fixing root nodule symbiosis occurs in ten taxonomic lineages from four related orders - Fagales, Fabales, Rosales and Cucurbitales- that together are called the nitrogen-fixing clade (NFC). Nodulating plants within the NFC are scattered by non-nodulating species, as well as can interact either with rhizobia or *Frankia* bacteria. To establish such an endosymbiosis, two processes are essential: nodule organogenesis and intracellular bacterial infection. Despite a significant body of knowledge of the legume-rhizobium symbiosis, it remains elusive which signalling modules are shared between nodulating species in different taxonomic clades. Besides, it is generally assumed that nodulation evolved independently multiple times, though molecular genetic support for this hypothesis is lacking.

To answer these questions, comparative genomic and transcriptomic analysis has been conducted using *Parasponia* species (Cannabaceae), the only non-legumes that can establish nitrogen-fixing nodules with rhizobium. Comparative transcriptomics of *P. andersonii* and the legume *Medicago truncatula* revealed utilization of at least 290 orthologous symbiosis genes in nodules. Among these are key genes that in legumes are essential for nodulation, including *NODULE INCEPTION* (*NIN*) and *RHIZOBIUM-DIRECTED POLAR GROWTH (RPG)*. Comparative analysis of genomes from three *Parasponia* species and related non-nodulating plant species show evidence of parallel loss in non-nodulating species of putative orthologs of *NIN*, *RPG*, and *NOD FACTOR PERCEPTION*. Parallel loss of these symbiosis genes indicates that these non-nodulating lineages lost the potential to nodulate. By making use of the highly efficient Parasponia transformation platform, we conducted promoter:GUS expression analysis as well as CRISPR-Cas9 mutagenesis. Consistent with legumes, *P. andersonii PanNIN* and *PanNF-YA1* are co-expressed in nodules. By analyzing single, double and higher-order CRISPR-Cas9 knockout mutants, we show that nodule organogenesis and early symbiotic expression of *PanNF-YA1* are PanNIN-dependent and that PanNF-YA1 is specifically required for intracellular rhizobium infection. This demonstrates that *NIN* and *NF-YA1* commit conserved symbiotic functions in non-elgume plant species. As Rosales, Fabales and Fagales diverged soon after the birth of the nodulation trait, we argue that NIN and NF-YA1 represent core transcriptional regulators in this symbiosis. Taken together, these results challenge the view that nodulation evolved in parallel and raises the possibility that nodulation originated ~100 million years ago in a common ancestor of all nodulating plant species, but was

subsequently lost in many descendant lineages. This will have profound implications for translational approaches aimed at engineering nitrogen-fixing nodules in crop plants.

The F1 hybrid between diploid *Parasponia andersonii* and tetraploid *Trema tomentosa* can form nodules, whereas it is devoid of intracellular infection when inoculated with either *Mesorhizobium plurifarium* BOR2 or *Bradyrhizobium elkanii* WUR3. Based on its genetic composition and symbiotic phenotype, we argue that the F1 hybrid may mimic future engineer results. Therefore we aimed to obtain a better understanding of the deviation in nodulation phenotype of wild type *P. andersonii* and F1 hybrid plants. To do so, we compared nodulation efficiencies and intracellular infection within nodule cells upon inoculation with a range of rhizobium strains, as *Parasponia* can interact with a wide range of rhizobia. This revealed that the host range of hybrid plants is narrower when compared to *P. andersonii*. We also show that the block in intracellular infection within hybrid nodules is consistent for all nodulating strains identified, cannot be overcome by increased LCO biosynthesis nor by mutating the type III or IV secretion systems of nodulating strains. The hybrid plants can establish arbuscular mycorrhization effectively, suggesting that the block of intracellular infection is rhizobium specific. Taken together, this indicates the occurrence of a yet unknown mechanism leading to an impaired host range and block of intracellular infection of hybrid plants. We noticed that nodule formation and intracellular infection in *P. andersonii* correlates with the presence of N-methylation conferred by the bacterial *nodS* gene, which encodes an N-methyl transferase that methylates non-reducing terminal residue of LCOs. The importance of *nodS* is shown by demonstrating that LCO signalling induced by *Rhizobium tropici* CIAT899 is abolished when *nodS* is mutated. We conclude that the N-methyl decoration of the non-reducing terminal residue of LCOs is essential for establishing successful nitrogen-fixing nodule symbiosis between rhizobium and *Parasponia andersonii*.

Acknowledgements

The past few years have helped shape my future, for both my career and personal life, and I am grateful for everything I have experienced. I'd like to give my sincerest acknowledgement to all the people who have been there along the way to help and support me.

Rene, thank you for recruiting me into your team. The memories of our greeting in the early morning at Schiphol are still fresh in my mind. You have helped me so much during the last few years, probably more than I realize. Thank you for all you provided, including the discussions, advice, encouragement, and not to forget, the criticisms. I appreciated the level of freedom you gave to me to explore my interests, I have learned so much from you about presentation, and I thank you for teaching me how to deal with conflict. I will treasure these skills for the rest of my life. To add to this, your support in extending my contract so that I could continue my PhD research and thesis writing was invaluable. Ton, thank you for being my promoter. Thanks for the inspiring discussions. I admire your passion for science and life, and will never forget the weekend we worked together in your garden.

I would like to give my special acknowledgements to Henk. Henk, without your help and support, I would not have been able to meet the deadline. And thank you for all the encouragement you offered, it means a lot to me.

I would also like to thank all members of the Parasponia team. Thanks, Marijke, for introducing me to Molbi, and for generously helping with my experiments at the beginning of my PhD. Thanks Robin and Rens for establishing the bioinformatic platform for the whole team which we all dependent on. Thank you Luuk for all your help with learning phylogenetic analysis and Golden Gate ligation system. Thanks to Sidney for your time in taking care of my tissue culture during my vacation, and also your help with sampling. Thank Wouter for your feedback during work discussions. Yuda, thank you for helping me analyzing the mycorrhization phenotype of the Parasponia mutant lines. Titis, thanks for all your knowledge and help with how to set up the LCO early induction experiments. Jan V, thank you for your assistance with molecular work. Thanks Elena for your guiding me when exploring the world of phenotyping the Parasponia nodules F1

hybrid nodules, and also for your aid in preparing the Parasponia nodule transcriptome samples. Arjan, thank you for all your help and encouragement through my PhD. All of our discussions about PhD life, projects, experiments, and results, helped me get through all of the hard times. It has been a pleasure to work with you to get our manuscript published, and thank you for your generous time in improving the story of other chapters.

I greatly enjoyed the dynamic and interactive work environment at Molbi. and first, I would like to thank Marie-Jose and Maria for always being so nice and helping me whenever I needed. In addition to this, I am grateful for the hard work of the lab management team - Jan Ho, Jan V, Carolien, Olga, Henk, Viola, Marijke, and Sidney. Olga, thank you for helping me with the in-situ hybridization experiments. Guling, Tingting, and Rik your help in microtome was indispensable. Renze, thank you for helping me with flow cytometry. Thank all other Molbi members for all the wonderful time we had together.

My Chinese colleagues in Molbi. Tingting, Huchen, Defeng, Tian, Guiling, Xu, Peng, Jing, Wenkun, Jieyu, Zhichun……. Thank you for all the help, encouragement, and all the nice activities. My other Chinese companions in Wageningen: Jinbin, Haikun, Kaile, Yiqian, Xiao Lin, Jinling, Xiaomei, Tao Zhao, Zhang Cheng… Forgive me, as I cannot mention all of your names here. Thank you all for your help and the nice times we had together. Thank visiting scientists Dr. Laiye Qu, Dr. Kai Cheng, Dr. Xingchun Tang, Dr. Jianming Bian, and Dr. Yanming Deng, for your support, help, and criticism. I also want to thank Dr. Sanwen Huang for all the training and support provided during my studies in Beijing. And thank you to all other lab members for your support during that time.

I would like to give my sincerest acknowledgement to my collaborators. Prof. Katharina Pawlowski, thank you for hosting me as a guest in your group. The discussions we had helped me to learn and understand the world of Frankia. And I also thank you for sharing your house with me. I enjoyed all our evening discussions about science and life. Thanks to Dr. Minggang Wang and Prof Martijn Bezemer for giving me the opportunity to participate in your plant microbe interaction publication. To Dr. Fredy Altpeter, thank you very much for allowing me the opportunity to work in your lab without a Dr. title. Thanks to all the lab members in the Altpeter

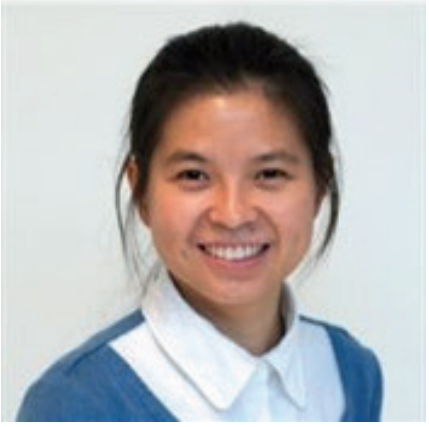
Lab. In particular, thank you Guangbin and Baskaran, for helping me to settle down in the lab. Sofia, I enjoyed your company very much while I wrote my thesis.

Lastly, Evelyn, thank you very much for your selfless and continuous support, company, and comfort. I will always remember how lucky I was to have you in my life at UF.

亲爱的爸爸妈妈, 谢谢你们一直以来给予我的无私的爱和鼓励！你们的爱给予我坚持的力量和克服困难的勇气！我永远爱你们！哥哥，谢谢你陪伴在爸妈身边，使我可以安心在外求学！感谢我所有的家人朋友对我的关爱！

28 Feb, 2020
UF, Gainesville, FL

Curriculum vitae



Fengjiao Bu was born on 4th Feb, 1988 in Hunan, China. She finished her Bachelor degree in 2010 at Hunan Agricultural University, followed by a Master's degree at the Institute of Vegetables and Flowers at the Chinese Academy of Agricultural Science, Beijing, in July 2013. The same year, she moved to Wageningen, Netherlands, to pursue a PhD with financial support from the Chinese Scholarship Council. There she worked on the only non-legume rhizobium host Parasponia, with the aim of understanding the evolution and genetics of nitrogen fixing root nodule symbiosis.

Publication list

Fengjiao Bu, Arjan van Zeijl, Sidney Linders, Marijke Hartog, Elena Fedorova, Robin van Velzen, Ton Bisseling, Rene Geurts. Rhizobium n*odS* gene mediated N-methylated LCOs is essential for triggering nodule organogenesis and intracellular infection on *Parasponia andersonii*. (manuscript in preparation)

Fengjiao Bu, Arjan van Zeijl, Luuk Rutten, Elena Fedorova, Ton Bisseling, Rene Geurts. Characterization of intergeneric hybrid between *Parasponia andersonii* and *Trema tomentosa* reveals new insight into engineering nitrogen fixing symbiosis. (manuscript in preparation)

Luuk Rutten, Kana Miyata, Yuda Roswanjaya, Rik Huisman, **Fengjiao Bu**, Marijke Hartog, Sidney Linders, Robin van Velzen, Ton Bisseling, Wouter Kohlen, Rene Geurts (under review). The duplication of two symbiotic LysM-receptors predates the evolution of Nitrogen fixing symbiosis.

Fengjiao Bu, Luuk Rutten, Yuda Roswanjaya, Olga Kulikova, Marta Rodriguez-Franco, Thomas Ott, Ton Bisseling, Arjan van Zeijl, Rene Geurts (2019) Mutant analysis in the non-legume *Parasponia andersonii* identifies NIN and NF-YA1 transcription factors as a core genetic network in nitrogen-fixing nodule symbioses. **New Phytologist.**

Robin van Velzen[#], Rens Holmer[#], **Fengjiao Bu**[*], Luuk Rutten[*], Arjan van Zeijl, Wei Liu, Luca Santuari, Qingqin Cao, Trupti Sharma, Defeng Shen, Yuda Roswanjaya, Titis A. K. Wardhani, Maryam Seifi Kalhor, Jolle Jansen, Johan van den Hoogen, Berivan Güngör, Marijke Hartog, Jan Hontelez, Jan Verver, Elio Schijlen, Rimi Repin, Menno Schilthuizen, M Eric Schranz, Renze Heidstra, Kana Miyata, Elena Fedorova, Wouter Kohlen, Ton Bisseling, Sandra Smit, and Rene Geurts (2018) Comparative genomics of the nonlegume Parasponia reveals insights into evolution of nitrogen-fixing rhizobium symbioses. **Proc. Natl. Acad. Sci. USA,** 115, E4700-E4709. ([#], [*] Equally contribution)

Minggang Wang, Wei-bin Ruan, Olga Kostenko, Sabrina Carvalho, S. Emilia Hannula, Patrick J. Mulder, **Fengjiao Bu**, Wim van der Putten, Martijn Bezemer (2018) Removal of soil biota alters soil feedback effects on plant growth and defense chemistry. **New Phytologist**.

Arjan van Zeijl, Titis A. K. Wardhani, Maryam Seifi Kalhor, Luuk Rutten, **Fengjiao Bu**, Marijke Hartog, Sidney Linders, Elena Fedorova, Ton Bisseling, Wouter Kohlen, Rene Geurts (2018) CRISPR/Cas9-Mediated Mutagenesis of Four Putative Symbiosis Genes of the Tropical Tree Parasponia andersonii Reveals Novel Phenotypes. **Front. Plant Sci**. 9, 284:1-14.

Fengjiao Bu[#], Huimin Chen[#], Qiuxiang Shi[#], Qian Zhou, Dongli Gao, Zhonghua Zhang, Sanwen Huang (2016). A major quantitative trait locus conferring subgynoecy in cucumber. **Theoretical and Applied Genetics**. 129,1: pp97-104. ([#] Equally contribution)

Zhonghua Zhang[#], Lin Mao[#], Huiming Chen[#], **Fengjiao Bu**[#], G Li[#], Jinjing Sun, Shuai Li, Honghe Sun, Chen Jiao, Rachel Blakely, Junsong Pan, Run Cai, Ruibang Luo, Yves Van de Peer, Evert Jacobsen, Zhangjun Fei, Sanwen Huan (2015). Genome-wide Mapping of Structural Variations Reveals a Copy Number Variant that Determines Reproductive Morphology in Cucumber. **The Plant Cell**. 27: 1595-1604. ([#] Equally contribution)

This work was performed in Laboratory of Molecular Biology, Wageningen University, with support from European Research Council (ERC-2011- AdG-294790), from a NWO-VICI grant (865.13.001), and from China Scholarship Council (201303250067).

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design: Vera van Beek

Layout: Fengjiao Bu

Printed by: ProefschriftMaken