Duplication in LysM-receptors predate nodulation

corresponding author: Rene Geurts rene.geurts@wur.nl

Duplication of symbiotic Lysin Motif-receptors predates the evolution of nitrogen-fixing nodule symbiosis

Luuk Rutten¹, Kana Miyata¹, Yuda Purwana Roswanjaya¹,², Rik Huisman¹, Fengjiao Bu¹, Marijke Hartog¹, Sidney Linders¹, Robin van Velzen¹,³, Arjan van Zeijl¹, Ton Bisseling¹, Wouter Kohlen¹ and Rene Geurts¹,*

# These authors contributed equally to this work.

1. Laboratory of Molecular Biology, Department of Plant Sciences, Wageningen University, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands. 2. Centre of Technology for Agricultural Production, Agency for the Assessment and Application of Technology (BPPT), Jakarta, Indonesia. 3. Biosystematics Group, Department of Plant Sciences, Wageningen University, 6708 PB, Wageningen, The Netherlands.

ONE SENTENCE SUMMARY:
Four lysin motif receptor kinases controlling rhizobium nodule formation in the non-legume Parasponia evolved after two ancient duplications.

AUTHOR CONTRIBUTIONS:

RG, LR and KM designed the research; LR, KM, YPR, RH, FB, MH, and SL performed research; LR, KM, YPR, RvV, WK and RG analysed data; and LR, KM, TB and RG wrote the manuscript.

FUNDING INFORMATION:

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, NWO-VENI grant.
Keywords:

nitrogen-fixing root nodules, Fabaceae, Parasponia, rhizobium, Nod factors, LysM-type receptor
Rhizobium nitrogen-fixing nodule symbiosis occurs in two taxonomic lineages: legumes (Fabaceae) and Parasponia (Cannabaceae). Both symbioses are initiated upon the perception of rhizobium-secreted lipo-chitooligosaccharides (LCOs), called Nod factors. Studies in the model legumes Lotus japonicus (lotus) and Medicago truncatula (medicago) showed that rhizobium LCOs are perceived by a heteromeric receptor complex of distinct Lysin Motif (LysM)-type transmembrane receptors named NOD FACTOR RECEPTOR1 (LjNFR1)-LjNFR5 (lotus) and LYSM DOMAIN CONTAINING RECEPTOR KINASE3 (MtLYK3)-NOD FACTOR PERCEPTION (MtNFP) (medicago). Recent phylogenomic comparative analyses indicated that the nodulation trait of legumes, Parasponia, as well as of so-called actinorhizal plants that establish a symbiosis with diazotrophic Frankia bacteria, share an evolutionary origin about 110 million years ago. However, the evolutionary trajectory of LysM-type LCO receptors remains elusive. By conducting phylogenetic analysis, trans-complementation studies, and CRISPR-Cas9 mutagenesis in Parasponia andersonii (parasponia), we obtained insight to the origin of LCO receptors essential for nodulation. We identified four LysM-type receptors controlling nodulation in parasponia: PanLYK1, PanLYK3, PanNFP1 and PanNFP2. These genes evolved upon ancient duplication events predating and coinciding with the origin of nodulation. Phylogenetic and functional analysis associated the occurrence of a functional NFP2-orthologous receptor to LCO-driven nodulation. Legumes and Parasponia use orthologous LysM-type receptors to perceive rhizobium LCOs, suggesting a shared evolutionary origin of LCO-driven nodulation. Furthermore, we found that both PanLYK1 and PanLYK3 are essential for intracellular arbuscule formation of mutualistic endomycorrhizal fungi. PanLYK3 also acts as a chitin oligomer receptor essential for innate immune signalling, demonstrating functional analogy to CHITIN ELECITOR RECEPTOR KINASE (CERK)-type receptors.
INTRODUCTION

Nitrogen availability is a critical factor for plant growth, but fixed nitrogen in the form of nitrate or ammonia in soils is limited. Plants have acquired different strategies to overcome this limitation. One such strategy is establishing a nodule endosymbiosis with nitrogen-fixing *Frankia* or rhizobium bacteria. Inside nodules, physiological conditions are created that allow the bacteria to convert atmospheric dinitrogen (N$_2$) into ammonia that can be used by the plant. Carbohydrates of plant origin fuel this energy demanding process. The unique character of nitrogen-fixing nodule symbiosis has raised the interest of plant researchers for more than a century, ultimately aiming to transfer this trait to non-leguminous crop species (Burrill and Hansen, 1917; Rogers and Oldroyd, 2014; Huisman and Geurts, 2019).

The *Frankia* and rhizobium nitrogen-fixing nodulation trait occurs in ten paraphyletic lineages within the orders Fabales, Fagales, Cucurbitales and Rosales, collectively known as the nitrogen-fixing clade (Soltis et al., 1995). Based on phylogenomic comparisons of nodulating and non-nodulating plant species, it is hypothesized that the nitrogen-fixing nodule symbiosis with rhizobium or *Frankia* bacteria has a shared evolutionary origin, dating to about 110 million years ago (Griesmann et al., 2018; van Velzen et al., 2018a; van Velzen et al., 2018b). Subsequently, the nodulation trait most probably was lost multiple times, which is associated with pseudogenization of two key genes essential for nodule organogenesis and bacterial infection; the transcription factor *NODULE INCEPTION* (*NIN*) and the coiled-coil protein-encoding gene *RHIZOBIUM POLAR GROWTH* (*RPG*) (Griesmann et al., 2018; van Velzen et al., 2018b). These two genes likely experienced genetic adaptations, allowing them to function exclusively in nodulation. However, insight into the evolutionary trajectory of signalling receptors involved in recognition of bacterial signals and subsequent activation of the pathways leading to nodule organogenesis and bacterial infection remains elusive.

The nitrogen-fixing nodulation trait is best studied in the legume models *Lotus japonicus* (lotus) and *Medicago truncatula* (medicago) (Fabaceae, Fabales). Both these legumes recognize their rhizobium microsymbionts by the structural characteristics of secreted lipo-chitoooligosaccharides (LCOs, also known as Nod factors). Perception of these molecules triggers nodule development (Wang et al., 2012). LCO signalling is also the basis of rhizobium-induced nodulation in the non-legume *Parasponia* (Cannabaceae, Rosales) (Marvel et al., 1987; Op den Camp et al., 2011; van Velzen et al., 2018b). Additionally, it was found that diazotrophic *Frankia* strains of a basal taxonomic lineage (so-called cluster-II strains) possess LCO biosynthesis genes, but the nodulating strains of two other taxonomic clusters do not (Pawlowski and Demchenko, 2012; Persson et al., 2015; Nguyen et al., 2016; Van Nguyen et al., 2019). LCOs, as well as chitin oligomers (COs), are
also used by arbuscular mycorrhiza (AM) fungi to signal their hosts (Maillet et al., 2011; Genre et al., 2013). Perception of these AM signals requires a plant LysM-type receptor that also is essential for chitin innate immune signalling; e.g. CHITIN ELECTOR RECEPTOR KINASE1 (OsCERK1) in rice (Oryza sativa) (Miyata et al., 2014; Zhang et al., 2015; He et al., 2019). This suggests that nodulating bacteria co-opted LCO signalling from the widespread AM symbiosis and/or innate immune signalling (Parniske, 2008; Gough and Cullimore, 2011; Geurts et al., 2012).

Genetic and biochemical studies in lotus and medicago demonstrated that rhizobium LCOs are perceived specifically by a heteromeric complex containing two distinct LysM-type receptors, named NOD FACTOR RECEPTOR1 (LjNFR1) and LjNFR5 in lotus, and LYSM DOMAIN CONTAINING RECEPTOR KINASE3 (MtLYK3) and NOD FACTOR PERCEPTION (MtNFP) in medicago (Limpens et al., 2003; Madsen et al., 2003; Radutoiu et al., 2003; Arrighi et al., 2006; Radutoiu et al., 2007; Broghammer et al., 2012). Other receptors may modulate the LCO response, such as LjNFRe, a homolog of LjNFR1 in lotus (Murakami et al., 2018). The LysM-type receptor family can be divided into two subclasses; named LYK and LYR, characterized by having a functional or dead kinase domain (Arrighi et al., 2006). Together these make up 11 orthogroups, two of which include legume LCO receptors (Buendia et al., 2018). Within legumes, the orthogroup that includes LjNFR1/MtLYK3 (named LYK-I clade) expanded upon gene duplications, allowing functional separation of rhizobium-induced signalling, AM symbiosis and chitin-triggered innate immune responses (De Mita et al., 2014; Bozsoki et al., 2017; Buendia et al., 2018; Gibelin-Viala et al., 2019). Likewise, the LjNFR5/MtNFP (orthogroup LYR-IA) experienced a gene duplication early in the legume clade (Young et al., 2011; Buendia et al., 2018).

Data on symbiotic LysM-type receptors in nodulating non-legumes are scarce. Only in Parasponia andersonii (parasponia) has a receptor functioning in nodulation been identified; named PanNFP1, which is a close homolog of LjNFR5/MtNFP (Op den Camp et al., 2011). Besides PanNFP1, Parasponia species possess a homologous receptor, named NFP2, which is more closely related to LjNFR5/MtNFP and transcriptionally activated in root nodules. Interestingly, this receptor is pseudogenized in non-nodulating Rosales species (van Velzen et al., 2018b). To obtain insight into the evolution of LysM-type LCO receptors that are essential for nodulation, we used parasponia as a comparative system to legumes. The genus Parasponia represents five tropical tree species, which form nitrogen-fixing nodules with LCO producing rhizobium species that also nodulate legumes (van Velzen et al., 2018b). Parasponia and legumes diverged at the root of the nitrogen-fixing clade >100 million years ago (Li et al., 2015; van Velzen et al., 2018a). The microbial symbionts of the ancestral nodulating plants remain elusive, and it is probable that Parasponia and legumes accepted rhizobium as a microbial partner in parallel (van Velzen et al., 2018a). In line with this, Parasponia
provides a unique comparative system to obtain insight into evolutionary trajectories of different LCO receptors that are essential for nodulation.
RESULTS

Phylogeny reconstruction of orthogroups representing LysM-type LCO receptors

To obtain insight into the LysM-type receptor family of parasponia, we analysed it phylogenetically. We identified 16 parasponia genes encoding putative LysM-type receptors that all grouped in known orthogroups except one (Figure S1; Table S1). Genetic studies in legumes uncovered only two orthogroups that contain proteins with a known function in rhizobium LCO signalling; these are named LYK-I and LYR-IA (Buendia et al., 2018). Parasponia has two gene copies in both these orthogroups.

LYK-I is the largest orthogroup, containing the functional legume LCO receptors *MtLYK3/LjNFR1* and *LjNFRe* (Limpens et al., 2003; Radutoiu et al., 2003; Murakami et al., 2018). Besides these, the LYK-I orthogroup also includes chitin innate immune receptors of medicago *MtLYK9/MtCERK1*, lotus *LjCERK6*, arabidopsis (*Arabidopsis thaliana*) *AtCERK1*, tomato (*Solanum lycopersicum*) *SlLYK1* and rice *OsCERK1* (Limpens et al., 2003; Miya et al., 2007; Wan et al., 2008; Shimizu et al., 2010; Miyata et al., 2014; Zhang et al., 2015; Bozsoki et al., 2017; Liao et al., 2018; Gibelin-Viala et al., 2019; He et al., 2019). *OsCERK1* and *MtLYK9/MtCERK1* have also been found to function in AM symbiosis (Miyata et al., 2014; Zhang et al., 2015; Feng et al., 2019; Gibelin-Viala et al., 2019). Two parasponia genes are part of this orthogroup; named *PanLYK1* and *PanLYK3*.

A more exhaustive phylogenetic reconstruction was conducted using gene orthologs of additional species to obtain insight into the evolutionary relationships of these genes when compared to LCO and CO receptors. Notably, LysM-type receptors of the recently sequenced nodulating actinorhizal plants and non-nodulating relatives were included (Griesmann et al., 2018). The resulting phylogeny largely resembled Rosid species trees as reconstructed on the basis of plastid-coding genes (Wang et al., 2009; Gonçalves et al., 2019). Our analysis revealed that *PanLYK1* and *PanLYK3* originated from an ancient duplication, dividing this orthogroup into two subgroups that we named LYK-Ia and LYK-Ib. This duplication does not coincide with the birth of the nitrogen-fixing clade, but rather has occurred in an ancestral eudicot (Figure 1; Data set S1). The only studied member in the LYK-Ia orthogroup is tomato *SILYK12*, and knockdown of this gene by virus-induced gene silencing (VIGS) substantially reduces mycorrhizal colonization (Liao et al., 2018). The LYK-Ib clade represents several functionally characterized genes, including the chitin innate immune receptors and legume rhizobium LCO receptors. Legumes exhibit an increased number of genes in the LYK-Ib subclade, which are the result of tandem duplications (Limpens et al., 2003; Radutoiu et al., 2003; Zhu et al., 2006). These duplications may have driven
neofunctionalization of LCO receptors in legumes (De Mita et al., 2014). In Parasponia, no gene

Figure 1. Phylogeny reconstruction of the LKY-I orthogroup, containing known CO and LCO receptors, based on 127 sequences from 47 species. Two main subgroups are recognized in Eudicots: LKY-la (blue) and LKY-lb (green). Note the presence of both variants in Aquilegia coerulea, a basal Eudicot in the Ranunculaceae. A subset of proteins is unresolved. Parasponia andersonii proteins are in bold. Parasponia and Trema LYK3.1 and LYK3.2 represent protein variants of LYK3.3. Deduced pseudo-proteins are depicted in grey/strikethrough. Proteins with known functions in nodulation, mycorrhization, and/or chitin-innate immune signaling are indicated. Bootstrap values indicate IQ-tree UF-bootstrap support%, values >80 are not shown. Scale bar represents substitutions per site. A complete list of species and accession numbers can be found in Data Set S1.
duplications have occurred in the LYK-Ib clade (represented by PanLYK3) nor in the LYK-Ia clade (represented by PanLYK1). In contrast, parasponia PanLYK3 experienced a duplication of exclusively the first exon. To determine whether this duplication is specific for the Parasponia...
We analysed the LYK3 genomic region of two additional *Parasponia* and three non-nodulating species of the closely related genus *Trema*. This revealed that the duplication of LYK3 exon 1 is present in all species investigated and occurred twice, where the most distal exon 1 copy was lost in parasponia (Figure 2A, Figure S2A). The encoded pre-mRNAs both splice into a shared second exon (Figure 2). Each exon 1 copy contains a putative transcription and translation start site, which allows for differential expression of the variants (Figure 2B-C). Genes of the LYK-I clade have a highly conserved intron-exon structure (Zhang et al., 2009). In most cases, the first exon encodes the extracellular domain comprising the signal peptide and three LysM motifs. So, the parasponia *PanLYK3* gene encodes two protein variants, named *PanLYK3.1* and *PanLYK3.2*, that differ in their extracellular domain (Figure S2B).

The LYR-IA orthogroup represents the legume LCO receptors MtNFP, LjNFR5 and pea (*Pisum sativum*) PsSYM10 (Madsen et al., 2003; Arrighi et al., 2006; Buendia et al., 2016; Miyata et al., 2016). Previously, we have shown that *Parasponia* species harbour two genes in this orthogroup, *PanNFP1* and *PanNFP2* in parasponia, of which the latter is more closely related to *MtNFP/LjNFR5* (van Velzen et al., 2018b). *PanNFP1* and *PanNFP2* originated from an ancient duplication. Phylogenetic reconstruction, including additional nodulating and non-nodulating species, supported the occurrence of NFP-I and NFP-II subclades in the LYR-IA orthogroup and showed that this duplication associates with the origin of the nitrogen-fixing clade (Figure 3; Data set S2). Several Actinorhizal species possess gene copies in both NFP subclades; including *Datisca glomerata*, *Dryas drummondii*, and *Ceanothus thyrsiflorius*. All these species nodulate with diazotrophic *Frankia* species of taxonomic cluster-II, which possess LCO biosynthesis genes. An NFP-II-type orthologous gene is notably absent in Actinorhizal species that are exclusively nodulated by *Frankia* species of cluster-I or cluster-III that lack LCO biosynthesis genes; e.g. *Alnus glutinosa* and *Casuarina glauca* (Figure 3) (Pawlowski and Demchenko, 2012; Griesmann et al., 2018; Salgado et al., 2018; Van Nguyen et al., 2019). In line with what was reported for the non-nodulating Rosales species (van Velzen et al., 2018b), NFP-II-type pseudogenes can be found in the genomes of the non-nodulating Fagales species *Castanea mollissima* and *Quercus fagus*. This shows a strict association of the presence of a functional NFP-II-type gene and LCO-driven nodulation, suggesting that the NFP-II subclade represents LCO receptors that function exclusively in nodulation.

*Parasponia PanNFP1, PanNFP2, PanLYK1 and PanLYK3 can perceive rhizobium LCOs*
Based on the orthologous relation to legume LCO receptors, we considered *PanLYK3* (both

---

**Figure 3.** Phylogeny reconstruction of LYRI-A orthogroup, containing known legume LCO receptors, based on 122 sequences from 87 species. A gene duplication in the root of the nitrogen-fixing clade is recognized, resulting in two subclades named NFP-I (blue) and NFP-II (green). The symbiotic capacities of the species are marked by filled (positive) and unfilled (negative) symbols: AM symbiosis (blue squares), ectomycorrhizal symbiosis (purple circles) and nodulation (green stars). *Parasponia PanNFP1* and *PanNFP2* are in bold. Deduced pseudo-proteins are depicted in grey strikethrough. Values indicate IQ-tree UF-bootstrap support %, values >98 are not shown. Branch support for the nitrogen-fixing clade indicates a SH-aLRT / UF-Bootstrap / approximate MrBayes support, respectively. Scale bar represents substitutions per site. A list of species and accession numbers can be found in Data set S2.
variants) and PanNFP2 as the most likely candidates to encode rhizobium LCO receptors in parasponia. We noted that, in contrast to PanLYK3, PanLYK1 is exclusively expressed in roots and nodule tissue (Figure 2B), suggesting this gene may also function in a symbiotic context. Therefore, we decided to include this gene in further studies. Finally, we included also PanNFP1, since an earlier study based on RNA interference (RNAi) in transformed parasponia roots showed that this gene functions in nodulation (Op den Camp et al., 2011). To test whether these four parasponia genes can function as rhizobium LCO receptors, we conducted two complementary experiments. First, we introduced parasponia receptor pairs into a lotus Ljnfr1;Ljnfr5 double mutant aiming to determine whether these parasponia receptors can trans-complement for LCO-induced Ca\(^{2+}\) oscillation. Second, we generated CRISPR-Cas9 knockout mutants in parasponia to study their role in nodulation.

We selected lotus for trans-complementation studies as its microbial host Mesorhizobium loti strain R7A can also nodulate parasponia (Figure S3A-C). By using A. rhizogenes-mediated root transformation, we tested six combinations of parasponia heterodimeric receptor pairs under control of the promoter and terminator of lotus LjNFR1 and LjNFR5 (Figure 4). These promoters showed to be functional in complementation of the lotus Ljnfr1-1;Ljnfr5-2 double mutant (Figure S3D-H). For the trans-complementation constructs, we included the nuclear localized calcium sensor R-GECO1.2, allowing visualization of nuclear Ca\(^{2+}\) oscillations (Zhao et al., 2011). In wild-type lotus roots, Ca\(^{2+}\) oscillation was most strong in young root hair cells, whereas this response is not recorded in the Ljnfr1-1;Ljnfr5-2 double mutant (Figure S3I,J; movie S1) (Miwa et al., 2006). Analysing the transgenic roots expressing parasponia receptor combinations revealed that nine out of eleven tested combinations elicit Ca\(^{2+}\) oscillation, although less regular in shape and frequency when compared to the positive control (Figure 4B; movie S2). Interestingly, the receptor combinations PanLYK1;LjNFR5 and LjNFR1;PanNFP2 did not elicit any Ca\(^{2+}\) oscillation response, whereas both parasponia receptors are -at least partially- functional as an M. loti LCO receptor when combined with a parasponia counterpart (Figure 4B). Upon inoculation with M. loti R7A, only nodule-like structures were observed on roots trans-complemented with different parasponia receptor combinations (4 weeks post-inoculation), but not with heterologous receptor pairs (Table S2). We sectioned the largest nodule-like structures, which were present on PanLYK3.2;PanNFP2 and PanLYK1;PanNFP1 transformed plants. This showed the absence of intracellular rhizobium infections (Figure S3K-P). Taken-together, the trans-complementation studies of a lotus Ljnfr1;Ljnfr5 mutant indicated that all four parasponia receptors -PanLYK1, PanLYK3, PanNFP1 and PanNFP2- have the potential to function as receptors for M. loti LCOs, but none could fully trans-complement a lotus Ljnfr1-1;Ljnfr5-2 double mutant for nodulation.
Parasponia PanNFP1, PanNFP2, PanLYK1 and PanLYK3 complement a lotus Ljnfr1::Ljnfr5 mutant for rhizobium-induced Ca\(^{2+}\) oscillation.

We recently established an efficient Agrobacterium tumefaciens-mediated transformation protocol for parasponia, which allows the generation of CRISPR-Cas9 mutant plantlets in a ~3 month timeframe (van Zeijl et al., 2018; Wardhani et al., 2019). This enabled us to test by mutagenesis whether PanLYK1, PanLYK3, PanNFP1 and PanNFP2 are essential for rhizobium-induced nodule formation. We aimed to generate small deletions of 100-300 bp in the area covering the LysM domains by using two or three single guide RNAs (sgRNAs) that have no potential high identity off-targets. In the case of PanLYK3 the transmembrane domain was targeted in order to mutate both alternative start variants. Additionally, we targeted specifically PanLYK3.1 and PanLYK3.2 by designing specific guides on the first exon. Selected single guides only had off-targets with at least three mismatches or two indels, based on alignments to the parasponia reference genome. Shoots regenerated after A. tumefaciens-mediated co-cultivation were genotyped using PCR and

**Figure 4.** Parasponia PanNFP1, PanNFP2, PanLYK1 and PanLYK3 function in nodulation

We recently established an efficient Agrobacterium tumefaciens-mediated transformation protocol for parasponia, which allows the generation of CRISPR-Cas9 mutant plantlets in a ~3 month timeframe (van Zeijl et al., 2018; Wardhani et al., 2019). This enabled us to test by mutagenesis whether PanLYK1, PanLYK3, PanNFP1 and PanNFP2 are essential for rhizobium-induced nodule formation. We aimed to generate small deletions of 100-300 bp in the area covering the LysM domains by using two or three single guide RNAs (sgRNAs) that have no potential high identity off-targets. In the case of PanLYK3 the transmembrane domain was targeted in order to mutate both alternative start variants. Additionally, we targeted specifically PanLYK3.1 and PanLYK3.2 by designing specific guides on the first exon. Selected single guides only had off-targets with at least three mismatches or two indels, based on alignments to the parasponia reference genome. Shoots regenerated after A. tumefaciens-mediated co-cultivation were genotyped using PCR and
subsequent sequence analysis to detect potential mutations at the CRISPR target sites. Only T0 shoots with a >75 bp deletion between the two target sites or edits generating a frameshift were considered for propagation and subsequent further evaluation. At least two independent mutant alleles were generated per gene, with the exception of Panlyk3.1 for which only a single suitable allele could be identified (Data set S3). Putative off-target sites that occur in coding sequence regions were amplified by PCR and subsequently sequenced by sanger sequencing. Subsequently, PanNFP1 was sequenced in PanNFP2 lines, and PanNFP2 in PanNFP1 lines. (Data set S3). No off-target mutations at these locations were identified. The selected tissue culture lines were propagated and rooted, so they could be used for experimentation.

We compared the nodulation phenotype of Panlyk1, Panlyk3, PanNFP1 and PanNFP2 knockout mutants in independent experiments, using empty vector (EV) transformed lines as control (Figure 5; Figure S4). All three independent PanNFP2 mutant lines showed to be unable to form nodules or nodule-like structures (5 weeks post inoculation, wpi) with strain Mesorhizobium plurifarium BOR2, demonstrating the requirement for this gene in the nodulation trait (Figure 5A). Additionally, we noted a reduced nodulation efficiency of all three independent PanNFP1 mutant lines. This is in line with earlier findings using RNAi to target PanNFP1 in A. rhizogenes-transformed parasponia roots (Op den Camp et al., 2011), demonstrating that PanNFP1 controls nodulation efficiency, but is not essential for rhizobium intracellular infection. Previously, we reported that PanNFP1 RNAi-nodules have a strong infection phenotype when inoculated with the Sinorhizobium fredii strain NGR234 (Op den Camp et al., 2011). We did not observe such an infection phenotype in nodules induced by M. plurifarium BOR2 on PanNFP1 knockout mutant plants (Figure S4). In order to determine whether the PanNFP1 infection phenotype is strain dependent, we nodulated plants, also with S. fredii NGR234. This strain showed to be less optimal under the chosen conditions (agroperlite supplemented with EKM medium and S. fredii NGR234,pHC60 at OD 0.05). In an effort to optimize nodulation efficiency with this strain, we used river sand and scored nodulation 8 weeks post-inoculation. Under these conditions, no difference between PanNFP1 and EV-control was observed. Nodules formed on PanNFP1 mutant plants were infected normally (Figure S4).

Similarly to PanNFP1 mutant plants inoculated with M. plurifarium BOR2, we found a reduced nodulation efficiency in parasponia Panlyk3 knockout mutants, but not in Panlyk3.1 and Panlyk3.2 variant specific mutant alleles, nor in Panlyk1 mutants (Figure 5; Figure S4). To determine whether nodules formed on Panlyk1 and Panlyk3 mutants have an infection phenotype, we analysed thin sections. In contrast to legumes, parasponia doesn’t guide rhizobia in infection threads towards the nodule primordia. Instead, rhizobia enter via apoplastic cracks in epidermis and cortex, and only
form infection threads to penetrate nodule cells. Once inside, infection threads develop into fixation threads, which are wider -having two phyla of bacteria aligned compared to one in infection threads- and possess a thinner cell wall (Lancelle and Torrey, 1984; Lancelle and Torrey, 1985). Panlyk1 mutant nodules showed no defects in infection thread structure or the transition from fixation threads to penetration threads.
infection threads to fixation threads. In the case of Panlyk3, nodules were relatively small and had diverse phenotypes. Out of 45 sectioned nodules of the line Panlyk3-e2, 22 were infected like wild type, 15 contained only infection threads, but no fixation threads, and 8 showed an intermediate phenotype with few infected cells (Figure 5 F-I, Figure S4). To confirm that the infection phenotype is a result of a full Panlyk3 knockout mutation, we sectioned 28 nodules of the independent knockout line Panlyk3-c3. This revealed similar results; 11 nodules normally infected, 11 contained only infection threads, and 6 nodules with an intermediate phenotype. Next, we determined whether this infection phenotype is controlled specifically by either PanLYK3.1 or PanLYK3.2, which showed not to be the case (Figure S4). As ~50% of the nodules formed on the parasponia Panlyk3 mutant plants displayed a wild-type phenotype, it suggests redundancy in gene functioning. Interestingly, S. fredii NGR234 could not nodulate Panlyk3 mutants, which suggest this strain is fully dependent on PanLYK3 controlled signal transduction (Figure S4).

As parasponia did not experience any gene duplication events in the LYK-Ib clade, PanLYK1 in the LYK-Ia clade is the closest homolog of PanLYK3. In order to investigate whether the PanLYK1 gene is functionally redundant with PanLYK3 in cases of M. plurifarium BOR2 inoculation, we generated a Panlyk1;Panlyk3 double mutant. To do so, a binary construct with the two sgRNAs targeting PanLYK1 was used for re-transformation of the Panlyk3 mutant (line a2). We obtained three independent Panlyk1;Panlyk3 mutants (data set S3). M. plurifarium BOR2 inoculation experiments revealed that all Panlyk1;Panlyk3 double mutant lines were unable to form any nodule or nodule-like structure (Figure 6). To confirm that the nodulation minus phenotype in the Panlyk1;Panlyk3 lines is not due to any off-target mutation, we conducted complementation studies using A. rhizogenes-mediated root transformation. As the putative promoter of PanLYK3 is rather complex due to the occurrence of alternative transcriptional start sites (Figure 2), we used the LjNFR1 promoter, as well as the constitutive AtUBQ10 and CaMV35S promoters, to drive a CRISPR-resistant allele of PanLYK3.1 (PanLYK3cr). Compound plants carrying transgenic roots expressing PanLYK3cr could be nodulated by M. plurifarium BOR2 (Figure S5). Together, this showed that in parasponia, PanLYK1 and PanLYK3 act redundantly in root nodule formation. (For complementation studies of Pannf2, see below).

The results demonstrate that parasponia PanLYK1, PanLYK3, PanNFP1 and PanNFP2 function in rhizobium LCO-driven nodulation. PanLYK3 and PanNFP2 are orthologous to legume LjNFR1/MtLYK3 and LjNFR5/MtNFP, indicating a shared evolutionary origin of LCO-driven nodulation in both taxonomic lineages. As PanLYK1 and PanLYK3 evolved from a duplication predating the emergence of the nitrogen-fixing clade, it suggests that LCO signalling is an ancestral function of these LYK-I receptors.
Figure 6. Parasponia *PanLYK1* and *PanLYK3* act redundantly in nodulation. 

(A) Average nodule numbers per plant in EV control 1 (n=11) and re-transformed *Panlyk3* a2 line (n=12) and *Panlyk1;Panlyk3* double mutant lines b3 (n=10), b7 (n=5) and b13 (n=10), 5 wpi with *Mesorhizobium plurifarium* BOR2. Data are represented as mean ± SE, dots represent individual data points. Letters denote statistical significance based on one-way ANOVA and Tukey post-hoc contrasts P>0.05. (B) Roots with nodules of EV control 1, 5 wpi with *M. plurifarium* BOR2. Scale bar: 5 mm. (C) Roots without nodules of the *Panlyk1;Panlyk3* double mutant (line b3) 5 wpi with *M. plurifarium* BOR2. Scale bar: 5 mm.

320

A repaired *Trema levigata NFP2* pseudogene, but not *PanNFP1*, can functionally complement a parasponia *nfp2* mutant.
PanNFP1 and PanNFP2 differ in expression pattern. Whereas both genes are expressed in root tissue, only PanNFP2 is upregulated in nodules (Figure 2) (Van Velzen et al., 2018b). We questioned whether the difference in symbiotic functioning between both genes is the result of regulatory evolution. To test this, we first identified a functional promoter region of PanNFP2. A. tumefaciens mediated transformation showed that a 2.75 kbps PanNFP2 upstream region can be used to functionally complement the parasponia Pannfp2 mutant when using a PanNFP2 CRISPR-resistant allele (PanNFP2cr). Two independent lines formed 7±7 and 4±1 nodules 5 weeks post inoculation with M. plurifarium BOR2 (Figure 7). However, when we used PanNFP1 driven by the PanNFP2 promoter, no trans-complementation of the parasponia nfp2 mutant phenotype was
observed. This suggests that there is a functional difference in the encoded PanNFP1 and PanNFP2 receptors.

Next, we questioned whether the nfp2 pseudogene as present in several non-nodulating Rosales species may have encoded a functional symbiosis receptor. To test this, we focussed on the the nfp2 pseudogene of Trema levigata, as it has only three mutations that cause a disturbance of the open reading frame (Figure 7). We repaired these three mutations, using PanNFP2 as a template, resulting in an engineered CRISPR resistant TleNFP2cr that encodes for a LysM-type receptor protein of 582 amino acids, similar to PanNFP2 of parasponia. We tested whether TleNFP2cr driven by the PanNFP2 promoter can trans-complement the parasponia Pannfp2 mutant. A.
*Pseudomonas tumefaciens* transformation resulted in two lines that can form functional root nodules 5 weeks post-inoculation with *M. plurifarium* BOR2. This supports the hypothesis that *T. levigata nfp2* encoded a functional symbiosis receptor prior to the pseudogenization of this gene.

**Parasponia PanLYK3 is essential for chitin triggered immune responses and controls AM symbiosis in coherence with PanLYK1**

Next, we aimed to determine whether the parasponia LysM-type receptors that control nodulation are also involved in other processes, as this may provide insights into ancestral functions of these genes. Some LysM-type receptors of the LYK-I clade are known to function in chitin-triggered immunity and/or the arbuscular mycorrhizal symbiosis; e.g. *lotus* *LjCERK6*, *medicago* *MtLYK9/MtCERK1*, *arabidopsis* *AtCERK1*, *tomato* *SILYK1* and *rice* *OsCERK1* (Figure 1) (Miya et al., 2007; Wan et al., 2008; Shimizu et al., 2010; Bozsoki et al., 2017; Liao et al., 2018; Feng et al., 2019; Gibelin-Viala et al., 2019a; He et al., 2019). Similarly, some experimental evidence using transient silencing assays indicated that LysM-type receptors of the LYR-IA clade function in mycorrhization, including parasponia *PanNFP1* (Op den Camp et al., 2011). In line with this, we aimed to confirm this phenotype in stable *Pannfp1* knockout mutants, and determine whether other parasponia symbiotic LysM-type receptors may function also in AM symbiosis and/or chitin-induced innate immunity signalling.

First, we investigated whether the parasponia LysM-type receptors mutants are affected in chitin-triggered immunity responses. To do so, two complementary assays were used; a chitin-induced ROS-burst production and MITOGEN-ACTIVATED PROTEIN KINASE3 (MAPK3) / MAPK6 phosphorylation assay. Chitin heptamers (CO7) effectively induced a ROS burst parasponia root segments at concentrations of ≥ 1 μM when incubated at 28°C, the regular growth temperature of *Parasponia* species. (Figure 8A, Figure S6B). To test whether ROS bursts can also be triggered by rhizobium LCOs, we used the extracts of *M. loti* R7A and *Rhizobium tropici* CIAT899. These two strains can nodulate parasponia but produce structurally different LCOs (López-Lara et al., 1995; Folch-Mallol et al., 1996). However, neither triggered a ROS burst in parasponia roots, (Figure S6A). To determine whether CO7-induced ROS bursts were associated with phosphorylation of parasponia MAPK3 and MAPK6 homologs, we used an anti–phospho-p44/42 HsMAPK antibody, which detects phosphorylated MAPK3 and MAPK6 of different plant species (Yamaguchi et al., 2013; Bozsoki et al., 2017). Parasponia possesses a single *PanMAPK3* and a single *PanMAPK6* gene, which each encodes a protein with a conserved Thr202/Tyr204 phosphorylation site (Figure S6C). Upon CO7 application (100 μM, 10 min.), a MAPK3/6 phosphorylation pattern can be
detected, which is not observed upon application of *M. loti* or *R. tropici* LCO extracts (Figure 8C; Figure S6D). Next, we determined whether parasponia LysM-type receptor mutants are affected in responses to chitin CO7 oligomers. *Pannfp1*, *Pannfp2* and also a newly created *Pannfp1;Pannfp2* double mutant showed a wild-type ROS-burst and MAPK3/6 phosphorylation profile (Figure S6; Data set S3). Similarly, the *Panlyk1* mutant showed a ROS burst and MAPK3/6 phosphorylation profile, as did wild-type root segments (Figure S6E,F). In contrast, parasponia *Panlyk3* mutant lines lacked a chitin triggered ROS-burst and showed no p44/42 MAPK phosphorylation (Figure 8). Individual exon knockout *Panlyk3.1* or *Panlyk3.2* mutants both showed ROS production and MAPK3/6 phosphorylation upon application of 100 μM CO7, however at reduced levels (Figure 8).
Taken together, these data show that *PanLYK3* - which is the only parasponia gene in the LYK-Ib clade is essential for chitin innate immune signalling in roots.

Studies in parasponia, tomato, medicago and rice revealed that LYR-IA and LYK-I putative orthologous genes have functions in AM symbiosis (Miyata et al., 2014; Zhang et al., 2015; Buendia et al., 2016; Miyata et al., 2016; Carotenuto et al., 2017; Liao et al., 2018; Feng et al., 2019; Gibelin-Viala et al., 2019; He et al., 2019). Interestingly, we noted that the NFP-I-type gene is pseudogenized in European beech (*Fagus sylvatica*) and Chinese chestnut (*Castanea mollissima*). Both species have lost AM symbiosis in favour of an ectomycorrhizal symbiosis (Figure 3) (Werner et al., 2018). We conducted an RNA-seq experiment on parasponia roots mycorrhized by *Rhizophagus irregularis* strain DOAM197198. Several marker genes for mycorrhization showed to be enhanced in expression in mycorrhized parasponia root samples; including *PanSTR1, PanSTR2, PanPT4, PanVPY, PanD27, PanRAD1* and *PanRAM1* (Figure S7). Also, this suggested that *PanNFP1* is expressed higher than *PanNFP2* under these conditions (Figure S7). However, no significant differential regulation of any of the studied LysM-type receptor encoding genes was detected between phosphate starved control roots and mycorrhized root samples (Figure S7). To determine whether parasponia symbiotic LysM-type receptors also function in AM symbiosis, we conducted three independent experiments using *in vitro* propagated mutant plantlets inoculated with 250 spores of *R. irregularis* DOAM197198. The average colonization and arbuscule formation frequency were scored 6 weeks post-inoculation. These experiments revealed substantial variation in mycorrhization efficiency between replicates, though no clear impaired AM symbiosis phenotype could be observed in any of the single mutants, including *Pannfp1*. Strikingly, *Panlyk1* showed a significant increase in colonization and arbuscule frequency (Figure S8ABC). Analysing both double mutants - *Pannfp1;Pannfp2* and *Panlyk1;Panlyk3* - revealed a strong AM symbiosis phenotype only in the latter (Figure 9, Figure S8). The fungal colonization of the *Panlyk1;Panlyk3* mutant was severely affected, with only a few infections observed. Confocal imaging of WGA-alexa488 stained roots showed that besides the level of colonization, also the morphology of the few arbuscules that were formed was affected in *Panlyk1;Panlyk3* plants. In wild type plants, many cortical cells were filled with arbuscules that were finely branched and occupied most of the cell. In contrast, the few hyphae that enter cortical cells in the *Panlyk1;Panlyk3* mutant were unable to form mature arbuscules, either because the fungus fails to switch to fine branching, or because a limited number of fine branches is made (Figure 9). As both *Panlyk1* and *Panlyk3* single mutant plants do not show such impaired mycorrhizal phenotype, we conclude that both genes function in conjunction to control mycorrhizal infection.
Taken together, these experiments revealed that PanLYK1 and PanLYK3 can function in multiple
processes, including rhizobium nodulation (PanLYK1 and PanLYK3), arbuscular mycorrhizal symbiosis (PanLYK1 and PanLYK3), and chitin innate immune signalling (PanLYK3). This suggests that no subfunctionalization of these receptors is required to allow functioning in the rhizobium nitrogen-fixing nodulation trait.
DISCUSSION

We used parasponia as a comparative system to legumes to obtain insight into the evolutionary trajectory of LysM-type rhizobium LCO receptors. By conducting phylogenetic analysis, trans-complementation studies in a lotus LCO receptor double mutant, and CRISPR-Cas9 mutagenesis in parasponia, we identified four LysM-type receptors that function in LCO-driven nodulation in a non-legume. Two of these, PanLYK3 and PanNFP2, are putative orthologs to known legume rhizobium LCO receptors LjNFR1/MtLYK3 and LjNFR5/MtNFP, respectively. As the Parasponia and legume lineages diverged early in the nitrogen-fixing clade (>100 MYA), the use of orthologous genes for rhizobium LCO perception supports the hypothesis of a shared evolutionary origin of LCO-driven nodulation. In contrast to legumes, symbiotic LysM-type receptors in Parasponia did not experience recent duplication events. Instead, the Parasponia symbiotic LysM-type LCO receptors evolved following two ancient duplications. We hypothesize that the PanNFP1-PanNFP2 duplication associates with the origin of the nitrogen-fixation clade, whereas in case of PanLYK1 and PanLYK3, the duplication occurred prior to the birth of the nitrogen-fixing clade. This makes it most probable that the capability of these receptors to perceive LCOs predates the evolution of the nitrogen-fixing nodulation trait.

Currently, the NFP1-NFP2 duplication cannot be precisely dated because legumes do not possess an NFP-I-type gene. This can be explained in two scenarios. (i) The NFP1-NFP2 duplication occurred in the root of the nitrogen-fixing clade, and subsequently, the NFP-I-type gene got lost in the Fabales lineage. (ii) The NFP1-NFP2 duplication occurred in an ancestor of the Fagales-Cucurbitales-Rosales lineages after the divergence of the Fabales order. The recent finding that ectopic expression of the NFP-type gene of two species outside of the nitrogen fixing clade (Petunia hybrida PhLYK10 and tomato SLYK10) can -at least partially- trans complement the medicago and lotus Mtnfp and Ljnfr5 mutant phenotypes demonstrates that LCO receptor functionality is ancestral to the NFP1-NFP2 duplication (Girardin et al., 2019). The putative promoters of PhLYK10 and SLYK10 show a nodule-enhanced expression profile similar to that reported for PanNFP2 (Girardin et al., 2019), which may support the second scenario, where the duplication of NFP1-NFP2 has occurred only after the divergence of the Fabales clade. However, for such a scenario, it is essential that Fabales represents the most basal lineage in the nitrogen-fixing clade. To date, this remains unknown. For example, a recent phylogenetic study suggests, although with limited statistical support, that Fabales is sister to Fagales (Koenen et al., 2019). The phylogenetic analysis presented here (Figure 3) suggests that the first scenario is most probable (aSH-aLRT/UF-Bootstrap/approximation with Mr.Bayes support 76.4/77/0.859). Additionally, we searched for amino acid motifs in NFP-I and NFP-II type proteins and found an indel region in

Downloaded on November 5, 2020.
legume and non-legume NFP-II type proteins that is distinct from NFP-I (Figure S9). This also supports the hypothesis that NFP1-NFP2 duplicated at the root of the nitrogen-fixing clade. However, additional experiments are needed to definitively reject either scenario.

Trans complementation studies in a lotus Ljnfr1;Ljnfr5 double mutant showed that parasponiaLCO receptors can only partially restore LCO signalling. This only partial complementation we did not anticipate, because of the shared microsymbiont M. loti that can nodulate parasponiaas well as lotus. One explanation for this limited functionality may be that such receptors function in larger multiprotein membrane domain complexes. In such a case, the parasponiaLCO receptors are not adapted to interact with associated lotus proteins. Additionally, legumes and Parasponia have diverged in the mode of rhizobium infection. Whereas rhizobium penetrates Parasponia roots apoplastically by crack entry, legumes are generally infected intracellularly via curled root hair cells. Phenotypic analysis of rhizobium infection in legumes suggests that a specific LCO receptor is involved in this process, the so-called entry receptor (Ardourel et al., 1994). Such entry receptors have not yet been fully characterized, but MtLYK3 may carry out such functions, as they control rhizobium infection (Limpens et al., 2003; Smit et al., 2007). It remains elusive whether such entry receptor functioning requires specific adaptations that did not occur in the Parasponia LYK3 ortholog.

We showed that an engineered T. levigata TleNFP2 receptor can functionally complement the parasponiai Pannfp2 mutant, whereas PanNFP1 cannot. This suggests that the NFP1 and NFP2 receptor proteins have functionally diverged. Based on the finding that NFP orthologous protein of P. hybrida (PhLYK10) and tomato (SILYK10) can complement lotus Ljnfr5 and medicago Mtnfp mutants, it can be hypothesized that in parasponia especially PanNFP1 has experienced protein adaptations. However, it should be noted that the trans complementation studies presented here were conducted using the native PanNFP2 promoter, whereas studies conducted with PhLYK10 and SILYK10 were conducted with CAMV35S (Girardin et al., 2019). Such overexpression may mask differences in substrate affinity and/or specificity, under which native transcriptional regulation are biologically relevant. Our data demonstrate that the ancestor of T. levigata possessed a NFP2 receptor that can function in nodulation.

Mutant analysis in legumes demonstrated that rhizobium nodulation co-opted elements of an AM signalling pathway, including the LRR-type transmembrane receptor kinase lotus SYMBIOTIC RECEPTOR KINASE (LjSYMRK)/medicago DOES NOT MAKE INFECTIONS2 (MtDMI2), the nuclear envelope located cation ion channels LjCASTOR, LjPOLLUX/MtDMI1, the nuclear localized CALCIUM CALMODULIN KINASE LjCCaMK/MtDMI3, and the transcription factor LjCYCLOPS/medicago INTERACTING PROTEIN OF DMI3 (MtIPD3) (Geurts et al., 2012).
However, in legumes, rhizobium and AM fungi were shown to have independent perception mechanisms to activate this common symbiosis signalling pathway. In lotus and medicago, these consist of LjNFR1-LjNFR5 / MtLYK3-MtNFP for rhizobium LCOs and MtLYK9/MtCERK1 for AM signals (Geurts et al., 2012; Feng et al., 2019; Gibelin-Viala et al., 2019). MtLYK3 and MtLYK9/MtCERK1 both belong to the LYK-Ib clade and evolved upon legume specific duplication events (Figure 1) (De Mita et al., 2014). The strong phenotype in AM and nodule symbioses of the parasponiaPanlyk1;Panlyk3 knockout mutant demonstrates that such subfunctionalization is not causal for the evolution of rhizobium LCO receptors. In parasponia, both receptors function in conjunction to control AM and rhizobium nodulation. Additionally, PanLYK3 acts as a chitin innate immune receptor. Such multifunctionality has also been reported for MtLYK9/MtCERK1 in medicago and OsCERK1 in rice, which function both in AM symbiosis and chitin innate immune signalling (Miyata et al., 2014; Carotenuto et al., 2017; Feng et al., 2019; Gibelin-Viala et al., 2019). As monocots did not experience the LYK-Ia / LYK-Ib duplication, it demonstrates that committing multiple functions in symbioses and innate-immunity was ancestral to species of the nitrogen fixing clade but functionally diverted in the legume lineage.

The presence of NFP-type genes (LYR-IA orthogroup) in species outside of the nitrogen-fixing clade associates with the ability to establish an AM symbiosis (Figure 3) (Delaux et al., 2014; Gough et al., 2018). However, corresponding mutants have only a relatively weak phenotype in AM symbiosis (Buendia et al., 2016; Miyata et al., 2016; Girardin et al., 2019). Upon duplication of this gene, the NFP-I and NFP-II subclades may have inherited the ancestral function. As both the parasponia PanNFP1 and PanNFP2 receptors can partially complement LCO-induced Ca$^{2+}$ oscillation in the lotus Ljnfr1;Ljnfr5 double mutant, it supports that receptors of the NFP-I and NFP-II clades can act as an LCO receptor, which may reflect the ancestral function. Our observation that the presence of a functional gene in the NFP-II clade strictly associates with LCO-based nodulation suggests that this gene was co-opted to function in this trait. The importance of this LysM-type LCO receptor in the nitrogen-fixing nodulation trait is underlined by the complete block of nodulation in knockout mutants in legumes (e.g. lotus Ljnfr5, medicago Mtnfp and pea Pssym10) and Parasponia (parasponia Pannfp2) (Madsen et al., 2003; Arrighi et al., 2006). As Parasponia and legumes diverged at the root of the nitrogen-fixing clade, it suggests that the adaptations in the NFP-II clade are ancient and may have coincided with the birth of the nodulation trait.

The NFP-I type gene retained -at least in part- its ancestral function, indicated by its presence in non-nodulating species in the nitrogen-fixing clade that can establish an AM symbiosis. In cases where AM symbiosis is replaced by an ectomycorrhizal symbiosis such as in Fagus sylvatica or...
Castanea mollissima, the NFP-I type gene pseudogenized. However, phenotypic studies in stable parasponia mutants could not support the functioning of PanNFP1 in AM symbiosis. These findings contradict our earlier observation that this gene functions in arbuscule formation (Op den Camp et al., 2011). The reason for this discrepancy may be due to the RNAi construct used, which may have off-target effects (van Velzen et al., 2018b). To determine whether this is the case, we have studied the expression of LysM-type RLK genes in two independent PanNFP1 RNAi experiments. This revealed significant knockdown not only of PanNFP1, but also PanNFP2, which can explain the strong rhizobium nodulation and infection phenotype as reported by Op den Camp et al. (2011). We also found variable expression levels of other LysM-RLKs, including PanLYK1 and PanLYK3, which may explain the reported mycorrhization phenotype on PanNFP1 RNAi roots (Figure S10). Studies presented here using CRISPR-Cas9 knockout mutant lines revealed substantial biological variation in mycorrhization efficiency of parasponia roots, which may have hindered the observation of minor quantitative AM symbiosis phenotypes. To rule out that PanNFP1 and PanNFP2 may function redundantly to control AM symbiosis, we analysed a Panfp1;Panfp2 double mutant. Also, these lines showed to be effectively mycorrhized. Therefore, we conclude that our current mutant phenotype analysis does not find support for essential functioning of parasponia PanNFP1 and PanNFP2 in AM symbiosis by.

The study presented here provided insight into the evolutionary trajectory of symbiotic LCO LysM-type receptors. By using parasponia as a comparative system to legumes, we revealed two ancestral duplications of LysM-type LCO receptors that predate, and coincide with, the evolution of nitrogen-fixing nodules. The strict association of genes in the NFP-II clade with LCO-driven nodulation strongly suggests that this gene was co-opted to function specifically in this symbiosis, making NFP2 a target in approaches to engineer LCO-driven nodulation in non-leguminous plants.

MATERIALS AND METHODS

LysM-type receptor phylogeny reconstructions

Orthogroups containing LysM-type receptor kinases of parasponia, generated in a previous study (van Velzen et al., 2018b), were combined and re-aligned into a single alignment using MafftV7.017. MrBayes3.2.6 was used to calculate phylogenetic relations under default parameters in Geneious R8.1.9 (Biomatters Ltd, UK) (Huelsenbeck and Bollback, 2001). Clades were named as published previously (Huelsenbeck and Bollback, 2001; Buendia et al., 2018). For clade LYK-I and LYR-IA additional putative orthologs were collected from Phytozome and NCBI databases using BLAST with AtCERK and MtNFP protein sequences as query (Table S1). Available Genomes from
Fabales, Fagales, Cucurbitales and Rosales species were downloaded and local BLAST analysis was conducted using Geneious R8.1.9 (Biomatters Ltd, UK) to search for additional unannotated LYK-I and LYR-IA protein sequences. Pseudogenes were annotated manually based on the closest functional ortholog so that a protein sequence could be deduced. Correct protein sequences were aligned using MAFFTV7.017 and subsequently manually curated. The deduced amino acid sequence was subsequently added to the alignment if the alignment length was at least 70% of the parasponia protein. Phylogenetic analysis was performed using IQ-tree (Nguyen et al., 2015; Trifinopoulos et al., 2016), running the modelfinder extension to find the best substitution models (Kalyaanamoorthy et al., 2017). Branch support analysis was done using Sh.aLRT 1000 replicates, UF-BOOTSTRAP support 1000 iterations (Kalyaanamoorthy et al., 2017; Hoang et al., 2018) and approximate Bayes support. Branch supports shown are UF-Bootstrap support%. Best fit model for the LYK-I clade: JTT+I+G4. Best fit model for LYR-IA clade: JTT+I+G4. Resulting tree files were loaded into Interactive Tree Of Life (iTOL) v3 for editing (Letunic and Bork, 2016). The analysis was run at least three times. Trees were rooted to outgroup angiosperm species Amborella trichopoda. UF Bootstrap Branch supports >98 were omitted for visual clarity. Gene names, accession numbers and alignment file of identified homologs can be found in Data set S1 for LYK-I and Data set S2 for LYR-IA, and Table S1 for parasponia.

LYK3 alignment and variant detection

Genomic LYK3 regions of parasponia, Parasponia rigida, Parasponia rugosa, Trema orientalis RG16, T. orientalis RG33, and Trema levigata were extracted from the respective assemblies (van Velzen et al., 2018b) and Aligned using MAFFTV7.0.17 implemented in Geneious R8.1. Coding sequences of parasponia, P. rigida, P. rugosa LYK3 protein variants were translated and aligned using MAFFTV7.0.17 implemented in Geneious R8.1 (Data set S1).

Vector constructs

All vectors generated for this study were created using golden gate cloning (Engler et al., 2009). Backbones and binary vectors were derived from the golden gate molecular toolbox (Engler et al., 2014). Parasponia LysM-type receptor cDNA clones were sequence synthesized as level 0 modules, including silent mutations in golden gate BsaI or BpiI restriction sites. Golden gate compatible clones of LjNFR1 and LjNFR5 promoters, CDS and terminators were obtained from Arhus University, Denmark. The calcium signalling reporter pLjUBQ1:R-GECo1.2 was published previously (Kelner et al., 2018). The generation and assembly of parasponia CRISPR constructs were done as published previously (van Zeijl et al., 2018). For hairy root transformation, a modified level 2 standard vector carrying spectinomycin instead of kanamycin resistance was created. All
sgRNAs were expressed using the AtU6 promoter. All Golden Gate binary vectors were verified by restriction digestion and DNA sequencing before transformation. A list of primers and constructs can be found in Table S3 and S4.

Genotyping and off-target analysis

All sgRNA targets were designed using the Geneious R10 CRISPR design tool, which picks targets on the principles described in Doench et al. (2014). To be selected Guide RNAs must have no potential target sites in the genome with (i) Less than three mismatches or (ii) less than two indels. Known off-target locations in CDS regions were PCR amplified and sequenced. No off-target mutations at these sites were detected. Genotypes and known off-target locations of CRISPR mutants used in this study can be found in Data set S3. Primers used for the creation of sgRNAs and subsequent sequencing of mutants and off-targets are listed in Table S4.

Bacterial strains

We used Mesorhizobium plurifarium BOR2 (van Velzen et al., 2018b) and Sinorhizobium fredii NGR234.pHC60 expressing GFP (Trinick and Galbraith, 1980; Cheng and Walker, 1998; Op den Camp et al., 2011) for parasponia inoculation experiments. M. loti R7A.pHC60 (Cheng and Walker, 1998; Sullivan et al., 2002) was used for lotus inoculations. M. loti R7A and Rhizobium tropici CIAT899 (Martínez-Romero et al., 1991) containing plasmid pMP604 (Spanik et al., 1989) were used for LCO extraction. A. rhizogenes strain AR10 (Hansen et al., 1989b; Martínez-Romero et al., 1991) was used for lotus root transformation. Agrobacterium tumefaciens strain AGL-1 (Lazo et al., 1991) was used in parasponia transformation. Agrobacterium sp. MSU440 was used for parasponia hairy root transformations (Cao et al., 2012). The Escherichia coli strain DH5α was used to propagate plasmids and in all subsequent cloning steps.

Rhizobium LCO isolation

To isolate rhizobium LCOs the plasmid pMP604 containing an auto-active NodD protein was introduced in M. loti R7A and R. tropici CIAT899 (Spanik et al., 1989; López-Lara et al., 1995). LCOs were extracted from a 750 ml liquid culture, OD600=0.5, grown at 28°C in minimal medium (5.75mM K₂HPO₄, 7.35mM KH₂PO₄, 5.9mM KNO₃, 460 nM CaCl₂, 37.5μM FeCl₃, 2.07mM MgSO₄, 20.5nM biotin, 2.9nM Thiamine HCl, 8.1nM Nicotinic acid, 4.8nM Pyridoxine-HCl, 2.8nM Myo-inositol, 4.6nM Panthotenate and 1% w/v sucrose) by the addition of 150 mL 1-butanol and 1h shaking. The butanol phase was transferred and subsequently evaporated (water bath 40°C). Pellet was dissolved in 75ml methanol, tested for Nod-factor activity and stored at -20°C for later
use. The concentration of active LCOs was estimated by using *LjNIN* induction in lotus wild type Gifu roots, 3h post-application. The lowest active dilution was estimated to be \(10^{-10}\) M.

**Lotus japonicus Agrobacterium rhizogenes root transformation**

Lotus *Ljnfr1-1;Ljnfr5-2* double mutants (Madsen et al., 2003; Radutoiu et al., 2003) were used for LysM complementation assays and ‘Gifu’ wild-type as control. Seedlings for *A. rhizogenes* root transformation were moved to fresh half-strength B5 medium and co-cultivated for 1 weeks as described previously using *A. rhizogenes* strain AR10 (Stougaard et al., 1987; Hansen et al., 1989a; Stougaard, 1995). During root emergence plants were grown on 1% agar plates half-strength B5 media containing 0.03% w/v cefotaxime and 1% w/v sucrose. Plants were screened for transformed roots using nuclear-localized R.GECO1.2 fluorescence. Shoots with transformed roots were grown in Agroperlite (Maasmond-Westland, Netherlands) supplemented with modified ½ Hoagland’s medium (Hoagland et al., 1950) containing 0.56 mM NH₄NO₃ and inoculated with *M. loti* R7A.pHC60 (expressing GFP) at OD600= 0.05. Plants were grown at 21°C under a 16h light/8h dark regime. For calcium oscillation analysis transformed plants were grown on ½ Hoagland’s plates with 1% agar containing 0.56 mM NH₄NO₃ for 1 week. Plants were moved to N-free ½ hoaglands medium 1 week prior to imaging.

**Calcium oscillation quantification**

Calcium spiking experiments were performed on a Leica TCS SP8 HyD confocal microscope equipped with a water lens HC plan-Apochromat CS2 40x/1.0. Transformed root segments expressing R-GECO1.2 were selected and incubated with 500x diluted LCO extract (estimated to represent \(10^{-9}\) M) in nitrate-free ½ Hoagland’s medium (Hoagland et al., 1950) on a glass slide with coverslip. Images were taken at 5s intervals for a minimum of 20 minutes per sample using an excitation wavelength of 552 nm and emission spectrum 585-620 nm. It is possible to monitor a large number of nuclei per root sample. However, only epidermal and especially root hairs showed to be responsive. Therefore, total nuclei numbers vary largely between samples. Video recordings of imaged root samples were exported to ImageJ1.50i (Collins, 2007). The Geciquant ImageJ plugin was used for background subtraction and region of interest (ROI) selection (Srinivasan et al., 2015). Average pixel intensity of ROIs (individual nuclei) were measured. Average pixel values (0-255) per nucleus were plotted and a background R-GECO1.2 fluorescence baseline of 2x 1 minute (2 regions of 12 frames) was selected manually in a region of the trace where no spikes were occurring. Only nuclei with a minimum of three spikes with an amplitude of over 1.5 times background were considered as positive.

**Parasponia growth conditions for propagation, transformation, mycorrhization and nodulation**

---

*Published by https://plantphysiol.org Copyright (c) 2020 American Society of Plant Biologists. All rights reserved.*
Sequenced parasponia WU1 trees or their direct descendants, were used in all experiments (Op den Camp et al., 2011; van Velzen et al., 2018b). Prior to transformation or transfer to tissue culture parasponia trees are grown in a conditioned greenhouse at 28°C, 85% humidity and a 16/8 h day/night regime. Parasponia in vitro propagation, transformation, CRISPR-Cas9 mutagenesis and nodulation assays were done according to Van Zeijl et al 2018 (van Zeijl et al., 2018). Parasponia hairy root transformations were performed according to Cao et al 2012 (Cao et al., 2012).

**Parasponia Nodulation assay and analysis**

Rooted tissue culture plantlets for phenotyping assays were grown in crystal-clear polypropelene containers (1 L), with a gas exchange filter (OS140BOX, Duchefa Biochemie, Netherlands). Pots were half-filled with agraperlite (Maasmond-Westland, Netherlands) and watered with modified EKM medium [3 mM MES (C6H13NO4) pH 6.6, 2.08 mM MgSO4, 0.88 mM KH2PO4, 2.07 mM K2HPO4, 1.45 mM CaCl2, 0.70 mM Na2SO4, 0.375 mM NH4NO3, 15 μM Fe-citrate, 6.6 μM MnSO4, 1.5 μM ZnSO4, 1.6 μM CuSO4, 4 μM H3BO3, 4.1 μM Na2MoO4] (Becking, 1983)]. For nodulation assays, modified EKM medium (Becking, 1983) was inoculated with rhizobia (OD600 = 0.025) prior to planting the shoots. For inoculation with strain *S. fredii* NGR234.pHC60, containers were half-filled with sterilized river sand and watered with modified EKM-medium containing the bacteria at an OD600 = 0.05.

All Nodules were fixed in buffer containing 4% w/v paraformaldehyde mixed with 3% v/v glutaraldehyde in 50 mM phosphate (pH = 7.4). A vacuum was applied for 2 hours during a total 48h incubation. Fixed nodules were embedded in Plastic, Technovit 7100 (Heraeus-Kulzer, Germany), according to manufacturer’s recommendations. Sections (5 μm) were made using a R12035 microtome (Leica Microsystems). Sections were stained using 0.05% w/v Toluidine Blue O. Images were taken with a DM5500B microscope equipped with a DFC425c camera (Leica microsystems).

**Parasponia mycorrhization assay**

For mycorrhization experiments, pots were half-filled with sterilized river sand, watered with modified ½ strength Hoagland’s medium containing 20 μM potassium phosphate. Pots were inoculated with 250 spores of *Rhizopagus irregularis* (Agronutrien-DAOM197198). In all experiments, plantlets in pots with closed lids were placed in a climate room at 28°C, 16/8 h day/night. Plants were watered with sterilized, demineralized water. Plants were harvested 6 weeks post inoculation with *Rhizopagus irregularis* (Agronutrien-DAOM197198). Root segments were treated with 10% (w/v) KOH and incubated at 90°C for 20 minutes. The root samples were then rinsed 6 times with water and stained with trypan blue at 90°C for 5 minutes. For each mutant, ten
plants were assessed and from each plant 30 root segments (each segment of approx. 1 cm long) were examined and mycorrhizal structures (hyphae, vesicles and arbuscules) were determined using the magnified line intersect method (Trouvelot, A, Kough J L, Gianinazzi-Pearson V, 1986) using a Leica CTR6000 microscope. For staining with WGA- Alexafluor 488 (Molecular Probes, Thermofisher Scientific, Waltham, MA, USA), roots were incubated in 10% (w/v) KOH at 60°C for 3 h. Then roots were washed three times in phosphate-buffered saline (PBS) (150 mM NaCl, 10 mM Na₂HPO₄ and 1.8 mM KH₂PO₄, pH 7.4), and incubated in 0.2 μg ml⁻¹ WGA-Alexafluor 488 in PBS at room temperature for 16 h. For RNA-isolation, parasponia WT plants were grown according to conditions above. RNA was isolated according to protocols published in (Op den Camp et al., 2011; van Velzen et al., 2018b). Mock inoculated plants were harvested as control. Three independent biological replicates were taken per sample. Expression was determined using RNA-seq. Reads were mapped using kallisto (Bray et al., 2016). Expression values and differential expression were determined using sleuth (Pimentel et al., 2017). Differentially expressed genes (Benjamini-Hochberg multiple testing corrected q-value <= 0.05)

**qPCR analysis of panNFPi cDNA samples.**

PanNFPi cDNA samples were generated previously (Op den Camp et al., 2011). qPCR was performed in 10 μl reactions using 2x iQ SYBR Green Super-mix (Bio-Rad, United States). PCR reaction was executed on a CFX Connect optical cycler, according to the manufacturer’s protocol (Bio-Rad, United States). Three technical replicates per cDNA sample were used. Data analysis and statistical analysis of biological replicates was performed using CFX Manager 3.0 software (Bio-Rad, United States). Gene expression was normalized against Reference genes PanACTIN and PanEF1alpha. Primers can be found in Table S4.

**ROS assay**

Parasponia plantlets were grown on rooting medium (van Zeijl et al., 2018) for 4 weeks at 28°C before the treatment. Roots, submerged in water, were cut into approximately 1 cm pieces. Each well of a black 96 well flat bottom polystyrene plate (Nunc) was filled with 10 root pieces. 10 replicates per line were analysed. After filling the wells, the plate was kept 5 hours in 28°C. After incubation, the water was replaced with 100 μl of assay solution containing 0.5 μM L-012 (FUJIFILM Wako Chemicals), 10 μg/ml Horseradish peroxidase (Sigma), and respective elicitors (; CO7 (ELICITYL) or LCOs extracted from *M. loti* or *R. tropici* at described concentrations. As a mock treatment, 100 μl H₂O was added. The light emission was immediately measured at 30 second intervals for 30 minutes, using a Clariostar multi well-plate reader. All data are the average of at least three independent biological replicates.
Protein extraction from Parasponia and western blotting

Parasponia plantlets were grown on rooting medium (van Zeijl et al., 2018) for 4 weeks at 28°C before the treatment. About 200mg of roots were cut while submerged in water and collected in a PCR-tube. Root segments were incubated for 5 hours at 28 °C before treatment. Root pieces were treated with water containing 100 μM CO7 (ELICITYL) for 10 min. After incubation, roots were immediately frozen in liquid nitrogen. Samples were homogenized using metal beads. Total root protein was extracted in a buffer containing 50 mM Tris·HCl (pH 7.5), 150 mM KCl, 1 mM EDTA (pH 7.5), 0.1% w/v Triton X-100, 1 mM DTT, complete protease inhibitors (Roche), and phosstop (Roche). Amounts of extracted protein were measured with Qubit (Thermo fisher Scientific) and equal amounts of protein ~20 μg were electrophoresed by Mini-PROTEAN TGX stain free gels (BIORAD). A Trans-Blot Turbo Transfer system was used for blotting. To visualize phosphorylated MPK3/MPK6, the antibody for anti–phospho-p44/42 MAPK was used (no. 4370; Cell Signalling Technology). Anti-rabbit antibody (no. 7054; Cell Signalling Technology) were used as secondary antibody. Equal loading was confirmed by CBB staining.

Quantification and statistical analysis

Nodule number was quantified as Mean nodule number ± SE for all experiments. Replicate number is denoted in figure or figure legend. Additionally, all individual data points were plotted for graphical visualization of variation. Graphs and statistical analysis were performed using R studio 1.1.456 for nodulation experiments. Statistical tests on nodule numbers was done using One Way Analysis of variance (ANOVA) and a Tukey post-hoc test for multiple comparisons. Statistical significance was defined as a p<0.05. Levenes test for homogeneity of variance was used prior to running a one-way ANOVA. In cases where normality assumption was violated, alternative tests such as Mann–Whitney–Wilcoxon (MWW) were used as denoted in the figure legends. For the mycorrhization experiment a standard linear model was used to estimate the difference, and the corresponding least significant differences (LSD), of the knockout mutants with the wild type control. The LSD with respect to the control was Bonferroni adjusted to correct for multiple testing.

ACCESSION NUMBERS

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers as mentioned in Table S1.
**SUPPLEMENTAL DATA**

**Supplemental Figure S1.** Phylogeny reconstruction of orthogroups representing LysM-type receptors.

**Supplemental Figure S2.** Duplication of the LYK3 first exon is conserved among *Parasponia* and *Trema* subspecies.

**Supplemental Figure S3.** Complementation of a lotus *Ljnfr1;Ljnfr5* mutant for LCO-induced calcium oscillation.

**Supplemental Figure S4.** Nodulation is affected in parasponia *Pannfp1, Pannfp2* and *Panlyk3* CRISPR-Cas9 mutants.

**Supplemental Figure S5.** Complementation of parasponia *Panlyk1;Panlyk3* double mutant.

**Supplemental Figure S6.** CO7 triggered ROS production and MPK phosphorylation in parasponia mutant lines.

**Supplemental Figure S7.** Expression of parasponia LYSM-type receptors during mycorrhization.

**Supplemental Figure S8.** Parasponia LysM-type receptor mutants can establish arbuscular mycorrhizal symbiosis.

**Supplemental Figure S9.** Conserved indel in NFP-II type receptor proteins.

**Supplemental Figure S10.** The PanNFPi RNAi construct has off target activity on *PanNFP2* and other LysM-type receptor kinases.

**Supplemental Table S1.** *Parasponia andersonii* LysM-type receptors.

**Supplemental Table S2.** Trans-complementation of lotus *Ljnfr1;Ljnfr5* for nodulation.

**Supplemental Table S3.** Constructs generated in this study.

**Supplemental Table S4.** Primers used in this study.

**Supplemental Data Set S1.** Sequence alignment of LYK-I type receptors in fasta format.

**Supplemental Data Set S2.** Sequence alignment of Lyr-Ia type receptors in fasta format.

**Supplemental Data Set S3.** Genotyping and off target analysis of parasponia CRISPR-Cas9 mutants generated in this study.
Supplemental Movie S1. Calcium spiking in root hairs of lotus *Ljnfr1-1;Ljnfr5-2* double mutant complemented with *LjNFR1;LjNFR5*.

Supplemental Movie S2. Calcium spiking in root hairs of lotus *Ljnfr1-1;Ljnfr5-2* double mutant trans-complemented with *PanLYK3.1;LjNFR5*.

ACKNOWLEDGEMENTS

We thank Jens Stougaard and Simona Radutoiu (Arhus University, Denmark) for sharing the *Lotus japonicus* *Ljnfr1-1;Ljnfr5-2* double mutant and the *LjNFR1* and *LjNFR5* constructs, Maryam Charpentier (John Innes Centre, United Kingdom) for sharing R-GECO1.2, Jinling Li and Bart Thomma (Wageningen University, The Netherlands) for the use of Clariostar multi-well reader and Joost van Heerwaarden (Wageningen University, The Netherlands) for statistical support. Furthermore, we would like to thank Arianne van Lierop, Jens van Kempen, Stan van der Wal and Vince Brouwer who contributed as undergraduate students to this research.
FIGURE LEGENDS:

Figure 1. Phylogeny reconstruction of the LYK-I orthogroup, containing known CO and LCO receptors, based on 127 sequences from 47 species. Two main subgroups are recognized in Eudicots, LYK-Ia (blue) and LYK-Ib (green). Note the presence of both variants in *Aquilegia coeralia*, a basal Eudicot in the Ranunculales. A subset of proteins is unresolved. Parasponia proteins are in bold. *Parasponia* and *Trema* LYK3.1 and LYK3.2 represent protein variants of LYK3. Deduced pseudo-proteins are depicted in grey/strikethrough. Proteins with known functions in nodulation, mycorrhization, and/or chitin-innate immune signalling are indicated. Bootstrap values indicate IQ-tree UF-bootstrap support%, values >98 are not shown. Scale bar represents substitutions per site. A complete list of species and accession numbers can be found in Data set S1.

Figure 2. Gene structure and expression of *Parasponia* *PanLYK3*

(A) Structure of the *PanLYK3* gene model and encoded proteins. *PanLYK3* possesses two transcriptional start sites resulting in two protein variants, which differ in the extracellular region containing the LysM-domains and are encoded by exon1. Red cross indicates a third upstream copy of exon1 lost in *Parasponia andersonii*, maintained in other *Parasponia* and *Trema* species. (B) Expression profile of *PanNFP1*, *PanNFP2*, *PanLYK1* and *PanLYK3* in different plant tissues. Expression is given in DESeq2-normalized read counts; error bars represent SE of biological replicates. Dots represent individual expression levels. The analysis is based on data presented in Van Velzen *et. al.* 2018 (van Velzen et al., 2018b). (C) Relative expression of the *PanLYK3.1* and *PanLYK3.2* transcriptional variants based on RNA-seq reads splicing into the second exon. Data are represented as mean ± SE (n=3). The analysis is based on data presented in Van Velzen *et. al.* 2018 (van Velzen et al., 2018b).

Figure 3. Phylogeny reconstruction of LYRI-A orthogroup, containing known legume LCO receptors, based on 122 sequences from 87 species. A gene duplication in the root of the nitrogen-fixing clade is recognized; resulting in two subclades named NFP-I (blue) and NFP-II (green). The symbiotic capacities of the species are marked by filled (positive) and unfilled (negative) symbols: AM symbiosis (blue squares), ectomycorrhizal symbiosis (purple circles) and nodulation (green stars). *Parasponia* PanNFP1 and PanNFP2 are in bold. Deduced pseudo-proteins are depicted in grey/strikethrough. Values indicate IQ-tree UF-bootstrap support%. values >98 are not shown. Branch support for the nitrogen-fixing clade indicates aSH-aLRT / UF-Bootstrap / approximate Mr.Bayes support, respectively. Scale bar represents substitutions per site. A list of species and accession numbers can be found in Data set S2.
Figure 4. Parasponia PanNFP1, PanNFP2, PanLYK1 and PanLYK3 complement a lotus Ljnfr1;Ljnfr5 mutant for rhizobium-induced Ca\(^{2+}\) oscillation.

(A) Schematic representation of the T-DNA region of the binary construct used for *A. rhizogenes* based root transformation of a lotus *Ljnfr1;Ljnfr5* double mutant. cDNA clones of LYK-I (marked blue) or LYR-Ia type genes (marked red) were cloned in identical fashion. cDNA clones were inserted between native promoter (marked black) pLjNFR1 (4,171bp) or pLjNFR5 (1,314bp), and native terminator (marked gray) sequences ter-LjNFR1 (394 bp) or ter-NFR5 (432 bp). pLjUBQ1::R.GECO1.2-nls:CaMV35S-ter (marked orange) was used to visualize nuclear calcium oscillation. The left border (LB) and right border (RB) (marked green) flank the T-DNA region. (B) Representative traces of nuclear Ca\(^{2+}\) oscillation, as observed in different combinations of LYK-I (red) and LYR-Ia (blue) type receptors introduced in a lotus *Ljnfr1;Ljnfr5* double mutant. Note that the receptor combinations PanLYK1;LjNFR5 and LjNFR1;PanNFP2 didn’t complement for Ca\(^{2+}\) oscillation. Traces were recorded ~10 min post-application of LCOs extracted from *M. loti* R7A (~10\(^{-9}\) M). Numbers denote spiking roots vs the number of roots analyzed. The number in brackets denotes the total number of spiking nuclei observed. Scale bar = 10 minutes. Y-axis is the relative fluorescence intensity compared to defined baseline in arbitrary units.

Figure 5. Parasponia Pannfp1, Pannfp2 and Panlyk3 mutants are affected in nodulation. Data are represented as mean ± SE, dots represent individual data points. Letters denote statistical significance based on one-way ANOVA and Tukey post-hoc contrasts *P*>0.05.

(A) Nodule numbers in parasponia CRISPR-Cas9 mutant lines *Pannfp1* d2 (n=18), k2 (n=31) and m1 (n=19) and *Pannfp2* b1 (n=19), e2 (n=10) and g1 (n=9), 5 wpi with *M. plurifarium* BOR2. EV Control 1 (n=12) represents a positive control line transformed with a binary vector not containing sgRNAs. (B) Nodule numbers in parasponia CRISPR-Cas9 mutant lines *Panlyk1* a1 (n=19) and b1 (n=20), 5 wpi with *M. plurifarium* BOR2. EV control 1 (n=14) and EV control 2 (n=14) represent two independent positive control lines transformed with a binary vector not containing sgRNAs. WT (n=20) represent untransformed plantlets. (C) Nodule numbers in parasponia CRISPR-Cas9 mutant lines *Panlyk3* a2 (n=21), c3 (n=21) and e1 (n=19), 5 wpi with *M. plurifarium* BOR2. EV Control 1 (n=14). (D,E,F,G,H,I) Toluidine blue-stained section of representative nodules grown with *M. plurifarium* BOR2 (D) Wild type parasponia transformed with an EV-1 construct expressing Cas9. Scale bar: 100 μm. (E) Infected nodule cells containing fixation threads formed on EV-1 plants. Scale bar: 20 μm. (F) Infected nodule of *Panlyk3* line e2. Note patches of infected cells. Scale bar: 100 μm. (G) infected nodule cells of the *Panlyk3* line e2 containing fixation threads. Scale bar: 20 μm. (H) Empty nodule of *Panlyk3* line e2. Note the absence of fully infected...
cells. Scale bar: 100 μm. (I) Nodule cells of the Panlyk3 line e2 containing infection threads but no fixation threads. Scale bar: 20 μm.

**Figure 6.** Parasponia PanLYK1 and PanLYK3 act redundantly in nodulation. (A) Average nodule numbers per plant in EV control 1 (n=11) and re-transformed Panlyk3 a2 line (n=12) and Panlyk1;Panlyk3 double mutant lines b3 (n=10), b7 (n=5) and b13 (n=10), 5 wpi with Mesorhizobium plurifarium BOR2. Data are represented as mean ± SE, dots represent individual data points. Letters denote statistical significance based on one-way ANOVA and Tukey post-hoc contrasts P>0.05. (B) Roots with nodules of EV control 1, 5 wpi with *M. plurifarium* BOR2. Scale bar: 5 mm. (C) Roots without nodules of the Panlyk1;Panlyk3 double mutant (line b3) 5 wpi with *M. plurifarium* BOR2 Scale bar: 5 mm

**Figure 7:** A repaired *Trema levigata* nfp2 pseudogene can replace *PanNFP2* for nodule formation. (A) Schematic representation of NFP2 coding region with indicated replacements to avoid CRISPR targeting of inserted *NFP2* genes of *P. andersonnii* (*PanNFP2cr*) and a repaired *T. levigata* (*TleNFP2cr*). Blue arrows: Guide RNA target sites. Red lines: *T. levigata* mutations.Region 1. Replacement of six codons at the sg1 site. Region 2. Replacement of five codons at the sg3 site plus repair of the *T. levigata* indel (red line). Region 3. Repair of the double stop codon in *Trema levigata*. (red line, black asterisks). The replacement of five codons at the sg2 site is not shown. (B) *PanNFP2cr* and repaired *TleNFP2cr* can restore nodulation in the *Pannfp2* mutant line C3 when driven by the *PanNFP2* promoter, whereas *PanNFP1* cannot. Nodulation scored 5 wpi with *M. plurifarium* BOR2. Error bars represent the SD of the mean, statistical significance by Mann–Whitney–Wilcoxon (MWW). P>0.05 not significant (ns), P<0.05 *, P<0.01 **, P<0.001, ***P <0.0001, ****P <0.00001. (C,D) Nodule and section of a *pNFP2:PanNFP2cr* line 1. (E,F) Nodule and section of *pNFP2:TleNFP2cr* line 1. (C,E) scale bar 2mm (D,F) scale bar 100μm.

**Figure 8.** Parasponia PanLYK3 is essential for chitin triggered immunity responses in roots. (A,B) Production of ROS measured upon treatment with 100 μM CO7 (filled circle) or H2O (open circle) with (A) EV control 1 plants (black), Panlyk3 line a2 (red) and Panlyk3 line e1 (orange) (B) EV control 1 plants (black), Panlyk3.1 line a4 (blue), Panlyk3.2 line a1 (red) and Panlyk3.2 line c1 (orange). For A and B, data are the average of at least three independent biological replicates ±SE. Luminescence is measured in relative light units RLU (C,D) Phosphorylation of mitogen activated protein kinase (MAPK) analysed by immunoblot using an anti-p44/42 MAPK antibody upon treatment with 100 μM CO7 (upper panel). Equal loading was confirmed by CBB staining (bottom panel).
panel). Results shown are a representative out of three independent experiments. (C) MAPK phosphorylation in root pieces of EV control 1, Panlyk3 line a2 and Panlyk3 line e1. (D) EV control 1, Panlyk3.1 line a4, Panlyk3.2 line a1 and Panlyk3.2 line c1.

Figure 9. Parasponia PanLYK1 and PanLYK3 act redundantly in arbuscular mycorrhization.

(A) The parasponia Panlyk1;Panlyk3 double mutant shows a strongly reduced colonization compared to wild type and control parasponia roots. Parasponia Pannfp1;Pannfp2 mutants are not significantly affected. Frequency and Arbuscule abundance classes according to Trouvelot et al. 1986 (Trouvelot, A, Kough J L, Gianinazzi-Pearson V, 1986). F%: colonization frequency in the root system. M%: intensity of mycorrhizal colonization A%: Arbuscule abundance in the root system. Error bars represent the SE of 10 biological replicates Scored at 6 weeks post inoculation using 250 spores of Rhizophagus irregularis strain DOAM197198 (Trouvelot, A, Kough J L, Gianinazzi-Pearson V, 1986). (B) Highly branched arbuscules formed in EV-Control plants 6 weeks PI stained with WGA-alexa488. Scale bar 10 μm. (C) Phenotype of stunted arbuscules formed in the Panlyk1;Panlyk3 double mutant stained with WGA-alexa488. Scale bar 10 μm. (D,E) Statistical analysis of raw (observed) data. (D) Mean colonization frequency score (classes 0 to 5) and (E) Mean arbuscule score (classes 0 to 3). Classes presented in Trouvelot et. al. (Trouvelot, A, Kough J L, Gianinazzi-Pearson V, 1986). Reduced mycorrhizal colonization and arbuscule formation in Panlyk1;Panlyk3 mutants is considered significant compared to wild type. Error bars represent the Bonferroni corrected Least Significant Difference. Error bars non overlapping with mean wild type value are considered significant. Dashed line indicates mean wild type score.
Ardourel M, Demont N, Debelle F, Maillet F, de Billy F, Promé JC, Dénarié J, Truchet G (1994) Rhizobium meliloti lipooligosaccharide nodulation factors: different structural requirements for bacterial entry into target root hair cells and induction of plant symbiotic developmental responses. Plant Cell 6: 1357–1374

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

Becking JH (1983) The Parasponia parviflora-Rhizobium symbiosis. Host specificity, growth and nitrogen fixation under various conditions. Plant Soil 75: 309–342

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. Proc Natl Acad Sci U S A 114: E8118–E8127

Bray NL, Pimentel H, Melsted P, Pachter L (2016) Erratum: Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34: 888

Buendia L, Girardin A, Wang T, Cottret L, Lefebvre B (2018) LysM Receptor-Like Kinase and LysM Receptor-Like Protein Families: An Update on Phylogeny and Functional Characterization. Front Plant Sci 9: 1531

Buendia L, Wang T, Girardin A, Lefebvre B (2016) The Lysm receptor-like kinase SILYK10 regulates the arbuscular mycorrhizal symbiosis in tomato. New Phytol 210: 184–195

Cheng HP, Walker GC (1998) Succinoglycan is required for initiation and elongation of infection threads during nodulation of alfalfa by Rhizobium meliloti. J Bacteriol 180: 5183–5191

Collins TJ (2007) Introduction to ImageJ for Light Microscopy. Microsc Microanal 13: 1674–1675

Cao Q, den Camp RO, Kalhor MS, Bisseling T, Geurts R (2012) Efficiency of Agrobacterium rhizogenes-mediated root transformation of Parasponia and Trem a is temperature dependent. Plant Growth Regul 68: 459–465

Carotenuto G, Chabaud M, Miyata K, Capozzi M, Takeda N, Kaku H, Shibuya N, Nakagawa T, Barker DG, Genre A (2017) The rice LysM receptor-like kinase OsCERK1 is required for the perception of short-chain chitin oligomers in arbuscular mycorrhizal signaling. New Phytol 214: 1440–1446

Cheng HP, Walker GC (1998) Succinoglycan is required for initiation and elongation of infection threads during nodulation of alfalfa by Rhizobium meliloti. J Bacteriol 180: 5183–5191

Collins TJ (2007) Introduction to ImageJ for Light Microscopy. Microsc Microanal 13: 1674–1675

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 Genet 10: e1004487
Li H-L, Wang W, Mortimer PE, Li R-Q, Li D-Z, Hyde KD, Xu J-C, Solitis DE, Chen Z-D (2015) Large-scale phylogenetic analyses reveal multiple gains of actinorhizal nitrogen-fixing symbioses in angiosperms associated with climate change. Sci Rep 5: 14023

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, Berg JDJ, Thomas-Oates JE, Glushka J, Lugtenberg BJJ, Spaink HP (1995) Structural identification of the iipo-chitin oligosaccharide nodulation signals of Rhizobium loti. Mol Microbiol 15: 627–638

Madsen EB, Madsen LH, Radutoiu S, Olbryt M, Rakwalska M, Szczygłowski 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

Marcel DJ, Torrey JG, Ausubel FM (1987) Rhizobium symbiotic genes required for nodulation of legume and nonlegume hosts. Proc Natl Acad Sci U S A 84: 1319–1323

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

Miya A, Albert P, Shinya T, Desaki Y, Ichimura K, Shirasu K, Narusaka Y, Kawakami N, Kaku H, Shibuya N (2007) CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in Arabidopsis. Proc Natl Acad Sci U S A 104: 19613–19618

Miyata K, Hayafune M, Kobae Y, Kaku H, Nishizawa Y, Masuda Y, Shibuya N, Nakagawa T (2016) Evaluation of the Role of the LysM Receptor-Like Kinase, OsNFR5/OsRLK2 for AM Symbiosis in Rice. Plant Cell Physiol 57: 2283–2290

Miyata K, Kozaki T, Kouzai Y, Ozawa K, Ishii K, Asanizu E, Okabe Y, Umehara Y, Miyamoto A, Kobae Y, et al (2014) The bifunctional plant receptor, OsCERK1, regulates both chitin-triggered immunity and arbuscular mycorrhizal symbiosis in rice. Plant Cell Physiol 55: 1864–1872

Murakami E, Cheng J, Gyssel K, Bozsoki Z, Kawaharada Y, Hjuler CT, Sørensen KK, Tao K, Kelly S, Venice F, et al (2018) Epidermal LysM receptor ensures robust symbiotic signalling in Lotus japonicus. Elife. doi: 10.7554/eLife.33506

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32: 268–274

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

Op den Camp R, Streng A, De Mita S, Cao Q, Polowe E, Liu W, Ahrendts JRUS, Kudrna D, Wang R, Unterlass A, et al (2011) LysM-type
mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume Parasponia. Science 331: 909–912

Parniske M (2008) Arbuscular mycorrhiza: the mother of plant root endosymbioses. Nat Rev Microbiol 6: 763–775

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

Persson T, Battenberg K, Demina IV, Vigil-Stenman T, Vanden Heuvel B, 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

Pimentel H, Bray NL, Puente S, Melsted P, Pachter L (2017) Differential analysis of RNA-seq incorporating quantification uncertainty. Nat Methods 14: 687–690

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

Radutoiu S, Madsen LH, Madsen EB, Jurkiewicz A, Fukai E, Quistgaard EMH, Albrektsen AS, James EK, Thirup S, Stougard J (2007) LysM domains mediate lipochitin–oligosaccharide recognition and Nfr genes extend the symbiotic host range. EMBO J 26: 3923–3935

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

Salgado MG, van Velzen R, Van Nguyen T, Battenberg K, Berry AM, Lundin D, Pawlowski K (2018) Comparative Analysis of the Nodule Transcriptomes of Ceanothus thyrsiflorus (Rhamnaceae, Rosales) and Datisca glomerata (Datiscaceae, Cucurbitales). Front Plant Sci 9: 1629

Shimizu T, Nakano T, Takamizawa D, Desaki Y, Ishii-Minami N, Nishizawa Y, Minami E, Okada K, Yamane H, Kaku H, et al (2010) Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice. Plant J 64: 204–214

Smit P, Limpens E, Geurts R, Fedorova E, Dolgikh E, Gough C, Bisseling T (2007) Medicago LKY3, an entry receptor in rhizobial nodulation factor signaling. Plant Physiol 145: 183–191

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. Proc Natl Acad Sci U S A 92: 2647–2651

Spaenk HP, Wijffelman CA, Okker RJ, Lugtenberg BE (1989) Localization of functional regions of the Rhizobium nodD product using hybrid nodD genes. Plant Mol Biol 12: 59–73

Srinivasan R, Huang BS, Venugopal S, Johnston AD, Chai H, Zeng H, Golshani P, Khakh BS (2015) Ca(2+) signaling in astrocytes from Ipr3r2(-/-) mice in brain slices and during startle responses in vivo. Nat Neurosci 18: 708–717

Stougard J (1995) Agrobacterium rhizogenes as a vector for transforming higher plants: Application in Lotus corniculatus transformation. Plant gene transfer and expression protocols 49–61

Stougard J, Abildsten D, Markert K. a. (1987) The Agrobacterium rhizogenes pRl TL-DNA segment as a gene vector system for transformation of plants. Mol Gen Genet 207: 251–255
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. J Bacteriol 184: 3086–3095

Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research 44: W232–W235

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

Trouvelot, A, Kough J L, Gianinazzi-Pearson V (1986) Mesure du taux de mycorhization VA d’un systeme radiculaire. Recherche de methodes d’estimation ayant une signification fonctionnelle. In GS Gianinazzi-Pearson V, ed, Physiological and Genetic Aspects of Mycorrhizae. INRA Press, Paris, pp 217–221

Van Nguyen T, Wibberg D, Vigil-Stenman T, Berckx F, Battenberg K, Demchenko KN, Blom J, Fernandez MP, Yamanaka T, Berry AM, et al (2019) Frankia-enriched metagenomes from the earliest diverging symbiotic Frankia cluster: They come in teams. Genome Biol Evol 11: 2273–2291

van Velzen R, Doyle JJ, Geurts R (2018a) A Resurrected Scenario: Single Gain and Massive Loss of Nitrogen-Fixing Nodulation. Trends Plant Sci. doi: 10.1016/j.tplants.2018.10.005

van Velzen R, Holmer R, Bu F, Rutten L, van Zeijl A, Liu W, Santuari L, Cao Q, Sharma T, Shen D, et al (2018b) Comparative genomics of the nonlegume Parasponia reveals insights into evolution of nitrogen-fixing rhizobium symbioses. Proc Natl Acad Sci U S A 115: E4700–E4709

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. Front Plant Sci 9: 284

Wang D, Yang S, Tang F, Zhu H (2012) Symbiosis specificity in the legume–rhizobial mutualism. Cell Microbiol 14: 334–342

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. Proc Natl Acad Sci U S A 106: 3853–3858

Wan J, Zhang X-C, Neece D, Ramonell KM, Clough S, Kim S-Y, Stacey MG, Stacey G (2008) A LysM receptor-like kinase plays a critical role in chitin signaling and fungal resistance in Arabidopsis. Plant Cell 20: 471–481

Wardhani TAK, Rosawanjaya 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. J Vis Exp. doi: 10.3791/59971

Werner GDA, Cornelissen JHC, Cornwell WK, Soudzilovskaia NA, Kattge J, West SA, Kiers ET (2018) Symbiont switching and alternative resource acquisition strategies drive mutualism breakdown. Proc Natl Acad Sci U S A doi: 10.1073/pnas.1721629115

Yamaguchi K, Yamada K, Ishikawa K, Yoshimura S, Hayashi N, Uchihashi K, Ishihama N, Kishi-Kaboshi M, Takahashi A, Tsuge S, et al (2013) A receptor-like cytoplasmic kinase targeted by a plant pathogen effector is directly phosphorylated by the chitin receptor and mediates rice immunity. Cell Host Microbe 13: 347–357
Young ND, Debellé F, Oldroyd GED, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KFX, Gouzy J, Schoof H, et al (2011) The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature 480: 520–524

Zhang X-C, Cannon SB, Stacey G (2009) Evolutionary genomics of LysM genes in land plants. BMC Evol Biol 9: 183

Zhang X, Dong W, Sun J, Feng F, Deng Y, He Z, Oldroyd GED, Wang E (2015) The receptor kinase CERK1 has dual functions in symbiosis and immunity signalling. Plant J 81: 258–267

Zhao Y, Araki S, Wu J, Teramoto T, -F. Chang Y, Nakano M, Abdelfattah AS, Fujiwara M, Ishihara T, Nagai T, et al (2011) An Expanded Palette of Genetically Encoded Ca2 Indicators. Science 333: 1888–1891

Zhu H, Riely BK, Burns NJ, Ané J-M (2006) Tracing nonlegume orthologs of legume genes required for nodulation and arbuscular mycorrhizal symbioses. Genetics 172: 2491–2499