

# Pseudogenization of the rhizobium-responsive *EXOPOLYSACCHARIDE RECEPTOR* in *Parasponia* is a rare event in nodulating plants

**Simon Dupin**

Vrije Universiteit Amsterdam

**Joël Klein**

Wageningen University & Research

**Luuk Rutten**

Wageningen University & Research

**Rik Huisman**

Wageningen University & Research

**Rene Geurts** (✉ [rene.geurts@wur.nl](mailto:rene.geurts@wur.nl))

Wageningen University & Research

---

## Research Article

**Keywords:** EXOPOLYSACCHARIDE RECEPTOR, EPR3, PARASPONIA, TREMA, NODULATION,

**Posted Date:** February 25th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1359063/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** Nodule symbiosis with diazotrophic *Frankia* or rhizobium occurs in plant species belonging to ten taxonomic lineages within the related orders Fabales, Fagales, Cucurbitales, and Rosales. Phylogenomic studies indicate that this nitrogen-fixing nodulation trait has a single evolutionary origin. In legume model plants, the molecular interaction between plant and rhizobium microsymbiont is mapped to a significant degree. A specific LysM-type receptor kinase, LjEPR3 in *Lotus japonicus* and MtLYK10 in *Medicago truncatula*, was found to act in a secondary identity-based mechanism, controlling intracellular rhizobium infection. Furthermore, LjEPR3 showed to bind surface exopolysaccharides of *Mesorhizobium loti*, the diazotrophic microsymbiont of *L. japonicus*. *EPR3* orthologous genes are not unique to legumes. Surprisingly, however, its ortholog *EXOPOLYSACCHARIDE RECEPTOR (EPR)* is pseudogenized in *Parasponia*, the only lineage of non-legume plants that nodulate also with rhizobium.

**Results:** Analysis of genome sequences showed that *EPR3* orthologous genes are highly conserved in nodulating plants. We identified a conserved retrotransposon insertion in the *EPR* promoter region in three *Parasponia* species, which associates with defected transcriptional regulation of this gene. Subsequently, we studied the *EPR* gene of two *Trema* species as they represent the sister genus of *Parasponia* for which it is assumed it lost the nitrogen-fixing nodulation trait. Both *Trema* species possess apparently functional *EPR* genes that have a nodulation-specific expression profile when introduced into a *Parasponia* background. This indicates the *EPR* gene functioned in nodulation in the *Parasponia-Trema* ancestor.

**Conclusion:** We conclude that nodule-specific expression of *EPR3* orthologous genes is shared between the legume and *Parasponia-Trema* lineage, suggesting an ancestral function in the nitrogen-fixing nodulation trait. Pseudogenization of *EPR* in *Parasponia* is an exceptional case in nodulating plants. We speculate that this may have been instrumental to the microsymbiont switch -from *Frankia* to rhizobium- that has occurred in the *Parasponia* lineage and the evolution of a novel crack entry infection mechanism.

## Introduction

The ability to engage in a nodule endosymbiosis with diazotrophic *Frankia* or rhizobium soil bacteria is a trait present in ten plant lineages within the taxonomic orders Fabales, Fagales, Cucurbitales and Rosales [1]. These four orders are commonly known as the nitrogen-fixing clade (NFC), but also represent multiple lineages of non-nodulating plants [2]. Recent phylogenomic studies indicated that the absent-present pattern of nitrogen-fixing root nodules in the NFC is the result of a single evolutionary gain of the nodulation trait, followed by multiple parallel losses [3–5]. In such a scenario, switches of microbial partners may have occurred.

A key feature of the nodulation trait is the potential to form a partnership with diazotrophic bacteria, in which the bacteria are carried to a newly formed root organ -the nodule- to establish an endosymbiosis.

This bacterial infection is typically supported by plant-derived tubular structures, called infection threads, that transport bacteria to dividing root cortical cells that form the nodule primordium. Finally, infection threads penetrate into nodule cells allowing bacteria to fill most of the cytoplasmic space. The plant host provides carbohydrates to symbiotic bacteria that then fix di-nitrogen gas (N<sub>2</sub>) to ammonia, which is metabolized by the plant.

Nodule formation relies on a complex cross-talk between plant and microbial partners. This cross-talk can vary in its specificity depending on which partners are involved. For example, in *Lotus japonicus* – an important legume model species - infection thread progression and cell infection are granted by recognition of compatible rhizobia surface exopolysaccharides (EPS) by the host's trans-membrane lysin motif (LysM) receptor kinase EXOPOLYSACCHARIDE RECEPTOR 3 (LjEPR3) [6, 7]. LjEPR3 harbours a singular configuration of its three LysM domains (LysM1-LysM2-LysM3) due to the atypical topology of LysM1 [6, 8]. As a result, the extracellular domain of LjEPR3 is specific to EPS and does not bind to fungal and rhizobia chitooligosaccharide and lipo-chitooligosaccharide signal molecules (COs and LCOs) [8]. Thus, LjEPR3 works as a secondary identity-based mechanism in the establishment of nitrogen-fixing nodule symbiosis between *L. japonicus* and its microsymbiont *Mesorhizobium loti*.

Studies on EPR3-type LysM receptors in species other than *L. japonicus* are limited. In *Medicago truncatula*, the *LjEPR3* ortholog *MtLYK10* is crucial for the progression of the infection thread to the nodule primordia. But recognition of succinoglycan -the surface EPS of the *M. truncatula* compatible microsymbiont *Sinorhizobium meliloti*- was not found [9]. EPR3-type receptors do occur also in non-nodulating plant species [8, 9], however, surprisingly was lost in the nodulating Cannabaceae species *Parasponia* [4].

*Parasponia* is the only taxonomic lineage outside the legume clade that can establish nitrogen-fixing root nodules with rhizobium. *Parasponia* represents five nodulating tropical tree species growing on volcanic islands of Indonesia and Papua New Guinea [10, 11]. *Parasponia* is closely related to the genus *Trema*, which includes 18 species that do not nodulate [4, 12]. Comparative analysis of *Trema* and *Parasponia* species showed that loss of the EPR3-type receptor *EPR* is specific to *Parasponia* species [4]. Here we aim to characterise the evolutionary trajectory of *EPR* in the *Parasponia* - *Trema* lineage. Specifically, we ask whether *EPR* may have functioned in nodulation in an ancestral *Parasponia* - *Trema* species, and how common loss of *EPR3* orthologous genes is in nodulating species. Furthermore, we discuss whether the loss of *EPR* in *Parasponia* was instrumental to the microsymbiont switch that occurred in this lineage.

## Results

### A retrotransposon insertion caused *epR* pseudogenization in *Parasponia* species

*Parasponia* represents five species, three for which genome sequence data have been generated; *P. andersonii*, *P. rigida*, and *P. rugosa*, respectively [4]. Earlier analysis revealed that these *Parasponia* species, as well as close relatives of the genus *Trema*, possess a single *LjEPR3/MtLYK10* orthologous

gene named *EXOPOLYSACCHARIDE RECEPTOR (EPR)*. *P. andersonii* and *P. rigida* *EPR* accumulated different mutations in the first exon causing a disruption of the predicted open reading frame (ORF), whereas *P. rugosa* *EPR* experienced a large deletion affecting exon 1 to 5 (table 1). As these mutations in *EPR* are not shared between the three *Parasponia* species, they must have occurred in parallel. This may suggest that the loss of *EPR* in *Parasponia* is the result of genetic erosion rather than specific selection. Alternatively, a shared, but yet unknown mutation occurred in the non-coding region of the gene affecting its functioning.

**Table 1: Independent mutations in the presumed coding region of *epr* pseudogene of three *Parasponia* species**

species gene name mutation in cds encoded protein GeneBank

species	gene name	mutation in cds	encoded protein	GeneBank
<i>P. andersonii</i>	<i>Panep</i>	TAA <sup>508-510</sup> stop codon	169 AA	KY786136.1
<i>P. rigida</i>	<i>Priepr</i>	A <sup>214</sup> insertion	72 AA	KY786146.1
<i>P. rugosa</i>	<i>Pruepr</i>	362-1217bp inframe deletion	326 AA	KY786178.1
<i>T. orientalis</i>	<i>TorEPR*</i>	no mutations	610 AA	JXTC01000021.1
<i>T. levigata</i>	<i>TleEPR</i>	no mutations	610 AA	KY786208.1

\*Note: *TorEPR* is named *TorLYK2* in GeneBank

To find evidence for this latter scenario, we investigated the putative promoter region of *EPR* in *Parasponia* and *Trema* species. In *L. japonicus* the functional promoter region of *LjEPR3* is relatively short, spanning only 329 bp upstream of the translational start codon [7]. We analysed the *EPR* promoter region in three *Parasponia* and two *Trema* species. The alignment of these promoters revealed a large 5,7 kb insertion in all three *Parasponia* species, just 154 bp upstream of the predicted translational start codon (Fig. 1A; Supplemental data file 1). Homology searches using BLAST revealed that this insertion represents a unique TY3-GYPSY-type retrotransposon element, which occurs only as a single copy in the genomes of the three *Parasponia* species, whereas it is absent in *Trema*. We compared the expression of the *P. andersonii* *epr* pseudogene to close homologs of the LysM-type receptor kinase (*LYK*) family [13]. This uncovered that in none of the samples, expression was observed, including roots and nodules at different stages of development (Figure S1). This further supports that the retrotransposon insertion in the putative regulatory region of *EPR* could have been instrumental for the pseudogenization of this gene in the *Parasponia* lineage.

## Trema EPR is expressed in rhizobium-induced nodules

In *L. japonicus*, the *LjEPR3* promoter possesses putative binding sites for the nodulation-specific transcription factors NIN and ERN1 [7]. We analysed the putative promoter region of *Trema* and *Parasponia EPR* using MEME combined with manual curation [14, 15]. This predicted the occurrence of conserved putative transcription factor binding sites for ERN1 (1x) and NIN (3x), both in *Trema* and *Parasponia EPR* promoters in a confined ~ 500bp region (Fig. 1B). This may suggest that transcriptional regulation of *EPR3* ortholog genes is conserved in legumes and non-legumes. As the putative NIN and ERN1 binding sites are present also in the *T. orientalis EPR* promoter, we questioned whether *Trema EPR* still possesses a nodule-enhanced expression profile, despite the loss of the nodulation trait.

To find support for the functioning of *EPR* in nodulation in a *Trema-Parasponia* ancestor, we generated transgenic *P. andersonii* lines carrying a *TorEPR* promoter GUS reporter construct (pTorEPR:*GUS*). As a putative promoter, a fragment of 1,730 bp upstream of the translational start codon was used, which includes the putative NIN and ERN1 binding sites. Two independent transgenic lines were studied. GUS staining of root tissue did not reveal any blue staining. Subsequently, plantlets (2x n = 10) were inoculated with the compatible strain *Bradyrhizobium elkanii* WUR3 [16] and studied 4 and 8 weeks post-inoculation. *TorEPR* promoter GUS activity was observed in rhizobium-induced cell divisions (Fig. 2A,B), which in *P. andersonii* occur in the root epidermis and outer cortical cell layers [17]. In mature nodules, pTorEPR:*GUS* induced blue staining is confined to the meristematic zones in the apex of the nodule (Fig. 2C-E). In both cases, the *GUS* expressing cells were yet to be infected by rhizobium.

To find additional support of *Trema EPR* expression in nodules, we studied gene expression in an intergeneric F1 hybrid of the cross *P. andersonii* x *Trema tomentosa*. Earlier studies showed such hybrid plants can be nodulated, but are hampered in hosting rhizobium intracellularly [4]. *T. tomentosa* is an allotetraploid. We analysed available genome sequence data and identified two *T. tomentosa EPR* genes, which were named *TtoEPRa* and *TtoEPRb* (Supplemental data file 2). Next, we studied *EPR* allele-specific expression in *P. andersonii* x *T. tomentosa* F1 hybrid roots and nodules. This revealed a nodule-specific expression of *TtoEPRa* and *TtoEPRb* whereas no expression of *P. andersonii EPR* was detected (Fig. 3).

Taken together, expression analysis of the *P. andersonii* x *T. tomentosa* F1 hybrid as well as *TorEPR:GUS* reporter studies in *P. andersonii* confirm that *Trema EPR* possesses essential cis-regulatory elements allowing nodule specific expression. This suggests that in a *Trema-Parasponia* ancestor, *EPR* functioned in nodulation.

## The loss of EPR in nodulating species is specific to the Parasponia lineage

In *L. japonicus* and *M. truncatula*, *LjEPR3* and *MtLYK10* commit essential functions in rhizobium infection, whereas in *Parasponia* the orthologous gene is pseudogenized. Earlier studies showed that also

in the legume *Aeschynomene evenia* the *LjEPR3/MtLYK10* orthologous gene is absent [18]. However, this species possesses a close paralog, possibly evolved as a result of a legume-specific duplication event that may commit a similar function [9, 18]. To determine whether loss of *EPR3* occurred more often in nodulating species, we analysed genome sequences of 34 species; 26 legumes (including *A. evenia*, *L. japonicus*, and *M. truncatula*), 7 actinorhizal plant species that nodulate with *Frankia*, and *P. andersonii*. In all species, 1 to 4 putative *ERP3* orthologous genes were identified. Many of these gene models have been predicted based on automated bioinformatics, without manual curation. As *LjERP3*, *MtLYK10* and *TorEPR/Panepnr* have a conserved gene structure consisting of 10 exons, we used these to manually curate the gene models in other species (Table S1). This revealed that all species investigated possess at least one gene copy that can encode a LysM-type receptor kinase that in length and structure is comparable to *LjEPR3/MtLYK10/TorEPR*. Subsequent phylogenetic reconstruction, based on a coding sequence alignment and using close paralogs LjLYS4, LjLYS5, MtLYK11, and PanLYK4 as an outgroup, supported the orthologous relation (Fig. 4; Supplemental data file 3). Also, it supports the occurrence of a duplication event in the legume Papilionoid subfamily, and the subsequent loss of one copy in the clade formed by *Cicer*, *Medicago*, *Trifolium*, *Vicia*, and *Pisum*. As all analysed plant genomes -except *Parasponia*- possess an *EPR3*-type gene, we conclude that loss of this gene in nodulating plant species is uncommon.

## Discussion

The legumes *L. japonicus* and *M. truncatula* use a specific LysM receptor kinase, namely LjEPR3 and MtLYK10, as a secondary identity-based mechanism to control rhizobium infection [6, 9]. We showed that the occurrence of *LjEPR3/MtLYK10* orthologous genes is highly conserved in nodulating plants, including actinorhizal plants that interact with diazotrophic *Frankia*. This suggests that a secondary identity-based mechanism allowing diazotrophic microsymbionts to infect is more generic. Strikingly, however, this mechanism seems to be mutated in *Parasponia* due to the pseudogenization of *LjEPR3/MtLYK10* orthologous gene *EPR*. This raises the question of why this gene was lost in the only non-legume lineage that nodulates with rhizobium?

First, we found that *epr* pseudogenization in the *Parasponia* lineage is associated with a unique retrotransposon insertion near the transcriptional start of the gene. Second, we studied the transcriptional regulation of the *EPR* gene in a nodulation context of the nearest non-nodulating sister species of *Parasponia*; namely *Trema* spp. It is hypothesized that *Trema* spp. lost the nodulation trait after the divergence of *Parasponia*, which is supported by the pseudogenization of several key nodulation genes in *Trema* [4]. We anticipated that *Trema EPR* may still possess the cis-regulatory elements critical for expression upon rhizobium-induced signalling since *T. orientalis EPR* contains putative ERN1 and NIN binding sites in its promoter region, similar as was reported for *L. japonicus LjEPR3* and *M. truncatula MtLYK10* [7, 9, 19–22]. Expression analysis using a *T. orientalis EPR* putative promoter GUS (pTorEPR:GUS) reporter construct in *P. andersonii* supported this view. The *TorEPR* putative promoter showed to be induced in the first dividing epidermal and cortical cells that occur upon rhizobium inoculation, whereas in the nodule expression is confined in the cell clusters that are about to be infected.

Likewise, enhanced expression of the *Trema EPR* alleles was found in nodules formed on *T. tomentosa* x *P. andersonii* F1 hybrid plants, whereas the *Parasponia* allele is not expressed. Together, these findings support the hypothesis that *EPR* committed a function in nodulation in the last *Trema-Parasponia* ancestor.

Loss of *EPR* in the *Parasponia* lineage may have been instrumental for the microsymbiont switch from *Frankia* to rhizobium. We speculated earlier that such a switch occurred at the base of the *Parasponia* lineage, based on evolutionary signatures in the nodule-specific haemoglobin gene [4, 23]. The *EPR* receptor of the nodulating *Parasponia-Trema* ancestor could have been co-evolved with its (ancestral) *Frankia* microsymbiont. Such *EPR* receptor may have hampered the interaction of ancestral *Parasponia* with rhizobium in a somewhat similar manner as observed in *L. japonicus* where *LjEPR3* hampers infection of the *M. loti* *exoU* exopolysaccharide mutant [6]. We initiated the first experiment to find evidence for this hypothesis and introduced the *T. orientalis EPR* gene in *P. andersonii* and quantified the nodulation efficiency of these transgenic lines. Though, despite nodule-specific expression of *TorEPR* (Figure S2), no phenotypes in nodulation were observed (Figure S3 and S4). This suggests that the effect of *trans TorEPR* in *P. andersonii* nodulation is only subtle and difficult to observe under given laboratory conditions. Alternatively, *trans TorEPR* in *P. andersonii* may contribute to the interaction potential to certain *Frankia* species. Although it is a tempting hypothesis, we considered it experimentally extremely challenging to prove. This is because we anticipate that the ancestral *Frankia* microsymbiont of the nodulating *Trema-Parasponia* ancestor most probably belonged to the taxonomic cluster-2 [13], which possess LCO biosynthesis genes [24–26]. *Frankia* cluster-2 strains are notoriously difficult to culture [27], and current (non-sterile) inocula only have limited compatibility with actinorhizal plants of the Southern hemisphere [26, 28, 29].

*M. truncatula* and *L. japonicus*, knock mutants in *Mtlyk10* and *Ljepr3* are affected in the progression of root hair-based infection threads [6, 9]. This results in a reduced number of trans-cellular cortical infection threads. In the *L. japonicus Ljepr3* mutant, successful infection often occurs from large intercellular pockets of bacteria from which subsequently cell penetration can occur (so-called peg infections) [7]. *Parasponia* does not support a root hair-based infection mechanism. Instead, rhizobium enters the roots by a novel crack entry mechanism, not found in legumes or actinorhizal plants. Upon inoculation, root epidermal and most outer cortical cells will divide. The newly formed daughter cells remain only loosely attached creating openings that are colonized by rhizobium [17, 30, 31]. From such an infection pocket, infection threads are formed to enter the nodule primordial cells. In comparison, crack entry infection is also found in some legume species. However, in these cases rhizobium exploits disruptions in the epidermis, e.g. due to later root emergence, rather than actively inducing the formation of such openings [32]. The evolution of this unique crack entry mechanism in *Parasponia*, may have coincided with the loss of *epr* and the acceptance of rhizobium as a microsymbiont.

Taken-together, this study highlights that *LjEPR3/MtLYK10* controlled secondary identity-based mechanism may predate the legumes, as cis-regulatory elements essential for a nodulation associated expression are present in the orthologous gene of *Trema*. Studies in *Parasponia* show, however, that the

occurrence of a LjEPR3/MtLYK10-orthologous gene is not essential to allow effective rhizobium nodulation.

## Materials And Methods

### Plant materials and nodulation

Experiments were conducted using *P. andersonii* WU1 or its offspring [33] and the interspecific hybrid *P. andersonii* x *T. tomentosa* line H9 [4]. *P. andersonii* plantlets used for nodulation experiment and qRT-PCR analysis were grown in 1 L clear polypropylene containers allowing for gas exchange (Duchefa Biochemie, The Netherlands). Pots were filled with agraperlite type 3 (Maasmond-Westland, The Netherlands), saturated with EKM nutrient solution (3mM MES (C<sub>6</sub>H<sub>13</sub>NO<sub>4</sub>) pH 6.6, 0.88 mM KH<sub>2</sub>PO<sub>4</sub>, 2.07 mM K<sub>2</sub>HPO<sub>4</sub>, 2.08 mM MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.7 mM Na<sub>2</sub>SO<sub>4</sub>, 0.375 mM NH<sub>4</sub>NO<sub>3</sub>, 1.45 mM CaCl<sub>2</sub>, 54 μM Fe-citrate, 6.6 μM MnSO<sub>4</sub>, 1.5 μM ZnSO<sub>4</sub>, 1,6 μM CuSO<sub>4</sub>, 4 μM H<sub>3</sub>BO<sub>3</sub>, 4.1 μM Na<sub>2</sub>MoO<sub>4</sub>). For nodulation assay, EKM was inoculated with *B. elkanii* WUR3 (OD<sub>600</sub> 0.05) [16]. Plants were placed in a conditioned climate room set at 28°C and a 16/8h day/night regime.

### Phylogenetic reconstruction

Protein sequences of publicly available genomes belonging to the Fabid clade (a taxonomic clade within the clade eurosids) were clustered into orthologous groups using Orthofinder (v2.5.1, Emms & Kelly, 2015). The orthogroup containing the EPR3 orthologues was extracted by searching for the *Lotus japonicus* LjEPR3 (Lj2g3v14154105) gene name. The EPR3 orthologous proteins were aligned using MUSCLE, and a phylogenetic tree based on this alignment was made using RAxML on the CIPRES Science Gateway version 3.3 [34]. EPR protein alignment was then used to manually curate the data set by assessing the protein model integrity based on the MUSCLE alignment. A protein model was scored complete when all three key EPS receptor domains were present (LysM motifs, trans-membrane and kinase domains). Otherwise, it was scored truncated when missing part of a domain or elongated when it had additional amino acids at the N or C terminal. Any species without a complete orthologous EPR protein model was annotated as a putative gene loss.

### Plant transformation

*Parasponia* stable transformation was conducted as described in Wardhani et al., 2019 [35]. For the *T. orientalis* promoter the 1,730 bp upstream region was cloned (Supplemental Data File 1) in a Golden Gate compatible level 0 clone (EC74289). This clone was subsequently used to assemble a pTorEPR:GUS level 2 binary vector EC74794. using the Moclo backbone pICH86966. As empty vector control, the binary vector EC74842 was used, expressing only a kanamycin resistance gene. Golden Gate constructs used in this study are listed in Table S2. Genotyping of transgenic lines was conducted using the Phire Plant Direct PCR Kit (Thermo Scientific, USA) and specific primers for *T. orientalis* pTorEPR. Amplicons were subsequently confirmed by sequencing.

### Microtome sectioning

Longitudinal sections of root nodules were made from root nodules 5 weeks post-inoculation. Plant tissue was fixed and embedded in technovit 7100 as previously described [35]. Thin sections (5 µm) were cut using a microtome (Leica Microsystems, Germany) placed on a glass slide and stained with 0.05% Toluidine blue for imaging. Pictures were taken using a Leica DM5500B microscope coupled with a DFC425C camera (Leica Microsystem, Germany).

## Library preparation and RNA sequencing

Nodules from *Parasponia* stable transformation lines containing the p*TorEPR:TorEPR trans* gene were harvested and flash-frozen in liquid nitrogen.

Frozen samples were homogenized for 2 min with a bead beater at 2000 rpm and the homogenized sample was immediately resuspended in modified RB buffer, 500µl RB buffer, 10 µl beta-mercaptoethanol and 50µl Plant RNA isolation aid (Thermo Fisher Scientific, USA). Then, RNA was extracted using the E.Z.N.A.® Plant RNA Kit (Omega Bio-tek, USA) following manufacturer recommendations. Library preparation and sequencing were performed by Novogene (Cambridge, UK). In short, the mRNA is fragmented randomly by adding a fragmentation buffer, then the cDNA is synthesized by using mRNA template and random hexamers primer. Samples were barcoded and pooled according to their effective concentration determined with qPCR and expected data volume. The resulting libraries were sequenced on the Illumina NovaSeq 6000s platform.

## RNA-seq data analysis and quantification

Paired-end 150 bp “raw reads” quality was assessed with FastQC (v0.11.9). Remaining adaptor sequences were removed from the reads and low-quality reads were filtered out (Q-score < Q20) with Trimmomatic (v0.39). Quantification of transcripts abundance was done with Kallisto (v0.46.1) [36] by pseudo-mapping the cleaned reads to *P. andersonii* reference genome [4, 36] with 500 bootstrap replicates other values were left to default. The Kallisto abundance files were loaded into R version 4.0.2 using tximport [37] and the transcript abundance was normalised using DESEQ2 version 1.34.0 [38].

A draft *T. tomentosa* genome assembly was soft masked using RepeatMasker version 4.0.7 [[39]. After softmasking *T. tomentosa* RNA-Seq samples (PRJNA388567) were aligned to the assembly using Hisat version 2.2.1. Gene models for this assembly were generated using BRAKER2 version 2.1.5 [40]. BRAKER2 was trained using the RNA-seq alignment and the proteome of *T. orientalis* and *P. andersonii*. Transcript abundance of the *T. tomentosa* x *P. andersonii* F1 hybrid was quantified using Kallisto version 0.46.2 [36] by simultaneously pseudo mapping with 500 bootstrap resampling; otherwise default values were used. Each RNA-seq sample was separately (PRJNA388743) mapped to the assembled genome of *P. andersonii* and *T. tomentosa*. The Kallisto abundance files were loaded into R version 4.0.2 using tximport [37] and the transcript abundance was normalised using DESEQ2 version 1.34.0 [38]. Differential expression analysis was done by combining each root and nodule sample and comparing the expression between root and nodules.

## Declarations

## **Ethics approval and consent to participate**

*Parasponia andersonii* seeds were obtained from P. Hadobas, and T. J. Higgins (as described in Op den Camp et al., (2011) *Science* DOI: 10.1126/science.1198181).

## **Consent for publication**

Not applicable.

## **Availability of data and materials**

RNA-seq datasets that were specifically generated and analysed during the current study are available in the NCBI SRA repository bioproject number PRJNA809142. Additionally, we analysed datasets available in bioprojects PRJNA272473 and PRJNA272482 generated by Van Velzen et al. 2018 [4]. Accession numbers of genes analysed can be found in Table 1.

## **Competing interests**

The authors declare no conflict of interest.

## **Funding**

This work was supported by NWO open competition grant to E. Toby Kiers (819.01.007) and the ENSA project funded by the Bill & Melinda Gates Foundation to the University of Cambridge (to R.G.).

## **Authors contribution**

Data presented in Figure 1: SD, JK, and RH, Figure 2: SD and LR, Figure 3: SD, JK and LR, Figure 4: JK and RG, Table 1: LR Figure S1: JK Figure S2-S4: SD, Table S1: JK and RG, Table S2: LR, Manuscript preparation: RG. All authors reviewed the manuscript.

## **Acknowledgements**

We thank Toby Kiers for feedback and lively discussions.

## **Authors' information**

Simon Dupin, Joël Klein, and Luuk Rutten contributed equally to this work.

Laboratory of Molecular Biology, Department of Plant Science, Wageningen University, Droevendaalsesteeg 1, 6703PB Wageningen, The Netherlands.

Simon Dupin, Joël Klein, Luuk Rutten, Rik Huisman, and Rene Geurts

2: Department of Ecological Science. Faculty of Earth and Life Sciences, Vrije Universiteit, De Boelelaan 1085, 1081HV Amsterdam, The Netherlands

## References

1. Doyle JJ. Phylogenetic perspectives on the origins of nodulation. *Mol Plant Microbe Interact.* 2011;24:1289–95.
2. Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JM, et al. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc Natl Acad Sci U S A.* 1995;92:2647–51.
3. Griesmann M, Chang Y, Liu X, Song Y, Haberer G, Crook MB, et al. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science.* 2018;361:eaat1743.
4. van Velzen R, Holmer R, Bu F, Rutten L, van Zeijl A, Liu W, et al. Comparative genomics of the nonlegume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proceedings of the National Academy of Sciences.* 2018;115:E4700–9.
5. van Velzen R, Doyle JJ, Geurts R. A Resurrected Scenario: Single Gain and Massive Loss of Nitrogen-Fixing Nodulation. *Trends Plant Sci.* 2019;24:49–57.
6. Kawaharada Y, Kelly S, Nielsen MW, Hjuler CT, Gysel K, Muszyński A, et al. Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature.* 2015;523:308–12.
7. Kawaharada Y, Nielsen MW, Kelly S, James EK, Andersen KR, Rasmussen SR, et al. Differential regulation of the Epr3 receptor coordinates membrane-restricted rhizobial colonization of root nodule primordia. *Nat Commun.* 2017;8:14534.
8. Wong JEMM, Gysel K, Birkefeldt TG, Vinther M, Muszyński A, Azadi P, et al. Structural signatures in EPR3 define a unique class of plant carbohydrate receptors. *Nat Commun.* 2020;11:3797.
9. Maillet F, Fournier J, Mendis HC, Tadege M, Wen J, Ratet P, et al. *Sinorhizobium meliloti* succinylated high-molecular-weight succinoglycan and the *Medicago truncatula* LysM receptor-like kinase MtLYK10 participate independently in symbiotic infection. *Plant J.* 2020;102:311–26.
10. Becking JH. The *Rhizobium* symbiosis of the nonlegume *Parasponia*. In: Stacey G, Burris RH, Evans HJ, editors. *Biological nitrogen fixation.* New York: Routledge, Chapman and Hall; 1992. p. 497–559.
11. Ishaq RM, Hairiah K, Alfian I, van Noordwijk M. Natural Regeneration After Volcanic Eruptions: Resilience of the Non-legume Nitrogen-Fixing Tree *Parasponia rigida*. *Frontiers in Forests and Global Change.* 2020;3:139.
12. Yang M-Q, Van Velzen R, Bakker FT, Sattarian A, Li D-Z, Yi T-S. Molecular phylogenetics and character evolution of Cannabaceae. *Taxon.* 2013;62:473–85.
13. Rutten L, Miyata K, Roswanjaya YP, Huisman R, Bu F, Hartog M, et al. Duplication of Symbiotic Lysin Motif Receptors Predates the Evolution of Nitrogen-Fixing Nodule Symbiosis. *Plant Physiol.* 2020;184:1004–23.
14. Bailey TL, Williams N, Misleh C, Li WW. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 2006;34 Web Server issue:W369–73.

15. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. *Nucleic Acids Res.* 2015;43:W39–49.
16. Op den Camp RHM, Polone E, Fedorova E, Roelofsen W, Squartini A, Op den Camp HJM, et al. Nonlegume *Parasponia andersonii* deploys a broad rhizobium host range strategy resulting in largely variable symbiotic effectiveness. *Mol Plant Microbe Interact.* 2012;25:954–63.
17. Shen D, Xiao TT, van Velzen R, Kulikova O, Gong X, Geurts R, et al. A homeotic mutation changes legume nodule ontogeny into actinorhizal-type ontogeny. *Plant Cell.* 2020;32:1868–85.
18. Quilbé J, Lamy L, Brottier L, Leleux P, Fardoux J, Rivallan R, et al. Genetics of nodulation in *Aeschynomene evenia* uncovers mechanisms of the rhizobium–legume symbiosis. *Nat Commun.* 2021;12:1–14.
19. Breakspear A, Liu C, Roy S, Stacey N, Rogers C, Trick M, et al. The Root Hair “Infectome” of *Medicago truncatula* uncovers changes in cell cycle genes and reveals a requirement for auxin signaling in rhizobial infection. *Plant Cell.* 2014;26:4680–701.
20. Van Zeijl A, Op den Camp RHM, Deinum EEE, Charnikhova T, Franssen H, Op den Camp HJM, et al. Rhizobium Lipo-chitooligosaccharide Signaling Triggers Accumulation of Cytokinins in *Medicago truncatula* Roots. *Mol Plant.* 2015;8 August:1213–26.
21. Jardinaud M-F, Boivin S, Rodde N, Catrice O, Kisiala A, Lepage A, et al. A Laser Dissection-RNAseq Analysis Highlights the Activation of Cytokinin Pathways by Nod Factors in the *Medicago truncatula* Root Epidermis. *Plant Physiol.* 2016;171:2256–76.
22. Liu C-W, Breakspear A, Guan D, Cerri MR, Jackson K, Jiang S, et al. NIN Acts as a Network Hub Controlling a Growth Module Required for Rhizobial Infection. *Plant Physiol.* 2019;179:1704–22.
23. Sturms R, Kakar S, Trent J, Hargrove MS. Trema and parasponia hemoglobins reveal convergent evolution of oxygen transport in plants. *Biochemistry.* 2010;49:4085–93.
24. Persson T, Battenberg K, Demina IV, Vigil-Stenman T, Heuvel BV, Pujic P, et al. Candidatus *Frankia datisca* Dg1, the actinobacterial microsymbiont of *Datisca glomerata*, expresses the canonical nod genes nodABC in symbiosis with its host plant. *PLoS One.* 2015;10:e0127630.
25. Nguyen TV, Wibberg D, Battenberg K, Blom J, Vanden Heuvel B, Berry AM, et al. An assemblage of *Frankia* Cluster II strains from California contains the canonical nod genes and also the sulfotransferase gene nodH. *BMC Genomics.* 2016;17:796.
26. Van Nguyen T, Wibberg D, Vigil-Stenman T, Berckx F, Battenberg K, Demchenko KN, et al. *Frankia*-enriched metagenomes from the earliest diverging symbiotic *Frankia* cluster: They come in teams. *Genome Biol Evol.* 2019;11:2273–91.
27. Gtari M, Ghodhbane-Gtari F, Nouioui I, Ktari A, Hezbri K, Mimouni W, et al. Cultivating the uncultured: growing the recalcitrant cluster-2 *Frankia* strains. *Sci Rep.* 2015;5:13112.
28. Silvester WB. Dinitrogen fixation by plant associations excluding legumes [Nitrogen]. *Treatise on Dinitrogen Fixation.* 1977.
29. Benson DR, Silvester WB. Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. *Microbiol Rev.* 1993;57:293–319.

30. Lancelle SA, Torrey JG. Early development of Rhizobium-induced root nodules of *Parasponia rigida*. I. Infection and early nodule initiation. *Protoplasma*. 1984;123:26–37.
31. Lancelle SA, Torrey JG. Early development of Rhizobium-induced root nodules of *Parasponia rigida*. II. Nodule morphogenesis and symbiotic development. *Can J Bot*. 1985;63:25–35.
32. Sprent JI. Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation. *New Phytol*. 2007;174:11–25.
33. Op den Camp R, Streng A, De Mita S, Cao Q, Polone E, Liu W, et al. LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume *Parasponia*. *Science*. 2011;331:909–12.
34. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: 2010 Gateway Computing Environments Workshop (GCE). [ieeexplore.ieee.org](http://ieeexplore.ieee.org); 2010. p. 1–8.
35. Wardhani TAK, Roswanjaya YP, Dupin S, Li H, Linders S, Hartog M, et al. Transforming, Genome Editing and Phenotyping the Nitrogen-fixing Tropical Cannabaceae Tree *Parasponia andersonii*. *J Vis Exp*. 2019. <https://doi.org/10.3791/59971>.
36. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;34:525–7.
37. Sonesson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res*. 2015;4:1521.
38. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
39. Smit AFA, Hubley R, Green P. 2013–2015. RepeatMasker Open-4.0. 2021.
40. Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP + and AUGUSTUS supported by a protein database. *NAR Genom Bioinform*. 2021;3:lqaa108.

## Figures

### Figure 1

**Gene structure of *Parasomnia* and *Trema* *EPR*.** (A) Gene structure of *L. japonicus* *LjEPR3*, *T. orientalis* *TorEPR*, and the *epr* pseudogene in *P. andersonii*, *P. rigida*, and *P. rugosa*. Note retrotransposon insertion (annotated in orange) in the putative promoter region of *Parasomnia* species. Promoter (light green), gene (green), CDS = coding DNA sequence (yellow). (B) LOGOs of putative NIN and ERN1 binding motifs in the promoter region of *T. orientalis* *TorEPR* and the *epr* pseudogenes of *P. andersonii*, *P. rigida*, and *P. rugosa*. TSS: translational start site.

## Figure 2

**temporal-spatial expression of *Trema orientalis* TorEPR promoter-GUS reporter constructs in *Parasponia andersonii*.** (A) GUS staining in young a nodule primordium. (B) Longitudinal section of a rhizobium-induced young nodule primordium formed in the outermost cell layers of the root. (C). GUS staining in a narrow zone in the apical region of a nodule. (D) Longitudinal section of a mature nodule with TorEPR:GUS activity in cells just below the meristem. (E) Enlargement of *pTorEPR:GUS* expressing cells in a nodule. Note absence of intracellular infection in cells showing TorEPR promoter activity.

## Figure 3

**Nodule-enhanced expression of *Trema* EPR in *Parasponia andersonii* (2n) x *Trema tomentosa* (4n) F1 hybrid plants.** Hybrid plants are triploid possessing two *T. tomentosa* gene (*TtoEPRa* and *TtoEPRb*) and one *P. andersonii* epr pseudogene (*Panep*). Expression is given in DESeq2 normalized read counts, error bars represent standard error of three biological replicates, dots represent individual expression levels.

## Figure 4

**Phylogeny reconstruction of the LjEPR3/MtLYK10 orthogroup of nodulating species.** Color coding indicates taxonomic orders; Cucurbitales (green, 1 species), Fagales (yellow, 3 species), Rosales (red, 4 nodulating species and two non-nodulating *Trema* species), Fabales (purple, 26 species). Tree rooted on outgroup (*L. japonicus* LjLYS4 and LjLYS5, *M. truncatula* MtLYK11, and *P. andersonii* PanLYK4). Asterisk indicates a duplication event in the legume Papilionoideae subfamily. Bootstrap values indicate IQ-tree UF-bootstrap support%; values. The scale bar presents substitutions per site. A complete list of species and accession numbers can be found in Table S1 and Supplemental Data file 3.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [DupinetalFigS1.pdf](#)
- [DupinetalFigS2.pdf](#)
- [DupinetalFigS3.pdf](#)
- [DupinetalFigS4.pdf](#)
- [DupinetalSupplementaldatafile1.docx](#)
- [DupinetalSupplementaldatafile2.docx](#)
- [DupinetalSupplementaldatafile3.docx](#)

- [DupinetalTableS1.xlsx](#)
- [DupinetalTableS2.docx](#)