Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A rare non-canonical splice site in Trema orientalis SYMRK does not affect its dual symbiotic functioning in endomycorrhiza and rhizobium nodulation

Sultan Alhusayni
King Faisal University
Yuda Purwana Roswanjaya
National Research and Innovation Agency
Luuk Rutten
Wageningen University & Research
Rik Huisman
Wageningen University & Research
Simon Bertram
Wageningen University & Research
Trupti Sharma
Wageningen University & Research
Michael Schon
Wageningen University & Research
Wouter Kohlen
Wageningen University & Research
Joël Klein (joel.klein@wur.nl)
Wageningen University & Research
Rene Geurts
Wageningen University & Research

Research Article

Keywords: Non-canonical splice site, SYMRK, LRR-type transmembrane receptor kinase, Mutualistic endosymbiosis, Nitrogen-fixing nodulation symbiosis, Arbuscular mycorrhizal symbiosis, Parasponia andersonii, Trema orientalis, Common symbiosis signalling pathway, plant evolutio

Posted Date: April 18th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2757645/v1

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

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Plant Biology on November 24th, 2023. See the published version at https://doi.org/10.1186/s12870-023-04594-0.

Abstract

Background

Nitrogen-fixing nodules occur in ten related taxonomic lineages interspersed with lineages of non-nodulating plant species. Nodules result from an endosymbiosis between plants and diazotrophic bacteria; rhizobia in the case of legumes and *Parasponia* and *Frankia* in the case of actinorhizal species. Nodulating plants share a conserved set of symbiosis genes, whereas related non-nodulating sister species show pseudogenization of several key nodulation-specific genes. Signalling and cellular mechanisms critical for nodulation have been co-opted from the more ancient plant-fungal arbuscular endomycorrhizal symbiosis. Studies in legumes and actinorhizal plants uncovered a key component in symbiotic signalling, the LRR-type SYMBIOSIS RECEPTOR KINASE (*SYMRK*). *SYMRK* is essential for nodulation and arbuscular endomycorrhizal symbiosis. To our surprise, however, despite its arbuscular endomycorrhizal symbiosis capacities, we observed a seemingly critical mutation in a donor splice site in the *SYMRK* gene of *Trema orientalis*, the non-nodulating sister species of *Parasponia*. This led us to investigate the symbiotic functioning of *SYMRK* in the *Trema-Parasponia* lineage and to address the question of to what extent a single nucleotide polymorphism in a donor splice site affects the symbiotic functioning of *SYMRK*.

Results

We show that *SYMRK* is essential for nodulation and endomycorrhization in *Parasponia andersonii*. Subsequently, it is revealed that the 5'-intron donor splice site of *SYMRK* intron 12 is variable and, in most dicotyledon species, doesn't contain the canonical dinucleotide 'GT' signature but the much less common motif 'GC'. Strikingly, in *T. orientalis*, this motif converted into a rare non-canonical 5'-intron donor splice site 'GA'. This *SYMRK* allele, however, is fully functional and spreads in the *T. orientalis* population of Malaysian Borneo. A further investigation into the occurrence of the non-canonical GA-AG splice sites confirmed that these are extremely rare.

Conclusion

SYMRK functioning is highly conserved in legumes, actinorhizal plants, and *Parasponia*. The gene possesses a non-common 5'-intron donor GC donor splice site in intron 12, which is converted into a GA in *T. orientalis* accessions of Malaysian Borneo. The discovery of this functional GA-AG splice site in *SYMRK* highlights a gap in our understanding of splice donor sites.

Background

Plants have evolved a range of mutualistic endosymbiotic partnerships with microbes to enhance nutrient uptake. The most ancient mutualistic endosymbiosis is the interaction between plant roots and Glomeromycota fungi, also known as arbuscular mycorrhizal (AM) fungi, which evolved over 400 million years ago [1]. Even today, AM endosymbiosis still occurs in ~ 72% of all higher plants [2]. Besides AM symbiosis, several plant lineages evolved additional or even alternative mutualistic endosymbiotic interactions, like orchid mycorrhiza, ericoid mycorrhiza, and diazotrophic rhizobia or *Frankia* bacteria hosted in root nodules. Interestingly, the evolution of these mutualistic endosymbiotic partnerships co-opted a signalling pathway critical for AM symbiosis. This pathway, known as the common symbiosis signalling pathway, is highly conserved and can be found in angiosperms, gymnosperms, monilophytes, and bryophytes [3].

The common symbiosis signalling pathway was first discovered in pea (*Pisum sativum*), showing to be critical for AM symbiosis and rhizobium-induced nodulation [4]. The subsequent molecular genetic characterisation in the legume models *Lotus japonicus* and *Medicago truncatula* revealed the pathway consists of four conserved components stretching from an LRR-type transmembrane receptor kinase down to the transcription factor *LjCYCLOPS/MtIPD3* [5, 6]. The LRR-type receptor kinase is generally named *SYMRK* (SYMBIOSIS SIGNALLING RECEPTOR KINASE), except for pea, *M. truncatula*, and *Medicago sativa*, where it is named PsSYM19, MtDMI2, and MsNORK, respectively [7]. The *SYMRK* extracellular structure varies between species, but in incase of eudicots possess a malectin domain, a conserved GDPC motif, and 2–3 LRR domains linked to a canonical intracellular serine-threonine kinase domain [7–10]. The malectin domain is cleaved in the absence of symbiotic signalling [11, 12]. Studies in *L. japonicus* showed that the remaining part of the SYMRK protein interacts with the LysM-type transmembrane receptor LjNFR5 [11, 12]. LjNFR5 is part of the receptor complex essential for recognising rhizobium-secreted lipo-

chitooligosaccharide (LCO) signal molecules [13, 14]. Legume *symrk* knockout mutants are blocked in rhizobium LCO-induced signalling through the common symbiosis signalling pathway. Subsequently, nodule formation is not initiated, nor is *Rhizobium* infection initiated in *symrk* mutants [7, 8, 15, 16]. *Lj*SYMRK also interacts with the innate immune receptor LjBAK1 (BRASSINOSTEROID INSENSITIVE 1-ASSOCIATED RECEPTOR KINASE 1), which may allow repression of immune responses upon symbiotic infection [17]. Such a role is supported by *symrk* mutant analysis, revealing fortification of the plant cell wall upon infection with *Glomus mosseae* AM fungus or rhizobium in mutant or RNAinterference (RNAi) lines [18, 19].

Studies on *SYMRK* in non-legumes are limited. RNAi Knockdown studies in the actinorhizal plants *Datisca glomerata* and *Casuarina glauca* showed that, like in legumes, *SYMRK* is essential for nodulation [9, 20]. These findings demonstrate that the common symbiosis signalling pathway defines a conserved genetic basis for nodulation with rhizobia or *Frankia*.

More recent phylogenomic studies support the hypothesis that the nodulation trait has a single evolutionary origin in the last common ancestor of the orders Fabales, Fagales, Cucurbitales and Rosales, representing all ten nodulating plant lineages [20–22]. The occurrence of non-nodulating lineages in these four taxonomic orders allowed the identification of nodulation-specific genes, as such genes are prone to pseudogenization from the moment a plant lineage lose the nodulation trait. We identified seven of such nodulation-specific genes by comparing nodulating *Parasponia* species to their non-nodulating sister species of the genus *Trema* [23]. Among these is an *NFR5* orthologous LysM-type receptor named *NFP2*, essential for nodulation in *Parasponia* [24]. To our surprise, however, we also identified a seemingly critical mutation in *SYMRK* of *Trema orientalis* (accession RG33; *TorSYMRK*^{RG33}), originating from the Sabah Provence in Malaysian Borneo [25]. It suggests that the *TorSYMRK*^{RG33} allele experiences pseudogenization, despite the fact *T. orientalis* accession RG33 can still establish an AM symbiosis [23].

TorSYMRK ^{RG33} has a conserved gene structure, though has a mutation in the conserved dinucleotide motif in the 5'-donor splice site of intron 12, converting this generally highly conserved dinucleotide motif into 'GA'. This led us to investigate the symbiotic functioning of *SYMRK* in the *Trema-Parasponia* lineage and investigate the impact of a seemingly critical SNP in an intron donor splice site in this gene.

Results

Trema orientalis **and** Parasponia andersonii **differ in** Rhizophagus irregularis **colonization**

Since *SYMRK* is known to be important for arbuscular mycorrhization in a range of species [7, 8, 20, 26], we first questioned whether *T. orientalis* accession RG33 can be effectively mycorrhized. To investigate this, we compared the mycorrhization dynamics of *T. orientalis* RG33 to *P. andersonii* (accession WU1). Both species are close relatives that diverged less than 20 million years ago [22], though have a somewhat different root architecture. *T. orientalis* plantlets have a shorter main root, whereas its lateral roots are longer when compared to *P. andersonii* (Figure S1).

To compare the mycorrhization efficiency, seedlings of both species were inoculated with 125 spores of *Rhizophagus irregularis* DOAM197198. Mycorrhization was quantified for 6 weeks, focussing on the frequency of mycorrhizal presence in the root (F%), the intensity of mycorrhization in infected root segments (M%), the arbuscule abundance in randomly selected segments (A%), and the averaged arbuscule abundance in infected root segments (a%) [27]. This revealed a clear difference in mycorrhization colonization dynamics between both species. The root system of *P. andersonii* is broadly colonized, showing an abundant presence of hyphae 4 weeks post-inoculation (F% >80%, M% >50%, Fig. 1A, C). In contrast, *T. orientalis* RG33 showed a reduced mycorrhizal infection and a low abundance of mycorrhizal hyphae in the root (F% <20%, M% <10%, 4 weeks post-inoculation) (Fig. 1A, D). These reduced mycorrhizal infection rates of *T. orientalis* RG33 were also reflected in a reduced number of arbuscules found in the root system (A%). However, when evaluating the infected root segments, the arbuscule abundance (a%) was comparable to *P. andersonii* (Fig. 1B). This indicates that *T. orientalis* RG33 is infected less frequently by *R. irregularis* DOAM197198 when compared to *P. andersonii*. But once infected, the number of arbuscules formed in the infected root segment is similar between both species.

Parasponia andersonii SYMRK **is essential for arbuscular mycorrhization and nodulation**

As *T. orientalis* RG33 can establish an arbuscular mycorrhizal symbiosis, we questioned whether *SYMRK* represents a single copy gene in the *Trema-Parasponi*a taxonomic lineage. We analysed genome sequences of 20 species representing monocots and major clades of dicots, including Fabales, Fagales, Cucurbiales, and Rosales species. The closest SYMRK paralogs of *P. andersonii* and *T. orientalis* were included as an outgroup. This revealed that *SYMRK* is a single-copy gene in the *Parasponia - Trema* lineage (Figure S2).

Mutant analysis In legumes and the actinorhizal species *C. glauca* and *D. glomerata* showed that *SYMRK* commits a dual role in establishing arbuscular mycorrhizal symbiosis and nodulation [7–9, 15, 16, 20, 28]. Furthermore, studies in *L. japonicus* revealed that ectopic expression of *LjSYMRK* results in the spontaneous onset of nodule organogenesis in absence of rhizobia [29]. To determine whether *SYMRK* in the *Parasponia-Trema* lineage fulfils a similar symbiotic role, two experiments were conducted. We generated CRISPR/Cas9 *symrk* knockout mutants in *P. andersonii* and conducted *PanSYMRK* ectopic expression studies in roots.

In total, three *Pansymrk* knockout mutant lines (homozygous line *Pansymrk-4* and the bi-allelic mutant lines *Pansymrk-5* and *Pansymrk-6*) were obtained by targeting the fourth and fifth coding exon using two single guide RNAs (sgRNAs) (Figure S3A). All mutant alleles represent large deletions, only encoding a fragment of the extracellular domain (Figure S3B). To determine whether *SYMRK* commits a key symbiotic function in *P. andersonii*, we first studied the nodulation phenotype of the *Pansymrk* mutants. *Pansymrk-4*, Pan*symrk-5* and Pan*symrk-6* plantlets were inoculated with *Mesohizobium plurifarium* BOR2, and the nodulation phenotypes were examined six weeks post-inoculation. The transgenic empty vector control plants (EV) were effectively nodulated, having nodule numbers ranging from 25 to 61 per plant. In contrast, the three *Pansymrk* mutant lines were unable to nodulate (Fig. 2A).

Next, we investigated the role of *PanSYMRK* in arbuscular mycorrhizal symbiosis. *Pansymrk-4, Pansymrk-5, Pansymrk-6*, and EV control plantlets were inoculated with an *R. irregularis* DAOM197198 spore suspension. Mycorrhization phenotypes were examined six weeks post-inoculation by quantifying four parameters; F%, M%, a%, and A%, as described above. The EV control plants interacted normally with the applied symbiont, with F%, M%, a%, and A% of 65,4%, 36,8%, 77,1%, and 26,1%, respectively (Fig. 2B, C). Although some intraradical hyphae were observed in a minority of the *Pansymrk* root segments (7 out of 417, 6 out of 760, and 9 out of 1085 segments) (Fig. 2B, D-F), generally, no arbuscules were observed in any of the tested *Pansymrk* mutant plantlets. This demonstrates that *SYMRK* is essential for nodulation and arbuscular mycorrhization of *P. andersonii* roots.

Next, we questioned whether the ectopic expression of *PanSYMRK* is sufficient for spontaneous formation of nodule-like structures. We employed *Agrobacterium rhizogenes*-mediated (*A. rhizogenes*) root transformation to introduce *PanSYMRK* driven by the *L. japonicus UBIQUITIN 1* (*LjUBI1*) promoter. This revealed spontaneous formation of nodule-like structures on roots ectopically expressing *PanSYMRK* (n = 5/25) (Fig. 3A-C). Longitudinal sections revealed that these nodule-like structures originate from dividing cortical and pericycle cells, similar to genuine *Parasponia* nodules (Fig. 3D). This led us to conclude that *SYMRK* is an essential key regulatory LRR-type receptor kinase for the onset of the nodule developmental program in the non-legume *P. andersonii.*

The GA mutation of the 5'-donor splice site of intron 12 doesn't affect SYMRK **functionally**

As *T. orientalis* RG33 -possessing a single *SYMRK* gene copy- can be mycorrhized effectively, it suggests that the *TorSYMRK*^{RG33} allele encodes a functional protein to support this plant-fungus symbiosis. Earlier studies in *M. truncatula* revealed that the *SYMRK* requirements differ between mycorrhizal colonization and rhizobium nodulation [7]. The *M. truncatula* R38 *dmi2* mutant possesses a missense mutation converting a glycine to glutamic acid mutation at position 794 of the protein [7]. This mutation affects the kinase phosphorylation activity and the capacity of the protein to interact with it's downstream target 3-HYDROXY-3-METHYLGLUTARYL COENZYME A REDUCTASE1 (MtHMGR1) [30, 31]. *M. truncatula* R38 *dmi2* is affected in nodulation but not in mycorrhization, suggesting an SYMRK functional kinase domain is less critical for the latter interaction [7]. As the *T. orientalis SYMRK*^{RG33} may encode -at least in part- a truncated SYMRK protein lacking essential domains of the kinase motif (Fig. 4A), we question to what extent this allele could function in nodulation.

To investigate this, first the native promoter region of *P. andersonii SYMRK* was identified. We used *A. rhizogenes* root transformation to show that a ~ 3 kb upstream region including the 5'-UTR driving the *PanSYMRK* gene functionally

complemented the *Pansymrk*-5 mutant (4.9 nodules/plant at 8 wpi) (Fig. 4B; Figure S4). Next, we used this promoter to drive a *PanSYMRK* gene mutant harbouring a GA at the donor site of intron 12, mimicking the *TorSYMRK*^{RG33} allele to determine its functionality in the *P. andersonii Pansymrk*-5 mutant background. Using *A. rhizogenes* root transformation, we found full complementation of the *Pansymrk* mutant phenotype (Fig. 4C-E; Figure S4). On average, 13 nodules per plant were formed at 8 wpi. Sections of these nodules revealed a wild type cytoarchitecture, including a large zone of cells possessing fixation threads. This shows that the GA point mutation at the donor site of intron 12 is not affecting *SYMRK* gene functionality.

A GA 5'-donor splice site is very rare, though effectively spliced in TorSYMRK ^{RG33}

We question how effective an intron that possesses a GA as the first two nucleotides of a donor splice site is spliced. To determine this, we aimed to compare the coverage of RNAseq reads of the 15 exons and 14 introns of the *SYMRK* gene of *T. orientalis* and *P. andersonii*. *SYMRK* is highly similar in both species, though introns show some variation in length (Table 1). *SYMRK* is known to be expressed in the root [8]. We grow *T. orientalis* and *P. andersonii* seedlings *in vitro* on a low nitrate medium and subsequently isolated 1 cm regions of roots just above the root meristemic zone. RNA extracted from these samples was sequenced (in triplicates), mapped, and analysed (Figure S5A). When focussing on intron 12, we found a per base mean coverage of 4.6 ± 1.0 for *TorSYMRK*^{RG33}, whereas in *P. andersonii*, the coverage of this intron is only 0.2 ± 0.3 mean per base coverage (Table 1). Comparing the splice site efficiency of intron 12, we observe that GA splice site in *T. orientalis* splices efficiently at approximately 95%, while the GT splice site in *P. andersonnii* shows an efficiency of 99.9%. This difference in intron retention between *PanSYMRK* and *TorSYMRK*^{RG33} was also observed by qRT-PCR on root mRNA (Figure S5B). These data suggest that *SYMRK* intron 12 is spliced less efficiently in *T. orientalis* when compared to *P. andersonii.* However, a similar variance is observed for other introns, which possess canonical donor and acceptor splice sites; e.g. *PanSYMRK* intron 11 (Table 1), suggesting some intron retention is not hampering gene function. Therefore, we conclude that SYMRK^{RG33} is fully functional, despite a non-canonical GA dinucleotide motif in the donor splice site.

Table 1

Splicing efficiency of Trema orientalis **and** Parasponia andersonii SYMRK **in the root susceptible zone for symbiotic engagement.**

SYMRK	*T. orientalis*	root susceptible zone		*P. andersonii*	root susceptible zone	
	length (bp)	coverage	s.d.	length (bp)	coverage	s.d.
exon1	99	28.1	8.9	99	94.6	39.6
intron1	765	0.0	0.1	912	0.1	0.1
exon2	525	35.2	13.2	525	96.5	17.6
intron2	339	0.0	0.0	341	0.1	0.2
exon3	475	51.2	9.1	475	114.6	17.1
intron3	351	0.0	0.0	364	0.0	0.0
exon4	150	59.3	23.4	150	127.0	24.9
intron4	188	0.1	0.1	175	0.1	0.2
exon5	71	89.2	27.6	71	166.1	29.4
intron5	103	0.3	0.2	102	1.0	0.7
exon6	68	84.4	20.7	68	133.1	22.2
intron6	202	2.7	0.5	319	0.2	0.2
exon7	71	70.7	18.3	71	125.0	13.2
intron7	133	0.0	0.0	134	0.3	0.4
exon8	71	62.2	16.9	71	138.8	15.5
intron8	97	0.9	1.1	86	0.1	0.1
exon9	193	76.1	16.1	193	142.9	22.8
intron9	98	1.9	0.6	98	0.9	0.9
exon10	235	79.6	13.5	235	155.3	18.6
intron10	83	0.1	0.1	90	0.3	0.6
exon11	126	100.8	13.3	126	158.0	35.0
intron11	65	1.3	1.9	83	10.5	4.3
exon12	86	91.8	19.3	86	158.9	34.5
intron12	**118**	**4.6**	**1.0**	**118**	**0.2**	**0.3**
exon13	189	109.4	13.8	189	159.3	16.9
intron13	276	1.6	1.0	273	0.7	0.6
exon14	132	155.1	7.4	132	255.8	22.9
intron14	112	4.2	4.3	112	0.4	0.7
exon15	326	135.4	9.4	326	207.6	13.0

RNA-seq quantification for each intron and exon in the *SYMRK* gene of *Trema orientalis* and *Parasponia andersonii* is determined by the mean per base coverage of three biological replicates.

Next, we questioned how unique a GA donor splice site is in plants. For this, we analysed all annotated introns in *T. orientalis*, *P. andersonii*, and the model plant species *L. japonicus, M. truncatula*, and *Arabidopsis thaliana* [23, 32–34]. This showed that a GA donor splice site is extremely rare, varying from none in the annotated gene models of *M. truncatula* to 14 in *A. thaliana* (Table 2, Table S1).

Table 2
Frequency of predicted canonical and non-canonical donor splice sites.

	T. orientalis		*P. andersonii*		*M.truncatula*		*L. japonicus*		*A. thaliana*	
Splice motif	**Total**	**%**	**Total**	**%**	**Total**	**%**	**Total**	**%**	**Total**	**%**
GT-AG	102107	98.09%	100852	98.36%	213612	95.94%	148570	97.99%	159839	98.80%
GC-AG	537	0.52%	567	0.55%	8962	4.03%	2193	1.45%	1742	1.08%
AT-AC	13	0.01%	14	0.01%	79	0.04%	656	0.43%	90	0.06%
GA-AG	**11**	**0.01%**	**8**	**0.01%**	**0**	**0.00%**	**6**	**0.00%**	**14**	**0.01%**
Others	1426	1.37%	1096	1.07%	0	0.00%	190	0.13%	89	0.06%
Canonical	102657	98.62%	101433	98.92%	222653	100.00%	151419	99.87%	161671	99.94%
Non-canonical	1437	1.38%	1104	1.08%	0	0.00%	196	0.13%	103	0.06%
Grand Total	104094	100.00%	102537	100.00%	222653	100.00%	151615	100.00%	161774	100.00%

Splice site occurrences are based on existing gene models predictions for *Trema orientalis*, *Parasponia andersonii*, *Lotus japonicus*, *Medicago truncatula* and *Arabidopsis thaliana*.

Trema orientalis SYMRK ^{RG33} **GA donor splice site is geographically limited**

As *T. orientalis* RG33 possesses an extremely rare GA motif at the donor splice site of intron 12, we question to what extent such polymorphism is unique in *SYMRK*. First, we analysed SYMRK orthologs in a broad phylogenetic context. This showed that a non-canonical GC donor splice site is common in *SYMRK* intron 12 of dicotyledon species (Fig. 5).

However, none of the analysed *SYMRK* genes possesses a GA motif at this position. Subsequently, we analysed *SYMRK* of the *Parasponia-Trema* species complex. Among others, *T. orientalis* accession RG33 was collected during an expedition in Sabah Provence, Malaysian Borneo, in 2012 [23, 25]. We analyzed 27 additional *T. orientalis* individuals collected from five distinct locations in Malaysian Borneo (Fig. 6A). All possess the rare GA intron 12 donor splice, whereas this mutation is absent in *Trema* and *Parasponia* accessions sampled outside Borneo (Fig. 6; Table S2). This demonstrates that the *SYMRK^{RG3}* allele is not unique, though it associates with the Borneo *T. orientalis* population.

Discussion

The LRR-type receptor kinase *SYMRK* is a critical component in the common symbiosis signalling pathway controlling endosymbioses. In legumes, *SYMRK* is essential for rhizobium LCO-induced signalling. We identified a seemingly critical mutation of the conserved dinucleotide motif in the 5'-donor splice site in *T. orientalis SYMRK* accession RG33. *T. orientalis* is a non-nodulating relative of nitrogen-fixing *Parasponia* species and has experienced pseudogenization of several key nodulation genes [23]. Here we show that despite a mutation in a splice site motif, *TorSYMRK^{RG33}* remains a functional allele that can be effectively spliced. The dominant occurrence of the *TorSYMRK^{RG33}* allele in the Malaysian Borneo *T. orientalis* population underlines the splice site mutation is not affecting the fitness of the tree species.

Splicing is a highly conserved process in eukaryotes, requiring a spliceosome complex consisting of five small nuclear RNAs and several proteins. The vast majority of introns are spliced by the so-called U2-type spliceosome, recognizing two highly conserved di-nucleotide motifs at the start and end of the intron sequence, namely GT-AG. Bioinformatic studies in plant, animal, and fungal species indicate that alternative dinucleotide motifs are used in less than 2% of cases, among which GC-AG is the most abundant non-canonical splice motif representing 1.5% of all introns annotated in plant gene models [35, 36]. The GA-AG splicing motif, as found in *TorSYMRK*^{RG33} intron 12, is reported to occur in > 0.03% of the cases [36].

The mechanism driving the evolution of rare non-canonical splice sites remains elusive. The GA-AG dinucleotide splicing motif was found in higher frequency in two non-related animal species; the copepod *Eurytemora affinis* and the tunicate *Oikopleura dioica* [36–38]. However, it remains unknown whether both species have gained these by convergent evolution or, alternatively, it is an ancestral trait preserved in only a few species [36]. In the case of *SYMRK*, we noted that in related species, *SYMRK* intron 12 possesses the more common non-canonical GC-AG dinucleotide splice motif. This may lead to the hypothesis that such a GC-AG motif is the ancestral state allowing the evolution of the even more rare GA-AG motif. We inserted the GC to GA mutation in the *P. andersonii SYMRK* gene and showed that this variant is fully functional when expressed under its native promoter. This suggests that a simple single nucleotide polymorphism is sufficient to allow the evolution of the GA-AG dinucleotide splicing motif in *TorSYMRK*^{RG33}. We analysed genomes of five plant species for gene models possessing a GA dinucleotide motif in the donor splice site. We found that the GA motif is indeed present in the annotated gene models, albeit at very low frequency in the analysed species.

Using CRISPR-Cas9 technology in *P. andersonii*, we demonstrated for the first time by mutant analysis that *SYMRK* commits a dual symbiotic role in essential nodulation and AM symbiosis in a non-legume. Earlier studies using RNAi in *C. glauca* and *D. glomerata* provided evidence that *SYMRK* is required for *Frankia*-induced nodulation and mycorrhization [9, 20]. *Parasponia*, *Casuarina*, and *Datisca*, together with legumes, represent all four taxonomic orders that contain nodulating species and for which *SYMRK* is an essential symbiotic gene. It supports the hypothesis that *SYMRK* -and other components of the common symbiosis signalling pathway- have been recruited to function in nodulation in a common ancestor that lived before the divergence of the Fabales, Fagales, Cucurbitales, and Rosales orders.

Conclusions

This study of the LRR-type receptor kinase SYMRK in the non-nodulating relative of nitrogen-fixing *Parasponia* species, *T. orientalis*, led to the identification of a functional splice site mutation in the gene. The discovery of this rare non-canonical GA-AG splice site motif in *SYMRK* raises questions about the evolution of such motifs and the mechanisms driving their occurrence. Furthermore, this study demonstrates the conservation of SYMRK functioning in nodulation and AM symbiosis in both legumes and non-legumes. The *Parasponia-Trema* comparative system is established to obtain insight into the evolutionary trajectory of the nodulation trait. It uncovered several genes critical for rhizobium-induced nodulation in a non-legume [23, 24, 39]. Eventually, *Trema* species can serve as an experimental test system to uncover essential genes to rebuild the nodulation trait. Additionally, we demonstrated that the *Parasponia-Trema* comparative system is equally valuable to uncovering the functionality of rare non-canonical splicing motifs. Overall, this study contributes to our understanding of both the common symbiosis signalling pathway and the mechanisms of gene splicing in plants.

Material And Methods

Plant materials and growth conditions

T. orientalis plants were sampled between September 10th and 25th, 2012, by the participants of the Crocker Range/Kinabalu Scientific Expedition, which was conceived, organized, funded, and conducted jointly by Sabah Parks (Malaysia) and Naturalis Biodiversity Center (The Netherlands) [25]. Taxonomic analysis of samples has been published previously [23]. *P. andersonii* WU1 and *T. orientalis* RG33 were grown and maintained as described previously [40, 41]. Plantlets for nodulation and mycorrhization assay were vegetatively propagated *in vitro* and rooted [40, 41].

Mycorrhization assays and trypan blue staining

Mycorrhization assays were performed using a commercial spore of *Rhizopagus irregularis* (Agronutrion-DAOM197198, Carbonne, France). Spores inoculum, inoculation, and trypan blue staining were prepared and performed as described previously [41].

To quantify mycorrhization, a minimum of ~ 50 cm roots for each sample were cut into 1 cm fragments. 25–30 root fragments were placed on a single microscope slide, and 30% glycerol was added. Roots were covered with a cover glass and pressed until root fragments became flat. The frequency of mycorrhiza (%F), the intensity of mycorrhizal colonization (%M), and arbuscules abundance (%A) in the root system was scored and calculated according to Trouvelot et al. [27].

Nodulation assay

P. andersonii plantlets for nodulation were inoculated with *Mesorhizobium plurifarium* BOR2 (OD600 = 0.05) [23, 40, 41]. Plants were removed from the pots six weeks post-inoculation, roots were washed with running water to remove perlite, and nodules were counted. In (*trans*) complementation studies, plant roots were examined under fluorescent stereo microscopy, and nodule number was quantified for each transgenic root (eight weeks post-inoculation with *Mesorhizobium plurifarium* BOR2 (OD600 = 0.025)).

Root growth assay

Five seedlings of *P. andersonii* and *T. orientalis* RG33 were grown on ½ strength modified Hoagland medium in 12 cm square plates. Plants were grown vertically at a 60-degree angle for 21 days at 28^{o}C, 16/8h day-night regime. The primary root was determined as the main root that emerged from cotyledon, whereas lateral roots were determined as roots that emerged from the primary root. Per plants, primary root length, the average number of lateral roots, and lateral root density (per cm main root) were determined 21 days post germination. Primary root growth was measured by following its development every day for 21 days post germination. The average lateral root length was determined by measuring its size in five selected lateral roots 21 days post germination.

Vectors and Constructs

Single-guide RNAs (sgRNAs) were designed using the 'Find CRISPR Targets' function implemented in Geneious software version 9.1.5 (Biomatters, New Zealand) and subsequently checked against the *P. andersonii* genome for high identity off-targets. For CRISPR/Cas9-mediated mutagenesis and complementation studies, binary transformation constructs were created using Golden Gate assembly as described previously [40, 41], and a list of constructs generated from both studies is listed in Table S3. For CRISPR/Cas9-mediated mutagenesis, two sgRNAs were used to target the fourth and the fifth coding exon of *PanSYMRK* (Figure S3). Selected sgRNAs were amplified using sequence-specific forward primers and a universal reverse primer (Table S4), using Addgene plasmid no. 46966 as template [42]. To allow for Golden Gate cloning, BpiI and BsaI restriction sites in the putative promoter sequence of *PanSYMRK* were mutated by introducing single nucleotide substitution [43]. For the complementation study, the sequence of *P. andersonii SYMRK* promoter, 5' untranslated region (5' UTR), genomic DNA, 3' untranslated region (3' UTR), and terminator were synthesized. Also, a modified version of *P. andersonii SYMRK* genomic DNA was synthesized harbouring a point mutation at the donor splice site of the 12th intron, mimicking *T. orientalis SYMRK*^{RG33}. (Invitrogen, Thermo Fisher Scientific, United States).

Plant transformation

Agrobacterium tumefaciens-mediated transformation and genotyping were done based on previously published protocols [40, 41]. Primers used for genotyping are listed in Table S4. Hairy root transformations were performed according to Cao *et al.* [44], where *A. rhizogenes* MSU440 or AR1193 harbouring plasmid DNA of interest were used to infect micro-propagated plants wounded on their base. Infected plants were grown on agar plates of Schenk and Hildbrandt medium (SH medium) [45] and incubated at 21°C for one week on a 16/8h light/dark regime. Transformed plants were transferred to agar plates of SH medium supplemented with 10 g sucrose/L, cefotaxime 100 µg/mL, and kanamycin 50 µg/mL and subsequently incubated at 21°C for one week followed by 28°C for two weeks. Plants were checked for transgenic roots using a fluorescence stereo microscope.

RNA Sequencing

For RNA isolation, tissue was harvested from a ~ 1 cm region just above the meristematic zone of young growing roots and snap-frozen in liquid nitrogen. Material from ~ 5 plants was combined to form a single biological replicate. RNA was isolated in triplicate as previously described [23]. Library preparation and RNA sequencing was conducted by BGI (Schenzhen, China). Mapped RNA-sequencing reads covering the *SYMRK* gene in *P. andersonii* and *T. orientalis* were visualized using Integrative Genomics Viewer (IGV) [46]. Based on the different splice sites, two *SYMRK* splice variants were manually constructed. Functional protein domains for these variants were annotated using InterProScan 5 [47].

Phylogenetic reconstruction

Orthologs of SYMRK were identified among 49 publicly available proteomes by applying a Reciprocal Best Hits (RBH) approach, using *L. japonicus* SYMRK (Lj2g3v1467920.1) as the query sequence. Identified orthologous proteins were aligned using Clustal Omega 1.2.3. [48]. A phylogenetic SYMRK tree was constructed using PhyML 3.0 [49] with LG substitution model 1,000 bootstrap replicates and rooted on the two Poales outgroup species. The tree was visualized using the Interactive Tree Of Life (iTOL) tree viewer [50]. A sub-selection of 20 species was extracted from the SYMRK orthogroup, and a tree was constructed using the same methods described above. Based on the SYMRK gene models for these 20 species, the splice site at intron 12 for each SYMRK ortholog was added.

Statistical analysis

Graphs and statistical analysis for mycorrhization quantification were performed using RStudio version 1.1.456. The Ramf R package was used to analyze and display quantitative AM fungal root colonization data [51]. Statistical tests on three classes of mycorrhization efficiency were done using Kruskal-Wallis test in combination with the post-hoc test using Fisher's least significant difference criterion. Statistical significance was defined as a $p < 0.01$. A statistical test on root growth assays and for nodules number quantification on complementation study was done using a student t-test. Statistical significance for these parameters was defined as a $p < 0.05$.

Declarations

Availability of data and materials

The datasets analysed during the current study are available in the NCBI SRA repository under BioProject numbers PRJNA272473 and PRJNA272482. Plant material and seeds used in this study can be obtained upon request from the corresponding author.

Funding

The project was funded by the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia (grant no. 8245–ID to YPR), the Ministry of Education, King Faisal University, Saudi Arabia (grant no. 10598 to SA) and the ENSA project funded by the Bill & Melinda Gates Foundation to the University of Cambridge (OPP1172165 to RG).

Author information

Sultan Alhusayni and Yuda Purwana Roswanjaya contributed equally to this work.

Authors'Contributions

Data presented in Fig. 1: YPR; Fig. 2: YPR; Fig. 3: SA; Fig. 4: SA, RH, LR; Fig. 5: JK, SA; Fig6: RG, YR, WK; Fig S1: TS, WK; Fig. S2: JK; Fig. S3: YPR, LR, RH; Fig. S4: SA, Fig. S5: JK, SA; Table 1: JK; Table 2: JK; Table S1: JK, MS; Table S2: YPR, TS, WK, RG; Table S3: YPR, SA. Manuscript preparation: RG, JK. All authors reviewed the manuscript. The authors read and approved the final manuscript.

Corresponding authors

Correspondence to Joël Klein or Rene Geurts.

Ethics approval and consent to participate

This study was conducted in accordance with local legislation and the

Convention on the Trade in Endangered Species of Wild Fauna and Flora (CITES). No specific permits were required.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflict of interest.

References

1. Remy W, Taylor TN, Hass H, Kerp H. Four hundred-million-year-old vesicular arbuscular mycorrhizae. Proc Natl Acad Sci U S A. 1994;91:11841–3.
2. Brundrett MC, Tedersoo L. Evolutionary history of mycorrhizal symbioses and global host plant diversity. New Phytol. 2018;220:1108–15.
3. Radhakrishnan GV, Keller J, Rich MK, Vernié T, Mbadinga Mbadinga DL, Vigneron N, et al. An ancestral signalling pathway is conserved in intracellular symbioses-forming plant lineages. Nat Plants. 2020;6:280–9.
4. Duc G, Trouvelot A, Gianinazzi-Pearson V, Gianinazzi S. First report of non-mycorrhizal plant mutants (Myc–) obtained in pea (Pisum sativum L.) and fababean (Vicia faba L). Plant Sci. 1989;60:215–22.
5. Parniske M. Arbuscular mycorrhiza: the mother of plant root endosymbioses. Nat Rev Microbiol. 2008;6:763–75.
6. Oldroyd GED. Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. Nat Rev Microbiol. 2013;11:252–63.
7. Endre G, Kereszt A, Kevei Z, Mihacea S, Kaló P, Kiss GB. A receptor kinase gene regulating symbiotic nodule development. Nature. 2002;417:962–6.
8. Stracke S, Kistner C, Yoshida S, Mulder L, Sato S, Kaneko T, et al. A plant receptor-like kinase required for both bacterial and fungal symbiosis. Nature. 2002;417:959–62.
9. Markmann K, Giczey G, Parniske M. Functional adaptation of a plant receptor-kinase paved the way for the evolution of intracellular root symbioses with bacteria. PLoS Biol. 2008;6:0497–506.
10. Kosuta S, Held M, Hossain MS, Morieri G, Macgillivary A, Johansen C, et al. Lotus japonicus symRK-14 uncouples the cortical and epidermal symbiotic program. Plant J. 2011;67:929–40.
11. Antolín-Llovera M, Ried MK, Parniske M. Cleavage of the SYMBIOSIS RECEPTOR-LIKE KINASE ectodomain promotes complex formation with Nod factor receptor 5. Curr Biol. 2014;24:422–7.
12. Pan H, Stonoha-Arther C, Wang D. Medicago plants control nodulation by regulating proteolysis of the receptor-like kinase DMI2. Plant Physiol. 2018.
13. Madsen EB, Madsen LH, Radutoiu S, Olbryt M, Rakwalska M, Szczyglowski K, et al. A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. Nature. 2003;425:637–40.
14. Gysel K, Laursen M, Thygesen MB, Lironi D, Bozsóki Z, Hjuler CT et al. Kinetic proofreading of lipochitooligosaccharides determines signal activation of symbiotic plant receptors. Proc Natl Acad Sci U S A. 2021;118.
15. Catoira R, Galera C, de Billy F, Penmetsa RV, Journet EP, Maillet F, et al. Four genes of Medicago truncatula controlling components of a nod factor transduction pathway. Plant Cell. 2000;12:1647–66.
16. Indrasumunar A, Wilde J, Hayashi S, Li D, Gresshoff PM. Functional analysis of duplicated Symbiosis Receptor Kinase (SymRK) genes during nodulation and mycorrhizal infection in soybean (Glycine max). J Plant Physiol. 2015;176:157–68.
17. Feng Y, Wu P, Liu C, Peng L, Wang T, Wang C, et al. Suppression of LjBAK1-mediated immunity by SymRK promotes rhizobial infection in Lotus japonicus. Mol Plant. 2021. https://doi.org/10.1016/j.molp.2021.07.016.

18. Calantzis C, Morandil D, Arnouldl C, Gianinazzi-Pearsonl V. Cellular Interactions between G. mosseae and a Myc" dmi2 mutant in Medicago truncatula. 2001. https://dalspace.library.dal.ca/bitstream/handle/10222/77857/VOLUME%2030-NUMBERS%202-3-2001-PAGE%2097.pdf?sequence=1. Accessed 25 Jul 2022.
19. Limpens E, Mirabella R, Fedorova E, Franken C, Franssen H, Bisseling T, et al. Formation of organelle-like N2-fixing symbiosomes in legume root nodules is controlled by DMI2. Proc Natl Acad Sci U S A. 2005;102:10375–80.
20. Gherbi H, Markmann K, Svistoonoff S, Estevan J, Autran D, Giczey G, et al. SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and Frankiabacteria. Proc Natl Acad Sci U S A. 2008;105:4928–32.
21. 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.
22. 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.
23. 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.
24. 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.
25. Merckx VSFT, Hendriks KP, Beentjes KK, Mennes CB, Becking LE, Peijnenburg KTCA, et al. Evolution of endemism on a young tropical mountain. Nature. 2015;524:347–50.
26. Capoen W, Goormachtig S, De Rycke R, Schroeyers K, Holsters M. SrSymRK, a plant receptor essential for symbiosome formation. Proc Natl Acad Sci U S A. 2005;102:10369–74.
27. Trouvelot A, Kough JL, Gianinazzi-Pearson V. Mesure du taux de mycorhization VA d'un systeme radiculaire. Recherche de methodes d'estimation ayant une signification fonctionnelle. In: Gianinazzi-Pearson GSV, editor. Physiological and Genetic Aspects of Mycorrhizae. INRA press; 1986. pp. 217–21.
28. Demchenko K, Winzer T, Stougaard J, Parniske M, Pawlowski K. Distinct roles of Lotus japonicus SYMRK and SYM15 in root colonization and arbuscule formation. New Phytol. 2004;163:381–92.
29. Ried MK, Antolín-Llovera M, Parniske M. Spontaneous symbiotic reprogramming of plant roots triggered by receptor-like kinases. Elife. 2014;3:1–17.
30. Kevei Z, Lougnon G, Mergaert P, Horváth GV, Kereszt A, Jayaraman D, et al. 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase1 Interacts with NORK and Is Crucial for Nodulation in Medicago truncatula. Plant Cell. 2007;19:3974–89.
31. Jayaraman D, Richards AL, Westphall MS, Coon JJ, Ané J-M. Identification of the phosphorylation targets of symbiotic receptor-like kinases using a high-throughput multiplexed assay for kinase specificity. Plant J. 2017;90:1196–207.
32. Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, et al. Genome structure of the legume, Lotus japonicus. DNA Res. 2008;15:227–39.
33. Pecrix Y, Staton SE, Sallet E, Lelandais-Brière C. Whole-genome landscape of Medicago truncatula symbiotic genes. Nat Plants. 2018.
34. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 2012;40:1202–10. Database issue:D.
35. Parada GE, Munita R, Cerda CA, Gysling K. A comprehensive survey of non-canonical splice sites in the human transcriptome. Nucleic Acids Res. 2014;42:10564–78.
36. Frey K, Pucker B. Animal, Fungi, and Plant Genome Sequences Harbor Different Non-Canonical Splice Sites. Cells. 2020;9.
37. Lee CE. Evolutionary mechanisms of habitat invasions, using the copepod Eurytemora affinis as a model system. Evol Appl. 2016;9:248–70.
38. Robertson HM. Noncanonical GA and GG 5' Intron Donor Splice Sites Are Common in the Copepod Eurytemora affinis. G3 Genes|Genomes|Genetics. 2017;7:3967–9.

39. Bu F, Rutten L, Roswanjaya YP, Kulikova O, Rodriguez-Franco M, Ott T et al. Mutant analysis in the non-legume Parasponia andersonii identifies NIN and NF-YA1 transcription factors as a core genetic network in nitrogen-fixing nodule symbioses. New Phytol. 2020.
40. van Zeijl A, Wardhani TAK, Seifi Kalhor M, Rutten L, Bu F, Hartog M, et al. CRISPR/Cas9-Mediated Mutagenesis of Four Putative Symbiosis Genes of the Tropical Tree Parasponia andersonii Reveals Novel Phenotypes. Front Plant Sci. 2018;9:284.
41. 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.
42. Nekrasov V, Staskawicz B, Weigel D, Jones JDG, Kamoun S. Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nat Biotechnol. 2013;31:691–3.
43. Engler C, Youles M, Gruetzner R, Ehnert T-M, Werner S, Jones JDG, et al. A golden gate modular cloning toolbox for plants. ACS Synth Biol. 2014;3:839–43.
44. Cao Q, den Camp RO, Kalhor MS, Bisseling T, Geurts R. Efficiency of Agrobacterium rhizogenes–mediated root transformation of Parasponia and Trema is temperature dependent. Plant Growth Regul. 2012;68:459–65.
45. Schenk RU, Hildebrandt AC. Medium and techniques for induction and growth of monocotyledonous and dicotyledonous plant cell cultures. Can J Bot. 1972;50:199–204.
46. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24–6.
47. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.
48. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539.
49. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59:307–21.
50. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 2007;23:127–8.
51. Chiapello M, Das D, Gutjahr C. Ramf: An Open-Source R Package for Statistical Analysis and Display of Quantitative Root Colonization by Arbuscular Mycorrhiza Fungi. Front Plant Sci. 2019;10:1184.

Figures

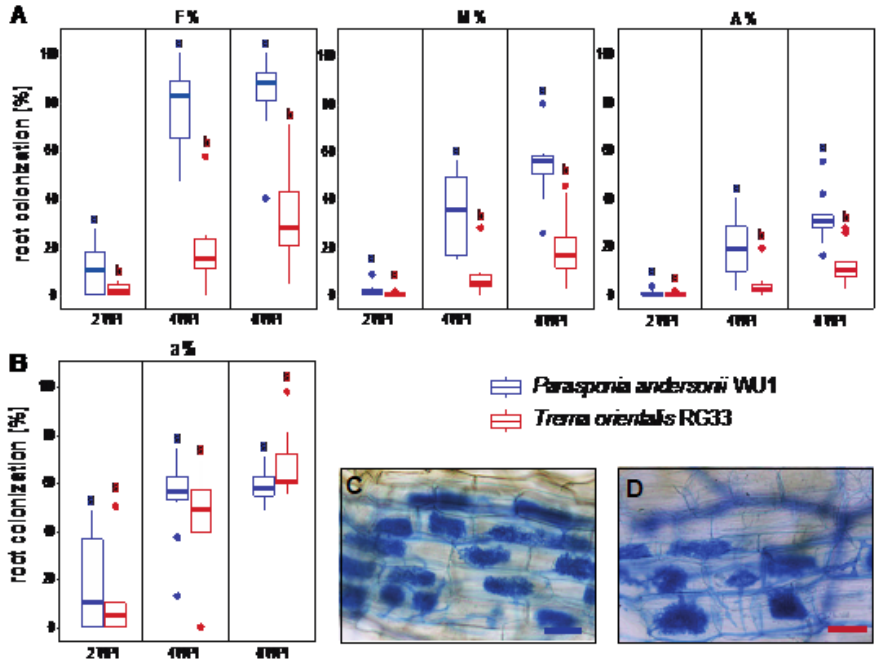


Figure 1

Trema orientalis accession RG33 and *Parasponia andersonii* accession WU1 differ in mycorrhizae colonisation. (A) Comparison of mycorrhization efficiency in the root system of *P. andersonii* WU1 (blue) and *T. orientalis* RG33 (red) at 2, 4 and 6 weeks post-inoculation with *Rhizophagus irregularis* DOAM197198. F%: The frequency of mycorrhiza detected in 50 randomly selected 1 cm segments of a root system. M%: the intensity of mycorrhizal colonisation in 50 randomly selected 1 cm segments of the root system. A%: Arbuscule abundance in 50 randomly selected 1 cm segments of the root system. (B) a%: Averaged arbuscule abundance in infected root segments. Error bars represent the SE of 10 biological replicates for each 50x 1cm root segment that has been analysed. Analysis was done according to Trouvelot *et al.* (1986) [26]. (C) Toluidine blue-stained *P. andersonii*and (D) *T. orientalis* root segment visualising *R. irregulates* arbuscules 6 weeks post-inoculation. Size bar = 10 µm.

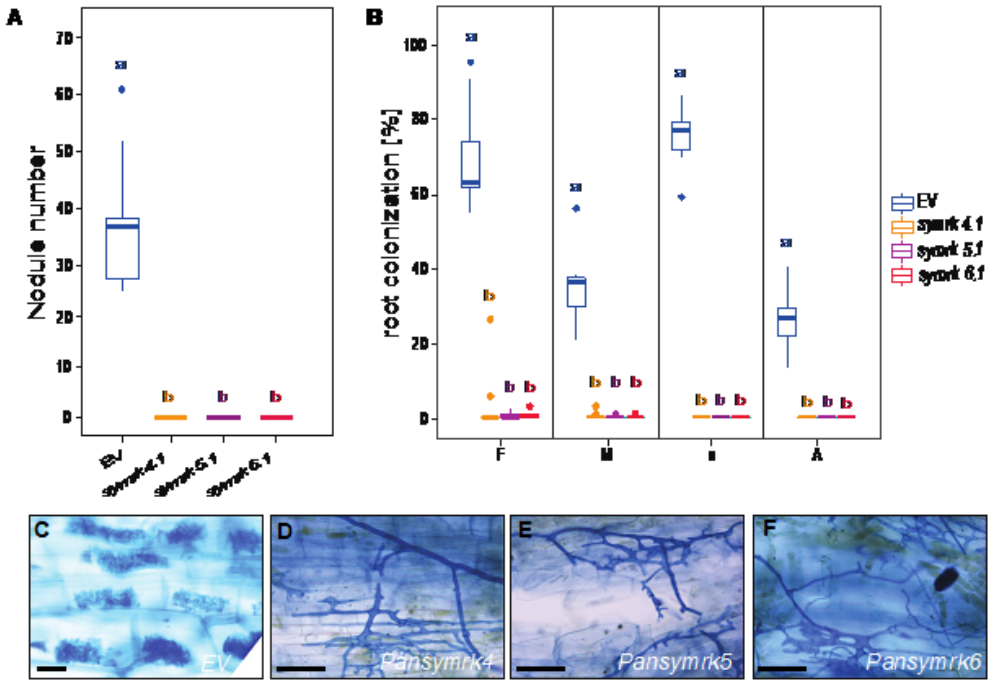


Figure 2

Parasponia andersonii SYMRK is essential for mycorrhization and nodulation. (A) Nodule numbers formed in *P. andersonii* empty vector control line (EV) and three *Pansymrk* mutant lines, 6 weeks post-inoculation with *Mesorhizobium plurifarium* BOR2. (B) mycorrhization efficiency in the root system of *P. andersonii* ev-control and three independent *Pansymrk* mutant lines 6 weeks post-inoculation with *Rhizophagus irregularis* DOAM197198. F%: The frequency of mycorrhiza detected in 50 randomly selected 1 cm segments of a root system. M%: the intensity of mycorrhizal colonization in 50 randomly selected 1 cm segments of the root system. A%: Arbuscule abundance in 50 randomly selected 1 cm segments of the root system. a%: Averaged arbuscule abundance in infected root segments. Error bars represent the SE of 10 biological replicates, for each 50x 1cm root segment that has been analyzed. Analysis was done according to Trouvelot *et al.* (1986) [26] (C-F): Toluidine blue-stained *P. andersonii* EV-control (C), *Pansymrk-4* (D), *Pansymrk-5* (E), and *Pansymrk-6* (F) root segment visualizing *R. irregulates* infections 6 weeks post-inoculation. Size bar = 10 µm.

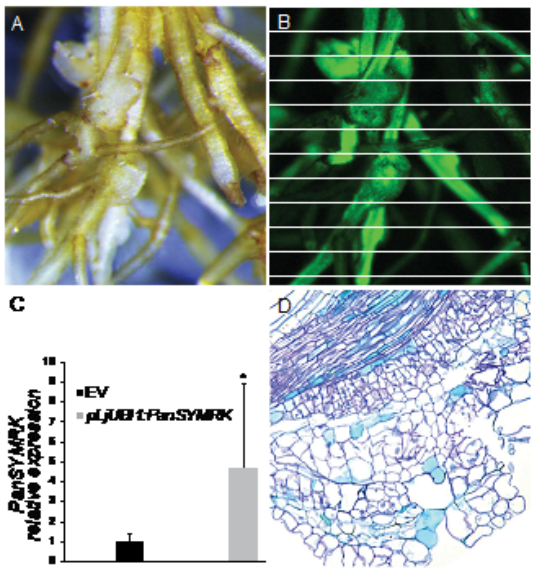


Figure 3

PanSYMRK ectopic expression induces spontaneous nodulation in *Parasponia andersonii*. (A, B) Bright-field (A) and green fluorescent image (B) of *P. andersonii A. rhizogenes*-transformed roots expressing *GFP* and *PanSYMRK* under control of the *pLjUBI1* promoter showing spontaneously formed nodule-like structures (6 weeks post planting). (C) Relative gene expression of *PanSYMRK* in *P. andersonii A. rhizogenes*-transformed roots containing an empty vector (EV) or *pLjUBI1:PanSYMRK*(n=3). (D) Longitudinal section of a spontaneously formed nodule-like structure visualizing cortical and pericycle cell divisions.

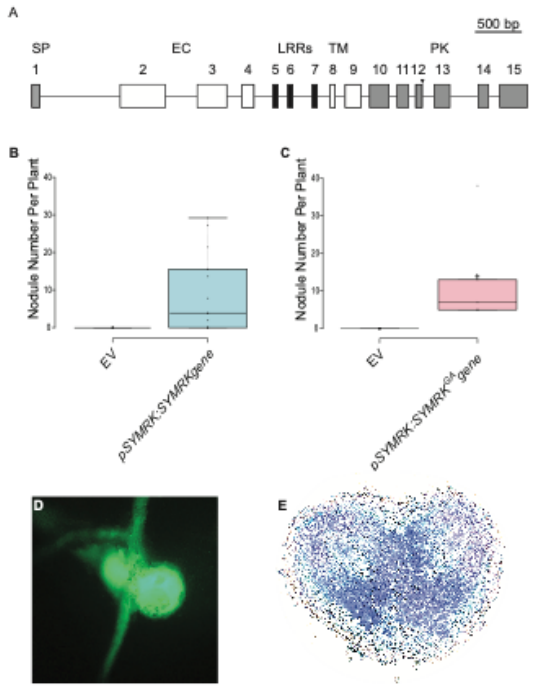


Figure 4

Parasponia symrk-5 mutant trans-complementation of root nodule symbiosis. (A) Schematic representation of *P. andersonii SYMRK* gene structure. Arrowhead points to the location of the introduced GA mutation in *PanSYMRK* at the 5'-donor splice site of intron 12. (B) Nodule number per plant formed on *Pansymrk-5 A. rhizogenes*transformed root with *pPanSYMRK:PanSYMRK*gene. (C-E) Nodule number per plant (C), representative image of green flurescence protein (GFP) nodule (D), and a section through a mature nodule (E) of *Pansymrk-5 A. rhizogenes* transformed root with *pPanSYMRK:PanSYMRKGA* carrying a GA mutation at the 5'-donor splice site of intron 12. Nodules were harvested and analysed at 8 weeks post inoculation with *Mesorhizobium plurifarium* BOR2 (OD600 = 0.025).

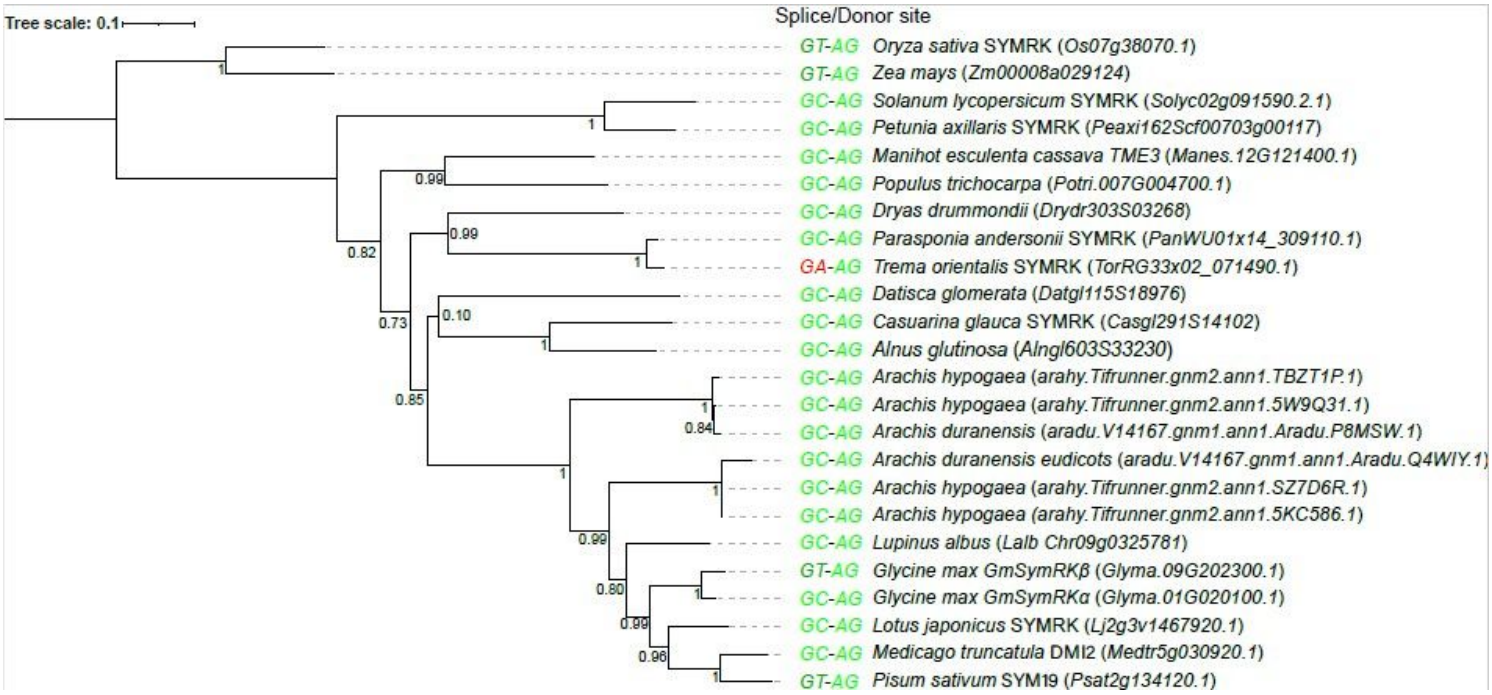


Figure 5

Phylogeny of SYMRK including the splice site dinucleotide motifs for intron 12. Phylogeny was reconstructed based on an alignment of SYMRK orthologous proteins from 19 species. Leaves are labelled by their respective species, gene name if available) and gene identifier. The non-canonical GC donor splice site is common in *SYMRK* intron 12 of dicotyledon species, except in *Glycine max SYMRKβ* and *Pisum sativum SYM19*, where GC is substituted by GT. In contrast, only *Trema orientalis* RG33 possesses a GA motif in this position (highlighted in red).

Supplementary Files

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

- FigureS1TremaParasponiarootgrowthdifferences.pdf
- FigureS2SYMRKorthologsphylogeny.pdf
- FigureS3CRISPRtargets.pdf
- FigureS4complementationmycphenotype.pdf
- FigureS5RNAseqmappingtoSYMRKintron12.pdf
- SupplementalfiguresandtablesRoswanjaya2023SYMRKMS.docx
- TableS1FrequencyofintronwithGAAGintronsplicesitesinplantspecies.xlsx
- TabelS2TwentyeightTremaorientalisindividualscollectedallpossessanoncanonicaldonorsplicesiteatintron12.xlsx
- TableS3ListofGoldenGateconstructs.xlsx
- TableS4Primersusedinthisstudy.xlsx