

Widely-conserved transcription factors regulate symbiosis-specific NCR genes

Senlei Zhang

Biological Research Centre

Ting Wang

Biological Research Centre

Rui Lima

Biological Research Centre

Aladár Pettkó-Szandtner

Biological Research Centre, Hungarian Academy of Sciences

Attila Kereszt

Biological Research Centre

J. Allan Downie

John Innes Centre <https://orcid.org/0000-0003-1011-4960>

Eva Kondorosi (✉ eva.kondorosi@gmail.com)

Hungarian Academy of Sciences, Hungary

Article

Keywords:

Posted Date: July 29th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1827226/v1>

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

[Read Full License](#)

Title: Widely-conserved transcription factors regulate symbiosis-specific *NCR* genes

Authors: Senlei Zhang¹, Ting Wang¹, Rui M. Lima¹, Aladár Pettkó-Szandtner², Attila Kereszt¹, J. Allan Downie^{1,3}, Eva Kondorosi^{1*}

Affiliations:

¹Institute of Plant Biology, Biological Research Centre, Temesvári krt. 62, 6726 Szeged, Hungary.

²Laboratory of Proteomic Research, Biological Research Centre, Temesvári krt. 62, 6726 Szeged, Hungary.

³John Innes Centre, Norwich Research Park, Colney Ln, Norwich, NR4 7UH, UK.

*Corresponding author. Email: eva.kondorosi@gmail.com

Symbiotic nitrogen fixation by rhizobium bacteria within the cells of legume root nodules alleviates the need for nitrogen fertilizers. Nitrogen fixation requires the endosymbionts to differentiate into bacteroids and this can be reversible or terminal. The latter is controlled by the plant, is more beneficial and has evolved in a large clade of the legume family. The plant effectors of terminal differentiation are Nodule-specific Cysteine-Rich NCR peptides, which are absent in legumes such as soybean where there is no terminal differentiation of rhizobia. It was assumed that *NCRs* coevolved with specific transcription factors, but our work demonstrates that expression of *NCR* genes does not require *NCR*-specific transcription factors. Introduction of the *Medicago truncatula* *NCR169* gene under its own promoter into soybean roots resulted in nodule-specific expression of *NCR169*, leading to bacteroid changes associated with terminal differentiation. We identified two AT-Hook

Motif Nuclear Localized (AHL) transcription factors from both *M. truncatula* and soybean nodules that bound to AT-rich sequences in the *NCR169* promoter inducing its expression. Whereas mutation of *NCR169* arrested bacteroid development at a late stage, the absence of MtAHL1 or MtAHL2 completely blocked bacteroid differentiation indicating that they also regulate other *NCR* genes required for development of nitrogen-fixing nodules. Regulation of *NCRs* by orthologous transcription factors in non-IRLC legumes opens the possibility of increasing the efficiency of nitrogen fixation in legumes lacking *NCRs*.

Legume plants limited for nitrogen establish an intracellular symbiosis with bacteria called rhizobia, resulting in the formation of nitrogen-fixing root nodules. Nodules can be of indeterminate or determinate types, respectively depending on whether the nodule meristem is persistent or limited during development. In nodule cells, the intracellular bacteria are present in organelle-like structures, called symbiosomes in which they differentiate into nitrogen fixing bacteroids. Depending on the legume, this differentiation can be reversible or terminal, In the determinate nodules of soybean or the model legume *Lotus japonicus*, bacteroids are similar to the free-living bacteria and, as they can regrow from nodules, their fate is reversible. In the indeterminate nodules of the Inverted Repeat-Lacking Clade (IRLC) legumes that includes peas, clovers, and medics there is an age and differentiation gradient from the apex to the nodule base resulting in the formation of a meristematic zone at the apex (ZI), an invasion zone (ZII) where rhizobia infect nodule cells, an interzone (IZ) where endosymbiont differentiation starts, and a nitrogen-fixing zone (ZIII) containing fully differentiated bacteroids. In IRLC legumes, such bacteroids cannot return to the free-living rhizobial lifestyle as their genome amplification results in the formation of large, polyploid, uncultivable, terminally differentiated bacterial cells that have altered morphology and physiology¹. This terminal bacteroid differentiation is governed by

plant-made nodule-specific cysteine-rich (NCR) peptides and it is more beneficial for the host plant in nitrogen fixation^{2,3}. *NCR* genes have uniquely evolved in the IRLC legumes and are absent from other legumes and other plants.

In the model legume *Medicago truncatula*, ~700 genes encode secreted NCR peptides. The mature peptides are mostly 30-50 amino-acid-long, highly divergent and characterized by 4 or 6 conserved cysteines⁴. The *NCR* genes are exclusively expressed in the symbiotic nodule cells in consecutive waves (*M. truncatula* Gene Expression Atlas (MtGEA), <https://lipm-browsers.toulouse.inra.fr/pub/expressionAtlas/app/mtgeav3/>)⁵. This results in the delivery of different sets of NCRs to bacteroids as they develop. The rhizobial *BacA* gene encoding a peptide transporter is required for NCR-induced bacteroid differentiation and symbiotic nitrogen fixation^{6,7}, but *BacA* (or its orthologue, *BclA*) is not required in rhizobia infecting legumes that lack NCR peptides⁸.

The function of only a few NCR peptides has been elucidated. Among them, NCR169 is essential for full differentiation of bacteroids and development of nitrogen-fixing nodules in *M. truncatula*⁹. *NCR169* is expressed in the interzone and the nitrogen fixing zone and its absence in the *M. truncatula dnf7-2* mutant provoked degradation of immature bacteroids and the absence of a nitrogen-fixing zone. Since the *NCR169* promoter shows conserved features found in many *NCR* genes¹⁰, *NCR169* was ideally suited for analysis of *NCR* gene regulation. The *dnf7-2* mutant was successfully complemented with a 1178 bp promoter region upstream of the translational start site⁹. In this work we identified *cis*-acting elements and nodule-expressed DNA-binding proteins belonging to the Type-I AT-Hook Motif Nuclear Localized (AHL) transcription factor family, which are essential for the induction and proper expression of *NCR169* for the development of nitrogen-fixing root nodules. These AHLs are conserved in non-IRLC legumes and can induce expression of active NCR169 in symbiotic nodule cells.

Results

***NCR169* is expressed under its own promoter in soybean nodules and elicits features of terminal bacteroid differentiation**

The 1181 bp promoter region of *NCR169* known to be sufficient for normal expression in *M. truncatula*⁹ was fused with the *GUS* reporter gene and introduced into soybean roots by *Agrobacterium rhizogenes*-mediated root transformation. Unexpectedly, the reporter gene was expressed, as detected by blue staining due to the GUS enzyme activity, in the infected symbiotic soybean nodule cells (Fig. 1a), revealing that transcription factors in soybean nodules can induce the promoter of this *Medicago*-specific gene. Furthermore, the GUS activity increased with maturation of the soybean symbiotic cells. In *M. truncatula* nodules *NCR169* expression also reached full expression in the nitrogen-fixing zone after initial expression in the interzone⁹.

NCR169 preceded by this 1181 bp promoter region was used to transform soybean roots to test if *NCR169* affected soybean bacteroid morphology. Nodule sections were stained with SYTO 13 to visualize bacteroids by confocal microscopy. Comparison of nodules on roots transformed with the empty vector (Fig. 1b) with nodules expressing *NCR169* (Fig. 1c) indicated *NCR169*-induced elongation of bacteroids. This was confirmed by scanning electron microscopy of bacteroids isolated from the control (Fig. 1b) and *NCR169*-expressing (Fig. 1c) nodules. The average length of bacteroids in the nodules expressing *NCR169* increased by about 25% (Fig. 1d). This was associated with a small but significant increase in the DNA content measured by flow cytometry (Fig. 1e).

In *Sinorhizobium meliloti*, *BacA* is required for NCR-mediated differentiation of bacteroids and symbiotic nitrogen fixation in *M. truncatula*; the homologous gene in *Bradyrhizobium* (*BclA*) is not required for nitrogen fixation in soybean nodules but is essential in *Aeschynomene* legumes

producing NCR-like peptides governing their terminal differentiation^{8,11}. To assess the requirement of BclA for NCR169-induced bacteroid elongation, wild type *B. japonicum* USDA110 and its *bclA* mutant were inoculated onto soybean roots transformed with *NCR169* or the vector control (Fig. 1f) and the bacteroids in transgenic nodules were visualized with SYTO 13. *NCR169*-induced cell elongation occurred in nodules infected by wild-type *B. japonicum* but was absent in the *bclA* mutant. This result supports the known interdependence of BclA/BacA and NCR functions. Moreover, the *NCR169*-induced increase in the size of soybean bacteroids is in line with the role of *NCR169* in *M. truncatula*, confirming that the *NCR169* peptide is expressed in, and can function in soybean nodules.

Delimitation of the minimal *NCR169* promoter region

Previous bioinformatic analysis of 209 *NCR* gene promoters from *M. truncatula* revealed five motifs in *NCR* promoters¹⁰. In the *NCR169* promoter, sequences with similarity to motifs 2, 1, 4, 5 and 3 were present in that order in ~ 400 bp upstream of the translation start, with motif 3 closest to the translational start. However, potential motifs were also found further upstream (Extended Data Fig. 1).

To test if the ~ 400 bp promoter region was sufficient for *NCR169* expression, 432 bp upstream of the translation start (including all five motifs) was fused with *GUS* and introduced into *M. truncatula* and *G. max* roots using *A. rhizogenes*-mediated root transformation. The expression of this reporter fusion in both plants (Fig. 2a) was similar to that seen with the longer promoter (Fig. 1a). In *M. truncatula*, the spatial expression pattern of the 432 bp promoter::*GUS* was identical to that seen with the 1181 bp promoter::*GUS*⁹ indicating that the 432 bp DNA fragment confers the correct, *NCR169*-specific expression pattern. In line with this, a construct containing the *NCR169* coding region downstream of the 432 bp fragment also complemented *M.*

truncatula dnf7-2 (ncr169 mutant) to a similar extent as the 1181 bp promoter region, resulting in nitrogen fixation, and normal plant development (Fig. 2b) and plant mass (Fig. 2c) under nitrogen limitation. This confirms that the 432 bp fragment is sufficient for proper production of *NCR169*.

Identification of proteins interacting with the *NCR169* promoter region

In initial experiments, the 1181 bp promoter region was used as the bait target DNA in a yeast one hybrid screen in which the prey proteins were from a *M. truncatula* nodule cDNA library expressing the proteins in fusion with the yeast GAL4 transcription activation domain (GAL4 AD)¹². This screen resulted in the identification of seven putative DNA-binding proteins from *M. truncatula* nodules (Table 1). These were: an AT hook motif DNA-binding family protein, MtAHL1; a bHLH domain class transcription factor; a MYB-like transcription factor family protein; a TCP family transcription factor; a transcription factor VOZ1-like protein; a BEL1-related homeotic protein; and a linker histone H1 and H5 family protein. Comparing the expression of these genes in different plant organs and in different nodule zones^{5,13} revealed that while each of them was expressed in both roots and nodules, only two of them, *MtAHL1* and the *BEL1*-related homeotic protein gene, were upregulated in nodules. Unlike *MtAHL1*, the *BEL1* transcripts were also detectable in leaves and petioles while the transcripts of the other five genes were present in all plant organs.

As a complementary approach to identify potential *NCR169* transcription factors, DNA affinity chromatography pull-down experiments were carried out with a 382 bp DNA fragment that extended from -4 bp to -386 bp upstream of the translation start encompassing all five potential promoter motifs described above. Nuclear protein extracts from *M. truncatula* and *G. max* nodules were separately added to this DNA fragment and proteins that bound were identified by

mass spectrometry. From the *M. truncatula* nuclear extracts, eight putative DNA-binding proteins were obtained: two of these were AT-hook-motif nuclear proteins including MtAHL1 (as identified with the Y1H screen), and one which we refer to as MtAHL2; two were PHD finger alfin-like proteins (MtPHD1 and MtPHD2); two were Myb/SANT-like DNA-binding domain proteins and the others were, a Pura α protein and a Zinc finger C-x8-C-x5-C-x3-H type family protein (Table 1). With the *G. max* nuclear extracts, five putative DNA-binding proteins were identified (Table 1): based on phylogeny (Extended Data Fig. 2) one of these was GmAHL1, a probable orthologue of MtAHL1 (76% identity); one was a PHD finger alfin-like protein (GmPHD1), a probable orthologue of MtPHD1 (90% identity); one was a SHOOT2-like protein; one was a tri-helix-like protein and one was GmPura α , a probable homologue of MtPura α identified above (Table 1).

MtAHL1 was detected with both the Y1H screen and by DNA affinity chromatography and its orthologue was identified by DNA affinity pulldown from *G. max*. This together with the nodule-specific expression pattern of *MtAHL1*, suggested that MtAHL1 regulates *NCR169* expression. The DNA pull-down experiments using *M. truncatula* nodule nuclear extracts also identified MtAHL2, which is 75% identical to MtAHL1. The expression patterns of *MtAHL1* and *MtAHL2* were different, with *MtAHL2* showing higher expression in roots with lower expression in nodules than MtAHL1 (Extended Data Fig. 3). Since the AHL family of regulators can act as heterotrimers¹⁴, we thought that MtAHL2 as well as its *G. max* orthologue, GmAHL2 (*Glyma.01G198800*), which we identified via phylogeny (Extended Data Fig. 2) could also be involved in *NCR169* regulation. The binding of MtAHL1, MtAHL2, GmAHL1 and GmAHL2 to the *NCR169* promoter was confirmed with Y1H assays (Extended Data Fig. 4). Probable orthologues of these AHL proteins from *M. truncatula* and *G. max* were also found in *L. japonicus* (Extended Data Fig. 2).

MtAHL1 and MtAHL2 form homomeric and heteromeric complexes in the nucleus and bind to AT-rich sequences required for proper expression of *NCR169*

A. tumefaciens-mediated infiltration in *Nicotiana benthamiana* leaves with constructs encoding green fluorescence protein (GFP) fused to MtAHL1 or MtAHL2 demonstrated nuclear localization of both proteins (Extended Data Fig. 5a). As the AHL proteins are known to act in homo- or hetero-trimeric forms¹⁴, we tested if MtAHL1 interacts with itself and with MtAHL2 using bimolecular fluorescence complementation (BiFC) assays in *N. benthamiana* leaves. Complementation of fluorescence was observed in assays using nYFP-MtAHL1 with both cYFP-MtAHL1 (self-interaction) and cYFP-MtAHL2 (heterologous interaction) (Extended Data Fig. 5b), indicating that MtAHL1 can participate in formation of homotrimeric and heterotrimeric complexes.

The interaction between the minimal promoter region and the nodule-enhanced MtAHL1 protein was assayed with electrophoretic mobility shifts (EMSA) using purified MtAHL1 protein produced in *Escherichia coli* and seven overlapping DNA probes of ~100 bp (A-G) covering the 436 bp promoter region (Fig. 2d). MtAHL1 formed low mobility complexes with the overlapping fragments C and G, but weaker complexes also formed with the other five fragments (Fig. 2e) indicating other possible binding sites.

A 72 bp fragment (H, nucleotides -192 to -121) corresponding to the common region of fragments C and G was generated; MtAHL1 strongly bound to fragment H (Fig. 2e). Within fragment H there is an AT-rich region between nucleotides -157 and -121 that showed high similarity to the *NCR*-specific Motif 4¹⁰. This 37 bp sequence (TTAATAATTATTTTAATTTTCAAAAAGTATTAATTAC), named H0 formed low mobility complexes with MtAHL1, MtAHL2 and with GmAHL1 (Fig. 2f). Analyses of RNA sequencing

reads from public databases such as the Sequence Read Archive (SRA) in NCBI (<https://www.ncbi.nlm.nih.gov/sra>) revealed that transcription of *NCR169* starts 23 nucleotides upstream of the translation start site. The end of the H0 fragment is 98 bp from this deduced transcription start, a good location for a site of regulation. A further seven AT-rich regions with p-values varying from 5.82×10^{-5} to 3.25×10^{-15} were identified on the 436 bp fragment and these could explain the secondary binding of MtAHL1 implied from the retardation of the other five DNA fragments (Fig. 2e).

The 432 bp *NCR169* promoter region fused to *GUS* was infiltrated into *N. benthamiana* leaves either alone or with *MtAHL1* or *MtAHL2* expressed from the ubiquitin promoter. Expression of *GUS* was observed only in the presence of *MtAHL1* or *MtAHL2* (Fig. 2g), confirming that either *MtAHL1* or *MtAHL2* can induce the *NCR169*. The 37 bp H0 sequence between -157 and -121 was deleted from the 432 bp promoter region and the deleted promoter (432 bp Δ H0) was fused to the *NCR169* coding sequence. Unlike the intact 432 bp promoter, which was fully effective for complementing the *ncr169* deletion mutant *dnf7-2*, deletion of the H0 sequence resulted in poor complementation; the plants were only slightly larger and greener than the *dnf7-2* mutant transformed with the empty vector (Fig. 2b, c).

Both MtAHL1 and MtAHL2 are crucial for normal nodule development

The *MtAHL1* and *MtAHL2* genes were mutated in transformed roots of *M. truncatula* using CRISPR/Cas9 genome editing. Genomic modifications were detected by targeted sequencing of the transgenic nodules and this revealed deletions and/or insertions in the *MtAHL1* and *MtAHL2* genes. However, there were large variations in the efficacy of mutagenesis in the different transgenic lines. Each nodule contained some wild-type DNA sequences and the mutant sequence reads varied from 12.4% to 93.4% and from 25.7% to 98.8% in genome-edited

MtAHL1 and *MtAHL2*, respectively (Extended Data Fig. 6a, b). Despite the apparent mosaic nature of these nodules that appear to contain wild-type and knock-out mutant cells, only small, white non-fixing nodules developed on the *MtAHL1* and *MtAHL2* mutant lines (Fig. 3). SYTO 9 (live)/PI (dead) staining of nodule sections revealed successful infection of young nodule cells in zone II, but blocked bacteroid differentiation. Therefore, the interzone and nitrogen fixing zones were absent and the nodule cells contained only dead non-enlarged bacteria. A loss of bacteroid viability was also observed in *M. truncatula dnf7-2* nodules⁹, but in the *dnf7-2* nodules the dead bacteroids were almost fully elongated.

As the CRISPR/Cas9 system did not provide 100% knock-out mutant nodules, expression of both *MtAHL1* and *MtAHL2* was down-regulated with RNA interference (RNAi) using *A. rhizogenes*-mediated root transformation. This resulted in ~70-90% down-regulation of the *MtAHL1* and 65-80% down-regulation of the *MtAHL2* measured by quantitative reverse transcription PCR in the transgenic roots (Extended Data Fig. 6d, e). Nodules that developed on the RNAi lines were small, white, non-fixing, had the same nodule structure as the CRISPR/Cas9 mutant nodules and were devoid of differentiated live bacteroids (Fig. 3, Extended Data Fig. 6c).

Mutation or down-regulation of *MtAHLs* in *M. truncatula* blocked bacteroid development at an earlier stage than observed in *dnf7-2* nodules lacking *NCR169*⁹. This could be due to lack of induction of both *NCR169* and other *NCR* genes. We searched for the H0 motif in 1 kb regions upstream of nodule-expressed *M. truncatula* genes using FIMO in the MEME suite (<https://meme-suite.org/meme/tools/fimo>). The identified genes are highly expressed in the proximal part of zone II, interzone and zone III of the nodules and most of them are *NCRs* (Supplementary Table 1) suggesting that nodule expression of these genes may also require *MtAHLs*. We conclude that knock-out and down regulation of *MtAHLs* may block induction of

other *NCR* genes as well as *NCR169*, causing the observed earlier arrest of bacteroid development.

Non-*NCR* genes required for establishment of the symbiosis could also be regulated by AHL transcription factors. A way of assessing if this is likely is to determine whether mutating *AHL* genes affects the nitrogen-fixing symbiosis in a legume lacking *NCR* peptides. As *ahl* mutants are available in *L. japonicus*, we firstly confirmed that, as in soybean, the *pNCR169::GUS* fusion is expressed in *L. japonicus* (Extended Data Fig. 7a). We then identified *L. japonicus* lines carrying homozygous mutations in the genes (*LjAHL1* and *LjAHL2*) orthologous to *MtAHL1* and *MtAHL2* (Extended Data Fig. 7b). These *L. japonicus* mutants both formed normal pink nodules and the nodulated plants grew well in the absence of added nitrogen (Extended Data Fig. 7c, d). This shows that neither of the *L. japonicus* *AHL* genes regulates a gene required for nitrogen fixation suggesting that the symbiotic defects caused by mutation of *AHL* genes in *M. truncatula* is primarily due to effects on regulation of *NCRs*.

Discussion

Since *NCR* genes are only found in the IRLC legumes, and their expression is developmentally regulated in nodules, we anticipated that they might be regulated by IRLC-specific transcription factors. However, we show that *M. truncatula* *NCR169* is induced in nodules of soybean and *L. japonicus*, in which no *NCR* homologs could be found. Furthermore, the changes in soybean bacteroid size and DNA content induced by expression of *NCR169* are consistent with the *NCR169* peptide being processed and delivered to the soybean bacteroids where it is taken up with the help of the BclA transporter. The initial acquisition of *NCR* genes in IRLC legumes involved the recruitment of the existing nodule-specific protein secretory pathway across the plant-made symbiosome membrane to deliver the mature *NCR* peptides³. We can now infer that

acquisition of *NCR* genes also involved recruitment of nodule-specific promoters regulated by existing transcription factors already present in nodules. A relatively short (436 bp) promoter upstream of the *NCR169* coding region is sufficient for nodule-specific expression of *NCR169* in all three legumes.

We identified two closely-related transcription factors MtAHL1 and MtAHL2 that bound to the previously-identified Motif 4 in the *NCR169* promoter¹⁰; within this motif, we identified a MtAHL1 binding site, that contains an AT-rich region of 37 bp (referred to as H0) required for normal *NCR169* expression. Deletion of this MtAHL binding site severely affected complementation of the *NCR169*-defective *M. truncatula dnf7-2* mutant for symbiotic nitrogen fixation. This H0 motif is conserved in the promoters (1 kb upstream regions) of many *M. truncatula NCR* genes that are highly expressed in the proximal part of zone II, the interzone and zone III of nodules (Supplementary Table 1).

The absence of either of the MtAHL1 and MtAHL2 transcription factors abolished the formation of nitrogen-fixing nodules in *M. truncatula* demonstrating their importance in symbiosis. Although essential for nodule induction of *NCR169*, they might also induce other *NCR* genes. In the *dnf7-2* (*ncr169*) mutant nodule, bacteroid differentiation is arrested late, after significant elongation and enlargement; these bacteroids are unable to fix nitrogen and are rapidly killed resulting in the absence of nitrogen fixation zone III⁹. However, in the absence or with low levels of MtAHL1 and MtAHL2, the nodule bacteria do not show any signs of bacteroid differentiation, are only viable in zone II and are already eliminated in the interzone (Fig. 3). This leads to early arrested growth of the nodules resulting in a spherical shape. These phenotypes imply that MtAHL1 and MtAHL2 play crucial roles in regulation of other *NCR* genes required for full bacteroid differentiation and the development of fully functioning nitrogen-fixing nodules.

A role for AHL family proteins in nodule symbiosis has not been reported previously, although they have been recognized to be important for organ development in other plants. AHL family proteins contain one or two AT-hook(s) and a Plant and Prokaryote Conserved (PPC/DUF296) domain responsible for their interaction with themselves, other AHL proteins and non-AHL proteins¹⁴. They form homo- and hetero-trimeric complexes and through their AT-hook domain(s) bind to DNA resulting in both repression and induction of genes and biological pathways. The *Arabidopsis* AHL15 protein is associated with multiple processes and its loss accelerates axillary meristem (AM) maturation, while its overexpression suppresses AM maturation¹⁵ and, in the absence of hormone treatment, induces somatic embryogenesis that is often accompanied by heterochromatin decondensation and endomitosis¹⁶. Other AHLs are required for the repression of hypocotyl elongation¹⁷, innate immunity¹⁸, the induction of transcriptional regulators involved in patterning and differentiation of reproductive organs¹⁹ and the activity of various transposable elements (TEs) and TE-like repeat-containing genes such as the central floral repressor FLOWERING LOCUS C²⁰. Binding of the AHLs to promoter elements rapidly changed histone H3 acetylation and methylation of the H3K9 residue via forming a complex with proteins participating in histone deacetylation²¹. Other AHLs are transcriptional regulators of the negative feed-back mechanism of gibberellin synthesis²² and regulate responses to (methyl-) jasmonic acid²³.

Given these roles in organ development, AHL family proteins also may be involved in the coordination of nodule development in legumes and the induction of nodule-specific genes such as *NCR* genes or possibly even their repression in other tissues. Of the more than 25 *AHL* genes of *M. truncatula*, at least 12 were expressed in nodules with variable patterns in the different nodule zones¹³. This might promote the formation of different AHL trimers that could differentially regulate different groups of genes. The lack of redundancy of *MtAHL1* and

MtAHL2 in *M. truncatula* nodules could be explained by each protein acting in different complexes or by the formation of different heterotrimeric complexes required for gene expression at different stages of nodule development. It seems likely that the symbiotic defects observed in the *MtAHL* knockdown lines and mutants were due to effects on regulation of both *NCR169* and some of the many other genes (primarily *NCRs*) induced in nodules²⁴. If correct, this could explain why we were unable to distinguish different phenotypes after knocking down or mutating *MtAHL1* or *MtAHL2*.

The lack of an effect on symbiotic nitrogen fixation in the *L. japonicus* *Ljahl1* and *Ljahl2* mutants could have different explanations. Possibly in *L. japonicus* these AHLs do not induce genes required for nitrogen fixation. Alternatively, in *L. japonicus* they may be functionally redundant in contrast to what we observed in *M. truncatula*. Such a difference could reflect the different patterns of nodule development in *L. japonicus* and *M. truncatula*. In *M. truncatula*, *MtAHL2* expression in roots decreased as nodules developed and conversely, low expression of *MtAHL1* in roots increased strongly during nodule development. This opens the possibility that different ratios of AHL1 and AHL2 in trimeric complexes could differentially regulate *NCR169* expression at different stages during nodule development. Possibly in determinate *L. japonicus* nodules such complementary expression is not required. In soybean and *L. japonicus*, the expression of *AHL2* is higher in both roots and nodules than that of *AHL1* (Extended Data Fig. 8) and their pattern of expression would not appear to fit with differential expression during nodule development.

The identification of two new transcription factors required for development of symbiotic nitrogen fixation in legume nodules opens a new phase in the analysis of the development of symbiotic nitrogen-fixing nodules. What other genes are regulated by *MtAHL1* and *MtAHL2* during nodule development? Is their essential role in nitrogen fixation limited to expression of

NCR genes as implied from the preliminary observation that mutation of each gene in *L. japonicus* did not block symbiotic nitrogen fixation? Do MtAHL1 and/or MtAHL2 regulate other genes in *M. truncatula* but not in *L. japonicus* nodules? Does MtAHL2 play some role in root development based on the observation that it is expressed in non-nodulated roots? Do other AHLs expressed in non-nodule tissues bind to the identified promoter element to suppress *NCR169* expression? Do these AHLs play a role during the development of determinate nodules with terminally differentiated bacteroids (i.e. on *Aeschynomene*, *Arachis*) or indeterminate nodules with non-terminally differentiated bacteroids (for example, on *Leucaena*)? Answering these questions will shed light on the mechanisms governing the development of symbiotic nitrogen fixation in legumes.

Interacting proteins	Gene ID	UniProt ID
<i>Proteins interacting with 1181 bp NCR169 promoter in Y1H screening</i>		
MtAHL-1	Medtr4g098450	G7JGG2
BHLH domain class transcription factor	Medtr4g081370	B7FHK4
MYB-like transcription factor	Medtr5g037080	G7K6Z9
TCP family transcription factor	Medtr7g028160	G7L283
transcription factor VOZ1-like protein	Medtr4g088125	A0A072UMQ9
BEL1-related homeotic protein	Medtr8g098815	A0A072TUA6
Linker histone H1 and H5 family protein	Medtr6g079520	A0A072UBH0
<i>M. truncatula proteins interacting with 382 bp NCR169 promoter in DNA pull-down assay</i>		
MtAHL1	Medtr4g098450	G7JGG2
MtAHL2	Medtr7g080980	G7KSI4
MtPHD1	Medtr4g015830	A0A072UGM4
MtPHD2	Medtr1g015185	A0A072VDQ0
Myb protein	Medtr7g020870	G7KSV7
Myb protein	Medtr7g103390	A0A072U4N3
MtPurα	Medtr4g083230	G7JHJ7
Zinc finger protein	Medtr1g053960	A0A072VIQ3
<i>G. max proteins interacting with 382 bp NCR169 promoter in DNA pull-down assay</i>		
GmAHL1	Glyma.18G247200	C6TMY6
GmPHD1	Glyma.09G107000	C6T7X8
GmPurα	Glyma.07G225200	I1KMC5
SHOOT2 protein	Glyma.03G189600	C6TI90
Trihelix protein	Glyma.18G281800	I1N4Y2

Table 1 | Interactors of *NCR169* promoter regions identified in Y1H library screens (1181 bp) and DNA affinity chromatography (382 bp) from both *M. truncatula* and *G. max*.

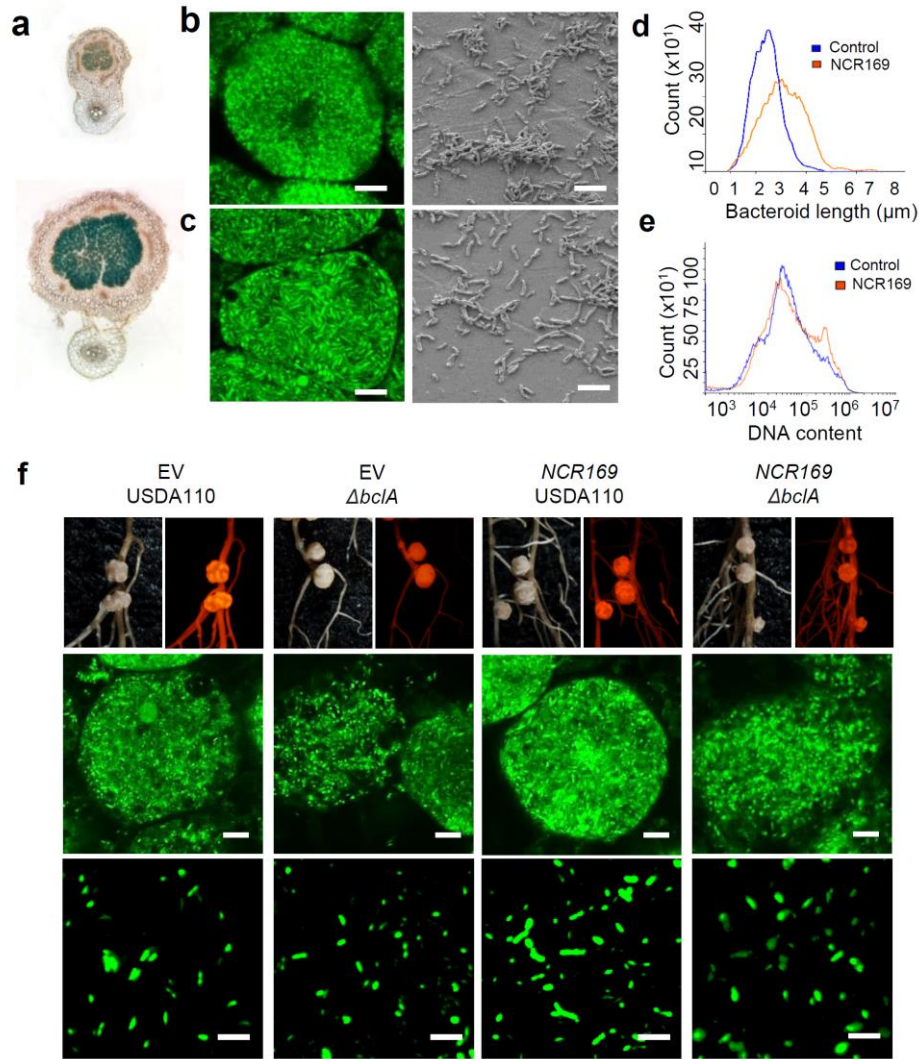


Fig. 1 | Analysis of the effects of expression of the *M. truncatula* NCR169 gene in soybean nodules. **a**, Expression of the 1181 bp *NCR169* promoter::*GUS* reporter in young (Upper) and mature (Lower) soybean nodules detected with blue staining due to GUS enzyme activity. Unlike the bacteroids in control soybean nodules formed on roots transformed with the empty vector (**b**), bacteroids in nodules formed on roots transformed with *NCR169* (**c**) were elongated as assessed by confocal microscopy of SYTO 13-stained soybean nodule sections (Left) and scanning electron microscopy of isolated bacteroids (Right). **d**, The lengths of 500 bacteroids isolated from control (blue trace) and *NCR169*-expressing (orange trace) nodules were measured with a Nano Measurer and plotted as frequency distributions. **e**, The DNA content of isolated and

349 SYTO 13-stained bacteroids from the control (blue trace) and the *NCR169* expressing (orange
350 trace) nodules were measured by flow cytometry and plotted as frequency distributions. **f**,
351 Morphologies of bacteroids were compared in control nodules formed on roots transformed with
352 the empty vector (EV) and in *NCR169*-expressing (*NCR169*) nodules in soybean inoculated with
353 wild-type *B. japonicus* USDA110 or its *bclA* mutant ($\Delta bclA$). The upper images show the
354 nodules on transformed roots expressing the vector-derived red fluorescence. The middle and
355 bottom images show confocal fluorescence microscopy images of SYTO 13-stained nodules
356 sections and isolated bacteroids, respectively. Scale bars indicate 4 μ m.

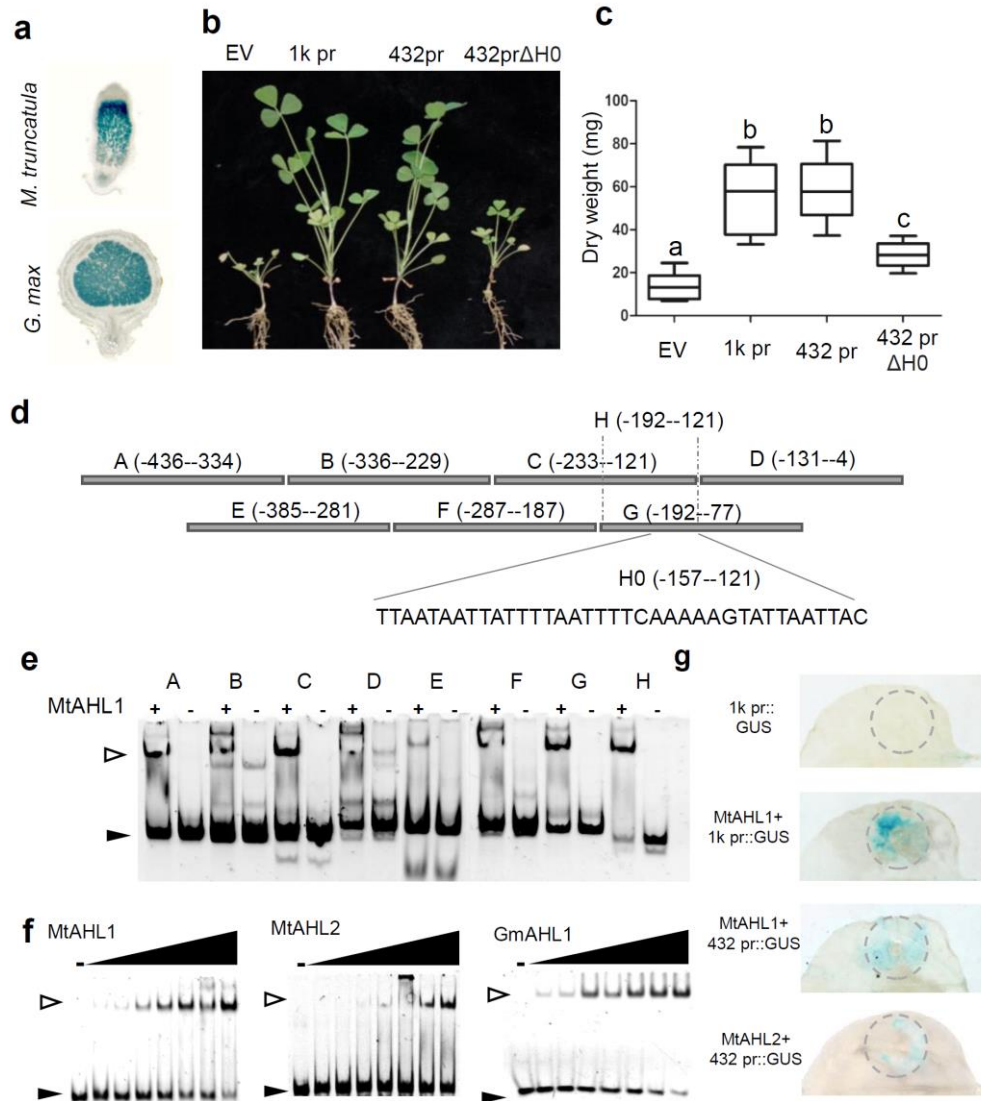


Fig. 2 | Identification of the region of the *NCR169* promoter essential for expression in nodules and for binding of MtAHLs. a, GUS enzyme activity (blue staining) induced by the 432 bp promoter region in nodules formed on hairy roots of *M. truncatula* and *G. max*. **b**, Complementation of the *NCR169* mutant *dnf7-2* with the *NCR169* coding sequence preceded either by the 1181 bp (1k pr) or the 432 bp promoter region (432 pr) or the 432 bp promoter region deleted for the H0 sequence (432 pr ΔH0). **c**, Average dry weights of plants grown as shown in **(b)** were calculated (n=10). In the boxplot, the centre black line represents the median; the box limits are the upper and lower quartiles; the whiskers represent the lowest or highest data point within the 1.5 interquartile range of the lower or upper quartile. The different letters above

the bars indicate significant differences (two-tailed unpaired t test, $P < 0.0006$). **d**, The locations of probes used for EMSA assay are shown. **e**, Interactions of probes A-H with purified His-tagged MtAHL1. **f**, H0 region was used to assess its binding to purified MtAHL1, MtAHL2, and GmAHL1. The black arrow indicates the free probe and the hollow arrow indicates retarded bands (**e** and **f**). **g**, Induction of the *NCR169* promoter in *N. benthamiana* was assayed. The promoters used were either the 1181 bp (1k pr::GUS) or 432 bp (432 pr::GUS) regions upstream of the gene encoding GUS. *A. tumefaciens* carrying *MtAHL1* or *MtAHL2* was co-infiltrated as indicated and the leaves were stained with X-Gluc for 72 h after infiltration. Dashed circle indicates the infiltrated area.

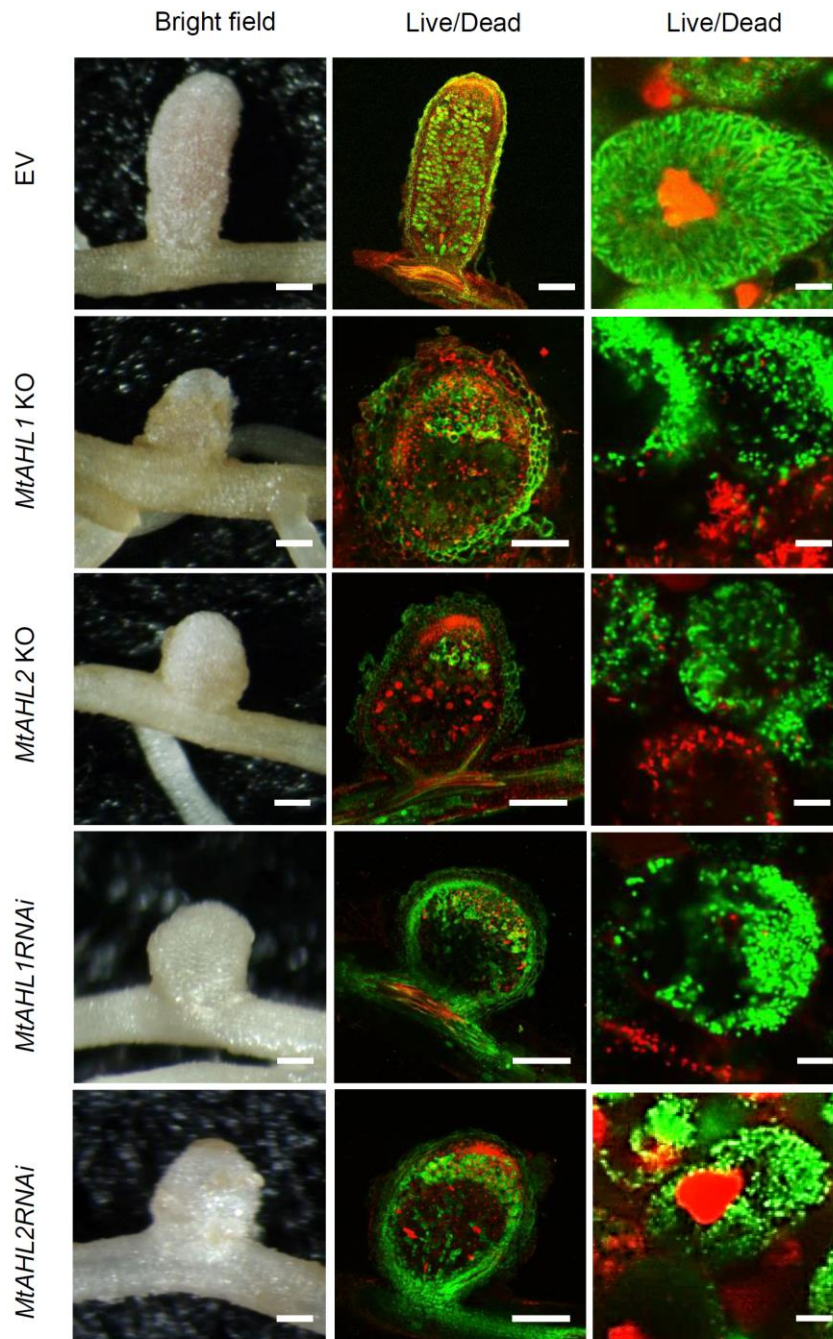


Fig. 3 | Assays of the effects of CRISPR/Cas9-mediated knockout mutations (KO) and RNAi-mediated gene silencing (RNAi) of *MtAHL1* or *MtAHL2* in *A. rhizogenes* transformed roots of *M. truncatula*. CRISPR/Cas9 knock-out mutations were induced using guide sequences described in Extended Data Fig. 6. RNA interference was achieved by expressing hairpin constructs specifically targeting the *MtAHL1* or *MtAHL2* genes. Sections of transformed root

nodules were examined by confocal microscopy after Live/Dead staining using SYTO 9 for staining live bacteroids green and propidium iodide (PI), which stains dead bacteroids and the nuclei of plant cells red. Scale bar indicates 1 mm on whole nodule images and 4 μm on the symbiotic cell images (right column).

Methods

Plant Materials, Hairy Root Transformation and Nodulation Assay

The plant materials were *M. truncatula* A17, soybean (*G. max* cv. Williams 82) and *L. japonicus* Gifu. Roots of these legumes were transformed using *Agrobacterium rhizogenes* strain ARqua-1 or K599 carrying specific vectors^{25,26}. Nodules were induced on soybean by *B. japonicum* CB1809 when the *NCR169* promoter activity was investigated, and by *B. japonicum* USDA110 wild-type its *AbcIA* mutant, when the effects of *NCR169* on bacteroids was tested. *L. japonicus* was inoculated with *M. loti* R7A and *M. truncatula* by *S. medicae* WSM419. Plants for nodulation were grown in vermiculite and fertilized once per week with 1g L⁻¹ Plant-Prod fertilizer (0-15-40, N-P-K, Brampton, ON, Canada) in a growth chamber programmed for 16 h light and 8 h dark, at 28 °C/22 °C day/night for soybean and at 22 °C/20 °C day/night for *M. truncatula* and *L. japonicus*.

Plasmid and Vectors

The activity of the *NCR169* promoter was analyzed using pCAMBIA3301 (CAMBIA, Canberra, Australia) modified by replacement of the CaMV35S promoter upstream of the GUS reporter either with the 1181 bp or 432 bp promoter regions from *NCR169*, or with the deleted derivative of the 432 bp fragment lacking nucleotides -157 bp to -121 bp. To assess the effect of *NCR169* on bacteroids in soybean nodules, the gene including the 1181 bp promoter and all the exons and introns was firstly introduced into pENTR2B (Thermo Fisher Scientific) donor vector and then ligated into pKGW-RR-MGW²⁷ destination vector via LR-clonase reaction. For complementation of the *M. truncatula* *dnf7-2* mutant, the genomic fragment of *NCR169* was cloned into pCAMBIA2201 (CAMBIA, Canberra, Australia) downstream of the 1181 bp or the 432 bp promoter regions, or the 432 bp region deleted for nucleotides -157 to -121 bp. For generating bait strain used in the yeast one hybrid assays, the 1181 bp promoter region of

NCR169 was cloned into the XbaI and EcoRI sites of pHIS3NB^{28,29} and then cloned together with *HIS3* gene into the NotI and BamHI sites of pINT1^{28,29} to make pINT1-*NCR169pr-HIS3*.

For CRISPR/Cas9 gene knock-out of the *MtAHL* genes, oligonucleotides MtAHL1gRNA/MtAHL1gRNARc and MtAHL2gRNA/MtAHL2gRNARc were annealed and ligated into the BsaI-digested pKSE401³⁰. For RNA interference, cDNA fragments were cloned into pCR8/GW/TOPO and then recombined into the destination vector pUB-GWS-GFP³¹.

Proteins were expressed in *E. coli* strain BL21 carrying the constructs with the full-length gene coding sequences inserted into EcoRI/SalI sites of the pET28a (+) plasmid (Novagen, USA).

The pUBC vector system³² was used for cellular localization and for detection of protein-protein interactions using bimolecular fluorescence complementation (BiFC). Gene coding regions were PCR amplified from cDNA and the products were ligated into pCR8/GW/TOPO via TOPO cloning (Invitrogen, USA). The resulting donor clones were used in LR-mediated recombination into pUBC-GFP-DEST for analyses of cellular localization and pUBC-nYFP-DEST/pUBC-cYFP-DEST for BiFC assays.

Primers used for creating the above constructs are shown in Supplementary Table 2.

Protein Purification and EMSA assays

Genes encoding His-tagged MtAHL1, MtAHL2 and GmAHL1 were generated by PCR using primers shown in Supplementary Table 2, cloned into pET28a (+) and introduced into *E. coli* BL21. Overnight cultures (2 ml) grown at 37 °C were inoculated into 200 mL LB medium, and grown in a shaking incubator to OD₆₀₀ 0.5. Protein expression was then induced overnight at room temperature in shaken flasks by adding 0.05% L-arabinose and 0.25 mM isopropyl β-D-thiogalactoside. Bacteria were then pelleted by centrifugation (4000 × g, 10 minutes, 4 °C), washed once with ice-cold water, resuspended in 4 mL of BS/THES buffer³³ (22 mM Tris HCl, pH 7.5, 4.4 mM EDTA, 8.9% (w/v) sucrose, 62 mM NaCl, 10 mM HEPES, 5 mM CaCl₂, 50

mM KCl, and 12% glycerol) supplied with 0.3% cOmplete protease inhibitor cocktail (Roche Diagnostics Ltd, Mannheim, Germany). The cells were disrupted on ice with cyclic sonication generated by UIS250v ultrasonic processor (Hielscher Ultrasonics GmbH, Germany) with a 5 mm sonotrode (0.9 second sonication, 0.1 second pause, 90% amplitude (15W) for 2 minutes) and then centrifuged at $16,000 \times g$ at 4 °C for 2 minutes and the His-tagged proteins were purified from the supernatant using a column of HisPur cobalt resin (Thermo Fisher Scientific, Rockford, IL). Bound proteins were eluted with 1 mL BS/THES buffer containing 150 mM imidazole as described by the manufacturer. EMSA assays were conducted as described by Chen³⁴, except that the bands were visualized by SYBR Gold staining. In the EMSA assay, a typical amount of 10 ng probe and 200 ng purified protein was used for one reaction. Titration EMSA was performed with fixed 10 ng of probe and purified protein in the range of 0, 25, 50, 100, 150, 200, 250 and 300 ng.

DNA affinity pull-down assays

Nodule nuclei were isolated from *M. truncatula* and *G. max* by chopping nodules with a razorblade in pre-chilled nuclear isolation buffer (45 mM MgCl₂, 30 mM trisodium citrate, 20 mM MOPS, 0.1% Triton X-100, pH 7.2-7.4). The suspensions were filtered firstly through a 100 µm pore size nylon mesh to remove cell debris. Then the nuclei going through a 30 µm filter were collected by centrifugation at $1500 \times g$ for 10 minutes at 4 °C and resuspended in BS/THES buffer containing 0.3% cOmplete protease inhibitor cocktail. Nuclear proteins were released by vortexing 5 seconds every 3 minutes for 15 minutes.

For the DNA affinity pull-down experiment, 382 bp of *NCR169* promoter sequence (-386 bp to -4 bp) was amplified by two PCR reactions in 1.5 ml volume to generate probes with biotin label on either end. PCR fragments were precipitated by adding 1/10 volume of 3 M sodium-acetate and 1 volume of isopropanol at -80 °C overnight. Precipitated DNA was collected by

centrifugation at 13,000 rpm for 10 minutes and washed 3 times with 70% ethanol, then dissolved in nuclease free H₂O. The purity and concentration of the DNA fragments were checked by agarose gel electrophoresis and optical density measurements using Nanodrop.

DNA affinity pull-down assays used 200 µL of Dynabeads M280 (binding 40-80 µg DNA) to bind the bait fragment pair essentially as described³³ except that in the final step, the proteins bound to the beads were digested with trypsin to release peptides that were identified by mass spectrometry. The proteins originating from *M. truncatula* and *G. max* nodules were aligned with BLAST on NCBI (<http://blast.ncbi.nlm.nih.gov/>) to find homologues.

Yeast One Hybrid Assays

The pINT1-*NCR169pr-HIS3* Y1H assay plasmid was linearized by EheI digestion and introduced by transformation²⁹ into *S. cerevisiae* strain Y187 to generate the bait strain which was then transformed with plasmids generated from a *M. truncatula* EST library¹². Transformants were plated on SD-Leu-His plates. Prey sequences were identified by sequencing and searches were done by BLAST on the Phytozome website (<http://phytozome.jgi.doe.gov/>) using the *M. truncatula* database.

Transient protein expression in *N. benthamiana* leaves

A. tumefaciens AGL-1 strains transformed with indicated constructs were grown and prepared for transient expression as previously described³⁵. The cultures were resuspended at OD₆₀₀ 0.2 in infiltration buffer (10 mM MES pH 5.7, 10 mM MgCl₂ and 100 mM acetosyringone). For co-expression and BiFC assays, suspensions of multiple constructs were mixed in equal ratios and infiltrated into expanding leaves of four-week-old *N. benthamiana* plants. The samples for microscopic imaging or GUS staining were collected 3 days after infiltration.

Microscopy

Nodules after SYTO 13 or live/dead staining were observed by confocal microscopy as

described⁹ using a Leica SP5 laser scanning confocal microscope (Leica Microsystems GmbH, Wetzlar, Germany). Transient GFP and YFP signals in *N. benthamiana* were observed by FluoView FV1000 (Olympus Corporation, Tokyo, Japan) confocal microscope.

GUS staining

GUS staining of nodules and nodule sections was carried out as described⁹. Samples were fixed in ice-cold 90% acetone for 1 hour, and then stained overnight at 37 °C with X-Gluc staining solution (containing 50 mM phosphate buffer pH 7.2, 0.5 mM K₃Fe(CN)₆ (potassium ferricyanide), 0.5 mM K₄Fe(CN)₆ (potassium ferro-cyanide) and 2 mM X-Gluc (Thermo Fisher Scientific)). *A. tumefaciens*-treated *N. benthamiana* leaves were collected 3 days after infection and were vacuum infiltrated for 15 minutes with X-Gluc staining solution containing 0.1% Triton X-100 and then incubated at 37 °C for 24 hours. Chlorophyll was removed by washing the leaves with absolute ethanol before photography.

Data availability

The data supporting the findings of this study are available within the paper and its supplementary information files.

References

1. Mergaert, P. et al. Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*–legume symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5230–5235 (2006).
2. Oono, R. & Denison, R. F. Comparing symbiotic efficiency between swollen versus nonswollen rhizobial bacteroids. *Plant Physiol.* **154**, 1541–1548 (2010).
3. Van de Velde, W. et al. Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science* **327**, 1122–1126 (2010).

4. Mergaert, P. et al. A novel family in *Medicago truncatula* consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. *Plant Physiol.* **132**, 161–173 (2003).
5. He, J. et al. The *Medicago truncatula* gene expression atlas web server. *BMC Bioinform.* **10**, 1–9 (2009).
6. Glazebrook, J., Ichige, A. & Walker, G. C., A *Rhizobium meliloti* homolog of the *Escherichia coli* peptide-antibiotic transport protein SbmA is essential for bacteroid development. *Genes Dev.* **7**, 1485–1497 (1993).
7. Haag, A. F. et al. Protection of *Sinorhizobium* against host cysteine-rich antimicrobial peptides is critical for symbiosis. *PLoS Biol.* **9**, e1001169 (2011).
8. Barrière, Q. et al. Integrated roles of BclA and DD-carboxypeptidase 1 in *Bradyrhizobium* differentiation within NCR-producing and NCR-lacking root nodules. *Sci. Rep.* **7**, 1–13 (2017).
9. Horváth, B. et al. Loss of the nodule-specific cysteine rich peptide, NCR169, abolishes symbiotic nitrogen fixation in the *Medicago truncatula* *dnf7* mutant. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15232–15237 (2015).
10. Nallu, S. et al. Regulatory patterns of a large family of defensin-like genes expressed in nodules of *Medicago truncatula*. *PLoS One* **8**, e60355 (2013).
11. Guefrachi, I. et al. *Bradyrhizobium* BclA is a peptide transporter required for bacterial differentiation in symbiosis with *Aeschynomene* legumes. *Mol. Plant-Microbe Interact.* **28**, 1155–1166 (2015).
12. Györgyey, J. et al. Analysis of *Medicago truncatula* nodule expressed sequence tags.

Mol.Plant-Microbe Interact. **13**, 62–71 (2000).

13. Roux, B. et al. An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J.* **77**, 817–837 (2014).
14. Zhao, J., Favero, D. S., Qiu, J., Roalson, E. H. & Neff, M. M. Insights into the evolution and diversification of the AT-hook Motif Nuclear Localized gene family in land plants. *BMC Plant Biol.* **14**, 1–19 (2014).
15. Karami, O. et al. A suppressor of axillary meristem maturation promotes longevity in flowering plants. *Nat. Plants* **6**, 368–376 (2020).
16. Karami, O. et al. An *Arabidopsis* AT-hook motif nuclear protein mediates somatic embryogenesis and coinciding genome duplication. *Nat. Commun.* **12**, 1–13 (2021).
17. Zhao, J., Favero, D. S., Peng, H. & Neff, M. M. *Arabidopsis thaliana* AHL family modulates hypocotyl growth redundantly by interacting with each other via the PPC/DUF296 domain. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E4688–E4697 (2013).
18. Lu, H., Zou, Y. & Feng, N. Overexpression of *AHL20* negatively regulates defenses in *Arabidopsis*. *J. Integr. Plant Biol.* **52**, 801–808 (2010).
19. Ng, K.-H., Yu, H. & Ito, T. AGAMOUS controls *GIANT KILLER*, a multifunctional chromatin modifier in reproductive organ patterning and differentiation. *PLoS Biol.* **7**, e1000251 (2009).
20. Xu, Y. et al. A matrix protein silences transposons and repeats through interaction with retinoblastoma-associated proteins. *Curr. Biol.* **23**, 345–350 (2013).
21. Yun, J., Kim, Y.-S., Jung, J.-H., Seo, P. J. & Park, C.-M. The AT-hook motif-containing

protein AHL22 regulates flowering initiation by modifying *FLOWERING LOCUS T* chromatin in *Arabidopsis*. *J. Biol. Chem.* **287**, 15307–15316 (2012).

22. Matsushita, A., Furumoto, T., Ishida, S. & Takahashi, Y. AGF1, an AT-hook protein, is necessary for the negative feedback of *AtGA3ox1* encoding GA 3-oxidase. *Plant Physiol.* **143**, 1152–1162 (2007).

23. Vom Endt, D., e Silva, M. S., Kijne, J. W., Pasquali, G. & Memelink, J. Identification of a bipartite jasmonate-responsive promoter element in the *Catharanthus roseus ORCA3* transcription factor gene that interacts specifically with AT-Hook DNA-binding proteins. *Plant Physiol.* **144**, 1680–1689 (2007).

24. Mergaert, P., Kereszt, A. & Kondorosi, E. Gene expression in nitrogen-fixing symbiotic nodule cells in *Medicago truncatula* and other nodulating plants. *Plant Cell* **32**: 42–68 (2020).

25. Kereszt, A. et al. *Agrobacterium rhizogenes*-mediated transformation of soybean to study root biology. *Nat. Protoc.* **2**, 948–952 (2007).

26. Boisson-Dernier, A. et al. *Agrobacterium rhizogenes*-transformed roots of *Medicago truncatula* for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. *Mol. Plant-Microbe Interact.* **14**, 695–700 (2001).

27. Op den Camp, R. et al. LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume *Parasponia*. *Science* **331**, 909–912 (2011).

28. Meijer, A. H., Ouwerkerk, P. B. & Hoge, J. H. C. Vectors for transcription factor cloning and target site identification by means of genetic selection in yeast. *Yeast* **14**, 1407–1416 (1998).

29. Ouwerkerk, P. B. & Meijer, A. H. Yeast one-hybrid screening for DNA-protein interactions. *Curr. Protoc. Mol. Biol.* **55**, 12–12 (2001).
30. Xing, H.-L. et al. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol.* **14**, 1–12 (2014).
31. Maekawa, T. et al. Polyubiquitin promoter-based binary vectors for overexpression and gene silencing in *Lotus japonicus*. *Mol. Plant-Microbe Interact.* **21**, 375–382 (2008).
32. Grefen, C. et al. A ubiquitin-10 promoter-based vector set for fluorescent protein tagging facilitates temporal stability and native protein distribution in transient and stable expression studies. *Plant J.* **64**, 355–365 (2010).
33. Jutras, L., Verma, A. & Stevenson, B. Identification of novel DNA-binding proteins using DNA-affinity chromatography/pull down. *Curr. Protoc. Microbiol.* **24**, 1–1F (2012).
34. Chen, R. A general EMSA (gel-shift) protocol. *Bio. Protoc.* e24–e24 (2011).
35. Wroblewski, T., Tomczak, A. & Michelmore, R. Optimization of *Agrobacterium*-mediated transient assays of gene expression in lettuce, tomato, and *Arabidopsis*. *Plant Biotechnol. J.* **3**, 259–273 (2005).
36. Bailey, T. L. & Gribskov, M. Combining evidence using p-values: application to sequence homology searches, *Bioinformatics* **14**, 48–54 (1998).
37. Mun, T., Bachmann, A., Gupta, V., Stougaard, J. & Andersen, S. U. *Lotus* Base: An integrated information portal for the model legume *Lotus japonicus*. *Sci. Rep.* **6**, 1–18 (2016).
38. Winter, D. et al. An “Electronic Fluorescent Pictograph” browser for exploring and analyzing large-scale biological data sets. *PLOS ONE* **2**, e718 (2007).
39. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence

analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).

40. Małolepszy, A. et al. The *LOREI* insertion mutant resource. *Plant J.* **88**, 306–317 (2016).

41. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).

Acknowledgments: We thank X. Li from the Huazhong Agricultural University for providing seeds of *G. max* cv. Williams 82 and for providing aid in the screen of target proteins in soybean. This work was supported by the Frontline Research project (KKP129924) to E. K.; OTKA grant (K128486) from the Hungarian National Research, Development and Innovation Office to A. K.; the Balzan Prize (2018) from the International Balzan Foundation to E. K.; and the China Scholarship Council fellowship to S. Z. and T. W.

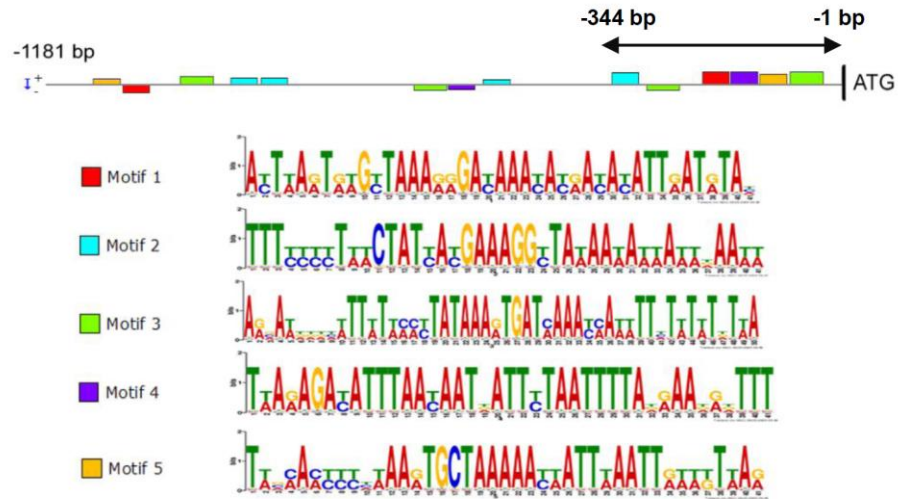
Author contributions: S. Z., A. K., J. A. D., and E. K. conceived the project, designed the experiments, analyzed the data and wrote the manuscript. S. Z. and T. W. conducted most of the experiments. R. M. L. performed microscopic and bioinformatic analysis. A. P.-S. performed DNA affinity pull-down assay and analyzed the result together with S. Z.

Competing interests: Authors declare that they have no competing interests.

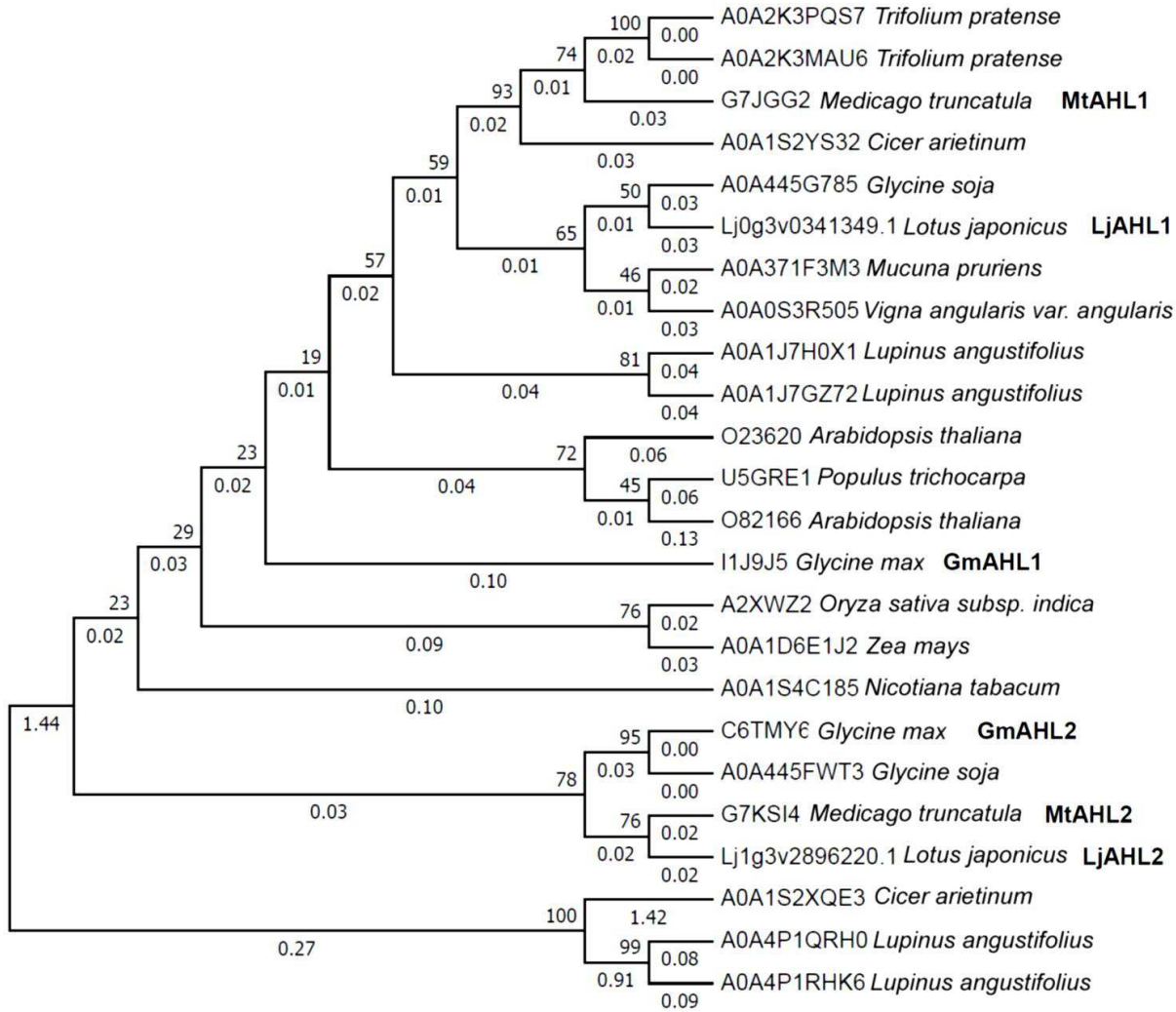
Additional information

Supplementary Information is available for this paper.

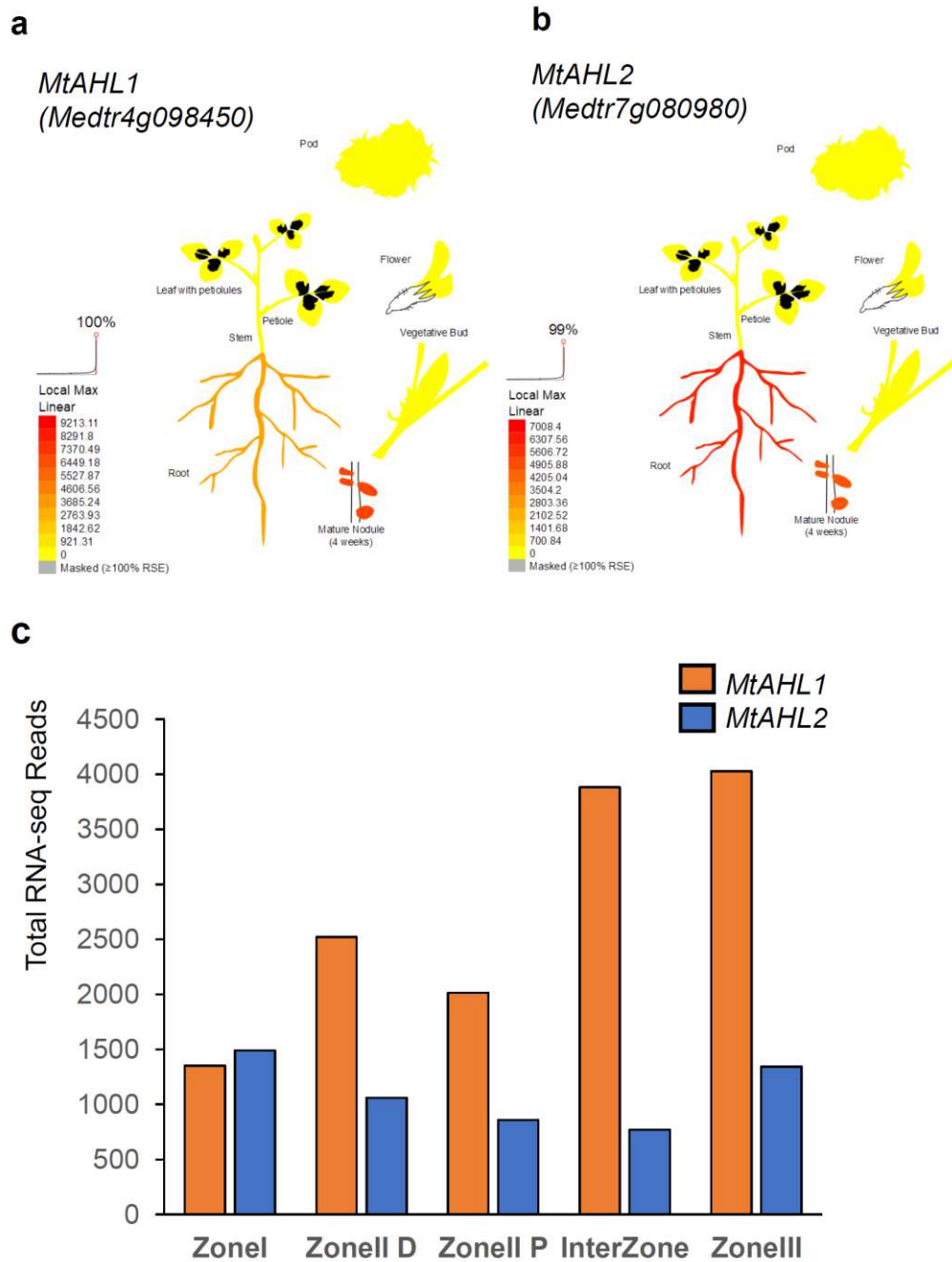
Correspondence and requests for materials should be addressed to Eva Kondorosi. Email: eva.kondorosi@gmail.com



Extended Data Fig. 1 | Occurrence of the five reported motifs in the 1181 bp promoter preceding the *NCR169* translation start. The analysis was conducted with MAST tool in the MEME Suite³⁶ (<https://meme-suite.org/meme/index.html>). Sequences matched to the five motifs conserved in promoters of *NCR* genes as reported¹⁰ are highlighted. The heights of the rectangles represent the levels of similarity. The sequences of the conserved five motifs, are densely clustered in the upstream 344 bp region, which is indicated. The conservation of nucleotides in the sequences of the five motifs are shown by letter height.

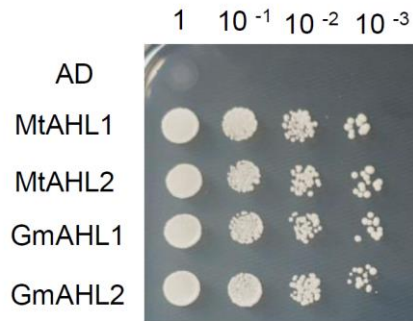


Extended Data Fig. 2 | Phylogenetic tree of predicted AHL family proteins from selected plant species. Protein sequences were acquired from UniProt data base, except for those of *L. japonicus*, which were obtained from *Lotus* Base³⁷. Protein sequences were aligned in MEGA 7³⁸ with default settings and phylogenetic tree was generated with the aligned sequences using the neighbour-joining distance method. Stability of the branch topology in phylogenetic trees was tested using 1,000 bootstrap replicates. Bootstrap values are indicated. The *M. truncatula* (Mt), *G. max* (Gm) and *L. japonicus* (Lj) AHL1 and AHL2 proteins described in the text are indicated in bold.

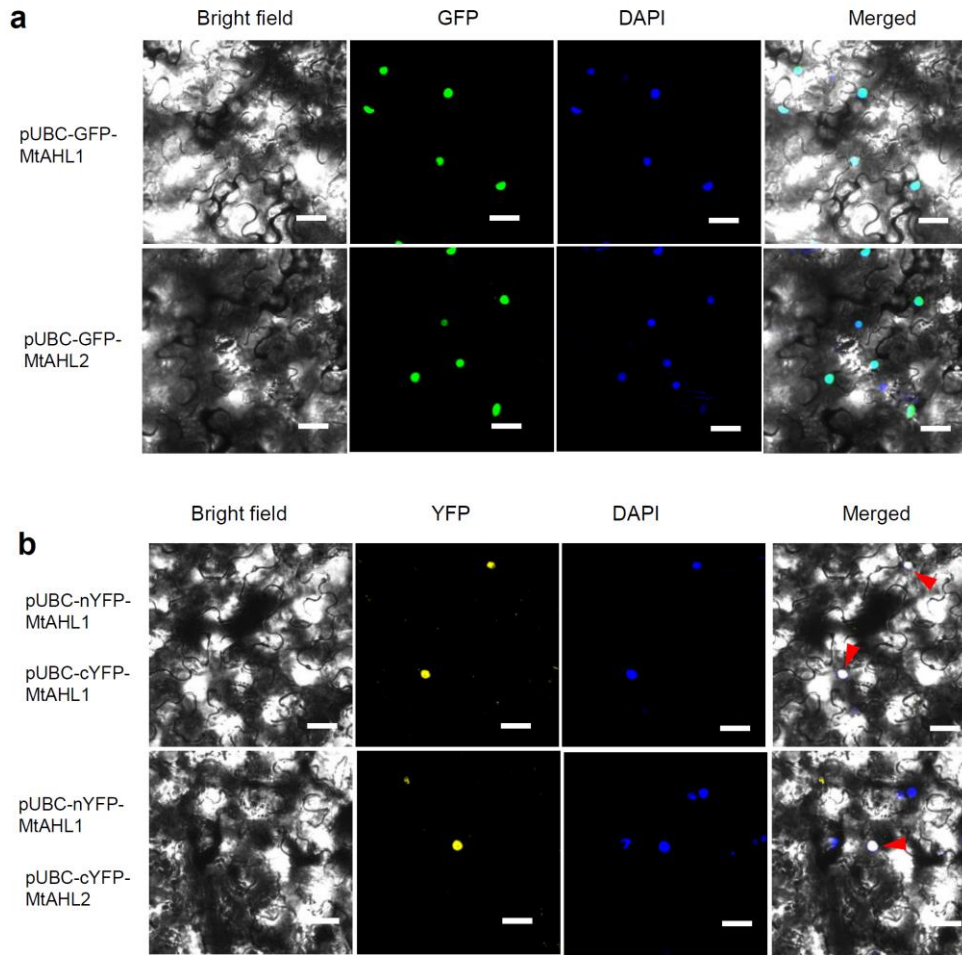


Extended Data Fig. 3 | Diagram showing expression levels of *MtAHL1* and *MtAHL2* in different tissues and in different zones of the root nodule of *M. truncatula*. **a**, The expression level of *MtAHL1* (*Medtr4g098450*) is represented in different tissues by a color gradient, showing strong expression in nodules and lower expression in roots. **b**, The expression level of *MtAHL2* (*Medtr7g080980*) is stronger in roots and weaker in nodules. The images were

generated using ePlant browser³⁹ (http://bar.utoronto.ca/eplant_medicago/). The colour gradient is generated in “Local Max” mode to discern expression pattern of the specific gene, rather than to indicate the actual expression level. **c**, The expression levels of *MtAHL1* and *MtAHL2* in the root nodule zones Zone I, Zone II distal (D) and proximal (P) regions, Inter Zone and Zone III as defined previously¹³.

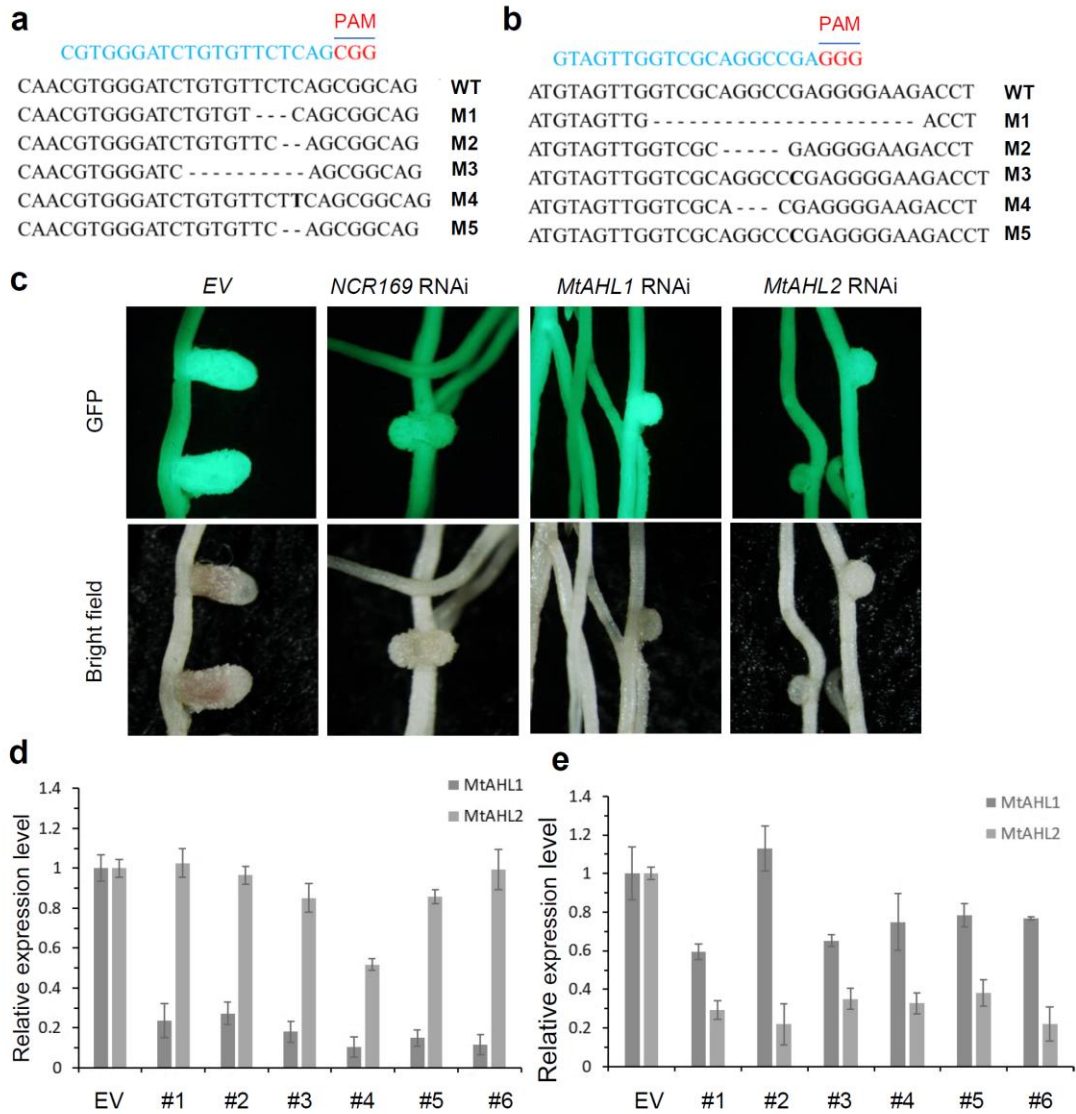


Extended Data Fig. 4 | Detection of binding of the 1181 bp promoter region of *NCR169* to MtAHL1, MtAHL2, GmAHL1 or GmAHL2 in yeast one hybrid (YH1) assays. The yeast bait strain *S. cerevisiae* Y187 carrying pINT1-*NCR169* *pr-HIS3* was transformed with plasmid pAD-GAL4 carrying *MtAHL1*, *MtAHL2*, *GmAHL1* or *GmAHL2*. The transformants were then diluted in 10-fold steps before cultivation on SD/Leu-His plates where growth indicates a positive interaction. The control strain carries the empty pAD-GAL4 vector (AD).



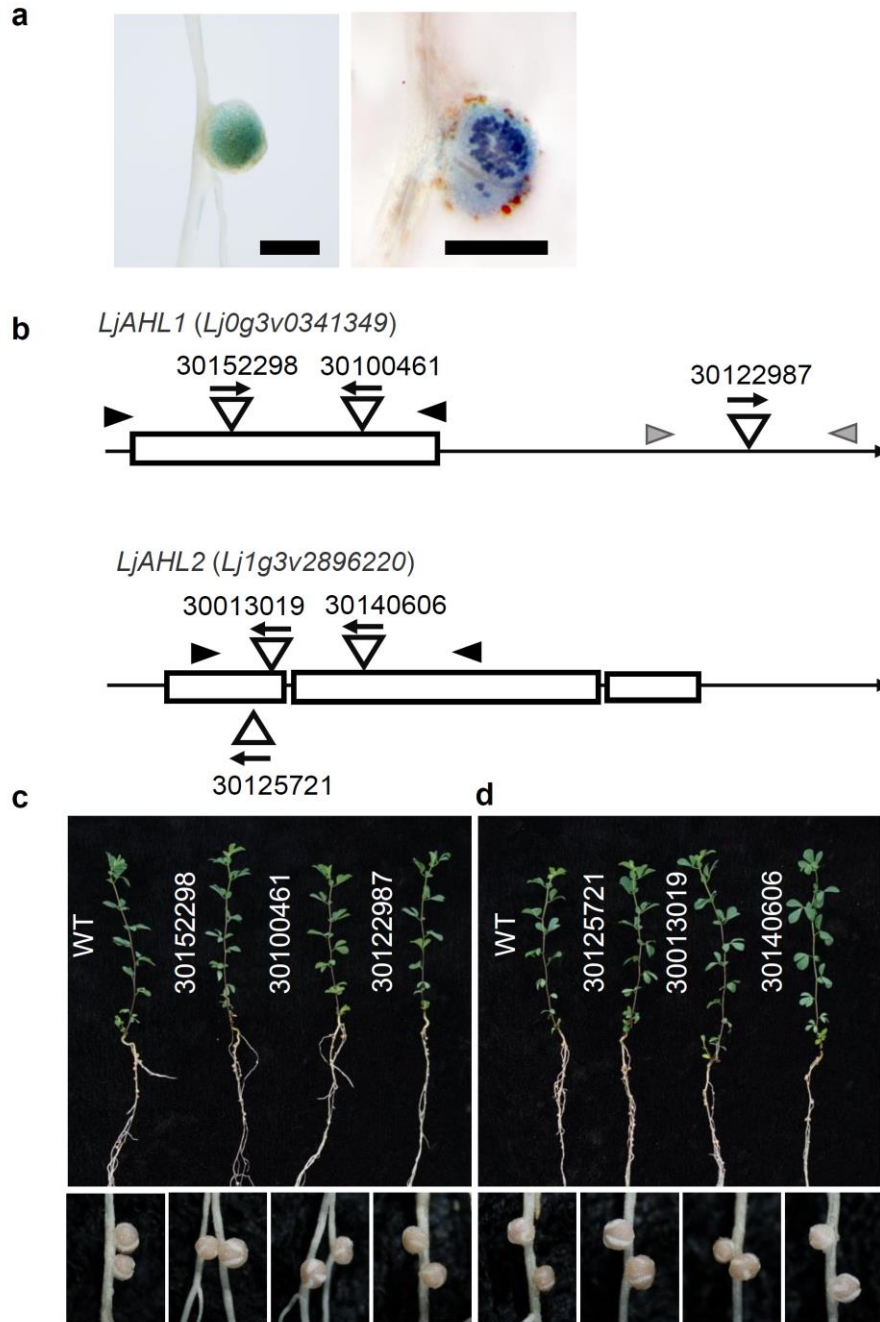
Extended Data Fig. 5 | Detection of subcellular localization and interaction of MtAHLs expressed transiently in *N. benthamiana* leaves. **a**, GFP expressed from the ubiquitin promoter (pUBC-GFP) was fused to the N-terminus of MtAHL1 or MtAHL2 and infiltrated into *N. benthamiana* leaves. The fluorescence images from the GFP fusion proteins and DAPI-stained nuclei were obtained by confocal microscopy. The fluorescence images were then combined with the bright field images to generate merged images showing sub-cellular localization. **b**, BiFC between MtAHL1 or MtAHL2 fused to the C-terminal half of YFP (pUBC-cYFP-MtAHL1 and pUBC-cYFP-MtAHL2) and MtAHL1 tagged with the N-terminal half of YFP (pUBC-nYFP-MtAHL1) was assayed in *N. benthamiana* leaves after co-infiltration. The fluorescence images from the reconstituted YFP fusion and DAPI-stained nuclei were imaged by

653 confocal microscopy and then combined with the bright field images to generate merged images.
654 Complemented YFP signal co-localized with DAPI is marked with red arrows. Scale bars
655 indicate 25 μm .



Extended Data Fig. 6 | Analyses of the efficiency of inactivation and down-regulation of *MtAHLs* in hairy roots of *M. truncatula*. **a** and **b**, CRISPR/Cas9 mediated mutations introduced into *MtAHL1* and *MtAHL2* were investigated using Illumina sequencing, typically with about 5000 reads for each sampled nodules and representative modified sequences in individual nodules are shown. The data were analyzed on the CRISPResso2 website⁴⁰ (<https://crispresso.pinellolab.partners.org/>). Representative sequences of mutated alleles of *MtAHL1* and *MtAHL2* are shown in **a** and **b**, respectively. The sgRNA sequences used are indicated in blue and the PAM site is in red. Nucleotide insertions are marked in bold. **c**,

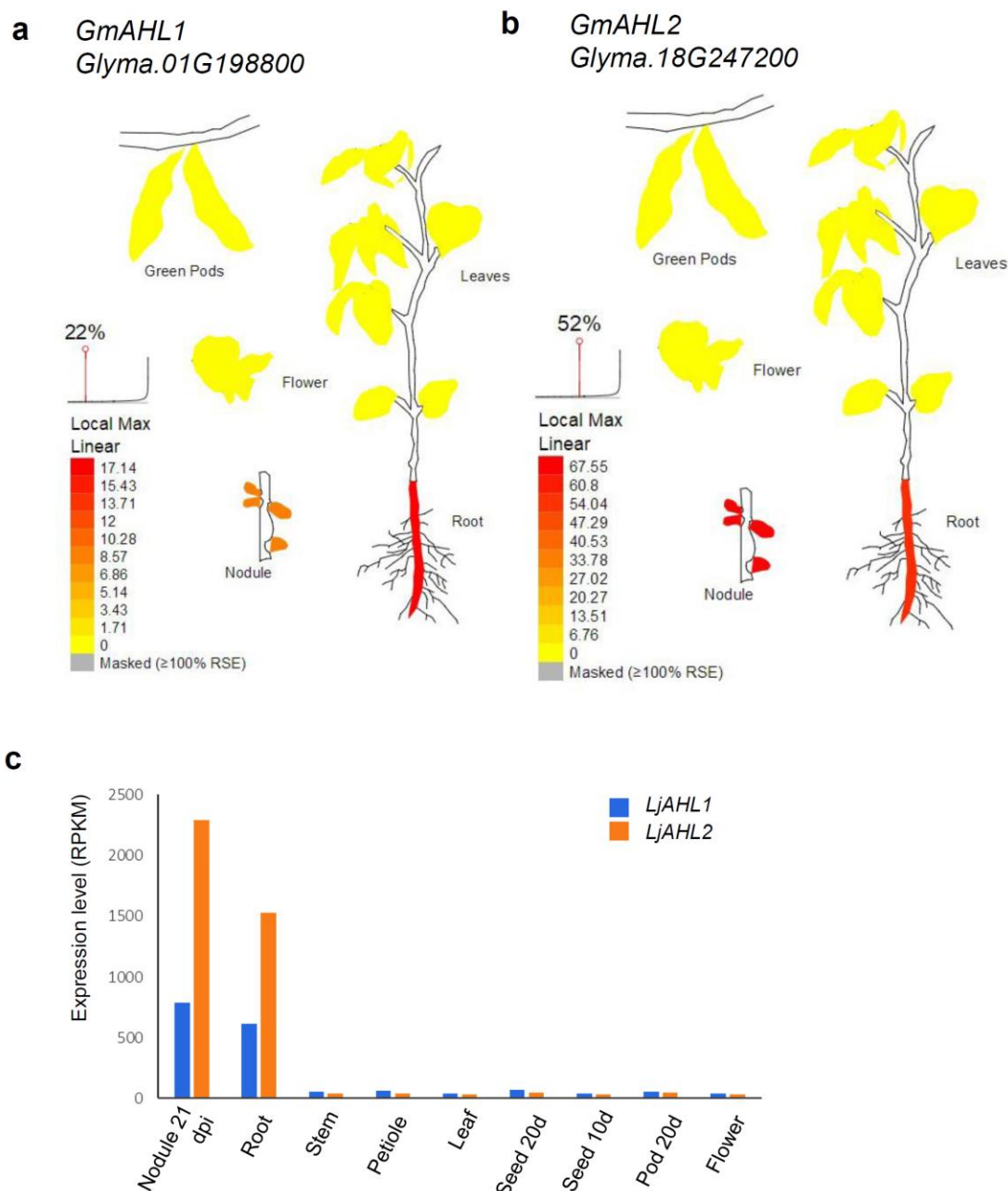
Photomicrographs of nodules formed on transgenic roots in which RNA interference was induced using RNAi constructs *MtAHL1* RNAi or *MtAHL2* RNAi; *NCR169* RNAi used as a phenotype control. The upper images were imaged to capture GFP fluorescence and the lower images were taken with a bright field. **d** and **e**, Expression levels of both *MtAHL1* and *MtAHL2* were measured by quantitative reverse transcription PCR in nodules formed on roots transformed with *MtAHL1* RNAi (**d**) or *MtAHL2* RNAi (**e**). Six roots were chosen for each analysis and experiments were repeated three times. Expression levels were normalized against the *Mt40S* ribosomal gene.



Extended Data Fig. 7 | Phenotyping of *L. japonicus* AHL mutants obtained from *Lotus*

Base⁴¹ (<https://lotus.au.dk/>). **a**, Nodules formed on the roots of *L. japonicus* transformed with the 1181 bp promoter of *NCR169* fused to the reporter gene resulted in nodule-specific GUS expression (left). The image on the right shows a nodule section to reveal that β -glucuronidase activity is restricted to the central nitrogen-fixing tissues. Scale bars indicate 2 mm. The

locations and orientations (arrows) of the *LORE1* insertions (**b**) in *LjAHL1* (*Lj0g3v0341349*) and in *LjAHL2* (*Ljlg3v2896220*) are indicated. The open boxes represent exons and the pairs of primers used for genotyping to identify the homozygous mutants are indicated with dark or gray arrowheads. Shoot growth and nodule morphology of the (**c**) *LjAHL1* and (**d**) *LjAHL2* mutant seedlings grown under N-limitation were indistinguishable from WT.



Extended Data Fig. 8 | Patterns and levels of expression of *AHL1* and *AHL2* genes in soybean and *L. japonicus*. **a**, The expression level of *GmAHL1* (*Glyma.01G198800*) is represented in different tissues by a color gradient, showing strong expression in roots and lower expression in nodules. **b**, The expression level of *GmAHL2* (*Glyma.18G247200*) is strong in both nodules and roots. The images in **a** and **b** were generated using ePlant browser³⁹ (http://bar.utoronto.ca/eplant_soybean/). The colour gradient is generated in “Local Max” mode

to discern expression pattern of the specific gene, rather than to indicate the actual expression level. **c**, The expression levels of *LjAHL1* (probe ID: Ljwgs_027513.1_at) and *LjAHL2* (probe ID: Ljwgs_062607.1_at) in different organs is plotted as reads per kilobase of transcript per million reads mapped (RPKM) using raw data acquired from ExpAt in *Lotus* Base³⁷ (<https://lotus.au.dk/expat/>).

Supplementary Files

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

- [SupplementaryTable1ListofgenescontainingH0Motif4intheir1Kbpromoter.xlsx](#)
- [SupplementaryTable2PrimerList.xlsx](#)