Microscopic and Transcriptomic Analyses of Dalbergoid Legume Peanut Reveal a Divergent Evolution Leading to Nod-Factor-Dependent Epidermal Crack-Entry and Terminal Bacteroid Differentiation

Bikash Raul,[1] Oindrila Bhattacharjee,[1,2] Amit Ghosh,[1] Priya Upadhyay,[1] Kunal Tembhare,[1] Ajeet Singh,[1] Tarannum Shaheen,[1] Asim Kumar Ghosh,[1] Ivone Torres-Jerez,[3] Nick Krom,[3] Josh Clevenger,[4] Michael Udvardi,[3] Brian E. Scheffler,[5] Peggy Ozias-Akins,[4] Ravi Datta Sharma,[2] Kaustav Bandyopadhyay,[2] Vineet Gaur,[1] Shailesh Kumar,[1] and Senjuti Sinharoy[1,†]

[1] National Institute of Plant Genome Research (NIPGR), Aruna Asaf Ali Marg, New Delhi 110067, India
[2] Amity University Haryana, Amity Education Valley, Manesar, Panchgaon, Haryana 122412, India
[3] Noble Research Institute, 2510 Sam Noble Pkwy, Ardmore, OK 73401, U.S.A.
[4] University of Georgia, Institute of Plant Breeding, Genetics and Genomics and Department of Horticulture, Tifton, GA 31793, U.S.A.
[5] United States Department of Agriculture–Agricultural Research Service Jamie Whitten Delta States Research Center (JWDSRC) Stoneville, JWDSRC, Bldg.1, Room 229, Experiment Station Road, PO Box 36, Stoneville, MS 38776-0036, U.S.A.

Root nodule symbiosis (RNS) is the pillar behind sustainable agriculture and plays a pivotal role in the environmental nitrogen cycle. Most of the genetic, molecular, and cell-biological knowledge on RNS comes from model legumes that exhibit a root-hair mode of bacterial infection, in contrast to the Dalbergoid legumes exhibiting crack-entry of rhizobia. As a step toward understanding this important group of legumes, we have combined microscopic analysis and temporal transcriptome to obtain a dynamic view of plant gene expression during *Arachis hypogaea* (peanut) nodule development. We generated comprehensive transcriptome data by mapping the reads to *A. hypogaea,* and two diploid progenitor genomes. Additionally, we performed BLAST searches to identify nodule-induced yet-to-be annotated peanut genes. Comparison between peanut, *Medicago truncatula*, *Lotus japonicus*, and *Glycine max* showed upregulation of 61 peanut orthologs among 111 tested known RNS-related genes, indicating conservation in mechanisms of nodule development among members of the Papilionoid family. Unlike model legumes, recruitment of class 1 phytoglobin-derived symbiotic hemoglobin (SymH) in peanut indicates diversification of oxygen-scavenging mechanisms in the Papilionoid family. Finally, the absence of cysteine-rich motif-1-containing nodule-specific cysteine-rich peptide (*NCR*) genes but the recruitment of *defensin-like NCRs* suggest a diverse molecular mechanism of terminal bacteroid differentiation. In summary, our work describes genetic conservation and diversification in legume–rhizobia symbiosis in the Papilionoid family, as well as among members of the Dalbergoid legumes.

Keywords: *Arachis hypogaea*, bacteria–plant symbiosis, bacteroid, crack entry, Dalbergoid legumes, nitrogen fixation, nod factors, nodule development, peanut, *Rhizobium*–legume symbiosis, root nodule symbiosis

[†]Corresponding author: S. Sinharoy; ssinharoy@nipgr.ac.in

B. Raul and O. Bhattacharjee contributed equally to this work.

The author(s) declare no conflict of interest.

Root nodule development in legumes merits scientific attention due to its environmental and agricultural importance. *Arachis hypogaea* (peanut) is an economically important allotetraploid legume, with an annual production of approximately 46 million metric tons in 2018 (according to the United Nations Food and Agriculture Organization [FAO]). Peanut belongs to the Dalbergoid clade along with *Aeschynomene* spp. Dalbergoid legumes diverged early from other members of the Papilionoid family, between 50 and 55 million years ago (MYA) (Bertioli et al. 2009; Sprent and James 2007). Two model legumes, *Medicago truncatula* (inverted-repeat-lacking clade [IRLC] clade), and *Lotus japonicus* (Robinoid clade), have been the focus of the majority of the genetic studies of nodule development. In almost 75% of legumes, including these model species, nodule development starts after recognition of rhizobial nodulation factors (NFs) by the host plant. In these legumes, rhizobia enter plant roots via epidermal root hairs. The bacteria travel down the length of root hair cells via infection threads (ITs), which are tubular invaginations of the cell wall and plasma membrane. ITs guide rhizobia

to underlying cortical cells. Ramified ITs eventually deliver rhizobia to cortical cells, where the bacteria are released into the host cells surrounded by a plant membrane called the symbiosome membrane, forming a symbiotic "organelle" known as a symbiosome. Ultimately, mature nodules contain a mosaic infection and nitrogen-fixing zone where infected and uninfected cells coexist together (Roy et al. 2020). Nodules in Dalbergoid plants are developmentally and morphologically different from nodules in model legumes. In particular, rhizobia enter via fissures called "cracks" and generate "aeschynomenoid" nodules without ITs in which all cells of the fixation zone contain rhizobia (Boogerd and van Rossum 1997; Sprent and James 2007). *Aeschynomene* spp. can develop functional nodules without NFs (Giraud et al. 2007). Nonetheless, nodule development in peanut usually initiates following an NF-dependent pathway (Ibáñez and Fabra 2011). Sporadic reports of NF-lacking bacteria successfully nodulating peanut exist in the literature (Guha et al. 2016; Teamtisong et al. 2014). Bacteroids in Dalbergoid legumes are terminally differentiated and are either elongated (*A. afraspera*) or spherical (*A. indica* and *Arachis* sp.) (Bonaldi et al. 2011; Boogerd and van Rossum 1997; Raul et al. 2019). Peanut and *Aeschynomene* plants diverged approximately 49 MYA (Chaintreuil et al. 2013) and show substantial differences in growth habitats. *Aeschynomene* spp. grow in a hydrophytic condition, in contrast to semiarid peanut. Soil flooding reduces oxygen availability to roots and generally decreases both nodule initiation and nitrogen-fixation capacity of legumes, including peanut (Pucciariello et al. 2019). *Aeschynomene* spp. are flooding tolerant and, in contrast to other legumes, can fix high amounts of nitrogen under hypoxic conditions (Tobisa et al. 2014). Thus, although the Dalbergoid clade is unique in terms of the evolution of nodulation and symbiotic nitrogen fixation, the individual members of this lineage show significant differences among themselves.

Peanut is an allotetraploid legume (AABB-type genome; $2n = 4x = 40$), derived from a single recent hybridization event between two diploid species (*A. duranensis* and *A. ipaensis*). Diploid peanut lines have significant sequence similarity with *A. hypogaea* 'Tifrunner' (99.96% with *A. ipaensis* and 98.66% *A. duranensis*). The genome sequences of *A. duranensis*, *A. ipaensis*, and tetraploid *A. hypogaea* 'Tifrunner' and the transcriptome sequence of *A. hypogaea* are available at PeanutBase (Bertioli et al. 2016, 2019; Clevenger et al. 2016). These resources enable a variety of research. To date, most of the research on nodule development of peanut has focused on understanding the status of genes known to be involved in symbiosis in model legumes (Sharma et al. 2020). In model species, NF activates the common symbiosis signaling pathway (CSP). The CSP initiates a cascade of events that lead to nodule development (Geurts et al. 2016). Two CSP pathway genes, *calcium calmodulin-dependent protein kinase* and *CYCLOPS*, play vital roles in peanut nodule initiation and symbiosome development (Das et al. 2019; Sinharoy and DasGupta 2009). CRISPR/Cas9-based transgenic hairy roots of the *nod factor receptor* mutant *Ahnfr5B* showed a *Nod−* phenotype, indicating that the NF signaling pathway is active in peanut (Shu et al. 2020). Natural polymorphisms in *AhNSP2* fail to generate nodules in peanut (Peng et al. 2021). Furthermore, *cytokinin receptor histidine-kinase 1* (*AhHK1*) is associated with nodule initiation and symbiosome development (Kundu and DasGupta 2018). These studies highlight the conservation of signaling modules among peanut and model legumes. Well-known nodulation genes involved in IT formation such as *exopolysaccharide receptor* (*EPR3*) and *flotillins* (*MtFLOT2/4*) are highly expressed in *A. hypogaea* nodules, whereas these genes are absent in the *Aeschynomene evenia* genome (Gully et al. 2018; Karmakar et al. 2019; Quilbé et al. 2021), indicating diversification of the early signaling pathway in the Dalbergoid lineage.

Selective breeding has given rise to four major cultivars in peanut. They are 'Runner', 'Virginia', 'Spanish', and 'Valencia'. Two transcriptomic studies of nodule stages have been conducted using Virginia (Peng et al. 2017) and Spanish (Karmakar et al. 2019) botanical types. It is noteworthy that the above studies relied on the genome sequences of the diploid progenitors. The peanut genome presents a complex situation with its approximately 2.7-GB allotetraploid genome. The above studies presented a partial picture of peanut nodule development, due to the complexity of the peanut genome and unavailability of a tetraploid reference genome during that time. In this study, we have undertaken transcriptional profiling at various time points of developing nodules in *Arachis hypogaea* 'Tifrunner', a runner-type peanut, along with detailed microscopic analysis. Our results show that a two-step process leads to the aeschynomenoid nodule development of peanut. In the first step, rhizobia activate cortical cell division followed by endocytosis but, in an unknown way, infection is later spread into every cell. In the second step, infected cells divide further to attain final aeschynomenoid nodule characteristics. To obtain a comprehensive list of differentially expressed genes (DEGs), we have mapped the *A. hypogaea* transcriptome data against three independent peanut genomes (*A. duranensis*, *A. ipaensis*, and *A. hypogaea*). The data are available at the *AhNGE* web portal. We have observed enrichment of diverse groups of cysteine-rich protein-encoding genes including *motif 2 nodule specific cysteine-rich peptides (NCRs),* or defensin-like *NCRs*, and *cysteine-rich secretory proteins, antigen 5,* and *pathogenesis-related 1 proteins (CAPs)*. Like *Aeschynomene* spp., class 1 phytoglobin-derived symbiotic hemoglobins (SymHs) have been recruited in peanut nodule development. Additionally, we have noticed upregulation of *nod factor hydrolase 1* (*NFH1*) in peanut; the ortholog of this gene is absent in *Aeschynomene* legumes. Collectively, our data show a shift in the molecular paradigm adopted by peanut. These are highlighted by (i) peanut employing the ortholog of genes associated with the IT mode of entry in model legumes during the NF-dependent "crack entry" of rhizobia, whereas corresponding orthologs are absent from the genome of *Aeschynomene* legumes; (ii) upregulation of defensin-like NCRs and CAPs along with an absence of motif 1 NCRs during terminal differentiation of bacteroids; and (iii) recruitment of class 1 phytoglobin-derived SymH in oxygen scavenging.

RESULTS

Nodule development in *A. hypogaea* 'Tifrunner'.

We characterized *A. hypogaea* 'Tifrunner' nodule development following inoculation with *Bradyrhizobium* sp. SEMIA 6144, a strain that secretes NF (Ibáñez and Fabra 2011). Root and nodule samples were harvested from 0 to 21 days postinoculation (dpi). On uninoculated roots (12 to 14 days postgermination, designated as 0 dpi), tuft-like clusters of multicellular root hairs had emerged from the lateral root base but rarely from the scar sites (Supplementary Fig. S1A and B). It is noteworthy that, in contrast to the majority of plants, peanut develops multicellular root hairs (MRHs) only from the base of the lateral roots (Boogerd and van Rossum 1997). Likewise, on inoculated plants, nodules emerged mostly from the base of lateral roots (Supplementary Fig. S1C) and rarely from the scar sites containing root hairs (Supplementary Fig. S1D). Nodule primordia were visible by 6 dpi (Supplementary Fig. S1C and D). Curling and typical "shepherd's crook" structures were observed in the root hairs (Fig. 1A, inset). Bacteria appeared to have reached the developing nodule primordium through intercellular cracks (Supplementary Fig. S2A). Microscopic images of 6-dpi nodule primordia showed that cortical cell division started before bacterial endocytosis (Fig. 1A to C). Most of the

plant cells at this stage were uninfected, except a few infected cells densely packed with small undifferentiated bacteroids (U-type) (Fig. 1C; Supplementary Fig. S2B). Dividing cells surrounding an infected cell acquired bacteroids, via an unknown mechanism (Supplementary Fig. S2B).

By 10 dpi, all plant cells in the nodule cortex were infected, a typical characteristic of aeschynomenoid nodules (Fig. 1D and E). At this stage, we have noticed division of the rhizobia containing cells (Supplementary Fig. S2C and D). Most nodules at 10 dpi did not fix nitrogen and are white (only 3 of 10 plants were acetylene reduction assay [ARA] positive) (Supplementary Fig. S3). Infected cells were filled with U-type bacteroids (Fig. 1E and F). Nodules were fully mature by 15 dpi and were ARA positive for all of the tested samples (Supplementary Fig. S3). These nodules were
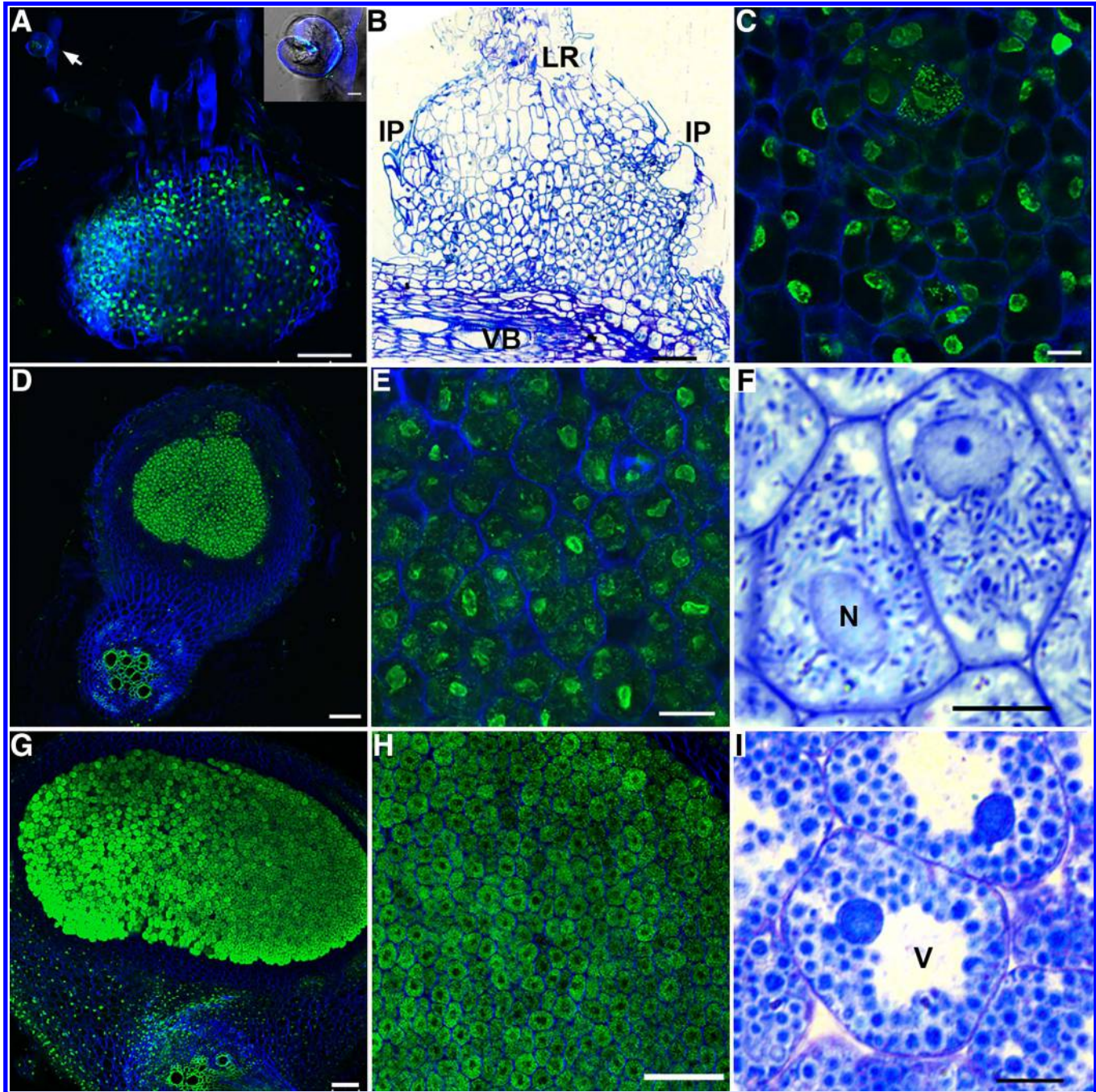

Fig. 1. Histological characterization of the progression of nodule development in peanut. **A to C,** Micrograph of nodule primordia at 6 days postinoculation (dpi). **A,** Longitudinal nodule section with a curled root hair. Inset: the enlarged view of the curled root hair. **B,** Longitudinal semithin section of nodule primordium with infection pockets. **C,** Enlarged view of A, mosaic-infected cell containing infection zone. **D to F,** White nonnitrogen-fixing nodule at 10 dpi. **D,** transverse section of nodule with uniformly infected central zone. **E,** Enlarged view of D, with uniformly infected infection zone. **F,** Semithin section showing undifferentiated (U-type) bacteroid in the infection zone. **G to I,** Pink nitrogen-fixing nodule at 15 dpi. **G,** Transverse section of a nodule showing uniformly infected central infection zone. **H,** Enlarged view of G. **I,** Semithin section showing spherical (S-type) bacteroids in the infection zone. For A, C, D, E, G, and H, 30-μm nodule sections were stained with nucleic-acid-specific dye SYTO13 (green) and cell-wall specific calcofluor white (blue). For B, F, and I, 1-μm sections were stained with toluidine blue-O. Scale bar = 100 μm in A, D, G, and H; 200 μm in B; 10 μm in C and inset of A; 20 μm in E; and 5 μm in F and I. Arrow indicates the curl root hair, IP = infection pocket, VB = vascular bundle, LR = lateral root, N = nucleus, and V = vacuole.

enlarged and retained their aeschynomenoid characteristics (Fig. 1G to I). All of the bacteroids at this stage were uniformly spherical (S-type) (Fig. 1G to I). Therefore, the transition from U-type to S-type took place between 10 and 15 dpi. The 21-dpi nodules were larger than 15-dpi nodules and showed a sharp increase in ARA (Supplementary Figs. S1G and S3).

Transcriptomic analysis of different stages of nodule development.

We selected five time points for RNA sequencing (RNA-seq) analysis to get a dynamic view of nodule development using *A. hypogaea*. These time points were selected based on the above microscopic analysis: 0 dpi for uninfected root control, 6 dpi for early response after rhizobial infection, 6 to 10 dpi for nodule organogenesis, 10 to 15 dpi for bacteroid maturation, and functional nitrogen fixation at 15 and 21 dpi. To obtain a comprehensive list of DEGs, we mapped the *A. hypogaea* transcriptome data against three independent peanut genomes (*A. duranensis*, *A. ipaensis*, and *A. hypogaea*). A summary of the RNA-seq data is given in Supplementary Table S1. On average, 94% of reads remained after quality filtering. Among these, 72, 75, and 90% of reads were unambiguously aligned to *A. duranensis*, *A. ipaensis*, and *A. hypogaea* genomes, respectively (Supplementary Tables S1 and S2). Principal component analysis (PCA) showed good reproducibility between biological replicates (Supplementary Fig. S4A). To obtain a dynamic view of nodule development, we conducted a sequential comparison between two consecutive time points while calculating the DEGs (Supplementary Table S2). Additionally, we also conducted DEG analysis of all time points in comparison with 0 dpi. Only the tetraploid *A. hypogaea* genome was used for this analysis (Supplementary Table S2). Prediction of orthologous proteins in *A. hypogaea*, *A. duranensis,* and *A. ipaensis* was done using OrthoFinder (Emms and Kelly 2015, 2019). Analysis of DEGs was performed using the orthologous relationship (Supplementary Table S3). Not all DEGs had corresponding orthologs annotated in all three peanut genomes (Supplementary Table S2). DEGs from 15,504 orthogroups had orthologs in all three peanut genomes whereas, in individual context, 1,148, 2,281, and 10,969 unique DEGs were identified using *A. duranensis*, *A. ipaensis*, and *A. hypogaea* genomes, respectively. Additionally, in total, 3,095 DEGs from 619 orthogroups were common to both the progenitor genomes but were absent in *A. hypogaea* (Supplementary Fig. S4B; Supplementary Table S3). Thus, use of parental genomes allowed us to identify a total of 6,524 additional DEGs. Because the actual samples were only collected from *A. hypogaea*, unique genes identified from the parental genotypes must also be present in the former but not yet annotated in the current version of the genome release. Furthermore, nodule-induced transcripts without a corresponding annotated gene in *A. hypogaea* genome were annotated using a BLAST search against the *Medicago* proteome (Supplementary Table S4). Finally, we could not find parental orthologs of some genes which are only present in the *A. hypogaea* genome. The apparent absence of these genes from *A. duranensis* or *A. ipaensis* was mainly due to an artifact of genome sequencing and assembly (see Supplementary Fig. S9 for an example of artifact). Transcriptome sequencing of nodules from progenitor varieties will be required to accurately determine the presence or absence of genes in these species.

Transcriptomic studies on nodule development in *A. hypogaea* ecotype JL24 were published previously (Karmakar et al. 2019). We normalized that data using LIMMA (Ritchie et al. 2015) to identify ecotype-independent (i.e., common) nodule-induced genes in *A. hypogaea*. Based on microscopic analysis of JL24 and Tifrunner (described in the Materials and Methods section), stages of nodule development were divided into four groups:

(i) control (i.e., uninfected) root, (ii) infection stage, (iii) developmental phase, and (iv) functional nodule. DEGs were calculated for every category with respect to the control 0-dpi group (Supplementary Fig. S4C). Most DEGs were detected at the functional nodule stage, followed by development, then infection (Supplementary Table S5). Gene ontology (GO) enrichment analysis using AgriGOv2 (Tian et al. 2017) revealed enrichment of genes in "cellular response to a stimulus", and "root hair cell maturation and differentiation" at the infection stage (Supplementary Fig. S4D). Enrichment of the genes for root hair cell maturation and differentiation was consistent with the observed root hair cell curling adjacent to the developing nodule primordia (Fig. 1A, inset). At the nodule development stage, DEGs in both ecotypes were enriched in "cell division", "microtubule movement", and "carbohydrate metabolism" (Supplementary Fig. S4E). Finally, at the functional nodule stage, DEGs were enriched for "transcription", "amino acid and organic acid transport activity", and "oxidation reduction activity" (Supplementary Fig. S4F). DEGs for "tetrapyrrole binding", "heme binding", "iron binding", and "transition metal ion binding" were enriched compared with other DEGs, specifically for the Tifrunner genotype (Supplementary Fig. S4F). Quercetin glycosylation pathway genes were also enriched among the DEGs of Tifrunner.

Status of the peanut orthologs of previously known nodulation genes.

OrthoFinder analysis (Emms and Kelly 2015, 2019) was conducted to identify putative peanut orthologs of genes known to control nodule development. A database was generated integrating the *A. hypogaea* RNA-seq data and the OrthoFinder results (AhNGE). Orthologs of 111 known nodulation genes from *Medicago* and *Lotus* spp., soybean, and three peanut genomes were downloaded from the database. Their expression status is shown in Figure 2A to G and Supplementary Table S6. These genes were assigned to seven different categories based on their role during nodule development. A cartoon representation summarizes the behavior of symbiotic genes along the different nodule developmental stages (Fig. 2H).

In the early signaling category, *NODULATION SIGNALING PATHWAY* (*NSP2*), and *ethylene response factor required for nodulation* (*ERN1/ERN2*) (Andriankaja et al. 2007; Kaló et al. 2005) are yet to be annotated in peanut genomes (Fig. 2A). Using a local BLAST search, we identified transcripts corresponding to *NSP2* and *ERN1/ERN2* (Fig. 3; Supplementary Table S4). Among the surface receptors, a putative ortholog of *exopolysaccharide receptor* (*EPR3*) (Kawaharada et al. 2017) showed significant induction but neither of the nod factor receptors showed induction. As in *A. evenia,* four copies of putative orthologs of *symbiosis receptor kinase* (*SymRK*) are present in *A. hypogaea* (Endre et al. 2002; Stracke et al. 2002). Among these, one pair was nodule specific (*Arahy.SZ7D6R.1* and *Arahy.5KC586.1*, renamed *AhSymRK1*), and the other pair showed low but ubiquitous expression (*Arahy.TBZT1P.1* and *Arahy.5W9Q31.1*, renamed *AhSymRK2*). Other early signaling genes such as *3-hydroxy-3-methylglutaryl coenzyme A reductase 1* (*HMGR1*) (Kevei et al. 2007), *NSP1,* and *PLANT U-BOX protein* (*PUB1*) (Smit et al. 2005; Mbengue et al. 2010) showed induction in peanut nodules (Fig. 2A). Notably, unlike *Aeschynomene evenia*, peanut genomes contain only one copy of *PUB1* (Quilbé et al. 2021). Among the genes required for nodule primordia formation and early infection, the majority of the known transcriptional regulators showed significant upregulation in peanut nodules. In addition, *DELLA1/3, chalcone synthase,* and *flavone synthase-II* (Roy et al. 2020) genes also showed significant upregulation (Fig. 2A).

The rhizobial infection category contains genes required for IT formation. The majority of genes in this category did not
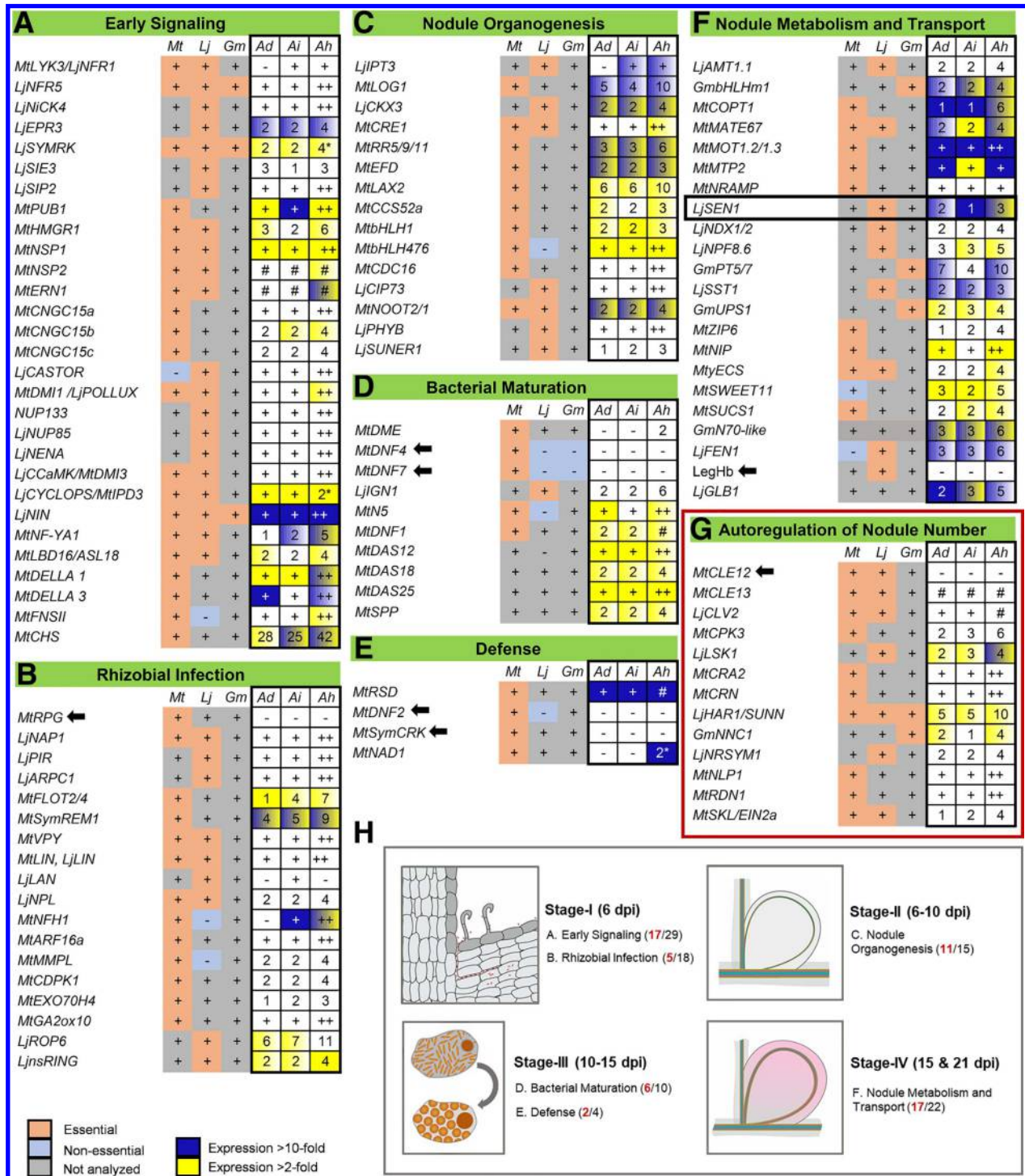
Fig. 2. Expression profile of putative orthologs of symbiotic nitrogen fixation (SNF) genes in peanut. The figure is partially adapted from Roy et al. (2020) and modified based on available literature. In addition to already known genes from *Medicago truncatula* (Mt), *Lotus japonicus* (Lj), and *Glycine max* (Gm), orthologous genes from peanut are shown with their expression status. The SNF-related genes were clustered in seven different categories: **A,** early signaling; **B,** rhizobial infection; **C,** nodule organogenesis; **D,** bacterial maturation; **E,** defense; **F,** nodule metabolism and transport; and **G,** autoregulation of nodule number. Presence or absence of genes representing a particular orthogroup in one peanut species are denoted by +, and −, respectively; ++ represents homologous sets of genes in the tetraploid peanut genome. Numerical values are used in case of more than two genes from the same orthogroup being present in *Arachis hypogaea*. Fold change was calculated by comparison with 0 days postinoculation (dpi). Expression values are color coded for >2-fold upregulation (yellow), >10-fold upregulation (blue), and constitutive expression or downregulation (white). Color gradients signify variable expression among members of the same orthogroup. Symbols: # denotes SNF-related genes not annotated in any of the three *Arachis* genomes and * shows presence and induction when analyzed against diploid progenitor genomes but not using the *A. hypogaea* genome. Quantitative reverse-transcription PCR validation of * and # category genes are given in Supplementary Figure S5. Also, see the Supplementary Table S6 for expression values. The "absent" calling shown in *A. duranensis* or *A. ipaensis* is mainly due to artifacts of genome sequencing and assembly. Confirmed absence call was made only when genes are not annotated in all three peanut genomes and our BLAST search analysis failed to confirm their presence (indicated by arrows). **H,** Cartoon representation of the nodule developmental stages of peanut along with categories of SNF-related genes. Total number of genes tested in each category along with number of upregulated genes are highlighted. The red box indicates genes belonging to the autoregulation of nodule number category which usually do not show induction during nodule development and, hence, were excluded from cartoon representation.

show induction in peanut. As reported previously, we were unable to detect an ortholog of the *RHIZOBIUM-DIRECTED POLAR GROWTH* (*RPG*) in peanut genomes (Griesmann et al. 2018). Apart from *RPG*, all other tested genes have annotated orthologs in at least one peanut genome. Two ancillary genes, *ROP6* and *nsRING*, an E3 ubiquitin ligase (Ke et al. 2012; Shimomura et al. 2006), showed significant induction in peanut nodules. Seven putative orthologs of *flotillin-like* genes are present in the *Arachis hypogaea* genome. Among these, four genes were upregulated (Supplementary Table S5). Notably, four orthologs of *nodule-specific pectate lyase*, which is involved in plant cell wall remodeling during rhizobial infection (Liu et al. 2019; Xie et al. 2012), are present in *A. hypogaea*. However, their expression was negligible. In contrast, an ortholog of *NFH1* (Cai et al. 2018) showed significant upregulation in peanut (Fig. 2B).

In the nodule organogenesis category, peanut genes related to biosynthesis, homeostasis, and signaling of cytokinin (CK) were induced (Fig. 2C). These genes are putative orthologs of *isopentenyl transferase* (*IPT3*), *LONELY GUY* (*LOG1* and *LOG 2*), *cytokinin oxidase/dehydrogenase* (*CKX3*), *response regulator* (*RR5*, *RR9*, and *RR11*), and *ethylene response factor required for nodule differentiation* (*MtEFD*) (Szczyglowski and Ross 2021; Vernié et al. 2008). Among two orthologs of CK receptor genes (*Arahy.IUYB4A.1* and *Arahy.XFB661.1*), *Arahy.IUYB4A* showed twofold induction. A CK-inducible bHLH transcription factor (TF) (*bHLH476*) (Ariel et al. 2012) also showed significant upregulation. Additionally, genes belonging to the orthogroup of putative auxin influx carrier genes (*MtLAX2*) (Roy et al. 2017) were induced. *NODULE ROOT1* (*NOOT1*) (Couzigou et al. 2012; Magne et al. 2018a) and *MtNOOT2* (Magne et al. 2018b) control nodule identity. Four

orthologs of *MtNOOT1* and *MtNOOT2* showed significant upregulation along with the TF *MtbHLH1*, which controls *Medicago* nodule vascular bundle development (Godiard et al. 2011).

In the bacteroid maturation category, the expressed gene *DEMETER* (*MtDME*), a DNA methylase (Satgé et al. 2016), was only detected in the *A. hypogaea* genome (Fig. 2D). *MtDME* regulates nodule-specific gene expression, including *NCRs* (Satgé et al. 2016). The *Defective in nitrogen fixation 1* (*DNF1*) gene NCR delivery to symbiosomes and, thereby, controls terminal bacteroid differentiation in IRLC and Aeschynomenoid plants (Czernic et al. 2015; Wang et al. 2010). The plant signal peptidase complex (SPC) is composed of four subunits. Those are MtDNF1, MtDAS18 (DNF1-associated SPC18), MtDAS12, and MtDAS25. A *signal peptide peptidase* (*SPP*) is also coexpressed with *MtDNF1* complex genes. All five of these genes showed strong upregulation in peanut, suggesting that ER-mediated protein delivery to symbiosomes is conserved in peanut. It is noteworthy that any putative ortholog of *DNF1* is yet to be annotated in the *A. hypogaea* genome (Fig. 2D). Quantitative reverse-transcription PCR (qRT-PCR) analysis established the presence and induction of *AhDNF1* (Supplementary Fig. S5).

A noncanonical *phosphatidylinositol-specific phospholipase C* (*DNF2*), *symbiotic cysteine-rich receptor-like kinase* (*SymCRK*), *nodules with activated defense 1* (*NAD1*), and *regulator of symbiosome differentiation* (*RSD*) are the genes that appear to control nodule defense responses (Benezech et al. 2020; Bourcy et al. 2013; Sinharoy et al. 2013; Wang et al. 2016). *RSD* is not annotated in the *A. hypogaea* genome. We cloned *AhRSD*-cDNA to establish its presence and expression in peanut. The putative ortholog of *MtNAD1* is not annotated in either progenitor genome but is annotated in the *A. hypogaea* genome,
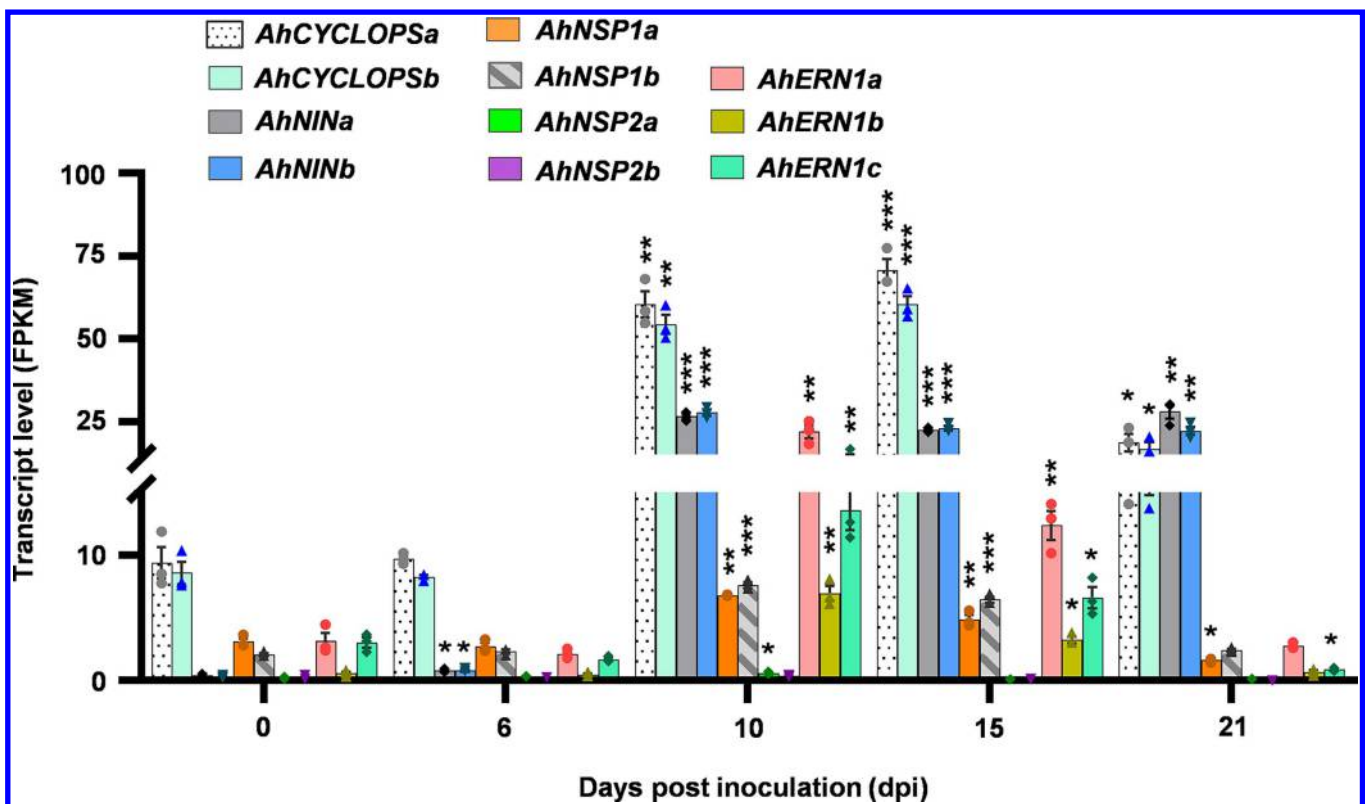

Fig. 3. Expression of selected transcription factors (TFs) during peanut nodule development. Normalized fragments per kilobase of transcript per million mapped reads (FPKM) counts for five TFs. Data represent mean values of three biological replicates. Error bars represent ± standard error of the mean and shapes represent individual expression values. Asterisks indicate the significant difference within a *P* value in *t* tests with respect to 0 days postinoculation; *, **, and *** indicate *P* < 0.05, 0.01, and 0.001, respectively.

where it shows significant induction according to qRT-PCR analysis (Fig. 2E; Supplementary Fig. S5). *DNF2* and *SymCRK* are absent in peanut genomes (Fig. 2E). We could not identify these genes even with our local BLAST search approach.

Most of the transporters required for nitrogen fixation in other legumes have putative orthologs in peanut. Apart from three transporters, all others showed induction in the nodule metabolism and transport category (Fig. 2F). Nodule-induced transporters include putative orthologs of copper transporter *MtCOP1* (Senovilla et al. 2018), iron-activated citrate efflux transporter *MtMATE67* (Kryvoruchko et al. 2018), vacuolar iron transporter *LjSEN1* (Hakoyama et al. 2012), molybdenum transporter *MtMOT1.2/1.3,* Zn efflux transporter *MtMTP2* (León-Mediavilla et al. 2018), nitrate/peptide transporter *LjNPF8.6* (Valkov et al. 2017), phosphate transporters *GmPT5* and *GmPT7* (Qin et al. 2012), sulfate transporter *LjSST1* (Krusell et al. 2005), allantoin and allantoic acid transporter *GmUPS1* (Collier and Tegeder 2012), and sucrose transporter *MtSWEET11* (Kryvoruchko et al. 2016) (Fig. 2F). Apart from the transporters, γ-glutamyl cysteine synthetase *MtγECS* (El Msehli et al. 2011), which controls redox status, and sucrose synthase *MtSucS1* (Hohnjec et al. 2003) were upregulated in peanut. Of three SEN1 proteins belonging to the same orthogroup with LjSEN1 (Fig. 2, black box), two showed nodule-specific expression. We have also identified a fourth *SEN1-like* gene (*Arahy.8KMA2D*) with nodule-specific expression (Supplementary Fig. S6). Gene *2-isopropylmalate synthase* (*LjFEN1*) catalyzes the synthesis of homocitrate, an essential cofactor of the bacterial nitrogenase complex. Among six orthologs of *LjFEN1* found in the *A. hypogaea* genome, two (*Arahy.W53L34* and *Arahy.Y8K7VT*) showed nodule-specific expression (Supplementary Table S5). As in *Aeschynomene evenia*, we found upregulation of putative orthologs of *LjGlb1* (a class 1 phytoglobin) and the absence of *leghemoglobin* (*LegH*) genes. The putative orthologs of *LjGlb1* showed induction in an ecotype-independent manner (Supplementary Table S6).

Legumes control nodule numbers through a complex root-to-shoot-to-root feedback pathway called autoregulation of nodulation (AON). In the AON category, we tested 13 genes (Fig. 2G). *CLAVATA3/ENDOSPERM SURROUNDING REGION* (*CLE*)-related small peptide gene *CLE13* is not annotated in any peanut genome (Mortier et al. 2014). We validated its expression by qRT-PCR (Supplementary Fig. S5). Orthologs of *SUNN/HAR1*, *CLV2, CRN,* and *NLP1* (Kang et al. 2016; Nishida et al. 2018) are present in peanut genomes. It is noteworthy that *CLV2* is yet to be annotated in the *Arachis hypogaea* genome, and its expression was validated through qRT-PCR (Supplementary Fig. S5). Ten copies of putative orthologs of *SUNN/HAR1* are present in the *A. hypogaea* genome (Fig. 2G), among which only one showed induction in nodules. An AP2-TF named *NODULE NUMBER CONTROL1* (*GmNNC1*) (Wang et al. 2020) is a negative regulator of nodule number. This showed significant upregulation in peanut. Additionally, *glycogen synthase kinase LjLSK1* (Garagounis et al. 2019), which controls *NIN* expression under low-nitrogen conditions, was also upregulated (Fig. 2G).

Early transcriptional control during peanut nodule development.

Five TFs—NSP1, NSP2, ERN1, IPD3/Cyclops, and NIN—control IT propagation and nodule organogenesis through their complex regulatory circuit in model legumes (Kang et al. 2016). We thoroughly examined the expression of these TFs in peanut. By 10 dpi, we observed upregulation of putative orthologs of all five TFs except *AhNSP2b*. The only TF that showed significant upregulation by 6 dpi was *AhNINa/b* (Fig. 3). We

observed a macroscopic primordium by 10 dpi, indicating the roles of these TFs in peanut nodule organogenesis. *AhNSP1* and *AhERN1* retained high expression up to 15 dpi, followed by their decline. *AhNIN* and *AhIPD3/Cyclops* maintained relatively high expression up to 21 dpi (Fig. 3). It is noteworthy that we used root zones susceptible to nodulation as the 6-dpi sample for our RNA-seq. Dissected nodules were used from 10 dpi onwards. Therefore, dilution of the RNA from nodule primordium by RNA from roots cannot be ruled out in the 6 dpi samples.

Expression of a diverse group of cysteine-rich protein and peptide genes.

Rhizobia in peanut symbiosomes are terminally differentiated (Oono and Denison 2010). NCRs control terminal differentiation of bacteroids in IRLC and *Aeschynomene* legumes (Czernic et al. 2015; Van de Velde et al. 2010). *Aeschynomene evenia* nodules express two types of *NCR-like* genes: cysteine-rich motif-1 containing NCRs (AeNCR motif 1), with six conserved cysteines, and cysteine-rich motif-2 containing NCRs (defensin-like or AeNCR motif 2), with eight conserved cysteines. AeNCR motif 1 resembles *Medicago* NCRs with six conserved cysteines (Quilbé et al. 2021). To understand the evolution of terminal bacteroid differentiation, we investigated the status of *NCR-like* genes in peanut. Initial analysis identified 55 *NCR-like* genes in *Arachis hypogaea*. Among these, 25 showed specific expression in nodules and the remaining 30 did not express in nodule (Supplementary Table S7). Next, we performed a local BLAST search using AhNCRs and AeNCR motif 1 as query to search *A. hypogaea* transcriptome data, which yielded 30 yet-to-be-annotated AhNCRs (Supplementary Table S7). Among the total 55 nodule-specific AhNCRs, 50 AhNCR*s* have the cysteine arrangement pattern ($CX_nCX_nCX_3CX_nCX_nCX_1CX_3C$) (i.e., AeNCR-motif 2) (Fig. 4A). In all, 5 nodule-induced NCRs and the majority of the remaining NCRs that did not express in nodule had diverse patterns of cysteine numbers and arrangement (Supplementary Fig. S7). In summary, we found only NCR motif 2 or defensin-like NCR genes being expressed in peanut nodules.

Interestingly, the transcriptome revealed a second category of cysteine-rich proteins; namely, CAP. According to our RNA-seq data, 48 CAP genes were expressed in nodules. Among these, 40 were nodule exclusive. AhCAPs are diverse in size (from 100 to 384 amino acids). Their protein sequence homologies vary between 11 and 99%, and isoelectric points range between 4.4 and 9.6 (Supplementary Table S7). Despite large diversity in sequence, a common feature of these proteins is the CAP domain. CAPs contain six conserved cysteines (Fig. 4B; Supplementary Fig. S8). OrthoFinder analysis of the proteomes of the three peanut lines, seven other legumes (including *Aeschynomene evenia, Chamaecrista arietinum, C. fasciculata, Glycine max, L. japonicus, M. truncatula,* and *L. angustifolius*), and *Arabidopsis* revealed that the majority of nodule-induced CAPs do not have any ortholog in the tested plant genomes (Supplementary Table S7), indicating possible peanut-specific evolution. It is noteworthy that the N-terminal signal peptide of the peanut CAPs showed a high degree of conservation with AhNCR signal peptides (Fig. 4C). *AhNCRs* and *AhCAPs* are induced late during nodulation, between 10 and 15 dpi (Fig. 4D).

Recruitment of class I hemoglobin in peanut nitrogen fixation.

According to OrthoFinder analysis, a canonical LegH is absent in all three peanut genomes (Fig. 2F). On the other hand, we found five putative hemoglobin genes in *Arachis hypogaea* which clustered with *LjGlb1*. Among these, three [*Arahy.C5DNVC* (*AhSymH1*), *Arahy.9R3RK8* (*AhSymH2*), and *Arahy.SU9WSI* (*AhSymH3*)] showed nodule-specific expression and, hence, were
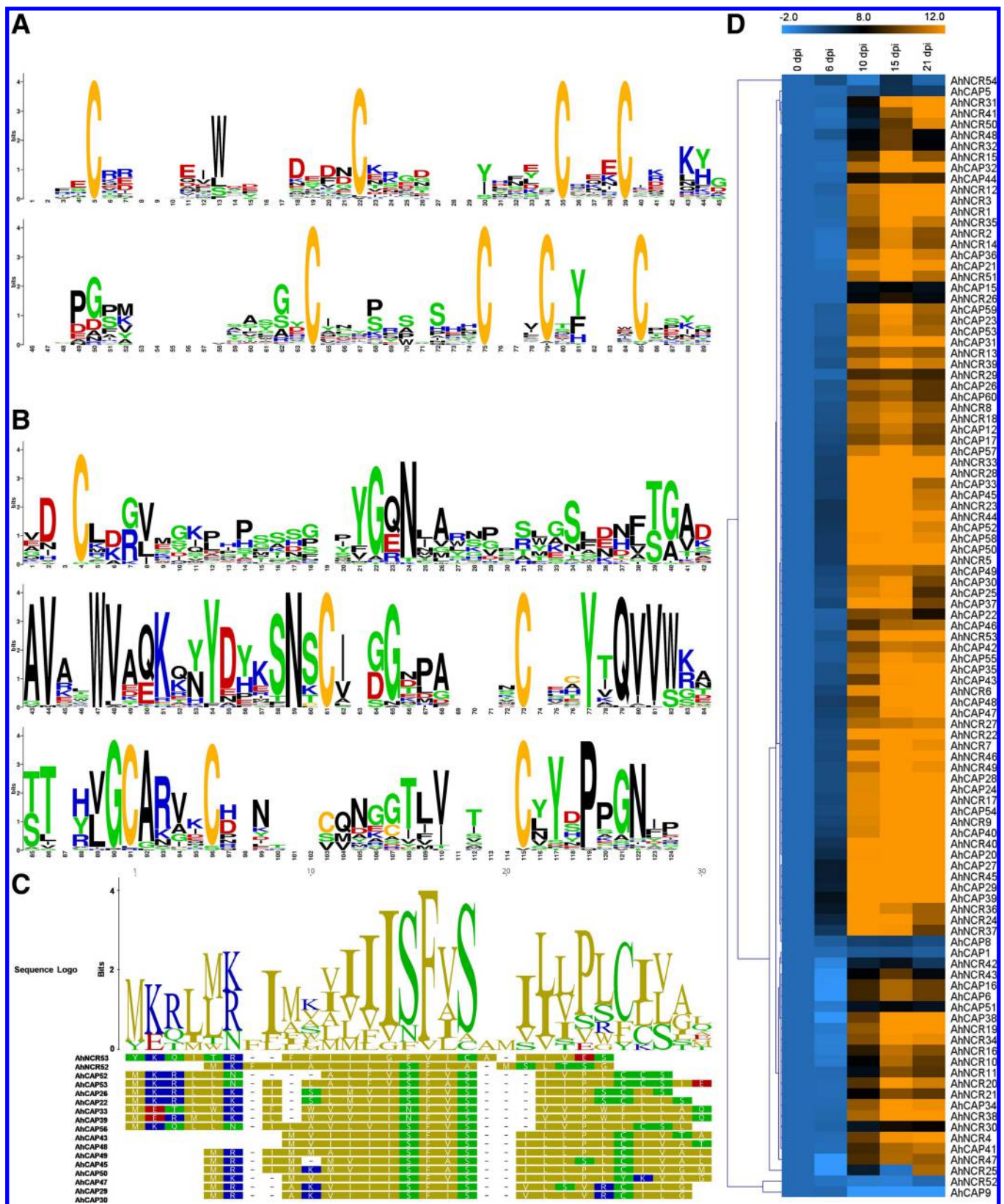
Fig. 4. Recruitment of cysteine-rich proteins or peptides in peanut nodule transcriptome. **A,** Sequence logo showing eight conserved cysteines in peanut defensin-like nodule-specific cysteine-rich peptides (NCRs). **B,** Sequence logo of nodule-induced cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins (CAPs), with six conserved cysteines. **C,** Alignments of signal peptides of NCRs and CAPs. Amino acids are colored based on polarity. **D,** Heat map with hierarchical cluster analysis of genes among the nodule-induced *NCRs* and *CAPs*; dpi = days postinoculation. Multiple sequence alignments for NCRs and CAPs which are used for the generation of the sequence LOGO are given in Supplementary Figs. S7 and S8. Expression data are available at Supplementary Table S7.
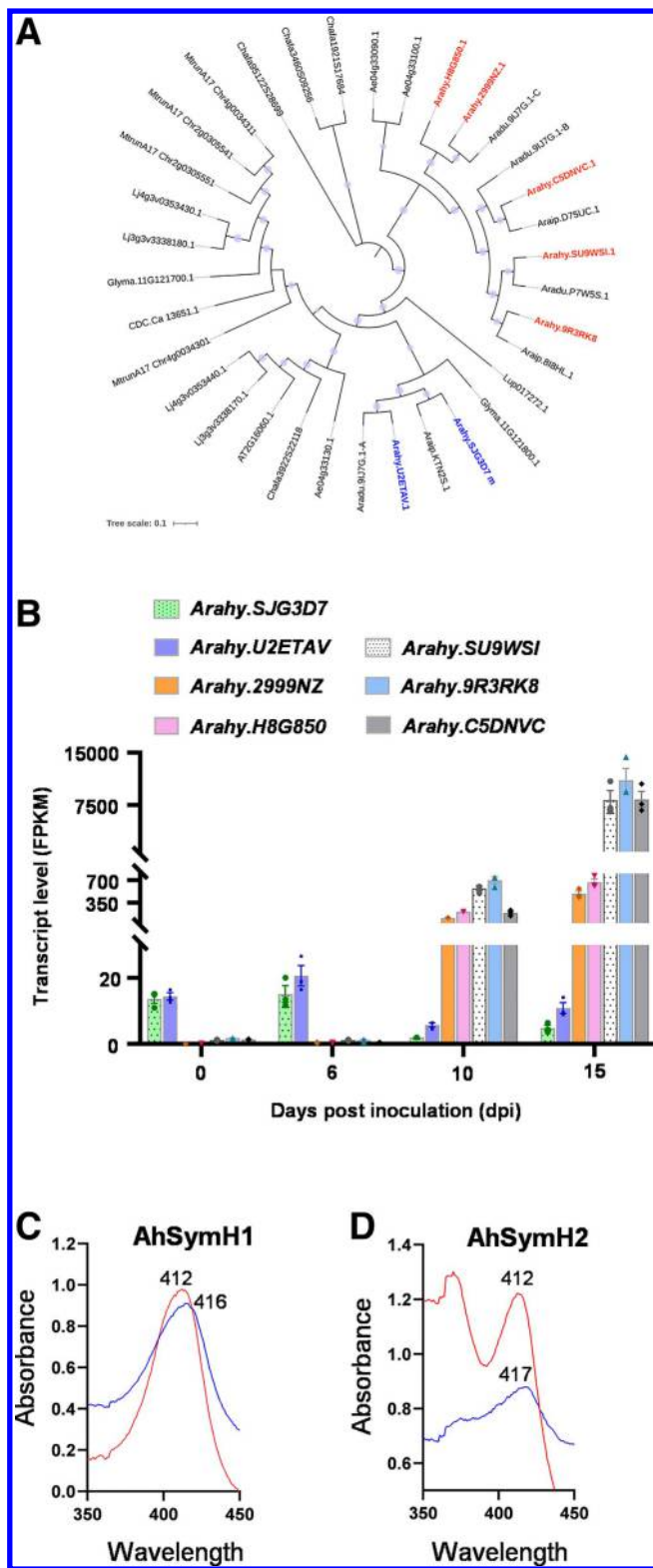
Fig. 5. Recruitment of class I-derived symbiotic hemoglobin in peanut nitrogen fixation. **A,** Phylogenetic analysis of class 1 phytoglobins belong to the same orthogroup. Peanut AhSymHs are highlighted in red and AhGlbs are highlighted in blue. Bootstrap values ($n = 100$) are shown using purple circles. **B,** Expression profiles of *AhGlobins* in root (0 days postinoculation [dpi]) and successive nodule developmental stages (6 to 15 dpi). Normalized fragments per kilobase of transcript per million mapped reads (FPKM) counts are shown. Data represent mean values of three biological replicates. Error bars represent ± standard error of the mean and shapes represent individual expression values. UV-visible spectra of **C,** AhSymH1 and **D,** AhSymH2; the oxidized forms are shown in red and reduced forms are shown in blue.

annotated as SymH. Two other genes, *Arahy.U2ETAV* (*AhGlb1*), and *Arahy.SJG3D7* (*AhGlb2*), also encode class 1 hemoglobins, which showed ubiquitous expression throughout the plant. Hence, they are designated as phytoglobins. Synteny analysis among the three peanut genomes revealed that *AhSymH1, AhSymH5* (*Arahy.2999NZ*), *AhSymH3*, and *AhGlb1* are clustered together on chromosome (Chr.) 3 of *A. hypogaea* and are syntenic to *A. duranensis*. In *A. duranensis*, *Aradu.9IJ7G* is annotated as a single gene but contains three individual genes (Supplementary Fig. S9). *Aradu.9IJ7G-A* is syntenic to *AhGlb1* (incorrectly annotated as plastid division protein), *Aradu.9IJ7G-B* is syntenic to *AhSymH1*, and *Aradu.9IJ7G-C* is syntenic to *AhSymH3* (Supplementary Fig. S9). *AhSymH2, AhSymH4* (*Arahy.H8G850*), two more truncated hemoglobin (*Arahy.SLEA2F* and *Arahy.ZEE1Y7*) having nodule-specific expression, and *AhGlb2* (incorrectly annotated as plastid division protein) were clustered together on Chr. 13, and all of the corresponding genes have orthologs in the *A. ipaensis* genome (Supplementary Fig. S9). Phylogenetic analysis clustered *AhSymHs* together with class 1 phytoglobins. *AhSymHs* clustered with two *AeLegHs* (Fig. 5A). In the phylogenetic tree, *C. fasciculata* hemoglobins are present in this outgroup. Among them, *Chafa1921S17684* was reported to play the role in symbiotic oxygen scavenging (Gopalasubramaniam et al. 2008). According to the phylogenetic analysis, *AhGlb1* and *AhGlb2* were clustered together with soybean hypoxia-induced *GmGlb* (*Glyma.11G121800*), which is expressed specifically in root and nodule (Nakayama et al. 2017). *AhSymH1* to *AhSymH3* showed the highest expression in nodules, followed by *AhSymH4* and *AhSymH5* (Fig. 5B). Alternatively, spliced isoforms of *AhSymH2* and *AhSymH3* showed induction in nodules, although they were not annotated in the peanut genome (Supplementary Table S6). *AhSymH2* and *AhSymH3* are located on Chr. 13 and Chr. 3, respectively. There is a high sequence identity between these proteins, with only a single amino acid difference (Supplementary Fig. S10). AhSymH1 to AhSymH5 contain all eight canonical loops (A to H) required for LegH function (Supplementary Fig. S10). To confirm that AhSymH1 and AhSymH2 (spliced variant) are porphyrin ring-containing proteins, we measured the soret spectra of AhSymH1 and AhSymH2 (Fig. 5C). Soret spectra shows an absorption peak around 400 nm, due to the porphryin ring (Wittenberg 1978). Purified recombinant AhSymH1 and AhSymH2 exhibited the soret spectra characteristic of bona fide hemoglobins (Fig. 5C and D).

DISCUSSION

In this study, we performed detailed microscopic and transcriptomic analyses of nodule development in allotetraploid peanut to facilitate future research in this field. Integration of three genomes for transcriptome analysis, local BLAST searches to identify previously unannotated genes, and genome-wide orthology analysis enabled us to draw important new inferences. One major finding is the apparent recruitment of class 1 phytoglobins in oxygen homeostasis during nitrogen fixation. Additionally, we identified a diverse group of nodule-induced cysteine-rich proteins or peptides in peanut.

Our microscopic data shed light on details of nodule initiation and development in peanut. Unlike other plants, peanut develops MRHs only from the base of the lateral roots (Boogerd and van Rossum 1997). These root hair cells show typical "shepherd's crook" structure in response to rhizobia (Fig. 1A). The nonlegume *Parasponia* contains similar MRHs at the infection site, which clump in response to rhizobium inoculation. Concomitantly with clumping of MRH, cell division begins at the pericycle or cortex in *Parasponia* roots (Bu et al. 2020; Lancelle and Torrey 1984). Curling of MRH is also associated with cortical

or pericyclic cell division in peanut (Fig. 1A). It is hypothesized that nodulation had a single origin approximately 110 MYA in a common ancestor of extant legumes (Griesmann et al. 2018; van Velzen et al. 2018). Most likely, MRHs have an evolutionarily conserved role in the bacterial entry in *Parasponia* and peanut. Related to this, a peanut variety that does not form root hairs also fails to develop nodules (Nambiar et al. 1983). We measured high levels of expression of *AhNFH1* in peanut. Medicago *MtNFH1* preferentially cleaves NFs (Cai et al. 2018). Peanut and *Aeschynomene* legumes belong to the Dalbergoid clade, and the latter exhibits NF-independent nodulation. *NFH1* is absent in the *Aeschynomene evenia* genome (Quilbé et al. 2021). In contrast, the presence and expression of *AhNFH1* in peanut points to its involvement in NF metabolism and signaling during the crack invasion in this species. Indeed, the *Bradyrhizobium* strain used in this study secretes NF (Ibáñez and Fabra 2011). Genes such as *EPR3* and *Flotillins* are present and induced in peanut. Implicated in IT formation, these genes are absent in the *A. evenia* genome, highlighting mechanistic differences in crack invasion among the members of the Dalbergoid clade. Sporadic reports highlighted NF-independent nodule development in peanut (Guha et al. 2016; Teamtisong et al. 2014). In such cases, whether peanut would use a molecular path similar to NF-dependent nodule development remains an open question.

Mature peanut nodules contain a uniformly infected central infection zone. According to our microscopic observation, the development of this aeschynomenoid feature in peanut nodules is a two-step process. We observed cell division at the base of lateral roots (primordium formation) prior to endocytosis of rhizobia (up to 6 dpi) (Fig. 1). Initially, rhizobia are seen only in intercellular spaces within the developing nodule primordium. Eventually, rhizobia are endocytosed by a few mitotically active cells, followed by the spread of infection (between 6 and 10 dpi). Repeated divisions of the infected cells take place only at the later stage (Fig. 1; Supplementary Fig. S2). Our data clearly show that aeschynomenoid nodule development passes through a step of the mosaic-infected nodule cortex. This finding adds a new dimension to the present hypothesis that rhizobial endocytosis is followed by repeated division of the infected cells to develop aeschynomenoid nodules (Ibáñez et al. 2017; Sharma et al. 2020). Intuitively, uptake of rhizobia into the mitotically active cells is rather easier because these cells contain thinner cell walls and could easily be modified by enzymes.

We have investigated 111 well-known symbiotic genes involved in seven different stages of nodule development (Fig. 2). Transcriptome analysis of developing peanut nodules was carried out previously, using the JL24 ecotype (Karmakar et al. 2019). Overall, our results are consistent with previous work with respect to genes involved in early signaling and rhizobial infection, with a few exceptions. Comparing the data from two different ecotypes, we identified more DEGs in JL24 in the infection category (Supplementary Fig. S4C). Karmakar et al. (2019) performed RNA-seq analysis on 4-dpi samples, whereas our earliest time point was 6 dpi. Therefore, our dataset is more suitable for analyzing later stages of nodule development. Nonetheless, we found induction of *AhNINa/b* by 6 dpi, whereas Karmakar et al. (2019) found it to be expressed later. Failure to detect *AhNIN* expression at 4 dpi in JL24 could reflect slower signaling and transcriptional responses in this ecotype, or transcript dilution of responsive cells among the background of normal root cells.

We investigated 22 genes in the "nodule metabolism and transport" category. Most of the transporters in this category were induced during nodule development. Thus, plant proteins that deliver metal ions and other nutrients to bacteroids, or that transport substrate between nodule tissues and cells, likely have conserved roles in the papilionoid lineage. One highlight

of our data is the absence of *LegH* in peanut. Recruitment of class 1 phytoglobin-derived *SymH* during nodulation has been shown before in *Aeschynomene* spp. (Quilbé et al. 2021). Clustering of *Arachis* spp., *A. evenia,* and *Chamaecrista* SymHs in the same orthogroup suggests that recruitment of class 1 hemoglobins in root nodule symbiosis (RNS) occurred after the ancestors of the above species diverged from other mainstream legumes. Subsequently, class 2 hemoglobins evolved in the latter for this purpose. The prevailing hypothesis that nodulation evolved once (Griesmann et al. 2018) is consistent with subsequent recruitment of different families of hemoglobins for oxygen homeostasis and protection of oxygen-labile nitrogenase in different legume lineages. *Frankia* bacteria have an internal mechanism to protect nitrogenase from oxidation. Therefore, a symbiotic oxygen-scavenging mechanism is not needed in most actinorhizal legumes (van Velzen et al. 2019). In this context, recruitment of class 1 phytoglobin in two distinct plant lineages, including the nonlegume *Parasponia* and the legumes *Chamaecrista*, *Aeschynomene*, and peanut, and use of class 2 phytoglobin in modern papilionoid legumes indicates three independent evolutionary pathways leading to oxygen homeostasis in nodules. Furthermore, flooded soil habitats of *Aeschynomene* spp. may have favored the evolution of class 1 hemoglobins under relatively hypoxic conditions. Adapted to semiarid environments, peanut is generally subjected to normoxic conditions. Hence, a lineage-specific evolution of AhSymH probably occurred in peanut. Detailed biochemical or transgenic analysis may clarify whether this class of hemoglobins is better suited to control oxygen in peanut nodules than class 2 hemoglobin.

NCR-mediated terminal differentiation of bacteroids evolved separately in IRLC and *Aeschynomene* legumes (Czernic et al. 2015; Van de Velde et al. 2010). We found only *NCR motif 2* or *defensin-like NCR*s in peanut, whereas both types are present in *Aeschynomene* legumes. Absence of *NCR motif 1* in *Arachis* spp. highlights diversification in the bacteroid terminal differentiation among the Dalbergoid legumes. The presence of a signal peptide in these AhNCRs and conservation of the DNF1 signal peptidase complex suggest that AhNCRs are targeted to symbiosomes and play a role in bacteroid terminal differentiation. Lineage-specific evolution of motif 2 NCRs appears to have occurred in peanut to control bacteroid maturation. The presence of diverse motif 1 *AhNCR*s in peanut but the absence of their transcripts in nodules supports this idea. U- to S-shaped bacteroid differentiation occurs between 10 and 15 dpi. Coexpression of motif 2 *AhNCR*s with 48 *CAPs* having NCR-like signal peptides indicates that CAPs may also play important roles in bacteroid differentiation, especially given the fact the PR1 (CAP) proteins possess broad antimicrobial activity (Breen et al. 2017), as do *NCRs*. The role of *AhCAPs* in bacteroid development is an interesting subject for further investigation.

In summary, we have presented a detailed microscopic study and robust transcriptomic analysis of developing peanut nodules that reveal common and unique aspects of nodule development and symbiosis in peanut compared with other species. The information presented here will serve as a platform for future research and development on RNS of this economically important species and basic research on the evolution of nitrogen-fixing symbioses in plants.

MATERIALS AND METHODS

Plant growth and inoculation.

RNA-seq analysis and ARA were conducted at the Noble Research Institute (NRI), Ardmore, OK, U.S.A. All further experiments were conducted at National Institute of Plant Genome Research (NIPGR), New Delhi, India. At NRI, peanut

seed pods were treated with 30% Clorox for 5 min and thoroughly rinsed with water. The seeds were then coated with antifungal Vitavax for 4 h and plated onto a wet filter paper in dark at 26 to 28°C. After 4 days, the germinated seeds were sown in twice-autoclaved turface/vermiculite (2:1) (thoroughly washed) at 28 to 30°C in a 3-in. pot. Plants were under a regime of 14 h of light and 10 h of darkness with 250 µE m^{-2} s^{-1} light irradiance at 28 and 22°C (day and night, respectively) and 40% relative humidity. Plants were inoculated with a 50-ml culture of *Bradyrhizobium* sp. SEMIA6144 (grown in tryptone-yeast extract [TY] broth at 28°C) at an optical density at 600 nm (OD$_{600}$) of 0.02 to 0.04 at 12 days post potting. SEMIA6144 were grown up to an OD$_{600}$ of 0.7 to 0.9 and diluted in B&D medium without nitrogen before inoculation (Broughton and Dilworth 1971). Experiments were repeated three times with freshly grown bacteria under identical growth condition for RNA-seq analysis. Plants were bottom watered for subsequent time points.

At NIPGR, seeds were treated with 30% commercial bleach (RNA integrity number [RIN]; Hindustan Unilever), and 1% Tween 20 for 5 min. Seeds were then thoroughly washed with water and kept in an antifungal (Vitavax; Dhanuka) solution for 8 to 12 h in dark. Surface-sterilized seeds were kept in Phyta jars (Himedia) with wet filter paper in darkness at 26 to 28°C for 3 to 4 days. Germinated seeds were transferred to nitrogen-free, sterile soil made up of leca, Agropeat, and germination sand in a 1:1:1 ratio. Plants were cultivated in a growth chamber with a cycle of 14 h of light and 10 h of darkness at 28 to 22°C, 40 to 60% relative humidity, and 250 µE m^{-2} s^{-1} light. Inoculation was performed similarly to the way described above, in three biological replicates.

ARA.

ARA was conducted as described previously (Sinharoy et al. 2013). Entire root systems at 10, 15, and 21 dpi were placed into sealed tubes and 0.5% acetylene gas was injected. The data were collected between 16 and 18 h.

RNA isolation.

For the RNA-seq analysis, samples were harvested at different time-points of 0, 6, 10, 15, and 21 dpi. For 0 and 6 dpi, root susceptible zones were harvested (root tips were eliminated) and, from 10 dpi onward, only nodules were harvested. Total RNA was extracted from three independent biological replicates using TRIZOL reagent (Life Technologies), followed by genomic DNA removal by DNase 1 (Ambion) and additional column purification with the RNeasy MinElute CleanUp kit (Qiagen). The RNA quality was checked using the Agilent 2100 Bioanalyzer and samples having a RIN number between 8.9 and 10 were used for RNA-seq. The Kapa stranded RNA-seq kit (Kapa Bio Systems) was used to generate paired-end libraries at the Georgia Genomics Facility (Athens, GA, U.S.A.). For qRT-PCR analysis, samples were harvested from three time points (0, 6, and 10 dpi) for RNA isolation. RNA was isolated in three biological replicates using an RNA purification kit (Nucleospin RNA Plant; Macherey-Nagel). Each biological replicate consisted of samples from four or five plants. RNAs were treated with DNase (TURBO DNA-free kit; Thermo Fisher) and were quality checked. Samples with a RIN value higher than 8 were used for cDNA preparation.

RNA-seq data analysis.

Demultiplexed paired-end reads were checked for quality using FastQC. These were checked for base composition and *k*-mer overrepresentation. After inspection, 10 bases from the 5′ end of each read were trimmed due to nonrandom base composition. All reads maintained high quality (>30) over the total length of the read. After QC, reads were mapped to a ribosomal RNA (rRNA) database using Bowtie v 1.1.0 to remove any

rRNA contamination. The RNA-seq data then were analyzed using the TopHat/Cufflinks pipeline (Trapnell et al. 2012). The assembly process was carried out thrice independently using *Arachis duranensis*, *A. ipaensis*, and *A. hypogaea* genomes (the statistics of the data analysis are given in Supplementary Table S1) and, for the analysis, default parameters were used. For analysis of the ortholog expression, differentially expressed genes ($P < 0.05$) have been taken from the abovementioned three independent analysis and merged with orthogroups using python. Venny (Oliveros 2007) was used to generate Venn diagrams (Fig. 2B). Additionally, the RNA-seq data submitted by Karmakar et al. (2019) was downloaded from NCBI and used for calculation of DEGs in ecotype context. LIMMA (Ritchie et al. 2015) was used to normalize the data. For the LIMMA analysis, only the *A. hypogaea* genome was used. A read count matrix was generated through the feature counts (Liao et al. 2014) method. Once a matrix of read count was generated, data were filtered out for zero counts and data was normalized using the TMM normalization method. Normalized read counts were then converted to logCPM and the mean-variance relationship was modeled using voom (Law et al. 2014), and LIMMA was used for differential expression analysis. Based on the microscopic features of both JL24 and Tifrunner, data were divided into four groups: (i) JL24 UNI (uninoculated) and Tifrunner 0-dpi samples were treated as the control; (ii) JL24 4-dpi and Tifrunner 6-dpi samples were considered to be the infection stage; (iii) JL24 8- and 12-dpi and Tifrunner 10-dpi samples were considered as developmental phase (white nodule); and (iv) JL24 NOD and Tifrunner 15- and 21-dpi samples were considered as functional nodule. The normalized data from both experiments are given in the Supplementary Table S5. We calculated DEG ($P < 0.05$) for control versus infection stage, development phase, and functional nodule separately. DEG was used to generate a bar chart for comparing numbers of genes which were differentially expressing in JL24, Tifrunner, or both. After Cuffdiff suite analysis, a matrix of replicate fragments per kilobase of transcript per million mapped reads (using the *A. hypogaea* genome) values was retrieved by using CummeRbund package and used to generate a PCA plot using python. GO enrichment analysis of DEGs was performed using AgriGOv2 (Tian et al. 2017) with the *A. hypogaea* gene model. Furthermore, GO terms for each time point were taken together, and the enriched GO terms were visualized using REVIGO. The differential coloring was performed using Cytoscape v3.8.2 (Shannon et al. 2003). Heatmaps were created using Multi Experiment Viewer (MeV 4.9.0) (Chu et al. 2008).

qRT-PCR analysis.

Total RNA (1 µg) from each sample was used to synthesize cDNA using the Verso cDNA synthesis kit (Thermo Fisher). qRT-PCR was performed using the CFX96 Touch Real-Time PCR Detection System (Bio-Rad) and PowerUp SYBR Green master mix (Thermo Fisher). All of the reactions were carried out according to the standard cycling mode (primer temperature > 60°C) with cycle. Transcripts were normalized by ubiquitin carrier-like protein (*UBC1*) (Reddy et al. 2013), which showed constant expression across the samples and exhibited amplification efficiency of approximately 1. Primer sequences are provided in Supplementary Table S8. A correlation analysis among the RNA-seq data and qRT-PCR data is given in Supplementary Figure S5.

Local BLAST search and annotation of nodule-induced transcripts.

Peanut nodule-induced NCRs (those are already annotated in the genome) and AeNCR motif 1 were used to search the peanut

transcriptome data using the tBLASTn program to obtain a putative NCR list. The lists thus obtained were used for signal peptide (SP) prediction using SignalP 5.0. The final AhNCR call was based on three parameters: (i) size of the protein needs to be less than 200 amino acids in length, (ii) at least three cysteines need to be present in the sequence, and (iii) presence of SP. For annotation of the nodule-induced but yet-to-be-annotated *A. hypogaea* genes, transcripts were used for BLAST search against *M. truncatula* proteome v5 using tBLASTn with an e-value cutoff of 10^{-3}. The corresponding annotation as shown in the *M. truncatula* was used as an annotation of the corresponding transcript (Supplementary Table S4).

Phylogenetic analysis.

Multiple sequence alignments (MSAs) were created in GUIDANCE2 (Penn et al. 2010) using MAFFT. Based on the GUIDANCE2 column score ≤70, final MSAs were generated. The maximum-likelihood phylogenetic tree was constructed using the PhyMLv3.0 SMS web server (Guindon et al. 2010) using Bayesian Information Criterion. The tree is based on the JTT substitution matrix, model of equilibrium frequencies, fixed proportion of invariable sites, four substitution rate categories, and γ shape parameter of 1.361. The visualization was performed using Interactive Tree of Life (iTOLv5) (Letunic and Bork 2007). The consistency of branches was calculated using a bootstrapping method (100 bootstraps). The following protein sequences were used to create MSA (given in Supplementary Fig. S10): Aradu.9IJ7G.1 (divided into three proteins A, B, and C); Aradu.P7W5S.1 from *A. duranensis* (v1.0); Araip.8I8HL.1, Araip.D75UC.1, Araip.KTN2S.1, and Araip.R1RQ4.1 from *A. ipaensis* (v1.0); Arahy.2999NZ.1, Arahy.9R3RK8.1, Arahy.C5DNVC.1, Arahy.H8G850.1, Arahy.SJG3D7.2, Arahy.SU9WSI.1, and Arahy.U2ETAV.1 from *A. hypogaea* (v1.0); Ae04g33090.1, Ae04g33100.1, and Ae04g33130.1 from *Aeschynomene evenia* (v1.0); AT2G16060.1 from *Arabidopsis thaliana* (v11); Glyma.11G121700.1, Glyma.11G121700.2, and Glyma.11G121800.1 from *G. max* (v1); Lj3g3v3338170.1, Lj3g3v3338180.1, Lj4g3v0353430.1, and Lj4g3v0353440.1 from *L. japonicus* (v3.0); MtrunA17_Chr2g0305541, MtrunA17_Chr2g0305551, MtrunA17_Chr4g0034301, and MtrunA17_Chr4g0034311 from *M. truncatula* (v5.0); Chafa1921S17684, Chafa3460S09256, and Chafa95122S28699 from *C. fasciculata*; and lupan.Tanjil.gnm1.ann1.Lup017272.1 from *L. angustifolius* (v9).

Tissue embedding, histological analysis, and microscopy.

For a macroscopic view, pictures were taken under a Nikon SMZ25 stereomicroscope and images were processed using NIS Elements AR 5.10.00. For semithin sectioning, peanut root nodules were fixed at 4°C overnight with 3% paraformaldehyde (wt/vol) and 2.5% glutaraldehyde (vol/vol) in 0.05 M sodium cacodylate buffer (pH 7.2). The fixed samples were dehydrated in ethanol gradient and subsequently embedded in LR white (Electron Microscopy Science), and semithin sections were generated according to the manufacturer's protocol. Sections (1 μm) were cut with an RJ2035 microtome (Leica Microsystems). For Toluidine Blue staining, the sections were stained for 5 min in 0.05% Toluidine Blue O (Sigma) and analyzed using a DM5500B microscope equipped with a DFC425C camera (Leica Microsystems). For confocal microscopy, sections were cut using a vibratome (Leica Microsystems, VT-1000S) and stained with Calcofluor white (Sigma) and SYTO13 (Life Technologies). Confocal images were captured with a TCS-SP8 (Leica) using excitation (ex) of 488 nm and emission (em) of 500 to 530 nm for SYTO13 and 405 nm ex and 410 to 430 nm em for Calcofluor white. Images were processed with LAS AF Lite version software and prepared for presentation with Photoshop CS6 (Adobe Systems Inc.).

Constructs generation.

The *SymH* genes and *RSD* were cloned from a 15-dpi *Arachis* nodule cDNA sample using primers listed in the Supplementary Table S8. *AhSymH1* and *AhSymH2* were cloned in a modified *pET-28a* vector (Novagen) with N-terminal 6xHis-SUMO tag (Gaur et al. 2019).

Protein purification and soret spectra.

For protein purification, the bacteria were grown in TY broth and purified using the His-Trap HP column (GE Healthcare), with binding buffer containing 20 mM HEPES (pH 7), 5 mM β-mercaptoethanol, 5 mM imidazole, 5% glycerol, 500 mM NaCl, 0.1 μg of lysozyme, 1 mM phenylmethylsulfonyl fluoride, and protease inhibitor cocktail (Sigma); elution buffer contained the same composition as lysis buffer except for 300 mM imidazole. The eluted protein was subjected to overnight SUMO protease digestion in the elution buffer. SUMO protease was purified in the lab. The digested products were further purified using the His-Trap HP column (GE Healthcare), and the purified tag less protein was concentrated using Amicon ultraprotein concentrator (Millipore). The proteins were treated separately with potassium ferrocyanide for oxidation and sodium dithionite for reduction. Proteins were desalted using the NAP-5 Sephadex column (Sigma). Proteins were used for spectral analysis using UV-Vis spectroscopy in lysis buffer without imidazole.

Data availability.

The complete transcriptome data was submitted to NCBI (accession ID: PRJNA596350). Additionally, the data are available from the *Arachis hypogaea* Nodule Developmental Gene Expression (AhNGE). The *AhRSD* sequence was submitted to NCBI (MW848437).

AUTHOR-RECOMMENDED INTERNET RESOURCES

AhNGE: http://14.139.61.8/AhNGE/index.php
FAOSTAT: https://www.fao.org/faostat/en/
PeanutBase: https://peanutbase.org

LITERATURE CITED

Andriankaja, A., Boisson-Dernier, A., Frances, L., Sauviac, L., Jauneau, A., Barker, D. G., and de Carvalho-Niebel, F. 2007. AP2-ERF transcription factors mediate Nod factor dependent Mt ENOD11 activation in root hairs via a novel cis-regulatory motif. Plant Cell 19: 2866-2885.

Ariel, F., Brault-Hernandez, M., Laffont, C., Huault, E., Brault, M., Plet, J., Moison, M., Blanchet, S., Ichanté, J. L., Chabaud, M., Carrere, S., Crespi, M., Chan, R. L., and Frugier, F. 2012. Two direct targets of cytokinin signaling regulate symbiotic nodulation in *Medicago truncatula*. Plant Cell 24:3838-3852.

Benezech, C., Berrabah, F., Jardinaud, M. F., Le Scornet, A., Milhes, M., Jiang, G., George, J., Ratet, P., Vailleau, F., and Gourion, B. 2020. *Medicago-Sinorhizobium-Ralstonia* co-infection reveals legume nodules as pathogen confined infection sites developing weak defenses. Curr. Biol. 30:351-358.e4.

Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G., Farmer, A. D., Cannon, E. K. S., Liu, X., Gao, D., Clevenger, J., Dash, S., Ren, L., Moretzsohn, M. C., Shirasawa, K., Huang, W., Vidigal, B., Abernathy, B., Chu, Y., Niederhuth, C. E., Umale, P., Araújo, A. C., Kozik, A., Kim, K. D., Burow, M. D., Varshney, R. K., Wang, X., Zhang, X., Barkley, N., Guimarães, P. M., Isobe, S., Guo, B., Liao, B., Stalker, H. T., Schmitz, R. J., Scheffler, B. E., Leal-Bertioli,

S. C. M., Xun, X., Jackson, S. A., Michelmore, R., and Ozias-Akins, P. 2016. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat. Genet. 48: 438-446.

Bertioli, D. J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., Leal-Bertioli, S. C. M., Ren, L., Farmer, A. D., Pandey, M. K., Samoluk, S. S., Abernathy, B., Agarwal, G., Ballén-Taborda, C., Cameron, C., Campbell, J., Chavarro, C., Chitikineni, A., Chu, Y., Dash, S., El Baidouri, M., Guo, B., Huang, W., Kim, K. D., Korani, W., Lanciano, S., Lui, C. G., Mirouze, M., Moretzsohn, M. C., Pham, M., Shin, J. H., Shirasawa, K., Sinharoy, S., Sreedasyam, A., Weeks, N. T., Zhang, X., Zheng, Z., Sun, Z., Froenicke, L., Aiden, E. L., Michelmore, R., Varshney, R. K., Holbrook, C. C., Cannon, E. K. S., Scheffler, B. E., Grimwood, J., Ozias-Akins, P., Cannon, S. B., Jackson, S. A., and Schmutz, J. 2019. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. Nat. Genet. 51: 877-884.

Bertioli, D. J., Moretzsohn, M. C., Madsen, L. H., Sandal, N., Leal-Bertioli, S. C., Guimarães, P. M., Hougaard, B. K., Fredslund, J., Schauser, L., Nielsen, A. M., Sato, S., Tabata, S., Cannon, S. B., and Stougaard, J. 2009. An analysis of synteny of *Arachis* with *Lotus* and *Medicago* sheds new light on the structure, stability and evolution of legume genomes. BMC Genomics 10:45.

Bonaldi, K., Gargani, D., Prin, Y., Fardoux, J., Gully, D., Nouwen, N., Goormachtig, S., and Giraud, E. 2011. Nodulation of *Aeschynomene afraspera* and *A. indica* by photosynthetic *Bradyrhizobium* sp. strain ORS285: The nod-dependent versus the nod-independent symbiotic interaction. Mol. Plant-Microbe Interact. 24:1359-1371.

Boogerd, F. C., and van Rossum, D. 1997. Nodulation of groundnut by *Bradyrhizobium*: A simple infection process by crack entry. FEMS Microbiol. Rev. 21:5-27.

Bourcy, M., Brocard, L., Pislariu, C. I., Cosson, V., Mergaert, P., Tadege, M., Mysore, K. S., Udvardi, M. K., Gourion, B., and Ratet, P. 2013. *Medicago truncatula* DNF2 is a PI-PLC-XD-containing protein required for bacteroid persistence and prevention of nodule early senescence and defense-like reactions. New Phytol. 197:1250-1261.

Breen, S., Williams, S. J., Outram, M., Kobe, B., and Solomon, P. S. 2017. Emerging insights into the functions of pathogenesis-related protein 1. Trends Plant Sci. 22:871-879.

Broughton, W. J., and Dilworth, M. J. 1971. Control of leghaemoglobin synthesis in snake beans. Biochem. J. 125:1075-1080.

Bu, F., Rutten, L., Roswanjaya, Y. P., Kulikova, O., Rodriguez-Franco, M., Ott, T., Bisseling, T., van Zeijl, A., and Geurts, R. 2020. Mutant analysis in the nonlegume *Parasponia andersonii* identifies NIN and NF-YA1 transcription factors as a core genetic network in nitrogen-fixing nodule symbioses. New Phytol. 226:541-554.

Cai, J., Zhang, L. Y., Liu, W., Tian, Y., Xiong, J. S., Wang, Y. H., Li, R. J., Li, H. M., Wen, J., Mysore, K. S., Boller, T., Xie, Z. P., and Staehelin, C. 2018. Role of the nod factor hydrolase MtNFH1 in regulating nod factor levels during rhizobial infection and in mature nodules of *Medicago truncatula*. Plant Cell 30:397-414.

Chaintreuil, C., Arrighi, J. F., Giraud, E., Miché, L., Moulin, L., Dreyfus, B., Munive-Hernández, J. A., Villegas-Hernandez, M. C., and Béna, G. 2013. Evolution of symbiosis in the legume genus *Aeschynomene*. New Phytol. 200:1247-1259.

Chu, V. T., Gottardo, R., Raftery, A. E., Bumgarner, R. E., and Yeung, K. Y. 2008. MeV+R: Using MeV as a graphical user interface for Bioconductor applications in microarray analysis. Genome Biol. 9:R118.

Clevenger, J., Chu, Y., Scheffler, B., and Ozias-Akins, P. 2016. A developmental transcriptome map for allotetraploid *Arachis hypogaea*. Front. Plant Sci. 7:1446.

Collier, R., and Tegeder, M. 2012. Soybean ureide transporters play a critical role in nodule development, function and nitrogen export. Plant J. 72:355-367.

Couzigou, J. M., Zhukov, V., Mondy, S., Abu el Heba, G., Cosson, V., Ellis, T. H., Ambrose, M., Wen, J., Tadege, M., Tikhonovich, I., Mysore, K. S., Putterill, J., Hofer, J., Borisov, A. Y., and Ratet, P. 2012. *NODULE ROOT* and *COCHLEATA* maintain nodule development and are legume orthologs of *Arabidopsis BLADE-ON-PETIOLE* genes. Plant Cell 24:4498-4510.

Czernic, P., Gully, D., Cartieaux, F., Moulin, L., Guefrachi, I., Patrel, D., Pierre, O., Fardoux, J., Chaintreuil, C., Nguyen, P., Gressent, F., Da Silva, C., Poulain, J., Wincker, P., Rofidal, V., Hem, S., Barrière, Q., Arrighi, J. F., Mergaert, P., and Giraud, E. 2015. Convergent evolution of endosymbiont differentiation in Dalbergioid and inverted repeat-lacking clade legumes mediated by nodule-specific cysteine-rich peptides. Plant Physiol. 169:1254-1265.

Das, D. R., Horváth, B., Kundu, A., Kaló, P., and DasGupta, M. 2019. Functional conservation of CYCLOPS in crack entry legume *Arachis hypogaea*. Plant Sci. 281:232-241.

El Msehli, S., Lambert, A., Baldacci-Cresp, F., Hopkins, J., Boncompagni, E., Smiti, S. A., Hérouart, D., and Frendo, P. 2011. Crucial role of (homo)glutathione in nitrogen fixation in *Medicago truncatula* nodules. New Phytol. 192:496-506.

Emms, D. M., and Kelly, S. 2015. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16:157.

Emms, D. M., and Kelly, S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. Genome Biol. 20:238.

Endre, G., Kereszt, A., Kevei, Z., Mihacea, S., Kaló, P., and Kiss, G. B. 2002. A receptor kinase gene regulating symbiotic nodule development. Nature 417:962-966.

Garagounis, C., Tsikou, D., Plitsi, P. K., Psarrakou, I. S., Avramidou, M., Stedel, C., Anagnostou, M., Georgopoulou, M. E., and Papadopoulou, K. K. 2019. Lotus SHAGGY-like kinase 1 is required to suppress nodulation in *Lotus japonicus*. Plant J. 98:228-242.

Gaur, V., Ziajko, W., Nirwal, S., Szlachcic, A., Gapińska, M., and Nowotny, M. 2019. Recognition and processing of branched DNA substrates by Slx1-Slx4 nuclease. Nucleic Acids Res. 47:11681-11690.

Geurts, R., Xiao, T. T., and Reinhold-Hurek, B. 2016. What does it take to evolve a nitrogen-fixing endosymbiosis? Trends Plant Sci. 21:199-208.

Giraud, E., Moulin, L., Vallenet, D., Barbe, V., Cytryn, E., Avarre, J. C., Jaubert, M., Simon, D., Cartieaux, F., Prin, Y., Bena, G., Hannibal, L., Fardoux, J., Kojadinovic, M., Vuillet, L., Lajus, A., Cruveiller, S., Rouy, Z., Mangenot, S., Segurens, B., Dossat, C., Franck, W. L., Chang, W. S., Saunders, E., Bruce, D., Richardson, P., Normand, P., Dreyfus, B., Pignol, D., Stacey, G., Emerich, D., Verméglio, A., Médigue, C., and Sadowsky, M. 2007. Legumes symbioses: Absence of *Nod* genes in photosynthetic bradyrhizobia. Science 316:1307-1312.

Godiard, L., Lepage, A., Moreau, S., Laporte, D., Verdenaud, M., Timmers, T., and Gamas, P. 2011. MtbHLH1, a bHLH transcription factor involved in *Medicago truncatula* nodule vascular patterning and nodule to plant metabolic exchanges. New Phytol. 191:391-404.

Gopalasubramaniam, S. K., Kovacs, F., Violante-Mota, F., Twigg, P., Arredondo-Peter, R., and Sarath, G. 2008. Cloning and characterization of a caesalpinoid (*Chamaecrista fasciculata*) hemoglobin: The structural transition from a nonsymbiotic hemoglobin to a leghemoglobin. Proteins 72:252-260.

Griesmann, M., Chang, Y., Liu, X., Song, Y., Haberer, G., Crook, M. B., Billault-Penneteau, B., Lauressergues, D., Keller, J., Imanishi, L., Roswanjaya, Y. P., Kohlen, W., Pujic, P., Battenberg, K., Alloisio, N., Liang, Y., Hilhorst, H., Salgado, M. G., Hocher, V., Gherbi, H., Svistoonoff, S., Doyle, J. J., He, S., Xu, Y., Xu, S., Qu, J., Gao, Q., Fang, X., Fu, Y., Normand, P., Berry, A. M., Wall, L. G., Ané, J. M., Pawlowski, K., Xu, X., Yang, H., Spannagl, M., Mayer, K. F. X., Wong, G. K. S., Parniske, M., Delaux, P. M., and Cheng, S. 2018. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. Science 361:eaat1743.

Guha, S., Sarkar, M., Ganguly, P., Uddin, M. R., Mandal, S., and DasGupta, M. 2016. Segregation of *nod*-containing and *nod*-deficient bradyrhizobia as endosymbionts of *Arachis hypogaea* and as endophytes of *Oryza sativa* in intercropped fields of Bengal Basin, India. Environ. Microbiol. 18:2575-2590.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst. Biol. 59:307-321.

Gully, D., Czernic, P., Cruveiller, S., Mahé, F., Longin, C., Vallenet, D., François, P., Nidelet, S., Rialle, S., Giraud, E., Arrighi, J. F., DasGupta, M., and Cartieaux, F. 2018. Transcriptome profiles of Nod factor-independent symbiosis in the tropical legume *Aeschynomene evenia*. Sci. Rep. 8:10934.

Hakoyama, T., Niimi, K., Yamamoto, T., Isobe, S., Sato, S., Nakamura, Y., Tabata, S., Kumagai, H., Umehara, Y., Brossuleit, K., Petersen, T. R., Sandal, N., Stougaard, J., Udvardi, M. K., Tamaoki, M., Kawaguchi, M., Kouchi, H., and Suganuma, N. 2012. The integral membrane protein SEN1 is required for symbiotic nitrogen fixation in *Lotus japonicus* nodules. Plant Cell Physiol. 53:225-236.

Hohnjec, N., Perlick, A. M., Pühler, A., and Küster, H. 2003. The *Medicago truncatula* sucrose synthase gene MtSucS1 is activated both in the infected region of root nodules and in the cortex of roots colonized by arbuscular mycorrhizal fungi. Mol. Plant-Microbe Interact. 16:903-915.

Ibáñez, F., and Fabra, A. 2011. Rhizobial Nod factors are required for cortical cell division in the nodule morphogenetic programme of the Aeschynomeneae legume *Arachis*. Plant Biol. 13:794-800.

Ibáñez, F., Wall, L., and Fabra, A. 2017. Starting points in plant-bacteria nitrogen-fixing symbioses: Intercellular invasion of the roots. J. Exp. Bot. 68:1905-1918.

Kaló, P., Gleason, C., Edwards, A., Marsh, J., Mitra, R. M., Hirsch, S., Jakab, J., Sims, S., Long, S. R., Rogers, J., Kiss, G. B., Downie, J. A., and Oldroyd, G. E. 2005. Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. Science 308:1786-1789.

Kang, Y., Li, M., Sinharoy, S., and Verdier, J. 2016. A snapshot of functional genetic studies in *Medicago truncatula*. Front. Plant Sci. 7:1175.

Karmakar, K., Kundu, A., Rizvi, A. Z., Dubois, E., Severac, D., Czernic, P., Cartieaux, F., and DasGupta, M. 2019. Transcriptomic analysis with the progress of symbiosis in 'crack-entry' legume *Arachis hypogaea* highlights its contrast with 'infection thread' adapted legumes. Mol. Plant-Microbe Interact. 32:271-285.

Kawaharada, Y., Nielsen, M. W., Kelly, S., James, E. K., Andersen, K. R., Rasmussen, S. R., Füchtbauer, W., Madsen, L. H., Heckmann, A. B., Radutoiu, S., and Stougaard, J. 2017. Differential regulation of the Epr3 receptor coordinates membrane-restricted rhizobial colonization of root nodule primordia. Nat. Commun. 8:14534.

Ke, D., Fang, Q., Chen, C., Zhu, H., Chen, T., Chang, X., Yuan, S., Kang, H., Ma, L., Hong, Z., and Zhang, Z. 2012. The small GTPase ROP6 interacts with NFR5 and is involved in nodule formation in *Lotus japonicus*. Plant Physiol. 159:131-143.

Kevei, Z., Lougnon, G., Mergaert, P., Horváth, G. V., Kereszt, A., Jayaraman, D., Zaman, N., Marcel, F., Regulski, K., Kiss, G. B., Kondorosi, A., Endre, G., Kondorosi, E., and Ané, J. M. 2007. 3-hydroxy-3-methylglutaryl coenzyme a reductase 1 interacts with NORK and is crucial for nodulation in *Medicago truncatula*. Plant Cell 19:3974-3989.

Krusell, L., Krause, K., Ott, T., Desbrosses, G., Krämer, U., Sato, S., Nakamura, Y., Tabata, S., James, E. K., Sandal, N., Stougaard, J., Kawaguchi, M., Miyamoto, A., Suganuma, N., and Udvardi, M. K. 2005. The sulfate transporter SST1 is crucial for symbiotic nitrogen fixation in *Lotus japonicus* root nodules. Plant Cell 17:1625-1636.

Kryvoruchko, I. S. I. S., Routray, P., Sinharoy, S., Torres-Jerez, I., Tejada-Jiménez, M., Finney, L. A. L. A., Nakashima, J., Pislariu, C. I. C. I., Benedito, V. A. V. A., González-Guerrero, M., Roberts, D. M. D. M., and Udvardi, M. K. 2018. An iron-activated citrate transporter, MtMATE67, is required for symbiotic nitrogen fixation. Plant Physiol. 176:2315-2329.

Kryvoruchko, I. S. I. S., Sinharoy, S., Torres-Jerez, I., Sosso, D., Pislariu, C. I. C. I., Guan, D., Murray, J., Benedito, V. A., Frommer, W. B. W. B., and Udvardi, M. K. M. K. 2016. MtSWEET11, a nodule-specific sucrose transporter of *Medicago truncatula*. Plant Physiol. 171:554-565.

Kundu, A., and DasGupta, M. 2018. Silencing of putative cytokinin receptor histidine kinase 1 inhibits both inception and differentiation of root nodules in *Arachis hypogaea*. Mol. Plant-Microbe Interact. 31:187-199.

Lancelle, S. A., and Torrey, J. G. 1984. Early development of *Rhizobium*-induced root nodules of *Parasponia rigida*. I. Infection and early nodule initiation. Protoplasma 123:26-37.

Law, C. W., Chen, Y., Shi, W., and Smyth, G. K. 2014. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15:R29.

León-Mediavilla, J., Senovilla, M., Montiel, J., Gil-Díez, P., Saez, Á., Kryvoruchko, I. S., Reguera, M., Udvardi, M. K., Imperial, J., and González-Guerrero, M. 2018. MtMTP2-Facilitated zinc transport into intracellular compartments is essential for nodule development in *Medicago truncatula*. Front. Plant Sci. 9:990.

Letunic, I., and Bork, P. 2007. Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. Bioinformatics 23:127-128.

Liao, Y., Smyth, G. K., and Shi, W. 2014. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923-930.

Liu, C. W., Breakspear, A., Guan, D., Cerri, M. R., Jackson, K., Jiang, S., Robson, F., Radhakrishnan, G. V., Roy, S., Bone, C., Stacey, N., Rogers, C., Trick, M., Niebel, A., Oldroyd, G. E. D., de Carvalho-Niebel, F., and Murray, J. D. 2019. NIN acts as a network hub controlling a growth module required for rhizobial infection. Plant Physiol. 179:1704-1722.

Magne, K., Couzigou, J. M., Schiessl, K., Liu, S., George, J., Zhukov, V., Sahl, L., Boyer, F., Iantcheva, A., Mysore, K. S., Wen, J., Citerne, S., Oldroyd, G. E. D., and Ratet, P. 2018b. MtNODULE ROOT1 and MtNODULE ROOT2 are essential for indeterminate nodule identity. Plant Physiol. 178:295-316.

Magne, K., George, J., Berbel Tornero, A., Broquet, B., Madueño, F., Andersen, S. U., and Ratet, P. 2018a. *Lotus japonicus NOOT-BOP-COCH-LIKE1* is essential for nodule, nectary, leaf and flower development. Plant J. 94:880-894.

Mbengue, M., Camut, S., de Carvalho-Niebel, F., Deslandes, L., Froidure, S., Klaus-Heisen, D., Moreau, S., Rivas, S., Timmers, T., Hervé, C., Cullimore, J., and Lefebvre, B. 2010. The *Medicago truncatula* E3 ubiquitin ligase PUB1 interacts with the LYK3 symbiotic receptor and negatively regulates infection and nodulation. Plant Cell 22:3474-3488.

Mortier, V., Wasson, A., Jaworek, P., De Keyser, A., Decroos, M., Holsters, M., Tarkowski, P., Mathesius, U., and Goormachtig, S. 2014. Role of *LONELY GUY* genes in indeterminate nodulation on *Medicago truncatula*. New Phytol. 202:582-593.

Nakayama, T. J., Rodrigues, F. A., Neumaier, N., Marcolino-Gomes, J., Molinari, H. B. C., Santiago, T. R., Formighieri, E. F., Basso, M. F., Farias, J. R. B., Emygdio, B. M., de Oliveira, A. C. B., Campos, Â. D., Borém, A., Harmon, F. G., Mertz-Henning, L. M., and Nepomuceno, A. L. 2017. Insights into soybean transcriptome reconfiguration under hypoxic stress: Functional, regulatory, structural, and compositional characterization. PLoS One 12:e0187920.

Nambiar, P. T. C., Nigam, S. N., Dart, P. J., and Gibbons, R. W. 1983. Absence of root hairs in non-nodulating groundnut, *Arachis hypogaea* L. J. Exp. Bot. 34:484-488.

Nishida, H., Tanaka, S., Handa, Y., Ito, M., Sakamoto, Y., Matsunaga, S., Betsuyaku, S., Miura, K., Soyano, T., Kawaguchi, M., and Suzaki, T. 2018. A NIN-LIKE PROTEIN mediates nitrate-induced control of root nodule symbiosis in *Lotus japonicus*. Nat. Commun. 9:499.

Oliveros, J. C. 2007. VENNY. An interactive tool for comparing lists with Venn diagrams. https://bioinfogp.cnb.csic.es/tools/venny/index.html

Oono, R., and Denison, R. F. 2010. Comparing symbiotic efficiency between swollen versus nonswollen rhizobial bacteroids. Plant Physiol. 154:1541-1548.

Peng, Z., Chen, H., Tan, L., Shu, H., Varshney, R. K., Zhou, Z., Zhao, Z., Luo, Z., Chitikineni, A., Wang, L., Maku, J., López, Y., Gallo, M., Zhou, H., and Wang, J. 2021. Natural polymorphisms in a pair of NSP2 homoeologs can cause loss of nodulation in peanut. J. Exp. Bot. 72:1104-1118.

Peng, Z., Liu, F., Wang, L., Zhou, H., Paudel, D., Tan, L., Maku, J., Gallo, M., and Wang, J. 2017. Transcriptome profiles reveal gene regulation of peanut (*Arachis hypogaea* L.) nodulation. Sci. Rep. 7:40066.

Penn, O., Privman, E., Ashkenazy, H., Landan, G., Graur, D., and Pupko, T. 2010. GUIDANCE: A web server for assessing alignment confidence scores. Nucleic Acids Res. 38:W23-W28.

Pucciariello, C., Boscari, A., Tagliani, A., Brouquisse, R., and Perata, P. 2019. Exploring legume-rhizobia symbiotic models for waterlogging tolerance. Front. Plant Sci. 10:578.

Qin, L., Zhao, J., Tian, J., Chen, L., Sun, Z., Guo, Y., Lu, X., Gu, M., Xu, G., and Liao, H. 2012. The high-affinity phosphate transporter GmPT5 regulates phosphate transport to nodules and nodulation in soybean. Plant Physiol. 159:1634-1643.

Quilbé, J., Lamy, L., Brottier, L., Leleux, P., Fardoux, J., Rivallan, R., Benichou, T., Guyonnet, R., Becana, M., Villar, I., Garsmeur, O., Hufnagel, B., Delteil, A., Gully, D., Chaintreuil, C., Pervent, M., Cartieaux, F., Bourge, M., Valentin, N., Martin, G., Fontaine, L., Droc, G., Dereeper, A., Farmer, A., Libourel, C., Nouwen, N., Gressent, F., Mournet, P., D'Hont, A., Giraud, E., Klopp, C., and Arrighi, J.-F. 2021. Genetics of nodulation in *Aeschynomene evenia* uncovers mechanisms of the rhizobium-legume symbiosis. Nat. Commun. 12:829.

Raul, B., Kryvoruchko, I., Benedito, V. A., Bandyopadhyay, K., and Sinharoy, S. 2019. Root nodule development in model versus non-canonical plants. Pages 397-428 in: Plant Biotechnology: Progress in Genomic Era. S. Khurana and R. Gaur, eds. Springer, Singapore.

Reddy, D. S., Bhatnagar-Mathur, P., Cindhuri, K. S., and Sharma, K. K. 2013. Evaluation and validation of reference genes for normalization of quantitative real-time PCR based gene expression studies in peanut. PLoS One 8:e78555.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K. 2015. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43:e47.

Roy, S., Liu, W., Nandety, R. S., Crook, A., Mysore, K. S., Pislariu, C. I., Frugoli, J., Dickstein, R., and Udvardi, M. K. 2020. Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. Plant Cell 32:15-41.

Roy, S., Robson, F., Lilley, J., Liu, C. W., Cheng, X., Wen, J., Walker, S., Sun, J., Cousins, D., Bone, C., Bennett, M. J., Downie, J. A., Swarup, R., Oldroyd, G., and Murray, J. D. 2017. MtLAX2, a functional homologue of the Arabidopsis auxin influx transporter AUX1, is required for nodule organogenesis. Plant Physiol. 174:326-338.

Satgé, C., Moreau, S., Sallet, E., Lefort, G., Auriac, M. C., Remblière, C., Cottret, L., Gallardo, K., Noirot, C., Jardinaud, M. F., and Gamas, P. 2016. Reprogramming of DNA methylation is critical for nodule development in *Medicago truncatula*. Nat. Plants 2:16166.

Senovilla, M., Castro-Rodríguez, R., Abreu, I., Escudero, V., Kryvoruchko, I., Udvardi, M. K., Imperial, J., and González-Guerrero, M. 2018. *Medicago truncatula* copper transporter 1 (MtCOPT1) delivers copper for symbiotic nitrogen fixation. New Phytol. 218:696-709.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 13:2498-2504.

Sharma, V., Bhattacharyya, S., Kumar, R., Kumar, A., Ibañez, F., Wang, J., Guo, B., Sudini, H. K., Gopalakrishnan, S., DasGupta, M., Varshney, R. K., and Pandey, M. K. 2020. Molecular basis of root nodule symbiosis between *Bradyrhizobium* and 'crack-entry' legume groundnut (*Arachis hypogaea* L.). Plants 9:276.

Shimomura, K., Nomura, M., Tajima, S., and Kouchi, H. 2006. LjnsRING, a novel RING finger protein, is required for symbiotic interactions between *Mesorhizobium loti* and *Lotus japonicus*. Plant Cell Physiol. 47:1572-1581.

Shu, H., Luo, Z., Peng, Z., and Wang, J. 2020. The application of CRISPR/Cas9 in hairy roots to explore the functions of AhNFR1 and AhNFR5 genes during peanut nodulation. BMC Plant Biol. 20:417.

Sinharoy, S., and DasGupta, M. 2009. RNA interference highlights the role of CCaMK in dissemination of endosymbionts in the Aeschynomeneae legume *Arachis*. Mol. Plant-Microbe Interact. 22:1466-1475.

Sinharoy, S., Torres-Jerez, I., Bandyopadhyay, K., Kereszt, A., Pislariu, C. I., Nakashima, J., Benedito, V. A., Kondorosi, E., and Udvardi, M. K. 2013. The C_2H_2 transcription factor regulator of symbiosome differentiation represses transcription of the secretory pathway gene *VAMP721a* and promotes symbiosome development in *Medicago truncatula*. Plant Cell 25:3584-3601.

Smit, P., Raedts, J., Portyanko, V., Debellé, F., Gough, C., Bisseling, T., and Geurts, R. 2005. NSP1 of the GRAS protein family is essential for rhizobial Nod factor-induced transcription. Science 308:1789-1791.

Sprent, J. I., and James, E. K. 2007. Legume evolution: Where do nodules and mycorrhizas fit in? Plant Physiol. 144:575-581.

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

Szczyglowski, K., and Ross, L. 2021. Baring the roots of nodulation. Nat. Plants 7:244-245.

Teamtisong, K., Songwattana, P., Noisangiam, R., Piromyou, P., Boonkerd, N., Tittabutr, P., Minamisawa, K., Nantagij, A., Okazaki, S., Abe, M., Uchiumi, T., and Teaumroong, N. 2014. Divergent nod-containing *Bradyrhizobium* sp. DOA9 with a megaplasmid and its host range. Microbes Environ. 29:370-376.

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., and Su, Z. 2017. agriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res. 45:W122-W129.

Tobisa, M., Shimojo, M., and Masuda, Y. 2014. Root distribution and nitrogen fixation activity of tropical forage legume American jointvetch (*Aeschynomene americana* L.) cv. Glenn under waterlogging conditions. Int. J. Agron. 2014:507405.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L., and Pachter, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7:562-578.

Valkov, V. T., Rogato, A., Alves, L. M., Sol, S., Noguero, M., Léran, S., Lacombe, B., and Chiurazzi, M. 2017. The nitrate transporter family protein LjNPF8.6 controls the N-fixing nodule activity. Plant Physiol. 175:1269-1282.

Van de Velde, W., Zehirov, G., Szatmari, A., Debreczeny, M., Ishihara, H., Kevei, Z., Farkas, A., Mikulass, K., Nagy, A., Tiricz, H., Satiat-Jeunemaître, B., Alunni, B., Bourge, M., Kucho, K., Abe, M., Kereszt, A., Maroti, G., Uchiumi, T., Kondorosi, E., and Mergaert, P. 2010. Plant peptides govern terminal differentiation of bacteria in symbiosis. Science 327:1122-1126.

van Velzen, R., Doyle, J. J., and Geurts, R. 2019. A resurrected scenario: Single gain and massive loss of nitrogen-fixing nodulation. Trends Plant Sci. 24:49-57.

van Velzen, R., Holmer, R., Bu, F., Rutten, L., van Zeijl, A., Liu, W., Santuari, L., Cao, Q., Sharma, T., Shen, D., Roswanjaya, Y., Wardhani, T. A. K., Kalhor, M. S., Jansen, J., van den Hoogen, J., Güngör, B., Hartog, M., Hontelez, J., Verver, J., Yang, W. C., Schijlen, E., Repin, R., Schilthuizen, M., Schranz, M. E., Heidstra, R., Miyata, K., Fedorova, E., Kohlen, W., Bisseling, T., Smit, S., and Geurts, R. 2018. Comparative genomics of the nonlegume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobium symbioses. Proc. Natl. Acad. Sci. U.S.A. 115:E4700-E4709.

Vernié, T., Moreau, S., de Billy, F., Plet, J., Combier, J. P., Rogers, C., Oldroyd, G., Frugier, F., Niebel, A., and Gamas, P. 2008. EFD is an ERF transcription factor involved in the control of nodule number and differentiation in *Medicago truncatula*. Plant Cell 20:2696-2713.

Wang, C., Yu, H., Luo, L., Duan, L., Cai, L., He, X., Wen, J., Mysore, K. S., Li, G., Xiao, A., Duanmu, D., Cao, Y., Hong, Z., and Zhang, Z. 2016. *NODULES WITH ACTIVATED DEFENSE 1* is required for maintenance of rhizobial endosymbiosis in *Medicago truncatula*. New Phytol. 212:176-191.

Wang, D., Griffitts, J., Starker, C., Fedorova, E., Limpens, E., Ivanov, S., Bisseling, T., and Long, S. 2010. A nodule-specific protein secretory pathway required for nitrogen-fixing symbiosis. Science 327:1126-1129.

Wang, Z., Wang, L., Wang, Y., and Li, X. 2020. The NMN module conducts nodule number orchestra. iScience 23:100825.

Wittenberg, J. B. 1978. Leghemoglobin. Low temperature optical spectra of acid and alkaline forms of leghemoglobin (IV). Configuration of the heme. J. Biol. Chem. 253:5690-5693.

Xie, F., Murray, J. D., Kim, J., Heckmann, A. B., Edwards, A., Oldroyd, G. E. D., and Downie, J. A. 2012. Legume pectate lyase required for root infection by rhizobia. Proc. Natl. Acad. Sci. U.S.A. 109:633-638.