Transcriptomic Analysis With the Progress of Symbiosis in ‘Crack-Entry’ Legume *Arachis hypogaea* Highlights Its Contrast With ‘Infection Thread’ Adapted Legumes

Kanchan Karmakar,¹ Anindya Kundu,¹ Ahsan Z Rizvi,² Emeric Dubois,³ Dany Severac,³ Pierre Czernic,² Fabienne Cartieaux,² and Maitrayee DasGupta^{1,†}

¹Department of Biochemistry, University of Calcutta, Kolkata 700019, India; ²LSTM, Univ. Montpellier, CIRAD, INRA, IRD, SupAgro, Montpellier, France; and ³Montpellier GenomiX (MGX), c/o Institut de Génomique Fonctionnelle, 141 rue de la cardonille, 34094 Montpellier Cedex 05, France

Accepted 13 August 2018.

In root-nodule symbiosis, rhizobial invasion and nodule organogenesis is host controlled. In most legumes, rhizobia enter through infection threads and nodule primordium in the cortex is induced from a distance. But in dalbergoid legumes like *Arachis hypogaea*, rhizobia directly invade cortical cells through epidermal cracks to generate the primordia. Herein, we report the transcriptional dynamics with the progress of symbiosis in *A. hypogaea* at 1 day postinfection (dpi) (invasion), 4 dpi (nodule primordia), 8 dpi (spread of infection in nodule-like structure), 12 dpi (immature nodules containing rod-shaped rhizobia), and 21 dpi (mature nodules with spherical symbiosomes). Expression of putative ortholog of symbiotic genes in ‘crack entry’ legume *A. hypogaea* was compared with infection thread-adapted model legumes. The contrasting features were i) higher expression of receptors like *LYR3* and *EPR3* as compared with canonical Nod factor receptors, ii) late induction of transcription factors like *NIN* and *NSP2* and constitutive high expression of *ERF1*, *EIN2*, *bHLH476*, and iii) induction of divergent pathogenesis-responsive *PR-1* genes. Additionally, symbiotic orthologs of *SymCRK*, *ROP6*, *RR9*, *SEN1*, and *DNF2* were not detectable and microsynteny analysis

indicated the absence of a *RPG* homolog in diploid parental genomes of *A. hypogaea*. The implications are discussed and a molecular framework that guides crack-entry symbiosis in *A. hypogaea* is proposed.

Nitrogen-fixing root-nodule symbiosis (RNS) allows plants to house bacterial diazotrophs in an intracellular manner (Kistner and Parniske 2002). The establishment of RNS involves rhizobial invasion in the root epidermis and nodule organogenesis in the root cortical cells. The most common invasion strategy is through root-hair curling and infection thread (IT) formation, in which the nodule primordia are induced from a distance (Gage 2004). Invasion through the IT is widespread among temperate legumes like *Vicia*, *Trifolium*, and *Pisum* species. Model legumes like *Lotus japonicus* and *Medicago truncatula* also undertake IT-mediated rhizobial invasion (Geurts and Bisseling 2002; Oldroyd and Downie 2004, 2006). The alternate mode of rhizobial invasion is known as ‘crack entry’, in which the rhizobia enter through natural cracks at the lateral root base in an intercellular manner. Approximately 25% of all legume genera are adapted to crack-entry and it is a characteristic feature of some subtropical legumes belonging to dalbergoid/genistoid clades, like *Arachis*, *Aeschynomene*, and *Stylosanthes* species (Sprent and James 2007). In these legumes, rhizobia directly access the cortical cells for development of their nodule primordia and the infected cells repeatedly divide to develop the mature nodule with an aescynomenoid infection zone (IZ) devoid of uninfected cells (Boogerd and Rossum 1997; Fabre et al. 2015). Very limited information is available for crack-entry legumes from the dalbergoid/genistoid clade that are basal in their divergence within the papilionoideae.

Investigations on model legumes that are adapted to IT-mediated rhizobial invasion have unraveled the molecular basis of RNS. The host responses are initiated by Nod factor (NF) receptors (NFRs) LjNFR1/MtLYK3 and LjNFR5/MtNFP (Arrighi et al. 2006; Madsen et al. 2003; Radutoiu et al. 2003; Smit et al. 2007). Another NF-induced receptor LjEPR3 was shown to monitor rhizobial exopolysaccharide (EPS) in *L. japonicus*, indicating a two-stage mechanism involving sequential receptor-mediated recognition of NF and EPS signals to ensure host symbiont compatibility (Kawaharada et al. 2015). Downstream to NFRs is the ‘SYM pathway’ consisting of the receptor kinase LjSYMRK/MtDMI2 (Endre et al. 2002; Stracke et al. 2002), the predicted ion-channel proteins LjCASTOR and LjPOLLUX/MtDMI1 (Ané et al. 2004; Imaizumi-Anraku et al.

Kanchan Karmakar and Anindya Kundu contributed equally to this work.

The raw FASTQ files for the 18 libraries were deposited in the Gene Expression Omnibus database of the National Center for Biotechnology Information under accession number GSE98997.

Current address for Ahsan Z Rizvi: 112, INSERM U981, Bâtiment Médecine Moléculaire (B2M), Gustave Roussy, 114, rue Edouard Vaillant, 94805 Villejuif Cedex-France

[†]Corresponding author: Maitrayee DasGupta;
E-mail: maitrayee_d@hotmail.com

Funding: This work was funded by grants from the government of India: IFCPAR/CEFIPRA (IFC/5103-4/2014/543); DBT-CEIB (Centre of Excellence and Innovation in Biotechnology, BT/01/CEIB/09/VI/10); DBT-IPLS (BT/PR14552/INF/22/123/2010); DST (EMR/2015/001006); fellowship to Kanchan Karmakar and Ahsan Z Rizvi (IFCPAR/CEFIPRA: IFC/5103-4/2014/543); fellowship to Anindya Kundu (Council of Scientific and Industrial Research, CSIR-09/028[0756]/2009-EMR-I).

*The e-Xtra logo stands for “electronic extra” and indicates that figures and tables showing more complete results of our analyses and lists of the genes analyzed are available online.

2005), the nucleoporins LjNUP85 and LjNUP133 (Kanamori et al. 2006; Saito et al. 2007), the Ca^{2+} /calmodulin-dependent protein kinase LjCCaMK/MtDMI3 (Lévy et al. 2004; Tirichine et al. 2006), and the transcription factor (TF) LjCYCLOPS/MtIPD3 (Messinese et al. 2007; Singh et al. 2014; Yano et al. 2008). Nodulation-specific TFs, such as MtNSP1/LjNSP1, MtNSP2/LjNSP2, MtERF1, and MtNIN/LjNIN function downstream of the SYM pathway and are involved in transcriptional reprogramming for initiation of RNS (Kaló et al. 2005; Middleton et al. 2007; Schauser et al. 1999; Smit et al. 2005).

Transcriptome analyses during symbiosis in several legumes like *M. truncatula*, *L. japonicus*, *Glycine max*, and *Cicer arietinum* provided molecular insights into symbioses that use the ‘root hair’ mode of entry (Limpens et al. 2014; Kant et al. 2016; Takanashi et al. 2012; Yuan et al. 2017). These resources have been useful in determining gene expression associated with different stages of RNS, for example, symbiont recognition, onset of infection, meristem activity for nodule organogenesis, and functional differentiation of infected cells (Breakspear et al. 2014; Roux et al. 2014). The outcome, in most cases, highlighted numerous novel genes coexpressed with known markers for symbiosis. Nodule transcriptome analyses in *M. truncatula* highlighted that TF genes essential for the control of the root apical meristem were also expressed in the nodule meristem (Roux et al. 2014). Transcriptomic analysis of infected root hairs further revealed involvement of genes regulating cell cycle, flavonoid biosynthesis, and auxin signaling during IT initiation (Breakspear et al. 2014). It also enabled identification of genes and processes previously undetected in whole-root studies or in forward genetic analyses. In *M. truncatula* transcriptomic analysis of symbiotic mutants further improved our understanding of gene expression associated with specific processes of symbiosis (Jardinaud et al. 2016; Moreau et al. 2011). But unlike in model legumes, transcriptomic profiling for species that utilize crack entry has rarely been reported to date and the symbiotic pathway of crack entry nodulation is largely unknown.

An earlier report listed several differentially expressed genes (DEGs) at an early stage of symbiosis in a nodulation-competent *A. hypogaea* in comparison with an incompetent variety (Peng et al. 2017). Our manuscript undertakes a systematic effort to study the transcriptomic dynamics with the onset and advancement of symbiosis at five different nodule developmental stages in *A. hypogaea*, using uninfected roots (UI) as a reference. Primarily, we focused on symbiotic genes that were characterized in model legumes and could be broadly classified according to their functional involvement in bacterial recognition, early signaling, infection, nodule organogenesis, functional differentiation, and nodule number regulation. Expression of putative orthologs of these symbiotic genes in crack-entry legume *A. hypogaea* is compared with the corresponding expression profiles in *M. truncatula* and *L. japonicus* that undertake IT-mediated symbiosis. And the outcome of the analysis is summarized in a model that can be further utilized for understanding nodulation dynamics in basal legumes.

RESULTS

Progress of symbiosis in *Arachis hypogaea*.

Within 3 weeks of infection with *Bradyrhizobium* sp. strain SEMIA 6144, *A. hypogaea* roots developed spherical functional nodules. We followed the progress of symbiosis in *A. hypogaea* for 21 days, to identify the distinct stages of development by ultrastructure analysis. There are rosettes of root hairs in the junction of taproot and lateral root that are reported to be important for bacterial invasion in *A. hypogaea* (Boogerd and Rossum 1997). Within 1 day postinfection (dpi), rhizobia

was found to be adhered to these root hairs (Fig. 1A to C). Within 4 dpi, bump-like primordial structures were noted at the lateral root bases (Fig. 1D). The longitudinal sections of these primordia revealed one or more centrally located defined pockets of rhizobia-infected cells that were surrounded by uninfected cells (Fig. 1E). These pockets of rhizobia-infected cells were distinct by having reduced Calcofluor-binding ability, indicating that they are thin-walled. The intracellularized rhizobia within the infection pockets were undifferentiated and rod-shaped (Fig. 1F). The infection pockets observed at 4 dpi act as IZ founder cells, and it is their uniform division and differentiation that give rise to the distinct aescynomenoid-type IZ in mature nodules. There has not been a single case in which uninfected primordium was noted, which is in accordance with the proposition of infection preceding development of aescynomenoid nodules (Fabre et al. 2015). By 8 dpi, there was a visible nodule-like structure at the lateral root base (Fig. 1G). Ultrastructure analysis revealed that, by 8 dpi, the compactness of the primordial structure with defined pockets of infected cells was lost and the IZ started growing by division of the infected cells (Fig. 1H and I). By 12 dpi, there were white spherical nodules (Fig. 1J). At this stage, the tissue organization turned aescynomenoid where there were no uninfected cells in the IZ and the endocytosed rhizobia remained undifferentiated and rod-shaped (Fig. 1K and L). At 21 dpi, the nodules were mature and functional where the rhizobia differentiated within the plant-derived peribacteroid membranes to develop spherical symbiosomes (Fig. 1M to O).

Transcriptome analysis with the progress of symbiosis in *Arachis hypogaea*.

Ultrastructural analysis revealed five distinct stages during the progress of symbiosis in *A. hypogaea* 1 dpi (recognition and invasion), 4 dpi (primordia formation), 8 dpi (nodule-like structure), 12 dpi (immature nodules with rod-shaped rhizobia), and 21 dpi (mature nodules with spherical symbiosomes). To probe into the expression of genes associated with the progress of symbiosis, RNA was extracted from these stages. As a reference (0 dpi), the UI roots were used. RNA-seq was done in triplicate for these six stages, using Illumina single-end sequencing technology (Illumina HiSeq 2000 SR50). The genomic data from *Arachis duranensis* (AA) and *Arachis ipaensis* (BB), which are two wild diploid parents of *A. hypogaea*, were used to assess the quality and coverage of the assembled transcriptomes. A total of 1,429,876,614 raw reads of 50 bp (approximately 71.5 Gb) were generated with an average of 88,029,386 reads per library. This was 600 times the total size of transcript sequences (109.0 Mb) of *A. hypogaea* for both AA and BB genomes and gave an average coverage of 36 times per library. The proportion of clean reads among the total acquired reads was more than 91.34% (Table 1). The filtered reads were simultaneously mapped to the AA and BB genomes, in which the overall accepted mapping rate per library ranged from 80.15 to 89.98%, with an average mapping rate of 86.42% with *A. duranensis* (AA) and 86.65% with *A. ipaensis* (BB). For both AA and BB genomes, about 66% of reads aligned to a gene exon in an unambiguous way, whereas the remaining 33% of reads aligned outside exons.

The expression level of each assembled transcript sequence was measured through FPKM (fragments per kilobase per million reads) values. The DEGs in the five different stages of symbiosis were evaluated by the significance of differences in their expression with respect to UI roots, using a false discovery rate (FDR) < 0.05, *P* value < 0.05, and fold change $\log_2$ ratio ≥ 1 . Since a single UI was used for all the stages, there could be representations of some DEGs that are associated with normal root development within the experimental timeframe.

Comparison between upregulated and downregulated DEGs at different stages is shown in a Venn diagram in Figure 2A. A total of 2,745 genes were up-regulated (1,297 in AA, 1,448 in BB) and a total 20,259 genes are down-regulated (9,552 in AA, 10,707 in BB), with the progress of symbiosis in *A. hypogaea* of

which 59 genes (29 in AA, 30 in BB) were commonly up-regulated and 2,054 genes (1,015 in AA, 1,039 in BB) were commonly down-regulated in all five of the stages of symbiosis. From the Venn diagram, we identified those genes that were first up- or down-regulated at a particular stage, though their

Fig. 1. Progress of symbiosis in *Arachis hypogaea* infected with *Bradyrhizobium* sp. strain SEMIA6144. **A**, Root was harvested at 1 day postinfection (dpi) at the root junction. **B**, Scanning electron microscopy of inoculated root hair at lateral root junction and **C**, confocal laser scanning microscopy of lateral root junction. **D** to **F**, Section of nodule primordia at 4 dpi (dashed lines indicate the infection zone in **E** and asterisks indicate the identical position in **E** and **F**) and **G** to **I**, 8 dpi (asterisks indicate the identical position in **H** and **I**). **J** to **L**, Section of nodules at 12 and **M** to **O**, 21 dpi. **A**, **D**, **G**, **J**, and **M**, stereomicroscopy; **C**, **E**, **H**, **K**, and **N**, bright field and propidium iodide (PI) merged; **F**, **I**, and **L**, PI + Calcofluor merged; and **O**, PI + Calcofluor + syto9 merged. Arrows indicate rod shaped rhizobia in **F**, **I**, and **L**. Scale bar = 500 µm (**A**, **D**, **G** and **J**), 1 mm (**M**), 2 µm (**B**), 100 µm (**E**, **H**, **K**, and **N**), and 10 µm (**C**, **F**, **I**, **L**, and **O**).

subsequent regulations could be different. Differentially expressed genes show clear transcriptional shifts at these stages, and the diverse expression patterns of these genes are indicated in a heatmap (Fig. 2B). Hierarchical clustering (Fig. 2B) and principal component analysis (PCA) both suggested that there is substantial overlap between the transcriptome of infected roots at 1 dpi and at 4 dpi, when rhizobial invasion and primordia formation occurs. Similarly, there was overlap between transcriptome of immature nodules at 8 and 12 dpi, when the primordia structurally develop into a nodule. Finally, the transcriptome from mature nodules at 21 dpi was distinct from all other stages in which a nodule matures to its functional form. These expression waves indicated that there could be three major transcriptional programs during the inception and maturation of symbiosis in *A. hypogaea*. The actual number of genes upregulated or downregulated from either the AA or BB genome at 1 dpi was 337 and 8,046 (156 upregulated and 3,826 downregulated in AA; 181 upregulated and 4,219 downregulated in BB), at 4 dpi was 388 and 4,643 (180 upregulated and 2,167 downregulated in AA; 208 upregulated and 2,476 downregulated in BB), at 8 dpi was 641 and 1,587 (297 upregulated and 735 downregulated in AA; 344 upregulated and 852 downregulated in BB), at 12 dpi was 502 and 1,548 (240 upregulated and 751 downregulated in AA; 262 upregulated and 797 downregulated in BB), and for mature nodules at 21 dpi was 877 and 4,435 (423 upregulated and 2,141 downregulated in AA; 454 upregulated and 2,294 downregulated in BB). This analysis shows that, within 1 to 4 dpi, the number of downregulated DEGs were approximately 18-fold higher than the number of upregulated DEGs, whereas in other stages the difference was approximately three- to fivefold. Previous reports on symbiosis-associated transcriptome of other legumes also indicated the prevalence of a higher number of downregulated transcripts (Boscari et al. 2013; Yuan et al. 2017). But the significantly high number of downregulated transcripts in early stages of nodule inception could be a unique feature for *A. hypogaea* RNS.

Functional analysis of DEGs.

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) categories were searched for terms that are significantly enriched in our DEGs. Among the 1,248 enriched

GO terms, there was a major representation of defense response genes; 470 and 31 such defense-related GO terms were enriched in down- and upregulated DEGs, respectively. Accordingly, KEGG analysis of plant-pathogen interaction pathways show that most genes involved in pattern-triggered immunity was notably down-regulated (Fig. 3A and B). The flagellin-sensing receptor FLS2 mediated the mitogen-activated protein kinase pathway (MAPK), however remained active along with a subset of cyclic nucleotide-gated ion channel genes (CNGCs) and genes encoding respiratory burst oxidase (*Rboh*) proteins. A subset of genes involved in the effector-triggered immunity also remained active during symbiosis, for example, genes encoding resistance proteins, like *RPM1*, *RPS2*, *RPS5*, *Pti1* kinase, and pathway regulators, like *SGT1*, *HSP90*, and *EDS1*. Intriguingly there was a significant upregulation of genes encoding pathogenesis-responsive (PR)-1 proteins, which are members of the CAP (cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins) superfamily (Breen et al. 2017). The PR-1 proteins upregulated during symbiosis clustered away from the PR-1 proteins that were reported to be upregulated in defense responses, indicating the symbiosis-associated PR-1 proteins were divergent in nature (Fig. 3C). There are two PR-1 proteins that clustered with defense-responsive PR proteins and these *PR-1* genes were not upregulated during symbiosis, further confirming the symbiotic PR-1 proteins to be distinct. PR-1 proteins harbor an embedded defense signaling peptide (CAP-derived peptides or CAPE) in which CNYxPxGNxxxxxPY is considered as a functional motif that marks cleavage of these bioactive peptides (Breen et al. 2017). The cleavage site is conserved in both classes of PR-1 proteins, suggesting the CAPE peptides could also be generated from the divergent PR-1 proteins synthesized during symbiosis (Fig. 3C and D). Since genes encoding CAP proteins are marker genes for the salicylic acid (SA) signaling pathway and systemic acquired resistance, we also checked the SA/JA (jasmonic acid) pathways to further understand the symbionts responsive signaling in *A. hypogaea*. As shown in Figure 3A, the JA pathway was completely downregulated but the SA-responsive genes, like *TGA1* and *NPR1*, were up-regulated. Thus, symbiotic *PR-1* gene expression could be justified by the activation of the SA-mediated signaling.

Several genes are reported to be expressed in nodulating roots by comparing the transcriptome profiles of nonnodulating and nodulating lines of *A. hypogaea* (Peng et al. 2017). The list includes known symbiotic genes, like *NIN* and *CLE13*, and other genes, like *NF-YA10*, *Myb127*, a receptor kinase, a soluble kinase, a F-BOX protein, TFs of the SHI family, and a lectin. All these genes were represented in our upregulated transcriptome, which thereby revalidates the importance of expression of these genes during the onset of symbiosis in *A. hypogaea*.

Expression profiles of putative orthologs of symbiotic genes.

Our final objective was to understand the expression of the putative orthologs of symbiotic genes in *A. hypogaea* that are characterized in model legumes like *M. truncatula* and *L. japonicus*. A total of 71 such genes were chosen and were functionally classified on the basis of their demonstrated role during nodulation in model legumes (Fig. 4). A BLAT search on *A. ipaensis* and *A. duranensis* genomes identified 63 genes that were annotated in PeanutBase and six genes that are yet to be annotated, i.e., *NSP2*, *ERN1*, *ERF1*, *VPY*, *ENOD40*, and *RSD*. A homolog for *MIRPG* could not be detected by BLAT search in either of the two parental genomes of *A. ipaensis* and *A. duranensis*, which could be due to an incomplete genome sequence or the quality of the gene model predictions.

Table 1. Summary of raw Illumina sequencing and filtered reads after trimming and alignment of reads to AA (*Arachis duranensis*) and BB (*A. ipaensis*) genomes in each library

Samples	Raw reads	Filter reads (%)	% of total mapped reads	
			AA (<i>A. duranensis</i>) (%)	BB (<i>A. ipaensis</i>) (%)
UI	57,913,214	95.08	82.96	86.71
	65,769,652	94.88	87.78	84.34
	85,558,982	90.59	85.29	86.19
1 dpi	61,747,256	95.00	88.80	87.51
	62,925,897	91.38	88.78	89.98
	74,542,718	91.68	87.85	88.76
4 dpi	96,807,458	89.21	88.46	84.42
	97,275,402	90.46	86.61	87.71
	79,984,160	91.21	85.24	86.25
8 dpi	89,987,084	90.09	86.03	87.19
	11,535,441	89.85	86.20	83.03
	77,031,650	91.50	87.81	88.95
12 dpi	57,198,145	90.28	86.77	88.08
	74,743,668	92.33%	88.47%	89.71%
	96,613,689	89.38%	86.12%	87.20%
21 dpi	65,329,428	90.03%	87.81%	86.81%
	100,254,251	90.71%	84.45%	85.42%
	70,839,550	90.46%	80.15%	81.44%

Microsynteny analysis in Phytozome indicated the absence of *MtRPG* orthologs in *A. ipaensis*, whereas, in *A. duranensis*, the entire gene block corresponding to *MtRPG* was not detectable. In a recent report, Griesmann et al. (2018) also indicated the loss of *RPG* gene in the *A. ipaensis* genome. Similar microsynteny analysis also revealed the absence of *MtDNF2* in *A. ipaensis* genome but indicated the presence of a singleton in *A. duranensis*. This singleton has 16% amino acid sequence homology with *MtDNF2* and phylogenetic analysis

(maximum likelihood) reveals that the *A. duranensis* sequence is highly divergent from both legume and nonlegume sequences. *MtDNF2* codes for a phosphatidylinositol phospholipase C-like protein, but a BLAST search indicated this singleton to code for an unknown protein. While the singleton showed very low transcript density, we could not identify any homologous contigs corresponding to *MtRPG* and *MtDNF2* in our RNA-seq data, using Burrow-Wheeler Aligner (Fig. 4G).

Fig. 2. Comparison of differentially expressed genes (DEGs) at five different stages during the progress of symbiosis in *Arachis hypogaea*. The total number of DEGs at 1, 4, 8, 12, and 21 days postinfection (dpi) was identified by EdgeR analysis from both AA and BB genomes. **A**, Venn diagram showing comparison between upregulated and downregulated DEGs. **B**, Heat map of the hierarchical cluster analysis of the same DEGs. Columns represent different timepoints and the arborescence above them indicates similarity. Scale bar represents log2-fold change in gene expression. Numbers beside the heat map indicate the number of DEGs at different timepoints.

Orthology for the identified symbiotic genes were checked by reciprocal BLAST and sequence alignments. Symbiotic orthologs were identified for 65 of 70 genes, for which the *A. hypogaea* homologs clustered with other legumes in the corresponding gene

trees. Exceptions were the homologs for *MtSymCRK*, *LjROP6*, *MtRR9*, and *LjSEN1*, for which the *A. hypogaea* genes clustered with nonlegumes, and *MtDNF2*, which was highly divergent and did not cluster with either legumes or nonlegumes. For

Fig. 3. Differentially expressed genes (DEGs) associated with plant-pathogen interaction and jasmonic acid and salicylic acid signaling pathways. **A**, Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of upregulated and downregulated DEGs in the respective pathways. **B**, Heat map of the DEGs that are annotated in KEGG pathway analysis. **C**, Neighbor-joining phylogenetic tree of all the annotated CAP (cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins), using nontruncated amino acid sequences when a branch denotes divergent upregulated CAP-PR1 proteins. **D**, CLUSTALW sequence alignment of CAPE (CAP-derived) peptides using Multalin; an arrow indicates cleavage site for CAPE peptide; CBM and CAP domain are annotated by shaded boxes, and bullets indicate *Arachis hypogaea* CAP peptides.

Fig. 4. Comparative expression pattern of 71 symbiotic genes in *Arachis hypogaea*, *Medicago truncatula*, and *Lotus japonicus*. For *A. hypogaea*, histograms represent normalized RNA-seq reads (fragments per kilobase per million reads) of symbiotic orthologous genes aligned to *A. duranensis* and *A. ipaensis* genomes. For *M. truncatula* and *L. japonicus*, histograms represent microarray data retrieved from the respective Gene Expression Atlas Affymetrix database (i.e., MtGEA and LjGEA) (Benedito et al. 2008; Verdier et al. 2013). Expression values of symbiotic genes (in arbitrary units [a.u.]) during nodulation are grouped as **A**, bacterial recognition, **B**, SYM pathway and early signaling, **C**, interactors, **D**, early transcription factors, **E**, infection, **F**, nodule organogenesis, **G**, nodule differentiation, **H**, nodule number regulation (dpi: days postinfection by *Bradyrhizobium* sp. strain SEMIA6144 for *A. hypogaea*, *Sinorhizobium meliloti* for *M. truncatula*, and *Mesorhizobium loti* R7A for *L. japonicus*). a, b, c, d, e, and f in the x axis represent 0, 1, 4, 8, 12, and 21 dpi, respectively for *A. hypogaea*; 0, 3, 6, 10, 14, and 20 dpi, respectively, for *M. truncatula*; 0, 1, 3, 7, 14, and 21 dpi, respectively, for *L. japonicus*. An asterisk indicates genes that are differentially expressed in *A. hypogaea*; a diamond indicates expression of a nonsymbiotic ortholog in *A. duranensis* and *A. ipaensis*, and an X indicates a gene absent in the diploid genome.

comparison of expression between the symbiotic or nonsymbiotic orthologs in *Arachis* sp. with their counterpart in model legumes, we used the microarray data derived from the *M. truncatula* gene expression atlas (MtGEA) (Benedito et al. 2008) and *L. japonicus* gene expression atlas (LjGEA) (Verdier et al. 2013). For the symbiotic genes that are characterized from one of these model legumes, reciprocal BLAST was done to identify the ortholog in the other. The pattern of expression of these genes with the progress of symbiosis was compared between *A. hypogaea* and model legumes. Both absolute expression (Fig. 4) and relative (log2-fold) expression values were analyzed so that high and constitutively expressed genes were not ignored. The notable exceptions are mentioned below, and the analyzed data are summarized in a model in Figure 5. The contrasting genes are highlighted along with the genes that are absent or divergent in sequence.

In the recognition module, expression of lipo-chitooligosaccharide (LCO)-binding *AhLYR3* and EPS-binding *AhEPR3* were significantly high as compared with the orthologs of canonical NF LysM domain-containing receptors *AhNFR1* and *AhNFR5* (Figs. 4A and 5). Most of the SYM pathway members and their interactors and the early TFs showed similar trends in expression in both *A. hypogaea* and in model legumes for both inducible and constitutive genes (Fig. 4B to D). Exception was the gene encoding an ortholog of cyclic nucleotide-gated channel *MtCNGC* (Charpentier et al. 2016), which was significantly upregulated in *A. hypogaea* (Fig. 4B). Among the interactors, the ortholog of an E3 ubiquitin ligase, *MtPUB1*, was constitutively expressed in *A. hypogaea*, whereas in the model legumes its expression is up-regulated only upon bacterial invasion (Fig. 4C). In model legumes, TFs like *NIN* and *NSP2* are upregulated with bacterial invasion in early stages (Kaló et al. 2005; Schauser et al. 1999). In contrast, upregulation of *AhNIN* was noted at 8 dpi onwards, and *AhNSP2* expression was induced only at 21 dpi in mature nodules (Figs. 4D and 5). The expression of ethylene responsive TF, *AhERF1*

appeared to be very high and constitutive with the progression of symbiosis. On the other hand, in model legumes, the expression of *ERF1* was significantly reduced (*L. japonicus*) or upregulated (*M. truncatula*) upon bacterial invasion.

Among the genes associated with rhizobial invasion, the homolog for *MtRPG* (Arrighi et al. 2008), a factor responsible for rhizobium-directed polar growth of Its, was absent in *A. hypogaea* (Fig. 5). Additionally, symbiotic orthologs of *LjROP6*, a Rho-like GTPase that is responsible for IT elongation in the cortex was not detected in *Arachis* sp., though the expression of its nonsymbiotic homolog was significantly high during symbiosis (Fig. 4C). Symbiotic orthologs of flotillins-like *MtFLOT2* and *MtFLOT4* (Haney and Long 2010; Ke et al. 2012), which are required for lipid raft formation during infection, were present in *A. hypogaea*, but only the expression of *AhFLOT2* was detectable during symbiosis, suggesting its adaptation during crack-entry nodulation (Fig. 4E). The expression of the ortholog of *LjCERBERUS*, a WD-40 repeats U-box protein essential for the progress of infection in *L. japonicus* (Yano et al. 2009), was significantly high in *A. hypogaea* as compared with its expression pattern in model legumes (Fig. 4E).

Among the factors associated with nodule development, the transcript for *MtDNF2* homolog encoding a phosphatidylinositol phospholipase C-like protein (Bourcy et al. 2013), was not detectable in *A. hypogaea*, but syntenic analysis revealed a highly divergent copy of DNF2, represented by a singleton in our RNA-seq data from a progression of symbiotic samples (Fig. 4G). Using the Burrow-Wheeler Aligner, we found that the singleton showed very low transcript density in our RNA-seq data, and intriguingly, PeanutBase also does not show any expression from this locus. In addition, symbiotic orthologs of non-RD receptor kinase *MtSymCRK* (Berrabah et al. 2014), type A response regulator *MtRR9*, and metal transporter *LjSEN1* was also not detectable in *A. duranensis* and *A. ipaensis* genomes (Fig. 4F and G). But significant expression of the nonsymbiotic homologs of these genes was noted during the

Fig. 5. Simplified model for the symbiotic signaling pathway in *Arachis hypogaea*. Symbiotic genes identified in *Medicago truncatula* and *Lotus japonicus* are listed and classified according to their main symbiotic functions. The homologs of these genes in *A. hypogaea* that are distinct from model legumes are marked as follows: an X (absent in diploid genome), shaded diamonds (symbiotic ortholog absent), an upward arrow (upregulation), and black diamonds (divergent expression).

progress of symbiosis in *A. hypogaea*. On the other hand, the expression of orthologs for cytokinin inducible TF *MtHLH476* (Ariel et al. 2012) and the ethylene-responsive TF *MtEIN2* (Penmetsa et al. 2008) (*sickle*) was found to be significantly higher in *A. hypogaea* as compared with their expression pattern in model legumes (Figs. 4F and H and 5). Apart from the indicated differences, relative expression patterns of all other symbiotic genes were found to be comparable between *A. hypogaea* and model legumes during progress of symbiosis, indicating their conserved roles.

Quantitative reverse transcription-polymerase chain reaction (qRT-PCR) validation of the RNA-seq results.

To validate the RNA-seq results, we did qRT-PCR analysis using *AhActin* as the reference gene. The stability of expression of *AhActin* during symbiosis in *A. hypogaea* was reported earlier (Kundu and DasGupta 2018; Morgante et al. 2011). The expression was evaluated for 10 symbiotic genes from different modules that are described in Figure 5. Additionally, expression was evaluated for five more genes that were described to be

important for *A. hypogaea* symbiosis (Peng et al. 2017). The results were in agreement with the transcriptional profile data for 83 of 90 (92%) datapoints. The fold change of expression did not always match between RNA-seq and qRT-PCR, but the trend of expression was similar in most of the timepoints for all the genes. No change of expression was noted for any of these genes in uninoculated roots corresponding to different infection timepoints, further confirming the expression of these genes to be specifically associated with symbiosis.

PCA of symbiotic gene expression in *Arachis hypogaea* and model legumes.

PCA was done to check if there was a signature, in the expression pattern of symbiotic genes in crack-entry legume *A. hypogaea*, that contrasts with the model legumes in which rhizobial entry is IT-mediated. Figure 6 is a projection of differential expression of symbiotic genes from *A. hypogaea*, *M. truncatula*, and *L. japonicus* for all the infection timepoints into the first two principal components. Altogether, expression of around 87% of the genes were accommodated along

Fig. 6. Principal component analysis (PCA) of differentially expressed symbiotic genes. Principal components of the variability in expression values for the symbiotic genes at all the infection timepoints in *Arachis ipaensis*, *Medicago truncatula*, and *Lotus japonicus* was calculated using R-Package. The two principal components and their fraction of the overall variability of the data (%) are shown on the x axis (PC1 68.40%) and the y axis (PC2 19.70%).

dimension 1 (PC1) and dimension 2 (PC2) in the analysis. Expressions of symbiotic genes that are constitutive or show minimal change are clustered near the origin for both *Arachis* sp. and the model legumes. These include the SYM pathway and most early signaling genes that are likely to be regulated at the posttranscriptional level. The genes that were noted in Figure 4 to have a contrasting trend in expression between *A. hypogaea* and the model legumes were all placed in opposing or adjacent quadrants in the PCA. For example, *AhCNGC*, *AhNIN*, *AhNSP2*, *AhPUB1*, and *AhCERBERUS* were clustered separately from their counterparts in model legumes. On the other hand, there are some factors whose contrasting trends in expression between *A. hypogaea* and model legumes was not obvious in Figure 4, but PCA highlighted their differences. For example, *AhVPY*, *AhENOD40*, *AhWOX5*, *AhRR1*, and *AhRR4* were placed in opposing or adjacent quadrants from their counterparts in model legumes. These factors highlighted by PCA can also be considered to be differentially adapted for crack-entry mediated rhizobial invasion and aeschynomenoid nodule development in *A. hypogaea*.

DISCUSSION

This is the first systematic effort toward transcriptome profiling with the progress of symbiosis in a crack-entry legume, *A. hypogaea*. We could correlate the hierarchical shift in upregulated DEGs to the morphological changes occurring during different timepoints of nodule development, and the process appeared to be governed by three major transcriptional programs. The first program is associated with rhizobial recognition and generation of nodule primordia by 1 to 4 dpi, the second program is associated with structural development of nodules by 8 to 12 dpi, and the third program is associated with functional maturation of nodules at 21 dpi (Figs. 1 and 2). The comparison between expression patterns of putative orthologs of symbiotic genes with progress of symbiosis in *A. hypogaea* with model legumes highlighted the genes that are important or dispensable for the crack entry-mediated root nodule symbiosis (Figs. 4, 5, and 6). Additionally, absence of symbiotic orthologs of certain symbiotic genes in *A. hypogaea* was highlighted. The probable implications of these distinguishing features in *A. hypogaea* are discussed in the light of their importance during crack-entry nodulation.

The most significant observation in the *A. hypogaea* symbiotic transcriptome was the significantly high expression of a group of genes encoding a divergent form of cysteine-rich PR-1 proteins (Fig. 3). Upregulation of *AhCNGC* and *AhERF1* in the symbiotic transcriptome (Fig. 4) could explain the expression of *PR1* genes, as their role in SA-induced PR gene expression is well-documented (Asamizu et al. 2008; Moeder et al. 2011; Solano et al. 1998). PR-1 proteins are ubiquitous across plant species and are among the most abundantly produced proteins in plants in response to pathogen attack (Breen et al. 2017). As a result, it is used as a marker for SA-mediated disease resistance in plants. Although differential expression of defense response genes belonging to GO:0006952 (defense related) and *PR-1* and *PR-10* protein families has previously been reported for *M. truncatula* RNS (Jardinaud et al. 2016), this is the first report in which a divergent group of PR-1 proteins is shown to be associated with nodule development (Fig. 3). PR-1 proteins harbor a caveolin binding motif (CBM) that binds sterol and the embedded Prorich C-terminal peptide (CAPE) is involved in plant immune signaling (Breen et al. 2017). All the symbiotic PR-1 proteins in *A. hypogaea* have both these conserved features, but whether these CAPE peptides are actually derived from PR-1 proteins during symbiosis remains to be understood. It may be noted that, in the IRLC clade of legumes as well as in

the dalbergioid legume *Aeschynomene evenia*, another group of cysteine-rich peptides belonging to the NCR family is responsible for bacterial differentiation associated with morphological change of symbionts (Van de Velde et al. 2010). Intriguingly, morphological change of symbionts from rod to spherical shape is also noted in *A. hypogaea* (Kundu and DasGupta 2018). Whether the antimicrobial CAPE peptides are enrolled as symbiosis effectors in *A. hypogaea* in a manner similar to NCRs remains to be understood.

Bacterial recognition is NF-dependent in *A. hypogaea* and NF-independent for *Aeschynomene evenia*, though rhizobial invasion in both these legumes is mediated through epidermal cracks (Chandler et al. 1982; Giraud et al. 2007; Ibáñez and Fabra 2011). In accordance, the canonical NF receptor NFR1 was absent in *Aeschynomene evenia* (Chaintreuil et al. 2016) but was expressed in *A. hypogaea* (Figs. 4A and 5). Intriguingly, the expression of *AhLYR3*, ortholog of LCO-binding LysM receptor *MtLYR3* (Fliegmann et al. 2016), was higher in *A. hypogaea* as compared with canonical *NFR1/5*, which is unlike model legumes, in which the expression of canonical *NFR* genes were much higher than orthologs of *MtLYR3* (Fig. 4A). Since *LjLYS12*, an ortholog of *MtLYR3*, is important for recognition of pathogens (Kelly et al. 2017), it is tempting to suggest a possible involvement of *AhLYR3* in symbiont recognition in *A. hypogaea*. *PUB1*, a substrate of *LjNFR1*, has a role in determining partner specificity (Mbengue et al. 2010). The constitutive expression of *AhPUB1* in *A. hypogaea* suggests that it might have such a role during the choice of symbionts (Fig. 4C). Another step in bacterial recognition is through perception and recognition of surface molecules present on the bacterial cell wall, like EPS, LPSs, cyclic glucans, and capsular polysaccharides (Via et al. 2016). EPS-binding LysM receptor *MtEPR3* has a role in bacterial recognition followed by IT elongation (Kawaharada et al. 2015), and its ortholog in *A. hypogaea* showed significant expression with progress of symbiosis (Fig. 4A). This receptor in *Aeschynomene evenia* transcriptome (Chaintreuil et al. 2016), indicating that it may not be important during crack-entry but, rather, it might have a role in symbiont recognition during *A. hypogaea* nodulation. Thus, similar to IT-adapted model legumes, expression of LysM receptors may also be important for symbiont recognition during intercellular invasion of rhizobia in a crack-entry legume like *A. hypogaea*.

Most of the genes that are important for IT progression in model legumes are significantly expressed during the progress of symbiosis in the crack-entry legume *A. hypogaea* (Fig. 4E). An apparent exception is *MtRPG*, which has a role in polar growth of ITs in *M. truncatula* (Arrighi et al. 2008; Murray et al. 2011). Our results indicate the absence of a *RPG* homolog in *Arachis* sp., which is in accordance with a recent report that loss of *RPG* gene is indicated in the *A. ipaensis* genome (Griesmann et al. 2018). Since *RPG* is present in the nonlegumes, it is expected to be present in the legume ancestor that has been selectively lost in *A. hypogaea*. Whether this loss of *RPG* precedes or follows the evolution of nodulation in *A. hypogaea* remains to be understood. Along with the absence of *RPG*, the symbiotic ortholog of *MtFLOT4*, which is important for IT initiation and elongation (Haney and Long 2010), has very low expression in *A. hypogaea* (Fig. 4E) and was absent in *Aeschynomene evenia* (Chaintreuil et al. 2016). Other factors that are important for IT progression in model legumes are *MtFLOT2* (Haney and Long 2010), *MtVPY* (Murray et al. 2011), *LjCERBERUS* (Yano et al. 2009), and *LjARPC1* (Hossain et al. 2012). Symbiotic orthologs for all these genes are present in *A. hypogaea* and they demonstrated similar expression pattern during symbiosis in comparison with the IT-adapted model legumes, indicating the importance of their

additional roles during crack-entry nodulation (Fig. 4). For example, *LjCERBERUS* is functionally associated with nodule organogenesis and *LjARPC1* with cytoskeletal rearrangement during bacterial endocytosis (Hossain et al. 2012; Yano et al. 2009) and these additional functions of *CERBERUS* and *ARPC1* could be conserved in crack-entry legumes.

Though cytokinin signaling is important for nodule development in both crack-entry and IT-adapted legumes (Fabre et al. 2015; Gonzalez-Rizzo et al. 2006; Kundu and DasGupta 2018), PCA revealed a distinct expression pattern for several downstream factors in *A. hypogaea* as compared with model legumes. For example, a homolog of *MtRR1*, a type B response regulator (RR) has a distinct expression pattern in *A. hypogaea* and, unlike model legumes, its downstream TF, *MtbHLH476* (Ariel et al. 2012), was highly expressed with the progress of infection in *A. hypogaea* (Figs. 4F and 6). The symbiotic ortholog of *MtRR9* (Op den Camp et al. 2011), a type A RR, is absent in *A. hypogaea*, but the expression pattern of its non-symbiotic homolog indicates its possible recruitment during symbiosis (Figs. 4F and 5). Apart from the RRs, the cytokinin-responsive TFs like *LjNIN* (Tirichine et al. 2007) and *MtNSP2* (Plet et al. 2011), which are central regulators for both nodule organogenesis and rhizobial infection, do not show early expression during nodulation in *A. hypogaea* as compared with model legumes. Thus, unlike in model legumes, the first transcriptional changes that are associated with rhizobial recognition and generation of nodule primordia in *A. hypogaea* may not involve *NIN* and *NSP2*. Thus, cytokinin signaling pathway and its effectors appeared to be differentially adapted for regulating nodule development in crack-entry legumes.

Several contrasting features were also noted among the symbiotic genes involved in nodule differentiation. For example, symbiotic orthologs for *MtDNF2* and *MtSymCRK* were not detectable in both *A. hypogaea* (Figs. 4G and 5) and *Aeschynomene evenia* transcriptome (Chaintreuil et al. 2016). As both these factors are required for suppressing defense responses during bacteroid differentiation (Berrabah et al. 2014; Bourcy et al. 2013), the divergence of *DNF2* and *SymCRK* in crack-entry legumes further indicate that the developmental program for aeshynomonoid legumes is distinct from that of the IT-adapted legumes. *LjSUNERGOS* and *LjVAG1* codes for topoisomerases and are required for endoreduplication of plant cells (Suzaki et al. 2014; Yoon et al. 2014). Although the orthologs of *LjSUNERGOS* and *LjVAG1* were missing in *Aeschynomene evenia* transcriptome, they are expressed in *A. hypogaea* highlighting the differences in nodule differentiation program between the Nod-dependent (Fig. 4G) and -independent crack-entry legumes (Chaintreuil et al. 2016).

A previous report on *A. hypogaea* transcriptomics (Peng et al. 2017) as well as our transcriptome highlighted the up-regulation of several AP2 domain-containing ethylene-responsive TFs during nodulation. *LjERF1* downregulates the expression of defense gene *LjPR-10* during symbiosis (Asamizu et al. 2008), whereas, in *A. hypogaea*, high expression of *AhERF1* is associated with high expression of a divergent group of *PR-I* genes, indicating the differences in ethylene signaling between the IT-adapted legumes and *A. hypogaea* (Fig. 4D). Also, the expression of *AhEIN2*, the master regulator of ethylene signaling, was found to be constitutively high in *A. hypogaea* at all postinfection timepoints (Figs. 4H and 5). This predominance of ethylene signaling during *A. hypogaea* nodulation is consistent with the previous observations in *Sesbania rostrata*, in which, under submerged conditions, ethylene signaling inhibited intracellular infection via ITs while promoting intercellular infection via crack entry (D'Haeze et al. 2003; Guinel and Geil 2002).

In summary, the comparative transcriptomic analyses have shed light on the contrasting expression profiles of DEGs in crack-entry versus IT-adapted symbiosis. Genes like *VPY*, *FLOT2*, *CERBERUS*, and *ARPC1* that are primarily known for their role in IT elongation are significantly expressed in *A. hypogaea*, making it imperative to investigate their role in crack entry-adapted legumes. Similarly, the nonsymbiotic homolog of genes like *SymCRK*, *ROP6*, *RR9*, and *SEN1* are significantly expressed during symbiosis in *A. hypogaea*, and it is important to know whether these genes were co-opted for symbiosis. Further investigations are also required to understand how cytokinin and ethylene signaling may have been differentially adapted in crack-entry legumes. Finally, in the absence of early induction of *NIN* and *NSP2*, it is important to identify the factors that coordinate the rhizobial invasion and nodule primordia formation in crack-entry legumes. The highlighted molecular intricacies can be further investigated for better understanding of crack-entry nodule development. But overall, our investigations have revealed a considerable overlap in expression profiles of most of the symbiotic genes between a crack-entry legume *A. hypogaea* and the IT-adapted model legumes, suggesting a functional conservation of most of the factors that govern the process of nitrogen-fixing symbiosis irrespective of the mode of rhizobial entry in their host plants.

The supplementary files published online provide lists and expression patterns of i) all DEGs demonstrating major shifts in the transcriptome during symbiosis in *A. hypogaea* (additive to Figure 2), ii) nodulation-specific DEGs in *A. hypogaea* L. (Peng et al. 2017) and DEGs involved in plant-pathogen interaction and JA/SA pathway (additive to Figure 3), and iii) orthologs of symbiotic genes in *A. duranensis*, *A. ipaensis*, *M. truncatula*, and *L. japonicus* demonstrating their comparative expression profile (additive to Figures 4, 5, and 6). Orthology prediction of symbiotic genes in *A. hypogaea* and validation of their expression is also included.

MATERIALS AND METHODS

Plant materials and sample preparation.

Five different developmental stages of *A. hypogaea* total infected roots, primordia/nodules, and uninfected roots were used in this study (UI at 0 dpi and at 1, 4, 8, 12, and 21 dpi). *A. hypogaea* JL24 strain seeds (ICRISAT, Hyderabad, India) were surface-sterilized with sodium hypochlorite solution (3% active chlorine) containing 0.01% Tween 20 for 15 min and were rinsed five times with sterile deionized water; the sterile seeds were soaked in sterile deionized water for germination. Germinated seeds were then transferred in pots containing sterile vermiculite and soilrite at 25°C in a growth room for 7 days before inoculation with *Bradyrhizobium* sp. strain SEMIA 6144 (Sinharoy et al. 2009). *Bradyrhizobium* sp. strain SEMIA 6144, obtained from MIRCEN, Porto Alegre, Brazil and gifted by A. Fabra, Universidad Nacional de Rio cuarto, Cordoba, Argentina. Samples were harvested, cleaned, and frozen in liquid nitrogen. Frozen samples were stored at -80°C for RNA isolation.

Phenotypic analysis and microscopy.

Images of whole-mount nodulated roots were captured using a Leica stereo fluorescence microscope M205FA equipped with a Leica DFC310FX digital camera (Leica Microsystems). Detached nodules were embedded in Shandon cryomatrix (Thermo scientific) and were sliced into 30-µm thick sections with a rotary cryomicrotome CM1850 (Leica Microsystems). For confocal microscopy, sample preparation was done according to Haynes and associates (2004). Sections were stained with Calcofluor (Life Technologies), propidium iodide

(Life Technologies), and Syto9 (Life Technologies). Images were acquired with a Leica TCS SP5 II AOBs confocal laser scanning microscope (Leica Microsystems). For confocal and scanning electron microscopy, sample preparation was done according to (Kundu and DasGupta 2018). All digital micrographs were processed using Adobe Photoshop CS5.

Isolation of total RNA.

A total 100 mg of frozen plant root was ground in liquid nitrogen, and total RNA was isolated using Trizol reagent (Invitrogen). RNA degradation and contamination was detected on 1% agarose gels. RNA concentration was then measured using NanoDrop spectrophotometer (ThermoScientific). Additionally, RNA integrity was assessed using the Bioanalyzer 2100 system (Agilent Technologies). Finally, the samples with RNA integrity values above 8 were used for library construction.

Library construction and sequencing.

Eighteen RNA libraries were prepared, using an Illumina-TruSeq stranded mRNA sample preparation kit, by the MGX-Montpellier GenomiX core facility, Montpellier, France (MGX). The protocol first required the selection of polyadenylated RNAs on oligodT magnetic beads. Selected RNAs were chemically fragmented and the first-strand cDNA was synthesized in the presence of actinomycin D. The second-strand cDNA synthesis incorporated dUTP in place of dTTP in order to quench the second strand during amplification. A 3' ends adenylation was used to prevent fragments from ligating to one another during the adapter ligation process. The quantitative and qualitative validation of the library was performed by qPCR, Roche Light Cycler 480 and cluster generation, and primary hybridization are performed in the cBot with an Illumina cluster generation kit. The sample libraries were sequenced on an IlluminaHiSeq 2000, sequencing by synthesis technique performed by MGX, and 50-bp single-end reads for each library were generated (Fuller 1995).

Illumina reads mapping and assembly.

Quality control and assessment of raw Illumina reads in FASTQ format were done by FastQC software (version 0.11.5) to obtain per-base quality, GC content, and sequence-length distribution. Clean reads were obtained by removing the low-quality reads, adapters, and poly-N-containing reads, using Trimmomatic v0.36 software (Bolger et al. 2014). Clean Reads were simultaneously aligned to the two wild peanut diploid ancestors *A. duranensis* (AA) and *A. ipaensis* (BB) reference genomes, using TopHat2 version 2.0.13, which is a fast splice junction mapper for RNA-Seq reads, for which up to three mismatches per read were authorized (Bertoli et al. 2016; Trapnell et al. 2010). It aligns RNA-Seq reads, using the ultrahigh-throughput short read aligner Bowtie2 version 2.2.3, and then analyzes the mapping results to identify splice junctions between exons (Langmead et al. 2009). Alignment files were combined and analyzed into Trinity for genome-guided assembly (Grabherr et al. 2011).

The reference-based assembly was compared with its respective transcript files from annotated reference genomes, using BLAT (Kent 2002). An e-value cutoff of $1e^{-05}$ was used to determine a hit. The annotated hits were further analyzed in this study. Genome annotation files in generic feature format (GFF) were downloaded from PeanutBase (Dash et al. 2016). Estimation of the gene expression level of each annotated transcript was performed by StringTie v1.3.3, which takes a sorted sequence alignment map or binary file for each sample along with genome annotation files (Pertea et al. 2015). Counting reads in unannotated genes is performed with HTseq-count after converting the coordinates of these genes (chromosomal positions) to a GFF file. As the data were from a

strand-specific assay, the read was mapped to the opposite strand of the gene. The resulting gene transfer format, normalized gene locus expression level as FPKM, transcripts per million, and count files for each sample were further analyzed for fold change analysis in gene expression levels. RNA-seq read alignments have been performed using the Burrow-Wheeler Aligner (version 0.7.15-r1140) (Li and Durbin 2009) against *MtRPG*, *MtDNF2*, and singleton sequences..

Identification of DEGs and

functional GO and KEGG pathway analyses of the DEGs.

Before statistical analysis, genes with less than two values lower than one count per million were filtered out. EdgeR 3.6.7 package was used to identify the DEGs (Robinson et al. 2010). Data were normalized using the trimmed mean of M values method. Genes with an adjusted *P* value less than 5% (according to the FDR method, using Benjamini-Hochberg correction) and log2-fold change > 1 was called differentially expressed. Venn diagrams were generated using InteractiVenn (Heberle et al. 2015) and the hierarchical heatmap was generated using TM4-MeV (Howe et al. 2011) and the values from the Venn diagram. PCA was performed by web-tool ClustVis (Metsalu and Vilo 2015). Principal components of the variability of expression data were calculated using the PCA method R-Package. Detailed functional annotation and explanations of DEGs were extracted from the GO database (Ashburner et al. 2000) and GO functional classification analysis was done using software WEGO (Ye et al. 2006). The GO terms for DEGs in genome annotation were also retrieved from the GFF file downloaded from the PeanutBase website. To identify important and enriched pathways involved by the DEGs, the DEGs were assigned to KEGG pathways, using the GenomeNet database KAAS job request webserver (Kanehisa and Goto 2000) against *A. duranensis* and *A. ipaensis* gene datasets. Enriched KO and GO terms were obtained by a developed Python script that uses hypergeometric test and Bonferroni corrected *P* value < 0.05.

Identification of symbiotic orthologous gene in *A. hypogaea*.

Candidate symbiotic genes were identified in *A. duranensis*, *A. ipaensis*, *L. japonicus*, and *M. truncatula* using BLASTN searches with reported nucleotide sequence of genes from *L. japonicus* and *M. truncatula*. The homologous genes were searched in *A. duranensis* and *A. ipaensis* at PeanutBase, the *M. truncatula* *Mt4.0v1* genome was searched at MtGEA, and the *L. japonicus* v3.0 genome was searched at LjGEA. Initial searches were conducted with E value = e^{-5} . The results were manually validated for the presence of an orthologous gene in an open reading frame and were searched for orthologs using BLASTP. Orthology of the genes were validated by generating maximum likelihood (SymCRK, RR9, ROP6, SEN1, FLOT4, NOOT, and DNF2) and a neighbor-joining phylogenetic tree, using amino acid sequences and nucleotide sequences (*ENOD40*) in MEGA 6.0, using 1,000 bootstrap replications (Tamura et al. 2013). Syntenic analysis of *RPG* and *DNF2* between the legumes was performed using the Legume Information System genome context viewer.

qRT-PCR validation.

Total RNA (500 ng) was reverse-transcribed by using SuperScriptIII RT (Life Technologies) and oligo(dT). RNA quantity from each sample in each biological replicate was standardized prior to first-strand cDNA synthesis. qRT-PCR was performed using Power SYBR green PCR master mix (Applied Biosystems), with primers as designed using software Oligoanalyzer (IntegratedDNA Technology). Calculations were done using the $\Delta\Delta$ cycle threshold method, using *AhActin* as the

endogenous control. The reactions were run in the Applied Biosystems 7500 Fast HT platform using the following protocol: 1 cycle at 50°C for 2 min, 1 cycle at 95°C for 5 min, 40 cycles at 95°C for 30 s, 54°C for 30 s, 72°C for 30 s, followed by melt curve analysis for 1 cycle at 95°C for 1 min, 55°C for 30 s, and 95°C for 30 s. A negative control without cDNA template was checked for each primer combination that was designed using OligoAnalyzer 3.1. Results were expressed as means $\pm$ standard error of the number of experiments.

Data availability.

The raw FASTQ files for the 18 libraries were deposited in the Gene Expression Omnibus database of the National Center for Biotechnology Information under accession number GSE98997.

LITERATURE CITED

- Ané, J.-M., Kiss, G. B., Riely, B. K., Penmetsa, R. V., Oldroyd, G. E., Ayax, C., Lévy, J., Debelle, F., Baek, J.-M., Kalo, P., Rosenberg, C., Roe, B. A., Long, S. R., Dénarié, J., and Cook, D. R. 2004. *Medicago truncatula* DMI1 required for bacterial and fungal symbioses in legumes. *Science* 303:1364-1367.
- 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.
- Arrighi, J. F., Barre, A., Ben Amor, B., Bersoult, A., Soriano, L. C., Mirabella, R., de Carvalho-Niebel, F., Journet, E. P., Ghérardi, M., Hugué, T., Geurts, R., Dénarié, J., Rougé, P., and Gough, C. 2006. The *Medicago truncatula* lysin [corrected] motif-receptor-like kinase gene family includes *NFP* and new nodule-expressed genes. *Plant Physiol.* 142:265-279.
- Arrighi, J. F., Godfroy, O., de Billy, F., Saurat, O., Jauneau, A., and Gough, C. 2008. The *RPG* gene of *Medicago truncatula* controls *Rhizobium*-directed polar growth during infection. *Proc. Natl. Acad. Sci. U.S.A.* 105:9817-9822.
- Asamizu, E., Shimoda, Y., Kouchi, H., Tabata, S., and Sato, S. 2008. A positive regulatory role for *LjERF1* in the nodulation process is revealed by systematic analysis of nodule-associated transcription factors of *Lotus japonicus*. *Plant Physiol.* 147:2030-2040.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. 2000. Gene ontology: Tool for the unification of biology. *Nat. Genet.* 25:25-29.
- Benedito, V. A., Torres-Jerez, I., Murray, J. D., Andriankaja, A., Allen, S., Kakar, K., Wandrey, M., Verdier, J., Zuber, H., Ott, T., Moreau, S., Niebel, A., Frickey, T., Weiller, G., He, J., Dai, X., Zhao, P. X., Tang, Y., and Udvardi, M. K. 2008. A gene expression atlas of the model legume *Medicago truncatula*. *Plant J.* 55:504-513.
- Berrabah, F., Bourcy, M., Eschstruth, A., Cayrel, A., Guefrachi, I., Mergaert, P., Wen, J., Jean, V., Mysore, K. S., Gourion, B., and Ratet, P. 2014. A nonRD receptor-like kinase prevents nodule early senescence and defense-like reactions during symbiosis. *New Phytol.* 203:1305-1314.
- Bertioli, D. J., Cannon, S. B., Froeniche, L., Huang, G., Farmer, A. D., Cannon, E. K., 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., 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.
- Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120.
- Booger, F. C., and Rossum, D. 1997. Nodulation of groundnut by *Bradyrhizobium*: A simple infection process by crack entry. *FEMS Microbiol. Rev.* 21:5-27.
- Boscari, A., Del Giudice, J., Ferrarini, A., Venturini, L., Zaffini, A. L., Delledonne, M., and Puppo, A. 2013. Expression dynamics of the *Medicago truncatula* transcriptome during the symbiotic interaction with *Sinorhizobium meliloti*: Which role for nitric oxide? *Plant Physiol.* 161:425-439.
- 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.
- Breakspear, A., Liu, C., Roy, S., Stacey, N., Rogers, C., Trick, M., Morieri, G., Mysore, K. S., Wen, J., Oldroyd, G. E. D., Downie, J. A., and Murray, J. D. 2014. The root hair "infectome" of *Medicago truncatula* uncovers changes in cell cycle genes and reveals a requirement for Auxin signaling in rhizobial infection. *Plant Cell* 26:4680-4701.
- 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.
- Chaintreuil, C., Rivallan, R., Bertioli, D. J., Klopp, C., Gouzy, J., Courtois, B., Leleux, P., Martin, G., Rami, J.-F., Gully, D., Parrinello, H., Séverac, D., Patrel, D., Fardoux, J., Ribiére, W., Boursot, M., Cartiaux, F., Czernic, P., Ratet, P., Mournet, P., Giraud, E., and Arrighi, J.-F. 2016. A gene-based map of the Nod factor-independent *Aeschynomene evenia* sheds new light on the evolution of nodulation and legume genomes. *DNA Res.* 23:365-376.
- Chandler, M. R., Date, R., and Roughley, R. 1982. Infection and root-nodule development in *Stylosanthes* species by *Rhizobium*. *J. Exp. Bot.* 33:47-57.
- Charpentier, M., Sun, J., Vaz Martins, T., Radhakrishnan, G. V., Findlay, K., Soumpourou, E., Thouin, J., Véry, A.-A., Sanders, D., Morris, R. J., and Oldroyd, G. E. 2016. Nuclear-localized cyclic nucleotide-gated channels mediate symbiotic calcium oscillations. *Science* 352:1102-1105.
- D'Haese, W., De Rycke, R., Mathis, R., Goormachtig, S., Pagnotta, S., Verplancke, C., Capoen, W., and Holsters, M. 2003. Reactive oxygen species and ethylene play a positive role in lateral root base nodulation of a semiaquatic legume. *Proc. Natl. Acad. Sci. U.S.A.* 100:11789-11794.
- Dash, S., Cannon, E., Kalberer, S., Farmer, A., Cannon, S., Wilson, R., and Stalker, T. 2016. PeanutBase and other bioinformatic resources for peanut. Pages 241-252 in: *Peanuts: Genetics, Processing, & Utilization*. AOCS Press, Urbana, IL, U.S.A.
- 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.
- Fabre, S., Gully, D., Poitout, A., Patrel, D., Arrighi, J. F., Giraud, E., Czernic, P., and Cartiaux, F. 2015. Nod factor-independent nodulation in *Aeschynomene evenia* required the common plant-microbe symbiotic toolkit. *Plant Physiol.* 169:2654-2664.
- Fliegmann, J., Jauneau, A., Pichereaux, C., Rosenberg, C., Gascioli, V., Timmers, A. C., Burlet-Schiltz, O., Cullimore, J., and Bono, J. J. 2016. LYR3, a high-affinity LCO-binding protein of *Medicago truncatula*, interacts with LYK3, a key symbiotic receptor. *FEBS Lett.* 590:1477-1487.
- Fuller, C. W. July 1995. Cycle sequencing with non-thermostable DNA polymerases. U.S. patent 5432065A.
- Gage, D. J. 2004. Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiol. Mol. Biol. Rev.* 68:280-300.
- Geurts, R., and Bisseling, T. 2002. *Rhizobium* nod factor perception and signalling. *Plant Cell* 14 (Suppl):S239-S249.
- Giraud, E., Moulin, L., Vallenet, D., Barbe, V., Cytryn, E., Avarre, J. C., Jaubert, M., Simon, D., Cartiaux, 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 Sadovsky, M. 2007. Legumes symbioses: Absence of Nod genes in photosynthetic bradyrhizobia. *Science* 316:1307-1312.
- Gonzalez-Rizzo, S., Crespi, M., and Frugier, F. 2006. The *Medicago truncatula* CRE1 cytokinin receptor regulates lateral root development and early symbiotic interaction with *Sinorhizobium meliloti*. *Plant Cell* 18:2680-2693.
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29:644-652.
- Griesmann, M., Chang, Y., Liu, X., Song, Y., Haber, G., Crook, M.B., Billault-Penneteau, B., Lauressergues, D., Keller, J., and Imanishi, L. 2018. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* 361:eaat1743.
- Guinel, F. C., and Geil, R. 2002. A model for the development of the rhizobial and arbuscular mycorrhizal symbioses in legumes and its use to understand the roles of ethylene in the establishment of these two symbioses. *Can. J. Bot.* 80:695-720.

- Haney, C. H., and Long, S. R. 2010. Plant flotillins are required for infection by nitrogen-fixing bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 107: 478-483.
- Haynes, J. G., Czymbek, K. J., Carlson, C. A., Veereshlingam, H., Dickstein, R., and Sherrier, D. J. 2004. Rapid analysis of legume root nodule development using confocal microscopy. *New Phytol.* 163: 661-668.
- Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P., and Minghim, R. 2015. InteractiVenn: A web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* 16:169.
- Hossain, M. S., Liao, J., James, E. K., Sato, S., Tabata, S., Jurkiewicz, A., Madsen, L. H., Stougaard, J., Ross, L., and Szczyglowski, K. 2012. *Lotus japonicus* *ARPC1* is required for rhizobial infection. *Plant Physiol.* 160: 917-928.
- Howe, E. A., Sinha, R., Schlauch, D., and Quackenbush, J. 2011. RNA-Seq analysis in MeV. *Bioinformatics* 27:3209-3210.
- 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 Stuttgart* 13:794-800.
- Imaizumi-Anraku, H., Takeda, N., Charpentier, M., Perry, J., Miwa, H., Umehara, Y., Kouchi, H., Murakami, Y., Mulder, L., Vickers, K., Pike, J., Downie, J. A., Wang, T., Sato, S., Asamizu, E., Tabata, S., Yoshikawa, M., Murooka, Y., Wu, G. J., Kawaguchi, M., Kawasaki, S., Parniske, M., and Hayashi, M. 2005. Plastid proteins crucial for symbiotic fungal and bacterial entry into plant roots. *Nature* 433:527-531.
- Jardinaud, M.-F., Boivin, S., Rodde, N., Catrice, O., Kisiala, A., Lepage, A., Moreau, S., Roux, B., Cottret, L., Sallet, E., Brault, M., Emery, R. J., Gouzy, J., Frugier, F., and Gamas, P. 2016. A laser dissection-RNAseq analysis highlights the activation of cytokinin pathways by nod factors in the *Medicago truncatula* root epidermis. *Plant Physiol.* 171:2256-2276.
- 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.
- Kanamori, N., Madsen, L. H., Radutoiu, S., Frantescu, M., Quistgaard, E. M., Miwa, H., Downie, J. A., James, E. K., Felle, H. H., Haaning, L. L., Jensen, T. H., Sato, S., Nakamura, Y., Tabata, S., Sandal, N., and Stougaard, J. 2006. A nucleoporin is required for induction of Ca²⁺ spiking in legume nodule development and essential for rhizobial and fungal symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 103:359-364.
- Kanehisa, M., and Goto, S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28:27-30.
- Kant, C., Pradhan, S., and Bhatia, S. 2016. Dissecting the root nodule transcriptome of chickpea (*Cicer arietinum* L.). *PLoS One* 11:e0157908.
- Kawaharada, Y., Kelly, S., Nielsen, M. W., Hjuler, C. T., Gysel, K., Muszyński, A., Carlson, R. W., Thygesen, M. B., Sandal, N., Asmussen, M. H., Vinther, M., Andersen, S. U., Krusell, L., Thirup, S., Jensen, K. J., Ronson, C. W., Blaise, M., Radutoiu, S., and Stougaard, J. 2015. Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* 523:308-312.
- 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.
- Kelly, S., Radutoiu, S., and Stougaard, J. 2017. Legume LysM receptors mediate symbiotic and pathogenic signalling. *Curr. Opin. Plant Biol.* 39: 152-158.
- Kent, W. J. 2002. BLAT—The BLAST-like alignment tool. *Genome Res.* 12:656-664.
- Kistner, C., and Parniske, M. 2002. Evolution of signal transduction in intracellular symbiosis. *Trends Plant Sci.* 7:511-518.
- Kundu, A., and DasGupta, M. 2018. Silencing of putative cytokinin receptor histidine kinase1 inhibits both inception and differentiation of root nodules in *Arachis hypogaea*. *Mol. Plant-Microbe Interact.* 31:187-199.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25.
- Lévy, J., Bres, C., Geurts, R., Chalhoub, B., Kulikova, O., Duc, G., Journet, E.-P., Ané, J.-M., Lauber, E., Bisseling, T., Dénarié, J., Rosenberg, C., and Debelle, F. 2004. A putative Ca²⁺ and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. *Science* 303:1361-1364.
- Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-1760.
- Limpens, E., Moling, S., Hooiveld, G., Pereira, P. A., Bisseling, T., Becker, J. D., and Küster, H. 2014. Correction: Cell- and tissue-specific transcriptome analyses of *Medicago truncatula* root nodules. *PLoS One* 9.
- Madsen, E. B., Madsen, L. H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., Sato, S., Kaneko, T., Tabata, S., Sandal, N., and Stougaard, J. 2003. A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. *Nature* 425:637-640.
- 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.
- Messinese, E., Mun, J.-H., Yeun, L. H., Jayaraman, D., Rougé, P., Barre, A., Loughnon, G., Schornack, S., Bono, J.-J., Cook, D. R., and Ané, J. M. 2007. A novel nuclear protein interacts with the symbiotic DMI3 calcium- and calmodulin-dependent protein kinase of *Medicago truncatula*. *Mol. Plant-Microbe Interact.* 20:912-921.
- Metsalu, T., and Vilo, J. 2015. ClustVis: A web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic Acids Res.* 43:W566-W570.
- Middleton, P. H., Jakab, J., Penmetsa, R. V., Starker, C. G., Doll, J., Kaló, P., Prabhu, R., Marsh, J. F., Mitra, R. M., Kereszt, A., Dudas, B., VandenBosch, K., Long, S. R., Cook, D. R., Kiss, G. B., and Oldroyd, G. E. 2007. An ERF transcription factor in *Medicago truncatula* that is essential for Nod factor signal transduction. *Plant Cell* 19:1221-1234.
- Moeder, W., Urquhart, W., Ung, H., and Yoshioka, K. 2011. The role of cyclic nucleotide-gated ion channels in plant immunity. *Mol. Plant* 4:442-452.
- Moreau, S., Verdenaud, M., Ott, T., Letort, S., de Billy, F., Niebel, A., Gouzy, J., de Carvalho-Niebel, F., and Gamas, P. 2011. Transcription reprogramming during root nodule development in *Medicago truncatula*. *PLoS One* 6:e16463.
- Morgante, C. V., Guimarães, P. M., Martins, A. C., Araújo, A. C., Leal-Bertioli, S. C., Bertioli, D. J., and Brasileiro, A. C. 2011. Reference genes for quantitative reverse transcription-polymerase chain reaction expression studies in wild and cultivated peanut. *BMC Res. Notes* 4:339.
- Murray, J. D., Muni, R. R. D., Torres-Jerez, I., Tang, Y., Allen, S., Andriankaja, M., Li, G., Laxmi, A., Cheng, X., Wen, J., Vaughan, D., Schultze, M., Sun, J., Charpentier, M., Oldroyd, G., Tadege, M., Ratet, P., Mysore, K. S., Chen, R., and Udvardi, M. K. 2011. Vapyrin, a gene essential for intracellular progression of arbuscular mycorrhizal symbiosis, is also essential for infection by rhizobia in the nodule symbiosis of *Medicago truncatula*. *Plant J.* 65:244-252.
- Oldroyd, G. E., and Downie, J. A. 2004. Calcium, kinases and nodulation signalling in legumes. *Nat. Rev. Mol. Cell Biol.* 5:566-576.
- Oldroyd, G. E., and Downie, J. A. 2006. Nuclear calcium changes at the core of symbiosis signalling. *Curr. Opin. Plant Biol.* 9:351-357.
- Op den Camp, R. H., De Mita, S., Lillo, A., Cao, Q., Limpens, E., Bisseling, T., and Geurts, R. 2011. A phylogenetic strategy based on a legume-specific whole genome duplication yields symbiotic cytokinin type-A response regulators. *Plant Physiol.* 157:2013-2022.
- 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.
- Penmetsa, R. V., Uribe, P., Anderson, J., Lichtenzeig, J., Gish, J.-C., Nam, Y. W., Engstrom, E., Xu, K., Sckisel, G., Pereira, M., Baek, J. M., Lopez-Meyer, M., Long, S. R., Harrison, M. J., Singh, K. B., Kiss, G. B., and Cook, D. R. 2008. The *Medicago truncatula* ortholog of Arabidopsis EIN2, *sickle*, is a negative regulator of symbiotic and pathogenic microbial associations. *Plant J.* 55:580-595.
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33:290-295.
- Plet, J., Wasson, A., Ariel, F., Le Signor, C., Baker, D., Mathiesius, U., Crespi, M., and Frugier, F. 2011. MtCRE1-dependent cytokinin signalling integrates bacterial and plant cues to coordinate symbiotic nodule organogenesis in *Medicago truncatula*. *Plant J.* 65:622-633.
- Radutoiu, S., Madsen, L. H., Madsen, E. B., Felle, H. H., Umehara, Y., Grönlund, M., Sato, S., Nakamura, Y., Tabata, S., Sandal, N., and Stougaard, J. 2003. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* 425:585-592.
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. 2010. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139-140.
- Roux, B., Rodde, N., Jardinaud, M. F., Timmers, T., Sauviac, L., Cottret, L., Carrère, S., Sallet, E., Courcelle, E., Moreau, S., Debelle, F., Capela, D., de Carvalho-Niebel, F., Gouzy, J., Bruand, C., and Gamas, P. 2014. 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.
- Saito, K., Yoshikawa, M., Yano, K., Miwa, H., Uchida, H., Asamizu, E., Sato, S., Tabata, S., Imaizumi-Anraku, H., Umehara, Y., Kouchi, H.,

- Murooka, Y., Szczygłowski, K., Downie, J. A., Parniske, M., Hayashi, M., and Kawaguchi, M. 2007. NUCLEOPORIN85 is required for calcium spiking, fungal and bacterial symbioses, and seed production in *Lotus japonicus*. *Plant Cell* 19:610-624.
- Schauser, L., Roussis, A., Stiller, J., and Stougaard, J. 1999. A plant regulator controlling development of symbiotic root nodules. *Nature* 402:191-195.
- Singh, S., Katzer, K., Lambert, J., Cerri, M., and Parniske, M. 2014. CYCLOPS, a DNA-binding transcriptional activator, orchestrates symbiotic root nodule development. *Cell Host Microbe* 15:139-152.
- Sinharoy, S., Saha, S., Chaudhury, S. R., and Dasgupta, M. 2009. Transformed hairy roots of *Arachis hypogaea*: A tool for studying root nodule symbiosis in a non-infection thread legume of the Aeschynomeneae tribe. *Mol. Plant-Microbe Interact.* 22:132-142.
- Smit, P., Limpens, E., Geurts, E., Fedorova, E., Dolgikh, E., Gough, C., and Bisseling, T. 2007. Medicago LYK3, an entry receptor in rhizobial nodulation factor signaling. *Plant Physiol.* 145:183-191.
- Smit, P., Raedts, J., Portyanko, V., Debelle, 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.
- Solano, R., Stepanova, A., Chao, Q., and Ecker, J. R. 1998. Nuclear events in ethylene signaling: A transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. *Genes Dev.* 12:3703-3714.
- 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., Szczygłowski, K., and Parniske, M. 2002. A plant receptor-like kinase required for both bacterial and fungal symbiosis. *Nature* 417:959-962.
- Suzaki, T., Ito, M., Yoro, E., Sato, S., Hirakawa, H., Takeda, N., and Kawaguchi, M. 2014. Endoreduplication-mediated initiation of symbiotic organ development in *Lotus japonicus*. *Development* 141:2441-2445.
- Takanashi, K., Takahashi, H., Sakurai, N., Sugiyama, A., Suzuki, H., Shibata, D., Nakazono, M., and Yazaki, K. 2012. Tissue-specific transcriptome analysis in nodules of *Lotus japonicus*. *Mol. Plant-Microbe Interact.* 25:869-876.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30:2725-2729.
- Tirichine, L., Imaizumi-Anraku, H., Yoshida, S., Murakami, Y., Madsen, L. H., Miwa, H., Nakagawa, T., Sandal, N., Albrechtsen, A. S., Kawaguchi, M., Downie, A., Sato, S., Tabata, S., Kouchi, H., Parniske, M., Kawasaki, S., and Stougaard, J. 2006. Dereglulation of a Ca^{2+} /calmodulin-dependent kinase leads to spontaneous nodule development. *Nature* 441:1153-1156.
- Tirichine, L., Sandal, N., Madsen, L. H., Radutoiu, S., Albrechtsen, A. S., Sato, S., Asamizu, E., Tabata, S., and Stougaard, J. 2007. A gain-of-function mutation in a cytokinin receptor triggers spontaneous root nodule organogenesis. *Science* 315:104-107.
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28:511-515.
- 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.
- Verdier, J., Torres-Jerez, I., Wang, M., Andriankaja, A., Allen, S. N., He, J., Tang, Y., Murray, J. D., and Udvardi, M. K. 2013. Establishment of the *Lotus japonicus* Gene Expression Atlas (LjGEA) and its use to explore legume seed maturation. *Plant J.* 74:351-362.
- Via, V. D., Zanetti, M. E., and Blanco, F. 2016. How legumes recognize rhizobia. *Plant Signal. Behav.* 11:e1120396.
- Yano, K., Shibata, S., Chen, W. L., Sato, S., Kaneko, T., Jurkiewicz, A., Sandal, N., Banba, M., Imaizumi-Anraku, H., Kojima, T., Ohtomo, R., Szczygłowski, K., Stougaard, J., Tabata, S., Hayashi, M., Kouchi, H., and Umehara, Y. 2009. CERBERUS, a novel U-box protein containing WD-40 repeats, is required for formation of the infection thread and nodule development in the legume-*Rhizobium* symbiosis. *Plant J.* 60:168-180.
- Yano, K., Yoshida, S., Müller, J., Singh, S., Banba, M., Vickers, K., Markmann, K., White, C., Schuller, B., Sato, S., Asamizu, E., Tabata, S., Murooka, Y., Perry, J., Wang, T. L., Kawaguchi, M., Imaizumi-Anraku, H., Hayashi, M., and Parniske, M. 2008. CYCLOPS, a mediator of symbiotic intracellular accommodation. *Proc. Natl. Acad. Sci. U.S.A.* 105:20540-20545.
- Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Bolund, L., and Wang, J. 2006. WEGO: A web tool for plotting GO annotations. *Nucleic Acids Res.* 34:W293-W297.
- Yoon, H. J., Hossain, M. S., Held, M., Hou, H., Kehl, M., Tromas, A., Sato, S., Tabata, S., Andersen, S. U., Stougaard, J., Ross, L., and Szczygłowski, K. 2014. *Lotus japonicus* SUNERGOS1 encodes a predicted subunit A of a DNA topoisomerase VI that is required for nodule differentiation and accommodation of rhizobial infection. *Plant J.* 78:811-821.
- Yuan, S. L., Li, R., Chen, H. F., Zhang, C. J., Chen, L. M., Hao, Q. N., Chen, S. L., Shan, Z. H., Yang, Z. L., Zhang, X. J., Qiu, Z., and Zhou, X. A. 2017. RNA-seq analysis of nodule development at five different developmental stages of soybean (*Glycine max*) inoculated with *Bradyrhizobium japonicum* strain 113-2. *Sci. Rep.* 7:42248.

AUTHOR-RECOMMENDED INTERNET RESOURCES

ClustVis: <https://biit.cs.ut.ee/clustvis>
 Gene Ontology Consortium database: <http://www.geneontology.org>
 InteractVenn: <http://www.interactivenn.net>
 GenomeNet database KAAS job request webserver: http://www.genome.jp/kaas-bin/kaas_main
 LIS-Legume Information System genome context viewer: https://legumeinfo.org/lis_context_viewer
Lotus japonicus gene expression atlas (LjGEA): <https://ljgea.noble.org/v2>
Medicago truncatula gene expression atlas (MtGEA): <https://mtgea.noble.org/v2>
 MGX-Montpellier GenomiX: <https://www.mgx.cnrs.fr>
 Integrated DNA Technologies OligoAnalyzer 3.1: <https://www.idtdna.com/calc/analyzer>
 PeanutBase database: <https://peanutbase.org>
 TM4MeV application: <http://mev.tm4.org>
 WEGO annotation plotting tool: <http://wego.genomics.org.cn>