Title: Transcriptomic analyses with the progress of symbiosis in 'crack-entry' legume *Arachis hypogaea* highlight its contrast with 'Infection thread' adapted legumes

Authors: 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*]

Author information

[1] Department of Biochemistry, University of Calcutta, Kolkata 700019, India

[2]LSTM, Univ. Montpellier, CIRAD, INRA, IRD, SupAgro, Montpellier, France

[3]Montpellier GenomiX (MGX), c/o Institut de Génomique Fonctionnelle, 141 rue de la cardonille, 34094 Montpellier Cedex 05, France

[†]Present address: 112, INSERM U981, Bâtiment Médecine Moléculaire (B2M), Gustave Roussy, 114, rue Edouard Vaillant, 94805 Villejuif Cedex-France

[‡] Equal contribution

[*]**Author for correspondence:**
Maitrayee DasGupta
Address: Department of Biochemistry, University of Calcutta,
35, Ballygunge Circular Road,
Kolkata-700019,
West Bengal, India
Phone No.: +91-33-2475-4680; Fax: +91-33-2476-4419
Email ID: maitrayee_d@hotmail.com

ABSTRACT

In root-nodule symbiosis, rhizobial invasion and nodule organogenesis are host controlled. In most legumes, rhizobia enter through infection-threads and nodule primordium in the cortex is induced from 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* by profiling the transcriptome at 1dpi: invasion; 4dpi: nodule primordia; 8dpi: spread of infection in nodule-like structure; 12dpi: immature nodules containing rod-shaped rhizobia; and 21dpi: mature nodules with spherical symbiosomes. Differentially expressed genes show clear transcriptional shifts at these stages. Expressions of putative orthologues of symbiotic genes in 'crack-entry' legume *A.hypogaea* were compared with their expression in model legumes where rhizobia invade through infection-threads. The notable contrasting features were (i) absence of early induction of *NIN* and *NSP2*, (ii) insignificant expression of *VPY* and (iii) significantly high expression of *ERF1*, *bHLH476*, *EIN2* and divergent *PR-1* genes that produce CAPE peptides. Additionally, homologues for *RPG*, *SymCRK* and *DNF2* were absent in *A. hypogaea* genome and for *FLOT4*, *ROP6*, *RR9*, *NOOT*, and *SEN1*, their symbiotic orthologues were not detectable. A molecular framework that may guide symbiosis in *A. hypogaea* is proposed.

INTRODUCTION

Nitrogen fixing root-nodule symbiosis (RNS) allows plants to house bacterial diazotrophs in an intracellular manner (Kistner and Parniske, 2002). RNS occurs in two major forms: legume–rhizobia (Fabaceae) and actinorhizal symbiosis (Fagaceae, Rosaceae, Cucurbitaceae)(Pawlowski and Bisseling, 1996). The Leguminosae (Fabaceae) is the third largest family of flowering plants and are agriculturally and economically important, being second only to the Poaceae (e.g. cereals). This economic importance of the Leguminosae is mainly due to RNS that allows the plant to grow well and produce protein rich seeds in the absence of nitrogen fertilizer in soils.

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 where the nodule primordia are induced from a distance (Sprent and James, 2007). Invasion through IT is adapted mostly by temperate legumes e.g. *Vicia sp., Trifolium sp., Pisum sp.*.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' where the rhizobia enter through natural cracks at the lateral root base in an intercellular manner. This is a characteristic feature of some subtropical legumes (e.g. *Arachis sp., Aeschynomene sp., Stylosanthes sp.*) belonging to dalbergoids/genistoids and accounts for approximately 25% of all legume genera (Gage, 2004; Giraud et al., 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 (Boogerd and Rossum, 1997; Fabre et al., 2015).

Investigations on model legumes have unravelled the molecular basis of RNS. The host responses are initiated by Nod-factor (NF) receptors *Lj*NFR1/*Mt*LYK3 and *Lj*NFR5/*Mt*NFP (Madsen et al., 2003; Radutoiu et al., 2003; Arrighi, 2006; Smit et al., 2007). Another NF induced receptor*Lj*EPR3 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 *Lj*SYMRK/*Mt*DMI2(Endre et al., 2002; Stracke et al., 2002), the predicted ion-channel proteins *Lj*CASTOR and *Lj*POLLUX/*Mt*DMI1(Ané et al., 2004; Imaizumi-Anraku et al., 2005), the nucleoporins *Lj*NUP85 and *Lj*NUP133(Kanamori et al., 2006; Saito et al., 2007), the Ca2+/calmodulin-dependent protein

kinase *Lj*CCaMK/*Mt*DMI3(Lévy et al., 2004; Tirichine et al., 2006), and the transcription factor *Lj*CYCLOPS/*Mt*IPD3(Messinese et al., 2007; Yano et al., 2008). Nodulation-specific transcription factors (TFs), such as *Mt*NSP1/*Lj*NSP1, *Mt*NSP2/*Lj*NSP2, *Mt*ERF1 and *Mt*NIN/*Lj*NIN function downstream of the 'SYM pathway' and are involved in transcriptional reprogramming for initiation of RNS (Schauser et al., 1999a; Kaló et al., 2005; Smit et al., 2005; Middleton et al., 2007). Very limited information is available for crack-entry legumes from Dalbergioid/Genistoid clade which are basal in their divergence within the Papilionoideae even though they contain important crop legumes such as *Lupinus angustifolius* and *Arachis hypogaea*. Transcriptome analysis in legumes has been a valuable resource for understanding symbiosis-related genes in *M. truncatula, L. japonicus, Glycine max, and Cicer arietinum.* An earlier report have listed several differentially expressed genes (DEGs) at an early stage of symbiosis in *A. hypogaea* (Peng et al., 2017). The conservativeness among DEGs identified in such studies has implied common genetic mechanisms of RNS in legume species. Herein we report the transcriptome dynamics with the onset and advancement of symbiosis in *A. hypogaea* using uninfected roots (UI) as a reference. The transcription profile of the putative orthologues of symbiotic genes in crack-entry legume *A. hypogaea* is compared and contrasted with the corresponding expression profiles in *M. truncatula* and *L. japonicus* that undertake root hair mediated symbiosis.

RESULTS

Progress of symbiosis in *Arachis hypogaea*

Within three weeks after infection with *Bradyrhizobium sp.* 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 post infection (1dpi) rhizobia was found to be adhered to these root hairs (Fig. 1A-C).Within 4dpi, 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 intracellularised rhizobia within the infection pocket was undifferentiated and rod-shaped (Fig.

1F). The infection pockets observed at 4dpi act as infection zone (IZ) founder cells and it is their uniform division and differentiation that give rise to the distinct aeschynomenoid type IZ in mature nodules. There has not been a single case where uninfected primordium was noted, which is in accordance with the proposition of infection preceding development of aeschynomenoid nodules (Fabre et al., 2015). By 8dpi there was visible nodule-like structure at the lateral-root base (Fig. 1G). Ultrastructure analysis revealed that by 8dpi, 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-I). By 12dpi there were white spherical nodules (Fig. 1J). At this stage the tissue organization turned aeschynomenoid where there were no uninfected cells in the infection zone (IZ) and the endocytosed rhizobia remained undifferentiated and rod-shaped (Fig. 1K-L). At 21dpi the nodules were mature and functional where the rhizobia differentiated within the plant derived peribacteroid membranes to develop spherical symbiosomes (Fig. 1M-O).

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

Ultrastructural analysis revealed 5 distinct stages during the progress of symbiosis in *A. hypogaea*: 1dpi: recognition and invasion; 4dpi: primordia formation; 8dpi: nodule-like structure; 12dpi: immature nodules with rod-shaped rhizobia; and 21dpi: mature nodules with spherical symbiosomes. To probe into the expression of genes associated with the progress of symbiosis, RNA was extracted from these stages along with UI roots. RNA-seq was done in triplicate for these six stages using Illumina single-end sequencing technology (IlluminaHiseq 2000 SR50). The genomic data from *Arachis duranensis* (AA) and *Arachis ipaensis* (BB) that 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 50bp (~71.5Gb) 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 where 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 genome about 66% reads aligned to a gene exon in an unambiguous way, whereas the rest 33% reads aligned outside exon.

The expression level of each assembled transcript sequence was measured through FPKM (Fragments per kilo-base per million reads) values. The DEGs in the 5 different stages of symbiosis were evaluated by the significance of differences in their expression with respect to UI roots using false discovery rate (FDR) < 0.05, P-value <0.05 and fold change |log2 ratio| ≥ 1 (Supplementary Table 1). Comparison between upregulated and downregulated DEGs at different stages is shown in a Venn-Diagram in (Fig.2A). A total of 2745 genes were up-regulated (↑1296:AA, ;↑1449:BB) and a total of 20415 genes are down-regulated (↓9709:AA;↓10706:BB) during symbiosis of which 59 genes (33:AA;26:BB) were upregulated and 2095 genes (1056:AA;1039:BB) were downregulated in all the 5 stages of symbiosis. From the Venn-diagram we identified those genes that were first upreregulated or downregulated at a particular stage though their subsequent regulations could be different. The number of such genes upregulated or downregulated at each stage from AA or BB genome is shown in Supplementary Table 2. 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). The major expression profiles are shown in expanded heatmaps and line graphs in (Supplementary fig.1). Hierarchical clustering as well as PCA analysis (Supplementary fig.1C) of the transcriptome indicated 3 distinct expression waves. Cluster 1 consists of 1dpi-4dpi transcripts where rhizobial invasion and primordia formation occurs. Cluster 2 consists of 8dpi-12dpi transcripts where the primordia structurally develop into a nodule and cluster 3 consists of the 21dpi transcripts where nodule matures to its functional form (Fig.2B; Supplimentary fig.1).

Functional analysis of DEGs

GO and KEGG terms that are significantly enriched in our DEGs are indicated in Supplementary fig. 2. Among the 1248 enriched GO terms there was a major representation of defense response genes. 470 and 31 such defense related GO terms were enriched in downregulated and upregulated DEGs respectively (Supplementary Table 1). Accordingly, KEGG analysis of plant-pathogen interaction pathways show that most genes involved in pattern-triggered immunity (PTI) was notably down-regulated (Fig. 3A-B; Supplementary Table 3). The FLS2 mediated MAPK pathway however remained active along with a subset of *CNGC*s and genes encoding Rboh proteins. A subset of genes involved in the effector triggered immunity (ETI) also remained active during symbiosis, for example the genes encoding R proteins like RPM1, RPS2, RPS5, Pti1 kinase, and the pathway regulators like SGT1, HSP90 and EDS1. Intriguingly there was a significant upregulation ofgene encoding PR-1 proteins which are members of Cysteine-rich

secretory proteins, Antigen 5, and Pathogenesis-related 1 proteins (CAP) 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 to be divergent in nature (Fig. 3C). There are two PR-1 proteins that clustered with defense responsive PRs and these *PR-1* genes were not upregulated during symbiosis further confirming the symbiotic PR-1s to be distinct. PR-1 proteins harbour an embedded defence signalling peptide (CAP-derived peptides or CAPE) where CNYxPxGNxxxxxPY is considered as a functional motif that mark 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-D, Supplementary Table 7).Since genes encoding CAP proteins are marker genes for the salicylic acid signaling pathway and systemic acquired resistance we also checked the SA/JA pathways to further understand the symbiont responsive signaling in *A. hypogaea*. As shown in Fig. 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. It needs to be mentioned that our analysis could not locate genes encoding NODULE SPECIFIC CYSTEINE-RICH (NCR) peptides in the DEGs that occurs in legumes belonging to the inverted repeat-lacking clade (IRLC) (e.g. *M. truncatula*, *Pisum sp.*, and *Trifolium sp.*) and recently demonstrated in *Aeschynomene sp.* as well (Van de Velde et al., 2010; Czernic et al., 2015).

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*, *NF-YA*, *Myb* and *CLE13* and other genes encoding a receptor kinase, a soluble kinase, a F-BOX protein, transcription factors of SHI-family and a lectin (Supplementary fig. 3; Supplementary Table 4). 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 orthologues of symbiotic genes

Our final objective was to understand the expression of the putative orthologues of symbiotic genes in *A. hypogaea* that are characterized in the model legumes *M. truncatula* and *L. japonicus*. A total of 71 genes were chosen and classified on the basis of their primary

association (Fig. 4; Supplementary Table 5). BLAST search on *A.ipanensis* and *A. duranensis* genome identified 68 (63 annotated) out of 71 genes for which the putative orthology was checked by reciprocal BLAST and sequence alignment. No orthologous gene for *MtRPG, MtSymCRK* and *MtDNF2* could be detected in either of these two parental genomes or in our transcriptome. For 63 out of 68 genes the symbiotic orthologue could be identified where the *A. hypogaea* sequences clustered with other legumes in the corresponding gene trees (Supplementary fig. 4). But in most cases the *A. hypogaea* genes were placed at the point of divergence of legumes from nonlegumes which is similar to what has been reported for *Ah*SYMRK, *Ah*CCaMK and *Ah*HK1 in respective distance trees(Sinharoy and DasGupta, 2009; Saha et al., 2014; Kundu and DasGupta, 2017b). Separation between the *A. hypogaea* genes and the other legume genes correlates with the rhizobial colonization by crack-entry and ITs. Exceptions were genes like *MtFLOT4, LjROP6, MtRR9, MtNOOT* and *LjSEN1* where the protein sequences from *A. hypogaea* were divergent and clustered with nonlegumes. However the expression of all these divergent genes was found to be significantly high during symbiosis indicating that they might have a role in *A.hypogaea* nodulation.In several cases we noted genomic bias in expression; for example genes like *LjNF-YC, MtFLOT2, MtFLOT4, LjROP6, MtbHLH476* and *MtRR4* had AA biased expression whereas expression of *MtDELLA, MtERF1, LjCHC1* and *LjASTRAY* was BB biased (Fig. 4). For comparison of expression of different symbiotic geneswe used the microarray data derived from the *M. truncatula* gene expression atlas (MtGEA)(http://mtgea.noble.org/v2/)(Benedito et al., 2008)&*L. japonicus* gene expression atlas (LjGEA) (https://ljgea.noble.org/v2/) (Verdier et al., 2013). If a symbiotic gene is characterised from one of these model legumes reciprocal BLAST was done to identify the orthologue in the other (Supplementary Table 5). Both absolute and the relative expression values (log_2 fold) were analysed so that high and constitutively expressed genes are not ignored (Fig. 4; Supplementary fig. 5).

In the recognition module, expression of genes encoding LCO-binding*LYR3*(Fliegmann et al., 2013) and EPS binding *EPR3*(Kawaharada et al., 2015) was significantly higher in *A. hypogaea* than the classical NF receptors (Fig. 4A). Whereas, in the model legumes the classical NF receptors like *LjNFR1/MtLYK3* and *LjNFR5/MtNFP* have a higher expression than these receptors. In the SYM pathway and early signaling most members had constitutive expression in all the 3 legumes irrespective of their mode of bacterial colonisation (Fig. 4B). Exception was gene encoding orthologue of cyclic nucleotide-gated channel *Mt*CNGC (Charpentier et al., 2016)

which was significantly upregulated in *A. hypogaea*. Most of the interactors of *NFR*s and *SYMRK* were also constitutively expressed (Fig. 4C). Expression of genes encoding ubiquitin ligase SIE3 (Yuan et al., 2012) (SYMRK interactor) and a UBQ superfamily protein CIP73 (Kang et al., 2011) (CCaMK interactor) was constitutively expressed in *A.hypogaea*. Genes encoding E3 uniquitin ligase *Mt*PUB1(Mbengue et al., 2010) (NFR1 interactor) and a symbiotic remorin *Mt*SYMREM (Lefebvre et al., 2010) (SYMRK and NFR1 interactor) that were upregulated in all 3 legumes highlighting their importance in nodulation. Among the TFs, upregulation of *NIN* (Singh et al., 2014) (target of CYCLOPS) expression was noted in the second transcriptional wave at 8dpi and expression of *NSP2* was only detectable in mature nodules of *A. hypogaea* (Fig. 4D). In model legumes expression of *NIN* and *NSP2* were upregulated on bacterial invasion(Schauser et al., 1999b; Kaló et al., 2005). Unlike model legumes, expression of *Mt*ERF1 as compare to other TFs was very high in *A. hypogaea*. The Expression pattern of *Mt*NSP1, *Lj*NFYA/ *Lj*NFYC (McDowell et al., 2013) (target of NIN), *Lj*ERN1(Cerri et al., 2017) (target of CYCLOPS), *Mt*DELLA (Jin et al., 2016) (bridge between IPD3/CYCLOPS and NSP2), *Lj*IPN2 (Kang et al., 2014) (NSP2 interactor) and *Lj*SIN1(Battaglia et al., 2014) (NF-YC interactor) was similar in *A. hypogaea* and in model legumes, suggesting that the basic transcriptional network could be conserved between these legumes. In the infection module *Mt*RPG is responsible for rhizobium-directed polar growth of ITs and gene encoding it's orthologue was not detected in *A. hypogaea* (Arrighi et al., 2008). *Mt*VPY (ankyrin repeat) is important for infection progression in model legumes (Murray et al., 2011a) and it had significantly low expression in *A. hypogaea*. On the other hand expression of factors like *LjARPC1*(Hossain et al., 2012) and *LjCEREBERUS* (Yano et al., 2009) that are important for the progress of infection was significantly higher in *A. hypogaea* (Fig. 4E).Other indicated factors that are required for bacterial invasion like *LjNAP1*(Yokota et al., 2009), *LjPIR1*(Yokota et al., 2009), *LjnsRING* (Shimomura et al., 2006), *MtFLOT4* (Haney and Long, 2010) and *MtFLOT2* (Haney and Long, 2010) were induced in all 3 legumes suggesting their analogous purposes. In the nodule organogenesis module, expression of cytokinin receptor *MtCRE1/LjHK1/AhHK1*(Gonzalez-Rizzo et al., 2006b; Murray et al., 2007; Kundu and DasGupta, 2017a) and the cytokinin inducible Type-A RRs like *MtRR1*(Ariel et al., 2012), *MtRR4*(Gonzalez-Rizzo et al., 2006b; Op den Camp et al., 2011), *LjRR5*(Murray et al., 2007) and *MtRR9* (Op den Camp et al., 2011)was high in all 3 legumes (Fig. 4F). The expression of cytokinin inducible TF *MtbHLH476* (Ariel et al., 2012) was however significantly high in *A. hypogaea* (Fig. 4F). All other factors that are required for the establishment and maintenance of nodule meristems like *MtNIP/LATD* (Yendrek et al., 2010), *MtWOX5* (Osipova et al., 2012), *MtNOOT* (Couzigou et al.,

2012)and *MtENOD40*(Crespi et al., 1994) have comparable expression pattern between *A. hypogaea* and model legumes (Fig. 4F). In the differentiation module, both *MtDNF2* and *MtSymCRK* are reported for supressing defense responses during nodulation were not detected in *A. hypogaea* genome (Bourcy et al., 2013; Berrabah et al., 2014). All other factors that are required for bacteroid differentiation like *LjSUNERGOS1(Yoon et al., 2014), LjVAG1(Suzaki et al., 2014), MtDNF1(Wang et al., 2010), MtRSD(Sinharoy et al., 2013), LjSEN1(Hakoyama et al., 2012), LjSST1(Krusell, 2005)* and *LjFEN1*(Hakoyama et al., 2009) have similar expression in all 3 legumes and may have conserved function (Fig. 4G). Among the nodule number regulators, expression of *MtEIN2*(Varma Penmetsa et al., 2008) (*sickle*) was distinct in *A. hypogaea*(Fig. 4H). It plays a key role in a range of plant–microbe interactions and had a significantly high expression in *A. hypogaea.* All other regulators like *MtSUNN* (Elise et al., 2005), *LjKLAVIER* (Miyazawa et al., 2010), *MtEFD* (Vernie et al., 2008a), *LjASTRAY* (Nishimura et al., 2002) and *MtRDN* (Kassaw et al., 2017) has comparable expression pattern in all 3 legumes (Fig. 4H). Quantitative reverse transcription-polymerase chain reaction (qRT-PCR) was done for11 symbiotic genes to prove the reliability of the RNAseq data (Supplementary Fig. 6). For few time-points the fold change of qRT-PCR and DEG analysis did not exactly match but mostly the results highlighted their consistency.

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

PCA analysis was done to check if there was a signature in the pattern of expression of symbiotic genes in crack-entry legume *A. hypogaea* that contrasts with the model legumes where rhizobial entry is IT mediated. Fig. 5 is a projection of differential expression of symbiotic genes from *A. hypogaea*, *M.truncatula* and *L.japonicus* into first two principal components. Altogether, expression of around 87% genes were found to be aligned along dimension1 (dim1) and dimension 2 (dim2) in the analysis. Expressions of symbiotic genes that show minimal change in expression and are likely to be regulated at post transcriptional level are clustered near the origin. Only for select genes, there were contrasting trends in differential expression between *A.hypogaea* and both the model legumes together placing them in opposing or adjacent quadrants. These contrasts were interpreted as significant for *A.hypogaea* symbiosis (Supplementary fig. 7). For example, among the early signaling and SYM pathway genes *AhCNGC* was distinctly placed away from their counterparts in model legumes. Among early TFs, *NIN*, *NSP2*, and *SIN1* were distinct. In the infection module, *AhVPY* and *AhCERBERUS* were distinct and placed in opposing quadrants with respect to model legumes. Among the interacting

proteins, *PUB1* scores in both dimensions in model legumes whereas in *A.hypogaea* it clusters near the origin. In the organogenesis module cytokinin inducible *RR1* and *ENOD40* and in the nodule differentiation module *SST1*havea contrasting trend in expression. Among the nodule number regulators expression of *EFD* was distinct. These factors highlighted by PCA analysis appear to be differentially adapted in *A. hypogaea* symbiosis.

DISCUSSION

This is the first systematic effort towards transcriptome profiling with the progress of symbiosis in a crack-entry legume *A. hypogaea*. 3 major transcriptional programs appear to govern the process. The first program is for rhizobial recognition and generation of nodule primordia by 1-4 dpi, the 2ndprogram is for structural development of nodules by 8-12dpi and the 3rd program is for functional maturation of nodules at 21dpi (Fig. 1-2). The comparison of expression of putative orthologues of symbiotically important genes in *A. hypogaea* with model legumes highlighted the genes that are important or disposable for its crack-entry mediated root nodule symbiosis (Fig. 4-6).

The most significant observation in *A. hypogaea* symbiotic transcriptome was the over expression of a group of genes encoding a divergent form of cysteine rich PR-1 proteins during the structural and functional development of nodules 8dpi onwards (Fig. 3). PR-1 proteins are ubiquitous across plant species and are among the most abundantly produced proteins in plants in response to pathogen attack. It is used as a marker for salicylic acid-mediated disease resistance in plants(Breen et al., 2017). Although differential expression of defense response genes belonging to GO:0006952 (defense related) and *PR-1/PR-10* protein families has previously been reported for *M.truncatula* RNS(Jardinaud et al., 2016), this is the first report where 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 an embedded Pro-rich C-terminal peptide (CAPE) that is involved in plant immune signaling(Breen et al., 2017). All the symbiotic PR-1s in *A. hypogaea* has both these conserved features but whether these CAPE peptides are actually derived from PR-1 proteins during symbiosis remains to be understood. It is relevant to mention here that NCR family of peptides are very highly expressed during nodulation in *M. truncatula* (Van de Velde et al., 2010). These peptides evolved from defensin ancestors and until recently was assumed to be specific to legume species belonging to the IRLC clade where they are

responsible for bacterial endoreduplication (Mergaert et al., 2006). Recently, divergent form of NCR peptides were reported to be essential for bacterial endoreduplication associated shape change in Nod-factor independent crack-entry legume *A.evenia* (Czernic et al., 2015). Intriguingly NCRs were absent in crack-entry legume *A. hypogaea* where similar to *A. evenia*, rhizobia change from a rod to spherical shape but unlike *Aeschynomene sp.* the symbiosis in *A. hypogaea* is NF-dependent (CHANDLER et al., 1982; Ibáñez and Fabra, 2011). Thus, it is imperative to investigate whether the antimicrobial CAPE peptides were enrolled as symbiosis effectors in *A. hypogaea* in place of NCRs.

Based on the comparative expression analysis of symbiotic genes, we propose a simple molecular framework where we highlighted those genes in *A. hypogaea* that are either conserved or divergent from the model legumes be it in sequence or in expression pattern (Fig. 6). The high expression of LCO binding receptor *LYR3* as compared to the classical NF receptors indicated NF signalling could be mediated through this receptor in *A. hypogaea* (Fig. 4). Intriguingly *NFR1* and *LYR3* were not reported in *A. evenia*, which is a NF-independent crack-entry legume as opposed to *A. hypogaea* which is NF-dependent (Ibáñez and Fabra, 2011; Fabre et al., 2015; Chaintreuil et al., 2016). Expression pattern of genes belonging to SYM pathway and early signalling in *A. hypogaea* were found to be similar to model legumes with only exception being *AhCNGC*. Significant upregulation of *AhCNGC* suggests its possible importance in mediating symbiotic calcium oscillations in the SYM pathway of crack-entry legumes.

Several observations indicated change in expression pattern of symbiotic genes in *A. hypogaea* in the absence of epidermal IT formation. For example TFs like *AhNIN* and *AhNSP2* are only expressed at the later stages of symbiosis in *A. hypogaea* indicating that unlike in model legumes these TFs may not have an early role in bacterial entry (Fig. 4). However, the cortical roles for these TFs could be conserved between IT and crack-entry legumes. Factors like *MtVPY* and *MtRPG* have a role in polar growth process of IT in model legumes (Arrighi et al., 2008; Murray et al., 2011b). That explains the absence of *RPG* and insignificant expression of *VPY* in *A. hypogaea* transcriptome (Fig. 4). The contrasting expression pattern of *LjCERBERUS* in *A. hypogaea* indicated its divergent function during rhizobial invasion through epidermal cracks. In NF-dependent symbiosis, membrane raft proteins like MtFLOT2 and MtFLOT4 are important for IT initiation and elongation(Haney and Long, 2009). While orthologues of both these *FLOT*s are absent in NF-independent crack-entry legume *A.evenia* (Chaintreuil et al., 2016), substantial expression was detected in *A. hypogaea* transcriptome suggesting them to be recruited for other

purposes (Fig. 3). The orthologue of EPS binding receptor *MtEPR3* is absent in *A.evenia* but upregulated during symbiosis in *A. hypogaea* suggesting it to have functions other than regulating IT progression.

Apart from NCRs, there are other features that contrast the process of differentiation in *A. hypogaea* and *A.evenia*. For example, neither of the topoisomerases *LjSUNERGOS* and *LjVAG1* and the homocitrate synthase *LjFEN1* is detectable in *A.evenia* (Chaintreuil et al., 2016). On the other hand *MtDNF2*, a phospholipaseC and *MtSymCRK* a non-RD receptor kinase that are required for suppressing defense response during bacteroid differentiation are absent in both the legumes belonging to the dalbergoid clade (Fig.6), thus indicating these genes to be disposable for crack entry mediated root nodule symbiosis.

Comparative analysis of DEGs between *A. hypogaea* and model legume highlighted the predominance of cytokinin and ethylene signaling during *A. hypogaea* nodulation. Two component cytokinin receptor HK1 has a central role in nodule organogenesis of both *A. hypogaea* and model legumes (Gonzalez-Rizzo et al., 2006a; Murray et al., 2007; Kundu and DasGupta, 2017b). Although PCA analysis indicated *AhHK1*, *LjHK1 (LHK1*) and *MtCRE1* to have similar expression pattern, its downstream effectors showed altered pattern of expression during *A. hypogaea* symbiosis (Fig. 5). Expressions of type-B RR like *MtRR1*, which is cytokinin responsive transcription factor and responsible for modulating downstream factors like *MtNSP1*and *MtbHLH476*was found to have a distinct expression pattern in *A. hypogaea* in comparison to its model legume counterparts (Fig. 5). In accordance, *AhbHLH476* was found to be very highly expressed during *A. hypogaea* nodulation (Fig. 4). Another cytokinin responsive factor *AhENOD40* was also found to be distinctly placed in a different quadrant in PCA analysis (Fig. 5). The distinct role of cytokinin signaling during *A. hypogaea* nodulation is in accordance to the previous report where silencing of *AhHK1* resulted in delayed nodulation associated with problem in nodule differentiation (Kundu and DasGupta, 2017b).

During nodulation ethylene responsive transcription factors play a decisive role by controlling cell division and differentiation (Asamizu et al., 2008; Vernie et al., 2008b). Previous report on *A. hypogaea* transcriptomics highlighted the upregulation of several AP2-domain containing ethylene responsive TFs during nodulation (Peng et al., 2017). Similarly, our transcriptomic analysis also indicated significantly high expression of the symbiotic orthologue of *ERF1* (Fig.4). In *L. japonicus LjERF1* is a positive regulator of nodulation and downregulates the expression of

defense gene *LjPR-10* during symbiosis(Asamizu et al., 2008). Intriguingly in *A.hypogaea* high expression of *ERF1* is associated with high expression of *PR-1s* indicating that the ethylene signalling network is differently recruited during *A.hypogaea* symbiosis (Fig. 5).In consistence with such proposition the expression of *EIN2*, the master regulator of ethylene signalling was significantly high in *A.hypogaea*, and the pattern of expression of *EFD*, a negative regulator of nodulation was distinctly different from model legumes (Fig. 4-5). The differential role of ethylene signalling during crack entry nodulation strongly supports the fact that ethylene signalling inhibits intracellular infection via infection threads while promoting intercellular infection via crack-entry (Vernie et al., 2008b).

In summary, the transcriptional dynamics with the progress of symbiosis in *A. hypogaea* highlighted the factors that are disposable or essential for the inception and progress of symbiosis in a crack entry legume.

MATERIALS AND METHODS:

Plant Materials and Sample Preparation:

Five different developmental stages of *A. hypogaea* total infected roots, nodules and uninfected roots were used in this study (UI, 1DPI, 4DPI, 8DPI, 12DPI and 21DPI). *A. hypogaea* JL24 strain seeds (from ICRISAT, INDIA) were surface sterilized and soaked into sterile water for germination. Germinated seeds were then transferred in pots containing sterile vermiculite and soilrite at 25^{o}C growth room for 7 days before inoculation with *Bradyrhizobium sp.* SEMIA 6144 (from Adriana Fabra, Universidad Nacional de Rio cuarto, Cordoba, Argentina) grown in liquid Yeast-Mannitol broth supplemented with 100mM $CaCl_2$ at 28^{o}C(A_{600}= 0.5–0.7). Samples are harvested, cleaned and freezed in liquid nitrogen. Frozen samples are 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 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(Haynes et al., 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 et al. (Kundu and DasGupta, 2017b). All digital micrographs were processed using Adobe Photoshop CS5.

Isolation of total RNA:

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

Library construction and Sequencing:

18 RNA library was prepared using an IlluminaTruSeq stranded mRNA sample preparation kit by MGX-Montpellier GenomiX core facility (MGX) France (https://www.mgx.cnrs.fr/). The protocol first requires the selection of polyadenylated RNAs on oligodT magnetic beads. Selected RNAs are chemically fragmented and the first strand cDNA is synthesized in the presence of actinomycin D. The second strand cDNA synthesis is incorporating with dUTP in place of dTTP which quenches it to the second strand during amplification. A 3' ends adenylation is used to prevent fragments from ligating to one another during the adapter ligation process. The quantitative and qualitative validation of the library is 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 (SBS) technique performed by MGX, France and 50bp single-end reads for each library were generated (Fuller, 1995).

Illumina Reads Mapping and Assembly:

Quality control and assesment of raw Illumina reads in FASTQformat 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, poly-N containing reads by using Trimmomatic v0.36 software(Bolger et al., 2014). Clean Reads are simultaneously aligned to the two wild peanut diploid ancestors *A. duranensis*(AA) and *A. ipaensis*(BB) reference genome by using TopHat2 version 2.0.13 which is a fast splice junction mapper for RNA-Seq

reads (Trapnell et al., 2010; Bertioli et al., 2015). It aligns RNA-Seq reads using the ultra high-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). The alignment files were combined and analyzed into Trinity for genome-guided assembly (Grabherr et al., 2011). The reference based assembly was compared to its respective transcript files from annotated reference genomes by using BLAT(Kent, 2002). An e-value cutoff of '$1e^{-05}$' was used to determine a hit. The annotated hits were furthermore analysed in this study. Genome annotation files in generic feature format (GFF) are downloaded from peanut database (https://peanutbase.org/download)(Dash et al., 2016). Estimation of gene expression level of each annotated transcript was performed by StringTie v1.3.3 which takes sorted sequence alignment map (SAM) or binary (BAM) file for each sample along with genome annotation files (Pertea et al., 2015). Resulted gene transfer format (GTF), normalized gene locus expression level as fragments per kilobase million (FPKM), transcripts per million (TPM), and count files for each sample were further analyzed for fold change analysis in gene expression levels.

Identification of DEGs and functional Gene Ontology and KEGG pathway analyses of the DEGs:

Before statistical analysis, genes with less than 2 values lower than one count per million (cpm) were filtered out. EdgeR 3.6.7 package was used to identify the differentially expressed genes(Robinson et al., 2010). Data were normalized using "Trimmed mean of M-values (TMM)" method. Genes with 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-diagram are generated using (http://www.interactivenn.net/)(Heberle et al., 2015) and hierarchical heatmap is generated usingTM4MeV (http://mev.tm4.org and http://www.tigr.org/software/tm4/mev.html)(Howe et al., 2011) the values from the venn diagram (Supplementary Table 2).Detailed functional annotation and explanations of DEGs were extracted from gene ontology database (http://www.geneontology.org/)(Ashburner et al., 2000) and GO functional classification analysis was done using software WEGO (http://wego.genomics.org.cn/cgi-bin/wego/index.pl)(Ye et al., 2006). The GO terms for DEGs in genome annotation were also retrieved from the 'GFF' file downloaded at PeanutBase website (http://peanutbase.org). To identify important and enriched pathways involved by the DEGs, the DEGs were assigned to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using the web server (http://www.genome.jp/kaas-bin/kaas_main)(Kanehisa and Goto, 2000) against *A.*

duranensis and *A. ipaensis* gene datasets. Enriched KO and GO terms are obtained by a developed Python script which uses hypergeometric testand Bonferroni corrected P-Value < 0.05.

Identification of Symbiotic orthologous gene in *A. hypogaea*:

Candidate symbiotic genes were identified in *A. hypogaea*, *L. japonicus* and *M. truncatula* using BLASTN searches with reported nucleotide sequence of genes from *L. japonicus* and *M. truncatula*. The homologous genes of were searched in *A. duranensis* and *A. ipaensis* in *PeanutBase* (http://peanutbase.org), *M. truncatula Mt4.0v1* genome was searched in *M.truncatula* gene expression atlas(MtGEA) (http://mtgea.noble.org/v2/) and the *L. japonicus v3.0* genome was searched in *L. japonicus* gene expression atlas (LjGEA) (https://ljgea.noble.org/v2/). Initial searches were conducted with E-value = e^{-5}. The results were manually validated for the presence of an orthoologous gene in an open reading frame and searched for orthologues using BLASTP. Orthology of the genes were validated by generating neighbor joining phylogenetic tree using amino acid sequences in MEGA 6.0 obtained from BLASTP (Tamura et al., 2013).

qRT PCR validation:

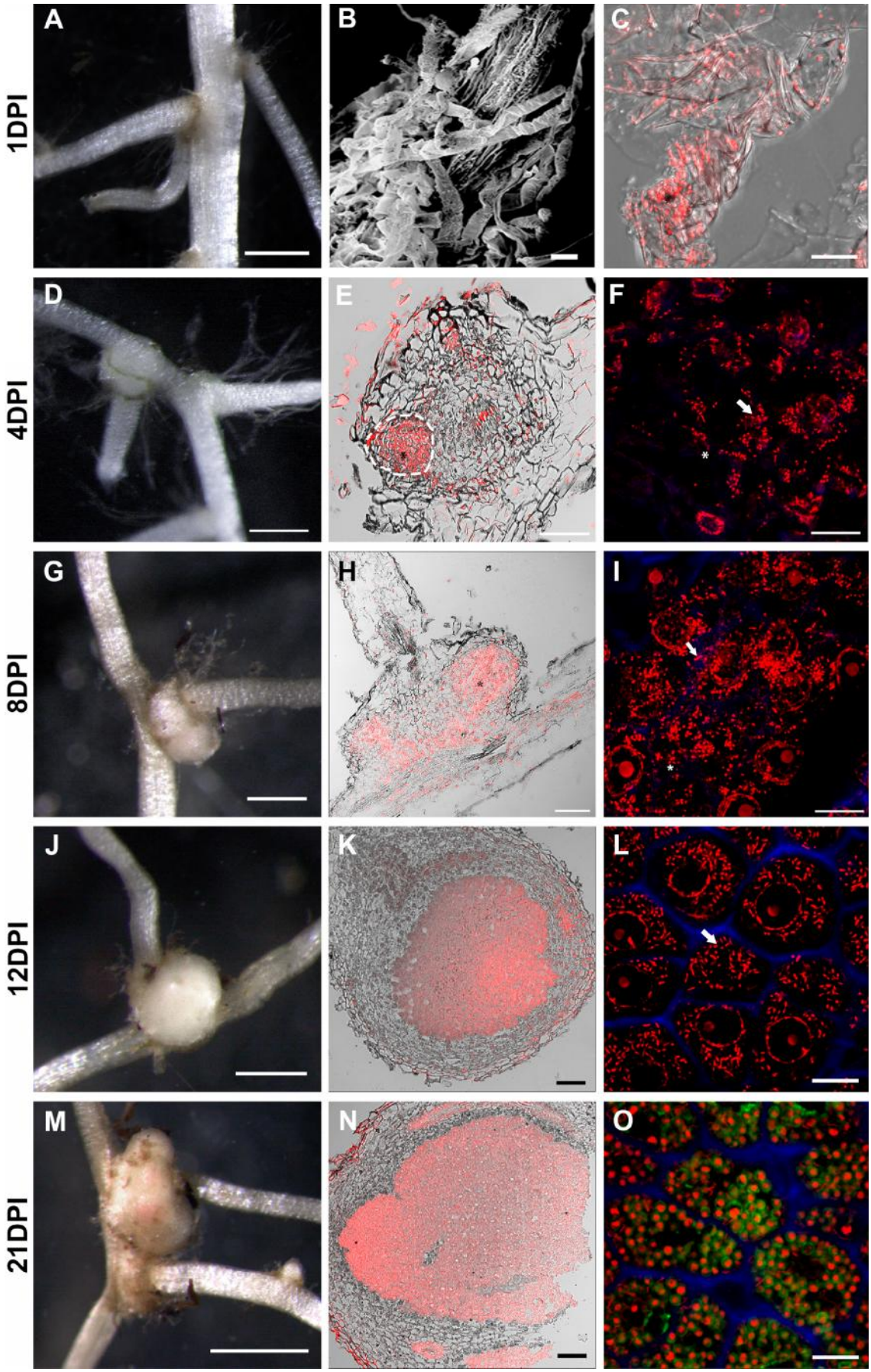
Total RNA (500 ng) was reverse-transcribed by using Super-ScriptIII RT (Life Technologies) and oligo (dT). RNA quantity from each sample in each biological replicate was standardized prior to first-strand cDNA synthesis. qRT-PCR was performed by using Power SYBR Green PCR Master Mix (Applied Biosystems) using primers as designed using software Oligoanalyser (Intergrated DNA Technology) (Supplementary Table S6). Calculations were done using the ΔΔ cycle threshold method using *AhActin* as the endogenous control. The reaction were run in Applied biosystems 7500 Fast HT platform using protocol: 1 cycle at 50°C for 2 mins, 1 cycle at 95°C for 5 min, 40 cycles at 95°C for 30 sec, 54°C for 30 sec, 72°C for 30 sec followed by melt curve analysis at 1 cycle at 95°C for 1 min, 55°C for 30 sec, and 95°C for 30 sec. A negative control without cDNA template was checked for each primer combination which was designed using OligoAnalyzer 3.1 (https://www.idtdna.com/calc/analyzer). Results were expressed as means standard error (SE) of the number of experiments.

Data Availability:

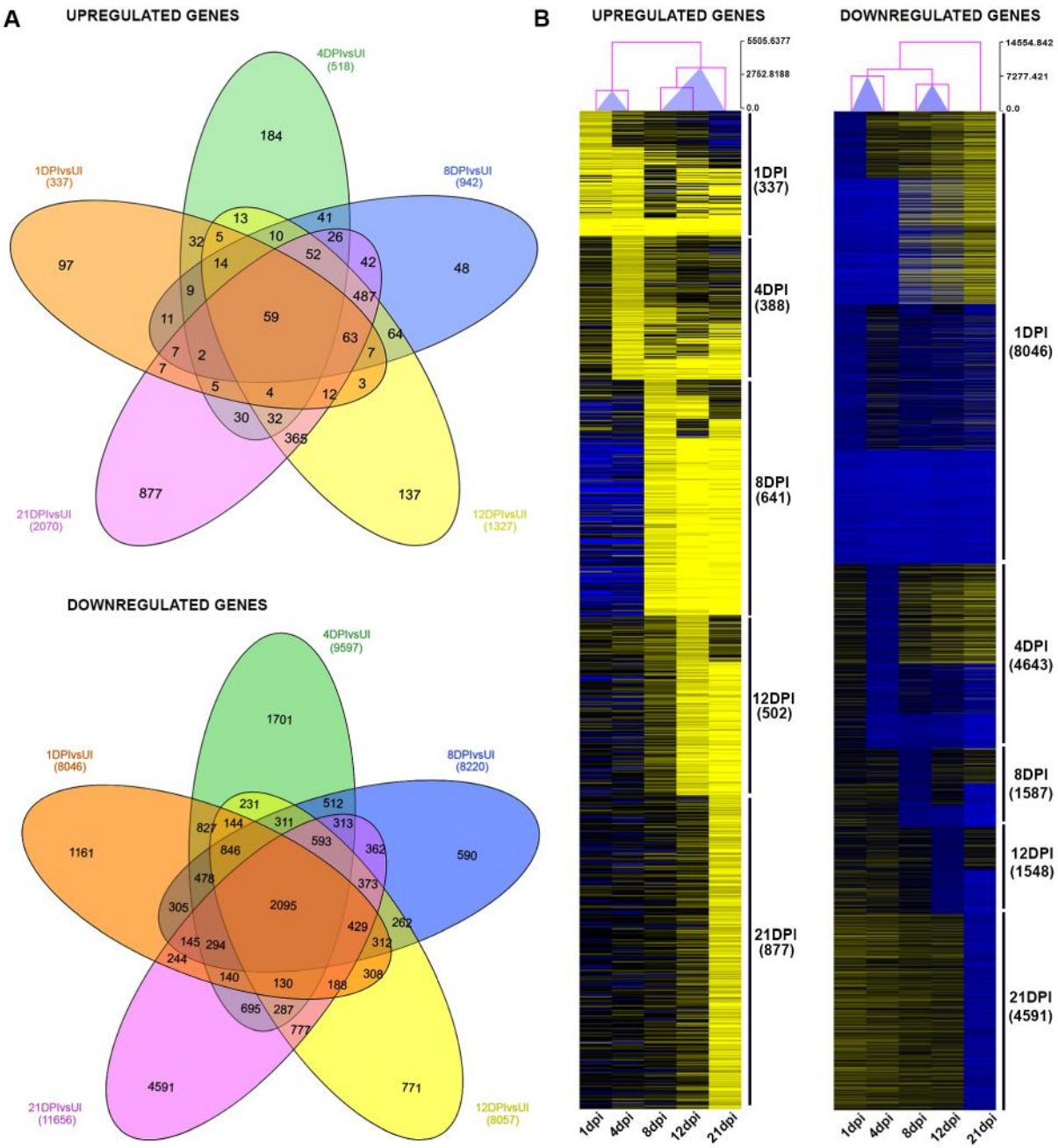
The raw FASTQ files for the 18 libraries were deposited in the Gene expression omnibus (GEO) of NCBI under accession number GSE98997.

530 **Figures:**

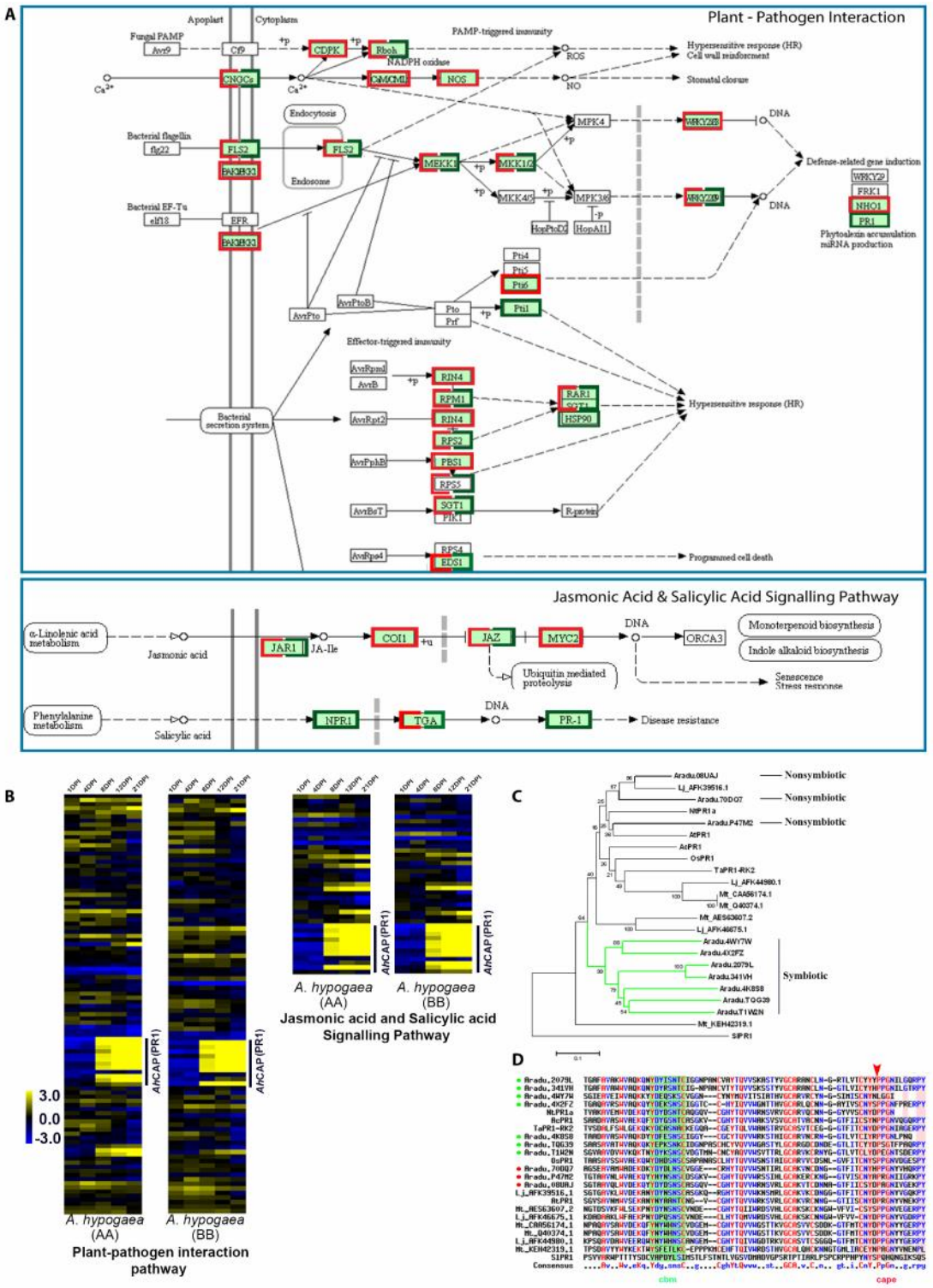
531 **Figure1:**



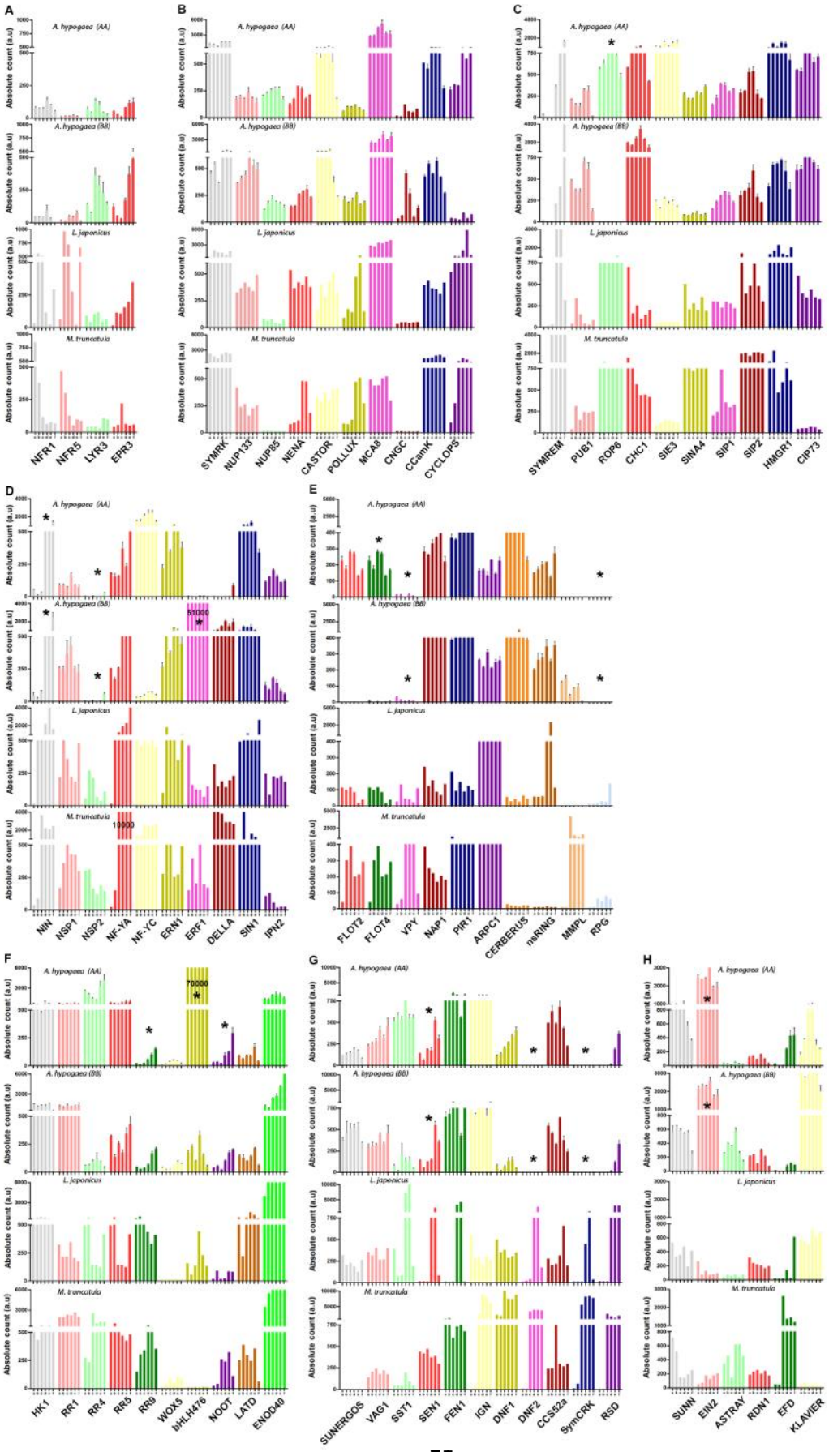
532 **Figure2:**



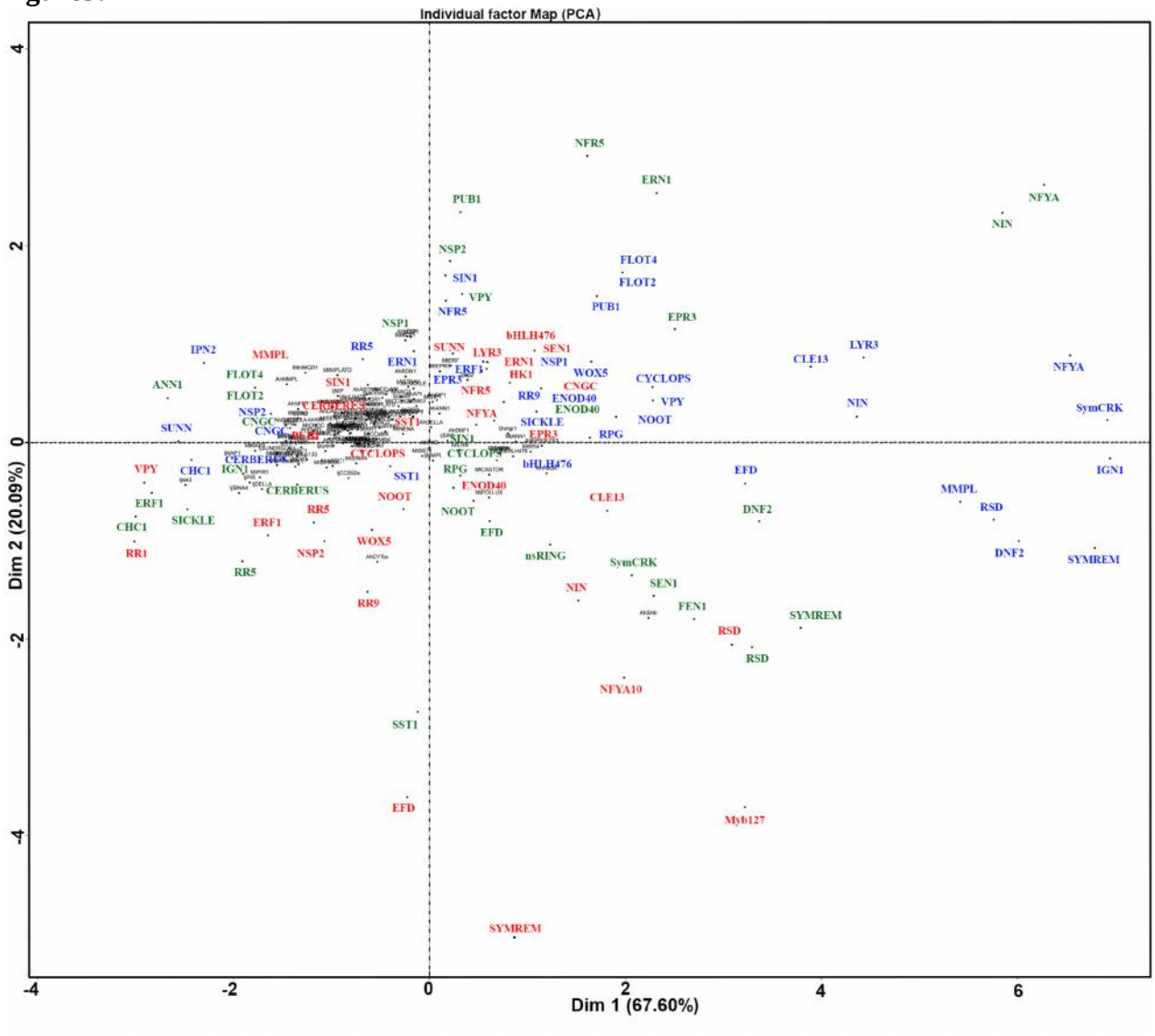
533 **Figure3:**



534 **Figure4:**



535 **Figure5:**



536 **Figure6:**

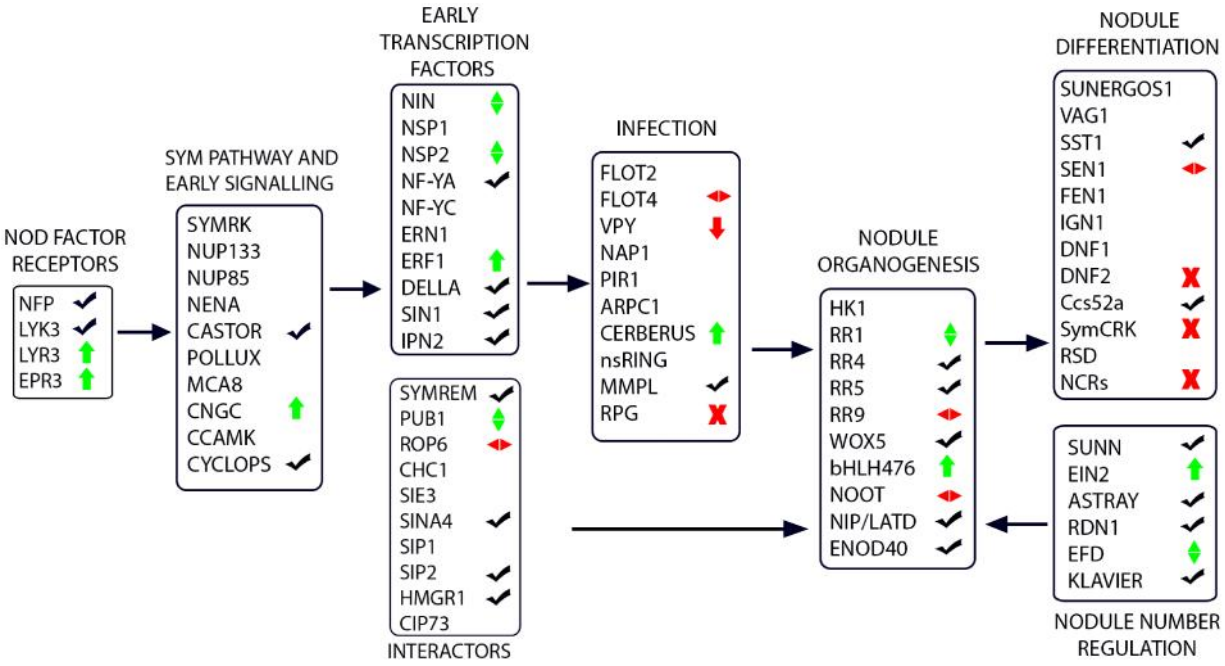


Table and Figure legends:

Table 1: Summary of Raw Illumina Sequencing and filtered reads after trimming and alignment of reads to AA (*Arachis duranensis*) and BB (*Arachis ipaensis*) genomes in each library

Samples	Raw Reads	Filter reads (%)	% of Total Mapped Reads	
			Arachis duranensis (AA Genome)	*Arachis ipaensis (BB Genome)*
UI	57913214	95.08%	82.96%	86.71%
	65769652	94.88%	87.78%	84.34%
	85558982	90.59%	85.29%	86.19%
1DPI	61747256	95.00%	88.80%	87.51%
	62925897	91.38%	88.78%	89.98%
	74542718	91.68%	87.85%	88.76%
4DPI	96807458	89.21%	88.46%	84.42%
	97275402	90.46%	86.61%	87.71%
	79984160	91.21%	85.24%	86.25%
8DPI	89987084	90.09%	86.03%	87.19%
	11535441	89.85%	86.20%	83.03%
	77031650	91.50%	87.81%	88.95%
12DPI	57198145	90.28%	86.77%	88.08%
	74743668	92.33%	88.47%	89.71%
	96613689	89.38%	86.12%	87.20%
21DPI	65329428	90.03%	87.81%	86.81%
	100254251	90.71%	84.45%	85.42%
	70839550	90.46%	80.15%	81.44%

FIGURE LEGENDS:

Figure 1: *Arachis hypogaea* nodule ontogeny at different time point post inoculation with *Bradyrhizobium* sp. SEMIA6144. (A-C) Root harvested 1 day post inoculation (1DPI) where (A) root junction, (B) SEM of inoculated root hair at lateral root junction and (C) CLSM image of lateral root junction. Section of nodule primordia at (D-F) 4DPI and (G-I) 8DPI. Section of nodule at (J-L) 12DPI and (M-O) 21DPI. Dashed lines delimit the infection zone in E and asterix indicate the precise position. Arrow indicate rod shaped rhizobia in F, I and L. Stereoimage (A, D, G, J and M); bright field and PI merged (C, E, H, K and N); PI + Calcofluor merged (F, I and L); PI + Calcofluor + syto9 merged (O). Scale bar: 500µm (A, D, G and J), 1mm (M), 2µm (B), 100µm (E, H, K and N) and 10µm (F, I, L and O). PI (red), Calcofluor (blue) and Syto9 (green).

Figure 2: Comparison of differentially expressed genes (DEGs) identified in the five time points 1DPI, 4DPI, 8DPI, 12DPI and 21DPI. The total DEGs were identified by EdgeR analysis where (**A**) Interactive Venn-diagram shows the comparison between Up-regulated genes and Down-regulated genes (from both AA and BB). (B) Heat map of the hierarchical cluster analysis of the Up-regulated genes and Down-regulated genes (both AA and BB). The columns indicating the different time points above the map and the arborescence indicate the similarity among transcriptomes. Below the heat map is a color coded scale bar for the relative expression levels of genes in log2 scale. Numbers beside the heat-map indicates the exclusive number of DEGs at different time point (see Supplementary Table 2).

Figure 3: **DEGs associated to plant-pathogen interaction and jasmonicacid-salicylic acid signaling across different time points post inoculation.** (A) KEGG analysis of upregulated (green) and downregulated (red) DEGs in the respective pathways. (B) Heat-map of the DEGs that are previously annotated in KEGG pathway analysis. (C) Neighbour-joining phylogenetic tree of all the annotated CAP proteins using non-truncated amino acid sequences where green branch denotes divergent upregulatedCAP-PR1 proteins. (D) CLUSTALW sequence alignment of CAPE peptides using Multalin where arrow indicates cleavage site for CAPE peptide, CBM and CAP domain are annotated by colored box and bullet indicates *A.hypogaea* CAP peptides (upregulated:green and downregulated:red).

Figure 4: Comparative *Expression Pattern of 71 Symbiotic genes in Arachis hypogaea, Medicago truncatula and Lotus japonicus.* For *A. hypogaea* histogram represent normalized RNA-seq reads (FPKM) of symbiotic orthologous genes aligned with AA and BB genome of *A.*

duranensis and *A. ipaensis* respectively. For *Medicago truncatula* and *Lotus japonicus* histogram represent microarray data retrieved from the respective Gene Expression Atlas Affymetrix database (MtGEA, LjGEA; Benedito et al., 2008). Relative expression level data of symbiotic genes (in arbitraty units, a.u.) during nodulation kinetics are grouped as (A) Nod factor receptors, (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 post inoculation by *Bradyrhizobium sp.* SEMIA6144 for *A. hypogaea*, *Sinorhizobium meliloti* for *M. truncatula* and *Mesorhizobium loti* strain R7A for *L. japonicus*). a,b,c,d,e and f in x-axis represented UI,1,4,8,12 and 21DPI respectively for *A. hypogaea*; UI,3,6,10,14 and 20DPI respectively for *M. truncatula*; UI,1,3.7,14 and 21DPI respectively for *L. japonicus.*

Figure 5: **Individual Factor map Principal Component Analysis (PCA) of differentially expressed Symbiotic genes.** *A. hypogaea* (red), *M. truncatula* (blue) and *L. japonicus* (green) genes are highlighted by a single acronym. Detail of the genes is mentioned in Supplementary Table 5. % of genes represented in Dim 1(dimention 1) and Dim 2 (dimention 2).

Figure 6: Simplified molecular model for the symbiotic signaling pathway in legume. Homologous symbiotic genes identified in *M. truncatula* and *L. japonicus* are listed and classified according to their main symbiotic functions. *A. hypogaea* putative orthologous genes are annotated by Black tick: Differentially expressed, Green up-headarrow: upregulation, Red down-headarrow: downregulation, Red Cross: absent in diploid genome, Green up-down triangle: divergent expression, Red sideway triangle: phylogenetically divergent.

Author's Contribution:

Project planning: A.K. and M.D.G. Sample preparation: K.K. and A.K.; Microscopy of symbiosis: A.K.; Preparation of RNA: K.K. Production of Illumina libraries, sequencing and transcriptome assembly: E.D, D.S.; Analysis of transcriptome: K.K. and A.Z.; Analysis of symbiotic transcriptome: K.K and A.K.; Critical analysis of data : P.C and F.C. Writing of the manuscript: A.K., K.K. and M.D.G. All authors approved the manuscript.

Acknowledgement:

This work was funded by Grants from Govt. 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; fellowship to K.K and A.Z.R (IFCPAR/CEFIPRA: IFC/5103-4/2014/543); fellowship to A.K (Council of Scientific and Industrial Research, CSIR-09/028[0756]/2009–EMR–I).

REFERENCES

Ané, J.-M., Kiss, G.B., Riely, B.K., Penmetsa, R.V., Oldroyd, G.E., Ayax, C., Lévy, J., Debellé, F., Baek, J.-M., and Kalo, P. 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., and Chabaud, M. 2012. Two direct targets of cytokinin signaling regulate symbiotic nodulation in Medicago truncatula. The Plant Cell 24:3838-3852.

Arrighi, J.F. 2006. The Medicago truncatula Lysine Motif-Receptor-Like Kinase Gene Family Includes NFP and New Nodule-Expressed Genes. Plant physiology 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. Proceedings of the National Academy of Sciences 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 physiology 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., and Eppig, J.T. 2000. Gene Ontology: tool for the unification of biology. Nature genetics 25:25-29.

Battaglia, M., Rípodas, C., Clúa, J., Baudin, M., Aguilar, O.M., Niebel, A., Zanetti, M.E., and Blanco, F.A. 2014. A nuclear factor Y interacting protein of the GRAS family is required for nodule organogenesis, infection thread progression, and lateral root growth. Plant physiology 164:1430-1442.

Benedito, V.A., Torres-Jerez, I., Murray, J.D., Andriankaja, A., Allen, S., Kakar, K., Wandrey, M., Verdier, J., Zuber, H., and Ott, T. 2008. A gene expression atlas of the model legume Medicago truncatula. The Plant Journal 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 Phytologist 203:1305-1314.

656 Bertioli, D.J., Cannon, S.B., Froenicke, L., Huang, G., Farmer, A.D., Cannon, E.K., Liu, X., Gao, D., Clevenger,
657 J., and Dash, S. 2015. The genome sequences of Arachis duranensis and Arachis ipaensis, the
658 diploid ancestors of cultivated peanut. Nature Genetics 47:438.
659 Bolger, A.M., Lohse, M., and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data.
660 Bioinformatics 30:2114-2120.
661 Boogerd, F.C., and Rossum, D. 1997. Nodulation of groundnut by Bradyrhizobium: a simple infection
662 process by crack entry. FEMS microbiology reviews 21:5-27.
663 Bourcy, M., Brocard, L., Pislariu, C.I., Cosson, V., Mergaert, P., Tadege, M., Mysore, K.S., Udvardi, M.K.,
664 Gourion, B., and Ratet, P. 2013. Medicago truncatulaDNF2 is a PI-PLC-XD-containing protein
665 required for bacteroid persistence and prevention of nodule early senescence and defense-like
666 reactions. New Phytologist 197:1250-1261.
667 Breen, S., Williams, S.J., Outram, M., Kobe, B., and Solomon, P.S. 2017. Emerging insights into the
668 functions of pathogenesis-related protein 1. Trends in plant science 22:871-879.
669 Cerri, M.R., Wang, Q., Stolz, P., Folgmann, J., Frances, L., Katzer, K., Li, X., Heckmann, A.B., Wang, T.L., and
670 Downie, J.A. 2017. The ERN1 transcription factor gene is a target of the CCaMK/CYCLOPS complex
671 and controls rhizobial infection in Lotus japonicus. New Phytologist 215:323-337.
672 Chaintreuil, C., Rivallan, R., Bertioli, D.J., Klopp, C., Gouzy, J., Courtois, B., Leleux, P., Martin, G., Rami, J.-F.,
673 Gully, D., Parrinello, H., Séverac, D., Patrel, D., Fardoux, J., Ribière, W., Boursot, M., Cartieaux, F.,
674 Czernic, P., Ratet, P., Mournet, P., Giraud, E., and Arrighi, J.-F. 2016. A gene-based map of the Nod
675 factor-independentAeschynomene eveniasheds new light on the evolution of nodulation and
676 legume genomes. DNA Research 23:365-376.
677 CHANDLER, M.R., Date, R., and Roughley, R. 1982. Infection and root-nodule development in Stylosanthes
678 species by Rhizobium. Journal of Experimental Botany 33:47-57.
679 Charpentier, M., Sun, J., Martins, T.V., Radhakrishnan, G.V., Findlay, K., Soumpourou, E., Thouin, J., Véry,
680 A.-A., Sanders, D., and Morris, R.J. 2016. Nuclear-localized cyclic nucleotide–gated channels
681 mediate symbiotic calcium oscillations. Science 352:1102-1105.
682 Couzigou, J.-M., Zhukov, V., Mondy, S., el Heba, G.A., Cosson, V., Ellis, T.N., Ambrose, M., Wen, J., Tadege,
683 M., and Tikhonovich, I. 2012. NODULE ROOT and COCHLEATA maintain nodule development and
684 are legume orthologs of Arabidopsis BLADE-ON-PETIOLE genes. The Plant Cell 24:4498-4510.
685 Crespi, M.D., Jurkevitch, E., Poiret, M., d'Aubenton-Carafa, Y., Petrovics, G., Kondorosi, E., and Kondorosi,
686 A. 1994. enod40, a gene expressed during nodule organogenesis, codes for a non-translatable
687 RNA involved in plant growth. EMBO J 13:5099-5112.

688 Czernic, P., Gully, D., Cartieaux, F., Moulin, L., Patrel, D., Pierre, O., Fardoux, J., Chaintreuil, C., Nguyen, P.,
689 and Gressent, F. 2015. Convergent evolution of endosymbiont differentiation in Dalbergioid and
690 IRLC legumes mediated by nodule-specific cysteine-rich peptides. Plant physiology:pp.
691 00584.02015.
692 Dash, S., Cannon, E., Kalberer, S., Farmer, A., Cannon, S., Wilson, R., and Stalker, T. 2016. PeanutBase and
693 other bioinformatic resources for peanut. Peanuts: Genetics, Processing, & Utilization. AOCS
694 Press, Urbana, IL:241-253.
695 Elise, S., Etienne-Pascal, J., de Fernanda, C.-N., Gérard, D., and Julia, F. 2005. The Medicago truncatula
696 SUNN Gene Encodes a CLV1-like Leucine-rich Repeat Receptor Kinase that Regulates Nodule
697 Number and Root Length. Plant Molecular Biology 58:809-822.
698 Endre, G., Kereszt, A., Kevei, Z., Mihacea, S., Kaló, P., and Kiss, G.B. 2002. A receptor kinase gene
699 regulating symbiotic nodule development. Nature 417:962-966.
700 Fabre, S., Gully, D., Poitout, A., Patrel, D., Arrighi, J.F., Giraud, E., Czernic, P., and Cartieaux, F. 2015. Nod
701 Factor-Independent Nodulation in Aeschynomene evenia Required the Common Plant-Microbe
702 Symbiotic Toolkit. Plant Physiol 169:2654-2664.
703 Fliegmann, J., Canova, S., Lachaud, C., Uhlenbroich, S., Gasciolli, V., Pichereaux, C., Rossignol, M.,
704 Rosenberg, C., Cumener, M., and Pitorre, D. 2013. Lipo-chitooligosaccharidic symbiotic signals are
705 recognized by LysM receptor-like kinase LYR3 in the legume Medicago truncatula. ACS chemical
706 biology 8:1900-1906.
707 Fuller, C.W. (1995). Cycle sequencing with non-thermostable DNA polymerases (Google Patents).
708 Gage, D.J. 2004. Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of
709 temperate legumes. Microbiol Mol Biol Rev 68:280-300.
710 Geurts, R., and Bisseling, T. 2002. Rhizobium Nod factor perception and signalling. The Plant Cell 14:S239-
711 S249.
712 Giraud, E., Moulin, L., Vallenet, D., Barbe, V., Cytryn, E., Avarre, J.C., Jaubert, M., Simon, D., Cartieaux, F.,
713 Prin, Y., Bena, G., Hannibal, L., Fardoux, J., Kojadinovic, M., Vuillet, L., Lajus, A., Cruveiller, S.,
714 Rouy, Z., Mangenot, S., Segurens, B., Dossat, C., Franck, W.L., Chang, W.S., Saunders, E., Bruce, D.,
715 Richardson, P., Normand, P., Dreyfus, B., Pignol, D., Stacey, G., Emerich, D., Vermeglio, A.,
716 Medigue, C., and Sadowsky, M. 2007. Legumes symbioses: absence of Nod genes in
717 photosynthetic bradyrhizobia. Science 316:1307-1312.
718 Gonzalez-Rizzo, S., Crespi, M., and Frugier, F. 2006a. The Medicago truncatula CRE1 Cytokinin Receptor
719 Regulates Lateral Root Development and Early Symbiotic Interaction with Sinorhizobium meliloti.
720 The Plant Cell Online 18:2680-2693.

721 Gonzalez-Rizzo, S., Crespi, M., and Frugier, F. 2006b. The Medicago truncatula CRE1 cytokinin receptor
722 regulates lateral root development and early symbiotic interaction with Sinorhizobium meliloti.
723 Plant Cell 18:2680-2693.
724 Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L.,
725 Raychowdhury, R., and Zeng, Q. 2011. Full-length transcriptome assembly from RNA-Seq data
726 without a reference genome. Nature biotechnology 29:644-652.
727 Hakoyama, T., Niimi, K., Yamamoto, T., Isobe, S., Sato, S., Nakamura, Y., Tabata, S., Kumagai, H., Umehara,
728 Y., Brossuleit, K., Petersen, T.R., Sandal, N., Stougaard, J., Udvardi, M.K., Tamaoki, M., Kawaguchi,
729 M., Kouchi, H., and Suganuma, N. 2012. The Integral Membrane Protein SEN1 is Required for
730 Symbiotic Nitrogen Fixation in Lotus japonicus Nodules. Plant and Cell Physiology 53:225-236.
731 Hakoyama, T., Niimi, K., Watanabe, H., Tabata, R., Matsubara, J., Sato, S., Nakamura, Y., Tabata, S., Jichun,
732 L., Matsumoto, T., Tatsumi, K., Nomura, M., Tajima, S., Ishizaka, M., Yano, K., Imaizumi-Anraku,
733 H., Kawaguchi, M., Kouchi, H., and Suganuma, N. 2009. Host plant genome overcomes the lack of
734 a bacterial gene for symbiotic nitrogen fixation. Nature 462:514-517.
735 Haney, C.H., and Long, S.R. 2009. Plant flotillins are required for infection by nitrogen-fixing bacteria.
736 Proceedings of the National Academy of Sciences 107:478-483.
737 Haney, C.H., and Long, S.R. 2010. Plant flotillins are required for infection by nitrogen-fixing bacteria.
738 Proceedings of the National Academy of Sciences 107:478-483.
739 Haynes, J.G., Czymmek, K.J., Carlson, C.A., Veereshlingam, H., Dickstein, R., and Sherrier, D.J. 2004. Rapid
740 analysis of legume root nodule development using confocal microscopy. New Phytologist
741 163:661-668.
742 Heberle, H., Meirelles, G.V., da Silva, F.R., Telles, G.P., and Minghim, R. 2015. InteractiVenn: a web-based
743 tool for the analysis of sets through Venn diagrams. BMC bioinformatics 16:169.
744 Hossain, M.S., Liao, J., James, E.K., Sato, S., Tabata, S., Jurkiewicz, A., Madsen, L.H., Stougaard, J., Ross, L.,
745 and Szczyglowski, K. 2012. Lotus japonicus ARPC1 Is Required for Rhizobial Infection. Plant
746 physiology 160:917-928.
747 Howe, E.A., Sinha, R., Schlauch, D., and Quackenbush, J. 2011. RNA-Seq analysis in MeV. Bioinformatics
748 27:3209-3210.
749 Ibáñez, F., and Fabra, A. 2011. Rhizobial Nod factors are required for cortical cell division in the nodule
750 morphogenetic programme of the Aeschynomeneae legume Arachis. Plant Biology 13:794-800.
751 Imaizumi-Anraku, H., Takeda, N., Charpentier, M., Perry, J., Miwa, H., Umehara, Y., Kouchi, H., Murakami,
752 Y., Mulder, L., and Vickers, K. 2005. Plastid proteins crucial for symbiotic fungal and bacterial
753 entry into plant roots. Nature 433:527-531.

754 Jardinaud, M.-F., Boivin, S., Rodde, N., Catrice, O., Kisiala, A., Lepage, A., Moreau, S., Roux, B., Cottret, L.,
755 and Sallet, E. 2016. A laser dissection-RNAseq analysis highlights the activation of cytokinin
756 pathways by nod factors in the medicago truncatula root epidermis. Plant physiology 171:2256-
757 2276.
758 Jin, Y., Liu, H., Luo, D., Yu, N., Dong, W., Wang, C., Zhang, X., Dai, H., Yang, J., and Wang, E. 2016. DELLA
759 proteins are common components of symbiotic rhizobial and mycorrhizal signalling pathways.
760 Nature communications 7.
761 Kaló, P., Gleason, C., Edwards, A., Marsh, J., Mitra, R.M., Hirsch, S., Jakab, J., Sims, S., Long, S.R., and
762 Rogers, J. 2005. Nodulation signaling in legumes requires NSP2, a member of the GRAS family of
763 transcriptional regulators. Science 308:1786-1789.
764 Kanamori, N., Madsen, L.H., Radutoiu, S., Frantescu, M., Quistgaard, E.M., Miwa, H., Downie, J.A., James,
765 E.K., Felle, H.H., and Haaning, L.L. 2006. A nucleoporin is required for induction of Ca2+ spiking in
766 legume nodule development and essential for rhizobial and fungal symbiosis. Proceedings of the
767 National Academy of Sciences of the United States of America 103:359-364.
768 Kanehisa, M., and Goto, S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research
769 28:27-30.
770 Kang, H., Zhu, H., Chu, X., Yang, Z., Yuan, S., Yu, D., Wang, C., Hong, Z., and Zhang, Z. 2011. A Novel
771 Interaction between CCaMK and a Protein Containing the Scythe_N Ubiquitin-Like Domain in
772 Lotus japonicus. Plant physiology 155:1312-1324.
773 Kang, H., Chu, X., Wang, C., Xiao, A., Zhu, H., Yuan, S., Yang, Z., Ke, D., Xiao, S., Hong, Z., and Zhang, Z.
774 2014. A MYB coiled-coil transcription factor interacts with NSP2 and is involved in nodulation
775 inLotus japonicus. New Phytologist 201:837-849.
776 Kassaw, T., Nowak, S., Schnabel, E., and Frugoli, J.A. 2017. ROOT DETERMINED NODULATION1 is required
777 for M. truncatula CLE12, but not CLE13 peptide signaling through the SUNN receptor kinase. Plant
778 physiology:pp. 00278.02017.
779 Kawaharada, Y., Kelly, S., Nielsen, M.W., Hjuler, C.T., Gysel, K., Muszyński, A., Carlson, R., Thygesen, M.B.,
780 Sandal, N., and Asmussen, M. 2015. Receptor-mediated exopolysaccharide perception controls
781 bacterial infection. Nature 523:308-312.
782 Kent, W.J. 2002. BLAT—the BLAST-like alignment tool. Genome research 12:656-664.
783 Kistner, C., and Parniske, M. 2002. Evolution of signal transduction in intracellular symbiosis. Trends in
784 plant science 7:511-518.
785 Krusell, L. 2005. The Sulfate Transporter SST1 Is Crucial for Symbiotic Nitrogen Fixation in Lotus japonicus
786 Root Nodules. The Plant Cell Online 17:1625-1636.

787 Kundu, A., and DasGupta, M. 2017a. Silencing of putative cytokinin receptor Histidine Kinase1 inhibits
788 both inception and differentiation of root nodules in Arachis hypogaea. Molecular plant-microbe
789 interactions.
790 Kundu, A., and DasGupta, M. 2017b. Silencing of Putative Cytokinin Receptor Histidine Kinase1 Inhibits
791 Both Inception and Differentiation of Root Nodules in Arachis hypogaea. Molecular Plant-Microbe
792 Interactions:MPMI-06-17-0144-R.
793 Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. 2009. Ultrafast and memory-efficient alignment of
794 short DNA sequences to the human genome. Genome biology 10:R25.
795 Lefebvre, B., Timmers, T., Mbengue, M., Moreau, S., Hervé, C., Tóth, K., Bittencourt-Silvestre, J., Klaus, D.,
796 Deslandes, L., and Godiard, L. 2010. A remorin protein interacts with symbiotic receptors and
797 regulates bacterial infection. Proceedings of the National Academy of Sciences 107:2343-2348.
798 Lévy, J., Bres, C., Geurts, R., Chalhoub, B., Kulikova, O., Duc, G., Journet, E.-P., Ané, J.-M., Lauber, E., and
799 Bisseling, T. 2004. A putative Ca2+ and calmodulin-dependent protein kinase required for
800 bacterial and fungal symbioses. Science 303:1361-1364.
801 Madsen, E.B., Madsen, L.H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., Sato, S., Kaneko, T.,
802 Tabata, S., and Sandal, N. 2003. A receptor kinase gene of the LysM type is involved in
803 legumeperception of rhizobial signals. Nature 425:637-640.
804 Mbengue, M., Camut, S., de Carvalho-Niebel, F., Deslandes, L., Froidure, S., Klaus-Heisen, D., Moreau, S.,
805 Rivas, S., Timmers, T., Hervé, C., Cullimore, J., and Lefebvre, B. 2010. TheMedicago truncatulaE3
806 Ubiquitin Ligase PUB1 Interacts with the LYK3 Symbiotic Receptor and Negatively Regulates
807 Infection and Nodulation. The Plant Cell 22:3474-3488.
808 McDowell, J.M., Soyano, T., Kouchi, H., Hirota, A., and Hayashi, M. 2013. NODULE INCEPTION Directly
809 Targets NF-Y Subunit Genes to Regulate Essential Processes of Root Nodule Development in Lotus
810 japonicus. PLoS Genetics 9:e1003352.
811 Mergaert, P., Uchiumi, T., Alunni, B., Evanno, G., Cheron, A., Catrice, O., Mausset, A.-E., Barloy-Hubler, F.,
812 Galibert, F., and Kondorosi, A. 2006. Eukaryotic control on bacterial cell cycle and differentiation
813 in the Rhizobium–legume symbiosis. Proceedings of the National Academy of Sciences of the
814 United States of America 103:5230-5235.
815 Messinese, E., Mun, J.-H., Yeun, L.H., Jayaraman, D., Rougé, P., Barre, A., Lougnon, G., Schornack, S.,
816 Bono, J.-J., and Cook, D.R. 2007. A novel nuclear protein interacts with the symbiotic DMI3
817 calcium-and calmodulin-dependent protein kinase of Medicago truncatula. Molecular Plant-
818 Microbe Interactions 20:912-921.

819 Middleton, P.H., Jakab, J., Penmetsa, R.V., Starker, C.G., Doll, J., Kaló, P., Prabhu, R., Marsh, J.F., Mitra,
820 R.M., and Kereszt, A. 2007. An ERF transcription factor in Medicago truncatula that is essential for
821 Nod factor signal transduction. The Plant Cell 19:1221-1234.
822 Miyazawa, H., Oka-Kira, E., Sato, N., Takahashi, H., Wu, G.J., Sato, S., Hayashi, M., Betsuyaku, S.,
823 Nakazono, M., Tabata, S., Harada, K., Sawa, S., Fukuda, H., and Kawaguchi, M. 2010. The receptor-
824 like kinase KLAVIER mediates systemic regulation of nodulation and non-symbiotic shoot
825 development in Lotus japonicus. Development 137:4317-4325.
826 Murray, J.D., Karas, B.J., Sato, S., Tabata, S., Amyot, L., and Szczyglowski, K. 2007. A cytokinin perception
827 mutant colonized by Rhizobium in the absence of nodule organogenesis. Science 315:101-104.
828 Murray, J.D., Muni, R.R.D., Torres-Jerez, I., Tang, Y., Allen, S., Andriankaja, M., Li, G., Laxmi, A., Cheng, X.,
829 and Wen, J. 2011a. Vapyrin, a gene essential for intracellular progression of arbuscular
830 mycorrhizal symbiosis, is also essential for infection by rhizobia in the nodule symbiosis of
831 Medicago truncatula. The Plant Journal 65:244-252.
832 Murray, J.D., Muni, R.R.D., Torres-Jerez, I., Tang, Y., Allen, S., Andriankaja, M., Li, G., Laxmi, A., Cheng, X.,
833 Wen, J., Vaughan, D., Schultze, M., Sun, J., Charpentier, M., Oldroyd, G., Tadege, M., Ratet, P.,
834 Mysore, K.S., Chen, R., and Udvardi, M.K. 2011b. Vapyrin, a gene essential for intracellular
835 progression of arbuscular mycorrhizal symbiosis, is also essential for infection by rhizobia in the
836 nodule symbiosis of Medicago truncatula. The Plant Journal 65:244-252.
837 Nishimura, R., Ohmori, M., Fujita, H., and Kawaguchi, M. 2002. A Lotus basic leucine zipper protein with a
838 RING-finger motif negatively regulates the developmental program of nodulation. Proceedings of
839 the National Academy of Sciences 99:15206-15210.
840 Oldroyd, G.E., and Downie, J.A. 2004. Calcium, kinases and nodulation signalling in legumes. Nature
841 reviews. Molecular cell biology 5:566.
842 Oldroyd, G.E., and Downie, J.A. 2006. Nuclear calcium changes at the core of symbiosis signalling. Current
843 opinion in plant biology 9:351-357.
844 Op den Camp, R.H., De Mita, S., Lillo, A., Cao, Q., Limpens, E., Bisseling, T., and Geurts, R. 2011. A
845 phylogenetic strategy based on a legume-specific whole genome duplication yields symbiotic
846 cytokinin type-A response regulators. Plant Physiol 157:2013-2022.
847 Osipova, M.A., Mortier, V., Demchenko, K.N., Tsyganov, V.E., Tikhonovich, I.A., Lutova, L.A., Dolgikh, E.A.,
848 and Goormachtig, S. 2012. WUSCHEL-RELATED HOMEOBOX5 Gene Expression and Interaction of
849 CLE Peptides with Components of the Systemic Control Add Two Pieces to the Puzzle of
850 Autoregulation of Nodulation. Plant physiology 158:1329-1341.

851 Pawlowski, K., and Bisseling, T. 1996. Rhizobial and actinorhizal symbioses: what are the shared features?
852 The Plant Cell 8:1899.
853 Peng, Z., Liu, F., Wang, L., Zhou, H., Paudel, D., Tan, L., Maku, J., Gallo, M., and Wang, J. 2017.
854 Transcriptome profiles reveal gene regulation of peanut (Arachis hypogaea L.) nodulation.
855 Scientific Reports 7:40066.
856 Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. 2015. StringTie
857 enables improved reconstruction of a transcriptome from RNA-seq reads. Nature biotechnology
858 33:290-295.
859 Radutoiu, S., Madsen, L.H., Madsen, E.B., Felle, H.H., Umehara, Y., Grønlund, M., Sato, S., Nakamura, Y.,
860 Tabata, S., and Sandal, N. 2003. Plant recognition of symbiotic bacteria requires two LysM
861 receptor-like kinases. Nature 425:585-592.
862 Robinson, M.D., McCarthy, D.J., and Smyth, G.K. 2010. edgeR: a Bioconductor package for differential
863 expression analysis of digital gene expression data. Bioinformatics 26:139-140.
864 Saha, S., Dutta, A., Bhattacharya, A., and DasGupta, M. 2014. Intracellular catalytic domain of symbiosis
865 receptor kinase hyperactivates spontaneous nodulation in absence of rhizobia. Plant Physiol
866 166:1699-1708.
867 Saito, K., Yoshikawa, M., Yano, K., Miwa, H., Uchida, H., Asamizu, E., Sato, S., Tabata, S., Imaizumi-Anraku,
868 H., and Umehara, Y. 2007. NUCLEOPORIN85 is required for calcium spiking, fungal and bacterial
869 symbioses, and seed production in Lotus japonicus. The Plant Cell 19:610-624.
870 Schauser, L., Roussis, A., Stiller, J., and Stougaard, J. 1999a. A plant regulator controlling development of
871 symbiotic root nodules. Nature 402:191-195.
872 Schauser, L., Roussis, A., Stiller, J., and Stougaard, J. 1999b. A plant regulator controlling development of
873 symbiotic root nodules. Nature 402:191-195.
874 Shimomura, K., Nomura, M., Tajima, S., and Kouchi, H. 2006. LjnsRING, a Novel RING Finger Protein, is
875 Required for Symbiotic Interactions Between

876 Mesorhizobium loti

877 and

878 Lotus japonicus. Plant and Cell Physiology 47:1572-1581.
879 Singh, S., Katzer, K., Lambert, J., Cerri, M., and Parniske, M. 2014. CYCLOPS, a DNA-binding transcriptional
880 activator, orchestrates symbiotic root nodule development. Cell Host & Microbe 15:139-152.

881 Sinharoy, S., and DasGupta, M. 2009. RNA interference highlights the role of CCaMK in dissemination of
882 endosymbionts in the Aeschynomeneae legume Arachis. Mol Plant Microbe Interact 22:1466-
883 1475.
884 Sinharoy, S., Torres-Jerez, I., Bandyopadhyay, K., Kereszt, A., Pislariu, C.I., Nakashima, J., Benedito, V.A.,
885 Kondorosi, E., and Udvardi, M.K. 2013. The C2H2 transcription factor regulator of symbiosome
886 differentiation represses transcription of the secretory pathway gene VAMP721a and promotes
887 symbiosome development in Medicago truncatula. Plant Cell 25:3584-3601.
888 Smit, P., Raedts, J., Portyanko, V., Debelle, F., Gough, C., Bisseling, T., and Geurts, R. 2005. NSP1 of the
889 GRAS protein family is essential for rhizobial Nod factor-induced transcription. Science 308:1789-
890 1791.
891 Smit, P., Limpens, E., Geurts, R., Fedorova, E., Dolgikh, E., Gough, C., and Bisseling, T. 2007. Medicago
892 LYK3, an entry receptor in rhizobial nodulation factor signaling. Plant physiology 145:183-191.
893 Sprent, J.I., and James, E.K. 2007. Legume evolution: where do nodules and mycorrhizas fit in? Plant
894 Physiol 144:575-581.
895 Stracke, S., Kistner, C., Yoshida, S., Mulder, L., Sato, S., Kaneko, T., Tabata, S., Sandal, N., Stougaard, J., and
896 Szczyglowski, K. 2002. A plant receptor-like kinase required for both bacterial and fungal
897 symbiosis. Nature 417:959-962.
898 Suzaki, T., Ito, M., Yoro, E., Sato, S., Hirakawa, H., Takeda, N., and Kawaguchi, M. 2014.
899 Endoreduplication-mediated initiation of symbiotic organ development in Lotus japonicus.
900 Development 141:2441-2445.
901 Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 2013. MEGA6: molecular evolutionary
902 genetics analysis version 6.0. Molecular biology and evolution 30:2725-2729.
903 Tirichine, L., Imaizumi-Anraku, H., Yoshida, S., Murakami, Y., Madsen, L.H., Miwa, H., Nakagawa, T.,
904 Sandal, N., Albrektsen, A.S., Kawaguchi, M., Downie, A., Sato, S., Tabata, S., Kouchi, H., Parniske,
905 M., Kawasaki, S., and Stougaard, J. 2006. Deregulation of a Ca2+/calmodulin-dependent kinase
906 leads to spontaneous nodule development. Nature 441:1153-1156.
907 Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M.J., Salzberg, S.L., Wold, B.J.,
908 and Pachter, L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated
909 transcripts and isoform switching during cell differentiation. Nature biotechnology 28:511-515.
910 Van de Velde, W., Zehirov, G., Szatmari, A., Debreczeny, M., Ishihara, H., Kevei, Z., Farkas, A., Mikulass, K.,
911 Nagy, A., and Tiricz, H. 2010. Plant peptides govern terminal differentiation of bacteria in
912 symbiosis. Science 327:1122-1126.

913 Varma Penmetsa, R., Uribe, P., Anderson, J., Lichtenzveig, J., Gish, J.-C., Nam, Y.W., Engstrom, E., Xu, K.,
914 Sckisel, G., Pereira, M., Baek, J.M., Lopez-Meyer, M., Long, S.R., Harrison, M.J., Singh, K.B., Kiss,
915 G.B., and Cook, D.R. 2008. TheMedicago truncatulaortholog of Arabidopsis EIN2,sickle, is a
916 negative regulator of symbiotic and pathogenic microbial associations. The Plant Journal 55:580-
917 595.
918 Verdier, J., Torres-Jerez, I., Wang, M., Andriankaja, A., Allen, S.N., He, J., Tang, Y., Murray, J.D., and
919 Udvardi, M.K. 2013. Establishment of the Lotus japonicus Gene Expression Atlas (LjGEA) and its
920 use to explore legume seed maturation. The Plant Journal 74:351-362.
921 Vernie, T., Moreau, S., de Billy, F., Plet, J., Combier, J.P., Rogers, C., Oldroyd, G., Frugier, F., Niebel, A., and
922 Gamas, P. 2008a. EFD Is an ERF transcription factor involved in the control of nodule number and
923 differentiation in Medicago truncatula. Plant Cell 20:2696-2713.
924 Vernie, T., Moreau, S., de Billy, F., Plet, J., Combier, J.P., Rogers, C., Oldroyd, G., Frugier, F., Niebel, A., and
925 Gamas, P. 2008b. EFD Is an ERF Transcription Factor Involved in the Control of Nodule Number
926 and Differentiation in Medicago truncatula. The Plant Cell Online 20:2696-2713.
927 Wang, D., Griffitts, J., Starker, C., Fedorova, E., Limpens, E., Ivanov, S., Bisseling, T., and Long, S. 2010. A
928 nodule-specific protein secretory pathway required for nitrogen-fixing symbiosis. Science
929 327:1126-1129.
930 Yano, K., Yoshida, S., Müller, J., Singh, S., Banba, M., Vickers, K., Markmann, K., White, C., Schuller, B., and
931 Sato, S. 2008. CYCLOPS, a mediator of symbiotic intracellular accommodation. Proceedings of the
932 National Academy of Sciences 105:20540-20545.
933 Yano, K., Shibata, S., Chen, W.L., Sato, S., Kaneko, T., Jurkiewicz, A., Sandal, N., Banba, M.,
934 Imaizumi-Anraku, H., and Kojima, T. 2009. CERBERUS, a novel U-box protein containing WD-40
935 repeats, is required for formation of the infection thread and nodule development in the legume–
936 Rhizobium symbiosis. The Plant Journal 60:168-180.
937 Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., and Bolund, L. 2006. WEGO: a
938 web tool for plotting GO annotations. Nucleic acids research 34:W293-W297.
939 Yendrek, C.R., Lee, Y.-C., Morris, V., Liang, Y., Pislariu, C.I., Burkart, G., Meckfessel, M.H., Salehin, M.,
940 Kessler, H., Wessler, H., Lloyd, M., Lutton, H., Teillet, A., Sherrier, D.J., Journet, E.-P., Harris, J.M.,
941 and Dickstein, R. 2010. A putative transporter is essential for integrating nutrient and hormone
942 signaling with lateral root growth and nodule development inMedicago truncatula. The Plant
943 Journal 62:100-112.

944 Yokota, K., Fukai, E., Madsen, L.H., Jurkiewicz, A., Rueda, P., Radutoiu, S., Held, M., Hossain, M.S.,
945 Szczyglowski, K., and Morieri, G. 2009. Rearrangement of actin cytoskeleton mediates invasion of
946 Lotus japonicus roots by Mesorhizobium loti. The Plant Cell 21:267-284.
947 Yoon, H.J., Hossain, M.S., Held, M., Hou, H., Kehl, M., Tromas, A., Sato, S., Tabata, S., Andersen, S.U.,
948 Stougaard, J., Ross, L., and Szczyglowski, K. 2014. Lotus japonicus SUNERGOS1encodes a
949 predicted subunit A of a DNA topoisomerase VI that is required for nodule differentiation and
950 accommodation of rhizobial infection. The Plant Journal 78:811-821.
951 Yuan, S., Zhu, H., Gou, H., Fu, W., Liu, L., Chen, T., Ke, D., Kang, H., Xie, Q., Hong, Z., and Zhang, Z. 2012. A
952 Ubiquitin Ligase of Symbiosis Receptor Kinase Involved in Nodule Organogenesis. Plant physiology
953 160:106-117.

954

955

956

957

958

959

960

961

962

963

964

965