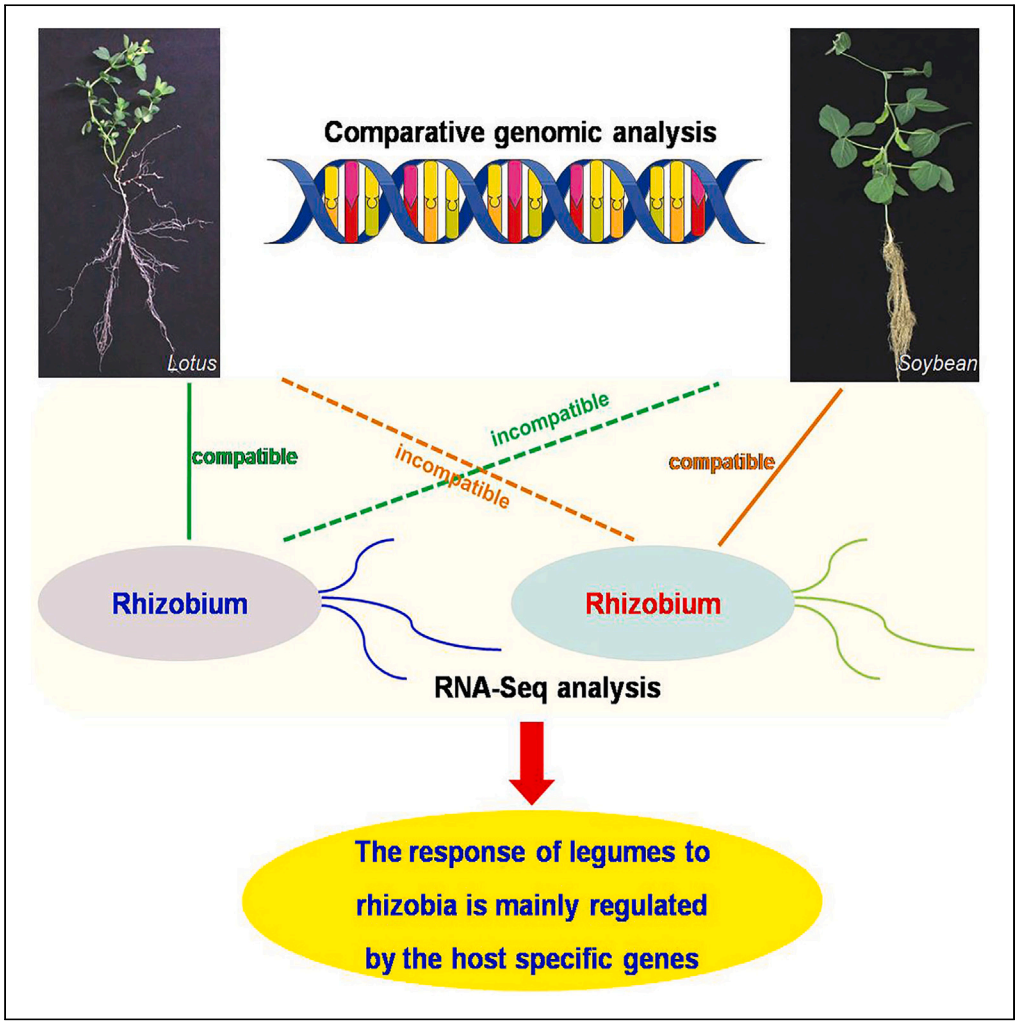
iScience

CellPress
OPEN ACCESS

Article

Comparative genomic and transcriptomic analyses provide new insight into symbiotic host specificity



Songli Yuan, Piao Leng, Yong Feng, ..., Hongli Yang, Haifeng Chen, Xinan Zhou

songliyuan@caas.cn (S.Y.)
chenhaifeng@caas.cn (H.C.)

Highlights

The response of legumes to rhizobia is mainly regulated by the host specific genes

GO terms and KEGG pathways distributions vary in different symbiotic systems

Host specific genes account for the majority of SNF genes and resistant genes

New SNF-related genes were identified from the selected SNF-related WGCNA modules

Yuan et al., iScience 27, 110207
July 19, 2024
https://doi.org/10.1016/j.isci.2024.110207

iScience

CellPress
OPEN ACCESS

Article

Comparative genomic and transcriptomic analyses provide new insight into symbiotic host specificity

Songli Yuan,[1,3,*] Piao Leng,[1] Yong Feng,[2] Fuxiao Jin,[1] Hui Zhang,[1] Chanjuan Zhang,[1] Yi Huang,[1] Zhihui Shan,[1] Zhonglu Yang,[1] Qingnan Hao,[1] Shuilian Chen,[1] Limiao Chen,[1] Dong Cao,[1] Wei Guo,[1] Hongli Yang,[1] Haifeng Chen,[1,*] and Xinan Zhou[1]

SUMMARY

Host specificity plays important roles in expanding the host range of rhizobia, while the genetic information responsible for host specificity remains largely unexplored. In this report, the roots of four symbiotic systems with notable different symbiotic phenotypes and the control were studied at four different post-inoculation time points by RNA sequencning (RNA-seq). The differentially expressed genes (DEGs) were divided into "found only in soybean or *Lotus*," "only expressed in soybean or *Lotus*," and "expressed in both hosts" according to the comparative genomic analysis. The distributions of enriched function ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways vary significantly in different symbiotic systems. Host specific genes account for the majority of the DEGs involved in response to stimulus, associated with plant-pathogen interaction pathways, and encoding resistance (R) proteins, the symbiotic nitrogen fixation (SNF) proteins and the target proteins in the SNF-related modules. Our findings provided molecular candidates for better understanding the mechanisms of symbiotic host-specificity.

INTRODUCTION

Engineering cereal crops with the capability of fix nitrogen like legumes or associate with nitrogen-fixing microbiomes could address the problems caused by excessive use of synthetic nitrogen fertilizer in agriculture.[1,2] The establishment of fully function nitrogen-fixing symbiosis in cereals will require nitrogen-fixing bacterial infection, nodule organogenesis and normal nitrogenase activity.[1,3] With the development of synthetic biology, various efforts have been undertaken to engineer Nod factor perception, the activation and nodulation-specific outputs of the common symbiotic signaling (SYM) pathway, and functional nitrogenase enzymes into cereal crops.[1–5] However, it is currently unclear whether these imported symbiotic system genes are compatible with the cereal host, meaning that host specificity may play key roles in the efficiency of this cross-kingdom collaboration.

Symbiotic host specificity always associated with distinct nodulation phenotype and/or symbiotic effects[6–8] and has led to the definition of different legume-rhizobium associations, for example, *Mesorhizobium japonicum* MAFF303099[9] only forms determinant-type globular nodules and performs nitrogen fixing on several host plants of *Lotus*,[10] *Mesorhizobium huakuii* 7653R forms specific symbiosis with *Astragalus sinicus*,[11] *Sinorhizobium meliloti* can form indeterminant-type nodules with alfalfa and *Medicago truncatula*,[12] and so on. Despite recent advances in our understanding of the symbiotic specificity between a legume plant and its different corresponding symbiotic rhizobia,[7,8,13–15] the host control of symbiotic specificity between different rhizobia corresponding to different legume plants remains poorly understood.

The symbiotic specificity is determined by a fine-tuned exchange of molecular signals between a host root and its inoculated rhizobial strains.[16] These signals include rhizobia utilizes Nod factors,[17,18] surface polysaccharides[6,19] and secreted proteins/type III secretion system (T3SS).[20–22] It has been proposed that Nod factors,[23,24] surface polysaccharides,[25,26] and T3SS[20,27] play roles in host defense responses, a feature that is shared by pathogenic and symbiotic bacteria. In contrast to pathogens, these rhizobial signals cannot cause disease and elicit the hypersensitive response in hosts.[24,28–30] For these rhizobial signals, several related receptor proteins or effector proteins were found in host plants,[12,13,31–34] while the host control of corresponding recognition mechanisms that control the compatibility of the legume-rhizobium interaction is yet unknown. To unravel such mechanisms, it is critical to investigate the genetic information responsible for host specificity.

[1]Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute of Chinese Academy of Agricultural Sciences, Wuhan 430062, China
[2]School of the Life Sciences, Jiangsu University, 301 Xuefu Road, Zhenjiang, Jiangsu Province 212013, China
[3]Lead contact
*Correspondence: songliyuan@caas.cn (S.Y.), chenhaifeng@caas.cn (H.C.)
https://doi.org/10.1016/j.isci.2024.110207

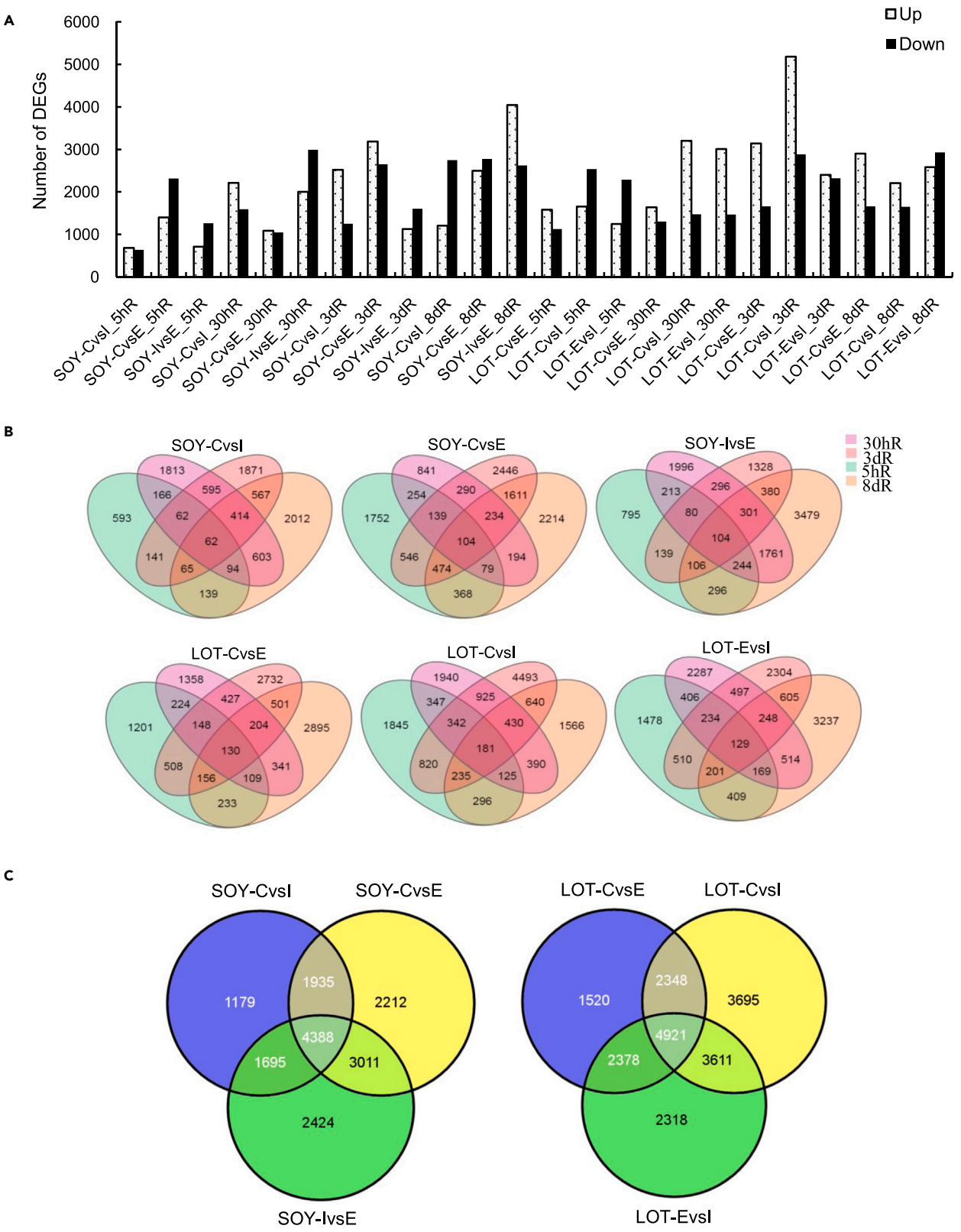

A
Up
Down
Number of DEGs
6000
5000
4000
3000
2000
1000
0
SOY-CvsI_5hR
SOY-CvsE_5hR
SOY-IvsE_5hR
SOY-CvsI_30hR
SOY-CvsE_30hR
SOY-IvsE_30hR
SOY-CvsI_3dR
SOY-CvsE_3dR
SOY-IvsE_3dR
SOY-CvsI_8dR
SOY-CvsE_8dR
SOY-IvsE_8dR
LOT-CvsE_5hR
LOT-CvsI_5hR
LOT-EvsI_5hR
LOT-CvsE_30hR
LOT-CvsI_30hR
LOT-EvsI_30hR
LOT-CvsE_3dR
LOT-CvsI_3dR
LOT-EvsI_3dR
LOT-CvsE_8dR
LOT-CvsI_8dR
LOT-EvsI_8dR
B
SOY-CvsI
SOY-CvsE
SOY-IvsE
30hR
3dR
5hR
8dR
LOT-CvsE
LOT-CvsI
LOT-EvsI
C
SOY-CvsI
SOY-CvsE
SOY-IvsE
LOT-CvsE
LOT-CvsI
LOT-EvsI

Figure 1. Genes differentially expressed in soybean and *Lotus* roots responding to *B. diazoefficiens* 113-2 or *M. japonicum* MAFF303099

(A) Genes differentially expressed in soybean and *Lotus* roots at different time points were separated into two groups according to whether they were significantly upregulated or downregulated.

(B) The numbers of differentially expressed genes in different gene sets in each group. Four different post-inoculation time points (5 h, 30 h, 3 days, and 8 days) are included and the division of DEGs into different gene sets depends on which time points (one or more) the DEGs were identified.

(C) The numbers of differentially expressed genes in different gene sets in soybean roots or *Lotus* roots. The division of DEGs into different gene sets depends on which groups (one or more) the DEGs were identified. SOY-CvsI, uninoculated control versus MAFF303099 (ineffective) in soybean roots; SOY-CvsE, uninoculated control versus 113-2 (effective) in soybean roots; SOY-IvsE, MAFF303099 (ineffective) versus 113-2 (effective) in soybean roots; LOT-CvsE, uninoculated control versus MAFF303099 (effective) in *Lotus* roots; LOT-CvsI, uninoculated control versus 113-2 (ineffective) in *Lotus* roots; LOT-EvsI, MAFF303099 (effective) versus 113-2 (ineffective) in *Lotus* roots. SOY, soybean; LOT, *Lotus*; C, control; I, ineffective inoculant; E, effective inoculant.

In the present study, we selected two legumes soybean and *Lotus japonicus* with genetic background of nodulation and nitrogen fixation, and performed comparative genomic and transcriptomic analyses to investigate the genetic determinants of host specificity. Firstly, we investigated the molecular events in the roots of four symbiotic systems (*Bradyrhizobium diazoefficiens* 113-2-soybean, *B. diazoefficiens* 113-2-*L. japonicus*, *M. japonicum* MAFF303099- soybean, and *M. japonicum* MAFF303099-*L. japonicus*); secondly, we identified a large number of differentially expressed genes (DEGs), and divided these DEGs into "DEGs found only in soybean," "DEGs found only in *Lotus*," "DEGs only expressed in soybean," "DEGs only expressed in *Lotus*," and "DEGs expressed in both hosts" gene sets according to the comparative genomic analysis; thirdly, we performed function ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis and gene co-expression network analysis; finally, we verified the RNA sequencing (RNA-seq) results by RT-qPCR and analyzed the DEGs involved in response to stimulus, associated with plant-pathogen interaction pathways, and encoding R proteins, the symbiotic nitrogen fixation (SNF) proteins and the target proteins in the SNF-related modules. Our results provided candidates genetic information responsible for symbiotic host specificity, and supplied fundamental clues to study the genetic determinants of non-legume-rhizobium symbiosis.

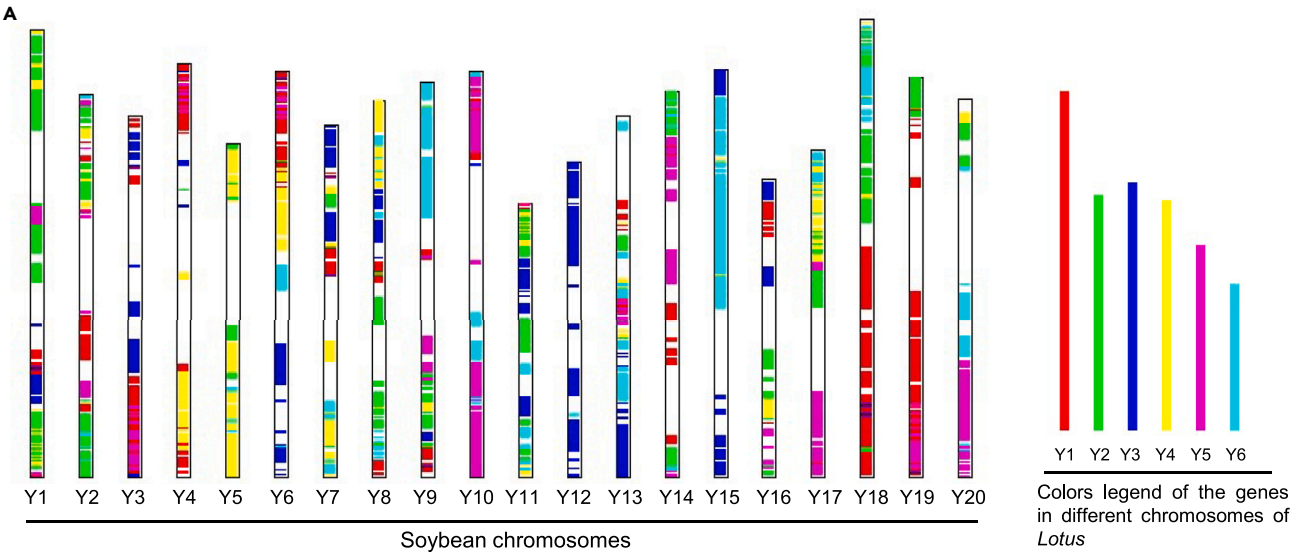
RESULTS

Identification of DEGs

M. japonicum MAFF303099 only forms specific symbiosis with several host plants of *Lotus*,[10] and *B. diazoefficiens* 113-2 is a highly efficient rhizobium of soybean Tian long No.1.[8,35] To investigate the causes of these different symbiotic phenotypes, RNA-seq was performed and the detailed information is shown in the materials and methods. The gene information, expression FPKM values and the other annotation information for all the detected genes in soybean and *Lotus* roots are shown in the core tables (Tables S1 and S2). The RT-qPCR analysis was used to verify the RNA-seq results, and the results agreed with the transcriptional profile data for 56 out of 72 (about 78%) data points (Figure S1). The numbers of upregulated and downregulated DEGs in each comparison are shown in Figure 1A. The DEG number of uninoculated control versus 113-2 (effective) in soybean roots (SOY-CvsE) was higher than uninoculated control versus MAFF303099 (ineffective) in soybean roots (SOY-CvsI), while the number of DEG of uninoculated control versus 113-2 (ineffective) in *Lotus* roots (LOT-CvsI) was higher than uninoculated control versus MAFF303099 (effective) in *Lotus* roots (LOT-CvsE), indicating the beginning of a series of new processes (not just nodulation), and more differential gene expression responses in soybean and *L. japonicus* roots to *B. diazoefficiens* 113-2 than *M. japonicum* MAFF303099. The numbers of DEGs found at one or more time points in the six groups are shown in Figure 1B. Among these DEGs, 710 genes were consistently found at the four time points (62, 104, 104, 130, 181, and 129 in SOY-CvsI, SOY-CvsE, SOY-IvsE, LOT-CvsE, LOT-CvsI, and LOT-EvsI, respectively). The DEG numbers of each gene set among SOY-CvsI, SOY-CvsE, and SOY-IvsE or LOT-CvsE, LOT-CvsI, and LOT-EvsI are shown, and a total of 16,844 and 20,791 DEGs in soybean and *L. japonicus* root samples, respectively (Figure 1C). The detailed gene ID information of DEGs is shown in Table S3. Among these DEGs, 2,212 soybean DEGs and 1,520 *Lotus* DEGs were found only in SOY-CvsE and LOT-CvsE, respectively, indicating that these DEGs may mainly play roles in nodule symbiosis. 1,935 soybean DEGs were found in both SOY-CvsI and SOY-CvsE, but not in SOY-IvsE, and 2,348 *Lotus* DEGs were found in both LOT-CvsI and LOT-CvsE, but not in LOT-IvsE, suggesting that these DEGs may mainly play roles in general response to rhizobia independent of nodule symbiosis. The gene ID information of the aforementioned DEGs is shown in Table S4.

Orthologs analysis of DEGs in soybean and *L. japonicus* root samples

The recent genome assemblies include 1017.57 Mb and 394.46 Mb for soybean and *L. japonicus*, respectively, with the *L. japonicus* genome being 61.2% smaller than the soybean one. We identified a set of 27,982 orthologous pairs between soybean and *L. japonicus* genomes, including 24,167 and 13,461 genes in soybean and *L. japonicus*, respectively (Figure 2A; Table S5). Then we divided the aforementioned DEGs into five gene sets: (1) the DEGs that have no homologous genes in the other legume ("DEGs found only in soybean" or "DEGs found only in *Lotus*"); (2) the DEGs that have homologous genes in the other legume, while the homologous genes have no differential expression in the RNA-seq analysis ("DEGs only expressed in soybean" or "DEGs only expressed in *Lotus*"); (3) the DEGs that have homologous genes in the other legume, and the homologous genes also have differential expression in the RNA-seq analysis ("DEGs expressed in both hosts"). The numbers of the "DEGs found only in soybean," "DEGs found only in *Lotus*," "DEGs only expressed in soybean," "DEGs only expressed in *Lotus*," and "DEGs expressed in both hosts" gene sets at each time point of the six groups are shown in Figure 2B, and the ID information of these DEGs is shown in Table S6. The "DEGs found only in soybean" and "DEGs found only in *Lotus*" gene sets account for the majority in all of the root samples, especially over three-quarters in *Lotus* root samples, suggesting that the response to rhizobia is mainly regulated by the host specific genes.



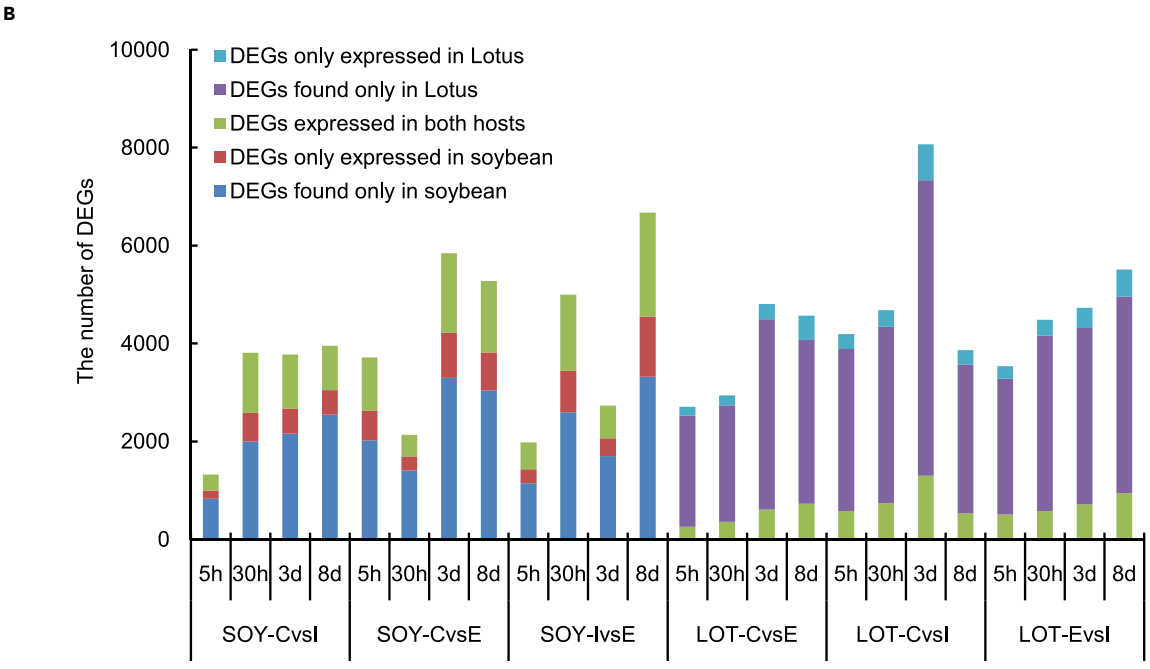


Figure 2. Orthologs analysis of DEGs in soybean and *L. japonicus*

(A) Synteny analysis of the orthologs of DEGs groupings in soybean and *L. japonicus*. The different colors on soybean chromosomes represent homologous genes of the genes in different chromosomes of *L. japonicus*.

(B) Statistics of different types (DEGs found only in soybean, DEGs only expressed in soybean, DEGs expressed in both hosts, DEGs found only in *Lotus*, and DEGs only expressed in *Lotus*) of DEGs at different time points of each group in soybean (SOY-CvsI, SOY-CvsE, and SOY-IvsE) and *L. japonicus* (LOT-CvsE, LOT-CvsI, and LOT-EvsI).

Function ontology enrichment analysis and DEGs involved in response to stimulus

To evaluate the potential functions of the DEGs between different symbiotic systems of soybean or *Lotus* roots, an internationally standardized gene function classification system, Gene Ontology (GO), was used to classify these DEGs to different terms. Only four enriched GO function

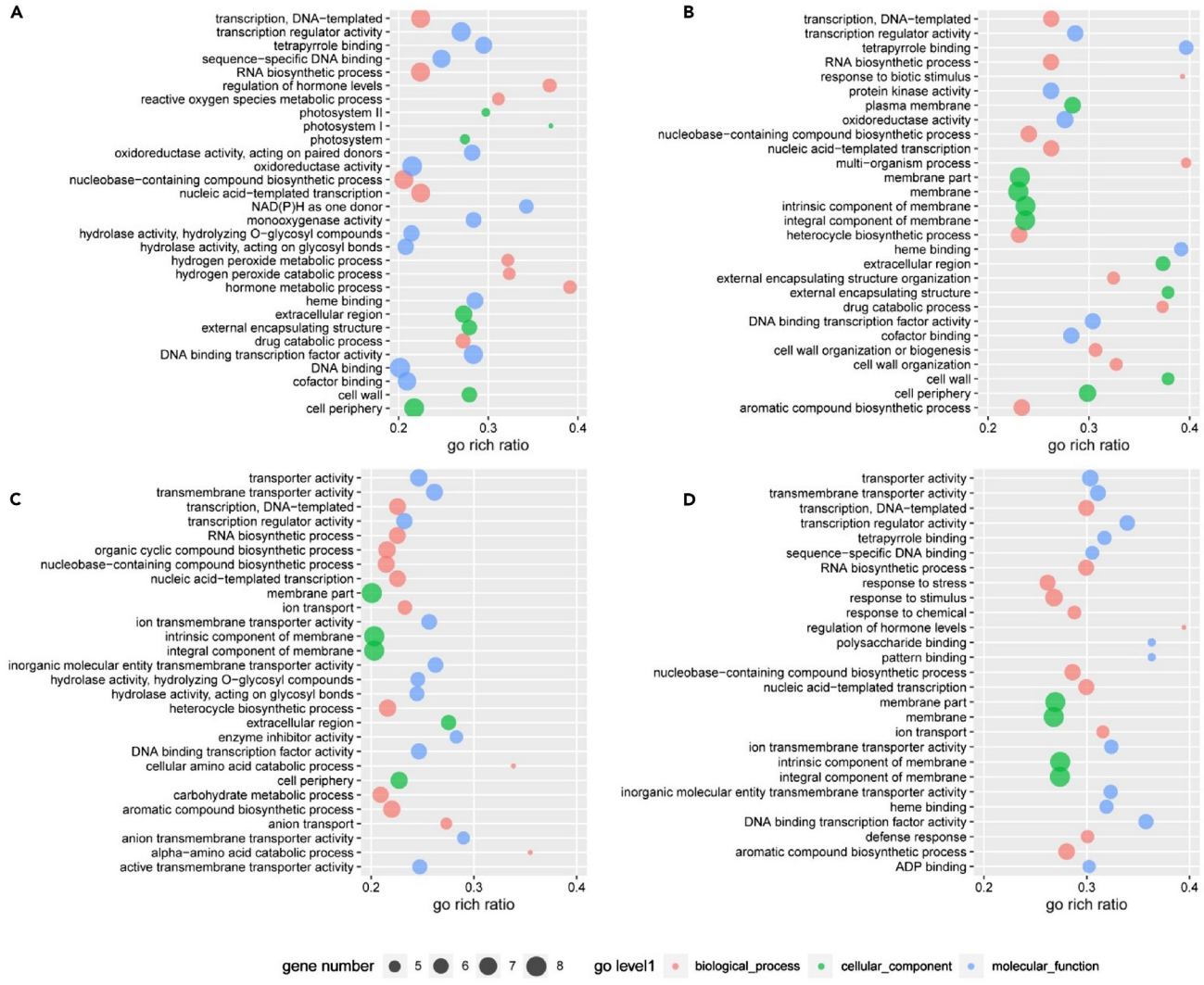


Figure 3. Gene ontology (GO) enrichment analyses of the "signature" genes in different groups

For each group, the smallest Q-value TOP 15 GO terms of each root sample (at each time point) were selected, and then the GO terms with *p* value less than 0.05 were used to perform GO term analysis, and the same GO terms from different root samples were integrated in each group.

(A) SOY-CvsI.

(B) SOY-CvsE.

(C) LOT-CvsE.

(D) LOT-CvsI.

terms were indicated at all the four groups (Figure 3), revealing high shift in the distribution of enriched GO terms among different symbiotic systems. Additionally, there are no unique GO function terms for effective or ineffective nodulation combination in soybean and *Lotus*. For example, in soybean effective nodulation combination (SOY-CvsE, Figure 3B), the cellular components associated with the DEGs mainly focused on membrane part, membrane, intrinsic or integral component of membrane, which is very different from the ineffective nodulation combination (SOY-CvsI, Figure 3A). While in *L. japonicus* root samples, no significant difference in the cellular components (Figures 3C and 3D).

Rhizobia have been shown to adopt a pathogenic system that stimulates their legume hosts to initiate symbiotic programs,[20] to evaluate the relative signaling events, the DEGs involved in response to stimulus and signaling were analyzed in more detail. Due to all of the signaling-related DEGs involved in response to stimulus, we here focused on the DEGs involved in response to stimulus (Table 1; Table S7). The numbers of DEGs involved in response to stimulus in the five gene sets ("DEGs found only in soybean," "DEGs found only in *Lotus*," "DEGs only expressed in soybean," "DEGs only expressed in *Lotus*," and "DEGs expressed in both hosts") of root samples are shown in Table 1, and the detailed gene ID information is shown in Table S7. The DEG numbers of roots inoculated with 113-2 (SOY-CvsE and LOT-CvsI) was higher than that with *M. japonicum* MAFF303099 (SOY- CvsI and LOT- CvsE) in soybean and *L. japonicus*, respectively, and the "DEGs found only in soybean" and "DEGs only expressed in *Lotus*" gene sets account for the majority number (Table 1).

Table 1. The numbers of DEGs involved in response to stimulus in the five gene sets of the root samples

	The numbers of DEGs								
	SOY-CvsI			SOY-CvsE			SOY-IvsE		
	DEGs found only in soybean	DEGs only expressed in soybean	DEGs expressed in both hosts	DEGs found only in soybean	DEGs only expressed in soybean	DEGs expressed in both hosts	DEGs found only in soybean	DEGs only expressed in soybean	DEGs expressed in both hosts
5 h	82	14	41	172	59	112	89	22	63
30 h	192	63	119	116	21	52	248	95	159
3 days	177	56	102	279	94	153	161	32	73
8 days	191	58	99	240	75	142	330	110	225
	LOT-CvsE			LOT-CvsI			LOT-EvsI		
	DEGs found only in *Lotus*	DEGs only expressed in *Lotus*	DEGs expressed in both hosts	DEGs found only in *Lotus*	DEGs only expressed in *Lotus*	DEGs expressed in both hosts	DEGs found only in *Lotus*	DEGs only expressed in *Lotus*	DEGs expressed in both hosts
5 h	92	8	24	219	25	51	186	20	43
30 h	124	14	32	201	26	54	236	2	11
3 days	205	22	61	351	71	126	191	31	71
8 days	192	30	52	257	5	12	231	41	83

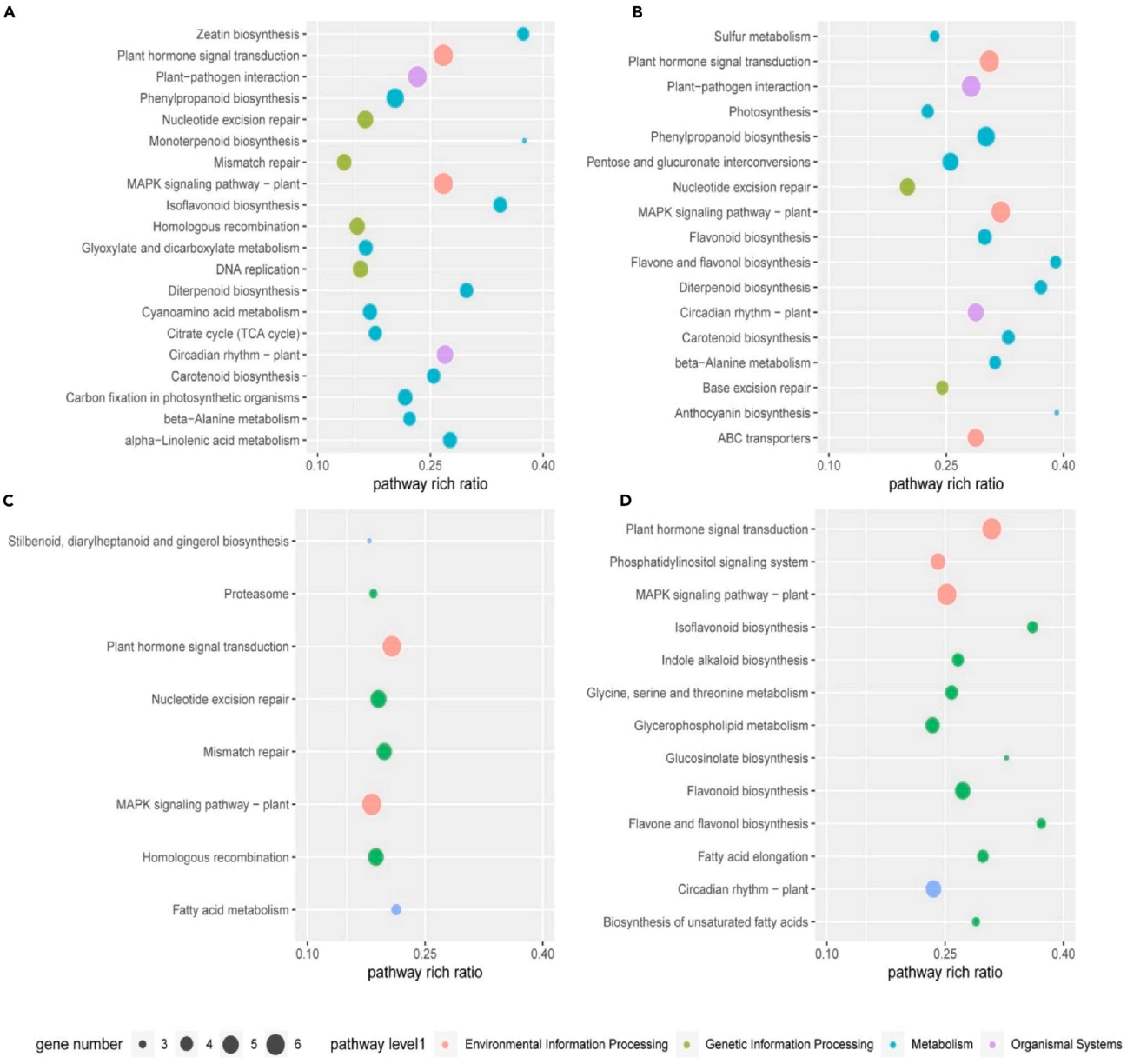


Figure 4. KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analyses of the "signature" genes in different groups

For each group, the smallest Q-value TOP 15 KEGG terms of each root sample (at each time point) were selected, and then the KEGG terms with *p* value less than 0.05 were used to perform KEGG terms analysis, and the same KEGG terms from different root samples were integrated in each group.

(A) SOY-CvsI.

(B) SOY-CvsE.

(C) LOT-CvsE.

(D) LOT-CvsI.

KEGG pathway enrichment analysis and DEGs associated with plant-pathogen interaction pathways

KEGG is the major public database for pathway enrichment analysis[36] and the enriched KEGG pathway subgroups associated with the DEGs between different symbiotic systems of soybean or *Lotus* roots are shown in Figure 4. Similar to GO enrichment analysis, high shift is existed in the distribution of enriched KEGG pathways among different symbiotic systems, and also no unique KEGG pathways for effective or ineffective nodulation combination in soybean and *Lotus* (Figure 4).

In the absence of NF signal, legume-derived flavonoid also can induce the pathogenic type III secretion system (T3SS) of rhizobia, which injects effector proteins into their legume hosts to initiate symbiotic programmes.[20,22] To explore the differential cell defense responses

Table 2. The numbers of DEGs associated with plant-pathogen interaction pathways in the five gene sets of the root samples

	The numbers of DEGs					
	SOY-CvsI			SOY-CvsE		
	DEGs found only in soybean	DEGs only expressed in soybean	DEGs expressed in both hosts	DEGs found only in soybean	DEGs only expressed in soybean	DEGs expressed in both hosts
5 h	26	6	5	38	14	16
30 h	67	18	32	32	4	12
3 days	61	16	25	102	23	39
8 days	69	11	27	89	20	32
	LOT-CvsE			LOT-CvsI		
	DEGs found only in *Lotus*	DEGs only expressed in *Lotus*	DEGs expressed in both hosts	DEGs found only in *Lotus*	DEGs only expressed in *Lotus*	DEGs expressed in both hosts
5 h	29	1	0	55	3	11
30 h	45	5	2	54	6	9
3 days	71	4	15	114	9	31
8 days	52	5	11	48	3	14

between soybean or *L. japonicus* roots inoculated with rhizobia strains 113-2 and *M. japonicum* MAF303099, the plant-pathogen interaction KEGG pathway was analyzed in more detail (Table 2; Table S8). The numbers of DEGs associated with plant-pathogen interaction pathways in the five gene sets ("DEGs found only in soybean," "DEGs found only in *Lotus*," "DEGs only expressed in soybean," "DEGs only expressed in *Lotus*," and "DEGs expressed in both hosts") of root samples are shown in Table 2, and the detailed information for these DEGs are shown in Table S8. The DEG numbers in the "DEGs found only in soybean" and "DEGs only expressed in *Lotus*" gene sets in root samples account for the majority, and the DEG numbers of roots inoculated with 113-2 (SOY-CvsE and LOT-CvsI) was higher than that with *M. japonicum* MAFF303099 (SOY-CvsI and LOT-CvsE) in soybean and *L. japonicus*, respectively (Table 2), which is similar to the DEGs involved in response to stimulus.

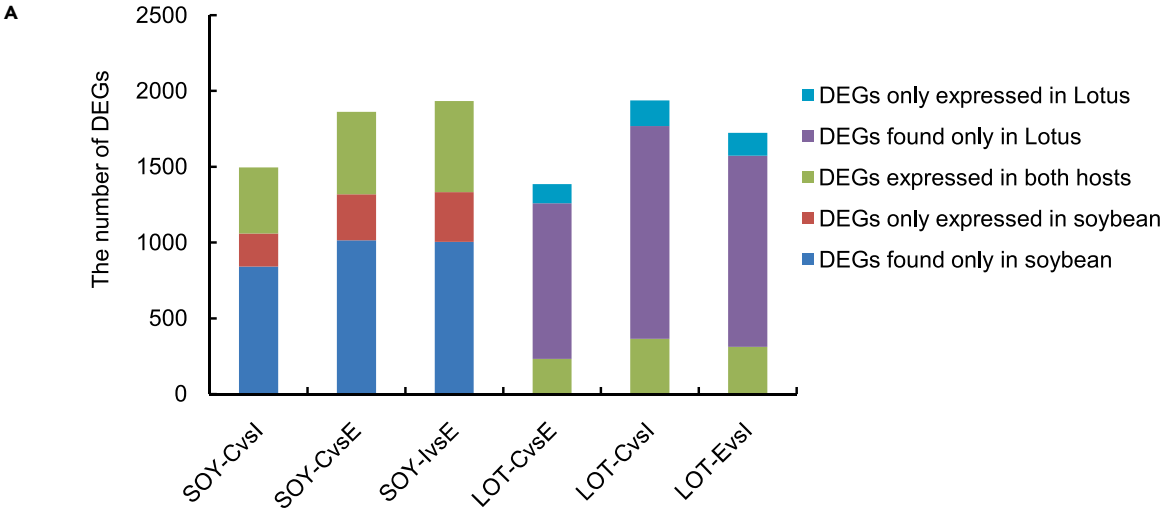
Analysis of DEGs encoding resistance proteins

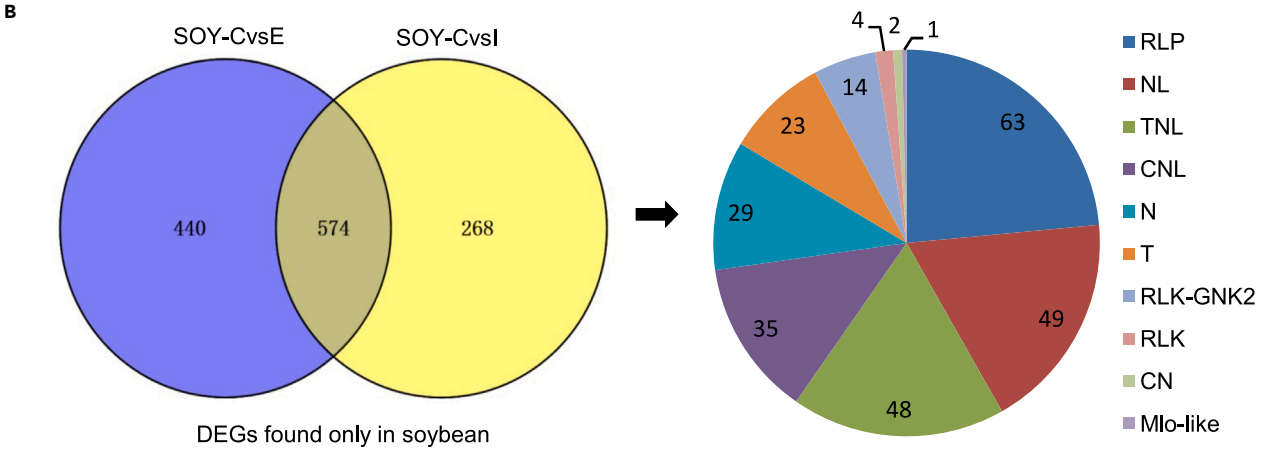
In a few cases, plant-encoded resistance (R) proteins can prevent nodulation in specific strains, presumably mediated by effectors recognition,[34] indicating the critical role of R genes in mediating genotype-specific nodulation. To investigate whether R proteins are involved in recognition of rhizobia 113-2 or MAFF303099 in soybean and *Lotus* roots, DEGs encoding R proteins were analyzed in more detail. The numbers of the DEGs encoding R proteins of the five gene sets ("DEGs found only in soybean," "DEGs found only in *Lotus*," "DEGs only expressed in soybean," "DEGs only expressed in *Lotus*," and "DEGs expressed in both hosts") in the six groups are shown in Figure 5A, and the detailed gene ID information of these DEGs is shown in Table S9. 268 unique soybean R genes were only identified in SOY-CvsI (not in SOY-CvsE), and 738 unique *Lotus* R genes were only identified in LOT-CvsI (not in LOT-CvsE). Based on the PRG database (http://prgdb.crg.eu/wiki/Category:Classes), 268 soybean R genes were classified into 10 subsets (Figure 5B) and 738 *Lotus* R genes were classified into 13 subsets (Figure 5C). The detailed classification information of these DEGs is shown in Table S10. Both in soybean and *Louts* R genes, the RLP (receptor-like protein), NL (NBS-LRR), TNL (TIR-NBS-LRR), and CNL (CC-NBS-LRR) are the four main types, supplying clues for identifying the R genes that regulate nodulation.

To identify the potential R genes that inhibit nodulation of the two symbiotic systems (*M. japonicum* MAFF303099-soybean and *B. diazoefficiens* 113-2-*L. japonicus*), we focused on analyzing the upregulated unique genes in SOY-CvsI and LOT-CvsI. For the 268 unique soybean R genes, 16, 54, 42, and 35 genes were upregulated in 5hR, 30hR, 3dR, and 8dR of SOY-CvsI, respectively. In LOT-CvsI, 53, 150, 188, and 100 unique *Lotus* R genes were upregulated in 5hR, 30hR, 3dR, and 8dR, respectively. We then randomly selected nine upregulated soybean R genes (more than 16-fold DEGs) and detected the expression of these DEGs in the soybean roots of control and two soybean symbiotic systems by RT-qPCR (Figure 6). Similar to the RNA-seq data, inoculation of *M. japonicum* MAFF303099 significantly increased the expression of these nine R genes in soybean roots at one or more time points. Additionally, the expression patterns of these R genes in ineffective nodulation combination (SOY-CvsI) are very different from effective nodulation combination (SOY-CvsE). The primer sets are listed in Table S11.

Analysis of soybean and *Lotus* functionally validated SNF genes

To confirm the opinions obtained from the RNA-seq, we analyzed 234 soybean and 197 *Lotus* functionally validated SNF genes selected from the previous studies[37] (Table 3; Table S12). 81 out 234 (35%) soybean SNF genes and 89 out 197 (45%) *Lotus* SNF genes are host specific genes (Table 3), meaning that a considerable proportion of host specific genes really involved in regulating nodulation. Not all of the selected SNF


A
2500
2000
1500
1000
500
0
The number of DEGs
SOY-CvsI
SOY-CvsE
SOY-IvsE
LOT-CvsE
LOT-CvsI
LOT-EvsI
DEGs only expressed in Lotus
DEGs found only in Lotus
DEGs expressed in both hosts
DEGs only expressed in soybean
DEGs found only in soybean


B
SOY-CvsE
SOY-CvsI
440
574
268
DEGs found only in soybean
63
49
48
35
29
23
14
4
2
1
RLP
NL
TNL
CNL
N
T
RLK-GNK2
RLK
CN
Mlo-like

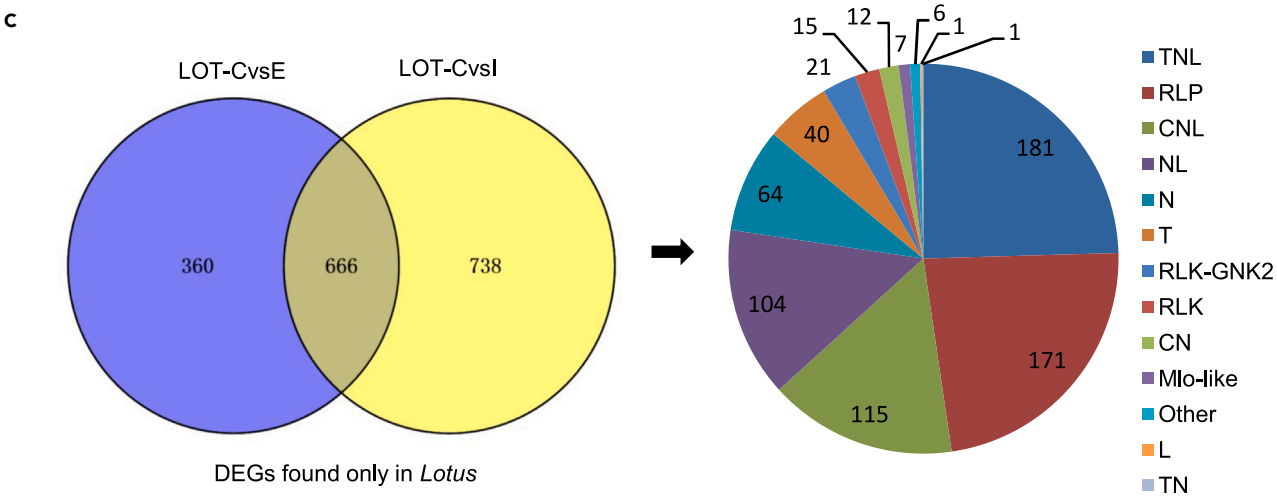

C
LOT-CvsE
LOT-CvsI
360
666
738
DEGs found only in Lotus
181
171
115
104
64
40
21
15
12
7
6
1
1
TNL
RLP
CNL
NL
N
T
RLK-GNK2
RLK
CN
Mlo-like
Other
L
TN

Figure 5. Analysis of DEGs encoding resistance proteins

(A) The numbers of DEGs encoding R genes in different types (DEGs found only in soybean, DEGs only expressed in soybean, DEGs expressed in both hosts, DEGs found only in *Lotus*, and DEGs only expressed in *Lotus*) in soybean and *L. japonicus*.

(B) Analysis of DEGs encoding R genes found only in soybean roots. The numbers in different gene sets (left image). Domain classification of 268 unique soybean R genes identified only in SOY-CvsI (right image).

(C) Analysis of DEGs encoding R genes found only in *Lotus* roots. The numbers in different gene sets (left image). Domain classification of 738 unique *Lotus* R genes identified only in LOT-CvsI (right image).

genes were identified in the RNA-seq (112 soybean SNF and 104 *Lotus* SNF), because of that our RNA-seq only analyzed the early stage of nodulation and not included the later stage of nodule development and nitrogen fixation. For the DEGs encoding SNF genes, the "DEGs found only in soybean" and "DEGs found only in *Lotus*" gene sets account for the majority both in soybean and *Lotus* root samples, and the numbers of DEGs in SOY-CvsE and LOT-CvsI were higher than SOY-CvsI and LOT-CvsE, respectively (Table 3). Additionally, most of aforementioned functionally validated SNF genes also different expressed in no fully function symbiotic systems (SOY-CvsI and LOT-CvsI). The detailed ID information for Table 3 is shown in Table S12. These results suggested that fully function symbiotic systems may mainly due to the symbiotic host specificity rather than symbiotic genes.

Gene co-expression network construction and SNF-related module analysis

The identified DEGs in soybean and *L. japonicus* root samples were imported into the WGCNA (Weighted Correlation Network Analysis) software package[38] for analysis. In the WGCNA Module Heatmap of soybean DEGs, 21 modules, which were distinguishable by different colors, were identified (Figure 7A), and the number of target genes for each module ranged from 20 to 177 (Table S13). WGCNA analysis for *Lotus* DEGs resulted in three modules that were distinguishable by different colors (Figure 7B), and the number of target genes for each module ranged from 126 to 204 (Table S14). Each module corresponded to each root sample and had its correlation. The sizes of the correlations for soybean WGCNA analysis are shown in Figure 7C, and for *L. japonicus* are shown in Figure 7D. Whether the correlation was positive or negative and the size of the correlation showed the degree of correlation with the target gene screened out by the RNA-seq data of this root sample.

According to the number of the functionally validated SNF genes in each module in soybean or *L. japonicus* DEGs WGCNA analysis, we identified two SNF-related modules (soybean pink module and *Lotus* turquoise module), and the numbers of target genes for these two modules in the five gene sets ("DEGs found only in soybean," "DEGs found only in *Lotus*," "DEGs only expressed in soybean," "DEGs only expressed in *Lotus*," and "DEGs expressed in both hosts") are shown in Figure 7E. The target gene number in "DEGs found only in soybean" or "DEGs found only in *Lotus*" gene set in each SNF-related module accounts for about 54% or 64% of the total in soybean and *L. japonicus*, respectively, indicating that the target genes in SNF-related modules mainly be composed of host specific genes. Additionally, we divided the DEGs in soybean pink module and *Lotus* turquoise module into "functionally validated SNF genes" and "new genes," among 57 soybean DEGs and 204 *Lotus* DEGs, only four soybean DEGs and 11 *Lotus* DEGs are the functionally validated SNF genes, the rest are the new genes that may play roles in SNF (Table 4).

DISCUSSION

Engineering nitrogen-fixing cereals is essential for sustainable food production for the projected global population of 9 billion people in 2050.[3] In addition to engineering nitrogen-fixing system in cereal crops, the host specificity is also a scientific issue that should be taken seriously. The genetic information responsible for host specificity remains largely unexplored. In the present study, to exclude the factors related to genetic background of nodulation and nitrogen fixation, we selected two legumes soybean and *L. japonicus* and performed comparative genomic and transcriptomic analyses to investigate the genetic determinants of host specificity. Our results for the first time divided the DEGs responding to different rhizobia in plant hosts into host specific genes and orthologous pair's genes, and found that host specific genes account for the majority of the DEGs both in soybean and *Lotus* roots samples.

Host-specificity plays a critical role in the host-rhizobia symbiosis

To expand the host range of rhizobia, synthetic biology was developed to engineer symbiosis signaling pathway in non-legume plants, and these SYM pathway genes can be normally expressed in non-legume host,[1–3,39] and cereal barley NFR1 and NFR5-like receptors were able to support *Lotus* root nodule organogenesis.[40] These results indicated that it should be feasible to engineer the SYM pathway for non-legume plants recognition of nitrogen-fixing bacteria. Rhizobia can inhabit the roots of non-legume and were identified as diazotrophic endophytes or N_2-fixing rhizobia.[41–43] Plant growth regulators such as 2,4-dichlorophenoxyacetic acid (2,4-D) can induce the formation of para-nodule in non-legume and rhizobia can inhabit in this para-nodule.[44] However, it is currently unclear how non-legume host responding to the imported SYM pathway, which is the critical factor in the efficiency of this cross-kingdom collaboration.

In this report, we focused on investigating the genetic determinants of host specificity without the influence of the genetic background of nodulation and nitrogen fixation, and utilized RNA-seq to analysis the causes of different symbiotic phenotypes in soybean and *Lotus* roots inoculated with compatible and incompatible rhizobia. RNA-seq, which is an effective method that produces quantitative data related to transcripts with greater sensitivity, higher repeatability, and wider dynamic range than conventional methods,[45] has been shown to have relatively little variation between technical replicates to identify DEGs.[46] A large number of DEGs were identified from RNA-seq data (Figure 1), and

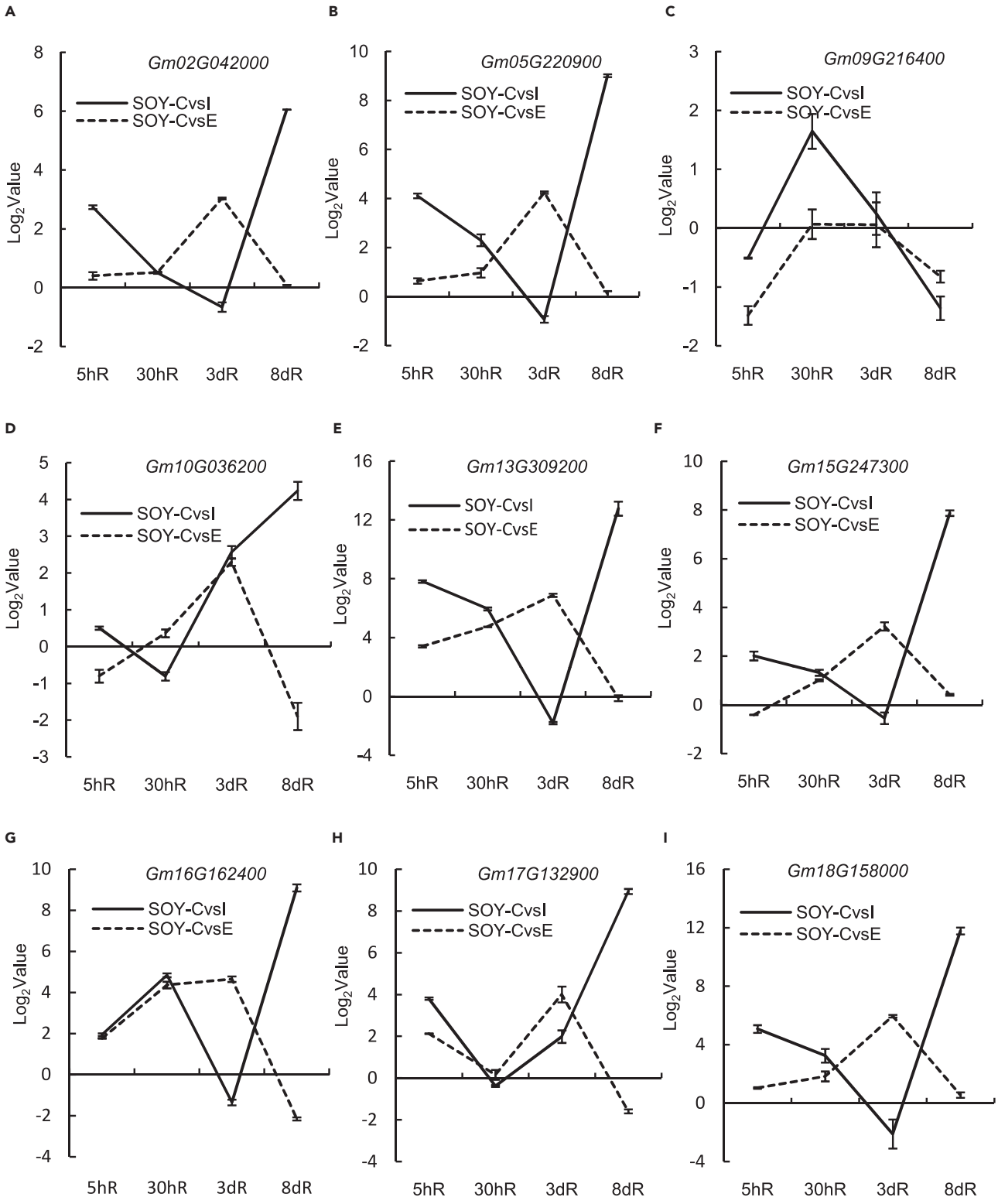


Figure 6. The RT-qPCR analysis of nine randomly selected upregulated soybean R genes

All RT-qPCR reactions were repeated three times and the data are presented as the mean ± SD.

(A) *Glyma.02G042000*.

(B) *Glyma.05G220900*.

Figure 6. *Continued*
(C) *Glyma.09G216400*.
(D) *Glyma.10G036200*.
(E) *Glyma.13G309200*.
(F) *Glyma.15G247300*.
(G) *Glyma.16G162400*.
(H) *Glyma.17G132900*.
(I) *Glyma.18G158000*.

then these DEGs were divided into five gene sets according to the results of comparative genomic analysis (Figure 2). The results showed that host specific genes account for the majority of the DEGs responding to compatible or incompatible rhizobia (Figure 2B). The numbers of the DEGs involved in response to stimulus, associated with plant-pathogen interaction pathways and encoding R proteins of the "only found in soybean" or "only found in *Lotus*" gene sets were far more than that of the other gene sets (Tables 1 and 2; Figure 5A), and the analyzed results of the functionally validated SNF genes and the target genes in SNF-related modules were consistent with the aforementioned DEGs (Table 3; Figure 7E). These results suggested that host-specificity plays a critical role in the symbiotic host-rhizobia associations.

High affinity rhizobia strains are the key to establish fully function nitrogen-fixing symbiosis

The activity of symbiotic nitrogen fixation greatly varies depending on the combination of both partners at species or genotypic levels. Strain-specific no function symbiotic phenotypes have been described in many legumes, such as soybean, pea, vetch. and *L. japonicus*,[13,47–50] and host secreted NCR peptides, nodule-specific aspartic peptidase, and plant-encoded R genes were responsible for the strain-specific nitrogen fixation activity.[13,47,51,52] Additionally, the cooperation between plant host and rhizobia increases as rhizobia adapt to their local host.[53] Our results also revealed that different rhizobia caused different symbiotic phenotypes with different host gene expression patterns. These results indicated that high affinity rhizobia strains are critical to establish fully function nitrogen-fixing symbiosis.

Different host specific R genes are induced at different time points after rhizobia inoculation

Rhizobia can adopt a pathogenic system for activating host symbiosis signaling to promote its infection.[20] Plant-encoded R genes can activate plant immune responses to prevent nodulation by effectors recognition,[34,54] and the host R genes were involved in the control of genotype-specific infection and nodulation.[13] In this reports, 268 unique soybean R genes and 738 unique *Lotus* R genes were identified only in the two non-nodulation symbiosis, respectively. The classification results showed that the R genes that regulate nodulation mainly be composed of RLP, NL, TNL, and CNL types R genes (Figures 5B and 5C). The analysis results of upregulated unique R genes in the two non-nodulation symbiosis demonstrated that different host R genes were induced at different time points after inoculation, and the expression patterns of the upregulated unique R genes in ineffective nodulation combination are very different from effective nodulation combination (Figure 6). These data declared that multiple host specific R genes involved in regulating nodulation at different stages during nodule formation.

In summary, to investigate the genetic determinants of host specificity without the influence of the genetic background of nodulation and nitrogen fixation, we analyzed the different gene expression responses in soybean and *Lotus* roots inoculated with compatible and incompatible rhizobia by comparative genomic and transcriptomic analyses. The DEGs uncovered in this study and their classification analyses provided a molecular basis for revealing the genetic determinants of host specificity, and shed new light on expanding the host range of rhizobia to non-legume plants.

Table 3. The numbers of functionally validated SNF genes in the five gene sets

The numbers of functionally validated SNF genes					
Total soybean genes	Genes found only in soybean	Total soybean DEGs	DEGs found only in soybean	DEGs only expressed in soybean	DEGs expressed in both hosts
234	81	112	48	21	43
DEGs found only in soybean in SOY-CvsE	DEGs only expressed in soybean in SOY-CvsE	DEGs expressed in both hosts in SOY-CvsE	DEGs found only in soybean in SOY-CvsI	DEGs only expressed in soybean in SOY-CvsI	DEGs expressed in both hosts in SOY-CvsI
41	15	31	21	11	22
Total *Lotus* genes	Genes found only in *Lotus*	Total *Lotus* DEGs	DEGs found only in *Lotus*	DEGs only expressed in *Lotus*	DEGs expressed in both hosts
197	89	104	51	15	38
Genes found only in *Lotus* in LOT-CvsE	DEGs only expressed in *Lotus* in LOT-CvsE	DEGs expressed in both hosts in LOT-CvsE	DEGs found only in *Lotus* in LOT-CvsI	DEGs only expressed in *Lotus* in LOT-CvsI	DEGs expressed in both hosts in LOT-CvsI
34	9	21	39	13	32

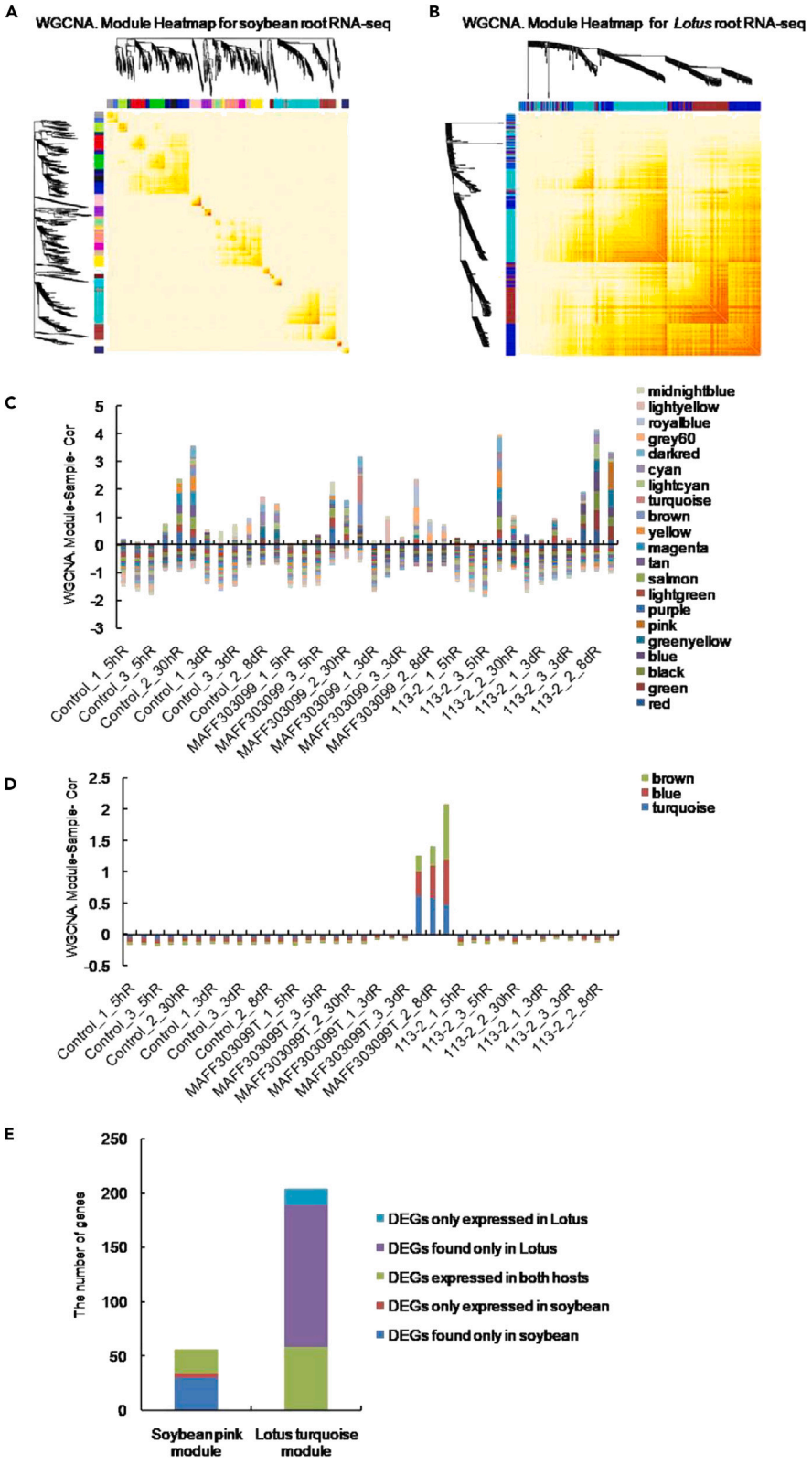

A
WGCNA. Module Heatmap for soybean root RNA-seq
B
WGCNA. Module Heatmap for Lotus root RNA-seq
C
WGCNA. Module-Sample- Cor
midnightblue
lightyellow
royalblue
grey60
darkred
cyan
lightcyan
turquoise
brown
yellow
magenta
tan
salmon
lightgreen
purple
pink
greenyellow
blue
black
green
red
Control_1_5hR
Control_3_5hR
Control_2_30hR
Control_1_3dR
Control_3_3dR
Control_2_8dR
MAFF303099_1_5hR
MAFF303099_3_5hR
MAFF303099_2_30hR
MAFF303099_1_3dR
MAFF303099_3_3dR
MAFF303099_2_8dR
113-2_1_5hR
113-2_3_5hR
113-2_2_30hR
113-2_1_3dR
113-2_3_3dR
113-2_2_8dR
D
WGCNA. Module-Sample- Cor
brown
blue
turquoise
MAFF303099T_1_5hR
MAFF303099T_3_5hR
MAFF303099T_2_30hR
MAFF303099T_1_3dR
MAFF303099T_3_3dR
MAFF303099T_2_8dR
E
The number of genes
DEGs only expressed in Lotus
DEGs found only in Lotus
DEGs expressed in both hosts
DEGs only expressed in soybean
DEGs found only in soybean
Soybean pink module
Lotus turquoise module

Figure 7. WGCNA analysis of the RNA-seq data of soybean and *L. japonicus* roots

(A) Heatmap plot of topological overlap in the gene network for soybean roots RNA-seq data.

(B) Heatmap plot of topological overlap in the gene network for *Lotus* roots RNA-seq data.

(A and B) Each row and column corresponds to a gene, light color denotes low topological overlap, and progressively darker red denotes higher topological overlap. Darker squares along the diagonal correspond to modules. The gene dendrogram and module assignment are shown along the left and top (each tree branch formed a module and each leaf in the branch represented a gene).

(C) Correlation between 21 modules and different soybean root samples.

(D) Correlation between 3 modules and different *Lotus* root samples.

(C and D) Positive value represents positive correlation, negative value represents negative correlation.

(E) The target gene numbers of soybean pink module and *Lotus* turquoise module in the five gene sets.

Limitations of the study

In this report, we described a viewpoint that the response of legumes to rhizobia is mainly regulated by the host specific genes by comparative genomic and transcriptomic analyses. Although we investigated the interest DEGs involved in response to stimulus, associated with plant-pathogen interaction pathways, and encoding R proteins, the SNF proteins and the target proteins in the SNF-related modules, we only analyzed these DEGs to confirm the point obtained from the RNA-seq data; future analysis should be performed to identify more new genes that may play a role in SNF or are a general response to rhizobia. Moreover, we classified the DEGs according to the comparative analyses of only two legumes genomes; future analyses using more legumes genomes should provide more accurate classification.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Soybean
  - Lotus
  - Rhizobial strains
  - Method details
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110207.

ACKNOWLEDGMENTS

This work was supported by funds from the National Natural Science Foundation of China (grant no. 32071964), Key Research and Development Plan Projects of Hubei Province (2022BBA0036), and Natural Science Foundation of Guangdong Province (2022A1515110240). The funding body played no role in the study design, sample collection, data analysis, data interpretation, and manuscript drafting.

AUTHOR CONTRIBUTIONS

Conceptualization, S.Y., H.C., and X.Z.; methodology and data analysis, S.Y., P.L., Y.F., F.J., H.Z., C.Z., Y.H., Z.S., Z.Y., Q.H., S.C., and L.C.; resources, D.C., W.G., and H.Y.; writing – original draft, S.Y.; writing – review & editing, S.Y. and H.C.; funding acquisition, S.Y. and Y. F.; supervision, H.C. and X.Z.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 24, 2023
Revised: February 3, 2024
Accepted: June 4, 2024
Published: June 6, 2024

Table 4. The gene ID information of the DEGs in soybean pink module and *Lotus* turquoise module

Soybean pink module					
Functionally validated SNF genes					
Glyma.10G122300	Glyma.13G093600	Glyma.17G202900	Glyma.09G129700		
New genes					
Glyma.01G037100	Glyma.03G164000	Glyma.07G088200	Glyma.09G188700	Glyma.13G300600	Glyma.17G133400
Glyma.02G044200	Glyma.05G051400	Glyma.07G249000	Glyma.10G135500	Glyma.13G310800	Glyma.18G012300
Glyma.02G051900	Glyma.05G223200	Glyma.08G037000	Glyma.10G170900	Glyma.14G035100	Glyma.18G018900
Glyma.02G076900	Glyma.06G061700	Glyma.08G097300	Glyma.11G222300	Glyma.15G048400	Glyma.18G216200
Glyma.02G278200	Glyma.06G187100	Glyma.08G162400	Glyma.11G227600	Glyma.16G201200	Glyma.02G052000
Glyma.03G104500	Glyma.06G315400	Glyma.09G149800	Glyma.12G053300	Glyma.17G073400	Glyma.06G269300
Glyma.03G246200	Glyma.06G319000	Glyma.08G102100	Glyma.12G115600	Glyma.17G066800	Glyma.19G165300
Glyma.05G041100	Glyma.07G230300	Glyma.08G131900	Glyma.16G137300	Glyma.18G124700	Glyma.19G251500
Glyma.06G004400	Glyma.13G348500	Glyma.14G126500	Glyma.19G105500	BGI_novel_G002083	
Lotus turquoise module					
Functionally validated SNF genes					
Lj4g3v1200890	Lj4g3v3099540	Lj5g3v0525260	Lj6g3v1055620	Lj2g3v1415410	Lj3g3v0014470
Lj5g3v1699100	Lj1g3v0264200	Lj3g3v2681670	Lj6g3v1537040	Lj5g3v0841080	
New genes					
Lj0g3v0045929	Lj0g3v0207559	Lj0g3v0364369	Lj2g3v1613100	Lj3g3v2296560	Lj4g3v1969860
Lj0g3v0084479	Lj0g3v0218349	Lj1g3v0752210	Lj2g3v1728960	Lj3g3v2309700	Lj4g3v1983330
Lj0g3v0086659	Lj0g3v0220499	Lj1g3v1363200	Lj2g3v1828460	Lj3g3v2315470	Lj4g3v2309300
Lj0g3v0093139	Lj0g3v0251019	Lj1g3v1784380	Lj2g3v1925790	Lj3g3v2719870	Lj4g3v2618540
Lj0g3v0102219	Lj0g3v0254099	Lj1g3v2264920	Lj2g3v2136010	Lj3g3v2920950	Lj4g3v3031650
Lj0g3v0123689	Lj0g3v0254249	Lj1g3v3444100	Lj2g3v2171710	Lj3g3v2986090	Lj4g3v3044980
Lj0g3v0142949	Lj0g3v0265069	Lj1g3v4104820	Lj2g3v2904980	Lj3g3v3714360	Lj5g3v0525250
Lj0g3v0151519	Lj0g3v0269259	Lj1g3v4275470	Lj2g3v3059650	Lj4g3v0336900	Lj5g3v0526350
Lj0g3v0151889	Lj0g3v0271059	Lj1g3v4446990	Lj2g3v3106320	Lj4g3v0336910	Lj5g3v0670650
Lj0g3v0152159	Lj0g3v0287589	Lj1g3v4955440	Lj3g3v0423980	Lj4g3v0575640	Lj5g3v2133790
Lj0g3v0164329	Lj0g3v0325719	Lj1g3v5061020	Lj3g3v0512940	Lj4g3v0679870	Lj6g3v0727870
Lj0g3v0178079	Lj0g3v0348869	Lj2g3v0911680	Lj3g3v0949000	Lj4g3v0911380	Lj6g3v1915950
Lj0g3v0190479	Lj0g3v0360189	Lj2g3v1068420	Lj3g3v1378140	Lj4g3v1212320	Lj6g3v2255710
Lj0g3v0207199	Lj0g3v0362469	Lj2g3v1352610	Lj3g3v2118200	Lj4g3v1616860	BGI_novel_G002356
Lj0g3v0008099	Lj0g3v0164339	Lj1g3v1318130	Lj2g3v0661570	Lj3g3v2520080	Lj5g3v1048210
Lj0g3v0055579	Lj0g3v0199229	Lj1g3v1932460	Lj2g3v0727510	Lj3g3v2769460	Lj5g3v1083950
Lj0g3v0055609	Lj0g3v0253889	Lj1g3v2001820	Lj2g3v0855300	Lj3g3v2888290	Lj5g3v1169780
Lj0g3v0055919	Lj0g3v0258549	Lj1g3v2011920	Lj2g3v1014480	Lj3g3v3751920	Lj5g3v2112290
Lj0g3v0059359	Lj0g3v0284719	Lj1g3v2432890	Lj2g3v1024320	Lj3g3v3752200	Lj6g3v0609710
Lj0g3v0064929	Lj0g3v0289579	Lj1g3v3779370	Lj2g3v1353620	Lj4g3v0793460	Lj6g3v0792120
Lj0g3v0070989	Lj0g3v0303549	Lj1g3v3918110	Lj2g3v1728900	Lj4g3v1399930	Lj6g3v1038780
Lj0g3v0074079	Lj0g3v0304969	Lj1g3v3975480	Lj2g3v1879770	Lj4g3v1855740	Lj6g3v1048890
Lj0g3v0075609	Lj0g3v0304989	Lj1g3v4082070	Lj2g3v1925800	Lj4g3v2022720	Lj6g3v1077220
Lj0g3v0083399	Lj0g3v0305209	Lj1g3v4154950	Lj2g3v2017510	Lj4g3v2133900	Lj6g3v1177330
Lj0g3v0108759	Lj0g3v0324769	Lj1g3v4155980	Lj2g3v2899940	Lj4g3v2140260	Lj6g3v2007310
Lj0g3v0115909	Lj0g3v0329029	Lj1g3v4447100	Lj2g3v3018620	Lj4g3v2253780	Lj6g3v2085690
Lj0g3v0119529	Lj0g3v0335709	Lj1g3v4452600	Lj2g3v3222870	Lj4g3v2376240	Lj6g3v2158610

(*Continued on next page*)

Table 4. *Continued*

Lotus turquoise module					
Lj0g3v0129439	Lj0g3v0342909	Lj1g3v4515410	Lj2g3v3339780	Lj4g3v2742940	Lj6g3v2274490
Lj0g3v0130479	Lj0g3v0346859	Lj1g3v4564810	Lj3g3v1428150	Lj4g3v3016310	BGI_novel_G000769
Lj0g3v0130729	Lj0g3v0349789	Lj1g3v4862990	Lj3g3v1618350	Lj4g3v3113860	BGI_novel_G000827
Lj0g3v0130739	Lj0g3v0357309	Lj1g3v5031290	Lj3g3v2261380	Lj5g3v0526340	BGI_novel_G000945
Lj0g3v0154319	Lj1g3v0416320	Lj2g3v0635390	Lj3g3v2385550	Lj5g3v0626670	BGI_novel_G001051
Lj0g3v0113159					

REFERENCES

1. Guo, K., Yang, J., Yu, N., Luo, L., and Wang, E. (2023). Biological nitrogen fixation in cereal crops: Progress, strategies, and perspectives. Plant Commun. *4*, 100499. https://doi.org/10.1016/j.xplc.2022.100499.
2. Rogers, C., and Oldroyd, G.E.D. (2014). Synthetic biology approaches to engineering the nitrogen symbiosis in cereals. J. Exp. Bot. *65*, 1939–1946. https://doi.org/10.1093/jxb/eru098.
3. Charpentier, M., and Oldroyd, G. (2010). How close are we to nitrogen-fixing cereals? Curr. Opin. Plant Biol. *13*, 556–564. https://doi.org/10.1016/j.pbi.2010.08.003.
4. He, W., Burén, S., Baysal, C., Jiang, X., Capell, T., Christou, P., and Rubio, L.M. (2022). Nitrogenase cofactor maturase NifB isolated from transgenic rice is active in FeMo-co synthesis. ACS Synth. Biol. *11*, 3028–3036. https://doi.org/10.1021/acssynbio.2c00194.
5. Ryu, M.H., Zhang, J., Toth, T., Khokhani, D., Geddes, B.A., Mus, F., Garcia-Costas, A., Peters, J.W., Poole, P.S., Ané, J.M., and Voigt, C.A. (2020). Control of nitrogen fixation in bacteria that associate with cereals. Nat. Microbiol. *5*, 314–330. https://doi.org/10.1038/s41564-019-0631-2.
6. Jones, K.M., Sharopova, N., Lohar, D.P., Zhang, J.Q., VandenBosch, K.A., and Walker, G.C. (2008). Differential response of the plant *Medicago truncatula* to its symbiont *Sinorhizobium meliloti* or an exopolysaccharide-deficient mutant. Proc. Natl. Acad. Sci. USA *105*, 704–709. https://doi.org/10.1073/pnas.0709338105.
7. Hayashi, M., Saeki, Y., Haga, M., Harada, K., Kouchi, H., and Umehara, Y. (2012). Rj (rj) genes involved in nitrogen-fixing root nodule formation in soybean. Breed Sci. *61*, 544–553. https://doi.org/10.1270/jsbbs.61.544.
8. Yuan, S., Li, R., Chen, S., Chen, H., Zhang, C., Chen, L., Hao, Q., Shan, Z., Yang, Z., Qiu, D., et al. (2016). RNA-Seq analysis of differential gene expression responding to different rhizobium strains in soybean (*Glycine max*) roots. Front. Plant Sci. *7*, 721. https://doi.org/10.3389/fpls.2016.00721.
9. Martínez-Hidalgo, P., Ramírez-Bahena, M.H., Flores-Félix, J.D., Igual, J.M., Sanjuán, J., León-Barrios, M., Peix, A., and Velázquez, E. (2016). Reclassification of strains MAFF 303099T and R7A into *Mesorhizobium japonicum* sp. nov. Int. J. Syst. Evol. Microbiol. *66*, 4936–4941. https://doi.org/10.1099/ijsem.0.001448.
10. Estrella, M.J., Muñoz, S., Soto, M.J., Ruiz, O., and Sanjuán, J. (2009). Genetic diversity and host range of rhizobia nodulating Lotus tenuis in typical soils of the *Salado River Basin* (Argentina). Appl. Environ. Microbiol. *75*, 1088–1098. https://doi.org/10.1128/aem.02405-08.
11. Wang, S., Hao, B., Li, J., Gu, H., Peng, J., Xie, F., Zhao, X., Frech, C., Chen, N., Ma, B., and Li, Y. (2014). Whole-genome sequencing of *Mesorhizobium huakuii* 7653R provides molecular insights into host specificity and symbiosis island dynamics. BMC Genom. *15*, 440. https://doi.org/10.1186/1471-2164-15-440.
12. Radutoiu, S., Madsen, L.H., Madsen, E.B., Jurkiewicz, A., Fukai, E., Quistgaard, E.M.H., Albrektsen, A.S., James, E.K., Thirup, S., and Stougaard, J. (2007). LysM domains mediate lipochitin-oligosaccharide recognition and *Nfr* genes extend the symbiotic host range. EMBO J. *26*, 3923–3935. https://doi.org/10.1038/sj.emboj.7601826.
13. Yang, S., Tang, F., Gao, M., Krishnan, H.B., and Zhu, H. (2010). R gene-controlled host specificity in the legume-rhizobia symbiosis. Proc. Natl. Acad. Sci. USA *107*, 18735–18740. https://doi.org/10.1073/pnas.1011957107.
14. Tang, F., Yang, S., and Zhu, H. (2016). Functional analysis of alternative transcripts of the soybean Rj2 gene that restricts nodulation with specific rhizobial strains. Plant Biol. *18*, 537–541. https://doi.org/10.1111/plb.12442.
15. Liu, J., Wang, T., Qin, Q., Yu, X., Yang, S., Dinkins, R.D., Kuczmog, A., Putnoky, P., Muszyński, A., Griffitts, J.S., et al. (2022). Paired Medicago receptors mediate broad-spectrum resistance to nodulation by *Sinorhizobium meliloti* carrying a species-specific gene. Proc. Natl. Acad. Sci. USA *119*, e2214703119. https://doi.org/10.1073/pnas.2214703119.
16. Perret, X., Staehelin, C., and Broughton, W.J. (2000). Molecular basis of symbiotic promiscuity. Microbiol. Mol. Biol. Rev. *64*, 180–201.
17. Lerouge, P., Roche, P., Faucher, C., Maillet, F., Truchet, G., Promé, J.C., and Dénarié, J. (1990). Symbiotic host-specificity of Rhizobium meliloti is determined by a sulphated and acylated glucosamine oligosaccharide signal. Nature *344*, 781–784. https://doi.org/10.1038/344781a0.
18. Schultze, M., Quiclet-Sire, B., Kondorosi, E., Virelizier, H., Glushka, J.N., Endre, G., Géro, S.D., and Kondorosi, A. (1992). Rhizobium meliloti produces a family of sulfated lipooligosaccharides exhibiting different degrees of plant host specificity. Proc. Natl. Acad. Sci. USA *89*, 192–196.
19. Skorupska, A., Janczarek, M., Marczak, M., Mazur, A., and Król, J. (2006). Rhizobial exopolysaccharides: genetic control and symbiotic functions. Microb. Cell Factories *5*, 7. https://doi.org/10.1186/1475-2859-5-7.
20. Okazaki, S., Kaneko, T., Sato, S., and Saeki, K. (2013). Hijacking of leguminous nodulation signaling by the rhizobial type III secretion system. Proc. Natl. Acad. Sci. USA *110*, 17131–17136. https://doi.org/10.1073/pnas.1302360110.
21. Fauvart, M., and Michiels, J. (2008). Rhizobial secreted proteins as determinants of host specificity in the rhizobium-legume symbiosis. FEMS Microbiol. Lett. *285*, 1–9. https://doi.org/10.1111/j.1574-6968.2008.01254.x.
22. Nelson, M.S., and Sadowsky, M.J. (2015). Secretion systems and signal exchange between nitrogen-fixing rhizobia and legumes. Front. Plant Sci. *6*, 491. https://doi.org/10.3389/fpls.2015.00491.
23. Day, R.B., Okada, M., Ito, Y., Tsukada, K., Zaghouani, H., Shibuya, N., and Stacey, G. (2001). Binding site for chitin oligosaccharides in the soybean plasma membrane. Plant Physiol. *126*, 1162–1173.
24. Cao, Y., Halane, M.K., Gassmann, W., and Stacey, G. (2017). The role of plant innate immunity in the legume-rhizobium symbiosis. Annu. Rev. Plant Biol. *68*, 535–561. https://doi.org/10.1146/annurev-arplant-042916-041030.
25. D'Haeze, W., and Holsters, M. (2004). Surface polysaccharides enable bacteria to evade plant immunity. Trends Microbiol. *12*, 555–561. https://doi.org/10.1016/j.tim.2004.10.009.
26. Soto, M.J., Domínguez-Ferreras, A., Pérez-Mendoza, D., Sanjuán, J., and Olivares, J. (2009). Mutualism versus pathogenesis: the give-and-take in plant-bacteria interactions. Cell Microbiol. *11*, 381–388. https://doi.org/10.1111/j.1462-5822.2008.01282.x.
27. Deakin, W.J., and Broughton, W.J. (2009). Symbiotic use of pathogenic strategies: rhizobial protein secretion systems. Nat. Rev. Microbiol. *7*, 312–320. https://doi.org/10.1038/nrmicro2091.
28. Lopez-Gomez, M., Sandal, N., Stougaard, J., and Boller, T. (2012). Interplay of flg22-induced defence responses and nodulation in *Lotus japonicus*. J. Exp. Bot. *63*, 393–401. https://doi.org/10.1093/jxb/err291.
29. Tellstrom, V., Usadel, B., Thimm, O., Stitt, M., Kuster, H., and Niehaus, K. (2007). The lipopolysaccharide of *Sinorhizobium meliloti* suppresses defense-associated gene expression in cell cultures of the host plant *Medicago truncatula*. Plant Physiol. *143*, 825–837. https://doi.org/10.1104/pp.106.090985.
30. Buttner, D., and He, S.Y. (2009). Type III protein secretion in plant pathogenic bacteria. Plant Physiol. *150*, 1656–1664. https://doi.org/10.1104/pp.109.139089.
31. 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., et al. (2015). Receptor-mediated

exopolysaccharide perception controls bacterial infection. Nature *523*, 308–312. https://doi.org/10.1038/nature14611.
32. Geurts, R., Heidstra, R., Hadri, A.E., Downie, J.A., Franssen, H., Van Kammen, A., and Bisseling, T. (1997). Sym2 of Pea is involved in a nodulation factor-perception mechanism that controls the infection process in the epidermis. Plant Physiol. *115*, 351–359.
33. Limpens, E., Franken, C., Smit, P., Willemse, J., Bisseling, T., and Geurts, R. (2003). LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. Science *302*, 630–633. https://doi.org/10.1126/science.1090074.
34. Yasuda, M., Miwa, H., Masuda, S., Takebayashi, Y., Sakakibara, H., and Okazaki, S. (2016). Effector-triggered immunity determines host genotype-specific incompatibility in legume-rhizobium symbiosis. Plant Cell Physiol. *57*, 1791–1800. https://doi.org/10.1093/pcp/pcw104.
35. 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., et al. (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. https://doi.org/10.1038/srep42248.
36. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. *44*, D457–D462. https://doi.org/10.1093/nar/gkv1070.
37. Roy, S., Liu, W., Nandety, R.S., Crook, A., Mysore, K.S., Pislariu, C.I., Frugoli, J., Dickstein, R., and Udvardi, M.K. (2020). Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. Plant Cell *32*, 15–41. https://doi.org/10.1105/tpc.19.00279.
38. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. *9*, 559. https://doi.org/10.1186/1471-2105-9-559.
39. Huisman, R., and Geurts, R. (2020). A roadmap toward engineered nitrogen-fixing nodule symbiosis. Plant Commun. *1*, 100019. https://doi.org/10.1016/j.xplc.2019.100019.
40. Rübsam, H., Krönauer, C., Abel, N.B., Ji, H., Lironi, D., Hansen, S.B., Nadzieja, M., Kolte, M.V., Abel, D., de Jong, N., et al. (2023). Nanobody-driven signaling reveals the core receptor complex in root nodule symbiosis. Science *379*, 272–277. https://doi.org/10.1126/science.ade9204.
41. Hara, S., Morikawa, T., Wasai, S., Kasahara, Y., Koshiba, T., Yamazaki, K., Fujiwara, T., Tokunaga, T., and Minamisawa, K. (2019). Identification of nitrogen-fixing Bradyrhizobium associated with roots of field-grown sorghum by metagenome and proteome analyses. Front. Microbiol. *10*, 407. https://doi.org/10.3389/fmicb.2019.00407.
42. Greetatorn, T., Hashimoto, S., Sarapat, S., Tittabutr, P., Boonkerd, N., Uchiumi, T., and Teaumroong, N. (2019). Empowering rice seedling growth by endophytic *Bradyrhizobium* sp. SUTN9-2. Lett. Appl. Microbiol. *68*, 258–266. https://doi.org/10.1111/lam.13114.
43. Rouws, L.F.M., Leite, J., de Matos, G.F., Zilli, J.E., Coelho, M.R.R., Xavier, G.R., Fischer, D., Hartmann, A., Reis, V.M., and Baldani, J.I. (2014). Endophytic *Bradyrhizobium* spp. isolates from sugarcane obtained through different culture strategies. Environ. Microbiol. Rep. *6*, 354–363. https://doi.org/10.1111/1758-2229.12122.
44. Liu, H., Wang, X., Qi, H., Wang, Q., Chen, Y., Li, Q., Zhang, Y., Qiu, L., Fontana, J.E., Zhang, B., et al. (2017). The infection and impact of *Azorhizobium caulinodans* ORS571 on wheat (*Triticum aestivum* L.). PLoS One *12*, e0187947. https://doi.org/10.1371/journal.pone.0187947.
45. Jiang, L., Romero-Carvajal, A., Haug, J.S., Seidel, C.W., and Piotrowski, T. (2014). Gene-expression analysis of hair cell regeneration in the zebrafish lateral line. Proc. Natl. Acad. Sci. USA *111*, E1383–E1392. https://doi.org/10.1073/pnas.1402898111.
46. Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. *18*, 1509–1517. https://doi.org/10.1101/gr.079558.108.
47. Shimoda, Y., Nishigaya, Y., Yamaya-Ito, H., Inagaki, N., Umehara, Y., Hirakawa, H., Sato, S., Yamazaki, T., and Hayashi, M. (2020). The rhizobial autotransporter determines the symbiotic nitrogen fixation activity of *Lotus japonicus* in a host-specific manner. Proc. Natl. Acad. Sci. USA *117*, 1806–1815. https://doi.org/10.1073/pnas.1913349117.
48. Caldwell, B.E. (1966). Inheritance of a Strain-Specific Ineffective Nodulation in Soybeans1. Crop Sci. *6*, 427–428. https://doi.org/10.2135/cropsci1966.0011183X000600050010x.
49. Lie, T.A. (1981). Gene centres, a source for genetic variants in symbiotic nitrogen fixation: Host-induced ineffectivity inPisum sativum ecotype fulvum. Plant Soil *61*, 125–134. https://doi.org/10.1007/BF02277369.
50. Duc, G., and Picard, J. (1986). Note on the presence of the sym-1 gene in Vicia faba hampering its symbiosis with Rhizobium leguminosarum. Euphytica *35*, 61–64. https://doi.org/10.1007/BF00028541.
51. Wang, Q., Yang, S., Liu, J., Terecskei, K., Ábrahám, E., Gombár, A., Domonkos, Á., Szűcs, A., Körmöczi, P., Wang, T., et al. (2017). Host-secreted antimicrobial peptide enforces symbiotic selectivity in *Medicago truncatula*. Proc. Natl. Acad. Sci. USA *114*, 6854–6859. https://doi.org/10.1073/pnas.1700715114.
52. Yang, S., Wang, Q., Fedorova, E., Liu, J., Qin, Q., Zheng, Q., Price, P.A., Pan, H., Wang, D., Griffitts, J.S., et al. (2017). Microsymbiont discrimination mediated by a host-secreted peptide in *Medicago truncatula*. Proc. Natl. Acad. Sci. USA *114*, 6848–6853. https://doi.org/10.1073/pnas.1700460114.
53. Batstone, R.T., O'Brien, A.M., Harrison, T.L., and Frederickson, M.E. (2020). Experimental evolution makes microbes more cooperative with their local host genotype. Science *370*, 476–478. https://doi.org/10.1126/science.abb7222.
54. Zhang, B., Wang, M., Sun, Y., Zhao, P., Liu, C., Qing, K., Hu, X., Zhong, Z., Cheng, J., Wang, H., et al. (2021). *Glycine max* NNL1 restricts symbiotic compatibility with widely distributed bradyrhizobia via root hair infection. Nat. Plants *7*, 73–86. https://doi.org/10.1038/s41477-020-00832-7.
55. Li, R., Yu, C., Li, Y., Lam, T.W., Yiu, S.M., Kristiansen, K., and Wang, J. (2009). SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics *25*, 1966–1967. https://doi.org/10.1093/bioinformatics/btp336.
56. Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. *40*, e49. https://doi.org/10.1093/nar/gkr1293.
57. Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402–408. https://doi.org/10.1006/meth.2001.1262.
58. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60. https://doi.org/10.1038/nmeth.3176.
59. Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics *26*, 136–138. https://doi.org/10.1093/bioinformatics/btp612.
60. Sanseverino, W., Roma, G., De Simone, M., Faino, L., Melito, S., Stupka, E., Frusciante, L., and Ercolano, M.R. (2010). PRGdb: a bioinformatics platform for plant resistance gene analysis. Nucleic Acids Res. *38*, D814–D821. https://doi.org/10.1093/nar/gkp978.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Mesorhizobium japonicum MAFF303099	Huazhong Agricultural University in China	N/A
Bradyrhizobium diazoefficiens 113-2	Oil Crops Research Institute of CAAS	N/A
Biological samples		
Soybean Tian long No.1	Oil Crops Research Institute of CAAS	N/A
Lotus japonicus (MG20)	Huazhong Agricultural University in China	N/A
Deposited data		
Soybean raw sequence reads	The SRA database	PRJNA1065945
Experimental models: Organisms/strains		
Soybean/ *Bradyrhizobium diazoefficiens* 113-2	This study	N/A
Soybean/ *Mesorhizobium japonicum* MAFF303099	This study	N/A
Lotus japonicus/ *Bradyrhizobium diazoefficiens* 113-2	This study	N/A
Lotus japonicus/ *Mesorhizobium japonicum* MAFF303099	This study	N/A
Software and algorithms		
SOAP aligner/ SOAP2	Li et al.[55]	N/A
MCScanX (v.0.8)	Wang et al.[56]	N/A
$2^{-\Delta\Delta CT}$ method	Livak et al.[57]	N/A
DIAMOND	Buchfink et al.[58]	N/A
Cluster Profiler package		https://git.bioconductor.org/packages/clusterProfiler
Enrich plot package		https://rdocumentation.org/packages/enrichplot/versions/1.13.1.994
Ggplot2 package		https://ggplot2.tidyverse.org.
WGCNA software package	Langfelder et al.[38]	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Songli Yuan (songliyuan@caas.cn).

Materials availability

The seeds of soybean Tian long No.1 and *Bradyrhizobium diazoefficiens* 113-2 were stored in Author's lab. The seeds of *Lotus japonicus* (MG20) were provided by Huazhong Agricultural University in China, Professor Cao Yangrong's laboratory.

Data and code availability

- All data supporting the findings of this study are available within the paper and within its supplementary data published online. The original sequencing data of Soybean roots RNA-seq have been submitted to NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under the assigned accession number PRJNA1065945, and the original expression data are shown in Table S1. The original

sequencing data of *Lotus* roots RNA-seq (72 documents) are unavailable because of the portable hard drive that stored the original sequencing data of *Lotus* roots RNA-seq was damaged and the data could not be repaired, while the original expression data are shown in Table S2.

- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Soybean

Seeds of Soybean Tian long No.1 (Oil Crops Research Institute of CAAS in China) were surface-sterilized and germinated on moistened filter paper for 5–7 d at 28°C in an incubator with 70% relative humidity (RH) and a 14-h light/10-h dark photoperiod.

Lotus

Seeds of wild-type *Lotus japonicus* 'MG-20' (Provided by Prof. Cao lab, Huazhong Agricultural University in China) were surface-sterilized and germinated on agar plates for 10-14d, then grown in a chamber in a 16-/8-h day/night cycle at 23°C.

Rhizobial strains

Rhizobia strains (*Mesorhizobium japonicum* MAFF303099 or *Bradyrhizobium diazoefficiens* 113-2) were streaked onto YMA plates surface at 28°C for 3 d, and then few colonies were picked and cultured in YMA liquid medium under 28°C for 3∼5 days by shaking cultivation. Centrifuged and collected the bacterial body, and resuspended the bacterial solution in sterile water until the OD value is 0.8∼1.0 for inoculation.

Method details

Plant materials and growth conditions

After a week of adaptation in the greenhouse, wild-type *Lotus japonicus* 'MG-20' plants were inoculated with *M. japonicum* MAFF303099 or *B. diazoefficiens* 113-2 and grown in the same medium without ammonium nitrate. 5–7 days seeds of Soybean Tian long No.1 were inoculated with *M. japonicum* MAFF303099 or *B. diazoefficiens* 113-2 and grown in pots filled with sterilized vermiculite supplemented with half-strength B&D medium under the same growth conditions. Samples for RNA isolation were collected from soybean and *Lotus* roots 1) at 5h; 2) 30h; 3) 3d and 4) 8d of post inoculation. The former two time points represent the period that root hairs recognize the rhizobium signals period, and the latter two time points represent two early nodule development periods. Each collection was performed with three biological replicates for subsequent library construction and sequencing.

RNA extraction and cDNA library preparation

We used TRIzol reagent (Invitrogen, USA) to isolated total RNA from the soybean and *Lotus* roots samples, and removed the potential genomic DNA by using RNeasy plant mini kit (QIAGEN, Germany). Then measured the RNA quantity and quality were measured by using an Epoch Multi-Volume Spectrophotometer system, NanoDrop and Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA).

The method of the cDNA library preparation as described previously8. We used oligo (dT) magnetic beads to enrich mRNAs, and then fragmented these mRNAs in fragmentation buffer to about 200 bp and reverse-transcribed into single strand cDNA by using random hexamer primers. After RNaseH digestion, the cDNA were converted into double strand cDNAs with DNA polymerase I and purified using magnetic beads. Subsequently, we performed end repair, addition of single nucleotide A (adenine) at 3'-end, ligation of cDNA to adaptors and PCR amplification. After qualification and quantification using an Agilent 2100 Bioanaylzer and ABI Step One Plus Real-Time PCR System, the libraries were subjected to sequencing on Illumina HiSeq™ 2000.

Clean reads library formation and quality assessment

The original data from Illumina Hi Seq™ 2000 were raw reads, which include partial adaptor sequences and/or low quality reads. High quality (clean) reads were obtained by trimming off the adaptor sequences and eliminating the reads with higher than 10% unknown bases and reads with higher than 50% low quality bases (base with quality value $\leq$ 5). The clean reads were then mapped to reference genes and genome (Soybean, https://phytozome-next.jgi.doe.gov/info/Gmax_Wm82_a2_v1; *Lotus*, http://www.kazusa.or.jp/lotus/) using SOAP aligner/SOAP2[55] with threshold that no more than two mismatches were permitted in the alignment.

The Q20 or Q30 value for the clean reads was more than 92.5%, and the proportion of clean reads among the total acquired reads was more than 82.97%, the mapping results of total clean reads, total mapping ratio and uniquely mapping ratio, sequencing saturation analysis and random nessassessment analysis indicated that the sequencing was of good quality and contained sufficient information for gene expression analysis. Additionally, in order to reflect the gene expression correlation between samples (especially for three biologically repeated root samples at each time point), we calculated the Pearson correlation coefficients for all gene expression levels between each two samples. For the three biologically repeated root samples at each time point in soybean or *L .japonicus*, most of the Pearson correlation coefficients values were more than 0.9 (94.4% in soybean, and 93.1% in *L. japonicus*).

Identification of DEGs

To judge the significance of differences in DEGs between roots inoculated with *B. diazoefficiens* 113-2 or *M. japonicum* MAFF303099 in soybean and *L. japonicus*, DEGseq method[59] was used and a fold change of $\geq$ 2 and Q-value$\leq$0.001 were used as criteria for the following six groups:

- SOY-CvsI (uninoculated control versus MAFF303099 in soybean roots): A comparison of soybean roots with no rhizobium inoculation versus soybean roots 5h, 30h, 3d and 8d after *M. japonicum* MAFF303099 (ineffective) inoculation.
- SOY-CvsE (uninoculated control versus 113-2 in soybean roots): A comparison of soybean roots with no rhizobium inoculation versus soybean roots 5h, 30h, 3d and 8d after *B. diazoefficiens* 113-2 (effective) inoculation.
- SOY-IvsE (MAFF303099 versus 113-2 in soybean roots): A comparison of soybean roots 5h, 30h, 3d and 8d after *M. japonicum* MAFF303099 (ineffective) inoculation versus soybean roots after *B. diazoefficiens* 113-2 (effective) inoculation.
- LOT-CvsE (uninoculated control versus MAFF303099 in *L. japonicus* roots): A comparison of *L. japonicus* roots with no rhizobium inoculation versus *L. japonicus* roots 5h, 30h, 3d and 8d after *M. japonicum* MAFF303099 (effective) inoculation.
- LOT-CvsI (uninoculated control versus 113-2 in *L. japonicus* roots): A comparison of *L. japonicus* roots with no rhizobium inoculation versus *L. japonicus* roots 5h, 30h, 3d and 8d after *B. diazoefficiens* 113-2 (ineffective) inoculation.
- LOT-EvsI (MAFF303099 versus 113-2 in *L. japonicus* roots): A comparison of *L. japonicus* roots 5h, 30h, 3d and 8d after *M. japonicum* MAFF303099 (effective) inoculation versus *L. japonicus* roots after *B. diazoefficiens* 113-2 (ineffective) inoculation.

Genomic syntenic analysis

Orthologous pair's genes between soybean and *L. japonicus* identified using BLASTP (E-value $\leq$ 1e-5). Syntenic blocks between two species were defined by MCScanX (v.0.8)[56] based on the orthologous pairs (the number of genes required to call a syntenic block $\geq$ 5).

RT-qPCR

We used RT-qPCR to further evaluate the DEGs. RNA samples were treated with DNase I (Takara) and reverse-transcribed using a Prime Script RT reagent Kit (Perfect Real Time) with gDNA Eraser (Takara Bio, Inc) and oligo (dT) as the primer. cDNA from the reverse transcription of approximately 1 ug of RNA was used as the template for RT-qPCR using primer sets listed in Table S5 and cycling conditions of 30 s at 95°C followed by 35 cycles of 5 s at 95°C, 30 s at 58°C and 12 s at 72°C and final 5 s at 72°C. The QACT and ubiquitin transcripts were used as the internal controls. Sample cycle threshold (CT) values were standardized for each template using the reference gene as control, and the $2^{-\Delta\Delta CT}$ method[57] was used to analyze the relative changes in gene expression from the RT-qPCR experiments. Three replicate reactions per sample were used to ensure statistical credibility.

Gene Ontology functional and KEGG pathway analyses of DEGs

The Gene ontology (GO) and KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis were performed as previously described,[35] then the enrichment analysis was implemented with "cluster Profiler", "enrich plot" and "ggplot2" packages.

PRG annotation

Plant Resistance Gene Database (PRGdb) (http://prgdb.crg.eu/) includes known and predicted disease resistance genes in various plants, and is a bioinformatics platform for plant resistance gene analysis.[60] Plant Resistance Gene (PRG) annotation and domain classification of DEGs were performed using software DIAMOND.[58]

Gene co-expression network analysis

Gene co-expression network analysis was performed by using WGCNA (Weighted Correlation Network Analysis) software package.[38]

QUANTIFICATION AND STATISTICAL ANALYSIS

All quantitative data are shown as mean $\pm$ SD, Details are provided in each figure legend (Figure 6; Figure S1). Besides, there are no statistical analyses in this study.