Pak. J. Bot., 50(1): 389-398, 2018.

TRANSCRIPTOMIC CHARACTERIZATION OF SOYBEAN (*GLYCINE MAX*) ROOTS IN RESPONSE TO *RHIZOBIUM* INFECTION BY RNA SEQUENCING

QINGYUAN HE[1,2,3], HONGYAN YANG[2,3],ZHENGPENG LI[1],SONGHUA WANG[1], CHANGWEI ZHU[1] AND SHOUCHENG HUANG[1]

[1]College of Life Sciences Anhui Science and Technology University, Fengyang 233100, Anhui, China
[2]Soybean Research Institute, Nanjing Agricultural University;
[3]National Center for Soybean Improvement, Ministry of Agriculture; Key Laboratory of Biology and Genetic Improvement of Soybean, Ministry of Agriculture; National Key Laboratory for Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University; Nanjing, Jiangsu 210095, China
[*]Corresponding author's email: huangsc@ahstu.edu.cn; Tel: +86-6732024

Abstract

Legumes interacting with *rhizobium* to convert N_2 into ammonia for plant use has attracted worldwide interest. However, the plant basal nitrogen fixation mechanisms induced in response to *Rhizobium*, giving differential gene expression of plants, have not yet been fully realized. The differential expressed genes of soybean between inoculated and mock-inoculated were analyzed by a RNA-Seq. The results of the sequencing were aligned against the Williams 82 genome sequence, which contain 55787 transcripts; 280 and 316 transcripts were found to be up- and down-regulated, respectively, for inoculated and mock-inoculated soybean roots at stage V1. Gene ontology (GO) analyses detected 104, 182 and 178 genes associated with the cell component category, molecular function category and biological process category, respectively. Pathway analysis revealed that 98 differentially expressed genes (115 transcripts) were involved in 169 biological pathways. We selected 19 differentially expressed genes and analyzed their expressions in mock-inoculated, inoculated USDA110 and CCBAU45436 using qRT-PCR. The results were in accordance with those obtained from *rhizobia* infected RNA-Seq data. These showed that the results of RNA-Seq had reliability and universality. Additionally, this study showed some novel genes associated with the nitrogen fixation process in comparison to previously identified QTLs.

Key words: Symbiotic nitrogen fixation, Soybean, transcriptomic, RNA-Seq analysis, Rhizobium.

Introduction

Nitrogen is the most limiting element for crop growth and is usually supplied by the application of fertilizer, which entails substantial costs to farmers and potentially has adverse effects on the environment. The leguminous plants establish a symbiotic relationship with *rhizobia* (symbiotic nitrogen fixation, SNF) to directly capture N_2 in nodulation to support plant growth. The development of root nodules commences with a molecular dialogue between the host plant and a compatible strain of *rhizobium*. Many molecular events are triggered in a coordinated manner, leading to morphological and physiological changes in the host plant (Oldroyd, 2011).

The plant excretes signal molecular, flavonoid, and phenolic compounds which induce the synthesis of specific *rhizobia*-produced-lipo-chito-oligosaccharides (Nod factors, NFs). NFs directly stimulate their putative receptor proteins (NFRs), such as LjNFR1/5 (Nod factor through receptor-like kinase) in *Lotus japonicus* and MtLYK3 (LysM-receptor like kinas 3)/NFP (Nod factor perception) in *Medicagorucatula* (Arrighi *et al*., 2006; Limpens *et al*., 2003; Madsen *et al*., 2003; Radutoiu *et al*., 2003), which are LysM (peptidogly can binding lysine motif) receptor-like of the legumes (Kouchi *et al*., 2010; Subramanian, 2006). Bacterial attachment to the root hair induces a calcium influx and membrane depolarization, subsequently leading to deformation and root hair curling, forming an infection thread. Concomitantly, certain cortical cells divide to form nodule primordia; further development gives rise to nodules that differ from tumors in having defined anatomical structures. *Rhizobia* are ramified into these tissues, with subsequent release to the bacterium into plant cells where they differentiate into bacteriods and begin to fix nitrogen (Day *et al*., 2000; Irving *et al*., 2000).

Nodule formation and accommodation of endosymbiotic *rhizobia* inside nodules are strictly controlled by host plant genes. Plant genes that show enhanced expression during nodulation are named "nodulins" (Bladergron & Spaime, 1998), like ENOD2, ENOD12, and ENOD40 (Kochi & Hata, 1995; Pichon *et al*., 1992; Papadopoulou *et al*., 1996). Several genes have been identified that play hinge functions in the perception and transduction of the bacterial NFs (Oldroyd & Downie, 2006; Oldroyd & Downie, 2008; Jones *et al*., 2007). NFRs are activated by NFs and subsequent stimulation of downstream signaling pathways through nuclear Ca^{2+} spiking. A perinuclear-anchoredcation channel, MtDMI1 (Does-not-make-infections1) LjCASTOR/ LjPOLLUX plays a critical role for upstream Ca^{2+} spiking during early *rhizobia* infection, and its function appears to be regulated by an upstream component, MtDMI2/LjSYMRK (Symbiosis receptor kinase)/MsNORK (Nodulation receptor kinase), which is a member of the LRR-RLK family (Leucine rich repeat-receptor like kinase) (Madsen *et al*., 2003; Catoira *et al*., 2000; Endre *et al*., 2002). DMI1 and DMI2 act downstream of NFRs in the Nod factor signaling pathway [5]. Ca^{2+}-related MtDMI3 is a nuclear-localized CCaMK (Ca^{2+}-calmodulin-dependent kinase) that functions downstream of Ca^{2+} spiking (Tirichine *et al*., 2006). Activated DMI3 directly interacts with and activates nodule-related transcription factors, including Nodulation signaling pathways (NSP1 and NSP2), the ethylene response factor (ERF) that is required for nodulation (ERNs) and the ERF that is required for nodule differentiation (EFD). These transcription factors enhance the expression of Nod factor-responsive genesby directly binding to the NF-box (Gleason *et al*., 2006; Murray *et al*., 2007; Tirichine *et al*., 2007; Hirsch *et al*., 2009). However, they are not distinct in how DMI3 regulates nodule-related nodule inceptions (NIN) and

ERN. How EFD inhibits nodulation is probably by disrupting cytokinin (Ryu *et al*., 2012).

Most *rhizoium* species are unable to fix nitrogen in the free-living state. Housing within the nodulation and differentiation into bacteroids is essential for *rhizobial* nitrogen fixation. Oxygen-binding leghemoglobin of the host maintains the microaerobic environment necessary for bacteriod nitrogenase activity. Ammonia (um) is the main product of N_2 fixation that is released from bacteroids and transported via an ammonium transporter across the peribacteroid membrane (PBM) to the plant where initial assimilation into amino acids (AAs) occurs (White *et al*., 2007; White *et al*., 2009). Four key enzymes are activated for the primary assimilation of NH_4^+ in nodules. Glutamine synthetase (GS) and glutamate synthase (GOGAT) are collectively referred to as the GS-GOGAT pathway, the conjunction of aspartate aminotransferase (AAT) and asparagine (ASN) synthetase (AS) (Sprent & Janes, 2007). So far, nitrogen fixation's genes of the host have been identified by analyses of Fix^- mutants, such as three *L. Japonicus* genes, SST1 (symbiotic sulfate transporter), FEN1 (fail in enlargement of infected cells) and IGN1 (ineffective greenish nodules 1) and one *M. truncatula* gene, DNF1 (defective in nitrogen fixation) (Wang *et al*., 2010).

Soybean (*Glycine max* (L.) Merr.) is one of the most important legume crops for seed protein and oil content. In the past, only a few genes of soybean have been identified through genetic analysis and positional cloning (Hayashi *et al*., 2012; Hayashi *et al*., 2014). As functional genomics and sequences develop, studies of transcriptional profiles during nitrogen fixation are important to gain a greater understanding of the related nitrogen fixation genes. Some studies use soybean evaluated nodulation gene-expression profiles in roots inoculated with *B. japonicum* and elucidated reduction of plant defenses (Brechenmacher *et al*., 2008; Libault *et al*., 2010; Carvalho *et al*., 2013). The bacteroid differentiation and their nitrogen fixation are under strict control with complex interactions between the host legume cells and the intracellular bacteria; however, the mechanisms underlying the differentiation of endosymbiotic *rhizobia* in symbiosomes to the bacteroid form are still largely unknown.

To improve our knowledge of genetic regulation during later differentiation nodulation and nitrogen fixation activation, we analyze the soybean's gene expression pattern of the R1 stage (first flower) inoculated with *rhizobia* or mock-inoculated. The R1 stage of soybean is the most powerful stage of nitrogen fixation. This study aimed to analyze the global expression change of genes in soybean roots. The results provide interesting insight into the evolution of genes specifically involved in the nitrogen fixation process.

Materials and Methods

Bacterial strains and plant material: *Bradyrhizobium japonicum* USDA110 and *Sinorhizobium fredii* CCBAU45436 were grown at 28°C in a darkroom in liquid yeast extract mannitol broth medium(YMB) (pH 6.8) with moderate shaking (120 rpm). After 6 d, cells of USDA110 and CCBAU45436 were amassed by centrifugation (4000 rpm, 10min), washed 3 times with sterile water, and diluted in water to an optical density at OD_{600} = 0.8.

Soybean (*Glycine max*) seeds of cultivar *Nannong 1138-2* were surface sterilized by soaking for 8 min in 0.1% $HgCl_2$ and rinsed 5 times with sterile water. Seeds were germinated and grown in sterile ceramic pots containing sterile soil with 5 seedlings per pot under greenhouse conditions and natural light. The plants were inoculated with inoculums containing USDA110, CCBAU45436 (50 ml of liquid suspension per pot at 0.8 OD_{600}) at stage V1 respectively. Mock-inoculated plants (CK) received the same amount of autoclaved inoculums. Soybeans continued to grow until R1 stage. Roots were harvested. Subsequently, the roots were extirpated nodulation and separated from shoots, then immediately frozen in liquid nitrogen and stored at -80ºC.

RNA extraction, isolation of mRNA and cDNA synthesis: Total RNA was isolated from the roots of each treatment using Trizol reagent (Tiangen, Beijing), according to the manufacturer's instructions. After extraction, the quality and quantity of the total RNA were analyzed by Nanodrop (Thermo Fisher Scientific Inc.) and Bioanalyser 2100 (Agilent Technologies). Pair-end index libraries were constructed according to the manufacture's protocol (NEBNext®Ultra™ RNA Library Prep Kit for Illumina®).

The isolation, fragmentation and priming of ploy(A)'s mRNA were performed using NEBNext Poly(A) mRNA Magnetic Isolation Module. The first and second strand cDNA were synthesized using ProtoScript Π Reverse Transcriptase and Second Strand Synthesis Enzyme Mix, respectively. The double-stranded cDNA was purified using AxyPrep Mag PCR Clean-up (Axygen) and treated with End Prep Enzyme Mix for end repairing. Then, the cDNA was added to 5'phosphorylation and dA-tailing in one reaction and was ligated to adaptors with a "T" base overhang.

Sequencing and sequence alignment: Approximate 400 bp fragments of adaptor-ligated DNA (with the approximate insert size of 250 bp) were recovered using AxyPrep Mag PCR Clean-Up (Axygen). Two samples of USDA110 and CK were amplified by PCR for 11 cycles using P5 and P7 primers which can anneal with the flow cell to perform bridge PCR and P7 primer carrying a six-base index allowing for multiplexing. The PCR productions were furthercleaned up using AxyPrep Mag PCR Clean-up (Axygen), validated using an Agilent 2100 Bioanalyzer (Agilent Technologies), and quantified by Qubit and real time PCR (Applied Biosystems). Then libraries with different indexes were multiplexed and loaded on an Illumina HiSeq (Genewiz, Beijing) instrument according to the manufacturer's instructions (Illumina, San Diego, CA, USA). Sequencing was carried out using a 2 × 100 paired-end (PE) configuration; image analysis and base calling were conducted by the HiSeq Control Software (HCS) + OLB + GAPipeline-1.6 (Illumina) on the HiSeq instrument.

The results of the sequencing were aligned against the Williams 82 genome sequence (ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Gmax/) using tophat-2.09 software, allowing a maximum of two mismatches.

Bioinformatics analysis: The gene expression level was calculated by FPKM

$$(\text{FPKM} = \frac{\text{Unique\_mapped fragment's number on a transcript per million reads}}{\text{Unique\_mapped fragment's number on whole genome} \times \text{base number of a gene}})$$

using Rsem (v1.2.4) software. The relation between the expressed difference and expressed level was calculated using a $\log_2^{\text{count}}$ (average value of $\log_2^{\text{counts per million reads of transcript expression}}$) and $\log_2^{\text{FC}}$ ($\log_2^{\text{fold change of transcript expression}}$). We calculated the differences in gene expression among different samples with EdgeR (v3.4.2) and the *p*-value of the significant difference. A *p*-value less than 0.05 was determined as a significant difference.

GO enrichment analysis of differentially expressed genes using default GO association files was performed with the "GO Term Finder" (http://go.princeton.edu/cgibin/GOTermFinder) where the statistical significance (*p*-value) was calculated based on a hypergeometric distribution with a bonferroni multiple testing correction and false discovery rate (FDR) calculation as described. The GO enrichment analysis was performed with adjusted GO association files, Ontologizer (http://www.charite.de/ch/medgen/ontologizer/), which used the"GO Term Finder" with a threshold *P* value of less than 0.01. The pathway annotation and enrichment of differentially expressed genes were performed using KAAS software.

Validation of RNA-Seq data by quantitative real-time PCR (qRT-PCR) and results according with other *rhizobium*: Quantitative real-time PCR experiments were performed to validate the RNA-Seq results for 19 transcripts whose expression differed by more than 2.0-fold among USDA110, CCBAU45436 and CK. Primers for qRT-PCR were designed using Primer 3 software (http://frodo.wi.nit.edu/primers). Approximately 2 μg of purified total RNA were reverse transcribed using M-MLV reverse transcriptase (Invitrogen) with Oligo $(dT)_{20}$ as the primer(Invitrogen), according to the manufacturer's instructions. Quantitative PCR was performed on an ABI ViiA™ 7 (Applied Biosystems) with the Light Cycler system. The PCR mixture (final volume 20 μl) included 0.4 μl (approximately 50 ng) of first strand cDNAs, 0.5 μM of each primer, and 10μl of 2 × SYBR® Green Realtime PCR Master Mix (TOYOBO). The cycling conditions were as follows: 2 min for denaturation at 95℃ followed by 40 cycles of 95℃ for 10 s, 60℃ for 15 s, and 72 ℃ for 25 s. Nineteen candidate gene's primers (Table 1) were designed according to every candidate gene sequence (http://soybase.org/gbrowse/cgi-bin/gbrowse/gmax1.01). Expression levels of these genes were normalized by tubulin (NCBI accession No. AY907703). Gene expression was quantified using the relative quantification ($\Delta\Delta C_T$) method and the data was compared with internal controls. Each sample was replicated three times.

Results and Discussion

RNA-Sequencing analysis: Level changes of transcripts in the plant root were analyzed between the inoculated with CK and USDA110 at R1 stage by RNA-Seq. 71.05 and 76.47 million reads were respectively generated by 101 bp sequencing from the cDNA libraries of CK and USDA110, constituting 7.18 and 7.72 Gb of cDNA sequence (raw data). The raw data were filtrated using the NGSQC Toolkit (V2.3) according to the following processes: (1) The joint sequences were removed; (2) These reads were retained if these loci of the quality value were higher than Q30 (percentage of error probability less than 0.1%) which accounted for 70% or more of the total length of reads; (3) A base of 3'-end was removed; (4) The reads of length less than 70 bp were removed. 84.56% and 85.44% of the sequenced reads were high quality filtered reads in CK and USDA110, respectively (Table 2). 91.94% and 92.99% clean data of CK and USDA110 were respectively successfully aligned to the soybean genome reference sequence (Glyma 1.01, ftp://ftp.jgi-psf.org/pub/compgen/ phytozome/v9.0/Gmax/) using Tophat-2.09 software. Therefore, the CK and USDA110's reads are uniquely aligned to the reference genome (unique mapped) and respectively accounted for 36.63% and 36.70% of the total clean data (Table 3).

Table 1. Primer sequences used for qRT-PCR.

Transcripts name	Forward	Reverse
Glyma01g24510.2	AGGGTGGTGGGAGACTACGT	GCGTGGCGATCTCCTTGATC
Glyma01g44400.3	AAATCTCGGGCTTCAACTGC	CCACAAGAGATGAAATAGCT
Glyma02g40820.2	TATCTGGAAGTCAATCAAGG	ACTTTATCATCGGTCTCATC
Glyma05g07020.1	TGGAGGTGTTATTCGGCAGT	GCCAATAGAGTCAGCACCAT
Glyma05g34960.7	GCTGTGGTGAAGGAGGAGAT	TCAAATGTGGAAGGGCAGAT
Glyma06g05280.1	CCAAAAAGGAGCAAAACACT	GGACGAAGCAGATTTAGAAC
Glyma08g19250.1	TAGGATGGTAATTGCGAGGC	GTTTACATTCCTGGGCACCT
Glyma10g04410.3	GGGCTGCTGGGAAGAAAAAG	TTTGCTGCTGCCACCTTCTG
Glyma11g27480.1	TGTGCTTGGTTGCTCTGATT	TCAACTATGGCTAACCTTTG
Glyma11g36000.2	TCTTCTTCTTGTTTGAGCAC	CAAGGAAGGACCAAGGAACT
Glyma13g23790.1	ACTAGCACAACCACACCTAG	GTAGAACACCCCCACAATTT
Glyma13g39450.1	ACTGGCTTCCTCGCAAAAAT	GTATCCCTGGCTCTCAAAAG
Glyma14g00760.2	CTCTGTCATCTCTATCTGCC	TTGTCTTGCAGGAGCTTGCC
Glyma14g06340.2	ACCATGAAAGAAGCCCTAAC	CTATTCTGTGGAGCGGTTTC
Glyma14g40320.2	TTCTCTCTACAGTAACAAGG	TTATCACCTTCTCCCAACAC
Glyma15g16470.2	CAGGGAATCGTTTGAAGAAT	CGAGAAAGCCACCTAACCAT
Glyma17g33050.1	TCTTGTTCTCTGTCCACCAC	GCCTTCAAACCGAGATACAT
Glyma17g34110.2	CTTTCTTCACACCCCTCCAT	GTTGGAAGAGGTTTGTAGGG
Glyma19g25360.2	CGTCCTCCTCTGCCGATTCT	CGCAGACGGTGGTTTTTCCC
Tubulin	GGAGTTCACAGAGGCAGAG	CACTTACGCATCACATAGCA

Table 2. Statistics of the Illumina HiSeq reads and comparison to the *G. max* reference genome (gmax1.01).

Samples		Number	Length	Size of bases (bp)	Q20 (%)	Q30 (%)	GC (%)	N(ppm)
CK	Raw data	71053298	101	7176383098	95	91	51	66
	Clean data	60681336	100	6068133600	99	97	50	35
USDA110	Raw data	76471968	101	7723668768	96	91	50	69
	Clean data	65992290	100	6599229000	99	97	50	36

Note: Number is number of reads. Length is average length of reads. Q20 is percentage of error probability less than 1%. Q30 is percentage of error probability less than 0.1%. GC = (G + C) number/ total bases number × 100%. N is content of unidentifiable bases per million bases

Table 3. Results of clean reads were aligned to reference genome.

Samples	Total_mapped	Unique_mapped	Multi_mapped	PE_mapped	Mapped percent (%)
CK	55793166	20436743	35356423	27327800	91.94
USDA110	61369446	22522473	38846973	30114865	92.99

Note: Total_mapped is the total number of reads aligned to the reference genome. Unique_mapped is the number of reads which were aligned to exclusive positions of the reference genome. Multi_mapped is the number of reads which were aligned to multi-positions of the reference genome. PE_mapped is the number of reads of PE sequencing double ends which were synchronously aligned to the reference genome

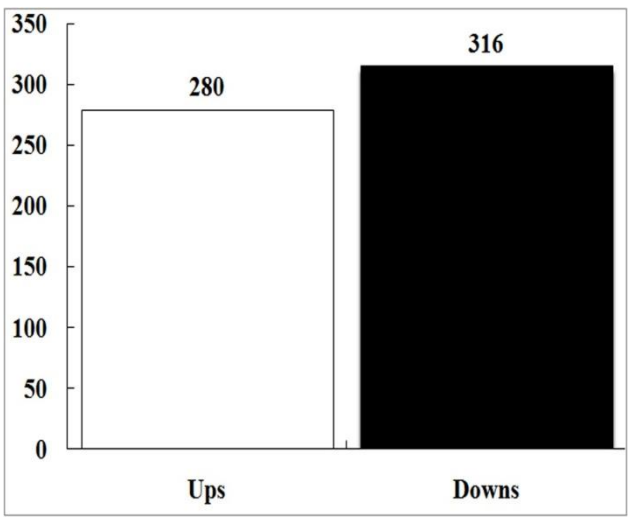

Fig. 1. Number of differentially expressed transcripts in soybean inoculated with CK VS USDA110.

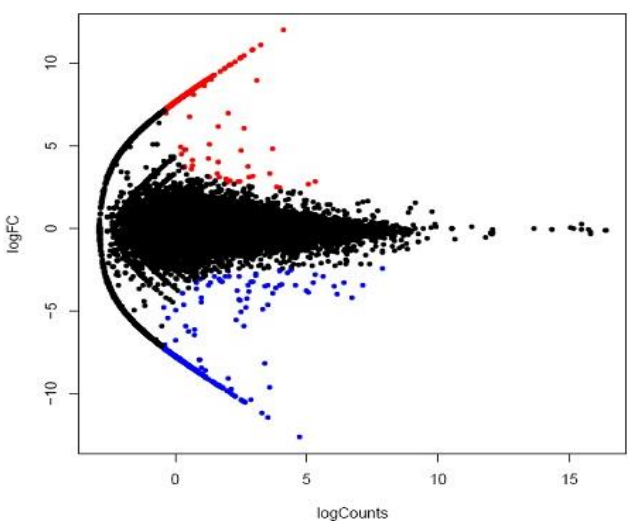

Fig. 2. Relation between transcripts expression difference and counts $\log^{FC}$ is the value of $\log_2$fold change of transcript expression. $\log^{Counts}$ is the average value of $\log_2$counts per million reads of transcript expression. The red dots represent the up-regulated transcripts. The blue dots represent the down-regulated transcripts. The black dots represent no significantly differential expressed transcripts.

The differentially expressed genes profiles responding to *rhizobium* inoculation: A total of 55787 transcripts were quantified based on the FPKM value of unique-mapped fragments for the analysis of gene expression. There are 596 differentially expressed transcripts (p-value ≤0.05 and fold-change ≥2.0) produced by the edger program of the Bioconductor software through using fragments counts. Of these, 280 transcripts were up-regulated and 316 were down-regulated in inoculated with USDA110 VS CK (Fig. 1) in the R1 stage. Most of the $\log_2^{Counts}$ values of transcripts expression level were less than 5.0, and the correlation relation was as shown in Fig. 2, representing the difference between transcripts expression levels. There were more down-regulated than up-regulated transcriptions in inoculated USDA110 VS mock-inoculated. There were many down-regulated transcripts which indicated that there were more negatively regulated genes than positively regulated ones with functions in the nitrogen fixation pathway.

Gene ontology category analysis: As a useful tool for gene functional annotation, WEGO (Web Gene Ontology Annotation Plot) has been widely used in many soybean studies (Li *et al*., 2013). It has become one of the most utilized tools for downstream gene annotation analysis studies. In this research, GO assignments were used to classify the functions of the predicted nitrogen fixation genes. A total of 39163 expressed genes of two materials were converted into GO-identities (IDs) (p-value ≤0.01) and classified into three functional categories; the cellular component, molecular function and biological process by mapping to the GO Term Finder (http://search.cpan.org/dist/GO-TermFinder/lib/GO/TermFinder.pm) (Fig. 3). The differentially expressed genes (DEGs) (464) according to three categories with a gene ontology annotation were further classified into subsets. There were 5 subsets within the cellular component category, 3 subsets within the molecular function category, and 11 subsets within the biological process category. The genes were classified as follows: 104 genes were mapped to the cell component category; 182 genes mapped to the molecular function category; and 178 genes mapped to the biological process category. Five subsets of the cellular component category are cell or cell part, organelle or organelle part, membrane-enclosed, membrane or membrane part, and non-

membrane-bounded organelle. Three subsets of molecular function category are catalytic activity, transporter, and binding. Eleven subsets of biological process category are metabolic process, biological regulation, cellular component organization or biogenesis, cellular process, reproduction, multicellular organismal process, signaling, developmental process, single-organism process, response to stimulus, and localization (Fig. 4). The most abundant unigenes were related to cell or cell parts in the cellular component category, catalytic activity and binding in the molecular function category, and metabolic process and response to stimulus and localization in the biological process category. The results showed that symbiotic fixation nitrogen brought on a series of plant gene changes.

Pathway enrichment analysis of DEGs: Genes usually interact with each other to carry out certain biological functions. Knowledge of the expressions of multiple genes and their regulation in symbiotic fixation nitrogen biosynthesis is required to further understand the regulatory mechanisms. Pathway-based analysis helps to clarify the biological functions of genes and signal transduction pathways associated with DEGs compared with the whole genome background. For our research, 169 biological pathways were identified by Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. There are 40 pathways related to carbon cycle and metabolism, 16 pathways related to amino acid synthesis and metabolism, 8 pathways related to hormones, 8 pathways related to signaling pathways, 6 pathways related to vitamins, 3 pathways related to alkaloids, and others related to RNA and viruses, etc. From this, a nitrogen metabolism and a NOD-like receptor (NLR) signaling pathway were identified. A total of 98 differentially expressed genes (115 transcripts) with pathway annotations were identified by the KAAS software. The expression level of 54 transcripts was up-regulated, and 61 transcripts were down-regulated. The number of transcripts associated with pathways ranges from 1 to 44, with an average number of 4.38 pathways per transcript. 23 transcripts were associated with one pathway, and 92 transcripts were associated with two or more pathways. The transcripts of Glyma10g04410.2 were involved in 44 pathways; also, Glyma14g00760.2 was involved in 17 pathways. However, 481 different expressed transcripts (80.70%) weren't found in any of the previous annotated pathways using KAAS software. The results may be due to nodulation and symbiotic nitrogen fixation of soybean evolved new pathways which are different to other biological models.

From those pathways, we selected the NOD-like receptor signaling pathway (Fig. 5) and nitrogen metabolism pathway (Fig. 6) for further analysis. There is only one up-regulated transcript (Glyma14g40320.4) in the NLR signaling pathway. There are seven differentially expressed transcripts associated with the nitrogen metabolism pathway. In all, 7 transcripts were down-regulated, which may be due to the symbiosis provided nitrogen nutrition for plants.Some nitrogen transformation genes don't need to activate their expression.

GO analysis and the pathway enrichment analysis of DEGs showed that most are related to carbon and amino acid synthesis or metabolism. In the symbiotic fixation nitrogen, the NOD-like receptor signaling pathway, the nitrogen metabolism and the plant hormone play an important role. The Hsp90 of the host plant was up-regulated in inoculated *rhizobium*. Hsp90 is an abundant, dimeric ATP-dependent molecular chaperone, and ATPase activity is essential for its functions (Retzlaff *et al*., 2009). The NLRs are classified as part of the signal transduction ATPases with numerous domains (STAND) clade within the AAA+ ATPase family (Proell *et al*., 2013). The interaction of heat shock protein 90 (Hsp90) and suppressor of the G2 allele of Skp1 (SGT1) activates Nod1 (Correia *et al*., 2007). The nodulation ability of host plants could be improved by boosting the expression level of Hsp90.

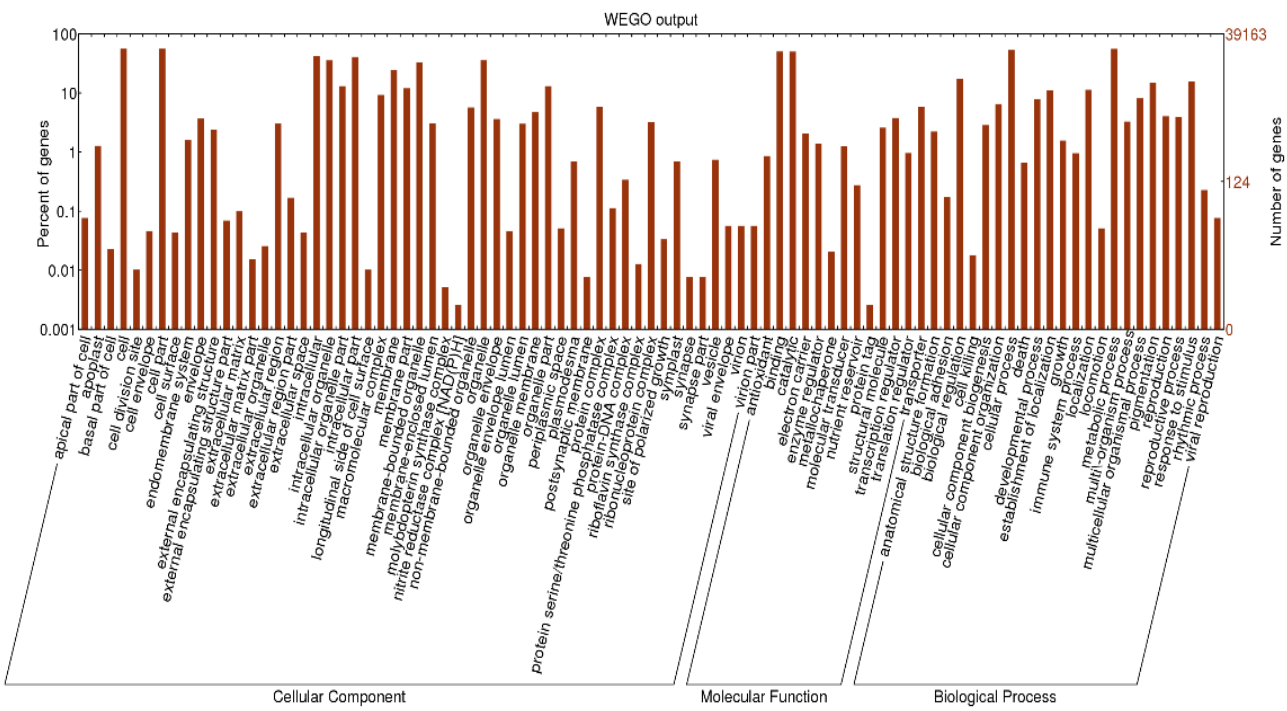

Fig. 3. Functional categorization of expressed genes in plant roots inoculated with CK and USDA110.

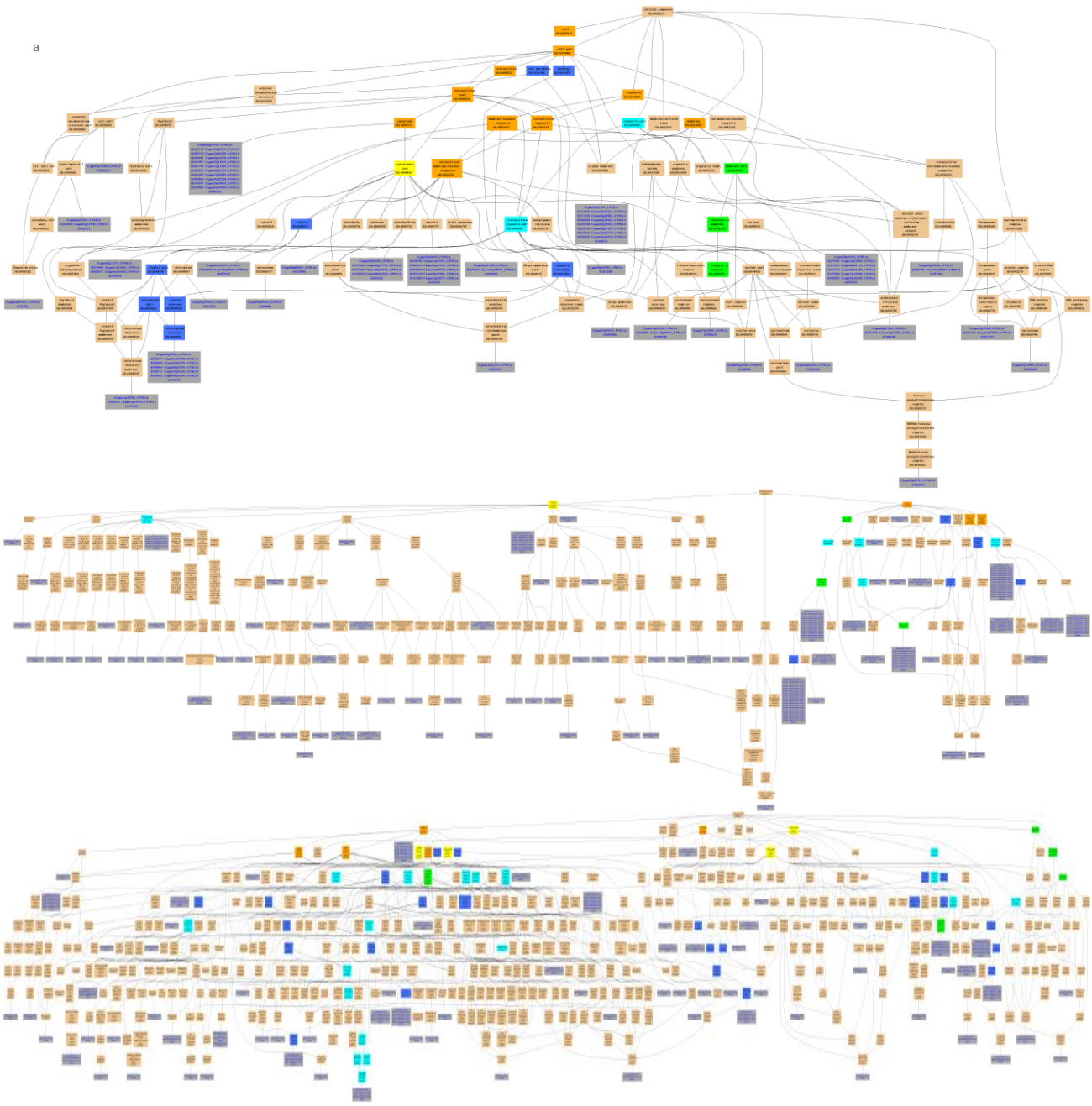

Fig. 4. Cellular Component, Molecular Function and Biological Process of differentially expressed genes in soybean roots inoculated with CK VS USDA110
a: Cellular Component; b: Molecular Function; c: Biological Process
Each level corresponds to a category level of GO. The differentially expressed genes were identified in different GO terms.

In nitrogen metabolism, the nitrate reductase (NAD (P)H), ferredoxin-nitrite reductase and carbonic anhydrase of the host plant were down-regulated by inoculating rhizobium. The enzyme nitrate reductase, ferredoxin-nitrite reductase and carbonic anhydrases respectively, catalyze the reduction of nitrate to nitrite (Okamoto & Marzluf, 1993), the six-electron reduction of nitrite (oxidation state +3) to ammonium (oxidation state -3) in the second step of the nitrate assimilation pathway (Takahashi *et al*., 2001) and the reversible hydration of CO_2 to form HCO_3^- and protons (Breton, 2001). The host plant demands ammonium which can be provided by symbiotic nitrogen fixation. Fewer nitrate nitrogen molecules were absorbed from the soil, so the activity of three reductases decreased.

The plant hormone regulates a number of developmental and physiological processes including nodulation. For example, a cytokine in receptor triggers spontaneous root nodule organogenesis (Murray *et al*., 2007; Tirichine *et al*., 2007). *Rhizobia* and external stress signals activate mitogen activated protein kinase (MAPK), signaling cascades and the action of plant hormones including ethylene, salicylic acid (SA), abscisic acid (ABA), and jasmine acid(JA) which were negatively regulated nodulation (Ryu *et al*., 2012). In this study, jasmonate ZIM domain-containing protein (JAZ) was down-regulated by inoculated *rhizobium*. The result is consistent with previous studies.

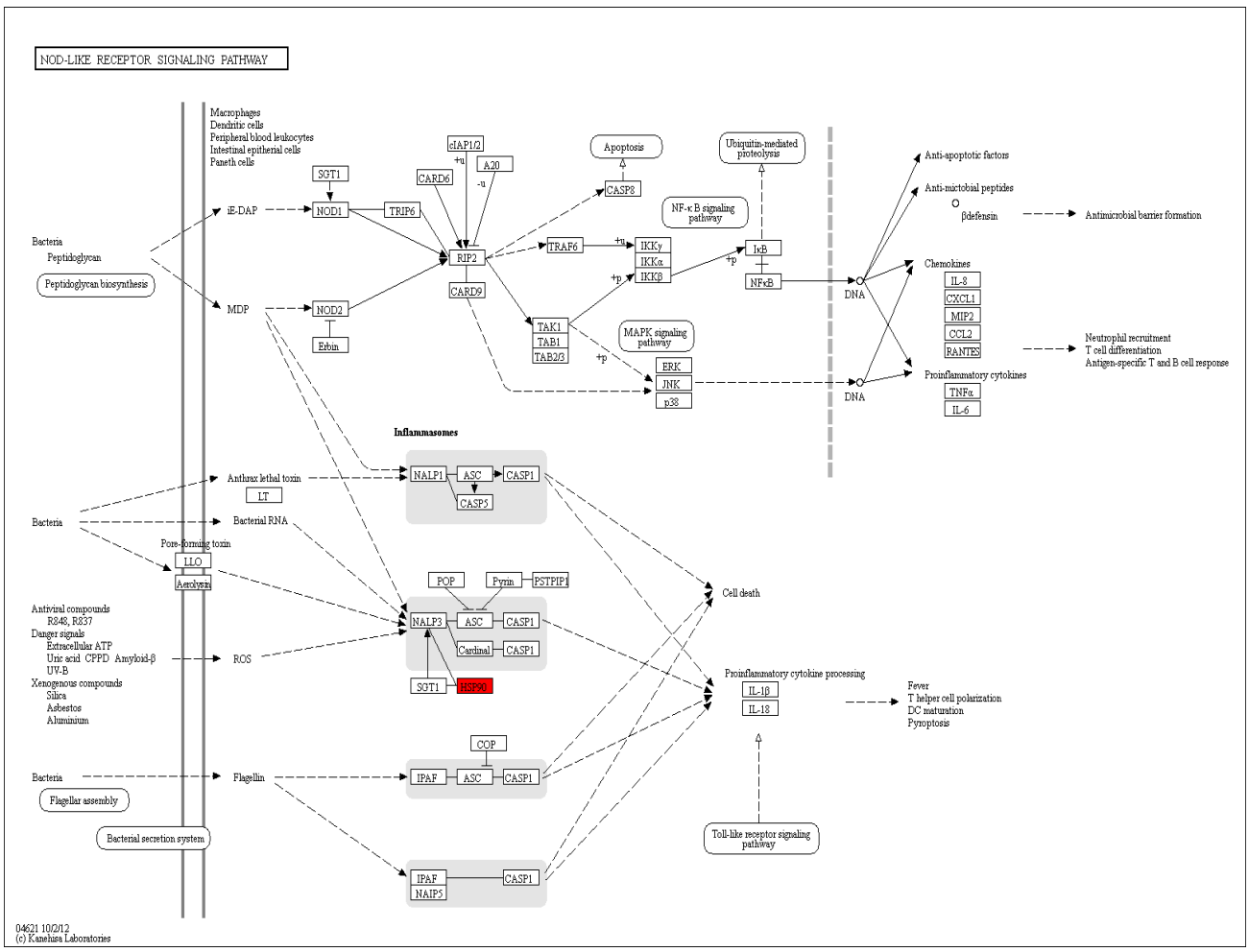

Fig. 5. NOD-like receptor signaling pathway

iE-DAP: γ-D-glutamyl-meso-diaminopimelic acid; MDP: muramyldipeptide; SGT1: suppressor of the G2 allele of Skp1; NOD1, NOD2: Nod-like receptor 1, 2; TRIP6: TAR RNA interacting protein 6; RIP2: receptor-interacting protein2; CARD6, 9: caspase recruitment domain family 6, 9; cIAP1/2: cellular inhibitor of apoptosis protein1/2; A20: ubiquitin-editing enzyme A20; CASP1, 5, 8: cysteine-aspartic acid protease1, 5, 8; TRAF6: eceptor associated factor6; TAK1: transforming growth factor-β-activated kinase1; TAB1, 2/3: TAK1, 2/3 binding protein; IKKα, β, γ: inhibitors of kappa B kinaseα, β, γ; ERK: extracellular signal-regulated kinase; JNK: c-Jun N-terminal kinase; IκB: inhibitors of kappa B; NFκB: nuclear factor kappa B; NALP1: nucleotide-binding oligomerization domain-like receptor family, pyrin domain-containing 1; ASC: apoptosis-associated speck-like protein containing; POP: persistent organic pollutant; PSTPIP1: proline-serine-threonine phosphatase-interacting protein 1; NALP3: leucine-rich repeat/pyrin domain-containing-3; HSP90: heat shock protein90; IPAF: interstitial pneumonia with autoimmune features.

Note: Red rectangles indicate up-regulated genes. Green rectangles indicate down-regulated genes. The same below

Validation of RNA-Seqdataby qRT-RCR and the result's universality in other *rhizobium*: To confirm results of the expression patterns validity by RNA-Seq in the soybeans inoculated with mock-inoculated, *B. rhizobium* USDA110 and the universality inoculated with other *rhizobium*, we used qRT-PCR to analyze the expression levels of 19 candidate genes of soybeans' roots inoculated with CK, USDA110 and CCBAU45436. Results showed that significant differences of the expression levels between the inoculated and mock-inoculated *rhizobium* for 19 transcripts (Fig. 7). Ten transcripts (Glyma 01g44400.3, Glyma02g 40820.2, Glyma 08g19250.1, Glyma11g 36000.2, Glyma13g 39450.1, Glyma14 g40320.2, Glyma15 g16470.2, Glyma17g 33050.1, Glyma17g34110.2 and Glyma19g 25360.2) were significantly up-regulated in soybean roots inoculated with *rhizobium*. The other nine transcripts were significantly down-regulated. Although the RNA-Seq values showed slight variations compared to that of values observed from the qRT-PCR results, the tendency of the expression level from RNA-Seq analysis was consistent with those obtained by qRT-PCR. The expression levels of 19 transcripts in soybean roots inoculated with USDA110 strain were similar to the soybean inoculated with CCBAU45436 strain. These results highlight the fidelity and reproducibility of the RNA-Seq analysis used in the present study and have a certain extent of universality in the different *rhizobium* strains.

Unique-mapped reads were selected to profile the soybean transcripome which would be propitious to decrease false positives for differentially expressed genes. 19 genes in different pathways were selected to validate the credibility of RNA-Seq using qPCR. The validation results of qPCR were consistent with the transcripts' expression patterns identified by RNA-Seq inoculation of USDA110 strains in soybean root. These two results are similar to the results of inoculation with CCBAU45436. These showed that the results of RNA-Seq have a certain extent of universality in soybean inoculated with different *rhizobium* strains.

Previously, QTL mapping was used to locate biological nitrogen fixation traits in soybean along with 16 gene/QTL regions of 12 chromosomes (Madsen *et al*., 2003; Hayashi *et al*., 2012; Yang *et al*., 2010; Searle *et al*., 2003; Tanya *et al*., 2005; Nicolás *et al*., 2006; Santos *et al*., 2013). There are significantly differential expressed genes of using RNA-Seq in these QTL regions. For example, three significantly differential expressed transcripts (Glyma11g08720.2, Glyma11g09250.2 and Glyma11g09630.2) between inoculated and mock-inoculated *rhizobium* were located between Satt509 and Satt251 on chromosome 11 in the study of *Santos*. However, the *Rj*/*rj* genes involved in nitrogen-fixation root nodule formation in soybean were not detected in the study. These results may be because the difference of the nodulation gene was not significantly different when comparing the inoculated and mock-inoculated *rhizobia* at the soybean's R1 stage. More novel genes were identified by RNA-Seq in the study and were not located by previous gene/QTL mapping studies. Their function needs to be further verified using other methods.

In the study, 280 and 316 transcripts were up- and down-regulated in the R1 stage of soybean, respectively. Gene Ontology analyses detected 5, 3 and 11 subsets within the cellular component, molecular function and biological process category, respectively. A total of 169 biological pathways were identified by KEGG pathway analysis. Putative functions for some of these genes were assigned for the first time in the *rhizobium*-soybean symbiosis. Novel genes were firstly described and could be related to the nitrogen fixation process.

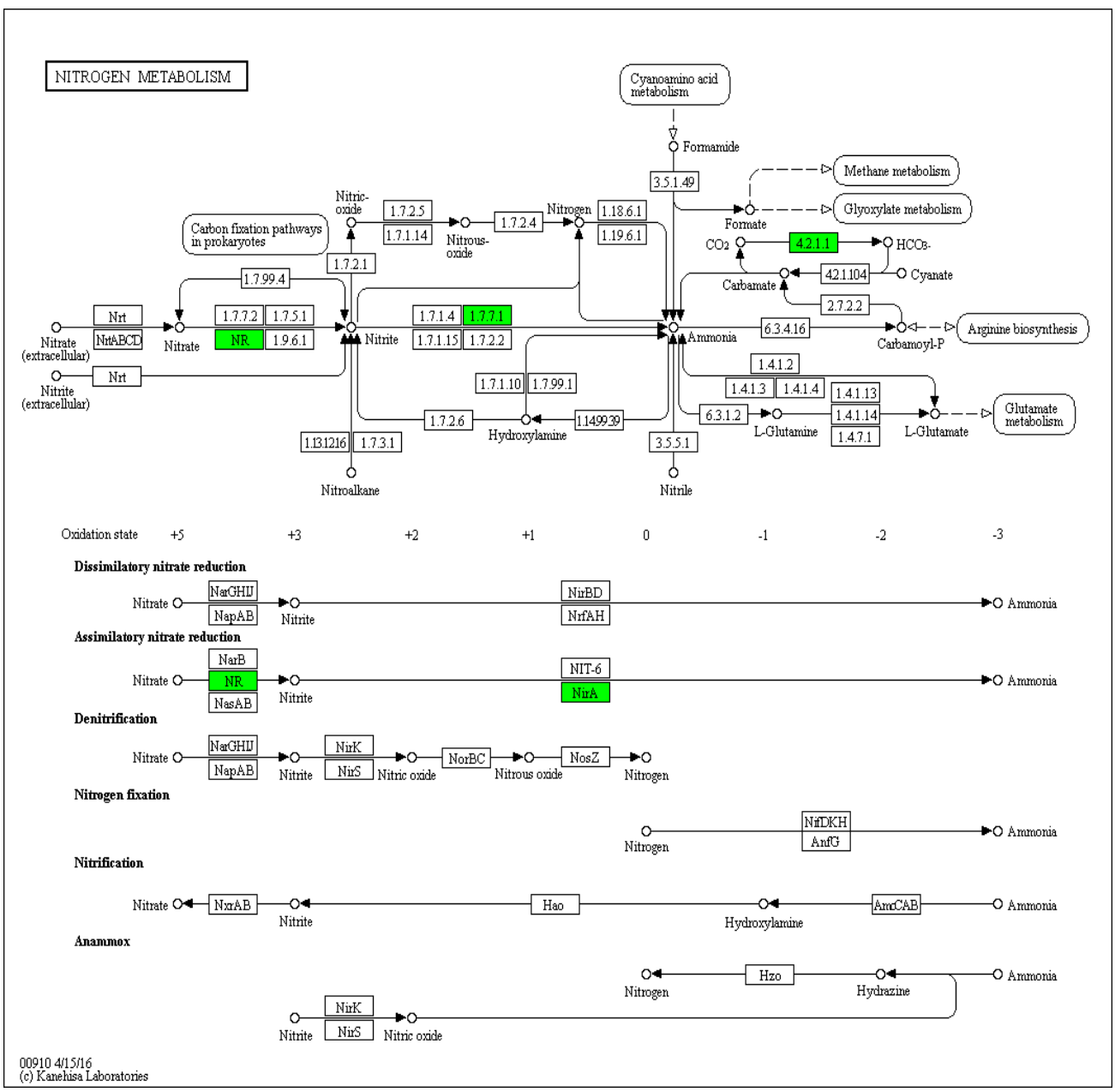

Fig. 6. Nitrogen metabolism pathway

Nrt: nitrate transporter; NR/Nar/Nir: nitrate reductase; Nap: periplasmic nitrate reductase; Nas: assimilatory nitrate reductase; Nxr: nitrite oxidoreductase; Nor: nitric oxide reductases; NosZ: N_2O reductase; Nif: nitrogen fixation; Anf: the structural genes of nitrogenase; Hao: hydroxylamine oxidoreductase; AmoCAB: ammonia monooxygenaseCAB; Hzo: hydrazine oxidoreductase.

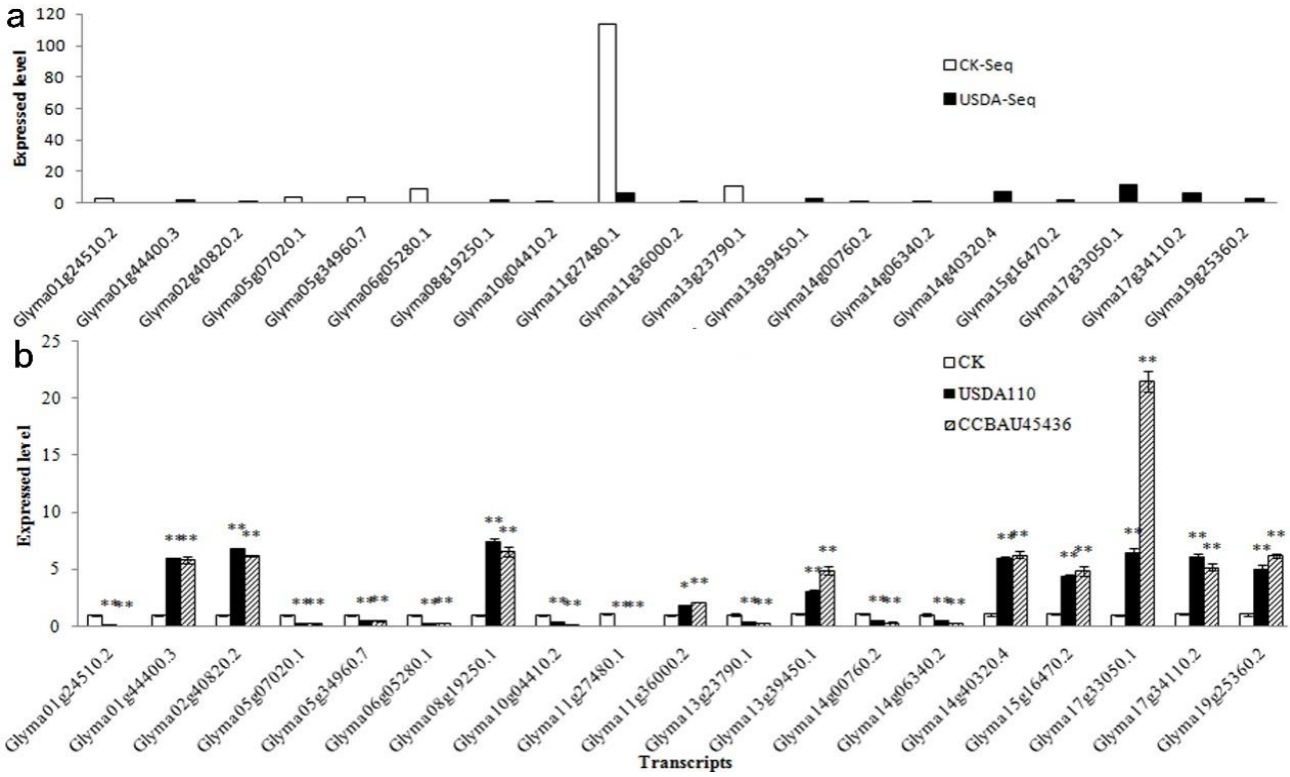

Fig. 7. Analysis of expressed level of 19 representative transcripts.
a: The expressed level of 19 differentially expressed transcripts using RNA-Seqin CK and USDA110, b: The expressed level of 19 differentially expressed transcripts using RT-PCR in CK, USDA110 and CCBAU45436

Acknowledgements

The authors are gratefully indebted to Prof. C. F. Tian, China Agricultural University, for donating *rhizobia* strains. This work was supported by the Anhui Provincial College Program for Natural Science (KJ2013A077, KJ2016A172), the Science and Technology Support Program of Sichuan Province (2015NZ0046), the National Natural Science Foundation of China (31300569) and the Key Disciplines and Program for Stable Talents of Anhui Science and Technology University (AKZDXK2015B02).

References

Arrighi, J.F., A. Barre, A. Ben, AB. Ben, A. Bersoult, LC. Soriano, R. Mirabella, F. Carvalho-Niebel, EP. Journet, M. Ghérardi, T. Huguet, R. Geurts, J. Dénarié, P. Rougé and C. Gough. 2006. The *Medicagob truncatula* lysine [corrected] mtif-receptor-like kinase gene family includes NFP and new nodule-expressed genes. *Plant Physiol*., 142(1): 265-279.

Bladergron, M.R. and H.P. Spaink. 1998. Genes and signal molecules involved in the *rhizobia*-leguminoseae symbiosis. *Curr. Opin*. *Plant Bio*., 1(4): 353-359.

Brechenmacher, L., M.Y. Kim, M. Benitez, M. Li, T. Joshi, B. Calla, MP. Lee, M. Libault, LO. Vodkin, D. Xu, S.H. Lee and S.J. Clough. 2008. Transcription profiling of soybean nodulation by *Bradyrhizobium japonicum*. *Mol. Plant Microbe In*., 21(5): 631-645.

Breton, S. 2001. The cellular physiology of carbonic anhydrases. *JOP*., 2(4):159-164.

Carvalho, G.A.B., J.S.S. Batista, F.C.M. Guimarães, L.C. Nascimento and M. Hungria. 2013. Transcriptional analysis of genes involved in nodulation in soybean roots inoculated with *Bradyrhizobium japonicum* strain CPAC 15. BMC. *Genomics*, 14: 153.

Catoira, R., C. Galera, F. Billy, RV. Penmetsa, EP. Journet, M. Fabienne, C. Rosenberg, D. Cook, C. Gough and J. Denarie. 2000. Four genes *Medicago truncatula* controlling components of a nod factor transduction pathway. *Plant Cell*, 12: 1647-1665.

Correia, J.D.S., Y. Miranda, N. Leonard and R. Ulevitch. 2007. SGT1 is essential for NOD1 activation. *P. Natl. Acad. Sci. USA*. 104(16): 6764-6769.

Day, R.B., J.T. Loh, J. Cohn and G. Stacey. 2000. Signal exchange involved in the establishment of the *Bradyrhizobium*-legume symbiosis, E.W. Triplett In Prokaryotic Nitrogen Fixation, A model system for analysis of a biological process. Horizon Scientific Press, Norfolk U. K., pp. 385-414.

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

Gleason, C., S. Chaudhuri, T.B. Yang, A. Munoz, B.W. Poovaiah and G.E.D. Oldroyd. 2006. Nodulation independent of *rhizobia* induced by a calcium-activated kinase lacking autoinhibition. *Nature*, 441: 1149-1152.

Hayashi, M., S. Shiro, H. Kanamori, S. Mori-hosokawa, H. Sasaki-Yamagata, T. Sayama, M. Nishioka, M. Takahashi, M. Ishimoto, Y. Katayose, A. Kaga, K. Harada, H. Kouchi, Y. Saeki and Y. Umehara. 2014. A thaumatin-like protein, Rj4, controls nodule symbiotic specificity in soybean. *Plant Cell Physiol*., 55(9): 1679-1689.

Hayashi, M., Y. Saeki, M. Haga, K. Harada, H. Kouchi and Y. Umehara. 2012. *Rj* (*rj*) genes involved in nitrogen-fixing root nodule formation in soybean. *Breed. Sci*., 61: 544-553.

Hirsch, S., J. Kim, A. Munoz, A.B. Heckmann, J.A. Downie and G.E. Oldroyd. 2009. GRAS proteins form a DNA binding complex to induce gene expression during nodulation signaling in *Medicago truncatula*. *Plant Cell*, 21(2): 545-557.

Irving, H.R., N.M. Boukli, M.N. Kelly and W.J. Broughton. 2000. Nod-factors in symbiotic development of root hairs. In Root Hairs: Cell and molecular biology. Ridge RW, Emons AMC Springer, Tokyo, Japan p241-265.

Jones, K.M., H. Kobayashi, B.W. Davies, M.E. Taga and G.C. Walker. 2007. How *rhizobial* symbionts invade plants: the *Sinorhizobium-Medicago* model. *Nat. Rev. Microbiol*., 5(8): 619-633.

Kochi, H. and S. Hata. 1995. GmN56, a novel-specific cDNA from soybean root nodules encodes a protein homologous to isopropylmalate synthase and homocitrate synthase. *Mol. Plant Microb*., 8(1): 172-176.

Kouchi, H. Imaizumi-Anraku, M. Hayashi, T. Hakoyama, T. Nakagawa, Y. Umehara, N. Suqanuma and M. Kawaquchi. 2010. How many peas in a pod? Legume genes responsible for mutualistic symbioses underground. *Plant Cell Physiol*., 51(9): 1381-1397.

Li, Q.G., L. Zhang, C. Li, J.M. Dunwell and Y.M. Zhang. 2013. Comparative genomics suggests that an ancestral polyploidy event leads to enhanced root nodule symbiosis in the papilionoideae. *Mol. Biol. Evol*., Doi:10.1093/molbev/ mst152.

Libault, M., A. Farmer, L. Brechenmacher, J. Drnevich, R.J. Langley, D.D. Bilgin, O. Radwan, D.J. Neece, S.J. Clough, G.D. May and G. Stacey. 2010. Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. *Plant Physiol*., 152: 541-552.

Limpens, E., C. Franken, P. Simit, J. Willemse, T. Bisseling and R. Geurts. 2003. LysM domain receptor kinases regulating *rhizobial* Nod factor-induced infection. *Science*, 303: 1361-1364.

Madsen, E.B., L.H. Madsen, S. Radutoiu, M. Olbryt, M. Rakwalska, K. Szczyglowski, S. Sato, T. Kaneko, S. Tabata, N. Sandal and J. Stougaard. 2003. A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. *Nature*, 425: 637-640.

Murray, J.D., B.J. Karas, S. Sato, S. Tabata, L. Amyot and K. Szczyglowski. 2007. A cytokinin perception mutant colonized by *Rhizobium* in the absence of nodule organogenesis. *Science*, 315: 101-104.

Nicolás, M.F., M. Hungria and C.A.A. Arias. 2006. Identification of quantitative trait loci controlling nodulation and shoot mass in progenies from two Brazilian soybean cultivars. *Field Crop. Res*., 95: 355-366.

Okamoto, P.M. and G.A. Marzluf. 1993. Nitrate reductase of Neurosporacrassa: the functional role of individual amino acids in the heme domain as examined by site-directed mutagenesis. *Mol. Gen. Genet*., 240: 221-230.

Oldroyd, G.E. and J.A. Downie. 2006. Nuclear calcium changes at the core of symbiosis signaling. *Curr. Opin. Plant Biol*., 9: 351-357.

Oldroyd, G.E. and J.A. Downie. 2008. Coordinating nodule morphogenesis with *rhizobial* infection in legumes. *Ann. Rev. Plant Biol*., 59: 519-546.

Oldroyd, G.E., J.D. Murray, P.S. Poole and J.A. Downie. 2011. The rules of engagement in the legume *rhizobial* symbiosis. *Annu. Rev. Genet*., 45: 119-144.

Papadopoulou, K., A. Roussis and P. Katinakis. 1996. Phaseolus ENOD40 is involved in symbiotic and non-symbiotic organogenetic processes: Expression during nodule and lateral root development. *Plant Mol. Biol*., 30(3): 403-417.

Pichon, M., E.P. Journet, A. Dedieu, F. de Billy, G. Truchet and D.G. Barker. 1992. *Rhizobium meliloti* elicits transient expression of the early nodulin gene ENOD12 in the differentiating root epidermis of transgenic alfalfa. *Plant Cell*, 4(10): 1199-1211.

Proell, M., S.J. Riedl, J.H. Fritz, A.M. Rojas and R. Schwarzenbacher. 2013. The Nod-like Receptor (NLR) family: a tale of similarities and differences. *Plos One* doi:10.1371/journal.pone. 0002119.

Radutoiu, S., L.H. Madsen, E.B. Madsen, H.H. Felle, Y. Umehara, M. Gronlund, S. Sato, Y. Nakamura, S. Tabata, N. Sandal and J. Stougaard. 2003. Plant recognition of symbiotic bacteria requires two LysM receptor-like Kinases. *Nature*, 425: 585-592.

Retzlaff, M., M. Stahl, HC.. Eberl, S. Lagleder, J. Beck, H. Kessler and J. Buchner. 2009. Hsp90 is regulated by a switch point in the C-terminal domain. *EMBO. Rep*., 10: 1147-1153.

Ryu, H., H. Cho, D. Choi and I. Hwang. 2012. Plant hormonal regulation of nitrogen-fixing nodule organogenesis. *Mol. Cells*., 34: 117-126.

Santos, M.A., I.O. Geraldi, A.A.F. Garcia, N. Bortolatto, A. Schiavon and M. Hungria. 2013. Mapping of QTLs associated with biological nitrogen fixation traits in soybean. *Hereditas*, 150: 17-25.

Searle, L.R., A.E. Men, T.S. Laniya, D.M. Buzas, I. Iturbe-Ormaetxe, B.J. Carroll and P.M. Gresshoff. 2003. Long-distance signaling in nodulation directed by a CLAVATA1-like receptor kinase. *Science*, 299: 109-112.

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

Subramanian, S., G. Stacey and O. Yu. 2006. Endogenous isoflavones are essential for the establishment of symbiosis between soybean and *Bradyrhizobium japonicum*. *Plant J*., 48: 261-273.

Takahashi, M., Y. Sasaki, S. Ida and H. Morikawa. 2001. Nitrite reductase gene enrichment improves assimilation of NO_2 in Arabidopsis. *Plant Physiol*., 126: 731-741.

Tanya, P., P. Srinives, T. Toojinda, A. Vanavichit and S.H. Lee. 2005. Identification of SSR markers associated with N_2-fixation components in soybean [*Glycine max* (L.) Merr.] *Korean J. Genetics*, 27(4): 351-359.

Tirichine, L., H. Imaizumi-Anraku, S. Yoshida, Y. Murakami, LH. Madsen, H. Miwa, T. Nakagawa, N. Sandai, AS. Albrektsen, M. Kawaguchi, A. Downie, S. Sato, S. Tabata, H. Kouchi, M. Parniske, S. Kawasaki and J. Stougaard. 2006. Deregulation of a Ca^{2+}/calmodulin-dependent kinase leads to spontaneous nodule development. *Nature*, 441: 1153-1156.

Tirichine, L., N. Sandal, L.H. Madsen, S. Radutoiu, A. Albrktsen, S. Sato, E. Asamizu, S. Tabata and J. Stougaard. 2007. A gain-of-function mutation in a cytokinin receptor triggers spontaneous root nodule organogenesis. *Science*, 315: 104-107.

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

White, J., J. Prell, E.K. James and P. Poole. 2007. Nutrient sharing between symbionts. *Plant Physiol*., 144: 604-614.

White, J.P., J. Prell, V.K. Ramachandran and PS. Poole. 2009. Characterization of a γ-aminobutyric acid transport system of *rhizobium leguminosarum* bv. viciae 3841. *J. Bacteriol*., 191(5): 1547-1555.

Yang, S.M., F. Tang, M.Q. Gao, H.B. Krishnan and H.Y. Zhu. 2010. R gene-controlled host specificity in the legume-*rhizobia* symbiosis. *P. Natl. Acad. Sci*., USA., 107: 18735-118740.

(Received for publication …… 2017)