bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Transcriptomic Characterization of Soybean Roots in Response to *Bradyrhizobium* Infection by RNA Sequencing

Qingyuan He[1,2], Yang Hongyan[2] Ping Wu[1], Zhengpeng Li[1], Songhua Wang[1], and Changwei Zhu[1]

[1] College of Life Sciences, Anhui Science and Technology University, Fengyang 233100, Anhui, China

[2]Soybean Research Institute/ National Center for Soybean Improvement/ MOA Key Laboratory for Biology and Genetic Improvement of Soybean/ National Key Laboratory for Crop Genetics and Germplasm Enhancement Nanjing Agricultural University, Nanjing 210095, Jiangsu, China

With 7 figures and 3 tables

Received /Accepted

Communicated by

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT Legumes interact with *rhizobium* convert N_2 into ammonia for plant use. To investigate the plant basal nitrogen fixation mechanisms induced in response to *Bradyrhizobium*, differential gene expression in root of inoculated and mock-inoculated soybean was analysed by RNA-Seq. A total of 55787 transcripts were aligned to soybean genome reference sequences, 280 and 316 transcripts were found to be up- and down-regulated, respectively, in inoculated relative to mock-inoculated soybean's root at V1 stage. Gene ontology (GO) analyses detected 104, 182 and 178 genes associated with cell component category, molecular function category and biological process category, respectively. Pathway analysis revealed that 98 differentiallly expressed genes (115 transcripts) 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 consistent with those obtained from *rhizobia* infected RNA-Seq data. These showed that the results of RNA-Seq have reliability and universality. Additionally, this study showed some novel genes associated with nitrogen fixation process comparison with the previously identified QTLs.

KEYWORDS Symbiotic nitrogen fixation; Soybean; Transcriptomic; RNA-Seq analysis; *Rhizobium*

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Nitrogen is the most limiting element for crop growth and usually supplied by application of fertilizer, which brings on substantial costs to farmers and with potentially adverse effects on the environment. The leguminous plants establish a symbiotic relationship with *rhizobia* (symbiotic nitrogen fixation, SNF) to directly capture N_2 to support plant growth. Nitrogen conversion takes place in a unique organ (root nodule). The development of root nodules commences with a molecular dialogue between the host plant and a compatible strain of *rhizobium*, involving a succession of complex process that lead to profound changes in both symbionts (Oldroyd *et al.*, 2011).

The plant excretes molecular signals, flavonoids, phenolic compounds which induce the synthesis of specific *rhizobia*-produced lipo-chito-oligosaccharides, called Nod factors (NFs). NFs directly stimulate their putative receptor protein (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 *Medicago Trucatula* (Arrighi *et al.*, 2006; Limpens *et al.*, 2003; Madsen *et al.*, 2003; Radutoiu *et al.*, 2003), which are LysM (peptidoglycanbinding lysine motif) receptor-like of the legumes (Kochi *et al.*, 2010; Subramanian *et al.*, 2006). Many molecular events are triggered in a coordinate manner, leading to morphological and physiological changes in the host plant, necessary for a successful symbiosis (Oldroyd *et al.*, 2011). Bacterial attachment to the root hair and induces a calcium influx and membrane depolarization, subsequently lead to deformation and root fair curling, forming infection thread. Concomitantly, certain cortical cells divide to form nodule primorda, further development gives rise to nodules that differ from tumours 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 and Spaink, 1998), Such as ENOD2, ENOD12, ENOD40 (Kochi and 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 and Downie, 2006, 2008; Jones *et al.*, 2007). NFRs are activated by NFs and subsequent stimulation of downstream signaling pathways through nuclear Ca^{2+} spiking. A

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

perinuclear-anchored cation channel, MtDMI1 (Does-not-make-infections 1)/ LjCASTOR/ LjPOLLUX plays a critical role for upstream of 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) (Catoira *et al.*, 2000; Endre *et al.*, 2002; Madsen *et al.*, 2003). DMI1 and DMI2 act downstream of NFRs in the Nod factor signaling pathway (Radutoiu *et al.*, 2003). 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 pathway (NSP1 and NSP2), Ethylene response factor (ERF) required for nodulation (ERNs) and ERF required for nodule differentiation (EFD). These transcription factors enhance the expression of Nod factor-responsive genes by 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 how DMI3 regulates nodule-related Nodule inception (NIN) and ERN, how EFD inhibits nodulation probably by disrupting cytokinin (Ryu *et al.*, 2012).

Nodulation is nitrogen fixation zone, Within the symbiosomes, the bacteria undergo a morphological terminal differentiation involving cell elongation and genome amplification, a process which is governed by numerous plant nodule-specific cysteine-rich peptides (NCR), which are addressed to the bacterium-containing compartments, where bacteroids reduce atmospheric nitrogen to ammonium. Oxygen-binding leghemoglobin functions to maintain a microaerobic environment necessary for bacteriod nitrogenase activity. Ammonia(um) is the main product of N_2 fixation that is released from bacteroid and transported via ammonium transporter across the peribacteroid membrane (PBM) to the plant where initial assimilation into amino acids (AAs) occurs (White *et al.*, 2007 and 2009). Four key enzymes were 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, conjunction with aspartate aminotransferase (AAT) and asparagine (ASN) synthetase (AS). Depending on the N-exporter nature, the determinate-nodule primarily transports allantoin (ALN) or/and allantoic acid (ALC) as fixed-N compounds (ureide-exporter), and the indeterminate-nodule export fixed-N

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

predominantly as ASN and glutamine (GLN) (amide-exporter) (Sprent and James, 2007). 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), have been identified by analyses of Fix^- mutants. SST1 is a sulfate transporter localized in the peribacteroid membrance (PBM) and transfers SO_4^{2-} from plant cytosol to bacteroids. FEN1 is homocitrate synthetase which supplies homocitrate to bacteroids to support synthesis of the nitrogenase complex. IGN1 is localized in the plasma membrane (PM) and plays important functions in symbiosome and/or bacteroid differentiation and maintenance. (Krusell *et al.*, 2005; Kumagai *et al.*, 2007; Hakoyama *et al.*, 2009; Kouchi *et al.*, 2010). The host plants control the differentiation of *rhizobia* into bacteria targeting nodule-specific cysteine-rich (NCR) peptides which direct symbiotic *rhizobia* into terminal bacteroid differentiation through the nodule-specific protein secretion system in indeterminate nodules (Wang *et al.*, 2010).

These studies in molecular genetics using two model legume plants, *Lotus japonicas* and *Medicago truncatula*, have identified a number of host genes thatessential for symbiotic nodule formation and nitrogen fixation activation. Soybean (*Glycine max* (L.) Merr.) is one of the most important legume crops for seed protein and oil content. In the past, soybean analysis of early gene expression in *rhizobia* symbioses focused on nodulin genes or only a few genes. These genetic loci (namely *Rj*/*rj*s) have been identified as controlling nodulation traits upon inoculation with compatible species (Hayashi *et al.*, 2012). *Rj* and/or *rj* genostypes were summarized into three types, as follows: (1) Recessive alleles at three loci, *rj1*, *rj5* and *rj6*, result in non-nodulation phenotypes (Williams and Lynch, 1954; Pracht *et al.*, 1993). (2) Recessive locus, Known as *rj7* or *nts1* (*nitrate-tolerant symbiosis*) causes a so-called 'hypernodulation' phenotype, the formation of an unusually large number of nodules (Caroll *et al.*, 1985; Akao and Kouchi, 1992). (3) Dominant alleles, *Rj2*, *Rj3*, *Rj4* and *Rfg1* have unique features that restrict nodulation with specific strains (or serogroups) of *Bradyrhizobium* or *Ensifer*/ *Sinorhizobium* (Cadwell, 1966; Vest and Caldwell, 1972; Weier *et al.*, 1990; Trase, 1995). These *Rj* and/or *rj* genes were map-based cloning using establishment of the resources for genomics studies of these model legumes and the soybean whole genomics sequencing (Schmutz *et al.*, 2010). For instance, *rj1* (*nod49*) and *rj5* (*nod139*), related lipo-oligochitin LysM-type receptor kinase genes, are characterized using positional cloning and candidate gene

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

approaches genes (Indrasumunar *et al.*, 2010 and 2011). *Rj2* (*Rfg1*) are allelic genes encoding a member of the Toll-interleukin receptor/ nucleotide-binding site/leucine-rich repeat (TIR-NBS-LRR) class of plant resistance (R) proteins using map-based cloning (Yang *et al.*, 2010).

As functional genomics and sequence develop, studies of transcriptional profiles during nitrogen fixation are important to gain greater understanding of the related nitrogen fixation genes. Some studies using 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 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 gene expression of R1 stage (first flower) in isolated soybean root inoculated or mock-inoculated with *rhizobium*. This study aimed at analyzing 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* 45436 were grown at 28 □ and darkroom in liquid yeast extract mannitol broth (YMB) medium (pH 6.8) with moderate shaking (120 rpm). After 6 d, *B. japonicum* cells were amassed by centrifugation (4000 rpm, 10 min), washed with sterile water 3 times, and diluted in water to an optical density at OD_{600} = 0.8.

Soybean 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 at the V1 stage of development, by respectively inoculated inoculums containing USDA110, CCBAU45436 (50 ml of liquid suspension per pot at 0.8 OD_{600}). Mock-inoculated plants (CK) received the same amount of autoclaved inoculums. Roots were harvested

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

at R1 inoculated with USDA 110, CCBAU45436 and CK. Subsequently, the roots were extirpated nodulation and separated from shoots, immediately frozen in liquid nitrogen and stored at -80 □.

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 Issolation Module. The first and second strand cDNA were synthesized using ProtoScript □ Reverst 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 5' phosphorylation and dA-tailing in one reaction and was ligated to adaptors with a "T" base overhang.

Sequencing and sequence aligment

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 flow cell to perform bridge PCR and P7 primer carrying a six-base index allowing for multiplexing. The PCR productions were further cleaned 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 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 reads from sequencing were aligned against the Williams 82 genome sequence (ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Athaliana/) using tophat-

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.09 software, allowing a maximum of two mismatches.

Bioinformatics analysis

A level of gene expression by expected number of fragments per kilobase of transcript sequence per million base pairs sequenced (FPKM value) was calculated using Rsem (v1.2.4) software. We calculated the differences in gene expression among different samples with EdgeR (v3.4.2) and *p*-value of significant difference. *P* value less than 0.05 was thought as significant difference.

GO enrichment analysis of differentially expressed genes using default GO association files was performed with "GO Term Finder" (http://go.princeton.edu/cgibin/GOTermFinder) where statistical significance (*p*-value) was calculated based on hypergeometric distribution with Bonferroni multiple testing correction and flase discovery rate (FDR) calculation as described. The GO enrichment analysis was performed with adjusted GO association files, Ontologizer, which (http://www.charite.de/ch/medgen/ ontologizer/) was using "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 with a threshold *P* value of FDR less than 0.05.

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 gene 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 primer (Invitrogen), according to the manufacturer's instruction. Quantitative PCR was performed on an ABI ViiA™ 7 (Applied Biosystems) with the LightCycler system. PCR mixtures (final volume 20 μl) including 0.4 μl (approximately 50 ng) of first strand cDNAs, 0.5 μM each primer, 10μl of 2 × SYBR ® Green Realtime PCR Master Mix (TOYOBO). The cycling conditions were as follows: 2 min denaturation at 95 □ followed by 40 cycles of 95 ℃ for 10 s, 60 ℃ for 15 s, and 72 ℃ for 25 s. Thirteen candidate gene's primers (Tab. 1) were designed according to every candidate gene sequence (http://soybase.org/gbrowse/cgi-bin/gbrowse/gmax1.01). Expression levels of these genes

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were normalized by tubulin (NCBI accession No. AY907703). Gene expression was quantified using the relative quantification ($\Delta\Delta C_T$) method and data was compared with internal controls. Each sample was replicated three times.

RESULTS

RNA-Sequencing analysis

Changes of plant roots' transcript levels were analyzed between the inoculated USDA110 and CK at R1 stage by RNA-Seq. A total of 147.5 million reads were generated by 101 bp single-end sequencing from the two cDNA libraries, constituting 14.9 Gb of cDNA sequence (raw data). The raw data were filtrated using NGSQCToolkit (V2.3) according to the following processes: (1) The joint sequences were removed; (2) These reads were retained that these loci of the quality value higher than Q30 (Percentage of error probability less than 0.1%) accounted for 70% or more of total length of reads; (3) A base of 3'-end was removed; (4) These loci of the quality value lower Q30 were removed; (5) The reads of length lowing 70 bp were removed. Approximately 85.41% of the sequenced reads were high quality filtered reads (Tab. 2, Fig. S1). 92.46% of the clean data were successfully aligned to the soybean genome reference sequence (Glyma 1.01, http://www.phytozome/v9.0/athaliana) using Tophat-2.09 software. Thereinto, the reads of uniquely aligned to the reference genome (unique_mapped) accounted for 36.67% of total clean data (Tab. 3).

The differentially expressed genes profiles responding to *rhizobium* inoculation

A total of 55787 transcripts were quantified based on FPKM value of unique_mapped fragments for the analysis of gene expression (Tab. S1). There are 596 differentially expressed transcripts (P value ≤ 0.05 and fold-change ≥ 2.0) by edgeR program of Bioconductor software through FPKM values (Tab. S2). Of these, 280 transcripts were up-regulated and 316 were down-regulated in inoculated USDA110 VS CK (Fig. 1). The most $\log_2^{CPM}$ value of transcripts expression level were less than 5.0, the correlation relation showed as figure 2 between transcripts expression level and expression difference.

Gene ontology category analysis

As a useful tool for gene functional annotation, WEGO (Web Gene Ontology

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Annotation Plot) has been widely used in many soybean studies (Li *et al.*, 2013). It has become one of the most 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 three functional categories; cellular component, molecular function and biological process by mapping to 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 mapped to cell component category; 182 genes mapped to molecular function category; and 178 genes mapped to biological process category. Five subsets of 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, binding, transporter, and binding. Eleven subsets of biological process category are metabolic process, biological regulation, cellular component organization or biogenesis, celluar process, reproduction, multicellular organismal process, signaling, developmental process, single-organism process, response to stimulus, and localization (Fig. 4 and supplementary dataset S1). The most abundant unigenes were related to cell or cell parts in cellular component category, catalytic activity and binding in molecular function category, and metabolic process, response to stimulus and localization in biological process category. The results showed that symbiotic fixation nitrogen brought on a series of plant's genes 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

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

pathways related to carbon cycle and metabolism, 16 pathways related to amino acid synthesis and metabolism, 8 pathways related to hormone, 8 pathways related to signaling pathway, 6 pathways related to vitamin, 3 pathways related to alkaloid, and related to RNA and virus etc. Thereinto, a nitrogen metabolism and a NOD-like receptor (NLR) signaling pathway were identified. These websites of pathway's map were also listed in table S3. A total of 98 difference expressed genes (115 transcripts) with pathway annotations were identified by KAAS software. The expressed level of 54 transcripts was up-regulated, and 61 transcripts were down-regulated. The number of transcript associated with pathway ranges from 1 to 44, with an average number of 4.38 pathways per transcript. 23 transcripts were associated with one pathway, 92 transcripts were associated with two or more pathway. The transcripts of Glyma10g04410.2 was involved in 44 pathways, secondly, Glyma14g00760.2 was involved in 17 pathways (Tab. S4).

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. Seven differentially expressed transcripts associated with the nitrogen metabolism pathway. The all 7 transcripts were down-regulated, which may be because symbiosis provided nitrogen nutrition for plant, some nitrogen transformation genes needn't over-expressed.

Validation of RNA-Seq data by qRT-RCR and result's university in other *rhizobium*

To confirm the expression patterns validity by RNA-Seq in the soybeans of mock-inoculated, inoculated *B. rhizobium* USDA110 and result's university in other *rhizobium*, we used qRT-PCR analyses to analyze expressions of 19 candidate genes of CK, USDA110 and CCBAU45436. Results showed significant differences in the expression levels between the inoculated and mock-inoculated *rhizobium* for 19 transcripts (Fig. 7). Ten transcripts (Glyma01g44400.3, Glyma02g40820.2, Glyma08g19250.1, Glyma11g36000.2, Glyma13g39450.1, Glyma14g40320.2, Glyma15g16470.2, Glyma17g33050.1, Glyma17g34110.2 and Glyma19g25360.2) were significantly up-regulated in soybean roots of inoculated *rhizobium*, other nine transcripts were significantly down-regulated. Although the RNA-Seq values showed slight variations compared with the corresponding values from the qRT-PCR analyses, the tendency of expression level from RNA-Seq analysis was consistent with those

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

obtained by qRT-PCR. The expression levels of 19 transcripts in soybean roots inoculated USDA110 strain were similar to inoculated CCBAU45436 strain. These results highlighted the fidelity and reproducibility of the RNA-Seq analysis used in the present study. The results had a certain extent's university in different *rhizobium* strains.

DISCUSSION

The role of biological nitrogen fixation (BNF) in world crop production has been emphasized for several decades. The molecular mechanisms involved in the interaction between legumes and *rhizobia* has greatly improved (Kouchi *et al.*, 2010), specifically expressed during the symbiotic nodule formation (Libault *et al.*, 2010; Hayashi *et al.*, 2012). The development of nitrogen-fixing nodules on roots of leguminous plants commences with a molecular dialogue between the host plant and a compatible strain of *rhizobia*. The identification of plants that spontaneously form nodules, together with the observations that ectopic application of cytokinins or auxin transport inhibitors to the root surface lead to the development of nodule-like structures, demonstrates that the machinery required for nodule primordium (NP) development is intrinsic to the plant (Gleason *et al.*, 2006; Murray *et al.*, 2007; Tirichine *et al.*, 2007; Van *et al.*, 2015). More recently, comprehensive analysis by means of proteomics and transcriptome have revealed that more than a thousand genes are specifically induced or highly enhanced in nodules (Wienkoop and Saalbach, 2003; Høgslund *et al.*, 2009; Carvalho *et al.*, 2013). However, few progresses have been achieved towards host plant to control bacteroid differentiation and activating *rhizobial* nitrogen fixation.

This study has provided a new data set identifying the expression of DEGs during symbiotic nitrogen fixation in R1 stage. Massive parallel sequencing identified 596 differentially expressed transcripts from two contrasting libraries. We examined gene expression levels in detail, and found significant difference between inoculated and mock-inoculated *rhizobia*. Among the differentially expressed genes, more were down-regulated than up-regulated, and some showed a differential expression pattern in all three contrast groups, indicating that there were overlaps at the transcriptional level. The fact that there were many down-regulated transcripts indicates that there are more negatively regulated genes than positively regulated ones with functions in the nitrogen fixation pathway. To explore the genes with unknown functions, the expression patterns of different genes were analyzed by according to similarities in

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expression profiles.

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, NOD-like receptor signaling pathway, the nitrogen metabolism and plant hormone play an important role. The Hsp90 of host plant was up-regulated by 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) activate Nod1 (Correia *et al.*, 2007). The nodulation ability of host plants could be improved by boosting up expression's level of Hsp90.

In nitrogen metabolism, the nitrate reductase (NAD(P)H), ferredoxin-nitrite reductase and carbonic anhydrase of host plant were down-regulated by inoculated *rhizobium*. The enzyme nitrate reductase, ferredoxin-nitrite reductase and carbonic anhydrases respectively catalyzes the reduction of nitrate to nitrite (Okamoto *et al.*, 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 mostly ammonium demand of host plant could be basically satisified by the nitrogen fixation ability of *rhizobium*, fewer nitrate nitrogen was absorbed from soil, so three reductases' activity decreased.

The plant hormone regulates a number of developmental and physiological processes including nodulation. For example, a cytokinin 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* (Tab. S3). The result is consistent with previous study.

The validation results of qPCR were consistent with the transcripts expression patterns identified by RNA-Seq and expression levels of transcripts in soybean roots inoculated USDA110 strain were similar to inoculated CCBAU45436 strain. These

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

showed that the results of RNA-Seq have a certain extent of universality in 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; Searle *et al.*, 2003; Tanya *et al.*, 2005; Nicolás *et al.*, 2006; Yang *et al.*, 2010; Hayashi *et al.*, 2012; Santos *et al.*, 2013) (Tab. S5). There are significantly different genes of using RNA-Seq in these QTL regions. For example, three significant different transcripts (Glyma11g08720.2, Glyma11g09250.2 and Glyma11g09630.2) between inoculated and mock-inoculated *rhizobium* were located between Satt509 and Satt251 on chromosome 11 in study of *Santos* (2013). However, the *Rj*/*rj* genes involved in nitrogen-fixation root nodule formation in soybean weren't detected in the study. These results may be because that the difference of nodulation gene wasn't significant difference between inoculated and mock-inoculated *rhizobia* in soybean's R1 stage. More novel genes were identified by RNA-Seq in the study and weren't located by previous gene/QTLs mapping study. Their function needs to be further verified using other methods.

In the study, 280 and 316 transcripts were up- and down-regulated in 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.

ACKNOWLEDGMENTS

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

LITERATURE CITED

Akao S. and H. Kouchi (1992) A supernodulating mutant isolated from soybean cultivar Enrei. Soil Sci. Plant Nut. 38:183-187.

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Arrighi J. F., A. Barre, A. Ben, A. B. Ben, A. Bersoult, L. C. Soriano, R. Mirabella, F. Carvalho-Niebel, E. P. Journet, M. Ghérardi, T. Huguet, R. Geurts, J. Dénarié, P. Rougé and C. Gough (2006) The *Medicago 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, M. P. Lee, M. Libault, L. O. 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.

Caldwell B. E. (1966) Inheritance of a strain-specific ineffective nodulation in soybeans [J]. Crop Sci. 6:427-428.

Carroll B. J., D. L. Mcneil and P. M. Gresshoff (1985) A supernodulation and nitrate-tolerant symbiotic (*nts*) soybean mutant. Plant Physiol. 78:34-40.

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, R. V. Penmetsa, E. P. 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.

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Endre G., A. Kereszt, Z. Kevei, S. Mihacea, P. Kalo and G. B. 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.

Hakoyama T., K. Niimi, H. Watanabe, R. Tabata, J. Matsubara, S. Sato, Y. Nakamura, S. Tabata, J. C. Li, T. Matsumoto, K. Tatsumi, M. Nomura, S. Tajima, M. Ishizaka, K. Yano, H. Imaizumi-Anraku, M. Kawaguchi and H. Kouchi (2009) Host plant genome overcomes the lack of a bacterial gene for symbiotic nitrogen fixation. Nature 462:514-517.

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. Breeding 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.

Høgslund N., S. Radutoiu, L. Krusell, V. Vorshilova, M. A. Hannah, N. Goffard, D. H. Sanchez, F. Lippold, T. Ott, S. Sato, S. Tabata, P. Liboriussen, G. V. Lohmann, L. Schauser, G. F. Weiller, M. K. Udvardi and J. Stougaard (2009) Dissection of symbiosis and organ development by integrated transcriptome analysis of Lotus japonicus mutant and wild-type plants. PLos one DOI:10.1371/journal.pone.0006556.

Indrasumunar A., A. Kereszt, I. Searle, M. Miyagi, D. X. Li, C. D. T. Nguyen, A. Men, B. J. Carroll and P. M. Gresshoff (2010) Inactivation of duplicated Nod factor receptor 5 (NFR5) genes in recessive loss-of-function non-nodulation mutants of allotetraploid soybean (*Glycine max* L. Merr.) Plant Cell Physiol. 51(2):201-214.

Indrasumunar A., I. Searle, M. H. Lin, A. Kereszt, A. Men, B. J. Carroll and P. M. Gresshoff (2011) Nodulation factor receptor kinase 1α controls nodule organ number in soybean (*Glycine max* L. Merr). *Plant J.* 65:39-50.

Irving H. R., N. M. Boukli, M. N. Kelly and W. J. Broughton (2000) Nod-factors in

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 Microbe. In. 8(1):172-176.

Kouchi H., 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.

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

Kumagai H., T. Hakoyama, Y. Umehara, S. Sato, T. Kaneko, S. Tabata and H. Kouchi (2007) A novel ankyrin-repeat membrane protein, IGN1, is required for persistence of nitrogen-fixing symbiosis in root nodules of Lotus japonicas. Plant Physiol. 143:1293-1305.

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)

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 Neurospora crassa: 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. Annu. 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.

Pracht J. E., C. D. Nickell and J. E. Harper (1993) Genes controlling nodulation in soybean: *Rj5* and *Rj6*. Crop Sci. 33:711-713.

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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, H. C. 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.

Schmutz J., S. B. Cannon, J. Schlueter, J. X. Ma, T. Mitro, W. Nelson, D. Hyten, Q. J. Song, J. J. Thelen, J. Cheng, D. Xu, U. Hellsten, G. D. May, Y. Yu, T. Sakurai, T. Umezawa, M. K. Bhattacharyya, D. Sandhu, B. Valliyodan, E. Lindgusit, M. Peto, D. Grant, S. Shu, D. Goodstein, et al. (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178-183.

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.

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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, L. H. Madsen, H. Miwa, T. Nakagawa, N. Sandai, A. S. 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.

Trase A. T. (1995) A single dominant gene in McCall soybean prevents effective nodulation with Rhizobium fredii USDA257. Euphytica 81:279-282.

Van Z. A., O. R. H. Camp, E. E. Deinum, T. Charnikhova, H. Franssen, O. H. J. Camp, H. Bouwmeester, W. Kohlen, T. Bisseling and R. Geurts (2015) Rhizobium lipo-chitooligosaccharide signaling triggers accumulation of cytokinins in *Medicago truncatula* roots. Mol. Plant doi:10.1016/j.molp.2015.03.010.

Vest G. and B. E. Caldwell (1972) *Rj*4-A gene conditioning ineffective nodulation in soybean. *Crop Sci*. 12:692-693.

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.

Weier G. V., H. D. Skipper and A. G. Wollum (1990) Exclusion of inefficient *Bradyrhizobium japonicum* serogroups by soybean. Plant Soil 121:99-105.

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 P. S. Poole (2009) Characterization of a γ-aminobutyric acid transport system of *rhizobium leguminosarum* bv. viciae 3841. *J. Bacteriol. 191(5):1547-1555.*

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Wienkoop S. and G. Saalbach (2003) Proteome analysis. Novel proteins identified at the peribacteroid membrane from *Lotus japonicus* root nodules. Plant Physiol. 131:1080-1090.

Williams L. F. and D. L. Lynch (1954) Inheritance of a non-nodulation character in the soybean. Agron. J. 46:28-29.

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.

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Legends to Tables and Figures

Table 1 Primer sequences used for qRT-PCR

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

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

Total_mapped is the total number of reads was aligned to reference genome. Unique_mapped is the number of reads which were aligned to exclusive position of reference genome. Muti_mapped is the number of reads which were aligned to multi-position of reference genome. PE_mapped is the number of reads of PE sequencing double end which were synchronously aligned to reference genome.

Figure 1 Number of differentially expressed transcripts in CK VS USDA110

Figure 2 Relation between transcripts expression difference and counts

$\log^{FC}$ is value of $\log_2^{\text{fold change of transcript expression}}$. $\log^{CPM}$ is average value $\log_2^{\text{counts per million reads of transcript expression}}$. Red dot represent up-regulated transcripts. Blue dot represent down-regulated transcripts. Black dot represent no significantly different transcripts.

Figure 3 Functional categorization of expressed genes in plant roots of inoculated CK and USDA110

Figure 4 Cellular Component, Molecular Function and Biological Process of differentially expressed genes between CK and USDA110

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 5 NOD-like receptor signaling pathway

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

Figure 6 Nitrogen metabolism pathway

Figure 7 Analysis of 19 differentially expressed genes using RNA-Seq by qRT-PCR in CK, USDA110 and CCBAU45436

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Legends to Supplementary Tables, Figures and Supplement dataset S1

Table S1 Expressed level of transcripts aligned to soybean genomic sequence

GeneLength is length of reference geness. Sample_reads is number of gene's exclusive alignment fragment. Sample_FPKM is gene's expected number of fragments per kilobase of transcript sequence per million base pairs sequenced. The same below

Table S2 Differentially expressed transcripts of roots between CK and USDA110

$\log^{FC}$ is value of $\log_2^{\text{fold change of transcript expression}}$. $\log^{CPM}$ is average value $\log_2^{\text{counts per million reads of transcript expression}}$. PValue is P value of significantly different gene. FDR is false discovery rate.

Table S3 The KEGG pathway of differentially expressed genes

CK-VS-USDA110 is differentially expressed gene's number of two samples in the pathway. All-Diffgene is all genes' number of gene's background in the pathway. Genes are differentially expressed genes in the pathway. KOs are terms of KEGG Orthology. PathwayURL is websites of pathway's map.

Table S4 Number of pathway of differentially expressed genes

Table S5 Previously genes/QTLs regions exist in differentially expressed genes

Note: NDW is nodule dry weight. SDW is shoot dry weight. NN is nodule number. NFW is nodule fresh weight, PDW is plant dry weight. NDW/NN is ratio nodule dry weight/NN.

Figure S1 Summary of quality check and filtering

Supplement dataset S1 GO analysis of differentially expressed genes

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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.

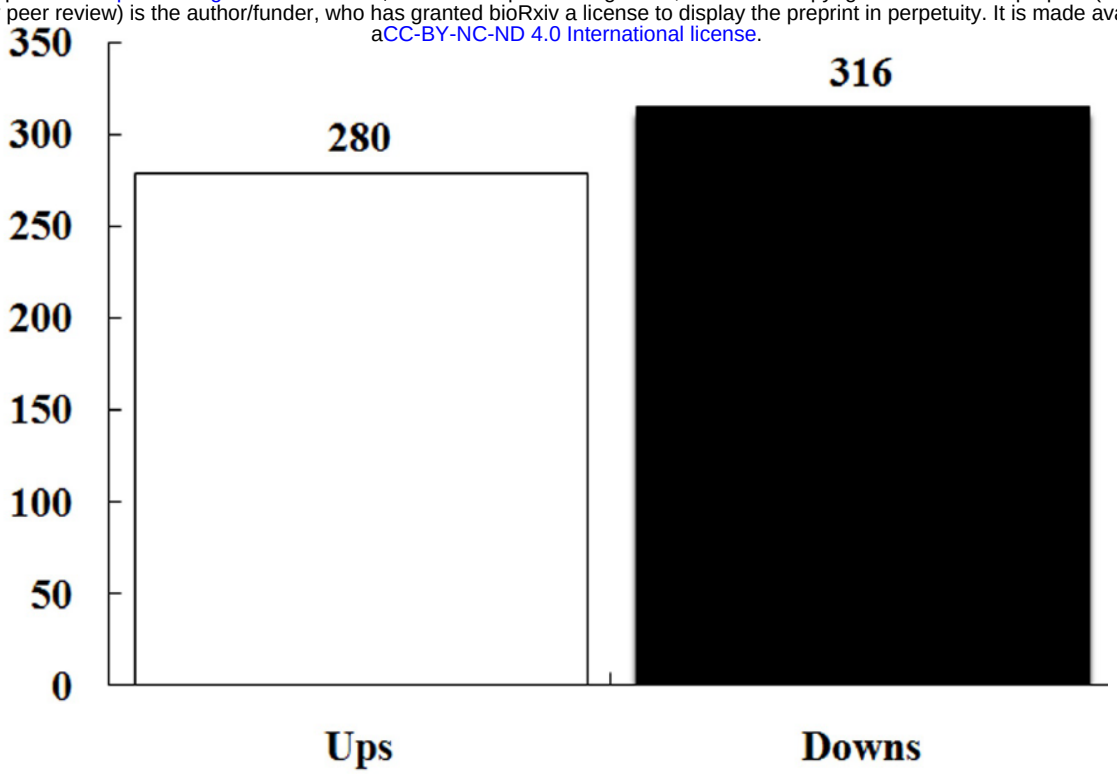
bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 3 Results of clean reads were aligned to reference genome

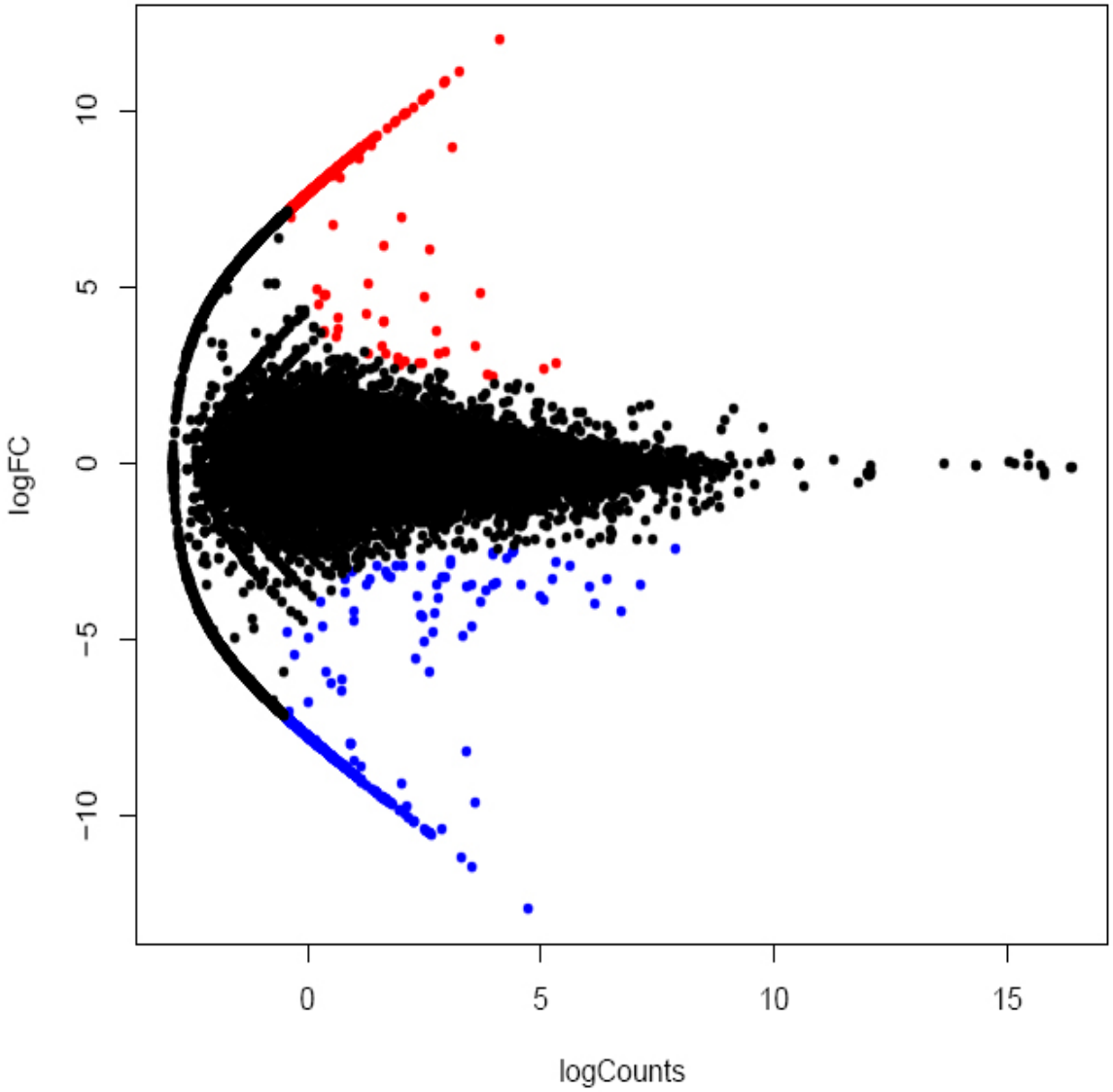
Samples	Total_mapped	Unique_mapped	Multi_mapped	PE_mapped	Mapped percent(%)
CK	55793166	20436743	35356423	27327800	91.94
USDA110	61369446	22522473	38846973	30114865	92.99

Total_mapped is the total number of reads was aligned to reference genome. Unique_mapped is the number of reads which were aligned to exclusive position of reference genome. Muti_mapped is the number of reads which were aligned to multi-position of reference genome. PE_mapped is the number of reads of PE sequencing double end which were synchronously aligned to reference genome.

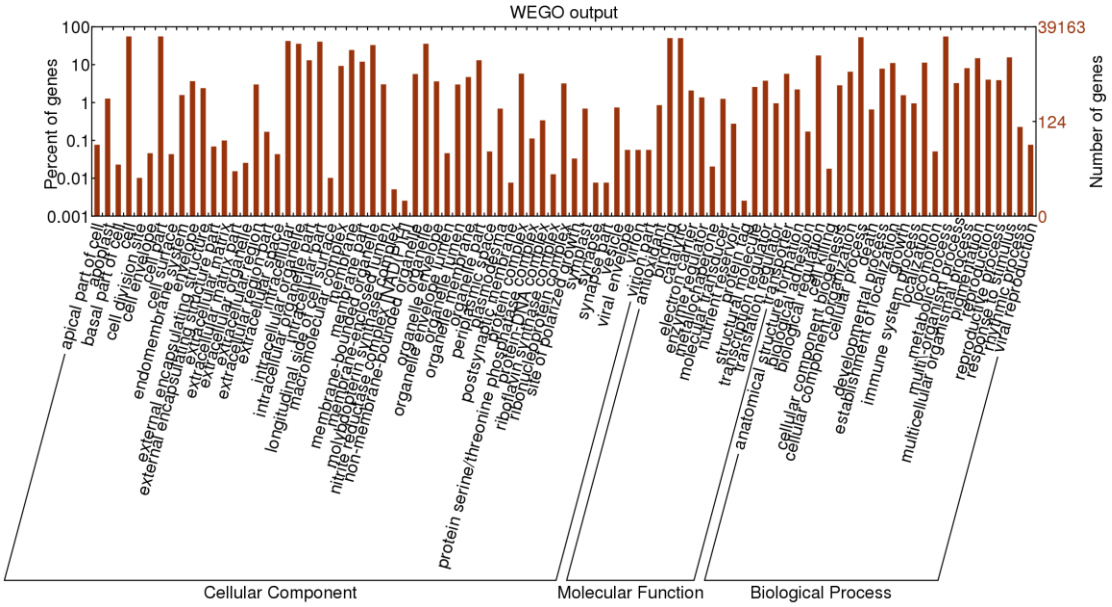
bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



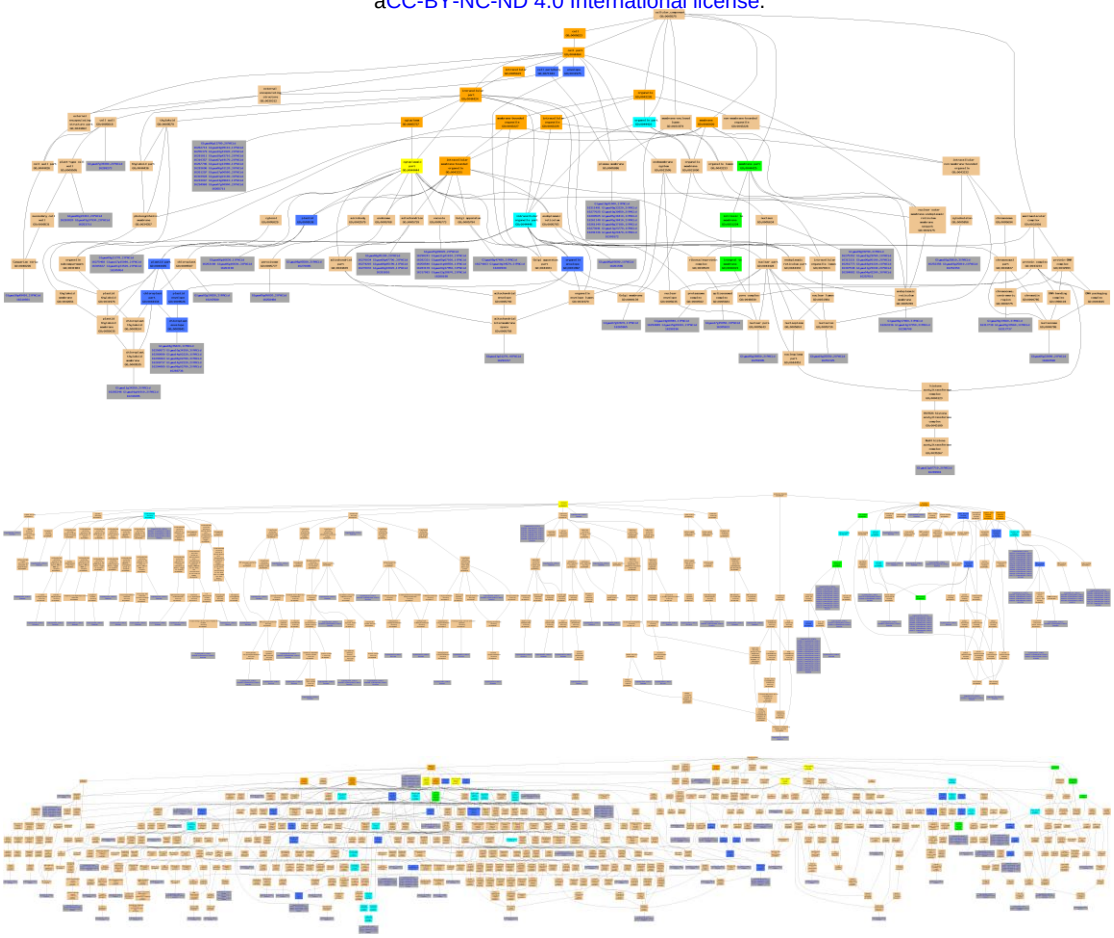
bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



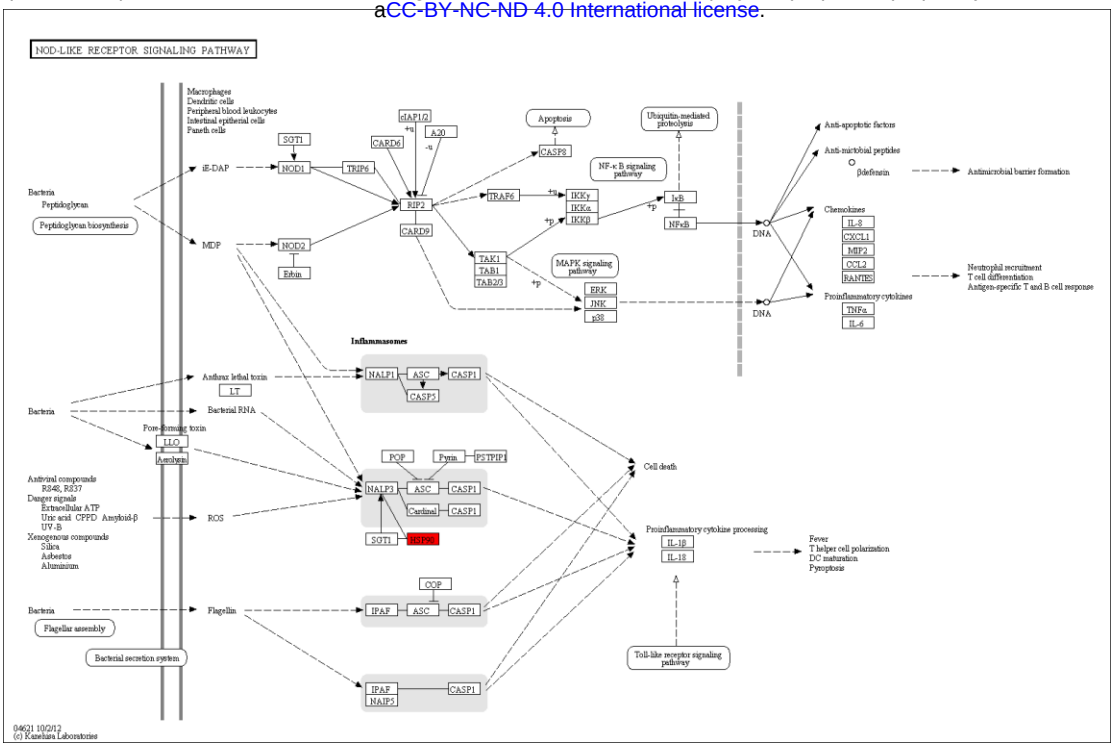
bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



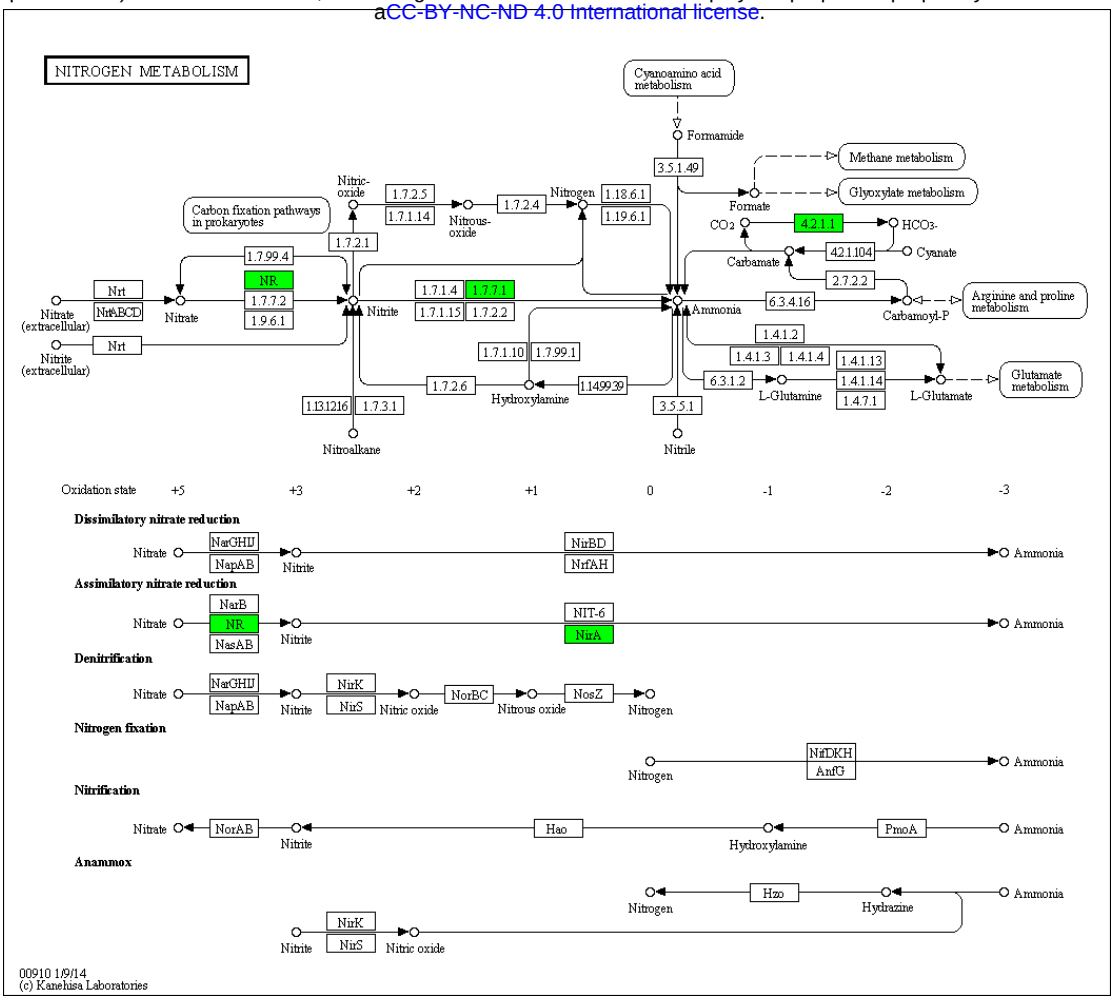
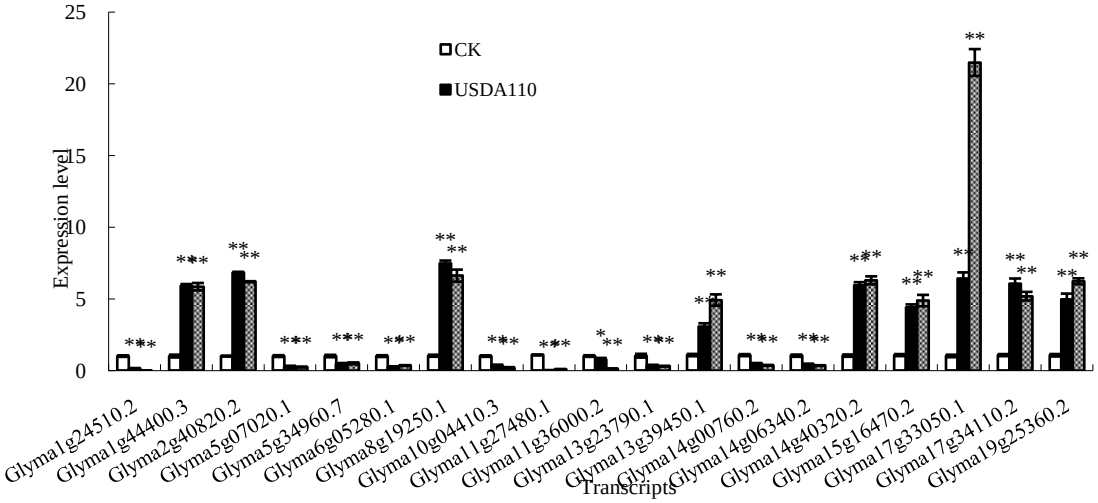
bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/024224; this version posted August 19, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.