Comparative profiling of the transcriptional response to soybean cyst nematode infection of soybean roots by deep sequencing

LI XiaoYan[1,2], WANG Xue[3], ZHANG ShaoPeng[1,2], LIU DaWei[3], DUAN YuXi[3*] & DONG Wei[4*]

[1] *Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100029, China;*

[2] *Graduate University of Chinese Academy of Sciences, Beijing 100049, China;*

[3] *Nematology Institute of Northern China, Shenyang Agricultural University, Shenyang 110866, China;*

[4] *Beijing Genomics Institute, Hangzhou 310008, China*

To gain insight into the changes in the transcriptome of soybean roots during soybean cyst nematode (SCN) infection, we conducted genome-wide gene expression profiling using serial analysis of gene expression (SAGE) combined with Solexa sequencing. More than 3 million tags were generated from the SCN-infected and uninfected roots, and 366941 and 314591 clean UniTags were obtained from SCN-infected and uninfected samples, respectively. In the SCN-infected sample, 48249 UniTags represented 18114 reference genes. In the uninfected control, 46290 UniTags represented 19323 reference genes. Comparison of tag frequencies identified 1405 genes that were expressed at greater levels in SCN-infected roots than in uninfected roots, and 1191 genes that were expressed at lower levels. Quantitative real-time PCR analyses confirmed the changes in mRNA levels observed in our sequencing analyses. A comparable number of genes were up- and down-regulated in response to nematode infection, indicating that down-regulation of some genes might be essential in the plant response to nematodes. Our SAGE results showed significant changes in expression of many unreported genes involved in nematode infection. Approximately 7% of tags mapped to the antisense strand of genes, indicating widespread antisense transcription.

***Glycine max*, soybean cyst nematode, gene expression profile, SAGE, Solexa sequencing**

Soybean (*Glycine max* (L.) Merrill) is one of the most widely planted dicot crops, and accounts for approximately 56% of global oil seed production. Soybeans are used as food, feed, and in industrial products. In addition, soybeans are a primary source of high-value secondary co-products such as lecithin, vitamins, nutraceuticals, and anti-oxidants [1]. The soybean cyst nematode (SCN), *Heterodera glycines*, is one of the most widespread and destructive soybean pathogens. In the top eight soybean producing countries, the soybean yield loss due to SCN was 263.7 million bushels in 2006. Measures to control the pest include the use of nematicides or planting resistant cultivars. However, the availability of chemical pesticides is decreasing and host resistance is limited. Thus, it is important to understand in detail the nature of SCN-disease progression [2].

Establishment and development of infection by SCN is a complicated process, which spans from the early migratory stages during penetration and migration through roots to the later sedentary stages of syncytium induction and feeding, which ultimately facilitates nematode development and reproduction. The magnitude of the changes in soybean during SCN-infection suggests that many genes are involved. Several genes related to SCN-infection have been identified in soybean, including those encoding polygalacturonases [3],

*Corresponding authors (email: duanyx6407@163.com; dongw@genomics.org.cn)

an ethylene-responsive element-binding protein [4], a phosphoribosylformyl-glycinamidine synthase [5], an extension [6], a cyclin [7], a GTP-binding protein [8], etc. In recent years, several studies profiling global soybean responses to SCN infection using differential display and microarray approaches have increased our understanding of changes in gene expression during SCN-infection [9–12]. However, the complexity of the soybean–SCN pathosystem suggests that there are many more genes expressed during the soybean response to nematode infection.

Serial analysis of gene expression (SAGE) is a high-throughput gene expression profiling technique that has been used to study global expression changes in a variety of organisms. In principle, SAGE and all its variants rely on the assumption that a small, defined part of a cDNA, a so-called "tag", characterizes that cDNA, and that the abundance of a particular tag in the tag population reflects the abundance its respective mRNA in the transcriptome. SAGE technology does not require prior knowledge of the transcriptome and is useful for discovery and annotation of novel transcripts. SAGE has been used successfully to investigate changes in gene expression during various host-pathogen interactions, for example, to monitor gene expression in rice leaves during infection by blast fungus, and to examine differences in gene expression between resistant and susceptible genotypes of cassava during viral infection with cassava mosaic disease. However, the Sanger sequencing method is expensive, and limits the application of this technology [13–15].

Next-generation sequencing (NGS) technologies including the Roche 454 Genome Sequencer, the Illumina Genome Analyzer, and Applied Biosystems' SOLiD have revolutionized SAGE and cDNA sequencing technology [16]. The combination of LongSAGE and Solexa sequencing, known as 3′ tag Digital Gene Expression (DGE), appears to be perfectly suited for deep transcriptome analysis. This technology generates up to 10–12 million short (36-bp) cDNAs from individual libraries. The increased throughput makes it possible to detect RNA molecules at very low-copy numbers in the cell. On the other hand, DGE also increases the tag count for each particular mRNA, enabling more precise expression profiling of transcripts expressed at low levels, and decreasing the error rate of the number of mappable tags [17,18].

The main aim of this study was to investigate the changes in gene expression patterns in response to SCN infection in soybean roots. We used the DGE approach to analyze gene expression profile in two root samples of the soybean cultivar 'Harbin xiaoheidou'; one collected at 30 d after infection by SCN race 3, and the other a control (uninfected). Quantitative real-time PCR (qRT-PCR) was performed for selected soybean mRNAs to evaluate the sensitivity, dynamic range, and accuracy of 3′ DGE sequencing. This study is the first analysis of the soybean-SCN interaction using open-architecture and provides data on the global profile of gene expression in soybean roots.

1 Material and methods

1.1 Plant growth and RNA preparation

The soybean (*Glycine max*) cultivar "Harbin xiaoheidou" was used to prepare control root (SHBC) and cyst-inoculated root samples (SHB). The plants were grown in greenhouse. Cyst inoculation samples were planted in soil infected with SCN race 3, which was collected from the pilot field of Shenyang Agriculture University. Control samples were grown in autoclaved field soil. Root samples were collected 30 d after emergence of seedlings and were ground to a powder in liquid nitrogen. Total RNA was extracted from powder with TRIzol reagent (Invitrogen, Carlsbad, CA, USA).

1.2 Library construction and sequencing

The 3′ tag DGE libraries were constructed from SHBC and SHB RNA following the Illumina DGE protocol. Briefly, mRNA was purified from total RNA by binding to magnetic oligo(dT) beads and then first and second strand cDNAs were synthesized. The bead-bound cDNA was digested by the restriction enzyme *Nla*III, and then the Illumina adapter 1, which includes a *Mme*I recognition site, was added. Digestion with *Mme*I yielded the adapter tag linked to 20 bp of cDNA and Illumina adapter 2 was then added to the 3' end of the tag. The acquired tags were amplified using PCR and then separated by TBE PAGE on 6% acrylamide gels. Cluster generation and sequencing were performed at the Beijing Genomics Institute using the Illumina Genome Analyzer II. Approximately 3 million 36-nt tags were generated for each sample. Verified data will be deposited into the public database at NCBI.

1.3 Tag annotation

Sequence data analysis was performed with the DGE analysis pipeline provided by BGI (Shenzhen, China). First, sequencing-received image data were transformed into raw sequence data by base calling. Second, low quality reads (tags with unknown nucleotide 'N'), empty tags (sequence with only the adaptor sequence), and low complexity tags were trimmed with our own perl script. Adaptor sequences were accurately clipped with the aid of a dynamic programming algorithm. A preprocessed database of all possible CATG+17 nucleotide tag sequences was created using the soybean UniGene set at NCBI. For annotation, all tags were mapped to the reference sequences and were allowed no more than one nucleotide mismatch. Tags mapped to multiple genes were excluded and tags that mapped to a unique reference gene were defined as unambiguous clean tags. When there were two or more unambiguous tags

mapped to the same gene, the sum of all tags represented the gene expression level. The remaining tags were aligned to the soybean genome assembly Glyma 1.0 and chloroplast sequences to identify candidates for new and non-annotated soybean genes.

1.4　Comparison of SHB and SHBC 3′ tag DGE profiles

The number of unambiguous tags for each gene was normalized to the number of transcripts per million clean tags (TPM) to compensate for variable numbers of tags generated for each sample [19]. Significant values for differences in expression levels were determined using a modified exact test, similar to Fisher's exact test. The FDR (false discovery rate) was used to determine the threshold P value in multiple comparisons and analyses [20].

　　All differentially expressed genes were mapped using the GO database, and compared with the genome background. GO terms more common in these genes than expected by chance (adjusted Fisher's $P<0.05$) were identified [21]. For pathway enrichment analysis, we mapped all differentially expressed genes to terms in the KEGG database and searched for KEGG terms that were significantly enriched compared with the genome background.

1.5　Real-time quantitative PCR

RNAs for quantitative PCR were handled in the same way as those extracted for construction of DGE libraries. Total RNA (1 μg) was treated by DNase (New England Biolabs, USA) and then converted to cDNA using superscript II reverse transcriptase (Invitrogen). For real-time PCR, sequences of selected genes were obtained from the NCBI UniGene database. Primers were designed using Primer v5. Real-time PCR for 20 randomly selected genes was performed using a StepOne RT-PCR machine (Applied Biosystems, Foster City, CA, USA) using SYBR green detection. Relative quantification of amplified genes was determined by the comparative $\Delta\Delta C_t$ method. GAPDH was used as the internal control [22]. All runs included a negative control without cDNA template, and three replicates were used for each reaction.

2　Results and discussion

2.1　Characterization of sequenced DGE libraries

Two 3′ tag DGE libraries were constructed; one from the SCN-infected root and the other from control (uninfected) soybean roots. A total of 6.3 million 36-bp reads were generated from one lane of Illumina sequencing; 3278391 from the SCN-infected sample and 3031188 from the uninfected control. These represented 1565157 unique transcripts, designated as UniTags. Solexa sequencing technology greatly improved the output, which was more than 100-fold greater than that obtained using the primary SAGE strategy [23,24].

　　Low quality tags were filtered using our perl script. From the uninfected root (SHBC) library, we obtained 366 941 UniTags that appeared more than twice and 498 361 singleton UniTags. From the SCN-infected root (SHB) library, we obtained 314591 multiple-frequency tags and 316042 singleton tags. UniTags that appeared more than once were defined as clean tags. These clean tags were used for further analyses because singleton tags can contain sequence errors and artificial sequences. Singletons accounted for 58.54% and 44.28% of tags in the infected and uninfected sample libraries, respectively. These results are consistent with those obtained from maize [25]. Tags are summarized according to frequency in Table 1. The significant sequence output greatly improved the tags sequencing depth. Tags with a low copy number in the uninfected root library increased significantly in the SCN-infected root library. Most of the clean UniTags (530852, or 77.8%) were present at a copy number of less than five. Only a small percentage of clean UniTags (5475, or 0.8%) were present at copy numbers greater than 100.

　　To identify genes corresponding to the clean UniTags

Table 1　Summary of tags obtained from SCN-infected and uninfected soybean root libraries

Library	SCN-infected	Uninfected control	Total
Total tags	3278391	3031188	6309579
Number of unique tags in raw data (UniTags)[a]	851365	713792	1565157
Number of unique tags in clean data (UniTags)[b]	314591	366941	681532
Abundance classes			
Copy number 1	498361	316042	814403
Copy number 2–5	245987	284865	530852
Copy number 5–10	35781	48602	84383
Copy number 10–20	16242	19011	35253
Copy number 20–50	9891	9388	19279
Copy number 50–100	3554	2736	6290
Copy number >100	3136	2339	5475

a) Singleton tags are included; b) only tags with high quality and copy number >1 are included.

detected in this study, an essential dataset containing 33001 reference soybean genes from the NCBI UniGene Build 38# was prepared. In total, 29425 genes (89.16%) contained CATG sites, allowing construction of a virtual tag database containing 115954 tags. All possible SAGE tags from each cDNA were recorded, progressing in the 3′ to 5′ direction. The experimental UniTags were matched against the virtual tag set, and 48249 UniTags from SHB and 46290 UniTags from SHBC were found to represent 18114 and 19323 genes, respectively (Table 2). Some reference genes that lacked CATG sites (10.84% of all reference genes) were not included in virtual tag database, transcripts which mapped to the those genes were excluded in this study. Thus, approximately 50% of reference genes were annotated by the experimental tags. The Illumina/Solexa DGE system is a combination of longSAGE with Solexa sequencing technology, generating 21-nt tags for mapping to annotated transcripts. These tags are more specific than the 14-nt or 19-nt tags acquired using the conventional SAGE strategy [26–29].

Approximately 7% of tags mapped to the antisense strand, demonstrating that these regions might be sense-antisense transcribed. Among the tags mapped to one reference gene, approximately 13% were mapped to multiple locations, including low complexity tags with poly-(A) tails and tags derived from repetitive sequences.

SOAP was used to map tags to soybean genome sequences. This allowed mapping of 52.21% of UniTags from SHB and 57.44% of those from SHBC to sequences in the soybean genome. Among the tags, 28.6% of UniTags from SHB and 27.7% of those from SHBC were unknown. This resulted from incomplete soybean genome annotation and the significant sequencing depth of this research. These values were higher than that in Li's study (14.7%) [25]. This difference may be related to the quality of the RNA samples. SAGE allows researchers to obtain a short sequence of any cDNA of interest, including genes for which there is no sequence information, and the full-length sequence of the cDNA of interest can be obtained by rapid amplification of cDNA ends (RACE). This is the main advantage of this strategy over other systems, such as microarray and quantitative RT-PCR. Some tags mapped to the soybean genome sequence, but were not aligned to any reference gene. These might represent non-annotated genes or non-coding transcripts derived from intergenic regions. In this study, approximately 50% of UniTags mapped to unannotated regions of the soybean genome, indicating transcription of as-yet-unidentified genes.

2.2 Comparison of expression profiles between SCN-infected and uninfected soybean roots

Normalization of the tag frequency between the two libraries was performed for comparative analysis. The distribution of tag copy numbers in the two libraries is shown in Figure 1. More than 97.24% of tags were expressed with a fold change of less than 5 between the two samples. At a statistically significant value ($P < 0.01$), we detected 11658 differentially expressed tags between infected and uninfected root samples.

The 15820 genes represented by more than two UniTags were selected for comparative analysis. Changes in gene expressions between the SCN-infected soybean root and uninfected root were quantified. A test was performed to identify statistically significant differences in gene expression. We used FDR < 0.001 and fold change values of $\geqslant 2$ or $\leqslant 0.5$ as the threshold. We identified 1405 genes that were expressed at higher levels in the SCN-infected soybean root than in the control root, and 1191 genes that were expressed at higher levels in the control root than in the SCN-infected root. More than 79.8% of genes showing frequency differences between the two libraries showed less than 5-fold differences. A total of 195 genes showed frequencies 20-fold greater in the SCN-infected root than in the uninfected control root, and 43 genes were more than 50-fold enriched in the infected root. In contrast, 149 genes were more than 20-fold enriched in the uninfected root, and 26 genes were enriched more than 50-fold. The 20 genes with the greatest differences in expression between and SHB and SHBC libraries are shown in S3. The gene showing the greatest increase in frequency in the infected sample compared with that in the control sample was a DNA-binding protein gene, followed by a tag with no similarity to known sequences in the database, and then a kinase gene. The gene showing the greatest increase in frequency in the uninfected sample compared with that in the infected sample was the gene encoding the photosystem I subunit PsaD, followed by a gene related to photosynthesis.

Almost all of the differentially expressed genes in the infected sample corresponded to low- and mid-abundance UniTags (<1000 copies), indicating that the greatest changes in expression levels occurred for genes expressed at low to intermediate levels.

To examine the characteristics of differentially expressed genes, functional classification of up- and down-regulated genes was performed using the gene ontology (GO) tool.

Table 2 Summary of UniTag mapping to gene data

Tag mapping	SCN-infected	Uninfected control
Sense		
Perfect match	14958	16905
1-bp mismatch	8549	7433
Antisense		
Perfect match	13686	15175
1-bp mismatch	5753	5577
All tags mapping to gene	46290	48249
All tag mapped genes	18114	19323
Tags mapping to genome	100709	135257
No matched tags	104062	107913

For this analysis, we performed BLASTX searches in the non-redundant GenBank and UniProtKB/TrEMBL protein databases. Of the 2596 differentially expressed genes, 2295 had an associated nucleotide sequence and 1236 matched to functionally annotated genes in the databases. Differentially expressed genes were involved in metabolism, stress or defense responses, phytohormone responses, cellular signaling, and other functions. GO annotations (biological processes) were assigned to 300 of these 1236 genes. Among the differentially expressed genes, 403 and 382 were annotated with terms from the Component Ontology and the Function Ontology, respectively. Comparison of statistically enriched GO terms for differentially expressed genes with the entire genome frequency revealed their involvement in cofactor binding CoA-ligase activity, acid-thiol ligase activity, structural molecule activity, coenzyme binding oxidoreductase activity, electron carrier activity, sucrose synthase activity, and as structural constituents of ribosomes (Figure 2).
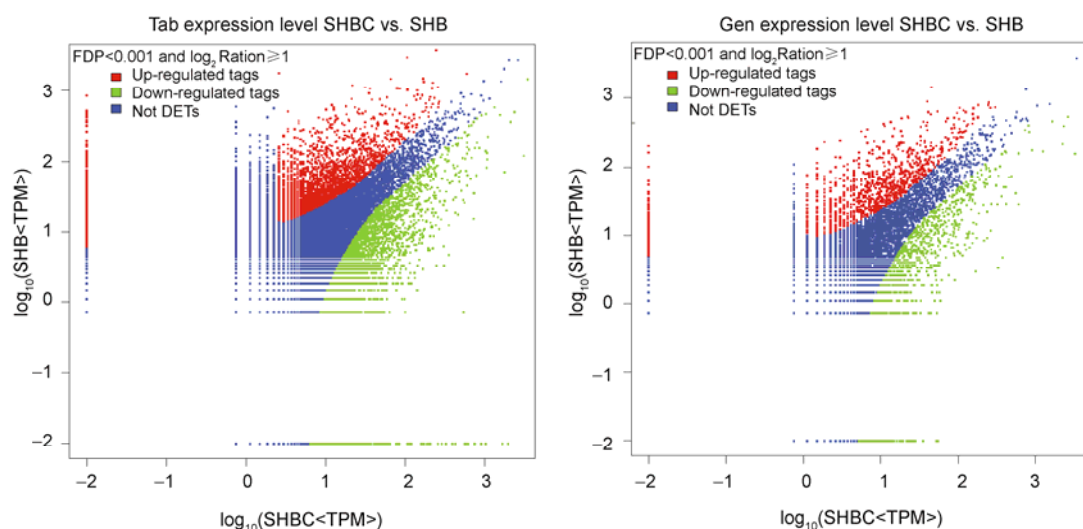

Figure 1   Distribution of expression levels in SCN-infected (SHB) and uninfected (SHBC) roots.
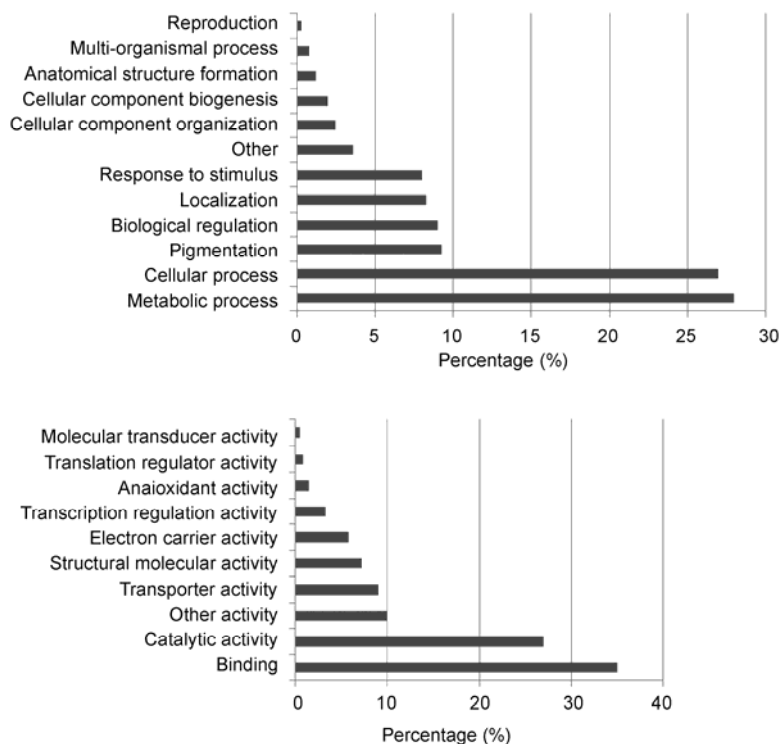

Figure 2   Distribution of annotated genes in Gene Ontology (GO) categories based on biological processes and molecular function. (a) Biological process category; (b) molecular function category.

It is difficult to compare these data directly with those from other studies, because of differences in soybean cultivars, pathosystems, time point selection, tissue sampling methods, and analytical technologies. However, several genes that were reported as up- or down-regulated in root-SCN interactions in previous studies were also detected in our study.

A major reprogramming of plant metabolism appears to occur during nematode-plant interactions. In this study, we found numerous metabolic and energy-related genes that were differentially expressed between SCN-infected and uninfected root samples. There were both up- and down-regulated genes in the SCN-infected root. For example, 21 genes involved in photosynthesis were differentially expressed between the two samples; 20 of them were down-regulated in the SCN-infected root. Both photosystems I and II were suppressed in SCN-stressed soybean. In *Arabidopsis,* a similar trend was observed in the responses of expressions of photosynthesis genes to 22 different forms of biotic damage. That is, the transcript levels of genes associated with the light reactions of photosynthesis, the carbon reduction cycle, and pigment synthesis decreased, regardless of the type of biotic attack [30]. The expression profiles of photosynthesis-related genes suggest that suppression of photosystem activity is an adaptive response to biotic attack. Genes encoding enzymes involved in secondary metabolism, including 4-coumarate-CoA ligase, chalcone flavonone isomerase 1A, chalcone and stilbene synthases, P450 monooxygenase CYP93C1v2p, CYP94C9, and CYPRO4 were expressed at high levels in SCN-infected roots. In addition, infected roots showed up-regulation of genes in the phenylpropanoid pathway, which leads to accumulation of flavonoids or isoflavonoid phytoalexins.

Several types of transcription factors (TF) showed substantial changes in expression in this study. TFs belonging to the MYB, AP2/EREBP, WRKY, NAC, and bZIP families have roles in various abiotic and biotic stress signal transduction pathways [31–33]. In the SCN-infected root, 8 out of 10 identified bZIP genes were up-regulated. The gene gi|37995943 (ID), the most up-regulated gene in this study, is an ERF transcription factor. WRKY transcripts appear to be a general characteristic of plant defense in the response to pathogens [34]. WRKY34, 23, 35, and 13, and an unclassified WRKY gene were induced by SCN infection, while two WRKY genes were down-regulated. Some TFs associated with control of plant development, for example, MADS box genes, were also identified here.

Changes in the expression of cell-wall-modifying proteins (CWMPs) such as endoglucanases, pectate lyases, and expansins are important in cyst nematode-plant interactions [35,36]. Plant CWMPs are activated after nematode infection, as shown in microarray analyses [11,35]. Differential regulation of CWMPs was also observed in the present study. Research on the *Arabidopsis* response to root-knot nematode infection showed that genes involved in cell-wall

metabolism were more induced than repressed [37]. Several genes related to cell-wall metabolism such as an α-expansin 4 precursor and endo-1,4-β-glucanase were down-regulated in the SCN-infected sample in the present study. One explanation for this phenomenon may be the time point we selected. Plant CWMPs play an important role in the nematode–plant interaction during syncytium formation. Our plant tissues were sampled 30 d after infection and at that point syncytium formation was already completed.

In many plant families, pathogenesis-related (PR) and similar proteins are induced by infection with various types of pathogens. The *PR-5* gene was induced by SCN infection in the present study. In addition, genes encoding chitinase III-A, a thaumatin-like protein, β-1,3-glucanase-related proteins, and phenylalanine ammonia-lyase 1 were up-regulated in the SCN-infected soybean root. These genes are homologous to *PR* genes in other plant species. Current views about PRs are based largely on *Arabidopsis,* because most microarray and other data are largely limited to *Arabidopsis* [38]. General stress-related genes such as those encoding glutathione-S-transferases and disease resistance responsive proteins (e.g. MsR1, SR1) were also up-regulated in SCN-infected roots.

2.3 Validation of SAGE results by qRT-PCR

To evaluate the validity of Solexa analysis of differential gene expression, 20 candidate genes were selected and their expressions detected by qRT-PCR. As shown in Table 3, the expression patterns of these genes were generally consistent with the sequencing data. The discrepancies between values of the ratios result from the different algorithms used in the two technologies [39]. In SAGE analyses, sampling effects have large impacts on the extent to which sequencing is carried out, which affects differential expression data. Deep sequencing methods overcome this disadvantage by improving sequence depth, thus generating absolute gene expression numbers. The fold change determined by qRT-PCR is a logarithmic measure of relative levels of expression; thus, ratios of gene expression differences are not generated by simple division of threshold cycle (Ct) values, and even slight differences in the PCR reaction conditions can greatly affect Ct values [40]. Most of the differently expressed genes had low-frequency tags at least in one library, and some of them were beyond the detection range of qRT-PCR. Genes with mid-frequency copy numbers in the two libraries matched better than others. For example, the Solexa sequence data indicated that the C2H2 zinc finger protein gene was 2.7-fold up-regulated, whereas qRT-PCR analysis indicated a 3.5-fold change.

3   Conclusions

This is the first report of the use of the Illumina/Solexa

Table 3 Confirmation of expression profiles of selected genes by qRT-PCR

| Gene ID | Description | Solexa sequence data | | | qRT-PCR |
		SCN-infected	Un-infected	SHB/SHBC[a]	SHB/SHBC
gi|19338625	Nodulin 6l	119	0	116.8	89.3
gi|151402393	Amino acid transporter	27	0	26.5	9.8
gi|51336897	Unknown	14	0	13.7	6.2
gi|210142157	Dof21	166	12	13.6	10.2
gi|213607948	Unknown	1573	118	13.1	12.2
gi|38679416	MADS-box protein	612	46	13.1	15.3
gi|113367245	bZIP121	52	7	7.3	12
gi|59668409	β-amylase	458	94	4.8	6.2
gi|22934160	C2H2 zinc finger protein	467	168	2.7	3.5
gi|58014611	β-1,3 glucanase	116	54	2.1	4.9
gi|192324880	Harpin-induced 1	199	399	0.489	0.26
gi|22934101	Glycosyltransferase	81	187	0.425	0.33
gi|710431	HS22M	19	50	0.373	0.15
gi|210145445	Ribosomal protein SA	33	116	0.279	0.48
gi|207693778	Unknown	7	31	0.222	0.09
gi|40457260	Proline dehydrogenase	5	53	0.0926	0.02
gi|210141004	Phenylalanine ammonia-lyase 2	427	8720	0.048	0.04
gi|151394854	Unknown	9	880	0.01	0.005
gi|58023824	Thr protein kinase	0	20	0.049	0.2
gi|210143403	Unknown	0	146	0.007	0.003

a) To avoid division by 0, we assigned a value of 1 to tags that were not detected in the sample. The sequence gene number was normalized to TPM.

DGE system to determine differences between the gene expression profiles of SCN-infected and healthy soybean roots. Our results showed that under SCN stress: (i) photosystems were repressed; (ii) transcription factors were up- or down-regulated; (iii) many *PR* genes were up-regulated; and (iv) many genes involved in the stress response were up-regulated while others were down-regulated. Although there are numerous un-annotated and unknown tags, our results highlight the usefulness of next-generation sequence data to investigate expression profiles in the genome, as well as those of specific candidate genes. The utilization rate of tags will increase as more soybean genomic and cDNA sequences become available. Further functional analysis of the differentially expressed genes and comparison of our data with future findings from genome-wide expression profiling obtained using other systems will provide deeper insight into the molecular mechanism of plant–nematode interactions.

1 Stacey G. Genetics and Genomics of Soybean. New York: Springer, Science and Business Media, 2008. 3–17

2 Wrather A, Shannon G, Balardin R, et al. Effect of diseases on soybean yield in the top eight producing countries in 2006. Plant Health Prog, 2010, 125: 1–3

3 Ramamurthy M, Wang G J, Halina T K. Polygalacturonase and polygalacturonase inhibitor protein: Gene isolation and transcription in *Glycine max*-Heterodera glycines interactions. Mol Plant-Microbe Interact, 1999, 12: 490–498

4 Mitra M, David P P, Jennifer K H, et al. Identification and characterization of a soybean ethylene-responsive element binding protein gene whose mRNA expression changes during soybean cyst nematode infection. Mol Plant-Microbe Interact, 2002, 15: 577–586

5 Zarir E V, Jessica A S, Randy C S, et al. Soybean FGAM synthase promoters direct ectopic nematode feeding site activity. Genome, 2004, 47: 404–413

6 Vander E W, Almeida E J, Inze D, et al. A molecular study of root-knot nematode-induced feeding sites. Plant J, 1996, 9: 45–54

7 Vaghchhipawala Z, Bassuner R, Clayton K, et al. Modulations in gene expression and mapping of genes associated with cyst nematode infection of soybean. Mol Plant-Microbe Interact, 2001, 14: 42–54

8 Hermsmeier D, Mazarei M, Baum T J. Differential display analysis of the early compatible interaction between soybean and the soybean cyst nematode. Mol Plant-Microbe Interact, 1998, 11: 1258–1263

9 Nadim A, Rana K, Benjamin M. Analysis of expressed sequence tags from roots of resistant soybean infected by the soybean cyst nematode. Genome, 2004, 47: 380–388

10 Nadim A, Vincent P K, Imed B C, et al. Time course microarray analyses reveal global changes in gene expression of susceptible *Glycine max* (soybean) roots during infection by *Heterodera glycines* (soybean cyst nematode). Planta, 2006, 224: 838–852

11 Nagabhushana I, Justin R, Dan N, et al. Parallel genome-wide expression profiling of host and pathogen during soybean cyst nematode infection of soybean. MPMI, 2007, 20: 293–305

12 Nagabhushana I, Justin R, Dan N, et al. Developmental transcript profiling of cyst nematode feeding cells in soybean roots. Mol Plant-Microbe Interact, 2007, 20: 510–525

13 Hideo M, Stefanie R, Ito A, et al. Gene expression analysis of plant host-pathogen interactions by SuperSAGE. Proc Natl Acad Sci USA, 2003, 100: 15718–15723

14 Jung S H, Lee J Y, Lee D H. Use of SAGE technology to reveal changes in gene expression in *Arabidopsis* leaves undergoing cold

stress. Plant Mol Biol, 2003, 52: 553–567

15 Taketo U, Sugiyama S P, Chikara M. Comparative serial analysis of gene expression of transcript profiles of tomato roots infected with cyst nematode. Plant Mol Biol, 2007, 63: 185–194

16 Mardis E R. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genetics, 2008, 9: 387–402

17 Yan W A, Eric W K, Aubrey T, et al. 3′ tag digital gene expression profiling of human brain and universal reference RNA using Illumina Genome Analyzer. BMC Genomics, 2009, 10: 531–542

18 Lesley J C, Patrick J B, Voelckel C, et al. An approach to transcriptome analysis of non-model organisms using short-read sequences. Genome Inform, 2008, 21: 3–14

19 Mark D R, Alicia O S. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol, 2010, 11: R25

20 Leonardo V, Arthur G, Carlos A B P. Significance tests for comparing digital gene expression profiles. Nat Prec, 2008, 3: 10101–10115

21 Drew R E, Walter L, Alan E P, et al. SAGE analysis of transcriptome responses in arabidopsis roots exposed to 2,4,6-trinitrotoluene. Plant Physiol, 2003, 13: 1397–1406

22 Thomas D S, Kenneth J L. Analyzing real-time PCR data by the comparative CT method. Nat Protoc, 2008, 3: 1101–1113

23 Matsumura H, Nirasawa S, Terauchi R. Transcript profiling in rice (*Oryza sativa L.*) seedlings using serial analysis of gene expression (SAGE). Plant J, 1999, 20: 719–726

24 Fregene M, Matsumura H, Akano A, et al. Serial analysis of gene expression (SAGE) of host-plant resistance to the cassava mosaic disease (CMD). Plant Mol Biol, 2004, 56: 563–571

25 Li Y J, Fu Y R, Huang J G, et al. Transcript profiling during the early development of the maize brace root via Solexa sequencing. FEBS J, 2011, 278: 156–166

26 Hu M, Kornelia P. Serial analysis of gene expression. Nat Protoc, 2006, 269: 1743–1760

27 Lee J Y, Lee D H. Use of serial analysis of gene expression technol-

ogy to reveal changes in gene expression in *Arabidopsis* pollen undergoing cold stress. Plant Physiol, 2003, 132: 517–529

28 Poroyko L G, Hejlek W, Spollen G, et al. The maize root transcriptome by serial analysis of gene expression. Plant Physiol, 2005, 138: 1700–1710

29 Fregene M, Matsumura H, Akano A, et al. Serial analysis of gene expression (SAGE) of host–plant resistance to the cassava mosaic disease (CMD). Plant Mol Biol, 2004, 56: 563–571

30 Damla D B, Jorge A Z, Zhu J, et al. Biotic stress globally down regulates photosynthesis genes. Plant Cell Environ, 2010, 33: 1597–1613

31 Karam B S, Rhonda C F, Luis O S. Transcription factors in plant defense and stress responses. Curr Opin Plant Biol, 2002, 5: 430–436

32 Rushton P J, Somssich I E, Ringler P, et al. WRKY transcription factors. Trends Plant Sci, 2010, 15: 247–258

33 Christian D, Ralf S, Erich G, et al. Transcription factors in *Arabidopsis*. Trends Plant Sci, 2010, 15: 10573–10581

34 Eulgem T. Regulation of the *Arabidopsis* defense transcriptome. Trends Plant Sci, 2005, 10: 71–78

35 Goellner M, Wang X H, Davis E L. Endo-beta-1,4-glucanase expression in compatible plant-nematode interactions. Plant Cell, 2001, 13: 2241–2255

36 Qin L, Kudla U, Roze E H A, et al. Plant degradation: A nematode expansin acting on plants. Nature, 2004, 427: 30

37 Jammes F, Lecomte P, Almeida J, et al. Genome wide expression profiling of the host response to root-knot nematode infection in *Arabidopsis*. Plant J, 2005, 44: 447–458

38 Sels J, Mathys J, DeConinck B M A, et al. Plant pathogenesis-related (PR) proteins: A focus on PR peptides. Plant Physiol Biochem, 2008, 46: 941–950

39 Dmitri V G, John J D, Sean R M, et al. Transcript profiling of human platelets using microarray and serial analysis of gene expression. Blood, 2003, 101: 2285–2293

40 Thomas D S, Kenneth J L. Analyzing real-time PCR data by the comparative CT method. Nat Protoc, 2008, 3: 1101–1113