Indian J. Hort. 68(4), December 2011: 522-528

Application of digital gene expression tag profiling on differential gene expression of two developmental stages in bulbs of *Lycoris sprengeri*

Le Chang, Yi-ping Xia*, Jing-jue Chen and Yu-mian Xiao
Department of Horticulture, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China

ABSTRACT

Gene expression-profiling data are necessary to understand more clearly the differential gene expression that occurs in biological developmental process. This study is aimed to compare the differential gene expression between bulb and bulblet of *Lycoris sprengeri*, as to reveal the biological and molecular mechanisms in bulb development. Method: Two Digital Gene Expression tag profiling (DGE) libraries of different developmental stages of *Lycoris sprengeri* were built according to Illumina sequencing and compared using bioinformatics methods. The altered expression levels of these genes, as well as the enrichment of gene ontology terms and pathways in the KEGG database were then analyzed in detail. Result: A total number of 3,476,452 and 3,597,890 of raw tags were generated for the mature and juvenile DGE libraries, and 4,389 significantly differentially expressed genes were obtained. 45,181 (13.95%) and 41,093 (32.25%) distinct tags for the mature and juvenile bulbs were mapped to a unique sequence in the reference database, respectively. For the DGE data quality check, the lowly-expressed tags (the copy number < 10) comprised most of the distinct clean tags, and the number of detected genes kept stable when the tag number larger than 2 million. Most genes expressed in fewer than 10 copies (log 10 < 1). The metabolic process, the catalytic activity, cell and cell part were the most significantly enriched GO term in the three categories, consisting of 74, 67.7, and 98.6% of their differentially expressed gene groups, respectively. Conclusion: DGE is proved to be effective method for differentially expressed gene analysis on bulb development of *Lycoris sprengeri*.

Key words: Digital Gene Expression tag profiling (DGE), *Lycoris sprengeri*, developmental stage, differential gene expression.

INTRODUCTION

Lycoris is a genus of flowers with remarkable ornamental, medical, ecological, and application value (Bryan, 3). However, the hereditary mechanism of this plant is very complex resulting in an extremely low propagation rate and a quite long life cycle. Only one or two bulblets are produced per year, and four to five years are required before the onset of blooming (Zhang *et al*., 22). These factors greatly restrict the propagation, breeding, and commercial development of *Lycoris*. The periods and mechanisms of *Lycoris* bulb development have been discussed in detail to study these plants.

Genomic information for *L. sprengeri* was unavailable before the authors' construction of the transcriptome library of this plant using Illumina GA II RNA-Seq technology. To comprehend the biological and molecular mechanisms of the bulb development of *Lycoris*, a tag-based Digital Gene Expression tag profiling (DGE) was used in the present study. DGE is based on the serial analysis of gene expression transcript profiling method. This method generates absolute, rather than relative, gene expression measurements and eradicates many "inherent limitations" of the traditional methods like microarray analysis (Feng *et al*., 7; Wang *et al*., 16). DGE is powerful in identifying differential genes and determining the abundance of gene expression in a sample tissue at different developmental stages or in different organs (Vizoso *et al*., 15). Xue *et al*. (19) obtained comprehensive gene expression information of brown plant hoppers using DGE. This study provided useful information for the identification of genes involved in the development modulation, wing dimorphism, and sex difference of the brown plant hoppers. Wang *et al.* (17) used DGE to compare the differentially expressed genes during the developmental stages of a whitefly and obtained an invaluable resource for the identification of genes involved in the insect development and virus transmission.

Two key developmental stages of *L. sprengeri* bulb were selected in the current study, and two DGE libraries were built. Then, bioinformatics technology was used to create an assembly and annotation according to the transcriptome library that the authors built in a prior study. Gene expression variations were compared, and the significant enrichment of gene ontology (GO) terms and pathways was analyzed. This process

*Corresponding author's E-mail: ypxia@zju.edu.cn

laid the foundation of the screening of the important regulating genes during bulb development, revealing the regulation mechanisms of gene expression and future molecular breeding of *L. sprengeri*.

MATERIALS AND METHODS

Four- to five-year-old mature bulbs of *L. sprengeri* at the dormancy period, 15.02 cm in circumference and 69.70 g in weight, were collected from the Zhoushan Islands in China. One-year-old juvenile bulbs (bulblets), 4.91 cm in circumference and 4.20 g in weight, were obtained from scale cutting in the Huajiachi Campus of Zhejiang University in Hangzhou, China. The fresh samples were used immediately to extract their total RNA. The total RNA was isolated from the bulbs using the SV Total RNA Isolation System (Promega) according to the manufacturer's protocol. DGE libraries were constructed using the Illumina Gene Expression Sample Prep Kit (Xue *et al*., 17) after the quality check of total RNA. Purified mRNA was captured from 6 μg of the total RNA using oligo (dT) magnetic bead adsorption. Oligo (dT) was also used as primer to synthesize the double-stranded cDNA through reverse transcription. The bead-bound cDNA was later digested with restriction enzyme NlaIII. The fragments apart from the 3′ cDNA fragments were purified by the magnetic bead and the Illumina adaptor 1 was ligated to the sticky 5′ ends. MmeI recognized the junction of Illumina adaptor 1 and CATG site, and cut at 17 bp downstream of the CATG site to generate tags containing adaptor 1. The tag library with different adaptors at both ends was finally finished by removing the 3′ fragments and adding Illumina adaptor 2. This tag library was amplified by 15 cycles of linear polymerase chain reaction, and 95 bp fragments were purified using 6% Tris/borate/EDTA polyacrylamide gel electrophoresis. Then, the single-strand molecules were fixed onto the Illumina Sequencing Chip (flow cell). Four types of nucleotides, which were marked in four colors, were added, and sequencing was performed through sequencing by synthesis. Each tunnel of flow cell generated millions of raw tags with sequencing length of 35 bp (Wang *et al*., 16; Zhang *et al*., 22).

Clean tags were obtained after filtering dirty tags from the generated raw tags by removing the adaptor sequence, empty reads (reads with no tags), low quality tags (tags with 'N'), tags too long or too short, and tags with a copy number of 1. Based on the authors' previous transcriptome database, Virtual libraries containing all the possible CATG+17 nucleotide tag sequences were generated (Wang *et al*., 17). For annotation, all clean tags were mapped to the reference database, and no more than 1 bp mismatch was allowed. After filtering the clean tags mapped to the reference sequences, the remaining were designed as unambiguous tags. The unambiguous clean tags for each gene was then normalized to transcripts per million clean tags (TPM) ('At Hoen *et al*., 1).

To compare the altered gene expression levels, Audic and Claverie's method (3) was used to analyze the frequency of each tag in the various cDNA libraries. In their method, False discovery rate (FDR) value was manipulated to determine the threshold of *P*-values, which corresponds to differential gene expression test, in the multiple tests. In the present study, significantly differentially expressed genes were characterized by an FDR value ≤ 0.001 and an absolute value of log 2 ratio>1. In the DGE analysis, the hypergeometric test is applied in GO and Kyoto encyclopedia of genes and genomes (KEGG) enrichment analyses of functional significance. So that all differentially expressed genes were mapped to the terms in the GO database and pathways in the KEGG database, facilitating in searching for significantly enriched GO terms and KEGG pathways compared with the genomic background (Xue *et al*., 19).

RESULTS AND DISCUSSION

The total number of raw tags generated from the Illumina sequencing of the mature and juvenile DGE

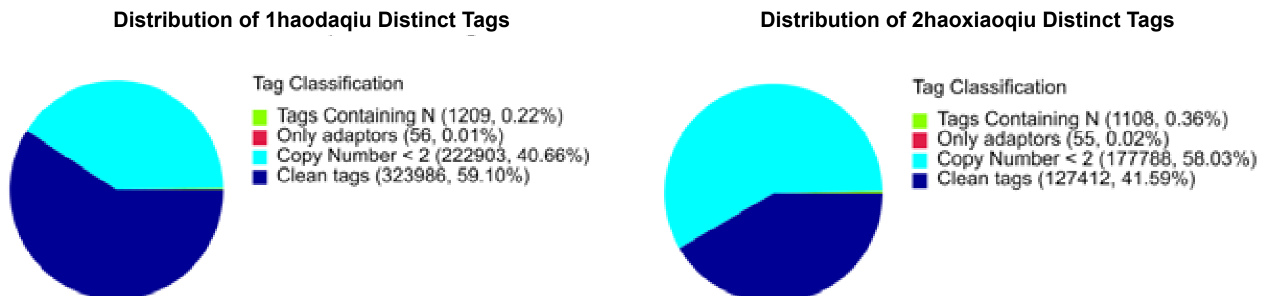

Fig. 1. Different components of the raw tags in mature and juvenile phases in *L. sprengeri* bulbs.

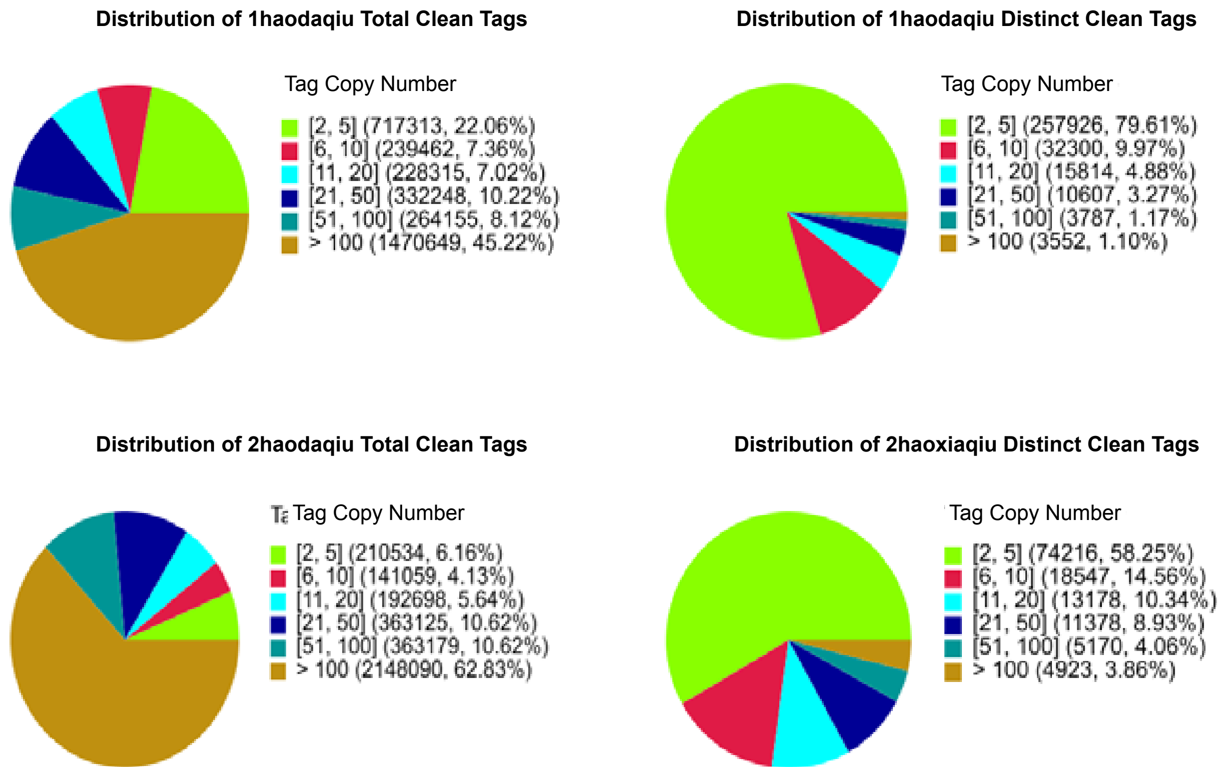

Fig. 2. Distribution of total clean tags and distinct clean tags in each sample of lycoris bulbs.

libraries was 3,476,452 and 3,597,890 (Table 1). And the total distinct tags of raw data were 548,154 and 306,363. After cleaning and quality checks, the tags containing 'N', the tags with only adaptors, and the too short tags were removed (Fig. 1). As a result, 323,986 (59.10%) and 127,412 (41.59%) distinct tags were obtained, respectively (Table 1).

Heterogeneity is one of the chief features of mRNA expression. The minority of mRNA has high expression abundance, while most of the others are lowly expressed. Therefore, the tag copy number indicates the expression of related genes, and the tag abundance categories is usually used to evaluate the normality of the DGE data (Fig. 2). Significant difference turned up between the distribution of total clean tags and the distribution of distinct clean tags for both mature and juvenile DGE. Overall, in the total clean tags, the highly-expressed tags (the copy number > 100) were dominant in the distribution. And the lowly-expressed tags (the copy number < 10) comprised most of the distinct clean tags.

There are insufficient studies of the genetic structure and function of *Lycoris*. Although the expressed sequence tags of *L. longituba* and *L. aurea* were sequenced by Cui (6) and He (9), respectively, little gene information on *Lycoris* could be found in public databases, which are also very difficult to be referenced. Thus, the authors established a transcriptome database library of *L. sprengeri* bulbs in a previous study using Illumina GA II based RNA-Seq technology, which provides a surprisingly large scale of short reads. These relative short reads could be effectively assembled and used to discover new genes as well as compare gene expression profiles (Rosenkranz *et al.*, 13; Hegedus *et al.*, 10), and to facilitate investigation of regulatory mechanisms for bulb development (Wu *et al.*,18; Lee *et al.*, 11).

This newly built transcriptome sequence data of *L. sprengeri* contains 98,150 distinct sequences with 167,641 unambiguous reference tags. Based on this, virtual libraries containing all the possible CATG+17 nucleotide tag sequences were generated. In the current study, among all the distinct clean tags generated from the DGE sequencing, 46,656 (14.40%) and 41,871 (32.86%) distinct tags for the mature and juvenile bulbs were successfully mapped to the reference sequences, respectively (Table 1). Of them, 45,181 (13.95%) and 41,093 (32.25%) distinct tags for the mature and juvenile bulbs were mapped to a unique sequence in the reference database, respectively

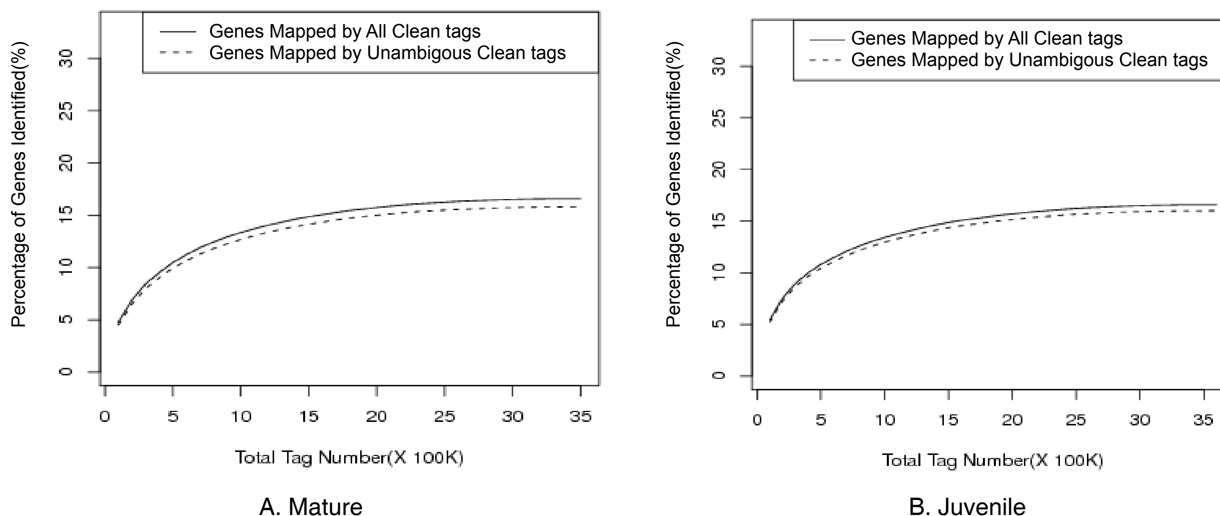

Fig. 3. Relationship between the number of detected genes and sequencing amount (total tag number) in lycoris.

(Table 1). These tags are extremely important, being properly identify a transcript. At the same time, in the reference database, 22,860 (23.29%) and 22,469 (22.89%) genes can be distinguished by unique tags of the DGE libraries (Table 1).

To ensure that we performed enough sequencing amount, a saturation analysis was carried out to check if any gene is missing (Fig. 3). The number of detected genes almost ceases to increase when the tag number reaches 2 million, which is far less than what we obtained (more than 3 million).

Using DGE sequencing, 15,512 and 15,719 TPM of the mature and juvenile bulbs were obtained, respectively. Gene expression level was then evaluated by calculating the number of TPM unambiguous tags. Figure 4 demonstrates that most genes express in fewer than 10 copies (log 10 < 1). Meanwhile, only small ratios of the genes are highly expressed (log 10 > 1). Basing on the next generation of high-throughput technology, the newly-developed DGE sequencing system is available for detecting the extremely lowly-expressed tags. In the present study, using the DGE libraries, most of the genes express in fewer than 10 copies (log 10 < 1). The similar results could be found in the report of different growth stages of whitefly (Wang *et al*., 16).

The 20 most abundantly expressed genes during these two stages were focused on (data not shown). In mature bulbs, 50% of the 20 most abundant genes were annotated, which were mapped to *Glycine max*, *Vitis vinifera*, *Ricinus communis*, *Sorghum bicolor*, *Picea sitchensis*, and *Oryza sativa* (Indica group). In juvenile bulbs, 60% of the 20 most abundant genes were annotated, homologous to *Musa acuminata, S. bicolor, Populus trichocarpa, V. vinifera, G. max,* and *O. sativa* (Japonica group). Concerned about only NCBI nr, GO, and KEGG database for the annotation, the annotated results of most of highly regulated genes are orphan sequences. It indicates that the development-related genes of *L. sprengeri* are slightly homologous with those of the other species. We suppose it probably because the underground bulbs were seldom used as samples to build cDNA library. In the other hand, the genus *Lycoris* has not been successfully domesticated yet, so the genome information is not that easy to be confirmed as the cultivated plants. Following work will be focused on searching with other databases, such as InterPro, trying to ascertain more clarifications about the functional information. Further studies on the functions of most abundant genes should be conducted.

To determine the differentially expressed genes during the *L. sprengeri* bulb development, the tag frequency in each DGE library was analyzed. A total of 21,712 significantly differentially expressed tags were detected and mapped to 4,389 genes. Most of these genes (2,748) were upregulated, whereas 1641 genes were down regulated (Fig. 5). Given that the TPM of juvenile/mature bulbs > 0 was defined as upregulated and that the TPM of juvenile/mature < 0 was defined as downregulated, most of the gene expression was restrained during bulb development (He *et al*., 8).

A total of 2,406 out of the 4,389 differentially expressed genes that were mapped in the KEGG database and compared with the transcriptome library were annotated, and the remaining 1,983 genes (no homologues found in the databases) were probably newly discovered genes. Five out of the ten most differentially upregulated genes were not annotated,

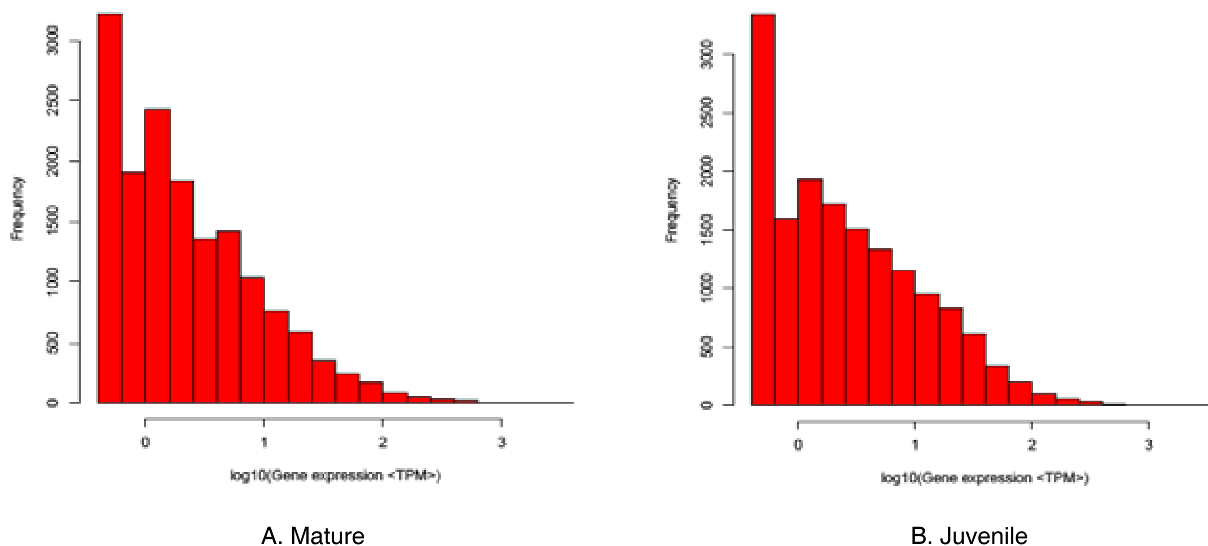

Fig. 4. Levels of expression for each gene.

whereas six out of the ten most downregulated genes were likewise not annotated. These 11 newly discovered genes are listed in Table 2.

Using aforementioned annotations, the BLAST2GO program (Conesa *et al.*, 5) was employed to obtain the GO annotation of terms. Then, the WEGO software (Ye *et al.*, 20) was used to perform GO function classification. Based on the GO database, all the tags were classified into three categories, namely, biological process, cellular component, and molecular function. Under biological process, the metabolic process was the most significantly enriched GO term, consisting of 74% of all the differentially expressed genes under this group. Under the molecular function group, terms with the catalytic activity comprise 67.7%, whereas under cellular component, cell and cell part equally occupied 98.6% of the genes. The intracellular,

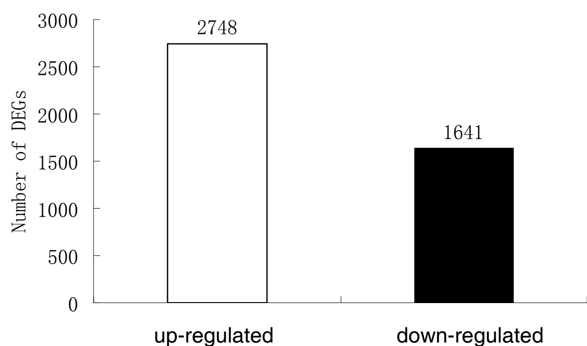

Fig. 5. Variations in the gene expression profile during the different developing stages of lycoris bulb.

intracellular part, organelle, intracellular organelle, membrane-bounded organelle, and intracellular membrane-bounded organelle functions were all greater than 50%. Twenty of the most differentially expressed genes that were annotated against GO database are listed in Additional file 2. Ten of these genes were upregulated whereas the other ten were downregulated.

To explore further the differentially expressed genes involved in the enriched metabolic or signal transduction pathways, all the genes were mapped to the terms in the KEGG database and annotated by the reference transcriptome database. A total of 17,159 genes and 1,164 differentially expressed genes were annotated (Table 3) and participated in 117 pathways. Seven of these pathways were significantly enriched (Q-value ≤ 0.05). The most specific enriched genes appeared in the metabolic pathways (up to 31.53% of all the annotated differentially expressed genes), followed by the biosynthesis of plant hormones, ribosomes, oxidative phosphorylation, carbon fixation in photosynthetic organisms, photosynthesis, and alpha-linolenic acid metabolism.

Among the seven pathways, genes expressed during photosynthesis and alpha-linolenic acid metabolism were all upregulated (Fig. 6). More downregulated genes than upregulated genes were found in the ribosome pathways. Additionally, upregulated and downregulated genes all existed in the metabolic pathways, biosynthesis of plant hormones, oxidative phosphorylation, and carbon fixation in photosynthetic organisms. However, upregulated genes were absolutely dominant.

Table 2. The 11 most significantly differentially expressed genes in the mature and juvenile lycoris bulbs.

Gene (upregulated)	Log 2 ratio (Juvenile/mature)	Gene (down regulated)	Log 2 ratio (Juvenile/ mature)
Unigene32854	15.71277	*Unigene13268*	-13.8433
Unigene74113	15.51246	*Unigene42422*	-13.6395
Unigene44040	14.5499	*Unigene44933*	-13.3476
Unigene49016	14.03531	*Unigene34900*	-13.2779
Unigene18407	13.52858	*Unigene43908*	-13.176
		Unigene36784	-12.9649

Table 3. Pathway enrichment between differentially expressed genes from the two developmental stages in lycoris.

Pathway	DEGs with pathway annotation (1164)	Genes with pathway annotation (17159)	P-value	Q-value	Pathway ID
Metabolic pathways	367 (31.53%)	4114 (23.98%)	7.51E-10	4.39E-08	ko01100
Biosynthesis of plant hormones	86 (7.39%)	921 (5.37%)	0.001475012	2.88E-02	ko01070
Ribosome	58 (4.98%)	348 (2.03%)	1.51E-10	1.77E-08	ko03010
Oxidative phosphorylation	36 (3.09%)	301 (1.75%)	0.000654867	1.53E-02	ko00190
Carbon fixation in photosynthetic organisms	27 (2.32%)	195 (1.14%)	0.000316744	9.26E-03	ko00710
Photosynthesis	18 (1.55%)	72 (0.42%)	1.00E-06	3.92E-05	ko00195
α-Linolenic acid metabolism	18 (1.55%)	124 (0.72%)	0.001749842	2.92E-02	ko00592

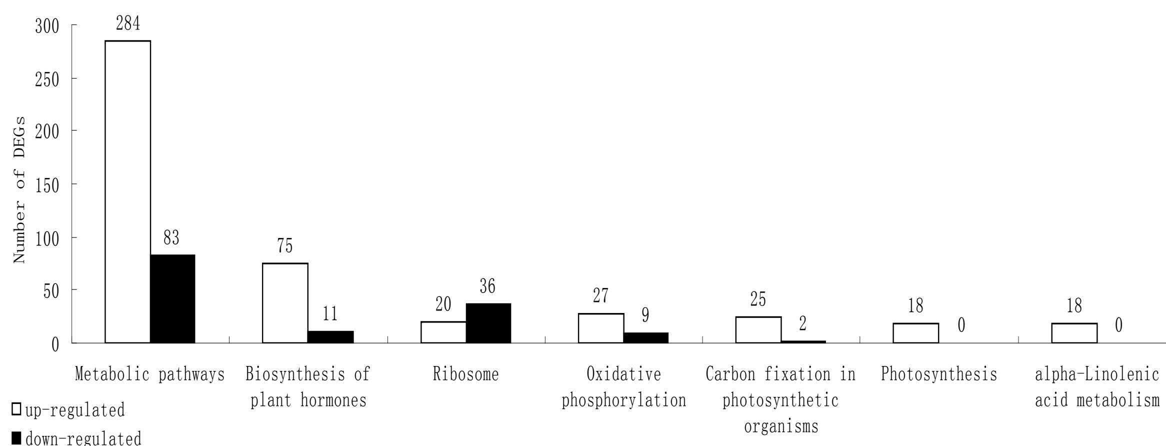

Fig. 6. Summary of the upregulated and downregulated genes of the seven significantly enriched pathways.

REFERENCES

1. At Hoen, P.A.C., Ariyurek, Y., Thygesen, H.H., Vreugdenhil, E., Vossen, R.H.A., Menezes, R.X., Boer, J.M., Ommen, G.J.B. and Dunnen, J.T. 2008. Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Res.* **36**: 141-44.

2. Audic, S. and Claverie, J.M. 1997. The significance of digital gene expression profiles. *Genome Res.* **7**: 986-95.

3. Bryan, J. 1992. *Manual of Bulbs*. Timber Press, Portland Oregon, 155 p.

4. Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M. and Robles, M. 2005. Blast2GO: a universal

tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**: 3674-76.

5. Cui, Y.L. 2004. Construction of cDNA library and analysis on expressed sequence tags of genus of *Lycoris* plant. Ph.D thesis, Nanjing Forestry University, Nanjing, China (*in Chinese*).

6. Feng, L., Liu, H., Liu, Y., Lu, Z.K., Guo, G.W., Guo, S.P., Zheng, H.W., Gao, Y.N., Cheng, S.J., Wang, J., Zhang, K.T. and Zhang, Y. 2010. Power of deep sequencing and agilent microarry for gene expression profiling study. *Mol Biotechnol*. **45**: 101-10.

7. He, K., Zhao, H.B., Bai, C.Y., Wang, Q.S. and Pan, Y.C. 2010. Study on regulatory mechanism of bovin early embryos development based on the analysis of gene expression data. *Hereditas*, **32**: 725-31 (*in Chinese*).

8. He, Q.L. 2006. Construction of cDNA library and analysis on expressed sequence tags of genus of *Lycoris longituba*. MS thesis, Nanjing Forestry University, Nanjing, China (*in Chinese*).

9. Hegedus, Z., Zakrzewska, A., Agoston, V.C., Ordas, A., Racz, P., Mink, M., Spaink, H.P. and Meijer, A.H. 2009. Deep sequencing of the zebra fish transcriptome response to *Mycobacterium* infection. *Mol. Immunol*. **46**: 2918-30.

10. Lee, J.J., Hassan, O.S.S., Gao, W.X., Wei, N.E., Kohel, R.J., Chen, X.Y., Payton, P., Sze, S.H., Stelly, D.M. and Chen, Z.J. 2006. Developmental and gene expression analyses of a cotton naked seed mutant. *Planta*, **223**: 418-32.

11. Morrissy, A.S., Morin, R.D., Delaney, A., Zeng, T., McDonald, H., Jones, S., Zhao, Y.J., Hirst, M. and Marra, M.A., 2009. Next-generation tag sequencing for cancer gene expression profiling. *Genome Res*. **13**: 187-94.

12. Sun, C.R., Shinozaki, K. and Sugiura, M. 1986. Molecular cloning of the genes for the large subunit of ribulose-1,5-disphosphate carbo-xylase/oxygenase from broad bean chloroplast DNA. *Chinese J. Biotech.* **2**: 19-24 (*in Chinese*).

13. Vizoso, P., Meisel, L.A., Tittarelli, A., Latorre, M., Saba, J., Caroca, R., Maldonado, J., Cambiazo, V., Campos-Vargas, R. and Gonzalez, M. 2009. Comparative EST transcript profiling of peach fruits under different post-harvest conditions reveals candidate genes associated with peach fruit quality. *BMC Genomics*, **10**: 423.

14. Wang, Q.Q., Liu, F., Chen, X.S., Ma, X.J., Zeng, H.Q. and Yang, Z.M. 2010. Transcriptome profiling of early developing cotton fibre by deep-sequencing reveals significantly differential expression of genes in a fuzzless/ lintless mutant. *Genomics*, doi:10.1016/j.ygeno.2010.08.009.

15. Wang, X.W., Luan, J.B., Li, J.M., Bao Y.Y., Zhang C.X. and Liu S.S. 2010. De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics*, **11**: 400.

16. Wu, Y.R., Machado A.C., White, R.G. Llewellyn, D.J. and Dennis, E.S. 2006. Expression profiling identifies genes expressed early during lint fibre initiation in cotton. *Plant Cell Physiol*. **47**: 107-27.

17. Xue, J., Bao, Y.Y., Li, B.L., Cheng, Y.B., Peng, Z.Y., Liu, H., Xu, H.J., Zhu, Z.R., Lou, Y.G., Cheng, J.A. and Zhang, C.X. 2010. Transcriptome analysis of the brown planthopper *Nilaparvata lugens*. *PLOS one.* **5**: 14233-38.

18. Ye, J., Fang, L., Zheng, H.K., Zhang, Y., Chen, J., Zhang, Z.J., Wang, L., Li, S.T., Li, R.Q., Bolund, L. and Wang, J. 2006. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res*. **34**: W293-97.

19. Zhang, G.J., Guo, G.W., Hu, X.D., Zhang, Y., Li, Q.Y., Li, R.Q., Zhuang, R.H., Lu, Z.K., He, Z.Q. and Fang, X.D. 2010. Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res*. **20**: 646-54.

20. Zhang, L., Wang, G.P. and Cao, F.L. 2002. Studies on vegetative propagation in the genus *Lycoris*. *J. Nanjing Forest. Univ. (Natural Sciences)*. **26**:1-5 (*in Chinese*).

Received: January, 2011; Revised: September, 2011; Accepted : October, 2011