韓藥作誌(Korean J. Medicinal Crop Sci.) 22(5) : 339 – 348 (2014)
ISSN(Print) 1225-9306
ISSN(Online) 2288-0186
http://www.medicinalcrop.org
http://dx.doi.org/10.7783/KJMCS.2014.22.5.339

차세대염기서열 분석을 이용한 고려인삼과 미국삼의 전사체 분석

조익현* · 김영창* · 이승호* · 김장욱* · 김선태** · 현동윤*
김동휘* · 김기홍* · 김홍식*** · 정종욱****† · 방경환*†

*농촌진흥청 국립원예특작과학원 인삼특작부, **부산대학교 식물생명과학과,
***충북대학교 식물자원학과, ****농촌진흥청 국립유전자원센터

Characterization of Root Transcriptome among Korean Ginseng Cultivars and American Ginseng using Next Generation Sequencing

Ick Hyun Jo*, Young Chang Kim*, Seung Ho Lee*, Jang Uk Kim*, Sun Tae Kim, Dong Yun Hyun*, Dong Hwi Kim*, Kee Hong Kim*, Hong Sig Kim***, Jong Wook Chung****† and Kyong Hwan Bang*†**

**Ginseng Research Division, Department of Herbal Crop Research, NIHHS, RDA, Eumseong 369-873, Korea.*
***Department of Plant Bioscience, Pusan National University, Miryang 627-706, Korea.*
****Department of Plant Resources, Chungbuk National University, Cheongju 361-763, Korea.*
*****National Agrobiodiversity Center, National Academy of Agricultural Science, RDA, Suwon 441-853, Korea.*

ABSTRACT : The transcriptomes of four ginseng accessions such as Cheonryang (Korean ginseng cultivar), Yunpoong (Korean ginseng cultivar), G03080 (breeding line of Korean ginseng), and *P. quinquefolius* (American ginseng) was characterized. As a result of sequencing, total lengths of the reads in each sample were 156.42 Mb (Cheonryang cultivar), 161.95 Mb (Yunpoong cultivar), 165.07 Mb (G03080 breeding line), and 166.48 Mb (*P. quinquefolius*). Using a BLAST search against the Phytozome databases with an arbitrary expectation value of 1E-10, over 20,000 unigenes were functionally annotated and classified using DAVID software, and were found in response to external stress in the G03080 breeding line, as well as in the Cheonryang cultivar, which was associated with the ion binding term. Finally, unigenes related to transmembrane transporter activity were observed in Cheonryang and *P. quinquefolius*, which involves controlling osmotic pressure and turgor pressure within the cell. The expression patterns were analyzed to identify dehydrin family genes that were abundantly detected in the Cheonryang cultivar and the G03080 breeding line. In addition, the Yunpoong cultivar and *P. quinquefolius* accession had higher expression of heat shock proteins expressed in *Ricinus communis*. These results will be a valuable resource for understanding the structure and function of the ginseng transcriptomes.

Key Words **:** *Panax ginseng*, *P. quinquefolius*, Transcriptomes, Next Generation Sequencing

INTRODUCTION

Next-generation sequencing (NGS) is greatly expanding sequencing depth and coverage (Heng and Homer, 2010; Margulies *et al*., 2005). This has great significance for non-model organisms, which are of interest to respective scientific communities, as development of non-model species has lagged relative to model species (Ellegren, 2008; Meyer *et al*., 2008; Vera *et al*., 2008).

RNA sequencing (RNA-seq), or deep-sequencing of RNAs, is a new technology for transcriptome profiling using NGS. RNA-Seq has been widely used to identify and quantify transcriptomes with unprecedented high resolution and low cost (Wang *et al*., 2009). RNA sequencing (RNA-seq), which is more cost-effective and more feasible than whole genome sequencing, is becoming a powerful tool in medicinal plant research and has accelerated the investigation of the plant gene expression (Egan *et al*., 2012). Previous

†Corresponding author: (Phone) +82-43-871-5531, +82-63-238-4871 (E-mail) bang31@korea.kr, jwchung72@korea.kr
Received 2014 June 25 / 1st Revised 2014 July 7 / 2nd Revised 2014 July 18 / 3rd Revised 2014 August 6 / Accepted 2014 August 6

Table 1. *P. ginseng* and *P. quinquefolius* accessions used in this study.

Name	Voucher No.	Classification	Remarks
Cheonryang	MPS002425	Korean cultivar	High-yielding
Yunpoong	MPS002380	Korean cultivar	High-yielding
G03080	MPS002431	Korean breeding line	Salinity tolerance
P. quinquefolius	MPS002310	American accession	Heat tolerance

studies of transcriptome in ginseng have focused on discovery of candidate genes associated with ginsenoside biosynthesis (Kim *et al*., 2013; Park *et al*., 2013). Nevertheless, gene discovery and expression profiling in ginseng are still very limited.

Most *Panax* species transcriptome studies have been performed on the 454-pyrosequencing platform because of the longer read and length possibilities compared with those of other platforms (Sun *et al*., 2010; Chen *et al*., 2011; Li *et al*., 2013). Notably, the transcriptome sequence for *P. quinquefolius* has been generated and assembled successfully using the Roche (454) GS FLX Genome Analyzer (Wu *et al*., 2013). The *P. ginseng* transcriptome has also been characterized using the 454 sequencing platform (Wu *et al*., 2012), and 454 GS-FLX Titanium technology provides around 1,000,000 sequences in a single 10-hour run. These sequences, with an average read length of 330 bp, may be up to 500 bp under shotgun libraries conditions, and are much longer than those obtained with other approaches (Gilles *et al*., 2011). This makes mapping easier, particularly for repetitive regions, and facilitates de novo transcriptome sequencing, exome capture, metagenomics, and amplicon sequencing (Metzker, 2009).

In this study, we analyzed transcriptome sequencing of 4-year-old *P. ginseng* and *P. quinquefolius* roots. Novel and tissue-specific *P. ginseng* and *P. quinquefolius* unigenes were obtained. We also performed annotation and functional category classification for these unigenes using the public databases (http://phytozome.net, http://david.abcc.ncifcrf.gov/ and http://ncbi.nlm.nih.gov/-COG/).

MATERIALS AND METHODS

1. Plant materials

The experiment was conducted at the field of National Institute of Horticultural and Herbal Science (NIHHS) of Rural Development Administration (RDA) in Eumsung (127º 45' 13.14" E, 36º 56' 36.63" N), Republic of Korea. *P. ginseng*, Cheonryang and Yunpoong cultivars, have been bred by pure line selection among landrace populations in Korea. Cheonryang and Yunpoong were developed at NIHHS of RDA and Korea Tomorrow & Global Central Research Institute, respectively.

The *P. ginseng* breeding line G03080, was selected for salinity tolerance plant through simple test of mature leaves (Hyun, 2012) at the NIHHS of the RDA during 2010 ~ 2012. *P. quinquefolius* accession was preserved and cultivated at the experimental field of the NIHHS, RDA, Chungbuk Province, Korea. These samples were deposited at Korean medicinal herbarium in NIHHS (Table 1).

At the maturity of each accession (Cheonryang, Yunpoong, G03080 and *P. quinquefolius*) were harvested from cultivated field of NIHHS in Eumsung, Korea in November, 2012. All the materials were planted in the same environment. The weather on the day of the samples was collected, November 3, 2012, was as follows: average temperature 3.4℃, max temperature 11.5℃, min temperature −2.5℃, and minimum grass temperature −3.7℃.

2. RNA isolation

Fresh roots of 4-year-old plants from ginseng accessions (*P. ginseng* and *P. quinquefolius*) were quickly cut the tissue into small pieces with a sterile razor blade, freeze in liquid nitrogen and grind well with use of mortar and pestle. In total, 3 g of root tissue for each sample was used to isolate total RNA. Total RNA extraction was performed using a Trizol kit (Invitrogen, Carlsbad, CA) according to the manufacturer's manual. After the RNA samples had been isolated and dried, they were dissolved in diethylpyrocarbonate-treated H_2O, and a Biospec-Nano spectrophotometer (Shimadzu, Kyoto, Japan) was used to determine the concentration of the solution.

The mRNA was purified from total RNA using the PolyATract mRNA Isolation Kit (Promega, Madison, WI,

USA). Quantity and quality (purity and integrity) of mRNA were assessed by two methods. First, the absorbance at different wavelengths was measured with a NanoDrop ND/1000 UV spectrophotometer (Thermo Fisher Scientific). The ratios of absorbance at 260/280 nm and 260/230 nm were used to assess RNA purity. A 260/230 ratio was used to estimate the presence of contaminants such as salts, carbohydrates, or peptides, among others, while an A260/280 ratio was used to estimate the purity of mRNA. Both ratios should show values close to 2.0 ~ 2.2 for pure mRNA. Second, capillary electrophoresis in an RNA Pico 6000 chip was performed using an Agilent BIOANALYZER 2100 (Agilent Technologies). Integrity of mRNA was estimated by the electropherogram profile.

3. Library construction

A cDNA library was constructed from isolated mRNAs using a ZAP-cDNA Synthesis Kit (Stratagene, Santa Clara, CA, USA) following the manufacturer's protocol and then cDNA was fragmented by nebulization for library construction. For 454 pyrosequencing, the GS FLX general library preparation method manual workflow (Roche Diagnostics) was followed. First-strand cDNA synthesis was performed with oligo (dT) primer as described in the provided protocol using 500 ng total RNA. Then 454 sequencing adapters A (5'-GCCTCCCTCGCGCCATCAG-3') and B (5'-CTGAGCGGGCTGGCAAG-GC-3') were ligated to the 5' and 3' cDNA ends. The cDNA carrying both, adapter A and adapter B attached to its ends was selected and then amplified with PCR using a proof reading enzyme (20 cycles) according to the 454 Life Science sequencing kit (454 Life Sciences, Branford, USA). Normalization was carried out by one cycle of denaturation and reassociation of the cDNA. Reassociated dscDNA was separated from the remaining ss-cDNA (normalized cDNA) by passing the mixture over a hydroxylapatite column. After hydroxylapatite chromatography, the ss-cDNA was amplified with 8 PCR cycles. Finally, the quality of this single-stranded template DNA (sstDNA) library was assessed using a 2100 BioAnalyzer (Agilent, Waldbronn, Germany). Approximately 10 g of normalized double stranded cDNA was used for 454 pyrosequencing.

4. Pyrosequencing and *de novo* assembly

The four captured cDNA-sequencing libraries were prepared separately and each one loaded on one of the four-lane gasket PicoTiterPlate device (PTP; 7075 mm; 454 Life Sciences, Branford, USA), respectively. Each library was quantitated by Pico-Green (Quant-iT, Molecular Probes, Invitrogen), diluted to 1×10^5 molecules/ul and used for emulsion polymerase chain reaction (emPCR) at a ratio of 0.1 copies of library fragments per DNA capture bead. The loaded PTP was then inserted into the GS-FLX Titanium instrument (Roche, Mannheim, Germany) for pyrosequencing, and sequencing reagents were sequentially flowed over the plate. Information from all the wells of the PTP is captured simultaneously by a camera and can be processed in real time by the onboard computer.

The data from the 454 read sequences of each sample were assembled into contigs using the proprietary Roche 454 Newbler Assembler software (version 2.3; Roche). SolexaQA package software (version 1.13) were used to remove the low-quality sequences, end regions that were rich in ambiguous nucleotides, very short reads (< 50 bp), poly (A/T) tails, adaptors for cDNA synthesis, primers and potential contaminating vector sequences. After removing low quality sequences, filtering for adaptors and primers, curated raw 454 read sequences were assembled by using the Roche 454 Newbler Assembler software.

5. Transcriptome analysis

Functional annotation was performed by sequence comparison with public database. All unique assembled sequences (unigenes) were sequentially compared using blast (cutoff E-value of 1e-10) with the sequence in Phytozome web database (http://www.phytozome.net). To further annotate the unigenes in this study, Gene Ontology (GO) term were assigned to the set of unigenes that showed hits against the TAIR (www.arabidopsis.org/) database using the functional annotation tool, available on the National Institute of Allergy and Infectious Diseases (NIAID) website http://david.abcc.ncifcrf.gov/.

RESULTS

1. Pyrosequencing and *de novo* assembly

The quarter-plate 454 pyrosequencing reaction of normalized cDNA was done using the GS FLX titanium platform. A total of 649 Mb of raw sequence data were generated, of which 546 Mb remained after trimming (84%

Table 2. Data regarding ginseng expressed sequence tags generated by the 454 GS-FLX platform.

Sample	No. of reads†	Raw data size‡ (bp)	After trim§		After trim data /raw data (ratio)
			Avg. read length	Total size (bp)	
Cheonryang	297,170	156,420,089	441	130,946,840	83.71%
Yunpoong	305,673	161,956,527	435	132,877,608	82.05%
G03080	311,861	165,070,255	451	140,752,190	85.27%
P. quinquefolius	308,313	166,489,892	460	141,795,358	85.17%
Total	1,223,017	649,936,763		546,371,996	

†No. of reads; the reads used in the assembly computation, ‡Raw data size; the reads' bases used in the assembly computation, §Trim; application of a quality filter to raw sequencing data.

Table 3. De novo reference assembly of transcriptome sequences from four raw sequences reads.

	No. of unigenes†	Total size (bp)	N size‡ (bp)
Reference§	57,272	61,943,470	2,161

†No. of unigenes; The number of unigenes annotated in the protein database (http://www.phytozome.com/poplar.php), ‡N size; Half of all bases reside in contigs of this size, §Reference; Assembled reads from raw reads of all four samples.

Table 4. Annotation results of reference unigenes against Phytozome database sequence.

No. of reference unigenes	57,272
Amino acids with start and end codon	40,510
Amino acids with start or end codon	16,762
Annotated transcripts	32,542
full-length transcripts	5,159 (15.85%)
partial transcripts	27,383 (84.15%)
No. of matched reference genes	20,901

of the sequenced bases). A summary of the 454 sequencing data for ginseng samples is presented in Table 2. As a result of pyrosequencing, 1,223,017 raw reads were generated, including 297,170 reads for Cheonryang, 305,673 for Yunpoong, 311,861 for the G03080 breeding line, and 308,313 for *P. quinquefolius*. The total and mean lengths of the reads were 156,420,089 bp and 441 bp for Cheonryang, 161,956,527 bp and 435 bp for Yunpoong, 165,070,255 bp and 451 bp for G03080, and 166,489, 892 bp and 460 bp for *P. quinquefolius* (Table 2).

References required for mapping were obtained by assembling raw reads of all four samples. The number of reference unigenes was 57,272, and total length was 61,943,470 bp, which included 2,161 bp of N (Table 3). The translation results of the reference unigenes showed that there were 40,510 cases in which the longest sequence included both a start and an end codon, and 16,762 cases included either a start or an end codon. The results of annotating the reference unigenes from the phytozome database (http://www.phytozome.net) showed that only 32,542 bp were annotated of the total 57,272 bp. Of these, 5,159 were estimated to be full length and the remaining 16,762 were partial length (Table 4).

2. The most abundant transcripts among the four samples

The most abundant transcripts in each ginseng root are listed in Table 5. All samples had high uncharacterized protein contents, which showed high similarity with 30S ribosomal protein S11, ATP synthase subunit alpha, and cell wall-associated hydrolase (*Medicago truncatula*). It was also confirmed that Cheonryang and the G03080 cultivar showed higher expression rates for dehydrin 1 and dehydrin 7 genes, which appeared specifically in *P. ginseng* com-pared to other samples. In addition, isotig 07914, encoding a heat shock protein of *Ricinus communis*, seemed to be expressed higher in Yunpoong and *P. quinquefolius* (Table 5). The most abundant category of transcripts had hits to uncharacterized proteins. That role is still largely unknown but they may play a regulatory role in gene expression.

3. Comparison of dehydrin gene expression

Based on the transcriptome analysis results, the dehydrin gene in P. ginseng and matching unigenes for each sample were summarized using the NCBI database in Table 6. The dehydrins in *P. ginseng* were divided into nine groups. The Dhn 9 gene was specifically expressed in seeds. In the case of Dhn1, 742, 632, 318, and 106 unigenes were confirmed in Cheonryang, G03080, Yunpoong, and *P. quinquefolius*, respectively. Dhn 1 transcript levels in Cheonryang and G03080 samples were higher than those

Table 5. Summary of most abundant unigenes in the transcriptome of each ginseng samples.

Unigene ID	No. of reads	Accession No.	E-value	Annotation	Source
P. ginseng cv. Chunryang					
contig00502	15,269	XP_003544036	3e-73	uncharacterized protein	*Glycine max*
isotig02853	1,576	XP_003606056	6e-168	30S ribosomal protein S11	*Medicago truncatula*
contig01049	1,503	XP_003588326	1e-87	ATP synthase subunit alpha	*Medicago truncatula*
isotig05031	1,115	XP_003637074	6e-97	Cell wall-associated hydrolase	*Medicago truncatula*
isotig11429	1,082	XP_003614394	1e-43	hypothetical protein MTR	*Medicago truncatula*
isotig02859	905	XP_002283383	2e-83	scarecrow-like protein 14-like	*Vitis vinifera*
isotig00651	742	ABF48474	1e-56	dehydrin 1	*Panax ginseng*
isotig00646	735	ABF48480	2E-27	dehydrin 7	*Panax ginseng*
isotig06456	470	HQ171905	2e-62	PR10-1 mRNA	*Panax ginseng*
Contig01470	444	XP_002280381	2e-61	F-box protein At2g16365-like	*Vitis vinifera*
P. ginseng cv. Yunpoong					
contig00502	15,047	XP_003544036	3e-73	uncharacterized protein	*Glycine max*
isotig02853	1,614	XP_003606056	6e-168	30S ribosomal protein S11	*Medicago truncatula*
isotig02863	1,460	XP_003588355	2e-127	Mitochondrial protein	*Medicago truncatula*
isotig05031	1,426	XP_003637074	6e-97	Cell wall-associated hydrolase	*Medicago truncatula*
isotig11429	1,228	XP_003614394	1e-43	hypothetical protein MTR	*Medicago truncatula*
contig01049	1,189	XP_003588326	1e-87	ATP synthase subunit alpha	*Medicago truncatula*
isotig02859	869	XP_002283383	2e-83	scarecrow-like protein 14-like	*Vitis vinifera*
isotig07914	558	XP_002527736	0	heat shock protein	*Ricinus communis*
Isotig04918	501	XP_003614380	8e-58	Cytochrome P450 likeTBP	*Medicago truncatula*
isotig06456	440	HQ171905	2e-62	PR10-1 mRNA	*Panax ginseng*
P. ginseng breeding line G03080					
contig00502	17,387	XP_003544036	3e-73	uncharacterized protein	*Glycine max*
isotig05031	2,132	XP_003637074	6e-97	Cell wall-associated hydrolase	*Medicago truncatula*
isotig02853	1,486	XP_003606056	6e-168	30S ribosomal protein S11	*Medicago truncatula*
isotig02863	1,402	XP_003588355	2e-127	Mitochondrial protein	*Medicago truncatula*
isotig11429	1,324	XP_003614394	1e-43	hypothetical protein MTR	*Medicago truncatula*
contig01049	1,176	XP_003588326	1e-87	ATP synthase subunit alpha	*Medicago truncatula*
isotig02859	834	XP_002283383	2e-83	scarecrow-like protein 14-like	*Vitis vinifera*
isotig00651	632	ABF48474	1e-56	dehydrin 1	*Panax ginseng*
isotig00646	627	ABF48480	2E-27	dehydrin 7	*Panax ginseng*
isotig06456	560	HQ171905	2e-62	PR10-1 mRNA	*Panax ginseng*
P. quinquefolius accession					
contig00481	7,826	XP_003544036	2e-61	uncharacterized protein	*Glycine max*
isotig02853	2,361	XP_003606056	6e-168	30S ribosomal protein S11	*Medicago truncatula*
isotig06456	1,519	HQ171905	2e-62	PR10-1 mRNA	*Panax ginseng*
isotig02859	1,252	XP_002283383	2e-83	scarecrow-like protein 14-like	*Vitis vinifera*
isotig05031	1,236	XP_003637074	6e-97	Cell wall-associated hydrolase	*Medicago truncatula*
contig01049	1,072	XP_003588326	1e-87	ATP synthase subunit alpha	*Medicago truncatula*
isotig04918	1,044	XP_003614380	8e-58	Cytochrome P450 likeTBP	*Medicago truncatula*
isotig11429	956	XP_003614394	1e-43	hypothetical protein MTR	*Medicago truncatula*
isotig07914	437	XP_002527736	0	heat shock protein	*Ricinus communis*
isotig03088	347	ABW74471		auxin-repressed protein	*Paeonia suffruticosa*

Table 6. Identification of dehydrin genes from the ginseng root transcriptomes and compared among their unigenes.

Gene name	Accession No.	No. of unigenes			
		Chunryang	Yunpoong	G03080	P. quinquefolius
P. ginseng dehydrin 1 (Dhn1)	DQ487106	742	318	632	106
P. ginseng dehydrin 2 (Dhn2)	DQ487107	892	124	248	212
P. ginseng dehydrin 3 (Dhn3)	DQ487108	221	140	168	44
P. ginseng dehydrin 4 (Dhn4)	DQ487109	258	225	225	241
P. ginseng dehydrin 5 (Dhn5)	DQ487110	5	3	4	0
P. ginseng dehydrin 6 (Dhn6)	DQ487111	0	0	5	0
P. ginseng dehydrin 7 (Dhn7)	DQ487112	735	306	627	106
P. ginseng dehydrin 8 (Dhn8)	DQ487113	0	2	0	0

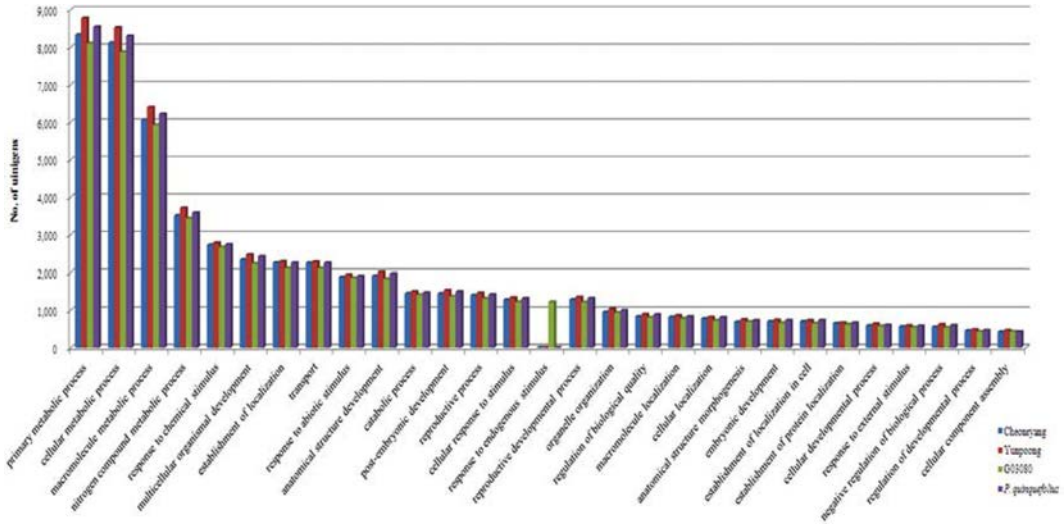

Fig. 1. A comparative biological process in GO analysis of four samples.

of the other two. The number of unigenes for the Dhn 2 gene in Cheonryang was 892, which was 47 times higher than that in the other samples. Dhn 3 gene also had the highest number of unigenes for Cheonryang but the difference with the other two samples was minimal except for *P. quinquefolius*. Dhn 4 gene expression was similar in all samples. The Dhn 5, 6, and 8 genes were small in number or there were no unigenes; therefore, there was a limited number of unigenes to compare with limited expression. Finally, the Dhn 7 gene had 537, 510, 209, and 50 unigenes in Cheonryang, G03080, Yunpoong and *P. quinquefolius*, respeEctively, whereas Cheonryang and G03080 showed relatively higher (2.5 times) expression compared to that in the other samples.

4. Comparative analysis of GO data among the four samples

The results of a GO analysis among the four samples are shown in Figures 1 ~ 3. Three main biological process category terms accounted for metabolic processes, cellular metabolic processes, and macromolecule metabolic processes. In the biological processes category, primary metabolic process was the most abundant, followed by 'cellular metabolic process', 'macromolecule metabolic process', 'nitrogen compound metabolic process' and 'response to chemical stimulus' (Fig. 1). In the molecular functions category, 'nucleotide binding' was the most abundant, followed by 'protein binding', 'ion binding' and 'nucleoside binding' (Fig. 2). A detailed GO analysis at the cellular component category, sorted all transcripts from 4 ginseng accessions into 5 groups being the most the most abundant GO term was 'cell part' followed by 'intracellular', 'intracellular part', 'intracellular organelle' and 'membrane-bounded organelle' (Fig. 3).

Among these, a higher frequency of unigenes related to the three main terms was found in Yunpoong than that in

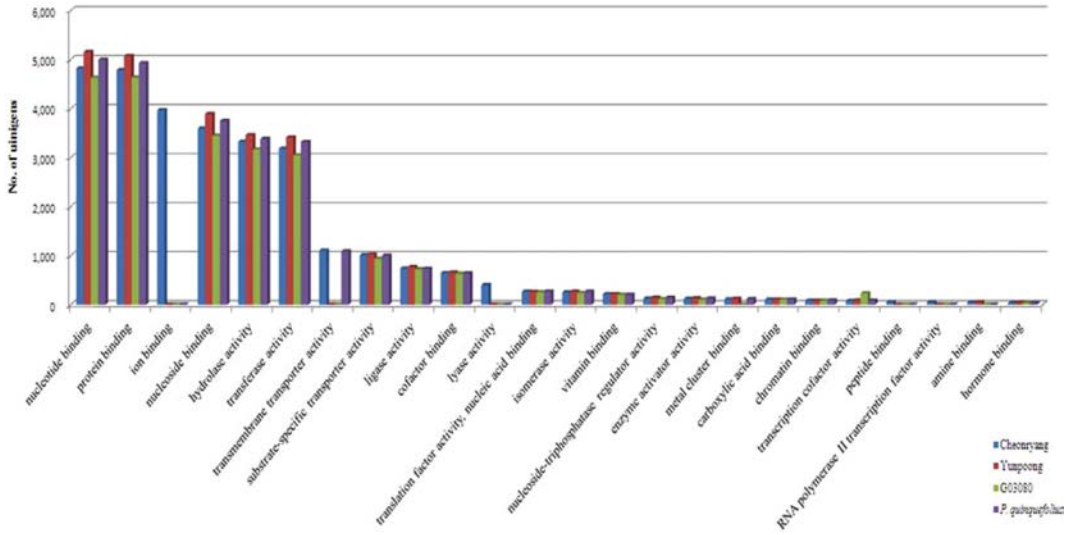

Fig. 2. A comparative molecular functions in GO analysis of four samples.

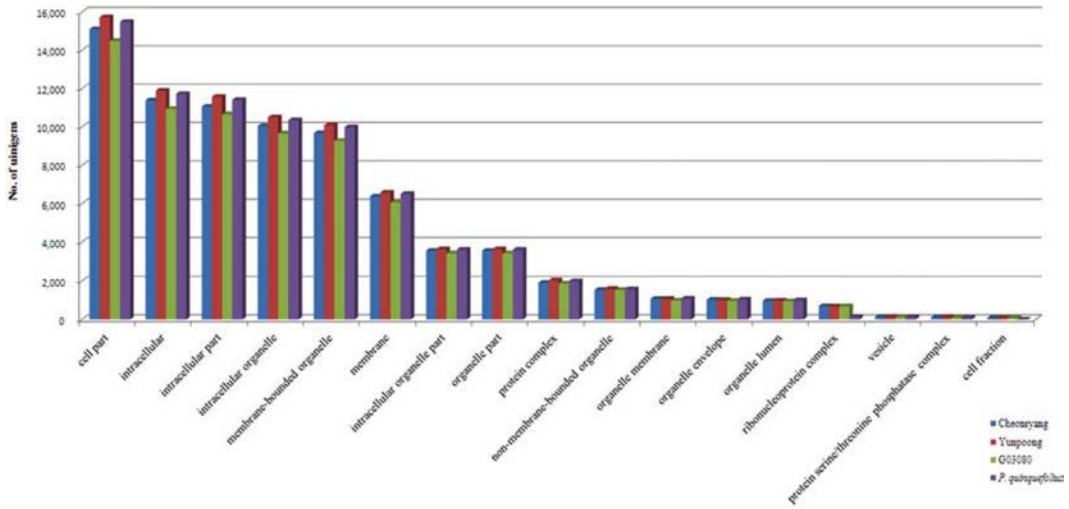

Fig. 3. A comparative cellular components in GO analysis of four samples.

the other samples. However, no differences among the four samples were observed if the rate of unigenes matched to primary metabolic processes, cellular metabolic processes, and macromolecule metabolic processes for the total number of unigenes of each sample were compared. Therefore the response to endogenous stimulus term existed only in G03080. Genes related to the response to endogenous stimulus repair damaged DNA genes was related to response to external stressors.

In the biological process category (Fig. 1), the most abundant processes were nucleotide binding and protein binding, which were not different between samples. Ion binding was only found in the Cheonryang accession (Fig. 2). The metal ions can enter the CusA pump via the periplasmic cleft as well as the cytoplasm. It is believed that CusA predominately carries out metal ion efflux through the periplasm. Specifically, charged amino acids (R83, E567, D617, E625, R669, and K678) of CusA are involved in the maintenance of homeostasis within cells and detoxification reactions (Su *et al*., 2012).

The transmembrane transporter activity related genes were identified only in the Cheonryang and *P. quinquefolius* accessions. Transmembrane transporter activity is the capability of transferring inorganic materials or ions from one side of the membrane to the other side by controlling osmotic and turgor pressure within cells (Bindra

et al., 1993). Finally, the cellular components category (Fig. 3) of GO terms showed slight differences in unigenes for the four samples, but the differences between samples were minimal if converted to a percent ratio.

DISCUSSION

We carried out a transcriptome analysis to investigate the characterization of gene expression in 4-year-old *P. ginseng* and *P. quinquefolius* roots. Based on the NGS analysis results, dehydrin genes in *P. ginseng* and matching unigenes for each sample were identified using the NCBI database, as shown in Table 5. The dehydrins are a family of late embryo abundant (LEA D-11) proteins that accumulate during the late stages of seed development or under low temperature or water deficient conditions. They are believed to play a protective role in freezing and drought tolerance. Dhn genes exist as multi-gene families in plants, and have been fairly well-characterized, particularly in barley.

Thirteen Dhn genes have thus far been identified in that species, with evidence of different expression patterns under different stress treatments and in various organs (Choi *et al*., 1999, 2000; Rodrguez *et al*., 2005). Previous studies indicated a consistent association between dehydrins and stress responses with robust representation of Dhn genes on the Affymetrix Barley1 GeneChip (Close *et al*., 2004) among about 22,000 other barley genes. This finding raised our interest in further investigating the relationships between dehydrin gene expression and global responses to drought and low-temperature stress.

The first study of Dhn genes in ginseng was conducted by Ha *et al*. (2006) who presented results of experiments by comparing the extent of expression of Dhn genes with cold stress and abscisic acid treatment. Their results confirmed that there are nine Dhn genes expressed in *P. ginseng*, among which Dhn 9 is a unique gene expressed only in seeds. In particular, the Dhn 1, 6, and 7 genes are expressed when seeds are exposed to = 5C for > 10 hours.

Our results also show that Cheonryang and G03080 had higher expression of Dhn 1 and 7 genes compared to that of the other two samples. It was speculated that the weather conditions on the day of sampling caused expression of Dhn genes in *P. ginseng*. The average temperature was 3.4℃, the maximum temperature was 11.5℃, the minimum temperature was −2.5℃, and the minimum grass temperature −3.7℃. Therefore, if Cheonryang and G03080 are exposed to cold stress for a long time, they are likely to survive in unfavorable conditions better than Yunpoong and *P. quinquefolius* due to the high expression of Dhn 1 and Dhn 7 genes. In addition, the Dhn 2 transcript levels were 47 times higher in Cheonryang than those in the other samples (Table 6). It was confirmed that the Dhn 2 gene was expressed when seeds are exposed to 5℃ for 310 hours.

Development of Cheonryang and G03080 was a major breakthrough for breeding salinity tolerant ginseng in Korea. Cheonryang and G03080 can tolerate EC 1 to 5 dS m^{-1} salinity stress during the five days (Hyun, 2012).

These results suggested that increased expression of Dhn genes in Cheonryang and G03080 not only occurred in response to salinity stress, but also in response to cold stress. From these results, we expect that Cheonryang and G03080 are resistance to cold and drought compared to that of the other two samples.

Heat shock proteins (HSPs) are a class of functionally related proteins involved in the folding and unfolding of other proteins. Their expression increases when cells are exposed to elevated temperatures or other stressors. This increase in expression is transcriptionally regulated. The dramatic upregulation of HSPs is a key part of the heat shock response and is induced primarily by the heat shock factor (Benjamin and Randy, 1998) In the cases of Yunpoong and *P. quinquefolius*, they retained more unigenes associated with HSPs compared to those of Cheonryang and G03080; thus, they are expected to be more resistant to high temperature. This result is significant in terms of nurturing species that are resistant to high temperature, and beneficial when selecting breeding material.

The GO classification separates genes involved into different biological processes, molecular functions, and cellular components. Biological processes include broad biological events, such as mitosis or purine metabolism, which are accomplished by ordered assembly of molecular functions. The molecular functions describe the task performed by individual gene products, and examples are transcription factors and DNA helicase. The cellular component classification type involves subcellular structures, locations, and macromolecular complexes, such as the nucleus, telomeres, and origin recognition complex (Dennis *et al*., 2003;

Huang *et al*., 2007). Each unigene is classified into a functional category based on the putative function played by the gene product.

The results of functional classification using GO analysis confirmed that G03080 contained unigenes that respond to external stress and repair damaged DNA. Unigenes were also found in Cheonryang and were associated with ion binding involved in the homeostasis within cells and detoxification. Finally, unigenes related to transmembrane transporter activity were observed in Cheonryang and *P. quinquefolius* that involved controlling osmotic and turgor pressure within the cell. The results of the GO analysis showed that Cheonryang and G03080 are protected against external stress. Comparative analysis of the transcriptomes among 4 ginseng accessions can lead to the identification of evolutionarily conserved and unique stress defence mechanisms.

We believe these results provide valuable information in nurturing species that are resistant to the environment. An important step in de novo assembly will be defining the gene set, and the availability of transcriptome sequencing data will greatly aid in gene prediction and validation, as well as the development of functional markers for improved ginseng cultivar breeding.

ACKNOWLEDGMENTS

This work was carried out with the support of "Co-operative Research Program for Agriculture Science & Technology Development(Project No. PJ010104)", Rural Development Administration, Republic of Korea. This study was supported by 2014 Postdoctoral Fellowship Program of National Institute of Horticultural and Herbal Science, Rural Development Administration, Republic of Korea.

REFERENCES

Benjamin I and Randy MD. (1998). Stress(Heat shock) proteins molecular chaperones in cardiovascular biology and disease. Circulation Research. 83:117-132.

Bindra PS, Knowles R and Buckley KM. (1993). Conservation of the amino acid sequence of SV2, a transmembrane transporter in synaptic vesicles and endocrine cells. Gene. 137:299-302.

Chen S, Luo H, Li Y, Sun Y, Wu Q, Niu Y, Song J, Lv A, Zhu Y, Sun C, Steinmetz A and Qian Z. (2011). 454 EST analysis detects genes putatively involved in ginsenoside biosynthesis in *Panax ginseng*. Plant Cell Reports. 30:1593-1601.

Choi DW, Zhu B and Close TJ. (1999). The barley(*Hordeum vulgare* L.) dehydrin multigene family: Sequences, allelic types, chromosome assignments, and expression characteristics of 11 Dhn genes of cv. Dicktoo. Theoretical and Applied Genetics. 98:1234-1247.

Choi DW and Close TJ. (2000). A newly identified barley gene, Dhn12 encoding a YSK2 DHN, is located on chromosome 6H and has embryo-specific expression. Theoretical and Applied Genetics. 100:1274-1278.

Close TJ, Wanamaker SI, Caldo RA, Turner SM, Ashlock DA, Dickerson JA, Wing RA, Muehlbauer GJ, Kleinhofs A and Wise RP. (2004). A new resource for cereal genomics: 22K barley genechip comes of age. Plant Physiology. 134:960-968.

Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC and Lempicki RA. (2003). DAVID: Database for annotation, visualization, and integrated discovery. Genome Biology. 4: R60.

Egan AN, Schlueter J and Spooner DM. (2012). Applications of next-generation sequencing in plant biology. American Journal of Botany. 99:175-785.

Ellegren H. (2008). Comparative genomics and the study of evolution by natural selection. Molecular Ecology. 17:4586-4596.

Gilles A, Meglcz E, Pech N, Ferreira S, Malausa T and Martin JF. (2011). Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. BMC Genomics. 12:245.

Ha YI, Lim JM, Ko SM, Liu JR and Choi DW. (2006). Sequence variability and expression characteristics of the ginseng(*Panax ginseng* C. A. Meyer) dehydrin gene family. Journal of Plant Biology. 49:205-211.

Heng L and Homer N. (2010). A survey of sequence alignment algorithms for next-generation sequencing. Brefings in Bioinformatics. 2:473-483.

Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC and Lempicki RA. (2007). The DAVID gene functional clas-sification tool: A novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biology. 8:R183.

Hyun DY. (2012). Screening method for salinity tolerance in *Panax ginseng* C. A. Meyer. Korea. Patent. 10-1128911.

Kim OT, Yoo NH, Kim GS, Kim YC, Bang KH, Hyun DY, Kim SH and Kim MY. (2013) Stimulation of Rg3 ginsenoside biosynthesis in ginseng hairy roots elicited by methyl jasmonate. Plant Cell, Tissue and Organ Culture. 112:87-93.

Li C, ZhuY, GuoX, Sun C, Luo H, Song J, Li Y, Wang L, Qian J and Chen S. (2013). Transcriptome analysis reveals ginsenosides biosynthetic genes, microRNAs and simple sequence repeats in *Panax ginseng* C. A. Meyer. BMC Genomics. 14:245.

Margulies M, Egholm M, Altman WE, Attiya S, Joel S. Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth

GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF and Rothberg JM. (2005). Genome sequencing in micro fabricated high-density picolitre reactors. Nature. 437:376-380.

Metzker ML. (2009). Sequencing technologies-the next generation. Nature Reviews Genetics. 11:31-46.

Meyer M, Stenzel U and Hofreiter M. (2008). Parallel tagged sequencing on the 454 platform. Nature Protocols. 3:267-278.

Park HW, Kim OT, Hyun DY, Kim YB, Kim JU, Kim YC, Bang KH, Cha SW and Choi JE. (2013). Overexpression of farnesyl diphosphate synthase by introducing *CaFPS* gene in *Panax ginseng* C. A. Mey. Korean Journal of Medicinal Crop Science. 21:32-38.

Rodrguez EM, Svensson JT, Malatrasi M, Choi DW and Close TJ. (2005). Barley Dhn13 encodes a KS-type dehydrin with constitutive and stress responsive expression. Theoretical and Applied Genetics. 110:852-858.

Su CC, Long F, Lei HT, Bolla JR, Do SV, Rajashankar and Yu EW. (2012). Charged amino acids(R83, E567, D617, E625, R669, and K678) of CusA are required for metal ion transport in the Cus efflux system. Journal of Molecular Biology. 422:429-441.

Sun C, Li Y, Wu Q, Luo H, Sun Y, Song J, Lui EMK and Chen S. (2010) De novo sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. BMC Genomics. 11:262.

Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I and Marden JH. (2008). Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Molecular Ecology. 17:1636-1647.

Wang Z, Gerstein M and Snyder M. (2009). Innovation: RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics. 10:57-63.

Wu B, Wang M, Ma Y, Yuan L and Lu S. (2012). High-throughput sequencing and characterization of the small RNA transcriptome reveal features of Novel and Conserved MicroRNAs in *Panax ginseng*. PLoS ONE 7:e44385.

Wu D, Austin RS, Zhou S and Brown D. (2013). The root transcriptome for North American ginseng assembled and profiled across seasonal development. BMC Genomics. 14:564.