

Research article

Transcriptome profiling and comparative analysis of *Panax ginseng* adventitious roots



Murukarthick Jayakodi^{1,☆}, Sang-Choon Lee^{1,☆}, Hyun-Seung Park¹, Woojong Jang¹, Yun Sun Lee¹, Beom-Soon Choi², Gyoung Ju Nah¹, Do-Soon Kim¹, Senthil Natesan³, Chao Sun⁴, Tae-Jin Yang^{1,*}

¹ Department of Plant Science, Plant Genomics and Breeding Institute, and Research Institute for Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul, Korea

² Phyzen Genomics Institute, Seoul, Korea

³ Genomics and Proteomics Laboratory, Centre for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India

⁴ Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

ARTICLE INFO

Article history:

Received 21 March 2014

Received in Revised form

21 May 2014

Accepted 22 May 2014

Available online 5 June 2014

Keywords:

adventitious root

de novo assembly

next-generation sequencing

Panax ginseng

transcriptome

ABSTRACT

Background: *Panax ginseng* Meyer is a traditional medicinal plant famous for its strong therapeutic effects and serves as an important herbal medicine. To understand and manipulate genes involved in secondary metabolic pathways including ginsenosides, transcriptome profiling of *P. ginseng* is essential. **Methods:** RNA-seq analysis of adventitious roots of two *P. ginseng* cultivars, Chunpoong (CP) and Cheongsun (CS), was performed using the Illumina HiSeq platform. After transcripts were assembled, expression profiling was performed.

Results: Assemblies were generated from ~85 million and ~77 million high-quality reads from CP and CS cultivars, respectively. A total of 35,527 and 27,716 transcripts were obtained from the CP and CS assemblies, respectively. Annotation of the transcriptomes showed that approximately 90% of the transcripts had significant matches in public databases. We identified several candidate genes involved in ginsenoside biosynthesis. In addition, a large number of transcripts (17%) with different gene ontology designations were uniquely detected in adventitious roots compared to normal ginseng roots.

Conclusion: This study will provide a comprehensive insight into the transcriptome of ginseng adventitious roots, and a way for successful transcriptome analysis and profiling of resource plants with less genomic information. The transcriptome profiling data generated in this study are available in our newly created adventitious root transcriptome database (<http://im-crop.snu.ac.kr/transdb/index.php>) for public use.

Copyright © 2014, The Korean Society of Ginseng, Published by Elsevier. All rights reserved.

1. Introduction

Panax ginseng Meyer (ginseng) has widely been used as a source of medicine in eastern Asia and North America [1,2]. Ginseng serves as an adaptogen, with effects on immune system stimulation, anticancer activity, and antihyperlipidemic effects [3–6]. To date, 10 varieties have been bred by the Korea Ginseng Corporation and

registered as commercial cultivars with the Korea Seed and Variety Service (<http://www.seed.go.kr>). Among those, Chunpoong (CP) is a very pure inbred line with relatively low heterozygosity, high yield, and superior quality [7]. Genetic study on ginseng has been challenging because of its long generation time, the small numbers of seeds it sets, and the difficulty of maintaining ginseng in the field. As an alternative, various tissue culture methods, including callus,

* Corresponding author. Department of Plant Science, Plant Genomics and Breeding Institute, College of Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Korea.

E-mail address: tjyang@snu.ac.kr (T.-J. Yang).

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

☆ These authors contributed equally to this work.

hairy root, and adventitious root culture systems, have been adapted for mass production of ginseng. Among these, adventitious root culture has been a promising alternative for production of ginsenoside, because the total saponin contents of the adventitious roots are comparable to those of field-grown roots and higher than those of callus and hairy roots [8]. Moreover, mass production of adventitious roots is well established through a balloon-type bubble bioreactor system [8].

To date, for *P. ginseng*, a few expressed sequence tag (EST) libraries have been generated and 17,114 EST sequences have been deposited in the dbEST database at the National Center for Biotechnology Information (NCBI). Most of the ESTs have been generated with the aim of identifying genes involved in ginsenoside biosynthesis and developing molecular markers [9–12]. Although transcriptome sequencing and assembling of plants with large and complex genomes such as *P. ginseng* are still difficult, next-generation sequencing (NGS) technologies made it affordable to sequence cDNA (RNA-Seq) and examine cellular transcriptomes along with high-throughput gene expression analysis [13]. To date, a few studies have applied NGS technology to transcriptome analysis of *Panax* species, including *Panax notoginseng*, *Panax quinquefolius*, and *P. ginseng* [14–16]. These studies used the 454 sequencing platform mainly to identify ginsenoside biosynthetic genes in the normal root transcriptome. Nevertheless, gene discovery and expression profiling in ginseng are still very limited.

Here, we used the Illumina sequencing platform for large-scale transcriptome analysis, and present *de novo* adventitious root transcriptome assemblies for CP, which is the oldest elite cultivar in Korea, and Cheongsun (CS), which is a superior cultivar for adventitious root production. We assembled CP and CS transcriptomes from millions of short sequence reads generated by Illumina paired-end transcriptome sequencing. After annotation, we conducted gene expression profiling, as well as identification of candidate genes involved in ginsenoside biosynthesis. This work provides the first transcriptome profiles of *in vitro*-grown adventitious roots of two ginseng cultivars. It also describes an advanced method for transcriptome assembly and validation in nonmodel plant species and for the study of genes related to secondary

metabolites, which can be affected greatly by small modifications in environmental conditions.

2. Materials and methods

2.1. Plant material and RNA isolation

Stratified seeds of Korean ginseng cultivars CP and CS from Korea Ginseng Corporation (Daejeon, Korea) were sterilized by immersion in 70% ethanol for 1 minute and 2% sodium hypochlorite for 10 minutes. After each step, seeds were gently washed with distilled water three times. All procedures were performed aseptically in a laminar hood. To induce adventitious roots, cotyledons separated from sterilized stratified seeds were cultured on a solid Schenk and Hildebrandt (SH) medium containing 2.0 mg/L indole butyric acid, 3% sucrose, and 0.23% Gelrite. After 1 month, induced adventitious roots were separated from cotyledon explants and cultured again for secondary growth on the same medium. Then, the roots were transferred to a 30 mL liquid SH medium supplemented with 3.0 mg/L indole butyric acid and 5% sucrose, and maintained on a rotary shaker (100 rpm) at 25°C in the dark. For further mass production, 12 g fresh adventitious roots in suspension culture were inoculated into a 2 L airlift balloon-type bioreactor (Biopia, Korea) containing 1 L of the same SH medium as that used for liquid suspension culture (Fig. 1). The medium was replaced with a fresh medium after 2 weeks, and 4 weeks later, 12 g adventitious roots were subcultured into a new bioreactor. After 10 days of cultivation, the subcultured adventitious roots were used for total RNA extraction with the Plant RNeasy mini kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. Approximately 2 µg total RNA from each cultivar was used for sequencing on the Illumina platform after the quality and quantity were checked using spectrophotometry.

2.2. Illumina sequencing and quality control

Paired-end reads with an average length of 101 bp were generated for CP and CS using the Illumina HiSeq2000 platform. Library construction and sequencing were performed by the

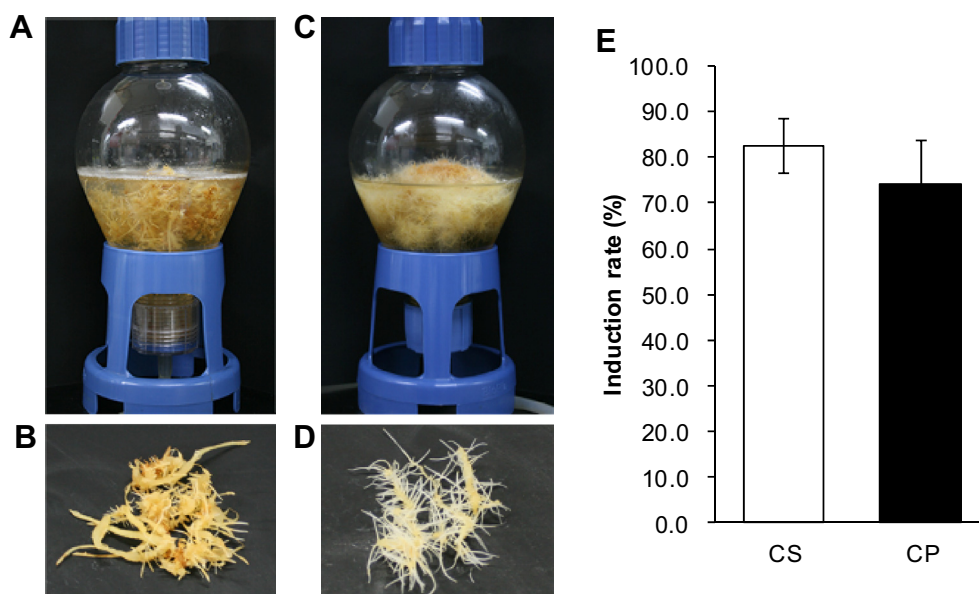


Fig. 1. Adventitious roots of CP and CS at 2 weeks after inoculation in 2 L BTBBs. (A, B) Adventitious roots of CP are dark yellow, with callus-like morphology. (C, D) whereas those of CS show bright yellow, root hair-like morphology. (E) The rate of adventitious root induction from cotyledons. Data are represented as means with standard deviation from four independent experiments. BTBB, balloon-type bubble bioreactor; CP, Chunpoong; CS, Cheongsun.

National Instrumentation Center and Environmental Management (NICEM), Seoul National University, Seoul, South Korea. The sequence data generated in this study have been deposited in the Short Read Archive (SRA) of the NCBI under the accession number SRA061905. The sequencing reads underwent various stringent quality controls, such as filtering of high-quality reads and removal of reads with an adaptor or primer-contaminated sequence using the NGS QC toolkit [17].

2.3. De novo assembly

All *de novo* assemblies were performed on a server with 48 cores and 512 GB random access memory. Publicly available transcriptome and genome assemblers were used to assemble the paired-end reads. Among the transcriptome assemblers, the open source program, Oases [18] (version: 0.2.06; <http://www.ebi.ac.uk/~zerbino/oases>), which uploads a preliminary assembly produced by Velvet, was validated for *k*-mer optimization. Various assembly parameters were also examined to yield statistically as well as biologically significant results. In addition, other publicly available transcriptome assemblers were used to determine the best assembler for the CP data set, which included Trinity [19] (package: trinityrnaseq_r2012-04-27; <http://trinityrnaseq.sourceforge.net>) and SOAPdenovo-Trans [20] (version: 1.01; <http://soap.genomics.org.cn/SOAPdenovo-Trans.html>); genome assemblers were also used for *de novo* transcriptome assembly, such as ABySS [21] (version: 1.3.3; <http://www.bcgsc.ca/platform/bioinfo/software/abyss>) and commercially available CLC Genomics Workbench (version 5.1; CLCbio, Denmark). The data for CS cultivar were assembled using the assembler that was identified as the best from the CP cultivar assembly. Transcriptome profiling data generated in this study are publicly accessible through our adventitious root transcriptome database (<http://im-crop.snu.ac.kr/transdb/index.php>).

2.4. Functional annotation and analysis

The assembled CP and CS transcript sequences were annotated by sequence comparison with well-annotated protein databases. All assembled transcripts were searched against the NCBI non-redundant protein (nr) database (<ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz>) using BLASTX with an *E*-value cutoff of $1E-05$. In addition, CP and CS transcripts were searched against the Uniprot (TrEMBL and SwissProt; ftp://ftp.expasy.org/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot.fasta.gz) and TAIR (The Arabidopsis Information Resource; ftp://ftp.arabidopsis.org/home/tair/Proteins/TAIR10_protein_lists/TAIR10_pep_20101214) databases using the BLASTX search with cutoff *E*-values of $1E-05$ and $1E-10$, respectively. Transcripts were functionally classified following the gene ontology (GO) scheme (<http://www.geneontology.org>). The Blast2GO program [22] was used to determine the molecular function, biological process, and cellular component categories associated with the best BLASTX hit in the nr database for the corresponding CP and CS transcripts.

2.5. Expression profiling

Trimmed raw reads were mapped onto their assembled transcripts to quantify transcript abundance using the CLC Genomics Workbench (version 5.1). The number of reads and reads per million were determined using the CLC mapping program. Further, reads per kilobase per million (RPKM) for each transcript and average RPKM were determined [23]. In addition, expression of transcripts related to ginsenoside biosynthesis was determined by mapping reads of CP and CS on CP transcripts as references.

2.6. Identification of candidate transcripts involved in ginsenoside biosynthesis

P. ginseng gene sequences that were reported to be involved in the biosynthesis of ginsenosides were collected from GenBank. The amino acid sequences of these genes were used as queries to search for homologous sequences in the CP and CS assembled transcript datasets using the TBLASTN program. Candidate transcripts were identified based on *E*-value, bit score, alignment length, and further validation using BLASTP.

3. Results

3.1. Adventitious root growth and morphology of two ginseng cultivars

We obtained adventitious roots from the cotyledons of CP and CS cultivars. Although the same culture conditions were used for both cultivars, they showed different adventitious root morphology during proliferation in bioreactor culture. Adventitious roots of CP appeared to be dark-yellow, callus-like clumps (Fig. 1B), whereas those of CS showed typical adventitious root morphology and were bright yellow (Fig. 1D). In addition, the adventitious root induction rate of CS was 9% higher than that of CP (Fig. 1E). This indicates that CS is better suited for adventitious root induction and growth under these conditions.

3.2. De novo assembly and validation of Illumina paired-end sequences

We generated a total of 90,242,024 and 82,011,294 raw reads from CP and CS, respectively (Table 1). After trimming the low-quality reads with Phred quality scores of ≤ 25 and removing primer/adaptor sequences, we obtained 85,335,736 (94.5%) and 77,583,736 (94.6%) high-quality reads, with an average read length of 99 bp, in CP and CS, respectively (Table 1). To obtain high-quality assemblies, we tested several algorithms for *de novo* assembly with different options. We used several criteria to determine the desirable assembly: number of reads used in the assembly, total length of transcriptome, average contig length, N50, and annotation by BLASTX against the TAIR protein database.

Using Velvet followed by Oases, we compared assembly results with randomly selected *k*-mer lengths of 31, 39, 41, 49, 51, 59, 61, 69, 71, and 79. The best assembly was obtained at *k* = 69, as it resulted in the highest total length (~ 138 Mbp), the largest N50 length (1,092 bp), the largest average contig length (19,999 bp), and a

Table 1
Summary statistics of CP and CS sequencing, assembly, and validation

Illumina sequencing data	CP	CS
Number of raw reads	90,242,024	82,011,294
Number of trimmed reads	85,335,736	77,583,736
Trinity assembly		
Total number of transcripts	35,527	27,716
Total transcriptome size (bp)	70,295,564	54,892,571
Non-redundant set (without isoforms)	14,051	11,209
Small transcript length (bp)	410	411
Large transcript length (bp)	15,918	15,980
Average transcript length (bp)	1,978	1,980
N50 length (bp)	2,274	2,277
Validation by BLASTX similarity searches (<i>E</i> -value cutoff of $1E-05$)		
Nr (NCBI)	33,718 (94%)	26,513 (95%)
TAIR (<i>Arabidopsis</i>)	32,996 (92%)	25,996 (93%)
SwissProt	27,745 (78%)	21,923 (79%)

CP, Chunpoong; CS, Cheongsun; NCBI, National Center for Biotechnology Information; Nr, nonredundant protein; TAIR, The Arabidopsis Information Resource

significant number of TAIR hits (74.79%). In addition to Oases assembly, we also used Trinity ($k = 25$ as a fixed option), SOAP-Trans, ABySS, and the CLC Genomics Workbench with default parameters. We also compared the assembly results by mapping all raw reads onto each assembly, in order to determine the read usage. We obtained the best assembly results from Oases and Trinity, as they showed the largest assembled transcriptome sizes, numbers of mapped reads, average contig lengths, and numbers of TAIR hits (BLASTX; data not shown).

For further evaluation of the accuracy of the datasets, we compared both against *P. ginseng* full-length gene sequences retrieved from GenBank. Large numbers of full-length sequences (including untranslated regions) were found in the Trinity dataset, with 95–100% identity. We found that many truncated transcripts (without the start and stop codons) were included in the Oases dataset. The extracted dataset sequences were also mapped successfully onto our ongoing *P. ginseng* draft genome sequence assembly using the BLAST algorithm. The Trinity dataset showed more hits and a higher percentage of identity than the Oases dataset, demonstrating that Trinity was the best assembler for our transcriptome assembly.

Using Trinity, we obtained 35,527 CP transcripts with an average length of 1,978 bp and 27,716 CS transcripts with an average length of 1,980 bp (Table 1). The lengths of the assembled transcripts ranged from 400 bp to 15,980 bp, with a large number of transcripts in the range of 1,000–2,000 bp in CP as well as in CS. We identified sets of 14,051 and 11,209 nonredundant transcripts from the CP and CS assembled datasets, respectively, by selecting only the longest sequence among isoforms that included alternatively spliced forms predicted by the Trinity assembler. We used the total assembled transcripts including isoforms for further analysis, because it was difficult to select the optimal representative nr dataset among various isoforms without a *P. ginseng* reference sequence.

3.3. Functional annotation and classification

For further validation and annotation of assembled transcripts, sequence similarity searches were conducted against the TAIR and Uniprot (SwissProt and TrEMBL) protein databases using the BLASTX algorithm. The significant hits were identified based on an E -value threshold of 10^{-5} . Impressively, the results indicated that more than 90% CP and CS transcripts showed significant similarity to proteins in the TAIR database. In addition, approximately 70% of CP and CS transcripts had significant matches among Uniprot proteins (Table 1). We also compared the CP and CS transcripts against the NCBI nr protein database using BLASTX, finding that a total of 33,718 (94%) CP transcripts and 26,513 (95%) CS transcripts had significant hits.

To classify the predicted functions of the transcripts, GO terms were assigned to CP and CS transcripts using Blast2GO, based on their similarity to the nr database. A total of 26,423 (74.37%) CP transcripts were assigned to GO classes. Of those, assignments to the cellular component class ranked the highest (22,706; 63.91%), followed by biological process (22,215; 62.53%) and molecular function (21,560, 60.68%). In CS, a total of 21,096 (76.11%) transcripts were assigned at least one GO term, and among them, 17,512 (63.18%), 17,249 (62.23%), and 18,178 (65.58%) were assigned at least one GO term in the biological process, molecular function, and cellular component category, respectively. Binding was the most abundant GO Slim within the molecular function category (Fig. 2A). Reproductive development, cellular process, and stress response were most abundant among various biological processes (Fig. 2B). Intracellular membrane-bound organelle and membrane were the most highly represented GO terms in the cellular component category (Fig. 2C).

3.4. Gene expression profiling

We mapped all the CP and CS reads onto their respective assembled transcripts in order to determine the RPKM. For the CP transcripts, the RPKM ranged from 0.16 to 4609, with an average of 15.93, and the RPKM for CS ranged from 0.22 to 4118, with an average of 19.90. This indicates that both CP and CS transcripts showed a wide range of expression levels (from very low to strong expression). However, over 97% of transcripts were in the RPKM range less than 100 (Fig. 3A), of which 1,244 (3.5%) and 585 (2.1%) had RPKM values below 1.0 for CP and CS, respectively. We compared the expression patterns of the transcripts in CP and CS cultivars based on the differences in RPKM value for each transcript in the cultivar datasets. As evident in Fig. 3B, 3C, the transcripts showed very similar expression patterns in both cultivars, which was also confirmed by the high correlation coefficient value of more than 0.9 between both datasets.

To identify highly expressed transcripts and their putative functions, we selected the 100 most abundant transcripts based on their RPKM values in the CP and CS datasets, and investigated the biological processes in which those transcripts might be involved. Although many transcripts (15 in CP and 23 in CS) could not be assigned to known biological processes, most (52 in CP and 51 in CS) were involved in stress response and protein metabolism, including pathogenesis-related proteins, antioxidant enzymes, heat-shock proteins, and metallothionein-like proteins in the stress response category, and translation- and protein degradation-related proteins in the protein metabolism category (Fig. 4). After these, transcripts related to lipid metabolism, such as fatty acid desaturases and lipid transfer proteins, were most abundant.

3.5. Identification of candidate genes involved in ginsenoside biosynthesis

Ginsenosides are the most important phytochemicals in ginseng and are known to be synthesized through the mevalonic acid pathway [24]. We focused on downstream enzymes from farnesyl diphosphate synthase (FDS) to UDP-glycosyltransferase (UGT) in the mevalonic acid pathway (Fig. 5A). In previous studies, 17 genes for the seven downstream enzymes (FDS to protopanaxatriol synthase) have been reported in *P. ginseng* [25–32] (Table 2). We used amino acid sequences of the 17 genes as queries for TBLASTN searches against transcript datasets of the CP cultivar, resulting in the identification of 10 genes encoding the seven downstream enzymes. Of them, a single transcript for FDS was identified with 15 isoforms in the CP dataset (Table 2). Squalene synthase, dammar-enediol synthase, β -amyirin synthase, protopanaxadiol synthase (CYP716A47), and protopanaxatriol synthase (CYP716A53v2) were also identified to be encoded by single transcripts with several isoforms. Exceptionally, four transcripts were identified for squalene epoxidase. Although we identified the isoforms using a reliable algorithm (Trinity assembler), the forthcoming *P. ginseng* genome sequence will provide more solid information about them. Based on our analysis, we considered that the isoforms are likely to originate from a single gene.

To investigate the expression levels of the transcripts, the RPKM values of isoforms from the same transcripts were averaged and compared (Fig. 5B). All showed similar expression levels between CP and CS cultivars, with transcripts encoding cytochrome P450 for protopanaxatriol synthase showing the highest expression in both cultivars.

Three UGT proteins, SvUGT74M1, MtUGT73K1, and MtUGT71G1, were used as queries for TBLASTN searches, because UGT genes for ginsenoside biosynthesis had not been identified in *P. ginseng*. Three UGT proteins were reported to function in triterpene saponin

biosynthesis in *Medicago truncatula* and *Saponaria vaccaria* [33,34]. Among the transcript hits, a total of 42 isoforms were selected based on their BLAST bit score of more than 200 (Table 3). Of those, 21 transcripts remained after removing redundant isoforms on the basis of their similarity at the amino acid level. Phylogenetic analysis of the deduced protein sequences of the 21 transcripts revealed

that some transcripts were closely grouped with the three reported UGT proteins (Fig. 6A). In particular, three transcripts, CP_comp126017, CP_comp142900, and CP_comp82124, showed much higher similarity to the reported UGT proteins than the other transcripts. Overall, the expression patterns of the 21 transcripts were similar between CP and CS, with the exception of

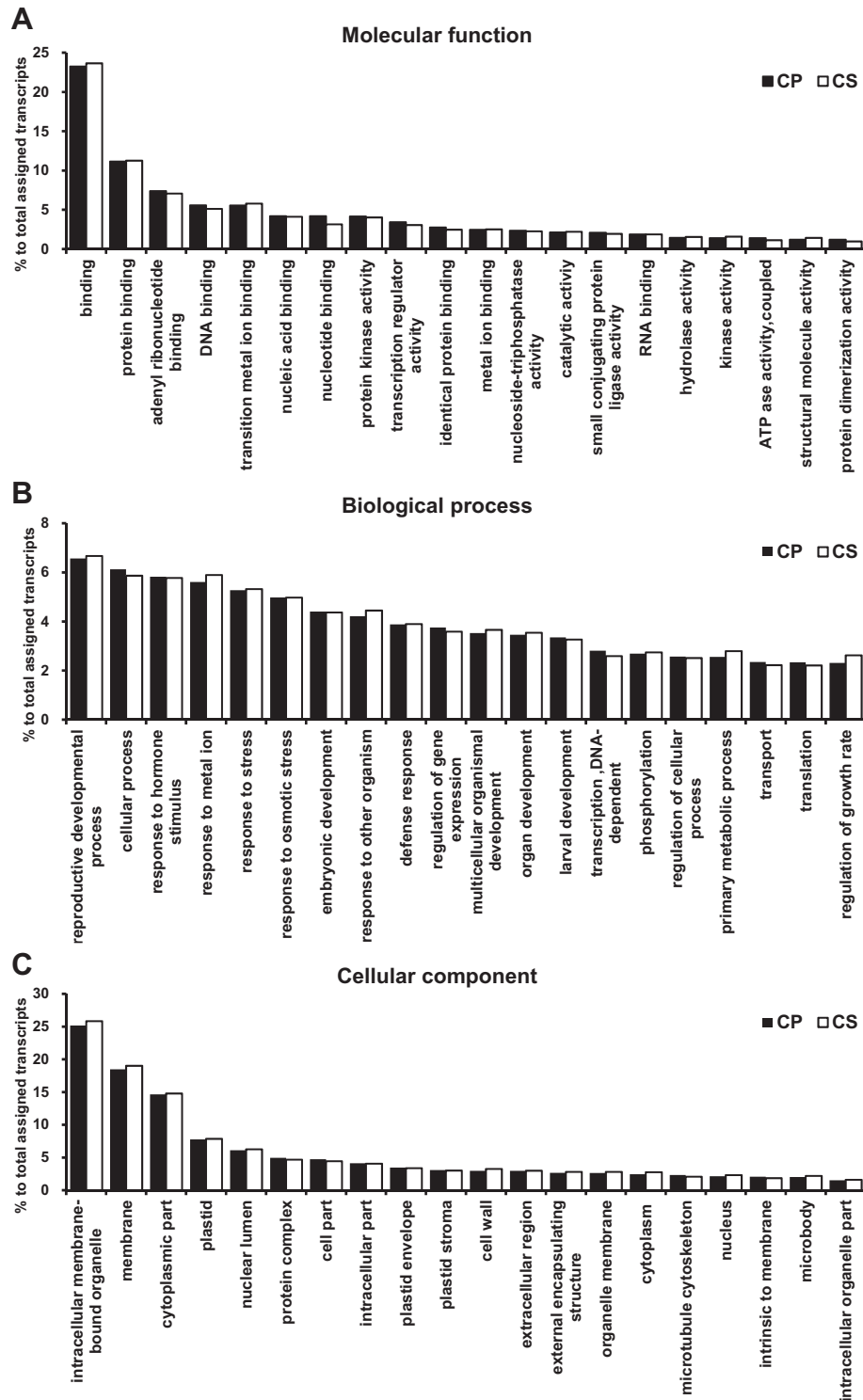


Fig. 2. GO analysis of transcripts expressed in CP and CS adventitious roots. Terms in the (A) molecular function, (B) biological process, and (C) cellular component categories are shown. A total of 26,423 CP and 21,096 CS transcripts were assigned to GO terms using Blast2GO with whole assembled transcripts of CP (35,527) and CS (27,716). CP, Chunpoong; CS, Cheongsun; GO, gene ontology.

CP_comp144124, which showed about two-fold higher expression in CP (2.5 in CP vs. 1.3 in CS; Fig. 6B).

3.6. Comparative analysis of the transcriptomes of adventitious and normal roots

To investigate the transcript expression differences between adventitious roots and primary roots, we compared 35,527 CP reference transcripts with 38,966 transcripts from 11-year-old ginseng primary roots, after assembly of 454 reads from the NCBI SRA database (accession no. SRX017443) [35]. When their sequence similarity was analyzed, 6,057 (17.0%) transcripts in adventitious roots and 6,354 (16.3%) in primary roots were found to be uniquely expressed. A total of 62,082 transcripts, 29,470 (83.0%) from adventitious roots and 32,612 (83.7%) from primary roots, were commonly expressed. GO analysis of unique transcripts was performed to characterize their functional category. As shown in Fig. 7, more transcripts from adventitious roots were assigned GO terms than from normal roots. Overall, the proportion of GO assignment in adventitious root transcriptomes was two-fold higher than that in normal roots, although the most frequent GO terms such as binding, response to other organisms, and nuclear lumen were

generally similar between both datasets. In particular, 11 out of 20 GO terms for biological processes had more transcripts in adventitious roots than in normal roots. Terms such as response to metal ion, transcription, multicellular organismal development, and reproductive developmental process showed more than eight-fold higher proportions compared to those in normal roots. By contrast, only two biological process terms, regulation of growth rate and response to stress, accounted for higher proportions in normal roots than in adventitious roots.

4. Discussion

4.1. Assembly of Illumina transcriptome sequences

Transcriptome profiling using NGS technology, the so-called RNA-Seq, is one of the most efficient tools for gene discovery and various functional studies. Illumina transcriptome sequencing and assembly have been used successfully for several nonmodel organisms [36–39], but transcriptome assembly has many challenges, including misassembled or chimeric contigs (i.e., assembled contigs containing reads from different transcripts [40]). Here, we describe a method to choose the best assembly result for both

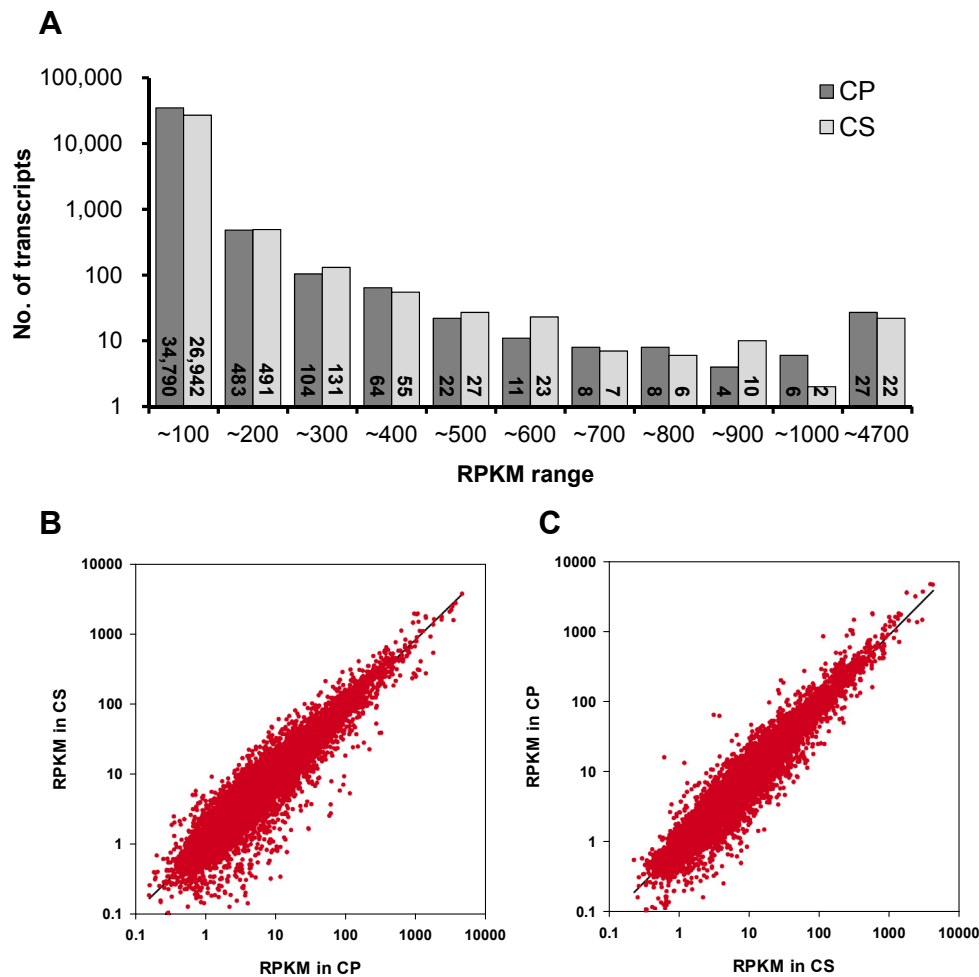


Fig. 3. RPKM values of transcripts expressed in CP and CS adventitious roots. RPKM values were calculated after reads of both cultivars were mapped onto the respective assembled transcripts. (A) Distribution of RPKM values. The number of transcripts belonging to the RPKM range defined on the X axis is shown at the bottom of each bar. Transcripts showing RPKM values of < 1.0 were included. (B) RPKM value comparison calculated by mapping reads of both cultivars onto CP transcripts as references. (C) RPKM value comparison calculated by mapping reads of both cultivars onto CS transcripts as references. Red spots indicate transcripts expressed in both cultivars. Regression lines [$r^2 = 0.92$ in (B) and $r^2 = 0.94$ in (C) at 99% confidential level] were calculated and depicted using Sigmaplot s/w. X and Y axes are in log-scale. CP, Chunpoong; CS, Cheongsun; RPKM, reads per kilobase per million.

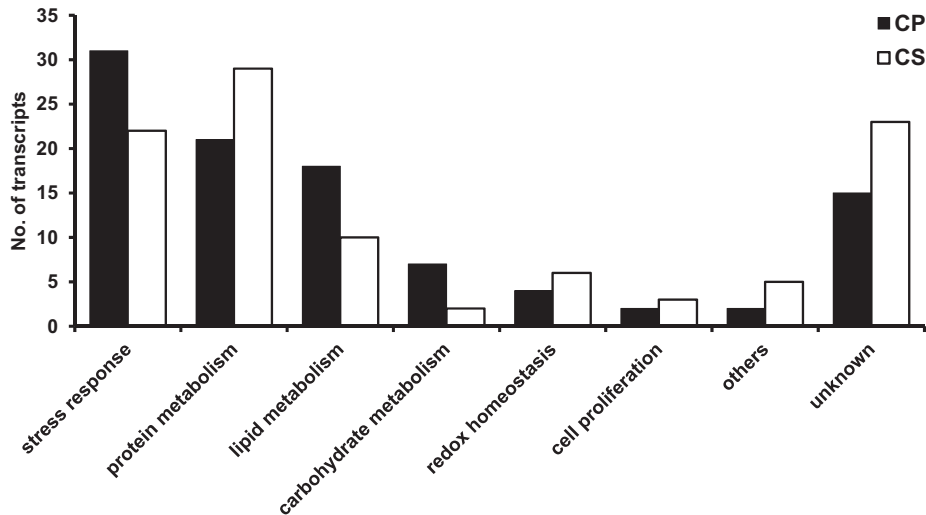


Fig. 4. Functional distribution of the 100 most highly expressed transcripts in CP and CS adventitious roots. The top 100 transcripts were chosen on the basis of their expression level determined by RPKM values, and then biological processes in which those transcripts are putatively involved were predicted based on GO assignment and BLASTX searches. CP, Chunpoong; CS, Cheongsun; GO, gene ontology; RPKM, reads per kilobase per million.

biologically and computationally meaningful results. We used metrics such as reads used in assembly, and average length and number of annotated proteome hits (TAIR) as indicators of assembly quality.

Our results show that the quality of a *de novo* transcriptome assembly is not highly dependent on the user-defined single or multiple *k*-mer length, because we found the best assembly in Oases at *k*-mer 69 (user defined: substantially higher length) and in Trinity at *k*-mer 25 (fixed: lower length), based on the assembly

indicators. Due to algorithmic differences between these two assemblers, they produced almost the same quantity of contigs with different proportions of accuracy. Thus, validation of *de novo* transcriptome assembly is highly challenging, and there is no standard method or criteria to identify misassembly or chimeric assembly. Through analyses using full-length *P. ginseng* genes retrieved from GenBank and our draft genome sequence, we found that the accuracy was greater in the transcripts assembled with Trinity compared to those with Oases. Our BLASTX annotation against the

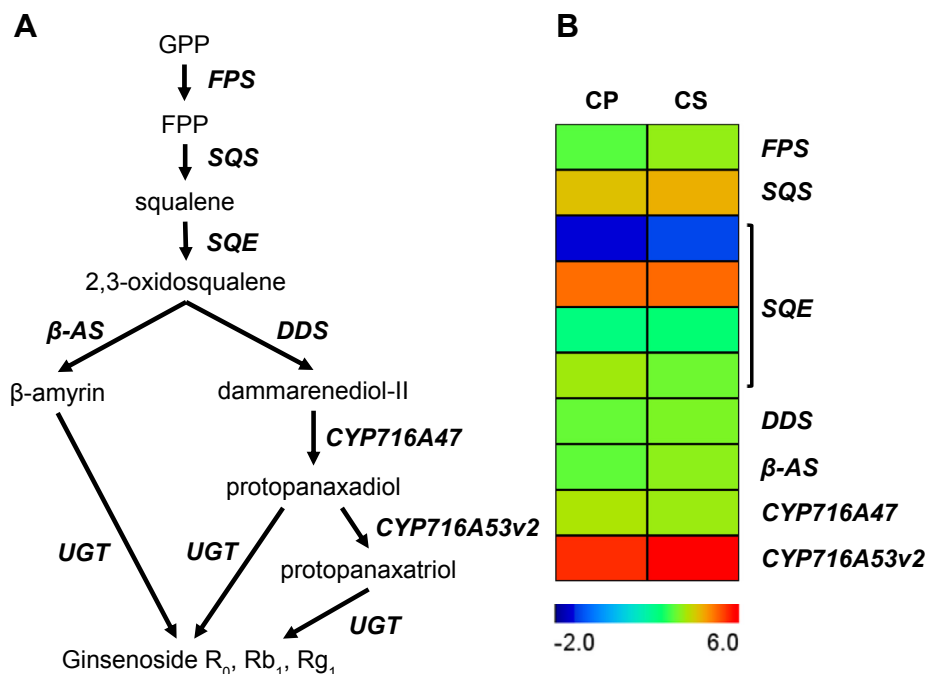


Fig. 5. Putative ginsenoside biosynthesis pathway and expression level of ginsenoside biosynthesis genes found in transcript datasets of CP and CS adventitious roots. (A) Candidate genes identified in this study, which are shown in bold. (B) Expression levels of the candidate genes in both cultivars. Several isoforms of the candidate genes were identified, and RPKM values were averaged and represented with the color scale shown at the bottom. The expression map was generated using MeV s/w (<http://www.tm4.org/mev/>) with \log_2 (RPKM) values. β -AS, beta-amyrin synthase; CP, Chunpoong; CS, Cheongsun; CYP716A47, cytochrome P450 for protopanaxadiol synthase; CYP716A53v2, cytochrome P450 for protopanaxatriol synthase; DDS, dammarenediol synthase; FPP, farnesyl diphosphate; FPS, farnesyl diphosphate synthase; GPP, geranyl diphosphate; RPKM, reads per kilobase per million; SQE, squalene epoxidase; SQS, squalene synthase; UGT, UDP-glycosyltransferase.

Table 2
List of candidate transcripts involved in ginsenoside biosynthesis

Gene name	Used query (acc. no.)	Transcripts (no. of isoforms) ¹⁾	Average RPKM (SD) ²⁾	
			In CP	In CS
Farnesyl diphosphate synthase	PgFDS (AAY87903)	CP_comp143639 (15)	9.0 (± 5.3)	11.8 (± 6.8)
Squalene synthase	PgSS1 (AB115496)	CP_comp140105 (3)	18.6 (± 1.0)	21.2 (± 0.7)
	PgSS2 (GQ468527)			
	PgSS3 (GU183406)			
Squalene epoxidase	PgSQE1 (AB122078)	CP_comp86463 (0)	0.4 (n.a.)	0.5 (n.a.)
	PgSQE2 (FJ393274)	CP_comp129934 (0)	32.4 (n.a.)	33.1 (n.a.)
	PgSE (AB003516)	CP_comp136313 (0)	5.5 (n.a.)	6.1 (n.a.)
		CP_comp137048 (3)	12.6 (± 5.6)	10.0 (± 2.2)
Dammarenediol synthase	PgPNA (AB265170)	CP_comp138594 (3)	9.5 (± 1.6)	10.6 (± 3.1)
	PgDDS (JN596111)			
	PgDDS (GU183405)			
β-amyrin synthase	PgOSCPNY1 (AB009030)	CP_comp138978 (0)	9.2 (n.a.)	11.6 (n.a.)
	PgOSCPNY2 (AB014057)			
Protopanaxadiol synthase	CYP716A47-1 (AEY75212)	CP_comp137446 (4)	13.6 (± 1.7)	12.4 (± 1.5)
Protopanaxatriol synthase	CYP716A53v2 (AFO63031)	CP_comp110990 (0)	45.3 (n.a.)	65.0 (n.a.)

CP, Chunpoong; CS, Cheongsun; n.a., not available; RPKM, reads per kilobase per million; SD, standard deviation

¹⁾ Candidate transcripts and their isoforms were identified by TBLASTN searches using protein sequences of the mentioned queries

²⁾ Average RPKM value and SD were calculated using RPKM values of isoforms from the same transcripts

NCBI nr protein database yielded more than 90% hits in the CP and CS assembly sets. This is similar to a previous ginseng EST study in which 90% of the ESTs had hits with the nr database [41]. This high percentage is probably due to the removal of misassembled sequences and the high frequency of long sequences (approximately 1.9 kb average length) in our assembled transcripts. Our work shows that it is possible to obtain reliable transcriptome sequences in nonmodel species by performing resequencing and read-depth analysis.

4.2. Comparative analysis of transcriptomes in adventitious roots

Increased pharmacological efficiency is a main goal for genomic studies of ginseng. Under field conditions, a normal ginseng root is affected by biotic and abiotic factors and becomes vulnerable to

many diseases. To analyze differences between the cultivars, the effects caused by environmental factors need to be controlled as much as possible because cultivar-specific characteristics could be masked by environmental variation. The adventitious root culture system described herein provided useful materials for a comparative analysis of the transcriptomes of two cultivars, because the adventitious roots were cultured under more controlled environmental conditions than can be obtained in the field. Although tissue culture represents a stress condition for plants due to the high concentration of plant growth regulators like auxin or a lack of proper nutrients for growth, the observed differences between cultivars mostly represent the unique characteristics of each cultivar in a tightly controlled environment. From our experience, adventitious roots are easy to handle and their transcriptomes are highly reproducible.

As ginseng research requires highly reliable reference sequences for functional genomics, we have created a reference sequence from CP, a highly desirable cultivar, because of its superior quality [7]. Various root transcriptome studies have been reported using NGS technologies [15,42,43], but not for the adventitious root transcriptome. Our successful comparative analysis of the transcriptomes of two cultivars using adventitious roots will promote systematic approaches for functional genomics and metabolomics in medicinal plants.

To investigate the expression level of the assembled transcripts, we determined the RPKM value for each transcript. According to GO categorization of the resulting 100 most abundant transcripts, stress response-related transcripts were the most highly expressed categories in the adventitious roots (Figs. 3, 4). Consistent with our results, stress proteins have also been reported to be highly expressed in calli of other plant species [44,45]. Furthermore, proteomic analysis of ginseng hairy roots revealed that stress response-related proteins are the mostly highly expressed [46]. In normal ginseng plants, stress-responsive proteins are induced to high levels upon exposure to abiotic and biotic stresses [47,48]. Therefore, our results strongly suggest that *in vitro* culture conditions represent a stress on adventitious roots of ginseng plants, with stress response-related transcripts induced to protect cells from harmful conditions.

Although two cultivars showed different characteristics in adventitious root culture (Fig. 1), no significant difference of transcriptome profile was found between two cultivars. GO assignments and gene expression showed almost similar patterns

Table 3
List of candidate UGT transcripts involved in ginsenoside biosynthesis

Gene name	Used query (acc. no.)	Transcripts (no. of isoforms) ¹⁾	Average RPKM (SD) ²⁾		
			In CP	In CS	
UDP-glycosyl transferase	SvUGT74M1 (ABK76266)	CP_comp61978 (0)	31.0 (n.a.)	27.6 (n.a.)	
		CP_comp82119 (0)	277.8 (n.a.)	256.1 (n.a.)	
		CP_comp82124 (2)	16.9 (±0.5)	17.2 (±0.4)	
		CP_comp136300 (0)	6.8 (n.a.)	6.0 (n.a.)	
		CP_comp141169 (2)	30.8 (±0.5)	18.9 (±0.3)	
		CP_comp142539 (0)	23.2 (n.a.)	19.9 (n.a.)	
		CP_comp143836 (2)	24.3 (±0.2)	28.7 (±0.8)	
		CP_comp151861 (0)	10.1 (n.a.)	12.1 (n.a.)	
		MtUGT73K1 (AAW56091)	CP_comp49402 (0)	19.3 (n.a.)	22.4 (n.a.)
			CP_comp111429 (2)	35.6 (±3.5)	34.0 (±3.4)
			CP_comp126017 (0)	13.3 (n.a.)	21.9 (n.a.)
			CP_comp129770 (0)	4.6 (n.a.)	4.9 (n.a.)
			CP_comp135237 (3)	5.3 (±0.2)	6.8 (±0.1)
	CP_comp141064 (0)	16.8 (n.a.)	23.2 (n.a.)		
	CP_comp144124 (13)	2.5 (±0.3)	1.3 (±0.2)		
MtUGT71G1 (AAW56092)	CP_comp129842 (0)	1.3 (n.a.)	1.0 (n.a.)		
	CP_comp133432 (0)	45.1 (n.a.)	75.5 (n.a.)		
	CP_comp135281 (0)	23.7 (n.a.)	22.7 (n.a.)		
	CP_comp142900 (2)	16.2 (±1.3)	19.0 (±0.7)		
	CP_comp143506 (3)	5.9 (±1.5)	7.2 (±1.2)		
	CP_comp143570 (0)	169.0 (n.a.)	122.5 (n.a.)		

CP, Chunpoong; CS, Cheongsun; n.a., not available; RPKM, reads per kilobase per million; SD, standard deviation

¹⁾ Candidate transcripts and their isoforms were identified by TBLASTN searches using protein sequences of the mentioned queries

²⁾ Average RPKM value and SD were calculated using RPKM values of isoforms from the same transcripts

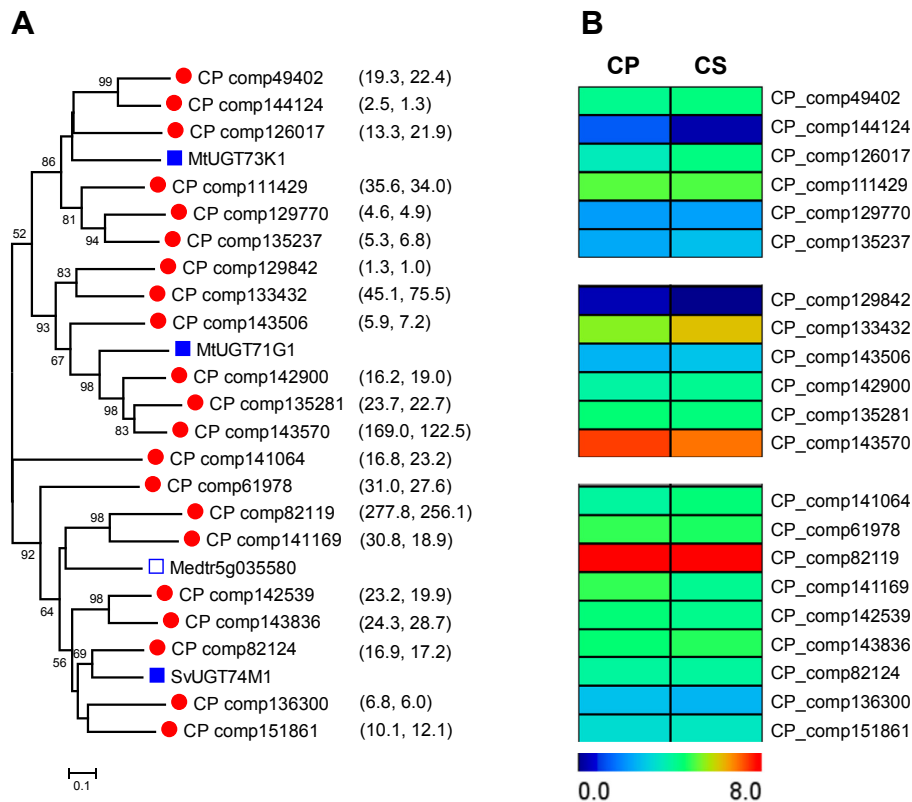


Fig. 6. Phylogenetic analysis of putative UGT proteins and comparison of their expression levels. (A) Twenty-one transcripts predicted to encode proteins with high similarity to UGT proteins involved in ginsenoside biosynthesis, SvUGT74M1 (ABK76266), MtUGT73K1 (AAW56091), and MtUGT71G1 (AAW56092), were identified in CP transcript dataset (Table 3). The deduced amino acid sequences were aligned using ClustalW, and the phylogenetic tree was generated using Poisson correction and the neighbor-joining method in MEGA5. Bootstrap values calculated for 1,000 replicates are shown on the branches; values < 50% are not shown. Average RPKM values of transcripts are shown in parentheses next to the transcript name. An *Medicago truncatula* homolog of SvUGT74M1, Medtr5g035580, was identified in the *M. truncatula* genome database (<http://medicago.jcvi.org/cgi-bin/medicago/overview.cgi>) and then included in the tree. (B) Expression comparison of the UGT genes in both cultivars. The RPKM values of transcripts closely grouped in the phylogenetic tree were compared and depicted by the expression map using MeV s/w (<http://www.tm4.org/mev/>) with \log_2 (RPKM) values. CP, Chunpoong; CS, Cheongsun; RPKM, reads per kilobase per million; UGT, UDP-glycosyltransferase.

between both cultivars (Figs. 2–6, Tables 2 and 3). Further comparative transcriptomics will be necessary to identify which gene makes a difference in adventitious root growth.

4.3. Genes related to ginsenoside biosynthesis

Ten transcripts encoding enzymes involved in ginsenoside biosynthesis were also identified through similarity searches with reported genes. Most of the ginsenoside biosynthesis transcripts were highly expressed in both cultivars, including transcripts for dammarenediol synthase or β -amyrin synthase, which catalyze the rate-limiting step for ginsenoside biosynthesis [49,50] (Fig. 5). This implies that the content and composition of ginsenosides may not be different between the cultivars under *in vitro* culture conditions. In addition, 21 transcripts related to UGT proteins were identified in the datasets of both cultivars (Fig. 6). Among those, three transcripts were closely related to MtUGT73K1, MtUGT71G1, and SvUGT74M1, which function in triterpene saponin biosynthesis. Therefore, these transcripts most likely encode UGTs involved in the last step of ginsenoside biosynthesis in *P. ginseng*. Simultaneous analysis of metabolite profiles and the transcriptome may promote in-depth understanding of the ginsenoside biosynthesis pathway.

4.4. Comparative analysis of transcriptomes between roots and adventitious roots

Through a comparative analysis of the transcriptomes of adventitious and normal ginseng roots, more than 6,000 transcripts were

identified to be unique to adventitious or normal roots, whose functional differences were characterized using GO analysis (Fig. 7). Although almost the same numbers of unique transcripts were analyzed for each tissue, transcripts unique to adventitious roots were more abundant for each individual GO term than those of normal roots. This indicates that a broader range of transcripts may be actively expressed in adventitious roots than in normal roots. In fact, more of the abundant transcripts in the adventitious root dataset were involved in transcription, cell proliferation, reproductive developmental processes, and multicellular organismal development.

5. Conclusion

In this study, we have generated a gene catalog for ginseng adventitious roots via *de novo* transcriptome assembly, which served as a useful resource for gene discovery in the ginsenoside pathway. In addition, we established an evaluation process to enhance assembly quality. To the best of our knowledge, this is the first report precisely categorizing the adventitious root transcriptome of *P. ginseng*. The approach we used to obtain the final transcriptome can be adopted for transcriptome assembly of other nonmodel species. Our work also reveals that adventitious roots are advantageous for transcriptome profiling analysis for genes related to secondary metabolites. If metabolite profiling is conducted along with transcriptome analysis, we may obtain more knowledge about complex metabolic pathways. In this work, we also developed an open web database for access and retrieval of our analyzed data. We anticipate that this study will take ginseng research to the next

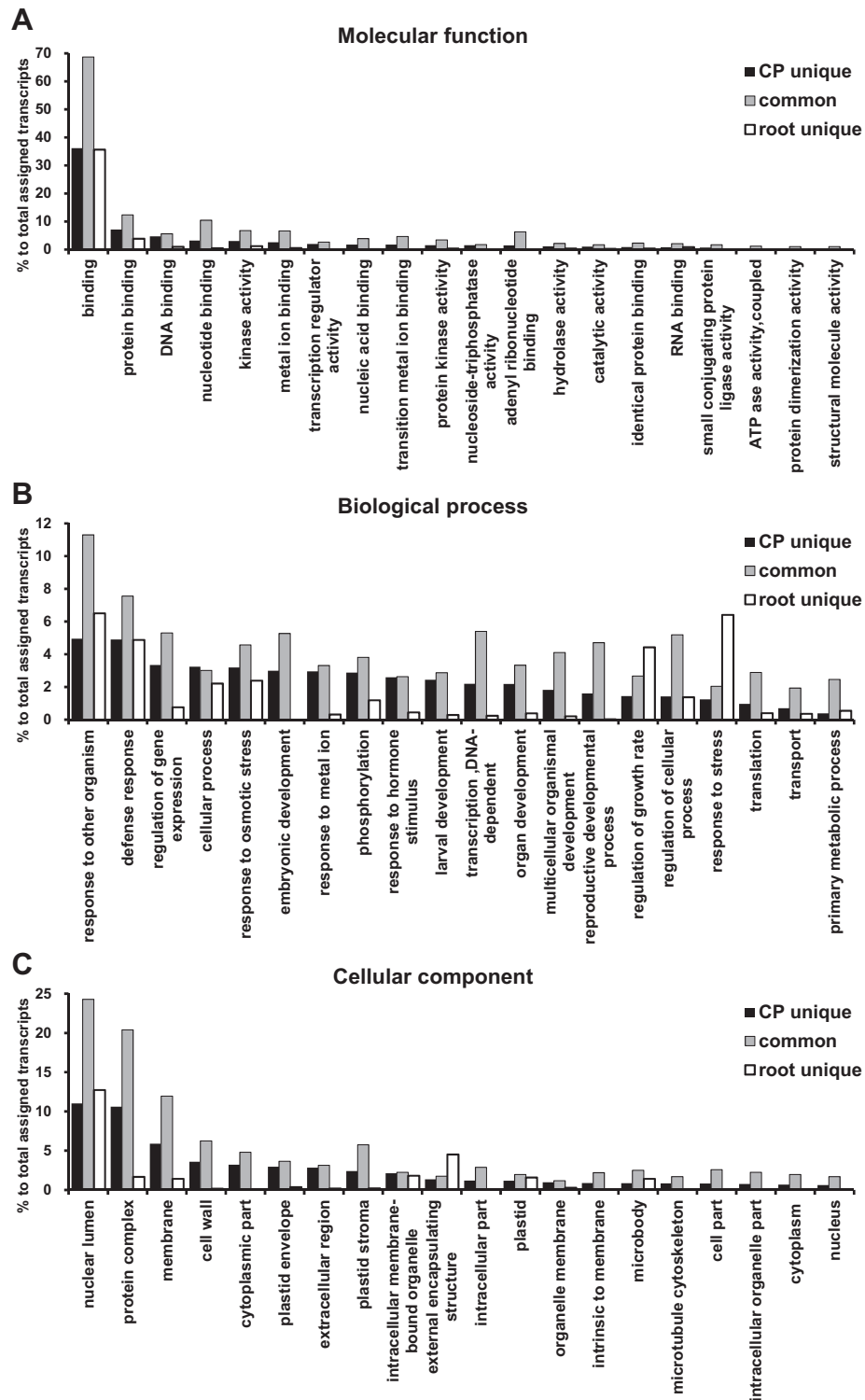


Fig. 7. GO analysis of transcripts common and unique to CP adventitious roots and normal roots. Terms are presented based on (A) molecular function, (B) biological process, and (C) cellular component. A total of 3,241 CP-unique and 2,378 root-unique transcripts were assigned GO terms using Blast2GO. GO assignment of transcripts common to CP adventitious roots and normal roots was analyzed for 23,675 common transcripts from the CP dataset. CP, Chunpoong; GO, gene ontology.

level, facilitating identification of additional ginsenoside genes and functional markers, as well as promoting understanding and engineering of complex metabolic pathways.

Conflict of interest

All authors declare no conflicts of interest.

Acknowledgments

We thank all members of the Laboratory of Functional Crop Genomics and Biotechnology, Seoul National University, for their technical assistance. This study was supported by the Next-Generation BioGreen 21 Program (No. PJ008202) of the Rural Development Administration, Republic of Korea.

References

- [1] Wen J, Zimmer EA. Phylogeny and biogeography of *Panax* L. (the ginseng genus, Araliaceae): inferences from ITS sequences of nuclear ribosomal DNA. *Mol Phylogenet Evol* 1996;6:167–77.
- [2] Hu SY. The genus *Panax* (ginseng) in Chinese medicine. *Econ Bot* 1976;30:11–28.
- [3] Lee FC. Facts about ginseng: the elixir of life. New Jersey: Hollym International Corporation Press; 1992.
- [4] Liu J, Wang S, Liu H, Yang L, Nan G. Stimulatory effect of saponin from *Panax ginseng* on immune function of lymphocytes in the elderly. *Mech Ageing Dev* 1995;83:43–53.
- [5] Shin HR, Kim JY, Yun TK, Morgan G, Vainio H. The cancer preventive potential of *Panax ginseng*: a review of human and experimental evidence. *Cancer Causes Control* 2000;11:565–76.
- [6] Kim SH, Park KS. Effects of *Panax ginseng* extract on lipid metabolism in humans. *Pharmacol Res* 2003;48:511–3.
- [7] Kwon WS, Chung CM, Kim YT, Lee MG, Choi KT. Breeding process and characteristics of KG101, a superior line of *Panax ginseng* C.A. Meyer. *Korean J Ginseng Sci* 1998;22:11–7.
- [8] Wang J, Man S, Gao W, Zhang L, Huang L. Cluster analysis of ginseng tissue cultures, dynamic change of growth, total saponins, specific oxygen uptake rate in bioreactor and immuno-regulative effect of ginseng adventitious root. *Ind Crops Prod* 2013;41:57–63.
- [9] Jung JD, Park HW, Hahn Y, Hur CG, In DS, Chung HJ, Liu JR, Choi DW. Discovery of genes for ginsenoside biosynthesis by analysis of ginseng expressed sequence tags. *Plant Cell Rep* 2003;22:224–30.
- [10] Sathiyamoorthy S, In JG, Gayathri S, Kim YJ, Yang DC. Generation and gene ontology based analysis of expressed sequence tags (EST) from a *Panax ginseng* C. A. Meyer roots. *Mol Biol Rep* 2010;37:3465–72.
- [11] Choi HI, Kim NH, Kim JH, Choi BS, Ahn IO, Lee JS, Yang TJ. Development of reproducible EST-derived SSR markers and assessment of genetic diversity in *Panax ginseng* cultivars and related species. *J Ginseng Res* 2011;35:399–412.
- [12] Lee JW, Kim YC, Jo IH, Seo AY, Lee JH, Kim OT, Hyun DY, Cha SW, Bang KH, Cho JH. Development of an ISSR-derived SCAR marker in Korean ginseng cultivars (*Panax ginseng* C. A. Meyer). *J Ginseng Res* 2011;35:52–9.
- [13] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009;10:57–63.
- [14] Luo H, Sun C, Sun Y, Wu Q, Li Y, Song J, Niu Y, Cheng X, Xu H, Li C, et al. Analysis of the transcriptome of *Panax notoginseng* root uncovers putative triterpene saponin-biosynthetic genes and genetic markers. *BMC Genomics* 2011;12(Suppl. 5):S5.
- [15] Sun C, Li Y, Wu Q, Luo H, Sun Y, Song J, Lui EM, Chen S. *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* 2010;11:262.
- [16] Li C, Zhu Y, Guo X, Sun C, Luo H, Song J, Li Y, Wang L, Qian J, Chen S. Transcriptome analysis reveals ginsenosides biosynthetic genes, microRNAs and simple sequence repeats in *Panax ginseng* C. A. Meyer. *BMC Genomics* 2013;14:245.
- [17] Patel RK, Jain M. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 2012;7:e30619.
- [18] Schulz MH, Zerbino DR, Vingron M, Birney E. *Oases*: Robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 2012;28:1086–92.
- [19] Grabherr MG, Haas BJ, Yassour M. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011;15:644–52.
- [20] Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, et al. SOAPdenovo-Trans: *de novo* transcriptome assembly with short RNA-Seq reads. *Bioinformatics* 2014;30:1660–6.
- [21] Birol I, Jackman SD, Nielsen CB, Qian JQ, Varhol R, Stazyk G, Morin RD, Zhao Y, Hirst M, Schein JE, et al. *De novo* transcriptome assembly with ABySS. *Bioinformatics* 2009;25:2872–7.
- [22] Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 2005;21:3674–6.
- [23] Velculescu VE, Kinzler KW. Gene expression analysis goes digital. *Nat Biotechnol* 2007;25:878–80.
- [24] Haralampidis K, Trojanowska M, Osbourn AE. Biosynthesis of triterpenoid saponins in plants. *Adv Biochem Eng Biotechnol* 2002;75:31–49.
- [25] Kushihiro T, Shibuya M, Ebizuka Y. Beta-amylin synthase—cloning of oxidosqualene cyclase that catalyzes the formation of the most popular triterpene among higher plants. *Eur J Biochem* 1998;256:238–44.
- [26] Lee MH, Jeong JH, Seo JW, Shin CG, Kim YS, In JG, Yang DC, Yi JS, Choi YE. Enhanced triterpene and phytosterol biosynthesis in *Panax ginseng* over-expressing squalene synthase gene. *Plant Cell Physiol* 2004;8:976–84.
- [27] Han JY, Kwon YS, Yang DC, Jung YR, Choi YE. Expression and RNA interference-induced silencing of the dammarenediol synthase gene in *Panax ginseng*. *Plant Cell Physiol* 2006;47:1653–62.
- [28] Tansakul P, Shibuya M, Kushihiro T, Ebizuka Y. Dammarenediol-II synthase, the first dedicated enzyme for ginsenoside biosynthesis in *Panax ginseng*. *FEBS Lett* 2006;580:5143–9.
- [29] Han JY, In JG, Kwon YS, Choi YE. Regulation of ginsenoside and phytosterol biosynthesis by RNA interferences of squalene epoxidase gene in *Panax ginseng*. *Phytochemistry* 2010;71:36–46.
- [30] Han JY, Kim HJ, Kwon YS, Choi YE. The Cyt P450 enzyme CYP716A47 catalyzes the formation of protopanaxadiol from dammarenediol-II during ginsenoside biosynthesis in *Panax ginseng*. *Plant Cell Physiol* 2011;52:2062–73.
- [31] Kim TD, Han JY, Huh GH, Choi YE. Expression and functional characterization of three squalene synthase genes associated with saponin biosynthesis in *Panax ginseng*. *Plant Cell Physiol* 2011;1:125–37.
- [32] Han JY, Hwang HS, Choi SW, Kim HJ, Choi YE. Cytochrome P450 CYP716A53v2 catalyzes the formation of protopanaxatriol from protopanaxadiol during ginsenoside biosynthesis in *Panax ginseng*. *Plant Cell Physiol* 2012;53:1535–45.
- [33] Achnine L, Huhman DV, Farag MA, Sumner LW, Blount JW, Dixon RA. Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. *Plant J* 2005;41:875–87.
- [34] Meesapyodsuk D, Balsevich J, Reed DW, Covello PS. Saponin biosynthesis in *Saponaria vaccaria*. cDNAs encoding beta-amylin synthase and a triterpene carboxylic acid glucosyltransferase. *Plant Physiol* 2007;143:959–69.
- [35] Chen S, Luo H, Li Y, Sun Y, Wu Q, Niu Y, Song J, Lv A, Zhu Y, Sun C, et al. 454 EST analysis detects genes putatively involved in ginsenoside biosynthesis in *Panax ginseng*. *Plant Cell Rep* 2011;30:1593–601.
- [36] Garg R, Patel RK, Tyagi AK, Jain M. *De novo* assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Res* 2011;18:53–63.
- [37] Li D, Deng Z, Qin B, Liu X, Men Z. *De novo* assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg. *BMC Genomics* 2012;13:192.
- [38] Xu DL, Long H, Liang JJ, Zhang J, Chen X, Li JL, Pan ZF, Deng GB, Yu MQ. *De novo* assembly and characterization of the root transcriptome of *Aegilops variabilis* during an interaction with the cereal cyst nematode. *BMC Genomics* 2012;13:133.
- [39] Zhang J, Liang S, Duan J, Wang J, Chen S, Cheng Z, Zhang Q, Liang X, Li Y. *De novo* assembly and characterisation of the transcriptome during seed development, and generation of genic-SSR markers in peanut (*Arachis hypogaea* L. *BMC Genomics* 2012;13:90.
- [40] Vijay N, Poelstra JW, Künstner A, Wolf JBW. Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive *in silico* assessment of RNA-seq experiments. *Mol Ecol* 2013;22:620–34.
- [41] Choi HI, Kim NH, Lee J, Choi BS, Kim KD, Park JY, Lee SC, Yang TJ. Evolutionary relationship of *Panax ginseng* and *P. quinquefolius* inferred from sequencing and comparative analysis of expressed sequence tags. *Genet Resour Crop Evol* 2013;60:1377–87.
- [42] Wang Z, Fang B, Chen J, Zhang X, Luo Z, Huang L, Chen X, Li Y. *De novo* assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (*Ipomoea batatas*). *BMC Genomics* 2010;11:726.
- [43] Daniel RR, Felipe HB, Natasha TH, Louise EJ, Daniel PS. Tomato root transcriptome response to a nitrogen-enriched soil patch. *BMC Plant Biol* 2010;10:75.
- [44] Ge XX, Chai LJ, Liu Z, Wu XM, Deng XX, Guo WW. Transcriptional profiling of genes involved in embryogenic, non-embryogenic calluses and somatic embryogenesis of Valencia sweet orange by SSH-based microarray. *Planta* 2012;236:1107–24.
- [45] Arican E, Albayrak G, Gozukirmizi N. Calli cultures from *Abies equi-trojani* (Aschers et Sinten) and changes in antioxidant defense system enzymes. *J Environ Biol* 2008;29:841–4.
- [46] Kim SI, Kim JY, Kim EA, Kwon KH, Kim KW, Cho K, Lee JH, Nam MH, Yang DC, Yoo JS, et al. Proteome analysis of hairy root from *Panax ginseng* C. A. Meyer using peptide fingerprinting, internal sequencing and expressed sequence tag data. *Proteomics* 2003;3:2379–92.
- [47] Lee OR, Sathiyaraj G, Kim YJ, In JG, Kwon WS, Kim JH, Yang DC. Defense genes induced by pathogens and abiotic stresses in *Panax ginseng* C. A. Meyer. *J Ginseng Res* 2011;1:1–11.
- [48] Sathiyaraj G, Lee OR, Parvin S, Khorolragchaa A, Kim YJ, Yang DC. Transcript profiling of antioxidant genes during biotic and abiotic stresses in *Panax ginseng* C. A. Meyer. *Mol Biol Rep* 2011;38:2761–9.
- [49] Kushihiro T, Ohno Y, Shibuya M, Ebizuka Y. *In vitro* conversion of 2,3-oxidosqualene into dammarenediol by *Panax ginseng* microsomes. *Biol Pharm Bull* 1997;20:292–4.
- [50] Hu W, Liu N, Tian Y, Zhang L. Molecular cloning, expression, purification, and functional characterization of dammarenediol synthase from *Panax ginseng*. *Biomed Res Int* 2013;2013:1–7. Article ID 285740.