Research Article
J. Ginseng Res. Vol. 35, No. 1, 21-30 (2011)
DOI:10.5142/jgr.2011.35.1.021

JGR Journal of Ginseng Research

Insilico Analysis for Expressed Sequence Tags from Embryogenic Callus and Flower Buds of *Panax ginseng* C. A. Meyer

Subramaniyam Sathiyamoorthy, Jun-Gyo In, Byum-Soo Lee, Woo-Seang Kwon, Dong-Uk Yang, Ju-Han Kim, and Deok-Chun Yang*

Korean Ginseng Center and Ginseng Genetic Resource Bank, Kyung Hee University, Yongin 449-701, Korea

Panax ginseng root has been used as a major source of ginsenoside throughout the history of oriental medicine. In recent years, scientists have found that all of its biomass, including embryogenic calli and flower buds can contain similar active ingredients with pharmacological functions. In this study, transcriptome analyses were used to identify different gene expressions from embryogenic calli and flower buds. In total, 6,226 expressed sequence tags (ESTs) were obtained from cDNA libraries of *P. ginseng*. *Insilico* analysis was conducted to annotate the putative sequences using gene ontology functional analysis, Kyoto Encyclopedia of Genes and Genomes orthology biochemical analysis, and interproscan protein functional domain analysis. From the obtained results, genes responsible for growth, pathogenicity, pigments, ginsenoside pathway, and development were discussed. Almost 83.3% of the EST sequence was annotated using one-dimensional *insilico* analysis.

Keywords: Ginsenoside pathway, *Panax ginseng*, Expressed sequence tags, Flower buds

INTRODUCTION

Panax ginseng C. A. Meyer is a perennial herb from the *Araliaceae* family. This plant is used as a medicine in Oriental countries such as China, Korea, and Japan, with more than a thousand years of history as a natural adaptogen [1]. The name *Panax* is derived from '*Panacea*,' which means "cure-all" in Greek. According to the literature, ginseng has been used to treat many challenging diseases like cancer, Alzheimer's, *Diabetes mellitus*, and HIV [1,2]. Activity has also been demonstrated against headache, fatigue, dizziness, nausea, and asthma. For the most part, ginseng extracts have been used alone as a tonic medicine or mixed with other plant extracts [3,4]. Additionally, plant material such as fresh root, dried leaves, dried flower buds, and red ginseng (boiled and dried roots) are consumed as components of chicken soup, green tea, tablets, capsules, and in medicinal powder form [5].

Ginseng has the ability to revitalize the body and mind, with numerous pharmacological effects having been shown on human metabolic systems [6]. Ginsenosides are the major therapeutic components, and special group of triterpeniod saponin components are present only in the panax family, in addition other antioxidants like polysaccharides, flavonoids, peptides, polyacetylenic alcohols, and fatty acids. More than 150 ginsenosides have been isolated from the panax family, of which 40 are present in Korean ginseng [4,7]. Ginseng, not only familiar to Asian countries, is also used as an herbal medicinal remedy by American and European consumers. Ginseng root is highly recognized throughout the world and has a worldwide market around US$3.5 billion. Ginseng roots are mainly used as a major source of ginsenoside [7,8]. For a few decades, the

Received 24 Aug. 2010, **Revised** 1 Nov. 2010, **Accepted** 17 Jan. 2011

***Corresponding author**
E-mail: dcyang@khu.ac.kr
Tel: +82-31-201-2100, Fax: +82-31-205-2688

other parts of the ginseng plant were used as a source of ginsenoside to balance the needs for effective utilization of biomass [9].

Every year, the amount of ginseng plant material consumed worldwide is enormous. However, the cultivation of ginseng is time-consuming and difficult, so alternative propagation methods through plant tissue culture and somatic embryogenesis were being studied [10,11]. Somatic embryogenesis has been successfully induced on solid and liquid media, directly from roots, leaves, and flower buds [10]. The culture of ginseng tissues in bioreactors was developed in order to produce fresh material containing ginsenoside saponins. Mass-production of embryogenic calli in *P. ginseng* is of special importance because the cellular and callus materials may be used as alternatives to naturally-grown ginseng roots [5]. *In vitro* culture is an advantageous alternative of ginseng production since it has no seasonal or regional restrictions [12,13]. There is no clear difference in saponin production between naturally-grown roots and *in vitro*-produced cells and roots. Regeneration of plant organs from embryogenic calli has also been successful [12]. Recently, ginseng callus has been used as an industrial source of ginseng saponins and extract. Flower buds are also a source of ginsenoside. Recently, scientists have reported different types of dammarane-type saponins present in flower buds. The major use of these buds is ginseng green tea [3,14-16].

In the present study, we generated expressed sequence tags (ESTs) for flower buds and embryogenic calli from *P. ginseng* to determine their gene expression levels. ESTs are short, unedited, randomly selected single-pass sequence reads derived from cDNA libraries. ESTs are also called the "poor man's genome" [17]. For the most part, the development of distinct tissues and cell-types and the production of primary and secondary metabolites are highly dependent on specific patterns of gene expression and transcript accumulation. The analysis of an EST offers a complete overview of the genes expressed in a certain organ and their relative expression levels [17]. The technique has proven to be a rapid and efficient way of obtaining information on gene diversity and mRNA expression patterns from a wide variety of tissues, cell types, or developmental stages.

In the past few years, there has been an increasing interest in the application of the EST approach to understand the molecular mechanisms associated with tissue culture. In this paper, our objectives are to generate the transcripts used to find the gene responsible for formation of secondary metabolites, particularly in the ginsenoside pathway, as well as to predict other secondary metabolite genes. Researchers have been studying the *Panax* species for many years, however it has only been a decade since scientists began using genome level analysis to improve plant breeding. Historically, no genome sequences were available, so the gene EST transcripts were used to aid in the study of plant development.

MATERIALS AND METHODS

Plant materials

Korean ginseng (*P. ginseng* C. A. Meyer) seeds containing extremely immature embryos were obtained just after harvest from the Ginseng Genetic Resource Bank, Kyung Hee University, Korea. The stratification of embryo maturation was completed in humidified sand. After the seed coats had been husked, the seeds were immersed in 70% alcohol for 1 min, surface-sterilized in 1% NaOCl for 20 min, and washed three times with sterilized distilled water, before the immature zygotic embryos were then dissected out. For the induction of embryogenic calli, the cotyledon bases of the zygotic embryos were transversally cut, and the excised cotyledons were incubated on Murashige and Skoog (1962) medium containing 3% sucrose, 1 ppm 2,4- dichlorophenoxyacetic acid, and 0.7% agar. The culture room was maintained at 24±2°C with a 16-h photoperiod of 24 $\mu mol\ m^{-2}\ s^{-1}$ under cool-white fluorescent tubes. Ginseng embryogenic calli were sub-cultured every four weeks in the same media. Embryogenic calli (white to light yellowish in color, compact, and friable) were selected, and non-embryogenic calli (mucilaginous and smooth) were discarded. Ginseng flower buds were harvested from four-year-old ginseng at the cultivation field in Daejeon, South Korea. Harvested flower buds were immediately frozen in liquid nitrogen and stored at -70°C until use.

RNA isolation and construction of the cDNA library

Total RNA was isolated from the flower buds and embryogenic calli of mature seeds of *P. ginseng* using the aqueous phenol extraction procedure [18]. The tissue was frozen and ground using liquid nitrogen prior to RNA extraction. Poly $(A)^+$ RNA was isolated using the Poly (A) quick mRNA isolation kit (Stratagene, La Jolla, CA, USA). The SMART cDNA library construction kit (Clonetech, Palo Alto, CA, USA) was used to synthesize cDNA from poly $(A)^+$ RNA. The fractions containing cDNA longer than 400 bp were recovered and used in the cDNA library construction. Size-selected cDNA was ligated to λTriplEx2 vector arms and packaged into

phage particles with Gigapack III Gold packaging extract (Stratagene). The cDNA library was amplified only once, yielding a primary titer of 10^6 pfu ml^{-1}. The plasmid library was then plated onto 70-mm LB media agar plates treated with ampicillin. Individual colonies were propagated and stored at -80˚C.

Nucleotide sequencing and sequence analysis

PTriplEx2 phagemids were converted from the λTriplEx2 library to *E. coli* strain BM25.8. The phagemids harboring inserts were selected via blue and white color screening on IPTG/X-GAL/ampicillin plates. Single-run partial sequencing was then conducted for the randomly selected cDNA clones. The 5' ends of the cDNA inserts were sequenced in an ABI Prism 3700 DNA sequencer (Perkin-Elmer, Norwalk, CT, USA), according to the thermal cycling protocol of the BigDye terminator cycle sequencing kit.

Generation of the unigene dataset

The unigene dataset contains a set of non-redundant sequences composed of singlets and contigs. Generating a unigene dataset requires several steps for sequence processing; all of them are integrated and freely accessible on the web-server EGassembler (EG) [19]. Data was subjected to the server with default parameters. The bioinformatics process starts with sequence cleaning, followed by masking the low quality sequence, removal of repetitive elements, and organelle sequence masking. The resulting sequences were subjected to CAP3 software to generate contigs [20]. The sequences not included in the contigs were called singlets.

Functional analysis and pathway assignments

Blast2go (B2G) is a high-throughput sequence analysis tool that was used to subject unigenes to BLASTX against the public non-redundant (NR) database in the NCBI. Based on the BLASTX result, sequences were putatively named using the BLAST description annotator (BDA) tool embedded in B2G. More collective logic models were embedded in B2G to retrieve gene ontology (GO), EC numbers, and Kyoto Encyclopedia of Genes and Genomes (KEGG) maps. GO terms were obtained from sequence similarity and BLAST score ($\leq E=10^{-3}$) with default parameters. Those annotations were simplified to plant functional categories using the plant GOslim [21].

KEGG orthology and protein domain assignments

KEGG orthology (KO) is an alternative control vocabulary for annotation. Unigenes were subjected to KO identity retrieval from KEGG using the freely accessible web server KO-based annotation system (KOBAS), with default parameters [22]. To obtain the protein domain information for the putative sequences, InterProScan embedded with B2G was utilized [21].

RESULTS AND DISCUSSION

Sequencing and assembly of expressed sequence tags

The Sanger sequencing method was followed to generate expressed sequence tags. The cloned cDNA library of *P. ginseng* was constructed using flower buds and embryogenic calli derived from flower buds. In total, 6,226 cDNA clones were collected based on gel fingerprinting and were subsequently sequenced. Among them, 3,067 were from flower buds, and 3,159 were from embryogenic calli. Unique sequences were generated using the EG bioinformatics pipeline. Out of 6,226 EST sequences, 3,845 unique sequences were used for annotation, after removing the non-informative sequence stretches such as reteroelements, LTR elements, TY1/Copia, Gypsy/DIRST1, Gyspy/DIRST1, DNA transposons, hodo-activators, MuDR-IS905 elements, and low quality sequences. The CAP3 software, which was embedded with blast2Go, was used for sequence assembly. The CAP3 results included 819 contigs and 3,026 singlets (all analyses were performed with default parameters) from the 3,845 cleaned sequences (Table 1). Sequences included in the contigs were similar, encompassing 75% of the EST library. All EST sequences are available at http://www.bioherbs.khu.ac.kr/Ginseng.

Comparing expressed sequence tags with public non-redundant databases

Homology-based functional assignment was used for putative sequences through BLASTX against the non-redundant database. The parameters used were an E-value equal to or less than 10^{-3}, HSP cut-off=33, and a maximum of 20 BLAST hits per sequence [21]. The average

Table 1. Unigene cluster report for *Panax ginseng* organs

Descriptive category	Flower buds	Embryogenic calli
Total cDNAs sequenced	3,067	3,159
No. of clustered sequences	1,860	1,985
No. of contigs	377	442
No. of singleton	1,483	1,543
No BLAST hits	255	370

Table 2. Classification based on species from BLAST hit sequences

Species	Flower buds	Embryogenic calli
Vitis vinifera	4,782	4,460
Arabidopsis thaliana	4,021	3,713
Populus trichocarpa	3,872	3,623
Oryza sativa	3,402	3,441
Ricinus communis	2,085	1,996
Zea mays	2,384	2,494
Physcomitrella patens	893	936
Picea sitchensis	806	819
Medicago truncatula	576	640
Solanum tuberosum	306	293

result of BLASTX for 83.4% of the EST library was 18 sequences per EST, with the remaining 625 sequences (16.5% of the EST library) not showing any meaningful matches (Table 1). 77.5% of BLASTX sequences were from the following plants: *Vitis vinifera* (15.70%), *Arabidopsis thaliana* (13.15%), *Populus trichocarpa* (12.75%), *Oryza sativa* (11.64%), *Ricinus communis* (6.94%), *Zea mays* (8.30%), *Physcomitrella patens* (3.11%), *Picea sitchensis* (2.76%), *Medicago truncatula* (2.07%), and *Solanum tuberosum* (1.01%), with the remaining coming from other plant species (Table 2). All of the ten plant species above are comparatively well studied and have more available experimental data than does *P. ginseng*. Annotations were given through the BDA tool, which was based on the homology comparison against the NR database and known sequence description. The annotations for putative sequences were used in the subsequent experimental analysis (Table 2).

Functional analysis based on Gene Ontology

GO is one of the best functional analysis tools used to describe gene functions and classification of gene families based on control vocabulary [23]. Control vocabulary schemas are designed using previous research evidence and are functionally grouped according to a hierarchy model based on a direct acyclic graph. Control vocabularies are grouped into three major categories, molecular function, biological processes, and cellular components. EST sequences were grouped in terms of their GO vocabularies, belonging to only one, a combination of two, or all three vocabularies, and were organized in a Venn diagram (Fig. 1). The total EST sequences included 1,847 cellular compounds (CC), 2,488 molecular functions (MF), and 2,061 biological processes (BP). Of these EST sequences, 1,342 were mapped under CC and MF, 1,343 were mapped under CC and BP, 1,734 were mapped under BP and MF, and 1,147 EST sequences were annotated according to all three GO vocabularies. All annotations were simplified using plant-specific GOslim.

Individual Venn diagrams were given for the flower buds (Fig. 1A) and embryogenic calli (Fig. 1B). In this report, we relied on well-annotated GO information of *Arabidopsis* and other plants [24]. A large number of unique sequences from our study were mapped under molecular functions such as protein binding, nucleotide binding, transporter, and structural molecule activity. The second most represented category was biological processes, including unique sequences associated with translation, cellular component organization and biogenesis, response to stress, transport, catabolic process, and metabolic process. Finally, the cellular compound

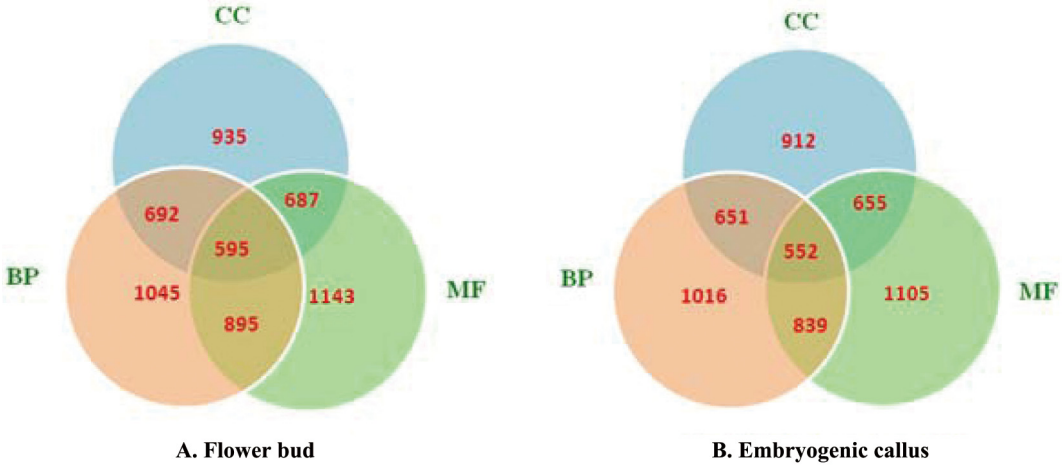

Fig. 1. Venn diagram of the INSAM data set. Annotated to one, a combination of two, and all three gene ontology vocabularies. CC, cellular compound; BP, biological process; MF, molecular function.

category was mapped for plastid, mitochondria, ribosome, cytosol, and endoplasmic reticulum (Tables 3-5). We studied very few genes related to plant growth, pigments, or pathogen-related functions.

Plant growth and identification of responsible genes

In plants, polyamines play an important role in morphogenesis, flower differentiation and initiation, pollen viability, root growth, somatic embryogenesis, anti-senescence, and the response to various biotic and abiotic stresses. Very few transcripts show sequence similarity to spermine synthase and spermidine synthase. The spermine synthase (PgSPS) (DC04011G05) gene was mapped with transferase activity (GO:0006810), transport (GO:0007275), multi-cellular organismal development (GO:0016740), and cytoplasm (GO:0005737), none of which have yet been characterized in *P. ginseng*. The four transcripts present in the library for spermidine synthase (PgSPD) were functionally characterized with polyamine analysis, and its role was also examined in *in-vivo* culture [25,26]. Lignins are also important components for plant growth and development, responsible for water, nutrient transport during plant growth and development, and in plant defense against biotic and abiotic stresses. Coniferrin beta-glucosidase (CBG) is involved in hydrolysis activity on monolignol glucosides in order to release cinnamyl alcohols for lignin oxidative polymerization [27]. The CBG (DC04031B10, DC04023D03) gene was

Table 3. Functional classification of biological processes based on gene ontology

Biological process	Flower buds (no. of ESTs)	Embryogenic calli (no. of ESTs)
Translation	144	141
Cellular organization and biogenesis	127	105
Response to stress	99	84
Transport	115	114
Catabolic process	69	61
Carbohydrate metabolic process	71	48
Protein modification process	69	56
Generation of precursor metabolites and energy	68	28
Amino acid and derivative metabolic process	48	45
Transcription	58	60
Signal transduction	48	59
Lipid metabolic process	43	33
Response to abiotic stimulus	53	47
Electron transport	33	34
Response to endogenous stimulus	21	32
Photosynthesis	49	8
Response to biotic stimulus	24	18
Secondary metabolic process	20	8
Response to external stimulus	12	9
DNA metabolic process	14	10
Cellular homeostasis	12	13
Anatomical structure morphogenesis	13	6
Embryonic development	11	12
Cell growth	7	8
Flower development	9	6
Cell cycle	15	6
Cell differentiation	0	9
Cell death	7	0
Regulation of gene expression, epigenetic	8	6

EST, expressed sequence tag.

Table 4. Functional classification of cellular compounds based on gene ontology

Cellular compound	Flower buds (no. of ESTs)	Embryogenic calli (no. of ESTs)
Plastid	229	196
Mitochondrion	183	172
Ribosome	116	116
Cytosol	76	69
Endoplasmic reticulum	26	32
Plasma membrane	26	20
Nucleoplasm	11	9
Golgi apparatus	19	20
Thylakoid	29	8
Extracellular region	15	11
Cytoskeleton	12	10
Peroxisome	9	10
Cell wall	13	6
Vacuole	12	0

EST, expressed sequence tag.

mapped with hydrolase activity (GO:0016787), binding (GO:0005975), and carbohydrate metabolic process (GO:0005488). The function of cinnamyl alcohol dehydrogenase (CAD) (DC06010F02 DC04012F05) in plants has been shown to catalyze the reversible conversion of p-hydroxycinnamaldehydes to their corresponding alcohols, leading to lignin biosynthesis. The CAD gene was mapped with binding (GO:0005488), catalytic activity (GO:0003824), and metabolic process (GO:0008152) [28].

Drought is a major abiotic stress affecting all levels of plant organization, it affects various metabolic processes but has only been suggested to be related to lignin biosynthesis. In maize, caffeate O-methyltransferase was overexpressed in the drought condition, affecting leaf growth. In our transcript, caffeate O-methyltransferase (DC04023B04) was mapped with transferase activity (GO:0016740) [29]. Cinnamate 4-hydroxylase is a cytochrome P450 that catalyzes the second step of the main phenylpropanoid pathway, leading to the synthesis of lignin, pigments, and many defense molecules. The cinnamate 4-hydroxylase (DC04007C10, DC04036C02) gene was mapped with binding (GO:0005488), catalytic activity (GO:0003824), and electron transport (GO:0006118) [30].

Pathogenicity and color pigment genes

Plant immunity to pathogen infections is believed to have contributed to the remarkable success of land plants on earth. Fungal pathogen infections can decrease plant productivity. In our library, very few genes are overexpressed or stimulated by external stimuli. Pathogenesis-related (PR) proteins encoded by plants are considered important for their roles in the plant defense mechanism. They are induced specifically in pathological or related situations. PR proteins are grouped into 17 gene families (PR-1 to PR-17) based on their structures. Our libraries show the presence of PR-1, PR-2, PR-5, and PR-10 subfamilies, that few were functionally characterized in *P. ginseng*, and that only a few other transcripts have been mapped for pathogen-related proteins [31,32]. The anthocyanin acyltransferase gene responsible for catalyzing the acyl group from acyl-CoA to a sugar moiety of an anthocyanin [33,34] is part of a class of flavonoids which are responsible for color discrimination (orange to blue) in flowers, playing an important role in plant reproduction and survival. For instance, the anthocyanidin 3-O-glucosyltransferase (DC04012A08) gene is responsible for red color [35].

Table 5. Functional classification of molecular function based on gene ontology

Molecular function	Flower buds (no. of ESTs)	Embryogenic calli (no. of ESTs)
Nucleotide binding	131	120
Protein binding	165	144
Structural molecule activity	121	116
Transporter activity	63	54
RNA binding	42	47
Kinase activity	50	43
Translation factor activity, nucleic acid binding	18	17
Transcription factor activity	20	17
Enzyme regulator activity	15	17
Nuclease activity	7	0
Lipid binding	11	9

EST, expressed sequence tag.

Biochemical analysis based on KEGG pathways

The KEGG is a major biochemical pathway database that includes many bioinformatics algorithms and is considered to be the most accessible database for the biological community. All putative transcripts were subjected to the KEGG database with a BLAST score, yielding KEGG enzyme codes and pathway maps. The putative sequences resulted in 236 EC numbers and 176 unique KEGG pathways. In total, 864 (22.48%) unique sequences were assigned to a KEGG map (Table 3) with sub-categories of carbohydrate metabolism, amino acid metabolism, energy metabolism, lipid metabolism, and secondary metabolism. From this, all of the enzymes were grouped into seven sub-groups of plant secondary metabolites, which were specially designed based on the KEGG-PLANT pathways. All enzyme codes were assigned a BLAST similarity score, while KEGG orthology was also assigned with the help of KOBAS tools.

All results were manually verified to validate the annotations assigned to the putative sequences. For example, squalene epoxidase (DC04007A02), with enzyme codes EC:1.14.99.7 and EC:1.11.1.7 and KEGG orthology K00511, is involved in the terpenoid biosynthesis pathway. The Delta12-fatty acid acetylenase (DC06009A12), with KEGG orthology K10256 and K10257 and enzyme code EC:1.3.1.35, is involved in fatty acid metabolism and is also responsible for cold stress.

Protein family and functional domains

Genome sequencing and structural genomics projects are providing new insights into the evolutionary history of protein domains. When comparing the sequence to distant structures, related domains are shown to be homologous due to an increasing realization that comparing a protein's structure or sequence with domains often precedes reliable and accurate predictions of molecular function. Based on these scenarios, structural proteomics scientists have developed interproscan for assigning homologous domains to putative sequences. Here, all putative sequences have been subjected to an interproscan search to assign a protein domain. Sequences, which did not contain gene ontologies, also contained the interpro identification based on the interproscan search. The unknown proteins also have some putative functional annotation. Interproscan is the final step annotation schema for unknown putative EST sequences (Table 6).

Our analysis also included secondary protein database identifications like PFAM, SMART, GENE3d, PROSITE, PROFILE, PRODOM, SUPERFAMILY, PANTHER, PIR, PRINTS, and TIGRFAMs. The EST sequence results for PFAM (1,082), SMART (224), GENE3d (681), PROSITE (373), PROFILE (274), PRODOM (69), SUPERFAMILY (767), PANTHER (1,413), PIR (50), PRINTS (209), and TIGRFAMs (90) were determined for the entire database.

EST protein descriptions are functionally unknown or hypothetical proteins. In this case, the interpro identifications can provide some insight to those protein functions using the domain homologs. For example, the protein (DC04004G05) EST assigned with IPR003333 has the domain for cyclopropane-fatty-acyl-phospholipid synthase. The domain function of a methyl transfer from the ubiquitous S-adenosyl-L-methionine (AdoMet) to either nitrogen, oxygen, or a carbon atom is frequently employed in diverse organismal proteins ranging from bacteria to plants and mammals. ESTs (DC04011E06) were assigned for YABBY (IPR006780) proteins, a group of plant-specific transcription factors involved in the specification of abaxial polarity in lateral organs such as leaves and floral organs. However, another EST, IPR013097, for the Dabb protein with a functionally unknown domain is upregulated in response to salt stress in *Populus balsamifera* (balsam poplar).

Putative ginsenoside pathway genes

Ginsenosides are the major therapeutic components in ginseng. Isoprenoid (also called terpenoid) pathways in plants are the backbones of ginsenoside synthesis. In plants, two isoprenoid pathways are present, one is the cytosol MVA pathway with an end product of IPP, while

Table 6. Gene ontology, enzymes, and Kyoto encyclopedia of genes and genomes maps

Descriptive category	Flower buds		Embryogenic calli	
	Count	ESTs	Count	ESTs
No. of GO annotations	4,428	1,239	3,650	1,125
No. of interpro annotations	607	1,035	517	1,079
No. of enzymes	548	464	473	403
No. of pathway maps	166	426	150	385

EST, expressed sequence tag; GO, gene ontology.

the other is the plastidial DXP pathway with end products of IPP and DMAPP [36]. In this pathway, the triterpene aglycone of ginsenoside, protopanaxadiol, is synthesized from 2, 3-oxidosqualene. Protopanaxadiol is then converted to protopanaxatriol by a cytochrome P450. Finally, different ginsenosides are synthesized through the addition of one or several monosaccharides to these triterpene aglycones by glycosyltransferase. One EST sequence is determined for the putative pathway genes, which are 1-deoxy-d-xylulose 5-phosphate reductoisomerase (EC 1.1.1.267), 1-deoxy-D-xylulose-5-phosphate synthase (EC 2.2.1.7), farnesyl diphosphate synthase (EC 2.5.1.1), geranyl-diphosphate synthase (EC 2.5.1.29), undecaprenyl pyrophosphate synthetase (EC 2.5.1.31), 2c-methyl-d-erythritol-cyclodiphosphate synthase (EC 4.6.1.12), and squalene epoxidase (EC:1.14.99.7). Thirty-one ESTs for cytochrome P450 and seven ESTs for glycosyltransferase were present in our library (Fig. 2). Very few genes were functionally characterized in the *P. ginseng* genome. Therefore, the above one-dimensional data will help to improve genome annotation and the determination of differential gene expression in the various parts of *P. ginseng*.

P. ginseng is a traditional medicinal plant in Asian countries, which has been commercialized all over the world. Recently, scientists have been interested in finding a novel gene involved in the ginsenoside pathway through whole genome analysis. However, very little transcriptome data is available in the gene bank for ginseng root and leaf material. To the best of our knowledge, this is

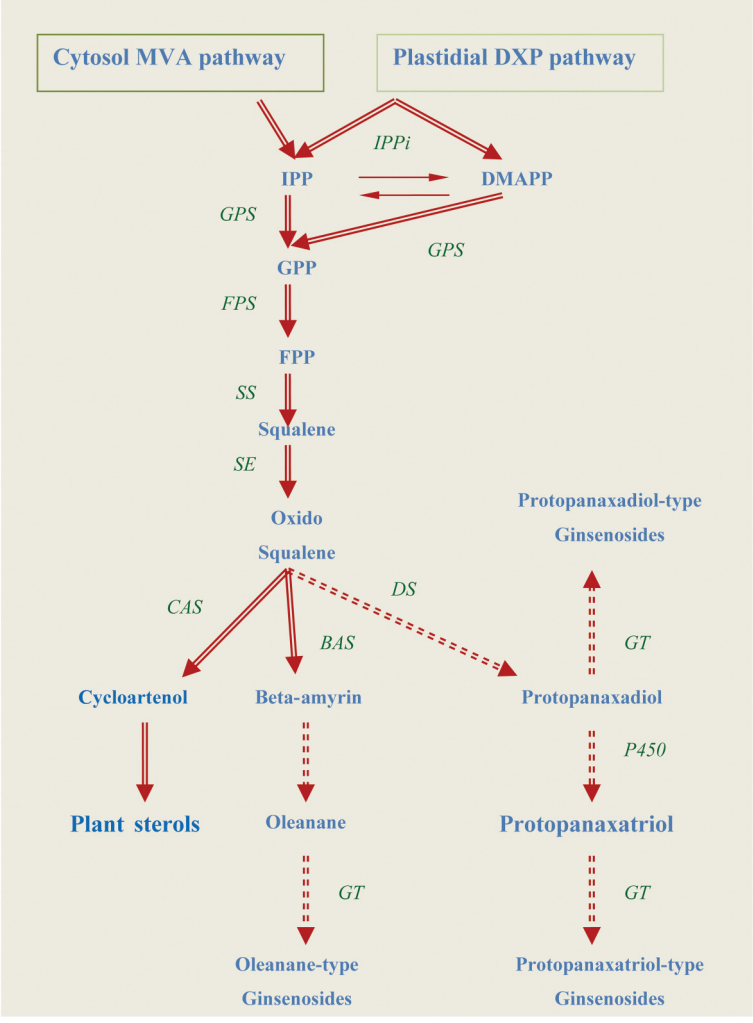

Fig. 2. Putative biosynthetic pathway of ginsenoside. Dotted lines represent the uncharacterized region of the pathway. MVA, mevalonate; DXP, 1-deoxyxylulose-5-phosphate; IPPi, isopentenyl diphosphate isomerase; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; GPP, geranyl diphosphate; FPP, farnesyl diphosphate; GPS, geranyl diphosphate synthase; FPS, farnesyl diphosphate synthase; SS, squalene synthase; SE, squalene epoxidase; CAS, cycloartenol synthase; BAS, beta-amyrin synthase; DS, dammarenediol synthase; GT, glucosyltransferase.

the first data from embryogenic calli and flower buds to provide a complete view of gene expression in different organs of the *P. ginseng* plant, aiding in the development of the *P. ginseng* gene library. This study demonstrated that one-dimensional data can benefit the analysis of multidimensional data and comparative genomics.

ACKNOWLEDGEMENTS

This study was supported by the KGCMVP for Technology Development program of the Ministry for Food, Agriculture, Forestry and Fisheries, Republic of Korea.

REFERENCES

1. Choi KT. Botanical characteristics, pharmacological effects and medicinal components of Korean *Panax ginseng* C A Meyer. Acta Pharmacol Sin 2008;29:1109-1118.
2. Park JD, Rhee DK, Lee YH. Biological activities and chemistry of saponins from *Panax ginseng* C. A. Meyer. Phytochem Rev 2005;4:159-175.
3. Wang H, Peng D, Xie J. Ginseng leaf-stem: bioactive constituents and pharmacological functions. Chin Med 2009;4:20.
4. Lee SY, Kim YK, Park NI, Kim CS, Lee CY, Park SU. Chemical constituents and biological activities of the berry of *Panax ginseng*. J Med Plants Res 2010;4:349-353.
5. Asaka I, Ii I, Hirotani M, Asada Y, Furuya T. Production of ginsenoside saponins by culturing ginseng (*Panax ginseng*) embryogenic tissues in bioreactors. Biotechnol Lett 1993;15:1259-1264.
6. Shi W, Wang Y, Li J, Zhang H, Ding L. Investigation of ginsenosides in different parts and ages of *Panax ginseng*. Food Chem 2007;102:664-668.
7. 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;37:3465-3472.
8. Sathiyamoorthy S, In JG, Gayathri S, Kim YJ, Yang DCh. Gene ontology study of methyl jasmonate-treated and non-treated hairy roots of *Panax ginseng* to identify genes involved in secondary metabolic pathway. Genetika 2010;46:932-939.
9. Christensen LP. Ginsenosides chemistry, biosynthesis, analysis, and potential health effects. Adv Food Nutr Res 2009;55:1-99.
10. Kevers C, Gal NL, Monteiro M, Dommes J, Gaspar T. Somatic embryogenesis of *Panax ginseng* in liquid cultures: a role for polyamines and their metabolic pathways. J Plant Growth Regul 2000;31:209-214.
11. Kim YJ, Kim MK, Shim JS, Pulla RK, Yang DC. Somatic embryogenesis of two new *Panax ginseng* cultivars Yun-Poong and Chun-Poong. Russian J Plant Physiol 2010;57: 283-289.
12. Lee JH, Kim YJ, Jung DY, Shim JS, Kim IH, Yang DC. In vitro induction of tetraploid roots by various pretreatments from anther of *Panax ginseng* C. A. Meyer. J Ginseng Res 2009;33:65-71.
13. Yang DC, Choi KT. Origin and development of single- and poly-embryos formed directly on excised cotyledons of ginseng zygotic embryos. J Ginseng Res 1999;23:74-80.
14. Nakamura S, Sugimoto S, Matsuda H, Yoshikawa M. Medicinal flowers. XVII. New dammarane-type triterpene glycosides from flower buds of American ginseng, *Panax quinquefolium* L. Chem Pharm Bull (Tokyo) 2007;55: 1342-1348.
15. Nguyen HT, Song GY, Kim JA, Hyun JH, Kang HK, Kim YH. Dammarane-type saponins from the flower buds of *Panax ginseng* and their effects on human leukemia cells. Bioorg Med Chem Lett 2010;20:309-314.
16. Tung NH, Song GY, Nhiem NX, Ding Y, Tai BH, Jin LG, Lim CM, Hyun JW, Park CJ, Kang HK, et al. Dammarane-type saponins from the flower buds of *Panax ginseng* and their intracellular radical scavenging capacity. J Agric Food Chem 2010;58:868-874.
17. Nagaraj SH, Gasser RB, Ranganathan S. A hitchhiker's guide to expressed sequence tag (EST) analysis. Brief Bioinform 2007;8:6-21.
18. Morris PC, Kumar A, Bowles DJ, Cuming AC. Osmotic stress and abscisic acid induce expression of the wheat Em genes. Eur J Biochem 1990;190:625-630.
19. Masoudi-Nejad A, Tonomura K, Kawashima S, Moriya Y, Suzuki M, Itoh M, Kanehisa M, Endo T, Goto S. EGassembler: online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments. Nucleic Acids Res 2006;34(Web Server issue):W459-W462.
20. Huang X, Madan A. CAP3: A DNA sequence assembly program. Genome Res 1999;9:868-877.
21. Conesa A, Gotz S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics 2008;2008:619832.
22. Mao X, Cai T, Olyarchuk JG, Wei L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics 2005;21:3787-3793.
23. Gene Ontology Consortium. The Gene Ontology in 2010: extensions and refinements. Nucleic Acids Res 2010;38(Database issue):D331-D335.
24. Horan K, Jang C, Bailey-Serres J, Mittler R, Shelton C, Harper JF, Zhu JK, Cushman JC, Gollery M, Girke T. Annotating genes of known and unknown function by large-

scale coexpression analysis. Plant Physiol 2008;147:41-57.
25. Monteiro M, Kevers C, Dommes J, Gaspar T. A specific role for spermidine in the initiation phase of somatic embryogenesis in *Panax ginseng* CA Meyer. Plant Cell Tissue Org Cult 2002;68:225-232.
26. Parvin S, Kim YJ, Pulla RK, Sathiyamoorthy S, Miah MG, Kim YJ, Wasnik NG, Yang DC. Identification and characterization of spermidine synthase gene from *Panax ginseng*. Mol Biol Rep 2010;37:923-932.
27. Dharmawardhana DP, Ellis BE, Carlson JE. cDNA cloning and heterologous expression of coniferin beta-glucosidase. Plant Mol Biol 1999;40:365-372.
28. Pulla RK, Shim JS, Kim YJ, Jeong DY, In JG, Lee BS, Yang DC. Molecular cloning and characterization of the gene encoding cinnamyl alcohol dehydrogenase in *Panax ginseng* C. A. Meyer. Korean J Med Crop Sci 2009;17:266-272.
29. Vincent D, Lapierre C, Pollet B, Cornic G, Negroni L, Zivy M. Water deficits affect caffeate O-methyltransferase, lignification, and related enzymes in maize leaves. A proteomic investigation. Plant Physiol 2005;137:949-960.
30. Lu S, Zhou Y, Li L, Chiang VL. Distinct roles of cinnamate 4-hydroxylase genes in Populus. Plant Cell Physiol 2006;47:905-914.
31. Kim YJ, Lee JH, Jung DY, Sathiyaraj G, Shim JS, In JG, Yang DC. Isolation and characterization of pathogenesis-related protein 5 (*PgPR5*) gene from *Panax ginseng*. Plant Pathol J 2009;25:400-407.
32. Pulla RK, Lee OR, In JG, Kim YJ, Senthil K, Yang DC. Expression and functional characterization of pathogenesis-related protein family 10 gene, *PgPR10-2*, from *Panax ginseng* C. A. Meyer. Physiol Mol Plant Pathol 2010;74:323-329.
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;41:875-887.
34. Griesser M, Hoffmann T, Bellido ML, Rosati C, Fink B, Kurtzer R, Aharoni A, Munoz-Blanco J, Schwab W. Redirection of flavonoid biosynthesis through the down-regulation of an anthocyanidin glucosyltransferase in ripening strawberry fruit. Plant Physiol 2008; 146:1528-1539.
35. Unno H, Ichimaida F, Suzuki H, Takahashi S, Tanaka Y, Saito A, Nishino T, Kusunoki M, Nakayama T. Structural and mutational studies of anthocyanin malonyltransferases establish the features of BAHD enzyme catalysis. J Biol Chem 2007;282:15812-15822.
36. Yan LI, Hong W, He-Chun YE, Guo-Feng LI. Advances in the plant isoprenoid biosynthesis pathway and its metabolic engineering. J Integr Plant Biol 2005;47:769-782.