Korean J. Medicinal Crop Sci. 12(2) : 154 − 162 (2004)

Analysis and Identification of Expressed Sequence Tags in Hairy Root Induced from Korean Ginseng (*Panax ginseng* C. A. Meyer)

Deok Chun Yang*† and Jun Gyo In**

*Department of Oriental Medicine Material and Processing, Kyung-Hee University, Suwon 449-701, Korea.
** Phytopia Research Institute, Biopia Co., Ltd, Yeongi, Chungnam 339-813, Korea

ABSTRACT : Hairy roots were induced from Korean ginseng (*Panax ginseng* C. A. Meyer) root explants and studied for their gene expression. A total of 3,000 ESTs (expressed sequence tags) from ginseng hairy root were determined and about 2,700 ESTs have a length of readable sequence, which result in 1,352 unique ESTs sequences. The 879 ESTs showed significant similarities to known nucleotide or amino acid sequences in other plant species, which were divided into eleven categories depending upon gene function. The remaining 473 sequences showed no significant matches, which are likely to be transcripts or to be matched to other organisms. The results indicated that the analysis of the ginseng hairy root ESTs by partial sequencing of random cDNA clones may be an efficient approach to isolate genes that are functional in ginseng root in a large scale. Our extensive EST analysis of genes expressed in ginseng hairy root not only contributes to the understanding of the dynamics of genome expression patterns in root organ but also adds data to the repertoire of all genomic genes.

Key words : *Panax ginseng,* expressed sequence tags, cDNA, function, hairy root

INTRODUCTION

Field cultivation of ginseng is a time-consuming and labor-intensive process: it takes five to seven years from seeding to final harvesting, during which great care is needed since growth is subject to several conditions such as soil, climate, pathogens and pests (Proctor, 1996; Sitcher, 1998). Thus the use of ginseng hairy root has been sought as a potential alternative for more efficient production of ginseng and its active ingredients such as ginseng saponins. With the ginseng hairy root, product quality and quantity can be more easily controlled, because there are no limitations by natural factors such as seasonal climate and geographical environment. In addition, the culture conditions and process variables can be more easily optimized. Increment of root yield is of special importance in *P. ginseng* because root growth is very slow and, consequently, highly expensive to maintain. In contrast, ginseng hairy roots induced by introducing Ri-plasmid of *Agrobacterium rhizogenes* into genomic DNA of plant cells show vigorous growth and the hairy roots produce the same or more saponins than natural ginseng roots (Yoshigawa & Furuya, 1987). Thus, ginseng hairy root may be a means for producing transformed ginseng plants, in particular with respect to increasing the root yield.

Recent rapid progress in genetic engineering and the availability of various automated genetic analysis instruments have made it possible to perform large-scale isolation and partial sequencing of anonymous cDNA clones (Yamamoto & Sasaki, 1997). Expressed sequence tags (ESTs) are short sequences, a few hundred base pairs in length, which are derivated by

† Corresponding author : (Phone) +82-31-201-2688 (E-mail) dcyang@khu.ac.kr
Received March 31, 2004 / Accepted April 19, 2004.

partial, single pass sequencing of the inserts of randomly selected cDNA clones (Tyra & David, 1997). These ESTs analyses are now a widely used tool in genome research (Adams *et al.*, 1991). ESTs assist in the quick identification of functions of expressed genes and in the understanding of the complexity of gene expression. ESTs continue to be the major source of new sequence records and genes, and also continue to serve as the major source of new gene discoveries. As part of its daily processing of EST data, National Center for Biotechnology Information (NCBI) identifies all homologies for new EST sequences through BLAST searches and incorporates that information into the companion dbEST database (Boguski *et al.*, 1994). In the current era of large scale genomic sequencing identifying new genes can require as little time as the few minutes taken to perform a computer-driven search of dbESTs. GenBank is a public database of all known nucleotide and protein sequences with supporting bibliographic and biological annotation, built and distributed by the NCBI (Benson *et al.*, 2000). Sequence comparisons of genomes or ESTs from related organisms provide insight into functional conservation and diversification (Swanson *et al.*, 2001). Since the first random sequencing of cDNA clones have been performed by the use of a human brain library (Adams *et al.*, 1991). Although ESTs from many plant species, including *Arabidopsis thaliana* (Hofte *et al.*, 1993; Newman *et al.*, 1994), *Oryza sativa* (Uchimiya *et al.*, 1992), *Zea mays* (Keith *et al.*, 1993), *Brassica napus* (Park *et al.*, 1993), *Brassica campestris* (Lim *et al.*, 1996) and *Medicago truncatula* (Covitz *et al.*, 1998) are available in dbEST, the gene analysis of *P. ginseng*, the Korean native plant, has not done till now. In this report we focus on the ginseng hairy root ESTs. The ultimate aim of our large-scale cDNA analysis of ginseng hairy root is to categorize all the expressed genes and to identify the valuable gene and physiological roles in ginseng. Furthermore the identified genes of hairy roots will be used in breeding a new species of ginseng.

ESTs analysis of ginseng hairy root in this report is a part of the ginseng genome project in Korea. This large-scale EST project was conducted to provide better understanding of the dynamics of genomic expression patterns of ginseng hairy root tissue. In addition, the tissue-specific EST information supplies supplementary data to the repertoire of all expressed genomic genes, because cDNA from whole plant bodies is much less likely to contain rare genes showing tissue-specific expression. Here we report the partial sequencing and database comparison of randomly selected cDNA clone of ginseng hairy root. These data identify a large number of genes that are expressed in this important medicinal plant.

MATERIALS AND METHODS

Plant materials

Four-year old Korean ginseng (*Panax ginseng* C. A. Meyer) roots were sterilized with NaOCl and pieced to root disks. The root disks were co-cultivated with *Agrobacterium rhizogenes* R1000 or *A. tumefaciens* A4T, and then transferred to hormone-free MS medium (Murashige & Skoog, 1962) with 500 ㎍/㎖ carbenicillin at 25℃ in dark conditions according to the procedure of Yang *et al.* (1998).

RNA isolation and construction of a cDNA library

Total RNA was isolated from ginseng hairy root tissues using aqueous phenol extraction procedure as described by Morris *et al.* (1990). The tissue was frozen and ground in liquid nitrogen prior to extraction of RNA. Poly $(A)^+$ RNA was isolated by oligo-dT-cellulose chromatography using the "mRNA purification kit (Amersham pharmacia, UK). A commercial cDNA synthesis kit (Uni-ZAP XR) and GigapackIII Gold packaging extract were used to construct library according to the manufacture's instruction manual (Stratagene, USA). Fractions containing cDNA greater than 500 bp were recovered, and this library was amplified once to yield a primary titre of 2.5×10^6 pfu ml^{-1}. The plasmid library was plated on 70 ㎜ LB media agar plates with ampicillin. Individual colonies were propagated and saved at -80℃ until further use.

Nucleotide sequencing and sequence analysis

pBluescript $SK^{(+)}$ phagemids were excised from the

Uni-ZAP XR library using ExAssist helper phage in *E. coli* XL-1 blue MRF'. Phagemids containing inserts were selected by blue and white color screening on IPTG/X-GAL/ampicillin plates as described in Sambrook *et al.* (1989). Single-run partial sequencing of such randomly selected cDNA clone was performed. The 5' ends of the cDNA inserts were sequenced using the T3 primers by an automatic DNA sequencer (ABI prism 377 DNA sequencer, Perkin-Elmer, USA) according to the thermal cycling protocol of the BigDye Terminator Cycle Sequencing kit.

A comparison of each analyzed EST clone to DNA and protein databases at NCBI was performed using the BLAST algorithm of Altschul *et al.* (1990). The edited sequences were searched against non-redundant nucleotide (EMBL, GenBank, DDJB and PDB) and protein (PIR and SwissProt) databases using BLASTX program.

RESULTS

Characterization of the EST sequences

To characterize the ginseng hairy root ESTs generated in this experiment, the deduced amino acid sequences were compared with protein sequences in the existing databases such as PIR, SwissProt and others. In ginseng hairy roots, the sites of infection by *A. rhizogenes* represent only a small proportion of the total mass of tissue in the root. We postulated that genes unique of preferentially expressed in hairy roots might be critical for metabolism. We therefore constructed a cDNA library from ginseng hairy root to increase the proportionate representation of ginseng hairy root-specific genes. As shown in Table 1, we classified the protein sequences that have homologies to sequences in the databases according to putative functions. The results indicate that genes involved in metabolic pathways produced the most abundant transcripts in the ginseng hairy root. Transcripts for the translational apparatus (especially ribosomal proteins) ranked next in abundance.

The similarity was considered significant when the percentage of the amino acid identity between two sequences was more than 40% or the significance value of the similarity was more than 8.0 and when further manual examination confirmed that the computer comparison was meaningful. The ESTs without significant similarity to the sequences in the PIR database were further compared to the sequences in the other databases.

Putative identification of genes

Each EST was compared against all sequences in the nonredundant database at the NCBI using the program BLASTX, which compares translated nucleotide sequences with protein sequences. Sequences that had no homology to any protein in the database were then reanalyzed using the program BLASTN, which compares nucleotide sequences with nucleotide sequences. The results of each comparison were screened manually. After screening and editing we retained 1,352 ESTs. Out of these, 879 had significant homology to previously identified genes, as indicated by their putative identifications. These ESTs were grouped into 11 functional categories (Table 1). Although the BLAST scores and P values were considered, the assessment of whether or not a given homology was significant was determined by investigator's judgment, not by absolute numerical cutoffs. The annotations of genes with similarities to an EST were used to assign a putative identification. In cases in which the annotation was vague, information contained in MEDLINE abstracts related to the gene was used to assign a putative identification

Abundantly expressed genes

The relative abundance of the mRNA in a tissue is approximately reflected in the abundance of its corresponding cDNA in non-normalized libraries. Random sequencing of cDNAs therefore yields information about the relative expression levels of the genes represented by the ESTs. As expected, many of the identified cDNAs were housekeeping genes, which are related to metabolism (Table 1). The highly expressed transcripts of the ginseng hairy root cDNA library are listed in Table 2. Clones encoding RNA-binding protein are the most abundant clones in this experiment. Ubiquitin conjugation enzyme E2, and elongation factor are critical for protein metabolism.

Table 1. ESTs identified putatively from the ginseng hairy root.

Metabolism

1-aminocyclopropane-1-carboxylate oxidase
2-oxoglutarate dehydrogenase e2 subunit
3-hydroxybutyryl-coa dehydrogenase
3-isopropylmalate dehydratase-like protein
Acetyl-coa synthetase
Acetyltransfersase
Acyl-coa oxidase
Adenylate kinase b
Allene oxide cyclase
Alpha/beta hydrolase
Alpha-mannosidase
Aminotransferase-like protein
Amp deaminase homolog
Anthranilate synthase beta chain
Apk1 gene for protein tyrosine-serine-threonine kinase
Argininosuccinate synthase
Aspartate aminotransferase
Atp citrate lyase
Beta-amylase enzyme
Beta-tubulin cofactor-like protein
Blue copper-binding protein
Branched-chain amino acid aminotransferase
Cab expression 1-like protein
Chloroplast thylakoidal processing peptidase
Cinnamoyl-coa reductase
Citrate synthase
Ctp synthase
Cytidine deaminase
Cytochrome c oxidase subunit 6b-1
Cytochrome p450
Cytosolic aldehyde dehydrogenase
Dehydrogenase-like protein
Diaminopimelate decarboxylase
Dihydroflavonol 4-reductase-like
Dihydrolipoamide s-succinyltransferase
Dihydroorotase precursor
Dihyrdolipoamide acetyltransferase
Dtdp-d-glucose 4,6-dehydratase
Endo-1,3-1,4-beta-d-glucanase
Enolase
Enolase-phosphatase
Enoyl coa hydratase-like protein
Fibrillarin
Fumarate hydratase
Gamma response i protein
Glucan synthase component
Glucosyl transferase
Glutathione synthase
Glycine hydroxymethyltransferase
Growth-regulating factor 1
Homocysteine s-methyltransferase athmt-
Isomerase like protein
L-asparaginase
Leukotriene-a4 hydrolase-like protein
Lipase
L-lactate dehydrogenase
Long-chain-fatty-acid-coa ligase
Lysophospholipase
Methionine aminopeptidase 2
Methionine synthase
Methionyl-trna synthetase
Methylenetetrahydrofolate dehydrogenase
Precursor
Molybdenum cofactor sulfurase
Monodehydroascorbate reductase
Nadh dehydrogenase
N-carbamyl-l-amino acid amidohydrolase-like protein
Oligopeptidase a
Oxidoreductase
Oxoglutarate dehydrogenase
Pectate lyase
Phosphatidic acid phosphatase
Phosphomannomutase
Phytochrome a
Phzf, catalyzing the hydroxylation of phenazine -1-carboxylic acid to 2-hydroxy-phenazine-1-carboxylic acid
Poly-ribose polymerase
Pyridoxamine 5-phosphate oxidase
Pyrophosphate-fructose-6-phosphate 1-ph osphotransferase
Pyruvate dehydrogenase e1 alpha subunit
Quinone reductase
Respiratory burst oxidase protein a
Ribophorin i
Rudimentary enhancer
S-adenosylmethionine:2-demethylmenaquinone methyltransferase
Short-chain dehydrogenase
Sorbitol dehydrogenase-like protein
Starch synthase
Sulfite oxidase
Thioredoxin reductase
Translationally controlled tumor protein
Trehalose-6-phosphate synthase
Tyrosine phosphatase-like protein
Urease accessory protein g homolog
Vacuolar atp synthase subunit c
Vetispiradiene synthase
Xylan endohydrolase

Protein synthesis

Ribosomal protein l14
Ribosomal protein l18a
Ribosomal protein l19
Ribosomal protein l26
Ribosomal protein l35
Ribosomal protein l7
Ribosomal protein s15
Ribosomal protein s19
Ribosomal protein s27
Ribosomal protein s30
Ribosomal s29-like protein
26s proteasome regulatory subunit
30s ribosomal protein s1
30s ribosomal protein s17
40s ribosomal protein s10
40s ribosomal protein s18
40s ribosomal protein s3
40s ribosomal protein s4
40s ribosomal protein s7
40s ribosomal protein s8
40s ribosomal protein s9
40s ribsomomal protein
50s ribosomal protein l18
50s ribosomal protein l33
60s acidic ribosomal protein p3
60s ribosomal protein l22
60s ribosomal protein l13
60s ribosomal protein l17
60s ribosomal protein l18
60s ribosomal protein l21
60s ribosomal protein l27
60s ribosomal protein l36
60s ribosomal protein l38
60s ribosomal protein l44
60s ribosomal protein l7
60s ribosomal protein l7a
Acidic ribosomal protein p1a
Cullin
Cysteine proteinase a494 precursor
Elongation factor
Eukaryotic initiation factor
Iceosomal-like protein
Initiation factor 4f p28 subunit
Initiation factor 3g
L3 cytoplasmic ribosomal protein
Mitochondrial processing peptidase
Molybdopterin biosynthesis cnx1 protein
Multicatalytic endopeptidase
Nclpp2
Osnac6 protein
Palmitoyl-protein thioesterase
Polyubiquitin 4 - common sunflower
Proteasome subunit alpha type 2
Protein disulfide-isomerase
Protein phosphatase
Protein-tyrosine-phosphatase-like protein
Serine protease-like protein
Subtilisin-like proteinase
Ubiquitin conjugating enzyme e2
Zinc metalloproteinase

DNA, RNA realated and gene expression

Alanyl t-rna synthetase
Apoptosis antagonizing transcription factor
At-hook dna-binding protein
Atp-dependent rna helicase-like protein
Bzip dna-binding protein-like
Caax processing zinc-metallo endoprotease
Chp-rich zinc finger protein-like
Clathrin coat assembly protein
Cobw-like protein
Dead box rna helicase
Dna topoisomerase i
Dna-binding protein-like
Dna-directed rna polymerase ii
Dna-directed rna polymerase ii subunit
Dsrna-binding protein odb1
Exonuclease exoi
Glutamyl aminopeptidase
Hd-zipper protein
Helicase
Histon acetyltransferase hat1

Table 1. Continued.

Histone deacetylase
Histone h2a
Histone h2b
Isoleucine-trna ligase
Kinesin-like protein b
Leucine zipper-containing protein
Mutt protein
Mybst1
N2,n2-dimethylguanosine trna methyltransferases
Nuclear movement protein nudc
Nuclear ribonucleoprotein g
Nucleoid dna-binding protein
Nucleotide sugar epimerase-like protein
Pfam family pf00400 -wd domain
Prematurely terminated mrna decay factor-like protein
Prolyl trna synthetase
Pspzf zinc finger protein-like
Rad23 protein
Remorin
Replication protein a1
Ribonucleoside-diphosphate reductase large subunit
Ring finger protein 14
Rna 3'-terminal phosphate cyclase-like protein
Rna helicase
Rna methyltransferase
Rna polymerase transcriptional regulation mediator
Rna-binding protein
Scarecrow-like 3 protein
Serine/threonine-protein kinase
Serine/threonine protein phosphatase
Small nuclear ribonucleoprotein
Splicing factor
Transcription factor btf3
Transcription initiation factor iib
Wd-40 repeat protein-like

Signal transduction

Adp-ribosylation factor
Atp-dependent clp protease regulatory subunit clpx
Calmodulin-domain protein kinase
Casein kinase ii alpha subunit
Galactokinase-like protein
Gtp-binding protein
Guanine nucleotide-binding protein
Guanylate kinase
Kinase-like protein
Mitogen-activated protein kinase
Mutator-like transposase
Phosphatidylinositol-4-phosphate 5-kinase
Protein kinase
Ras-related protein rab7
Receptor protein kinase
Rp42 is a member of the transposase pf
Shaggy kinase
Signal peptidase
Signal recognition particle receptor-like protein
Small ras-like gtp-binding protein
Somatic embryogenesis receptor-like kinase-like protein

Cytoskeleton

Ankyrin-like protein
Arp2
Kinesin heavy chain-like protein
Myosin heavy chain kinase
T-complex protein 1

Membrane and transport

Abc transporter-like protein
Acetyltransfersase
Adp,atp carrier protein precursor
Amino acid transporter
Atp synthase epsilon chain
Aux1-like permease
Cationic amino acid transporter-like protein
Chloride channel protein homolog clc1
F1f0-atpase inhibitor-like protein
Intrinsic protein
Leucine-rich repeat transmembrane protein kinase 1
Ligand-gated ion channel subunit
Membrane transporter
Mitochondrial dicarboxylate carrier protein
Mitochondrial import inner membrane translocase
Mitochondrial inner membrane translocatin g protein
O-linked glcnac transferase
Outer mitochondrial membrane protein porin
Peptide transporter
Phosphate transport protein
Phospholipid-transporting atpase 1
Phosphoribosylanthranilate transferase-like protein
Plasma membrane intrinsic protein
Potassium transport protein
Potential cation-transporting atpase
Protein transport protein sec61
Proton pump interactor
Rattus o-glcnac transferase
Sm protein
Sulfate adenylyltransferase
Translocase 7k chain tom7
Transport inhibitor response 1
Transporter protein
Vacuolar proton-atpase subunit 1
Water-stress inducible protein

Defense

2-cys peroxiredoxin-like protein
Bacillus cota
Berberine bridge enzyme-like protein
Chaperonin
Competence-damage protein
Dehydration-induced protein erd15
Disease resistance protein,
Glutaredoxin-like protein
Glutathione-regulated potassium-efflux System protein
Glutathione s-transferase t3
Heat shock protein
Low-temperature-induced cysteine proteinase
Metallothionein-I gene transcription activator
Multidrug resistance-associated protein
Noi protein
Pathogenesis-related protein
Peroxidase
Peroxiredoxin
Proline-rich protein
Proteinase inhibitor ii
Selenium-binding protein
Sudd-like protein
Uvb-resistance protein-like
Wound-inducible carboxypeptidase

Hormone

Abscisic acid responsive elements-binding factor
Auxin response factor 8
Auxin-induced protein

Cell wall metabolism

Beta-galactosidase
Cellulose synthase catalytic subuni
Glucosidase i
Hydroxyproline-rich glycoprotein
Integral membrane protein
Pectin methylesterase-like protein
Pectinesterase
P-glycoprotein
Polygalacturonase inhibitor-like protein
Porin-like protein

Cell division

Cdc2 protein kinases
Cdc6 protein
Cell division protein ftsh
Cyclin
Skp1
Wee1-like protein

Others

12s seed storage globulin precursor
Aim1 protein
Argonaute protein
Arm repeat-containing protein
Atm-like protein
Beta-adaptin-like protein b
Brittle-1 protein precursor
Burp domain containing protein
Bzip protein
Calnexin homolog precursor
Centrin
Centromere/kinetochore protein zw10 homolog
Cer1 protein
Chromatin remodeling protein syd
Chromosome condensation protein
Cim1 protein
Clathrin-associated adaptor protein
Cw14

Table 1. Continued.

Dad-1 protein	Lamin	Ripening-related protein-like
Dag protein	Minor allergen	Rub1 conjugating enzyme
Delta-cop	Mitotic spindle checkpoint protein mad2	Sah7 protein
Dipeptidyl peptidase IV-like protein	Nicotiana lesion-inducing orf	Sec13-related protein
Disulfide bond formation protein	Nodulin-like protein	Sec61p
Dynamin-like protein dlp2	Not56-like protein	Separation anxiety protein-like
Early nodule-specific protein-like	Nucleosome assembly protein	Similar to protein disulfide isomerase
Er lumen protein retaining receptor-like protein	Peptidyl-prolyl cis-trans isomerase	Sin3 associated polypeptide p18
Er33 protein	Peptidylprolyl isomerase	Snap25a protein
Eza1	Pescadillo-like protein	Starch associated protein r1
Fertilization-independent seed 2 protein	Phosphoribosylformylglycinamidine cyclo-ligase precursor	Su3-9 protein homolog f2714.7
Fkbp-like	Pms10 protein	Sucrose cleavage protein
Fkf1-like protein 2	Prli-interacting factor	Sync1 protein
Fusca protein fus6	Prmc3	Syntaxin protein
Glycosylation enzyme-like protein	Progesterone-binding protein	Tuberisation-related protein
Gmfp4	Pur alpha-1	Uclacyanin i
Gtpase activator protein	Px domain	Virf-interacting protein fip1
Isomerase like protein	Rad23 protein	Wo8e3.3
Kin17 protein	Rap8	Xrn3
Ku70-like protein	Rga-like	

Table 2. Most abundant mRNA. Values in parentheses indicate percentage of total.

Putative identification	No. of ESTs	Percentage of total (%)
Protein kinase	18	1.76
RNA-binding protein	11	1.08
Ubiqutin conjucating enzyme	10	0.98
Cytochrome P450	6	0.59
ARP2 (Actin-related protein 2)	6	0.49
ABC transporter	5	0.49

The mitogen-activated protein kinase (MAPK) including p42/44 ERK, p38 MAPK, a serine-threonine protein kinase, has been suggested to play a role in apoptosis (Cross *et al.*, 2000). ARP2 (actin-related protein 2) is involved in cytoskeleton formation. The ABC transporter is a membrane protein and is related to materials transport when the expression is subject to a complex hormonal and environmental regulation. They thus seem to represent a highly conserved molecular mechanism for the directed transport of specific molecules against a concentration gradient. The abundant expression of these genes reflects the actively growing state of the source tissue used to generate library.

DISCUSSION

We analyzed ESTs using ginseng hairy root which are fast productive than normal ginseng to search for beneficial genes related to secondary metabolite synthesis and further genes will be selected for breeding in ginseng. As a part of our effort to identify genes that functions in ginseng hairy root, the randomly selected 3,000 cDNA clones of the ginseng hairy root were partially sequenced by single-run sequencing reaction and generated 1,352 ESTs. These EST sequences were characterized mainly by comparison with the sequences in the PIR, SwissPort and other databases. 879 cDNA clones matched the protein-coding sequences from other plant species in the databases. These sequences include housekeeping genes, stress- or defense related genes, cell wall formation genes and others. The nature of the cDNA library seems to be important in generating ESTs, depending on the purpose of EST generation. It is obvious that a cDNA library that is devoid of noncoding regions and allows sequencing of the coding regions is important for putative identification of the EST sequences by database search. If we consider the rapid increase of the number of genes isolated from other organisms, including human, mouse, *Drosophila,*

Caenorhabditis elegans, Saccharomyces cerevisiae, and *E. coli* (Maddox, 1991), it is very likely that the identifiable genes by database search will increase substantially. In addition, development of a better computer algorithm for identifying distantly related genes that show weak amino acid identity but contain structurally or evolutionally related sequences will be necessary to increase the level of identification of the EST sequences. The sequencing and the database search of the random cDNA clones as described in this experiment should be able to provide an idea about what types of genes are functioning in ginseng hairy root.

Genes of Interest

This is the first EST analysis performed on ginseng, the most important medicinal plant in the oriental, and provides information including genetic composition and physiological changes in molecular level. The genes involved in metabolic pathways were most abundant in the ginseng hairy roots; genes involved in protein synthesis and processing ranked next in abundance. Our RNA preparation included actively growing and differentiating cells. Therefore we expected to find a diverse expressed sequences representing major root functions such as membrane transport, cell growth and division and cytoskeleton. However, since no cells from other portions of the plant were present, we presumed that the functional genes typically associated with shoot and leaf would not be found. We found that a number of sequences showed significant homologies with plant specific genes including cell wall protein and cell wall synthetic enzyme. The strongest matches were also homologous with house-keeping genes including ribosomal proteins, translation initiation factors, and the cellular cytoskeleton (Table 1). Recently the other roles of the house-keeping genes in plant development were reported. For example recent findings indicate that genes encoding eukaryotic translation initiation factors such as eIF 2a may have a role in stress responses probably via phosphorylation (Langland *et al.*, 1997). Also, a number of gene products with homology to proteins involved in signal transduction and protein synthesis were identified. Genes categorized under defense and metabolism functions could also be appropriate targets of study the host defenses mechanisms during infection by *A. rhizogenes*. Several other sequences are especially interesting because of their homology to genes with known functions in other systems. Various ribosomal protein genes were especially abundant in the ginseng hairy roots, suggesting that these hairy root cells were metabolically active. This observation is consistent with previous EST data from *Arabidopsis* (Hofte *et al.*, 1993) and Chinese cabbage (Lim *et al.*, 1996). According to their putative functions, the sequences in the ginseng hairy root cDNA library were further characterized by grouping the cDNA clones into 11 distinct functional classes. Details of the gene species included in each class are given in the legend to figure 1. Like a normal root, ginseng hairy root possibly play many essential roles as specific root organs in a plant: they absorb and transport water and dissolved ions, and they serve as the place of synthesis and storage for plant growth regulators such as auxin and cytokinins. Among the clones, ATP-binding cassette (ABC) transporter gene is abundantly expressed. Its superfamily contains membrane proteins that translocate a variety of substrates across extra- and intra-cellular membranes (Dean *et al.*, 2001). It has been implicated in the active movement of a variable organisms, both prokaryotic and eukaryotic (Higgins, 1992). This result is consistent with the physiological role of the hairy root in mineral and water uptake like the normal root. In comparison with ESTs from ginseng hairy root, there was a remarkably high proportion of ESTs that were similar to transcripts of stress response genes in the ginseng hairy root library. Such genes accounted for 6% of the 879 cDNA clones identified (Fig. 2). *In vitro* culture condition gave rise to stress in the plant organs, which resulted in active expression of stress response genes. It is interesting that 25 defense- or stress- related genes were identified among the database-matched ESTs. The matched sequences include glutathione S-transferase, glutaredoxin, metallothionein, pathogenesis-related protein genes and heat-shock protein family.

The results suggest that wounding and pathogen *in*

vitro culture condition are major stress affecting the ginseng hairy root. It is known that genes responsive to wounding and pathogen attack often show shared expression patterns. A wounding response gene, metallothioneins, encodes small cystein−rich proteins with a high metal−binding capacity. The members of these gene families are differentially regulated and probably have overlapping but distinct functions. It was shown that they are essential for cellular process of metal uptake and detoxification (Hamer *et al.*, 1986). Also, sequences with homology to drought and cold−response genes were identified. The high frequency of these transcripts greatly contributed to the high proportion of putative stress response gene transcripts. Other clones showed significant similarity to pathogenesis related genes encoding endohydrolytic or peroxidase protein (Van Loon, 1985), and to the heat shock protein family. This result indicates that like a plant root system, hairy root may express many defense− or stress−related genes to cope with the intensive environmental interactions mentioned above. Expression cataloging of the ginseng hairy root ESTs as performed in our experiment along with the putative identification of the ESTs by database search may provide the information that are needed to dissect the molecular processes of ginseng hairy root.

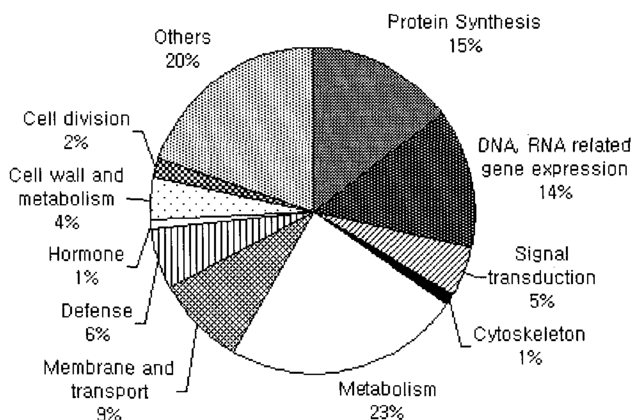

Fig. 1. Functional classification of ginseng hairy root ESTs. The ESTs that had sequence similarity to only known proteins were classified based on their biological functions except unknown function protein.

Uses for the ESTs

In order to identify the functions, EST sequences that encode proteins of known or predicted function may be used to create peptide antigens for generating antibodies. The protein products of cytoskeletal protein and cell wall enzyme genes may be good candidates for this approach. Thus the EST database may be useful for the scientists who have biochemically purified proteins of interest from *P. ginseng.* The partial peptide sequence of a purified protein could be compared against translated EST sequences. The related and more extensive cDNAs could then be readily identified and used as tools for additional studies.

Results described here indicate that EST analysis is a useful approach for isolation novel genes as well as ginseng hairy root homologous to known genes. Of course, we cannot exclude the possibilities that the homologous genes in ginseng do not have the same function compared to other plants. However, the homology may be very useful in formulating predictions about their functions. The further analysis of ESTs from *P. ginseng* will be extremely helpful for understanding the biological characteristics of ginseng species. Moreover, considering the rapid development of gene characterization in other species, further cDNA screening will facilitate the isolation of many agronomically and medicinally important genes in ginseng.

ACKNOWLEDGEMENT

This work was supported by a grant from the Plant Diversity Research Center of the 21st Century Frontier Research Program (code # PF003101−01) funded by Ministry of Science and Technology of the Korean government.

LITERATURE CITED

Adams MD, Kelly JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombine WR, Venter JC (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252: 1651-1656.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J. Mol. Biol. 215:377-403-

410.

Benson DA, Mizrachi IK, Lipman DJ, Ostell J, Rapp BA, Wheeler DL (2000) GenBank 28(1):15-18.

Boguski MS, Tolstoshev CM, Bassett DE (1994) Gene discovery in dbEST. Science 265:1993-1994.

Covitz PA, Smith LS, Long SR (1998) Expressed sequence tags from a root-hair-enriched *Medicago truncatula* cDNA library. Plant Physiol. 117:1325-1332.

Cross TG, Scheel-Toellner D, Henriquez NV, Deacon E, Salmon M, Lord JM (2000) Serine/Threonine Protein Kinases and Apoptosis. Exp. Cell Res. 10:256(1) 34-41.

Dean M, Rzhetsky A, Allikmets R (2001). The human ATP-binding cassette (ABC) transporter superfamily. Genome Res. 11(7):1156-1166.

Hamer DH (1986) Metallothionein. Annu. Rev. Biochem. 55:913-951.

Higgins CF (1992) ABC transporters: from microorganisms to man. Annu. Rev. Cell Biol. 8, 67-113.

Hofte H, Desperez T, Amselem J, Chiapello H, Caboche M, Moisan A, Jourjon MF, Charpenteau JL, Berthomieu P, Guerrier D, Giraudat J, Quigley F, Thomas F, Yu DY, Mache R, Ryanal M, Cooke R, Grellet F, Delseny M, Parmentier Y, de Marcillac G, Gigot C, Fleck J, Philips G, Axelos M, Bardet C, Tremousaygue D, Lescure B (1993) An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. Plant J. 4:1051-1061.

Keith CS, Hoang DO, Barret BM, Feigelman B, Nelson MC, Thai H, Baysdorfer C (1993) Partial sequence analysis of 130 randomly selected maize cDNA clones. Plant Physiol. 101:329-332.

Langland JO, Landland L, Zeman C, Saha D, Roth D (1997) Developmental regulation of a plant encoded inhibitor of eukaryotic initiation factor 2 α phosphorylation. Plant J. 12:393-400.

Lim CO, Kim HY, Kim MG, Lee SI, Chung WS, Park SH, Hwang IH, Cho MJ (1996) Expressed sequence tags of Chinese cabbage flower bud cDNA. Plant Physiol. 111:557-588.

Maddox JC (1991) The case for the human genome. Nature 352:11-14.

Morris PC, Kumar A., Bowles, DJ, Cuming AC (1990) Osmotic stress and abscisic acid regulate the expression of the *Em* gene of wheat. Eur. J. Biochem. 190:625-630.

Murashige T, and Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol. Plant. 15:473-497.

Newman T, de Bruijn FJ, Green P, Keegstra K, Kende H, Mclntosh L, Ohlrogge J, Raikhel M, Somerville S, Thomashow M, Retzel E, Sommervile C (1994) Genes galore: a summary of methods for accessing results from large-scale sequencing of anonymous Arabidopsis cDNA clone. Plant Physiol. 103:359-370.

Park YS, Kwak JM, Kwon OY, Kim YS, Lee DS, Cho MJ, Lee HH, Nam HG (1993) Generation of expressed sequence tags of random root cDNA clones of *Brassica napus* by single-run partial sequencing. Plant Physiol. 103:359-370.

Proctor JTA (1996) Ginseng: old crop, new directions. In: Janick J, editior. Progress in new crop. Alington, VA: ASHS 565-577.

Sambrook H, Fritch EF, Maniatis T (1989) Molecular cloning: a Laboratory Manual. Ed. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sitcher O (1998) Getting to the root of ginseng. Chemtech. 28:26-32.

Swanson WJ, Clark AG, Waldrip H, Wolfner M, Aquadro CF (2001) Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc. Natl. Acad. Sci. 98(13):7375-7379.

Tyra GW, David L (1997) A comparison of expressed sequence tags (ESTs) to human genomic sequences. Nucleic Acids Res. 25:1626-1632.

Uchimiya H, Kidou S, Shimazaki T, Takamatsu S, Hashimoto H, Nishi R, Aotsuka S, Matsubayashi Y, Kidou N, Umeda K, Kato A (1992) Random sequencing of cDNA libraries reveals a variety of expressed genes in cultured cells of rice (*Oryza sativa* L.). Plant J. 2:1005-1009.

Van Loon LC (1985) Pathogenesis-related proteins. Plant Mol. Biol. 4:111-116.

Yamamoto K, Sasaki T (1997) Large-scale EST sequencing in rice. Plant Mol. Biol. 35:135-144.

Yang DC, Kim YH, Yang DC, Min BH, Shin SL, Choi KT (1998) Selection of active grow hairy root lines in ginseng. Korena J. Plant Tissue Culture. 25(6):525-530.

Yoshikawa T, Furuya T (1987) Saponin production by cultures of *Panax ginseng* transformed with *Agrobacterium rhizogenes*. Plant Cell Rep. 6:449-453.