African Journal of Biotechnology Vol. 8 (22), pp. 6082-6088, 16 November, 2009
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB09.630
ISSN 1684–5315 © 2009 Academic Journals

Full Length Research Paper

Construction and analysis of a cDNA library from yellow-fruit ginseng (*Panax ginseng* C.A.Meyer.) leaf tissue

Chengjun Yang[1,2], Jun Wang[2,3]*, Chuanping Yang[2] and Guifeng Liu[2]

[1]Key Laboratory of Forest Tree Genetic Improvement and Biotechnology of Ministry of Education, Northeast Forestry University, Harbin 150040, P. R. China.
[2]College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083, P. R. China.

Accepted 9 July, 2009

The total RNA was isolated from yellow-fruit ginseng (*Panax ginseng* C.A. Meyer) leaf tissue. A cDNA library of panax ginseng leaves was constructed by using pDNR-LIB vector according to the SMART cDNA library construction kit protocol. We obtained 378 high quality sequences (GenBank accession number: ES672876-ES673253). ESTs were annotated, analyzed by BlastX and functional classified based on gene ontology, the results showed that 221 ESTs showed significant similarities to gene sequences in Nr database and were known genes, 21 ESTs were non-significance and unknown function genes, and 136 ESTs were considered novel genes. Most of the ESTs appeared to be related to physiological and cellular processes.

Key words: cDNA library, expressed sequence tags (EST), *Panax ginseng*.

INTRODUCTION

Ginseng (*Panax ginseng* C. A. Meyer), a perennial herb from the Araliaceae family, is one of the most commonly utilized medicinal plants. Ginseng is considered to be one of the most potent medicinal plants that have been used to bolster immunity, provide nutrition, ameliorate fatigue and enhance resistance to stress, disease and exhaustion. Ginsenosides, which are triterpene glycosides (saponins), are believed to be the main active compounds in ginseng tissues (Choi et al., 2005; Kim et al., 2006). More than 30 different ginsenosides have been isolated from ginseng plants (Sun et al., 2001; Zheng et al., 2001). Despite the considerable commercial interest in ginsenosides, little is known about the genes and biochemical pathways of ginsenoside biosynthesis. 'Jilin yellow-fruit ginseng' is identified as a homozygous recessive mutant from ordinary panax ginseng and contains high levels of ginsenosides, rich proteins, amino acids and other nutrients (Zhao et al., 1998).

cDNA library is one of the basic means to study functional genomics. Expressed sequence tags (ESTs) are partial sequences of randomly selected complementary DNA (cDNA) clones; automated sequencing techniques make it possible to generate large numbers of EST at one time. Expressed sequence tags (EST) analysis is an effective method to discover novel genes and investigate gene expression in different organs and tissues (Wang et al., 2006). The generation and analysis of expressed sequence tags provides useful information on development, metabolism and signaling in various organisms. Expressed sequence tags have applications in the discovery of new genes, mapping of the genome and identification of coding regions in genomic sequences (Miyahara et al., 2000).

In the study, we constructed a cDNA library of yellow-fruit ginseng leaf tissue and obtained 378 high quality EST sequences. Through EST analysis and gene ontology, this will help us to understand the gene expression pattern, discover novel genes and study the biochemical pathways of ginsenoside biosynthesis.

*Corresponding author. E-mail: junwang1966@yahoo.com.cn. Tel: 86-045-82191829(O). Fax: 86-451-82190607-11.

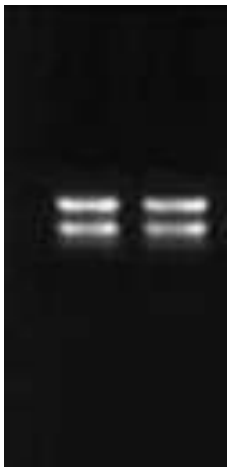

Figure 1. Agarose gel electrophoresis of total RNA from yellow-fruit ginseng leaf.

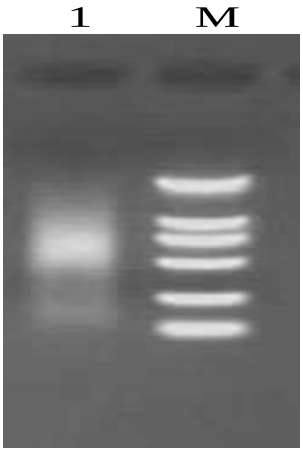

Figure 2. Agarose gel electrophoresis of double-stranded cDNA. Lane 1: double-strand cDNA; lane M: DL2000 marker.

MATERIALS AND METHODS

Plant materials

Actively growing 2 year old yellow-fruit ginseng (*Panax ginseng* C. A. Meyer) leaves were obtained from ZuoJia, Jilin province, China, on June 20, 2006. The harvested leaves were immediately frozen in liquid nitrogen and then stored at -70°C until RNA isolation.

Isolation and quantification of RNA

The total RNA was isolated from ginseng leaf tissue using SDS methods, as described by Yang et al. (2008). Total RNA was quantified by measuring the optical density of a dilute RNA solution. The integrity of the RNA was analyzed using 1.1% agarose/EtBr gel electrophoresis. The purity of the RNA was checked by the ratio of OD_{260}/OD_{280}.

cDNA synthesis and library construction

In accordance with the creator^{tm} SMART^{TM} cDNA library construction kit user manual provided by the manufacturers (Clontech), total RNA (1.0 µg) as starting material was reverse transcribed to synthesize first-strand cDNA, and double-strand cDNA was synthesis by LD-PCR. 5 µg of the double-stranded cDNA were taken for analysis by electrophoresis on a 1.1% agarose/EtBr gel. Then the amplified double-strand cDNA was digested with proteinase K and Sfi I. After digestion, cDNA size fractionation was performed using chroma spin-400 columns to collect cDNA large than 400 bp and checked the profile of fractions on a 1.1% agarose/EtBr gel. The cDNA was ligated to the Sfi I-digested dephosphorylated pDNR-LIB vector provided with the kit and electroporated into DH5α *Escherichia coli* bacteria to develop the cDNA library of ginseng leaf tissue. To make a large, stable quantity of a high-titer stock of the library, we amplified the primary cDNA library.

Identification of the cDNA library

According to the library tittering protocol, the unamplified and amplified cDNA library were tittered. To identify the cDNA inserts of the recombinants and determined the percentage of recombinant clones, 16 plaques were randomly picked from plate. Then PCR was performed with M13 primers provided by the advantage 2 PCR kit. The PCR products checked on 1.2% TAE/agarose gel with DNA size markers.

Library sequencing and analysis

The cDNA library clones were plated into LB agar plate containing 30 µg/ml of chloramphenicol, white clones were picked in to 96 well plates randomly, plasmids DNA of each clone were prepared by standard alkaline lysis preparation protocol and then sequenced by Beijing genomics institute. Sequencing was performed in a Mega BACE1000 DNA capillary sequence machine. The raw expressed sequence tags (EST) were edited to remove vector and poor quality sequences. The remaining sequences were subjected to blast analysis against the non-redundant database on the GeneBank (http://www.ncbi.nlm.nih.gov/blast) for similarity. The confirmed sequences were submitted to the dbEST database of GeneBank. Identified genes were classified according to gene ontology. The network information of the gene ontology database is categorized into 3 groups: cellular component, molecular function and biological process.

RESULTS AND DISCUSSION

High-quality total RNA was isolated from the leaf tissue of ginseng. Electrophoresis of the total RNA on 1.1% agarose/EtBr showed distinct 28S and 18S rRNA bands and the ratio of intensities of 28S and 18S was about 2:1 (Figure 1), the ratio of OD_{260}/OD_{280} to the total RNA was 1.90. The total RNA isolated was integrated and suitable for constructing the cDNA library. The double-strand cDNA synthesized using LD-PCR was analyzed on a 1.1% agarose/EtBr gel and the product showed a smear from 0.2 to 2.5 kb (Figure 2). The double-strand cDNA fractionated using CHROMA SPIN-400 columns was

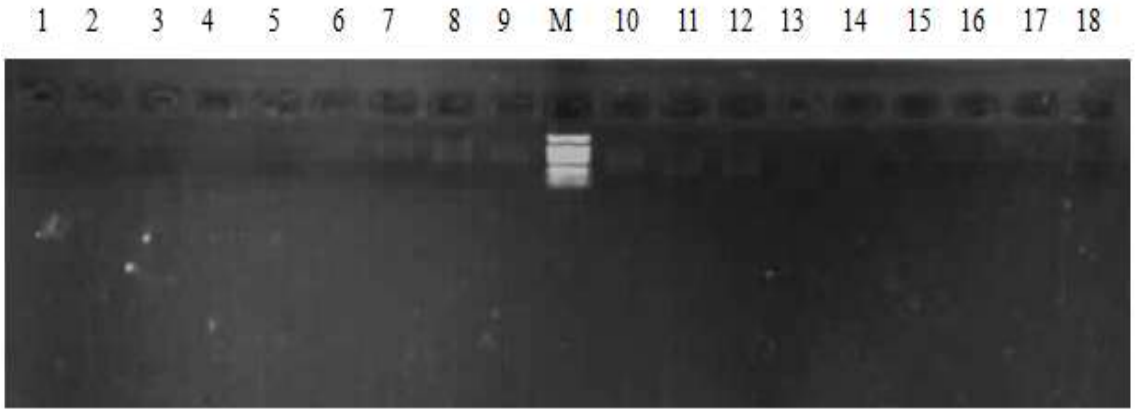

Figure 3. Double-stranded cDNA fractiionated on a 1.1% agarose/EtBr gel 1-18: double-stranded cDNA; M: DL2000 marker.

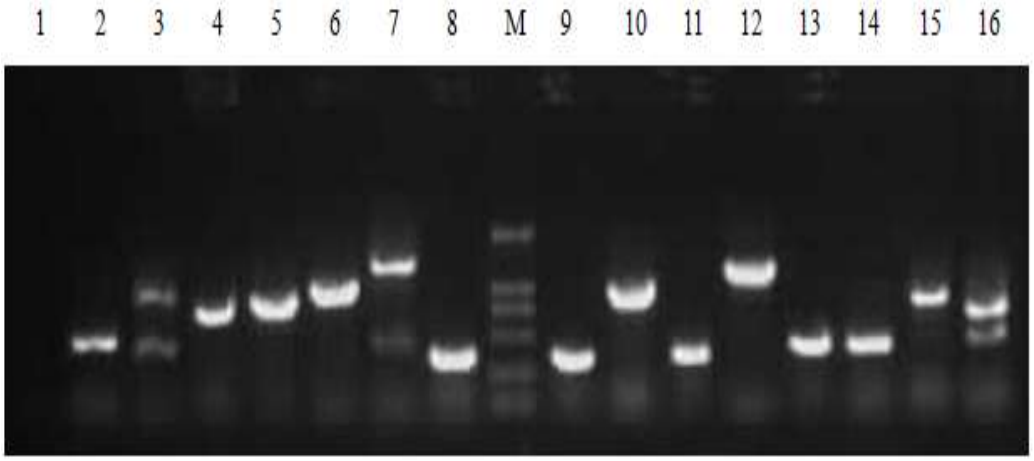

Figure 4. 16 clones in the cDNA library were selected randomly to evaluate their insert sizes. Lanes 1-16: Products of inserts; lane M: DL2000 marker.

collected and pipette 3 µl of cDNA to run on a 1.1% agarose/EtBr gel at 150 V for 10 min. Electrophoresis results showed the cDNA in lanes 7 - 11 was larger than 400 bp (Figure 3) and was ready to be ligated to the Sfi I-digested, dephosphorylated pDNR-LIB vector.

The titer of the un-amplified constructed cDNA libray was approximately 1.01×10^6 pfu/ml, the titer of the amplified cDNA library was 2.16×10^9 pfu/ml. 16 independent clones were selected randomly from unamplified cDNA library and were amplified by PCR to check the percentage of recombinants and the insert size of the recombinants. The results showed the percentage of recom-binants was 93.75%, the cDNA inserts ranged from 0.3 to 2 kb (Figure 4), with an average size of 700 bp. The results indicated that the quality of cDNA library should be sufficient to identify the expressed genes in yellow-fruit ginseng.

A total of 400 randomly selected clones were sequenced from the library. Of these, 378 high quality sequences were obtained after deletion of the vector sequences and sequences with short base pairs. The obtained 378 EST sequences were submitted to GeneBank (accession No: ES672876-ES673253) and dbEST (ID: 46881386 46881763). Through BlastX analysis, 221ESTs showed significant similarities to gene sequences in Nr database and were known genes, 21 ESTs were non-significance and unknown function genes, and 136 ESTs, which have no matched were considered novel genes in *P. giseng*. The results of BLASTX search and annotation are shown in Table 1. The generated EST were categorized using gene ontology terms as shown in Table 2, which provide a structured vocabulary to describe a sequence according to its cellular component status, molecular function and biological process. Most EST appeared to be related to physiological and cellular processes.

In conclusion, we described the construction of a cDNA library from yellow-fruit ginseng leaf tissue, EST analysis and gene ontology. The results of analysis of 378 cDNA

Table 1. List of BlastX searched and annotated expressed sequence tags from yellow-fruit ginseng leaf cDNA library.

Accession oo.	Identity (%)	Score (bit)	E-value	Annotation	No of EST
ES672914	92	310	1.00E-83	(S)-2-hydroxy-acid oxidase (EC 1.1.3.15) - cucurbit	1
ES673235	87	61.2	2.00E-10	19S proteasome regulatory complex subunit S6A	1
ES673125	81	66.6	3.00E-10	1-aminocyclopropane-1-carboxylate oxidase	1
ES673214	80	279	3.00E-74	1-deoxyxylulose 5-phosphate synthase	1
ES672956	86	303	4.00E-81	60S ribosomal protein L13	1
ES673025	97	186	3.00E-46	60S ribosomal protein L37a	1
ES672934	79	183	3.00E-45	GTP binding	1
ES673044	84	183	2.00E-45	AT-HF	1
ES672970	100	253	1.00E-66	ATP synthase beta subunit	1
ES673152	93	63.2	8.00E-12	ATP synthase CF-0 subunit I	1
ES672877	100	174	6.00E-43	ATPase epsilon subunit	3
ES672903	73	227	7.00E-59	putative AtRer1A protein	1
ES672922	83	150	1.00E-35	auxin response factor 3	1
ES672960	68	184	1.00E-45	blight-associated protein p12 precursor	2
ES673234	78	241	5.00E-63	carbonate dehydratase/ zinc ion binding	1
ES673212	94	133	1.00E-30	CAB-like protein [Ipomoea nil]	1
ES673022	71	163	1.00E-39	calcium ion binding	2
ES673029	60	156	4.00E-37	calcium-binding protein	1
ES672945	48	98.2	7.00E-20	Ribonuclease Mc1	1
ES672901	67	151	1.00E-35	chitinase-like protein	1
ES672917	99	275	8.00E-73	chlorophyll a/b binding protein	3
ES673014	99	295	4.00E-79	chlorophyll a/b binding protein of LHCII type I precursor	5
ES673007	94	193	2.00E-48	Chlorophyll a-b binding protein 13, chloroplast precursor	1
ES673114	92	112	9.00E-49	Chlorophyll a-b binding protein, chloroplast precursor	1
ES672969	90	134	9.00E-31	chloroplast hypothetical protein	2
ES672953	64	217	2.00E-55	chloroplast oxygen-evolving enhancer protein	2
ES673194	86	304	1.00E-81	chloroplast pigment-binding protein CP24	1
ES673148	56	75.1	7.00E-13	conserved hypothetical protein	1
ES672983	70	213	2.00E-54	CONSTANS-like protein	1
ES673055	78	132	5.00E-30	copper chaperone	1
ES672886	70	270	7.00E-72	CPD photolyase	1
ES673171	89	296	3.00E-79	cyclophilin	1
ES673041	100	303	1.00E-81	cytoplasmic ribosomal protein S13	1
ES672961	64	132	5.00E-30	dehydrin 4	1
ES673219	76	85.1	7.00E-16	delta 12 oleic acid desaturase FAD2	1
ES673113	67	199	3.00E-50	Desiccation protectant protein Lea14 homolog	2
ES673180	95	316	3.00E-85	DSK2	1
ES673046	77	195	7.00E-49	electron transporter/ thiol-disulfide exchange intermediate	1
ES672999	77	164	1.00E-39	elongation factor 1-beta	1
ES672981	71	115	5.00E-25	enoyl-CoA-hydratase	1
ES673040	71	54.3	1.00E-06	expressed protein	1
ES672915	82	253	9.00E-67	ferritin	1
ES673057	75	253	2.00E-66	galactinol synthase, isoform GolS-1	1
ES673002	100	95.5	5.00E-19	GBR3	3
ES673076	39	84.7	1.00E-15	GBR5	5
ES673205	73	192	3.00E-48	GDSL-lipase protein	1
ES673079	94	105	3.00E-22	geranylgeranyl reductase	1
ES673091	83	60.1	3.00E-08	glossy1 homolog	1

Table 1. contd.

ES672967	88	215	7.00E-55	glutathione peroxidase	1
ES673220	84	66.6	2.00E-10	glyceraldehyde 3-phosphate dehydrogenase	1
ES673215	78	220	2.00E-56	glyceraldehyde 3-phosphate dehydrogenase B subunit	1
ES673075	84	278	5.00E-74	glyceraldehyde-3-phosphate dehydrogenase	1
ES672997	52	109	5.00E-23	Harpin-induced 1	1
ES673176	81	171	1.00E-41	heat-shock protein 80	1
ES673042	46	145	9.00E-34	histone acetyltransferase complex component	1
ES673020	75	124	1.00E-27	hypothetical protein	1
ES672963	67	164	2.00E-39	hypothetical protein	1
ES673016	47	65.9	4.00E-10	hypothetical protein Afu4g09870	1
ES673132	85	52	1.00E-07	hypothetical protein CIMG_06034	1
ES673048	92	123	4.00E-27	hypothetical protein MtrDRAFT_AC151668g11v1	17
ES673230	92	106	2.00E-37	hypothetical protein MtrDRAFT_AC151668g27v1	1
ES673043	88	131	8.00E-30	hypothetical protein OeelhCp020	2
ES673035	81	91.3	1.00E-17	hypothetical protein OeelhCp021	1
ES673010	97	169	4.00E-41	hypothetical protein PhapfoPp090	2
ES673039	63	121	1.00E-26	hypothetical protein SNOG_04123	1
ES673162	73	77	2.00E-13	late embryogenis abundant protein 5	1
ES673151	99	340	3.00E-92	light harvesting chlorophyll a /b binding protein	7
ES673197	88	167	2.00E-40	Single-stranded DNA-binding protein	1
ES672906	58	57.4	1.00E-07	low temperature and salt responsive protein	1
ES673232	78	206	1.00E-52	ly200 protein	1
ES673061	56	151	9.00E-36	Major sperm protein	1
ES673185	74	243	2.00E-63	malate dehydrogenase	1
ES673017	43	62	1.00E-08	metal ion binding	1
ES672955	75	86.7	4.00E-16	metallothionein-1 like protein	1
ES672883	82	82.4	6.00E-15	mRNA-binding protein precursor	1
ES673068	61	77.4	1.00E-13	NAK-type protein kinase	1
ES673078	64	136	7.00E-31	nonspecific lipid transfer protein 1	2
ES672938	74	143	2.00E-33	O-GlcNAc-transferase-like protein	1
ES673139	93	61.2	1.00E-08	putative GMPase	1
ES673119	86	193	2.00E-54	ALM beta-like	1
ES673000	50	158	8.00E-38	putative adapter-related protein complex 4 epsilon 1 subunit	1
ES673030	78	53.1	3.00E-06	putative lipase	1
ES673155	61	147	2.00E-34	early flowering 4	1
ES673242	50	62.4	6.00E-16	hypothetical protein	1
ES673141	76	75.1	9.00E-13	oxidoreductase	1
ES673071	94	296	1.00E-79	oxygen evolving complex 33 kDa photosystem II protein	1
ES673240	39	59.7	3.00E-08	16 kDa protein of the photosynthetic oxygen- evolving protein	1
ES673161	89	54.7	1.00E-06	PAP fibrillin	1
ES673051	62	75.5	6.00E-13	peptidase/ threonine endopeptidase	1
ES673237	88	270	1.00E-71	permease	1
ES673033	90	117	1.00E-25	peroxiredoxin Q	1
ES673054	35	58.9	7.00E-08	phloem protein 2-1	4
ES672921	39	66.2	5.00E-10	phloem protein 2-2	9
ES673208	89	238	7.00E-62	Phosphoribulokinase	1
ES673053	100	183	1.00E-45	photosynthetic electron transfer-like protein	1
ES673207	80	196	2.00E-49	photosystem I subunit XI	1
ES673104	45	75.1	1.00E-12	photosystem II	1
ES672994	71	168	2.00E-41	photosystem II 10 kDa protein	3

Table 1. contd.

ES673116	41	72.8	5.00E-12	Photosystem II 5 kDa protein, chloroplast precursor (PSII-T)	1
ES673227	100	56.6	3.00E-07	photosystem II CP47 protein	1
ES673177	100	39.7	3.00E-06	photosystem II M protein	2
ES673239	87	200	1.00E-50	photosystem II protein D1	2
ES673252	94	105	7.00E-22	photosystem II protein K	1
ES673146	63	125	4.00E-28	plastidic aldolase NPALDP1	1
ES672927	52	95.1	1.00E-18	plastoquinol-plastocyanin reductase	2
ES673221	92	140	2.00E-32	poly(A)-binding protein	2
ES672984	84	79	5.00E-14	Polygalacturonase-1 non-catalytic subunit beta precursor (AroGP1)	1
ES673015	81	184	2.00E-45	protein binding / ubiquitin-protein ligase/ zinc ion binding	1
ES672980	84	144	2.00E-33	protein disulfide isomerase	1
ES673203	86	59.7	4.00E-08	PsbC	1
ES673005	98	178	4.00E-44	PSI 9 kDa protein	1
ES673027	68	53.5	2.00E-06	PSI-H precursor	1
ES672992	100	75.5	7.00E-13	PSII 44 kDa protein	1
ES672925	100	96.3	3.00E-19	PSII K protein	1
ES672896	52	112	8.00E-24	structural constituent of ribosome	1
ES672978	72	221	2.00E-56	putative 60S ribosomal protein L7-like protein	1
ES673059	82	232	6.00E-60	Putative auxin efflux carrier component 8 (AtPIN8)	1
ES673135	91	161	9.00E-39	putative chlorophyll A-B binding protein of LHCI type II precursor	1
ES673200	98	115	4.00E-25	putative chloroplast chlorophyll A-B binding protein type I	1
ES672923	94	77.4	1.00E-13	putative chloroplast thiazole biosynthetic protein	1
ES672946	76	157	3.00E-37	putative E2, ubiquitin-conjugating enzyme UBC7	1
ES672943	41	93.6	2.00E-18	putative F-box and leucine-rich repeat protein	1
ES673154	76	213	2.00E-54	putative L24 ribosomal protein	1
ES673236	61	140	2.00E-32	putative phosphatidylcholine-sterol acyltransferase	1
ES672880	80	169	4.00E-41	putative phosphatidylinositol- phosphatidylcholine transfer protein SEC14	1
ES673217	74	187	7.00E-47	putative photosystem I reaction centre PSI-D subunit precursor	2
ES673011	76	172	5.00E-42	putative rubisco subunit binding-protein alpha subunit	2
ES673112	54	164	1.00E-39	receptor-like protein kinase homolog RK20-1	1
ES673244	81	202	5.00E-51	Ribonuclease 1	3
ES672907	100	181	7.00E-45	ribosomal protein L16	1
ES672876	83	106	3.00E-22	Ribosomal protein S30	1
ES673003	78	114	1.00E-24	Ribosomal protein S5, bacterial and organelle form	1
ES672940	94	99.8	4.00E-20	ribosomal protein small subunit 28	2
ES673123	94	298	5.00E-83	ribulose-1,5-bisphosphate carboxylase/oxygenase activase	2
ES673006	98	342	3.00E-93	ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	3
ES673175	45	135	5.00E-31	RNA-binding protein-like	1
ES673201	90	80.1	3.00E-14	rRNA intron-encoded homing endonuclease	1
ES673065	93	67.8	9.00E-16	S-adenosylmethionine synthetase	1
ES673173	93	69.3	1.00E-15	selenium binding	1
ES672919	88	209	3.00E-53	small GTP-binding protein	1
ES672924	75	88.6	8.00E-17	sorbitol related enzyme	1
ES673074	72	102	5.00E-21	structural constituent of ribosome	2
ES673143	60	94.7	1.00E-18	T-complex protein 1 epsilon subunit	1
ES673064	46	103	3.00E-25	TGF-beta receptor, type I/II extracellular region	1
ES672993	49	118	1.00E-25	thioredoxin H	2
ES673225	74	150	1.00E-35	Thioredoxin H-type (TRX-H) emb|CAA94534.1| thioredoxin	1
ES672979	60	128	6.00E-29	TIR-NBS disease resistance-like protein [Populus trichocarpa]	1

Table 1. contd.

ES672944	80	224	1.00E-57	tonoplast intrinsic protein	1
ES673110	63	156	4.00E-37	vacuolar ATPase subunit E-like protein	1
ES673164	76	80.9	2.00E-14	type 2 metallothionein	1
unknown					21
no matched					136
Total					378

Table 2. Gene ontology of expressed sequence tags from yellow-fruit ginseng leaf cDNA library.

Gene ontology term	No. of genes
Cell component	
Cell	46
Cell part	46
Organelle	20
Organelle part	8
Protein complex	26
Molecular function	
Antioxidant activity	1
Binding	30
Catalytic activity	25
Enzyme regulator activity	2
Molecular function unknown	2
Obsolete molecular function	5
Signal transducer activity	1
Structural molecule activity	13
Translation regulator activity	1
Transporter activity	8
Biological process	
Cellular process	65
Physiological process	75
Regulation of biological process	2
Response to stimulus	4

demonstrated EST sequencing and data analysis as a useful and efficient approach to identifying novel genes and for functional genes expression, which would help us understand the mechanisms of ginseng plants during development stages and enrich our knowledge of functional genomic research in ginseng.

ACKNOWLEDGEMENT

This work was supported by a grant from State Forestry Administration, P.R. China (010-413255).

REFERENCES

Choi DW, Jung JD, Ha YI, Park HW, In DS, Chung HJ, Liu JR (2005). Analysis of transcripts in methyl jasmonate-treated ginseng hairy roots to identify genes involved in the biosynthesis of ginsenosides and other secondary metabolites Plant Cell Rep. 23: 557-566

Kim MK, Lee BS, In JG, Sun H, Yoon JH, Yang DK (2006). Comparative analysis of expressed sequence tags (ESTs) of ginseng leaf. Plant Cell Rep. 25: 599-606

Miyahara T, Hirono I, Aoki T (2000). Analysis of expressed sequence tags from a Japanese eel *Anguilla japonica* spleen cDNA library. Fish. Sci. 66: 257-260

Sun YL, Xue WZ (2001). Study on the main chemical constituents in ginseng. Chin. J. Health Lab. Technol. 5: 555-556. (in chinese.)

Wang YC, Yang CP, Liu GF, Jiang J, Wu JH (2006). Generation and analysis of expressed sequence tags from a cDNA library of Tamarix androssowii. Plant Sci. 170: 28-36

Yang CJ, Wang J, Zhou L, Liu GJ (2008).Extraction of total RNA from Panax ginseng C.A. Meryer. cv. Hongguo leaves. Biotechnol. Bull. 1: 136-139. (in chinese.)

Zhao SJ, Liu YZ, Zhao YH, Li FY, Huang ZH, Wu LJ, Guo J, Liu JY (1998). Comprehensive evaluation on characters of Jilin yellow-fruit ginseng. Special wild economic animal plant res. 4: 1-6. (in chinese.)

Zheng YL, Zhang CX, Li XG (2001). Indicators and methods of quality evaluation to Jilin ginseng. Ginseng res. 2: 12-14. (in chinese.)