Journal of Plant Biology, October 2006, 49(5) : 364-370

Identification and Characterization of Highly Expressed Genes in Suspension-Cultured Cells of Sweet Potato

Young Hwa Kim[1,2], Cheol Goo Hur[3], Yun Hee Shin[3], Jung Myung Bae[4], Young Sun Song[2], and Gyung Hye Huh[1]*

[1]College of General Education, Inje University, Gimhae 621-749, Korea
[2]School of Food and Life Science, BPRC, Inje University, Gimhae 621-749, Korea
[3]Plant Genome Laboratory, Genome Research Center, KRIBB, Daejeon 305-333, Korea
[4]School of Life Sciences and Biotechnology, Korea University, Seoul 136-701, Korea

Cultured plant cells have high potential in providing efficient and low-cost molecular farming systems for the large-scale production of commercially valuable recombinant proteins. As an initial aim at establishing an efficient expression system for suspension-cultured cells of food crops, we first obtained 1411 expressed sequence tags from a sweet potato cDNA library of exponential phase cells, and assembled them into 156 contigs and 1039 singletons. Five ESTs were selected as the most significantly abundant genes. These were transcripts for a cell wall development protein (cinnamyl alcohol dehydrogenase), a carbohydrate metabolite protein (glyceraldehyde-3-phosphate dehydrogenase), and a cell cycle regulator protein (small GTP binding protein Ran), as well as inorganic pyrophosphatase and cyclophilin. Comparisons were then made with the root and leaf EST libraries of sweet potato. Northern blot and RT-PCR analyses revealed that these five genes were strongly expressed in the suspension cells but not in the roots and leaves, thereby supporting the data obtained from the comparative analysis. This is the first reported comparison of the various EST libraries isolated from different cell types of the same plant species. These genes can now contribute to an applicable promoter for devising an efficient expression system from suspension-cultured cells.

Keywords: cinnamyl alcohol dehydrogenase, expressed sequence tags, glyceraldehyde-3-phosphate dehydrogenase, small GTP binding protein Ran, suspension-cultured cells, sweet potato

Cultured plant cell systems are being actively studied for their potential as low-cost and convenient 'farming factories' for the large-scale commercial production of valuable recombinant proteins. Plant cells, like microbes, are easy to maintain and grow on inexpensive synthetic media and, because they are higher eukaryotes, they can synthesize complex molecules and carry out many of the post-translational modifications that occur in human cells. Plant cell-suspension cultures are also intrinsically safe and require relatively simple and inexpensive downstream processing schemes for product recovery. Suspension cells are also the most amendable to Good Manufacturing Practice (GMP) procedures, and many industrial-based systems, such as large-scale bioreactors, are now efficiently producing recombinant proteins (Doran, 2000; Twyman et al., 2003; Fischer et al., 2004; Hellwing et al., 2004).

However, although these plant cell factories are capable of generating recombinant proteins derived from animals and plants, biopharmaceutical proteins have not yet been produced commercially using plant cell-suspension cultures. One of the most important factors determining the commercial viability of molecular farming is an adequate yield of the recombinant protein being harvested. Absolute yields depend on the species and expression system; an example of the latter is the design of the construct for expressing a particular recombinant protein. The choice of promoter, leader peptide, and enhancer are all critical aspects of this construct design. Therefore, analyzing the database expressed sequence tags (dbESTs) of suspension-cultured cells, as a step towards developing an effective molecular farming system in plants, would provide very useful data.

Rapidly growing cultured cells of the tobacco BY-2 cell line are currently the most popular hosts for recombinant protein production. Other plant suspension cultures, including rice, have also been used, partly because they may prove more favorable than tobacco in terms of product levels and because they are derived from food crops (Torres et al., 1999; Smith et al., 2002; Kwon et al., 2003; Shin et al., 2003; Hellwing et al., 2004). Consequently, choosing the most appropriate expression system and cell line for the commercial production of biomolecules must be done on a case-by-case basis. To develop an efficient expression system in suspension-cultured cells, it is necessary to identify highly abundant and cell culture-specific genes. However, no reports have yet been made of the comparative analysis of ESTs produced by different cell types and/or organs within a single plant. Matsuoka et al. (2004) recently have examined the dbESTs of cultured BY-2 cell lines with the aim of determining, via microarray analysis, any global changes in gene expression. The objective of that study was to identify genes that show elevated levels during the growth phase compared with other phases. They also have compared the expression pattern of BY-2 genes with that of genes

*Corresponding author; fax +82-55-339-3734
e-mail igehuh@inje.ac.kr

Abbreviations: CAD, cinnamyl alcohol dehydrogenase; DAS, days after subculture; dbEST, database expressed sequence tag; EST, expressed sequence tag; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; MIPS, Munich Information Center for Protein Sequences; NR protein, non-redundant protein

expressed in cultured *Arabidopsis* root apex cells (Birmbaum et al., 2003), and have found that log-phase BY-2 cells exhibit a pattern of gene expression similar to that of cultured root meristematic cells. Their data analyses have focused on identifying genes with significantly different expression patterns at each growth phase. Still to be investigated are the dbEST libraries from suspension-cultured cells of *Solanum tuberosum*, *Zea mays*, *Arabidopsis thaliana*, and *Lycopersicon esculentum* (http://www.ncbi.nlm.nih.gov/UniGene).

The aim of our investigation described here was to identify specific genes that are highly expressed in suspension-cultured cells of sweet potato (*Ipomoea batatas* Lam. cv. White Star), using large-scale, single-pass cDNA sequencing with cDNA clones at the exponential phase. The ESTs that were generated were then analyzed with BLASTX against the NCBI EST database, and compared against the sweet potato root and leaf EST database in order to mine highly expressed genes in suspension-cultured cells.

MATERIALS AND METHODS

Plant Material and Cell Culture

We used the SP-47 cell line of sweet potato suspension cultures [*I. batatas* (L.) Lam. cv. White Star] (Kim et al., 1994). Cells (1 g fresh weight) were sub-cultured at 14-d intervals and inoculated into 50 mL of MS basal medium (Murashige and Skoog, 1962) supplemented with 1 mg L^{-1} of 2,4-dichlorophenoxyacetic acid and 30 g L^{-1} of sucrose. They were then cultured at 25°C in the dark (100 rpm). Our previous collection had enabled us to determine the cell growth stage of sweet potato suspension-cultured cells (Kim et al., 2004). For the northern blot analysis, different stages of the cells were used, sampled at 5, 10, 15, and 20 d after subculture (DAS). The log phase cells (15 DAS) were used to make a cDNA library.

cDNA Library Construction and Sequencing

To construct the cDNA library, total RNA was extracted from suspension cultures of 15 DAS cells, using TRIzol® reagent (Invitrogen, USA). Poly (A) RNA was prepared with the polyA Track mRNA isolation system (Promega, USA). Our cDNA library was constructed with a uni-ZAP cDNA Synthesis kit and the ZAP-cDNA Gigapack® III Gold packing extract (Stratagene, USA), according to the manufacturer's instructions. After *in vivo* mass-excision of the library, plasmid DNA from 1700 randomly selected colonies was prepared and subjected to fluorescence cycle sequencing, using the ABI Big Dye Cycle Sequencing kit (PE Applied Biosystems, USA) and the T7 primer (5'-GTAATACGACTCACTATAGGG-3'). The reactions were run and analyzed on an ABI Prism 3700 DNA analyzer (PE Applied Biosystems).

Functional Classification and Comparative Analysis

Plasmid DNA was automatically prepared from more than 1700 cDNA clones, and 5'-end sequences of 1626 cDNA clones were obtained using the T7 primer. To convert those EST sequences from ABI files and base calling, we used the Phred package (http://www.phrap.org) (Ewing et al., 1998), with a trim cutoff value of 0.05 for selecting high-quality sequences. The vector sequences were first masked by cross-matching (minmatch = 12, minscore = 20) followed by removal of the poly A tail and linker sequences. Sequences longer than 100 bp (excluding vector sequences) were selected, for a total of 1411 high-quality ESTs.

For the comparative analysis, a file of 3159 ABI-formatted EST sequences from a sweet potato early storage root library was provided by Dr. Jung Myung Bae of Korea University (You et al., 2003). In addition, 1079 EST sequences from the sweet potato leaf library were downloaded from the NCBI dbEST (CB329881-CB330959). These were then processed in the same way as for the library of ESTs generated from our suspension cell cultures. After screening, the sequences were clustered using the software package StackPACK v2.1 (Christoffels et al., 2001), which compared the sequences and grouped them into clusters and singletons (i.e., ungrouped ESTs). The d2 cluster function (Hide et al., 1994; Burke et al., 1999; Christoffels et al., 2001) was applied as a clustering algorithm with option values (word size = 6, similarity cutoff = 0.96, minimum sequence size = 50, window size = 100, reverse comparison = 1). The sequences in the loosely grouped clusters were further aligned and analyzed to obtain contigs using the Phred/Phrep function, with a trim score of 20 (Burke et al., 1999; Miller et al., 1999). The contigs and singletons that were obtained from this StackPACK analysis were used as the query in BLASTX searches against the MIPS functional catalog of *A. thaliana* (http://mips.gsf.de/) (Mewes et al., 2002), with an E-value $<1.0\ E^{-10}$. The five highest ranked matches were then retrieved for annotation. Contigs and singletons were assigned to MIPS functional categories by multiple annotation. To increase the accuracy of annotation by homology search (Hide et al., 1994), we constructed a 'plant NR database' by selecting plant genes from the NR database of NCBI. BLASTX was performed with an E-value $<1.0\ E^{-10}$, and the five highest ranked matches were retrieved for annotation.

Northern Blot Analysis

Ten μg of total RNA were denatured, separated by electrophoresis, and transferred onto a Tropilon-Plus' nylon membrane (Tropix, USA). The probes were biotin (biotin-14-dCTP, Invitrogen) labeled by PCR amplification using cDNA as template. PCR was performed in a 20 μL volume containing 1.25 U Ex *Taq* DNA polymerase (Takara, Japan), 2 μL of 10× *Taq* buffer, 4 μL of 5× dNTPs mix (0.25 mM biotin-14-dCTP, 0.25 mM dCTP, 0.5 mM dATP, 0.5 mM dGTP, and 0.5 mM dTTP), and 10 pMol of the T3 and T7 primers. Conditions included the following: 94°C for 5 min; then 30 cycles of 30 s at 94°C, 30 s at 60°C, and 60 s at 72°C; and a final extension step of 7 min at 72°C. Labeled probes were purified using a PCR Purification kit (Qiagen, USA), according to the manufacturer's instructions. The membrane was hybridized at 65°C for 16 h, then washed twice with 2× SSC/1% SDS at room temperature (RT) for 10 min, twice with 0.1× SSC/1% SDS at 65°C for 20 min, and twice with 1× SSC at RT for 10 min. Hybridized signals were detected using the Southern-Star™ System (Tropix).

Reverse Transcription (RT)-PCR

Semi-quantitative RT-PCR was conducted to analyze the expression pattern of the five most abundant genes from our suspension-cultured cells, as well as the young storage roots and leaves. Reverse transcription was performed using oligo $(dT)_{20}$ (Invitrogen) as the primer and SuperScript™ II reverse transcriptase (Invitrogen). Aliquots consisting of one-tenth of the RT product were used as template for each amplification. PCR primers were designed for gene-specific amplification on the 5′- and 3′-UTRs of the genes. Sequences for the forward (F) and reverse (R) primers included: CAD (C0500393): F, 5′-ATCTTGATTGTCTCAATCTA-3′; R, 5′-GGACATTATTACATTACAC-3′, Ran GTPase (C0500253): F, 5′-GACCTAACCAAGCAACGACG-3′; R, 5′-CAAGTACTCG-TGTAAGTACC-3′, GAPDH (C0500859): F, 5′-TCTAGAA-GCTTCAAGCCTCC-3′; R, 5′-GCTGCCAATATGTTGGAGGC-3′, inorganic pyrophosphatase (C0500483): F, 5′-AACAAAG-CTCCCAAGGCCAT-3′; R, 5′-GAATATATCATCCACCACCC-3′, cyclophilin (C0500766): F, 5′-CCGTATAGTGATGGAGT-TATTC-3′; R, 5′-GGTTCAGTAATCAGGCGATG-3′, and tubulin (quantitative control): F, 5′-CAACTACCAGCCACCAACTGT-3′; R, 5′-CAAGATCCTCACGAGCTTCAC-3′. PCR was carried out in a 20 μL reaction mixture containing 0.2 mM of each dNTP, 1 pMol of each forward and reverse primer, and 0.125 U of Super *Taq* polymerase (SUPER-Bio, Korea), including the cDNA products. After an initial denaturation at 94°C for 5 min, amplification was carried out at 94°C for 30 s, 50-60°C for 1 min, and 72°C for 1 min, followed by an incubation for 7 min at 72°C. The PCR cycles numbered 25 for CAD; 20 for inorganic pyrophosphatase; 23 for cyclophilin, Ran, and GAPDH; and 20 for tubulin. The amplified PCR products were separated on a 1% agarose gel.

RESULTS

Construction of the cDNA Library and Cloning of Suspension-Cultured Cell ESTs from Sweet Potato

We previously determined the cell growth stage of sweet potato suspension-cultured cells, based on the fresh weight of the cultures and their protein levels (Kim et al., 2004). The cDNA library for the present investigation was constructed with mRNA obtained from 15-d-old suspension cultures of sweet potato cells equivalent to those used previously. To determine the average size of the insert DNA, we randomly picked 20 plaques for PCR amplification with the T7 and T3 primers. Sizes ranged from 0.5 to 5.0 kbp. Plasmid DNAs were prepared in an automated system from 1700 cDNA clones and sequenced from the 5′-end. In the initial step, 5′-end sequences of 1626 cDNA clones were obtained, and the sequencing process then gave rise to 1411 high-quality ESTs. The average read length of these ESTs, after vector-trimming and removal of the low-quality sequences, was 593 bp. These 1411 sweet potato EST sequences obtained from the log phase suspension-cultured cells have now been submitted to the dbEST and GenBank database under the codes 24458268-24459678 and C0499854-C0501264, respectively.

Distribution of ESTs

The 1411 sweet potato ESTs were clustered using the StackPack software package to classify the redundant ESTs as well as the singletons. In all, 156 independent contigs from 372 ESTs and 1039 singletons were generated. These data suggest that about 84.7% of the total suspension cell ESTs represent independent sweet potato genes. The contigs and singletons were primarily analyzed against the non-redundant (NR) protein database of GenBank. To increase the accuracy of annotation by homology search (Spang and Vingron, 2000), we constructed a 'plant NR database' by selecting plant genes from the NCBI NR database. Of the ESTs obtained from the suspension cells, 73.4% showed significant (E-value $<1.0\ E^{-10}$) similarity to the translation products in the NR database. Despite this significant similarity, however, 7.5% of these were classified as "function unknown". The remainder (26.6%) were classified as 'no hits'.

As a first step to determine which ESTs are highly expressed and specific to cultured cells of sweet potato, we processed the 3159 ESTs from the early storage root library (R-library) and the 1079 ESTs from the leaf library (L-library) in the same manner as those generated from the suspension cells (S-library). The R- and L-libraries were clustered separately. Figure 1 shows the distribution of contigs and singletons

A

Sequence source	S-library	R-library	L-library	Total
Total number of sequences analyzed	1,411	3,054	1,079	5,541
Number of ESTs in contigs	372	2,592	580	3,774
Number of contigs	156	464	156	803
Number of singletons	1,039	462	499	1,770
Number of unique sequences	1,193	926	655	2,573
Redundancy (%)	26.4	84.9	53.8	68.1

B

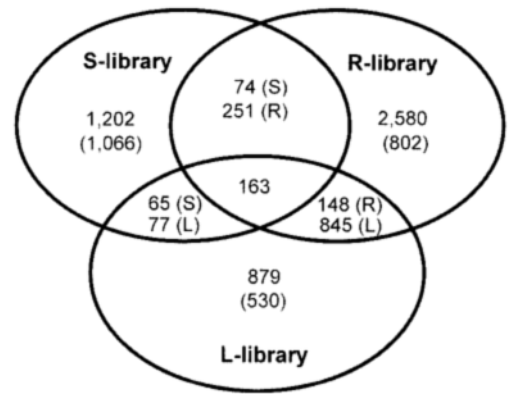

Figure 1. Distribution of contigs and unique ESTs among EST libraries of sweet potato suspension cells (S), roots (R), and leaves (L). (**A**) Summary of assembled results. Sequences obtained from three libraries were aligned both separately and as a group using Phred/Phrap function of StackPack software package. Redundancy was calculated as 100 × (number of ESTs assembled in contigs/total number of ESTs). (**B**) Distribution of ESTs among three EST libraries. Values in overlapped areas within each library indicate numbers of ESTs present in more than one library. Of 163 ESTs occurring in all three libraries, 50 came from S-library, 75 from R-library, and 38 from L-library. Values in parentheses represent number of unique genes in each library.

among these three libraries. The S-library yielded 1193 unique sequences, at a redundancy of 26.4% [100 × (number of ESTs assembled in contigs/total number of ESTs)]. In contrast, the R- and L-libraries yielded 926 and 655 unique sequences, at redundancy rates of 84.7 and 53.8%, respectively. A greater number of unique sequences were obtained from the S-library due to its lower redundancy. A total of 2573 unique sequences were obtained from the collective database of all three libraries.

Functional Annotation and Comparative Analysis of Sweet Potato ESTs from Suspension-Cultured Cells, Roots, and Leaves

The ESTs of each of the three libraries (S, R, and L) were classified into functional categories assigned by the Munich Information Center for Protein Sequences (MIPS) (http://mips.gsf.de) for *A. thaliana* (Table 1). This classification was based entirely on interference from BLASTX reports, such that some ESTs were placed into two or three functional categories. Patterns of gene expression were similar between the S- and L- libraries. However, for three MIPS categories, large differences in their patterns were found between the S-/L-libraries and the R-library. In the latter, a significantly higher percentage of ESTs fell into the categories of energy, protein synthesis, and protein activity regulation. When we compared functional subcategories among the libraries, the number of ESTs assigned to the categories of metabolism, and cell cycle and DNA processing differed significantly, with the number of ESTs encoding genes related to the cell cycle subcategory being higher in the S-library. In the category of metabolism, genes related to amino acid metabolism were more highly expressed in the suspension-cultured cells, while those associated with C-compound and carbohydrate metabolism were more highly expressed in the roots and leaves. These results reflect the inherent characteristics of each library, with the S-library ESTs having been derived from the exponential stage of suspension-culture cells, the L-library ESTs from photosynthetic tissue, and the R-library from early storage tissue.

Highly expressed genes from each library were analyzed, with the number of ESTs in a contig providing an estimate of the expression level for each gene. The most significantly abundant ESTs in the S-library were transcripts for cell-wall development protein (cl43ct46cn47; cinnamyl-alcohol dehydrogenase, CAD), carbohydrate metabolite protein (cl119ct126cn128; glyceraldehyde-3-phosphate dehydrogenase, GAPDH), cell cycle regulator protein (cl77ct83cn85; small GTP binding protein Ran), including inorganic pyrophosphatase (cl102ct109cn111), cyclophilin (cl11ct13cn14), and aquaporin (cl32ct34cn35) (Table 2). The most abundant ESTs in the S-library, however, were very low or absent in the L- and/or R- libraries, although the number of metallothionein ESTs was higher in the R-library. This indicates that the most abundant ESTs in the S-library are unique or highly expressed genes relative to the L- and R-libraries.

The most abundant ESTs in the young storage roots of sweet potato were transcripts for the GCN5-related N-acetyltransferase (cl6ct81cn83, 224 ESTs) gene, which regulates cell growth and development, as well as the xyloglucan endo-transglycosylase (cl106ct139cn141, 40 ESTs) gene,

Table 1. Distribution of ESTs into MIPS functional categories for each of the libraries[a,b].

MIPS No.	Functional category	% EST in the library		
		Suspension cells	Leaves	Roots
1	Metabolism	10.7	12.3	14.6
1.01	Amino acid metablism	2.1	1.2	1.6
1.06	C-compound and carbohydrate metabolism	4.5	10.4	7.6
2	Energy	4.8	5.1	12.8
10	Cell cycle and DNA processing	3.8	2.8	1.1
10.01	DNA processing	1.1	1.1	0.3
10.03	Cell cycle	2.6	1.4	0.8
11	Transcription	4.7	5.6	5.1
12	Protein synthesis	7.1	5.4	17.9
14	Protein fate (folding, modification destination)	8.2	14.0	8.5
18	Protein activity regulation	1.3	0.8	14.6
20	Cellular transport	4.5	4.1	2.0
30	Cellular communication/signal transduction	4.3	2.6	4.4
32	Cell rescue, defense and virulence	5.9	4.9	1.7
34	Interaction with the cellular environment	2.1	0.9	4.2
36	Interaction with the environment (systematic)	1.7	0.5	1.9
42	Biogenesis of cellular components	6.4	4.3	10.7
70	Subcellular localization	14.9	11.2	23.3
98	Classification not yet clear-cut	9.2	8.3	8.5
99	Unclassified protein	37.7	39.0	43.7

[a]Dual or triple functions were assigned to some ESTs.
[b]Percentages are calculated from the numbers of ESTs in each functional category per total EST numbers of each library to reflect the relative levels of expression of genes in each library.

Table 2. Comparison of the most abundant ESTs of the S-library with their counterparts of the L- and R-libraries. The top 17 contigs containing the most ESTs in the S-library were annotated using a BLASTX search against the MIPS functional catalog of *A. thaliana*. Contigs were generated by the StackPack software package.

Functional category[a]	Contig	Putative function	No. of ESTs			Hit ID of *Arabidopsis*	E-value
			S	L	R		
CW	cl43ct46cn47	cinnamyl-alcohol dehydrogenase (CAD)	12	1	-	at4g37970	3e-62
CM	cl119ct126cn128	glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	9	2	7	at3g04120	e-160
PS	cl11ct13cn14	cyclophilin	7	5	1	at4g38740	3e-76
EN	cl102ct109cn111	inorganic pyrophosphatase	5	-		at3g53620	2e-63
CC/DP	cl77ct83cn85	small GTP binding protein (Ran)	5	1	-	at5g55190	e-129
TF	cl32ct34cn35	putative aquaporin	5	-	-	at4g01470	9e-92
DT	cl2ct3cn3	metallothionein	5	2	10	at3g09390	9e-17
UC	cl3ct4cn4	unknown protein	5	-	1	at4g09580	1e-15
UC	c146ct48cn49	unknown protein	5	-	-	at5g65520	4e-27
RP	cl13ct15cn16	60S ribosomal protein L10A	4	1	-	at2g27530	6e-95
UC	cl1ct1cn1	tumor protein-like protein	4	-	5	at3g16640	1e-68
UC	cl21ct23cn24	unknown protein	4	-	6	–	–
UC	cl33ct35cn36	unknown protein	4	-	-	at1g73010	1e-91
TC	cl34ct36cn3	unknown protein	4	1	1	at3g12390	8e-59
RP	cl42ct44cn45	60S ribosomal protein L7	4	-	-	at2g01250	e-118
AM/DT	cl47ct50cn51	L-ascorbate peroxidase	4	-	4	at1g07890	e-110
CC	cl55ct60cn61	2S proteosome subunit PAE2	4	-	-	at3g14290	4e-81

[a]Functional categories: CW, cell wall; CM, C-compound and carbohydrate metabolism; PS, protein folding and stabilization; EN, energy; CC, cell cycle; DP, DNA processing; TF, transport facilitation; DT, detoxification; UC, unclassified; TC, translational control; RP, ribosomal protein; AM, amino acid metabolism.
'–', No ESTs were observed in the roots and leaves. S, L and R represent suspension cell, leaf and root library, respectively.

which participates in cell wall construction of growing tissues. In the leaf ESTs, the most abundant genes were for ribulose 1,5-biphosphate carboxylase (cl17ct41cn44, 71 ESTs), which encodes the Calvin-cycle enzyme that catalyzes CO_2 fixation during photosynthesis, and the light-harvesting complex II (LHCII) type III gene (cl16ct23cn24, 22 ESTs), corresponding to a major pigment-binding protein in plant chloroplast membranes.

Northern blot analysis of total RNA extracted from the suspension-cultured cells, leaves, and young storage roots was carried out to analyze the expression level of the most abundant ESTs in the S-library. The longest EST clone in each contig was used as a representative probe. Expressions of the top five genes were substantially higher in the suspension cells than in the leaves and roots (Fig. 2A), thereby supporting our conclusions from the comparative analysis. To verify gene-specific expression, 5'- and 3'-UTRs of each gene were amplified from cDNAs and subjected to quantitative RT-PCR (Fig. 2B). Those expression patterns were very similar to the results found with our northern analysis. Two abundant genes corresponding to CAD (C0500393) and Ran GTPase (C0500253) were uniquely expressed in the suspension cells. We also investigated the expression patterns of these five genes during the growth phase of the suspension-cultured cells, using total RNA purified from the same cells as those in our previous study (Kim et al., 2004). These genes were strongly expressed throughout the early log phase (5 DAS) to the late log phase (15 DAS). Expression of CAD (C0500393) increased

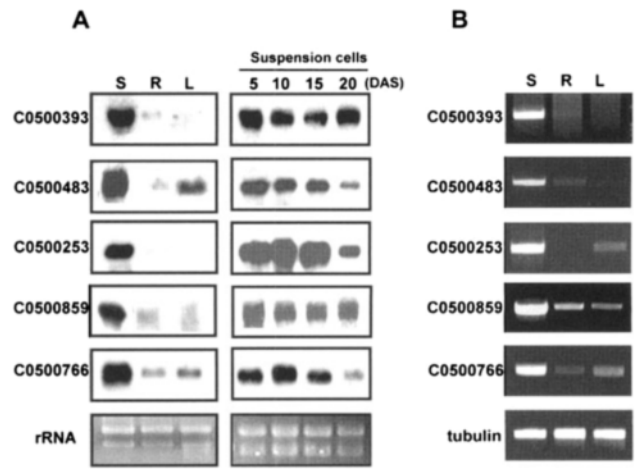

Figure 2. Expression patterns of five most abundant genes. (**A**) Northern blot analysis. Total RNA was isolated from suspension-cultured cells, young storage roots, and leaves. Suspension-cultured cells were collected at different stages during cell growth cycle (Kim et al., 2004). C0500393 (CAD), C0500483 (inorganic pyrophosphatase), C0500253 (Ran), C0500859 (GAPDH), and C0500766 (cyclophilin) are accession numbers from GenBank database. (**B**) Quantitative RT-PCR analysis for different tissue types, with tubulin transcripts amplified as qualitative control. S, suspension-cultured cells (15 DAS); R, young storage roots; L, leaves.

at the stationary phase (20 DAS), while those of Ran GTPase (C0500253), inorganic pyrophosphatase (C0500483), and cyclophilin (C0500766) decreased at that same phase (Fig. 2A).

DISCUSSION

Cultured plant cells may be ideal resources for large-scale production of commercially valuable recombinant proteins. EST sequencing has been widely used as an efficient approach in gene discovery (Suh et al., 2003; You et al., 2003; Kim, 2004). Here, we initially obtained 1411 expressed sequence tags (ESTs) from a sweet potato cDNA library of exponential phase cells. This led to the identification of 1193 putative genes. To determine the most significantly abundant genes in the suspension cells (S), we compared the ESTs of the S-library with those of the sweet potato root (R) (You et al., 2003) and leaf (L) libraries. This comparison of the three libraries proved especially interesting because they represent different cell types and/or organs within a particular plant system. This is the first comparative analysis of different EST libraries from a single species that has focused on the ESTs generated from suspension-cultured cells.

The S- and L-libraries showed similar patterns of gene expression, based on our functional analysis (Table 1). In three categories – energy, protein synthesis, and protein activity – the proportion of ESTs from the R-library differed significantly from that found in the S- and L-libraries. The S-library showed a low level of redundancy (26.4%), and the distribution of the functional categories did not severely skew toward several major categories, as had been the case with the R-library. In contrast, the R-library showed very high redundancy (84.9%), while that of the L-library was 53.8% (Fig. 1). The ESTs of a mature potato tuber, i.e., the storage organ, also have a high degree of redundancy (74.8%) (Crookshanks et al., 2001). Therefore, based on these results, we suggest that a relationship – the extent of which is still unknown – exists between the high level of redundancy and the extreme mal-distribution of expressed genes in the R-library.

Our comparative analysis revealed that five genes (CAD, GAPDH, Ran, inorganic pyrophosphatase, and cyclophilin) were the most abundant and specific in the suspension cells (Table 2). This was supported, as well, by data from our northern analysis and RT-PCR, thereby demonstrating that, for example, the cell-wall development protein CAD is strongly expressed throughout the growth stage of sweet potato suspension cells, especially during the early log (5 DAS) and stationary phases (20 DAS) (Fig. 2A). Furthermore, the cell-cycle regulator protein Ran (Robert and Gruissem, 1994; Li et al., 2003) is strongly expressed during the log phase, from 5 DAS to 15 DAS. Our results are in agreement with those of Matsuoka et al. (2004) concerning tobacco BY-2 cells, in which the genes related to cell-wall synthesis are more highly expressed during the log and stationary phases while those involved in the cell cycle are more highly expressed during the lag and log phases.

The most abundant gene, CAD (C0500393), is an important enzyme in the lignin biosynthetic pathway. Lignin is a major component of secondary cell walls. Because the cell wall is one of the determinants of plant cell morphology, one should expect the organizational genes related to cell-wall synthesis to be expressed during the cell growth and expansion phases. In monocots and dicots, CAD activity is regulated developmentally (Sauter and Kende, 1992) as well as by abiotic and biotic stresses (Mitchell et al., 1994). GAPDH (C0500859), the second most abundant gene, also shares 88% identity with an *Oryza sativa* GAPDH at the amino acid level. This NAD-dependent enzyme is involved in glycolysis, as shown by the "C" designation. Therefore, the genes that encode CAD, GAPDH, Ran, inorganic pyrophosphatase, and cyclophilin, and which are highly expressed in the suspension-cultured cells of sweet potato, will be good candidates for use as a strong promoter in developing an efficient expression system suitable for applications in industrially sized plant cell bioreactors.

ACKNOWLEDGEMENTS

This work was supported by a Korea Research Foundation Grant (KRF-2003-041-F20019), and by grants from the BioGreen 21 Program funded by the Rural Development Administration, Korea, and the Plant Signaling Network Research Center, the Korea Science and Engineering Foundation.

Received June 2, 2006; accepted September 4, 2006.

LITERATURE CITED

Birmbaum K, Shasha DE, Wang JY, Jung JW, Lambert GM, Galbraith DW, Benfey PN (2003) A gene expression map of the *Arabidopsis* root. Science 302: 1956-1960

Burke J, Davidson D, Hide W (1999) d2_cluster: A validated method for clustering EST and full length cDNA. Genome Res 9: 1135-1142

Christoffels A, Gelder A, Greyling G, Miller R, Hide T, Hide W (2001) STACK: Sequence tag alignment and consensus knowledgebase. Nucl Acids Res 29: 234-238

Crookshanks M, Emmersen J, Welinder KG, Nielsen KL (2001) The potato tuber transcriptome: Analysis of 6077 expressed sequence tags. FEBS Lett 506: 123-126

Doran PM (2000) Foreign protein production in plant tissue cultures. Curr Opin Biotechnol 11: 199-204

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using Pred. I. Accuracy assessment. Genome Res 8: 175-185

Fischer R, Stoger E, Schillberg S, Christou P, Twyman RM (2004) Plant-based production of biopharmaceuticals. Curr Opin Plant Biol 7: 152-158

Hellwing S, Drossard J, Twyman RM, Fischer R (2004) Plant cell cultures for the production of recombinant proteins. Nat Biotechnol 22: 1415-1422

Hide W, Burke J, Davidson D (1994) Biological evaluation of d2, an algorithm for high performance sequence comparison. J Comput Biol 1: 199-215

Kim DJ (2004) A study of cotyledon senescence in cucumber (*Cucumis sativus* L.) based on expressed sequence tags and gene expression. J Plant Biol 47:244-253

Kim SK, Kwak SS, Jung KH, Min SR, Park IH, Liu JR (1994) Selection of plant cell lines for high yields of peroxidase. Kor Biochem J 27: 132-137

Kim YH, Cho EH, Kwak SS, Kwon SK, Bae JM, Lee BR, Meen BI, Huh GH (2004) Alterations in intracellular and extracellular activities of antioxidant enzymes during suspension culture of

sweet potato. Phytochem 65: 2471-2476

Kwon TH, Kim YS, Lee JH, Yang MS (2003) Production and secretion of biologically active human granulocyte-macrophage colony stimulating factor in transgenic tomato suspension cultures. Biotechnol Lett 25: 1571-1574

Li HY, Cao K, Zheng Y (2003) Ran in the spindle checkpoint: A new function for a versatile GTPase. Trends Cell Biol 13: 553-557

Matsuoka K, Demura T, Galis I, Horiguchi T, Sasaki M, Tashiro G, Fukuda H (2004) A comprehensive gene expression analysis toward the understanding of growth and differentiation of tobacco BY-2 cells. Plant Cell Physiol 45: 1280-1289

Mewes HW, Frichman D, Guldener W, Mannhaupt G, Mayer K, Mokrejs M, Morgenstern B, Munsterkotter M, Rudd S, Weil B (2002) MIPS: A database for genomes and protein sequences. Nucl Acids Res 30: 31-34

Miller R, Christoffels A, Gopalakrishnan C, Burke J, Ptitsyn AA, Broveak TR, Hide W (1999) A comprehensive approach to clustering of expressed human gene sequence: The sequence tag alignment and consensus knowledgebase. Genome Res 9: 1143-1155

Mitchell HJ, Hall JL, Barber MA (1994) Elicitor-induced cinnamyl alcohol dehydrogenase activity in lignifying wheat (*Triticum aestivum* L.) leaves. Plant Physiol 104: 551-556

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

Robert A, Gruissem W (1994) A small nuclear GTP-binding protein from tomato suppresses a *Schizosaccharomyces pombe* cell-cycle mutant. Proc Natl Acad Sci USA 91: 5863-5867

Sauter M, Kende H (1992) Levels of α-glucan and lignin in elongating internodes of deepwater rice. Plant Cell Physiol 33: 1089-1097

Shin YJ, Hong SY, Kwon TH, Jang YS, Yang MS (2003) High level of expression of recombinant human granulocyte-macrophage colony stimulating factor in transgenic rice cell suspension culture. Biotechnol Bioengr 82: 778-783

Smith ML, Mason HS, Shuler ML (2002) Hepatitis B surface antigen (HBsAg) expression in plant cell culture: Kinetics of antigen accumulation in batch culture and its intracellular form. Biotechnol Bioengr 80: 812-822

Spang R, Vingron M (2000) Limits of homology detection by pairwise sequence comparison. Bioinformatics 17: 338-342

Suh MC, Kim MJ, Hur CG, Bae JM, Park YI, Chung CH, Kang CW, Ohlrogge JB (2003) Comparative analysis of expressed sequence tags from *Sesamum indicum* and *Arabidopsis thaliana* developing seeds. Plant Mol Biol 52: 1107-1123

Torres E, Vaquero C, Nicholson L, Sack M, Stoger E, Drossard J, Christou P, Fischer R, Perrin Y (1999) Rice cell culture as an alternative production system for functional diagnostic and therapeutic antibodies. Transgen Res 8: 441-449

Twyman RM, Stoger E, Schillberg S, Christou P, Fischer R (2003) Molecular farming in plants: Host systems and expression technology. Trends Biotechnol 21: 570-578

You MK, Hur CG, Ahn YS, Suh MC, Jeong BC, Shin JS, Bae JM (2003) Identification of genes possibly related to storage root induction in sweet potato. FEBS Lett 536: 101-105