Indian Journal of Experimental Biology
Vol. 51, July 2013, pp. 522-530

Gene ontology based characterization of Expressed Sequence Tags (ESTs) of *Brassica rapa* cv. Osome

Senthil Kumar Thamil Arasan[1,†], Jong-In Park[1,†], Nasar Uddin Ahmed[1], Hee-Jeong Jung[1], In-Ho Lee[1,2], Yong-Gu Cho[3], Yong-Pyo Lim[4], Kwon-Kyoo Kang[5] & Ill-Sup Nou[1,*]

[1]Department of Horticulture, Sunchon National University, 413 Jungangno, Suncheon, Jeonnam 540-742, Republic of Korea

[2]Asia Seed Co. 442-2 Ihwang-ri, Janghowon-eup, Icheon-si, Gyeonngi-do 467-906, Republic of Korea

[3]Department of Crop Science, Chungbuk National University, 410 Seongbongro, Heungdokgu, Cheongju 361-763, Republic of Korea

[4]Department of Horticulture, Chungnam National University, 96 Daehangno, Gung-dong, Yuseong-gu, Daejeon 305-764, Republic of Korea

[5]Department of Horticulture, Hankyong National University, 327 Chungangno, Anseong, Kyonggi 456-749, Republic of Korea

Received 26 December 2012; revised 18 April 2013

Chinese cabbage (*Brassica rapa*) is widely recognized for its economic importance and contribution to human nutrition but abiotic and biotic stresses are main obstacle for its quality, nutritional status and production. In this study, 3,429 Express Sequence Tag (EST) sequences were generated from *B. rapa* cv. Osome cDNA library and the unique transcripts were classified functionally using a gene ontology (GO) hierarchy, Kyoto encyclopedia of genes and genomes (KEGG). KEGG orthology and the structural domain data were obtained from the biological database for stress related genes (SRG). EST datasets provided a wide outlook of functional characterization of *B. rapa* cv. Osome. *In silico* analysis revealed % 83 of ESTs to be well annotated towards reeds one dimensional concept. Clustering of ESTs returned 333 contigs and 2,446 singlets, giving a total of 3,284 putative unigene sequences. This dataset contained 1,017 EST sequences functionally annotated to stress responses and from which expression of randomly selected SRGs were analyzed against cold, salt, drought, ABA, water and PEG stresses. Most of the SRGs showed differentially expression against these stresses. Thus, the EST dataset is very important for discovering the potential genes related to stress resistance in chinese cabbage, and can be of useful resources for genetic engineering of *Brassica* sp.

Keywords: *Brassica rapa* cv. Osome, ESTs, Gene ontology, Stress related genes

The genus *Brassica* includes many vegetable crops, such as broccoli, cabbage, Chinese cabbage, cauliflower, mustard, rape, kale and turnip[1]. Chinese cabbage, (*B. rapa* cv. Osome) is one of the most important vegetable crops and is cultivated widely in Asia, particularly in China, Korea, and Japan. Like many other crop plants, Chinese cabbage is challenged by many stresses including drought, high salt concentrations, low temperatures, pathogens and pests[2,3]. Efforts to develop plants resistant to biotic and abiotic stresses are under way and functional genomics study is one of the most important approaches for identifying potential genes related to stress resistance.

An efficient method of functional genomics is an analysis of the expressed sequence tags (EST). EST represents short, unedited randomly selected single pass sequences derived from the cDNA libraries, which are also known as a 'poor man's' genome and have proven to be valuable in molecular biology[4,5]. This technology is a commonly used approach to identify the genes involved in specific biological functions[6]. Currently, an interesting application of EST sequencing is a study of the gene expression pattern in response to a given environmental stimulus[7] and a trancriptomics study of a plant at various stages of development under different experimental conditions with various plant tissues (leaf stem, flower bud). EST is also used for single nucleotide polymorphism (SNP) characterization, proteomic exploration, CpG Island and promoter analysis[5,8].

Since EST technology was introduced to the human genome project, it has been used widely to

*Correspondent author
Fax: +82-61-750-3208
E-mail: nis@sunchon.ac.kr
†These two authors equally contributed to this work.

clone new genes, determine tissue specific gene expression profiles, annotate functional genome sequences etc,[9-12]. Manual annotation is not possible in annotating high-throughput sequences, such as ESTs. Computational methods are used to process EST sequences, such as sequence cleaning, vector masking, clustering, assembly and annotation to yield biological information to putative sequences. The use of a bioinformatics tool for homology-based functional annotation and statistical information is simple for ESTs from a range of organisms in the public non redundant databases[13,5,14]. Gene ontology (GO) is the most reliable biological annotation schema, in which globally biological communities have an approved structure. The GO project was developed with three structured controlled vocabularies (i.e. ontologies) to describe various genes or proteins in terms of their associated cellular component (CC), biological process (BP) and molecular function (MF)[15].

In this study, EST sequences of ***B. rapa*** cv. Osome have been applied to ***in silico*** approaches for functional classification based on GO vocabularies, transcript abundance estimation, biochemical classification based on Kyoto encyclopedia of genes and genomes (KEGG) orthology and analyzed protein function domains from a range of biological databases for the respective EST sequences. According to reeds words, all one dimensional data have been assigned to the putative sequences and those functional annotations helped to develop a further understanding of complex gene expression networks.

Materials and Methods

cDNA library construction—An FLcDNA library was constructed according to Sato *et al*.[19] which was reported previously[20] and here described briefly. It was done by biotinylated CAP trapper using trehalose-thermoactivated reverse transcriptase. The mRNA isolated from RNA samples was quality checked and used for first-strand cDNA synthesis. After oxidation, biotinylation and RNase digestion of first-strand cDNA/mRNA hybrids, FLcDNA/RNA hybrids were captured on magnetic beads. RNA was removed by alkaline treatment to collect first-strand FLcDNA. The oligo(dG)-tailed first-strand cDNA was used for second-strand cDNA synthesis. The cDNA was restricted with *Bam*HI and *Sal*I/*Xho*I. After purification, cDNA was cloned into a pFLC-III vector.

EST sequencing—Plasmid DNA was isolated from 3,429 selected colonies with 96 TurboFilter Miniprep kits (Qiagen, Valencia, CA) and a Qiagen BioRobot 9600. Sequencing of 3,429 clones was carried out using ABI 3700 automated capillary DNA sequencers (Applied Biosystems, Republic of Korea). Sequencing was performed using the M13 primer (5'-TGTAAAACGACGGCCAGT-3') to evaluate the clones from the 5' end of the full length cDNA inserts.

EST assembly and annotation—The 3,429 raw read sequences were composed of singlets, contigs and some short repetitive sequences. Obtaining high quality sequences requires several steps; all of them integrated in a freely accessible web-interface named an EGassembler[16,13]. For an EST assembly, an EGassembler was used with the default analysis parameters. This bioinformatics pipeline begins with a sequence cleaning stage, removing low quality sequence stretches, followed by the detection and removal of repetitive elements. The resulting output was then searched and cleaned from the organelle sequences. The sequences not included in the clusters were classified in the assembled dataset.

The assembled unique transcripts were compared with the public databases using Blast2go[17], a sequence-based tool to assign GO terms and annotation, extracting them for each BLAST hit obtained by mapping the extant annotation associations. The GO terms for each of the three main categories (biological process, molecular function and cellular component) were obtained from sequence similarity using the application default parameters. From these annotations, the Group 2nd level and multilevel GO terms were based on the biological process, molecular function and cellular component. This annotation was simplified and focused on the plant related to the functional categories using the Plant GOslim.

Pathway assignment with KEGG—Pathway assignments were mapped according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg/pathway.html)[18]. Enzyme commission (EC) numbers were assigned to the unique sequences with BLASTX scores with an ***E*** value cut-off of 10^{-10} after searching the KEGG protein databases. The unique sequences were mapped to the specific biochemical pathways according to the corresponding EC distribution in the KEGG database.

RNA extraction—For an expression study, Chinese cabbage (*Brassica rapa* 'SUN-3061') seeds were grown aseptically on half-strength MS agar medium in a culture room under a 16 h light photoperiod at 25 °C. After 3 weeks growth, the seedlings were transferred to fresh liquid MSH (half-strength MS medium without sucrose) medium containing 250 mM NaCl and 100 mM abscisic acid (ABA) for 24 h. To induce cold stress, the seedlings were maintained at 4 °C for 24 h. A drought treatment was applied by keeping the seedlings on the filter paper at 28 °C for 24 h. For water stress, plants were transferred to a dish containing water in submerged condition. The samples were treated with all stresses for 0 (wild type), 30 min, 1, 2, 4, 8 and 24. The samples were then frozen immediately in liquid nitrogen, and stored at –80 °C for RNA isolation. The total RNA was extracted from the frozen samples of the roots, stems, leaves and flower buds of healthy plants and those exposed to abiotic stress using a RNeasy mini kit (Qiagen, USA). RNA was treated with RNase-free DNase (Promega, USA) to remove the genomic DNA contamination.

RT-PCR analysis—RT-PCR was performed using an Avian Myeloblastosis Virus (AMV) one step RT-PCR kit (Takara, Japan). The gene specific primers for the stress responsive genes are listed in Supplementary data, Table S1. RT-PCR was performed using 50 ng cDNA of plants exposed to various abiotic stress, such as cold, salt drought, ABA, water, and PEG in different time intervals (0, 30 min, 1, 2, 4, 8, 24 h). In 0.5 mL PCR tubes, 20 *p*mol of each primer, 150 µM of each dNTP, 1.2 U of *Taq* polymerase, 1x *Taq* polymerase buffer, and double-distilled H_2O to a total volume of 20 µL were added and mixed. The PCR procedure involved pre-denaturing at 94 °C for 5 min followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 58 °C for 30 s and extension at 72 °C for 45 s, and was terminated with additional extension for 5 min at 72 °C.

Results and Discussion

Sequencing and assembly of ESTs—Full length cDNA library of *B. rapa* cv. Osome were constructed following the methods described by Sato *et al.*[19] and for that the plants were maintained in a growth chamber under normal conditions. Samples of leaves and stems were collected from five weeks old seedlings and flower buds were collected at flowering stage. These samples of three different organs were mixed for RNA isolation. A single-run partial sequencing of randomly selected cDNA clones was conducted and obtained 3,429 ESTs from this library[20]. With a view to identify stress related genes of *Brassica*, this annotation study was conducted based on gene ontology. After removing the repetitive sequence, short sequence and vector organelle sequence 3,284 ESTs were used for further analysis. The EGassembler online clustering pipeline provided 333 contigs and 2,446 singlets of ESTs giving a total of 3,284 unigenes (analysis preformed with default parameters) (Table 1). Sequences included in the contigs were similar, ranging from 9.7% for the EST library. This result revealed the detection of 0.05% reteroelements, 0.05% LTR elements, 0.02% TY1/Copia, 10.02% Gypsy/DIRS, 0.03% DNA transposons, and 0.03% MuDR-IS905 elements in these sequences. These elements have an important role in genome evolution in *Brassica* species[21].

ESTs against public non-redundant databases—A homology based functional annotation assignment for the putative sequence was accomplished using BLASTX, queries against non-redundant databases. 3,284 ESTs were used to search for non redundant protein databases using BLASTX in the blast2go suite with parameters with an E-value of 10^{-3} or below, a HSP cut-off of 33 and a maximum of 20 blast hits per sequence description tool[17,22]. The average result of BLASTX was 17 sequences per EST for 2,730 ESTs (83.13%). The remainder of the unique sequence 554 ESTs (16.8%) had no homology searches, which

Table 1—General clustering of *B. rapa* EST sequences

Descriptive category	EST sequence	Sources
Total ESTs sequenced	3,429	-
Number of cluster sequence	3,284	EGassembler http://egassembler.hgc.jp/
Number of contigs	333	
Number of singletons	2,446	
No blast hits	554	
Number of sequence with EC	1,197	Blast2go V.2.5.0 www.blast2go.com
Number of pathway maps	124	
Number of Interpro annotation	17,266	

suggests that the little or no similarity was caused mostly by the UTR sequence, or it could be due to the presence of genes that remained unassigned in the genome annotation effort. The total blast hits was 42,716 (90.09%) from 47,413 sequences from the 29 species and another 4,637 (9.78%) sequences from other species of a non redundant database (Supplementary data, Table S2). All the plant species have been well studied and have more experimental data. The use of a homology based functional annotation of those species with the BLAST score is acceptable[23].

Function analysis based on gene ontology—GO is used widely to simplify the annotation process, gene functional annotation and classification of functional genomics. GO describes the gene function using controlled vocabularies and hierarchy, including three major categories, namely molecular function, biological processes and cellular component[24]. The EST sequences were grouped in terms of the GO vocabularies belonging to only one combination of two, and all three vocabularies were organized in a Venn diagram. A total of EST sequences were classified into 1,986, 2,042 and 2,116 ESTs of Molecular Functoion (MF), Biological Process (BP) and Cellular Component (CC), respectively. In addition, 1,420, 1,608 and 168 ESTs were mapped under CC and BP, CC and MF, and MF and BP, respectively. 1,420 were annotated according to all GO sub-vocabularies (i.e. CC, BP, and MF) (Fig. 1). Plant-GOslim was used to screen the plant-specific GO vocabularies. A large number of unique sequences were grouped under the first category of MF with nucleotide binding, structural molecule activity, RNA binding, transcription factor activity, etc. The second category included BP with subcategories, such as the response to stress, response to biotic stimulus, response to abiotic stimuli, protein modification process, catabolic process, and lipid metabolic process, etc. The third category included CC with the subcategories of plastid, plasma membrane, mitochondrion, cytosol, and vacuole, etc. The GO results relied on well annotated GO information of other plants. A large number of unique sequences from this study mapped to MF included nucleotide binding, structural molecule activity, RNA binding, and transcription factor activity. 255 ESTs of nucleotide binding have a significant representation in this category. These ESTs are related to the disease resistance, pathogenesis, hypersensitivity response (HR) and plant defense responses to diseases. The catalytic activities are more important for a cell to survive major metabolic changes while the plant is under biotic and abiotic stress[25,26,23]. The second most represented category was BP, which includes the unique sequences associated with metabolic process, cellular process, stimulus localization, cell growth and development process. The growth and development of plants were altered by the application of external stimuli[27]. Although the BP category has an abundance of responses to stress 462 (SR), abiotic stimulus 415 (ABS) and biotic stimulus 140 (BS) of ESTs, 255 ESTs were involved in ABS and SR. 74 ESTs were involved in SR and BS. One EST was involved in ABS and BS. 50 ESTs were involved in all three stresses (Fig. 2). On the CC category, there is a plastid, plasma membrane, mitochondrion, nucleus, cytosol, thylakoid and endoplasmic reticulum. In this category,

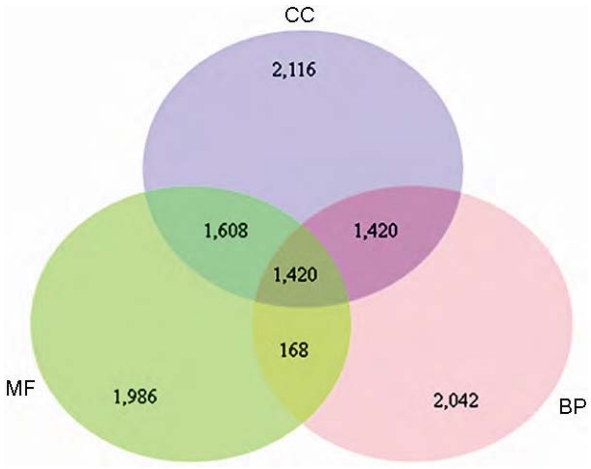

Fig. 1—Venn diagram of 3,284 ESTs dataset showing numbers annotated to a one a combination of two and all three GO vocabularies (MF: molecular function, CC: cellular component, BP: biological process)

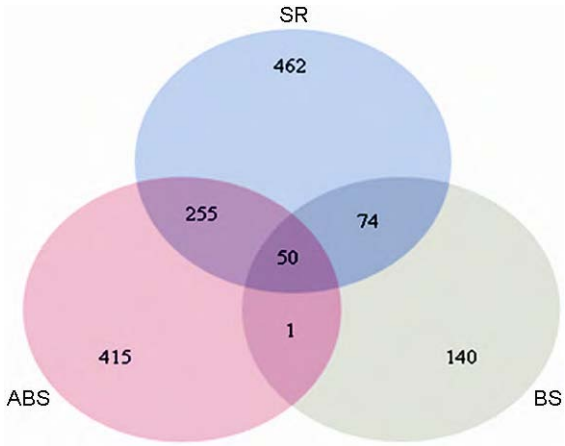

Fig. 2—Venn diagram of 1,017 stress related ESTs dataset showing numbers of annotated, response to stress (SR), response to abiotic stimulus (ABS), and response to biotic stimulus (BS)

there is a significant representation of plastid 784 ESTs. The plastids are involved in nucleus communication, which appears to be of particular importance during plant stress responses. The plastid signals identified thus far can be linked to specific stress conditions[28]. These annotations can be used to identify the genes in a particular component of the plant cells, the elemental activities of a gene product at the molecular level and the operations of molecular events with a defined beginning and end, and are pertinent to the functioning of integrated living units, such as cells, tissues, organs and organisms. An overview of the groups of genes cataloged by the specific annotated GO terms is given as Supplementary data, Table S3.

KEGG based biochemical analysis—Biochemical analysis of the ESTs was performed using KEGG. KEGG was created from bioinformatics algorithms and is the most accessible database of the major biochemical pathways and represents five categories and 32 sub categories for plants[18]. All putative transcripts were subjected to a KEGG database query with the BLAST score to retrieve the KEGG enzyme codes and pathway maps. A collection of 1,197 ESTs were assigned enzyme commission (EC) numbers and had 124 unique mapping to the KEGG biochemical pathway. The KEGG profiles of the transcripts from *B. rapa* cv. Osome are listed in Supplementary data, Table S4 and the metabolism was the most represented category with 12 sub categories representing the total number of corresponding ESTs. Among these, only stress related ESTs corresponding to each sub category of the metabolism category were presented in the Table 2. According to the electronic annotation, approximately 560 stress related enzyme activities of these transcripts were captured. An overview of the metabolic processes showed the highest transcripts (152) followed by carbohydrate metabolism (114) and energy metabolism (96) in the stress related genes. Under these sub categories, there are some pathways involved. These ESTs have been clustered according to the role in the respective pathways (Supplementary data, Table S5). In the cluster of glycolysis/glycogenesis 13 ESTs, of glyoxylate and dicarboxylate 36 ESTs, of carbon fixation and photosynthetic organism 47 ESTs were involved in the stress-related functions. Demonstrated direct involvement of the glycolysis/glycogenesis and glyoxylate and dicarboxylate metabolism in many biochemical adaptations of plant and non plant species to environmental stresses, such as drought, cold/freezing, osmotic stress and anoxia has been demonstrated[29,30]. In the energy metabolism, the mechanism of carbon fixation in photosynthetic organisms are involved in the efficient light harvesting mechanism, repression of cell growth, alteration in enzyme activity, increased respiration and accumulation of osmolytes and proteins and thereby increase stress tolerance[31,32].

KEGG orthology and protein domains—To improve the theoretical evidence of the annotation, some data mining was performed from the biological databases. An alternative control vocabulary of biochemical for genes is KEGG orthology (KO). The use of the KOBAS (http://kobas.cbi.pku.edu.cn/program.inputForm.do?program=Annotate) bioinformatics pipeline assigned KOs of 83, 281 and 303 to 140 ESTs of BS, 415 ESTs of ABS and 462 ESTs of SR genes in that clusters, respectively. The use of these 83, 281 and 303 putative sequences was assigned to KOs, and 65.5% of the annotation was improved for the putative sequence. Interproscan was provided by EBI web service[33]. Based on the GO

Table 2—Stress related genes based on KEGG biochemical analysis

Metabolism pathways	Stress related ESTs
Carbohydrate metabolism	114
Energy metabolism	96
Lipid metabolism	30
Nucleotide metabolism	21
Amino acid metabolism	52
Metabolism of other amino acids	32
Glycan biosynthesis and metabolism	2
Metabolism of cofactors and vitamins	12
Metabolism of terpenoids and polyketides	8
Biosynthesis of other secondary metabolites	16
Xenobiotics biodegradation and metabolism	29
Overview	152

terms, the functional domains from a range of primary and secondary protein databases were obtained for ESTs using Blast2Go bioinformatics tool. Databases, including PFAM[34], SMART[35], GENE3d[36], PROSITE[37], PROFILE[37], SUPERFAMILY[38], PANTHER[39], PIR[40], PRINTS[41] and TIGRFAMs[42], TMHMM[43] and SINGNALP[44] were used to identify the trans-membrane regions and signaling part of the EST sequence. The total unique ids were 1,378 PFAM, 306 SMART, 1,100 GENE3d, 291 PROSITE, 402 PROFILE, 688 SUPERFAMILY, 1,971 PANTHER, 52 PIR, 235 PRINTS, and 87 TIGRFAMs. For the transmembrane regions, 1,342 TMHMM and 1,274 SINGNALP were found. Functional domains were identified for the genes involved in the response to stress Glyoxalase/bleomycin resistance protein/dioxygenase PF00903 (PFAM), NAC (Nascent polypeptide-associated complex) (PF01849), PDZ binding protein (PF10235)[45-47]. Using these merging results with the GO annotation, a mean of 12.2% totally augmented the functional annotations for the putative sequences.

Expression analysis after abiotic stress treatments—This study was extended to analyze the responses of stress-responsive ESTs against corresponding stresses as experimental evidence of such EST clustering and 18 ESTs were randomly selected from 415 abiotic stress-responsive ESTs and sequenced as full length cDNA. In another study, 3 chitinase genes were identified from the 140 biotic stress-responsive ESTs and characterized their function against soft rot causing bacteria, *Pectobacterium carotovorum* subsp. *carotovorum*[48]. Here, the full length cDNA sequences were then blasted in the public database and found high homology with *Arabidopsis thaliana*. All these sequences were well annotated and relied on the abiotic stress related genes with a high percentage of identity, excluding *SRG*-11 (Table 3). In addition their expression was analyzed after applying cold, salt, drought, ABA, water and PEG stress treatments in *Brassica rapa*. Among these genes, 15 SRGs showed responsive expression against the abiotic stress factors applied in this study. The following genes were up-regulated from 0 h to 24 h: after cold *SRG, 14, 24, 47, 52, 58, 65* and *71*, after salt *SRG* 3, 24, 44 and 52, after drought *SRG* 3, 9, 14, 15, 24, 42, and 53, after ABA *SRG* 14, 42, 52 and 53, after water *SRG* 9, 11, 14, 28, 42, 44 and 53, and after PEG stress treatments, *SRG* 14, 24, 53 and 71 (Fig. 3). Other SRGs might be responsive to other stress factors and have different functional roles despite their expression in all time courses (unpublished data). Among these differentially expressed genes, *SRG14*, *42, 65* and *75* were observed in metabolic pathway which is regulated by drought stress[49]. *SRG14* and *42* also accounted in ascorbate and aldarate metabolism

Table 3—Sequence comparison of stress related genes with *Arabidopsis thaliana*

Gene	Top matched gene to *A. thaliana*	Identity (%)	Seq. description of *A. thaliana*
SRG-3	AT3G03270	90	universal stress protein family protein
SRG -9	AT3G09870	83	auxin-responsive family protein
SRG -11	AT5G53820	90	unknown protein
SRG -14	AT1G76160	86	copper ion binding
SRG -15	AT1G05180	90	auxin resistant 1
SRG -19	AT1G13450	87	DNA binding protein GT-1
SRG -24	AT5G15970	83	KIN2, COR6.6
SRG -27	AT1G30360	82	early-responsive to dehydration 4
SRG -28	AT2G47710	88	universal stress protein (USP) family protein
SRG -42	AT1G55570	80	copper ion binding /oxidoreductase
SRG -44	AT1G06040	88	salt tolerance
SRG -47	AT5G12020	86	HSP17.6II class II (heat shock protein)
SRG -52	AT1G78240	87	TSD2 (tumorous shoot development 2)
SRG -53	AT5G39730	84	avirulence-responsive protein-related
SRG -58	AT1G29420	85	response to auxin stimulus
SRG -65	AT3G30775	87	proline dehydrogenase
SRG -71	AT1G30360	82	early-responsive to dehydration 4
SRG -75	AT1G13280	85	allene oxide cyclase 4

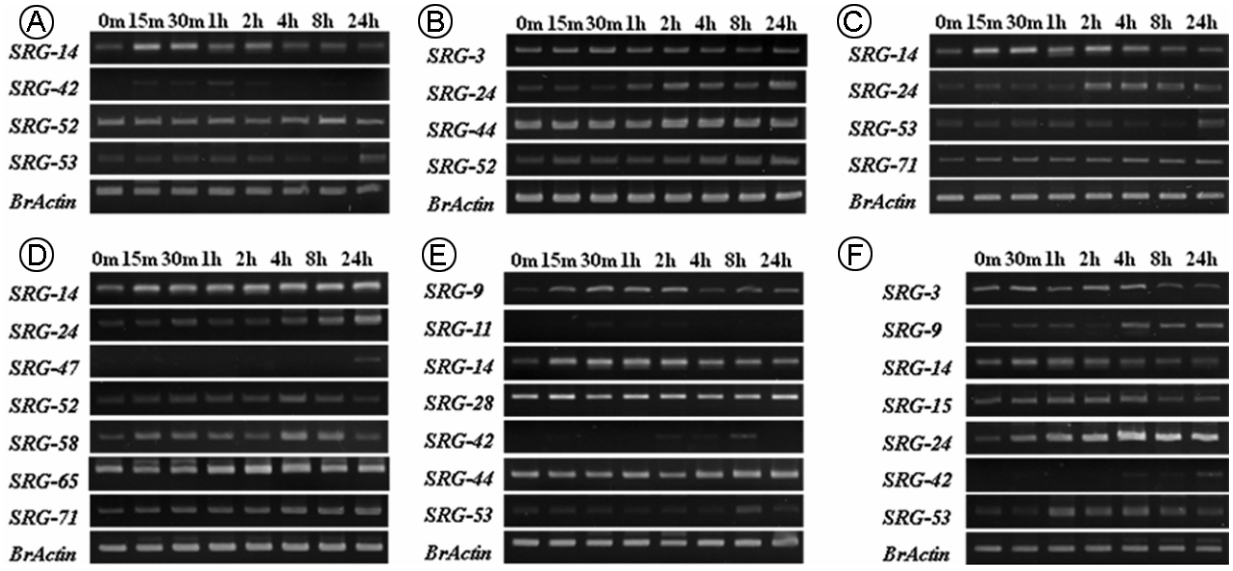

Fig. 3—RT-PCR analysis of stress related genes (SRG). Total RNA was isolated from 3 weeks old plants and treated under different stress condition (A) ABA, (B) salt, (C) PEG, (D) Cold, (E) water and (F) drought for different time courses from 0 m to 24 h (0m as wild type, m: minute, h: hour)

pathway which is regulated by UV-radiation and sulphur di oxide stress[50,51]. *SRG65* also accounted in arginine and proline metabolism and biosynthesis of secondary metabolites. These pathways were commonly regulated by environmental and biotic stress[52,53]. *SRG75* also involved in alpha-linolenic acid metabolism and biosynthesis of plant hormones pathways were regulate by temperature effects and plant growth and development[54,55]. This information provided here is the important to better understand the molecular mechanism involved in abiotic stresses. Other differentially expressed genes may be involved in other defense related pathways which require further study. These results showed that the ESTs were well clustered by such EST analysis.

Overall, the EST dataset provides significant resources for discovering the potential genes of interest and suggests a possible way to characterize them.

Supplementary data

Supplementary data (Tables S1–S5) associated with this article may be obtained from the correspondent author on request.

Acknowledgment

This work was supported by the Cabbage Genomics assisted breeding supporting center research program funded by Ministry for Food, Agriculture, Forestry and Fisheries of the Korean Government.

References

1 Abe H, Narusaka Y, Sasaki I, Hatakeyama K, Shin I S, Narusaka M, Fukami-Kobayashi K, Matsumoto S & Kobayashi M, Development of full-length cDNAs from Chinese cabbage (*Brassica rapa* Subsp. pekinensis) and identification of marker genes for defence response, *DNA Res*, 18 (2011) 277.

2 Xiang Y, Wong W K R, Ma M C & Wong R S C, *Agrobacterium*-mediated transformation of *Brassica campestris* ssp. *Parachinensis* with synthetic *Bacillus thuringiensis cry1Ab* and *cry1Ac* genes, *Plant Cell Rep*, 19 (2000) 251.

3 Zhuang J, Xiong A S, Peng R H, Gao F, Zhu B, Zhang J, Fu X Y, Jin X F, Chen J M, Zhang Z, Qiao Y S & Yao Q H, Analysis of *Brassica rapa* ESTs: gene discovery and expression patterns of AP2/ERF family genes, *Mol Biol Rep*, 37 (2010) 2485.

4 Adams M D, Kelley J M, Gocayne J D, Dubnick M, Polymeropoulos M H, Xiao H, Merril C R, Wu A, Olde B, Moreno R F & et al, Complementary DNA sequencing: expressed sequence tags and human genome project, *Science*, 252 (1991) 1651.

5 Nagaraj S H, Gasser R B & Ranganathan S, A hitchhiker's guide to expressed sequence tag (EST) analysis, *Brief Bioinform*, 8 (2007) 6.

6 Gueguena Y, Jean-Paul C, Didier F, Catherine B R, Anne L G, Julien G, Alixe H, Evelyne B & Michel Escoubas J, Immune gene discovery by expressed sequence tags generated from hemocytes of the bacteria-challenged oyster, *Crassostrea gigas*, *Gene*, 303 (2003) 139.

7 Zhang L, Ma X L, Zhang Q, Ma C L, Wang P P, Sun Y F, Zhao Y X & Zhang H, Expressed sequence tags from a NaCl-treated *Suaeda salsa* cDNA library, *Gene*, 267 (2001) 193.

8 Kim S I, Kim J Y, Kim E A, Kwon K H, Kim K W, Cho K, Lee J H, Nam M H, Yang D C, Yoo J S & Park Y M, Proteome analysis of hairy root from *Panax ginseng* C.A.

Meyer using peptide fingerprinting, internal sequencing and expressed sequence tag data, *Proteomics*, 3 (2003) 2379.
9 Ohlrogge J & Benning C, Unraveling plant metabolism by EST analysis, *Curr Opin Plant Biol*, 3 (2000) 224.
10 White J A, Todd J, Newman T, Focks N, Girke T, de Ilarduya O M, Jaworski J G, Ohlrogge J B & Benning C, A new set of *Arabidopsis* expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil, *Plant Physiol*, 124 (2000) 1582.
11 Gang D R, Wang J, Dudareva N, Nam K H, Simon J E, Lewinsohn E & Pichersky E, An investigation of the storage and biosynthesis of phenylpropenes in sweet basil, *Plant Physiol*, 125 (2001) 539.
12 Brandle J E, Richman A, Swanson A K & Chapman B P, Leaf ESTs from *Stevia rebaudiana*: a resource for gene discovery in diterpene synthesis, *Plant Mol Biol*, 50 (2002) 613.
13 Falgueras J**,** Antonio J L, Noe F P**,** Francisco R C**,** Guillermo P T & Claros M G, SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read, *BMC Bioinformatics*, (2010) 1471.
14 Pal D, On gene ontology and function annotation, *Bioinformation*, 1 (2006) 97.
15 Gene Ontology Consortium, The Gene Ontology (GO) project in 2006, *Nucleic Acids Res*, 34 (2006) D322.
16 Masoudi-Nejad A, Tonomura K, Kawashima S, Moriya Y, Suzuki M, Itoh M, Kanehisa M, Endo T & Goto S, EGassembler: online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments, *Nucleic Acids Res*, 34 (2006) W459.
17 Gotz S, Garcia-Gomez J M, Terol J, Williams T D, Nagaraj S H, Nueda M J, Robles M, Talon M, Dopazo J & Conesa A, High-throughput functional annotation and data mining with the Blast2GO suite, *Nucleic Acids Res*, 36 (2008) 3420.
18 Kanehisa M & Goto S, KEGG: Kyoto encyclopedia of genes and genomes, *Nucleic Acids Res*, 28 (2000) 27.
19 Sato K, Shin I T, Seki M, Shinozaki K, Yoshida H, Takeda K, Yamazaki Y, Conte M & Kohara Y, Development of 5006 Full-Length CDNAs in Barley: A tool for accessing cereal genomics resources, *DNA Res*, 16 (2009) 81.
20 Park J I, Thamil arasan S K, Ahmed N U & Nou I S, Construction of full length cDNA library and investigation of potential genes in *Brassica rapa, J Basic Life Res Sci*, 10 (2) (2010) 6.
21 Qiu D, Gao M, Li G & Quiros C, Comparative sequence analysis for *Brassica oleracea* with similar sequences in *B. rapa* and *Arabidopsis thaliana*, *Plant Cell Rep*, 28 (2009) 649.
22 Conesa A & Gotz S, Blast2GO: a comprehensive suite for functional analysis in plant genomics, *Int J Plant Genomics*, (2008) 619832.
23 Subramaniyam S, Jun G I, Sathiyaraj G, Kim Y J & Yang D C, Generation and gene ontology based analysis of expressed sequence tags (EST) from a *Panax ginseng* C. A. Meyer roots, *Mol Biol Rep*, 37 (2010) 3465.
24 The Gene Ontology Consortium The Gene Ontology in 2010: extensions and refinements, *Nucl Acids Res*, 38 (suppl 1) (2010) D331.
25 Samuel M A, Caixeta E T, Hufnagel B, Thiebaut F, Maciel-Zambolim E, Zambolim L & Sakiyama N S, *In silico* identification of coffee genome expressed sequences potentially associated with resistance to diseases, *Genet Mol Biol*, 33, 4 (2010) 795.
26 Di Baccio D, Galla G, Bracci T, Andreucci A, Baraccia G, Tognetti R & Sebastiani L, Transcriptome analyses of *Populus* × *euramericana* clone I-214 leaves exposed to excess zinc, *Tree Physiol,* 31 (2011) 1293.
27 Jeter C R & Roux S J, Plant responses to extracellular nucleotides: cellular processes and biological effects, *Purinergic Signal*, 2 (2006) 443.
28 Aurora P F & Strand A, Retrograde signaling and plant stress: plastid signals initiate cellular stress responses, *Curr* Opin Plant Biol, 11 (2008) 509.
29 William C P, The organization and regulation of plant glycolysis, *Annu Rev Plant Physiol Plant Mol Biol*, 47 (1996) 185.
30 Kristensen T N, Sorensen P, Pedersen K S, Kruhoffer M & Loeschcke V, Inbreeding by environmental interactions affect gene expression in *Drosophila melanogaster*, *Genetics*, 173 (2006) 1329.
31 Sakuma Y, Maruyama K, Osakabe Y, Qin F, Seki M, Shinozaki K & Yamaguchi-Shinozaki K, Functional analysis of an *Arabidopsis* transcription factor, DREB2A, involved in drought-responsive gene expression, *Plant cell*, 18 (2006) 1292.
32 Shinozaki K & Yamaguchi-Shinozaki K, Gene networks involved in drought stress response and tolerance, *J Exp Bot*, 58 (2007) 221.
33 Labarga A, Valentin F, Anderson M & Lopez R, Web services at the European bioinformatics institute, *Nucleic Acids Res*, 35 (2007) W6.
34 Finn R D, Mistry J, Schuster B B, Griffiths J S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy S R, Sonnhammer E L & Bateman A, Pfam: Clans, web tools and services, *Nucleic Acids Res*, 34 (2006) D247.
35 Letunic I, Doerks T & Bork P, SMART 6: Recent updates and new developments, *Nucleic Acids Res*, 37 (2009) D229.
36 Buchan D W, Shepherd A J, Lee D, Pearl F M, Rison S C, Thornton J M & Orengo C A, Gene3D: structural assignment for whole genes and genomes using the CATH domain structure database, *Genome Res*, 12 (2002) 503.
37 Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux P S, Pagni M & Sigrist C J, The PROSITE database, *Nucleic Acids Res*, 34 (2006) D227.
38 Gough J, Karplus K, Hughey R & Chothia C, Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure, *J Mol Biol*, 313 (2001) 903.
39 Mi H, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, Rabkin S, Guo N, Muruganujan A, Doremieux O, Campbell M J, Kitano H & Thomas P D, The PANTHER database of protein families, subfamilies, functions and pathways, *Nucleic Acids Res*, 33 (2005) D284.
40 Wu C H, Yeh L S, Huang H, Arminski L, Castro-Alvear J, Chen Y, Hu Z, Kourtesis P, Ledley R S, Suzek B E, Vinayaka C R, Zhang J & Barker W C, The protein information resource, *Nucleic Acids Res*, 31(2003) 345.
41 Attwood T K, Croning M D, Flower D R, Lewis A P, Mabey J E, Scordis P, Selley J N & Wright W, PRINTS-S: the database formerly known as PRINTS, *Nucleic Acids Res*, 28 (2000) 225.
42 Selengut J D, Haft D H, Davidsen T, Ganapathy A, Gwinn-Giglio M, Nelson W C, Richter A R & White O, TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes, *Nucleic Acids Res*, 35 (2007) D260.

43 Krogh A, Larsson B, von Heijne G & Sonnhammer E L, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, *J Mol Biol*, 305 (2001) 567.
44 Bendtsen J D, Nielsen H, von Heijne G & Brunak S, Improved prediction of signal peptides: SignalP 3.0, *J Mol Biol*, 340 (2004) 783.
45 Ananda M, Singh A K, Pareek A, Sopory S K & Pareek S LS, Genome-wide analysis of rice and *Arabidopsis* identifies two glyoxalase genes that are highly expressed in abiotic stresses, *Funct Integr Genomics*, 11 (2011) 293.
46 Jiang Y, Yang B, Harris N S & Deyholos1 M K, Comparative proteomic analysis of NaCl stress-responsive proteins in *Arabidopsis* roots, *J Exp Bot*, 58 (2007) 3591.
47 Schlieker C, Mogk A & Bukau B, A PDZ Switch for a Cellular Stress Response, *Cell*, 117 (2004) 417.
48 Ahmed N U, Park J I, Mi-Suk S, Thamilarasan S K, Lee I H, Park B S & Nou I S, Identification and expression analysis of chitinase genes related to biotic stress resistance in *Brassica*, *Mol Biol Rep*, 39 (2012) 3649.
49 Dias F G, Neves-Borges A C, Barbosa Viana A A *et al.*, Expression analysis in response to drought stress in soybean: Shedding light on the regulation of metabolic pathway genes, *Genetics and Molecular Biology*, 35 (1) (2012) 222.
50 Kubo A, Saji H, Tanaka K & Kondo N, Expression of Arabidopsis cytosolic ascorbate peroxidase gene in response to ozone or sulfur dioxide, *Plant Molecular Biology*, 29 (1995) 479.
51 Costa H, Gallego SM & Tomaro ML, Effect of UV-B radiation on antioxidant defense system in sunflower cotyledons, *Plant Science*, 162 (2002) 939.
52 Verslues P E & Sharma S, Proline metabolism and its implications for plant-environment interaction, The Arabidopsis Book, *American Society of Plant Biologists* (2010) 1.
53 Franca S C, Roberto P G, Marins M A, Puga R D, Rodrigues A & Pereira J O, Biosynthesis of secondary metabolites in sugarcane, *Genet Mol Biol*, 24 (2001)1.
54 Chapman J D, Felice J D, & Barber J, Growth temperature effects on thylakoid membrane lipid and protein content of pea chloroplasts, *Plant Physiol*, 72(1) (1983) 225.
55 Kim G T & Tsukaya H, Regulation of the biosynthesis of plant hormones by cytochrome P450s, *J Plant Res*, 115 (2002) 169.