Journal of Plant Biology, June 2003, 46(2) : 83-89

Cloning and Functional Annotation of Rare mRNA Species from Drought-Stressed Hot Pepper (*Capsicum annuum*)

Yoo Bok Cho, Hee Jin Jeon, and Choo Bong Hong*
Institute of Molecular Biology and Genetics and School of Biological Sciences, Seoul National University, Seoul 151-742, Korea

To better understand gene expression at very low levels, we have designed a method to eliminate cDNA clones representing abundant mRNAs. A cDNA library for drought-stressed hot pepper (*Capsicum annuum*) (Choi et al., 2002) underwent double-negative screening, once with probes made from a drought-stressed plant, the second time, with probes from a non-stressed plant. The cDNA clones that showed very weak or negative signals were isolated for further analysis, which resulted in 1399 cDNA clones from about 20,000 screened clones. When nucleotide sequences were determined, we obtained 1142 tentative unique genes, with a redundancy rate of 20.41%. An homology database search for the deduced amino acid sequences revealed that about 79% of the cDNA clones could not be matched for functioning with previously characterized sequences. However, when these uncategorized clones were subjected to classification based on functional domains, most could be cited. Notably, clones with possible functions in RNA transport, protein synthesis, and regulation of protein activity showed a dramatic increase in appearance while those coding for transposable elements, viral proteins, and plasmid proteins occupied a much smaller portion compared with those in the *Arabidopsis thaliana* genome. In addition, those coding for proteins targeted to the endoplasmic reticulum were dramatically more abundant in our clones compared with the *Arabidopsis* database.

Keywords: *Capsicum annuum*, cDNA clones, drought stress, functional annotation, hot pepper

Because abiotic stress dramatically lowers plant productivity, much effort has been given to improving stress resistance in crops. A molecular approach to deciphering stress perception and its signal transduction pathway has revealed that these events are very complex and widely interconnected (Cushman and Bohnert, 2000; Hasegawa et al., 2000; Knight and Knight, 2001; Xiong et al., 2002; Zhu, 2002). Drought stress alone accounts for losses exceeding 20 million tons of grain per year in the tropics, or about 17% of the well-watered production (Ribaut et al., 2002).

Abiotic stresses alter cellular metabolism and gene expression, often leading to adaptive responses. The components of this overall process, which involve interconnected pathways, fall into two groups: those that directly exhibit protective functions against environmental stresses, and those that regulate gene expression and signal transduction in the stress response. The first group comprises proteins that function as antioxidants, osmoprotectants, radical scavengers, detoxifiers, chaperones, etc. The second includes transcription factors, phosphoprotein cascade members, enzymes involved in phosphoinositide metabolism, and secondary signaling-molecule modulators (Yamaguchi-Shinozaki and Shinozaki, 2002).

Genomic approaches have greatly facilitated the discovery of relevant plant genes (Cooke et al., 1996; Martin, 1998; Asamizu et al., 2000; Ohlrogge and Benning, 2000; White et al., 2000; van der Hoeven et al., 2002). High throughput technologies, e.g., DNA microarrays, transposon/T-DNA tagging, and protein-protein interaction cloning (Cushman and Bohnert, 2000; Seki et al., 2001, 2002; Fowler and Thomashow, 2002), as well as the development of large Expressed Sequence Tag (EST) databases (Choi et al., 2002) and search software are important for envisioning genes whose products are essential either during development or as part of the response pathways to environmental stresses. In particular, the large-scale EST approach has greatly enhanced the rate of gene discovery, all at low cost. However, this approach inevitably leads to the over-representation of highly abundant transcripts while it under-represents rare transcripts. This drawback makes it difficult to isolate regulatory factors for low expression levels.

The objective of the current study was to isolate the regulatory factors associated with drought stress in hot pepper (*Capsicum annuum*). To preferentially identify the cDNA clones of rare mRNA species, we screened

*Corresponding author; fax +82-2-875-1157
e-mail hcb@snu.ac.kr

a previously constructed cDNA library (Choi et al., 2002) for clones showing weak or negative signals by both cDNA probes (i.e., 'double-negative screening'). After the selected clones were sequenced, previous annotations of *Arabidopsis thaliana* (Mysore et al., 2001) were used as comparison for putatively characterizing those rare transcript species.

MATERIALS AND METHODS

Preparation of cDNA Library

A previously constructed cDNA library for drought-stressed hot pepper (C. *annuum* cv. Bu Gang) was used (Choi et al., 2002). The cDNA library was prepared as follow: Hot pepper (C. *annuum* cv. Bu Gang) plants were raised in a growth chamber (16-h photoperiod; 25°C; 60% humidity; and 200 $\mu E/m^2 \cdot s$, from white fluorescent lamps). Dehydration stress was induced in 10-d-old plants by withholding water for 14 to 20 d. Total RNA was then isolated according to the method described by Park and Hong (1991). After the poly(A)$^+$RNA was purified, a unidirectional EcoRI/XhoI cDNA library was constructed in a Uni-λZAP XR vector (Stratagene, USA). The cDNA and vector recombinants were treated with GigapackIII gold packaging extracts (Stratagene).

Double-Negative Screening

Low-abundance mRNA species were selected via double-negative screening. After the cDNA library was amplified and titrated, 3000 to 4000 plaques were plated on *Escherichia coli* strain XL blue lawn in a 150-mm NZY agar plate. The plates were incubated for 9 h at 37°C, and the plaques on each plate were then transferred to two neutral nylon membranes (Hybond-N, Amersham, USA). Duplicate plaque lifts were denatured, neutralized, blotted dry, and exposed to UV irradiation to cross-link the DNA to the membranes. For hybridization, [α-^{32}P]dCTP (3,000 Ci/mmol)-labeled cDNA probes were prepared through random priming of the mRNAs from both drought-stressed plants and unstressed, well-irrigated stock. These probes were denatured by NaOH and added to fresh prehybridization buffer. One hour of prehybridization and >18 h of hybridization were performed at 65°C in 0.5 M sodium phosphate buffer. Afterward, the membranes were washed in 2x SSC and 0.1% SDS at room temperature for 10 min, 1x SSC and 0.1% SDS at 65°C for 10 min, and 0.5x SSC and 0.1% SDS at 65°C for 10 min. Membranes were then blotted dry and exposed to X-ray film overnight. Plaques that showed weak or negative signals from both probes were isolated and stored in SM buffer containing 0.05% chloroform (Sambrook et al., 1989). Approximately 250 to 300 plaques were picked up from each plate. Those that had been selected were then converted to pBluescript SK(-) by in-vivo excision, using the ExAssist helper phage (M13; Stratagene).

Sequencing and Sequence Editing

An automatic sequencer (ABI 3700, Perkin Elmer, USA) was used to sequence the selected cDNA inserts from the 5' end. Vector sequences were removed with a vectorstrip program in the EMBOSS package, and were edited manually to trim unusable data at the 3' end (Rice et al., 2000). The sequences were then assembled with the CAP EST Assembler at IFOM (http://bio.ifom-firc.it/ASSEMBLY/assemble.html) and, to estimate redundancy, were manually corrected if obviously unrelated sequences were clustered together (Huang, 1996). Further analysis was performed using BLAST to confer putative function and to verify the rate of novel gene discovery (Altschul et al., 1997). Our search results were formatted by MuSeqBox (Xing and Brendel, 2000).

Functional Classification and Prediction of Subcellular Localization

A functional catalog of TUGs (Tentative Unique Genes) was produced according to the manually assigned functional catalog of *Arabidopsis* at the MIPS *A. thaliana* Database (MAtDB). This followed a BLASTX similarity search of TUGs, in which *Arabidopsis* protein-coding sequences were downloaded from MAtDB (Schoof et al., 2002); the top scoring hits were automatically extracted (Frishman et al., 2001) via the *Arabidopsis* Genome Initiative (AGI, 2000) locus identifier. Enrichment was calculated by dividing the difference between genome and transcriptome contents by the genome content. Subcellular localization was predicted by extracting from the annotation of *Arabidopsis* functional categorization only those hits with such localization information.

RESULTS

Double-Negative Screening of cDNA Library

The cDNA library from drought-stressed hot pepper

had an average insert size of 0.89 kb (data not shown). By double-negative screening, 1399 clones were chosen and nucleotide sequenced. Among these, 1296 clones were selected for read length over 400 nucleotides. The 5'-EST sequences were then assembled to identify the number of TUGs. Out of 1296 clones, 266 had tentative consensus sequences (TC) and 1030 clones were revealed as singletons. After clustering by the CAP assembler and manual curating, 1142 TUGs were produced; redundancy was 20.41%. However, if all the clones assembled into TC were considered the same, the redundancy rate increased to 22.76% (Table 1). Among the 266 ESTs that were members of TC, 87 were represented twice, 16 three times, 7 four times, 1 five times, and another, 11 times. The most highly represented clone was a hypothetical protein followed by the small subunit of ribulose bisphosphate carboxylase. Moderately redundant clones included Pn47p (a lipase-like protein), the S-adenosylmethionine decarboxylase proenzyme, an NADH dehydrogenase-like protein, a BURP domain-containing protein, thioredoxin peroxidase, and a putative translocation transactivator (data not shown).

Similarity with Publicly Accessible Sequence Data

Most of the ESTs could be annotated for their molecular function based on amino acid-sequence similarity to proteins with known functions. Here, a BLASTX was used, with an expect value (E-value) of $1.0E^{-10}$. Of the 1142 TUGs obtained, 866 (75.83%) had hits when the sequences were searched against the non-redundant database (nr) of GenBank; 276 (24.17%) had no hits. Among those with hits, 55 expressed proteins with no known function, 30 were hypothetical, 33 were putative, 42 were unknown, and 10 had only a returned locus. If the top-scoring hits also were applied, the counts listed above could be increased to 82, 71, 70, 64, and 38, respectively. Regarding nucleotide-sequence similarity, 687 ESTs (60.16%) had no hits against nr, and 155 (13.57%) had no hits against dbEST. When the BLAST results of coding nucleotide sequences and peptide sequences were compared with the ESTs of *Arabidopsis*, 80.54% (294 of 365) of the coding nucleotide sequences and 94.12% (801 of 851) of the protein-coding sequences had EST matches. Tomato and potato ESTs produced fewer hits than dbEST (Table 2).

Table 1. Statistics for tentative unique genes (TUGs) of hot pepper.

Total number of ESTs	1296
Number of ESTs in TCs	266
Number of singleton ESTs	1030
Total number of TCs	112
Total number of TUGs	1142
Estimated average insert length	0.89 kb*
Redundancy	20.41%

*Insert length was calculated from 100 randomly chosen clones.

Functional Annotation and Classification of Tentative Unique Genes

We were unable to assign functional categories based on our sequence-analysis results because of the partiality that arose from the 5' single-pass sequencing. Therefore, functions were matched with corresponding *Arabidopsis* genes that had previously been annotated (Mewes et al., 2002). The unclassified category occupied the largest portion (78.97%), whereas the major components in the classified category (Table 3A) consisted of ESTs related to metabolism (4.47%), cellular organization (3.06%), cell rescue (2.47%), and protein synthesis (1.65%).

Although manual assignment of the MIPS catalog is a useful tool for determining the distribution of functional categories, this method was not suitable for our analysis because most of the ESTs were not categorized. Thus, we relied on the automatically derived catalog of *Arabidopsis* that can search for possible domains with suggested functions. This approach, however, often assigns a protein to multiple functions, which usually makes the sum of fractions >100% (Frishman et al., 2001). Nevertheless, when the automatic derivations

Table 2. Sequence similarity search from public databases.

Category showing similarity to our ESTs	BLASTN	TBLASTX
Number of nucleotide sequences in the coding region of *Arabidopsis*	365	ND[a]
Peptide sequences of *Arabidopsis*	NDa	851[b]
ESTs of *Arabidopsis*	294	801
ESTs of hot pepper at PDRC[c]	732	ND[a]
ESTs of tomato	884	852
ESTs of potato	832	903
Non redundant DB (nr)	455	866[b]
DbEST	987	ND[a]

[a]ND, not determined.
[b]BLASTX was carried out instead of TBLASTX
[c]PDRC, http://plant.pdrc.re.kr/blast/blast.html
Arabidopsis sequences were downloaded from MatDB of The Arabidopsis Information Resource (TAIR) home page (http://www.arabidopsis.org/info/genefamily/genefamily.html), and ESTs were downloaded from ZmDB web site (http://www.zmdb.iastate.edu: updated Sep. 17, 2002) except for hot pepper ESTs, which were searched on the web (http://plant.pdrc.re.kr/blast/blast.html).

were used, ESTs in the following categories were significantly increased compared with the distribution in the *Arabidopsis* database: RNA transport (9.81 times), protein activity regulation (2.24 times), and protein synthesis (2.05 times). In contrast, transposable elements, viral and plasmid proteins (0.09 times), and proteins with a binding function or cofactor requirement (0.42 times) were greatly depleted. Throughout the catalog, the largest portion contained categories of regulatory functions, e.g., protein activity regulation, systemic regulation, protein fate, cellular transport, cell growth, transcription, protein destination, transport facilitation, and cellular communication (Table 3B).

Prediction of Subcellular Localization and Membrane Proteins

In our analysis, cytoplasmic and nuclear proteins comprised >74% of the ESTs categorized for subcellular localization. When compared with the *Arabidopsis*

Table 3. Distribution and estimated enrichment of functional categories of hot pepper TUGs. A) Manually assigned functional catalog of pepper TUGs; B) Automatically derived functional catalog of pepper TUGs.

A)

Category	Pepper TUGs		Enrichment[a]	*Arabidopsis*
01 Metabolism	38	4.47%	43.58%	3.11%
02 Energy	7	0.82%	49.81%	0.55%
03 Cell growth, cell division and DNA synthesis	2	0.24%	−44.00%	0.42%
04 Transcription	9	1.06%	−54.83%	2.34%
05 Protein synthesis	14	1.65%	139.70%	0.69%
06 Protein destination	6	0.71%	−27.51%	0.97%
07 Transport facilitation	10	1.18%	12.64%	1.04%
08 Cellular transport and transport mechanisms	1	0.12%	−74.82%	0.47%
09 Cellular biogenesis	3	0.35%	−50.06%	0.71%
10 Cellular communication / signal transduction	13	1.53%	−22.56%	1.97%
11 Cell rescue, defense, cell death and aging	21	2.47%	88.95%	1.31%
30 Cellular organization	26	3.06%	94.27%	1.57%
50 Development	4	0.47%	59.80%	0.29%
90 Transposable elements, viral and plasmid protein	0	0.00%	−100.00%	0.55%
98 Classification not yet clear-cut	25	2.94%	7.32%	2.74%
99 Unclassified protein	672	78.97%	16.44%	67.82%
SUM	851	100.00%	0.00%	100.00%

B)

01 Metabolism	229	26.88%	34.18%	20.03%
02 Energy	76	8.92%	66.44%	5.36%
03 Cell cycle and DNA processing	86	10.09%	16.03%	8.70%
04.01 rRNA transcription	16	1.88%	17.15%	1.60%
04.03 tRNA transcription	9	1.06%	12.49%	0.94%
04.05 mRNA transcription	154	18.08%	16.82%	15.47%
04.07 RNA transport	22	2.58%	880.87%	0.26%
05 Protein synthesis	77	9.04%	105.00%	4.41%
06 Protein fate (folding, modification, destination)	133	15.61%	43.38%	10.89%
08 Cellular transport and transport mechanisms	100	11.74%	34.68%	8.71%
10 Cellular communication / signal transduction mechanism	114	13.38%	12.24%	11.92%
11 Cell rescue, defense and virulence	155	18.19%	31.20%	13.87%
13 Regulation of / interaction with cellular environment	38	4.46%	43.87%	3.10%
14 Cell fate	82	9.62%	38.23%	6.96%
20 Systemic regulation of / interaction with environment	24	2.82%	67.12%	1.69%
25 Development (Systemic)	35	4.11%	1.21%	4.06%
29 Transposable elements, viral and plasmid proteins	3	0.35%	−90.53%	3.72%
30 Control of cellular organization	92	10.80%	30.99%	8.24%
62 Protein activity regulation	3	0.35%	124.04%	0.16%
63 Protein with binding function or cofactor requirement	1	0.12%	−57.93%	0.28%
65 Storage protein	2	0.23%	−14.65%	0.28%
67 Transport facilitation	63	7.39%	39.61%	5.30%
SUM		177.70%	ND[b]	135.95%

[a]Enrichment (%) = (hot pepper TUGs - *Arabidopsis* genome)/*Arabidopsis* genome × 100.
[b]ND, Not determined.

Table 4. Prediction results of subcellular localization.

MIPS-localization	Pepper TUGs		Enrichment[a]	*Arabidopsis*
Cytoplasm	152	40.64%	37.20%	29.62%
Plasma membrane	26	6.95%	−29.23%	9.82%
Nucleus	126	33.69%	−22.10%	43.25%
Endoplasmic reticulum	51	13.64%	85.07%	7.37%
Plastid	42	11.23%	46.61%	7.66%
Cell wall	8	2.14%	−39.89%	3.56%
Golgi	22	5.88%	9.50%	5.37%
Mitochondrion	65	17.38%	22.83%	14.15%
Cytoskeleton	24	6.42%	17.37%	5.47%
Endosome	3	0.80%	1.27%	0.79%
Peroxisome	12	3.21%	21.13%	2.65%
Chromosome	12	3.21%	−10.58%	3.59%
Intracellular transport vesicles	6	1.60%	−33.05%	2.39%
Vacuole or lysosome	13	3.48%	−5.18%	3.67%
Other subcellular localization	1	0.27%	−60.29%	0.68%
Extracellular / secretion	1	0.27%	−65.38%	0.78%
Centrosome	3	0.80%	37.93%	0.58%
Extracellular matrix component	0	0.00%	−100.00%	0.04%
SUM	374	100.00%	ND[b]	100.00%

[a]Enrichment (%) = (hot pepper TUGs - *Arabidopsis* genome)/*Arabidopsis* genome × 100.
[b]ND, Not determined.

database, proteins targeted to ER were dramatically increased (85% enrichment; Table 4).

DISCUSSION

Large-scale, single-pass sequencing of cDNAs, prepared from specific plant species or tissues, has evolved as an inexpensive and efficient gene-discovery tool for identifying novel cDNAs that encode proteins in specific environments or at various developmental stages (Cooke et al., 1996; Soares, 1997; Asamizu et al., 2000; White et al., 2000; Zhang et al., 2001; Haas et al., 2002). However, because EST approaches generally identify genes with moderately or highly abundant transcripts, the extreme redundancy that can result is a drawback to this technology. However, although our cDNA library was not normalized, EST redundancy was only 20.41%. Despite the rather small population size examined here, that rate is even lower than those of normalized cDNA libraries, where estimates range from 28 to 35% when approximately 1400 clones are considered (Carninci et al., 2000). Therefore, we have proven that our double-negative screening approach avoids high redundancy while being efficient, economical, and, technically, very simple.

Analyses of yeast have shown that expression levels are correlated with amino acid composition, peptide length, functional category, and subcellular localization (Drawid et al., 2000; Jansen and Gerstein, 2000; Greenbaum et al., 2002). For regulatory proteins, such as signaling molecules, transcription factors, and transmembrane proteins, expression is usually low, and is greatly depleted in the transcriptome. In contrast, because of its ability to detect low expression levels, double-negative screening is especially useful in isolating the large number of regulatory proteins that are difficult to discover by conventional methods. Therefore, our categories of regulatory proteins were enriched by more than 30% in metabolism, energy, protein synthesis, protein fate, cellular transport, cell defense, regulation of cellular environment, cell fate, and cellular organization (Table 3). Likewise, our eight-fold enrichment in the RNA-transport category supports previous research that suggests the RNA binding protein may have an important role in responses to abiotic stress (Gong et al., 2002; Guo et al., 2002). Our hot pepper TUGs also showed a larger quantity of functional category than that reported for the *Arabidopsis* genome -- 177% versus 135%. This higher value reflects the large portion of isolated clones that may have multiple functions.

In conclusion, our novel method is suitable for isolating genes for rarely identified transcripts, thereby greatly enriching regulatory factors. The ESTs isolated in our study should be further investigated to determine whether they actually code for those proteins that participate in the reaction to drought stress. Nonetheless, we believe this approach may increase the chance of

identifying those genes important to regulating drought tolerance.

ACKNOWLEDGEMENTS

This work was supported by a grant (code #CG1231) to CBH from the Crop Functional Genomics Center of the 21st Century Frontier Research Program, as funded by the Ministry of Science and Technology, and the Technology Development Program for Agriculture and Forestry, funded by the Ministry of Agriculture and Forestry, Republic of Korea. HJJ was partially supported by a BK21 Research Fellowship from the Ministry of Education, Korea.

Received February 25, 2003; accepted March 28, 2003.

LITERATURE CITED

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402

Arabidopsis Genome Initiative -- AGI (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796-815

Asamizu E, Nakamura Y, Sato S, Tabata S (2000) A large scale analysis of cDNA in *Arabidopsis thaliana*: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries. DNA Res 7: 175-180

Carninci P, Shibata Y, Hayatsu N, Sugahara Y, Shibata K, Itoh M, Konno H, Okazaki Y, Muramatsu M, Hayashizaki Y (2000) Normalization and subtraction of cap-trapper selected cDNAs to prepare full-length cDNA libraries for high-rate new gene discovery. Genome Res 10: 1617-1630

Choi KH, Hong CB, Kim WT (2002) Isolation and characterization of drought-induced cDNA clones from hot pepper. J Plant Biol 45: 212-218

Cooke R, Raynal M, Laudie M, Grellet F, Delseny M, Morris PC, Guerrier D, Giraudat J, Quigley F, Clabault G, Li YF, Mache R, Krivitzky M, Gy IJ, Kreis M, Lecharny A, Parmentier Y, Marbach J, Fleck J, Clement B, Philipps G, Herve C, Bardet C, Tremousaygue D, Hofte H (1996) Further progress towards a catalogue of all *Arabidopsis* genes: Analysis of a set of 5000 non-redundant ESTs. Plant J 9: 101-124

Cushman JC, Bohnert HJ (2000) Genomic approaches to plant stress tolerance. Curr Opin Plant Biol 3: 117-124

Drawid A, Jansen R, Gerstein M (2000) Genome-wide analysis relating expression level with protein subcellular localization. Trends Genet 16: 426-430

Fowler S, Thomashow MF (2002) *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. Plant Cell 14: 1675-1690

Frishman D, Albermann K, Hani J, Heumann K, Metanomski A, Zollner A, Mewes HW (2001) Functional and structural genomics using PEDANT. Bioinformatics 17: 44-57

Gong Z, Lee H, Xiong L, Jagendorf A, Stevenson B, Zhu JK (2002) RNA helicase-like protein as an early regulator of transcription factors for plant chilling and freezing tolerance. Proc Natl Acad Sci USA 99: 11507-11512

Greenbaum D, Jansen R, Gerstein M (2002) Analysis of mRNA expression and protein abundance data: An approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts. Bioinformatics 18: 585-596

Guo Y, Xiong L, Ishitani M, Zhu JK (2002) An *Arabidopsis* mutation in translation elongation factor 2 causes superinduction of CBF/DREB1 transcription factor genes but blocks the induction of their downstream targets under low temperature. Proc Natl Acad Sci USA 99: 7786-7791

Haas BJ, Volfovsky N, Town CD, Troukhan M, Alexandrov N, Feldmann KA, Flavell RB, White O, Salzberg SL (2002) Full-length messenger RNA sequences greatly improve genome annotation. Genome Biol 3: 0029.1-0029.12

Hasegawa PM, Bressan RA, Zhu JK, Bohnert HJ (2000) Plant cellular and molecular responses to high salinity. Annu Rev Plant Physiol Plant Mol Biol 51: 463-499

Huang X (1996) An improved sequence assembly program. Genomics 33: 21-31

Jansen R, Gerstein M (2000) Analysis of the yeast transcriptome with structural and functional categories: Characterizing highly expressed proteins. Nucleic Acids Res 28: 1481-1488

Knight H, Knight MR (2001) Abiotic stress signalling pathways: Specificity and cross-talk. Trends Plant Sci 6: 262-267

Martin GB (1998) Gene discovery for crop improvement. Curr Opin Biotech 9: 220-226

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

Mysore KS, Tuori RP, Martin GB (2001) *Arabidopsis* genome sequence as a tool for functional genomics in tomato. Genome Biol 2: 1003.1-1003.4

Ohlrogge J, Benning C (2000) Unraveling plant metabolism by EST analysis. Curr Opin Plant Biol 3: 224-228

Park JK, Hong CB (1991) A cDNA clone for red pepper (*Capsicum annuum* L.) actin. Mol Cells 1: 301-308

Ribaut JM, Banziger M, Hoisington D (2002) Genetic dissection and plant improvement under abiotic stress conditions: Drought tolerance in maize as an example. JIRCAS Working Report 85-92

Rice P, Longden I, Bleasby A (2000) EMBOSS: The Euro-

pean Molecular Biology Open Software Suite. Trends Genet 16: 276-277

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: A Laboratory Manual. 2nd ed, Cold Spring Harbor Laboratory Press, New York

Schoof H, Zaccaria P, Gundlach H, Lemcke K, Rudd S, Kolesov G, Arnold R, Mewes HW, Mayer KF (2002) MIPS *Arabidopsis thaliana* Database (MAtDB): An integrated biological knowledge resource based on the first complete plant genome. Nucleic Acids Res 30: 91-93

Seki M, Narusaka M, Abe H, Kasuga M, Yamaguchi-Shinozaki K, Carninci P, Hayashizaki Y, Shinozaki K (2001) Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses by using a full-length cDNA microarray. Plant Cell 13: 61-72

Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama K, Taji T, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K (2002) Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. Plant J 31: 279-292

Soares MB (1997) Identification and cloning of differentially expressed genes. Curr Biol 8: 542-546

van der Hoeven RV, Ronning C, Giovannoni JJ, Martin G, Tanksley SD (2002) Deductions about the number, organization and evolution of genes in the tomato genome based on analysis of a large EST collection and selective genomic sequencing. Plant Cell 14: 1441-1456

White J, Todd J, Newman T, Focks N, Girke T, Martinez de Ilarduya OI, Jaworski J, Ohlrogge J, Benning C (2000) A new set of *Arabidopsis* expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil. Plant Physiol 124: 1582-1594

Xing L, Brendel V (2000) MuSeqBox: A program for multi-query sequence BLAST output examination. Bioinformatics 17: 744-745

Xiong L, Schumaker KS, Zhu JK (2002) Cell signaling during cold, drought, and salt stress. Plant Cell S165-S183

Yamaguchi-Shinozaki K, Shinozaki K (2002) Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. Novartis Found Symp 236: 176-186

Zhang L, Ma XL, Zhang Q, Ma CL, Wang PP, Sun YF, Zhao YX, Zhang H (2001) Expressed sequence tags from a NaCl-treated *Suaeda salsa* cDNA library. Gene 267: 193-200

Zhu JK (2002) Salt and drought stress signal transduction in plants. Annu Rev Plant Biol 53: 247-73