J. Plant Biol. 40(1) : 1~7 (1997)

# Partial Sequence Analysis of Randomly Selected Watermelon [*Citrullus lanantus* (Thunb.) Mansf.] cDNA Clones

**Ok, Sung Han, Jeong Sheop Shin*, Young Soo Chung[1] and Hyok Ji Kwon[1]**

*Division of Plant Technology, Graduate School of Biotechnology,*
*[1]Department of Agronomy, College of Natural Resources,*
*Korea University, Seoul 136-701, Korea*

**An expressed sequence tag (EST) is simply a segment of a sequence over 150 bp from a randomly selected cDNA. EST helps to quickly identify functions of expressed genes and to understand the complexity of gene expression with database comparison. Sequencing of random cDNA clones can be applicable to discovery of new genes, mapping of the genome, identification of coding regions in genomic sequences, and antisense method. To accomplish these goals, in this research, randomly selected cDNA sequencing was performed with watermelon plant. Among 30 clones picked up and analyzed, all clones had an insert length over 0.5 kb. The average size of insert was about 1.3 kb. The size range of most cDNA insert was 1.0-2.0 kb. For sequence comparison, data from the coding region at 5' end of selected cDNA should be much more informative than those from the untranslated 3' tail. Thirty clones were sequenced from one end (5' end). Of these, 29 had no poly (A) tail in this direction, while only one was inverted. Thus, this library is over 96% unidirectional. Two clones had homologies to ribulose bisphosphate carboxylase/oxigenase (Rubisco) small subunit precursor gene. Thirteen cDNAs had high degree of sequence similarity to genes from other organisms. The remaining cDNA clones seem to be new genes not only in watermelon but also in all organisms.**

*Keywords*: Expressed sequence tag, watermelon, cDNA, DNA sequencing, Homology

Genome biology has its major goal in the identification of all of the functional genes in the genome and the structure of genome. An expressed sequence tag (EST) is simply a segment of a sequence over 150 bp from a randomly selected cDNA that corresponds to an mRNA. The expressed genes are ultimately responsible for the characteristics of organisms. Specific genes are important for characterizing the species taxonomically (Sasaki *et al.*, 1994). ESTs have also served as molecular genetic markers in genomic mapping (Kurata *et al.*, 1994; Shen *et al.*, 1994). Adams *et al.* (1991, 1992) were the first to show that, in humans, even partial sequencing of randomly chosen cDNA clones can provide a large amount of information concerning the genetic makeup of an organism and can generate sequence-tagged markers for the genome mapping. With almost 3,400 cDNA clones reported from human brain, about 17% of the sequenced clones coded for already reported gene products, including genes reported from *Escherichia coli, Saccharomyces cerevisiae, Caenorhabditis elegans, Drosophila melanogaster* and several plants.

In the area of plant species, two major cDNA sequencing projects have been conducted in *Arabidopsis* and in rice (Uchiyama *et al.*, 1992; Hofte *et al.*, 1993; Newman *et al.*, 1994; Sasaki *et al.*, 1994). Over 21,044 *Arabidopsis* and almost 11,015 rice sequences had been registered in dbEST. Approximately 32% of EST cDNA clones of *Arabidopsis* and 35% rice have sequence similarity to known proteins from either microbes, plants, or animals. Plant ESTs also have been reported in the published literature for 200 maize cDNAs (Keith *et al.*, 1993) and 197 *Brassica napus* cDNAs (Park *et*

*Corresponding author: Fax +82-2-927-9028

*al.*, 1993). Various other organisms, such as nematode (McCombie *et al.*, 1992; Waterson *et al.*, 1992), mouse (Hoog., 1991) have also been examined by extensive sequencing of randomly selected cDNA clones. Recently, partial sequencing of 1,216 randomly selected cDNA clones from chinese cabbage flower buds and classification of these clones based on the biological functions of the encoded proteins were reported The enormous accumulation of ESTs has led to the establishment of dbEST (Boguski *et al.*, 1993, 1994). The processes of searching, retrieving, and submitting ESTs have been greatly facilitated by e-mail or internet file transfer protocol (Boguski *et al.*, 1993, 1994; Newman *et al.*, 1994).

Plants, in addition to being important targets for improvement by genetic engineering, provide opportunities to study distinct biological aspects, including plant-specific developmental patterns and intricate associations with many environmental factors such as plant-microbe interactions and unique defense processes. Despite of recent technological improvements and a large effort to isolate plant genes, plant gene resources are still very limited. Current approaches for isolating plant genes, such as chromosome walking or gene tagging, are, in most cases, fairly expensive and time-consuming and do not provide enough plant gene resources. To address genes of unknown function in the cDNA sequencing projects is to define highly conserved domains or structural motives among homologous genes from heterologous organisms. Such an approach may be possible once a large number of genes are compiled from many different species.

To accomplish these goals, in this research, randomly selected cDNA sequencing was performed with watermelon plant. Watermelon [*Citrullus lanantus* (Thunb.) Mansf.] is the one of important vegetables, which can compete with imported vegetables and has a best palatability for the most Korean. Considering the rapid development of gene characterization in various crops, this approach will contribute the completion of goals in genome biology of watermelon.

## MATERIALS AND METHODS

### Plant Materials

Commercial $F_1$ hybrid seeds of watermelon [*Citrullus lanantus* (Thunb.) Mansf.] were used. The plants of watermelon were grown in the pots. The day/night temperature was 32°C/24°C. Leaves of 8-week-old leaves were collected for isolation of mRNA to construct cDNA library.

### Total and poly(A)$^+$ RNA Isolation

Total RNA and poly(A)$^+$ RNA were extracted from leaf tissue according to the method described by Hong *et al.* (1996). The quality of RNA was checked by UV spectrophotometer ($OD_{260/280}$=1.7-2.0) and 1.0% agarose-formaldehyde gel (rRNA bands should be obvious and distinct).

### Construction of cDNA Library

A commercial complementary DNA synthesis kit purchased from Stratagene was used to construct a proper watermelon cDNA library. To produce single-stranded cDNA appropriate for directional cloning, 5 μg of poly(A)$^+$ RNA was primed with a oligo(dT) primer having a *Xho*I site at the 3' end. Double-stranded cDNA was synthesized by RNase H and *E. coli* DNA polymerase. A pool of synthesized double-stranded cDNAs was fractionated through sephacryl S-400 spin column. After ligation of *Eco*RI linkers onto the cDNA, it was digested with *Eco*RI and *Xho*I, and finally ligated into an *Eco*RI-*Xho*I-cut λ Uni-ZAP vector. Ligated DNA was packaged *in vitro* using a commercial extract. This cDNA library contained $1\times10^7$ primary plaques and was amplified to a titer of $8\times10^{10}$ pfu/ml.

### Template Preparation and Nuclotide Sequencing

Phagemid templates for the 30 clones from the cDNA library were produced by randomly picking each single plaques from the primary library onto a lawn of *E. coli* SOLR™ strain, which had been plated on LB agar containing 100 μg/ml ampicillin. The resulting amp$^r$ colonies were grown at 37°C in 5-ml cultures of LB medium containing 100 μg/ml ampicillin. Phagemids were extracted according to the method described by Hong *et al.* (1996). Single-stranded template was prepared using VCSM13 helper phage according to Stratagene's single-stranded rescue protocol.

Manual nucleotide sequencing was performed by the standard dideoxy chain termination method on single stranded DNA using the Sequenase version 2.0 DNA-sequencing kit (USB, U.S.A.) and [α-$^{35}$S] dATP (DuPont, U.S.A.) with $T_3$ promoter primer. For double strand phagemid sequencing, the sequencing reaction was performed with TOP™ DNA cycle sequencing system (Korea Biotech. Inc., Korea). Those nucleotide sequencing reaction were carried out according to the method described by Hong *et al.* (1996).

### Computer Analysis of Sequences

Sequences were edited manually to remove vector and ambiguous sequences at the ends. In general, the sequence information of insert was obtained from approximately 150-300 nuclotides of each clone. Nucleotide and amino acid sequence analysis was performed using the DNASIS (Hitachi, Japan) and PCGENE software (Intelligenetics, Switzerland). The insert sequence was compared to GenBank entries first at the amino acid level with the BLASTX subroutine and at the nucleotide level by the BLASTN subroutine (Altschul *et al.*, 1990). Similarity was considered significant when the percentage of identity was higher than 40% at the amino acid level or if the BLASTX PAM120 score was greater than 80.

## RESULTS AND DISCUSSION

### Characterization of cDNA library

A watermelon cDNA library was constructed using 5 μg of poly(A)$^+$ mRNA isolated from watermelon leaves. The recombination rate was higher than 95%. The primary library contained $1\times10^7$ recombinant phage and after plaque amplification, the serial titering of amplified plaques showed that the library contains approximately $8\times10^{10}$ pfu/ml SM buffer

**Table 1.** Putatively identified watermelon ESTs by the database search

| Clone | putative identification | species[a] | database[b] | No.[c] | length[d] | identity (%, a.a.[e]) | score | size of insert (kbp) |
|---|---|---|---|---|---|---|---|---|
| 1 | EDPG precursor | CA | GI | 285741 | 43 | 68 | 176 | 1.3 |
| 2 | carbonic anhydrase | A | SP | P27140 | 36 | 66 | 135 | 1.0 |
| 3 | ribulose bisphosphate carboxylase small chain precursor | CU | SP | P08474 | 39 | 82 | 175 | 1.8 |
| 4 | unmatched | | | | | | | 1.5 |
| 5 | unmatched | | | | | | | 1.0 |
| 6 | unmatched | | | | | | | 0.5 |
| 7 | unmatched | | | | | | | 0.5 |
| 8 | unmatched | | | | | | | 1.0 |
| 9 | ribulose bisphosphate carboxylase | CU | SP | P24007 | 65 | 72 | 253 | 2.0 |
| 10 | small chain precursor shaggy-like protein kinase | TO | PIR | S42085 | 58 | 74 | 242 | 1.5 |
| 11 | chloroplast elongation factor TuB | TO | GI | 218312 | 41 | 97 | 217 | 1.5 |
| 12 | unmatched | | | | | | | 0.5 |
| 13 | unmatched | | | | | | | 3.0 |
| 14 | unmatched | | | | | | | 0.6 |
| 15 | unmatched | | | | | | | 1.0 |
| 16 | unmatched | | | | | | | 1.5 |
| 17 | NAM gene product | PE | GN1/PID | e205713 | 25 | 68 | 101 | 1.0 |
| 18 | unmatched | | | | | | | 1.8 |
| 19 | unmatched | | | | | | | 1.5 |
| 20 | phospho-2-dehydro-3- deoxy heptonate aldolase 1 precursor | TOM | SP | P37215 | 32 | 71 | 134 | 2.0 |
| 21 | unmatched | | | | | | | 1.7 |
| 22 | unmatched | | | | | | | 1.1 |
| 23 | lipase | A | GI | 1145627 | 52 | 59 | 172 | 1.6 |
| 24 | unmatched | | | | | | | 1.2 |
| 25 | unmatched | | | | | | | 1.6 |
| 26 | 60S ribosomal protein LIB | A | GI | 17281 | 56 | 67 | 204 | 0.8 |
| 27 | putative small subunit of TFIIA gene | A | GN1/PID | e253416 | 77 | 84 | 334 | 1.2 |
| 28 | unmatched | | | | | | | 1.1 |
| 29 | receptor-like kinase | A | GI | 1405837 | 61 | 45 | 135 | 0.9 |
| 30 | ribulose bisphosphate carboxylase/ oxygenase activase precursor | CU | SP | Q01587 | 72 | 88 | 358 | 1.5 |

[a]Species abbreviations: A, *Arabidopsis*; CA, *Carrot*; CU, *Cucumber*; PE, *Petunia*; TO, *Tobacco*; TOM, *Tomato*.
[b]Data base (DB) abbreviations: GI, NCBI-gi; PIR, Protein Identification Resources; GN1/PID, EMBL; SP, SwissProt.
[c]Number indicates database accession number of the matched sequences.
[d]Length indicates length of identical or similar amino acid residues for peptide matches.
[e]a.a. means amino acid.

and was considered to have adequate representation of the expressed genes. A suitable library will have a low percentage of clones with no inserts or very short inserts. Among 30 clones picked up and analyzed, all clones had an insert length over 0.5 kb. This length was long enough to do similarity search after translating the nucleotide sequence to the amino acid sequences. The size distribution of the 30 clones was 0.5 to 3.0 kb, with an average value of 1.3 kb, and with 24 clones being larger than 1 kb. The size range of most cDNA inserted was 1.0-2.0 kb, probably reflecting the abundance of cDNAs of this size in the library used. The individual sizes of clones are listed in the Table 1, and presented in Fig. 1.

For sequence comparisons, data from the coding region at 5' end of a poly (A)-selected cDNA should be much more informative than those from the untranslated 3' tail. For all of the clones from this library, the sequences were obtained only from the putative 5' end of the cDNA to enhance the probability of obtaining coding sequence. To confirm the directionality of this library, 30 clones were sequenced from one end, using $T_3$ promoter primer. Of these, 29 had no poly (A) tail in this direction, while only one (clone 19) was inverted. Thus, this library is over 96% unidirectional. Potential contamination of a cDNA library with genomic sequence is a matter of concern (Burglin and Barnes, 1992), but in this experiment the presence of a poly (A) tail at the 3' end confirms that this library is essentially free of such sequences.

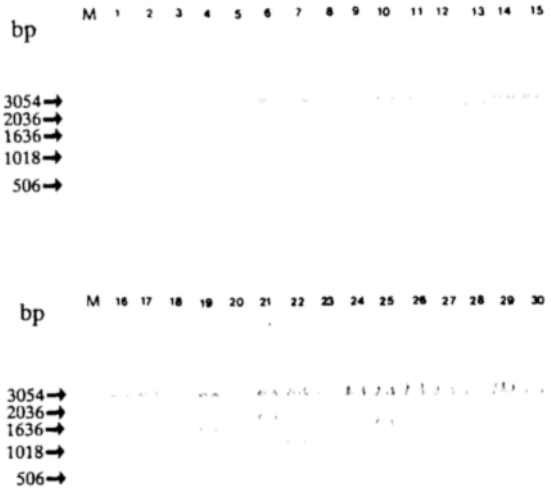


**Fig. 1.** Restriction digestion of 30 sequenced cDNAs with *BstX*I and *Xho*I. M: 1kb ladder molecular weight marker. Arrows indicate the molecular weight of DNA fragments.

## Estimation of Insert Redundancy

In any cDNA sequencing project, there will come a point at which the percentage of previously analyzed cDNAs makes further sequencing unproductive. The use of differentially selected clones as opposed to randomly selected clones has been shown to extend the usefulness of the library (Hoog, 1991), as has the preparation of "sorted" libraries (Waterson *et al.*, 1992). In this experiment, to identify insert dulpication, the clones which were assigned to have the putatively same function at GenBank and PIR database were analyzed by the program PCGENE. Two clones had a homology to ribulose bisphosphate carboxylase/oxigenase (Rubisco) small chain precursor gene in several plants. In analyzing the two clones, it was found that those clones were the same (Fig. not shown). Remaining 28 clones had no match of same gene. Although small number of clones were sequenced and analyzed, this result showed that this general cDNA library had no problem of insert redundancy.

## Sequencing errors

Single strand sequencing is not particularly accurate. Although it is not possible to determine the exact error rate for a specific single strand sequence without having the double strand sequence for comparison, error rate estimates of 2% (Adams *et al.*, 1991) and 0.5% (Hoog, 1991) have been made for entire projects.

In this research, each single strand cDNA sequence was not compared to double strand sequence and previously sequenced watermelon cDNA in the database. Only one single strand sequence (clone 1) was compared to the same double strand cDNA sequence (Fig. 2). In this comparison, 0.3% ambiguity was found. Since the database of watermelon cDNA clones is very limited, estimation of sequencing accuracy in comparison with the previously sequenced database was impossible. In the near future, watermelon cDNA clones large-scale sequencing will be performed in our lab. This means that another plant gene sequence database for genome biology will be added, and this will be able to compare the previously sequenced database.

## Sequence comparisons

GenBank and PIR searches were conducted with the "Basic Local Alignment Search Tool" programs for nucleotide (BLASTN) and peptide (BLASTX)

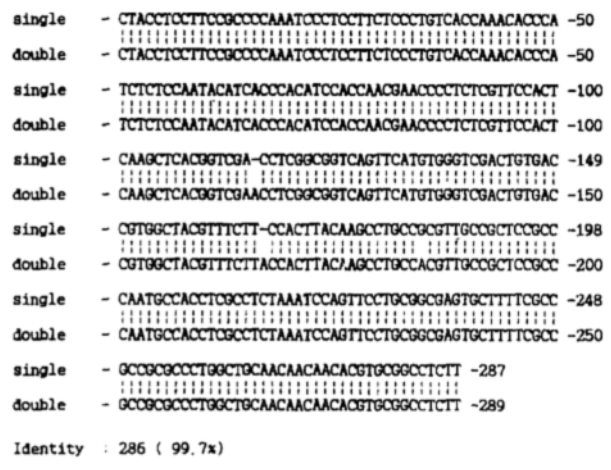


**Fig. 2.** Comparison of single- and double-stranded sequences of putatve EDPG precursor (clone 1).

comparisons (Altschul *et al.*, 1990). Sequence similarities identified by the BLAST programs were considered statistically significant with a Poisson p-value < 0.01. The Poisson p-value is the probability of as high a score occuring by chance, given the number of residues in the query sequence and the database. Deduced amino acid sequence homology between a cDNA and a known sequence was deemed significant if the BLAST PAM 120 score was greater than 80. All 30 sequences were automatically translated in the three reading frames and compared with the protein sequence database in GenBank using the subroutine BLASTX. If no significant homology was found, the sequences were subsequently compared at the nucleotide level using BLASTN. It was found that BLASTN subroutine useful for detecting similarities to randomly sequenced cDNAs from other organisms (Shen *et al.*, 1994).

Database comparisons of 30 cDNAs (using BLASTX) revealed that 13 cDNAs had a high degree of sequence similarity to genes from other organisms (Table 1). These cDNAs have been provisionally characterized based on the following criteria : (a) PAM 120 scores over 80, (b) the prescence of appropriate conserved residues and (c) the biological " appropriateness" of the identification. Although none of these criteria are absolute and biological appropriateness is not quantifiable, use of this approach is likely to yield fewer incorrect assignments (Keith *et al.*, 1993).

The cDNA clones with significant similarities to the protein-coding sequences in the PIR and GenBank databases are listed in the Table 1. Thus, 43% of the sequenced cDNAs showed clear similarity to identifiable genes. This percentage is slightly lower than that of reported for randomly selected human brain cDNAs (52%; Adams *et al.*, 1991) or randomly selected mouse cDNAs (44%; Hoog, 1991). However, differential selection from a mouse library to yield a "rare" group of cDNAs also gave a low percentage (25%) of identifiable genes (Hoog, 1991). Likewise, an analysis of over 2000 human genes from several types of libraries resulted in the identification of 17% of the genes (Adams *et al.*, 1992). Identification rate in *Caenorhabditis elegans* cDNA projects are also in the range of 25 to 30% (McCombie *et al.*, 1992; Waterson *et al.*, 1992). As more watermelon cDNAs are sequenced, can the percentage of identified cDNAs maintain or even increase? This may seem unlikely because the pool of identified plant genes is relatively small.

In looking through Table 1, it can be seen that a very diverse group of gene types have been identified, despite the somewhat lower number of clones actually sequenced. Two clones (clone 26 and 27) were identified as 'housekeeping' genes, i.e. genes that might be expected to be found in any cell type, with representatives from various metabolic pathways such as glycolysis, carbohydrate metabolism, transcription, translation and protein degradation, etc. Clone 26 had a 52.2% homology to 60S ribosomal protein. Ribosomal proteins are fundamental proteins for living systems and have been studied in detail for their amino acid sequences (Sasaki *et al.*, 1994). Clone 27 was identified to be the small subunit of transcription factor IIA (TFIIA). Initiation of mammalian mRNA synthesis requires at least eight general transcription initiation factors, termed TFIIA, B, D, E, F, G, H, and J. Three genes (clone 1, 10 and 29) were also identified to be involved in signal transduction. Clone 1 had a 42.6% similarity in EDGP (extracelluar dermal glycoprotein) precursor. This protein, previously called GP 57, is now referred to as EDGP. Clone 10 had a 42.3% homology to the shaggy like protein kinase. Shaggy belongs to the class of segment-polarity genes, necessary for correct pattern formation within each segment. It also plays a role in epidermal cell differentiation in the adult fly (Bourois *et al.*, 1989; Simpson *et al.*, 1988). Clone 29 showed 41.7% similarity to an *Arabidopsis* receptor-like serine/threonine protein kinase. Recent reports on the cloning and sequencing of plant genes showing sequence similarities to animal receptor kinases suggested that in plants, as like other eukaryotes, cell surface receptors of the protein kinase family most likely play a fundamental role in

signal transduction (Stone and Walker, 1995).

Other clones exhibited high similarity to genes which would be suspected to tissue- or situation-specific. Clone 2 had a 38.2% homology to carbonic anhydrase (CA). CA catalyzes the reversible hydration of $CO_2$ to bicarbonate and is one of the most abundant soluble proteins in the leaves of $C_3$ higher plants. Most localization studies indicate that CA is found in the chloroplast of $C_3$ plants and primarily within the cytosol of mesophyll cells of $C_4$ species. Clone 3 and 9 showed 47.3% similarity to a ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) small subunit as mentioned previously. Copy number of the gene is four to six. Clone 30 had a 67.4% homology to a ribulose-1,5-bisphosphate carboxylase/oxygenase activase gene. Rubisco activation state is regulated by Rubisco activase, a soluble chloroplast enzyme (Salvucci *et al.*, 1985; Portis *et al.*, 1987). Clone 11 showed a 84.2% similarity to chloroplast elongation factor Tu. Elogation factor Tu plays an essential role in protein synthesis; it binds aminoacyl tRNAs and GTP, associates with the A site of ribosomes, and consequently directs the elongation of polypeptides on ribosomes. Clone 17 had a 38.2% homology with 'No Apical Meristem' (*nam*) gene of Petunia. The *nam* gene is required for SAM formation during embryogenesis. Clone 20 showed a 44.3% similarity to 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase. DAHP synthase is the first enzyme of the shikimate pathway in plants and bacteria. Clone 23 had a 47.9% homology to lipase in *Arabidopsis thaliana*. True lipases, which attack the fatty acyl linkage of water-insoluble triacylglycerols, are present in high activities in oil seeds (Galliard, 1980). Results from this analysis are essentially hypotheses that must be tested by other criteria.

Our results indicated that most identified cDNAs with high similarity to known proteins in the abundantly expressed group than that of rare group. This is probably not too surprising in that the abundantly expressed genes are generally easier to isolate. In mouse, comparison of ESTs to the database found that 90% of abundantly expressed sequences detected a significant similarity while only 25% of sequences in the rarely expressed group yielded such a result (Hoog, 1991). One of the ways in which this way be accomplished is by the construction of a normalized library in which all cDNA clones are represented with similar abundance. The remaining 57% of the sequenced cDNAs seem to be new genes not only in watermelon but also in all organisms. It is necessary that further analysis has to be done to identify the function and other informations of the clones. The clones which are matched with the database in this experiment will be submitted to dbEST. This will contribute to accomplish the goal of genome biology.

## LITERATURE CITED

**Adams, M.D., M. Dubnick, A.R. Kerlavage, R. Moreno, J.M. Kelly, T.R. Utterback, J.W. Nagle, C. Fields and J.C. Venter.** 1992. Sequence identification of 2, 375 human brain genes. *Nature* **355**: 632-634.

**Adams, M.D., J.M. Kelly, J.D. Gocayne, M. Dubnick, M.H. Polymeropoulos, H. Xiao, C.R. Merril, A. Wu, B. Olde, R.F. Moreno, A.R. Kerlavage, W.R. McCombie and J.C. Venter.** 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**: 1651-1656.

**Altschul, S.F., W. Gish, W. Miller, E.W. Myers and D. Lipman.** 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 402-410.

**Boguski, M.S., T.M.J. Lowe and C.M. Tolstoshev.** 1993. dbEST-database for "expressed sequence tags." *Nature Genet.* **4**: 332-333.

**Boguski, M.S., C.M. Tolstoshev and D.E. Bassett.** 1994. Gene discovery in dbEST. *Science* **265**: 1993-1994.

**Bourois, M., M. Heitzler, M. El Messal and P. Simpson.** 1989. Mutant Drosophila embryo in which all cells adopt a neural fate. *Nature* **341**: 442-444.

**Burglin, T.R. and T.M. Barnes.** 1992. Introns in sequence tags. *Nature* **357**: 367.

**Galliard, T.** 1980. *In The Biochemistry of Plants.* Vol. 4. P.K. Stumpf and E.E. Conn (eds.). Academic Press, New york. pp. 85-116.

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

**Hong, S.H., J.S. Shin, J.U. Jeung and S.H. Ok.** 1996. Molecular cloning and nucleotide sequencing of a DNA clone encoding arginine decarboxylase in rice (*Oryza sativa* L.). *Agr. Chem. Biotech.* **39**: 112-117.

**Hoog, C.** 1991. Isolation of a large number of novel mammalian genes by a differential cDNA library screening strategy. *Nucleic Acids Res.* **19**: 6123-6127.

**Keith, C.S., D.O. Hoang, B.M. Barrett, B. Feigelman, M.C. Nelson, H. Thai and Baysdorfer, C.** 1993. Partial sequence analysis of 130 randomly selected maize cDNA clones. *Plant Physiol.* **101**: 329-332.

**Kurata, N., Y. Nagamura, K. Yamamoto, Y. Harushima, N. Sue, J., Wu, B.A. Antonio, A. Shormura, T.**

**Shimuzu, S-Y. Lin, T. Inoue, A. Fukuda, T. Shimano, Y. Kuboki, T. Toyama, Y. Miyamoto, K. Hayasaka, A. Miyao, L. Monna, H.S. Zhong, Y. Tamura, Z-X. Wang, T. Momma, Y. Umehara, M. Yano, T. Sasaki and Y. Minobe.** 1994. A 300 kilobase interval genetic map of rice including 883 expressed sequences. *Nature Genet.* **8**: 365-372.

**McCombie, W.R., M.D. Adams, J.M. Kelly, M.G. FitzGerald, T.R. Utterback, M. Khan, M. Dubnick, A.R. Kerlavage, J.C. Venter and C. Fields.** 1992. Caenorhabditis elegans expressed sequence tags identify gene families and potential disease gene homologoues. *Nature Genet.* **3**: 124-131.

**Newman, T., F.J. de Bruijn, P. Green, K. Keegstra, H. Kende, L. McIntosh, J. Ohlrogge, N. Raikhel, S. Somerville, M. Thomashow, E. Retzel and C. Sommerville.** 1994. Genes galore: a summary of methods for accessing results from large-scale sequencing of anonymous *Arabidopsis* cDNA clones. *Plant Physiol.* **103**: 359-370.

**Park, Y.S., J.M. Kwak, O.Y. Kwon, Y.S. Kim, D.S. Lee, M.J. Cho, H.H. Lee and H.G. Nam.** 1993. Generation of expressed sequence tags of random root cDNA clones of *Brassica napus* by single-run partial sequencing. *Plant Physiol.* **103**: 359-370.

**Portis, A.R. Jr., M.E. Salvucci and W.L. Ogren.** 1987. Activation of ribulose bisphosphate carboxylase/oxygenase at physiological $CO_2$ and ribulose bisphosphate concentrations by rubisco activase. *Plant Physiol.* **82**: 967-991.

**Salvucci, M.E., A.R. Portis Jr. and W.L. Ogren.** 1985. A soluble chloroplast protein catalyzes ribulose bisphosphate carboxylase/oxygenase activation *in vivo*. *Photosynth. Res.* **7**: 193-201.

**Sasaki, T., J. Song, Y. Koga-Ban, E. Matsui, F. Fang, H. Higo, H. Nagasaki, M. Hori, M. Miya, E. Murayama-Kayano, T. Takiguchi, A. Takasuga, T. Niki, K. Ishimaru, H. Ikeda, Y. Yamamoto, Y. Mukai, I. Ohta, N. Miyadera, I. Havukkala and Y. Minobe.** 1994. Toward cataloguing all rice genes: large-scale sequencing of randomly chosen rice cDNAs from callus cDNA library. *Plant J.* **6**: 615-624.

**Shen, B., N. Carneiro, I. Torres-Jerez, B. Stevenson, T. McCreery, T. Helentjaris, C. Baysdorfer, E. Almira, R.J. Ferl, J.E. Habben and B. Larkins.** 1994. Partial sequencing and mapping of clones from two maize cDNA libraries. *Plant Mol. Biol.* **26**: 1085-1101.

**Simpson, P., M. El Messai, M. Moscoso del Prado and P. Ripoll.** 1988. Stripes of positional homologies across the wing blade of *Drosophila melanogaster*. *Development* **103**: 391-401.

**Stone, J.M. and J.C. Walker.** 1995. Plant protein kinase families and signal transduction. *Plant Physiol.* **108**: 451-457.

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

**Waterson, R., C. Martin, M. Craxton, A. Coulson, L. Hillier, R. Durbin, P. Green, R. Shownkeen, N. Halloran, M. Metzstein, T. Hawkins, R. Wilson, M. Berks, Z. Du, K. Thomas, J. Thierry-Mieg and J. Sulston.** 1992. A survey of expressed genes in *Caenorhabditis elegans*. *Nature Genet.* **1**: 114-123.

(Received January 9, 1997)