GENETICS AND GENOMICS

S. Ok · Y.S. Chung · B.Y. Um · M.S. Park
J.-M. Bae · S.J. Lee · J.S. Shin

Identification of expressed sequence tags of watermelon 
(Citrullus lanatus) leaf at the vegetative stage

Received: 10 July 1999 / Revision received: 20 December 1999 / Accepted: 11 January 2000

Abstract A cDNA library was constructed using mRNA prepared from leaves of watermelon [Citrullus lanatus (Thunb.) Matsum&Nakai] at the vegetative stage. Randomly selected cDNA clones were sequenced in order to identify potentially informative genes. Database comparisons indicated that out of the 704 watermelon cDNA clones, 399 clones (56.7%) revealed a high degree of sequence similarity to genes from other organisms. These expressed sequence tag clones were divided into ten categories depending upon gene function. Since this kind of experiment has not previously been carried out in this genome, random nucleotide sequencing of these cDNAs could contribute considerable information concerning the novel genes in this organism.

Key words Expressed sequence tags · cDNA sequence · Watermelon · Citrullus lanatus

Introduction

An expressed sequence tag (EST) is a partial sequence from a randomly selected cDNA. EST data-bases have been constructed from many organisms (Adams et al. 1991; Hoog 1991; McCombie et al. 1992; Waterson et al. 1992; Boguski et al. 1993; Sasaki et al. 1994; Lim et al. 1996; Covitz et al. 1998; Wood et al. 1999; Machuka et al. 1999). Sequence analysis of these ESTs helps to identify the functions of expressed genes rapidly and to reveal the complexity of gene expression (Lim et al. 1996). cDNA clones can also be used as molecular markers in genomic mapping (Kurata et al. 1994; Shen et al. 1994).

EST databases have been established from many plant species, including Arabidopsis thaliana (Höfte et al. 1993; Newman et al. 1994), Oryza sativa (Uchimiya et al. 1992; Sasaki et al. 1994), Zea mays (Keith et al. 1993), Brassica napus (Park et al. 1993), Brassica campestris (Lim et al. 1996) and Medicago truncatula (Covitz et al. 1998). More recently, partial sequences of 754 randomly selected cDNA clones from the leaf of Brassica napus were reported (Lee et al. 1998), and these clones were classified based on the biological functions of the encoded proteins. In that report, 204 ESTs had significant amino acid sequence similarity to the sequences in the PIR and SwissProt protein databases. Since the dbEST was established (Boguski et al. 1993, 1994), 3011754 ESTs had been registered up to September 1999. This database makes it possible to compare a large number of genes and the proteins they encode between animal and plant species.

In this paper, we report the partial sequencing and database comparison of randomly selected vegetative leaf cDNA clones of watermelon. These data identify a large number of genes that are expressed in the vegetative leaves in this important crop species. Moreover, considering the rapid development of gene characterization in other species, further cDNA screening will facilitate the isolation of many agronomically important genes in watermelon.

Communicated by C.F. Quiros

All ESTs in this article are registered in dbEST (Accession No.: AA660011–AA660163, A1563040–A1563569)

S. Ok · M.S. Park · J.-M. Bae · S.J. Lee · J.S. Shin (✉)
Graduate School of Biotechnology, Korea University,
Seoul 136-701, Korea
e-mail: jsshin@mail.korea.ac.kr
Fax: +82-2-9279028

Y.S. Chung
Division of Bioresources, Faculty of Life Sciences and Resources, Dong-A University, Pusan 604-714, Korea
B.Y. Um
Department of Horticultural Science, Korea University,
Seoul 136-701, Korea
Materials and methods

Plant material and construction of the cDNA library

Seeds of the commercial F1 watermelon [Citrullus lanatus (Thunb.) Matsum&Nakai] hybrid Lucky were used. Plants were grown in pots under natural light in a greenhouse. The day/night temperature was 32/24°C. Eight-week-old leaves were harvested during the vegetative stage. Total RNA and poly(A)+ RNA were extracted from leaf tissue according to the method of Hong et al. (1996). The amount and quality of RNA was checked by UV spectrophotometer (OD260/280 = 1.7–2.0). A commercial cDNA synthesis kit was used to construct the library according to the manufacturer’s instructions (Stratagene, USA). To produce single-stranded cDNA appropriate for directional cloning, 5 μg of poly(A)+ RNA was primed with an oligo(dT) primer having a XhoI site at the 3’ end. Double-stranded cDNA was produced using RNase H and Escherichia coli DNA polymerase. The double-stranded cDNAs were then fractionated through a Sephacryl S-400 spin column. After ligation of EcoRI linkers, cDNA was digested with EcoRI and XhoI and ligated into an EcoRI-XhoI cut lambda Uni-ZAP vector. Ligated DNA was packaged in vitro using a commercial packaging extract (Stratagene). This cDNA library contained 1 × 108 primary plaques and was amplified to a titer of 8 × 1010 pfu/ml.

Template preparation and nucleotide sequencing

Plaques (740) were picked at random, the phagemids were excised and then plated onto LB agar containing 100 μg/ml ampicillin. The resulting AmpR colonies were grown at 37°C in 5-ml cultures of LB medium containing 100 μg/ml ampicillin. Phagemid DNA was purified according to the method described by Hong et al. (1996). Using a Perkin-Elmer 9600 thermal cycler and an ABI 310 Genetic Analyzer (Applied Biosystems, USA), the 5’ ends of the cDNA clones were sequenced using the T3 promoter primer according to the thermal cycling protocol of the Taq Dye Terminator Cycle Sequencing kit (Applied Biosystems).

Computer analysis of sequences and homology comparison

Sequences were edited (Navigator, Applied Biosystems) to remove vector and ambiguous sequences at either end. Approximately 250–500 nucleotides of insert sequence information were obtained from each clone. Sequences were compared with GenBank entries at the amino acid level with the BLASTX subrou- tine and at the nucleotide level by the BLASTN subroutine (Altschul et al. 1990). Matches were considered significant when the percentage identity was higher than 40% at the amino acid level or when the BLASTX PAM120 score was greater than 80 (Newman et al. 1994).

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 primary library contained 1 × 107 recombinant phages and after plaque amplification, serial tittering showed that the library contained approximately 8 × 1010 pfu/ml SM buffer. Libraries should have a low percentage of clones with no inserts or very short inserts. Of the 740 clones analyzed, all had an insert length greater than 0.5 kb. The size distribution of the 740 clones was 0.5–3.0 kb, with an average of 1.3 kb, and with 516 clones being larger than 1 kb. The most abundant size range of cDNA was 1.0–2.0 kb, probably reflecting the abundance of cDNAs of this size in the library.

For sequence comparisons, data from the coding region are more informative than data from the untranslated 3’ tail. Hence, all sequence data were obtained from the 5’ end of the cDNA to enhance the probability of obtaining coding sequence. Since all 34 clones of the Rubisco small subunit that were sequenced from both ends showed the appropriate orientation, it was concluded that this library is uni-directional. Potential contamination of a cDNA library by genomic DNA is a matter of concern (Burglin and Barnes 1992), but in this library the universal presence of a poly (A) tail at the 3’ end shows that it is essentially free of such sequences.

Characterization of ESTs

Among 740 cDNA clones, 704 gave usable sequencess. Database comparisons of the PIR and GenBank databases revealed that 399 cDNAs out of 704 (56.7%) showed a high degree of sequence similarity to genes from other organisms (Table 1). Of the 399 database-matched clones, 365 represented similarities to known plant genes and the remaining 34 clones matched non-plant genes. The remaining 43.3% (305) of the clones did not have a database match. Further analysis is required to identify the function of these clones.

A diverse group of genes has been identified in watermelon. As expected, many of the identified cDNAs were housekeeping genes which are related to metabolism (Table 2). Seventy were genes involved in photosynthesis. Clones encoding ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit (34 clones), which is the key regulatory enzyme of photosynthetic carbon assimilation were the most abundant clones in this experiment. Several other genes related to secondary metabolism, cell wall metabolism and protein biosynthesis were also present. These results are similar to that of Brassica (Lee et al. 1998) and show that the cells in vegetative leaves are metabolically quite active.

Defense-related genes are also highly expressed in watermelon leaves as reported by others (Park et al. 1993; Sasaki et al. 1994; Lim et al. 1996; Covitz et al. 1998; Lee et al. 1998). In addition, two clones (AA660070 and A1563221) matched genes that promote flowering, col2 and CONSTANS, respectively. The presumed reason for the appearance of these genes is that the sample leaves were at the initial stage