

A. Levi · A. Davis · A. Hernandez · P. Wechter ·  
J. Thimmapuram · T. Trebitsh · Y. Tadmor ·  
N. Katzir · V. Portnoy · S. King

## Genes expressed during the development and ripening of watermelon fruit

Received: 27 December 2005 / Revised: 1 March 2006 / Accepted: 25 March 2006 / Published online: 27 June 2006  
© Springer-Verlag 2006

**Abstract** A normalized cDNA library was constructed using watermelon flesh mRNA from three distinct developmental time-points and was subtracted by hybridization with leaf cDNA. Random cDNA clones of the watermelon flesh subtraction library were sequenced from the 5' end in order to identify potentially informative genes associated with fruit setting, development, and ripening. One-thousand and forty-six 5'-end sequences (expressed sequence tags; ESTs) were assembled into 832 non-redundant sequences, designated as “EST-unigenes”. Of these 832 “EST-unigenes”, 254 (~30%) have no significant homology to sequences published so far for other plant species. Additionally, 168 “EST-unigenes” (~20%) correspond to genes with unknown function, whereas 410 “EST-unigenes” (~50%) correspond to genes with known function in other plant species. These “EST-unigenes” are

mainly associated with metabolism, membrane transport, cytoskeleton synthesis and structure, cell wall formation and cell division, signal transduction, nucleic acid binding and transcription factors, defense and stress response, and secondary metabolism. This study provides the scientific community with novel genetic information for watermelon as well as an expanded pool of genes associated with fruit development in watermelon. These genes will be useful targets in future genetic and functional genomic studies of watermelon and its development.

**Keywords** cDNA · *Citrullus* · Expressed sequence tags · Gene expression · Watermelon

Communicated by P. Puigdoménech

A. Levi (✉) · P. Wechter  
USDA, ARS, U.S. Vegetable Laboratory,  
2700 Savannah Highway, Charleston, SC 29414, USA  
e-mail: alevi@saa.ars.usda.gov

A. Davis  
USDA, ARS,  
P.O. Box 159, Lane, OK 74555, USA

A. Hernandez · J. Thimmapuram  
University of Illinois at Urbana-Champaign,  
Roy J. Carver Biotechnology Center, W.M. Keck Center  
for Comparative and Functional Genomics,  
1201 W. Gregory Drive, Urbana, IL 61801, USA

T. Trebitsh  
Department of Life Sciences,  
Ben-Gurion University of the Negev,  
Beer-Sheva 84105, Israel

Y. Tadmor · N. Katzir · V. Portnoy  
Agricultural Research Organization,  
P.O. Box 1021, Ramat Yishay 30095, Israel

S. King  
Department of Horticulture, Texas A&M University,  
College Station, TX 77845, USA

### Introduction

Watermelon [*Citrullus lanatus* (Thunb.) Matsum. & Nakai var. *lanatus*] accounts for 2% of the world area devoted to vegetable production (FAO 1995). It is grown in 44 states in the United States, with an annual production of 2.1 million tons and a farm value of \$340 million (National Watermelon Promotion Board; www.watermelon.org). In recent years, there has been an increase in consumer demand for seedless watermelons and production of this watermelon type has increased significantly. During 2003, over 60% of the watermelons produced in the United States were seedless (U.S. Department of Agriculture 2004).

Watermelon fruits are diverse in shape and size, in rind and flesh color, and in flesh texture, aroma, flavor, and nutrient composition. Like fruits of most plant species (Seymour et al. 1993), ripening watermelons undergo changes in pigment accumulation, flavor and aromatic volatiles, conversion of starch to sugars, and in increased susceptibility to post-harvest pathogens (Karakurt and Huber 2004). In this respect, there are differences between seeded and seedless watermelon fruits. For example, seeded watermelons may have shorter shelf life than seedless watermelons due to faster degradation of the tissue surrounding the seeds (Maynard 2004). Seedless watermelons may have different

sugar and aromatic compound contents than seeded watermelons of similar genetic background. Identifying the genes that control watermelon fruit quality and analyzing their differential expression in seeded versus seedless watermelon will be useful in enhancing fruit quality, nutritional values, and shelf life of seedless watermelons, to make them suitable to consumer needs.

The high-throughput sequencing of cDNA clones (libraries) has produced extensive genomic databases and large numbers of expressed sequence tags (ESTs) for various plant species (Richmond and Somerville 2000; Alba et al. 2004). Extensive EST analyses have been conducted for *Arabidopsis* (*Arabidopsis thaliana* L. Heynh.) (The Arabidopsis Genome Initiative 2000), rice (*Oryza sativa* L.) (Yu et al. 2002), and tomato (Budiman et al. 2000; Fei et al. 2004). Significant knowledge has accumulated in a number of plant species with respect to genes associated with cell wall metabolism, ethylene biosynthesis, and hormones affecting fruit setting, growth, and ripening (Giovannoni 2001). However, there is little information on genes controlling these processes in the watermelon fruit. Identifying, mapping, and characterizing these genes will be useful to research and breeding efforts in this crop.

In this study, we report the development of 832 “EST-unigenes” for watermelon fruit, and their classification based on their putative function in other plant species. In addition, we show that a large number of these “EST-unigenes” have no significant homology to any sequences reported so far in other plant species.

## Materials and methods

### Plant material

Watermelon fruits at early development stage (white flesh; 12 days post-pollination), ripening stage (light pink flesh; 24 days post-pollination), and ripe fruit (red flesh; 36 days post-pollination) from the heirloom cultivar “Illini Red” were used for RNA isolation. Leaf and stem tissue came from the terminal end of the U.S. Plant Introduction (PI) 525088 grown in the greenhouse under natural light, and 26 and 21°C day and night temperature, respectively. The plants producing watermelon fruits were grown in a field plot located at the South Central Agricultural Research Laboratory at Lane, OK. Upon collection, fruits were rinsed with sterile de-ionized water in the field, followed by flesh-tissue excision and processing as described later.

### RNA isolation

Prior to use in the field, all glassware and utensils were treated with RNaseZap (Ambion Inc., Austin, TX, USA) to neutralize any RNAase activity, and rinsed with RNAase-free water before each use. Utensils used in the field were washed with 95% ethanol and were sterilized in the autoclave prior to use.

Upon excision, fruit flesh was immediately chopped, placed in a sterile 50 ml conical polypropylene tube, and then frozen with liquid nitrogen. These frozen sample tubes were partially sealed, and then transported in a liquid nitrogen container to a freeze-dryer (LabConco Co., Kansas City, MO, USA). The frozen samples were freeze-dried under shelf refrigeration at  $-35^{\circ}\text{C}$  (as described by Callahan et al. 1989). Once lyophilized, the samples were treated with TRIzol reagent (Invitrogen Life Technologies Inc., Carlsbad, California, USA). RNA was extracted according to the manufacturer’s protocols (1 ml TRIzol per 100 mg lyophilized tissue) as described by Chomczynski and Sacchi (1987). Leaf and stem tissues were processed in a similar manner after collection from the greenhouse. RNA quality and quantity were determined using a spectrophotometer and denaturing agarose gel electrophoresis (Levi et al. 1992).

### cDNA synthesis, size selection, and cloning

Poly(A) + mRNA was isolated from total RNA using the Oligotex Direct mRNA kit (Qiagen, CA), and converted to double-stranded cDNA using the “Superscript Choice System kit” (Invitrogen, CA). First-strand cDNA synthesis was primed using a modified oligo(dT) primer (5'-AACTGGAAGAAT TCGCGGCCGCACGCA(T)<sub>18</sub>V-3'; V: A, G, or C) designed to anchor initiation at the 5'-end of the poly(A) + tail and enable directional cloning. cDNA sequences greater than 400 base pairs were selected by agarose gel electrophoresis. *EcoRI* adaptors (Invitrogen, CA) were ligated to the cDNAs, followed by digestion with *NotI*, and then directionally cloned into the *EcoRI* and *NotI* sites of the pBluescript II SK + vector (Stratagene). Cloned cDNAs were transformed into *E. coli* DH10B electrocompetent cells (Invitrogen, CA) and amplified as previously described by Soares and Bonaldo (1998).

### Normalization of the primary library

The primary cDNA library was normalized as previously described (Bonaldo et al. 1996). Essentially, a single-stranded “tracer” version of the library was created by digestion with Gene II and Exonuclease III enzymes (Invitrogen). Contaminating double-stranded DNA was removed by hydroxyapatite (HAP) chromatography. The purified single-stranded library was used as a template for PCR amplification through the T7 and T3 priming sites that flank the cloned cDNA inserts. The purified PCR products, representing the entire cDNA population cloned, were used as a “driver” for subtractive hybridization. Essentially, 0.5  $\mu\text{g}$  of PCR-amplified cDNA inserts were denatured and mixed with 50 ng of purified single-stranded tracer-DNA, as well as 10  $\mu\text{g}$  each of 5' and 3' blocking oligonucleotides. The resulting solution (50% formamide, 0.12 M NaCl, 1% SDS) at a final volume of 20  $\mu\text{l}$  was overlaid with mineral oil and subtractive hybridization was carried out for 44 h at 30°C. Non-hybridized, single-stranded tracer-DNAs were

separated from hybridized DNA duplexes by HAP column chromatography. These purified, non-hybridized ssDNAs were rendered partially double-stranded by M13 reverse primer extension (only a small part of the second strand is synthesized by primer extension to improve transformation efficiency) and were electroporated into *E. coli* DH10B cells to generate the normalized library.

#### Subtraction of the normalized library

cDNAs generated from leaf tissue were subtracted to enrich the normalized library for genes differentially expressed in the fruit tissue. For this, a primary library was created from leaf mRNA as described earlier. Subtraction was performed essentially as described earlier, except that the driver consisted of PCR products from the primary leaf library. Essentially, 2.5 µg of PCR-amplified cDNA driver was combined with 75 ng of single-stranded tracer-DNA from the normalized library and 40 µg each of 5' and 3' blocking oligonucleotides. These were hybridized in a solution consisting of 50% formamide, 0.12 M NaCl, and 1% SDS at a final volume of 20 µl for 88 h at 30°C.

#### Sequencing of ESTs

Individual transformed bacteria colonies were robotically picked from agar plates and racked as LB media + 10% glycerol stocks (All clones are grown in LB media + 10% glycerol stocks for long-term storage at -80°C) in 384-well plates. After overnight growth, glycerol stocks were inoculated into LB medium amended with 100 µg/ml of carbenicillin in 96-well, deep-culture plates and grown for 16 h. Plasmid DNA was purified with Qiagen 8000 and 9600 BioRobots (Qiagen, CA) and associated chemistries. Sequencing of the 5'-ends was performed using standard T7 primer and ABI BigDye terminator chemistry on ABI 3700 and 3730xl capillary systems (Applied Biosystems, CA). All 384- and 96-well format plates were labeled with a barcode and a laboratory information management system (HTLims) was used to track the sample flow.

#### Sequencing and data analysis

Sequencing of random EST clones from normalized and subtracted libraries from various tissues of watermelon resulted in 1046 clean sequences. A sequence is considered clean when a minimum of 200 nucleotides remains after trimming vector and low-quality sequences. The average read length of clean sequences was 555 nucleotides with a minimum quality score of 20. The redundancy in the library was 20%. The final "clean" sequences were clustered and assembled using Paracel Transcript Assembler (PTA). Contaminant sequences like *E. coli*, mitochondrial, chloroplast, cloning vector, and RNA were filtered during the cleanup stage. Repeat sequences were masked and annotated. EST sequences were then clustered based on local

similarity scores of pairwise comparison using 88% similarity over 100 nt. Clusters containing only one sequence were grouped as singlets. The EST clusters were assembled into contigs (contiguous sequence) by multiple-sequence alignment, which generates a consensus sequence for each cluster with criteria of 95% identity over 30 nt overlap. Multiple contigs may be generated per cluster, since EST clusters may not share enough similarity over their entire length to be assembled as single contig. Multiple contigs may also be generated when ESTs in a cluster represent an alternative splice form of the gene. The ESTs remaining in a cluster after the formation of contigs, are designated as cluster singlets. The set of non-redundant sequences for the library includes the contigs, cluster-singlets, and singlets and was designated as "EST-unigenes".

These sequences were used to query the GenBank database for homologs using the Basic Local Alignment Search Tool (BLAST) (Altschul et al. 1990), using *E*-value of 0.01, to ensure a high level of confidence that each sequence represents a non-redundant gene transcript.

---

## Results and discussion

A large number of the ESTs identified in this study are homologous to genes previously reported to be important in fruit growth and ripening in other plant species. The 1046 random cDNA clones sequenced produced 832 "EST-unigenes". Of these 832 "EST-unigenes", 747 were single ESTs (singletons; non-assembled sequencing reads), and 85 were contigs generated by computer-based assembly of sequence fragments from several clones (contigs). Of the 832 "EST-unigenes", 578 have significant homology to amino acid sequences from the GenBank non-redundant protein database. A large number of these homologous sequences have previously been ascribed to Arabidopsis proteins, and were successfully annotated using gene ontology (GO) analysis. The length of the ESTs ranges from 338 to 699 bases, whereas contigs range from 555 to 2823 bases. The individual sequences of 1046 ESTs have been submitted to NCBI (Accession numbers: DV736965—DV738010, to be released on May 1, 2006).

A functional class was assigned to each "EST-unigene" based on the degree of similarity (*E*-value) to the closest counterpart sequence found in other plant species (Table 1). Of the 578 "EST-unigenes" that had significant homology to the nucleotide database (nt), 168 are homologous to genes with unknown function, while 410 are homologous to genes with known function (Table 1). These 410 "EST-unigenes", based on GO annotation, were assigned to one of the following functional classes: (1) primary metabolism (74 "EST-unigenes"), (2) amino acid synthesis and processing (57 "EST-unigenes"), (3) membrane and transport (66 "EST-unigenes"), (3) cell division, cell wall and metabolism, cytoskeleton, and cellular organization (41 "EST-unigenes"), (4) DNA/RNA transcription and gene expression (63 "EST-unigenes"), (5) cellular communication/signal transduction (70 "EST-unigenes"), (6) defense- and stress-related proteins (31 "EST-unigenes"),

**Table 1** ESTs with significant homology to genes with known function in other plant species

EST-unigene	Similar gene in other species	E-value	Plant species
Primary metabolism			
AL_01.86.C1.Contig82	2-Oxoglutarate dehydrogenase	e-140	<i>O. sativa</i>
AL01005A1C06.f1	3-Hydroxyisobutyryl-coenzyme A hydrolase protein	1.00E-85	<i>A. thaliana</i>
AL01005A1C06.f1	3-Hydroxyisobutyryl-coenzyme A hydrolase protein	1.00E-85	<i>A. thaliana</i>
AL01006A1F12.f1	Acyl CoA:diacylglycerol acyltransferase	6.00E-56	<i>Glycine max</i>
AL01005A1D09.f1	Acyl-ACP thioesterase	3.00E-23	<i>Garcinia mangostana</i>
AL_01.62.C1.Contig59	Acyl-CoA-binding protein	1.00E-30	<i>O. sativa</i>
AL01006B2A07.f1	Adenosine monophosphate binding protein	1.00E-57	<i>A. thaliana</i>
AL010002000D04	ADP-glucose pyrophosphorylase large subunit 1	3.00E-23	<i>C. lanatus</i>
AL_01.10.C1.Contig10	AIM1 protein	8.00E-93	<i>A. thaliana</i>
AL01006B1B11.f1	alpha-Hydroxynitrile lyase	2.00E-30	<i>A. thaliana</i>
AL01004X1F01.f1	AP-1 Golgi-related complex component	3.00E-67	<i>Camptotheca acuminata</i>
AL01005A1G02.f1	Chloroplast ribosome recycling factor protein	2.00E-24	<i>Spinacia oleracea</i>
AL01005B2B06.f1	Chloroplast small heat shock protein	3.00E-43	<i>Capsicum frutescens</i>
AL010001000F09	Copine-related	1.00E-43	<i>A. thaliana</i>
AL01004X1B04.f1	Cytochrome b5	8.00E-14	<i>Olea europaea</i>
AL010002000E09	Cytochrome b5	5.00E-16	<i>A. thaliana</i>
AL010002000B04	Cytochrome b5 reductase isoform II	3.00E-51	<i>Zea mays</i>
AL01006A1A09.f1	Cytochrome b561	1.00E-22	<i>C. lanatus</i>
AL01005B1D11.f1	Cytochrome c biogenesis protein-like	6.00E-58	<i>O. sativa</i>
AL01006B2C01.f1	Cytochrome c oxidase subunit 6b	1.00E-38	<i>A. thaliana</i>
AL01005A2D02.f1	Cytochrome c reductase subunit	4.00E-07	<i>Solanum tuberosum</i>
AL01004X1H09.f1	Cytochrome P450 85 (Dwarf protein)	5.00E-28	<i>Lycopersicon esculentum</i>
AL01005B1E08.f1	Cytochrome P450 protein	7.00E-27	<i>A. thaliana</i>
AL_01.73.C1.Contig70	Cytochrome P450-like protein	6.00E-77	<i>A. thaliana</i>
AL010001000G11	Dihydrolipoamide succinyltransferase	4.00E-35	<i>A. thaliana</i>
AL010001000G11	Dihydrolipoamide succinyltransferase	4.00E-35	<i>A. thaliana</i>
AL01005B1C08.f1	Dihydropyrimidinase	2.00E-12	<i>A. thaliana</i>
AL01005B1B12.f1	Disulfide bond formation protein	3.00E-37	<i>A. thaliana</i>
AL01005B2F05.f1	Enolase protein	5.00E-20	<i>A. thaliana</i>
AL010002000C03	ent-Kaurenoic acid oxidase	3.00E-70	<i>Pisum sativum</i>
AL010001000C11	Ferredoxin III, chloroplast precursor	4.00E-40	<i>Z. mays</i>
AL_01.41.C1.Contig39	Formate dehydrogenase, mitochondrial precursor	3.00E-76	<i>Hordeum vulgare</i>
AL01005B2F12.f1	Glutamine synthetase	3.00E-67	<i>Vitis vinifera</i>
AL01006A1D05.f1	Glycinamide ribonucleotide transformylase	6.00E-67	<i>Vigna unguiculata</i>
AL01005A1F08.f1	Glycine decarboxylase complex H-protein	6.00E-50	<i>A. thaliana</i>
AL01005A1H05.f1	GTP-binding protein	2.00E-61	<i>A. thaliana</i>
AL01006B1A05.f1	His-Asp phosphotransfer protein	1.00E-27	<i>P. sativum</i>
AL01004X1G06.f1	Hydrolase family protein	3.00E-31	<i>A. thaliana</i>
AL010001000C01	Isopentenyl-diphosphate delta-isomerase	3.00E-85	<i>C. acuminata</i>
AL_01.61.C1.Contig58	Lipase class 3 family protein	7.00E-44	<i>A. thaliana</i>
AL01006A2G06.f1	Malate dehydrogenase	e-103	<i>V. vinifera</i>
AL010001000H04	Malate dehydrogenase, glyoxysomal precursor	1.00E-72	<i>V. vinifera</i>
AL01004X1D04.f1	NADH-ubiquinone oxidoreductase-related	1.00E-39	<i>A. thaliana</i>
AL01006B1E07.f1	NADPH:quinone oxidoreductase	2.00E-23	<i>A. thaliana</i>
AL010001000D04	Nodule-enhanced malate dehydrogenase	3.00E-22	<i>P. sativum</i>
AL01006B2E04.f1	Peroxidase ATP2a	2.00E-19	<i>A. thaliana</i>
AL01005B2F03.f1	Phosphoenolpyruvate carboxylase	5.00E-18	<i>Cucumis sativus</i>
AL01005A2E12.f1	Phosphoglycerate kinase	4.00E-14	<i>O. sativa</i>
AL01006B2G02.f1	Phospholipase	8.00E-43	<i>C. lanatus</i>
AL01005B2E02.f1	Phospholipase D	2.00E-26	<i>Ricinus communis</i>
AL01006B2G09.f1	Phospholipase D delta isoform	1.00E-37	<i>Gossypium hirsutum</i>
AL01005A2D05.f1	Phospholipase D zeta1	2.00E-32	<i>A. thaliana</i>
AL01005A2D08.f1	Phospholipid cytidylyltransferase	4.00E-44	<i>A. thaliana</i>
AL01006A1A03.f1	Phosphoribosyl pyrophosphate synthase isozyme 3	3.00E-64	<i>S. oleracea</i>



**Table 1** Continued

EST-unigene	Similar gene in other species	E-value	Plant species
AL010001000H12	Phosphoribosylanthranilate transferase-like protein	2.00E-62	<i>A. thaliana</i>
AL01005A2D06.f1	Cytosolic factor	1.00E-85	<i>A. thaliana</i>
AL01004X1B09.f1	NADH-ubiquinone oxidoreductase subunit	1.00E-64	<i>A. thaliana</i>
AL_01.9.C1.Contig9	Pyruvate decarboxylase	e-132	<i>S. tuberosum</i>
AL01006A1G09.f1	Pyruvate dehydrogenase E1 alpha subunit	e-136	<i>Citrus x paradisi</i>
AL01005B2H10.f1	Ribosomal protein L11	9.00E-72	<i>A. thaliana</i>
AL01003X1F01.f1	Serine racemase,	8.00E-13	<i>A. thaliana</i>
AL_01.63.C1.Contig60	Specific tissue protein 2	7.00E-25	<i>Cicer arietinum</i>
AL_01.57.C1.Contig54	Specific tissue protein 2	2.00E-09	<i>C. arietinum</i>
AL01005B2B03.f1	Stem secoisolariciresinol dehydrogenase	3.00E-50	<i>Forsythia x intermedia</i>
AL_01.68.C1.Contig65	Steroid 5alpha-reductase-like protein	4.00E-85	<i>A. thaliana</i>
AL01004X1B11.f1	Succinate dehydrogenase iron-protein subunit-like	5.00E-66	<i>A. thaliana</i>
AL01005B1H04.f1	Sucrose-phosphatase	2.00E-85	<i>L. esculentum</i>
AL01003X1H11.f1	Thioredoxin	9.00E-36	<i>A. thaliana</i>
AL010001000D05	Thioredoxin h	2.00E-38	<i>Cucurbita maxima</i>
AL01006A1H11.f1	Ubiquinol-cytochrome C reductase complex	3.00E-28	<i>A. thaliana</i>
AL01005A2H02.f1	UDP-glucose 6-dehydrogenase	e-126	<i>G. max</i>
AL01006A2H06.f1	UDP-glucose:sterol glucosyltransferase	7.00E-62	<i>A. thaliana</i>
AL01006A2B05.f1	Uricase subunit	2.00E-50	<i>A. thaliana</i>
AL01005B1D12.f1	Xanthine dehydrogenase-like protein	3.00E-57	<i>A. thaliana</i>
Amino acid synthesis and processing			
AL01006B1E06.f1	20S proteasome alpha 6 subunit	9.00E-14	<i>Nicotiana benthamiana</i>
AL010001000F11	40S ribosomal protein S17	2.00E-58	<i>A. thaliana</i>
AL010002000F11	40S ribosomal protein S2	2.00E-49	<i>Picea abies</i>
AL_01.35.C1.Contig34	40S ribosomal protein S20-like protein	2.00E-55	<i>A. thaliana</i>
AL01006B2D08.f1	40S ribosomal protein S21	2.00E-35	<i>A. thaliana</i>
AL010002000F10	40S ribosomal protein S7 homolog	5.00E-75	<i>S. tuberosum</i>
AL010002000C04	40S ribosomal S4 protein	5.00E-28	<i>G. max</i>
AL01006A2D10.f1	60S ribosomal protein L24	8.00E-59	<i>C. arietinum</i>
AL01006A2D03.f1	60S ribosomal protein L34	3.00E-47	<i>Solanum demissum</i>
AL01004X1H05.f1	60S ribosomal protein L34	9.00E-38	<i>A. thaliana</i>
AL01004X1A03.f1	60S ribosomal protein L37a	1.00E-46	<i>G. hirsutum</i>
AL01005B1F01.f1	60S ribosomal protein L39	1.00E-23	<i>O. sativa</i>
AL01004X1F08.f1	Alanine:glyoxylate aminotransferase 2 homolog	3.00E-27	<i>A. thaliana</i>
AL01004X1C11.f1	Amino acid transport protein AAP2	8.00E-09	<i>R. communis</i>
AL01003X1E05.f1	Aminopeptidase C	8.00E-73	<i>O. sativa</i>
AL01004X1G02.f1	Aspartic proteinase	8.00E-20	<i>Theobroma cacao</i>
AL01003X1B01.f1	Chaperonin CPN10	1.00E-37	<i>A. thaliana</i>
AL010002000H04	Cyclophilin	7.00E-58	<i>R. communis</i>
AL01005B2E08.f1	Cysteine protease 1	2.00E-65	<i>Plantago major</i>
AL01006B2H02.f1	Cysteine proteinase	1.00E-46	<i>Petunia x hybrida</i>
AL010001000D09	Dormancy-associated protein	5.00E-35	<i>A. thaliana</i>
AL010002000H02	Elongation factor 1 alpha	2.00E-83	<i>Saccharum</i>
AL010002000B08	Elongation factor 1-alpha	2.00E-42	<i>Manihot esculenta</i>
AL010001000B09	HECT ubiquitin-protein ligase 3	3.00E-88	<i>A. thaliana</i>
AL01005B1F08.f1	Low molecular weight heat-shock protein	1.00E-25	<i>Corylus avellana</i>
AL01005A2G12.f1	OTU-like cysteine protease family protein	2.00E-49	<i>A. thaliana</i>
AL_01.49.C1.Contig46	Papain-like cysteine peptidase XBCP3	2.00E-26	<i>A. thaliana</i>
AL010001000E12	Profilin	5.00E-13	<i>C. lanatus</i>
AL01004X1B05.f1	Prolylcarboxypeptidase-like protein	1.00E-21	<i>A. thaliana</i>
AL01005A2F07.f1	Prolylcarboxypeptidase-like protein	2.00E-75	<i>A. thaliana</i>
AL_01.11.C1.Contig11	Proteinase inhibitor II	2.00E-09	<i>A. thaliana</i>
AL_01.7.C1.Contig7	Proteinase inhibitor II	9.00E-10	<i>G. max</i>
AL01006A2D11.f1	Pyrrrolidone carboxyl peptidase-like protein	1.00E-12	<i>O. sativa</i>

**Table 1** Continued

EST-unigene	Similar gene in other species	E-value	Plant species
AL01006A2D11.f1	Pyrrolidone carboxyl peptidase-like protein	1.00E-12	<i>O. sativa</i>
AL010001000B04	Ribosomal protein L12 family protein	5.00E-04	<i>A. thaliana</i>
AL01003X1C09.f1	Ribosomal protein L17	6.00E-75	<i>Castanea sativa</i>
AL010002000H03	Ribosomal protein L28-like	3.00E-56	<i>O. sativa</i>
AL01004X1E08.f1	Ribosomal protein L28-like	1.00E-57	<i>O. sativa</i>
AL01006A1C12.f1	Ribosomal protein L29	3.00E-27	<i>Panax ginseng</i>
AL010002000D10	Ribosomal protein L30	7.00E-56	<i>Lupinus luteus</i>
AL010002000B07	Ribosomal protein L31	9.00E-45	<i>A. thaliana</i>
AL010001000H08	Ribosomal protein S26	4.00E-48	<i>A. thaliana</i>
AL_01.75.C1.Contig72	Ribosomal protein S29	3.00E-29	<i>O. sativa</i>
AL_01.51.C1.Contig48	Ribosomal protein S29	3.00E-29	<i>O. sativa</i>
AL010002000E11	S-adenosylmethionine decarboxylase	2.00E-66	<i>x Citrofortunella mitis</i>
AL01005A1H10.f1	Subtilisin	5.00E-06	<i>Narcissus pseudonarcissus</i>
AL01005B2E05.f1	Subtilisin-like protease	1.00E-24	<i>G. max</i>
AL01006A1F06.f1	Subtilisin-like protease C1	2.00E-45	<i>G. max</i>
AL_01.74.C1.Contig71	Subtilisin-type protease precursor	e-105	<i>G. max</i>
AL_01.13.C1.Contig13	tRNA pseudouridine synthase family protein	6.00E-37	<i>A. thaliana</i>
AL01006B2A04.f1	Ubiquitin carboxyl-terminal hydrolase	3.00E-16	<i>A. thaliana</i>
AL01005A1A05.f1	Ubiquitin carboxyl-terminal hydrolase family	2.00E-08	<i>A. thaliana</i>
AL01006B1F08.f1	Ubiquitin carboxyl-terminal hydrolase family protein	4.00E-32	<i>A. thaliana</i>
AL01004X1G01.f1	Ubiquitin ligase SINAT5	6.00E-61	<i>O. sativa</i>
AL01006A1G08.f1	Ubiquitin (UBA)/TS-N domain-containing protein	4.00E-12	<i>A. thaliana</i>
AL010002000D06	Ubiquitin-conjugating enzyme 8	2.00E-82	<i>Capsicum annuum</i>
AL01005A1B08.f1	X-Pro dipeptidase-like protein	8.00E-93	<i>A. thaliana</i>
<b>Membrane and transport</b>			
AL01005B2A10.f1	26S proteasome non-ATPase regulatory subunit 3	2.00E-42	<i>Nicotiana tabacum</i>
AL010001000B01	26S proteasome particle non-ATPase subunit	8.00E-90	<i>O. sativa</i>
AL010002000B06	26S proteasome particle triple-A ATPase subunit	1.00E-36	<i>O. sativa</i>
AL01006A1F03.f1	ABC transporter	3.00E-55	<i>A. thaliana</i>
AL_01.12.C1.Contig12	ABC transporter protein	e-142	<i>A. thaliana</i>
AL_01.83.C1.Contig79	Annexin	3.00E-17	<i>A. thaliana</i>
AL01006A1H07.f1	arp3	5.00E-35	<i>O. sativa</i>
AL01005A2H10.f1	Autophagocytosis protein-like	5.00E-78	<i>A. thaliana</i>
AL_01.23.C1.Contig23	B12D protein	2.00E-35	<i>Ipomoea batatas</i>
AL01005B2A08.f1	Ca <sup>2+</sup> /H <sup>+</sup> antiporter [ <i>Zea mays</i> ]	1.00E-67	<i>Z. mays</i>
AL010002000B09	Cation transport protein	2.00E-27	<i>A. thaliana</i>
AL01004X1A07.f1	CD2-binding protein-related	1.00E-61	<i>A. thaliana</i>
AL010002000E06	CEF/transport protein	1.00E-13	<i>A. thaliana</i>
AL01006A1G06.f1	Citrate synthase	1.00E-17	<i>N. tabacum</i>
AL01005A1C07.f1	CMP-sialic acid transporter	2.00E-74	<i>A. thaliana</i>
AL01006B1D09.f1	Coated vesicle membrane protein	1.00E-89	<i>O. sativa</i>
AL01005B2B02.f1	Coatomer alpha subunit-like protein	3.00E-55	<i>Lotus corniculatus</i>
AL01006B1B09.f1	Complex I intermediate-associated protein	2.00E-55	<i>A. thaliana</i>
AL01003X1C05.f1	Copine-related protein	7.00E-17	<i>A. thaliana</i>
AL01006A2A03.f1	Cyclic nucleotide and calmodulin-regulated ion channel	2.00E-68	<i>A. thaliana</i>
AL01005B1G06.f1	Endomembrane protein	8.00E-13	<i>A. thaliana</i>
AL01005B2C07.f1	ent-Kaurene synthase	3.00E-89	<i>C. sativus</i>
AL010001000A12	Epsilon subunit of mitochondrial F1-ATPase	1.00E-30	<i>A. thaliana</i>
AL_01.50.C1.Contig47	Globulin-like protein	3.00E-27	<i>A. thaliana</i>
AL01005B1E10.f1	GONST3 Golgi Nucleotide sugar transporter	2.00E-39	<i>A. thaliana</i>
AL01006B1F06.f1	H <sup>+</sup> -transporting two-sector ATPase	2.00E-43	<i>I. batatas</i>
AL01006A2D04.f1	Low temperature and salt responsive protein	2.00E-21	<i>A. thaliana</i>
AL01006A1B03.f1	Indole-3-glycerol phosphate synthase	1.00E-22	<i>A. thaliana</i>

**Table 1** Continued

EST-unigene	Similar gene in other species	E-value	Plant species
AL_01.55.C1.Contig52	Integral membrane family protein	2.00E-28	<i>A. thaliana</i>
AL01006A1B12.f1	Integral membrane protein	2.00E-13	<i>A. thaliana</i>
AL01006B2B11.f1	Iron transporter-related	3.00E-29	<i>A. thaliana</i>
AL01006A2D06.f1	Lipid-transfer protein 3 precursor (LTP 3)	3.00E-29	<i>Prunus dulcis</i>
AL01004X1H06.f1	Magnesium transporter CorA-like family protein	2.00E-18	<i>A. thaliana</i>
AL01005A1D04.f1	Membrane protein	3.00E-58	<i>O. sativa</i>
AL010001000D07	Mitochondrial precursor	2.00E-36	<i>A. thaliana</i>
AL01006A1C01.f1	MRP-like ABC transporter	2.00E-47	<i>O. sativa</i>
AL01006A2C09.f1	Nonspecific lipid transfer; auxin induced gene	5.00E-12	<i>Zinnia elegans</i>
AL01006B1D05.f1	Oligopeptide transporter-like protein	6.00E-68	<i>A. thaliana</i>
AL_01.56.C1.Contig53	Oxysterol binding protein	2.00E-84	<i>O. sativa</i>
AL01003X1C12.f1	Permease	6.00E-73	<i>G. hirsutum</i>
AL01006B1G05.f1	Peroxin-3 family protein [ <i>Arabidopsis thaliana</i> ]	7.00E-19	<i>A. thaliana</i>
AL01003X1C03.f1	Peroxisomal targeting signal type 2 receptor	7.00E-96	<i>A. thaliana</i>
AL01005B2C12.f1	Plasma membrane H <sup>+</sup> -ATPase	7.00E-10	<i>Daucus carota</i>
AL01006B2F09.f1	Plastidic ATP/ADP transporter	8.00E-93	<i>Citrus hybrid cultivar</i>
AL_01.22.C1.Contig22	PM28B protein	8.00E-55	<i>S. oleracea</i>
AL010001000A07	PQ-loop repeat/transmembrane family protein	1.00E-14	<i>A. thaliana</i>
AL01003X1H10.f1	PQ-loop repeat/transmembrane family protein	3.00E-19	<i>A. thaliana</i>
AL01006A2G05.f1	Presenilin-like protein	1.00E-44	<i>A. thaliana</i>
AL01004X1H03.f1	RING finger protein	7.00E-30	<i>A. thaliana</i>
AL01003X1A08.f1	Stomatin-like protein	6.00E-13	<i>A. thaliana</i>
AL01006A1B09.f1	Sugar transporter	2.00E-37	<i>A. thaliana</i>
AL01005B2A04.f1	Sugar transporter-like protein	3.00E-26	<i>A. thaliana</i>
AL01006A1C11.f1	Sulphate transporter	9.00E-51	<i>Brassica napus</i>
AL01005B1D10.f1	Sulphate transporter	1.00E-13	<i>B. napus</i>
AL010001000G01	Transmembrane domains	3.00E-40	<i>A. thaliana</i>
AL01006A2F12.f1	Transmembrane MLO family protein	e-127	<i>A. thaliana</i>
AL01006A2B04.f1	Transmembrane protein	2.00E-36	<i>O. sativa</i>
AL01006A1H12.f1	Transport protein	1.00E-25	<i>A. thaliana</i>
AL01003X1B03.f1	Transporter-related	2.00E-65	<i>A. thaliana</i>
AL010002000F03	Type IIB calcium ATPase	2.00E-30	<i>Medicago truncatula</i>
AL01006B2B08.f1	Ubiquitin-transferase	1.00E-57	<i>A. thaliana</i>
AL01006B1F10.f1	Uncoupling protein	2.00E-59	<i>Dracunculus vulgaris</i>
AL01006B2A05.f1	Vacuolar H <sup>+</sup> -ATPase C subunit	1.00E-58	<i>P. major</i>
AL01005B2H04.f1	Vacuolar sorting/VPS9 domain-containing protein	2.00E-10	<i>A. thaliana</i>
AL010002000D09	Vacuolar V-H + ATPase subunit E	1.00E-73	<i>Citrus limon</i>
AL01006A2D01.f1	Vesicle transport v-SNARE	4.00E-86	<i>A. thaliana</i>
<b>Cell division</b>			
AL01006B1D02.f1	Allergen-like protein BRSn20	3.00E-21	<i>Sambucus nigra</i>
AL01005B2E10.f1	Anaphase promoting complex subunit 11	3.00E-44	<i>A. thaliana</i>
AL_01.1.C1.Contig2	ARD-like protein	5.00E-92	<i>Brassica juncea</i>
AL01005B1G12.f1	Cell cycle switch protein CCS52a	1.00E-85	<i>L. luteus</i>
AL01006B1B10.f1	Cullin	2.00E-28	<i>A. thaliana</i>
AL01006B1B06.f1	Cyclin-dependent kinase CDC2C	2.00E-27	<i>O. sativa</i>
AL01006B2D10.f1	Fasciclin-arabinogalactan precursor (AtAGP8)	5.00E-90	<i>A. thaliana</i>
AL010001000E02	F-box family protein	4.00E-35	<i>A. thaliana</i>
AL01005B2D08.f1	F-box family protein	4.00E-12	<i>A. thaliana</i>
AL01006B1D04.f1	Glycosyltransferase family protein	5.00E-29	<i>A. thaliana</i>
AL01006B1C07.f1	Kinase family protein	3.00E-66	<i>A. thaliana</i>
AL_01.53.C1.Contig50	N-acetylglucosaminyltransferase	6.00E-81	<i>A. thaliana</i>
AL_01.70.C1.Contig67	NAM (no apical meristem)-like protein	1.00E-12	<i>A. thaliana</i>
AL01005A1A01.f1	O-linked N-acetyl glucosamine transferase	4.00E-42	<i>A. thaliana</i>
AL01004X1A02.f1	Phagocytosis and cell motility protein ELMO1-related	5.00E-62	<i>A. thaliana</i>
AL010001000G09	Plastid division protein ftsZ	e-103	<i>Gentiana lutea</i>

**Table 1** Continued

EST-unigene	Similar gene in other species	E-value	Plant species
AL01005B1F07.f1	Skp1	2.00E-61	<i>Medicago sativa</i>
Cytoskeleton			
AL01005A2B04.f1	Alpha 1-tubulin	1.00E-18	<i>Z. mays</i>
AL01005A1A07.f1	Glycine-rich protein	7.00E-09	<i>A. thaliana</i>
AL010002000E05	Mo25 family protein	1.00E-30	<i>A. thaliana</i>
AL01005A1D06.f1	Myosin	2.00E-41	<i>A. thaliana</i>
AL.01.77.C1.Contig74	Tubulin folding cofactor E/Pfifferling (PFI)	2.00E-36	<i>A. thaliana</i>
Cell wall and metabolism			
AL01006B1D10.f1	Expansin-like protein	8.00E-32	<i>Quercus robur</i>
AL01004X1D12.f1	Fiber protein Fb19	1.00E-50	<i>Gossypium barbadense</i>
AL01005A1F03.f1	Glycosyl hydrolase family protein	1.00E-28	<i>A. thaliana</i>
AL01006A2E04.f1	Hydroxyproline-rich glycoprotein family protein	2.00E-23	<i>A. thaliana</i>
AL01006B1E09.f1	Hydroxyproline-rich glycoprotein family protein	7.00E-15	<i>A. thaliana</i>
AL01006A1H06.f1	LIM domain protein PLIM1	1.00E-58	<i>N. tabacum</i>
AL01006A1E07.f1	Microtubule associated protein	6.00E-22	<i>C. arietinum</i>
AL.01.15.C1.Contig15	Microtubule associated protein	7.00E-59	<i>C. arietinum</i>
AL01004X1B06.f1	Microtubule-associated protein	2.00E-48	<i>A. thaliana</i>
AL.01.38.C1.Contig37	Neurofilament protein-related	7.00E-12	<i>A. thaliana</i>
AL01005A2C07.f1	Nitrate-induced NOI protein	1.00E-22	<i>A. thaliana</i>
AL.01.34.C1.Contig33	Omega-6 fatty acid desaturase	9.00E-54	<i>Cucurbita pepo</i>
AL01005B2A05.f1	pEARLI 1/extensin-like protein	1.00E-23	<i>A. thaliana</i>
AL.01.84.C1.Contig80	Pectin methylesterase	3.00E-94	<i>N. tabacum</i>
AL.01.58.C1.Contig55	Phloem filament protein PP1	2.00E-21	<i>C. maxima</i>
AL01005A2E10.f1	Ripening-related protein [Vitis vinifera]	1.00E-45	<i>V. vinifera</i>
AL01005B1D06.f1	Tetratricopeptide repeat (TPR)-containing protein	9.00E-33	<i>A. thaliana</i>
AL01006B2D06.f1	Translocon Tic40-like/hydroxyproline-rich glycoprotein	3.00E-20	<i>A. thaliana</i>
AL.01.48.C1.Contig45	Caffeoyl-CoA O-methyltransferase	1.00E-69	<i>Populus kitakamiensis</i>
DNA, RNA related and gene expression			
AL01006A1E03.f1	AP2/EREBP transcription factor	2.00E-33	<i>A. thaliana</i>
AL01004X1G07.f1	Apoptosis-related/DNA-binding protein	9.00E-35	<i>A. thaliana</i>
AL01006A1E06.f1	bZIP family transcription factor	1.00E-63	<i>A. thaliana</i>
AL01005B2C10.f1	bZIP transcription factor	4.00E-20	<i>N. tabacum</i>
AL01005A2H06.f1	bZIP transcription factor	3.00E-19	<i>Phaseolus vulgaris</i>
AL01006A2E06.f1	CAAT box binding factor/ transcription factor Hap2a	5.00E-08	<i>A. thaliana</i>
AL01005A1E10.f1	CCAAT-box binding factor HAP5 homolog	4.00E-34	<i>D. carota</i>
AL01005B2E09.f1	CCCH-type zinc finger protein	6.00E-12	<i>A. thaliana</i>
AL01006B1D03.f1	Contains similarity to human PCF11p homolog	4.00E-43	<i>A. thaliana</i>
AL01006B1G09.f1	DNA-binding protein	2.00E-14	<i>A. thaliana</i>
AL01006A2B10.f1	DNA-binding protein; bZIP type	1.00E-10	<i>Petroselinum crispum</i>
AL01003X1E06.f1	Dof zinc finger protein	2.00E-38	<i>N. tabacum</i>
AL01006A2F05.f1	EIN3-like protein	6.00E-58	<i>Cucumis melo</i>
AL01006B2H09.f1	ERF-like protein	2.00E-65	<i>C. melo</i>
AL01006B1A09.f1	Exonuclease family protein	2.00E-43	<i>A. thaliana</i>
AL01005B2D03.f1	HB2 homeodomain protein	7.00E-68	<i>Populus tremula</i>
AL01005B2F02.f1	Homeobox HAT22/leucine zipper protein	5.00E-23	<i>A. thaliana</i>
AL01006B1A11.f1	Homeobox-leucine zipper transcription factor protein	7.00E-27	<i>A. thaliana</i>
AL01006B1C11.f1	Human RAN binding protein 16-like	1.00E-61	<i>A. thaliana</i>
AL01005B1D07.f1	Iron-deficiency protein Ids3	2.00E-20	<i>O. sativa</i>
AL01005B1G11.f1	KOW domain-containing transcription factor	5.00E-83	<i>A. thaliana</i>
AL01005B1C10.f1	Leucine-rich repeat family protein	7.00E-67	<i>A. thaliana</i>
AL01005A2B07.f1	mads-box transcription factor	3.00E-14	<i>Momordica charantia</i>
AL.01.71.C1.Contig68	Mei2-like protein	1.00E-06	<i>H. vulgare</i>
AL010002000H07	Minichromosome maintenance (MCM) protein	6.00E-31	<i>A. thaliana</i>
AL01004X1A12.f1	Mitotic control/ribonuclease II family protein	3.00E-49	<i>A. thaliana</i>



**Table 1** Continued

EST-unigene	Similar gene in other species	E-value	Plant species
AL01006B1A07.f1	myb-related protein	6.00E-22	<i>A. thaliana</i>
AL01004X1E11.f1	nam-like protein 14	1.00E-37	<i>Petunia x hybrida</i>
AL01006A1E08.f1	Nuclear movement family protein	1.00E-38	<i>A. thaliana</i>
AL01006B1F09.f1	orf265a	4.00E-19	<i>Beta vulgaris</i>
AL01004X1D10.f1	Phage-type RNA polymerase	3.00E-10	<i>N. tabacum</i>
AL01005A1D03.f1	pRGR1	3.00E-18	<i>O. sativa</i>
AL01005B1H11.f1	Probable CCCH-type zinc finger protein	8.00E-19	<i>A. thaliana</i>
AL01006A2E10.f1	Probable CCR4-associated factor	1.00E-57	<i>A. thaliana</i>
AL01003X1D01.f1	Proline-rich protein	2.00E-12	<i>G. max</i>
AL01006A1G05.f1	Retroelement	4.00E-11	<i>O. sativa</i>
AL01005A1G09.f1	Retrotransposon/reverse transcriptase protein	5.00E-39	<i>A. thaliana</i>
AL010002000C11	Ribonucleoprotein 29 kD	2.00E-68	<i>Nicotiana sylvestris</i>
AL01005A1A03.f1	RNA helicase	1.00E-48	<i>A. thaliana</i>
AL01006A1A06.f1	RNA recognition motif (RRM)-containing protein	3.00E-19	<i>A. thaliana</i>
AL01005B1C01.f1	RNA-binding protein	3.00E-41	<i>A. thaliana</i>
AL_01.29.C1.Contig28	Sd-RNase	2.00E-23	<i>Prunus salicina</i>
AL01006A2B01.f1	Similar to RIKEN cDNA 2010107D16 gene	e-100	<i>Mus musculus</i>
AL01005B1C09.f1	Small nuclear ribonucleoprotein	4.00E-23	<i>A. thaliana</i>
AL010002000G12	snRNP core Sm protein Sm-X5-like protein	4.00E-35	<i>A. thaliana</i>
AL01006B2B04.f1	Synovial sarcoma, X breakpoint 2 interacting protein	6.00E-60	<i>O. sativa</i>
AL010001000A04	THUMP domain-containing protein	2.00E-71	<i>A. thaliana</i>
AL010001000C06	tonneau 2	9.00E-85	<i>O. sativa</i>
AL010002000H11	Transcriptional activation factor	2.00E-41	<i>A. thaliana</i>
AL010002000G06	Tubby protein	6.00E-62	<i>O. sativa</i>
AL01005A2E09.f1	Uclease	e-100	<i>N. tabacum</i>
AL01006A2A05.f1	Regulator of chromosome condensation (RCC1)	3.00E-51	<i>A. thaliana</i>
AL01003X1D05.f1	WRKY transcription factor	2.00E-23	<i>G. hirsutum</i>
AL01005A2B12.f1	XAP-5 protein	9.00E-86	<i>A. thaliana</i>
AL01004X1B07.f1	Zinc finger (C2H2-type) family protein	3.00E-21	<i>A. thaliana</i>
AL01006B1C08.f1	Zinc finger (C3HC4-type RING finger) family protein	2.00E-37	<i>A. thaliana</i>
AL01003X1G11.f1	Zinc finger (GATA type) protein ZIM (ZIM)	6.00E-42	<i>A. thaliana</i>
AL01005A2H08.f1	Zinc finger homeobox family protein	2.00E-07	<i>A. thaliana</i>
AL01005B1E06.f1	Zinc finger protein	1.00E-20	<i>A. thaliana</i>
AL01005B2B12.f1	Zinc finger protein	1.00E-75	<i>A. thaliana</i>
AL010002000B03	Zinc finger protein	5.00E-20	<i>S. demissum</i>
AL01005B2C03.f1	Zinc finger protein	5.00E-14	<i>A. thaliana</i>
AL01006A2F08.f1	Zinc finger protein-related	3.00E-19	<i>A. thaliana</i>
Signal transduction			
AL010001000C07	Abscisic acid induced protein	8.00E-40	<i>A. thaliana</i>
AL_01.37.C1.Contig36	Abscisic acid response protein	1.00E-12	<i>C. melo</i>
AL01003X1E04.f1	Ankyrin protein kinase	2.00E-11	<i>A. thaliana</i>
AL01006B2D01.f1	ATMPK9	2.00E-08	<i>A. thaliana</i>
AL01006A2C04.f1	AtRab GTP-binding protein	3.00E-80	<i>A. thaliana</i>
AL_01.65.C1.Contig62	Auxin efflux carrier family protein	3.00E-80	<i>A. thaliana</i>
AL01005A2D01.f1	Avr9/Cf-9 rapidly elicited protein	6.00E-52	<i>N. tabacum</i>
AL_01.59.C1.Contig56	Avr9/Cf-9 rapidly elicited protein 146	6.00E-14	<i>N. tabacum</i>
AL01003X1F07.f1	Brassinosteroid LRR receptor kinase protein	2.00E-22	<i>O. sativa</i>
AL_01.30.C1.Contig29	BURP domain containing protein	3.00E-54	<i>A. thaliana</i>
AL01005A2E04.f1	C3HC4 zinc finger protein	1.00E-15	<i>A. thaliana</i>
AL_01.26.C1.Contig26	Calcineurin B-like protein 1	4.00E-47	Imported
AL01005A1D12.f1	Calcineurin-like phosphoesterase family protein	2.00E-06	<i>A. thaliana</i>
AL01005B2G02.f1	Calcineurin-like phosphoesterase family protein	6.00E-61	<i>A. thaliana</i>
AL01006B2F04.f1	Calcium dependent protein kinase	4.00E-24	<i>A. thaliana</i>
AL01005A1C09.f1	Calcium-binding EF-hand family protein	2.00E-31	<i>O. sativa</i>
AL010002000D03	Calmodulin	7.00E-25	<i>D. carota</i>

**Table 1** Continued

EST-unigene	Similar gene in other species	E-value	Plant species
AL01005A1H04.f1	Calmodulin-binding protein	7.00E-47	<i>A. thaliana</i>
AL01006A1F11.f1	Calmodulin-binding protein	1.00E-75	<i>A. thaliana</i>
AL01006B2D03.f1	Caltractin (Centrin)	8.00E-25	<i>A. thaliana</i>
AL01003X1A01.f1	Caltractin / centrin	4.00E-72	<i>A. thaliana</i>
AL010001000G04	CDC2	1.00E-95	<i>P. tremula</i>
AL010001000B12	cis-Zeatin O-glucosyltransferase	8.00E-09	<i>Sorghum bicolor</i>
AL.01.17.C1.Contig17	Copper chaperone	2.00E-25	<i>Populus alba</i>
AL01005B1A12.f1	Deoxyuridine triphosphatase	1.00E-41	<i>O. sativa</i>
AL01006A1H04.f1	Elicitor-inducible protein EIG-J7	5.00E-56	<i>C. annuum</i>
AL01006A2H11.f1	Ethylene receptor; Cm-ETR1	2.00E-95	<i>C. melo</i>
AL.01.20.C1.Contig20	Eukaryotic translation initiation factor 5A isoform II	8.00E-87	<i>Hevea brasiliensis</i>
AL01006B1A12.f1	F-box / SKP1 interacting protein	3.00E-66	<i>A. thaliana</i>
AL01003X1B07.f1	FCA protein	1.00E-15	<i>Triticum aestivum</i>
AL.01.25.C1.Contig25	Flowering locus T like protein	2.00E-15	<i>Populus nigra</i>
AL01005B1H05.f1	GRAM domain-/ABA-responsive protein	1.00E-47	<i>A. thaliana</i>
AL01003X1C10.f1	GSK-3-like protein MsK4	5.00E-41	<i>M. sativa</i>
AL010001000A01	GTPase activating-like protein	5.00E-28	<i>A. thaliana</i>
AL01005B2E11.f1	GTP-binding protein	2.00E-53	<i>A. thaliana</i>
AL01003X1G02.f1	GTP-binding protein	1.00E-52	<i>P. sativum</i>
AL01004X1E03.f1	Histidine acid phosphatase family protein	1.00E-45	<i>A. thaliana</i>
AL010002000F08	Immunophilin	9.00E-53	<i>Vicia faba</i>
AL01004X1F07.f1	Kinase family protein	1.00E-22	<i>A. thaliana</i>
AL01006B1E12.f1	Leucine-rich receptor-like protein kinase	3.00E-59	<i>A. thaliana</i>
AL01005B1A10.f1	Leucine-rich repeat receptor-like kinase	1.00E-34	<i>A. thaliana</i>
AL010001000D11	Microtubule-associated protein	6.00E-41	<i>A. thaliana</i>
AL01005B1E01.f1	Mitogen-activated protein kinase	6.00E-20	<i>M. sativa</i>
AL01005A2E08.f1	Mitogen-activated protein kinase MAPK	1.00E-39	<i>Prunus armeniaca</i>
AL01006B1H11.f1	PERK1-like protein kinase	1.00E-82	<i>N. tabacum</i>
AL01005A1E12.f1	Phosphatidylinositol 4-kinase PI4K	1.00E-50	<i>A. thaliana</i>
AL01006A2D02.f1	Phototropic-responsive NPH3 family protein	2.00E-37	<i>A. thaliana</i>
AL01004X1H04.f1	Phytoalexin-deficient 4 protein (PAD4)	2.00E-43	<i>A. thaliana</i>
AL.01.46.C1.Contig44	Pi starvation-induced protein	3.00E-34	<i>C. arietinum</i>
AL01006A2C08.f1	Protein kinase	7.00E-83	<i>C. sativus</i>
AL01005A1C11.f1	Protein kinase-like protein	e-105	<i>A. thaliana</i>
AL01005A2A06.f1	Protein kinases	7.00E-56	<i>A. thaliana</i>
AL010002000C07	Ras-related GTP-binding protein	2.00E-64	<i>C. sativus</i>
AL01005A2C01.f1	Receptor protein kinase	e-103	<i>A. thaliana</i>
AL.01.39.C1.Contig38	Receptor-like protein kinase	e-128	<i>A. thaliana</i>
AL01006B1E11.f1	S3 self-incompatibility locus-linked pollen protein	5.00E-12	<i>Petunia integrifolia</i>
AL01006B2E03.f1	Serine threonine kinase	8.00E-38	<i>Sandersonia aurantiaca</i>
AL01005B1F10.f1	Serine/threonine protein kinase	3.00E-52	<i>O. sativa</i>
AL01005A1D08.f1	Serine/threonine protein phosphatase	5.00E-11	<i>A. thaliana</i>
AL01005A1C03.f1	Similar to serine/threonine kinases	3.00E-31	<i>A. thaliana</i>
AL01005B1H04.f1	Sucrose-phosphatase	2.00E-85	<i>L. esculentum</i>
AL01005B1A05.f1	Tousled-like protein kinase	2.00E-27	<i>Nicotiana glutinosa</i>
AL01005A1H06.f1	Transducin family protein/WD-40 repeat protein	2.00E-34	<i>A. thaliana</i>
AL010001000C12	Translation factor	1.00E-54	<i>Pinus pinaster</i>
AL.01.5.C1.Contig6	Translationally controlled tumor protein	2.00E-71	<i>C. melo</i>
AL01005B2G01.f1	VP1/ABI3 family regulatory protein	7.00E-10	<i>O. sativa</i>
AL01005A2A11.f1	WD-40 repeat/transducin family protein	1.00E-63	<i>A. thaliana</i>
AL01006B1H12.f1	Enzyme-forming ethylene (ACC oxidase)	3.00E-55	<i>C. melo</i>
AL01005A1B06.f1	Ethylene insensitive (EIN3/EIL)-transcription regulator	4.00E-29	<i>Fagus sylvatica</i>
AL01003X1D12.f1	Diacylglycerol kinase	1.00E-23	<i>O. sativa</i>

**Table 1** Continued

EST-unigene	Similar gene in other species	E-value	Plant species
<b>Defense- and stress-related proteins</b>			
AL_01.8.C1.Contig8	Al-induced protein	e-108	<i>G. hirsutum</i>
AL_01.44.C1.Contig42	Dessication-related protein	e-109	<i>A. thaliana</i>
AL01005B1A06.f1	DnaJ protein	3.00E-16	<i>Malus x domestica</i>
AL01005B1D09.f1	DnaJ protein	3.00E-25	<i>Malus x domestica</i>
AL_01.0.C1.Contig1	Early nodulin	3.00E-27	<i>G. max</i>
AL010002000F09	EXS/ERD1/XPR1/SYG1 family protein	2.00E-22	<i>A. thaliana</i>
AL01005B2G11.f1	GEKO1	1.00E-34	<i>O. sativa</i>
AL01003X1A07.f1	Geraniol-responsible factor 15	3.00E-09	<i>Matricaria chamomilla</i>
AL01006A1D11.f1	Harpin-induced family protein	2.00E-51	<i>A. thaliana</i>
AL01006B1D07.f1	Integral membrane family protein	5.00E-50	<i>A. thaliana</i>
AL_01.21.C1.Contig21	Jasmonic acid 2	4.00E-34	<i>L. esculentum</i>
AL_01.82.C1.Contig78	Jasmonic acid regulatory protein	7.00E-14	<i>A. thaliana</i>
AL_01.69.C1.Contig66	Nodulin-like protein	7.00E-42	<i>O. sativa</i>
AL01006B1E05.f1	NOI protein	5.00E-17	<i>O. sativa</i>
AL01005B1D04.f1	Pathogenesis related homeodomain protein	2.00E-43	<i>A. thaliana</i>
AL_01.33.C1.Contig32	Pathogenesis-related protein	2.00E-55	<i>C. sativus</i>
AL01005A1G08.f1	Pentatricopeptide (PPR) repeat-containing protein	9.00E-26	<i>A. thaliana</i>
AL010002000A08	Plasma intrinsic protein	9.00E-79	<i>Juglans regia</i>
AL_01.4.C1.Contig5	PVR3-like protein	1.00E-12	<i>Ananas comosus</i>
AL010001000F05	Seed maturation protein PM39	2.00E-15	<i>G. max</i>
AL01005A1F10.f1	Silverleaf whitefly-induced protein	2.00E-41	<i>C. pepo</i>
AL010002000A03	snakin-1	3.00E-17	<i>S. tuberosum</i>
AL010002000B01	Superoxide dismutase	5.00E-17	<i>F. sylvatica</i>
AL_01.14.C1.Contig14	Type-2 metallothionein	5.00E-26	<i>C. lanatus</i>
AL01006A1C09.f1	Universal stress/early nodulin family protein	4.00E-58	<i>A. thaliana</i>
AL01006A2D05.f1	Universal stress/early nodulin family protein	3.00E-23	<i>A. thaliana</i>
AL01006B2E09.f1	Wound-induced protein-like	7.00E-08	<i>V. vinifera</i>
AL010002000C10	Wound-inducible basic protein	8.00E-13	<i>P. vulgaris</i>
AL010001000E10	Wound-inducible carboxypeptidase	5.00E-81	<i>L. esculentum</i>
AL010001000C04	Chain X, Ascobate Peroxidase	2.00E-22	<i>G. max</i>
AL01006B1B03.f1	WWE domain-containing protein / ceo protein	1.00E-15	<i>A. thaliana</i>
<b>Secondary metabolism</b>			
AL010002000H08	Glutathione S-transferase	7.00E-96	<i>C. maxima</i>
AL01004X1C05.f1	Glutathione S-transferase	2.00E-60	<i>G. max</i>
AL01005B2G03.f1	Glutathione S-transferase	1.00E-18	<i>C. maxima</i>
AL01005A2B03.f1	Glutathione S-transferase	3.00E-13	<i>P. sativum</i>
AL01006B1H02.f1	Phytoene synthase, chloroplast precursor (MEL5)	4.00E-61	<i>C. melo</i>
AL01005A2F12.f1	Phytoene desaturase	7.00E-58	<i>M. charantia</i>
AL01004X1H07.f1	relA/spoT homologous protein RSH2	8.00E-18	<i>O. sativa</i>
AL01006A1G03.f1	beta-Amilase	5.00E-18	<i>C. arietinum</i>

(7) secondary metabolism (8 “EST-unigenes”). These “EST-unigenes” may take part in fruit development (involving rapid cell division and differentiation, as well as rapid nutrient and carbohydrate translocation, synthesis, and accumulation), and fruit ripening (cell wall softening, and break down of carbohydrates and storage proteins) (Table 1; Fig. 1) (Giovannoni 2001).

A large number of the watermelon fruit “EST-unigenes” are associated with basic cell function (Table 1, Fig. 1) including respiration, photosynthesis, electron transfer (cytochrome, mitochondrial, and chloroplast proteins), or carbohydrate synthesis (as shown for ADP-glucose pyrophos-

phorylase in watermelon fruit; In-Jung et al. 1998). Others are associated with amino acid and protein synthesis and trafficking (ribosomal proteins, hydrolase, phospholipase, malate dehydrogenase, and ubiquitin proteins; Table 1). During fruit development, a considerable amount of energy is invested in the chemical reactions leading to synthesis of amino acids, and synthesis of functional and storage proteins. Later, during fruit ripening, the transient storage proteins serve as reservoirs for the amino acids used for the synthesis of “ripening-associated” proteins (Peumans et al. 2002). A few of the watermelon fruit “EST-unigenes” are homologous to proteases or peptidases (Table 1) that might

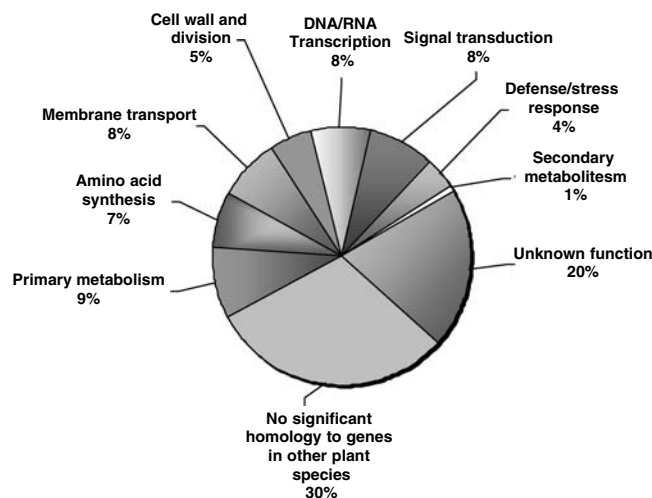
be associated with the release of amino acids from these storage proteins.

The developing fruit is a nutrient sink, which derives nutrients reallocated from other parts of the plant in support of its continued development. Indeed, a large number of the “EST-unigenes” are homologous to genes involved in membrane transport and cytosolic trafficking (Table 1). Upon reaching full size, the fruit enters the ripening phase through production of internal ethylene, followed by softening of cell walls, production of secondary compounds, as well as changes in sugar content, flavor, and aroma. This process involves a sequence of events that lead to the activation of transcription factors and signal transduction proteins associated with ripening (Giovannoni 2001). The ethylene biosynthesis genes (including the S-adenosylmethionine decarboxylase and ACC oxidase) and ethylene signal transduction genes (the ethylene receptor “Cm-ETR1”, or the transcription regulator “EIN3/EIL”; Table 1) take part in the ripening processes (Arif et al. 1994; Giles et al. 2001; Giovannoni 2001; Naoki et al. 2003). The abscisic acid- and auxin-induced proteins, DNA- and RNA-binding proteins, and a variety of protein kinases (Table 1) are also taking part in the signal transduction processes leading to fruit ripening.

Fruit softening is a result of enzymatic activity that impairs cell wall properties and dissolves chemical bonds between cell walls, leading to their separation (Rose and Bennett 1999). A few of the “EST-unigenes” are cell wall enzymes (the pectin-modifying protein expansin, the cell wall P8 protein, and the fasciclin-like arabinogalactan protein 8 precursor; AtAGP8; Table 1) required for cell surface adhesion and expansion (Brummell and Harpster 2001; Shi et al. 2003; Trainotti et al. 2003).

Transcription factors were also identified in this study (Table 1). These include the basic leucine-zipper transcription factor (bZIP family), a regulatory element for an abscission-specific cellulose enzyme involved in cell wall softening (Tucker et al. 2002), the CCAAT-binding transcription factor (CBF-B/NF-YA) protein, which binds to the CCAAT-box motif present in certain promoters of genes expressed in vegetative and reproductive plant tissues (Guerineau et al. 2003), and a CCCH-type zinc finger motif that plays a key regulatory role in flower and fruit development (Li et al. 2001).

The ability of a fruit to resist pathogen attack or environmental stress decreases with tissue softening (Brummell and Harpster 2001). Thus, cell defense and stress response genes might be programmed to be expressed in the fruit tissue to slow pathogen invasion and fruit tissue decline prior to full seed development and maturation. Among these, are the DNAJ heat shock chaperone proteins (Table 1), which protect the intracellular milieu proteins from irreversible aggregation during cellular stress. Other important stress and defense proteins include the exonuclease protein required for post-transcriptional silencing in *Arabidopsis* (Glazov et al. 2003), the hemoglobin (HB2) produced in plants growing under stress conditions (as shown for rice; Lira-Ruan et al. 2001), the zinc finger proteins that enhance disease resistance and drought tolerance in plants (Kim



**Fig. 1** Distribution of watermelon flesh ESTs according to their function

et al. 2004) and in the *A. thaliana* fruit (Balasubramanian and Schneitz 2002) (Table 1). “EST-unigenes” with a significant homology to ethylene-responsive element binding factor (ERF) proteins also exist in the watermelon fruit (Table 1). The ERF proteins are transcription factors linked to defense and stress response in plants (Oñate-Sánchez and Singh 2002).

Secondary compounds are produced mainly during fruit ripening. Carotenoids (including lycopene) are the main secondary compounds produced in watermelon fruit (Rodriguez-Amaya 1999). A group of “EST-unigenes” homologous to glutathione S-transferase (GST) genes were identified and classified in the category of secondary compounds (Table 1). Glutathione S-transferase proteins may perform a variety of functions in the binding of flavonoids and the deposition of these compounds in the vacuole (Board et al. 2000). GST proteins also take part in the detoxification of reactive electrophilic compounds by catalysing their conjugation to the tripeptide (gamma-glutamyl-cysteinyl-glycine) glutathione (Armstrong 1997). Recent studies implicated GSTs as signaling compounds leading to apoptosis (Dixon et al. 2002). The glutathione S-transferase domain is also found in elongation factors 1-gamma and the HSP26 family of stress-related proteins, which include auxin-regulated proteins in plants.

Of the 832 watermelon “EST-unigenes” analyzed, 254 (~30%) had no detectable homologs ( $E \geq 0.1$ ) to any other plant genomes or protein sequences reported so far in GenBank (Fig. 1). Some of these “EST-unigenes” may represent untranslated (UTR) 3' regions (Mignone et al. 2005). However, further studies are needed to determine if they are typical to watermelon and other cucurbit species. The majority of the watermelon fruit “EST-unigenes” reported in this study could be grouped into abundantly expressed gene families. However, a considerable number of the “EST-unigenes” could not be classified (Table 1; Fig. 1). Extensive genome sequencing is still needed for cucurbit species to identify the genes that may be distinct



to this family. Future work will include sequencing and microarray analysis of ESTs representing different fruit tissues and developmental stages.

## References

- Alba R, Fei Z, Payton P, Liu Y, Moore SL, Debbie P, Cohn J, D'Ascenzo M, Gordon JS, Rose JK, Martin G, Tanksley SD, Bouzayen M, Jahn MM, Giovannoni J (2004) ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development. *Plant J* 39:697–714
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815
- Arif SAM, Taylor MA, George LA, Butler AR, Burch LR, Davies HV, Stark MJR, Kumar A (1994) Characterization of the S-adenosylmethionine decarboxylase (SAMDC) gene of potato. *Plant Mol Biol* 26:327–338
- Armstrong RN (1997) Structure, catalytic mechanism, and evolution of the glutathione transferases. *Chem Res Toxicol* 10:2–18
- Balasubramanian S, Schneitz K (2002) NOZZLE links proximal–distal and adaxial–abaxial pattern formation during ovule development in *Arabidopsis thaliana*. *Development* 129:4291–4300
- Board PG, Coggan M, Chelvanayagam G, Easteal S, Jermini LS, Schulte GK, Danley DE, Hoth LR, Griffior MC, Kamath AV, Rosner MH, Chrnyk BA, Perregaux DE, Gabel CA, Geoghegan KF, Pandit J (2000) Identification, characterization, and crystal structure of the Omega class glutathione transferases. *J Biol Chem* 275:24798–24806
- Bonaldo MF, Lennon G, Soares MB (1996) Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res* 6:791–806
- Brummell DA, Harpster MH (2001) Cell wall metabolism in fruit softening and quality and its manipulation in transgenic plants. *Plant Mol Biol* 47:311–340
- Budiman MA, Mao L, Wood TC, Wing RA (2000) A deep coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing. *Genome Res* 10:129–136
- Callahan A, Morgens P, Walton E (1989) Isolation and in vitro translation of RNAs from developing peach fruit. *HortScience* 24:356–358
- Chomczynski P, Sacchi N (1987) Signal-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162:156–159
- Dixon DP, Laphorn A, Edwards R (2002) Plant glutathione transferases. *Genome Biol* 3:1–7
- FAO (1995) Production year book for 1994. No. 48. Food and Agricultural Organization of the United Nations, Rome
- Fei Z, Tang X, Alba RM, White JA, Ronning CM, Martin GB, Tanksley SD, Giovannoni JJ (2004) Comprehensive EST analysis of tomato and comparative genomics of fruit ripening. *Plant J* 40:47–59
- Giles EDO, Eric ME, Sharon RL (2001) Ethylene inhibits the nod factor signal transduction pathway of *Medicago truncatula*. *Plant Cell* 13:1835–1849
- Giovannoni J (2001) Molecular regulation of fruit ripening. *Annu Rev Plant Physiol Plant Mol Biol* 52:725–749
- Glazov E, Phillips K, Budziszewski GJ, Schob H, Meins F Jr, Levin JZ (2003) A gene encoding an RNase D exonuclease-like protein is required for post-transcriptional silencing in *Arabidopsis*. *Plant J* 35:342–349
- Guerineau F, Benjdia M, Zhou DX (2003) A jasmonate-responsive element within the *A. thaliana* vsp1 promoter. *J Exp Bot* 54:1153–1162
- Karakurt Y, Huber DJ (2004) Ethylene-induced gene expression, enzyme activities, and water soaking in immature and ripe watermelon (*Citrullus lanatus*) fruit. *J Plant Physiol* 161:381–388
- Kim IJ, Kahng HY, Chung WI (1998) Characterization of cDNAs encoding small and large subunits of ADP-glucose pyrophosphorylases from watermelon (*Citrullus vulgaris* S.). *Biosci Biotechnol Biochem* 62:550–555
- Kim SH, Hong JK, Lee SC, Sohn KH, Jung HW, Hwang BK (2004) CAZFP1, Cys2/His2-type zinc-finger transcription factor gene functions as a pathogen-induced early-defense gene in *Capsicum annuum*. *Plant Mol Biol* 55:883–904
- Levi A, Galau GA, Wetzstein HY (1992) A rapid procedure for the isolation of RNA from high-phenolic-containing tissues of pecan. *HortScience* 27:1316–1318
- Li J, Jia D, Chen X (2001) HUA1, a regulator of stamen and carpel identities in *Arabidopsis*, codes for a nuclear RNA binding protein. *Plant Cell* 13:2269–2281
- Lira-Ruan V, Sarath G, Klucas RV, Arredondo-Peter R (2001) Synthesis of hemoglobins in rice (*Oryza sativa* var. Jackson) plants growing in normal and stress conditions. *Plant Sci* 161:279–287
- Maynard DN (2004) University of Florida, Gulf Coast Research and Education Center—Watermelons Website. <http://watermelons.ifas.ufl.edu>
- Mignone F, Grillo G, Licciulli F, Iacono M, Liuni S, Kersey PJ, Duarte J, Saccone C, Pesole G (2005) UTRdb and UTRsite: a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res* 33:141–146
- Naoki Y, Sumiko T, Ryohei N, Akitsugu I, Yasutaka K (2003) Characterization of a novel tomato EIN3-like gene (LeEIL4). *J Exp Bot* 54:2775–2776
- Oñate-Sánchez L, Singh KB (2002) Identification of arabidopsis ethylene-responsive element binding factors with distinct induction kinetics after pathogen infection. *Plant Physiol* 128:1313–1322
- Peumans WJ, Proost P, Swennen RL, Van Damme EJM (2002) The Abundant Class III Chitinase Homolog in Young Developing Banana Fruits Behaves as a Transient Vegetative Storage Protein and Most Probably Serves as an Important Supply of Amino Acids for the Synthesis of Ripening-Associated Proteins. *Plant Physiol* 130:1063–1072
- Richmond T, Somerville S (2000) Chasing the dream: plant EST microarrays. *Curr Opin Plant Biol* 3:108–116
- Rose JK, Bennett AB (1999) Cooperative disassembly of the cellulose-xyloglucan network of plant cell walls: parallels between cell expansion and fruit ripening. *Trends Plant Sci* 4:176–183
- Rodriguez-Amaya DB (1999) Latin American food sources of carotenoids. *Arch Latinoam Nutr* 49:74–84
- Shi H, Kim YS, Guo Y, Stevenson B, Zhu JK (2003) The arabidopsis SOS5 locus encodes a putative cell surface adhesion protein and is required for normal cell expansion. *Plant Cell* 15:19–32
- Seymour GB, Taylor JE, Tucker GA (1993) Biochemistry of fruit ripening. Chapman & Hall, London, 442 pp
- Soares MB,onaldo MF (1998) Constructing and screening normalized cDNA libraries. In: Detecting genes, vol 2. Cold Spring Harbor Press, Cold Spring Harbor, pp 49–157
- Trainotti L, Zanin D, Casadoro G (2003) A cell wall-oriented genomic approach reveals a new and unexpected complexity of the softening in peaches. *J Exp Bot* 54:1821–1832
- Tucker ML, Whitelaw CA, Lyssenko NN, Nath P (2002) Functional analysis of regulatory elements in the GenePromoter for an abscission-specific cellulase from bean and isolation, expression, and binding affinity of three TGA-type basic leucine zipper transcription factors. *Plant Phys* 130:1487–1496
- United States Department of Agriculture (2004) Agricultural statistics. U.S. Department of Agriculture, National Agricultural Statistics Service, Washington, D.C.
- Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 296:79–92