

Expressed Sequence Tags of Chinese Cabbage Flower Bud cDNA¹

Chae Oh Lim, Ho Yeon Kim, Min Gab Kim, Soo In Lee, Woo Sik Chung, Sung Han Park, Inhwan Hwang, and Moo Je Cho*

Department of Biochemistry (C.O.L., M.G.K., S.I.L., W.S.C., S.H.P., M.J.C.), and Plant Molecular Biology and Biotechnology Research Center (C.O.L., H.Y.K., I.H., M.J.C.), Gyeongsang National University, Chinju, 660–701, Korea

We randomly selected and partially sequenced cDNA clones from a library of Chinese cabbage (*Brassica campestris* L. ssp. *pekinensis*) flower bud cDNAs. Out of 1216 expressed sequence tags (ESTs), 904 cDNA clones were unique or nonredundant. Five hundred eighty-eight clones (48.4%) had sequence homology to functionally defined genes at the peptide level. Only 5 clones encoded known flower-specific proteins. Among the cDNAs with no similarity to known protein sequences (628), 184 clones had significant similarity to nucleotide sequences registered in the databases. Among these 184 clones, 142 exhibited similarities at the nucleotide level only with plant ESTs. Also, sequence similarities were evident between these 142 ESTs and their matching ESTs when compared using the deduced amino acid sequences. Therefore, it is possible that the anonymous ESTs encode plant-specific ubiquitous proteins. Our extensive EST analysis of genes expressed in floral organs not only contributes to the understanding of the dynamics of genome expression patterns in floral organs but also adds data to the repertoire of all genomic genes.

Single-run partial sequencing of randomly selected cDNA clones is now a widely used tool in genome research (Adams et al., 1991; Boguski et al., 1993; Sasaki et al., 1994). ESTs help to quickly identify functions of expressed genes and to understand the complexity of gene expression. ESTs have also served as molecular genetic markers in genomic mapping (Kurata et al., 1994; Shen et al., 1994). Since the number of ESTs from various species has increased rapidly, it is now possible to compare a large number of genes and the proteins they encode between animals and plants. Genes expressed in different tissues within an organism have also been randomly sequenced (Höfte et al., 1993). Comparison of ESTs between different tissues yields information on the dynamics of genomic expression patterns. The first random sequencing of cDNA clones was performed utilizing a human brain library (Adams et al., 1991), and almost 3400 cDNA clones have been reported from human brain (Adams et al., 1991, 1992, 1993). Various other organisms, such as nematode (McCombie et al., 1992;

Waterston et al., 1992), mouse (Höög, 1991), and several plants (Uchimiya et al., 1992; Höfte et al., 1993; Keith et al., 1993; Park et al., 1993; Newman et al., 1994; Sasaki et al., 1994) have also been examined by extensive sequencing of randomly selected cDNA clones. The enormous accumulation of ESTs has thus 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).

In the area of plant science, two major cDNA sequencing projects have been conducted in *Arabidopsis* and in rice (Uchimiya et al., 1992; Höfte et al., 1993; Newman et al., 1994; Sasaki et al., 1994). As of September 14, 1995, 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% of rice have sequence similarity to known proteins from either microbes, plants, or animals. Functions of “unmatched” ESTs still await elucidation through genetic and biochemical studies. For example, generation of mutants or characterization of proteins encoded by such unmatched ESTs could provide a means to pinpoint functions of those or other genes. Another way 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. In plants, major efforts to generate ESTs have been restricted until now, mainly with regard to *Arabidopsis*, rice, and maize. Once expressed genes are sequenced from many different plant species, however, defining highly conserved domains within homologous ESTs will be possible for plant-specific genes.

In this report single-run partial sequencing of randomly selected cDNA clones from Chinese cabbage (*Brassica campestris* L. ssp. *pekinensis*) was performed as a part of the *Brassica* genome project in Korea. Chinese cabbage belongs to the genus *Brassica*, which comprises many economically important vegetable plants, especially in Korea, China, and Japan. In addition, *Brassica* has served as a favorite model system for various biological processes in plants (Park et

¹ This work was supported by a grant to the Plant Molecular Biology and Biotechnology Research Center from the Korea Science and Engineering Foundation.

* Corresponding author; e-mail mjcho@nongae.gsnu.ac.kr; fax 82–591–759–9363.

Abbreviations: dbEST, EST nucleotide database; EST, expressed sequence tag; PIR, Protein Identification Resource.

al., 1993). The relatively small genome size of Chinese cabbage (approximately 7.7×10^8 bp per haploid genome), only a few times larger than that of Arabidopsis, greatly simplifies both genetic and molecular analyses of genes (Croy et al., 1993).

A majority of the ESTs reported in Arabidopsis and rice were derived from cDNAs from a mixture of different tissues (Höfte et al., 1993; Newman et al., 1994) or from cultured cells (Uchimiya et al., 1992; Sasaki et al., 1994). Use of a whole plant body or suspension cells in generating ESTs is an efficient means to obtain a representative EST population from a given plant species. In our study, though, we chose flower buds of *Brassica*. A flower bud is one of the most complicated organs in plants. Many morphological and biochemical processes are unique to this young reproductive organ. This large-scale EST project was conducted to provide better understanding of the dynamics of genomic expression patterns of floral organs. In addition, the tissue-specific EST information supplies supplementary data to the repertoire of all expressed genomic genes, because cDNAs from whole plant bodies are much less likely to contain rare, tissue-specific, expressed genes. In this paper we report partial sequencing of 1216 randomly selected cDNA clones from Chinese cabbage flower buds and classification of these clones based on the biological functions of the encoded proteins.

MATERIALS AND METHODS

Plant Materials and cDNA Library

Flower buds of approximately 5 mm in length were harvested from *Brassica campestris* L. ssp. *pekinensis* grown in a greenhouse at Seoul Seed Co. (Seoul, Korea). Total RNA was isolated from flower buds as described previously (Ausubel et al., 1992). Poly(A)⁺ RNA was selected using a commercially available poly(A)⁺ RNA purification kit (Pharmacia). cDNA was synthesized using a λZapII cDNA synthesis kit (Stratagene) and was cloned into pBluescript II KS(+) (Stratagene) using unphosphorylated adaptors following, with slight modification, a published method (Stanley et al., 1988). The plasmid library was plated on 15-cm Luria Bertani agar plates with ampicillin. Individual colonies were propagated and saved at -80°C until further use.

Nucleotide Sequencing

The template DNAs for the sequencing reaction were prepared, with minor modifications, by an alkaline lysis method (Sambrook et al., 1989). Cells of 2-mL overnight cultures were collected and resuspended in 200 μL of a lysozyme buffer containing 2 mg/mL lysozyme and 2 mg/mL RNase A. The cells were subsequently lysed with 0.2 N NaOH/1.0% SDS and neutralized with 5 M KOAc. Plasmid DNA was precipitated with an equal volume of isopropanol, and the pellet was washed with 80% cold ethanol. The amount of isolated DNA template was estimated on a 1.0% agarose gel by comparison to serial dilutions of pBluescript II KS(+). Insert sizes were estimated by

agarose gel electrophoresis after restriction enzyme digestion with *Bam*HI and *Xho*I. Using a Perkin-Elmer 9600 thermal cycler and an ABI 373A sequencer (Applied Biosystems), the 5' ends of the cDNA clones were sequenced according to the thermal cycling protocol with a Taq Dye Primer Cycle Sequencing Kit (Applied Biosystems).

Sequence Analysis

The partial sequences were translated into three reading frames and then compared with sequences in the PIR (release 40.0, 1994) or SwissProt (release 28.0, 1994) databases using the FASTA algorithm (Pearson and Lipman, 1988). A match was declared when the optimized similarity score was greater than 120 and the sequence identity was greater than 35% when compared to the quarry sequence and a known sequence. From the proteins that ranked higher than 120 score and 35% identity, the sequence with the highest optimized similarity score sequence was chosen. Sequences that did not match with sequences in the protein databases were further analyzed by searching for homology in GenBank (release 82.0, 1994) and EMBL (release 38.0, 1994) at the nucleotide level using the FASTA algorithm (Pearson and Lipman, 1988). A match was declared when the score was higher than 120 (optimized similarity score) and 65% (sequence identity). The remaining unidentified sequences were compared to each other, and unique sequences were submitted to the Genome Sequence Database (Los Alamos, NM) and the dbEST.

RESULTS

cDNA Library and Nucleotide Sequencing

cDNAs from poly(A)⁺ RNA derived from flower buds of Chinese cabbage were directionally cloned into a plasmid vector. This cDNA library was the source of the ESTs in this study. The overall insert sizes of the total 1216 EST sequences ranged from 0.5 to 4.0 kb, with the majority (75%) falling between 0.6 and 1.0 kb. We sequenced the 5' ends of the inserts. After deletion of vector sequences and ambiguous bases, an average length of 320 bp was used in the database searches. To evaluate the quality of the library, both ends of 15 clones of *rbcS* (small subunit of Rubisco complex) were sequenced. They all had poly(A) tails, and translation initiation codons (ATG) were present in 10 clones. For a gene encoding histone H4, we found 4 clones that had poly(A) tails and translation initiation codons. In this library a high percentage of cDNAs of less than 1.5 kb had full-length coding regions.

Characterization of ESTs

We partially sequenced 1216 individual cDNA clones in a single run. The deduced amino acid sequences were compared with protein sequences in several databases, although we searched primarily in PIR. SwissProt was used also when no matching sequences were found in PIR. We found 588 ESTs (48.4%) that had significant amino acid sequence similarities to sequences registered in both protein databases, and 393 of those ESTs that could be func-

tionally identified are listed in Table I. When more than 1 EST showed homology to a gene registered in the databases, only one EST was included in Table I, even if the ESTs were not from the same gene. We observed that 269 ESTs encoded proteins previously identified in other plant species, and only 20 ESTs matched registered genes from the *Brassica* species (Table I). Five known flower-specific genes were identified: the anther-specific protein (Shen and Hsu, 1992), the bp4C protein (Albani et al., 1990), the C98 protein (Roberts et al., 1991), the fil 1 protein (Nacken et al., 1991), and the microspore-specific protein I3 (Roberts et al., 1991) (see Table I).

We also classified in Figure 1 the 588 protein sequences that have homologies to sequences in the databases according to putative functions. Genes involved in metabolic pathways (e.g. glycolysis or photosynthesis) produced the most abundant transcripts in the flower buds. Transcripts for the translational apparatus (especially ribosomal proteins) ranked next in abundance. One hundred twenty-four clones shared sequence homologies with nonplant sequences. Some of them included the FK506/rapamycin-binding protein, 26K antigen, spermatid-specific protein, placental protein 15, and valosin-containing protein. It was not possible to assign probable functions to these proteins in plants. The remaining ESTs had sequence similarities to proteins found in distantly related organisms, such as viruses, algae, bacteria, yeasts, and animals.

Out of 1216 ESTs, 904 were unique or nonredundant cDNA clones. Twenty-five percent redundancy was found in the cDNA library, and the redundant clones could be transcripts of the same gene or cognate genes. It is possible that the number of unique ESTs is overestimated, since they could be nonoverlapping cDNA fragments of the same gene. The most frequently represented genes are those encoding the microspore-specific protein I3 (GenBank accession no. S16569) (Roberts et al., 1991) and the lipid transfer protein (accession no. S22168) (Fleming et al., 1992), which appeared 19 and 17 times, respectively. Since an average of 320 bp were sequenced from the 5' ends of the cDNA inserts, the redundancy of a cDNA clone should not be thought to represent the expression level of the gene unless the size of the *in vivo* transcript is less than 0.6 kb. In a previous study of ESTs from flower buds of *Arabidopsis* (Höfte et al., 1993), the most redundant gene was the small subunit of Rubisco, which appeared 46 times in a total of 234 ESTs. The small subunit of Rubisco in our library was encountered 15 times in a total of 1216 ESTs. This disproportionate discrepancy can be explained by a difference in either species or developmental stages. Our EST analysis of flower buds resulted in a spectrum of genes expressed that differs from the spectrum found in *Arabidopsis*. Unique ESTs (904) were registered in Genome Sequence Data Base. The sequence data are also accessible in GenBank, EMBL, DNA Data Bank of Japan, and National Center for Biotechnology Information.

The deduced peptides of 628 ESTs that did not share homology with sequences in the protein databases were further examined at the nucleotide level using nucleotide sequence databases. Among the 628 cDNA clones, 184

showed significant sequence similarity to known nucleotide sequences in the databases, and 142 clones shared significant nucleotide sequence identity with plant ESTs previously reported from *Arabidopsis* (Höfte et al., 1993; Newman et al., 1994), maize (Keith et al., 1993), and rice (Uchimiya et al., 1992; Sasaki et al., 1994). To examine the significance of this finding, we again compared the *Brassica* cDNAs and the matching EST at the amino acid level using the TFASTA program. Of a total of 142 deduced peptides, 119 had more than 35% amino acid sequence similarity to peptide sequences of *Arabidopsis*, maize, and rice ESTs. In many instances, the sequence identity between the *Brassica* and other ESTs was less than 80% (100/119) at the amino acid level. Among these matched *Brassica* ESTs, 26 clones of sequences were present in both *Arabidopsis* and rice. Therefore, one can expect that these ESTs are ubiquitously present in both monocotyledonous and dicotyledonous plants, even though their biochemical and genetic functions are not yet known. Since they did not reveal significant homology to animal or microbe ESTs, they may be plant specific. For the remaining 444 clones (36%), we could not find significant similarities to sequences in either the protein or nucleotide databases.

DISCUSSION

We attempted to characterize expressed genes that were active during floral development. Partial sequences of 1216 randomly selected cDNA clones from developing flower buds of Chinese cabbage were obtained. Compared to a previous EST study of *Arabidopsis* flower buds with a total of 234 ESTs, this is a much more extensive EST analysis of flower organs. Our data, therefore, can supply significant information about the dynamics of genome expression during floral development.

Of the total ESTs, 48% (588/1216) carried cDNA with significant amino acid sequence similarities to previously identified genes deposited in protein databases. This is rather high for database matches. It was reported previously that the percentage of significant matches to known genes was 32% for *Arabidopsis* (Höfte et al., 1993). The lower percentage may be due to a more stringent cut-off score (greater than 120 of the sequence similarity) used during the database searches in that study. From the sequence analysis of these ESTs, we have identified 5 flower-specific ESTs. However, it is rather surprising to find only 5 ESTs out of a total of 1216 ESTs. One possible explanation would be that the current protein databases may have a very limited number of flower-specific protein sequences. In a previous study of ESTs with flower buds from *Arabidopsis*, Höfte et al. (1993) found no flower-specific ESTs from their 234 flower bud ESTs. This indicates that our cDNA library adequately represents transcript populations during floral development. Since many (118/589) of the identified cDNA clones from *Brassica* flower buds encode proteins with currently unknown functions, further study is required to determine how many flower-specific cDNAs are represented.

(Text continues on page 587.)

Table I. Putatively identified genes of *B. campestris* L. ssp. *pekinensis* flower bud cDNA ESTs

Clone	Putative Identification ^a	Organism	Percentage Id ^b	Overlap ^c	DB ^d	Acc. No. ^e
F1518	A6 protein (S31906)	Arabidopsis	86.9	99	P	L47870
F0083	Acetyl CoA-carboxylase (S35959)	Wheat	47.7	149	P	L33517
F0834	Acidic ribosomal protein PO (S37083)	Arabidopsis	73.5	68	P	L38531
F1875	Acidic ribosomal protein PO (S21519)	Red goosefoot	90.1	101	P	L47934
F0322	Actin (S31933)	Tobacco	91.8	73	P	L33569
F0453	Actin 1 (S10020)	Rice	58.8	102	P	L37615
F0266	Acyl carrier protein precursor (S00806)	Rape	96.9	128	P	L37453
F0904	Acyl carrier protein 1 (S14964)	Arabidopsis	76.6	47	P	L37463
F0942	Acyl carrier protein II (S12310)	Spinach	70.2	47	P	L37469
F0952	Acyl-[acyl-carrier-protein] desaturase (S31959)	Flax	69.5	131	P	L37473
F1649	ADP, ATP carrier protein (S29618)	Arabidopsis	69.2	104	P	L47890
F0591	Ala aminotransferase (P24298)	Human	45.2	73	S	L33632
F1867	Annexin (S30636)	Arabidopsis	69.6	56	P	L49930
F0350	Annexin VII (S14723)	Slime mold	58.4	77	P	L33574
F0061	Anther-specific protein (S26252)	Rape	60.1	138	P	L33509
F1040	Anther-specific protein Bcpl (JQ1327)	Field mustard	41.9	79	P	L37489
F1712	Anther-specific protein S18 (S38847)	Arabidopsis	41.3	80	P	L47901
F1960	AP3 (A42095)	Arabidopsis	77.8	108	P	L47953
F1736	APG protein (S21961)	Arabidopsis	52.9	70	P	L47906
F0893	Ara protein (JS0163)	Arabidopsis	45.0	129	P	L38533
F1941	Arg decarboxylase (JQ2341)	Tomato	91.2	34	P	L47947
F2007	L-Ascorbate peroxidase (S20866)	Arabidopsis	89.5	105	P	L47961
F1487	Asp transaminase (S18891)	Proso millet	68.3	82	P	L47863
F0624	Aspartic proteinase (S19697)	Barley	72.6	84	P	L33639
F1477	Auxin-induced protein 6B (S31098)	Arabidopsis	40.5	74	P	L47859
F1103	Auxin-induced protein Aux2-27 (S12244)	Arabidopsis	75.7	103	P	L37504
F0924	B2 protein (S32124)	Carrot	73.2	56	P	L37466
F0687	BBC 1 protein (S37271)	Arabidopsis	96.3	107	P	L33652
F1517	Gene bendless protein (S35793)	Fruit fly	76.8	82	P	L47869
F0167	bp4C protein (S12242)	Rape	52.5	80	P	L33534
F0965	Gene Bp10 protein (S24949)	Rape	98.6	70	P	L37479
F1789	Gene Bp10 protein (S24951)	Rape	73.0	111	P	L47918
F1656	Brittle-1 protein precursor (P29518)	Maize	47.2	89	S	L47893
F0955	Ca ²⁺ -transporting ATPase (A28065)	Rat	54.4	79	P	L37475
F1082	Caffeoyl-CoA 3-O-methyl-transferase (A40975)	Parsley	41.1	90	P	L37497
F0501	Calmodulin (S16138)	Carrot	77.2	114	P	L33613
F1469	Calmodulin (A49774)	<i>P. falciparum</i>	38.0	71	P	L47857
F1832	Calmodulin-like protein (S29595)	Arabidopsis	69.0	87	P	L47927
F0396	Calreticulin (S11205)	Rat	63.9	97	P	L33591
F2005	Carbonate dehydrogenase precursor (S28412)	Arabidopsis	63.0	92	P	L47959
F2020	Cellulase (S11946)	Avocado	38.8	98	P	L47965
F1556	Chalcone isomerase (JQ1687)	Arabidopsis	88.3	103	P	L47876
F0593	Chaperonin 10 protein (S29974)	Cattle	53.3	92	P	L33633
F1479	Casein kinase II (S31098)	Arabidopsis	95.2	83	P	L47860
F0027	Chlorophyll <i>a/b</i> -binding protein (S00442)	Petunia	96.8	93	P	L37608
F0143	Chlorophyll <i>a/b</i> -binding protein (S25435)	Arabidopsis	61.7	81	P	L37611
F0171	Chlorophyll <i>a/b</i> -binding protein (S06765)	White mustard	97.1	102	P	L37612
F0441	Chlorophyll <i>a/b</i> -binding protein (S22522)	Pine	87.1	116	P	L33601
F0468	Chlorophyll <i>a/b</i> -binding protein (S07408)	Tomato	41.1	107	P	L37616
F0507	Chlorophyll <i>a/b</i> -binding protein (S17737)	Tomato	45.8	203	P	L37618
F0682	Chlorophyll <i>a/b</i> -binding protein (A24717)	Petunia	78.8	85	P	L37621
F0784	Chlorophyll <i>a/b</i> -binding protein (S22511)	Wheat	53.7	149	P	L37628
F0810	Chlorophyll <i>a/b</i> -binding protein (S14306)	Tomato	58.7	104	P	L37631
F0829	Chlorophyll <i>a/b</i> -binding protein (S20917)	Cotton	77.4	53	P	L37632
F1012	Chlorophyll <i>a/b</i> -binding protein (S22511)	Mustard	95.4	109	P	L37645
F1658	Chlorophyll <i>a/b</i> -binding protein (S21386)	Barley	51.0	155	P	L47894
F1754	Chlorophyll <i>a/b</i> -binding protein (A30836)	White campion	80.8	78	P	L47911
F0582	Citrate (si)-synthase (JQ1392)	<i>C. burnetii</i>	44.4	81	P	L33630
F0641	Citrolysin-related protein 1 (S06446)	<i>C. freundii</i>	54.5	44	P	L33645

Table I. Continued

Clone	Putative Identification ^a	Organism	Percentage Id ^b	Overlap ^c	DB ^d	Acc. No. ^e
F0157	Clathrin-associated protein 19 (A40535)	Mouse	62.2	98	P	L33527
F0597	Clathrin heavy chain (A39941)	Rat	55.6	117	P	L33634
F0943	Cold-induced protein BnC24B (S37134)	Rape	51.3	199	P	L37470
F0040	COT 1 protein (S31302)	Yeast	35.9	64	P	L33498
F0633	Cyc07 protein, S-phase specific (JQ0939)	<i>M. periwinkle</i>	85.1	74	P	L33642
F0625	Cyclin (S16521)	Carrot	38.5	78	P	L33640
F0570	Cyclopropane fatty acid synthase (A44292)	<i>E. coli</i>	36.7	79	P	L33626
F0843	Probable Cys proteinase (S30149)	Tobacco	64.1	64	P	L38532
F0068	Cys proteinase inhibitor (S32164)	Cowpea	71.1	90	P	L33512
F1278	Cys proteinase inhibitor (S27239)	Maize	52.8	53	P	L47846
F1241	Cys proteinase tpp (S24602)	Pea	36.9	103	P	L47843
F0733	Cyt <i>b</i> ₅ (S33157)	Tobacco	66.7	105	P	L33668
F0325	Cyt <i>b</i> ₅ reductase (A23896)	Cattle	59.6	47	P	L33570
F0644	Cyt <i>b</i> ₆ - <i>f</i> complex Rieske iron-sulfur protein (S00454)	Spinach	85.1	87	P	L33646
F1983	Cyt <i>c</i> (A04613)	Cauliflower	97.7	87	P	L47957
F0380	Cyt <i>c</i> oxidase chain VIb (S03287)	Human	51.2	82	P	L33585
F1077	Cyt <i>c</i> reductase-processing peptidase (B48529)	Potato	59.1	66	P	L37495
F1723	Gene CYP 77A2 (S40266)	Eggplant	55.6	133	P	L47904
F1722	Cyt p450 Pbc2 (A00182)	Rabbit	37.8	74	P	L47903
F1512	Dehydrin dhn-cog (S25121)	Garden pea	56.1	66	P	L47867
F0301	2-Dehydro-3-deoxyphosphoheptonate aldolase 1 (A41370)	Arabidopsis	54.5	132	P	L33561
F1380	2-Dehydro-3-deoxyphosphooctonate aldolase (A30309)	<i>Escherichia coli</i>	44.2	77	P	L47850
F0417	3-Dehydroquinone synthase (A24863)	<i>E. coli</i>	54.5	110	P	L33595
F0392	Desiccation-related protein (D45509)	<i>C. plantagineum</i>	45.4	108	P	L33588
F0353	Dihydroflavonol 4-reductase (S34648)	Arabidopsis	53.8	65	P	L33576
F0148	Dihydroliipoamide S-succinyltransferase (A41015)	Rat	37.6	109	P	L33523
F0372	Disulfide-isomerase (A34930)	Mouse	44.0	109	P	L38527
F1083	DNA-binding E4 protein (JQ0988)	Tomato	46.7	90	P	L37498
F0218	DnaJ heat-shock protein (A47079)	<i>L. lactis</i>	35.2	88	P	L33548
F1546	dnaJ protein (S23509)	Human	49.0	49	P	L47873
F1687	DRT 112 protein (S33707)	Arabidopsis	72.4	98	P	L47896
F0439	EBER-associated protein (S13370)	Human	55.8	43	P	L33600
F0050	Elastin C (C26728)	Cattle	42.2	64	P	L33504
F0801	Elongation factor eEF-1 α (S17434)	Soybean	63.3	128	P	L37459
F1602	Elongation factor eEF-1 α (S08348)	Arabidopsis	89.2	111	P	L47882
F1968	Elongation factor eEF-1 α (S06724)	Arabidopsis	88.5	78	P	L47954
F0737	Elongation factor eEF-1 β -A1 chain (S37103)	Arabidopsis	94.5	55	P	L33669
F0379	Elongation factor Ts (A03525)	<i>E. coli</i>	55.8	86	P	L33584
F0584	Embryonic abundant protein precursor (S04136)	Tick bean	35.1	111	P	L33631
F1020	EMP protein (S25110)	Yeast	37.5	120	P	L37487
F0649	Epoxide hydrolase (S35587)	Human	43.6	110	P	L33647
F0047	Ethylene-forming enzyme (S22488)	Leaf mustard	86.8	106	P	L33502
F0514	Extensin (S14984)	Tomato	39.7	63	P	L33618
F0535	F59B2.7 protein (S31127)	<i>C. elegans</i>	64.0	75	P	L33621
F0153	Fd (S09979)	Arabidopsis	92.2	51	P	L33524
F1994	Fd (A00234)	Rape	97.5	80	P	L47958
F0019	Fibrillarlin (S33690)	Yeast	67.4	92	P	L33496
F0702	fil 1 protein (S17699)	Garden Snapdragon	53.3	45	P	L33655
F0207	FK506/rapamycin-binding protein FKBP13 (JC1365)	Human	63.0	81	P	L33547
F0166	Flavonol 3-O-glucosyltransferase (S01052)	Maize	35.7	115	P	L33533
F0054	Flavonoid 3',5'-hydroxylase (S33515)	Petunia	42.9	154	P	L33505
F1702	Flavonol-4'-sulfotransferase (A40216)	<i>F. chloraefolia</i>	51.4	107	P	L47897
F0452	β -Fructofuranosidase (S31925)	Potato	45.6	125	P	L33602
F0314	Fru-bisphosphate aldolase (S31091)	Spinach	72.7	143	P	L33566
F0612	Fru-bisphosphate aldolase (S29048)	Pea	81.7	71	P	L37620

Table I. Continued

Clone	Putative Identification ^a	Organism	Percentage Id ^b	Overlap ^c	DB ^d	Acc. No. ^e
F1513	fsh membrane protein (A43742)	Fruit fly	54.9	82	P	L47868
F0498	gag polyprotein (A41991)	Anemia virus	40.0	35	P	L33612
F0910	α -galactosidase (JQ1021)	Yeast	36.2	47	P	L37464
F0472	GAST 1 protein (S22151)	Tomato	56.6	53	P	L33606
F1645	Gastrula zinc finger protein (P18724)	Frog	38.6	70	S	L47889
F0308	Gene C98 protein (S24960)	Rape	88.0	133	P	L33564
F0958	Geranyltranstransferase (JX0257)	<i>B. stearrowthermophilus</i>	37.9	87	P	L37477
F1874	GF14/G box binding factor (A47237)	Arabidopsis	95.2	42	P	L47933
F1111	β -1,3-Glucanase (S31612)	Rape	86.1	72	P	L37506
F0092	Glc-6-P isomerase (A36567)	<i>P. falciparum</i>	40.0	110	P	L33519
F0928	Glc transport protein (S09705)	Rat	35.0	80	P	L37467
F1507	β -Glucosidase (S23940)	Cassava	35.7	98	P	L47866
F1889	β -Glucosidase (S16581)	White clover	60.6	66	P	L47939
F1000	Glutamate-ammonia ligase (A26025)	Alfalfa	75.0	84	P	L37483
F0079	Glutathione peroxidase (S20501)	Tobacco	66.0	106	P	L33515
F0393	Glutelin 2 precursor (A23014)	Maize	48.2	56	P	L38540
F0219	Glyceraldehyde-3-phosphate dehydrogenase (A24796)	White mustard	90.8	130	P	L33549
F0530	Glyceraldehyde-3-phosphate dehydrogenase (A24430)	Tobacco	86.0	57	P	L37619
F0383	Glyceraldehyde-3-phosphate dehydrogenase (S14243)	Pea	72.8	114	P	L37614
F1794	Gly-rich protein (S32123)	Carrot	53.2	111	P	L47919
F1927	Gly-rich protein (S14857)	Carrot	39.1	69	P	L47944
F1396	Gly-rich protein 2 (JQ1061)	Arabidopsis	58.6	111	P	L47853
F0487	Gly-rich protein 5 (JQ1064)	Arabidopsis	73.6	87	P	L33608
F0185	Gly-rich protein atGRP-6 (S19932)	Arabidopsis	54.8	84	P	L33543
F1942	Gly-rich protein atGRP-7 (S19933)	Arabidopsis	54.5	101	P	L47948
F1136	Gly-rich cell-wall structure protein (S17732)	Arabidopsis	81.4	86	P	L37511
F0159	Gly-rich RNA-binding protein (S31443)	Arabidopsis	82.1	78	P	L33528
F0927	Glycogen synthase (S16555)	Mouse	77.2	92	P	L38534
F0080	GTP-binding protein (S28875)	Arabidopsis	65.7	99	P	L33516
F0204	GTP-binding protein ara-3 (JS0640)	Arabidopsis	98.5	65	P	L33546
F1548	GTP-binding protein β chain (A33928)	Chicken	50.0	58	P	L47874
F1940	GTP-binding protein rab (S33531)	Garden pea	82.5	80	P	L47946
F1483	GTP-binding protein rpg 1 (S16554)	Rice	51.2	43	P	L47861
F0005	GTP-binding protein Sar 1 (S28603)	Arabidopsis	73.2	164	P	L33494
F1753	GTP-binding protein ypt (B38202)	Maize	84.7	98	P	L47910
F1570	Heat-shock protein (S00646)	Soybean	40.0	85	P	L47878
F0638	Heat-shock protein 26A (A33654)	Soybean	56.9	72	P	L33643
F0754	Heat-shock cognate protein 70 (S25005)	Bean	35.6	90	P	L33674
F0746	Heat-shock cognate protein 70 (S36623)	Arabidopsis	97.1	69	P	L33670
F0016	Heat-shock protein 82 (S25541)	Rice	92.2	115	P	L33495
F1024	H ⁺ -transporting ATPase (A40814)	Oat	91.6	95	P	L37488
F1261	H ⁺ -transporting ATP synthase (B39732)	Arabidopsis	96.6	58	P	L47845
F1820	H ⁺ -transporting ATP synthase (S34473)	Spinach	47.8	69	P	L47923
F2017	H ⁺ -transporting ATP synthase (A01028)	Spinach	87.0	69	P	L47964
F0564	H ⁺ -transporting ATP synthase β chain (JQ0230)	Rice	84.1	126	P	L33624
F1141	Histone H1 (S18053)	Arabidopsis	70.5	44	P	L37513
F0327	Histone H2A (JQ1183)	Pea	77.0	126	P	L33571
F1009	Histone H2A.IV (JQ0796)	<i>V. carteri</i>	64.3	129	P	L37485
F0046	Histone H2B (S30619)	Arabidopsis	68.2	66	P	L38525
F0193	Histone H2B (S28048)	Maize	91.6	95	P	L33545
F0512	Histone H3 (S06250)	Arabidopsis	77.4	137	P	L33616
F1533	Histone H3 (S04099)	Rice	77.7	130	P	L47872
F1506	Histone H3.1 (S28528)	Human	77.9	122	P	L47865
F0713	Histone H3.3-like protein (S24346)	Arabidopsis	93.3	90	P	L33660
F0373	Histone H4 (A25642)	Maize	100	103	P	L33581
F0059	HMG-1-like protein (S22309)	Soybean	53.0	83	P	L33508
F0778	Hyp-rich glycoprotein (S06733)	Tobacco	36.5	96	P	L37457
F0062	Hypothetical protein (S24835)	Arabidopsis	60.6	94	P	L33510

Table 1. Continued

Clone	Putative Identification ^a	Organism	Percentage Id ^b	Overlap ^c	DB ^d	Acc. No. ^e
F0069	Hypothetical protein 2 (S22515)	Barley	46.2	65	P	L33513
F0161	Hypothetical protein (S33464)	Arabidopsis	54.2	59	P	L33529
F0358	Hypothetical protein (S12209)	Tomato	51.2	123	P	L33578
F0575	Hypothetical protein pPLZ12 (S14688)	Lupine	62.4	85	P	L33627
F0692	Hypothetical protein (S12411)	Duckweed	83.1	65	P	L33653
F0720	Hypothetical protein (S11850)	Strawberry	45.6	90	P	L33662
F0726	Hypothetical protein (S10911)	Carrot	41.3	80	P	L33663
F1042	Hypothetical protein (S38378)	<i>M. periwinkle</i>	80.8	52	P	L37491
F1868	Hypothetical protein 17 (S11690)	<i>Bacillus subtilis</i>	40.0	95	P	L47931
F1741	L-Iditol 2-dehydrogenase (A45052)	<i>B. subtilis</i>	40.4	114	P	L47907
F0401	Initiation factor 5a (S31362)	Arabidopsis	90.3	72	P	L33592
F0260	Initiation factor eIF-2 α chain (A32108)	Yeast	45.4	66	P	L33555
F0712	Initiation factor eIF-5A.2 (S21059)	Tobacco	71.8	142	P	L33659
F1446	Isocitrate dehydrogenase (S33612)	Soybean	59.4	64	P	L47856
F0432	10-K protein (S04126)	Barley	50.7	69	P	L33598
F0155	26-K antigen (A33168)	<i>H. pylori</i>	100	28	P	L33526
F1787	Keratin, 67K type II (A44861)	Human	35.5	107	P	L47916
F1788	Keratin 3, type I (S01327)	Frog	37.0	100	P	L47917
F0602	Ketol-acid reductoisomerase (S30145)	Arabidopsis	91.9	99	P	L33635
F0503	KIN 1 protein (S29471)	Arabidopsis	38.5	65	P	L33614
F1389	D-Lactate dehydrogenase (S17556)	<i>L. delbrueckii</i>	51.5	66	P	L47851
F0534	Laminin receptor (S30570)	Arabidopsis	93.6	109	P	L33620
F1877	Laminin receptor (S31352)	Arabidopsis	98.4	61	P	L47935
F1873	LEA 76 protein (S38452)	Arabidopsis	37.5	64	P	L47932
F1398	Lipid transfer protein (S33461)	Sorghum	45.3	64	P	L47854
F0221	Lipid transfer protein (S22528)	Wheat	37.8	111	P	L33550
F0964	Lipid transfer protein (S07409)	Barley	46.7	105	P	L38535
F0734	Probable lipid transfer protein precursor (S20862)	Tomato	47.1	68	P	L38529
F0354	Major histocompatibility complex-encoded proteasome (B44324)	Human	36.6	123	P	L33577
F1980	Major latex protein (S38456)	Arabidopsis	81.0	79	P	L47956
F0580	Malate dehydrogenase (S28987)	Pig	67.2	125	P	L33629
F0766	Malate dehydrogenase (S10162)	Watermelon	91.9	111	P	L37624
F1909	Malate dehydrogenase (A34482)	Maize	66.7	42	P	L47942
F1086	Metallothionein I (S37234)	Arabidopsis	79.4	63	P	L37499
F0377	Metallothionein-like protein (S18069)	Arabidopsis	92.6	81	P	L33583
F0769	5-Methyltetrahydrofolate (A42863)	<i>E. coli</i>	41.6	113	P	L37456
F0463	Microspore-specific protein I3 (S16569)	Rape	80.2	81	P	L33603
F0065	Mov-34 protein (A40556)	Mouse	59.5	84	P	L33511
F1782	MSS 1 protein (S24353)	Human	89.8	88	P	L47915
F1521	Mucorpepsin (A29039)	<i>R. miehei</i>	35.5	62	P	L47871
F0966	myo-Inositol-1-phosphate synthase (B32209)	Yeast	62.9	97	P	L37480
F0413	NADH dehydrogenase 24-K chain (A30113)	Human	47.5	80	P	L33594
F0255	NADH dehydrogenase 39-K chain (S17676)	Cattle	43.8	121	P	L33553
F0163	NADPH dehydrogenase chain OYE2 (A46009)	Yeast	41.1	95	P	L33531
F1138	NAM8 protein (S22439)	Yeast	46.9	81	P	L37512
F2006	Naringenin 3-dioxygenase (S32154)	<i>M. incana</i>	95.7	46	P	L47960
F0022	Naringenin-chalcone synthase (S06877)	White mustard	84.5	58	P	L37607
F0756	Naringenin-chalcone synthase (S11876)	<i>M. incana</i>	67.0	106	P	L33675
F1882	NEDD-6 protein (S38851)	Arabidopsis	92.6	95	P	L47936
F0165	Nitrogen fixation protein nifU (D34443)	<i>Anabaena</i>	39.2	102	P	L33532
F0946	Nodulin-21 (S08632)	Soybean	55.6	36	P	L37472
F0577	Nonspecific lipid-transfer protein (P19656)	Maize	45.8	83	S	L33628
F1633	Nucleotide diphosphate kinase (S31444)	Arabidopsis	88.7	106	P	L47886
F1255	OEE 1 protein (S09383)	Arabidopsis	84.1	69	P	L47844
F1504	Oleosin (P29529)	Sunflower	71.1	114	S	L47864
F1059	ω fatty acid desaturase (A44227)	Rape	36.5	85	P	L37494

Table I. Continued

Clone	Putative Identification ^a	Organism	Percentage Id ^b	Overlap ^c	DB ^d	Acc. No. ^e
F0551	Oryzain α (JU0388)	Rice	40.0	60	P	L33623
F0604	P59 protein (P27124)	Rabbit	45.7	46	S	L33636
F1099	parC protein (S19185)	Tobacco	56.5	85	P	L37502
F0055	Pathogenesis-related protein 5 (JQ1695)	Arabidopsis	37.8	98	P	L33506
F0351	Pectate lyase LAT59 (S27098)	Tomato	43.6	94	P	L33575
F0727	Pectin esterase (S14952)	Rape	68.5	146	P	L33664
F0394	Pectin esterase-related protein (S14952)	Rape	49.3	134	P	L33589
F1392	Peptidylpropyl isomerase (A39252)	Tomato	82.1	95	P	L47852
F0256	Peptidylpropyl isomerase (B39252)	Rape	90.8	130	P	L33554
F1650	Peptidylpropyl isomerase (A40516)	Chicken	71.8	71	P	L47891
F1707	Peptidyl-prolyl- <i>cis</i> -transisomerase (P34791)	Arabidopsis	78.1	96	S	L47898
F0310	Phosphoglucomutase 1 (A41801)	Human	65.3	72	P	L33565
F0267	Phosphoglycerate kinase (S05966)	Wheat	75.0	144	P	L33558
F1239	Phosphoglycerate mutase (A33793)	Rat	39.7	58	P	L47842
F0361	Phospholipid transfer protein (S06427)	Rice	53.8	80	P	L33579
F0673	Phospholipid transfer protein (S21757)	Wheat	44.0	50	P	L38528
F0732	Phospholipid transfer protein 9C2 (JH0378)	Maize	43.8	73	P	L33667
F0964	Probable phospholipid transfer protein precursor (S07409)	Barley	46.7	105	P	L38535
F1710	Probable phospholipid transfer protein (S14610)	Barley	39.8	98	P	L47899
F1711	Phosphopyruvate hydratase (S07586)	Fruit fly	60.0	110	P	L47900
F0467	PSI 18K protein (A39759)	Barley	76.0	92	P	L33605
F0686	PSI protein psaH (S00453)	Spinach	76.6	107	P	L33651
F0879	PSI chain II (A60695)	Cucumber	93.2	88	P	L37461
F0336	PSI chain IV (S00450)	Spinach	77.4	62	P	L33573
F0174	PSI chain XI (S35151)	Spinach	60.9	92	P	L33537
F0085	PSII 5-K protein (S29447)	Arabidopsis	78.8	80	P	L33518
F0565	PSII 7-K protein (S29418)	Arabidopsis	82.1	112	P	L33625
F1554	PSII 10-K protein (S17430)	Arabidopsis	86.5	111	P	L47875
F0710	PSII 22-K protein (S26436)	Spinach	90.5	105	P	L33658
F0364	PSII oxygen-evolving complex protein (S00008)	Spinach	55.2	125	P	L33580
F1772	PSII oxygen-evolving complex protein 23K (S10016)	White mustard	91.5	82	P	L47897
F0954	Placental protein 15 (S00751)	Human	42.9	84	P	L37474
F0622	Pollen-preferential protein (S29611)	Lily	52.9	70	P	L33638
F1829	Pollen-specific protein precursor (S36466)	Arabidopsis	75.7	74	P	L47926
F1885	Pollen-specific protein precursor (S22495)	Tobacco	61.8	76	P	L47937
F0465	Polygalacturonase (S32008)	Tobacco	40.3	62	P	L33604
F1652	Polygalacturonase (S32010)	Tobacco	39.6	101	P	L47892
F0783	Polygalacturonase 1 beta-chain (JQ1670)	Tomato	66.3	104	P	L37458
F0298	Polygalacturonase P22 (JQ0992)	Evening primrose	67.1	76	P	L33560
F1151	Polygalacturanase-inhibiting protein (S23764)	Kidney bean	39.1	115	P	L37515
F0304	Porin (S34146)	Maize	38.0	187	P	L33562
F0493	Profilin 2 (S35797)	Maize	72.4	105	P	L33611
F0391	Pro-rich protein (S31096)	Tobacco	44.6	112	P	L33587
F1569	Pro-rich protein TPRP-F1 (S19129)	Tomato	61.2	116	P	L47877
F0044	Protease inhibitor II (S30578)	Arabidopsis	87.0	77	P	L33500
F1058	Protein kinase (A30311)	Arabidopsis	49.2	59	P	L37493
F1003	Protein kinase 6 (S27760)	Soybean	51.9	81	P	L37484
F0038	Protein kinase BKIN 12 (S24578)	Barley	41.2	68	P	L33497
F1814	Probable protein kinase cot-1 (S22711)	<i>N. crassa</i>	40.3	119	P	L47922
F0045	PRT1 protein (A29562)	Yeast	36.8	68	P	L33501
F1635	Receptor-like protein kinase (S27754)	Arabidopsis	46.3	54	P	L47887
F1747	Retrovirus-related polyprotein (A03324)	Fruit fly	47.6	105	P	L47908
F0764	Rho1Ps=ras-related small GTP-binding protein (A47525)	Garden pea	86.2	58	P	L37455
F0375	Ribosomal protein ML 16 (S28586)	Ice plant	74.5	137	P	L33582
F0731	Ribosomal protein L5 (JC1308)	Chicken	58.2	134	P	L33666
F0261	Ribosomal protein L5b (B33823)	Frog	67.8	90	P	L33556
F1857	Ribosomal protein L7.e.A (S22789)	Yeast	42.5	106	P	L47928

Table I. Continued

Clone	Putative Identification ^a	Organism	Percentage Id ^b	Overlap ^c	DB ^d	Acc. No. ^e
F0486	Ribosomal protein L9 (S19978)	Pea	75.3	97	P	L33607
F0747	Ribosomal protein L11 (S17351)	Rat	65.1	86	P	L33671
F0632	Ribosomal protein L17 (S31354)	Arabidopsis	91.2	68	P	L33641
F0182	Ribosomal protein L17-1 (S35101)	Barley	78.0	91	P	L33542
F1105	Ribosomal protein L18a (S37576)	Fruit fly	45.7	92	P	L37505
F0320	Ribosomal protein L18b (B25766)	Frog	58.5	135	P	L33568
F1081	Ribosomal protein L19 like (S30588)	Arabidopsis	82.0	82	P	L37496
F0154	Ribosomal protein L23 (S18815)	Human	76.0	125	P	L33525
F1949	Ribosomal protein L23 (JH0418)	Rat	83.3	108	P	L47951
F0424	Ribosomal protein L26 (S05024)	Rat	67.3	110	P	L33597
F0407	Ribosomal protein L27 (S26612)	Green alga	67.7	96	P	L33593
F0170	Ribosomal protein L27a (S29458)	Arabidopsis	98.9	87	P	L33536
F1134	Ribosomal protein L30 (S11622)	Mouse	71.9	89	P	L37510
F1424	Ribosomal protein L31 (A26417)	Rat	46.0	100	P	L47855
F1947	Ribosomal protein L31 (S24989)	<i>Chlamydomonas reinhardtii</i>	56.5	92	P	L47950
F1900	Ribosomal protein L34 (S04271)	Rat	41.8	79	P	L47940
F1726	Ribosomal protein L35 (A34571)	Rat	56.4	110	P	L47905
F1714	Ribosomal protein L36 (JN0483)	Rat	56.0	100	P	L47902
F1120	Ribosomal protein L37 (JN0478)	Rat	66.7	87	P	L37655
F0423	Ribosomal protein L37 (S21496)	Rat	53.2	94	P	L33596
F0306	Ribosomal protein L37a (S34661)	Turnip	94.6	93	P	L33563
F1098	Ribosomal protein L39 (A02780)	Rat	74.0	50	P	L37501
F0762	Ribosomal protein S2 (S18828)	Rat	47.8	113	P	L37454
F1486	Ribosomal protein S3 (S13109)	Human	82.8	64	P	L47862
F1077	Ribosomal protein S3a (S15665)	Frog	82.6	109	P	L33538
F2012	Ribosomal protein S5 (S14606)	Rat	90.5	63	P	L47962
F0979	Ribosomal protein S8 (S38421)	Yeast	56.4	55	P	L37481
F0987	Ribosomal protein S10 (S01881)	Rat	48.2	56	P	L37482
F0329	Ribosomal protein S11 (C35542)	Arabidopsis	89.0	118	P	L33572
F0781	Ribosomal protein S12 (S14482)	Human	60.7	56	P	L38530
F1769	Ribosomal protein S12 (S29454)	Arabidopsis	65.6	61	P	L47912
F0537	Ribosomal protein S13 (A35889)	Rat	78.9	95	P	L33622
F0718	Ribosomal protein S14 (A30097)	Maize	90.8	87	P	L33661
F1600	Ribosomal protein S14 (S05618)	<i>H. vannielli</i>	59.1	44	P	L47880
F1618	Ribosomal protein S15 (S34016)	Arabidopsis	83.3	114	P	L47883
F0043	Ribosomal protein S17 (JT0405)	Human	78.5	79	P	L33499
F0489	Ribosomal protein S19 (S10392)	Rat	57.4	122	P	L33610
F0676	Ribosomal protein S20 (S14682)	Rat	72.3	83	P	L33649
F1952	Ribosomal protein S20 (S38356)	Rice	72.2	108	P	L47952
F0706	Ribosomal protein S26 (S30652)	Arabidopsis	94.0	84	P	L33657
F0660	Ribosomal protein S28 (JQ1170)	Rat	77.0	61	P	L33648
F1472	Ribosomal protein S28.e (S30006)	Yeast	73.1	67	P	L47858
F0883	Ribosomal protein YL10 (S25633)	Midge	60.7	61	P	L37462
F0317	Rieske iron-sulfur protein (B41607)	Tobacco	86.3	80	P	L33567
F0250	Ripening associated membrane protein (S34651)	Tomato	81.2	138	P	L33552
F2015	RNA-binding protein RNP-T (S28057)	Arabidopsis	78.2	78	P	L47963
F0088	Rubisco small chain (S24794)	Kidney bean	43.2	132	P	L37610
F0309	Rubisco small chain precursor (S16253)	Rape	91.7	109	P	L37613
F0473	Rubisco small chain precursor (S00934)	Rape	95.7	115	P	L37617
F1369	Rubisco (S37575)	Rape	89.7	68	P	L47849
F1821	Rubisco (S04048)	Arabidopsis	84.3	102	P	L47924
F0048	Rubisco subunit-binding protein (S02119)	Castor bean	60.5	129	P	L33503
F0180	S-Adenosylhomocysteine hydrolase (A45569)	<i>L. donovani</i>	68.2	88	P	L33540
F0187	Salt-associated protein csaA (S33618)	Sweet orange	66.7	36	P	L33544
F0178	Ser-proteinase inhibitor (S21120)	White mustard	54.2	59	P	L33539
F1150	Ser-type carboxypeptidase (B29639)	Wheat	48.1	129	P	L37514
F0141	Signal recognition particle receptor (A24570)	Dog	44.8	125	P	L33521
F0071	snRNP-E related protein C29 (P24715)	Alfalfa	46.2	39	S	L33514
F0382	Spermatid-specific protein T2 (B40973)	Cuttlefish	44.1	59	P	L33586
F1801	SRP1 protein (S30884)	<i>Saccharomyces cerevisiae</i>	51.8	85	P	L47920

Table I. Continued

Clone	Putative Identification ^a	Organism	Percentage Id ^b	Overlap ^c	DB ^d	Acc. No. ^e
F0956	Starch branching enzyme RBE3 (A48537)	Rice	78.6	98	P	L37476
F1133	Stress-inducible protein sti35 (A37767)	<i>F. solani</i>	62.4	93	P	L37509
F1924	Strictosidine-synthase (S01325)	Serpentwood	36.8	87	P	L47943
F0695	Superoxide dismutase (Cu-Zn) (A25569)	Cabbage	88.1	101	P	L33654
F1326	Superoxide dismutase (Cu-Zn) (S12313)	Garden pea	60.6	94	P	L47847
F1628	Superoxide dismutase (Cu-Zn) (S19117)	Arabidopsis	94.9	78	P	L47885
F1620	Superoxide dismutase II (Cu-Zn) (S12313)	Garden pea	67.5	83	P	L47884
F0856	Superoxide dismutase (Mn) (S03639)	Tobacco	36.4	110	P	L37460
F1945	Tat-binding protein 1 (A34832)	Human	73.9	119	P	L47949
F0605	Thiamine biosynthetic enzyme (S35117)	<i>E. coli</i>	77.2	92	P	L33637
F0263	Thiolase (S33637)	Cucumber	85.5	124	P	L33557
F1644	Thionin (S22515)	Barley	45.9	98	P	L47888
F1121	Thioredoxin h1 (A28086)	Tobacco	40.0	60	P	L37508
F1677	Thioredoxin h1 (S16590)	Tobacco	40.7	91	P	L47895
F1057	Thioredoxin reductase (A28074)	<i>E. coli</i>	57.3	82	P	L37492
F1601	tma protein (S28533)	<i>L. lactis</i>	35.2	91	P	L47881
F0959	TobRB7-5A protein (JQ1011)	Tobacco	46.5	142	P	L37478
F0932	Tonoplast intrinsic protein (S22202)	Arabidopsis	84.3	70	P	L37468
F1572	Tonoplast intrinsic protein (S36463)	Arabidopsis	94.4	90	P	L47879
F1904	Tonoplast intrinsic protein (S30634)	Arabidopsis	78.0	59	P	L47941
F1803	Transcription factor UBF 2 (S17977)	Frog	35.7	84	P	L47921
F0640	Transforming protein (myb) (A25075)	Chicken	56.7	97	P	L33644
F0510	Transplantation antigen P198 (JL0149)	Mouse	63.2	87	P	L33615
F1334	Probable transcription factor DdTBP 2 (JN0611)	Slime mold	82.9	70	P	L47848
F0751	Probable transcription factor DdTBP10 (JN0610)	Slime mold	57.1	91	P	L33673
F0944	Triose-phosphate isomerase (A25501)	Maize	58.8	102	P	L37471
F1976	Triose phosphate/3-phosphoglycerate	Tobacco	36.5	85		
F0395	Tropomyosin-related protein (A60021)	Rat	37.3	67	P	L33590
F0168	Tryptophane synthase (S31843)	Red alga	47.3	112	P	L38526
F0911	TUB13 protein (S28047)	Potato	70.2	104	P	L37465
F1091	Tubulin α -5 chain (A32712)	Arabidopsis	88.6	105	P	L37500
F0748	Tubulin β -1 chain (S20868)	Pea	70.9	134	P	L33672
F1928	Tumor protein (S30551)	Arabidopsis	90.0	50	P	L47945
F1886	U1 snRNP 70K protein (S28147)	Arabidopsis	37.0	73	P	L47938
F0513	U2 snRNP A' (S30580)	Arabidopsis	91.3	46	P	L33617
F0181	Ubiquitin precursor (S01425)	Arabidopsis	100	123	P	L33541
F0076	Ubiquitin precursor (S06921)	Garden pea	75.9	116	P	L37609
F1780	Ubiquitin precursor (S03599)	Sunflower	95.3	86	P	L47914
F0522	Ubiquitin conjugating enzyme (S32674)	Arabidopsis	38.7	150	P	L33619
F1748	Ubiquitin conjugating enzyme (S31971)	Arabidopsis	90.9	55	P	L47909
F1114	Ubiquitin-conjugating enzyme UBC10 (S32672)	Arabidopsis	98.4	61	P	L37507
F1019	Ubiquitin fusion protein UBF9 (JS0657)	Maize	97.0	99	P	L37486
F0685	Ubiquitin/ribosomal protein CEP52 (S28420)	Tobacco	95.6	90	P	L33650
F1824	Ubiquitin/ribosomal protein CEP52.1 (A29456)	Yeast	62.5	128	P	L47925
F1859	Ubiquitin/ribosomal protein S27a	Tomato	100	91	P	L47929
F0096	UTP-Glc 1-P uridylyltransferase (JX0128)	Potato	47.0	185	P	L33520
F0162	Valosin-containing protein (S25197)	Mouse	57.9	114	P	L33530
F1101	Viscotoxin (S16099)	Mistletoe	50.0	72	P	L37503
F0297	Wilm's tumor suppressor (S29906)	Arabidopsis	95.2	124	P	L33559

^a The descriptions of the database match are listed together with the accession numbers of the match in parentheses. ^b Percentage Id, Percentage identity. ^c Overlap indicates the number of amino acid residues between a quarry sequence and its matched protein sequence. ^d DB, Database. Database abbreviations: P, Protein Identification Resource Data Bank; S, SwissProt. ^e Acc. No., Accession numbers of Chinese cabbage clones.

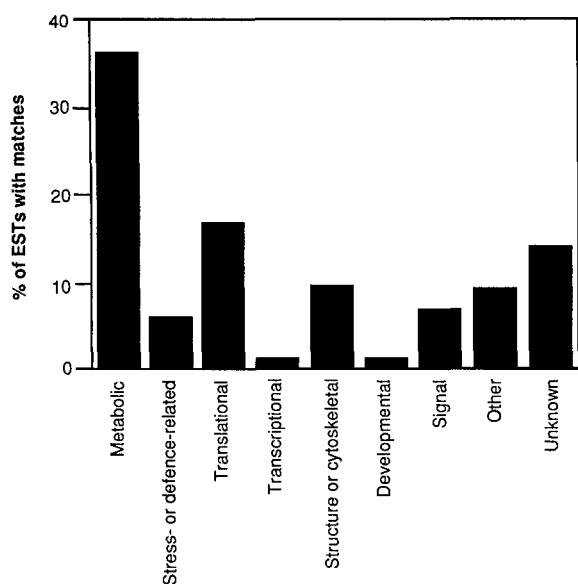


Figure 1. Functional classification of *B. campestris* L. ssp. *pekinensis* flower bud ESTs. The ESTs that had sequence similarity to known proteins were classified based on their biological functions.

(Text continues from page 579.)

Of the 588 ESTs that have sequence similarity to known proteins, 124 clones shared sequence homology with non-plant genes. At present it is not possible to assign functional roles for these proteins in plants. The functional classification of the putatively identified genes listed in Figure 1 shows that metabolism-related genes are the most prevalent among the identified cDNA clones. cDNA clones encoding various ribosomal proteins are also abundant. The data suggest that cells in flower buds are metabolically quite active, an observation also made in the cases of *Arabidopsis* (Höfte et al., 1993) and rice (Uchimiya et al., 1992; Sasaki et al., 1994).

Using the ESTs in this study, we could not find sequence similarities to known proteins in databases for more than 50% of the cDNAs. To define the functional identities of these unidentified genes will require extensive biochemical and genetic studies. When we compared our unidentified sequences with other plant ESTs, we found 142 clones with sequence similarity to other plant ESTs at the nucleotide level, indicating that they may possibly encode similar polypeptides. Of these, 119 ESTs showed similarity with ESTs from other plants at the amino acid level. One way to analyze a large number of unidentified clones is to define sequences highly conserved among homologous ESTs from various plants (Sasaki et al., 1994). If a large number of ESTs from various species were available, homologous peptides or nucleotides aligned with novel ESTs might help to classify anonymous genes. Furthermore, highly conserved domains that determine sequence homology also help to elucidate putative functions.

ACKNOWLEDGMENTS

The authors thank Dr. Do Hyun Lee for supplying Chinese cabbage flower buds and seeds and Sang Hyoung Lee and Ja

Choon Koo for technical help in the computer analysis of sequences. We also thank Dr. Chang-deok Han for critical reading of the manuscript.

Received October 6, 1995; accepted January 29, 1996.

Copyright Clearance Center: 0032-0889/96/111/0577/12.

The nucleotide sequence data reported in this article will appear in the Genome Sequence Data Base, the EMBL Data Library, the DNA Data Bank of Japan, and the National Center for Biotechnology Information under the following accession numbers: L33494–L33508, L33510–L33525, L33527–L33534, L33536–L33675, L35773, L35774, L35777–L35790, L35792, L35794–L35796, L35798–L35810, L35812–L35815, L35817–L35822, L35824–L35829, L35831–L35833, L35835–35843, L37453–L37515, L37607–L37659, L37974–L38233, L38525–L38543, L47842–L47912, L47914–L47929, L47931–L47965, L49930.

LITERATURE CITED

- Adams MD, Dubnick M, Kerlavage AR, Moreno R, Kelley JM, Utterback TR, Nagle JW, Fields C, Venter JC (1992) Sequence identification of 2,375 human brain genes. *Nature* 355: 632–634
- Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252: 1651–1656
- Adams MD, Kerlavage AR, Fields C, Venter JC (1993) 3,400 new expressed sequence tags identify diversity of transcripts in human brain. *Nature Genet* 4: 256–267
- Albani D, Robert LS, Donaldson PA, Altosaar I, Arnison PG, Fabijanski SF (1990) Characterization of a pollen-specific gene family from *Brassica napus* which is activated during early microspore development. *Plant Mol Biol* 15: 605–622
- Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1992) Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology, Ed 2. John Wiley & Sons, New York, pp 4.8–4.9
- Boguski MS, Lowe TMJ, Tolstoshew CM (1993) dbEST-database for "expressed sequence tags." *Nature Genet* 4: 332–333
- Boguski MS, Tolstoshew CM, Basset DE (1994) Gene discovery in dbEST. *Science* 265: 1993–1994
- Croy EJ, Ikemura T, Shirsat A, Croy RRD (1993) Plant nucleic acids. In RRD Croy, ed, *Plant Molecular Biology LABFAX*. BIOS Scientific Publishers, Durham, UK, pp 21–48
- Fleming A, Mandel T, Hofmann S, Sterk P, de Vries S, Kuhlemeier C (1992) Expression pattern of a tobacco lipid transfer protein gene within the shoot apex. *Plant J* 2: 855–862
- Höfte H, Desprez T, Amselem J, Amselem J, Chiapello H, Caboche M, Moisan A, Jourjon MF, Charpentier JL, Berthomieu P, Guerrier D, Giraudat J, Quigley F, Thomas F, Yu DY, Mache R, Raynal M, Cooke R, Grellet F, Delseny M, Parmentier Y, Matcillac GD, Gigot C, Fleck J, Philipps G, Axelos M, Bardet C, Tremousaygue D, Lescure B (1993) An inventory of 1,152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *Plant J* 4: 1051–1061
- Höög 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 CS, Hoang DO, Barrett BM, Feigelman B, Nelson MC, Thai H, Baysdorfer C (1993) Partial sequencing analysis of 130 randomly selected maize cDNA clones. *Plant Physiol* 101: 329–332
- Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J, Antonio BA, Shormura A, Shimizu T, Lin S-Y, Inoue T, Fukuda A, Shimano T, Kuboki Y, Toyama T, Miyamoto Y, Kirihara T, Fukuda A, Shimano T, Kuboki Y, Toyama T, Miyamoto Y, Kirihara T, Hayasaka K, Miyao A, Monna L, Zhong HS, Tamura Y, Wang Z-X, Momma T, Umehara Y, Yano M, Sasaki T, Minobe Y (1994) A 300 kilobase interval genetic map of rice including 883 expressed sequences. *Nature Genet* 8: 365–372

- McCombie WR, Adams MD, Kelley JM, FitzGerald MG, Utterback TR, Khan M, Dubnick M, Kerlavage AR, Venter JC, Fields C (1992) *Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease gene homologues. *Nature Genet* 1: 124–131
- Nacken WKF, Huijser P, Beltran JP, Saedler H, Sommer H (1991) Molecular characterization of two stamen-specific genes, *tap1* and *fil1*, that are expressed in the wild type, but not in the deficient mutant of *Antirrhinum majus*. *Mol Gen Genet* 299: 129–136
- Newman T, de Bruijin FJ, Green P, Keegstra K, Kende H, McIntosh L, Ohlrogge J, Raikhel N, Somerville S, Thomashow M, Retzel E, Somerville C (1994) Genes galore. A summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. *Plant Physiol* 106: 1241–1255
- Park YS, Kwak JM, Kwon OY, Kim YS, Lee DS, Cho MJ, Lee HH, Nam HG (1993) Generation of expressed sequence tags of random root cDNA clones of *Brassica napus* by single-run partial sequencing. *Plant Physiol* 103: 359–370
- Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 85: 2444–2448
- Roberts MR, Robson F, Foster GD, Draper J, Scott RJ (1991) A *Brassica napus* mRNA expressed specifically in developing microspores. *Plant Mol Biol* 17: 295–299
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*, Ed 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 1.25–1.28
- Sasaki T, Song J, Koga-Ban Y, Matsui E, Fang F, Higo H, Nagasaki H, Hori M, Miya M, Murayama-Kayano E, Takiguchi T, Takasuga A, Niki T, Ishimaru K, Ikeda H, Yamamoto Y, Mukai Y, Ohta I, Miyadera N, Havukkala I, Minobe Y (1994) Toward cataloguing all rice genes: large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library. *Plant J* 6: 615–624
- Shen B, Carneiro N, Torres-Jerez I, Stevenson B, McCreery T, Helentjaris T, Baysdorfer C, Almira E, Ferl RJ, Habben JE, Larkins B (1994) Partial sequencing and mapping of clones from two maize cDNA libraries. *Plant Mol Biol* 26: 1085–1101
- Shen JB, Hsu FC (1992) *Brassica* anther-specific genes: characterization and in situ localization of expression. *Mol Gen Genet* 234: 379–389
- Stanley KK, Herz J, Haymerle H (1988) Constructing expression cDNA libraries using unphosphorylated adaptors. In JM Walker, ed, *Methods in Molecular Biology: New Nucleic Acid Techniques*. Humana Press, Clifton, NJ, pp 319–328
- Uchimiya H, Kidou SI, Shimazaki T, Aotsuka S, Takamatsu S, Nishi R, Hashimoto H, Matsubayashi Y, Kidou N, Umeda M, Kato A (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
- Waterston R, Matin C, Craxton M, Huynh C, Coulson A, Hillier L, Durbin R, Green P, Shownkeen R, Halloran N, Metzstein M, Hawkins T, Wilson, R, Berks M, Du Z, Thomas K, Thierry-Mieg J, Sulston J (1992) A survey of expressed genes in *Caenorhabditis elegans*. *Nature Genet* 1: 114–123