Biotechnology Letters (2005) 27: 1517–1528
DOI 10.1007/s10529-005-1468-8

Analysis of organ-specific, expressed genes in *Oncidium* orchid by subtractive expressed sequence tags library

Jun Tan[1,3], Heng-Long Wang[2] & Kai-Wun Yeh[1,*]
[1]*Institute of Plant Biology, National Taiwan University, 106, Taipei, Taiwan*
[2]*Department of Life Science, National Kaohsiung University, 811, Kaohsiung, Taiwan*
[3]*College of Bioinformation, Chongqing University of Post and Telecom, 400065, Chongqing, China*
**Author for correspondence (Fax: +886-2-23622703; E-mail: ykwbppp@ntu.edu.tw)*

Received 18 May 2005; Revisions requested 13 June 2005; Revisions received 27 July 2005; Accepted 27 July 2005

Key words: developmental stage, expressed sequence tag (EST), *Oncidium* pseudobulb, subtractive cDNA library, transcriptome profile

Abstract

The pseudobulb of *Oncidium* orchid plays a key role in water, carbohydrate, and other nutrition support during floral development, yet a large scale of gene expression analysis involved in the metabolisms have not been evaluated. By subtracting *Rsa*I-digested cDNAs of leaf from those of psuedobulb, an efficient subtractive cDNA library was developed. In total, 1080 subtractive expressed sequence tags (ESTs) were obtained. Analysis revealed approximately 636 unique gene parts, 120 clusters and 516 singles. Of these sequences, 74.8% were annotated on the database of NCBI GenBank. Peroxidase, sodium/dicarboxylate cotransporter, and mannose-binding lectin were highly expressed. Some gene profiles were identified as related to carbohydrate metabolism involved in mannan, pectin, starch and sucrose biosynthesis. A large fraction of the ESTs (35%) were classified into transportation, stress-related, cell cycle, or regulatory functions. Most genes that were differentially expressed are important in early flowering development, carbohydrate metabolism and stress-response physiology. This efficient organ-specific EST library represented an explicit transcriptome profile of *Oncidium* pseudobulb.

Introduction

Orchids (*Orchidaceae*, L.) are the largest family of plants and the number of species may exceed 30 000. They comprise almost 30% of monocotydons or 10% of flowering plants. The genus *Oncidium* is commercially important as cut flowers and as houseplants (Liau *et al.* 2003). Consequently, considerable effort is now being made to improve the economic traits of this ornamental plant. After the method of efficient plant regeneration through somatic embryogenesis from *Oncidium* orchid callus, a routine procedure of transformation with *Agrobacterium tumefaciens* has also been established (Liau *et al.* 2003). A ferredoxin-like protein (pflp) from sweet pepper was used as a novel selection marker for orchid transformation (You *et al.* 2003) and was demonstrated to confer resistance against soft-rot disease in *Oncidium* orchid (Liau *et al.* 2003).

Oncidium orchid has an enlarged bulblike stem at the base of the second upper leaf, termed the pseudobulb. This is important for the storage and support of moisture, mineral nutrition, and carbohydrates during both auxiliary bud and inflorescence development. The carbohydrate pool of pseudobulb varies during inflorescence development (Hew & Ng 1996). From initiation to the end of inflorescence development, galactonio-1,4-lactone, mannan and hexoses, including glucose, fructose and galactose, gradually decreased in the pseudobulb but sucrose and

mannose remained almost constant. Before flowering, there was dramatic accumulation of mannan followed by degradation of starch (Wang *et al.* 2003).

To perform a functional genomic study of *Oncidium* pseudobulb, a subtractive expression sequence tag (EST) library was developed by subtracting *Rsa*I-digested cDNAs of upper leaf tissue from those of pseudobulb. The specific genes, such as the ones coding for peroxidase, sodium/dicarboxylate cotransporter, mannose-binding lectin, senescence, or resistance-associated proteins, have been obtained which cover most of the genes involved in the metabolism of *Oncidium* pseudobulb. All of them relate to the water, nutrition, and energy support of the pseudobulb during/before inflorescence development. Our results are therefore beneficial for the further functional determination of these genes and to identify the physiological linkage with floral time and qualities.

Materials and methods

Plant material and RNA isolation

Pseudobulbs were taken from of *Oncidium* cv. 'Gower Ramsey' grown locally. Total RNAs of the pseudobulb and its upper leaf were extracted following the pine tree method (Chang *et al.* 1993). In brief, fresh tissues were ground into fine powder in liquid N_2, and dissolved in approximately volume of extraction buffer (2% hexadecyltrimethylammonium bromide, 2% polyvinyl pyrrolidone K30, 100 mM Tris/HCl, 25 mM EDTA, 2 M NaCl, 0.05% spermidine, 2% β-mercaptoethanol). After heating for 15 min at 65 °C, the solution was centrifuged 20 000 × *g* for 2 min. The supernatant was treated with chloroform/isoamyl alcohol twice. LiCl (10 mM) was added to the clean supernatant at 1:4 (v/v), and held at 4 °C overnight. Total RNA was harvested after 20 000 × *g* centrifugation at 4 °C.

Construction of subtractive cDNA library and EST sequencing

mRNAs were purified from total RNAs with Oligotex mRNA Kit (Qiagene). After cDNA synthesis, *Rsa*I digestion and ligation of PCR adaptors, subtractive hybridization was performed by using pseudobulb's cDNAs as tester and leaf DNA as drivers. Following the protocol of PCR-select cDNA Subtraction Kit (Clontech), the cDNA mixture was amplified by PCR twice and the products were cloned into pGEM-T easy vector (Promega) with blue and white selection in *E. coli* XL1-Blue. 1248 white clones were Randomly selected and inoculated into 96-well plates with 1 mL 2 × YT broth (tryptone 16 g, yeast extract 10 g, NaCl 5g l^{-1}) and cultured in 37 °C overnight. Plasmids were extracted with Mini-M plasmid DNA Extraction System (Viogene). Using BigDye Terminator v3.1 Cycle Sequencing Kit (ABI) and SP6 as primer, subtractive cDNAs were determined in an automated sequencer ABI 3730.

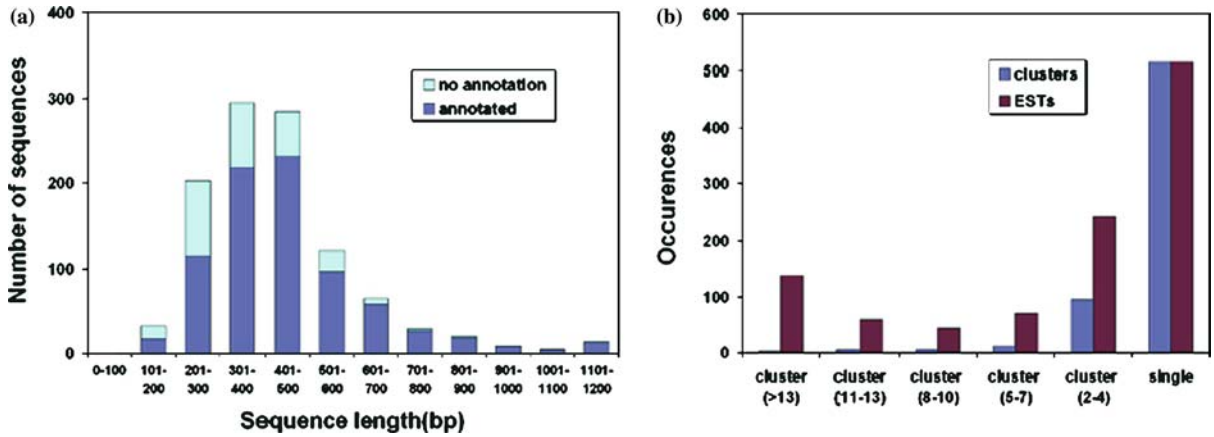

Fig. 1. Alignment and annotation of subtractive ESTs. Panel a: Length distribution of subtractive ESTs. Panel b: Prevalence distribution of subtractive EST cluster size. The numbers in parentheses are the EST copies of clusters.

Table 1. Assembled clusters that contain more than four ESTs.

Gene annotation	Reference organism	GI number	*E*-value	ESTs
Peroxidase (EC 1.11.1.7)	*Gossypium irsutum*	7433087	2.10E−47	92
Sodium/dicarboxylate cotransporter	*Arabidopsis thaliana*	15238130	5.10E−43	24
Peroxidase (EC 1.11.1.7) 2, cationic	*Glycine max*	7433098	6.10E−26	21
T23E23.17	*Arabidopsis thaliana*	9369404	7.00E−11	12
Similar to *Arabidopsis thaliana* T18N14.110	*Oryza sativa*	13486662	7.10E−29	12
Short-chain dehydrogenase/reductase	*Arabidopsis thaliana*	15224306	1.10E−76	11
No hits found				11
Mannose-binding lectin	*Cymbidium hybrid*	2144226	2.10E−51	10
BURP domain protein	*Vigna unguiculata*	7106540	6.00E−12	9
No hits found				9
Peroxidase (EC 1.11.1.7)	*Gossypium irsutum*	7433087	2.00E−17	8
Glycosyl hydrolase family 19 (chitinase)	*Arabidopsis thaliana*	15228911	1.10E−37	8
Lipid transfer protein isoform 4	*Vitis vinifera*	28194086	3.00E−16	7
Unknown	*Arabidopsis thaliana*	21553375	1.10E−22	7
DP-E2F-related protein 1	*Arabidopsis thaliana*	22331664	1.10E−48	7
Unknown protein	*Arabidopsis thaliana*	28393189	7.10E−70	6
Senescence-associated protein	*Pisum sativum*	13359451	3.10E−42	6
Mannose-binding lectin	*Cymbidium hybrid*	2144226	2.10E−21	6
Sucrose synthase	*Oncidium*	22347630	1.10E−95	6
Peroxidase (EC 1.11.1.7) 2, cationic	*Glycine max*	7433098	1.10E−90	5
GDSL-motif lipase/hydrolase protein	*Arabidopsis thaliana*	15228189	3.10E−90	5
Pollen allergen-like protein	*Arabidopsis thaliana*	21593946	6.10E−20	5
No hits found				5
Peroxidase	*Glycine max*	5002234	7.00E−18	5

Sequence annotation and classification

After vector and adaptors in raw sequences were trimmed using cross-match (Green 1993, http://www.phrap.org), the insert fragments were assembled to make contigs and singletons. In order to annotate the clusters and singles, they were aligned with the non-redundant protein sequence database in GenBank (NCBI) by BlastX (Altschul *et al.* 1997) with an *E*-value threshold of 1E−10. According to the GI numbers, presented in annotated subtractive ESTs, they were classified and distributed onto three GO trees. All the EST dataset information is accessible at the website of Institute of Plant Biology, National Taiwan University, *http://plantbio.lifescience.ntu.edu.tw/english/research/estdatabase.htm/*.

Analysis of gene expression profiles by Northern gel blot

Total RNAs (10 μg each) from pseudobulb and its upper leaf were loaded on 1% agrose/formaldehyde gels and transferred onto nylon membranes (Amersham). Based on the annotation of each subtractive EST, sequenced plasmids were selected out and amplified by PCR using both SP6 and T7 primers. The PCR-amplified insert fragment was randomly labeled with α-^{32}P-dCTP (R*edi*prime II Kit, Amersham Biosciences). ^{32}P-labeled probes were hybridized to membranes and membranes were washed following the standard protocol. The membrane was exposed to fluorescent plate for 12–72 h (Typhoon 9400, Amersham Biosciences).

Results and discussion

Overall distribution of subtractive ESTs from pseudobulb of Oncidium Gower Ramsey

In total, 1248 clones were sequenced. 1088 clones were readable and had an average length of 1031 bp. Among them, 1080 inserts with a length more than 100 bp were accepted as subtractive

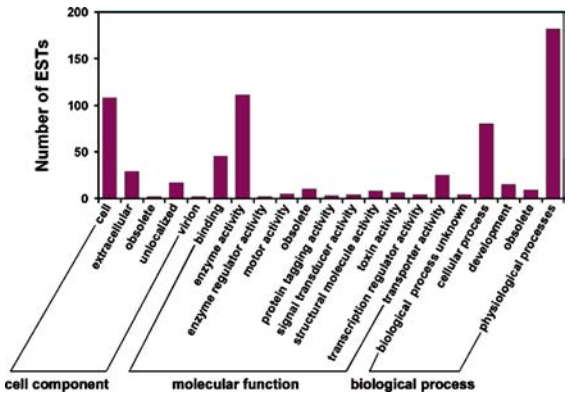

Fig. 2. Classification of subtractive ESTs with three GO trees.

ESTs for further study. Their length was 430 bp on average and in the range of 100–1200 bp, mainly from 200 to 800 bp (Figure 1a).

After Phrap, 1080 subtractive ESTs were assembled into 149 contigs and 543 singletons. Furthermore, they were aligned into 51 clusters and 585 singles by BLASTN. There were 69 contigs in the 585 singles, so 516 real singles had only one subtractive EST, 48.3% of total ESTs. The number of ESTs in clusters, also called cluster size, reflected the abundance of mRNAs. Including singles with one EST, the distribution of cluster size was indicated in Figure 1b, and clusters were divided into six classes. There were 95 clusters containing 2–4 ESTs, comprising 22.3% of the total accepted ESTs (241 of 1080 subtractive ESTs) and 79.1% of total clusters (95 of 120 clusters).

There were 12 clusters containing 5–7 ESTs, comprising 6.5% of the total accepted ESTs (70 of 1080 subtractive ESTs) and 10.0% of total clusters (12 of 120 clusters). These clusters represented lipid transfer protein, DP-E2F-related protein, senescence-associated protein, mannose-binding lectin, sucrose synthase, peroxidase, GDSL-motif lipase/hydrolase, pollen allergen, and two unknown genes. One cluster had no significant similarity to any protein sequence in the GenBank nr database.

There were five clusters containing 8–10 ESTs, comprising 4.1% of the total accepted ESTs (44 of 1080 subtractive ESTs) and 4.1% of total clusters (5 of 120 clusters). These clusters represented mannose-binding lectin, BURP domain protein, peroxidase, chitinase family 19 genes. One cluster had no significant similarity to any protein sequence in the GenBank nr database. There were another five clusters containing 11–13 ESTs comprising 5.5% of the total accepted ESTs (59 of 1080 subtractive ESTs). These clusters represented short-chain dehydrogenase/reductase and two unkown genes (*Arabidopsis* T23E23.17, T18N14.110). Two clusters had no significant similarity to any protein sequence in the GenBank nr database.

Three clusters were regarded as the most abundant transcripts. They were only 2.5% of total clusters (3 of 120 clusters), but comprised 12.7% of total subtractive ESTs (137 of 1080 subtractive ESTs). Two of them were peroxidase genes and sodium/dicarboxylate cotransporter genes. More details and the possible functions of these genes are shown in Table 1. Comparatively, mannose-binding lectin genes were expressed predominately in normal pseudobulb (data not shown).

Annotation and classification of the subtractive ESTs

Six hundred and thirty six clusters and singles including 1080 subtractive ESTs were submitted

Table 2. Selected examples of ESTs for genes with known or putative functions related to inflorescence.

Gene annotation	Reference organism	GI number	*E*-value	ESTs
Abnormal inflorescence meristem 1 (AIM1)	*Arabidopsis thaliana*	15235527	2.00E−36	1
Chalcone sythase C2	*Zea mays*	116380	6.00E−15	1
Chalcone-flavanone isomerase	*Arabidopsis thaliana*	18414838	3.00E−29	3
Cytochrom P450	*Arabidopsis thaliana*	21595357	2.00E−34	1
Cytochrome P450 71D2	*Catharanthus roseus*	28261339	2.00E−30	1
Dihydroflavonol reductase	*Arabidopsis thaliana*	18390863	1.00E−16	1
MADS box protein (DOMADS2)	*Dendrobium grex*	6467974	5.00E−24	1
Shaggy-like kinase	*Ricinus communis*	1877397	8.00E−99	1

to BLASTX for homologous searching with the nr. database in GenBank. Four hundred and thirty nine clusters and singles composed of 808 ESTs (484 contigs and singletons) were identified. Three hundred and nine of them comprising 614 ESTs were annotated with a gene name and could be analyzed further. From Figure 1a, the length and sequence quality of subtractive ESTs had a linear relationship with the number of ESTs that could be annotated. The higher the sequence quality, the greater the chance of being annotated, and vice versa.

Using the GO classification system, only 155 ESTs (25.2% of 614 annotated ESTs) were matched, protein sequences with special GI numbers which had been previously classified on three different GO trees (cell component, molecular function, and cellular process) (Figure 2). This

Table 3. Selected examples of ESTs for genes related to carbohydrate metabolisms.

Gene annotation	Reference organism	GI number	*E*-value	ESTs
ADP-glucose pyrophosphorylase	*Arabidopsis thaliana*	30699056	4.00E−66	2
Aldose 1-epimerase	*Arabidopsis thaliana*	15242099	3.00E−65	2
α 1,4-glucan phosphorylase L isozyme	*Oryza sativa*	13195430	1.00E−43	1
α-galactosidase	*Arabidopsis thaliana*	11264291	3.00E−33	1
β-1,3 glucanase	*Oryza sativa*	20161490	7.00E−39	1
β-fructofuranosidase 1	*Zea mays*	1352468	2.00E−31	1
β-galactosidase	*Oryza sativa*	20514290	7.00E−63	1
β-galactosidase	*Oryza sativa*	18461259	2.00E−49	1
β-mannosidase	*Lycopersicon esculentum*	17226270	1.00E−103	1
Cinnamyl alcohol dehydrogenase	*Populus balsamifera*	9998899	3.00E−38	1
dTDP-glucose 4−6-dehydratase	*Arabidopsis thaliana*	21594350	1.00E−16	1
Epimerase/dehydratase	*Oryza sativa*	20042976	1.00E−56	3
Glucosyltransferase	*Arabidopsis thaliana*	25408401	2.00E−26	1
Glycogenin glucosyltransferase (EC 2.4.1.186)	*Oryza sativa*	5441877	5.00E−84	1
Glycogenin glucosyltransferase (EC 2.4.1.186)	*Oryza sativa*	5441877	7.00E−93	1
Glycosyl hydrolase 1	*Arabidopsis thaliana*	15220627	6.00E−36	1
Granule-bound starch synthase	*Pisum sativum*	15626365	4.00E−86	1
Mannose-6-phosphate isomerase	*Arabidopsis thaliana*	15232927	8.00E−50	1
Mannose-6-phosphate isomerase	*Oryza sativa*	11275529	6.00E−45	1
Mannosyltransferase	*Arabidopsis thaliana*	22326970	4.00E−15	1
NAD-dependent epimerase/dehydratase	*Arabidopsis thaliana*	15231926	2.00E−63	1
Nucleoside-diphosphate-sugar pyrophosphorylase	*Oryza sativa*	29893646	6.00E−68	4
N-acetylglucosamine-phosphate mutase	*Arabidopsis thaliana*	30686654	8.00E−38	1
Pectate lyase	*Arabidopsis thaliana*	10177179	4.00E−23	1
Pectin esterase	*Oryza sativa*	20161185	1.00E−12	2
Pectinesterase 1	*Lycopersicon esculentum*	6174913	1.00E−13	4
Phosphoglucomutase, cytoplasmic	*Solanum tuberosum*	12585316	2.00E−84	1
Phosphoglucose isomerase	*Dioscorea septemloba*	2351056	2.00E−80	1
Phosphomannomutase	*Arabidopsis thaliana*	15225896	2.00E−16	2
Polygalacturonase	*Pisum sativum*	13958032	3.00E−57	3
Polygalacturonase	*Arabidopsis thaliana*	18412253	2.00E−22	1
Ripening-related protein	*Vitis vinifera*	7406669	4.00E−79	3
Starch phosphorylase	*Ipomoea batatas*	12658431	2.00E−31	1
Sucrose synthase	*Oncidium*	22347630	1.00E−96	6
Sucrose synthase	*Oncidium*	22347630	2.00E−27	2
Sucrose synthase	*Oncidium*	22347630	1.00E−38	2
Sucrose synthase	*Oncidium*	22347630	4.00E−30	1
Triosephosphate isomerase, cytosolic (TIM)	*Petunia x hybrida*	1351279	4.00E−27	1

represented three different points of view for classified genes. There were 104 classified ESTs belonging to cellular components, 29 ESTs belonging to extracellular components, and 19 ESTs unlocalized components. According to the molecular function tree, the majority (85 ESTs) were classified as enzymes; 44 ESTs as binding proteins, and 20 as transporters. One hundred and thirty ESTs belonged to physiological processes, and 42 were involved in cellular processes.

Northern blot analysis for the subtractive efficiency of the subtractive ESTs dataset

From the ESTs clones, 16 sequences, including eight abundant transcripts, were selected as

Table 4. Selected examples of ESTs for genes related to transportation.

Gene annotation	Reference organism	GI number	*E*-value	ESTs
Acyl-CoA-binding protein	*Panax ginseng*	19352190	3.00E−19	2
ADP-ribosylation factor	*Glycine max*	4324967	2.00E−52	2
ADP-ribosylation factor	*Oryza sativa*	18844784	2.00E−25	1
C2 domain-containing protein	*Arabidopsis thaliana*	15239959	2.00E−36	1
C2 domain-containing protein	*Arabidopsis thaliana*	15223764	5.00E−15	1
C2 domain-containing protein	*Arabidopsis thaliana*	15217968	5.00E−37	1
dynamin like protein 2a	*Arabidopsis thaliana*	19032337	2.00E−44	1
F-actin capping protein, α subunit	*Oryza sativa*	23617186	2.00E−12	1
γ-adaptin 1	*Oryza sativa*	19386749	1.00E−30	1
GDSL-like lipase/acylhydrolase	*Oryza sativa*	29837765	1.00E−26	4
GDSL-like lipase/acylhydrolase	*Oryza sativa*	29837765	9.00E−34	2
GDSL-motif lipase/hydrolase	*Arabidopsis thaliana*	15228189	3.00E−91	5
GDSL-motif lipase/hydrolase	*Arabidopsis thaliana*	21593518	2.00E−25	3
GDSL-motif lipase/hydrolase	*Arabidopsis thaliana*	15221260	6.00E−54	2
GDSL-motif lipase/hydrolase	*Arabidopsis thaliana*	18416824	2.00E−25	1
GDSL-motif lipase/hydrolase	*Arabidopsis thaliana*	15224201	4.00E−12	1
golgi-localized protein (GRIP)	*Oryza sativa*	22093862	2.00E−18	1
high mobility group protein 2	*Arabidopsis thaliana*	15231065	7.00E−31	1
kinesin	*Daucus carota*	15186760	3.00E−30	2
kinesin-related protein	*Arabidopsis thaliana*	22327641	3.00E−54	3
lipid transfer protein isoform 4	*Vitis vinifera*	28194086	3.00E−16	7
membrane bound O-acyl transferase (MBOAT)	*Arabidopsis thaliana*	22329514	9.00E−24	1
membrane bound O-acyl transferase (MBOAT)	*Arabidopsis thaliana*	22329514	1.00E−75	1
mitochondrial carrier protein	*Arabidopsis thaliana*	15240756	5.00E−39	1
myosin heavy chain	*Arabidopsis thaliana*	18402909	5.00E−17	1
permease	*Oryza sativa*	27545049	2.00E−17	3
peroxisomal targeting signal type 1 receptor	*Arabidopsis thaliana*	15241175	3.00E−15	1
PEX14 protein	*Arabidopsis thaliana*	30697742	3.00E−15	1
plasma membrane intrinsic protein	*Oryza sativa*	22831004	2.00E−44	4
Rer1A protein (AtRer1A)	*Oryza sativa*	10945247	2.00E−37	2
Sec31p	*Oryza sativa*	22831279	4.00E−63	1
Secretory carrier membrane protein	*Arabidopsis thaliana*	15222550	1.00E−30	1
Signal peptidase	*Arabidopsis thaliana*	15240934	7.00E−56	1
Sodium/dicarboxylate cotransporter	*Arabidopsis thaliana*	15238130	5.00E−44	24
Sodium/dicarboxylate cotransporter	*Arabidopsis thaliana*	15238130	2.00E−39	3
Sodium-dicarboxylate cotransporter	*Arabidopsis thaliana*	21536650	2.00E−35	1
Vesicle transport v-SNARE protein	*Oryza sativa*	19571103	8.00E−55	1
Villin 1 (VLN1)	*Arabidopsis thaliana*	26451417	2.00E−22	1

probes for Northern blot analysis (Figure 3). Peroxidase, sodium/dicarboxylate cotransporter, BURP domain protein (dehydration-responsive protein RD22), mannose-6-phosphate isomerase, proline-rich-like protease inhibitor, Na^+/H^+ antiporter isoform 2, invertase, and pectate lyase were highly expressed in pseudobulb at the early initiation of inflorescence, but they exhibited almost no expression in its upper leaf. The expression patterns of mannose-binding lectin, sucrose synthase, GDP-mannose pyrophosphorylase, and granule-bound starch synthase were similar with peroxidase genes, except they had very low expressions in leaf. Glycine-rich RNA binding protein and leucine-rich receptor-related protein kinase genes were expressed highly in

Table 5. Selected examples of ESTs for known or putative stress-related genes.

Gene annotation	Reference organism	GI number	*E*-value	ESTs
Acid phosphatase	*Arabidopsis thaliana*	22330531	1.00E−35	1
AP2 domain transcription factor	*Arabidopsis thaliana*	21593696	2.00E−18	1
AP2 domain transcription factor	*Arabidopsis thaliana*	21593696	5.00E−58	1
β-N-acetylhexosaminidase	*Arabidopsis thaliana*	21537026	4.00E−50	1
Biostress-resistance-related protein	*Triticum aestivum*	29409364	1.00E−61	1
bZIP DNA-binding protein	*Capsicum chinense*	4457221	3.00E−27	1
Chloroplastic light-induced, drought-induced stress protein	*Solanum tuberosum*	22261807	4.00E−40	1
Choline monooxygenase	*Suaeda liaotungensis*	21217447	6.00E−19	1
Dehydration-induced protein	*Arabidopsis thaliana*	18411430	2.00E−68	1
DHHC-type zinc finger domain-containing protein	*Arabidopsis thaliana*	18409331	2.00E−34	1
Disease resistance protein	*Arabidopsis thaliana*	15232373	9.00E−26	3
Disease resistance protein (NBS-LRR class)	*Arabidopsis thaliana*	15231860	4.00E−18	1
Extensin	*Populus nigra*	7484770	6.00E−42	4
Farnesyltranstransferase	*Oryza sativa*	20160508	7.00E−11	1
Glyceraldehyde 3-phosphate dehydrogenase, cytosolic	*Magnolia quinquepeta*	120669	2.00E−99	2
Glycosyl hydrolase family 19 (chitinase)	*Arabidopsis thaliana*	15228911	1.00E−38	8
Heat shock protein	*Arabidopsis thaliana*	15225377	3.00E−20	1
Heat shock protein	*Arabidopsis thaliana*	15225377	1.00E−40	1
Heat shock protein cognate 70	*Oryza sativa*	29124135	2.00E−42	1
Heat shock protein hsc70-3 (hsc70.3)	*Arabidopsis thaliana*	15232682	7.00E−29	1
Late embryogenesis abundant protein	*Arabidopsis thaliana*	15224810	2.00E−16	1
Leucine rich repeat protein	*Arabidopsis thaliana*	30686169	1.00E−37	1
Major intrinsic protein (MIP)	*Arabidopsis thaliana*	15236485	1.00E−81	2
Na + /H + antiporter 2	*Lycopersicon esculentum*	15982206	4.00E−17	1
Nodulin	*Oryza sativa*	11072005	9.00E−31	1
PDR-like ABC transporter	*Oryza sativa*	27368827	4.00E−35	1
Peroxidase	*Glycine max*	5002234	7.00E−18	5
Peroxidase (EC 1.11.1.7)	*Gossypium irsutum*	7433087	2.00E−48	92
Peroxidase (EC 1.11.1.7)	*Gossypium irsutum*	7433087	2.00E−17	8
Peroxidase (EC 1.11.1.7) 2, cationic	*Glycine max*	7433098	6.00E−27	21
Peroxidase (EC 1.11.1.7) 2, cationic	*Glycine max*	7433098	1.00E−91	5
Phosphoethanolamine methyltransferase	*Oryza sativa*	22535531	8.00E−13	1
Plastid-lipid associated protein PAP/fibrillin	*Arabidopsis thaliana*	18403751	2.00E−35	1
Proline rich protein 3	*Cicer arietinum*	21615411	5.00E−75	1
Proline-rich protein APG isolog	*Cicer arietinum*	10638955	4.00E−16	1
Proline-rich-like protein	*Asparagus officinalis*	1531756	2.00E−29	1
Senescence-associated protein	*Pisum sativum*	13359451	3.00E−43	6
Senescence-associated protein	*Arabidopsis thaliana*	18398417	2.00E−20	1
Wound-induced protein	*Arabidopsis thaliana*	15234987	2.00E−15	3

Table 6. Selected examples of ESTs for genes related to cell cycle.

Gene annotation	Reference organism	GI number	*E*-value	ESTs
26S proteasome non-ATPase, regulatory subunit 6	*Oryza sativa*	20978545	8.00E−79	1
3-Hydroxy-3-methylglutaryl-coenzyme A reductase 3 (HMG3.3)	*Solanum tuberosum*	11133016	5.00E−32	1
AUX1-like permease	*Arabidopsis thaliana*	5881784	2.00E−24	1
Auxin efflux carrier protein	*Arabidopsis thaliana*	15239215	3.00E−50	1
Biotin carboxyl carrier protein subunit	*Glycine max*	12006165	2.00E−29	1
Cyclic nucleotide-regulated ion channel (CNGC9)	*Arabidopsis thaliana*	15234769	8.00E−34	1
Cysteine proteinase AALP	*Arabidopsis thaliana*	23397070	6.00E−30	1
Cysteine proteinase mir3 (EC 3.4.22.-)	*Zea mays*	7435806	5.00E−47	1
DP-E2F-related protein 1	*Arabidopsis thaliana*	22331664	1.00E−49	7
Histone deacetylase 2 isoform b	*Zea mays*	7716948	6.00E−19	1
Homeobox 20	*Nicotiana tabacum*	4589882	1.00E−49	1
Homeobox protein knotted-1 2 (KNAP2)	*Malus x domestica*	6016217	8.00E−46	3
Homeobox-leucine zipper protein ATHB-13	*Arabidopsis thaliana*	15222452	9.00E−38	1
Homeotic protein knotted-1 (TKN1)	*Lycopersicon esculentum*	3023974	9.00E−16	1
Nucleolysin	*Oryza sativa*	4680340	1.00E−15	1

pseudobulb and slightly in leaf at the same time. Only AP2 domain transcription factor gene displayed the opposite pattern: the expression level in pseudobulb was high, but that in leaf seemed higher. In general, most ESTs had more specific expression profiles. Significantly, the results indicated the subtractive ESTs dataset did show pseudobulb-specific expression at the early initiation of the inflorescence. Therefore, these Northern blot data demonstrated that the EST subtraction was very precise and reliable.

Subtractive ESTs relevant to inflorescence, carbohydrate metabolisms, transportation, stress, cell cycle, and regulation

According to the annotations and references about their functions, the subtractive ESTs were gathered and analyzed manually. One cluster and 7 singles (10 ESTs in total) were related to specific flower genes (Table 2). Abnormal inflorescence meristem 1 (AIM1) could affect inflorescence and floral development in *Arabidopsis* (Richmond & Bleecker 1999). MADS box protein genes were expressed in different organs and mainly during floral development. DOMADS2 was expressed throughout the process of floral transition and development (Yu & Goh 2000). Shaggy-like kinase was flower-specific and responsible for osmotic changes and darkness (Charrier *et al*. 2002). The others were related to four kinds of genes within the synthetic pathway of flower pigment. Chalcone synthase and chalcone–flavanone isomerase were at upper stream of the pathway. More chalcone-flavanone isomerase gene seems expressed in pseudobulb at the initiation of inflorescence. Two ESTs annotated with cytochrome P450 genes also could be annotated as flavanone 3′ hydroxylase or flavanone 3′ 5′ hydroxylase. They catalyze dihydrokeampherol into dihydroquercetin or dihydromyricetin. Incorporated with UDP-glucose transferase, dihydroflavonol reductase was a downstream gene of the pathway and took part in the synthesis of anthocyanins.

In total we found 61 subtractive ESTs annotated with genes involved in the metabolism of saccharides, including mannose, glucose, fructose, galactose, sucrose, starch, pectin, and cellulose (Table 3). Based on this information, we can draw a draft of carbohydrate pathways to explain what happened in pseudobulb at the initiation of inflorescence development (data not shown).

Ninety one ESTs were thought to have probable relationships with transportation (Table 4). Most of the proteins the genes encoded were localized in the kinds of membranes or on cell matrixes that help material transportation. Sodium/dicarboxylate cotransporter and GDSL-motif lipase/hydrolase were most abundant. Sodium/dicarboxylate cotransporter was a single copy gene in *Arabidopsis* and localized on

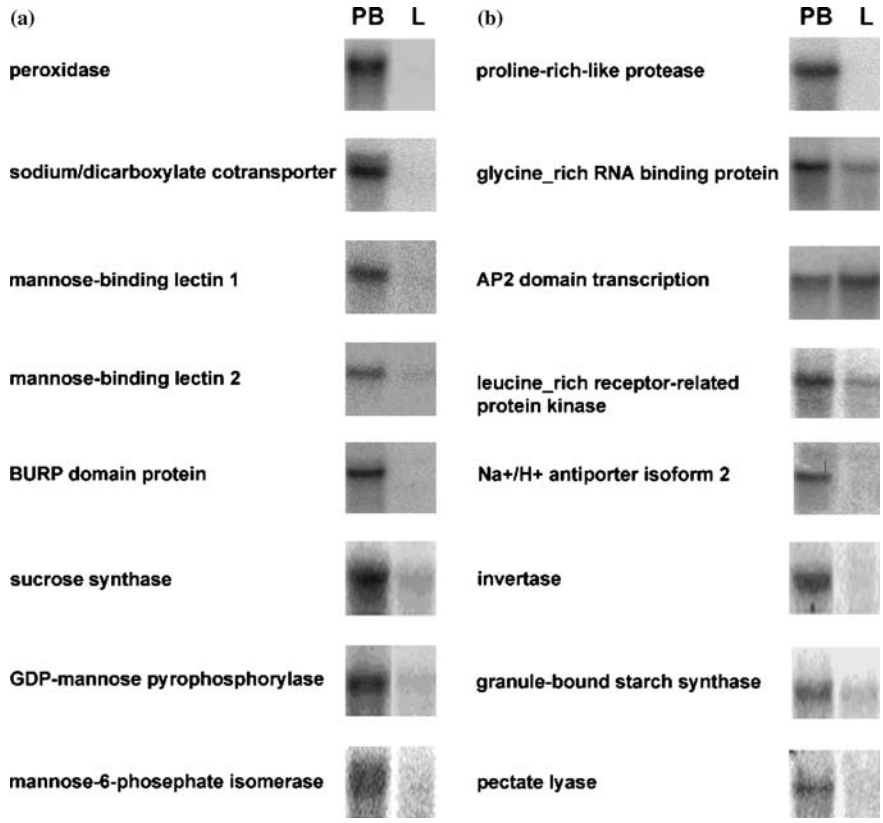

Fig. 3. Northern blot assay for gene expression pattern of 16 subtractive ESTs. Panel a: 8 subtractive ESTs with redundancy. RNA membranes were probed with annotated subtractive EST clones: peroxidase [*Gossypium irsutum*], sodium/dicarboxylate co-transporter [*Arabidopsis thaliana*], mannose-binding lectin [*Cymbidium hybrid*] 1, mannose-binding lectin [*Cymbidium hybrid*] 2, BURP domain protein [*Oryza sativa*], sucrose synthase [*Oncidium*], GDP-mannose pyrophosphorylase [*Oryza sativa*], mannose-6-phosephate isomerase [*Arabidopsis thaliana*]. Panel b: 8 subtractive ESTs without redundancy. RNA membranes were probed with annotated clones: proline-rich-like protease inhibitor [*Asparagus officinalis*], glycine-rich RNA binding protein [*Oryza sativa*], AP2 domain transcription factor [*Arabidopsis thaliana*], leucine-rich receptor-related protein kinase [*Arabidopsis thaliana*], Na + /H + antiporter isoform 2 [*Lycopersicon esculentum*], invertase [*Zea Mays*], granule-bound starch synthase [*Pisum sativum*], pectate lyase [*Arabidopsis thaliana*]. PB = pseudobulb; L = leaf.

vacuole membrane to transfer malate into vacuole (Emmerlich *et al*. 2003). GDSL-motif lipase/hydrolase was lipolytic enzyme, maybe related to a secretion mechanism (Wilhelm *et al*. 1999). ADP-ribosylation factor plays a critical role in intracellular trafficking and maintenance of endoplasmic reticulum morphology in *Arabidopsis* (Lee *et al*. 2002). γ-adaptin is involved in Golgi-endosome traffic, including the recruitment of accessory proteins, γ-synergin and Rabaptin-5 (Nogi *et al*. 2002). Golgi-localized protein (GRIP) could maintain normal Golgi morphology and function (Ungar *et al*. 2002). C2 domain-containing protein occurs in a large variety of membrane trafficking and signal transduction protein. Many of their biological roles have not been identified (Ochoa *et al*. 2002). Including transporters in membranes, others interacting with cell scaffolds were also expressed, such as dynamin-like protein, F-actin capping protein, kinesin, and villin.

We identified 186 ESTs as possible stress-related genes (Table 5). Amazingly, 131 ESTs were peroxidase genes. In *Arabidopsis*, they are a large gene family composed of 78 members with different expression profiles in different organs (Tognolli *et al*. 2002). Based on EST alignment, expressed peroxidase genes in pseudobulb belong to a large family too. The others were genes induced by different biotic and abiotic stresses,

Table 7. Selected examples of ESTs for known or putative regulatory functions.

Gene annotation	Reference organism	GI number	*E*-value	ESTs
Adapter protein SPIKE1	*Oryza sativa*	24899400	4.00E−56	1
Adenine phosphoribosyltransferase form 2	*Oryza sativa*	29826070	8.00E−75	1
Amidase	*Arabidopsis thaliana*	8163875	7.00E−31	1
Amidase	*Oryza sativa*	18542894	1.00E−12	1
BURP domain protein	*Vigna unguiculata*	7106540	6.00E−12	9
c-myc binding protein	*Arabidopsis thaliana*	22325671	2.00E−12	1
Cupin domain-containing protein	*Arabidopsis thaliana*	15226403	3.00E−29	2
DEAD/DEAH box helicase	*Arabidopsis thaliana*	15222526	5.00E−40	1
DEAD/DEAH box helicase	*Arabidopsis thaliana*	15219185	1.00E−26	1
DnaJ protein	*Salix gilgiana*	11277163	1.00E−106	2
DnaJ protein homolog 2	*Allium porrum*	1169382	4.00E−26	1
DNAJ-like protein	*Oryza sativa*	29367357	8.00E−33	1
Elongation factor 1-α	*Elaeis oleifera*	18419676	8.00E−40	1
GAMYB-binding protein	*Hordeum vulgare*	27948448	7.00E−51	1
GF14 protein	*Fritillaria agrestis*	2921512	2.00E−66	1
Glycine-rich RNA-binding protein	*Arabidopsis thaliana*	21553602	7.00E−21	1
HD-Zip transcription factor Athb-14	*Arabidopsis thaliana*	15226808	1.00E−96	1
helicase	*Arabidopsis thaliana*	18395518	4.00E−30	1
Homeobox-leucine zipper protein ATHB-13	*Arabidopsis thaliana*	15222452	9.00E−38	1
MuDR mudrA-like protein	*Oryza sativa*	5441874	4.00E−31	1
Phosphoprotein phosphatase (EC 3.1.3.16)	*Arabidopsis thaliana*	25513447	2.00E−94	1
Probable protein disulfide-isomerase	*Nicotiana tabacum*	7489183	1.00E−88	1
Ras-related protein Rab11C	*Nicotiana tabacum*	3024503	5.00E−31	1
Receptor-like kinase RHG4	*Glycine max*	21239384	2.00E−17	1
Receptor-like protein kinase	*Arabidopsis thaliana*	7487253	6.00E−41	1
Receptor-like protein kinase (EC 2.7.1.-)	*Oryza sativa*	7434420	3.00E−18	1
Receptor-related protein kinase	*Arabidopsis thaliana*	15240720	6.00E−45	1
RNA recognition motif (RRM)-containing protein	*Arabidopsis thaliana*	22328805	8.00E−18	1
RNA-binding protein	*Oryza sativa*	18087662	2.00E−37	2
RNA-binding protein	*Mesembryanthemum crystallinum*	1076251	4.00E−26	1
Serine/threonine kinase	*Arabidopsis thaliana*	25387051	2.00E−51	1
Serine/threonine protein kinase	*Nicotiana tabacum*	3811293	3.00E−19	1
Serine/threonine protein kinase (EC 2.7.1.-)	*Avena sativa*	7489361	5.00E−37	1
Serine/threonine-specific protein kinase	*Arabidopsis thaliana*	25751318	2.00E−18	1
SNF2 domain/helicase domain-containing protein	*Arabidopsis thaliana*	15226870	4.00E−42	1
Sphingosine kinase	*Oryza sativa*	13786462	3.00E−71	1
Transcription factor LIM	*Nicotiana tabacum*	18565124	5.00E−46	2
Transcription factor X1	*Oryza sativa*	6650526	2.00E−28	1
Transducin / WD-40 repeat protein	*Arabidopsis thaliana*	30682603	3.00E−21	1
Transfactor	*Arabidopsis thaliana*	6223653	2.00E−34	1
Translational activator	*Arabidopsis thaliana*	25404492	2.00E−18	1
Translational activator	*Arabidopsis thaliana*	15217742	5.00E−42	1
WD-40 repeat protein	*Arabidopsis thaliana*	30685408	5.00E−46	1
Zinc finger (C3HC4-type RING finger) protein	*Arabidopsis thaliana*	15233298	9.00E−28	1
Zinc finger protein	*Pisum sativum*	11288368	3.00E−77	1
Zinc finger protein 5 (ZFP5)	*Arabidopsis thaliana*	21592423	5.00E−12	1
Zinc-finger protein Lsd1	*Arabidopsis thaliana*	30685085	1.00E−22	1

including wounding, drought, and pathogens. Among them, AP2 domain transcription factor could be induced by cold, dehydration, and ABA stress, and was involved in regulation of low-temperature responsive genes in barley (Xue 2003).

Twenty three ESTs were found to have functions in the cell cycle (Table 6). That is to say, the cells in pseudobulb kept an active growing state and differentiating actively in this stage. DP-E2F-related protein 1 and homeobox genes were most abundant among this group. The E2F/DP protein family controls cell cycle progression by acting predominantly as an activator or repressor of transcription. *Arabidopsis* had more than 180 potential E2F target genes with various functions: cell cycle, transcription, stress and defense, or signaling (Ramirez-Parra *et al.* 2003). Homeobox 20 had a common motif and took part in xylem cell differentiation (Hertzberg & Olsson 1998). Homeobox-leucine zipper protein ATHB-13 was a transcription factor. It could specify the cell fate and body plan in early embryogenesis.

Fifty nine ESTs were annotated with known or putative regulatory functions (Table 7). It seemed that the regulation of the genes involved in the active material and energy metabolism in the pseudobulb was very complex.

Thus far, a large-scale analysis of gene expression related to physiological status of the *Oncidium* pseudobulb, particularly during the early floral stage, has not yet been reported. Therefore, the expressed gene catalogue presented here will provide the basal information to investigate the molecular genetics basis of the *Oncidium* pseudobulb's early flowering stage by transcriptome profiling. In this small-scale subtractive EST, a conclusive picture of the cellular processes of stress-response (Table 5), carbohydrate metabolism (Table 3), and transportation (Table 4) was obtained. Also, the RNA gel-blot expression data showed some evidence that this EST-set is indeed enriched with such genes (Figure 3), indicating the high efficiency of the cDNA subtraction strategy.

In summary, the subtractive EST approach is an efficient tool to overview gene expression profiles in the metabolically active tissue of the *Oncidium* pseudobulb. The EST data provides us with insight into a wide range of genes. These genes represent the physiological status in the pseudobulb during early inflorescence development. The abundant genes, e.g. peroxidase and sodium/dicarboxylate cotransporter, shown in the profile revealed some especially unexpected facts. These will make it possible to exploit flowering-related mechanisms for the benefit of mankind.

Acknowledgements

This work was financially supported by National Science Council, Taiwan, ROC under the Grant NSC 91-2317-B-002-041 to Professor Kai-Wun Yeh.

References

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

Chang S, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* **11**: 113–116.

Charrier B, Champion A, Henry Y, Kreis M (2002) Expression profiling of the whole *Arabidopsis* shaggy-like kinase multigene family by real-time reverse transcriptase-polymerase chain reaction. *Plant Physiol.* **130**: 577–590.

Emmerlich V, Linka N, Reinhold T, Hurth MA, Traub M, Martinoia E, Neuhaus HE (2003) The plant homolog to the human sodium/dicarboxylic cotransporter is the vacuolar malate carrier. *Proc. Natl. Acad Sci. USA* **100**: 11122–11126.

Hertzberg M, Olsson O (1998) Molecular characterisation of a novel plant homeobox gene expressed in the maturing xylem zone of *Populus tremula* × *tremuloides*. *Plant J.* **16**: 285–295.

Hew CS, Ng CKY (1996) Changes in mineral and carbohydrate content in psuedobulbs of the C_3 epiphytic orchid hybrid *Oncidium* Goldiana at different growth stages. *Lindleyana* **11**: 125–134.

Lee MH, Min MK, Lee YJ, Jin JB, Shin DH, Kim DH, Lee KH, Hwang I (2002) ADP-ribosylation factor 1 of *Arabidopsis* plays a critical role in intracellular trafficking and maintenance of endoplasmic reticulum morphology in *Arabidopsis*. *Plant Physiol.* **129**: 1507–1520.

Liau CH, Lu JC, Prasad V, Hsiao HH, You SJ, Lee JT, Yang NS, Huang HE, Feng TY, Chen WH, Chan MT (2003) The sweet pepper ferredoxin-like protein (pflp) conferred resistance against soft rot disease in *Oncidium* orchid. *Transgenic Res.* **12**: 329–336.

Nogi T, Shiba Y, Kawasaki M, Shiba T, Matsugaki N, Igarashi N, Suzuki M, Kato R, Takatsu H, Nakayama K, Wakatsuki S (2002) Structural basis for the accessory protein recruitment by the gamma-adaptin ear domain. *Nature Struct. Biol.* **9**: 527–531.

Ochoa WF, Corbalan-Garcia S, Eritja R, Rodriguez-Alfaro JA, Gomez-Fernandez JC, Fita I, Verdaguer N (2002) Additional binding sites for anionic phospholipids and calcium

ions in the crystal structures of complexes of the C2 domain of protein kinase calpha. *J. Mol. Biol.* **320**: 277–291.

Ramirez-Parra E, Frundt C, Gutierrez C (2003) A genome-wide identification of E2F-regulated genes in *Arabidopsis*. *Plant J.* **33**: 801–811.

Richmond TA, Bleecker AB (1999) A defect in beta-oxidation causes abnormal inflorescence development in *Arabidopsis*. *Plant Cell* **11**: 1911–1924.

Tognolli M, Penel C, Greppin H, Simon P (2002) Analysis and expression of the class III peroxidase large gene family in *Arabidopsis thaliana*. *Gene* **288**: 129–138.

Ungar D, Oka T, Brittle EE, Vasile E, Lupashin VV, Chatterton JE, Heuser JE, Krieger M, Waters MG (2002) Characterization of a mammalian Golgi-localized protein complex, COG, that is required for normal Golgi morphology and function. *J. Cell Biol.* **157**: 405–415.

Wang HL, Chung JD, Yeh KW (2003) Changes of carbohydrate and free amino acid pools in current pseudobulbs of *Oncidium* 'Gower Ramsey' during inflorescence development. *J. Agric. Assoc. China* **4**: 476–488.

Wilhelm S, Tommassen J, Jaeger KE (1999) A novel lipolytic enzyme located in the outer membrane of *Pseudomonas aeruginosa*. *J. Bacteriol.* **181**: 6977–6986.

Xue GP (2003) The DNA-binding activity of an AP2 transcriptional activator HvCBF2 involved in regulation of low-temperature responsive genes in barley is modulated by temperature. *Plant J.* **33**: 373–383.

You SJ, Liau CH, Huang HE, Feng TY, Prasad V, Hsiao HH, Lu JC, Chan MT (2003) Sweet pepper ferredoxin-like protein (pflp) gene as a novel selection marker for orchid transformation. *Planta* **217**: 60–65.

Yu H, Goh CJ (2000) Identification and characterization of three orchid MADS-box genes of the AP1/AGL9 subfamily during floral transition. *Plant Physiol.* **123**: 1325–1336.