Plant Cell Rep (2004) 22:839–847
DOI 10.1007/s00299-004-0764-4

GENETICS AND GENOMICS

S. K. Cho · S. H. Ok · J. U. Jeung · K. S. Shim · K. W. Jung · M. K. You · K. H. Kang · Y. S. Chung · H. C. Choi · H. P. Moon · J. S. Shin

Comparative analysis of 5,211 leaf ESTs of wild rice (*Oryza minuta*)

Received: 12 August 2003 / Revised: 7 January 2004 / Accepted: 8 January 2004 / Published online: 17 February 2004

Abstract The expressed sequence tags (ESTs) presented in this report are the first transcriptomes of wild rice. A cDNA library was constructed from 4-week-old leaf samples of greenhouse-grown *Oryza minuta*. The 5,211 cDNA clones of *O. minuta* represent 3,401 unique sequences, consisting of 2,787 singletons and 614 assembled sequences. Database comparisons of the cDNAs in GenBank's non-redundant databases using BLAST revealed that 4,957 of the 5,211 cDNAs (95.1%) showed a high degree of sequence homology to genes from other organisms. Most of the transcripts identified were genes related to metabolism, energy, protein biosynthesis and subcellular localization. The metabolism and energy categories of the *O. minuta* ESTs showed a considerably higher gene expression level than those of *O. sativa* ESTs. These data and genes can be utilized in rice breeding.

Keywords Wild rice · *Oryza minuta* · Expressed sequence tag · Functional classification · Vegetative stage.

Communicated by I.S. Chung

S.K. Cho and S.H. Ok contributed equally to the investigation reported here.

All EST data are publicly available through the National Center for Biotechnology Information (NCBI, USA; GenBank dbEST Accession No. CB209721~CB214919).

S. K. Cho · S. H. Ok · K. S. Shim · K. W. Jung · M. K. You · J. S. Shin (✉)
School of Life Sciences and Biotechnology,
Korea University, 136-701 Seou, Korea
e-mail: jsshin@korea.ac.kr
Tel.: +82-2-32903430
Fax: +82-2-9279028

J. U. Jeung
IRRI-Korea Office, National Crop Experiment Station,
Rural Development Administration,
441-100 Suwon, Korea

K. H. Kang · H. C. Choi · H. P. Moon
National Crop Experiment Station,
Rural Development Administration,
441-100 Suwon, Korea

Y. S. Chung
Life Science and Natural Resources,
Dong-A University, 604-714 Pusan, Korea

Introduction

Cultivated rice (*Oryza sativa* L.) is the world's most important food crop, and more than 90% of the world's rice production and consumption occurs in Asia. One-third of the world's population depends on rice as a primary food source (Khush 1997). However, the productivity of rice is continuously being threatened by various unfavorable environmental factors, such as salinity, drought and growing temperature, and by many kinds of pathogens.

Wild crop relatives are an important source of existing genetic variations, including resistance to various stresses (Vaughan 1994; Xiao et al. 1998). In the case of rice, wild relatives have diversified over a wide range of environments over 40 million years and have become major gene sources for various valuable economic characteristics that are not easily available in the cultivated germplasm (Khush 1997). The *Oryza* genus comprises 23 species and nine recognized genome types (AA, BB, BBCC, CC, CCDD, EE, FF, GG and HHJJ). These genomes are distantly related to *O. sativa* and could provide beneficial genes for cultivated rice (Ge et al. 1998; Vaughan 1994). In the case of *O. rufipogon* (AA genome), in spite of its overall inferior appearance, it is known to contain valuable traits with respect to salinity tolerance and cytoplasmic male sterility, and useful quantitative trait loci (QTL) with agronomically important traits. *O. officinalis* (CC genome) has also been reported to show resistance to vermin, such as the yellow stem borer, planthopper and leafhopper. These results suggest that the wild-type germplasms may provide new solutions to improve rice productivity, which may also be applicable

to other crops (Brar et al. 1991; Brar and Khush 1997; Vaughan 1994; Xiao et al. 1996, 1998).

Expressed sequence tags (ESTs), which is the single-pass sequencing of randomly chosen clones from the cDNA library, are molecular tools that are more than adequate in defining an expressed gene and reflecting transcript abundance. The global and comprehensive analyses of many species have been made possible by large-scale EST projects (Adams et al. 1991). Large-scale EST databases provide a great deal of information on the complexities of gene expression patterns, the functions of transcripts and the development of single nucleotide polymorphisms (SNPs) (Breyne and Zabwau 2001; Michalek 2002; Wu et al. 2002). In the plant kingdom, large-scale EST databases have been accumulated for model plants and crops—for example, *Arabidopsis thaliana*, *O. sativa*, *Zea mays*, *Brassica napus*, *B. campestris* and *Medicago truncatula*. In addition, various ESTs from diverse tissues, developmental stages, and stress-treated cDNA libraries have been compared and reported (Carson et al. 2002; Chen et al. 2002; Covitz et al. 1998; Höfte et al. 1993; Keith et al. 1993; Lim et al. 1996; Ok et al. 2000; Newman et al. 1994; Park et al. 1993; Qutob et al. 2000; Sasaki et al. 1994; Uchimiya et al. 1992; Ujino-Ihara et al. 2000; Wu et al. 2002). In the cases of *A. thaliana* and rice, full genome and draft sequences were reported recently (Goff et al. 2002; The Arabidopsis Genome Initiative 2000; Yu et al. 2002).

We describe here, for the first time, the expression patterns of novel genes expressed under normal growth conditions in a wild species of the *Oryza* genus. *O. minuta*, a wild relative of rice, contains the BBCC genome and has been used as a donor of resistance to blast and bacterial blight. Because of the potential importance of wild rice species in rice breeding, rice breeders have recently been paying them a great deal of attention (Vaughan 1994). We report on the partial sequences, database comparisons and functional categorization of 5,211 randomly collected vegetative-stage leaf cDNA clones of *O. minuta* (BBCC) based on the classification of the Munich Information Center for Protein Sequences (MIPS) for *Arabidopsis thaliana*. These data will be useful for those searching for novel genes expressed at the vegetative leaf stage and will contribute to the connection between EST databases designed to elucidate the gene expression profiles of plants, especially those of the Poaceae family.

Materials and methods

Plant material and construction of the cDNA library

Four-week-old leaf samples of glasshouse-grown *Oryza minuta* (accession no. 101144) were used for this study. Total RNA was isolated from leaves using TRIZOL reagent (Gibco/BRL, Gaithersburg, Md.) according to the manufacturer's instruction. The amount and quality of total RNA was checked by spectrophotometry ($OD_{260/280}$) and a formaldehyde-1% agarose gel electrophoresis system. Poly $(A)^+$ RNA was extracted from total RNA using the PolyATtract mRNA isolation system (Promega, Madison, Wis.) according to the manufacturer's protocol. A HybriZAP-2.1XR library construction kit and a HybriZAP-2.1 XR cDNA synthesis kit (Stratagene, La Jolla, Calif.) were used to construct the cDNA library at Eugentech (Korea). The library was packaged into Gigapack III Gold packaging extract; lambda ZAP yielded 6×10^6 primary plaques, which were then amplified to a titer of 3×10^{10} pfu/ml. cDNA-inserted pAD-GAL4-2.1 phagemid vectors were excised by mass in vivo excision using an ExAssist helper phage system (Stratagene). The titer of the resulting library was as 1.67×10^8 cfu/ml, and phagemids were used to infect *Escherichia coli* strain XLOLR according to the manufacturer's instructions.

Nucleotide sequencing and sequence data analysis

A total of 5,760 randomly collected clones were sequenced at Green Gene BioTech (Korea) from the 5′ ends using the 5′ AD primer (5-AGGGATGTTTAATACCACTAC-3′) of the pAD-GAL4-2.1 phagemid vectors. Ambiguous sequences of the 5′ and 3′ ends were removed, and vector sequences were trimmed automatically using a custom Python script. This script linked sequence backup, basecalling by Phred (trimming option on, cut-off set to 0.05; Green Gene BioTech). Sequences smaller than 200 bp or with more than 5% ambiguity were excluded. BLASTX searches and putative identifications were carried out automatically by Python script and the web BLAST program. Contigs were constructed with the edited sequences using CAP (Contig Assembly Program). All EST data are publicly available through the National Center for Biotechnology Information (NCBI, USA; GenBank dbEST accession nos. CB209721–CB214919).

Comparative analysis of the EST sequences

The MIPS functional categories applied to *Arabidopsis* genes were used for *O. minuta*. Translated *O. minuta* ESTs were categorized into 20 functional groups and an unclear classified group by sequence comparison with all *Arabidopsis* proteins using a *P*-value cut-off threshold of 10^{-5}. All *Arabidopsis* protein sequences were downloaded from the ftp site of MIPS and transformed to BLAST searchable data by the formatdb program (NCBI). Functional redundancies of *Arabidopsis* proteins were allowed in the classification.

Results and discussion

Characterization of the cDNA library and EST sequences

A cDNA library was constructed using mRNA isolated from *Oryza minuta* leaves. The primary library contained 6×10^6 recombinant phages, and after plaque amplification, the serial titers of the amplified plaques showed that the library contained approximately 3×10^{10} pfu/ml of SM buffer, which was considered to adequately represent gene expression. After the removal of vector sequences and ambiguous short sequences (<200 bp) from the 5' end sequences of 5,760 *O. minuta* cDNA clones, 5,211 clones revealed meaningful sequences (Table 1). These 5,211 *O. minuta* cDNAs produced a total of 3,401 unique sequences, which consist of 2,787 singletons and 614 assembled sequences. Redundancy (ESTs assembled in clusters/total ESTs) of an mRNA indicates the abundance of its corresponding cDNA in non-normalized libraries. In other words, information on randomly picked cDNAs represents the relative expression levels of the genes in a

Table 1 Summary of the *Oryza minuta* EST library

Total number of ESTs: 5,211	
Number of clusters	614 (2,424)
Number of singletons	2,787
Redundancy (%)[a]	46.5
Homologs (no. of ESTs)	
All organisms	4,957
Unknown	254
G+C content (%)	
Known ESTs	44
Unknown	46

[a] Redundancy is the percentage of ESTs assembled in clusters/total ESTs

library. In this project, among than 5,211 total cDNAs, 2,424 genes were assembled in 614 clusters, indicating a redundancy of 46.5%.

Database comparisons of cDNAs in GenBank non-redundant databases using BLAST revealed that 4,957 of the 5,211 cDNAs (95.1%) showed a high degree of sequence similarity to genes from other organisms (Table 1). The remaining 5.9% (254 clones of 5,211) of the sequenced cDNAs did not meet the criterion required for a match (E-value cut-off at 10^{-5}). It was clear that other methodologies, such as RNA blot analysis, linkage mapping or transformation into *Arabidopsis*, would be required to identify the functions of these clones and provide more information.

Expression profiles of leaf transcripts at the vegetative stage

Although relatively few clones (5,211) have actually been sequenced, diverse groups of genes have been identified in *O. minuta*. Approximately 73% of ESTs were assigned functions by alignment with *Arabidopsis* proteins, with E-values lower than 10^{-5} (Fig. 1). As expected, most of the identified transcripts appeared to be genes related to the metabolism and a subcellular localization (Fig. 1). The ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit (rbcS), which is the key enzyme of carbon assimilation, was the most frequently found gene (263 hits, Table 2). Several other genes related to energy and protein biosynthesis were also found in abundance (Fig. 1). These results are quite different from those of other plant EST projects in which different tissue materials were used (Crookshanks et al. 2001; Keith et al. 1993; Lim et al. 1996; Park et al. 1993; Sasaki et al. 1994; Uchimiya et al. 1992; Xiao et al. 1998), showing that the cells in vegetative leaves are metabolically active.

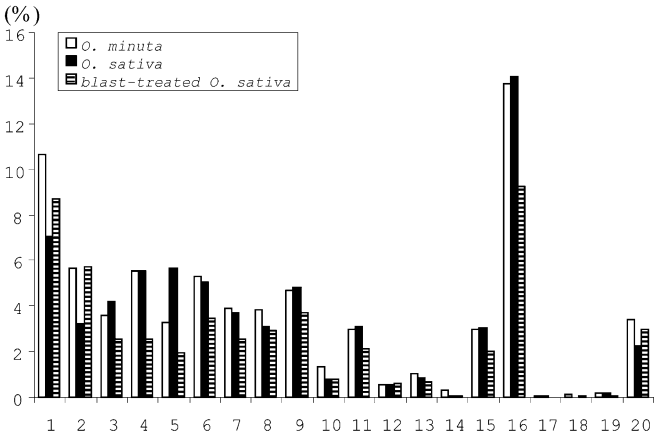

Fig. 1 Functional classifications and comparative analysis of the ESTs of *Oryza minuta*, *O. sativa* and blast-treated *O. sativa*. The ESTs were classified on the basis of their biological functions by alignment to *Arabidopsis* protein sequences using an E-value cutoff of 10^{-5}. For comparative analysis, 8,720 *O. sativa* immature leaf ESTs and 15,599 blast-treated *O. sativa* ESTs were retrieved from dbEST and grouped into functional categories. The *Y-axis* indicates percentage (%) of ESTs matched with *Arabidopsis* protein DB; the *X-axis*: *1* metabolism, *2* energy, *3* cell-cycle and DNA processing, *4* transcription, *5* protein synthesis, *6* protein fate, *7* cellular transport and transport mechanisms, *8* cellular communication/signal transduction, *9* cell rescue, defense and virulence, *10* regulation of interaction with cellular environment, *11* cell fate, *12* systemic regulation of interaction with environment, *13* development, *14* transposable element, viral and plasmid proteins, *15* control of cellular organization, *16* subcellular localization, *17* protein activity regulation, *18* protein with binding function or cofactor requirement, *19* storage protein, *20* transport facilitation

As described in other projects (Covitz et al. 1998; Lee et al. 1998; Lim et al. 1996; Ok et al. 2000 Park et al. 1993; Sasaki et al. 1994), defense-related genes exhibited high expression (4.7%) in our experiment (Fig. 1). The fact that defense-related genes can be found in an EST project with normal tissue suggests that these genes may help plants maintain a normal physiological condition in response to several stresses. It should be pointed out that the metallothionein gene, which binds to toxic heavy metals and makes it easier for them to be excreted from the cell, was the second most frequent gene (103 hits, Table 2). The reason why this defense-related gene was observed at this level in *O. minuta* requires further examination.

In our study, 254 clones (about 5% of our EST data) did not match with previously reported GenBank data-

Table 2 Most prevalent mRNAs as determined by EST redundancy

Putative identity	Accession number	Number of ESTs in cluster
Rubisco small chain	D00644	263 (5.04)[a]
Metallothionein-like protein	AF001396	103 (1.97)
Protein translation factor SUI1	AF380357	84 (1.61)
Light-regulated protein	AP002969	63 (1.21)
Malate dehydrogenase	BAA12870	27 (0.52)

[a] Values in parenthesis are the percentages of the total number of ESTs

Table 3 ESTs of *0. minuta*-specific novel transcripts

Accession no.[a]	Clone name	Accession no.	Clone name	Accession no.	Clone Name	Accession no.	Clone name
CB209739	KOR-06-F-A11	CB210278	KOR-03-F-J17	CB211064	KOR-10-F-P10	CB211643	KOR-11-F-L23
CB209748	KOR-06-F-B15	CB210279	KOR-07-F-O21	CB211100	KOR-10-F-E10	CB211685	KOR-12-F-C05
CB209768	KOR-06-F-E14	CB210307	KOR-08-F-C08	CB211117	KOR-10-F-F18	CB211688	KOR-12-F-C08
CB209794	KOR-06-F-H18	CB210362	KOR-08-F-L06	CB211128	KOR-10-F-H11	CB211721	KOR-12-F-G10
CB209795	KOR-06-F-I12	CB210365	KOR-08-F-M02	CB211161	KOR-10-F-K17	CB211737	KOR-12-F-I10
CB209803	KOR-06-F-J10	CB210388	KOR-03-F-C02	CB211171	KOR-10-F-L15	CB211748	KOR-12-F-K04
CB209964	KOR-06-F-B01	CB210390	KOR-08-F-O09	CB211181	KOR-10-F-M18	CB211752	KOR-12-F-K08
CB209872	KOR-06-F-C22	CB210407	KOR-08-F-B11	CB211182	KOR-10-F-M19	CB211808	KOR-12-F-C12
CB209882	KOR-06-F-E21	CB210430	KOR-08-F-E11	CB211184	KOR-10-F-N12	CB211811	KOR-12-F-C17
CB209919	KOR-06-F-J22	CD210450	KOR-08-F-G18	CD211201	KOR-03-F-G21	CB211850	KOR-12-F-H16
CB209922	KOR-03-F-D14	CB210485	KOR-08-F-M12	CB211214	KOR-10-F-B21	CB211851	KOR-12-F-H17
CB209941	KOR-06-F-N19	CB210503	KOR-08-F-P12	CB211216	KOR-10-F-B23	CB211852	KOR-12-F-H18
CB209957	KOR-07-F-A04	CB210506	KOR-08-F-P16	CB211239	KOR-10-F-F01	CB211880	KOR-12-F-L12
CB209961	KOR-07-F-A09	CB210550	KOR-08-FI23	CB211257	KOR-03-F-H19	CB211935	KOR-12-F-D24
CB209964	KOR-07-F-B05	CB210612	KOR-09-F-C09	CB211272	KOR-10-F-J24	CB211946	KOR-03-F-E09
CB210003	KOR-07-F-H03	CB210638	KOR-09-F-G08	CB211277	KOR-10-F-K24	CB211959	KOR-12-F-H20
CB210031	KOR-07-F-K06	CB210663	KOR-09-F-J08	CB211313	KON-03-F-I20	CB211968	KOR-12-F-I24
CB210054	KOR-07-F-N06	CB210664	KOR-09-F-K02	CB211338	KOR-11-F-D04	CB211979	KOR-12-F-K21
CB210061	KOR-07-F-O02	CB210703	KOR-09-F-O09	CB211343	KOR-11-F-D09	CB211983	KOR-12-F-K24
CB210069	KOR-07-F-P04	CB210741	KOR-09-F-D18	CB211347	KOR-11-F-E03	CB211990	KOR-12-F-M21
CB210081	KOR-07-F-A15	CB210753	KOR-09-F-F18	CB211393	KOR-11-F-J07	CB212006	KOR-12-F-O21
CB210104	KOR-07-F-D18	CB210769	KOR-09-F-H16	CB211422	KOR-11-F-N08	CB212031	KOR-13-F-C04
CB210122	KOR-03-F-G17	C9210790	KOR-09-F-K16	CB211435	KOR-03-F-J24	CB212043	KOR-13-F-D08
CB210123	KOR-07-F-G15	CB210835	KOR-03-F-B01	CB211448	KOR-11-F-A16	CB212061	KOR-13-F-F06
CB210124	KOR-07-F-G16	CB210844	KOR-09-F-C21	CB211469	KOR-11-F-D15	CB212081	KOR-13-F-H09
CB210135	KOR-07-F-I14	CB210878	KOR-09-F-I20	CB211470	KOR-11-F-D16	CB212082	KOR-13-F-I02
CB210140	KOR-07-F-J13	CB210882	KOR-09-F-I23	CB211495	KOR-11-F-G18	CB212140	KOR-13-F-O06
CB210148	KOR-07-F-K15	CB210892	KOR-09-F-K21	CB211618	KOR-11-F-J15	CB212144	KOR-13-F-P03
CB210150	KOR-07-F-K19	CB210894	KOR-09-F-K23	CB211529	KOR-11-F-L10	CB212235	KOR-04-F-F06
CB210168	KOR-07-F-M19	CB210939	KOR-10-F-B07	CB211538	KOR-11-F-M12	CB212263	KOR-13-F-M14
CB210170	KOR-07-F-N12	CB210947	KOR-10-F-C05	CB211561	KOR-11-F-O18	CB212328	KOR-14-F-E03
CB210177	KOR-07-F-O11	CB210972	KOR-10-F-F05	CB211580	KOR-11-F-C01	CB212341	KOR-14-F-G05
CB210223	KOR-07-F-E22	CB211022	KOR-10-F-K10	CB211620	KOR-11-F-I20	CB212344	KOR-14-F-H03
CB210226	KOR-07-F-F21	CB211051	KOR-10-F-O07	CB211628	KOR-11-F-J20	CB212346	KOR-04-F-G08
CB210234	KOR-07-F-H19	CB211060	KOR-10-F-P06	CB211633	KOR-11-F-K20	CB212384	KOR-14-F-M09
CB212405	KOR-14-F-P03	CB213080	KOR-04-F-P10	CB214430	KOR-04-F-H22	CB214808	KOR-05-F-O22
CB212432	KOR-14-F-D12	CB213098	KOR-16-F-C15	CB214448	KOR-04-F-K20	CB214816	KOR-06-F-P22
CB212457	KOR-04-F-I03	CB213142	KOR-16-F-I13	CB214468	KOR-04-F-N23	CB214820	KOR-08-F-A04
CB212492	KOR-14-F-M15	CB213203	KOR-16-F-P17	CB214482	KOR-05-F-A05	CB214824	KOR-06-F-B02
CB212501	KOR-03-F-F08	CB213251	KOR-16-F-I22	CB214504	KOR-05-F-D06	CB214864	KOR-06-F-F08
CB212535	KOR-04-F-J04	CB213257	KOR-04-F-C12	CB214535	KOR-05-F-H05	CB214892	KOR-06-F-J05
CB212557	KOR-04-F-J06	CB213265	KOR-16-F-K22	CB214540	KOR-05-F-I02	CB214894	KOR-06-F-J07
CB212574	KOR-14-F-K22	CB213268	KOR-04-F-C13	CB214559	KOR-05-F-KO4	C9214898	KOR-06-F-K04
CB212585	KOR-14-F-M20	CB213297	KOR-17-F-A09	CB214577	KOR-05-F-M10	CB214910	KOR-06-F-M04
CB212603	KOR-15-F-A03	CB213302	KOR-04-F-C16	CB214581	KOR-05-F-N04		
CB212607	KOR-15-F-A08	CB213371	KOR-17-F-K05	CB214584	KOR-05-F-N07		
CB212621	KOR-15-F-C07	CB213375	KOR-17-F-K09	CB214621	KOR-05-F-E10		
C8212632	KOR-15-F-E02	CB213427	KOR-17-F-B12	CB214624	KOR-05-F-E15		
CB212647	KOR-15-F-G03	CB213446	KOR-04-F-E13	CB214645	KOR-03-F-M07		
CB212652	KOR-15-F-G08	CB213486	KOR-17-F-F17	CB214668	KOR-06-F-K15		
CB212698	KOR-15-F-M09	CB213479	KOR-04-F-E17	CB214669	KOR-05-F-K16		
CB212779	KOR-04-F-M06	CB213487	KOR-17-F-I14	CB214672	KOR-05-F-L10		
CB212800	KOR-15-F-J16	CB213658	KOR-01-F-B06	CB214678	KOR-03-F-N03		
CB212801	KOR-04-F-M08	CB213993	KOR-01-F-L18	CB214681	KOR-05-F-M15		
CB212823	KOR-04-F-N02	CB214079	KOR-04-F-M16	CB214689	KOR-03-F-N04		
CB212847	KOR-15-F-P16	CB214101	KOR-04-F-M17	CB214690	KOR-05-F-N15		

Table 3 (continued)

Accession no.[a]	Clone name	Accession no.	Clone name	Accession no.	Clone Name	Accession no.	Clone name
CB212851	KOR-15-F-A18	CB214125	KOR-02-F-M04	CB214691	KOR-05-F-N16		
CB212853	KOR-15-F-A21	CB214180	KOR-04-F-N17	CB214693	KOR-05-F-O10		
CB212859	KOR-15-F-B21	CB214217	KOR-02-F-I13	CB214695	KOR-05-F-O12		
CB212860	KOR-15-F-B22	CB214239	KOR-02-F-K19	CB214728	KOR-05-F-C21		
CB212879	KOR-04-F-N08	CB214268	KOR-04-F-O17	CB214731	KOR-05-F-C24		
CB212885	KOR-15-F-F01	CB214301	KOR-02-F-D20	CB214747	KOR-05-F-E24		
CB212900	KOR-15-F-H21	CB214333	KOR-02-F-H24	CB214771	KOR-05-F-I23		
CB212944	KOR-15-F-P01	CB214361	KOR-02-F-L22	CB214772	KOR-05-F-I24		
CB212951	KGR-18-F-A03	CB214374	KOR-02-F-N22	CB214775	KOR-05-F-J20		
CB212961	KOR-16-F-B05	CB214396	KOR-04-F-B24	CB214778	KOR-03-F-O07		
CB212964	KOR-16-F-B09	CB214399	KOR-04-F-C19	CB214791	KOR-05-F-L24		
CB213026	KOR-16-F-J07	CB214404	KOR-04-F-D21	CB214794	KOR-05-F-M21		
CB213038	KOR-16-F-L05	CB214405	KOR-04-F-D22	CB214799	KOR-05-F-N20		
CB213042	KOR-16-F-M02	CB214418	KOR-04-F-F24	CB214808	KOR-05-F-O20		

[a] GenBank accession numbers and their clone names are given

base; hence, these ESTs were regarded as wild rice-specific novel transcripts (Table 3). While *O. minuta* may contain novel stress resistance genes that do not exist in *O. sativa*, this does not mean that all of the novel or new genes exist solely in the former—some were only expressed more highly in this wild rice species than in cultivated rice or even in other unrelated species. Therefore, in the near future, the blast-infected cDNA library of *O. minuta* should be analyzed for novel defense-related genes. The list of defense-related transcripts based on functional categorization is presented in Table 4.

Comparative analysis of leaf transcriptomes in wild (*O. minuta*) and cultivated rice (*O. sativa*)

The purpose of the EST analysis of *O. minuta* leaf was to find clues to the plant's resistance mechanisms and to characterize novel genes involved in stress resistance based on evidence that wild rice has a better resistance to abiotic and biotic stresses than cultivated rice (Brar and Khush 1997). Consequently, we compared *O. sativa* leaf ESTs with those registered at the GenBank EST database. To identify the differences between the expression profiles of *O. minuta* and *O. sativa*, we collected the immature leaf ESTs of *O. sativa* (8,720 ESTs) and categorized these in 20 functional groups by sequence comparison with all *Arabidopsis* proteins using a cut-off *P*-value threshold of 10^{-5}. ESTs of *Magnaporthe grisea* (rice blast)-infected *O. sativa* (15,599 ESTs) were also gathered for analysis. On comparing the expression profiles of these three transcriptomes (*O. minuta*, *O. sativa* and blast-treated *O. sativa*), we found that with respect to the metabolism and energy categories the gene expression levels of *O. minuta* and blast-infected *O. sativa* ESTs were considerably more elevated than that of normal *O. sativa* ESTs (Fig. 1). In terms of metabolism, the amino acid metabolism and nitrogen and sulfur metabolism sub-categories of *O. minuta* and blast-infected *O. sativa* exhibited elevated gene expression levels in comparison with those of normal *O. sativa*. Similarly, compared to *O. sativa* ESTs, *O. minuta* and blast-infected *O. sativa* ESTs showed higher gene expression levels in the sub-categories of photosynthesis, pentose-phosphate pathway, electron transport and membrane-associated energy conservation (Fig. 2A, B). In the case of protein synthesis, however, *O. minuta* showed a lower level (about one-half) of gene expression than that of *O. sativa* (Fig. 1), although *O. minuta* ESTs did show elevated levels in the sub-categories of translation, initiation, translational control, aminoacyl-tRNA-synthetases and other protein-synthesis activities, like those in blast-infected *O. sativa* (Fig. 2C). Based on these results, we suggest that the differences in gene expression between cultivated rice and their wild relatives rely on the regulation of gene expression, rather than novel sequences per se. Ongoing investigations on mutant and overexpression, as well as the promoter analyses will hopefully confirm this supposition.

It should be noted that 30 retrotransposons (30/5,211) were found in *O. minuta* ESTs, but only five (5/8,720) were found in *O. sativa* ESTs. Retrotransposons are known to be the most abundant and widely spread transposable elements in eukaryotes and are commonly found as multi-copies in plant genomes (Kumar and Bennetzen 1999). The majority of these transposable elements in plants are believed to be inactive and not transcribed, a few elements, such as barley *BARE-1* (Suoniemi et al. 1998), tobacco *Tnt1* (Grandbastien et al. 1989) and rice *Tos17* (Hirochika et al. 1996) have been reported to be activated under conditions of biotic and abiotic stresses. Therefore, future refined research on these retrotransposons in *O. minuta* should provide information on the relationships between active transposable elements and stress resistance.

Table 4 The putative identities of ESTs in functional categories of cell rescue, defense and virulence

Accession no.[a]	Putative identity	Accession no.[a]	Putative identity
NP_94858	20S Proteasome beta subunit A (PBA1)	T06413	Cathepsin B-like cysteine proteinase
Q08937	29-kDa Ribonucleoprotein B	NP_196647	CBS domain-containing protein
CAC48323	2-Cys peroxiredoxin	NP_194627	Cdk-activating kinase 1At
NP_568983	3′(2′),5′-Bisphosphate nucleotidase	O06431	Chaperone protein dnaJ
AAD37166	3-Phosphoinositide-dependent protein kinase-1	NP_197588	Chaperonin
BAA92724	60-kDa Chaperonin Beta subunit	PW0005	Chaperonine 60-K alpha chain
Q43207	70-kDa Peptidylprolyl isomerase	CAA09935	Chloroplast protease
NP_486690	ABC transporter ATP-binding protein	NP_177412	Cinnamyl-alcohol dehydrogenase
NP_187678	ABC transporter family protein	AAC04687	ClpC
CAB75508	AB13-interacting protein	AAF87849	Contains similarity to RNA binding protein PutA
P51823	ADP-ribosylation factor	NP_177261	Cooper chaperone (CCH)-related
AAB65432	ADP-ribosylation factor 1	AAK01931	Cu/Zn-superoxide dismutase copper chaperone precursor
NP_180050	ADP-ribosylation factor 3	CAB81029	Cyclic nucleotide and calmodulin-regulated ion channel-like prot
BAB08464	ADP-ribosylation factor-like protein	NP_173408	Cyclic nurcleotide-regulated ion channel
T02927	Alcohol dehydrogenase	AAA57046	Cyclophilin 2
P93436	Alcohol dehydrogenase class III	S30150	Cysteine proteinase
NP_175081	Aldehyde dehydrogenase	T02024	Cytochrome b245 beta chain homolog rbohA
BAB19052	Aldehyde dehydrogenase ALDH2b	Q43267	Cytochrome P450
NP_172907	Anionic peroxidase	NP_568463	Cytochrome P450 family
BAB07962	Anthocyanin 5-*O*-glucosyltransferase	BAB92262	Cytochrome P450-like protein
P92956	Arg/Ser-rich splicing factor RSP41	BAA96794	Cytosolic aldehyde dehydrogenase
BAB17666	Ascorbate peroxidase	BAA77214	Cytosolic monodehydroascorbate reductase
AAL09741	AT4g05320/C17L7_240	AAK70636	Cytosolic IRNA-Ala synthelase
T51174	Ataxia-telangiectasia mutated protein	BAB68072	Dehydration-responsive protein RD22 precursor
BAB10917	ATP binding protein associated with cell differentiation	BAB39219	Dihydrolipoamide dehydrogenase precursor
P31542	ATP-dependent clp protease ATP-binding subunit	NP_188404	Dihydroxyacetone kinase
BAB91828	B1103C09.18	AAK27812	Disease resistance protein
NP_564932	B-box zinc finger family protein	NP_195582	Disease resistance response protein-related
BAA96835	Beta 2 subunit of 20S proteasome	NP_199065	Disease resistance response protein-related
BAA96839	Beta 7 subunit of 20S proteasome	NP_568170	D-lactate dehydrogenase (D-LCR) (DLD) family
CAA41685	Beta-glucanase	BAB84605	DNA excision repair protein ERCC3-like
BAB86071	Beta-glucosidase	NP_176707	DnaJ domain-containing protein
AAL14713	Beta-glucosidase isozyme 2 precursor	NP_565982	DnaJ protein family
NP_197090	Carboxylesterase-like protein	AAD27555	DnaJ-like protein
P12365	Catalasa isozyme 2	T08901	DnaK-type molecular chaperone HSC70-11
NP_187332	Drought-induced protein-related	P14895	High molecular mass early light-inducible protein HV58
AAL66989	E2, ubiquitin-conjugating enzyme	BAA77822	Homeobox gene
AAM08126	Elicitor- and UV light-related transcription factor	AAD37698	Homeodomain Leu zipper protein
AAK62346	Elicitor-inducible cytochrome P450	NP_179233	Homeodomain *protein*
AAK11516	Ent-kaurenoic acid oxidase	NP_568314	HSP100/ClpB
T07044	Epoxide hydrolase	AAK53865	Hydrolase
T03439	Ethylene-response protein	NP_196884	Hydrolase alpha/beta fold family
AAD46405	Ethylene-responsive small GTP-binding protein	AAF68391	Hypersensitive-induced response protein
NP_172840	Expressed protein	AAF18727	Hypothetical protein
BAA88237	Ferredoxin	T00960	Hypothetical protein F20D22.10
O23877	Ferredoxin-NADP reductase	T47848	Hypothetical protein T8B10.30
AAL32441	Fe-superoxide dismutase precursor	NP_194140	Hypothetical protein; protein id: At4g24090.1
AAG17894	Genetic modifier	NP_568922	Imidazoleglycerol-phosphate synthase subunit H-like
BAB64202	Glutaredoxin	NP_191111	Immunophilin/FKBP-type peptidyl-prolyl *cis*-trans isomerase
CAB59895	Glutathione peroxidase-like protein GPX54Hv	NP_069681	Inosine monophosphate dehydrogenase (guaB-1)
AAG34826	Glutathione S-transferase GST 18	P52681	Isoflavone reductase homolog
AAC64007	Glutathione S-transferase II	T04375	Jacalin homolog
NP_171793	Glutathione transferase	BAB55659	*KNOX* family class 2 homeodomain protein
NP_182301	Glutathoine-conjugate transporter AtMRP4.	Q9THX6	L-ascorbate peroxidase
P93436	Glutathione-dependent formaldehyde dehydrogenase	NP_173168	Leu-rich repeat protein family
T03766	Glutathione-disulfide reductase	NP_177363	Leu-rich repeat transmembrane protein kinase
NP_565939	Glycerol-3-phosphate dehydrogenase	NP_567439	Light-induced protein
NP_189273	Gly-rich RNA-binding protein	BAB63876	Lipoamide dehydrogenase
NP_567787	Glycosyl hydrolase family 1	AAD38281	Low molecular early light-inducible protein
NP_194413	Glycosyl hydrolase family 17	P20721	Low-temperature-induced cysteine proteinase precursor
NP_172855	Growth regulator protein-related	NP_179721	Mannose 6-phosphate reductase
BAA90672	GSH-dependent dehydroascorbate reductase 1	AAF61238	MAP kinase MAPK2
BAB07961	GTP-binding protein	AAL82672	Membrane protein
NP_196709	GTP-binding protein-related	P26571	Metallothionein-like protein 1

Table 4 (continued)

Accession no.[a]	Putative identity	Accession no.[a]	Putative identity
NP_178487	Heat shock protein	AAK38824	Metallothionein-like protein type 2
AAF23074	Heat shock protein 70	AAK52524	MLA6 protein
NP_200414	Heat shock protein 81-2	CAA73621	Multicatalytic endopeptidase
T09295	Heat shock protein EMB1	NP_027544	Myb family transcription factor
T06102	Heat shock protein T5J17.130, dnaJ-type	BAA94084	NAD-dependent sorbitol dehydrogenase
BAA35120	NADH-dependent glutamate synthase	NP_180483	Pumilio-family RNA-binding protein
AAK93796	NBS-LRR-like protein	BAB21279	Putative ABC transporter protein
T05727	Nucleic acid-binding protein	AAC78265	Putative chloroplast outer envelope 86-like protein
NP_172633	Obtusifoliol 14-demethylase	AAG46147	Putative cytochrome P450-related protein
NP_173786	Oxidoreductase, zinc-binding dehydrogenase family	BAB90827	Putative histidine kinase
BAB92570	P0497A05, 14	AAK00972	Putative homeodomain protein
T02055	Pathogenesis related protein-5	BAB62557	Putative MRP-like ABC transporter
T04299	Pathogenesis-related protein class 1	CAC09354	Putative oryzain alpha precursor
T04165	Pathogenesis-related thaumatin-like protein	CAC34501	Putative protein
Q9SEC2	Peptide methionine sulfoxide reduclase,	AAL82527	Putative ribonucleoprotein
S48017	Peptidylprotyl isomerase	AAG13589	Putative ubiquitin protein
BAB39277	Peroxidase	AAL58113	Putative ubiqutin-conjugating enzyme
NP_189235	Peroxiredoxin	T03962	r40g3 protein
BAB62533	Peroxisome-type ascorbate peroxidase	BAB85469	Rad6
AAK01711	Phosphoinositide-specific phospholipase C	Q40522	Ras-related protein Rab 11 D
P10931	Phytochrome A	P40392	Ras-related protein RIC1
AAA34093	Phytochrome B	AAK18840	Receptor kinase
S25164	Polyubiquitin	AAM00988	Receptor protein kinase
NP_192994	Polyubiquitin-related protein	AAD43962	Receptor-like kinase ARK1AS
BAB90350	Pre-mRNA splicing factor SF2	NP_176009	Receptor-related protein kinase
CAB44983	Pre-pro-cysteine proteinase	BAB44007	Resistance gene analog PIC23
CAA06243	Pre-pro-TPE4A protein	T03983	rf2 nuclear restorer protein
G71400	Probable heat shock transcription factor	AAM01030	Riboflavin biosynthesis protein ribF
T02879	Probable plasma membrane intrinsic protein	T09557	Rieske iron-sulfur protein L73G19.30
AAF68384	Prohibitin	AAL25175	RING-H2 finger protein RHB1a
NP_197165	Pro-rich protein family	AAF87849	RNA binding protein PulA
Q9LSU3	Proteasoma subunit alpha type 6	NP_177494	RNA recognition motif (RRM)-containing protein
BAA75633	Protein abundantly expressed during apple fruit development	P08823	RuBisCO subunit binding-protein alpha subunit
AAG31752	Protein kinase AKINbetagamma-2	NP_200461	RuBisCO subunit binding-protein beta subunit; chaperonin
NP_564946	Protein kinase family	BAB92541	Rust resistance protein
BAB92400	Protein kinase Xa21	AAD51625	Seed maturation protein PM37
NP_197876	Protein phosphatase 2C	NP_565497	Senescence-associated protein
NP_192235	Protein phosphatase regulatory subunit	T14735	Ser/Thr kinase
AAM22720	Protein-Tyr-phosphatase	AAK38344	Seven transmembrane protein MIo8
NP_568 107	Pseudo-response regulator. APRR7 (APRR1/TOC1 family)	BAB66368	SHEPHERD
BAA92737	Similar to Zantedeschia aethiopica iron superoxide dismutase	AAD35009	Thioredoxin-like 5
AAK38149	Small GTP-binding protein	T07367	Thioredoxin-like protein CDSP32
AAL30396	Small Ras-related GTP-binding protein	NP_176610	Transcription factor inhibitor 1 kappa B-related
Q15393	Spliceosome associated protein 130	NP_187053	Transfactor-like
AAD52610	Splicing factor SR1B	NP_178216	Transfactor-related protein
S60767	S-receptor kinase	CAB50925	Transfocon Tic40
AAF68388	Stomatin-like protein	BA810596	Transporter-like protein
NP_191995	Stress-induced protein OZI1 precursor	NP_173799	Trehalose phosphatase family
Q08080	Stromal 70-kDa heat shock-related protein	NP_177979	Trehalose-6-phosphate synthase
NP_179435	Succinate dehydrogenase	AAK52534	U1 small nuclear ribonucleoprotein
NP_178062	Succinate-semialdehyde dehydrogenase	BAB39294	Ubiquitin/ribosomal protein S27a
AAA33917	Superoxide dismutese	CAA09619	Ubiquitin activating enzyme
AAF87039	T24P13.20	NP_197812	Ubiquitin family
AAK58370	T-cytoplasm male sterility restorer factor 2	NP_179311	Ubiquitin protein-related
AAK55325	Thaumatin-like protein TLP7	NP_567791	Ubiquitin-conjugating enzyme 9
NP_177735	Thioredoxin family	P25866	Ubiquitin-conjugating enzyme E2-17 kDa
Q81332	Thioredoxin F-type	S43786	Ubiquitin-protein ligase
P29449	Thioredoxin H-type	AAK00454	Unknown protein
Q9ZP20	Thioredoxin M-type	AAF70831	XIG
T00824	Thioredoxin reductase At2g41680	JE0113	Zinc-finger protein S3574

[a] GenBank accession number of the most similar sequence as identified by BLASTX alignment

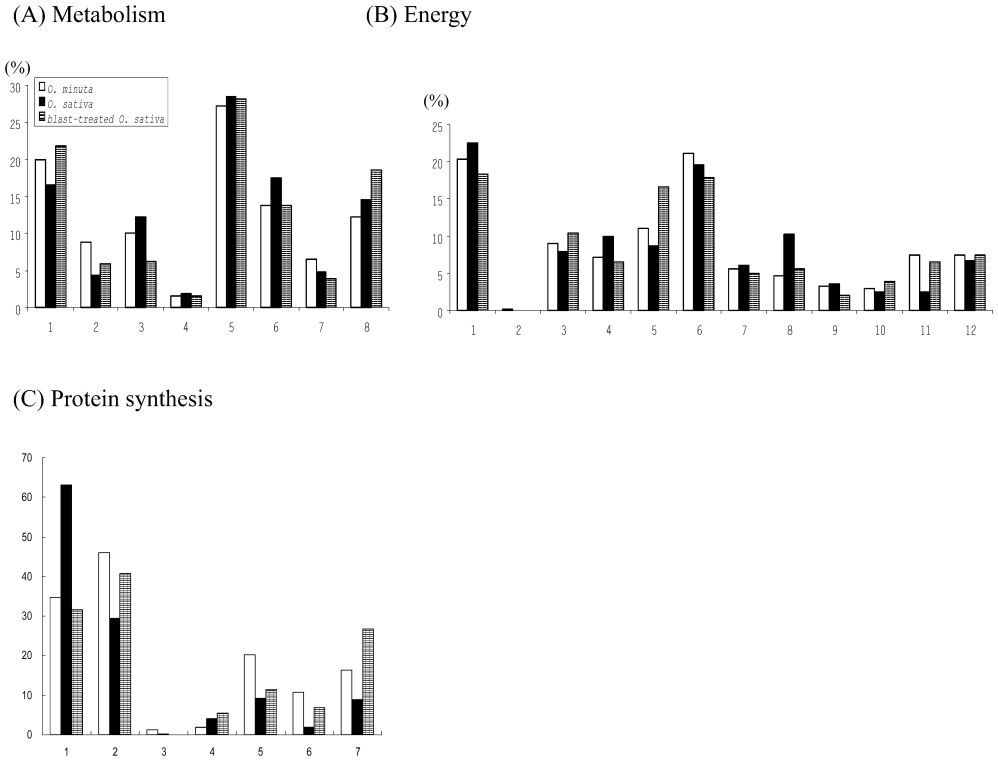

Fig. 2A–C Sub-functional classifications of ESTs matched to the categories of metabolism, energy and protein synthesis. The *Y-axis* indicates the percentage (%) of ESTs with *Arabidopsis* protein database matches. **A** Metabolism category: *1* amino acid metabolism, *2* nitrogen and sulfur metabolism, *3* nucleotide metabolism, *4* phosphate metabolism, *5* C-compound and carbohydrate metabolism, *6* lipid, fatty acid and isoprenoid metabolism, *7* metabolism of vitamins, cofactors and prosthetic groups, *8* secondary metabolism. **B** Energy category: *1* glycolysis and gluconeogenesis, *2* Entner-Doudoroff pathway, *3* pentose-phosphate pathway, *4* tricarboxylic acid pathway, *5* electron transport and membrane-associated energy conservation, *6* respiration, *7* fermentation, *8* metabolism of energy reserves, *9* glyoxylate cycle, *10* oxidation of fatty acids, *11* photosynthesis, *12* other energy generation activities. **C** Protein synthesis category: *1* ribosome biogenesis, *2* translation, *3* initiation, *4* elongation, *5* translational control, *6* aminoacyl-tRNA-synthetase, *7* other protein-synthesis activities

The ESTs described here are the first reported transcriptomes to be expressed at the vegetative leaf stage of wild rice. The gene expression profile of *O. minuta* was compared with that of cultivated rice, *O. sativa*. These genes and the different gene expression pattern can be used to unravel the regulatory networks of stress resistance in rice and possibly in other crops. The EST data provided also makes it feasible for molecular breeders to develop new varieties of cultivated rice with higher stress resistance.

Acknowledgements This project was funded by a grant from the Crop Functional Genomics Center of the 21st Century Frontier Research Program funded by the Ministry of Science and Technology and by a grant from the BioGreen21 Program of Rural Development Administration.

References

Adams MD, Kelly JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Meril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Ventor JC (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252:1651–1656

Brar DS, Khush GS (1997) Alien introgression in rice. Plant Mol Biol 35:35–47

Brar DS, Elloran R, Khush GS (1991) Interspecific hybrids produced through embryo rescue between cultivated and eight wild species of rice. Rice Genet Newsl 8:91–93

Breyne P, Zabwau M (2001) Genome-wide expression analysis of plant cell cycle modulated genes. Curr Opin Plant Biol 4:136–142

Carson DL, Huckett BI, Botha FC (2002) Sugarcane ESTs differentially expressed in immature and maturing internodal tissue. Plant Sci 162:289–300

Chen M et al. (2002) An integrated physical and genetic map of the rice genome. Plant Cell 14:537–545

Covitz PA, Smith LS, Long SR (1998) Expressed sequence tags from a root-hair-enriched *Medicago truncatula* cDNA library. Plant Physiol 117:1325–1332

Crookshanks M, Emmersen J, Welinder KG, Nielsen KL (2001) The potato tuber transcriptome: analysis of 6,077 expressed sequence tags. FEBS Lett 506:123–126

Ge S, Sang T, Lu BR, Hong DY (1999) Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proc Natl Acad Sci USA 96:14400–14405

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S. Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gurin A, Pruss D,

Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Reldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome. Science 296:92–100

Grandbastien MA, Spielmann A, Caboche M (1989) *Tnt1*, a mobile retroviral-like transposable elements of tobacco isolated by plant cell genetics. Nature 337:376–380

Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M (1996) Silencing of retrotransposons in *Arabidopsis* and reactivation by the *ddm1* mutation. Proc Natl Acad Sci USA 23:7783–7788

Höfte H, Desprez T, Amselem J, Chiapello H, Caboche M, Moisan A, Jourjon M, Charpenteau J, Berthomieu P, Guerrier D, Giraudat J, Quigley F, Thomas F, Yu D, Mache R, Raynal M, Cooke R, Grellet F, Delseny M, Parmentier Y, Marcillac G, 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 4:1051–1051

Keith CS, Hoang DO, Barrett BM, Feigelman B, Nelson MC, Thai H, Baysdorfer C (1993) Partial sequence analysis of 130 randomly selected maize cDNA clones. Plant Physiol 101:329–332

Khush GS (1997) Origin, dispersal, cultivation and variation of rice. Plant Mol Biol 35:25–34

Kumar A, Bennetzen JL (1999) Plant retrotransposons. Annu Rev Genet 33:479–532

Lee CM, Lee YJ, Lee MH, Nam HG, Cho TJ, Hahn TR, Cho MJ, Sohn U (1998) Large-scale analysis of expressed genes from the leaf of oilseed rape (*Brassica napus* L.). Plant Cell Rep 17:930–936

Lim CO, Kim HY, Kim MG, Lee SI, Chung WS, Park SH, Hwang I, Cho MJ (1996) Expressed sequence tags of Chinese cabbage flower bud cDNA. Plant Physiol 111:577–588

Michalek W, Weschke W, Pleissner KP, Graner A (2002) EST analysis in barley defines a unigene set comprising 4,000 genes. Theor Appl Genet 104:97–103

Newman T, de Bruijn 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

Ok SH, Chung YS, Um BY, Park MS, Bae JM, Lee SJ, Shin JS (2000) Identification of expressed sequence tags of watermelon (*Citrullus lanatus*) leaf at the vegetative stage. Plant Cell Rep 19:932–937

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

Qutob D, Hraber PT, Sobral BWS, Gijzen M (2000) Comparative analysis of expressed sequences in *Phytophthora sojae*. Plant Physiol 123:243–253

Sasaki T, Song J, Koga-Ban Y, Matsui E, Fang F, Higo H, Nagasaki H, Hori M, Miya M, Murayama-Kayano E, Takiguchi T, Tasasuga 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

Suoniemi A, Tanskanen J, Schulman AH (1998) Gypsy-like retrotransposons are widespread in the plant kingdom. Plant J 13:699–705

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Uchimiya Y, Lidou S, Shimazaki T, Takamatsu S, Hashimoto H, Nishi R, Aotsuka S, 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

Ujino-Ihara T, Yoshimura K, Ugawa Y, Yoshimura H, Nagasaka K, Tsumura Y (2000) Expression analysis of ESTs from the inner bark of *Cryptomeria japonica*. Plant Mol Biol 43:451–457

Vaughan DA (1994) The wild relatives of rice—a genetic resources handbook. IRRI, Philippines

Wu J, Maehara T, Shimokawa T, Yamamoto S, Harada C, Takazaki Y, Ono N, Mukai Y, Koike K, Yasaki J, Fujii F, Shomura A, Ando T, Kono I, Waki K, Yamamoto K, Yano M, matsumoto T, Sasaki T (2002) A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell 14:525–535

Xiao J, Grandillo S, Ahn SN, McCouch SR, Tanksley SD, Li J, Yuan L (1996) Genes from wild rice improved yield. Nature 384:223–224

Xiao J, Li J, Grandillo S, Ahn SN, Yuan L, Tanksley SD, McCouch SR (1998) Genes from wild rice improved yield. Genetics 150:899–909

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, He S, Zhang J, Xu H, Ahang K, Aheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, H;ao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296:79–92