View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Publications of the IAS Fellows

© Indian Academy of Sciences

Annotation and BAC/PAC localization of nonredundant ESTs from drought-stressed seedlings of an *indica* rice

P. RAVINDRA BABU, A. CHANDRA SEKHAR, NAGABHUSHANA ITHAL,
G. MARKANDEYA and ARJULA R. REDDY*

*Plant Molecular Genetics and Functional Genomics Laboratory, Department of Plant Sciences,
School of Life Sciences, University of Hyderabad, Hyderabad 500 046, India*

Abstract

To decipher the genes associated with drought stress response and to identify novel genes in rice, we utilized 1540 high-quality expressed sequence tags (ESTs) for functional annotation and mapping to rice genomic sequences. These ESTs were generated earlier by 3′-end single-pass sequencing of 2000 cDNA clones from normalized cDNA libraries constructed from drought-stressed seedlings of an *indica* rice. A rice UniGene set of 1025 transcripts was constructed from this collection through the BLASTN algorithm. Putative functions of 559 nonredundant ESTs were identified by BLAST similarity search against public databases. Putative functions were assigned at a stringency E value of 10^{-6} in BLASTN and BLASTX algorithms. To understand the gene structure and function further, we have utilized the publicly available finished and unfinished rice BAC/PAC (BAC, bacterial artificial chromosome; PAC, P1 artificial chromosome) sequences for similarity search using the BLASTN algorithm. Further, 603 nonredundant ESTs have been mapped to BAC/PAC clones. BAC clones were assigned by a homology of above 95% identity along 90% of EST sequence length in the aligned region. In all, 700 ESTs showed rice EST hits in GenBank. Of the 325 novel ESTs, 128 were localized to BAC clones. In addition, 127 ESTs with identified putative functions but with no homology in IRGSP (International Rice Genome Sequencing Program) BAC/PAC sequences were mapped to the Chinese WGS (whole genome shotgun contigs) draft sequence of the rice genome. Functional annotation uncovered about a hundred candidate ESTs associated with abiotic stress in rice and *Arabidopsis* that were previously reported based on microarray analysis and other studies. This study is a major effort in identifying genes associated with drought stress response and will serve as a resource to rice geneticists and molecular biologists.

[Babu P. R., Sekhar A. C., Ithal N., Markandeya G. and Reddy A. R. 2002 Annotation and BAC/PAC localization of nonredundant ESTs from drought-stressed seedlings of an *indica* rice. *J. Genet.* **81**, 25–44]

Introduction

Rice, the most important food crop of the world, is the staple food for more than half the world's population and has become a model cereal because of its relatively small genome, availability of high-density genetic and physical maps, and now the complete draft genome sequence published by public (Yu *et al*. 2002) and private (Goff *et al*. 2002) enterprise. Rice has become central to cereal genetics and functional genomics. It is the favoured cereal in investigation of structure, function and regulation of genes. In particular, the rice genome now offers remarkable advantages in elucidating genetic and molecular bases of complex traits such as drought resistance. Rapid advances in genome technologies have made it possible to take a 'many genes at a time' approach instead of the usual 'one gene at a time' approach.

Broadly defined, expressed sequence tags (ESTs) are single-pass partial sequences of cDNA clones sequenced

---

*For correspondence. E-mail: arjulsl@uohyd.ernet.in.

Keywords. ESTs; annotation; *Oryza sativa*; *indica* rice; cDNA libraries; rice genome; drought tolerance.

from one end. A large number of them have been deposited in public databases such as the US National Center for Biotechnology Information (NCBI) database of expressed sequence tags (dbEST), and also in private domains. Large-scale EST sequencing has emerged as a direct way to catalogue the expressed genes in rice and many other organisms. EST clones are used in more than one way in molecular-biological and genetic analyses. In particular, ESTs have been used as molecular markers for construction of high-density genetic linkage maps (Harushima *et al.* 1998) and physical maps (Kurata *et al.* 1997) of rice. ESTs are found to be useful in identification of members of gene families (Cooke *et al.* 1997; Epple *et al.* 1997), uncovering long exons, and detection of splice variants; they also form the basis for large-scale identification of single nucleotide polymorphisms (SNPs) (Cho *et al.* 1999). ESTs are the central resources in studies of global gene expression through high-density microarrays and analysis of complex traits such as drought and salinity tolerance governed by multiple genes, as demonstrated in *Arabidopsis*, barley and rice (Kawasaki *et al.* 2001; Seki *et al.* 2001; Öztürk *et al.* 2002). The current dbEST release 040502 has 104,594 sequences from *Oryza sativa* (http://www.ncbi.nlm.nih.gov/dbEST/index.html). Recently, different groups generated a large number of UniGene sets of EST collections for different cereals such as barley (Michalek *et al.* 2002), maize (Fernandes *et al.* 2002) and wheat (Echenique *et al.* 2002) that will serve as powerful tools for comparative analysis and study of evolution of gene families among cereals (Bennetzen 2002). This will be an integral part of comparative genomics of cereals.

Localizing ESTs to genomic sequences provides a powerful route for gene prediction and understanding gene structure (Kan *et al.* 2001). This simplifies the cloning of agronomically important genes through use of the physical map of the rice genome (Chen *et al.* 2002). The International Rice Genome Sequencing Program (IRGSP) group isolated 29,000 cDNA clones from *japonica* rice (Yamamoto and Sasaki 1997), partially sequenced them from both 5′ and 3′ ends, and, using similarity analysis, classified them into a set of nonredundant unique sequences. These unique sequences have been used to generate clone-specific primers and then to construct a comprehensive rice transcript map using 3′-end sequences of rice cDNA clones (Wu *et al.* 2002).

Large-scale ESTs were earlier generated from cDNA libraries of different callus tissues or organs such as panicle, root and leaf at different developmental stages and following hormone treatments. In general these libraries represent abundantly synthesized transcripts and are likely to have missed transcripts that are synthesized in low copy number. Owing to complex expression patterns, such libraries may not include transcripts associated with specific stress response. To overcome these inherent problems and increase the likelihood of finding novel genes, ESTs were generated from normalized cDNA libraries (Bonaldo *et al.* 1996). To capture low-abundance mRNAs by reducing the redundancy in cDNA libraries and to increase efficiency of novel gene discovery, we have, as reported earlier (Reddy *et al.* 2002), constructed normalized cDNA libraries from drought-stressed rice seedlings. Here we describe the annotation of this nonredundant set of ESTs generated from drought-stressed rice seedlings

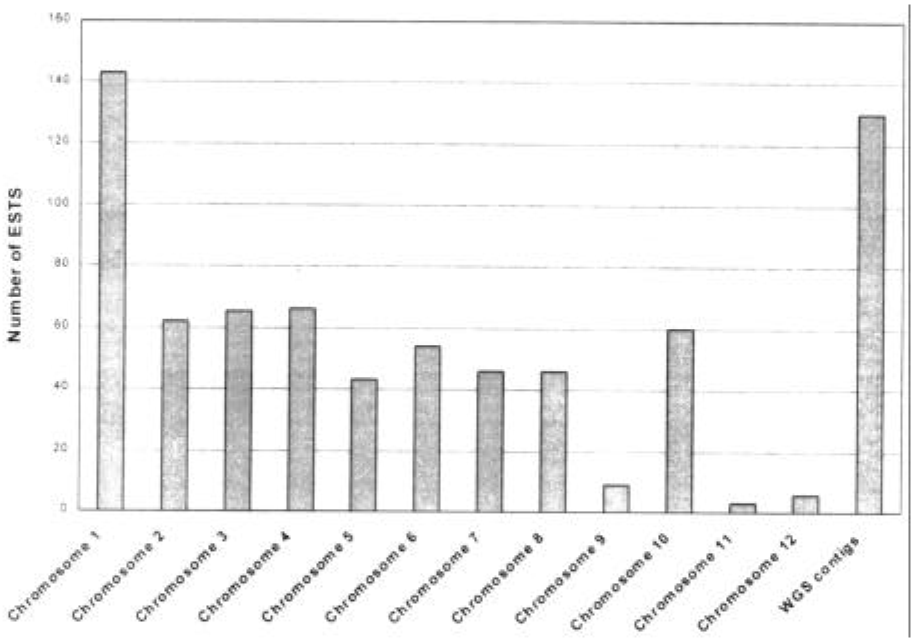

Figure 1. Localization of rice ESTs to rice chromosomes and the draft rice genome contigs.

and their utility in genetic and molecular analyses of drought tolerance in rice. The objective of this effort is to accelerate the process of identification of the genes associated with diverse stress response pathways in rice, define their function, and elucidate their coordinated regulation.

Materials and methods

EST source: High-quality ESTs deposited in the NCBI EST database dbEST generated from normalized cDNA libraries constructed from drought-stressed seedlings of N22 rice (Reddy *et al.* 2002) (http://www.ncbi.nlm.nih.gov/dbEST; GenBank Accession Numbers: BI305180–BI306756) were utilized for annotation in this study. These ESTs are from leaf and root libraries of drought-stressed *indica* rice N22 seedlings designated in GenBank as Library.73 (http://www.ncbi.nlm.nih.gov/UniGene/lib.cgi?ORG=Os&LID=73) and Library.72 (http://www.ncbi.nlm.nih.gov/UniGene/lib.cgi?ORG=Os&LID=72) respectively.

Annotation tools: Homology search was done against nonredundant (nr) nucleotide and protein sequence databases using BLASTN 2.2.2 and BLASTX 2.2.2 versions of the BLAST programs (Altschul *et al.* 1997) through BLAST 2.0 network client software with the dnatools interface (http://www.crc.dk/dnatools). The BLASTN program was used to identify rice EST hits and rice BAC clones in the High Throughput Genomic Sequences (HTGS) division of GenBank and the Chinese WGS (whole genome shotgun contigs) draft sequence of *indica* rice genome in the NCBI database. CLUSTALX version 1.8 (Thompson *et al.* 1994, 1997) was used to determine similarity against the database hits through pairwise alignment between EST sequence and database sequence. Genchek^{TM} (http://www.ocimumbio.com/nh/products_gcH.htm) was also used in annotation.

Results and discussion

The EST resource used in this study represents high-quality sequences with a Phred score of greater than 20 and showed a sequence similarity of above 95% with database hits and genomic regions. The 3′-end sequencing gave low BLASTX scores but clearly identified paralogues in the genome. The nonredundant (nr) set of ESTs was developed using the criteria that the ESTs do not show similarity to the same accession in nr nucleotide and protein sequence databases and do not fall in the same group against rice EST database by use of the BLASTN algorithm. The ESTs were mapped to BAC/PAC (BAC, bacterial artificial chromosome; PAC, P1 artificial chromosome) clones and the *indica* draft rice genome (http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/riceWGS.html). The results are summarized in figure 1. The summary of the annotation is given in table 1.

The proportions of ESTs with identified putative functions showing sequence homologies in various cereals and in *Arabidopsis* are shown in figure 2. Annotations of the rice leaf and root ESTs and BAC/PAC localization are shown in table 2a and table 2b, respectively. The leaf and root ESTs with putative functions localized to WGS *indica* rice genome draft sequence contigs are shown in table 3a and table 3b respectively. Interestingly, 175 ESTs that were mapped to rice genomic sequences have no orthologous sequences in other species. This may be because the ESTs often include short sequences of coding

(*text continues on page 42*)

Table 1. Summary of EST analysis.

Total number ESTs analysed	1545
ESTs from leaf cDNA library	1345
ESTs from root cDNA library	200
Nonredundant (nr) EST hits in nr nucleotide and protein database	762
nr ESTs showing similarity to rice ESTs	700
Novel ESTs identified	325
BAC/PAC clones assigned to nr ESTs	603
Novel ESTs assigned to BAC/PAC clones	129
nr ESTs with putative functions assigned to BAC/PAC clones	429
Total putative functions assigned	559
ESTs with putative functions mapped to WGS contigs	130
ESTs mapped to BAC/PAC clones without putative functions	175
Abiotic stress response genes uncovered	100

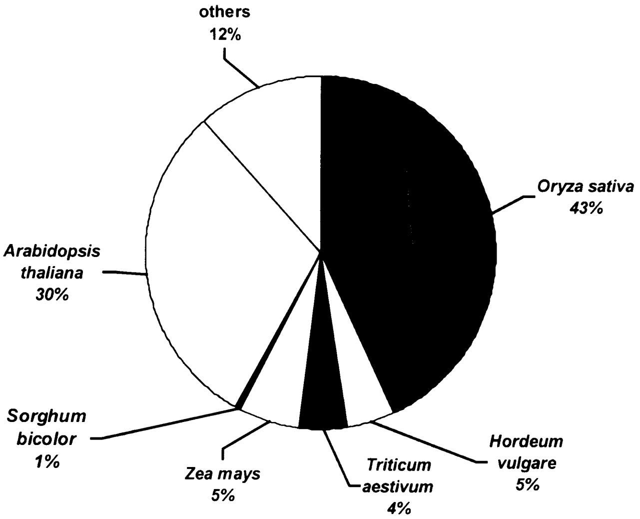

Figure 2. Proportions of the 559 rice ESTs assigned putative functions showing homologies in other cereals, *Arabidopsis* and other plants.

Table 2a. Annotation of ESTs from leaf cDNA library.

	Accn No.[a]	Putative function[b]	Identical Accn No.[c]	BAC/PAC clone[d]	BAC/PAC clone Accn No.	BAC/PAC E-value[e]
1	BI305482	PKF1	X97547	Chr 1 BAC clone:B1015E06	AP003197	0.0
2	BI305481	metallothionein-like protein	AF001396	Chr 1 BAC clone:B1015E06	AP003197	0.0
3	BI306401	NA	NH	Chr 1 BAC clone:B1064G04	AP003924	3e-95
4	BI306675	Ras-related GTP binding protein possessing GTPase activity	S66160	Chr 1 BAC clone:B1064G04	AP003924	5e-82
5	BI305955	NA	NH	Chr 1 BAC clone:B1065E10	AP003561	2e-53
6	BI306054	33 kDa oxygen-evolving protein of photosystem II	X57408	Chr 1 BAC clone:B1080D07	AP003203	0.0
7	BI306731	NA	NH	Chr 1 BAC clone:B1088C09	AP003734	4e-96
8	BI306220	hypothetical protein	AP003331	Chr 1 BAC clone:B1088D01	AP003331	0.0
9	BI305697	NA	NH	Chr 1 BAC clone:B1108H10	AP003562	2e-84
10	BI305715	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	M80912	Chr 1 BAC clone:B1148D12	AP003411	0.0
11	BI306556	signal recognition particle receptor alpha	AF360125	Chr 1 BAC clone:OJ1294_F06	AP004326	5e-88
12	BI306500	dTDP-glucose 4-6-dehydratase	AC005824	Chr 1 BAC clone:OJ1656_A11	AP003448	0.0
13	BI305740	putative sterol-C5(6)-desaturase	AF099969	Chr 1 BACclone:OSJNBa0083M16	AP003214	0.0
14	BI306421	NA	NH	Chr 1 BACclone:OSJNBa0083M16	AP003214	5e-66
15	BI306073	IAI2 for wound-induced protein homologue	AB059238	Chr 1 BACclone:OSJNBa0083M16	AP003214	0.0
16	BI305781	NA	NH	Chr 1 BACclone:OSJNBa0090K04	AP003216	2e-61
17	BI306460	NA	NH	Chr 1 BACclone:OSJNBb0021A09	AP003218	e-145
18	BI306660	NA	NH	Chr 1 BAC clone:OSJNBb0093M23	AP003854	0.0
19	BI305761	sucrose-6F-phosphate phosphohydrolase SPP3	AY029159	Chr 1 BACclone:B1111C09	AP003204	e-155
20	BI305874	AP2 domain-containing protein	AF071893	Chr 1 clone B1011A07	AP003722	4e-66
21	BI305434	acyl-CoA:1-acylglycerol-3- phosphate acyltransferase	AP002039	Chr 1 clone B1100D10	AP003368	0.0
22	BI306595	NA	NH	Chr 1 clone B1114B07	AP003334	0.0
23	BI306666	unknown protein	AP001299	Chr 1 clone B1144G04	AP003335	0.0
24	BI306591	HMG protein	Y08807	Chr 1 clone OSJNBa0047D12	AP003516	7e-47
25	BI305562	NA	NH	Chr 1 clone OSJNBa0052O12	AP004330	e-134
26	BI306502	NA	NH	Chr 1 clone OSJNBa0085D07	AP004331	2e-44
27	BI306229	NA	NH	Chr 1 clone OSJNBa0085D07	AP004331	0.0
28	BI305964	L24 ribosomal protein	X94296	Chr 1 clone OSJNBa0085D07	AP004331	1e-78
29	BI306283	NA	NH	Chr 1 clone OSJNBb0053G03	AP003377	0.0
30	BI305604	NA	NH	Chr 1 clone P0019E03	AP004363	1e-51
31	BI306474	purple acid phosphatase	AJ006224	Chr 1 clone P0413G02	AP003344	0.0
32	BI306608	unknown protein	AC015446	Chr 1 clone P0415C01	AP003243	e-162
33	BI306081	NA	NH	Chr 1 clone P0439E07	AP003768	e-174
34	BI306170	expressed protein	NM_129142	Chr 1 clone P0460C04	AP004366	e-178
35	BI305405	putative protein	AL390921	Chr 1 clone P0468B07	AP003260	0.0
36	BI306083	NA	NH	Chr 1 clone P0470A12	AC091088	e-154
37	BI305620	unknown protein	AY050948	Chr 1 clone P0471B04	AP003261	0.0
38	BI306648	peroxiredoxin	AF203879	Chr 1 clone P0485G01	AP003264	0.0
39	BI305524	fructose-1,6-bisphosphatase (cytosolic)	AB007193	Chr 1 clone P0505D12	AP003270	e-175
40	BI306255	NifU-like protein	AL021712	Chr 1 clone P0671D01	AP003284	e-130
41	BI306051	5 S ribosomal RNA	X64622	Chr 1 clone P0674H09	AP003349	2e-53
42	BI306592	putative 3-hydroxybutyryl-CoA dehydrogenase	NM_112392	Chr 1 clone P0699H05	AP003299	0.0
43	BI305530	NA	NH	Chr 1 clone P0699H05	AP003299	6e-85
44	BI306056	similarity to methyltransferase and transcriptional regulators	AL049481	Chr 1 clone:P0011D01	AP000969	0.0
45	BI306615	ribosomal protein L30	AF034949	Chr 1 clone:P0038F12	AP000836	e-116
46	BI305804	NA	NH	Chr 1 clone:P0499C11	AP001080	2e-37
47	BI306326	hypothetical protein	AP001081	Chr 1 clone:P0693B08	AP001081	e-146
48	BI306668	NA	NH	Chr 1 clone:P0708G02	AP001539	e-129
49	BI305815	polypeptide deformylase	AF269165	Chr 1 PAC clone:P0004A09	AP003607	3e-73

(**Table 2a,** *continued*)

	Accn No.[a]	Putative function[b]	Identical Accn No.[c]	BAC/PAC clone[d]	BAC/PAC clone Accn No.	BAC/PAC E-value[e]
50	BI305919	eukaryotic initiation factor 4A-3	X61206	Chr 1 PAC clone:P0004A09	AP003607	0.0
51	BI305981	group 4 late embryogenesis-abundant protein	M88321	Chr 1 PAC clone:P0006C01	AP002744	5e-41
52	BI306737	unknown protein	AP003223	Chr 1 PAC clone:P0007F06	AP003223	0.0
53	BI306739	unknown protein	AB028619	Chr 1 PAC clone:P0013F10	AP002523	e-130
54	BI305483	NA	NH	Chr 1 PAC clone:P0014E08	AP004194	0.0
55	BI305846	putative peroxisomal Ca-dependent solute carrier protein	AP002483	Chr 1 PAC clone:P0024G09	AP003311	2e-56
56	BI306432	unknown protein	AP003504	Chr 1 PAC clone:P0025A05	AP003504	e-128
57	BI306644	hypothetical protein	AP001072	Chr 1 PAC clone:P0025D05	AP001072	0.0
58	BI305601	40 S ribosomal protein S26	AF457935	Chr 1 PAC clone:P0031D02	AP003230	2e-28
59	BI305836	unknown protein	AP002881	Chr 1 PAC clone:P0035H10	AP002881	2e-56
60	BI306504	NA	NH	Chr 1 PAC clone:P0038D11	AP003234	0.0
61	BI305821	22 kDa protein of photosystem II	D84392	Chr 1 PAC clone:P0039A07	AP003235	0.0
62	BI306282	oligopeptide transporter	AP003235	Chr 1 PAC clone:P0039A07	AP003235	3e-95
63	BI306654	NA	NH	Chr 1 PAC clone:P0039A07	AP003235	e-101
64	BI305994	NA	NH	Chr 1 PAC clone:P0043B10	AP003236	e-106
65	BI305473	ribosomal protein L18a, cytosolic	D21301	Chr 1 PAC clone:P0046E05	AP003237	0.0
66	BI305571	malate dehydrogenase	AF195869	Chr 1 PAC clone:P0047B08	AP003053	2e-94
67	BI306411	glutathione S-transferase II	AF062403	Chr 1 PAC clone:P0403C05	AP003239	0.0
68	BI305667	putative GTP-binding protein	X63278	Chr 1 PAC clone:P0406H10	AP002524	4e-54
69	BI306049	60 S ribosomal protein L36 homologue	AL132960	Chr 1 PAC clone:P0408C03	AP003241	6e-53
70	BI306695	subtilisin-chymotrypsin inhibitor 2	Y08625	Chr 1 PAC clone:P0410E01	AP002866	0.0
71	BI305960	unknown protein	AP002866	Chr 1 PAC clone:P0410E01	AP002866	8e-59
72	BI305552	small GTP-binding protein (Ran1)	AB015971	Chr 1 PAC clone:P0410E03	AP002844	e-141
73	BI305911	NA	NH	Chr 1 PAC clone:P0415C01	AP003243	2e-28
74	BI305756	lysine decarboxylase-like protein	AB006700	Chr 1 PAC clone:P0415C01	AP003243	0.0
75	BI306729	unknown protein	AY039545	Chr 1 PAC clone:P0423A12	AP003246	6e-86
76	BI306329	response regulator 5	AB042267	Chr 1 PAC clone:P0431G06	AP003683	0.0
77	BI306138	unknown protein	AP003683	Chr 1 PAC clone:P0431G06	AP003683	0.0
78	BI306425	putative beta-glucosidase homologue	AP003272	Chr 1 PAC clone:P0432B10	AP003570	e-112
79	BI305763	triosephosphate isomerase (Rictipi2) gene	L04967	Chr 1 PAC clone:P0434B04	AP002540	0.0
80	BI305843	triosephosphate isomerase (Rictpi)	M87064	Chr 1 PAC clone:P0434B04	AP002540	4e-58
81	BI305638	unknown protein	AY065150	Chr 1 PAC clone:P0435B05	AP003249	e-161
82	BI306231	NA	NH	Chr 1 PAC clone:P0435B05	AP003249	0.0
83	BI306677	light-regulated protein	X68807	Chr 1 PAC clone:P0436E04	AP002818	4e-82
84	BI306548	16.9 kDa heat shock protein	AP003250	Chr 1 PAC clone:P0443D08	AP003250	e-105
85	BI305582	putative lipase	AF026480	Chr 1 PAC clone:P0456F08	AP002901	0.0
86	BI305978	hypothetical protein	AP002901	Chr 1 PAC clone:P0456F08	AP002901	0.0
87	BI305796	RicMT	AB002820	Chr 1 PAC clone:P0459B04	AP003627	e-174
88	BI305628	hypothetical protein	AP003256	Chr 1 PAC clone:P0460E08	AP003256	0.0
89	BI306132	NA	NH	Chr 1 PAC clone:P0470A12	AP003436	e-154
90	BI305583	ATP-dependent RNA helicase-like protein	AB008265	Chr 1 PAC clone:P0470A12	AP003436	e-123
91	BI305838	ribosomal protein L26	AF093540	Chr 1 PAC clone:P0480E02	AP002913	e-158
92	BI305742	NA	NH	Chr 1 PAC clone:P0481E12	AP003076	3e-15
93	BI306702	1-deoxy-D-xylulose 5-phosphate reductoisomerase	AF367205	Chr 1 PAC clone:P0482C06	AP002845	0.0
94	BI306379	peroxidase BP 1	M73234	Chr 1 PAC clone:P0483G10	AP003263	e-176
95	BI305484	lipid transfer protein	AF109195	Chr 1 PAC clone:P0485B12	AP003348	0.0
96	BI306561	NA	NH	Chr 1 PAC clone:P0489G09	AP002745	6e-20
97	BI305792	histone H4	X00043	Chr 1 PAC clone:P0506B12	AP003271	e-103
98	BI305390	glutaminyl-tRNA synthetase	P52780	Chr 1 PAC clone:P0510F03	AP002486	1e-54
99	BI305381	NA	NH	Chr 1 PAC clone:P0514H03	AP003275	5e-047
100	BI305521	ribosomal protein S4	Y15009	Chr 1 PAC clone:P0514H03	AP003275	e-100
101	BI306183	putative lipase	U38916	Chr 1 PAC clone:P0515G01	AP001633	e-118
102	BI306587	NA	NH	Chr 1 PAC clone:P0518C01	AP003277	1e-22

(**Table 2a,** *continued*)

	Accn No.[a]	Putative function[b]	Identical Accn No.[c]	BAC/PAC clone[d]	BAC/PAC clone Accn No.	BAC/PAC E-value[e]
103	BI306012	unknown protein	AP003277	Chr 1 PAC clone:P0518C01	AP003277	0.0
104	BI306292	type I light-harvesting chlorophyll a/b-binding protein	D00641	Chr 1 PAC clone:P0518F01	AP003278	e-117
105	BI306549	light harvesting chlorophyll a/b-binding protein	X13909	Chr 1 PAC clone:P0518F01	AP003278	e-123
106	BI306395	q group of receptor for activated C-kinase	D38231	Chr 1 PAC clone:P0519D04	AP003455	5e-20
107	BI305395	NA	NH	Chr 1 PAC clone:P0519D04,	AP003455	e-152
108	BI305819	putative 26 S proteasome subunit	AB070262	Chr 1 PAC clone:P0520B06	AP003077	e-106
109	BI306324	unknown protein	AP003213	Chr 1 PAC clone:P0520B06	AP003077	5e-83
110	BI306311	hypothetical protein	AC004146	Chr 1 PAC clone:P0520B06	AP003077	5e-51
111	BI305822	LOB domain 41	AF447895	Chr 1 PAC clone:P0520B06	AP003077	0.0
112	BI306497	ubiquitin (mub1) gene	M60175	Chr 1 PAC clone:P0537A05	AP002971	0.0
113	BI306002	NA	NH	Chr 1 PAC clone:P0554D10	AP002869	0.0
114	BI305427	hypothetical 12 k protein (trnA intron)	JQ0280	Chr 1 PAC clone:P0557A01	AP003280	0.0
115	BI305581	beta-tubulin 1	U76744	Chr 1 PAC clone:P0581F09	AP003631	3e-59
116	BI306285	unknown protein	AP002972	Chr 1 PAC clone:P0638D12	AP002972	0.0
117	BI306349	putative serine proteinase	AP003106	Chr 1 PAC clone:P0665A11	AP003106	0.0
118	BI305943	hypothetical protein	AP003106	Chr 1 PAC clone:P0665A11	AP003106	5e-50
119	BI305947	gigantea-like protein	AJ133787	Chr 1 PAC clone:P0666G04	AP003047	0.0
120	BI306510	putative protein	AL021710	Chr 1 PAC clone:P0678F11	AP003437	e-162
121	BI305890	NA	NH	Chr 1 PAC clone:P0681B11	AP003022	e-147
122	BI305958	unknown protein	AP003289	Chr 1 PAC clone:P0683F02	AP003289	e-169
123	BI305825	unknown protein	AP003023	Chr 1 PAC clone:P0684B02	AP003023	e-168
124	BI306124	vacuolar proton-translocating ATPase subunit E	U84268	Chr 1 PAC clone:P0684E06	AP003291	e-148
125	BI305968	hypothetical protein	AP002897	Chr 1 PAC clone:P0686E09	AP002897	3e-90
126	BI305625	floral homeotic protein HUA1	AY024357	Chr 1 PAC clone:P0686E09	AP002897	e-108
127	BI306752	cytochrome P450	AP002744	Chr 1 PAC clone:P0688A04	AP002839	0.0
128	BI305422	chlorophyll a/b-binding protein	U74295	Chr 1 PAC clone:P0690B02	AP003292	0.0
129	BI305408	PSST subunit of NADH: ubiquinone oxidoreductase	X82274	Chr 1 PAC clone:P0690B02,	AP003292	0.0
130	BI305588	NA	NH	Chr 1 PAC clone:P0699D11	AP002817	1e-50
131	BI306712	putative protein synthesis inhibitor II	AP002912	Chr 1 PAC clone:P0701D05	AP003301	1e-62
132	BI305642	phytochrome-associated protein	AF088281	Chr 1 PAC clone:P0710E05	AP002743	0.0
133	BI306394	unknown protein	AC013258	Chr 1 PAC clone:P0401G10	AP003238	0.0
134	BI305760	putative protein	AL034567	Chr 1 PAC clone:P0481E12	AP003076	e-157
135	BI306727	NA	NH	Chr 2 clone B1079H01	AP004766	e-133
136	BI306560	heat shock protein 169C (hsp169C), 3′-end	L14444	Chr 2 clone B1079H01	AP004766	0.0
137	BI305558	carbamoyl phosphate synthetase small subunit	U73175	Chr 2 clone OJ0003_C09	AP004076	e-155
138	BI305764	putative zinc finger protein	AF466199	Chr 2 clone OJ1448_G06	AP004853	e-165
139	BI305683	root-specific RCc3	L27208	Chr 2 clone OJ1001_D02	AP004037	0.0
140	BI305463	hypothetical protein	NM_103154	Chr 2 clone OJ1004_H01	AP004038	0.0
141	BI305498	expressed protein	NM_102910	Chr 2 clone OJ1007_D04	AP004150	e-147
142	BI306575	NA	NH	Chr 2 clone OJ1011_C09	AP004077	0.0
143	BI305755	succinate dehydrogenase subunit 3 (sdh3) gene	AF362741	Chr 2 clone OJ1020_C02	AP004078	0.0
144	BI305468	cyclophilin 2 (Cyp2) gene	L29469	Chr 2 clone OJ1020_C02	AP004078	0.0
145	BI305622	NA	NH	Chr 2 clone OJ1020_C02	AP004078	e-132
146	BI305572	small GTP-binding protein RACDP (RACD)	AF218381	Chr 2 clone OJ1020_C02	AP004078	2e-35
147	BI306125	shaggy-like kinase etha	Y13437	Chr 2 clone OJ1077_A12	AP003991	e-154
148	BI305578	gamma-tocopherol methyltransferase	AF213481	Chr 2 clone OJ1111_E07	AP003994	e-100
149	BI306105	NA	NH	Chr 2 clone OJ1112_G07	AP004156	4e-89

(**Table 2a,** *continued*)

	Accn No.[a]	Putative function[b]	Identical Accn No.[c]	BAC/PAC clone[d]	BAC/PAC clone Accn No.	BAC/ PAC E-value[e]
150	BI306222	NA	NH	Chr 2 clone OJ1113_G05	AP004018	e-123
151	BI305824	fibrillarin 2	NM_118695	Chr 2 clone OJ1119_A01	AP004020	2e-75
152	BI305758	sucrose-regulated, 3′-end sequence	U16257	Chr 2 clone OJ1136_C04	AP004026	0.0
153	BI305912	NA	NH	Chr 2 clone OJ1148_D05	AP004118	0.0
154	BI305665	NA	NH	Chr 2 clone OJ1175_B01	AP004159	e-110
155	BI306519	putative pumilio/Mpt5 family RNA-binding protein	NM_128471	Chr 2 clone OJ1202_E07	AP004048	0.0
156	BI306003	NA	NH	Chr 2 clone OJ1212_C01	AP004083	3e-74
157	BI305577	NA	NH	Chr 2 clone OJ1212_E12	AP004050	0.0
158	BI306402	annexin p35	X98245	Chr 2 clone OJ1288_G09	AP004119	0.0
159	BI306534	NA	NH	Chr 2 clone OJ1293_E04	AP004120	0.0
160	BI306584	putative protein	NM_123095	Chr 2 clone OJ1293_E04	AP004120	e-147
161	BI306111	60 S ribosomal protein L37	AF401593	Chr 2 clone OJ1293_E04	AP004120	2e-96
162	BI305841	NA	NH	Chr 2 clone OJ1297_C09	AP004087	e-124
163	BI306641	NA	NH	Chr 2 clone OJ1369_G08	AP004257	e-173
164	BI305600	possible apospory-associated protein	Z36546	Chr 2 clone OJ1399_D07	AP004138	0.0
165	BI305865	NA	NH	Chr 2 clone OJ1399_D07	AP004138	1e-52
166	BI306374	NA	NH	Chr 2 clone OJ1435_F07	AP004187	e-140
167	BI306198	putative kinetochore protein	AJ277096	Chr 2 clone OJ1435_F07	AP004187	2e-44
168	BI305883	NA	NH	Chr 2 clone OJ1486_E07	AP004139	e-144
169	BI306691	cinnamoyl CoA reductase	AJ428493	Chr 2 clone OJ1520_C09	AP004064	7e-86
170	BI305818	fumarylacetoacetate hydrolase-like protein	NM_101077	Chr 2 clone OJ1524_D08	AP004191	e-151
171	BI306185	NA	NH	Chr 2 clone OJ1626_B09	AP004069	e-112
172	BI306754	NA	NH	Chr 2 clone OJ1643_A10	AP004192	0.0
173	BI305504	NA	NH	Chr 2 clone OJ1705_E12	AP004070	e-108
174	BI306121	gamma hydroxybutyrate dehydrogenase	AY044183	Chr 2 clone OJ1712_E04	AP004144	e-162
175	BI305829	NA	NH	Chr 2 clone OJ2056_H01	AP004098	e-140
176	BI306468	NA	NH	Chr 2 clone OJ9003_G05	AP004126	e-130
177	BI305917	NA	NH	Chr 2 clone P0006C08	AP004683	0.0
178	BI306623	putative protein	AL353814	Chr 2 clone P0409F01	AP004748	0.0
179	BI306491	integral membrane protein	NM_105398	Chr 2 clone P0413A11	AP004771	0.0
180	BI305485	NA	NH	Chr 2 clone P0452F04	AP004776	9e-47
181	BI305743	18 S small subunit ribosomal RNA gene	AF069218	Chr 2 clone P0459B01	AP004778	0.0
182	BI305486	25 S ribosomal RNA gene	M11585	Chr 2 clone P0459B01	AP004778	0.0
183	BI306689	18 S ribosomal RNA gene	U42796	Chr 2 clone P0459B01	AP004778	0.0
184	BI306714	phosphoribulokinase	X51608	Chr 2 clone P0459B01	AP004778	0.0
185	BI306000	hypothetical protein	AP000367	Chr 2 clone:P0437H03	AP000367	0.0
186	BI305734	phosphoenolpyruvate carboxylase kinase	AF399915	Chr 2 OSJNBa0049O12	AC069158	8e-22
187	BI306412	NA	NH	chr 2 clone OJ1124_H01	AP004003	1e-70
188	BI306264	cytosolic glutamine synthetase	X14245	chr 2 clone P0487D09	AP004880	e-147
189	BI305607	NA	NH	Chr 3 BAC clone OSJNBa0010I09	AC084748	0.0
190	BI305570	Rubisco subunit binding-protein alpha subunit	X07851	Chr 3 BAC clone OSJNBa0033P04	AC092263	5e-20
191	BI305650	beta-D-glucan exohydrolase, isoenzyme ExoII	U46003	Chr 3 BAC clone OSJNBa0069E14	AC091811	0.0
192	BI306485	exoglucanase precursor	T04414	Chr 3 BAC clone OSJNBa0069E14	AC091811	e-134
193	BI305605	putative vesicle soluble NSF attachment protein receptor	AC082644	Chr 3 BAC OSJNBa0013M12	AC082644	0.0
194	BI305828	putative salt-induced protein	AC084295	Chr 3 BAC OSJNBa0015K03	AC084295	0.0
195	BI305502	brain-specific protein	D16140	Chr 3 BAC OSJNBa0018H01	AC087181	0.0
196	BI306092	unknown protein	AC087181	Chr 3 BAC OSJNBa0018H01	AC087181	0.0
197	BI305784	unknown protein	AC084404	Chr 3 BAC OSJNBa0026A15	AC084404	0.0
198	BI305946	translation initiation factor eIF-5A	AJ252135	Chr 3 BAC OSJNBa0040E01	AC079887	0.0

(**Table 2a,** *continued*)

	Accn No.[a]	Putative function[b]	Identical Accn No.[c]	BAC/PAC clone[d]	BAC/PAC clone Accn No.	BAC/PAC E-value[e]
199	BI306077	inosine monophosphate dehydrogenase	AF421559	Chr 3 BAC OSJNBa0091J19	AC084320	0.0
200	BI306079	actin-depolymerizing factor	AF112887	Chr 3 BAC OSJNBa0091J19	AC084320	e-158
201	BI305876	unknown protein	AC084320	Chr 3 BAC OSJNBa0091J19	AC084320	e-166
202	BI305833	NA	NH	Chr 3 BAC OSJNBb0033N16	AC082645	2e-52
203	BI305812	NA	NH	Chr 3 BAC OSJNBb0033N16	AC082645	0.0
204	BI305977	Rab28 protein	X59138	Chr 3 clone OJ1134F05	AC099401	e-180
205	BI305677	ubiquitin protein fused to a ribosomal protein	D12629	Chr 3 clone OJ1175C11	AC103891	e-105
206	BI305771	beta-glucosidase	U28047	Chr 3 clone OJ1212_C08	AC091670	0.0
207	BI306662	reversibly glycosylated polypeptide	Y18624	Chr 3 clone OJ1523_A02	AC090874	e-126
208	BI306210	reversibly glycosylated polypeptide	AF294725	Chr 3 clone OJ1523_A02	AC090874	0.0
209	BI306687	C2H2 zinc finger protein	AY077725	Chr 3 clone OJ1754_E06	AC104433	e-177
210	BI305487	Ras-like GTP-binding protein	NM_111825	Chr 3 clone OJ1754_E06	AC104433	e-133
211	BI306007	NA	NH	Chr 3 clone OJ1754_E06	AC104433	e-130
212	BI306492	NA	NH	Chr 3 clone OJ1781E12	AC105927	e-137
213	BI305440	UDP-glucuronic acid decarboxylase	AB079064	Chr 3 clone OSJNBa0002D01	AC083942	0.0
214	BI306356	UDP-glucuronic acid decarboxylase	AB079064	Chr 3 clone OSJNBa0002D01	AC083942	6e-54
215	BI305527	NA	NH	Chr 3 clone OSJNBa0008D12	AL607101	0.0
216	BI306269	possible apospory-associated protein	U13149	Chr 3 clone OSJNBa0015N08	AC096688	0.0
217	BI306726	enoyl CoA hydratase	AJ275305	Chr 3 clone OSJNBa0016B07	AF461424	e-150
218	BI305445	3-deoxy-D-arabino-heptulosonate 7-phosphate synthase	Y14797	Chr 3 clone OSJNBa0017N12	AC092075	0.0
219	BI305428	putative protein	AL031394	Chr 3 clone OSJNBa0022C08	AC097277	0.0
220	BI305569	alpha-tubulin	X91807	Chr 3 clone OSJNBa0024O21	AC104474	2e-49
221	BI306414	NA	NH	Chr 3 clone OSJNBa0027J18	AC096689	0.0
222	BI305799	expressed under carbonate stress	AB053296	Chr 3 clone OSJNBa0036E17	AC099041	3e-99
223	BI305623	NA	NH	Chr 3 clone OSJNBa0047E24	AC092556	9e-66
224	BI306467	putative strictosidine synthase-like	AL589883	Chr 3 clone OSJNBa0047E24	AC092556	e-119
225	BI306044	ribosomal protein S15	D10962	Chr 3 clone OSJNBa0052F07	AC104321	0.0
226	BI305493	NA	NH	Chr 3 clone OSJNBa0052F07	AC104321	0.0
227	BI305800	chloroplast ribosomal protein S22	X59270	Chr 3 clone OSJNBa0059G06	AC096690	0.0
228	BI306530	NA	NH	Chr 3 clone OSJNBa0059G06	AC096690	e-110
229	BI305505	gamma-Tip	D25534	Chr 3 clone OSJNBa0067N01	AC090485	0.0
230	BI305615	NA	NH	Chr 3 clone OSJNBa0090D11	AC105732	0.0
231	BI305516	ribulose-5-phosphate-3-epimerase	AF047444	Chr 3 clone OSJNBa0091P11	AC073556	e-152
232	BI306258	NA	NH	Chr 3 clone OSJNBb0006P09	AC104429	e-103
233	BI306110	NA	NH	Chr 3 clone OSJNBb0021G19	AC092076	0.0
234	BI305388	putative *N*2, *N*2-dimethylguanosine tRNA methyltransferase	AC009755	Chr 3 clone OSJNBb0024J04	AC084296	0.0
235	BI306204	NA	NH	Chr 3 clone OSJNBb0043C10	AC105733	0.0
236	BI305520	NA	NH	Chr 3 clone OSJNBb0043C10	AC105733	e-110
237	BI306722	NA	NH	Chr 3 clone OSJNBb0043P23	AC099324	e-136
238	BI306157	small subunit ribosomal RNA gene	AF161089	Chr 3 clone OSJNBb0079B16	AC092780	4e-97
239	BI306482	NA	NH	Chr 3 clone OSJNBb0081B07	AC093018	e-111
240	BI305560	ribosomal S29-like protein	AF457936	Chr 3 clone OSJNBb0094O03	AC092781	3e-27
241	BI306268	unknown protein	AC010657	Chr 3 clone OSJNBb0096M04	AC092559	0.0
242	BI306228	NA	NH	Chr 3 OSJNBa0002D01	AC083942	1e-33
243	BI305599	histone H2A	D38090	Chr 3 OSJNBa0002D01	AC083942	0.0
244	BI305429	putative protein	AL391143	Chr 3 OSJNBa0021B21	AC104179	0.0
245	BI306120	EREBP-like protein	AC079633	Chr 3 OSJNBa0032G08	AC079633	2e-83
246	BI305497	ethylene responsive protein (ebp-89 gene)	AJ304840	Chr 3 OSJNBa0032G08	AC079633	1e-22
247	BI305553	putative protein	AL022605	Chr 3 OSJNBa0048F08	AC091733	5e-78
248	BI305412	EF-1 alpha	D63583	Chr 3 OSJNBa0061L19	AC090484	0.0
249	BI306102	EF-1 alpha	D63580	Chr 3 OSJNBa0061L19	AC090484	0.0
250	BI305793	NA	NH	Chr 3 OSJNBa0067N01	AC090485	0.0
251	BI305801	putative gag-pol protein	AC084767	Chr 4 BAC clone B0311F12	AL512548	3e-90
252	BI305696	NA	NH	Chr 4 BAC cloneH0609A12	AL512544	e-108

(**Table 2a,** *continued*)

	Accn No.[a]	Putative function[b]	Identical Accn No.[c]	BAC/PAC clone[d]	BAC/PAC clone Accn No.	BAC/PAC E-value[e]
253	BI306436	*N*-hydroxycinnamoyl/benzoyl transferase	AL442115	Chr 4 BAC cloneH0711G06	AL442115	0.0
254	BI306609	NA	NH	Chr 4 BAC cloneH0806H05	AL442113	6e-24
255	BI306189	mitochondrial gene for 18 S rRNA	X00794	Chr 4 BAC clone:t17804	AL117265	0.0
256	BI305404	NA	NH	Chr 4 BAC clone:t17804	AL117265	2e-28
257	BI306457	cytosolic glyceraldehyde-3-phosphate dehydrogenase GAPDH	AF251217	Chr 4 clone H0302E05	AL627350	e-152
258	BI306291	NA	NH	Chr 4 clone OSJNBa0010D21	AL606635	0.0
259	BI306327	protein phosphatase 2C-like protein	NM_122403	Chr 4 clone OSJNBa0011F23	AL662953	0.0
260	BI305933	cell division protein FtsH-like protein	AC021640	Chr 4 clone OSJNBa0016O02	AL606588	e-161
261	BI305728	indole-3-glycerol phosphate synthase	AB023039	Chr 4 clone OSJNBa0016O02	AL606588	7e-97
262	BI305629	NA	NH	Chr 4 clone OSJNBa0017B10	AL606628	0.0
263	BI305855	HMG1 protein	AJ006708	Chr 4 clone OSJNBa0017B10	AL606628	1e-37
264	BI306117	elongation factor EF-2	AY054461	Chr 4 clone OSJNBa0020P07	AL606450	0.0
265	BI306384	unknown protein	AY045923	Chr 4 clone OSJNBa0029H02	AL606594	0.0
266	BI306131	NA	NH	Chr 4 clone OSJNBa0029H02	AL606594	5e-52
267	BI306388	jasmonate-induced protein	X98124	Chr 4 clone OSJNBa0033H08	AL662942	0.0
268	BI306466	expressed protein	NM_127785	Chr 4 clone OSJNBa0036B21	AL606636	0.0
269	BI306512	TMV-induced protein	AF242731	Chr 4 clone OSJNBa0036B21	AL606636	0.0
270	BI305474	NA	NH	Chr 4 clone OSJNBa0041A02	AL606638	e-151
271	BI305995	NA	NH	Chr 4 clone OSJNBa0043A12	AL606619	1e-82
272	BI305983	PBng143	AB020610	Chr 4 clone OSJNBa0043A12	AL606619	e-122
273	BI306017	NA	NH	Chr 4 clone OSJNBa0043A12	AL606619	3e-50
274	BI306360	ubiquitin-conjugating enzyme	D17786	Chr 4 clone OSJNBa0043A12	AL606619	0.0
275	BI305745	NA	NH	Chr 4 clone OSJNBa0043L24	AL662969	e-152
276	BI306037	NA	NH	Chr 4 clone OSJNBa0043L24	AL662969	1e-29
277	BI306058	glutaredoxin	X77150	Chr 4 clone OSJNBa0043L24	AL662969	e-142
278	BI305810	NA	NH	Chr 4 clone OSJNBa0045O17	AL663018	0.0
279	BI306288	putative copper amine oxidase	NM_129810	Chr 4 clone OSJNBa0053D18	AL606591	8e-66
280	BI306165	NA	NH	Chr 4 clone OSJNBa0053D18	AL606591	0.0
281	BI305803	symbiosis-related protein	NM_104884	Chr 4 clone OSJNBa0053K19	AL606645	e-126
282	BI306441	expressed protein	NM_114220	Chr 4 clone OSJNBa0060D06	AL606691	6e-81
283	BI306722	NA	NH	Chr 4 clone OSJNBa0060N03	AL606690	0.0
284	BI305935	DNA-binding protein, putative	NM_103643	Chr 4 clone OSJNBa0064H22	AL606448	e-161
285	BI306544	unknown protein	NM_111635	Chr 4 clone OSJNBa0067K08	AL606627	0.0
286	BI306353	quinone oxidoreductase-like protein	AL391141	Chr 4 clone OSJNBa0067K08	AL606627	0.0
287	BI305663	10 kDa phosphoprotein potential component of photosystem II	X12695	Chr 4 clone OSJNBa0079C19	AL662978	0.0
288	BI305725	expressed protein	NM_105860	Chr 4 clone OSJNBa0084A10	AL606458	0.0
289	BI305534	transporter-like protein	AL356014	Chr 4 clone OSJNBa0084K20	AL606613	e-158
290	BI305568	Glossy2 locus DNA	X88779	Chr 4 clone OSJNBa0085I10	AL606684	0.0
291	BI305769	NA	NH	Chr 4 clone OSJNBa0085I10	AL606684	e-164
292	BI305548	histone H3	AF109910	Chr 4 clone OSJNBa0086B14	AL606615	0.0
293	BI305706	NA	NH	Chr 4 clone OSJNBa0087O24	AL606646	5e-51
294	BI306067	OsCDPK7	AB042550	Chr 4 clone OSJNBa0088A01	AL662987	0.0
295	BI306723	histone H4	M12277	Chr 4 clone OSJNBa0088A01	AL662987	0.0
296	BI305753	NA	NH	Chr 4 clone OSJNBA0088I22	AL607001	0.0
297	BI305826	NA	NH	Chr 4 clone OSJNBa0094O15	AL662935	0.0
298	BI305391	3-oxyacyl-[acyl carrier protein] reductase	AJ243085	Chr 4 clone OSJNBB0006N15	AL607003	1e-75
299	BI306094	NA	NH	Chr 4 clone OSJNBb0012E08	AL606610	e-115
300	BI305565	partial ribosomal protein L7 gene	D29720	Chr 4 clone OSJNBb0015N08	AL662996	0.0
301	BI306589	glucose 6-phosphate/phosphate translocator, putative	NM_102035	Chr 4 clone OSJNBb0020J19	AL606656	0.0
302	BI306711	NA	NH	Chr 4 clone OSJNBB0026E15	AL607008	6e-78
303	BI306027	NA	NH	Chr 4 clone OSJNBb0026i12	AL663002	e-142
304	BI305478	22 kDa kafirin cluster	AC080019	Chr 4 clone OSJNBb0028M18	AL606654	6e-75
305	BI305699	remorin 1	AF123265	Chr 4 clone OSJNBb0039L24	AL663006	0.0
306	BI306393	NA	NH	Chr 4 clone OSJNBb0048E02	AL606653	0.0

(**Table 2a,** *continued*)

	Accn No.[a]	Putative function[b]	Identical Accn No.[c]	BAC/PAC clone[d]	BAC/PAC clone Accn No.	BAC/PAC E-value[e]
307	BI305872	NA	NH	Chr 4 clone OSJNBb0050O03	AL606631	0.0
308	BI306580	protein phosphatase	AY065090	Chr 4 clone OSJNBb0067G11	AL663014	0.0
309	BI306033	hydroxyproline-rich glycoprotein	X61280	Chr 4 clone OSJNBb0108J11	AL606618	2e-25
310	BI305708	class III chitinase homologue	AB027420	Chr 5 clone OJ1037G10	AC104270	0.0
311	BI305750	class III chitinase homologue (OsChib3H-h)	AB027426	Chr 5 clone OJ1037G10	AC104270	0.0
312	BI305853	putative mitochondrial carrier protein	AF372957	Chr 5 clone OJ1045C06	AC104272	0.0
313	BI305566	rd22	D10703	Chr 5 clone OJ1076H08	AC108498	0.0
314	BI306129	rd22 gene (D10703)	AP003560	Chr 5 clone OJ1076H08	AC108498	0.0
315	BI305776	ethylene responsive element binding factor3	AB036883	Chr 5 clone OJ1118C04	AC108523	0.0
316	BI306501	ADP-ribosylation factor	D17760	Chr 5 clone OJ1119H02	AC097175	e-145
317	BI305768	ribosomal protein L30 (rpl30)	AF034949	Chr 5 clone OJ1119H02	AC097175	e-131
318	BI305681	cytochrome P450	AB023038	Chr 5 clone OJ1174H11	AC104708	0.0
319	BI305591	proline-rich protein RiP-15	AF221552	Chr 5 clone OJ1212C10	AC104275	e-178
320	BI306524	ribosomal protein L35A	AF448416	Chr 5 clone OJ1214E03	AC104709	e-124
321	BI305603	putative protein	AL132959	Chr 5 clone OJ1214E03	AC104709	e-150
322	BI306476	putative protein	NM_120678	Chr 5 clone OJ1231F08	AC104276	e-144
323	BI305573	NA	NH	Chr 5 clone OJ1280A04	AC108500	0.0
324	BI305649	putative apoptosis related protein 19	AY045844	Chr 5 clone OJ1345B12	AC104278	e-149
325	BI305670	NA	NH	Chr 5 clone OJ1362G11	AC104713	0.0
326	BI305507	NA	NH	Chr 5 clone OJ1362G11	AC104713	0.0
327	BI305918	expressed protein	NM_130219	Chr 5 clone OJ1362G11	AC104713	e-151
328	BI305437	histone H4	M12277	Chr 5 clone OJ1387F08	AC108503	0.0
329	BI306527	NA	NH	Chr 5 clone OJ1504G04	AC105772	3e-56
330	BI306539	tonneau 2	F280057	Chr 5 clone OJ1504G04	AC105772	0.0
331	BI305584	U6 snRNA-associated Sm-like protein-like	AB012242	Chr 5 clone OJ1562H01	AC105773	1e-61
332	BI305492	NA	NH	Chr 5 clone OJ1576F01	AC097176	0.0
333	BI306418	R2R3MYB-domain protein	AF099432	Chr 5 clone OJ1654B10	AC108504	0.0
334	BI306300	putative legumin-like protein	AP003627	Chr 5 clone OJ1654B10	AC108504	0.0
335	BI306705	globulin-like protein	NM_100650	Chr 5 clone OJ1654B10	AC108504	e-162
336	BI305885	acetohydroxy acid isomeroreductase	AJ251333	Chr 5 clone OJ1735C10	AC104284	e-100
337	BI305543	mitochondrial malate dehydrogenase	P17783	Chr 5 clone OJ1735C10	AC104284	0.0
338	BI306331	putative CTP synthase	AP002897	Chr 5 clone OJ1735C10	AC104284	e-110
339	BI306174	NA	NH	Chr 5 clone P0008A07	AC079021	e-116
340	BI306659	metallothionein-like protein	U77294	Chr 5 clone P0016H04	AC079356	3e-44
341	BI305617	metallothionein-like protein	U57638	Chr 5 clone P0016H04	AC079356	0.0
342	BI305712	amino acid selective channel protein	AJ011921	Chr 5 clone P0016H04	AC079356	0.0
343	BI306005	expressed protein	NM_111934	Chr 5 clone P0419C04	AC084817	3e-85
344	BI305899	VIP2 protein	AJ251051	Chr 5 clone P0431G05	AC087551	0.0
345	BI305704	*S*-adenosylmethionine synthetase	AJ296743	Chr 5 clone P0519E07	AC087552	0.0
346	BI305442	putative tRNA synthase	AC079022	Chr 5 clone P0574H01	AC079022	0.0
347	BI305762	putative transcription factor IIA small subunit	AJ223634	Chr 5 clone P0574H01	AC079022	0.0
348	BI305631	cytochrome B5	X75670	Chr 5 clone P0574H01	AC079022	e-117
349	BI305945	NA	NH	Chr 5 clone P0676G05,	AC087425	e-121
350	BI305902	NA	NH	Chr 6 BAC clone OJ1540_H01	AC091774	0.0
351	BI306415	ribosomal protein S30 homologue	AL161574	Chr 6 BAC clone:OSJNBa0014B15	AP002854	0.0
352	BI306444	unknown	AF391808	Chr 6 BAC clone:OSJNBa0035I03	AP003019	e-166
353	BI305835	vacuolar H+pyrophosphatase	AB012766	Chr 6 BAC clone:OSJNBa0035I03	AP003019	0.0
354	BI306332	NA	NH	Chr 6 BAC clone:OSJNBa0041F13	AP003708	2e-49
355	BI305867	RING3-like bromodomain protein	AC004238	Chr 6 BAC clone:OSJNBa0041F13	AP003708	e-124
356	BI306553	NA	NH	Chr 6 clone OJ1008_D02	AP003938	e-164
357	BI305692	thionin	AF090836	Chr 6 clone OJ1536_A04	AP003959	0.0
358	BI306551	NA	NH	Chr 6 clone OSJNBb0026P21	AP004682	0.0
359	BI306555	partially homologous to chaperonin 10 gene	D29698	Chr 6 clone OSJNBb0026P21	AP004682	0.0

(**Table 2a,** *continued*)

	Accn No.[a]	Putative function[b]	Identical Accn No.[c]	BAC/PAC clone[d]	BAC/PAC clone Accn No.	BAC/PAC E-value[e]
360	BI305506	NAD-dependent formate dehydrogenase	AB019533	Chr 6 clone P0008F02	AP003518	e-168
361	BI306652	EF-hand Ca2+-binding protein CCD1	AF181661	Chr 6 clone P0009H10	AP003766	0.0
362	BI306443	catalase	D26484	Chr 6 clone P0017G10	AP004685	0.0
363	BI306486	DnaK-type molecular chaperone BiP-C – soybean	U08382	Chr 6 clone P0021C04	AP004687	0.0
364	BI305651	NA	NH	Chr 6 clone P0021H10	AP003520	0.0
365	BI305401	putative protein	NM_12301	Chr 6 clone P0036H01	AP004724	0.0
366	BI306343	hypothetical protein p85RF	AF139499	Chr 6 clone P0417G12	AP003711	0.0
367	BI306583	RNA-binding protein	NM_121073	Chr 6 clone P0436F11	AP003488	e-118
368	BI305580	NA	NH	Chr 6 clone P0436F11	AP003488	0.0
369	BI306541	putative protein	NM_125047	Chr 6 clone P0456F09	AP003762	7e-78
370	BI306328	wound-induced protein	X59882	Chr 6 clone P0462F09	AP004277	0.0
371	BI305410	NA	NH	Chr 6 clone P0470C02	AP003508	e-141
372	BI306483	NA	NH	Chr 6 clone P0486H12	AP003615	e-151
373	BI306245	unknown protein	NM_128223	Chr 6 clone P0486H12	AP003615	0.0
374	BI305831	sucrose synthase	Z15028	Chr 6 clone P0492A09	AP004697	3e-77
375	BI306369	peroxiredoxin Q	AB037598	Chr 6 clone P0492A09	AP004697	0.0
376	BI306424	glycine-rich RNA-binding protein grp1a, putative	NM_102006	Chr 6 clone P0523F01	AP003573	0.0
377	BI306753	NA	NH	Chr 6 clone P0531C01	AP003763	e-129
378	BI306431	NA	NH	Chr 6 clone P0542E10	AP003456	0.0
379	BI305948	NA	NH	Chr 6 clone P0556B08	AP004279	0.0
380	BI305630	NA	NH	Chr 6 clone P0583E12	AP004728	0.0
381	BI305970	xyloglucan endotransglycosylase-like protein	X93175	Chr 6 clone P0622F03	AP003771	0.0
382	BI306616	NA	NH	Chr 6 clone P0637D03	AP003633	0.0
383	BI306514	3-hydroxyisobutyryl-coenzyme A hydrolase	AL049608	Chr 6 clone P0652A05	AP004571	7e-81
384	BI305746	cyclophilin CYP5	AF020433	Chr 6 clone P0655A07	AP003634	0.0
385	BI305766	hypersensitivity-related gene	X95343	Chr 6 clone P0655A07	AP003634	2e-59
386	BI305621	expressed protein	NM_113124	Chr 6 clone P0686E06	AP003635	0.0
387	BI305923	putative protein translocation complex Sec61gamma chain	AY059131	Chr 6 clone P0709F06	AP003579	e-130
388	BI305640	similar to latex allergen	NM_129480	Chr 6 clone P0710B08	AP003728	0.0
389	BI306107	Similar to DREB1A	AB007787	Chr 6 clone:P0425F02	AP001168	2e-72
390	BI306078	RING-H2 finger protein RHA1a	AF078683	Chr 6 clone:P0514G12	AP000616	0.0
391	BI305540	NA	NH	Chr 6 clone:P0535G04	AP000399	3e-86
392	BI305557	ubiquinol--cytochrome c reductase	X79275	Chr 6 PAC clone:P0015E04	AP002069	e-150
393	BI305511	hypothetical protein	AL022580	Chr 6 PAC clone:P0541H01	AP001389	e-168
394	BI306208	60 kDa chaperonin beta subunit	Z68903	Chr 6 PAC clone:P0541H01	AP001389	1e-94
395	BI306276	pathogenesis-related protein	U20347	Chr 6 PAC clone:P0541H01	AP001389	e-151
396	BI305606	NA	NH	Chr 6 PAC clone:P0541H01	AP001389	1e-77
397	BI305433	Scl1 protein	AF149807	Chr 6 PAC clone:P0644B06	AP001129	0.0
398	BI306205	NA	NH	Chr 6 PAC clone:P0644B06	AP001129	3e-73
399	BI306302	putative RNA-binding protein	AC004561	Chr 6 PAC clone:P0675A05	AP002071	e-174
400	BI306642	unknown protein	AP002542	Chr 6 PAC clone:P0679C08	AP002542	0.0
401	BI306238	unknown protein	AB018109	Chr 6 PAC clone:P0679C08	AP002542	0.0
402	BI306179	NA	NH	Chr 7 clone OJ1008_F04	AP003939	e-132
403	BI306034	ribosomal protein S15	D10962	Chr 7 clone OJ1046_F10	AP003861	0.0
404	BI306617	expressed protein	NM_117530	Chr 7 clone OJ1046_F10,	AP003861	e-156
405	BI306213	aquaporin (PIP2a)	AF062393	Chr 7 clone OJ1047_A06	AP003802	e-164
406	BI305644	ribosomal protein L44	U64677	Chr 7 clone OJ1047_A06	AP003802	1e-91
407	BI305988	NA	NH	Chr 7 clone OJ1136_D11	AP003749	8e-66
408	BI305765	60 S ribosomal protein L27a	AB042856	Chr 7 clone OJ1154_D08	AP003814	0.0
409	BI306106	elongation factor 1 beta	D23674	Chr 7 clone OJ1154_D08	AP003814	e-122
410	BI305659	novel protein, osr40c1	X95402	Chr 7 clone OJ1240_G08	AP003929	8e-72
411	BI306274	expressed protein	NM_118229	Chr 7 clone OJ1316_A04	AP003822	e-128
412	BI305735	23 kDa polypeptide of photosystem II	AF052203	Chr 7 clone OJ1351_C05	AP004010	0.0

(**Table 2a,** *continued*)

	Accn No.[a]	Putative function[b]	Identical Accn No.[c]	BAC/PAC clone[d]	BAC/PAC clone Accn No.	BAC/ PAC E-value[e]
413	BI306429	ribosomal protein L41	D10406	Chr 7 clone OJ1372_D12	AP003827	e-129
414	BI306408	alpha-galactosidase-like protein	AL163763	Chr 7 clone OJ1409_C08	AP003757	e-162
415	BI305816	photosystem I chain IV precursor	Y00966	Chr 7 clone OJ1483_E04	AP003834	0.0
416	BI305386	histone H2A	D38091	Chr 7 clone OJ1582_D10	AP003838	0.0
417	BI306021	chlorophyll a/b-binding protein	AF058796	Chr 7 clone OSJNBa0006F16	AC087839	e-122
418	BI305984	unknown protein	AC084320	Chr 7 clone OSJNBa0087K02	AL607095	3e-99
419	BI306404	NA	NH	Chr 7 clone OSJNBa0087K02	AL607095	0.0
420	BI305925	NA	NH	Chr 7 clone OSJNBb0024A20	AC079038	e-128
421	BI306755	putative small nuclear ribonucleoprotein G	AC005170	Chr 7 clone OSJNBb0024A20	AC079038	e-113
422	BI305751	NA	NH	Chr 7 clone P0013G11	AP004261	e-154
423	BI305397	photosystem I PSI-K subunit	L12707	Chr 7 clone P0022E03	AP004263	0.0
424	BI305906	NA	NH	Chr 7 clone P0022E03	AP004263	2e-91
425	BI306416	nucleoside diphosphate kinase	D16292	Chr 7 clone P0038F10	AP004266	e-117
426	BI305415	NA	NH	Chr 7 clone P0039H02	AP004267	e-139
427	BI306148	similarity to protein translation inhibitor	AB024036	Chr 7 clone P0048D08	AP004269	e-138
428	BI305892	alpha 1 tubulin	Z11931	Chr 7 clone P0408B10	AP004271	e-116
429	BI305928	NA	NH	Chr 7 clone P0434A03	AP004299	0.0
430	BI305675	ribosomal protein S28	AJ001161	Chr 7 clone P0440B02	AP004313	e-122
431	BI306557	hypothetical protein	AP003223	Chr 7 clone P0453G03	AP004276	0.0
432	BI306155	unknown protein	AY059725	Chr 7 clone P0470D12	AP004300	e-116
433	BI305680	RNA polymerase II 136 kDa subunit	U28048	Chr 7 clone P0496D04	AP004670	e-113
434	BI305614	thioredoxin h	D26547	Chr 7 clone P0506C07	AP004384	e-146
435	BI305652	NA	NH	Chr 7 clone P0523A04	AP004340	0.0
436	BI306256	NA	NH	Chr 7 clone P0524G08	AP004671	e-144
437	BI305749	ribosomal protein	U86017	Chr 7 clone P0554D11	AP004569	e-146
438	BI306674	alanine aminotransferase-like protein	AY042902	Chr 7 clone P0585H11	AP004342	e-159
439	BI306295	cyclin D21 protein	AJ011892	Chr 7 clone P0594D10	AP004380	e-110
440	BI305450	translation initiation factor (GOS2)	AF094774	Chr 7 clone P0681F05	AP004674	e-152
441	BI305420	similarity to beta-1,3-glucanase-like protein	AB008265	Chr 7 clone P0711B07	AP004575	e-135
442	BI305464	mitochondrial ribosomal protein S14	AB017429	Chr 8 clone OJ1005_B05	AP003925	0.0
443	BI305448	manganese-binding protein PsbY precursor	AF060198	Chr 8 clone OJ1005_B05	AP003925	e-121
444	BI305526	NA	NH	Chr 8 clone OJ1113_A10	AP004643	e-104
445	BI306438	NA	NH	Chr 8 clone OJ1117_F10	AP003871	e-103
446	BI305878	NA	NH	Chr 8 clone OJ1117_F10	AP003871	1e-82
447	BI306060	GF14-c protein	U65957	Chr 8 clone OJ1124_B05	AP003881	5e-57
448	BI306103	papain-like cysteine protease	AF133839	Chr 8 clone OJ1150_A11	AP003928	e-124
449	BI305523	cp31BHv protein	AJ224324	Chr 8 clone OJ1150_A11	AP003928	2e-59
450	BI306673	glyceraldehyde-3-phosphate dehydrogenase	U31676	Chr 8 clone OJ1163_G08	AP003886	1e-65
451	BI306358	ferredoxin	AF010320	Chr 8 clone OJ1300_E01	AP003909	0.0
452	BI306681	tryptophan decarboxylase (EC 41128) DOPA	X67662	Chr 8 clone OJ1368_G08	AP003911	0.0
453	BI305664	NA	NH	Chr 8 clone OJ1381_H02	AP004164	0.0
454	BI305613	4-coumarate-CoA ligase	X52623	Chr 8 clone OJ1506_F01	AP004190	e-156
455	BI306350	NA	NH	Chr 8 clone OJ1506_F01	AP004190	4e-88
456	BI306522	aminotransferase 1	AY066012	Chr 8 clone OJ1506_F01	AP004190	0.0
457	BI305666	mitochondrial F0 ATP synthase D chain	AJ271469	Chr 8 clone OJ1666_A04	AP003917	0.0
458	BI306307	NA	NH	Chr 8 clone OJ1705_A03	AP003918	e-101
459	BI305657	NA	NH	Chr 8 clone OJ1734_E04	AP003919	2e-40
460	BI305503	cytochrome P450 monooxygenase CYP92A1	AY072297	Chr 8 clone P0005B10	AP004653	1e-79
461	BI306612	NADH dehydrogenase	AJ295997	Chr 8 clone P0045E02	AP004659	e-131
462	BI306688	NA	NH	Chr 8 clone P0047G03	AP004660	e-162
463	BI306089	aconitase-iron regulated protein 1 (IRP1)	D29629	Chr 8 clone P0047G03	AP004660	0.0

(**Table 2a,** *continued*)

	Accn No.[a]	Putative function[b]	Identical Accn No.[c]	BAC/PAC clone[d]	BAC/PAC clone Accn No.	BAC/PAC E-value[e]
464	BI306602	glyoxalase I	AB017042	Chr 8 clone P0047G03	AP004660	e-107
465	BI305894	NA	NH	Chr 8 clone P0048G02	AP004662	e-171
466	BI306312	ribonuclease, partial cds	AB052844	Chr 8 clone P0431A03	AP004666	0.0
467	BI305986	NA	NH	Chr 8 clone P0433E10	AP004667	0.0
468	BI306050	NA	NH	Chr 8 clone P0437G01	AP004690	0.0
469	BI306634	NADPH-cytochrome P450 oxydoreductase isoform 3	AF302498	Chr 8 clone P0437G01	AP004690	e-127
470	BI306601	monodehydroascorbate reductase	AF109695	Chr 8 clone P0443G08	AP004461	0.0
471	BI306299	NA	NH	Chr 8 clone P0451G12	AP004399	5e-47
472	BI305495	NA	NH	Chr 8 clone P0451G12	AP004399	0.0
473	BI306372	NA	NH	Chr 8 clone P0453D01	AP004691	e-148
474	BI306242	NA	NH	Chr 8 clone P0479C08	AP004617	0.0
475	BI306480	high mobility group I/Y-2	AF291748	Chr 8 clone P0479C08	AP004617	e-119
476	BI305648	NA	NH	Chr 8 clone P0498E12	AP004698	0.0
477	BI306627	RUB1 conjugating enzyme	AF202771	Chr 8 clone P0508B09	AP004631	0.0
478	BI306390	photosystem II 10 kDa polypeptide	U86018	Chr 8 clone P0556A11	AP004589	0.0
479	BI306639	putative 60 S ribosomal protein L37	NM_101393	Chr 8 clone P0582D05	AP004591	e-111
480	BI306682	putative snRNP splicing factor	AC007196	Chr 8 clone P0583B06	AP004619	e-168
481	BI306663	F12A2116	AL442114	Chr 8 clone P0682A06	AP004705	0.0
482	BI305387	NA	NH	Chr 8 clone P0686C03	AP004761	0.0
483	BI306593	pyrophosphate-F 6-phosphate 1-phosphotransferase-α-subunit	M55190	Chr 9 clone BAC0091I24	AC091687	e-112
484	BI306090	cold acclimation protein WCOR410b (Wcor410b)	U73210	Chr 9 clone OSJNBa0046G16	AC108756	8e-27
485	BI305595	heat shock protein 82 HSP82	D10427	Chr 9 clone OSJNBa0087J09	AC108761	0.0
486	BI305963	nucleoid DNA-binding protein	D26015	Chr 9 clone OSJNBa0087J09	AC108761	0.0
487	BI305633	unknown protein	AC004521	Chr 9 clone OSJNBa0087J09	AC108761	0.0
488	BI306657	hsp82 gene for heat shock protein 82	Z15018	Chr 9 clone OSJNBa0087J09	AC108761	e-170
489	BI305525	nucleoid DNA-binding protein cnd41-like protein	AL365234	Chr 9 clone OSJNBa0087J09	AC108761	e-131
490	BI305895	60 S ribosomal protein	AF140494	Chr 9 clone OSJNBb0004A05	AC108763	0.0
491	BI306147	putative Bci-5 protein	AJ250661	Chr 9 clone PAC0651G05	AC090055	0.0
492	BI305489	unknown protein	AB010069	Chr X clone OSJNBa0019H14	AL607096	0.0
493	BI306166	plastocyanin precursor	AF093636	Chr X clone OSJNBa0082A03	AL607097	0.0
494	BI306700	putative fatty acid elongase (fae2 gene)	AJ292770	Chr 10 BAC clone nbxb0032I20	AF229187	2e-44
495	BI306221	putative RING zinc finger protein	AY074275	Chr 10 BAC cloneOSJNBa0004P12	AC099040	0.0
496	BI306506	putative protein	AL392174	Chr 10 BAC cloneOSJNBa0029C15	AC087182	1e-57
497	BI305714	TGA-type basic leucine zipper protein TGA21	AF402608	Chr 10 BAC cloneOSJNBa0057L21	AC087599	0.0
498	BI306301	NA	NH	Chr 10 BAC nbxb0049A03	AC027661	1e-87
499	BI305501	polygalacturonase isoenzyme 1 beta subunit	AF251069	Chr 10 BAC nbxb0049A03	AC027661	0.0
500	BI305586	putative carnitine/acylcarnitine translocase	AB016882	Chr 10 BAC OSJNBa0003O19	AC060755	0.0
501	BI306647	carnitine/acylcarnitine translocase	AC060755	Chr 10 BAC OSJNBa0003O19	AC060755	0.0
502	BI306740	60 S ribosomal protein L17	AY054508	Chr 10 BAC OSJNBa0006L06	AC022457	e-119
503	BI305985	beta-expansin (EXPB7)	AF261275	Chr 10 BAC OSJNBa0010C11	AC069300	4e-35
504	BI306181	putative lipid transfer protein	AC026758	Chr 10 BAC OSJNBa0015J15	AC026758	0.0
505	BI305673	putative prolyl 4-hydroxylase, alpha subunit	AC068923	Chr 10 BAC OSJNBa0017E08	AC068923	e-140
506	BI305953	unknown protein	AC018929	Chr 10 BAC OSJNBa0027L23	AC018929	5e-28

(**Table 2a,** *continued*)

	Accn No.[a]	Putative function[b]	Identical Accn No.[c]	BAC/PAC clone[d]	BAC/PAC clone Accn No.	BAC/PAC E-value[e]
507	BI306515	NA	NH	Chr 10 BAC OSJNBa0027L23	AC018929	e-110
508	BI306535	hypothetical protein	AC084763	Chr 10 BAC OSJNBa0027P10	AC084763	1e-91
509	BI305458	nucleoside diphosphate kinase	AF271362	Chr 10 BAC OSJNBa0027P10	AC084763	e-161
510	BI306473	putative thioredoxin-like U5 small ribonucleoprotein	AC087182	Chr 10 BAC OSJNBa0029C15	AC087182	0.0
511	BI306011	putative CEO protein	AC027037	Chr 10 BAC OSJNBa0035H01	AC027037	0.0
512	BI305996	NA	NH	Chr 10 BAC OSJNBa0035H01	AC027037	0.0
513	BI306035	NA	NH	Chr 10 BAC OSJNBa0055P24	AC037425	e-160
514	BI306142	ubiquinone oxidoreductase subunit	AC018727	Chr 10 BAC OSJNBa0056G17	AC018727	e-170
515	BI306176	glycine decarboxylase subunit	AF022731	Chr 10 BAC OSJNBa0076F20	AC025296	5e-84
516	BI305608	NA	NH	Chr 10 BAC OSJNBa0079L16	AC026815	0.0
517	BI305731	putative RNA-binding protein	AC026815	Chr 10 BAC OSJNBa0079L16	AC026815	e-167
518	BI306649	membrane protein	L13655	Chr 10 BAC OSJNBb0028C01	AC079029	3e-24
519	BI306014	histone H4	M12277	Chr 10 BAC OSJNBb0064P21	AC073166	0.0
520	BI305471	NA	NH	Chr 10 BAC OSJNBb0089A17	AC079890	7e-57
521	BI305509	NA	NH	Chr 10 clone nbxb0018F16	AC025905	e-103
522	BI306376	ORF; able to induce HR-like lesions	U66269	Chr 10 clone nbxb0094K20	AC025907	0.0
523	BI305772	CYP18 gene	Y08273	Chr 10 clone OSJNBa0004E08	AC091724	0.0
524	BI306249	transcription factor BTF3	AF370253	Chr 10 clone OSJNBa0012L23	AC051632	0.0
525	BI306380	NA	NH	Chr 10 clone OSJNBa0031A07	AC084884	e-102
526	BI306066	minor allergen	Z99708	Chr 10 clone OSJNBa0041F04	AC026759	0.0
527	BI306141	membrane-associated protein	AC068923	Chr 10 clone OSJNBa0041P03	AC068950	0.0
528	BI305811	NA	NH	Chr 10 clone OSJNBa0049K09	AC074282	0.0
529	BI305713	anthocyanidin-3-glucoside rhamnosyltransferase	AB026639	Chr 10 clone OSJNBa0049K09	AC074282	0.0
530	BI306513	mitochondrial chaperonin-60	Z12115	Chr 10 clone OSJNBa0071K18	AC027038	e-125
531	BI306397	NA	NH	Chr 10 clone OSJNBa0073L20	AC099774	0.0
532	BI305535	NA	NH	Chr 10 clone OSJNBa0078O01	AC079888	e-105
533	BI306643	32 kDa protein jakalin homologue	AF021256	Chr 10 clone OSJNBb0004A06	AC099734	0.0
534	BI306032	NA	NH	Chr 10 clone OSJNBb0015K05	AC090870	0.0
535	BI305444	proline-rich protein	AC091665	Chr 10 clone OSJNBb0016M10	AC091665	0.0
536	BI306716	valyl-tRNA synthetase	NM_101328	Chr 10 clone OSJNBb0022I16	AC091238	1e-80
537	BI306024	class III chitinase	AF296279	Chr 10 OSJNBa0061K21	AC016780	0.0
538	BI306410	epimerase/dehydratase	NM_122767	Chr 10 OSJNBa0061K21	AC016780	0.0
539	BI305646	profilin (Hvpro1)	U49505	Chr 10 OSJNBa0065C16	AC074354	2e-19
540	BI306316	23 S ribosomal RNA	AF254866	Chr 10 OSJNBb0075K12	AC092750	2e-27
541	BI305844	45 S and 23 S ribosomal RNA	X01365	Chr 10 OSJNBb0075K12	AC092750	0.0
542	BI306542	Orf122	AF287482	Chr 10 OSJNBb0075K12	AC092750	e-107
543	BI306259	ORF85	X15901	Chr 10 OSJNBb0075K12	AC092750	e-157
544	BI306338	hypothetical protein	AJ271079	Chr 10 OSJNBb0075K12	AC092750	0.0
545	BI305891	hypothetical protein	AC078891	Chr 10 OSJNBb0081F12	AC090488	5e-71
546	BI306598	NA	NH	Chr 11 clone OSJNBb0005C17	AC112658	7e-51
547	BI306315	ATP citrate lyase b-subunit	AJ344108	Chr 11 clone OSJNBb0005C17	AC112658	e-157
548	BI305791	50 S ribosomal protein L4	AY072419	Chr 11 clone P0480H08	AC104847	1e-78
549	BI306099	arm repeat containing protein homologue	AY042791	Chr 12 clone OSJNBa0016C14	AL513403	e-114
550	BI305532	NA	NH	Chr 12 clone OSJNBa0016C14	AL513403	e-151
551	BI306697	chloroplast atpB for ATP synthase beta subunit	AB037543	Chr 12 clone OSJNBa0021D06	AL513004	e-135
552	BI305396	putative protein	NM_121022	Chr 12 clone OSJNBa0041K23	AL513404	2e-75
553	BI306498	NA	NH	Chr 12 clone OSJNBb0036J04	AL607103	0.0
554	BI306381	hypothetical protein	NM_105056	Chr12 clone OSJNBa0041K23	AL513404	0.0

[a] GenBank accession numbers of our ESTs generated earlier from leaf tissue of N22 rice line and analysed in this study. [b] Putative function assigned through BLASTN and BLASTX. NA, Not assigned. [c] GenBank accession numbers of homologous sequences. NH, No hit to the annotated sequences in the database. [d] Rice BAC/PAC clones to which the ESTs have been mapped. [e] E-value is a measure of the statistical significance of the alignment against BAC/PAC sequence; the lower the E-value, the higher the significance.

Table 2b. Annotation of ESTs from root cDNA library.

	Accn No.[a]	Putative function[b]	Identical Accn No.[c]	BAC/PAC clone[d]	BAC/PAC clone Accn No.	BAC/ PAC E-value[e]
1	BI305196	expressed protein	NM_128155	Chr 1 BAC clone:B1064G04	AP003924	0.0
2	BI305219	NA	NH	Chr 1 BAC clone:B1064G04	AP003924	3e-44
3	BI305202	aldolase	D10419	Chr 1 PAC clone:P0494A10	AP002541	e-150
4	BI305272	cytochrome P450	AB047400	Chr 1 PAC clone:P0688A04	AP002839	0.0
5	BI305262	putative lipase	AC007508	Chr 1 PAC clone:P0417G05	AP002835	e-172
6	BI305247	chloroplast RNA helicase VDL isoform 1	AF261020	Chr 1 PAC clone:P0518C01	AP003277	8e-96
7	BI305334	choline kinase	U43839	Chr 1 PAC clone:P0683B11	AP003288	e-131
8	BI305350	NA	NH	Chr 1 PACclone:P0028G04	AP003921	9e-77
9	BI305249	NA	NH	Chr 2 clone OJ1175_B01	AP004159	1e-88
10	BI305265	hypothetical protein	NM_106275	Chr 2 clone OJ1217_F02	AP004084	e-129
11	BI305325	methyljasmonate-inducible lipoxygenase 2	U56406	Chr 2 clone OJ1225_F07	AP004184	0.0
12	BI305328	lipoxygenase	AJ270938	Chr 2 clone OJ1225_F07	AP004184	0.0
13	BI305211	hydroxypyruvate reductase (HPR)	AB060810	Chr 2 clone OJ1435_F07	AP004187	5e-79
14	BI305192	NA	NH	Chr 2 clone OJ1435_F07	AP004187	2e-25
15	BI305352	*S*-adenosylmethionine decarboxylase 2	AJ251899	Chr 2 clone OJ1476_F05	AP004063	0.0
16	BI305252	molybdopterin synthase sulphurylase	AF124159	Chr 2 clone OJ1548_F12	AP004240	0.0
17	BI305213	heat stress transcription factor A3	AF208544	Chr 2 clone P0458B05	AP004777	0.0
18	BI305298	cytochrome P450 monooxygenase	AJ004810	Chr 2 clone P0459B01	AP004778	0.0
19	BI305222	putative ABC transporter	AC069158	Chr 2 OSJNBa0049O12	AC069158	0.0
20	BI305209	amino acid permease, putative	AC079041	Chr 3 clone OSJNBa0008D12	AL607101	0.0
21	BI305361	NA	NH	Chr 3 clone OSJNBa0015N08	AC096688	4e-57
22	BI305332	l13324	AJ243961	Chr 4 BAC l1332	AJ243961	0.0
23	BI305242	NA	NH	Chr 4 clone OSJNBa0052O21	AL606590	e-179
24	BI305369	histone H3	AF109910	Chr 4 clone OSJNBa0086B14	AL606615	0.0
25	BI305304	OsCDPK7	AB042550	Chr 4 clone OSJNBa0088A01	AL662987	0.0
26	BI305268	helicase-like transcription factor	AP003224	Chr 4 clone OSJNBa0089N06	AL662988	0.0
27	BI305289	NA	NH	Chr 4 clone OSJNBb0012E08	AL606610	e-109
28	BI305224	contains similarity to protein phosphatase-2c~gene	AB022217	Chr 4 clone OSJNBB0118P14	AL607005	0.0
29	BI305323	ethylene responsive element binding factor3	AB036883	Chr 5 clone OJ1118C04	AC108523	0.0
30	BI305256	vegetative storage protein	L20233	Chr 5 clone OJ1212C10	AC104275	e-166
31	BI305226	NA	NH	Chr 5 clone P0036D10	AC073405	0.0
32	BI305240	EF-hand Ca2+-binding protein CCD1	AF181661	Chr 6 clone P0009H10	AP003766	0.0
33	BI305317	copia-like retroelement pol polyprotein	AC005169	Chr 6 clone P0556B08	AP004279	0.0
34	BI305279	calcium-dependent protein kinase	AL133248	Chr 7 clone OJ1092_A07	AP003866	0.0
35	BI305199	ascorbate peroxidase	D45423	Chr 7 clone OJ1445_H10	AP003831	e-130
36	BI305275	root border cell-specific protein	AF139187	Chr 7 clone OJ1699_E05	AP003845	3e-49
37	BI305208	hypothetical protein	AC005662	Chr 7 clone P0496C02	AP004378	2e-97
38	BI305357	thioredoxin h	D26547	Chr 7 clone P0506C07	AP004384	e-105
39	BI305327	hypothetical protein	NM_127629	Chr 7 clone P0506H09	AP004400	1e-66
40	BI305273	cytochrome P450	X81831	Chr 8 clone OJ1300_E01	AP003909	0.0
41	BI305253	glyceraldehyde-3-phosphate dehydrogenase (Gpc)	U31676	Chr 8 clone OJ1499_A07	AP004166	6e-76
42	BI305233	fructose-bisphosphate aldolase	AJ133146	Chr 8 clone P0023G04	AP004374	0.0
43	BI305235	unknown protein	AB005242	Chr 8 clone P0456B03	AP004463	e-129
44	BI305318	nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase	X75326	Chr 8 clone P0528B09	AP004703	4e-80
45	BI305278	NA	NH	Chr 10 BAC OSJNBa0029C15	AC087182	e-145
46	BI305345	putative CEO protein	AC027037	Chr 10 BAC OSJNBa0035H01	AC027037	0.0
47	BI305181	NA	NH	Chr 10 clone OSJNAb0023M11	AC112514	e-131
48	BI305372	mitochondrion rrn26 gene for rRNA large subunit	Z11889	Chr 10 cloneOSJNBa0030B02	AC074105	0.0
49	BI305214	ORF185	X15901	Chr 10 clone OSJNBb0075K12	AC092750	0.0

[a] GenBank accession numbers of our ESTs generated earlier from root tissue of N22 rice line and analysed in this study. [b] Putative function assigned through BLASTN and BLASTX. NA, Not assigned. [c] GenBank accession numbers of homologous sequences. NH, No hit to the annotated sequences in the database. [d] Rice BAC/PAC clones to which the ESTs have been mapped. [e] E-value is a measure of the statistical significance of the alignment against BAC/PAC sequence; the lower the E-value, the higher the significance.

Table 3a. Leaf ESTs with putative functions mapped to WGS draft rice genome.

	Accn No.[a]	Putative function[b]	Identical Accn No.[c]	WGS contig[d]	E-value[e]
1	BI305950	50 S ribosomal protein L18	AF336922	AAAA01001147	e-139
2	BI306430	hypothetical protein	AJ012688	AAAA01000090	e-161
3	BI305393	ribosomal protein L29	AB042860	AAAA01010524	e-172
4	BI305518	transcriptional regulator	AF118223	AAAA01003689	e-156
5	BI305757	beta-amylase gene	L10346	AAAA01008982	e-147
6	BI305490	apospory-associated protein C-like	NM_126051	AAAA01000914	e-140
7	BI306517	beta-glucosidase isozyme 2 precursor	AY056828	AAAA01029659	0.0
8	BI306052	enolase	U09450	AAAA01004378	1e-52
9	BI305808	cytochrome P450-like sequence	AF088221	AAAA01005413	0.0
10	BI305417	elicitor-inducible cytochrome P450	AF368380	AAAA01003216	0.0
11	BI306386	vacuolar H+-ATPase (vatp-P1)	U27098	AAAA01002413	0.0
12	BI305850	calmodulin (CaM1)	AF042840	AAAA01015948	e-158
13	BI306475	calmodulin (CaM2)	AF042839	AAAA01000089	0.0
14	BI306631	myo-inositol monophosphatase 3	U39059	AAAA01000280	9e-41
15	BI305806	Nt-iaa28 deduced protein	AF123508	AAAA01001061	e-131
16	BI305653	ADP-ribosylation factor	X80042	AAAA01000560	e-121
17	BI305515	putative RAD23	NM_111121	AAAA01001258	3e-86
18	BI306038	putative eukaryotic translation initiation factor 3 subunit	AJ293728	AAAA01008134	0.0
19	BI305589	phenylalanine ammonia-lyase (EC 4315)	X16099	AAAA01007822	e-167
20	BI306528	positive element factor 1 (PF1)	L24390	AAAA01008657	1e-66
21	BI305709	GP28 gene (partial	Z15085	AAAA01002832	0.0
22	BI306365	glycine-rich RNA-binding protein	AJ302060	AAAA01002851	1e-72
23	BI306420	40 S ribosomal protein S3A	D26060	AAAA01008594	e-118
24	BI306254	glycine-rich protein	AF011331	AAAA01002851	0.0
25	BI306525	cyclophilin	AJ132763	AAAA01008298	0.0
26	BI306570	mitochondrial ATP synthase 6 kDa subunit	AB055076	AAAA01007916	e-110
27	BI306260	mitochondrial phosphate transporter	AB016065	AAAA01002624	0.0
28	BI305705	Dof zinc finger protein	AB028132	AAAA01004110	0.0
29	BI306362	zinc finger protein	AC079281	AAAA01023497	0.0
30	BI306250	glycine-rich protein (OSGRP1)	AF010579	AAAA01009071	e-133
31	BI305424	wound-induced basic protein	D30015	AAAA01033026	2e-71
32	BI306153	cytochrome c oxidase subunit 5c	AB027123	AAAA01011120	0.0
33	BI305862	peptidylprolyl cis-trans isomerase	X86903	AAAA01014778	e-117
34	BI306709	glyceraldehyde-3-phosphate dehydrogenase	M36650	AAAA01012992	4e-76
35	BI305402	*S*-adenosyl-L-homocysteine hydrolase	L11872	AAAA01000585	0.0
36	BI306290	Ca2+ sensitive 3′(2′),5-diphosphonucleoside 3′(2′) phosphohydrolase	U33283	AAAA01009369	e-165
37	BI306469	mitochondrial genes for NADH dehydrogenase subunits	D32052	AAAA01020774	e-125
38	BI305870	plasma membrane integral protein ZmPIP2-2	AF326491	AAAA01011307	e-112
39	BI306478	EPSPs, rps20 genes for 3-phosphoshikimate 1-carboxyvinyltransferase	AB052962	AAAA01000531	e-125
40	BI306279	putative senescence-associated protein	AB049723	AAAA01039016	e-124
41	BI306484	expressed protein	NM_100442	AAAA01002116	0.0
42	BI305508	putative protein	NM_114727	AAAA01022344	0.0
43	BI305632	putative protein	AL133315	AAAA01022344	0.0
44	BI306721	blight-associated protein p12 precursor	AF015782	AAAA01005435	e-155
45	BI306574	adh1-adh2 region	AF172282	AAAA01025101	e-109
46	BI306538	1-aminocyclopropane-1-carboxylate oxidase	X85747	AAAA01028331	0.0
47	BI305990	peroxidase	X98322	AAAA01028269	0.0
48	BI306152	26 S proteasome regulatory particle triple-A ATPase subunit 2b	AB037154	AAAA01005972	e-165
49	BI305500	putative coated vesicle membrane protein	NM_111647	AAAA01010440	e-150
50	BI306026	aldolase C-1	D50307	AAAA01028973	0.0
51	BI305736	transaldolase	U95923	AAAA01000199	e-164
52	BI305682	Sec61 alpha subunit	AY044237	AAAA01009041	0.0
53	BI306043	RNAase S-like protein	AY056038	AAAA01005125	0.0

(**Table 3a,** *continued*)

	Accn No.[a]	Putative function[b]	Identical Accn No.[c]	WGS contig[d]	E-value[e]
54	BI306573	glutathione-dependent dehydroascorbate reductase precursor	AF301597	AAAA01014616	e-175
55	BI306163	early nodulin	AB018377	AAAA01003189	e-146
56	BI306458	ubiquinol–cytochrome c reductase	X79276	AAAA01003222	e-178
57	BI305547	unknown protein	AB010700	AAAA01002357	0.0
58	BI305637	putative *N*-carbamyl-L-amino acid amidohydrolase	AY074343	AAAA01001368	e-161
59	BI305752	xyloglucan endotransglycosylase-related protein	U43487	AAAA01013384	0.0
60	BI305937	unknown protein	NM_102055	AAAA01006847	2e-93
61	BI306140	putative 6-phosphogluconate dehydrogenase, nuclear gene encoding putative plastid protein	AF061839	AAAA01006991	0.0
62	BI306278	unknown protein	AB018121	AAAA01003064	e-146
63	BI306016	small nuclear ribonucleoprotein	AC005724	AAAA01000535	e-130
64	BI306209	RING finger protein	Y09539	AAAA01001938	0.0
65	BI305739	abscisic acid- and stress-inducible protein (Asr1)	AF039573	AAAA01004684	0.0
66	BI306576	mitochondrial ribosomal protein S10	AB035348	AAAA01004333	0.0
67	BI306736	photosystem II D1 protein	D21291	AAAA01000758	e-119
68	BI306036	RuBisCO activase small isoforms	AB034748	AAAA01001004	e-165
69	BI305778	photoreceptor-interacting protein-like	AB013389	AAAA01002712	0.0
70	BI305786	thioredoxin h	AB053294	AAAA01001848	e-162
71	BI305574	ferredoxin	AB001386	AAAA01011395	0.0
72	BI306603	RNA polymerase subunit	AF266463	AAAA01001020	e-159
73	BI306059	OSMYB1	D88617	AAAA01000087	0.0
74	BI306554	serine carboxypeptidase	AC079632	AAAA01002789	e-153
75	BI305658	serine threonine kinase, putative	NM_099996	AAAA01000865	e-124
76	BI305554	seed maturation protein PM36	AF169021	AAAA01004382	0.0
77	BI305877	40 S ribosomal protein S19	P40978	AAAA01002116	6e-78
78	BI305467	ribosomal protein S8	U64436	AAAA01000403	0.0
79	BI305690	delta-type tonoplast intrinsic protein	U86763	AAAA01009344	0.0
80	BI306651	lipid transfer protein LPT III	AF017360	AAAA01018432	e-112
81	BI305398	metallothionein-like protein	AF009959	AAAA01001262	e-139
82	BI306339	protein transport protein subunit	NM_125439	AAAA01006368	e-155
83	BI306097	lipid transfer protein precursor	U29176	AAAA01005869	e-161
84	BI306344	lipid transfer protein LPT II	AF017359	AAAA01005869	0.0
85	BI306341	endosomal protein-like	AP000421	AAAA01021524	0.0
86	BI306352	lipid transfer protein LPT IV	AF017361	AAAA01028809	e-147
87	BI306717	cytosolic tRNA-Ala synthetase	AC091238	AAAA01003899	e-141
88	BI305514	photosystem I protein	M61146	AAAA01006595	0.0
89	BI306518	timing of CAB expression 1-like protein	AF272040	AAAA01012092	e-146
90	BI306046	mitochondrial ribulose bisphosphate carboxylase/oxygenase	L22155	AAAA01003705	0.0
91	BI306330	small subunit of ribulose-1, 5-bisphosphate carboxylase	D00643	AAAA01003705	0.0
92	BI305551	photosystem I subunit N	X66428	AAAA01002786	6e-66
93	BI305564	CP26	D85512	AAAA01000661	0.0
94	BI306661	chorismate synthase 2	Z21791	AAAA01003441	0.0
95	BI306085	hypothetical protein	AF114171	AAAA01004390	e-103
96	BI306751	18 S small subunit ribosomal RNA gene	U53380	AAAA01018453	0.0
97	BI305802	small GTP-binding protein (rab5A)	AY029301	AAAA01004497	0.0
98	BI306451	11-beta-hydroxysteroid dehydrogenase-like	AB023037	AAAA01004662	0.0
99	BI305837	small GTP-binding protein OsRac3	AB029510	AAAA01000231	0.0
100	BI306368	RNA polymerase I, II and III subunit RPB8	AF323605	AAAA01001020	e-149
101	BI305869	28 S ribosomal RNA gene	AY049041	AAAA01004991	2e-85
102	BI305767	expressed protein	NM_100905	AAAA01002293	e-159
103	BI305726	LMW heat shock protein precursor (hsp22)	AF035460	AAAA01002524	e-167
104	BI306310	NDR1/HIN1-like protein	AL589883	AAAA01002765	3e-80
105	BI306130	protein kinase, putative	AC027135	AAAA01000660	0.0
106	BI305561	small GTP-binding protein (ORRab-2)	L35845	AAAA01003359	0.0
107	BI306214	chaperonin 21 precursor	AF233745	AAAA01013252	0.0

(**Table 3a,** *continued*)

	Accn No.[a]	Putative function[b]	Identical Accn No.[c]	WGS contig[d]	E-value[e]
108	BI305794	ribosomal protein L17	D30026	AAAA01001010	e-121
109	BI305624	NAM (no apical meristem)-like protein	AC005312	AAAA01008859	2e-84
110	BI305618	chloroplast GrpE protein	NM_121777	AAAA01006836	0.0
111	BI305436	diacylglycerol kinase	D63787	AAAA01001561	0.0
112	BI306396	28 S ribosomal RNA gene	AY049041	AAAA01004991	3e-61
113	BI306507	putative coated vesicle membrane protein	AC009176	AAAA01010440	e-131
114	BI305479	25 S ribosomal RNA gene	M11585	AAAA01004991	0.0
115	BI305413	60 S ribosomal protein L9 induced by GA3	D83527	AAAA01013583	0.0
116	BI306594	*Hordeum vulgare* BAC 259I16	AF474373	NL	0.0

[a]GenBank accession numbers of our ESTs generated from leaf tissue of N22 rice line earlier and analysed in this study. [b]Putative function assigned through BLASTN and BLASTX. [c]GenBank accession numbers of homologous sequences. [d]Beijing Genomics Institute's whole genome shotgun *indica* rice genome draft sequence contigs to which the ESTs have been mapped. [e]E-value is a measure of the statistical significance of the alignment; the lower the E-value, the higher the significance.

Table 3b. Root ESTs with putative functions mapped to WGS draft rice genome.

	Accn No.[a]	Putative function[b]	Identical Accn No.[c]	WGS contig[d]	E-value[e]
1	BI305206	unknown protein	AC084221	AAAA01014668	0.0
2	BI305186	MEKK1/MAP kinase kinase kinase	AL161511	AAAA01003521	8e-91
3	BI305333	serine/threonine kinase	Y12465	AAAA01005657	3e-61
4	BI305315	CBL-interacting protein kinase 23	AY035226	AAAA01005657	5e-66
5	BI305216	putative protein kinase	AC005623	AAAA01005657	4e-39
6	BI305248	dehydrin	U60097	AAAA01012244	e-160
7	BI305212	respiratory burst oxidase protein D	AF055357	AAAA01008960	e-150
8	BI305245	unknown protein	AC004667	AAAA01020019	3e-84
9	BI305374	alanine:glyoxylate aminotransferase	AF251070	AAAA01004845	e-110
10	BI305259	nicotinamine aminotransferase A	D88273	AAAA01003256	0.0
11	BI305201	MAP3K beta 1 protein kinase	AJ010093	AAAA01003521	e-116
12	BI305302	putative 3′–5′ exoribonuclease	NM_106417	AAAA01007096	0.0
13	BI305215	plastid RNA polymerase sigma factor	AB005290	AAAA01004751	0.0
14	BI305346	unknown protein	AB024034	AAAA01011053	e-131
15	BI305336	ERD1 protein	D17582	NL	8e-25
16	BI305258	ATP-dependent Clp protease proteolytic subunit	NM_103884	AAAA01002736	e-116

[a]GenBank accession numbers of our ESTs generated earlier from root tissue of N22 rice line and analysed in this study. [b]Putative function assigned through BLASTN and BLASTX. [c]GenBank accession numbers of homologous sequences. [d]Beijing Genomics Institute's whole genome shotgun *indica* rice genome draft sequence contigs to which the ESTs have been mapped. [e]E-value is a measure of the statistical significance of the alignment; the lower the E-value, the higher the significance.

region and primarily represent 5′ and 3′ untranslated regions (UTR) of the genes. This is similar to the results of primary annotation of the draft *indica* rice genome, which indicate that approximately 50% of the predicted genes have no orthologous sequences (Yu *et al*. 2002).

Functional annotation of our nonredundant ESTs led to the definition of putative functions for 559 ESTs. The identified ESTs with putative functions cover major pathways associated with cellular metabolism (183) such as photosynthesis, amino acid metabolism, nucleic acid metabolism, protein synthesis, besides signal transduction pathways (16), transcription factors (17), cell defence (25) and stress response pathways (100). The present data will be useful in identifying orthologues in other grasses as well as paralogous sequences in the rice genome. In fact, the two rice whole genome sequencing projects—the Chinese (Yu *et al*. 2002) and the one by the Swiss company Syngenta (Goff *et al*. 2002)—and IRGSP have also used large sets of ESTs for annotation of the genome. Aligning ESTs on genomic sequences normally results in correct prediction of gene structure. One such example is the EST with GenBank accession number BI305796, which represents the coding sequence for a metallothionein-like protein. This full-length cDNA sequence is mapped to the chromosome 1 PAC clone P0459B04. However, the earlier annotation described it as an unknown protein since the gene prediction programs failed to identify the gene structure. One reason for this disparity is that the genes that encode metallothionein-like proteins have short exons (less than 100 bp), which escaped earlier annotations. Incidentally, genes for metallothionein-like proteins are the most abundant class in this collection of leaf ESTs. Five different classes of metallothionein are represented in this collection. The expression pattern of this class of genes has shown that their transcripts are more abundant in above-ground tissues such as mature leaf, panicle, endosperm and callus than in root tissue. This was also demonstrated earlier for rice metallothionein gene expression by Yu *et al*. (1998). Serial analysis of gene expression (SAGE) in rice clearly revealed the abundance of metallothionein gene transcripts (Matsumura *et al*. 1999). CLUSTALX sequence alignment between our EST sequences and database hit sequences uncovered full-length cDNA clones that have small coding regions for products such as metallothionein-like proteins, glutaredoxin and cytochrome P450.

The 3′-UTR regions of the EST clones have been used for gene-specific expression studies (Yazaki *et al*. 2000). A comprehensive rice transcript map has been constructed containing 6591 EST sites using 3′-end sequences of rice cDNA clones (Wu *et al*. 2002). Functional annotation of the full-length *Arabidopsis* cDNA Collection revealed important information on gene structure and pitfalls of gene prediction programs in the genome annotation (Seki *et al*. 2002). As more and more large-scale plant EST collections become available, gene prediction and functional definition will sharply improve across plant genomes. We are using ESTs as tools to characterize genes and their *cis* regulatory elements through mapping to genomic sequence (unpublished data from our laboratory). Access to the complete rice genome sequence will accelerate gene discovery and lead to a quantum jump in functional genomics.

We have described here a large number of rice ESTs, their predicted functions, and their localization on BAC/PAC clones and rice chromosomes. We have provided the GenBank accession numbers of our EST clones for ready reference to rice researchers. This comprehensive EST database will be periodically upgraded as we continue the sequencing and annotation of more clones from the leaf cDNA library from drought-stressed seedlings. These clones will be useful as reference markers and as probes in gene isolation and expression analysis for rice researchers in India and elsewhere. In particular, researchers working on genetic and molecular analysis of drought tolerance in rice may find this database useful. All ESTs are maintained as plasmid minipreps in the Plant Molecular Genetics and Functional Genomics Laboratory, University of Hyderabad.

Acknowledgements

We thank the Rockefeller Foundation, USA, for financial support to the Rice Functional Genomics Project of A. R. R. and fellowships to P. R. B., N. I. and A. C. S. We are grateful to Prof. Jeffrey L. Bennetzen and Dr W. Ramakrishna, Purdue University, USA, for their continued collaboration and interest in our work.

References

Altschul S. F., Madden T. L., Schäffer A. A., Zhang J., Zhang Z., Miller W. and Lipman D. J. 1997 Gapped BLAST and PSI-BLAST, a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402.

Bennetzen J. 2002 Opening the door to comparative plant biology. *Science* **296**, 60–62.

Bonaldo M. F., Lennon G. and Soares M. B. 1996 Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res*. **6**, 791–806.

Chen M., Presting G., Barbazuk W. B., Goicoechea J. L., Blackmon B., Fang G. *et al*. 2002 An integrated physical and genetic map of the rice genome. *Plant Cell* **14**, 537–545.

Cho R. J., Mindrinos M., Richards D. R., Sapolsky R. J., Anderson M., Drenkhard E. *et al*. 1999 Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nat. Genet.* **23**, 203–207.

Cooke R., Raynal M., Laudie M. and Delseny M. 1997 Identification of members of gene families in *Arabidopsis thaliana* by contig construction from partial cDNA sequences: 106 genes encoding 50 cytoplasmic ribosomal proteins. *Plant J.* **11**, 1127–1140.

Echenique V., Stamova B., Wolters P., Lazo G., Carollo V. and Dubcovsky J. 2002 Frequencies of Ty1-copia and Ty3-gypsy retroelements within the *Triticeae* EST databases. *Theor. Appl. Genet.* **104**, 840–844.

Epple P., Apel K. and Bohlmann H. 1997 ESTs reveal a multigene family for plant defensins in *Arabidopsis thaliana*. *FEBS Lett.* **400**, 168–172.

Fernandes J., Brendel V., Gai X., Lal S., Chandler V. L., Elumalai R. P. *et al*. 2002 Comparison of RNA expression profiles based on maize expressed sequence tag frequency analysis and micro-array hybridization. *Plant Physiol.* **1283**, 896–910.

Goff S. A., Ricke D., Lan T., Presting G., Wang R., Dunn M. *et al*. 2002 A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**, 92–100.

Harushima Y., Yano M., Shomura A., Sato M., Shimano T., Kuboki Y. *et al*. 1998 A high-density rice genetic linkage map with 2275 markers using a single F2 population. *Genetics* **148**, 479–494.

Kan Z., Rouchka E. C., Gish W. R. and States D. J. 2001 Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. *Genome Res.* **11**, 889–900.

Kawasaki S., Borchert C., Deyholos M., Wang H., Brazille S., Kawai K. *et al*. 2001 Gene expression profiles during the initial phase of salt stress in rice. *Plant Cell* **13**, 889–906.

Kurata N., Umehara Y., Tanoue H. and Sasaki T. 1997 Physical mapping of the rice genome with YAC clones. *Plant Mol. Biol.* **35**, 101–113.

Matsumura H., Nirasawa S. and Terauchi R. 1999 Transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression (SAGE). *Plant J.* **20**, 719–726.

Michalek W., Weschke W., Pleissner K.-P. and Graner A. 2002 EST analysis in barley defines a unigene set comprising 4,000 genes. *Theor. Appl. Genet.* **104**, 97–103.

Öztürk Z. N., Talamé V., Michalowski C. B., Gozukirmizu N., Tuberosa R. and Bohnert H. J. 2002 Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. *Plant Mol. Biol.* **48**, 551–573.

Reddy A. R., Ramakrishna W., Sekhar A. C., Ithal N., Babu P. R, Bonaldo M. F., Soares M. B. and Bennetzen J. L. 2002 Novel genes are enriched in normalized cDNA libraries from drought-stressed seedlings of rice (*Oryza sativa* L. subsp. *indica* cv. Nagina 22). *Genome* **45**, 204–211.

Seki M., Narusaka M., Abe H., Kasuga M., Yamaguchi-Shinozaki K., Carninci P., Hayashizaki Y. and Shinozaki K. 2001 Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses using a full-length cDNA microarray. *Plant Cell* **13**, 61–72.

Seki M., Narusaka M., Kamiya A., Ishida J., Satou M., Sakurai T. *et al.* 2002 Functional annotation of a full-length *Arabidopsis* cDNA Collection. *Science* **296**, 141–145.

Thompson J. D., Higgins D. G. and Gibson T. J. 1994 CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.

Thompson J. D., Gibson T. J., Plewniak F., Jeanmougin F. and Higgins D. G. 1997 The ClustalX Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res.* **24**, 4876–4882.

Wu J., Maehara T., Shimokawa T., Yamamoto S., Harada C., Takazaki Y. *et al.* 2002 A comprehensive rice transcript map containing 6591 expressed sequence tag sites. *Plant Cell* **14**, 525–535.

Yamamoto K. and Sasaki T. 1997 Large-scale EST sequencing in rice. *Plant Mol. Biol.* **35**, 135–144.

Yazaki J., Kishimoto N., Nakamura K., Fujii F., Shimbo K., Otsuka Y. *et al.* 2000 Embarking on rice functional genomics via cDNA Microarray: Use of 3′ UTR probes for specific gene expression analysis. *DNA Res.* **7**, 367–370.

Yu L.-H., Umeda M., Liu J.-Y., Zhao N.-M. and Uchimiya H. 1998 A novel MT gene of rice plants is strongly expressed in the node portion of the stem. *Gene* **206**, 29–35.

Yu J., Hu S., Wang J., Wong G. K., Li S., Liu B. *et al.* 2002 A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**, 79–92.

Received 25 April 2002