

Functional genomics of drought stress response in rice: Transcript mapping of annotated unigenes of an *indica* rice (*Oryza sativa* L. cv. Nagina 22)

Markandeya Gorantla¹, P. R. Babu¹, V. B. Reddy Lachagari¹, F. Alex Feltus², Andrew H. Paterson² and Arjula R. Reddy^{1,*}

¹Department of Plant Sciences, School of Life Sciences, University of Hyderabad, Hyderabad 500 046, India

²Plant Genome Mapping Laboratory, University of Georgia, Athens, Georgia 30602, USA

Rice being one of the widely cultivated cereals across diverse agroecological systems, is prone to high yield losses due to recurring droughts. In India, drought is a major constraint of rice production and accounts for as much as 15% of yield losses during some years. Conventional plant breeding techniques though cumbersome and time-consuming, have been immensely helpful in releasing drought-tolerant varieties. However, this is not adequate to cope up with the future demand for rice, as drought seems to spread to more regions and seasons across the country. Understanding the genes that govern rice plant architecture and response to drought stress is urgently needed to enhance breeding rice with improved drought tolerance. In order to identify genes associated with drought stress response and their temporal and spatial regulation, we took the genomic approach. By generating a large set of expressed sequence tags (ESTs) from cDNA libraries of drought-stressed seedlings and transcript profiling, we identified 589 genes presumed to be involved in drought stress. These 5814 ESTs are assembled into 2094 contigs and localized onto chromosome arms. We present here the physical map of the 2094 unigene set along with 589 annotated putative stress responsive genes of rice. Further, using ESTs, a few of drought quantitative trait loci (QTLs) have been dissected and putative candidate genes identified. This will be useful to rice researchers as ready reference source for breeding through developing candidate gene markers, molecular dissection of QTLs associated with drought stress and map-based cloning.

Keywords: Expressed sequence tags, physical mapping, QTL dissection, unigene.

RICE is an important cereal grown in about one-third of the world's total cereal crop area, providing staple food and 35–60% of the calories consumed by more than 2.7 billion people¹. However, rice consumes about 90% of the freshwater resources in Asia used for agriculture². About 80% of the

world's rice is grown under irrigated (55%) and rainfed lowland (25%) ecosystems, both of which depend on fresh-water resources. The limitation of water availability due to inadequate and erratic rainfall has contributed to as much as 15% loss in rice production during some years³. Drought stress has emerged as the major cause of rice yield instabilities across diverse crop-growing regions of Asia. Developing drought-tolerant rice lines by breeding was the most commonly used approach to combat the problem of drought stress-induced yield losses. However, breeding for drought tolerance in rice for such a complex trait has been rather slow because of lack of precise genetic and molecular information on associated genes and their regulation. Genomics has provided a new avenue to investigate and utilize allelic variation in target genomic segments associated with drought tolerance. A crucial step towards understanding the molecular genetic basis of drought tolerance in rice is identification of genes and their functions, which requires large-scale genomic and genetic resources.

With the complete genome sequence available, rice has emerged as the model crop system to understand the basis of complex traits such as yield, hybrid vigour, disease resistance and abiotic stress tolerance. Rice has become a powerful tool for comparative cereal genetic and genomic analysis in view of its small genome size (400–430 Mbp), high-resolution genetic maps, and well-established syntenic relationship with other agronomically important cereal species⁴. Numerous studies provided evidence for synteny between rice and other members of the grass family, which led to the suggestion that the grass genomes could be studied as a single genome^{5,6}. However, there is increasing evidence regarding differences among cereal genomes at the level of gene collinearity. Nevertheless, the rice system serves as a model for determining how genes interact in controlling complex traits in plants. The whole genome sequence analysis of rice has provided a valuable resource for gene prediction and analysis. However, identification of gene functions using such resources requires precise annotation tools. Though several gene-prediction algorithms have been used, no single

*For correspondence. (e-mail: arjuls1@uohyd.ernet.in)

algorithm is completely accurate and adequate⁷. As further reports on full-length cDNAs of rice become available, annotation of rice genome will be more accurate and precise.

With the advent of high-throughput genomic technologies, large amount of sequence information of rice has become available in different DNA databases (e.g. Genbank, EMBL, DDBJ). A major chunk of the data is on expressed sequence tags (ESTs) generated through large-scale cDNA sequencing projects. This has led to the identification of many novel genes from a wide range of species addressing tissue-specific gene expression⁸. EST resources have been extensively used to analyse changes in gene expression controlling physiological processes such as responses to biotic or abiotic stresses⁹.

The comprehensive physical map of rice genome based on large-insert bacterial artificial chromosome (BAC) and P1-derived artificial chromosome fingerprints has been constructed¹⁰. The Rice Genome Programme high-density genetic map¹¹ (<http://rgp.dna.affrc.go.jp/publicdata/geneticmap2000/index.html>) contains more than 3267 well-mapped genetic markers and has been integrated with the physical map. Many of these markers are conserved among the grass genomes, because they represent genes in the form of full-length cDNAs or ESTs. Further, the development of high-density maps constitutes an important step in the positional cloning of genes underlying complex traits¹². Utilizing the physical-genetic map of rice, high-resolution comparative physical maps of cereal species are being developed. This will facilitate map-based cloning of agronomically important genes¹⁰ in other crop species with considerably large genome sizes.

Comparison of the sequences of expressed genes in different plant species would allow identification of functional genes that are related by descent, though they might have become paralogous and have divergent functions. Transcript mapping of candidate genes is a powerful tool to elucidate the underlying mechanisms of drought tolerance in diverse crop species. Localization of trait-associated ESTs will lead to molecular dissection of (QTLs) through functional characterization of genes. In fact, localizing rice ESTs to rice physical map provides a direct route for gene discovery (as opposed to *in silico* gene prediction) and the elucidation of gene structure¹³.

In an effort to catalogue and categorize the expression of genes associated with drought stress response and to physically map them onto the rice genome, we have generated 5814 ESTs¹⁴ (G. Markandeya *et al.*, 2004, unpublished) by partial sequencing of randomly selected clones from a normalized cDNA library constructed from drought-stressed *indica* rice (Nagina 22; N 22) seedlings. N22 is a drought-tolerant *indica* rice cultivar that has been used as a donor parent in breeding for drought tolerance. This library served as a rich source of non-redundant cDNA clones and therefore is used on a continuous basis in our gene identification programme. We have assembled these ESTs into 2094 unigenes and identified 589 putative stress responsive genes

(Markandeya *et al.*, 2004, unpublished). We present here the physical chromosomal localization of 2094 unigenes along with 589 annotated putative stress responsive genes of rice. Sequence organization of these putative candidate genes for abiotic stress response in rice has been described. We expect this communication will be useful to rice researchers as a ready reference source for developing candidate gene markers, molecular dissection of QTLs associated with drought stress, map-based cloning and breeding of rice for drought tolerance.

Materials and methods

Sequence repositories and software resource used in EST analysis

ESTs are generated from drought-stressed seedlings of *O. sativa* ssp. *indica* cultivar N22 and deposited at dbEST division of NCBI (www.ncbi.nlm.nih.gov/dbest, GenBank accession numbers BI305180 to BI306756; BU672765 to BU673915; CB964418 to CB967504)¹⁴⁻¹⁶. Full-length cDNA sequences of possible candidate genes were derived from *Arabidopsis* expression profiling studies from The *Arabidopsis* Information Resource (TAIR – <http://www.arabidopsis.org>). The nucleotide, protein and EST databases at NCBI and TIGR were utilized for homology search using BLAST program¹⁷.

Phred¹⁸ and crossmatch were used for sequence processing like base-calling and vector-trimming. Homology search of the NCBI database was carried out using network client software with the DNATools interface (<http://www.crc.dk/dnatools>). Transcript mapping and localization of ESTs onto rice genome were done using cMap software¹⁹.

EST clustering, gene annotation and transcript mapping

CAP3 assembly algorithms²⁰ were used to assemble the individual ESTs into clusters of sequences deriving from the same transcript as tentative consensus sequences (TCs) and singletons representing unique transcripts. Homology search was done against non-redundant (nr) nucleotide and protein sequence databases using BLASTN 2.2.2 and BLASTX 2.2.2 versions of the BLAST programs¹⁷ through BLAST 2.0 network client software with the DNATools interface. Assembled N22 unigenes were aligned to the TIGR *japonica* rice assembly (ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_1.0) using BLASTN²¹. Unigenes were annotated if they met the following stringency criteria: *E*-value $\leq 1 \times 10^{-10}$; per cent identity $\geq 95\%$; and alignment length ≥ 50 bp. Functional annotations were then associated to mapped unigenes. These positions were displayed using cMap software¹⁹. Putative functions of the stress-responsive genes were also displayed along with the unigenes.

Results

From 8000 N22 cDNA clones, 5814 high quality ESTs were generated and deposited in the public domain through NCBI dbEST division. The raw sequences generated were base-called and the readable sequences of high quality with Phred score ≥ 20 were considered for further sequence analysis. The sequences were trimmed of leading primer, oligoDT adapter, poly-A tail. Poor-quality sequences were trimmed manually¹⁴⁻¹⁶ before submission to GenBank under accession numbers: BI305180-BI306756; BU672765-BU673915; CB964418-CB967504.

Classification of ESTs

Putative functions were identified by homology search against non-redundant protein and nucleotide databases. These ESTs were annotated and putative functions were attributed to each of them using BLASTX algorithm. Further, the ESTs involved in stress response were identified based on the putative functions and homology to known stress-responsive genes²² (Markandeya *et al.*, unpublished data).

Putative functions of the sequences have been categorized on the basis of cellular and physiological functions. To decipher the genes associated with drought-stress response, the available data from micro-array studies of *Arabidopsis*²³⁻²⁵ limited data from barley²⁶ and rice²⁷⁻²⁹ were utilized to compare the EST dataset using TBLASTX (E -value $\leq 1 \times 10^{-20}$) of different species and to categorize the stress responsive ESTs. Novel ESTs were identified by a search against dbEST division through BLASTN program. Chimeric clones and cDNA synthesis occurring from internal *NotI* site were observed during homology search against databases. These inherent problems of library construction and normalization are further complicated by the high GC content of rice genes.

CAP assembly and clustering analysis of ESTs

Assembly and clustering of N22 ESTs using CAP3 program²⁰ revealed 2094 (unigene set of sequences) unique transcripts represented by our ESTs. The assembly of 5814 sequences produced 1260 singletons and the remaining 4334 sequences were grouped into 834 contigs (Figure 1). Similar pattern of contigs was observed in NCBI unigene clustering, but more singleton sequences were assembled into contigs. Identification of transcripts that are highly represented among the analysed ESTs may provide information concerning processes important for acclimation to stress conditions.

Transcript mapping and chromosomal localization of N22 unigenes

The 2094 N22 unigenes were localized onto rice genome by a search against the TIGR *japonica* rice assembly through

BLASTN algorithm. Functional annotations were then associated to the mapped unigenes and displayed using cMap software¹⁹. Not all unigenes could be displayed in regions of high density due to pixel limitations of the image file (Figure 2). Further information and data will be available on request. Distribution of the identified chromosomal locations of ESTs among the 12 rice chromosomes was displayed. Annotations were abbreviated.

The chromosomal positions of unigenes and the location of identified stress-responsive genes are represented in Figure 3 *a-d*. The number of exons and exon sizes of these genes are given in Table 1. Gene organization analysis by aligning the unigene set sequences onto the genomic sequences revealed genes with single to twelve exons. Single-exon genes were mainly those having small coding sequences. Structural analysis of these 589 putative stress-responsive genes revealed extensive variation both in number and length of exons (20–1600 bp) and introns (37–2000 bp). We are currently analysing this structural information in detail to decipher the mechanisms of gene regulation governing the drought-stress response mechanisms in rice.

Putative known candidate genes spanning quantitative trait loci

Target QTL regions were extracted from reference genetic maps¹¹ and QTL studies associated with drought and yield traits³⁰⁻³². A few of the known putative stress-responsive genes identified in the QTL locations²² are given in Table 2.

Discussion

ESTs represent the largest constituent of DNA repositories in terms of sequence number and total nucleotide count. EST resources are exploited in genome annotation, gene discovery and comparative genomic analysis. ESTs

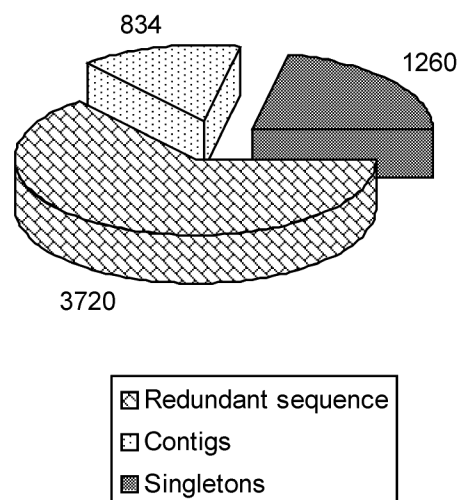


Figure 1. CAP3 assembly of 5814 ESTs.

Table 1. Physical location of N22 stress-responsive unigenes

Unigene	Putative function	CHR	NE	CHR_Start	CHR_End	Exon length
Contig197	33 kDa oxygen evolving protein of photosystem II	1	1	17160993	17160182	811
NL_13_49	Cellulose synthase Cesa-1	1	1	31213345	31212697	648
NL_9_92	Copper chaperone	1	1	17552370	17551808	562
Contig4	Ferredoxin	1	1	36975083	36974410	673
Contig473	Gigantea-like protein	1	1	4326219	4326637	418
NL_4_J01	Glutathione S-transferase II	1	1	31933083	31933609	526
NRS2R_1_O03GmCK2p		1	1	29630363	29630125	238
NL_6_54	Metallothionein-like protein type 2	1	1	2690421	2690033	388
NL_12_80	Pectinesterase	1	1	7427944	7428268	324
NL_1_N22	Photosystem II subunit (22 kDa) precursor	1	1	37432570	37431931	639
Contig540	PKF1	1	1	5507370	5507800	430
Contig246	Protein phosphatase 2C	1	1	36076874	36076404	470
NL_4_E16	Serine proteinase	1	1	32241424	32240776	648
NL_2_M24	Similar to lipase	1	1	15059074	15059278	204
NL_2_K13	Similar to RING-H2 finger protein RHA1a	1	1	38011907	38012018	111
Contig175	Sterol-C5(6)-desaturase	1	1	1889020	1888525	495
NL_5_M06	Subtilisin-chymotrypsin inhibitor 2	1	1	24168977	24168588	389
Contig600	Ubiquitin (mub1)	1	1	12567528	12566876	652
NL_3_D12	Wound induced protein homolog	1	1	1757007	1757273	266
NL_5_C09	16.9 kDa heat shock protein	1	1	1938369	1938175	194
NLP_1_E15	Acyl-CoA: 1-acylglycerol-3-phosphate acyltransferase	1	1	32939253	32939774	521
NL_12_60	Amino acid permease	1	1	39215922	39215688	234
Contig723	AP2 domain protein homolog	1	1	11705584	11705458	126
Contig302	dnaJ-like protein	1	1	20764665	20764935	270
NL_4_O24	dTDP-Glucose 4-6-dehydratase	1	1	42373576	42374130	554
Contig445	Iron(III)-zinc(II) purple acid phosphatase precursor	1	1	32633164	32633748	584
NL_0_A19	Lipid transfer protein	1	1	34862285	34862640	355
Contig613	Lysine decarboxylase-like protein	1	1	22730619	22731055	436
Contig504	Peroxidase BP 1	1	1	42064470	42064331	139
Contig14	Sucrose-6F-phosphate phosphohydrolase SPP3	1	1	15495362	15495640	278
Contig36	ATP-dependent RNA helicase-like protein	1	1	39339348	39339511	163
Contig449	Chloroplast carbonic anhydrase	1	1	25480854	25480767	87
Contig799	Fructose-1,6-bisphosphatase (cytosolic)	1	1	37252850	37252549	301
Contig808	Auxin-regulated protein (Aux28)	1	1	4072487	4072432	55
Contig783	Histone H4	1	1	35573196	35572992	204
Contig280	Chloroplast apocytochrome b6 (petB)	1	1	33283195	33282914	281
Contig61	ABC transporter	2	1	33958062	33958391	329
NL_4_I11	Annexin p35	2	1	31325252	31325739	487
Contig47	Arabinogalactan-like protein	2	1	11964480	11965179	699
NRS2R_1_FC	C3HC4-type RING zinc finger protein	2	1	31597138	31596948	190
Contig526	Dof zinc finger protein	2	1	29837894	29838252	358
NL_4_E04	Endosomal protein-like	2	1	33590253	33590733	480
Contig509	GP28	2	1	24622905	24623297	392
Contig170	GrpE protein	2	1	23761022	23761378	356
Contig76	Lipoxygenase	2	1	5226704	5226040	664
Contig354	Major intrinsic protein	2	1	7468895	7469277	382
NL_4_N20	Phosphoshikimate 1-carboxyvinyltransferase	2	1	12566220	12566360	140
NRS2R_1_P07	S-adenosylmethionine decarboxylase 2	2	1	23716218	23715582	636
NL_0_K03	Small GTP-binding protein (ORRab-2)	2	1	22314180	22313818	362

(contd...)

RESEARCH ARTICLES

Table 1. (contd...)

Unigene	Putative function	CHR	NE	CHR_Start	CHR_End	Exon lengths
Contig123	S-ribonuclease binding protein SBP1	2	1	1529662	1530259	597
NL7_G10	Argonaute protein	2	1	26966326	26966780	454
NL_5_M01	Cinnamoyl CoA reductase	2	1	34344046	34344301	255
NL49_D12	Cytochrome b5	2	1	25795706	25795638	68
Contig487	Cytochrome P450 monooxygenase	2	1	28338691	28338810	119
NL_15_8	Defensin	2	1	24826222	24826110	112
NL_3_J02	Early nodulin	2	1	7089600	7089662	62
Contig56	Heat stress transcription factor Spl7	2	1	19061804	19061604	200
Contig469	Phosphoenolpyruvate carboxylase kinase	2	1	24568976	24569735	759
NL_5_N17	Phosphoribulokinase	2	1	28326870	28326751	119
NL_5_A10	Pumilio/Mpt5 family RNA-binding protein	2	1	34750324	34750235	89
Contig661	Shoot GS1 for cytosolic glutamine synthetase	2	1	30297864	30298133	269
Contig470	Plastocyanin precursor	2	1	112061	112174	113
NL_1_P12	Small GTP-binding protein OsRac3	2	1	30711012	30711078	66
NL_12_32	ABC transporter family protein	2	1	34221596	34221461	135
Contig88	Auxin-regulated protein	2	1	35369658	35369943	285
NL_2_H17	Translocation complex Sec61gamma chain	2	1	4287267	4287359	92
Contig576	Water channel protein	2	1	24799243	24799139	104
Contig783	Histone H4	2	1	27617865	27618069	204
NL_24_11	Actin	3	1	27953160	27953222	62
NL_4_J14	Cyc07	3	1	5244178	5244110	68
NL14_F12	H ⁺ -transporting ATP synthase chain 9	3	1	9450893	9451421	528
NL36_D05	Nicotianamine synthase 2	3	1	10887027	10887213	186
NL_1_L04	Photoreceptor-interacting protein-like	3	1	28960019	28960353	334
NL_1_O09	Salt-induced protein, lectin	3	1	15837687	15837580	107
NL_1_G06	Sec61 alpha subunit	3	1	5856159	5855713	446
Contig372	Transcription factor Hap5a-like protein	3	1	7935867	7935295	572
NL_3_I13	26S proteasome regulatory particle triple-A ATPase subunit2b	3	1	10427411	10427265	146
Contig599	Adenylate kinase	3	1	1698539	1698481	58
Contig67	Amino acid permease, putative	3	1	20515128	20515435	307
Contig191	Beta-expansin (EXPB7)	3	1	169249	169197	52
NL15_B08	Chalcone isomerase	3	1	33013257	33012915	342
Contig395	Gamma-Tip	3	1	2544622	2545458	836
NL_12_16	IAA1 protein	3	1	29115446	29115499	53
NL_3_H02	Protein kinase, putative	3	1	8792282	8792145	137
Contig241	S18.A ribosomal protein	3	1	31670166	31669905	261
NL_4_N02	Strictosidine synthase-like	3	1	29561626	29561524	102
Contig549	Vesicle soluble NSF attachment protein receptor	3	1	14174872	14175249	377
Contig494	Zinc-finger-like protein	3	1	33046446	33046128	318
Contig719	Histone-like protein	3	1	31924004	31923901	103
NLP_1_G22	3-Deoxy-D-arabino-heptulosonate 7-phosphate synthase	3	1	15228517	15228876	359
Contig416	Actin depolymerizing factor	3	1	30983305	30983457	152
Contig178	Beta-glucosidase	3	1	27115305	27115822	517
NL_0_C19	Brain specific protein	3	1	27542944	27542819	125
Contig569	Beta-D-glucan exohydrolase, isoenzyme ExoII	3	1	29490502	29490989	487
Contig636	SR3 sucrose-regulated mRNA, 3'-end sequence	3	1	1450324	1450426	102
Contig808	Auxin-regulated protein (Aux28)	3	1	23371767	23371705	62
Contig783	Histone H4	3	1	1060860	1060660	200
Contig806	Chitinase-B	4	1	23683287	23683884	597

(contd...)

Table 1. (contd...)

Unigene	Putative function	CHR	NE	CHR_Start	CHR_End	Exon lengths
NL_4_A22	Copper amine oxidase	4	1	22821230	22821378	148
NL_3_G11	Elongation factor EF-2	4	1	1016986	1016564	422
Contig443	Glucose 6 phosphate/phosphate translocator, putative	4	1	34401639	34402111	472
NL_3_A20	Hydroxyproline-rich glycoprotein	4	1	19821585	19821517	68
Contig590	Plasma membrane major intrinsic 2 protein	4	1	8412989	8412555	434
Contig560	Quinone oxidoreductase-like protein	4	1	23849160	23848676	484
NL_13_44	Ascorbate peroxidase (TL29)	4	1	29376090	29375950	140
NL_2_I15	DNA binding protein, putative	4	1	21442320	21442225	95
Contig249	Fatty acid elongase-like protein	4	1	29986161	29986002	159
NL_24_12	gt-2	4	1	26065951	26065846	105
NL_2_P05	Heavy-metal-associated domain-containing protein	4	1	33062093	33062195	102
Contig68	Helicase-like transcription factor	4	1	30995206	30995674	468
Contig813	Jasmonate-induced protein	4	1	13244675	13245087	412
Contig404	Pollen allergen-like protein	4	1	22268197	22268014	183
Contig114	Transcription factor GT-3b	4	1	29382428	29381946	482
NL40_A04	Beta-carotene hydroxylase	4	1	28145854	28146197	343
Contig562	Cell division protein FtsH-like protein	4	1	22298877	22298740	137
Contig834	hmgcl	4	1	27281786	27281705	81
Contig576	Water channel protein	4	1	25087978	25087865	113
Contig66	Cytosolic glyceraldehyde-3-phosphate dehydrogenase GAPDH	4	1	23296354	23296452	98
Contig783	Histone H4	4	1	28477482	28477851	369
Contig186	Acetohydroxy acid isomeroreductase	5	1	27197708	27197894	186
NL_16_8	Glutaredoxin	5	1	2697529	2697888	359
NL_2_A13	Mitochondrial carrier protein	5	1	15810499	15810960	461
NL7_F02	Trehalose-6-phosphate synthase	5	1	24408996	24408688	308
NL_11_34	Unknown cold induced protein	5	1	24759821	24759238	583
Contig409	Disulfide isomerase A6 precursor (P5)	5	1	3129077	3129465	388
Contig475	Class III chitinase homologue (OsChib3H-h)	5	1	8263768	8264203	435
Contig706	Cytochrome B5	5	1	465090	465308	218
Contig153	Plasma membrane associated protein	5	1	18294090	18294026	64
NL32_E07	Protein kinase family	5	1	18215428	18215519	91
Contig772	R2R3MYB-domain protein	5	1	771224	771324	100
Contig496	Nonspecific lipid transfer protein	5	1	25970757	25970857	100
Contig216	ADP-ribosylation factor	5	1	22806969	22807267	298
NL_1_H24	Amino acid selective channel protein	5	1	570849	570797	52
Contig252	Chitinase	5	1	8263767	8263073	694
NL_4_H24	Guanine nucleotide-binding protein beta subunit	5	1	26099852	26099979	127
NL22_D06	Nuclear RNA binding protein A	5	1	27981214	27981280	66
Contig204	Oligopeptide transporter	5	1	19994684	19994736	52
Contig277	Gibberellin-20 oxidase (Sd-1)	5	1	19501251	19501320	69
Contig769	L24 ribosomal protein	5	1	22690197	22690116	81
NL_2_E21	VIP2 protein	5	1	3023917	3024451	534
Contig340	Phosphoethanolamine methyltransferase	5	1	25871773	25871662	111
Contig636	SR3 sucrose-regulated mRNA, 3'-end sequence	5	1	17732165	17732277	112
NL_13_48	Dehydrin-like protein	5	1	24846106	24846038	68
Contig77	Lipase	5	1	129354	129716	362
Contig265	Ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit	5	1	19783209	19783645	436

(contd...)

RESEARCH ARTICLES

Table 1. (contd...)

Unigene	Putative function	CHR	NE	CHR_Start	CHR_End	Exon lengths
NL_3_L24	60 kDa Chaperonin Beta subunit	6	1	776296	776084	212
Contig311	60S ribosomal protein L31-1	6	1	12159855	12160192	337
NL_1_G15	Delta-type tonoplast intrinsic protein	6	1	13159618	13159055	563
Contig701	Leaf-specific thionin precursor	6	1	17797496	17797575	79
NL_1_D15	Lipid acyl hydrolase	6	1	27057463	27056804	659
Contig26	RING-H2 finger protein RHA1a	6	1	497668	498125	457
Contig159	Scl1 protein	6	1	329169	329643	474
NL_1_O14	Sucrose synthase	6	1	4769083	4768936	147
NL23_F10	Thaumatococcus-like protein	6	1	27781479	27782108	629
NL_4_D06	Wound induced protein	6	1	27439953	27440569	616
NL47_G04	Amino acid transporter family	6	1	24638853	24638666	187
NL_4_J23	Glycine-rich RNA-binding protein grp1	6	1	24861058	24860863	195
Contig334	Luminal binding protein 2 precursor (BiP2)	6	1	5694870	5694172	698
NL37_G02	Mitogen-activated protein kinase	6	1	2761008	2761531	523
Contig470	Plastocyanin precursor	6	1	112190	112694	504
NL50_B10	40S subunit ribosomal protein	6	1	1795853	1795506	347
NL_1_J08	Cyclophilin CYP5	6	1	28815373	28815289	84
Contig223	Glutathione dependent dehydroascorbate	6	1	6859085	6859399	314
NL_1_P05	Low molecular weight heat shock protein precursor (hsp22)	6	1	6111230	6111309	79
Contig208	Glycine-rich protein	6	1	23496130	23495431	699
Contig636	SR3 sucrose-regulated mRNA, 3'-end sequence	6	1	19258285	19258415	130
Contig808	Auxin-regulated protein (Aux28)	6	1	22487523	22487578	55
Contig265	Ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit	6	1	22574476	22574040	436
NL32_B08	Anthranilate phosphoribosyl-transferase-like protein	7	1	17407476	17408075	599
NL24_E03	CONSTANS family zinc finger protein	7	1	27770404	27770632	228
Contig288	LRK1 protein	7	1	1729373	1730024	651
Contig205	RNA-binding protein	7	1	10607896	10608636	740
NL_1_F20	rpS28	7	1	4790517	4790772	255
Contig276	Alpha tubulin	7	1	22940618	22940934	316
Contig213	Alpha-galactosidase-like protein	7	1	28353792	28353650	142
NL_1_N15	Photosystem I chain IV precursor	7	1	14248521	14248453	68
Contig791	Zinc finger-like protein	7	1	3571758	3572469	711
Contig69	Calcium-dependent protein kinase	7	1	22556009	22555947	62
NL46_A03	Glutamate dehydrogenase	7	1	7617729	7617826	97
NL50_B05	Glycolate oxidase	7	1	25084829	25084766	63
Contig646	Hydrophobic protein RC12B	7	1	26076023	26076188	165
NL_15_7	p53 binding protein	7	1	17961947	17962234	287
NL_14_33	14-3-3 protein homolog GF14-12	8	1	22990197	22990254	57
Contig400	4-coumarate-CoA ligase	8	1	24101100	24100972	128
NL9_B04	Biotin synthase	8	1	26294292	26293787	505
Contig310	Cysteine endopeptidase precursor	8	1	27128756	27128530	226
NL_3_C11	GF14-c protein	8	1	20161163	20161050	113
NL_5_F18	Monodehydroascorbate reductase	8	1	2977634	2977280	354
NL_8_53	Phosphoglycerate dehydrogenase-like protein	8	1	21156756	21156559	197
NL14_B11	Fiber protein Fb14	8	1	27112301	27112126	175
Contig329	Manganese-binding protein PsbY precursor photosystem II-associated	8	1	1102021	1102148	127
Contig511	NAM (no apical meristem)	8	1	747538	747349	189
NL_15_53	Nuclear transport factor 2	8	1	25783485	25783186	299
Contig630	P450	8	1	297897	297736	161

(contd...)

Table 1. (contd...)

Unigene	Putative function	CHR	NE	CHR_Start	CHR_End	Exon lengths
Contig586	Ribonuclease	8	1	20399469	20399152	317
Contig44	RUB1 conjugating enzyme	8	1	16860533	16861122	589
NL_6_90	Caffeoyl CoA O-methyltransferase	8	1	23832663	23832133	530
Contig828	Homeodomain leucine zipper protein	8	1	23043325	23043509	184
Contig737	TGA-type basic leucine zipper protein	8	1	27411343	27411174	169
Contig193	Small nuclear ribonucleoprotein	8	1	2961117	2961064	53
Contig408	Metallothionein-like protein	8	1	8948336	8948407	71
NL16_A06	Glyoxalase II	9	1	18480620	18480169	451
Contig338	Calmodulin-like protein	9	1	15848498	15848081	417
Contig597	Nucleoid DNA-binding protein cnd41, chloroplast	9	1	16986752	16986828	76
NL_15_14	ADP-glucose pyrophosphorylase small subunit	9	1	6075403	6075141	262
Contig202	Chaperonin 21 precursor	9	1	14804179	14804561	382
Contig831	Cold acclimation protein WCOR410b	9	1	17121570	17121411	159
Contig551	Cytochrome P450 monooxygenase CYP92A1	9	1	14959013	14959206	193
NL_5_F07	HMG protein	9	1	20134956	20134876	80
NL_12_21	Protein kinase	9	1	18353296	18352812	484
Contig280	Chloroplast apocytochrome b6 (petB)	9	1	1915287	1915046	241
NL_6_92	Cap-binding protein p28	10	1	16492328	16492177	151
Contig9	Carnitine/acylcarnitine translocase	10	1	22003895	22004358	463
NL_1_K21	CYP18	10	1	3423331	3422801	530
Contig260	Ovule/fiber cell elongation protein Ghfe1	10	1	12643509	12644059	550
Contig164	Polygalacturonase isoenzyme 1 beta subunit	10	1	13402488	13402661	173
NLP_1_G18	Proline-rich protein	10	1	2881765	2882372	607
Contig13	Prolyl 4-hydroxylase, alpha subunit	10	1	18212109	18211833	276
Contig434	32 kDa protein jakalin homolog	10	1	1982985	1983464	479
Contig762	Hexose transporter	10	1	20275515	20275979	464
NL_3_C23	Minor allergen	10	1	14833391	14833280	111
Contig217	Mitochondrial chaperonin-60	10	1	16284267	16284187	80
Contig568	Nucleoside diphosphate kinase	10	1	21484580	21484291	289
Contig788	UDP-glucosyltransferase	10	1	1390999	1391363	364
Contig20	RNA binding protein	10	1	16664476	16664174	302
Contig737	TGA-type basic leucine zipper protein	10	1	21650887	21651327	440
Contig265	Ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit	10	1	10254331	10254767	436
Contig783	Histone H4	10	1	20245762	20245966	204
Contig489	6-phosphogluconate dehydrogenase	11	1	13570614	13571224	610
NL_4_H09	Vacuolar H ⁺ -ATPase (vatp-P1)	11	1	3391142	3391583	441
Contig723	AP2 domain protein homolog	11	1	3322244	3322318	74
Contig475	Class III chitinase homologue (OsChib3H-h)	11	1	24603831	24603890	59
Contig671	Cytokinin binding protein CBP57	11	1	11967874	11968373	499
Contig107	Hydroxymethyltransferase	11	1	11986753	11986653	100
Contig31	Protein transport protein subunit	11	1	162299	162022	277
Contig235	Disease resistance response protein	11	1	3929944	3929454	490
Contig530	50S ribosomal protein L4	11	1	18279527	18279648	121
Contig252	Chitinase	11	1	24584125	24584190	65
NL_0_K15	CP26, partial sequence	11	1	7577845	7578297	452
Contig601	Cytochrome P450-like sequence	12	1	9451082	9450372	710
NLP_1_E19	Diacylglycerol kinase	12	1	23203799	23203994	195
NL_18_30	Glutathione S-transferase OsGSTZ1	12	1	5690510	5690452	58

(contd...)

RESEARCH ARTICLES

Table 1. (contd...)

Unigene	Putative function	CHR	NE	CHR_Start	CHR_End	Exon lengths	
Contig803	Lipid transfer protein LPT III	12	1	693540	693732	192	
Contig698	Photosystem I protein (PSI-L)	12	1	12850587	12850439	148	
Contig31	Protein transport protein subunit	12	1	155922	155645	277	
Contig394	Ribulose biphosphate	12	1	9948024	9948674	650	
NL_0_L13	carboxylase/oxygenase Rubisco subunit binding- protein alpha subunit	12	1	10155739	10155557	182	
Contig829	Small subunit of ribulose- 1,5-biphosphate carboxylase	12	1	11122341	11122796	455	
NLP_1_A24	Wound-induced basic protein	12	1	11982435	11982348	87	
Contig259	Peroxidase	12	1	627888	627327	561	
NL22_D06	Nuclear RNA binding protein A	1	2	23519059	23518914	145	409
Contig204	Oligopeptide transporter	1	2	37525870	37525414	456	85
NL24_D05	GDSL-motif lipase/ hydrolase protein	1	2	30148718	30148351	367	79
NL_5_D17	Protein phosphatase	1	2	2717460	2717361	99	49
Contig737	TGA-type basic leucine zipper protein	1	2	38024783	38024535	248	200
NL_14_94	RING finger-like protein	1	2	210441	210303	138	384
Contig265	Ribulose 1,5-biphosphate- carboxylase/oxygenase large subunit	1	2	14063584	14063679	95	436
Contig408	Metallothionein-like protein	1	2	5475453	5475346	107	376
NL_1_P05	Low molecular weight heat shock protein precursor (hsp22)	2	2	31549440	31549144	296	284
Contig636	SR3 sucrose-regulated mRNA, 3'-end sequence	2	2	12953311	12952830	481	100
NL_14_94	RING finger-like protein	2	2	33440175	33440288	113	138
NL_1_K10	Zinc finger protein	2	2	32383552	32383634	82	182
Contig627	Membrane protein	3	2	912309	912091	218	163
Contig627	Membrane protein	3	2	913082	912916	166	141
Contig502	Translation initiation factor 4A	3	2	19911960	19912050	90	66
NL13_G12	Protein phosphatase 2C-like protein	3	2	2040594	2040140	454	81
Contig265	Ribulose 1,5-biphosphate carboxylase/oxygenase large subunit	3	2	10096108	10096218	110	435
Contig467	Zyloglucan endo-transglycosylase	3	2	7298798	7298860	62	57
Contig280	Chloroplast apocytochrome b6 (petB)	3	2	15806238	15805943	295	129
NL_5_D17	Protein phosphatase	4	2	10649268	10649169	99	337
NL_10_31	Zinc finger protein	4	2	23638559	23639022	463	239
Contig280	Chloroplast apocytochrome b6 (petB)	4	2	8598511	8598792	281	483
NL46_H02	Protein kinase	5	2	24355654	24355598	56	375
Contig694	S-adenosylmethionine synthetase	5	2	2019000	2018566	434	102
NL_1_G04	Cytochrome P450	5	2	18556499	18556628	129	471
Contig783	Histone H4	5	2	21492300	21492504	204	204
NL_1_G04	Cytochrome P450	5	2	18556499	18556628	129	471
Contig783	Histone H4	5	2	21492300	21492504	204	204
NL_2_H17	Translocation complex Sec61gamma chain	6	2	25731130	25730893	237	97
NL_4_F16	Zinc finger protein	6	2	8671098	8671675	577	106
Contig200	Lipase	6	2	2848295	2848344	49	49
Contig595	Zyloglucan endotransglycosylase	6	2	28093444	28093110	334	321
Contig280	Chloroplast apocytochrome b6 (petB)	6	2	22556820	22557101	281	483
Contig794	Ribosomal protein	7	2	21366075	21365811	264	186
Contig576	Water channel protein	7	2	15126880	15126707	173	415
NL42_H02	Germin-like protein 5	8	2	21899565	21899756	191	452
Contig52	Cytochrome P450	8	2	296921	296595	326	441
Contig280	Chloroplast apocytochrome b6 (petB)	8	2	9045885	9046166	281	484
Contig783	Histone H4	9	2	14502345	14502138	207	211
NL_3_M19	RING zinc finger protein	10	2	6945499	6945641	142	351
Contig280	Chloroplast apocytochrome b6 (petB)	10	2	10271705	10271221	484	280
Contig411	Lipid transfer protein precursor	11	2	682510	682158	352	172

(contd...)

Table 1. (contd...)

Unigene	Putative function	CHR	NE	CHR_Start	CHR_End	Exon lengths		
Contig303	Ribosomal protein	11	2	3311070	3310694	376	180	
Contig411	Lipid transfer protein precursor	12	2	703284	703456	172	352	
Contig284	Light-regulated protein	1	3	166189	166458	269	159	182
Contig312	nifU-like protein	1	3	26832887	26832454	433	55	244
Contig478	RicMT	1	3	42682110	42681969	141	68	316
Contig423	Beta-D-glucan exohydrolase, Isoenzyme ExoII	1	3	39333580	39333635	55	265	149
Contig277	Gibberellin-20 oxidase (Sd-1)	1	3	38118120	38118035	85	74	225
Contig769	L24 ribosomal protein	1	3	34415968	34416045	77	105	150
Contig21	VIP2 protein	1	3	26600252	26600053	199	100	280
Contig502	Translation initiation factor 4A	1	3	25415336	25415270	66	90	356
Contig78	Cytochrome P450	1	3	24821268	24821711	443	158	511
Contig54	Hydroxypyruvate reductase	2	3	74365	74497	132	164	115
Contig335	Integral membrane protein	2	3	9915926	9915810	116	169	451
NL35_C10	Pyruvate dehydrogenase E1 component alpha subunit, mitochondrial- precursor (PDHE1-A)	2	3	30543762	30543829	67	76	64
Contig176	Succinate dehydrogenase subunit 3 (sdh3)	2	3	1126169	1126102	67	341	95
Contig84	Succinic semialdehyde dehydrogenase	2	3	4003711	4003631	80	79	348
NL_3_D19	Inosine monophosphate- dehydrogenase	3	3	30988976	30989424	448	185	176
NL_1_E21	Novel protein, osr40c1	3	3	11917176	11917490	314	228	133
NL_15_62	P40-like protein	3	3	4318616	4318970	354	181	173
Contig314	Phosphoribosylanthranilate- transferase	3	3	1465702	1466069	367	128	126
Contig6	Ras-like GTP-binding protein	3	3	33028215	33028152	63	94	249
NL_15_51	Submergence induced protein 2A	3	3	3321916	3321855	61	72	352
Contig349	rab28 protein	3	3	3166690	3166816	126	321	251
Contig339	Phospholipid hydroperoxide gluta- thione peroxidase	3	3	13655076	13654987	89	169	309
Contig722	Putative anthocyanidin reductase	4	3	31043161	31043038	123	167	97
NL36_D01	Ribosomal protein L25	4	3	23998670	23998912	242	93	243
Contig433	Symbiosis-related protein	4	3	30703305	30703534	229	118	54
NL23_B03	OSMYB1	5	3	19908323	19908169	154	132	408
Contig677	Proline-rich protein RiP-15	5	3	7163526	7164180	654	419	140
Contig398	RD22	5	3	6603840	6603757	83	860	128
Contig396	Transcriptional regulator	5	3	18496165	18496365	200	156	308
NRS2R_1_F23	Vegetative storage protein	5	3	7192831	7192534	297	156	106
NL48_A04	Nucleic acid binding protein	5	3	19356393	19356597	204	131	270
Contig507	Peroxiredoxin Q	6	3	4861114	4860649	465	230	128
Contig454	EF-hand Ca ²⁺ -binding protein CCD1	6	3	27433046	27432735	311	129	155
NL_5_D22	RNA binding protein	6	3	6117964	6117894	70	120	265
NLP_1_A17	Similarity to beta-1,3- glucanase-like protein	7	3	20938061	20938277	216	254	217
NL52_D08	Tat binding protein	7	3	29022442	29022316	126	248	322
Contig783	Histone H4	7	3	21161260	21161200	60	65	208
Contig724	NAM-like protein	8	3	26022750	26022536	214	104	208
NL_2_B08	Peptidylprolyl <i>cis-trans</i> isomerase	8	3	25390187	25390317	130	120	211
Contig514	AP2 domain transcription factor	8	3	22590850	22591204	354	159	99
NL_4_N22	High mobility group I/Y-2	8	3	20005834	20005961	127	222	66
Contig362	Poly(A)-binding protein	8	3	12900903	12900697	206	328	70
NL_7_34	Mitogen-activated protein kinase homolog MMK2	10	3	19985246	19985115	131	181	310
Contig323	ORF; able to induce HR-like lesions	10	3	15326275	15326339	64	97	506
Contig237	Chlorophyll synthase	10	3	21709172	21709250	78	488	171
NL18_F12	Malate oxidoreductase	10	3	18451771	18451654	117	253	75
Contig574	RuBisCO activase large isoform precursor	11	3	24777472	24777770	298	86	204
NL_4_C14	Similar to ATP-citrate-lyase	11	3	24188244	24188336	92	120	282

(contd...)

RESEARCH ARTICLES

Table 1. (contd...)

Unigene	Putative function	CHR	NE	CHR_Start	CHR_End	Exon lengths			
Contig253	Wali7	11	3	2199685	2199987	302	67	113	
NL21_F12	Phospholipid hydroperoxide-gluthathione peroxidase	11	3	1983618	1983775	157	218	83	
Contig453	One helix protein	12	3	17181528	17181371	157	124	263	
NL_20_11	bZIP transcription factor PF100170 domain	12	3	25358673	25358807	134	70	49	
Contig707	Glycine-rich protein	12	3	26444609	26444724	115	163	244	
NL21_F12	Phospholipid hydroperoxide glutathione peroxidase	12	3	1864813	1864970	157	216	83	
Contig748	Phytochrome-associated protein	1	4	4814961	4815053	92	63	124	188
NL_6_64	RNA-binding glycine rich protein	1	4	42706351	42706461	110	69	110	303
NL_7_74	Sgt1	1	4	24717054	24716979	75	86	108	331
NL_0_J05	Small GTP-binding protein (Ran1)	1	4	23964572	23964510	62	182	79	263
Contig340	Phosphoethanolamine methyltransferase	1	4	28525412	28525540	128	139	141	147
Contig200	Lipase	1	4	6316182	6316508	326	216	150	426
Contig401	SR3 sucrose-regulated, 3'-end sequence	2	4	34957712	34957372	340	93	78	63
NL19_G12	Thylakoid-bound ascorbate peroxidase	2	4	20633306	20633204	102	123	79	68
Contig680	Xyloglucan endo-transglycosylase	2	4	28246081	28245844	237	101	567	123
Contig360	Cyclophilin	3	4	32616195	32616010	185	59	96	49
Contig635	Cysteine synthase	3	4	29396312	29396230	82	61	83	253
NL_5_O24	Photosystem II D1 protein	3	4	11633466	11633408	58	92	120	248
Contig660	Vacuolar membrane ATPase subunit G	4	4	29363095	29363451	356	101	98	61
NL50_C03	WRKY family transcription factor	4	4	29534308	29534436	128	106	55	71
Contig403	osERF3	5	4	23191453	23191155	298	256	118	384
Contig790	Ribosomal protein S31	5	4	22238778	22239059	281	132	87	94
NL_3_M01	RING finger protein	5	4	3548997	3549358	361	149	101	65
Contig703	Thioredoxin h	5	4	23853999	23853936	63	108	123	289
Contig686	Chaperonin	6	4	773874	773628	246	126	96	95
Contig414	Pathogenesis-related protein	6	4	795112	795390	278	148	103	121
NL_2_B20	RING3-like bromodomain protein	6	4	2001475	2001576	101	226	100	61
Contig383	RSZp22 splicing factor	6	4	4396135	4396044	91	235	124	266
Contig413	Ubiquinol-cytochrome c reductase	6	4	3782799	3782537	262	291	262	291
Contig410	Similar to DREB1A	6	4	1410465	1410546	81	140	106	112
Contig531	23 kDa polypeptide of photosystem II	7	4	2053881	2054010	129	355	54	169
Contig415	Elongation factor 1 beta	7	4	24986199	24986428	229	161	112	93
Contig483	Translation initiation factor (GOS2)	7	4	20454642	20454915	273	56	214	75
Contig382	Aconitase-iron regulated protein 1 (IRP1)	8	4	5417848	5417765	83	100	479	98
NL_5_A17	Aminotransferase 1	8	4	24074229	24074311	82	212	368	72
Contig554	Histidine-containing phosphotransfer protein	8	4	27157740	27157789	49	132	68	75
NRS2-7_8_M18	Plastid RNA polymerase sigma factor	8	4	3728469	3728845	376	199	96	97
Contig579	Beta-glucosidase-like protein	8	4	24506374	24506214	160	258	110	104
Contig227	Tryptophan decarboxylase	8	4	2253323	2253838	515	123	168	245
Contig389	Nucleic acid-binding protein	8	4	27133093	27132794	299	282	104	118
NL_5_O02	Blight-associated protein p12 precursor	9	4	16525847	16525570	277	252	125	252
Contig192	Peroxidase	9	4	8175970	8175844	126	399	402	126
Contig203	Transcription factor BTF3	10	4	17471874	17472237	363	58	151	69
Contig83	Aldolase	11	4	3502289	3502219	70	112	271	215
NL35_B11	Dormancy-associated protein	11	4	23058643	23059006	363	117	55	66
NL_4_B03	Ca ²⁺ sensitive 3'(2'),5-diphosphonucleoside 3'(2') phosphohydrolase	12	4	4196142	4196452	310	155	100	55
Contig336	Metallothionein 2a	12	4	22900909	22900824	85	79	171	196
Contig280	Chloroplast apocytochrome b6 (petB)	12	4	5580049	5579566	483	157	128	185

(contd...)

Table 1. (contd...)

Unigene	Putative function	CHR	NE	CHR_Start	CHR_End	Exon lengths														
Contig714	Aldolase (T25)	1	5	1012993	1013078	85	271	112	92	262										
NRS2R_1_F03	Chloroplast RNA helicase 1 VDL isoform	1	5	42458599	42458494	105	96	114	95	183										
NL_5_K18	Ras-related GTP binding protein	1	5	20919418	20919311	107	106	156	73	73										
NL_13_48	Dehydrin-like protein	1	5	30523740	30523864	124	70	182	143	69										
Contig193	Small nuclear ribonucleoprotein	2	5	1373643	1373771	128	100	79	63	84										
Contig15	Enolase	3	5	7806046	7806184	138	125	57	54	90										
Contig786	Beta-oxyacyl-[acyl-carrier protein] reductase	4	5	17494485	17494387	98	83	55	60	145										
NL_0_G19	Transporter-like protein	4	5	23506026	23506334	308	65	51	81	130										
NL_4_D05	Protein phosphatase 2C-like protein	4	5	32631502	32631101	401	76	94	88	72										
Contig306	Aspartic proteinase	5	5	2105603	2105670	67	143	122	93	291										
Contig780	Catalase	6	5	29636929	29636823	106	91	70	96	393										
Contig623	Chaperonin 10	6	5	4909234	4909158	76	159	81	365	76										
Contig480	HOS59	6	5	25358199	25358698	499	138	135	194	80										
NL_5_K17	Alanine aminotransferase-like protein	7	5	444751	445037	286	159	94	64	66										
Contig779	Photosystem II 10 kDa polypeptide	8	5	5895414	5895078	336	56	56	54	205										
Contig796	Wound inducive	8	5	15686219	15686390	171	59	94	120	63										
NL_0_B11	Apospory-associated protein C-like	9	5	8345299	8345558	259	81	109	98	101										
Contig797	Enolase	10	5	4297754	4297669	85	105	82	88	50										
NL_1_M20	Small GTP-binding protein (rab5A)	12	5	26407109	26407016	93	76	93	172	142										
Contig808	Auxin-regulated protein (Aux 28)	12	5	24691762	24691628	134	119	75	76	94										
Contig585	Signal recognition particle receptor alpha	1	6	41870396	41870325	71	122	62	75	107	166									
Contig513	Zinc finger protein, putative, 5' partial	1	6	24247888	24247783	105	147	53	68	204	71									
Contig263	Shaggy-like kinase etha	2	6	7674064	7673923	141	59	78	277	61	91									
Contig58	Serine/threonine kinase	3	6	11487632	11487567	65	127	82	107	74	57									
Contig190	Thioredoxin-like protein	4	6	33101408	33101541	133	110	57	70	62	261									
NRS2R_1_G06	ATP-dependent Clp protease proteolytic subunit	5	6	28120886	28120814	72	63	105	211	116	161									
Contig807	Metallothionein-like protein	5	6	576154	576052	102	105	75	433	51	156									
NRS2R_1_J03	Root border cell-specific protein	7	6	22890665	22890586	79	67	99	101	89	69									
Contig75	Nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase	8	6	20797542	20797625	83	131	122	139	157	136									
Contig221	Serine carboxypeptidase	11	6	11479781	11479857	76	273	115	123	72	64									
Contig609	Small GTP binding protein RACDP (RACD)	2	7	35464611	35464449	162	89	112	65	66	67	54								
Contig691	MAP3K beta 1 protein kinase	3	7	8533667	8533558	109	63	77	158	65	81	402								
Contig602	Indole-3-glycerol phosphate synthase	4	7	22353107	22353048	59	80	169	76	181	130	54								
Contig66	Cytosolic glyceraldehyde-3-phosphate dehydrogenase GAPDH	8	7	1524014	1524115	101	117	102	150	62	99	94								
Contig65	Dehydrin	11	8	11867675	11867735	60	84	86	56	336	102	120	69							
Contig402	Triosephosphate isomerase (Rictipi2)	1	9	2586567	2586925	358	84	80	97	130	87	124	72	61						
NL_3_H17	Membrane-associated protein	10	9	17945364	17945303	61	153	154	61	94	449	154	61	332						
NL_11_88	Elongation factor 1 alpha	3	12	4076966	4077270	304	142	125	300	142	125	304	142	125	304	142	125	304	142	125

CHR, Chromosome; NE, Number of exons predicted by EST BlastN hits.

CHR_Start, Chromosome physical positions of EST unigene starts.

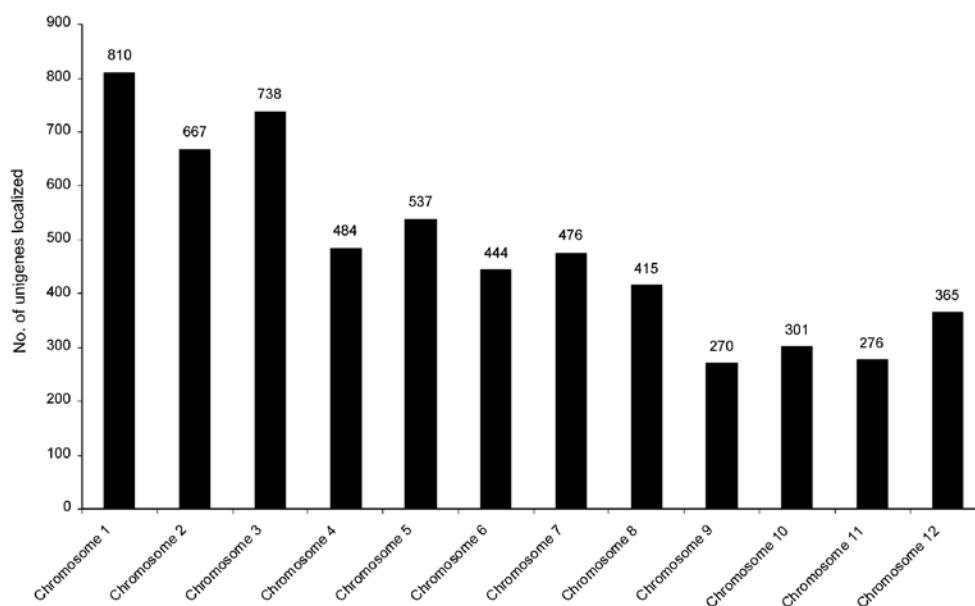
CHR_End, Chromosome physical position of EST unigene ends.

generated from a tissue-specific library constructed during a defined developmental stage and environmental conditions, have been used to get a snapshot of gene expression profile of that particular tissue. ESTs generated from N22 seedlings subjected to drought stress represented stress-induced gene expression profile. Data presented here give a clear description of the utility of ESTs in gene discovery, in chromosomal localization of candidate stress-responsive genes

and in high density physical maps of drought transcriptome, all contributing to molecular dissection of genetic determinants for drought tolerance (Figure 3 a-d). Further, sequence analysis of the targeted putative candidate genes of drought stress response revealed precise organization of these genes. This sequence and positional information at nucleotide level is expected to facilitate development of molecular probes and markers for drought tolerance in rice and likely

Table 2. Known stress-responsive genes at QTL locations

Clone ID	Putative function	Chromosome	cM	Flanking marker	QTL acc. no.
NL_7_34	Mitogen-activated protein kinase	10	61.7–68.6	E10477S, R716	CQE77
NL_1_M20	Small GTP-binding protein (Rab5a)	12	108.2	R2292	CQE87, AQDZ005, AQCI012, CQE83
NL_14_33	14-3-3 protein homologue GF14-12	8	92.2–96.6	R2382	CQE69, CQE70, CQN37, CQN38
NL46_H02	Protein kinase	5	104.7	E31112S	AQDZ013, AQDZ002
NL_5_B21	1-Aminocyclopropane-1-carboxylate oxidase	7	55.6	C60626SB	CQG5, CQN36, CQN35, CQN47
NL_3_G15	EREBP-like protein	3	20.3	S2769	CQE15
NL_0_C09	Ethylene-responsive protein (ebp-89 gene)	3	20.3	S2769	CQE15
NRS2R_1_H13	Helicase-like transcription factor	4	109.9	R78	CQE34, CQN32
NL_3_C10	OSMYB1	5	85.7	C308	CQE41
NL_3_B07	RNase S-like protein	9	82.4, 83.2	S10578, S955	CQE74, CQE76, CQE75
NL_1_A17	Vesicle-soluble NSF attachment protein receptor	3	65.4	C12845S	CQE18
NL_1_F05	Mitochondrial F0 ATP synthase D chain	8	92.2	S4036S	CQE69, CQE70, CQN37, CQN38
NL_1_N15	Photosystem I chain IV precursor	7	50.0	R658	AQM001, AQDN004, CQN36, CQN35

**Figure 2.** N22 unigene EST chromosome localization.

in other cereal crops. We have chosen to generate ESTs from the 3'-end aimed at the dissection of 3'UTR diversity and the possible presence of (SSRs). In fact, several SSRs have been identified in the study, which will serve as gene specific tags (unpublished data). These annotated unigenes of *indica* rice will serve as a reference source to rice researchers in particular, and cereal researchers in general.

A major constraint in map-based cloning approaches is the insufficient number of PCR-based molecular markers available to perform fine mapping³³. It is obvious that there is need for generating a large set of EST-based PCR markers, particularly those specific in determining intravarietal differences among Indian cultivars for drought tolerance. They can be directly deployed in marker-assisted selection for drought tolerance.

Particularly, the sequence and map data described here will help in designing gene-specific markers in rice. Further,

these resources can be used to convert candidate ESTs into PCR-based markers. (We have been doing this part of our research programme on functional genomics of drought tolerance in rice; data not shown.) These EST resources can be used to study syntenic relationships among cereals, as these are generally conserved among the grass genomes. We have used these ESTs as RFLP markers for synteny in rice and sorghum³⁴.

Genetic association studies, including the development of high-density maps constitute an important step in the positional cloning of genes underlying complex traits¹². The sequence information in public databases will provide necessary tools for the creation of new molecular markers and identification of (SNPs) and through SNP information, positional cloning in *Arabidopsis* is highly accelerating³⁵. Unlike other molecular markers, SNPs provide a way to generate highly saturated genetic maps and are amenable

Figure 3b. Transcript physical mapping on rice chromosomes IV-VI

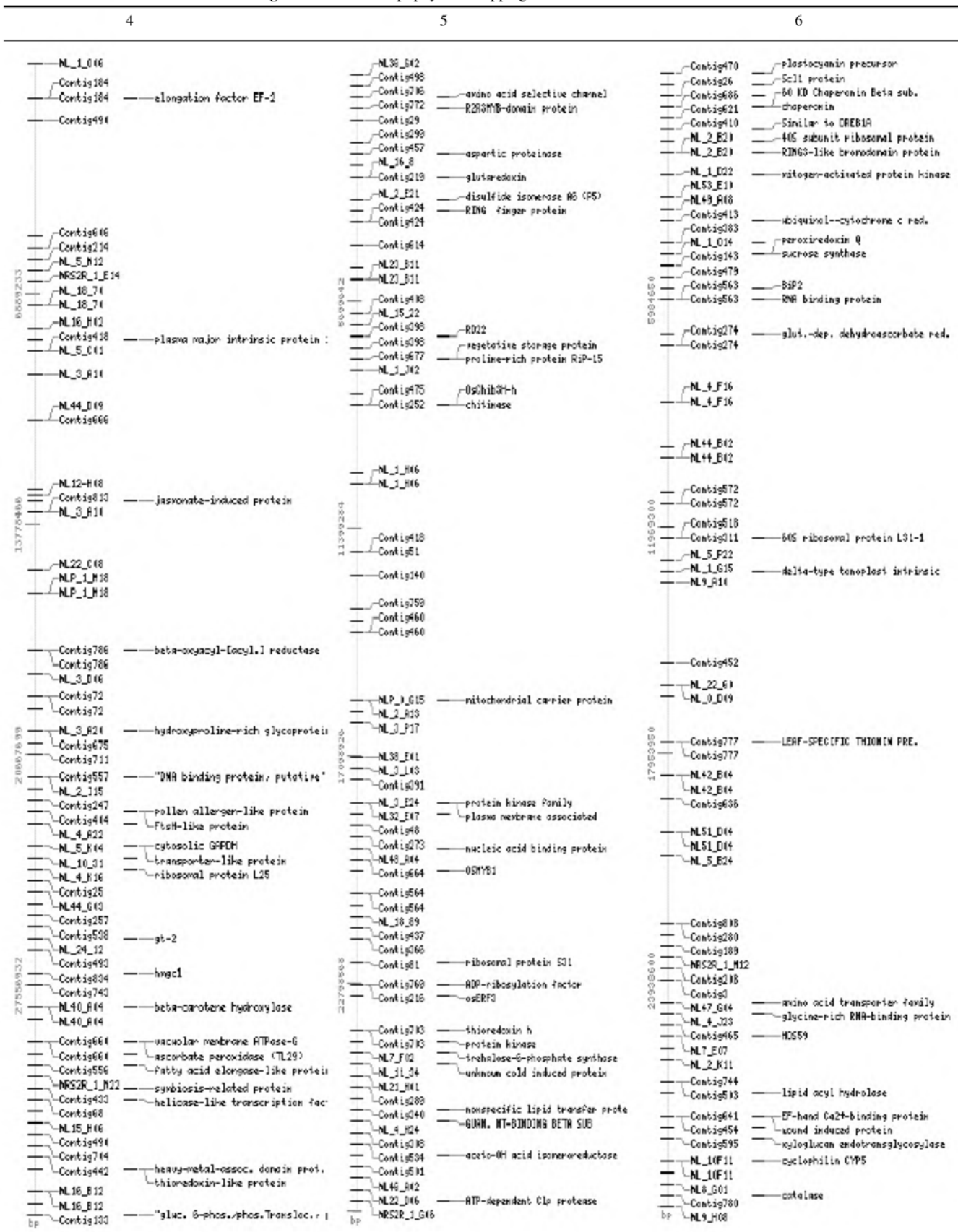


Figure 3 c. Transcript physical mapping on rice chromosomes VII-IX

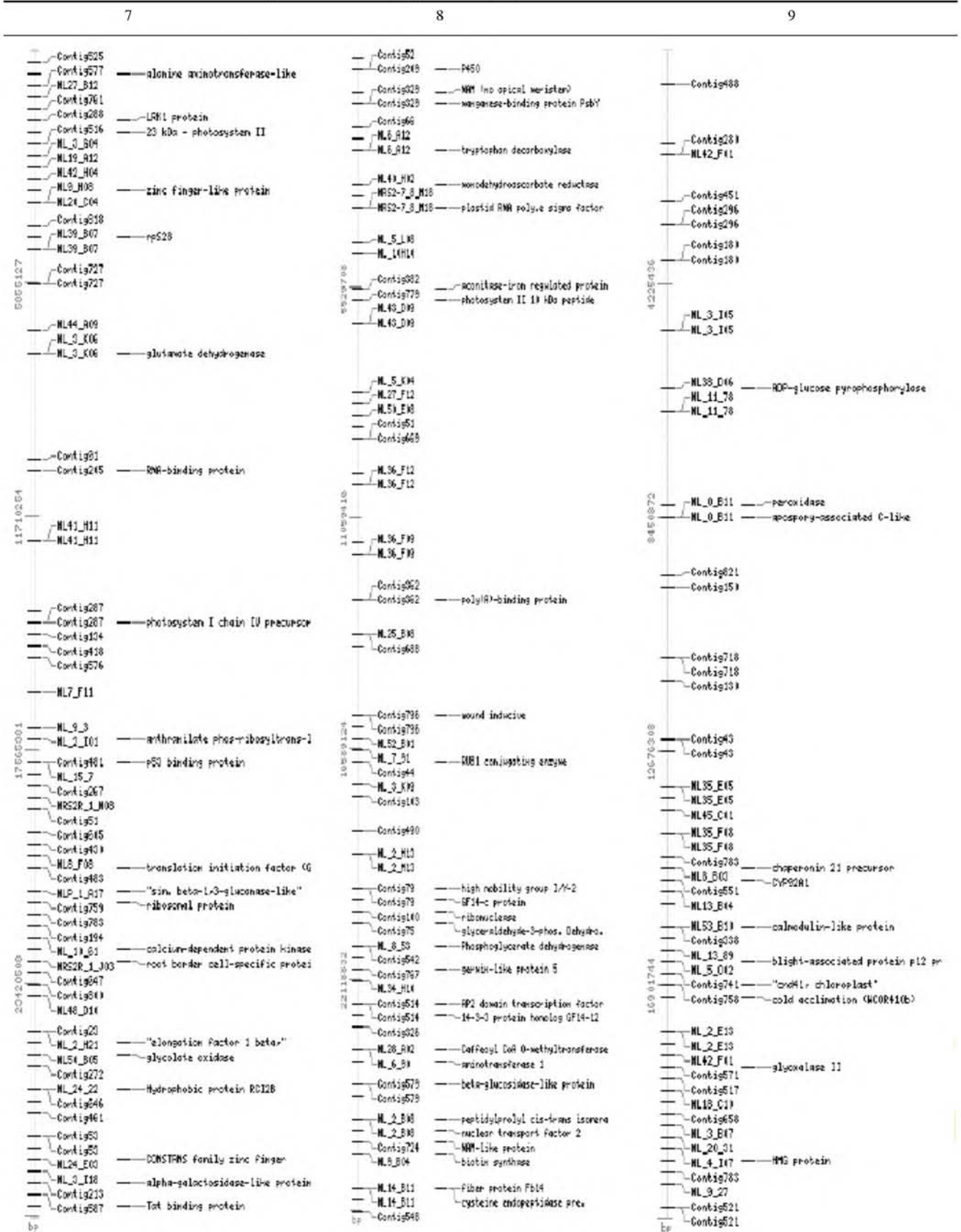
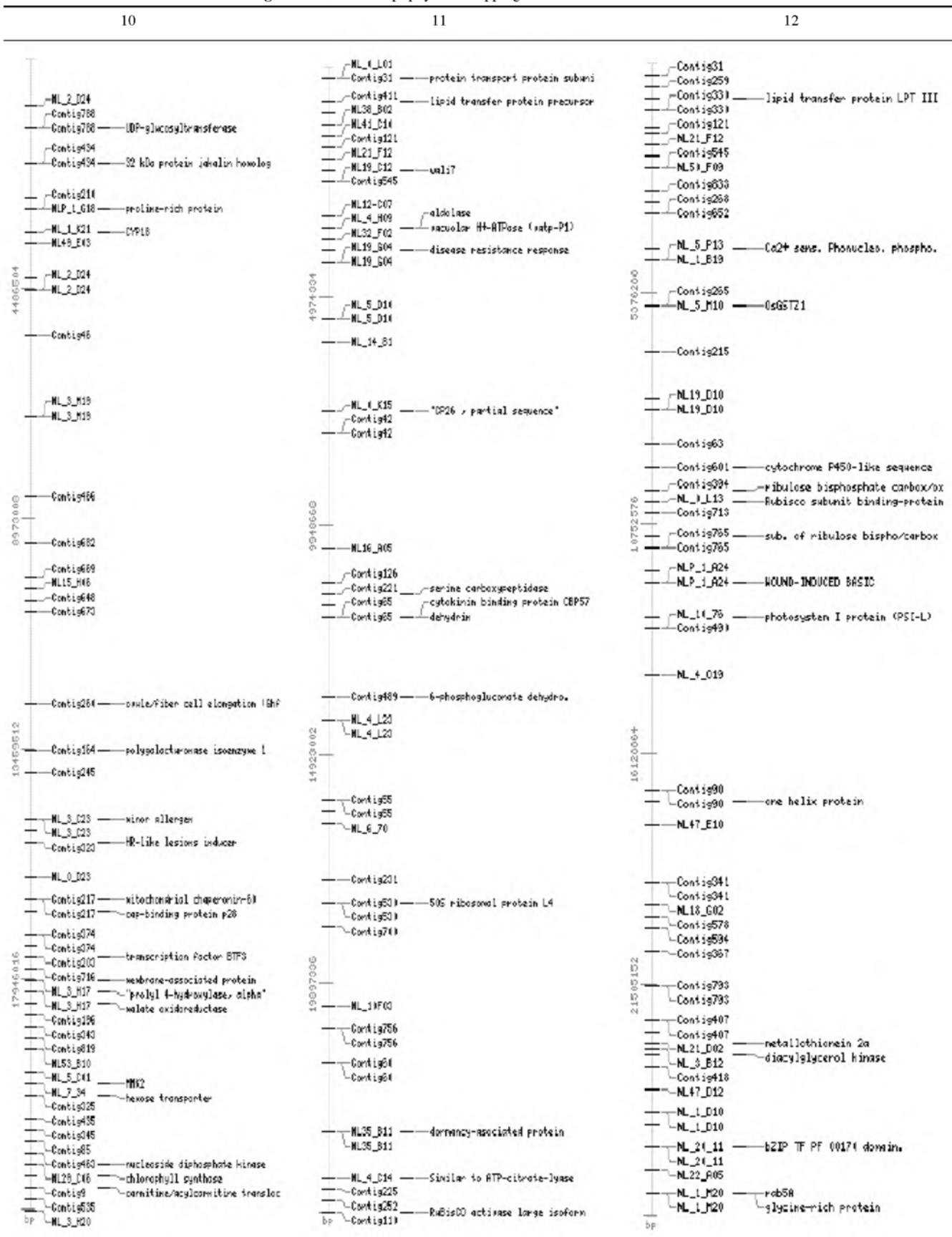


Figure 3 d. Transcript physical mapping on rice chromosomes X–XIII



for automation³⁶. The SNP frequency pattern along rice chromosomes shows an uneven distribution of polymorphism-rich and poor regions^{37,38}. Current estimates³⁷⁻³⁹ of SNP frequencies in rice range from 1.70 SNP/Kbp to 11.7 SNP/Kbp. Further, SNP frequency differs as much as 0.49% among *indica* genotypes³⁸, which is considerably higher than *japonica* genotypes (0.03 to 0.05%). The above data suggest that there exists a high degree of polymorphism in *indica* cultivars, which makes it possible to develop markers even between very closely related cultivars, this has been difficult to find by conventional methods such as RFLP. The SNPs of N22 putative candidate genes will help breeders in understanding genetic determinants of drought tolerance and implementing strategies for efficiently introgressing these genes across rice lines. We have analysed a few candidate genes for intraspecific sequence variation using the ESTs, for which both the structural and functional information has been described here (Reddy *et al.*, unpublished).

Data presented here can be used as a resource in identification and analysis of QTLs for drought tolerance in rice. Localizing rice ESTs onto genomic sequence provides a direct route for drought tolerance gene discovery. Structural information of putative candidate genes and genes linked to QTLs will be an important resource for such studies. Candidate QTL genes can also be identified from expression profiling experiments, under the assumption that genes that show genotype-specific differences could be the causative agents for variation in a trait. Physical map locations of rice ESTs observed by mapping EST sequences onto genetically anchored BAC/PAC clones of rice genomic sequences⁴⁰ revealed known stress responsive genes in QTLs associated with drought tolerance (Table 2). Further experiments are underway to elucidate the precise role of these putative candidate genes in drought stress response in rice.

1. Khush, G. S., Origin, dispersal, cultivation and variation of rice. *Plant Mol. Biol.*, 1997, **35**, 25–34.
2. Bhuiyan, S. I., Water management in relation to crop production: Case study on rice. *Outlook Agric.*, 1992, **21**, 293–299.
3. Dey, M. M. and Upadhaya, H. K., Yield loss due to drought, cold, and submergence in Asia. In *Rice Research in Asia, Progress and Priorities* (eds Evenson, R. E., Herdt, R. W. and Hossain, M.), Oxford University Press, NC, 1996, pp. 231–242.
4. Devos, K. M. and Gale, M. D., Comparative genetics in the grasses. *Plant Mol. Biol.*, 1997, **35**, 3–15.
5. Bennetzen, J. L. and Freeling, M., Grasses as a single genetic system: Genome composition, collinearity and compatibility. *Trends Genet.*, 1993, **9**, 259–261.
6. Bennetzen, J. L. and Freeling, M., The unified grass genome: Synergy in synteny. *Genome Res.*, 1997, **7**, 301–306.
7. Hogenesch, J. B. *et al.*, A comparison of the Celera and ensemble predicted gene sets reveals little overlap in novel genes. *Cell*, 2001, **106**, 413–415.
8. Adams, M. D. *et al.*, Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature*, 1995, **377**, 3–174.
9. Jiang, H., Kang, D., Alexandre, D. and Fisher, P. B., RaSH, a rapid subtraction hybridization approach for identifying and cloning differentially expressed genes. *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 12684–12689.
10. Chen, M. *et al.*, An integrated physical and genetic map of the rice genome. *Plant Cell*, 2002, **14**, 537–545.
11. Harushima, Y. *et al.*, A high-density rice genetic linkage map with 2275 markers using a single F2 population. *Genetics*, 1998, **148**, 479–494.
12. Wolford, J. K., Blunt, D., Ballecer, C. and Prochazka, M., High throughput SNP detection by DNA pooling and denaturing high performance liquid chromatography (DHPLC). *Hum. Genet.*, 2000, **107**, 483–487.
13. Kan, Z. *et al.*, Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. *Genome Res.*, 2001, **11**, 889–900.
14. Reddy, A. R. *et al.*, Novel genes are enriched in normalized cDNA libraries from drought-stressed seedlings of rice (*Oryza sativa* L. subsp. *indica* cv. Nagina 22). *Genome*, 2002, **45**, 204–211.
15. Markandeya, G. *et al.*, Novel EST enrichment with normalized cDNA libraries from drought-stressed rice (*Oryza sativa* L. cv Nagina 22), 2002; www.ncbi.nlm.nih.gov/dbest.
16. Markandeya, G. *et al.*, ESTs from a normalized cDNA library of drought-stressed rice seedlings (*Oryza sativa* L. cv Nagina 22), 2003; www.ncbi.nlm.nih.gov/dbest.
17. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J., Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.*, 1997, **25**, 3389–3402.
18. Ewing, B. and Green, P., Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.*, 1998, **8**, 186–194.
19. Fang, Z., Polacco, M., Chen, S., Schroeder, S., Hancock, D., Sanchez, H. and Coe, E., cMap: The comparative genetic map viewer. *Bioinformatics*, 2003, **19**, 416–417.
20. Huang, X. and Madan, A., CAP3: A DNA sequence assembly program. *Genome Res.*, 1999, **9**, 868–877.
21. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J., Basic local alignment search tool. *J. Mol. Biol.*, 1990, **215**, 403–410.
22. Ravindra Babu, P., Development and analysis of gene resources generated from drought-stressed seedlings of *indica* rice (*Oryza sativa* L.), Ph D thesis, University of Hyderabad, 2004.
23. Seki, M. *et al.*, Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses by using a full-length cDNA microarray. *Plant Cell*, 2001, **13**, 61–72.
24. Seki, M. *et al.*, Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J.*, 2002, **31**, 279–292.
25. Kreps, J. A., Wu, Y., Chang, H. S., Zhu, T., Wang, X. and Harper, J., Transcriptome changes for *Arabidopsis* in response to salt, osmotic, and cold stress. *Plant Physiol.*, 2002, **130**, 2129–2141.
26. Ozturk, Z. N. *et al.*, Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. *Plant Mol. Biol.*, 2002, **48**, 551–573.
27. Matsumura, H. *et al.*, Technical advance: Transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression (SAGE). *Plant J.*, 1999, **20**, 719–726.
28. Kawasaki, S. *et al.*, Gene expression profiles during the initial phase of salt stress in rice. *Plant Cell*, 2001, **13**, 889–906.
29. Rabbani, M. A. *et al.*, Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. *Plant Physiol.*, 2003, **133**, 1755–1767.
30. Price, A. H., Townend, J., Jones, M. P., Audebert, A. and Courtois, B., Mapping QTLs associated with drought avoidance in upland rice grown in the Philippines and West Africa. *Plant Mol. Biol.*, 2002, **48**, 683–695.

RESEARCH ARTICLES

31. Zhang, W. P. *et al.*, QTLs and epistasis for rice root elongation induced by different water conditions. *Theor. Appl. Genet.*, 2001, **103**, 118–1213.
32. Babu, R. C. *et al.*, Genetic analysis of drought resistance in rice by molecular markers: Association between secondary traits and field performance. *Crop Sci.*, 2003, **43**, 1457–1469.
33. Drenkard, E. *et al.*, A simple procedure for the analysis of single nucleotide polymorphisms facilitates map based cloning in *Arabidopsis*. *Plant Physiol.*, 2000, **124**, 1483–1492.
34. Seetharama, N. *et al.*, Progress in comparative genome mapping of rice and sorghum. An international workshop on progress towards developing resilient crops for drought-prone areas, Los Banos, Philippines, 27–30 May 2002.
35. Lukowitz, W., Gillmor, C. S. and Scheible, W. R., Positional cloning in *Arabidopsis*: Why it feels good to have a genome initiative working for you. *Plant Physiol.*, 2000, **123**, 795–805.
36. Wang, D. G. *et al.*, Large scale identification, mapping and genotyping of single-nucleotide polymorphisms in human genome. *Science*, 1998, **280**, 1077–1082.
37. Feltus, F. A., Wan, J., Schulze, S. R., Estill, J. C. and Paterson, A. H., A SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. *Genome Res.*, 2004, **14**, 1812–1819.
38. Nasu, S. *et al.*, Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. *DNA Res.*, 2002, **9**, 163–171.
39. Zhang, K., Qin, Z. S., Liu, J. S., Chen, T., Waterman, M. S. and Sun, F., Haplotype block partitioning and tag SNP selection using genotype data and their applications to association studies. *Genome Res.*, 2004, **14**, 908–916.
40. Babu, P. R., Sekhar, A. C., Ithal, N., Markandeya, G. and Reddy, A. R., Annotation and BAC/PAC localization of non-redundant ESTs from drought-stressed seedlings of an *indica* rice. *J. Genet.*, 2002, **81**, 25–44.

ACKNOWLEDGEMENTS. Financial grant from Rockefeller Foundation, USA to A.R.R. is acknowledged. M.G. thanks CSIR, New Delhi for award of SRF and V.B.R.L. for award of JRF.

Received 9 February 2005; accepted 18 May 2005
