



Evolutionary and expression study of the aldehyde dehydrogenase (ALDH) gene superfamily in rice (*Oryza sativa*)

Chenxi Gao^{a,b}, Bin Han^{a,*}

^a National Center for Gene Research and Shanghai Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 500 Caobao Road, Shanghai 200233, PR China

^b Graduate School of the Chinese Academy of Sciences, Beijing 100039, PR China

ARTICLE INFO

Article history:

Received 3 September 2008
Received in revised form 3 November 2008
Accepted 7 November 2008
Available online 21 November 2008

Received by I. King Jordan

Keywords:

Rice
Aldehyde dehydrogenase
Phylogeny
Expression profile
Abiotic stress

ABSTRACT

Aldehyde dehydrogenase (ALDH) superfamily represents a group of NAD(P)⁺-dependent enzymes that catalyze the oxidation of endogenous and exogenous aldehydes to the corresponding carboxylic acids. A total of twenty ALDH genes were identified in the rice genome. They were grouped into 10 distinct families based on protein sequence identity. The whole genome duplication (WGD) predating the divergence of cereals and tandem duplications represent the major mechanism for this superfamily expansion. Intron loss was found to accompany the recent evolution of four rice ALDH families. Quantitative RT-PCR analysis revealed that some of the rice ALDH genes were expressed in an organ-specific manner. Microarray data analysis indicated that expression of most duplicated rice ALDH genes showed high tissue specificities. Diverse expression patterns for duplicated genes were evaluated using both microarray and MPSS data. Expression levels of some ALDH genes were up-regulated by drought and high salinity stresses and the phytohormone abscisic acid (ABA) application, indicating that the products of these genes were potentially involved in rice osmotic stress adaptation. These results suggested that the specific rice ALDH genes might be potentially useful in rice genetic improvement.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Endogenous aldehydes are common intermediates of a number of metabolic pathways, including amino acids, protein, lipid and carbohydrates metabolism (Schauenstein et al., 1977), whereas xenobiotics are the major exogenous source of aldehydes (Lindahl, 1992). The resultant aldehydes can react with cellular nucleophiles because of the electrophilic nature of their carbonyl group (Lindahl, 1992). Excess aldehydes will cause deleterious effects on organism metabolism. Therefore, the selective elimination of aldehydes is essential for cellular function. ALDHs have been considered as general detoxifying enzymes which eliminate biogenic and xenobiotic aldehydes in an NAD(P)⁺-dependent manner (Yoshida et al., 1998). The presence of ALDH genes in majority of the species suggests that the ALDH superfamily has an ancient origin. More than 555 distinct ALDH genes have been identified throughout all taxa (Sophos and Vasiliou, 2003). Based on protein sequence identity, the ALDH

superfamily has been categorized into distinct families (Vasiliou et al., 1999).

The plant ALDHs are represented in 11 distinct families: ALDH2, ALDH3, ALDH5, ALDH6, ALDH7, ALDH10, ALDH11, ALDH12, ALDH18, and ALDH21. Three of them (ALDH11, ALDH19 and ALDH21) are unique to plants. A genome-wide analysis performed in *Arabidopsis* (Kirch et al., 2004) described 14 genes grouped into 9 families, one of which (ALDH22) was a novel one.

Compared to the comprehensive study of ALDHs in humans (Yoshida et al., 1998; Marchitti et al., 2008), only a small number of plant ALDHs have been functionally characterized (Kirch et al., 2004). To date, most of the studied plant ALDH genes are shown to be induced under high salinity or water deficit conditions, suggesting possible roles of these genes in improving the plant osmotic stress tolerance (Kotchoni and Bartels, 2003; Kirch et al., 2004; Kirch et al., 2005). Several studies indicate that over-expression of some plant ALDHs indeed enhances plant tolerance to diverse abiotic stresses (Kotchoni and Bartels, 2003; Sunkar et al., 2003; Kotchoni et al., 2006; Rodrigues et al., 2006; Huang et al., 2008). Furthermore, the first identified plant ALDH2 gene *rj2* is required for male fertility in maize (Liu et al., 2001). The study of rice ALDH2a shows that this enzyme might be responsible for efficient detoxification of acetaldehyde during re-aeration after submergence of rice plants (Tsuji et al., 2003a). ALDH2C4 in *Arabidopsis* is involved in ferulic acid and sinapic acid biosynthesis (Nair et al., 2004). ALDH5F1 encodes a succinic-

Abbreviations: ABA, abscisic acid; ALDH, aldehyde dehydrogenase; cDNA, DNA complementary to RNA; CDS, coding sequence; DNase, deoxyribonuclease; NJ, Neighbor-joining; PEG, polyethylene glycol; P5C, Δ^1 -pyrroline-5-carboxylate; P5CDH, Δ^1 -pyrroline-5-carboxylate dehydrogenase; P5CS, Δ^1 -pyrroline-5-carboxylate synthetase; SSADH, succinic-semialdehyde dehydrogenase; WGD, whole genome duplication.

* Corresponding author. Tel.: +86 21 64845260; fax: +86 21 64825775.

E-mail address: bhan@ncgr.ac.cn (B. Han).

semialdehyde dehydrogenase (SSADH) in *Arabidopsis*. T-DNA knock-out mutants of this gene result in dwarfed plants with necrotic lesions and are sensitive to both ultraviolet-B light and heat stress (Bouché et al., 2003). The mitochondrial Δ^1 -pyrroline-5-carboxylate dehydrogenase (P5CDH) in *Arabidopsis* (ALDH12A1) probably participates in preventing proline toxicity (Deuschle et al., 2001). The *Arabidopsis* P5CS2, which is a Δ^1 -pyrroline-5-carboxylate synthetase (P5CS) gene encoding member of ALDH18, plays essential role in embryo development (Székely et al., 2008).

However, the composition, evolution and structure of rice ALDH superfamily have not been investigated. Genome-wide expression analysis of plant ALDH genes in different organs under normal growth condition and/or under various abiotic stresses has not been reported so far. In the present work, twenty rice ALDH genes encoding members of ten families were identified. Subsequently, the expansion history of rice ALDH superfamily and the genomic structure evolution of ALDH genes were investigated. Moreover, we analyzed the mRNA abundance of rice ALDH genes in various organs, and their responses to drought, high salinity and ABA treatments. Data presented here suggested potential roles of some rice ALDH genes in the rice adaptation to environmental stressors. The current work aimed to provide a foundation for further functional characterization of this gene superfamily in rice and in angiosperms.

2. Materials and methods

2.1. Database search for ALDH genes

Protein sequences of six known plant ALDHs (ALDH2B4, AAM27003; ALDH3H1, AAL59944; ALDH7B4, AAK55676; ALDH12A1, AAK73756; OsP5CS, BAA19916; ALDH21A1, AAK59374) were used as queries to search against the protein database of rice in TIGR with BLASTP (TIGR Rice Annotation Release 4, <http://tigrblast.tigr.org/eukblast/index.cgi?project=osa1>). All sequences with an *E*-value <1e-6 were selected for manual inspection. Two ALDH active site signature sequences (Kirch et al., 2004) were also considered in this step: (1) the ALDH glutamic acid active site (PROSITE PS00687); (2) the ALDH cysteine active site (PROSITE PS00070). Pfam web service (<http://pfam.sanger.ac.uk/>) was employed to confirm the candidate sequences as ALDH proteins. The full-length (FL) cDNA sequences of rice ALDH genes were searched at the GenBank (<http://www.ncbi.nlm.nih.gov/>) with BLASTN program. Deduced rice ALDH polypeptides were analyzed using tools available at the Expasy Proteomics Server (<http://www.expasy.ch/tools/>).

2.2. Plant materials, growth conditions and treatments

Rice (*Oryza sativa*) variety Nipponbare (*japonica* rice) was used for all experiments. Rice seedlings were used for drought and high salinity stresses and ABA treatments. Seeds were germinated on wet paper at 28 °C for 2 days, and planted in hydroponic culture under controlled conditions for ten days (12 h light 30 °C/12 h dark 24 °C cycles). The seedlings were then transplanted to solutions containing 250 mM NaCl, 20% polyethylene glycol (PEG) -6000 and 100 μ M ABA for 24 h, 24 h and 6 h, respectively. The same staged seedlings incubated in water for 24 h and 6 h were used as controls, respectively. Only young leaf was harvested for RNA isolation. To simulate drought stress at the reproductive stage, rice plants grown in soil at four days before heading stage were kept in glasshouse under natural growth condition without irrigation until the leaf was completely rolled. Only flag leaf was harvested for RNA extraction. The flag leaf from the same staged rice with normal irrigation was simultaneously harvested as control. The drought-stressed plants at the seedling stage and the reproductive stage showed similar phenotype with leaf completely rolled. For detecting ALDH expression in normal tissues, the germinated seeds were grown in pots for ten days to harvest young

leaf and young root. Stem and panicle after heading were prepared from the same staged plants.

2.3. RNA extraction, quantitative RT-PCR analysis and cDNA cloning

Total RNA was extracted using the Trizol reagent (Invitrogen) according to the manufacturer's instructions. The DNase-treated RNA was reverse-transcribed using SuperScript™ II reverse transcriptase (Invitrogen). Quantitative PCR was performed on the Applied Biosystems 7500 real-time PCR System using SYBR Premix Ex Taq™ (TaKaRa). The PCR thermal cycle conditions were as follows: denature at 95 °C for 10 s and 40 cycles for 95 °C, 5 s; 60 °C, 34 s. The specificity of the PCR reactions was determined by melting curve analysis of the products. Two rice genes were selected as internal reference genes: *UBQ5* (AK061988) and *eEF-1 α* (AK061464) (Jain et al., 2006). The average PCR efficiency of each primer pair calculated by LinRegPCR for young leaf and root was applied in the normalization (Ramakers et al., 2003). Relative quantification was calculated as $\frac{(E_{\text{target}})^{\Delta C_{\text{Ptarget}}(\text{control-sample})}}{(E_{\text{ref}})^{\Delta C_{\text{Pref}}(\text{control-sample})}}$ (Pfaffl, 2001). Three biological replicates for each sample were used for real-time PCR analysis. The Student's *t*-test was performed as a test of significance. FL-cDNAs of *OsALDH5* and *OsALDH2-3* were isolated using primers that encompassed the translation start codons and stop codons of putative genes. PCR products were cloned into pGEM-T Easy vector (Promega) and sequenced automatically. The two sequences of *OsALDH5* (CU606989) and *OsALDH2-3* (CU607043) were deposited in the EMBL. Primers used in present study were listed in Supplementary Table S1.

2.4. Sequence and phylogenetic analysis

Multiple protein alignment was performed with ClustalX 1.81 (Thompson et al., 1997). Alignment was edited manually using GeneDoc (Nicholas et al., 1997). Identification of conserved motifs of rice and *Arabidopsis* ALDHs was accomplished with multiple sequence alignments and Multiple Em for Motif Elicitation (MEME) version 3.5.7 (<http://meme.sdsc.edu>). The sequence logos were generated using the online Weblogo platform (<http://weblogo.berkeley.edu/>) with default parameters. Phylogenetic tree was constructed with MEGA 4.0 (Tamura et al., 2007) by neighbor-joining (NJ) method and the bootstrap test was carried out with 500 replicates. Pairwise deletion opinion and Jones, Taylor, and Thornton (JTT) model for amino acid sequences were used.

2.5. Duplication history of rice ALDH genes

The chromosomal locations of rice ALDH genes were determined by inspection on TIGR rice genome browser based on the LOC number (http://www.tigr.org/tigr-scripts/osa1_web/gbrowse/rice/). For the detection of large-scale duplications, we consulted the duplicated blocks identified by Lin et al. (2006).

2.6. Evaluation of rice ALDH gene expression patterns using microarray and MPSS data

The expression behaviors of rice ALDH genes were examined in a set of rice microarray data (GSE 7951) downloaded from GEO at NCBI (<http://www.ncbi.nlm.nih.gov/geo>). This dataset was generated by hybridization of RNAs from unpollinated stigma at anthesis, seedling shoot, seedling root, mature anther, ovary at anthesis, seeds of five days after pollination, 10-day-old embryo, 10-day-old endosperm as well as suspension cell on 57K Affymetrix rice whole genome array. To evaluate the tissue specificity for ALDH genes, we calculated the *tissue specificity index*, τ as described (Shoja et al., 2007). For three genes (*OsALDH2-3*, *OsALDH2-4* and *OsALDH3-2*) with more than one unique probe, we selected the probe with higher intensity value for

calculation. We evaluated overlapping tissue expression for rice *ALDH* duplicates using MPSS data according to the reported method (Johnson and Thomas, 2007). Data from seventeen representing libraries was chosen in this analysis (Supplementary Table S2).

3. Results

3.1. Rice *ALDH* superfamily contains 20 members

In this study, we grouped the rice *ALDH* proteins that had more than 40% sequence identity with known *ALDH*s into the same family (Kirch et al., 2004). Twenty rice genes were identified to encode members of ten *ALDH* families (Table 1). Four families (families 2, 3, 10 and 18) contained multiple members, and each of the rest six families (families 5, 6, 7, 11, 12 and 22) was represented by a single gene. A unique identifier was assigned to each of the rice *ALDH*s. The number following the dash in rice *ALDH* name was used to distinguish multiple members contained in one family. Two genes, *OsALDH18-1* and *OsALDH18-2*, encoded the P5CS defined as *ALDH*-like protein (Sophos and Vasiliou, 2003) with both gamma-glutamyl kinase and gamma-glutamyl phosphate reductase activities. P5CS proteins that are traditionally grouped into a distinct *ALDH* family (*ALDH18*) have great sequence divergence with proteins in other *ALDH* families; especially they do not contain the conserved *ALDH* active sites. So they were excluded from the following protein alignment and phylogenetic analysis.

In an attempt to ensure the accuracy of the sequences used in the following work, database search was carried out to find the FL-cDNAs matching to the candidate *ALDH* sequences. Seventeen *ALDH* genes were found to have corresponding FL-cDNAs (Table 1). The FL-cDNAs for *OsALDH2-3* and *OsALDH5* were cloned in this study. The original annotated sequence for *OsALDH2-3* was revised according to the cDNA sequence in this analysis. A sequencing error in the reference genome, the missing of a nucleotide relative to AK121765 (*OsALDH12*), was corrected by re-sequencing. And the corrected sequence for *OsALDH12* was used in following analysis.

Nine conserved motifs were identified in the rice and *Arabidopsis* *ALDH*s (Supplementary Figs. 1 and 2). All of them have been reported in the study of 145 *ALDH*s (Perozich et al., 1999). Among the nine conserved motifs, two contained the *ALDH* active sites: motif 4 contained the *ALDH* glutamic acid active site; motif 5 contained the cysteine active site. Motif 3 contained the possible coenzyme-binding site. The remaining motifs resided around the active sites of *ALDH* proteins.

3.2. WGD and tandem duplication both contribute to the superfamily expansion

Chromosomal localizations of all members in this superfamily indicated that the 20 rice *ALDH* genes were randomly distributed on 10 rice chromosomes (Supplementary Fig. 3). Among them, nine genes were located in six pairs of duplicated genome regions. Three *ALDH*

Table 1
The *ALDH* superfamily in rice

LOC no. ^a	Protein name	Former name ^b	cDNA accession no.	Putative function	CDS (bp) ^c	ORF (aa) ^d	MW (kDa) ^e	Subcellular localization
Os01g40860	OsALDH2-1	ALDH1a	AK119571/AK121462 AK105638/AB037421 AK105638/AB037421	Aldehyde dehydrogenase	1509	502	54.2	Cytosol
Os01g40870	OsALDH2-2	ALDH1b	AK101427	Aldehyde dehydrogenase	1524	507	54.2	Cytosol
Os06g39230	OsALDH2-3		CU607043	Aldehyde dehydrogenase	1581	526	56.4	Cytosol
Os02g49720	OsALDH2-4	ALDH2a	AK121610/AB030939	Aldehyde dehydrogenase	1662	553	58.9	Mitochondria
Os06g15990	OsALDH2-5	ALDH2b	AK073079/AK100358 AK099248/AK120296 AF162665/AB044537 AK099248/AK120296 AF162665/AB044537	Aldehyde dehydrogenase	1650	549	59.3	Mitochondria
Os02g43194	OsALDH3-1		AK070741	Aldehyde dehydrogenase	1464	487	54.5	
Os02g43280	OsALDH3-2		AK120274/AK060824	Aldehyde dehydrogenase	1476	491	54.6	
Os04g45720	OsALDH3-3		AK071169/AF464932	Aldehyde dehydrogenase	1500	499	54.3	
Os11g08300	OsALDH3-4		AK104746/AK104603 AK068297	Aldehyde dehydrogenase	1449	482	52.5	
Os12g07810	OsALDH3-5			Aldehyde dehydrogenase	1455	484	52.4	
Os02g07760	OsALDH5		CU606989	Succinic semialdehyde dehydrogenase	1584	527	56.1	Mitochondria
Os07g09060	OsALDH6		AK121280/CT833977	Methylmalonate semi-aldehyde dehydrogenase	1605	534	57.2	Mitochondria
Os09g26880	OsALDH7	ALDH7B7	AK120185/AK060757 AK119560/AF323586 CT833156 AK119560/AF323586 CT833156	Aldehyde dehydrogenase	1530	509	54.5	
Os04g39020	OsALDH10-1	BADH	AK103582	Betaine aldehyde dehydrogenase	1518	505	54.6	Mitochondria
Os08g32870	OsALDH10-2		AK071221/AB096083 DQ910546 DQ910546	Betaine aldehyde dehydrogenase	1512	503	54.7	Mitochondria
Os08g34210	OsALDH11		AK071520/AK062559 AY224439	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase	1500	499	53.4	
Os05g45960	OsALDH12		AK121765	Δ^1 -pyrroline-5-carboxylate dehydrogenase	1653	550	60.5	Mitochondria
Os01g62900	OsALDH18-1		AK101230/AK069425	Δ^1 -pyrroline-5-carboxylate synthetase	2208	735	79.5	
Os05g38150	OsALDH18-2	P5CS	AK102633/AY574031 D49714 D49714	Δ^1 -pyrroline-5-carboxylate synthetase	2151	716	77.7	
Os07g48920	OsALDH22		AK068462	Aldehyde dehydrogenase	1794	597	66.0	Secretory pathway

^a Gene model at the TIGR rice database.

^b Protein names used in former publications: ALDH1a (Li et al., 2000); ALDH2a (Nakazono et al., 2000); ALDH1b and ALDH2b (Tsuji et al., 2003b); ALDH7B7 (Wu et al., 2007); BADH (Nakamura et al., 1997); P5CS (Igarashi et al., 1997).

^c Coding sequence.

^d Open reading frame.

^e Molecular weight.

genes (*OsALDH6*, *OsALDH7* and *OsALDH11*) were asymmetrically located in three duplicated block pairs, indicating the occurrence of gene loss following the gene duplication. The remaining duplicated block pairs contained *ALDH* genes (*OsALDH2-4/OsALDH2-5*, *OsALDH3-4/OsALDH3-5* and *OsALDH18-1/OsALDH18-2*) in both duplicated regions. The nine genes can also be mapped to the WGD regions that were dated back to a common time before the divergence of cereals (Yu et al., 2005) except for *OsALDH3-4* and *OsALDH3-5*. In addition, tandem duplications seemed to contribute to the rise of two gene pairs: *OsALDH2-1/OsALDH2-2* and *OsALDH3-1/OsALDH3-2*.

3.3. Phylogenetic analysis of ALDH proteins

In order to compare rice ALDH proteins with those of *Arabidopsis*, a phylogenetic tree was constructed from the alignment of

full-length protein sequences. The tree can be generally classified into 9 major groups, representing the 9 distinct ALDH families (Fig. 1). Similar to the result of previous work (Kirch et al., 2004), families 2, 5 and 10 clustered together. Families 22 and 3 were connected by a node with high bootstrap value, indicating close relationship between them. The predicted cytosolic and mitochondrial ALDHs in family 2 were clearly separated from each other (Fig. 1; Skibbe et al., 2002; Kirch et al., 2004). Family 3 contained two subgroups each containing ALDHs from rice and *Arabidopsis* respectively, indicating an old split for them. Six families (families 5, 6, 7, 11, 12 and 22) had a ratio of 1:1 orthologous relationship, suggesting well conserved functions for rice and *Arabidopsis* orthologous genes. Each of *ALDH2C4* and *ALDH3F1* clustered with three rice genes, suggesting diverse functions for rice genes.

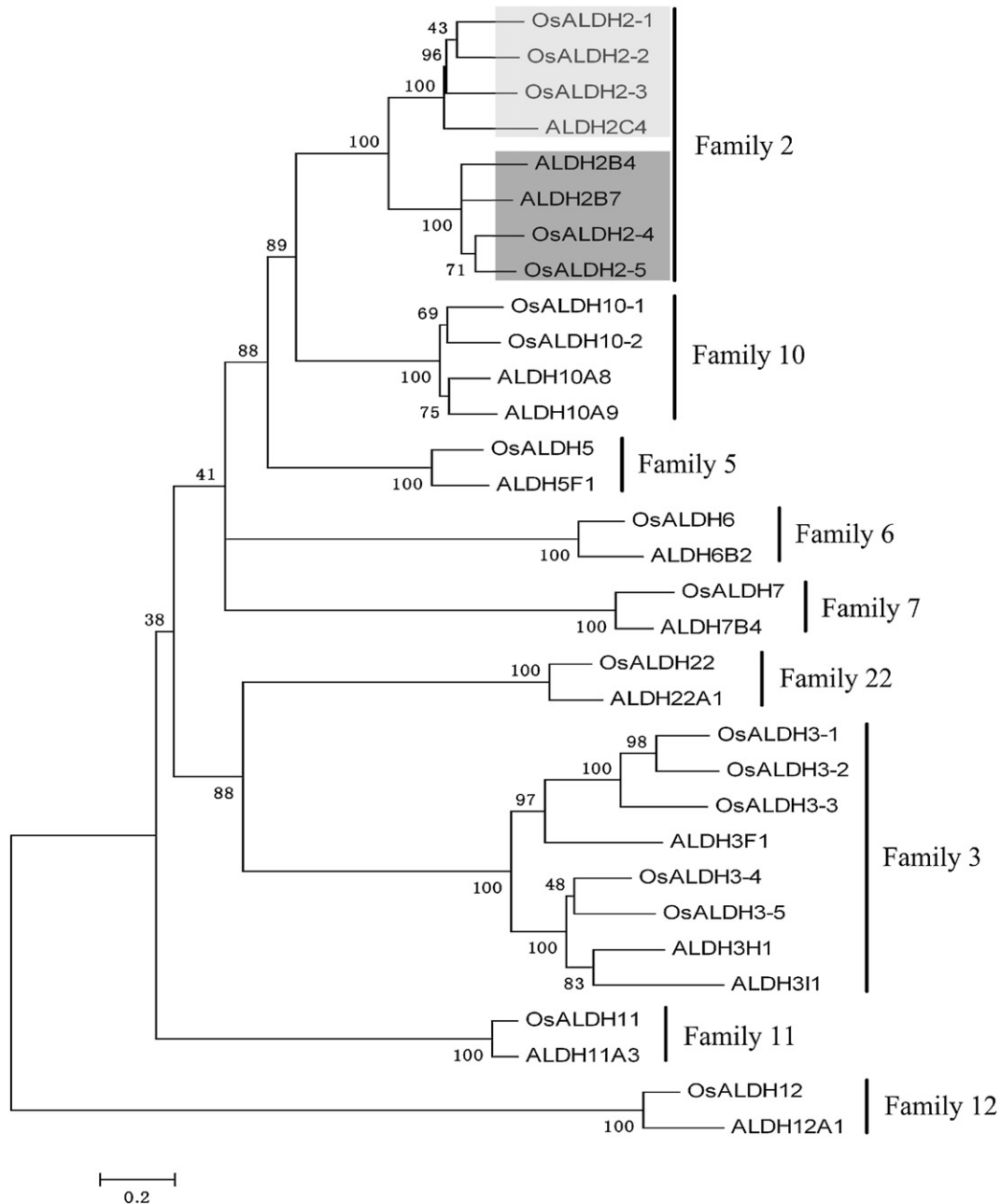


Fig. 1. Phylogenetic relationship of rice and *Arabidopsis* ALDH proteins. The unrooted NJ tree was generated with MEGA 4.0 program. Bootstrap values from 500 replicates are indicated at each branch. The nine major groups representing nine distinct ALDH families are indicated. The predicted cytosolic and mitochondrial ALDHs in family 2 were highlighted by grey and black background, respectively.

3.4. Intron loss might accompany the recent evolution of rice *ALDH* genes

We collected and compared the genomic structures of *ALDH* genes from rice and *Arabidopsis* to obtain further insight into the evolutionary history of rice *ALDH* superfamily. *ALDH* genes from the two genomes in each of families 5, 6, 7, 10, 11 and 22 had same exon numbers, and also possessed nearly identical exon lengths, except for the first exons and last exons (Supplementary Fig. 4). Intron loss was observed for rice *ALDH* genes in families 2, 3, 12 and 18 (Fig. 2). For confirming intron loss of *OsALDH12*, we also consulted the genomic structures of its orthologous genes from poplar (data not shown). In comparison with other genes in family 3, *ALDH311* had acquired two more exons in 5'-end during evolution (Fig. 2).

3.5. Rice *ALDH* genes were differentially expressed in various organs

To examine the expression profiles of rice *ALDH* genes, quantitative RT-PCR was carried out using the cDNA templates from four organs including young leaf, young root, stem and panicle. As shown in Fig. 3, *OsALDH2-3*, *OsALDH3-3* and *OsALDH18-1* were mainly expressed in stem and panicle. *OsALDH2-1*, *OsALDH10-1* and *OsALDH10-2* showed high expression levels in young root and panicle. *OsALDH2-2* was predominantly expressed in young root. *OsALDH3-5* was highly expressed in panicle. *OsALDH2-4* and

OsALDH3-2 showed relatively high expression level in stem. *OsALDH3-1* and *OsALDH3-4* were highly expressed in young leaf. Two other genes, *OsALDH2-5* and *OsALDH7*, exhibited low expression level in young root. *OsALDH6* was highly expressed in young root and stem. *OsALDH11* was mainly expressed in young leaf and panicle. *OsALDH12* had relatively high expression level in panicle, which was in accordance with the expression of its *Arabidopsis* orthologous gene (Deuschle et al., 2001). *OsALDH5*, *OsALDH18-2* and *OsALDH22* showed little expression variation in the four tissues. *OsALDH2-2*, and *OsALDH3-1* and *OsALDH3-2* were barely expressed in stem and young root, respectively.

3.6. Expression of duplicated *ALDH* genes showed high and distinct tissue specificities

The expression profiles based on microarray data for most *ALDH* genes matched well with our experimental result (Supplementary Fig. 5). The tissue specificity index (τ) for each of rice *ALDH* genes was calculated using microarray data. *OsALDH3-5* was excluded because of its extremely low expression signal. The τ value ranges from 0 to 1, with higher value indicating higher tissue specificity (Shoja et al., 2007). Most genes in multiple-member families exhibited high τ values, indicating high tissue specificities for their expression patterns (Fig. 4). *OsALDH2-2* and *OsALDH2-3* were predominantly expressed in



Fig. 2. Genomic structures of rice and *Arabidopsis* *ALDH* genes. The phylogenetic trees were constructed using MEGA 4.0. Five rice *ALDH* genes that experienced intron loss during recent evolution were highlighted in red color. Only coding exons, represented by black boxes, were drawn to scale. Numbers within boxes indicate the exon size in nucleotides. Line angles connecting two boxes represent introns.

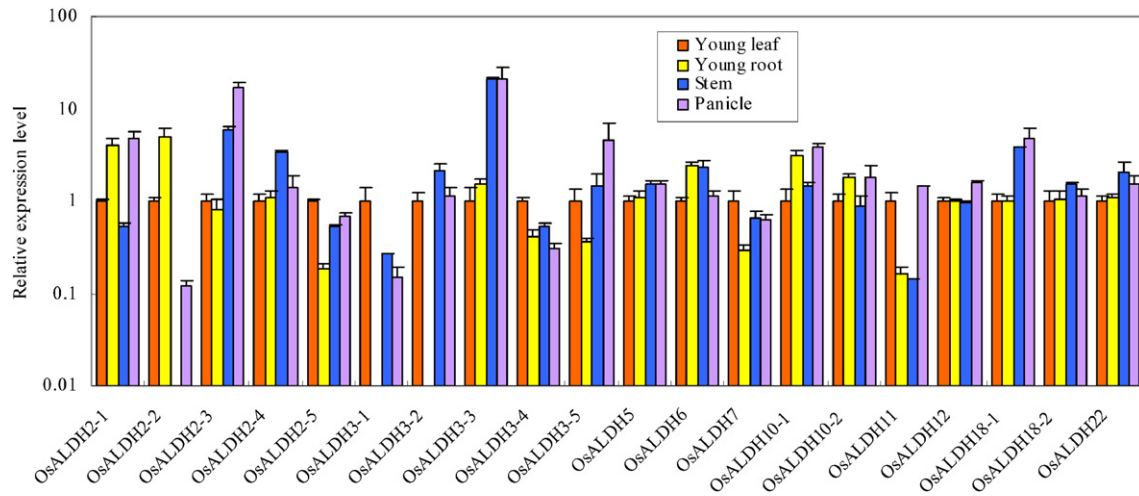


Fig. 3. Expression profiles of rice *ALDH* genes in four organs revealed by quantitative RT-PCR. Expression level is expressed relative to the young leaf on a logarithmic scale. Error bars indicate +SD of the mean of three biological samples.

root and ovary, respectively. *OsALDH2-1* was predominantly expressed in root according to the microarray data. But this gene was found to be also highly expressed in panicle in our analysis and previous study (Fig. 3; Li et al., 2000). *OsALDH2-4* was lowly expressed in stigma, shoot and root, while *OsALDH2-5* was highly expressed in most tissues, especially in shoot. Expression of *OsALDH3-1* can only be detected in shoot. *OsALDH3-3* was highly expressed in stigma and ovary, and *OsALDH3-2* was highly expressed in several tissues. *OsALDH18-1* was mainly expressed in anther, whereas *OsALDH18-2* was highly expressed in most tested tissues. *OsALDH10-1* and *OsALDH10-2* showed similar expression pattern. *OsALDH11* that was highly expressed in shoot, embryo and stigma had the highest τ value among the six genes in single-member families (Fig. 4). This gene also showed great expression divergence in four organs in our analysis with high expression levels in young leaf and panicle (Fig. 3). These data also indicated great expression divergence for most duplicated *ALDH* genes.

To get more evidence for the distinct expression patterns of duplicated *ALDH* genes, we also evaluated the overlapping tissue expression for 7 rice *ALDH* paralog pairs using MPSS data (Supplementary Table S2). The expression profiles retrieved from MPSS database for most genes generally matched with our experimental result. Six pairs showed no more than 50% overlapping tissue expression or even 0% overlap. Only one pair, *OsALDH2-4* and *OsALDH2-5*, had high overlap (88%).

3.7. Expression of most rice *ALDH* genes was regulated by abiotic stresses

Plant *ALDH* genes have been identified to play important roles in the adaptation of plant to various abiotic stresses. We therefore examined the responses of rice *ALDH* genes to drought and high salinity stresses in young leaf. Thirteen genes were found to be responsive to at least one stress ($P < 0.05$). Three genes (*OsALDH2-2*, *OsALDH3-1* and *OsALDH3-2*) were down-regulated by both drought and high salinity stresses (Fig. 5A). *OsALDH11* and *OsALDH22* were highly reduced by drought stress (Fig. 5B). Five genes (*OsALDH2-4*, *OsALDH3-4*, *OsALDH7*, *OsALDH10-1* and *OsALDH18-2*) were up-regulated by both drought and high salinity stresses (Fig. 5C). *OsALDH5* and *OsALDH12* were only induced by drought stress, while *OsALDH18-1* was only up-regulated by high salinity stress (Fig. 5D).

Rice is reported to be more sensitive to deleterious environmental factors at the seedling and the reproductive stage (Lutts et al., 1995). The responses to drought stress in flag leaf at the reproductive stage were examined for the thirteen genes that showed stress-regulated expression in young leaf. Five genes that were down- or up-regulated more than 2-fold in the preliminary analysis with one control sample and one drought-stressed sample were selected for further analysis. Finally, four of the 5 genes were found to be responsive to drought stress ($P < 0.05$, Fig. 5E). Among them, *OsALDH3-1* was down-regulated, while *OsALDH7*, *OsALDH18-1* and *OsALDH18-2* were up-regulated.

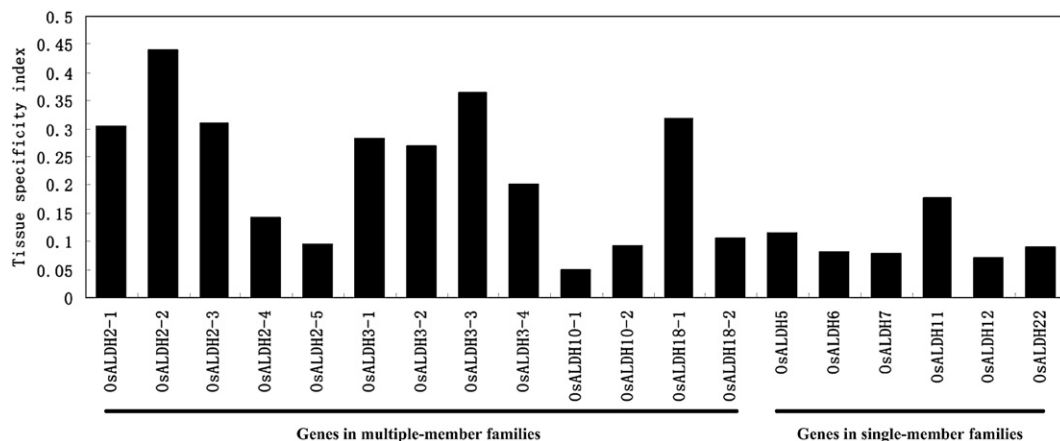


Fig. 4. Comparison of the tissue specificity index (τ) calculated from microarray data for rice *ALDH* genes. Thirteen genes in multiple-member families and six genes in single-member families are indicated.

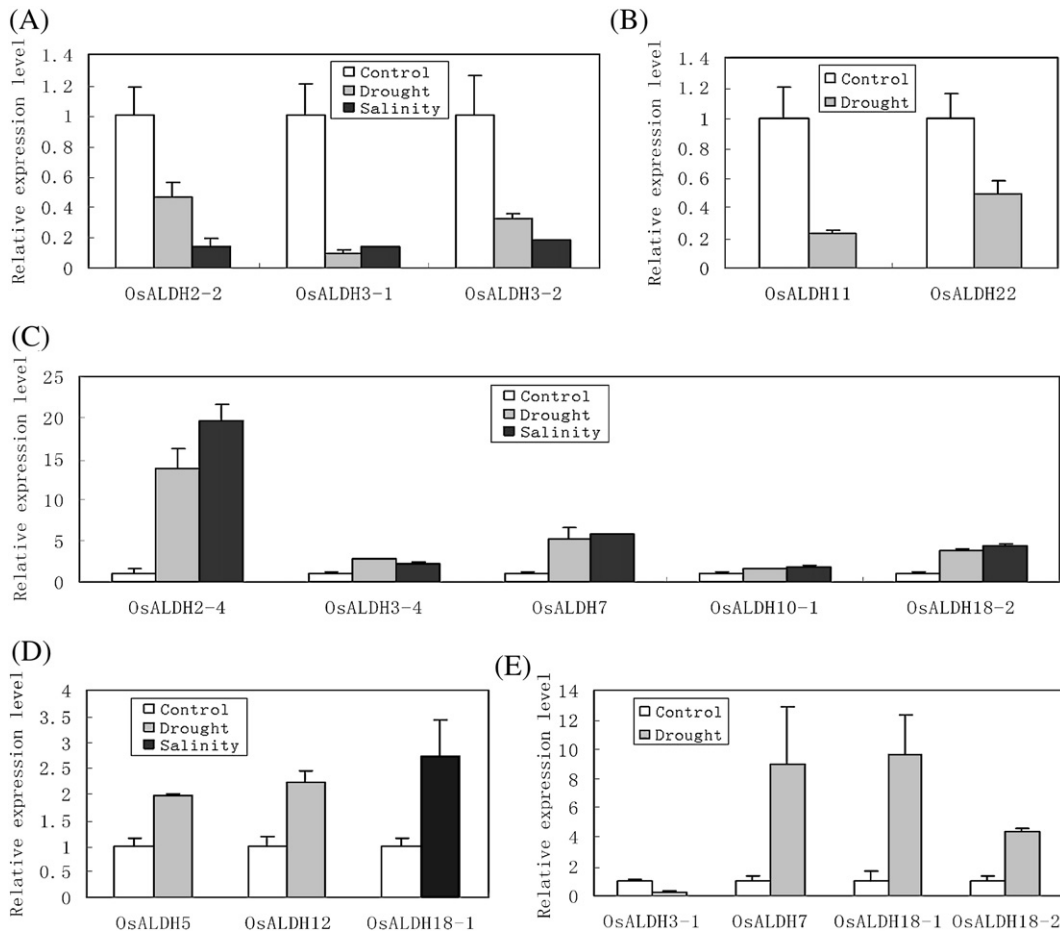


Fig. 5. Expression patterns of rice *ALDH* genes under stress treatments in young leaf at the seedling stage (A–D) and in flag leaf at the reproductive stage (E). The genes that showed difference in expression levels between control and stress-treated samples with statistical significance ($P < 0.05$) are shown in this figure. Expression level is shown relative to the level in control sample. Error bars indicate +SD of the mean of three biological replicates. (A) Genes down-regulated by both drought and high salinity stresses. (B) Genes down-regulated only by drought stress. (C) Genes up-regulated by both drought and high salinity stresses. (D) Genes up-regulated only by one tested stress (drought or high salinity). (E) Genes down- or up-regulated by drought stress in flag leaf.

In order to elucidate whether stress-responsiveness of the rice *ALDH* genes was mediated by ABA-dependent pathway, we also examined the responses of seven genes to ABA treatment in young leaf. Five genes were found to be significantly up-regulated (Fig. 6).

4. Discussion

4.1. Expansion of rice *ALDH* superfamily

ALDH proteins play essential roles in metabolism that are critical for development and response to environmental changes. Here, we

characterized the *ALDH* superfamily in rice. Twenty genes encoding members of 10 *ALDH* families were identified in rice. Database search and sequence analysis revealed that the same 10 families also presented in other examined angiosperm plants including maize, poplar and grape (data not shown).

The rice genome has endured two rounds of ancient large-scale genome duplications: one occurred before the divergence of dicots and monocots and the other occurred before the divergence of cereals (Vandepoel et al., 2003; Yu et al., 2005). Nine of the 20 rice *ALDH* genes can be mapped to the identified WGD regions (Supplementary Fig. 3). Tandem duplications were probably the cause for the rise of

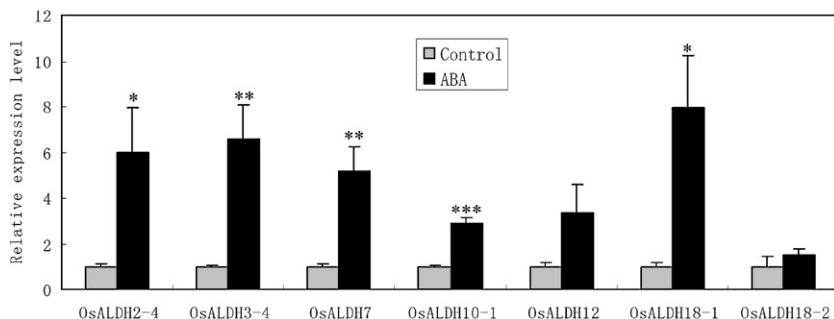


Fig. 6. Expression of rice *ALDH* genes under ABA treatment in young leaf. Expression level is shown relative to the level in control sample. Each bar represents mean +SD of three biological replicates. Single, double and triple asterisks indicate statistically significant difference with $P < 0.05$, $P < 0.01$ and $P < 0.001$ using Student's *t*-test, respectively.

two gene pairs (*OsALDH2-1/OsALDH2-2* and *OsALDH3-1/OsALDH3-2*). *OsALDH10-1* and *OsALDH10-2* were raised in the duplication postdating the divergence of rice and *Arabidopsis* (Fig. 1). But they cannot be mapped to the duplicated regions. The same duplications were also observed for *OsALDH2-1/OsALDH2-2/OsALDH2-3* and *OsALDH3-1/OsALDH3-2/OsALDH3-3*. The loss of conservation of gene content in the regions surrounding the paralogous genes implied that chromosome rearrangements should have occurred in rice genome. In fact, it has been suggested that massive gene losses and chromosome rearrangements, following large-scale genome duplications, have occurred in rice (Wang et al., 2005).

4.2. Diverse expression profiles of duplicated rice ALDH genes

Gene duplication is a major source of evolutionary novelty. Several models have been proposed to explain the evolutionary force and retention of duplicate genes (Sémon and Wolfe, 2007). The studies of some close-related ALDHs hint potential roles of functional specialization in the retention of duplicated genes (Liu and Schnable, 2002; Tsuji et al., 2003a, 2003b). Most rice ALDH genes were arisen from recent WGD or tandem duplications. Duplicated ALDH genes showed different tissue-specific expression patterns (Fig. 3; Supplementary Fig. 5), suggesting that the gene duplications supplied opportunities for the duplicates to be free from the functional constraint. We noticed that *OsALDH10-1* and *OsALDH10-2* exhibited similar tissue expression pattern both in our experimental result and the microarray data (Fig. 3; Supplementary Fig. 5). But only the expression of *OsALDH10-1* could be induced by stress treatments (Fig. 5C). Expression profiles for rice ALDH duplicates retrieved from MPSS database also showed great divergence (Supplementary Table S2). Based on these data, expression divergence and/or functional specialization were proposed to play an important role in the retention of rice ALDH duplicates.

Analysis based on microarray data indicated that expression of most duplicated genes in multiple-member families exhibited high tissue specificity (Fig. 4). The tissue-specific expression patterns for some genes implied their involvement in special developmental processes. Genes in single-member families tended to maintain invariable expression level across diverse organs because of the functional constraint, suggesting that these genes might participate in the basic metabolism of rice.

4.3. Potential rice ALDH genes involved in plant response to stresses

Environmental stresses including drought and high salinity are deleterious factors for crops survival and yield. These stressors can induce the rapid and excessive accumulation of reactive oxygen species (ROS) in plant cells (Sunkar et al., 2003). The ALDH activity increase is considered as an efficient defence strategy to eliminate the toxic aldehydes caused by ROS (Sunkar et al., 2003; Rodrigues et al., 2006). Here, we presented the first genome-wide expression study of rice ALDH genes under osmotic stress at the seedling stage and the reproductive stage. Previous study reveals that reprogramming of rice genome expression in response to drought and high salinity stresses is largely organ specific (Zhou et al., 2007). In our analysis, five genes (*OsALDH2-4*, *OsALDH3-4*, *OsALDH7*, *OsALDH18-2* and *OsALDH12*) were induced more than 2-fold in drought-stressed young leaf (Figs. 5C, D). Only three genes were up-regulated more than 2-fold in drought-stressed flag leaf (Fig. 5E). Among the eight genes that were induced by stress treatments in young leaf (Figs. 5C, D), none of them was up-regulated more than 2-fold by drought stress in panicle in the preliminary analysis (data not shown). These data suggested that the functions of ALDH genes in rice stress adaptation were organ specific.

The *OsALDH2-4* gene, which was previously reported to be significantly induced by submergence (Nakazono et al., 2000; Tsuji et al., 2003a), was up-regulated by stresses and ABA in young leaf,

indicating that this gene might participate in the detoxification of aldehyde generated in rice osmotic stress (Figs. 5C, 6). Previous study has established that over-expression of *ALDH3I1* could improve the tolerance of transgenic plant to diverse stresses (Sunkar et al., 2003). *OsALDH3-4*, which had the closest phylogenetic relationship with *ALDH3I1*, was up-regulated by drought, high salinity and ABA in the young leaf (Figs. 5C and 6). The two genes were suggested to function in a similar way in these two species after millions of years of divergence. *ALDH5F1* in *Arabidopsis* encodes SSADH that is involved in γ -aminobutyric acid (GABA) metabolism as part of the 'shunt' from Glu to GABA (Kirch et al., 2004). The activity of the GABA shunt in plants is rapidly enhanced in response to various biotic and abiotic stresses (Bouché et al., 2003). The drought-induced expression of *OsALDH5* hinted conserved function of it in rice (Fig. 5D). In angiosperms, genes that belong to family 7 could be induced by a wide range of stresses (Kirch et al., 2004; Kirch et al., 2005). Ectopical expression of ALDH7 genes in plants displays enhanced tolerance to drought, salinity and oxidative stress (Kotchoni et al., 2006; Rodrigues et al., 2006). The only gene of family 7 in rice, *OsALDH7*, was significantly induced by stresses at both the seedling stage and the reproductive stage (Figs. 5C and E). Another study has reported that expression of *OsALDH7* can be induced by blast fungus infection, ultraviolet and exogenous phytohormone application (Wu et al., 2007). These results implied that *OsALDH7* might function as a converging point for multiple signal transduction pathways. The proteins encoded by *OsALDH10-1* and *OsALDH10-2* are putative betaine ALDHs that catalyze the oxidation of betaine aldehyde to the compatible solute glycine betaine (Weretilnyk and Hanson, 1990). The expression of *OsALDH10-1* was induced by both drought and high salinity in young leaf, suggesting a possible role of it in rice stress adaptation (Fig. 5C). Proline metabolism and particularly Δ^1 -pyrroline-5-carboxylate (P5C) play a key role in stress response and accumulation of ROS (Borsani et al., 2005). *ALDH12A1* in *Arabidopsis* encodes P5CDH that catabolizes P5C. This gene is highly induced by exogenous proline application and salinity (Deuschle et al., 2001). Expression of *ALDH12A1* is regulated by a series of nat-siRNA processing steps in salt stress (Borsani et al., 2005). The drought-induced expression of *OsALDH12* suggested a potential role of it in rice stress adaptation (Fig. 5D). The exact role of *OsALDH12* in rice proline metabolism remains to be revealed. Genes in family 18 encode P5CS, an enzyme involved in the biosynthesis of proline (Igarashi et al., 1997). Recently, the role of *Arabidopsis P5CS1* in stress-induced proline synthesis and the function of *P5CS2* in embryo development are characterized in details (Székely et al., 2008). *OsALDH18-1* and *OsALDH18-2* showed different responses to drought stress and ABA treatment in young leaf. These two genes were suggested to function in partially overlapped pathways in response to osmotic stress. In addition, both of them showed drought-induced expression in flag leaf, supporting essential roles of them in drought stress adaptation at the reproductive stage (Fig. 5E). *ZmALDH22A1* is induced by abiotic stresses and ABA treatment in maize seedling root (Huang et al., 2008). Transgenic tobacco plants overexpressing *ZmALDH22A1* show elevated stress tolerance (Huang et al., 2008). But its orthologous genes in *Arabidopsis* and rice cannot be induced by osmotic stress (Kirch et al., 2005; Fig. 5B). More work is needed to reveal the exact role of *OsALDH22* in rice stress adaptation.

The stress-induced expression patterns of ALDH genes from different model plants have been studied. And it is also found that over-expression of some stress-induced ALDH genes could enhance the stress tolerance of transgenic plants. Increased attention is paid on the rice improvement. However, little has been done for rice ALDH genes. Our expression data suggest that some rice ALDH genes are potential candidates for improving rice tolerance to abiotic stresses, such as drought and high salinity. Extensive studies are also required to examine the exact biochemical roles of rice ALDHs in developmental process and stress tolerance in the future.

Acknowledgments

We thank Wenhao Hu, Xuehui Huang and Yongrui Wu for their technical assistance. This work was supported by grants from the Ministry of Science and Technology of China (2005CB120805 and 20060110A1031), the Chinese Academy of Sciences (KSCX2-YW-N-024), and the Shanghai Municipal Commission of Science and Technology.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2008.11.010.

References

- Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., Zhu, J.K., 2005. Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* 123, 1279–1291.
- Bouché, M., Fait, A., Bouchez, D., Møller, S.G., Fromm, H., 2003. Mitochondrial succinyl-semialdehyde dehydrogenase of the γ -aminobutyrate shunt is required to restrict levels of reactive oxygen intermediates in plants. *Proc. Natl. Acad. Sci. U. S. A.* 100, 6843–6848.
- Deuschle, K., Funck, D., Hellmann, H., Däschner, K., Binder, S., Frommer, W.B., 2001. A nuclear gene encoding mitochondrial Δ^1 -pyrroline-5-carboxylate dehydrogenase and its potential role in protection from proline toxicity. *Plant J.* 27, 345–355.
- Huang, W., et al., 2008. Significant improvement of stress tolerance in tobacco plants by overexpressing a stress-responsive aldehyde dehydrogenase gene from maize (*Zea mays*). *Plant Mol. Biol.* 68, 451–463.
- Igarashi, Y., Yoshida, Y., Sanada, Y., Yamaguchi-Shinozaki, K., Wada, K., Shinozaki, K., 1997. Characterization of the gene for Δ^1 -pyrroline-5-carboxylate synthetase and correlation between the expression of the gene and salt tolerance in *Oryza sativa* L. *Plant Mol. Biol.* 33, 857–865.
- Jain, M., Nijhawan, A., Tyagi, A.K., Khurana, J.P., 2006. Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochem. Biophys. Res. Commun.* 345, 646–651.
- Johnson, D.A., Thomas, M.A., 2007. The monosaccharide transporter gene family in *Arabidopsis* and rice: a history of duplications, adaptive evolution, and functional divergence. *Mol. Biol. Evol.* 24, 2412–2423.
- Kirch, H.H., Bartels, D., Wei, Y., Schnable, P.S., Wood, A.J., 2004. The ALDH gene superfamily of *Arabidopsis*. *Trends Plant Sci.* 9, 371–377.
- Kirch, H.H., Schlingensiepen, S., Kotchoni, S., Sunkar, R., Bartels, D., 2005. Detailed expression analysis of selected genes of the aldehyde dehydrogenase (ALDH) gene superfamily in *Arabidopsis thaliana*. *Plant Mol. Biol.* 57, 315–332.
- Kotchoni, S.O., Bartels, D., 2003. Water stress induces the up-regulation of a specific set of genes in plants: aldehyde dehydrogenase as an example. *Bulg. J. Plant Physiol. Special Issue*, 37–51.
- Kotchoni, S.O., Kuhns, C., Ditzer, A., Kirch, H.H., Bartels, D., 2006. Over-expression of different aldehyde dehydrogenase genes in *Arabidopsis thaliana* confers tolerance to abiotic stress and protects plants against lipid peroxidation and oxidative stress. *Plant Cell Environ.* 29, 1033–1048.
- Li, Y., Nakazono, M., Tsutsumi, N., Hirai, A., 2000. Molecular and cellular characterizations of a cDNA clone encoding a novel isozyme of aldehyde dehydrogenase from rice. *Gene* 249, 67–74.
- Lin, H., Zhu, W., Silva, J.C., Gu, X., Buell, C.R., 2006. Intron gain and loss in segmentally duplicated genes in rice. *Genome Biol.* 7, R41.
- Lindahl, R., 1992. Aldehyde dehydrogenases and their role in carcinogenesis. *Crit. Rev. Biochem. Mol. Biol.* 27, 283–335.
- Liu, F., Schnable, P.S., 2002. Functional specialization of maize mitochondrial aldehyde dehydrogenase. *Plant Physiol.* 130, 1657–1674.
- Liu, F., Cui, X., Horner, H.T., Weiner, H., Schnable, P.S., 2001. Mitochondrial aldehyde dehydrogenase activity is required for male fertility in maize. *Plant Cell* 13, 1063–1078.
- Lutts, S., Kinet, J.M., Bouharmont, J., 1995. Changes in plant response to NaCl during development of rice (*Oryza sativa* L.) varieties differing in salinity resistance. *J. Exp. Bot.* 46, 1843–1852.
- Marchitti, S.A., Brocker, C., Stagos, D., Vasiliou, V., 2008. Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. *Expert Opin. Drug Metab. Toxicol.* 4, 697–720.
- Nair, R.B., Bastress, K.L., Ruegger, M.O., Denault, J.W., Chapple, C., 2004. The *Arabidopsis thaliana* REDUCED EPIDERMAL FLUORESCENCE1 gene encodes an aldehyde dehydrogenase involved in ferulic acid and sinapic acid biosynthesis. *Plant Cell* 16, 544–554.
- Nakamura, T., et al., 1997. Expression of a betaine aldehyde dehydrogenase gene in rice, a glycinebetaine nonaccumulator, and possible localization of its protein in peroxisomes. *Plant J.* 11, 1115–1120.
- Nakazono, M., et al., 2000. Expression of a gene encoding mitochondrial aldehyde dehydrogenase in rice increases under submerged conditions. *Plant Physiol.* 124, 587–598.
- Nicholas, K.B., Nicholas Jr., H.B., Deerfield II, D.W., 1997. GeneDoc: analysis and visualization of genetic variation. *EMBNEW News* 4, 14.
- Perozich, J., Nicholas, H., Wang, B., Lindahl, R., Hempel, J., 1999. Relationships within the aldehyde dehydrogenase extended family. *Protein Sci.* 8, 137–146.
- Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29, e45.
- Ramakers, C., Ruijter, J.M., Deprez, R.H., Moorman, A.F., 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci. Lett.* 339, 62–66.
- Rodrigues, S.M., Andrade, M.O., Gomes, A.P., Damatta, F.M., Baracat-Pereira, M.C., Fontes, E.P., 2006. *Arabidopsis* and tobacco plants ectopically expressing the soybean antiquitin-like ALDH7 gene display enhanced tolerance to drought, salinity, and oxidative stress. *J. Exp. Bot.* 57, 1909–1918.
- Schauenstein, E., Esterbauer, H., Zollner, H., 1977. Aldehydes in Biological Systems: Their Natural Occurrence and Biological Activities. Pion, London.
- Sémon, M., Wolfe, K.H., 2007. Consequences of genome duplication. *Curr. Opin. Genet. Dev.* 17, 505–512.
- Shoja, V., Murali, T.M., Zhang, L., 2007. Expression divergence of tandemly arrayed genes in human and mouse. *Comp. Funct. Genom.* 2007, 60964.
- Skibbe, D.S., et al., 2002. Characterization of the aldehyde dehydrogenase gene families of *Zea mays* and *Arabidopsis*. *Plant Mol. Biol.* 48, 751–764.
- Sophos, N.A., Vasiliou, V., 2003. Aldehyde dehydrogenase gene superfamily: the 2002 update. *Chem. Biol. Interact.* 143–144, 5–22.
- Sunkar, R., Bartels, D., Kirch, H.H., 2003. Overexpression of a stress-inducible aldehyde dehydrogenase gene from *Arabidopsis thaliana* in transgenic plants improves stress tolerance. *Plant J.* 35, 452–464.
- Székel, G., et al., 2008. Duplicated P5CS genes of *Arabidopsis* play distinct roles in stress regulation and developmental control of proline biosynthesis. *Plant J.* 53, 11–28.
- Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.
- Tsuji, H., Meguro, N., Suzuki, Y., Tsutsumi, N., Hirai, A., Nakazono, M., 2003a. Induction of mitochondrial aldehyde dehydrogenase by submergence facilitates oxidation of acetaldehyde during re-aeration in rice. *FEBS Lett.* 546, 369–373.
- Tsuji, H., Tsutsumi, N., Sasaki, T., Hirai, A., Nakazono, M., 2003b. Organ-specific expressions and chromosomal locations of two mitochondrial aldehyde dehydrogenase genes from rice (*Oryza sativa* L.), ALDH2a and ALDH2b. *Gene* 305, 195–204.
- Vandepoele, K., Simillion, C., Van de Peer, Y., 2003. Evidence that rice and other cereals are ancient aneuploids. *Plant Cell* 15, 2192–2202.
- Vasiliou, V., Bairoch, A., Tipton, K.F., Nebert, D.W., 1999. Eukaryotic aldehyde dehydrogenase (ALDH) genes: human polymorphisms, and recommended nomenclature based on divergent evolution and chromosomal mapping. *Pharmacogenetics* 9, 421–434.
- Wang, X., Shi, X., Hao, B., Ge, S., Luo, J., 2005. Duplication and DNA segmental loss in the rice genome: implications for diploidization. *New Phytol.* 165, 937–946.
- Weretilnyk, E.A., Hanson, A.D., 1990. Molecular cloning of a plant betaine-aldehyde dehydrogenase, an enzyme implicated in adaptation to salinity and drought. *Proc. Natl. Acad. Sci. U. S. A.* 87, 2745–2749.
- Wu, C., Su, S., Peng, Y., 2007. Molecular cloning and differential expression of an aldehyde dehydrogenase gene in rice leaves in response to infection by blast fungus. *Biologia* 62, 523–528.
- Yoshida, A., Rzhetsky, A., Hsu, L.C., Chang, C., 1998. Human aldehyde dehydrogenase gene family. *Eur. J. Biochem.* 251, 549–557.
- Yu, J., et al., 2005. The genomes of *Oryza sativa*: a history of duplications. *PLoS Biol.* 3, e38.
- Zhou, J., et al., 2007. Global genome expression analysis of rice in response to drought and high-salinity stresses in shoot, flag leaf, and panicle. *Plant Mol. Biol.* 63, 591–608.