Gene 431 (2009) 86–94

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

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.

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 nucleophilies 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 *rf2* 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).

0378-1119/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2008.11.010

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/euk-blast/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 (*japanica* 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 μM 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-1α* (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 $\frac{(E_{\text{target}})^{\Delta CP_{\text{target}}(control-sample)}}{(E_{\text{ref}})^{\Delta CP_{\text{ref}}(control-sample)}}$ (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