Comparative study of the aldehyde dehydrogenase (ALDH) gene superfamily in the glycophyte Arabidopsis thaliana and Eutrema halophytes

Quancan Hou and Dorothea Bartels*
Institute of Molecular Physiology and Biotechnology of Plants, University of Bonn, Kirschallee 1, D-53315 Bonn, Germany
*For correspondence. E-mail dbartels@uni-bonn.de

Received: 13 May 2014 Returned for revision: 23 May 2014 Accepted: 28 May 2014

INTRODUCTION

Environmental stresses such as drought and high salinity induce the rapid generation of reactive oxygen species (ROS), which subsequently cause excessive accumulation of aldehydes in plant cells. Aldehydes are intermediates in a range of metabolic pathways, which may contribute to salinity tolerance. Aldehyde dehydrogenases are found throughout all taxa and are involved in a broad range of metabolic functions in-cluding participating in intermediary metabolism such as amino acid and retinoic acid metabolism or generating osmoprotec-tants, such as glycine betaine (Ishitani et al., 1995; Xing and Rajashekar, 2001). Aldehyde dehydrogenase enzymes produce NADPH and NADH in their enzymatic reactions and thus may contribute to balancing redox equivalents.

Aldehyde dehydrogenases are found throughout all taxa and have been classified into 24 distinct families based on protein sequence identities. These families are numbered according to the criteria from the ALDH Gene Nomenclature Committee (AGNC) (Vasiliou et al., 1999). The plant ALDH superfamily contains 14 distinct families: ALDH2, ALDH3, ALDH5, ALDH6, ALDH7, ALDH10, ALDH11, ALDH12, ALDH18, ALDH19, ALDH21, ALDH22, ALDH23 and ALDH24. The families ALDH10, ALDH11, ALDH12, ALDH18, ALDH19, ALDH21, ALDH22, ALDH23 and ALDH24 are specific to plants, whereas the remaining families have mammalian orthologues (Table 1). There are a few ALDH genes identified in algae species: seven ALDH genes in the colonial algae Volvox carteri (Brocker et al., 2013), and six and nine ALDH genes in the unicellular algae Ostreococcus tauri and Chlamydomonas reinhardtii, respectively (Wood and Duff, 2009). The ALDH gene numbers increased in the moss Physcomitrella patens, which contains 21 members including all plant ALDH gene families except for ALDH22 (Wood and Duff, 2009). Physcomitrella patens has gained two novel gene families, ALDH21 and ALDH23,
<table>
<thead>
<tr>
<th>Taxonomy</th>
<th>Species</th>
<th>ALDH family</th>
<th>Total no.</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herbaceous dicotyledon</td>
<td><em>A. thaliana</em></td>
<td>– 3 3 – 1 1 1 – – 2 1 1 – – – – 2 – – – 1 – –</td>
<td>16</td>
<td>Kirch <em>et al.</em> (2004)</td>
</tr>
<tr>
<td>Herbaceous dicotyledon</td>
<td><em>E. parvulum</em></td>
<td>– 3 3 – 1 1 1 – – 2 1 1 – – – – 2 – – – 1 – –</td>
<td>16</td>
<td>This review</td>
</tr>
<tr>
<td>Herbaceous dicotyledon</td>
<td><em>E. salsugineum</em></td>
<td>– 3 4 – 1 1 1 – – 2 1 1 – – – – 2 – – – 1 – –</td>
<td>17</td>
<td>This review</td>
</tr>
<tr>
<td>Herbaceous dicotyledon</td>
<td><em>G. max</em></td>
<td>– 5 1 – – – 4 – 6 2 – – – – – – – – – – – –</td>
<td>18</td>
<td>Kotchoni <em>et al.</em> (2012)</td>
</tr>
<tr>
<td>Herbaceous monocotyledon</td>
<td><em>O. sativa</em></td>
<td>– 5 5 – 1 1 1 – – 2 1 1 – – – – 2 – – – 1 – –</td>
<td>20</td>
<td>Gao and Han (2009)</td>
</tr>
<tr>
<td>Herbaceous monocotyledon</td>
<td><em>Z. mays</em></td>
<td>– 6 5 – 2 1 1 – – 3 1 1 – – – – 2 – – – 1 – –</td>
<td>23</td>
<td>Jimenez-Lopez <em>et al.</em> (2010); Zhou <em>et al.</em> (2012)</td>
</tr>
<tr>
<td>Herbaceous monocotyledon</td>
<td><em>S. bicolor</em></td>
<td>– 5 4 – 1 1 1 – – 2 1 1 – – – – 2 – – – 1 – –</td>
<td>19</td>
<td>Sophos and Vasilious (2003); Brocker <em>et al.</em> (2013)</td>
</tr>
<tr>
<td>Woody plant</td>
<td><em>V. vinifera</em></td>
<td>– 5 4 – 3 3 2 – – 2 2 1 – – – – 2 – – – 1 – –</td>
<td>25</td>
<td>Zhang <em>et al.</em> (2012)</td>
</tr>
<tr>
<td>Woody plant</td>
<td><em>M. domestica</em></td>
<td>– 13 7 – 2 2 2 – – 2 3 2 – – – – 4 – – – 2 – –</td>
<td>39</td>
<td>Li <em>et al.</em> (2013)</td>
</tr>
<tr>
<td>Woody plant</td>
<td><em>P. trichocarpa</em></td>
<td>– 4 6 – 1 4 2 – – 2 3 1 – – – – 2 – – – 1 – –</td>
<td>26</td>
<td>Brocker <em>et al.</em> (2013)</td>
</tr>
<tr>
<td>Fern</td>
<td><em>S. moellendorfii</em></td>
<td>– 6 2 – 1 1 1 – – 1 6 1 – – – – 1 – – 1 1 2 –</td>
<td>24</td>
<td>Brocker <em>et al.</em> (2013)</td>
</tr>
<tr>
<td>Moss</td>
<td><em>P. patens</em></td>
<td>– 2 5 – 2 1 1 – – 1 5 1 – – – – 1 – – 1 1 –</td>
<td>21</td>
<td>Wood and Duff (2009)</td>
</tr>
<tr>
<td>Algae</td>
<td><em>O. tauri</em></td>
<td>– – 1 – 1 – – – 1 1 – – – – 1 – – 1 – – 1 –</td>
<td>9</td>
<td>Wood and Duff (2009)</td>
</tr>
<tr>
<td>Algae</td>
<td><em>C. reinhardtii</em></td>
<td>– 1 – – 1 1 – – – 1 1 1 – – – – 1 – – 1 – –</td>
<td>9</td>
<td>Wood and Duff (2009)</td>
</tr>
<tr>
<td>Algae</td>
<td><em>V. carteri</em></td>
<td>– 1 – – 1 – – – 1 1 1 – – – – 1 – – 1 – –</td>
<td>7</td>
<td>Brocker <em>et al.</em> (2013)</td>
</tr>
<tr>
<td>Mammal</td>
<td><em>H. sapiens</em></td>
<td>6 1 4 1 1 1 1 1 1 – – – – 1 – – 1 – – – –</td>
<td>19</td>
<td>Yoshida <em>et al.</em> (1998); Jackson <em>et al.</em> (2011)</td>
</tr>
<tr>
<td>Mammal</td>
<td><em>M. musculus</em></td>
<td>7 1 5 1 1 1 1 1 1 – – – – 1 – – 1 – – – –</td>
<td>21</td>
<td>Jackson <em>et al.</em> (2011)</td>
</tr>
</tbody>
</table>

‘–’ no member of the ALDH gene family was identified in the corresponding species.
Eutrema species have a close phylogenetic relationship but with extremely divergent adaptations. The E. salsugineum and E. parvulum genome is approx. 50 and 15 % larger than that of A. thaliana, respectively. The higher content of transposable elements in E. salsugineum is the main reason for its genome expansion besides tandem duplications of single-copy genes. Comparative genomic analysis showed that A. thaliana and E. salsugineum share 95-2 and 93-7 % of all their gene families, respectively (Wu et al., 2012).

We carried out a genome-wide identification of ALDH genes, to study their evolutionary relationships and to investigate their expression profiles in response to stress. Comparing Eutrema and A. thaliana should provide us with some insight into the role of ALDH in abiotic stress tolerance. By searching public databases, we identified 16 ALDH genes from E. parvulum and 17 from E. salsugineum. ALDH genes from both species have been classified and have been assigned to ten subfamilies. We further surveyed genomic organization, sequence homology and subcellular localization of Eutrema ALDH genes. Expression profiles of stress-related ALDH genes were compared between A. thaliana and E. salsugineum under salinity stress conditions.

MATERIALS AND METHODS

Database searches and annotation of ALDH genes

The protein databases of Eutrema parvulum were searched in Thellungiella (http://thellungiella.org/) with BLASTP, and of Eutrema salsugineum in Phytozome (http://phytozome.jgi.doe.gov/) with TBALSTN. The following ALDH protein sequences were used as queries for the searches: 16 A. thaliana ALDHs (ALDH2B4, ALDH2B7, ALDH2C4, ALDH3H1, ALDH3I1, ALDH3F1, ALDH5F1, ALDH6B2, ALDH7B4, ALDH10A9, ALDH11A3, ALDH12A1, ALDH18B1, ALDH18B2 and ALDH22A1), three Selaginella moellendorffii ALDHs (ALDH21A1, ALDH23B1 and ALDH23B2), one Solanum lycoctonum ALDH (ALDH21A1), one Solanum lycopersicum ALDH (ALDH19) and one Chlamydomonas reinhardtii ALDH (ALDH24A1). In addition, the same queries were used to search against the E. salsugineum protein database obtained from the website http://omicslab.genetics.ac.cn/resources.php with BLASTP by using the perfectBLAST2.0 program (Santiago-Sotelo and Ramirez-Prado, 2012). All sequences with an E-value <1e-6 were selected for manual inspection. The Pfam domain PF00171 (ALDH family), PF00070 (ALDH cysteine active site) and PF00687 (ALDH glutamic acid active site) searches (http://pfam.xfam.org/) were performed to confirm the candidate sequences as ALDH proteins. The confirmed Eutrema ALDH proteins were annotated using the criteria established by the AGNC (Vasiliou et al., 1999). Amino acid sequences that are >40 % identical to identified ALDH sequences comprise a family, and sequences that are >60 % identical comprise a protein subfamily, while sequences with <40 % identity are considered to be a new family.

Sequence alignments and phylogenetic analyses

Multiple protein alignments of ALDH protein sequences from A. thaliana, E. parvulum and E. salsugineum were performed using the ClustalW program of MEGA5.0 (Tamura et al., 2011).
<table>
<thead>
<tr>
<th>Gene name (Arabidopsis)</th>
<th>Accession no.</th>
<th>Locus</th>
<th>Length (aa)*</th>
<th>Gene name (E. parvulum)</th>
<th>Accession no.</th>
<th>Locus</th>
<th>Length (aa)**</th>
<th>Gene name (E. salsugineum)</th>
<th>Accession no.</th>
<th>Locus</th>
<th>Phytome ID</th>
<th>Length (aa)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtALDH2B4</td>
<td>NP_190383.1</td>
<td>AT3G48000</td>
<td>538</td>
<td>EpALDH2B4</td>
<td>AFAN01000020-1</td>
<td>Tp5g13980</td>
<td>538</td>
<td>EsALDH2B4</td>
<td>XM_006404245.1</td>
<td>Ts5g15070</td>
<td>Thhalv10010271m</td>
<td>538</td>
</tr>
<tr>
<td>AtALDH2B7</td>
<td>NP_564204.1</td>
<td>AT1G23800</td>
<td>534</td>
<td>EpALDH2B7</td>
<td>AFAN01000001-1</td>
<td>Tp1g21180</td>
<td>540</td>
<td>EsALDH2B7</td>
<td>XM_006418675.1</td>
<td>Ts3g24340</td>
<td>Thhalv10002492m</td>
<td>500</td>
</tr>
<tr>
<td>AtALDH2C4</td>
<td>NP_566749.1</td>
<td>AT3G24503</td>
<td>501</td>
<td>EpALDH2C4</td>
<td>AFAN01000010-1</td>
<td>Tp3g22290</td>
<td>502</td>
<td>EsALDH2C4</td>
<td>XM_006393677.1</td>
<td>Ts1g33100</td>
<td>Thhalv10011438m</td>
<td>483</td>
</tr>
<tr>
<td>AtALDH2H1</td>
<td>NP_175081.1</td>
<td>AT1G44170</td>
<td>484</td>
<td>EpALDH2H1</td>
<td>AFAN01000005-1</td>
<td>Tp1g33460</td>
<td>482</td>
<td>EsALDH2H1</td>
<td>XM_006412189.1</td>
<td>Ts7g33420</td>
<td>Thhalv10024844m</td>
<td>547</td>
</tr>
<tr>
<td>AtALDH3H1</td>
<td>NP_567962.1</td>
<td>AT4G34240</td>
<td>530</td>
<td>EpALDH3H1</td>
<td>AFAN01000039-1</td>
<td>Tp7g32020</td>
<td>548</td>
<td>EsALDH3H1</td>
<td>XM_006411952.1</td>
<td>Ts7g35460</td>
<td>Thhalv10025045m</td>
<td>485</td>
</tr>
<tr>
<td>AtALDH3F1</td>
<td>NP_195348.2</td>
<td>AT4G36250</td>
<td>484</td>
<td>EpALDH3F1</td>
<td>AFAN01000040-1</td>
<td>Tp7g33990</td>
<td>484</td>
<td>EsALDH3F1</td>
<td>XM_006389827.1</td>
<td>Ts5g36850</td>
<td>Thhalv10018394m</td>
<td>527</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EsALDH5F1</td>
<td>XM_006409602.1</td>
<td>Ts3g28400</td>
<td>Thhalv10022637m</td>
<td>538</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EsALDH6B2</td>
<td>XM_006392654.1</td>
<td>Ts1g42160</td>
<td>Thhalv10011405m</td>
<td>508</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EsALDH7B4</td>
<td>XM_006411149.1</td>
<td>Ts4g25140</td>
<td>Thhalv10016322m</td>
<td>716</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EsALDH7B2</td>
<td>XM_006403379.1</td>
<td>Ts5g07350</td>
<td>Thhalv10010150m</td>
<td>733</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EsALDH7B2</td>
<td>XM_006407856.1</td>
<td>Ts3g05680</td>
<td>Thhalv10020345m</td>
<td>596</td>
</tr>
</tbody>
</table>

*Genome size in mega bases.
†Length of the deduced amino acid sequence.
‡Length of the amino acid deduced from two datasets (see Supplementary Data Fig. S2).
The phylogenetic tree was constructed in MEGA5.0 using the Neighbor-Joining method and the bootstrap test was replicated 500 times.

**Exon–intron structure analyses of ALDH genes**

The exon–intron structures of ALDH genes were determined from the alignments of the coding sequences to the corresponding genomic sequences using the program Spidey in NCBI (http://www.ncbi.nlm.nih.gov/spidey/). The diagrams of exon–intron structures were obtained using the program FancyGene (Rambaldi and Ciccarelli, 2009).

**Prediction of transcription factor binding sites**

Nucleotides –1000 to +200 relative to the transcription start site of the AtALDH10A8 and EsALDH10A8 genes were retrieved from the Phytozome v10 database (http://phytozome.jgi.doe.gov/) and taken as promoter sequences. The promoter sequences were scanned using the MATCH program of the TRANSFAC® Professional Suite subscribed from BIOBASE (http://www.biobase-international.com).

**Plant materials, growth conditions and treatments**

Arabidopsis thaliana (Col-0) and E. salsugineum (Shandong) were used for all experiments. Seeds for Arabidopsis were obtained from the Nottingham Arabidopsis Stock Centre, UK. Seeds for E. salsugineum were originally obtained from R. Bressan (Purdue University, USA). Seeds were germinated and plants were grown in plastic pots containing potting soil under 120–150 μE m⁻² s⁻¹ light at 22 °C with a day/night cycle of 8/16 h. Six-week-old Arabidopsis and E. salsugineum plants were regularly irrigated with NaCl solutions (100, 300 mM NaCl) for 5 d while well-watered plants served as controls. Leaves were collected 2 and 5 d after starting irrigation with NaCl. All samples were frozen in liquid nitrogen and stored at –80 °C until further analyses.

**RNA extraction and reverse transcription–PCT (RT–PCR) analysis**

Total RNA extraction was performed according to Valenzuela-Avendaño et al. (2005). For RT–PCR analyses, 2 μg of total RNAs were treated with 10 U of RNase-free DNase I (Roche, Mannheim, Germany) in a 10 μL reaction containing 1× DNase I buffer (20 mM Tris–HCl pH 8.4, 50 mM KCl and 2 mM MgCl₂) at 37 °C for 30 min. Then, 1 μL of 25 mM EDTA was added and the reaction was heated at 65 °C for 15 min to deactivate the DNase I. First-strand cDNA synthesis was performed using the RevertAid™ H Minus First Strand cDNA Synthesis Kit (Fermentas, Burlington, Canada) following the protocol provided by the supplier.

**RESULTS**

**Genome-wide identification of ALDH gene families in the Eutrema halophytes**

Here, we identified 16 ALDH genes from E. parvulum using the available genome sequence (Tables 1 and 2). In the case of E. salsugineum, the same ecotype (ecotype: Shandong) had been sequenced twice by two different groups (Wu et al., 2012; Yang et al., 2013). When ALDH genes were identified from the two genome datasets, different results were obtained. Seventeen putative ALDH genes were identified from the dataset of Yang et al. (2013) and 19 putative ALDH genes were identified from the dataset of Wu et al. (2012). The datasets differ in three putative orthologues of ALDH2C4 which are designated as Ts3g24340, T5g11880 and Ts5g11870 in the dataset of Wu et al. (2012) and only one orthologue found in the dataset of Yang et al. (2013) and designated as Thhalv10002492 m in Phytozome. The discrepancy is most probably due to an assembly error. The question was experimentally addressed by amplifying the ALDH2C4 gene from genomic DNA and cDNA of E. salsugineum. Only one amplicon was obtained from genomic DNA but three different fragments were amplified from cDNA. These results demonstrate that there is one ortholog of ALDH2C4 in the E. salsugineum genome, which generates three different transcripts most probably by alternative splicing. The existence of the transcripts was verified in RT–PCRs (Fig. 1). All transcripts are constitutively expressed and the expression does not change under different salt conditions. The EsALDH2C4.1 transcript is abundantly expressed and transcripts EsALDH2C4.2 and EsALDH2C4.3 are only weakly expressed (Figs 1 and 4).

Although Ts3g24340 and Thhalv10002492 m which represent the EsALDH2C4 gene have the same DNA sequence in the two datasets, the protein sequences are not identical, because different open reading frames (ORF) were predicted from the two datasets. Our PCR analysis suggests that Thhalv10002492 m is the correct prediction as the corresponding transcript was detected from the cDNA whereas a fragment corresponding to Ts3g24340 could not be amplified. Similar ORF prediction discrepancies were observed for the orthologues of ALDH11A3 and ALDH12A1 in E. salsugineum which are designated as Ts4g04290, Ts2g30990 and Thhalv1000267 m, Thhalv10003935 m, respectively, in the two datasets.
In addition, two fragments were identified that resemble part of ALDH7B4 and ALDH10A8 but with stop codons designated as Ts4q13390 and Ts7q03810, respectively, in the genome dataset of Wu et al. (2012), but no annotation exists in the genome dataset of Yang et al. (2013). It was possible to amplify two cDNA fragments using Ts4q13390-specific primers. DNA sequencing analysis showed that the two transcripts were derived by alternative splicing, but no correct ORFs exist. Our results do not support the prediction of Wu et al. (2012) (Supplementary Data Fig. S1), and the two fragments with homologies to ALDH7B4 and ALDH10A8 are considered as pseudogenes.

Using the nomenclature criteria (Vasiliou et al., 1999), the ALDH proteins from the two Eutrema species fall into ten families based on their sequence identities. In both Eutrema species, six families are represented by a single gene (families 5, 6, 7, 11, 12 and 18), whereas the remaining four families contain multiple members (families 2, 3, 10 and 18). To classify the ALDH genes, a formal name was given for each ALDH gene of the two Eutrema species following the suggested nomenclature system (Table 2). Following the nomenclature criteria, sequences that share ≥60% identity should be grouped into a subfamily (Vasiliou et al., 1999); therefore, the parologue of EsALDH3F1 was designated as EsALDH3F2.

**Evolutionary relationships of ALDH genes between Arabidopsis and Eutrema**

As organisms evolve, the genetic materials accumulate mutations over time, causing phenotypic changes and speciation. Therefore, molecular phylogenetics is an informative way to reconstruct evolutionary history. To explore the function and evolutionary process of Eutrema ALDH genes, an unrooted phylogenetic tree was constructed from the alignment of full-length protein sequences with Arabidopsis. The two pseudogenes were excluded from the phylogenetic tree construction as they do not encode complete proteins. All of the EpALDHs, EsALDHs and AtALDHs fall into ten families with well-supported bootstrap values (Fig. 2A). The topology is similar to those trees which were constructed with ALDH proteins from other plant species (Yang et al., 2013). Therefore, family 18 has a distant phylogenetic relationship between Arabidopsis and Eutrema. Longer introns were found in several Eutrema species (Fig. 3). Genes within the same subfamily such as subfamily ALDH2B, ALDH10A and ALDH18B show >60% sequence identity and have the same exon–intron structure in all three species (Fig. 2B). The high degree of sequence identity and similar exon–intron structures of ALDH genes within each family suggest that gene duplications most probably occurred before A. thaliana and Eutrema were separated. In contrast to the ALDH superfamily in other analysed species (Gao and Han, 2009; Zhang et al., 2012; Zhou et al., 2012), there is no intron loss in Eutrema, indicating a close evolutionary relationship between Arabidopsis and Eutrema. However, different intron lengths were observed in almost all families. Longer introns were found in several E. salsugineum ALDH genes, especially for EsALDH2C4 (Fig. 3). This finding is consistent with reports that E. salsugineum and A. thaliana genes have similar average exon lengths whereas the average intron length is about 30% larger in E. salsugineum than in Arabidopsis (Wu et al., 2012). The new gene EsALDH3F2 has one more exon than its parologue EsALDH3F1 and the orthologues AtALDH3F1 and EpALDH3F1, suggesting that EsALDH3F2 diverged from EsALDH3F1, which is consistent with the phylogenetic analysis (Fig. 2). The divergence in gene structure is mainly due to differences in intron sequences while the ALDH protein sequences are highly conserved in these three species, indicating that the enzymes kept their roles during evolution.

**Expression profiles of stress-related ALDH genes from A. thaliana and E. salsugineum under salt stress conditions**

Expression of the abi10 stress-related genes ALDH3H1, ALDH3I1, ALDH7B4 and two betaine aldehyde dehydrogenase
FIG. 3. Phylogenetic relationship and exon–intron structures of three selected *A. thaliana*, *E. parvulum* and *E. salsugineum* ALDH genes. The genes *ALDH2C4*, *ALDH6B2* and *ALDH3F* were selected for this analysis. (A) The phylogenetic relationship was constructed using MEGA 5.0. (B) Exons are represented by grey boxes and line angles connecting two boxes represent introns. The scale bars at the top are scaled by number of nucleotides. Exons and introns are drawn to scale by using FancyGene (http://bio.ieo.eu/fancygene/).
genes ALDH10A8 and ALDH10A9 has been investigated in Arabidopsis (Kirch et al., 2005; Missihoun et al., 2010). To obtain information of these ALDH genes in halophytes, the expression was examined in response to salt stress in both A. thaliana and E. salsugineum after 2 and 5 d of salt treatments (Fig. 4A) using RT–PCR with a variable number of amplification cycles. First, expression was recorded by using 25 cycles, and then the cycle numbers were decreased from 25 to 22 for highly expressed transcripts and increased from 25 to 30 cycles for weakly expressed transcripts (Fig. 4B).

Under the salt treatments used here, ALDH3H1 transcripts were constitutively expressed in Arabidopsis and Eutrema; this result is different from those of other reports (Kirch et al., 2005; Kotchoni et al., 2006). ALDH3H1 shows the lowest expression of all examined transcripts and it is induced at high salt concentrations in Arabidopsis and Eutrema, with a slightly lower level....
in Eutrema (Fig. 4B). ALDH7B4 accumulates in response to salt similarly to ALDH3H1, but it is more abundant than ALDH3H1. Expression of ALDH7B4 is induced by salt in Arabidopsis and Eutrema, but the induction is delayed and at a lower level in Eutrema (Fig. 4B).

Under the stress conditions used here, ALDH10A9 is constitutively expressed in both A. thaliana and E. salsugineum, although it was reported to be weakly induced by different abiotic stressors (Missihoun et al., 2010). ALDH10A8 shows a differential accumulation in A. thaliana and E. salsugineum. It was constitutively expressed under low salt stress conditions in A. thaliana and downregulated in high salinity conditions, while it was upregulated in 600 mM NaCl in E. salsugineum.

Genomic organization and in silico comparison of putative promoter regions of the ALDH10A8 gene in A. thaliana and E. salsugineum

Because ALDH10A8 is differentially expressed in Arabidopsis and E. salsugineum, the gene organization and putative promoter regions were examined. Similar spatial organizations of the ALDH10A8 gene were found in the genome of the two species. The upstream and downstream neighbouring genes of the ALDH10A8 gene encode an ERF/AP2 transcription factor and a homologue of ADP-glucose pyrophosphorylase, respectively, in both species (Fig. 5). The intergenic region between ALDH10A8 and the upstream gene is much longer in E. salsugineum than in Arabidopsis, with 3959 and 1665 bp, respectively (Fig. 5). In A. thaliana, the intergenic region comprises the gene At1g74929 which encodes a putative protein of 33 amino acids. The sequence is very similar in Eutrema but the stop codon is absent (Supplementary Data Fig. S3A).

To investigate whether sequence conservation exists in the promoter regions of ALDH10A8, the sequences of nucleotides −1000 to +200 relative to the transcription start site of the ALDH10A8 gene were compared in Arabidopsis and E. salsugineum. The sequences were retrieved from the Phytozome v10 database (http://phytozome.jgi.doe.gov/) and aligned using the Align X tool of the Vector NTI Advance® 11 software. No regions were observed that are conserved between the two sequences (Supplementary Data Fig. S3B).

Since ALDH10A8 showed different expression profiles in Arabidopsis and Eutrema under high salt concentrations (Fig. 4B), the 1000 bp promoter sequences upstream and 200 bp downstream of the transcription start site were scanned using the MATCH program of the TRANSFAC® Professional Suite to predict transcription factor binding sites and to compare their distribution in Arabidopsis and Eutrema. Forty-one transcription factor binding sites were predicted in the 1200 bp Arabidopsis sequence and 20 sites were predicted for the E. salsugineum promoter (Supplementary Data Fig. S4). Although identical sites are identified in both sequences (Fig. S4), the distribution is very different and no common promoter architecture can be predicted. The 1200 bp promoter sequence analysed here covers nearly the complete intergenic region between AtALDH10A8 and the upstream adjacent gene. However, E. salsugineum has a much longer intergenic region between EsALDH10A8 and the upstream adjacent gene. Therefore, there may be additional binding sites beyond the analysed 1200 bp. The non-conserved promoter region implies that
AtALDH10A8 and EsALDH10A8 are probably regulated in different ways, which may explain why AtALDH10A8 is differentially expressed in Arabidopsis and E. salsugineum under high salt concentrations.

Sub-cellular localization of ALDH proteins and overview of substrates used by ALDH enzymes

Aldehyde dehydrogenase enzymes are localized in different cellular compartments which may be correlated with functional specialization. We searched published data and used Euk-mPLoc 2.0 (Chou and Shen, 2010) to compile the subcellular localizations of the Arabidopsis and Eutrema ALDH proteins (Table 3).

According to predictions, ALDH2 (EC 1.2.1.3) family members are targeted to mitochondria and the cytoplasm (Skibbe et al., 2002; Kirch et al., 2004) which is consistent with human homologues (Braun et al., 1987). Arabidopsis and E. parvula as well as E. salsugineum contain two highly similar members of ALDH2B (ALDH2B4 and ALDH2B7) targeted to mitochondria, and ALDH2C4, a member of subfamily ALDH2C, is localized in the cytoplasm. In Arabidopsis, AtALDH2B4 was shown to be involved in the pyruvate dehydrogenase pathway (Wei et al., 2009). ALDH2C4 plays a role in ferulic and sinapic acid biosynthesis which are important compounds for maintaining cell wall strength (Nair et al., 2004). Kinetic studies of two maize enzymes ALDH2B2 and ALDH2B5 show that ALDH2B5 only oxidizes short-chain aliphatic aldehydes and ALDH2B2 oxidizes both aliphatic and aromatic aldehydes, indicating that ALDH2B2 has a broader substrate spectrum (Liu and Schnable, 2002; Jimenez-Lopez et al., 2010). Escherichia coli complementation assays showed that ALDH2C4 has a broader substrate spectrum than the two subfamily ALDH2B proteins, as ALDH2C4 oxidizes sinapaldehyde and coniferaldehyde as well as L-lactaldehyde, while ALDH2B does not use these substrates (Kirch et al., 2004; Nair et al., 2004).

ALDH2 (EC 1.2.1.3) enzymes both prefer medium to long-chain saturated aldehydes (C-6 to C-12) as substrates in enzymatic reactions (Stiti et al., 2004; 2006; Nair et al., 2004). Recombinant AtALDH3H1 and AtALDH3I1 proteins which are localized in the cytoplasm and in the chloroplast, respectively (Kotchoni et al., 2011). Reombinant AtALDH3H1 and AtALDH3I1 enzymes both prefer medium to long-chain saturated aldehydes as substrates in enzymatic reactions (Stiti et al., 2011). However, ALDH3I1 can use both NADP⁺ and NAD⁺ as co-enzyme, whereas ALDH3H1 only uses NAD⁺ as co-enzyme due to an isoecine instead of a valine in its co-enzyme binding site (Stiti et al., 2011). EsALDH3F1 and EpALDH3F1 and the new ALDH member EsALDH3F2 are predicted to localize to the cytoplasm, which is consistent with experimental data of the Arabidopsis orthologues (Stiti et al., 2011).

ALDH7, known as Δ-1-piperideine-6-carboxylate dehydrogenase (P6CDH; EC 1.2.1.31) or antiquitin, is highly conserved across all domains. Studies showed that it can metabolize various...
substrates such as acetaldehyde, glyceraldehyde and malondialdehyde in rice (Shin et al., 2008), or aminoacidic semialdehyde and guanidinobutyraldehyde in maize and pea (Brocker et al., 2013). The Arabidopsis enzyme is localized in the cytoplasm, which is consistent with the prediction for the E. parvulum homologue (Stiti et al., 2011). The homologue in E. salsugineum is predicted to target to mitochondria, however. Since ALDH7B is highly conserved throughout evolution, it will be interesting to investigate whether a minor sequence difference can change the localization from the cytoplasm to the mitochondria in E. salsugineum and whether this has any functional implications.

There are two members of the ALDH10 family in both Arabidopsis and Eutrema. They are known as aminoaidehyde dehydrogenases (AMADHs; EC 1.2.1.19) or betaine aldehyde dehydrogenases (BADHs; EC 1.2.1.9). Most of the members in this family show preference for aminoaidehyde substrates over betaine aldehyde. However, in vitro activity assays of recombinant ALDH10A9 exhibited both betaine aldehyde and aminoaidehyde dehydrogenase (4-aminobutanol and 3-aminopropanal) activities (Missihoun et al., 2010). It is not conclusive whether A. thaliana or E. salsugineum accumulate glycine betaine (Lugan et al., 2010; Chen and Murata, 2011); this suggests that ALDH10A8 and ALDH10A9 may be involved in as yet unknown metabolic pathways. In Arabidopsis, ALDH10A8 and ALDH10A9 proteins were shown to be localized in leucoplasts and peroxisomes, respective-ly (Missihoun et al., 2010).

The ALDH18 family contains Δ-1-pyrroline-5-carboxylate synthase enzymes (P5CSs; EC 1.2.1.41 and EC 2.7.2.11). Gene expression is induced by dehydration and salt stress, and subsequently proline accumulates (Hu et al., 1992; Yoshiba et al., 1997). In mammals, ALDH18A1 converts glutamate to pyrroline-5-carboxylate, which is an intermediate in the biosynthesis of ornithine, proline and arginine (Hu et al., 1992; Marchitti et al., 2008). In plants, ALDH18B catalyses the first two steps in proline biosynthesis in moth bean (Vigna aconitifolia) (Hu et al., 1992). All plant species analysed contain two ALDH18 members which are predicted to be localized in the cytoplasm, while the human homologue is localized in the inner mitochondrial membrane.

The ALDH families 5, 6, 11, 12 and 22 are represented by one member in Arabidopsis and Eutrema. ALDH5F1, a succinic semialdehyde dehydrogenase (SSADH; EC 1.2.1.24), catalyses the conversion of succinic semialdehyde to succinate in γ-aminobutyrate (GABA) and may be involved in stress protection. ALDH6B2, a methylmalonyl semialdehyde dehydrogenase (MM-ALDH; EC 1.2.1.27), is involved in valine and pyrimidine catabolism and can catalyse CoA-dependent conversion of methylmalonyl semialdehyde to propionyl-CoA and of malonate semialdehyde to acetyl-CoA. ALDH11A3, a non-phosphorylating glyceraldehyde 3-phosphate dehydrogenase (GAPN; EC 1.2.1), catalyses the irreversible NADPH-dependent oxidation of glyceraldehyde 3-phosphate to 3-phosphoglycerate and NADP+, which is the main source for mannitol biosynthesis in many plant species (Valverde et al., 1999; Gao and Loescher, 2000). ALDH12A1, a Δ-1-pyrroline-5-carboxylate dehydrogenase (P5CDH; EC 1.5.1.12), degrades the toxic proline intermediate Δ-1-pyrroline-5-carboxylate to glutamate. ALDH22A1, a plant-specific aldehyde dehydrogenase with an as yet unknown function is localized in the cytoplasm in Arabidopsis (Stiti et al., 2011) and in plastids in maize (Huang et al., 2008). In Arabidopsis, ALDH22A1 is constitutively expressed at a low level (Kirch et al., 2005), and in Z. mays it is induced by a variety of stressors (Huang et al., 2008). In Arabidopsis, ALDH5F1 and ALDH12A1 are targeted to mitochondria (Busch and Fromm, 1999; Deuschle et al., 2001), which is also predicted for the two Eutrema species. ALDH6B2 is predicted to localize in mitochondria and ALDH11A3 in the cytoplasm.

DISCUSSION

Arabidopsis is an excellent model to understand basic processes of flowering plants, but it is a stress-sensitive species. Therefore, Eutrema species as close relatives of Arabidopsis have been suggested as halophytic model species to understand adaptation mechanisms to high salinity (Bressan et al., 2001; Zhu, 2001; Amtmann et al., 2005). Recently the genome sequences of E. parvulum and E. salsugineum have become available, which facilitates gene discovery (Dassanayake et al., 2011; Wu et al., 2012; Yang et al., 2013). Dating analyses showed that E. salsugineum and E. parvulum diverged from each other about 38-4 million years ago, and both share a common ancestor with the Arabidopsis genus around 43.2 million years ago (Yang et al., 2013), indicating a closer relationship between E. salsugineum and E. parvulum than to Arabidopsis. The same relationship is reflected by the ALDH genes of the three species (Fig. 2). Sixteen and 17 genes encoding members of ten ALDH families were identified from E. parvulum and E. salsugineum, respectively (Tables 1 and 2). Eutrema parvulum has the same number of ALDH genes as Arabidopsis, while one more parologue designated as EsALDH3F2 was identified in E. salsugineum. EsALDH3F2 has probably arisen by gene duplication after E. salsugineum, E. parvulum and Arabidopsis had split off. Additionally, two pseudogenes resembling ALDH7B4 and ALDH10A8 were identified in E. salsugineum. According to Wu et al. (2012), these pseudogenes are located on different chromosomes from their corresponding paralogues. It is possible that they resulted from ectopic recombination of repetitive elements such as transposable elements as the genome of E. salsugineum contains a high amount of repetitive sequences. Families 2, 5 and 10 are clustered together, which suggests that these families have probably derived from a common ancestor. ALDH18 which does not contain the conserved ALDH active site, is the most distantly related family, as already observed in other plant species.

Except for the ALDH family 3 in E. salsugineum, copy numbers are identical in Arabidopsis and the two Eutrema species. This is in contrast to HKT1 (sodium–potassium co-transporter) genes related to the maintenance of ion equilibrium. These gene families are expanded in Eutrema, with three HKT1 genes in E. salsugineum and two HKT1 genes in E. parvulum, while only one copy is present in Arabidopsis (Dassanayake et al., 2011; Ali et al., 2012; Wu et al., 2012). However, another comparative genomic study on copy numbers in families among all the functional categories including transcription factors and SOS-like genes found that Arabidopsis and E. salsugineum contain equal or near-equal numbers and, in most cases, Arabidopsis has more copies than E. salsugineum (Yang et al., 2013). ALDH7 family genes are considered as ancient genomic DNA sequences as they are broadly present in plants and animals and highly conserved throughout evolution. Unlike soybean and woody plants Vitis vinifera and Populus...
trichocarpa which contain multiple members in family 7, there is only one member in each of the two Eutrema species as in many other analysed plants. It is not clear why this family has been lost from both the marine unicellular algal species O. tauri and the freshwater unicellular algal species C. reinhardtii, and the multicellular species V. carteri (Wood and Duff, 2009; Brocker et al., 2013). It suggests that this family of proteins may play a fundamental role in land organisms but it is not necessary for aquatic algae.

The localization of ALDH in different cellular compartments has been demonstrated mainly in Arabidopsis. The differential localization indicates that ALDH members have acquired functional specialization during evolution and that different compartments require specific ALDH enzymes to oxidize aldehydes which are generated from different pathways. Except for EsALDH7B4, the same sub-cellular localizations were predicted for the ALDH of the two Eutrema species as for the corresponding Arabidopsis orthologues. This result was expected due to the high degree of conservation of the ALDH proteins in the three species.

High-throughput transcript analysis revealed that a number of universally stress-responsive genes are highly expressed in Eutrema even in the absence of challenges (Inan et al., 2004); therefore, expression of stress-associated ALDH genes was compared under different salinity conditions in A. thaliana and E. salsugineum. No such expression pattern was observed from the selected ALDH genes. Most ALDH genes have a similar expression pattern in Arabidopsis and E. salsugineum, indicating that most ALDH genes are probably not functionally connected with the salinity tolerance of E. salsugineum. However, the two salt-responsive transcripts ALDH3H1 and ALDH7B4 increased in response to higher NaCl levels in E. salsugineum than in Arabidopsis, indicating that higher NaCl levels are required in E. salsugineum than in Arabidopsis to trigger defence reactions. A differential expression profile was only observed for ALDH10A8, which is downregulated in A. thaliana under 600 mM NaCl stress but upregulated in E. salsugineum. This differential regulation is probably due to differences in the promoter structures in the two species, but the functional promoter elements need to be identified experimentally. A comparison of the ALDH7 promoter between A. thaliana and E. salsugineum also revealed a longer putative promoter region in E. salsugineum than in Arabidopsis, with several conserved blocks of functional promoter motifs in both species (Missihoun et al., 2014). Alterations in the sequences and cis-element structures of promoters for other orthologous genes have also been observed in these two species (Wu et al., 2012). Studies carried out so far indicate that genes involved in response to salinity are almost identical in Arabidopsis and in Eutrema halophytes (Zhu, 2000). This implies that the halophytic characteristics of E. salsugineum may be due to specific regulatory mechanisms rather than to increased copy numbers of genes. The analyses of the ALDH genes in Arabidopsis and Eutrema are in agreement with Zhu (2001) who hypothesized that halophytes generally use salt tolerance effectors and regulatory pathways similar to glycophytes and that subtle differences in regulation account for variations in tolerance or sensitivity. Nevertheless there are some potential orphan genes in the genome of the Eutrema halophytes which also may contribute to tolerance mechanisms.

Conclusions

The ALDH superfamily represents a group of NAD(P)⁺-dependent enzymes that catalyse the oxidation of endogenous and exogenous aromatic and aliphatic aldehydes to the corresponding carboxylic acids. Although ALDH gene families have been reviewed in many plants, no attention was paid to halophytes. In the present study, we identified 16 and 17 ALDH genes from the genomes of the halophyte models E. parvulum and E. salsugineum, respectively. Analyses demonstrate that genomic organization, copy number, sub-cellular localization and expression profiles of ALDH genes are fairly conserved in Arabidopsis, E. parvulum and E. salsugineum. Except for the expression of ALDH7B4, ALDH3H1 and ALDH10A8, no differences were observed which may contribute to salinity tolerance. Transcripts of ALDH3H1 and ALDH7B4 increased in response to NaCl at higher levels in E. salsugineum than in Arabidopsis. In addition, ALDH10A8 showed a different expression pattern under high salt in these two species. This indicates that regulation of transcription may be better adapted to high salt in E. salsugineum than in Arabidopsis.

SUPPLEMENTARY DATA

Supplementary data are available online at www.aob.oxfordjournals.org and consist of the following. Table S1: primer sequences used for reverse transcription–PCR of ALDH genes analysed from Arabidopsis and E. salsugineum. Figure S1: PCR products of the pseudogene ALDH7B4 on an agarose gel, and nucleotide sequence alignment. Figure S2: exons—introns structures of predicted EsALDH11A3 ORFs in different genome datasets. Figure S3: nucleotide sequence alignment of putative ALDH10A8 gene promoter regions from Arabidopsis and E. salsugineum. Figure S4: predicted transcription factor binding sites in the 1000 bp promoter sequences upstream and 200 bp downstream of the transcription start sites of AaALDH10A8 and EsALDH10A8. Figure S5: ALDH protein sequences from Arabidopsis, E. parvulum and E. salsugineum.

ACKNOWLEDGEMENTS

Q.H. acknowledges a fellowship from the China Scholarship Council (CSC). The work was supported by the European COST project FA0901 Putting Halophytes to work: from Genes to Ecosystems. We thank Magdalena Grucza for some initial analysis during a COST STSM fellowship, and Tagnon D. Missihoun for advice and discussions. Requests for materials should be addressed to the corresponding author.

LITERATURE CITED


Hou & Bartels — ALDH gene families in Arabidopsis and Eutrema halophytes


