



Published in final edited form as:

Planta. 2013 January ; 237(1): 189–210. doi:10.1007/s00425-012-1749-0.

Aldehyde dehydrogenase (ALDH) superfamily in plants: gene nomenclature and comparative genomics

Chad Bocker,

Department of Pharmaceutical Sciences, Molecular Toxicology and Environmental Health Sciences, Program, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA

Melpomene Vasiliou,

Department of Pharmaceutical Sciences, Molecular Toxicology and Environmental Health Sciences, Program, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA

Sarah Carpenter,

Department of Pharmaceutical Sciences, Molecular Toxicology and Environmental Health Sciences, Program, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA

Christopher Carpenter,

Department of Pharmaceutical Sciences, Molecular Toxicology and Environmental Health Sciences, Program, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA

Yucheng Zhang,

Key Laboratory of Horticultural Plant Biology and Germplasm Innovation in Northwest China, College of Horticulture, Ministry of Agriculture, Northwest A&F University, Yangling, Shanxi 712100, People's Republic of China

Xiping Wang,

Key Laboratory of Horticultural Plant Biology and Germplasm, Innovation in Northwest China, College of Horticulture, Ministry of Agriculture, Northwest A&F University, Yangling, Shanxi 712100, People's Republic of China

Simeon O. Kotchoni,

Department of Biology, Center for Computational and Integrative Biology, Rutgers University, Camden, NJ 08102, USA

Andrew J. Wood,

Department of Plant Biology, Southern Illinois University, Carbondale, Carbondale, IL 62901, USA

Hans-Hubert Kirch,

Institute of Molecular Physiology and Biotechnology of Plants, (IMBIO), University of Bonn, 53115 Bonn, Germany

David Kopečný,

Faculty of Science, Department of Protein Biochemistry and Proteomics, Centre of the Region Haná for Biotechnological and Agricultural Research, Palacký University, Šlechtitelů 11, 783 71 Olomouc, Czech Republic

Daniel W. Nebert, and

© Springer-Verlag 2012

Correspondence to: Vasilis Vasiliou, vasilis.vasiliou@ucdenver.edu.

The authors have no conflicts of interest related to the work detailed within this manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Department of Environmental Health, University of Cincinnati, Medical Center, Cincinnati, OH 45267, USA

Vasilis Vasiliou

Department of Pharmaceutical Sciences, Molecular Toxicology and Environmental Health Sciences, Program, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA

Vasilis Vasiliou: vasilis.vasiliou@ucdenver.edu

Abstract

In recent years, there has been a significant increase in the number of completely sequenced plant genomes. The comparison of fully sequenced genomes allows for identification of new gene family members, as well as comprehensive analysis of gene family evolution. The aldehyde dehydrogenase (*ALDH*) gene superfamily comprises a group of enzymes involved in the NAD⁺- or NADP⁺-dependent conversion of various aldehydes to their corresponding carboxylic acids. *ALDH* enzymes are involved in processing many aldehydes that serve as biogenic intermediates in a wide range of metabolic pathways. In addition, many of these enzymes function as ‘aldehyde scavengers’ by removing reactive aldehydes generated during the oxidative degradation of lipid membranes, also known as lipid peroxidation. Plants and animals share many *ALDH* families, and many genes are highly conserved between these two evolutionarily distinct groups. Conversely, both plants and animals also contain unique *ALDH* genes and families. Herein we carried out genome-wide identification of *ALDH* genes in a number of plant species—including *Arabidopsis thaliana* (thale cress), *Chlamydomonas reinhardtii* (unicellular algae), *Oryza sativa* (rice), *Physcomitrella patens* (moss), *Vitis vinifera* (grapevine) and *Zea mays* (maize). These data were then combined with previous analysis of *Populus trichocarpa* (poplar tree), *Selaginella moellendorffii* (gemmiferous spikemoss), *Sorghum bicolor* (sorghum) and *Volvox carteri* (colonial algae) for a comprehensive evolutionary comparison of the plant *ALDH* superfamily. As a result, newly identified genes can be more easily analyzed and gene names can be assigned according to current nomenclature guidelines; our goal is to clarify previously confusing and conflicting names and classifications that might confound results and prevent accurate comparisons between studies.

Keywords

ALDH; Aldehyde dehydrogenase; Stress response; Gene family; Nomenclature

Introduction

The aldehyde dehydrogenase (*ALDH*) superfamily is composed of a wide variety of enzymes involved in endogenous and exogenous aldehyde metabolism. *ALDH* enzymes use either NAD⁺ or NADP⁺ as a cofactor to convert aldehydes to their corresponding carboxylic acids plus NADH or NADPH. Compounds with aldehydic functional groups are generated as important intermediates in many catabolic and biosynthetic pathways. Moreover, the *ALDH*-mediated generation of NADH/NADPH represents a major source of reducing equivalents required for maintaining cellular redox balance.

The cytoprotective action of these enzymes during oxidative stress represents another major function of the *ALDH* superfamily. Under conditions inducing oxidative stress, *ALDH* enzymes act as ‘aldehyde scavengers’ by metabolizing reactive aldehydes produced as a consequence of the oxidative degradation of lipid membranes, also known as lipid peroxidation (LPO). Many LPO-derived aldehydes such as 4-hydroxynonenal (4-HNE), 4-oxononenal (4-ONE) and malondialdehyde (MDA), to name a few, are potent electrophiles and readily form adducts with physiologically vital nucleophiles such as nucleic acids and proteins. Adduction of these molecules can cause mutations, hinder function, and lead to

significant perturbations in homeostasis. In plants, many studies have shown that ALDH up-regulation is a common target of stress response pathway activation. As such, there is significant economic interest in utilizing ALDH expression to improve both growth and quality in crops grown under harsh conditions such as poor soil quality or drought (Bartels and Sunkar 2005).

Aldehyde dehydrogenases are found throughout prokaryotic and eukaryotic organisms, and the ALDH superfamily is well represented within virtually all plant species. In addition, *ALDH* expression is variable and widespread throughout plant tissues and also developmentally regulated (Missihoun et al. 2011; Tsuji et al. 2003). Furthermore, plant ALDH proteins are found in numerous subcellular compartments—including cytosol, mitochondria, plastids (chloroplasts, chromoplasts and leucoplasts), peroxisomes and microsomes (Missihoun et al. 2011; Mitsuya et al. 2009). Many ALDH families are highly conserved between plants and animals. As predicted, numerous studies have shown that these enzymes share a number of aldehyde substrates.

ALDHs play a crucial role in many catabolic and bio-synthetic pathways including carnitine biosynthesis, glycolysis/gluconeogenesis, and amino-acid metabolism (Marchitti et al. 2008; Sophos and Vasilou 2003; Tylichova et al. 2010; Yang et al. 2011). ALDH expression is also responsible for the aroma associated with fragrant rice strains such as jasmine and basmati (Sakthivel et al. 2009). In plants, ALDH enzymes also play a role in seed development and maturation (Shin et al. 2009). The widespread expression and multifaceted functions of these enzymes in plants underscore their importance.

The plant ALDH superfamily contains 13 distinct families: ALDH2, ALDH3, ALDH5, ALDH6, ALDH7, ALDH10, ALDH11, ALDH12, ALDH18, ALDH21, ALDH22, ALDH23 and ALDH24. The ALDH10, ALDH12, ALDH21, ALDH22, ALDH23 and ALDH24 families are specific to plants, whereas the ALDH2, ALDH3, ALDH5, ALDH6, ALDH7 and ALDH18 families have mammalian orthologues. The total number of *ALDH* genes within a given plant species varies greatly and appears to increase as plants became more complex and moved from water onto land (Table 1).

In many cases, gene expansion commonly associated with increasing organismal complexity is associated with duplication and expansion of a specific lineage or subfamily of proteins. Gene duplication is generally accepted as the predominant evolutionary force behind generation of new genes and subsequent gene functions. In most plant species, gene families account for more than half of the genes within the genome (Flowers and Purugganan 2008). A gene duplication event is generally followed by either partitioning of gene function between the duplicated genes, known as subfunctionalization, or development of a new function, known as neofunctionalization (Force et al. 1999). The two algae analyzed within this study, *V. carteri* and *C. reinhardtii*, contain seven and nine *ALDH* genes, respectively. Mosses and vascular plants have on average twice as many ALDH family members.

The moss *P. patens* separated from flowering plants and unicellular algae more than 400 million years ago (MYA) (Rensing et al. 2008). As plants moved onto land there was a concomitant loss of many genes associated with aquatic life and expansion of genes required for adaptation to terrestrial stressors. In addition, acquisition of many genes is coupled to multicellular growth, vascularization, tissue distribution and differentiation, and the ability to coordinate organismal response via hormone-signaling pathways. Expansion of *ALDH* genes within these species may be due to additional protection offered against environmental stresses encountered during the transition to terrestrial life, as well as increasing structural and developmental complexity (Cronk 2001).

Many plant *ALDH* genes respond to stress, and changes in expression occur following exposure to a wide variety of stressors including dehydration, water logging, heavy metals, high salinity, heat, cold, oxidative stress, ultraviolet radiation (UVR) and many others (Chugh et al. 2011; Inostroza-Blancheteau et al. 2011; Sunkar et al. 2003). The fact that many *ALDH* genes in plants are stress-responsive underscores their importance in supporting environmental adaptability.

Increases in *ALDH* gene number may impart a selective advantage and enhance survival. The increased number of *ALDH* genes found in higher plants is usually the result of expansion of one or more *ALDH* families. For example, the moss *P. patens* and the lycophyte *S. moellendorffii* show an increase in *ALDH 10* homologues. In contrast, *V. vinifera* and *P. trichocarpa* have undergone expansion in the number of *ALDH6* family members.

The species in this manuscript were chosen because they allow for a relatively comprehensive look at *ALDH* superfamily expansion throughout plant evolution. There is a tremendous amount of diversity in plant genome size, structural organization, and chromosome number (Table 2). *Chlamydomonas reinhardtii* is a unicellular alga that belongs to the chlorophyte group, which primarily comprises aquatic photosynthetic eukaryotes. *Volvox carteri* is a multicellular green alga closely related to *C. reinhardtii* that also belongs to the chlorophytes. *Physcomitrella patens* is a non-vascular moss that is generally considered phylogenetically half-way between algae and seed plants—making genomic analyses very important when drawing evolutionary comparisons between species (Rensing et al. 2008). *Selaginella moellendorffii*, also known as the gemmiferous spikemoss, is considered one of the most basic vascular plants and belongs to the lycophytes. Lycophytes first appeared in the fossil record over 400 MYA, which makes present-day members of this family a great resource for comparative genomics. *Arabidopsis thaliana* is a small flowering plant that is commonly used as a model organism in plant biology. *Populus trichocarpa* represents the first tree to have its genome completely sequenced. The analysis of many species encompassing the full gamut of plant evolution from single-celled algae to flowering trees facilitates a comprehensive and informative comparison of *ALDH* evolution and superfamily expansion.

The purpose of this manuscript is to expand upon and also consolidate nomenclature currently found within the literature from a number of species-specific analyses in an effort to clarify *ALDH* nomenclature, as well as identify new *ALDH* genes within these species. In addition, recently sequenced genomes from additional species will be analyzed for *ALDH* superfamily members and the identified *ALDH* genes named, according to approved nomenclature guidelines.

Nomenclature

The nomenclature system for *ALDH* superfamily members was established in 1999 (Vasiliou et al. 1999). The system was developed according to the Human Gene Nomenclature guidelines (<http://www.genenames.org/guidelines.html>) and its use is recommended for all other species. All genes are given the same root symbol 'ALDH' which is followed by an Arabic numeral denoting the family. The family designation is followed by a letter representing the given subfamily and another number signifying the individual gene within that subfamily. The guidelines are based on pairwise alignments between amino-acid sequences for any given set of *ALDH*s. Sequences that share ~40 % identity or more fall into the same *ALDH* family. Sequences that share ~60 % identity or more are grouped into the same subfamily. Gene subfamilies and numbers are designated chronologically following their identification. Genes are italicized and capitalized, whereas

proteins are only capitalized. Additional information relating to the *ALDH* gene superfamily can be found online at the Aldehyde Dehydrogenase Superfamily Database (<http://www.aldh.org>).

ALDH identification and nomenclature has been reported previously by a number of groups for various plant species. The *Arabidopsis thaliana* and *Zea mays* ALDH superfamilies were described in 2004 and 2010, respectively, according to the nomenclature system detailed above (Jimenez-Lopez et al. 2010; Kirch et al. 2004). The *Oryza sativa* ALDH superfamily was originally described in 2009 and revised according to nomenclature guidelines outlined above in 2010 (Gao and Han 2009; Kotchoni et al. 2010). Original descriptions of the ALDH families have appeared for the moss *P. patens* and the algae *C. reinhardtii* and *Ostreococcus tauri* (Wood and Duff 2009), and in *V. vinifera* (Zhang et al. 2012).

In this report, we compile nomenclature information from the above-mentioned publications and use the information to classify a number of newly identified genes within these species. In addition, this information was combined with homology-based searches to identify and name *ALDH* genes within four as-yet-not-studied species—*S. moellendorffii*, *S. bicolor*, *P. trichocarpa* and *V. carteri*. The prefixes ‘arth’ (*A. thaliana*), ‘chre’ (*C. reinhardtii*), ‘hosa’ (*Homo sapiens*), ‘orsa’ (*O. sativa*), ‘phpa’ (*P. patens*), ‘potr’ (*P. trichocarpa*), ‘semo’ (*Selaginella moellendorffii*), ‘sobi’ (*Sorghum bicolor*), ‘vivi’ (*Vitis vinifera*), ‘voca’ (*V. carteri*) and ‘zema’ (*Z. mays*) are used for clarification when referring to ALDH proteins or genes found within different species. *ALDH* genes and associated sequence information used for analyses are listed in Tables 3, 4, 5, 6, 7, 8, 9, 10, 11, 12.

ALDH2 family

The *ALDH2* gene family comprises mitochondrial and cytosolic enzymes that exhibit relatively broad substrate specificity (EC 1.2.1.3) (Marchitti et al. 2008). Studies have shown that ALDH2 family proteins are physiologically active as homo-tetramers (Marchitti et al. 2008). The first plant ALDH2 family members were initially identified as genes that restored fertility in plants; a mutation was identified that causes cytoplasmic male sterility (CMS) and subsequently grouped with other fertility-restorer (RF) genes (Skibbe et al. 2002). However, RF genes are classified by function, not homology, and are therefore not necessarily evolutionarily related. The *Z. mays* *ALDH* genes originally identified as *RF* genes were named *RF2A*, *RF2B*, *RF2C* and *RF2D*; however, the genes were renamed *ALDH2B1*, *ALDH2B6*, *ALDH2C2* and *ALDH2C3*, respectively, by Skibbe et al. in 2002 (Cui et al. 1996; Skibbe et al. 2002). *RF2A*, *RF2B*, *RF2C* and *RF2D* were also independently renamed *ALDH2B2*, *ALDH2B5*, *ALDH2C1* and *ALDH2C2*, respectively, by Jimenez-Lopez et al. in 2010 after a comprehensive analysis of the entire *Z. mays* ALDH gene family by comparing structural 3-D conformations and specific epitope domain/cavities/tunnels found among many well-characterized ALDH crystal structures. Jimenez-Lopez et al. also used naming criteria corresponding to guidelines proposed by the ALDH Gene Nomenclature Committee (AGNC) (Jimenez-Lopez et al. 2010). This nomenclature scheme has therefore taken precedence over previous systems. Similarly, a number of *ALDH* genes identified in *Arabidopsis* have undergone a variety of nomenclature changes. Li et al. identified three *Arabidopsis* *ALDH* genes which were named *ALDH1a*, *ALDH2a* and *ALDH2b* (Li et al. 2000). The same genes were originally named *AtALDH2*, *AtALDH1* and *AtALDH3*, respectively, and then renamed according to approved nomenclature as *ALDH2C4*, *ALDH2B4* and *ALDH2B7* (Skibbe et al. 2002).

ALDH2 family members in plants metabolize acetaldehyde generated as a consequence of ethanolic fermentation (op den Camp and Kuhlmeier 1997; Wei et al. 2009). Metabolism of acetaldehyde produces acetate, which is subsequently used for CoA synthesis via acetyl-

CoA synthase activity; this pathway is known as the ‘pyruvate dehydrogenase (PDH) pathway’. In *Arabidopsis*, mitochondrial ALDH2B4 was shown to be the predominate ALDH isoform contributing to this reaction (Wei et al. 2009). Another study found that transgenic expression of the ALDH2B4 homologue identified in Chinese wild grapevine prevents mildew infection, although the exact means by which expression protects against pathogen infection is not yet understood (Wen et al. 2012). Studies characterizing a recently identified plant ALDH, ALDH2C4, suggested this enzyme plays a role in biosynthesis of ferulic acid and sinapic acid, important compounds contributing to cell wall strength (Grabber et al. 2000; Nair et al. 2004).

The plant ALDH2 family is relatively diverse and includes four distinct subfamilies, namely ALDH2B, C, D and E (Fig. 1). The ALDH2 family expanded significantly during evolution of terrestrial plants and the number of ALDH2 family members varies substantially between species. Two aquatic algae species that we have analyzed, *V. carteri* and *C. reinhardtii*, each contain a single ALDH2E family member; these genes compared between the two species encode proteins that share 69.2 % sequence identity and 79.0 % similarity. The algae ALDH2E proteins share, on average, a higher degree of sequence homology to ALDH2B proteins than ALDH2C or ALDH2D. The lycophyte *S. moellendorffii* genome contains four *ALDH2B* and two *ALDH2D* family members, whereas the moss *P. patens* contains only two *ALDH2B* genes. Lycophytes are considered a basal vascular plant; many novel physiological systems and pathways developed during evolution of mosses to lycophytes. For example, gibberellin-signaling pathways, which control aspects of growth and development, do not exist in *P. patens* but are present in *S. moellendorffii* (Aya et al. 2011). It is possible that expansion of the ALDH2 family occurred in concert with, or as a result of, novel pathway development when basal land plants evolved into basal vascular plants.

ALDH3 family

Studies in mammals have shown that ALDH3 homologues are found within both the cytosol and mitochondria (EC 1.2.1.5) (Marchitti et al. 2008). Bioinformatic analysis of plant proteins predicts diverse subcellular localization— including cytosol, chloroplasts and endoplasmic reticulum (Jimenez-Lopez et al. 2010; Kirch et al. 2004; Stiti et al. 2011). Expression of many of the genes within this family is believed to be regulated by the abscisic acid (ABA) stress-response pathway. The plant ALDH3 family has diverged significantly into six subfamilies, namely ALDH3E, 3F, 3H, 3I, 3J and 3K.

The three *A. thaliana* ALDH isoforms—ALDH3F1, ALDH3H1 and ALDH3I—show distinct expression and response patterns. For example, ALDH3I1 expression is restricted to leaves and responds to treatment with ABA exposure, salinity, dehydration, heavy metals, oxidants and pesticides (Kirch et al. 2001; Stiti et al. 2011; Sunkar et al. 2003). ALDH3H1 was also found to be stress-responsive and constitutively expressed at low levels throughout tissues. A complex expression pattern of *ALDH3H1* gene locus, regulated by gene-splicing or alternative promoters, may be responsible for the abundance of ALDH3H1 (Missihoun et al. 2012). In contrast, ALDH3F1 expression does not change in response to any of the treatments listed above (Kirch et al. 2004). Based on very distinct expression and response patterns noted in *A. thaliana*, it was suggested that ALDH3 isoforms have evolved as a consequence of functional specialization within specific tissues and subcellular organelles (Kirch et al. 2004).

ALDH3 proteins make up one of the most expanded and diverse groups of plant *ALDH* gene families (Fig. 2). *Z. mays*, *O. sativa* and *P. patens* genomes each contain five *ALDH3* homologues. *Sorghum bicolor* and *V. vinifera* each contain four *ALDH3* members, whereas *S. moellendorffii*, *A. thaliana* and *P. trichocarpa* each contain two, three and six *ALDH3*

genes, respectively. It had previously been reported that the unicellular algae *C. reinhardtii* lacked ALDH3; our analyses also were not successful in identifying an ALDH3 homologue within the *C. reinhardtii* genome. Furthermore, the colonial algae *V. carteri* lacks ALDH3—suggesting that expansion and diversification of the ALDH3 gene family occurred in conjunction with the evolutionary movement of aquatic plants onto land.

ALDH5 family

The ALDH5 gene family is made up of succinic semialdehyde dehydrogenases (SSADHs; EC 1.2.1.24), which catalyze the conversion of succinic semialdehyde (SSA) to succinate during the last step of γ -aminobutyrate (GABA) catabolism. ALDH5 participates in the GABA ‘shunt’ pathway found in bacteria, plants and animals—which allows these organisms to metabolically bypass the tricarboxylic acid pathway. GABA in mammals plays a very important role as a neurotransmitter. In plants, GABA is associated with pollen–pistil interactions, herbivore deterrence, oxidative stress and hypoxia (Fait et al. 2008; Palanivelu et al. 2003). ALDH5 in *Arabidopsis* was found to be localized to the mitochondria and is also predicted to be a mitochondrial protein in other plant species (Bouche et al. 2003; Gao and Han 2009). *ALDH5* mutations in plants have been shown to cause enhanced accumulation of reactive oxygen intermediates and cell death in response to light and heat stress (Bouche et al. 2003). Recently, it was also shown that SSA or its derivatives affect adaxial-abaxial polarity and thus leaf patterning (Toyokura et al. 2011).

ALDH5 orthologues were identified in all species with the exception of *V. carteri* (Fig. 3a). We identified an unknown *ALDH* gene within the *V. carteri* genome that was phylogenetically most similar to members of the *ALDH5* gene family; pairwise comparisons revealed less than 35 % sequence identity with other family members. Moreover, the encoded protein is predicted to contain a truncated ALDH domain and is therefore considered to be a pseudogene. The *ALDH5* gene identified in the other algae species analyzed in this study, *C. reinhardtii*, is predicted to encode a protein exhibiting significant sequence divergence from the terrestrial plant species and was placed into a new subfamily, namely ALDH5G. All other newly identified ALDH5 genes from *P. trichocarpa*, *S. moellendorffii* and *S. bicolor* showed a high degree of sequence identity to previously identified genes that currently make up the ALDH5F subfamily and were named accordingly.

ALDH6 family

Members of the *ALDH6* gene family are also known as methylmalonyl semialdehyde (MMS) dehydrogenases (EC 1.2.1.27). These enzymes facilitate reactions associated with both valine and pyrimidine catabolism. To date, ALDH6 homologues are the only ALDH superfamily members to use coenzyme A (CoA) as a cofactor; they catalyze the CoA-dependent conversion of MMS to propionyl-CoA (Marchitti et al. 2008). Studies have also shown that ALDH6 enzymes are capable of metabolizing malonate semialdehyde to acetyl-CoA and revealed up-regulation of *ALDH6* in rice by treatment with the plant hormones auxin and gibberellin (Marchitti et al. 2008; Oguchi et al. 2004). Unfortunately, the exact functions of the ALDH6 family in plants are yet to be thoroughly elucidated.

All plant species analyzed herein contain at least a single *ALDH6* gene (Fig. 3b). Furthermore all plant ALDH6 orthologues identified share >60 % sequence identity and therefore fall into the same subfamily, namely ALDH6B. In *P. trichocarpa* and *V. vinifera*, *ALDH6* has undergone a series of gene duplications resulting in four and three copies, respectively. Recent studies indicated that *ALDH6B3* and *ALDH6B5* are the result of a tandem duplication event in *V. vinifera* (Zhang et al. 2012). The third *V. vinifera* ALDH6 isoform identified, ALDH6B7, is predicted to encode a protein with 1031 amino acids—

making it approximately twice the size of other ALDH6 family members. It does, however, share a very high degree of sequence identity. To date, there is no EST or cDNA sequence data for ALDH6B7 which verify whether the predicted sequence accurately reflects the full-length peptide so it is possible that there might be errors in gene prediction.

Microarray studies, however, have found ALDH6B7 to be up-regulated, along with ALDH6B3, after long-term salinity and dehydration treatments, which provides support for physiological function (Zhang et al. 2012). *ALDH6B3* and *ALDH6B4* from *P. trichocarpa* are predicted to encode proteins that share >94 % sequence identity—supporting a relatively recent gene duplication event. The remaining *P. trichocarpa* isoforms share ~70 % sequence identity with one other, as well as the other plant homologues, and their physiological function has yet to be determined.

ALDH7 family

Members of ALDH7 family (EC 1.2.1.31) are also known as Δ^1 -piperidine-6-carboxylate dehydrogenases (P6CDH), α -amino adipic semialdehyde dehydrogenases or antiquitins. The *ALDH7* gene family is highly conserved throughout evolution. A high degree of conservation observed between evolutionarily distant species implies that physiological function may also be conserved. Studies examining plant ALDH7 expression have identified responsiveness to a wide variety of insults, and expression is thought to function as a part of general stress-response pathways. ALDH7B up-regulation occurs in response to many stressors including ultraviolet radiation, dehydration, increased salinity, low temperature, heat shock and ABA treatment (Kotchoni et al. 2006; Rodrigues et al. 2006).

A recent study also found that ALDH7B in *O. sativa* is required for seed viability and maturation (Shin et al. 2009). Mutant seeds accumulate malondialdehyde and yellow pigment named oryzamutic acid A, a product of amino adipic semialdehyde polymerization (Shen et al. 2012; Shin et al. 2009). Unfortunately, studies have yet to look at the comprehensive substrate specificity of plant ALDH7 homologues. A study in rice found that purified ALDH7B was capable of metabolizing MDA, acetaldehyde and glyceraldehyde (Shin et al. 2009). Analysis of ALDH7B from *Z. mays* and *P. sativum* shows that both enzymes exhibit highest rates with amino adipic semialdehyde followed by guanidinobutyraldehyde. Human ALDH7A1 has wider substrate specificity than both maize and pea enzymes (Kopečný et al. unpublished). In mammals, ALDH7A1 was shown to have relatively broad substrate specificity and was found to play a major role in metabolizing amino adipic semialdehyde, betaine aldehyde, and many lipid peroxidation-derived aldehydes. Human ALDH7A1 shares slightly less than 60 % amino-acid identity with plant ALDH7B orthologues. The high degree of identity suggests that orthologous proteins might metabolize the same physiological substrates; however, further studies will be needed to confirm this supposition.

Similar to other *ALDH* gene families, ALDH7 proteins are highly conserved between individual plant species (Fig. 3c). Surprisingly, both algae analyzed in this study lack an *ALDH7* gene. It is uncertain whether the gene has been lost through evolutionary pressure or gene deletion within these species. *Vitis vinifera* and *P. trichocarpa* contain two *ALDH7* genes. *Vitis vinifera* *ALDH7B4* and *ALDH7D1* are located on chromosomes 11 and 9, respectively; the two genes share ~57 % sequence identity and 69 % similarity. *ALDH7D1* shares between 50 and 57 % identity with other plant ALDH7B proteins. There is currently no information relating to the function or up-regulation of *ALDH7D1* in plants.

ALDH10 family

Members of the ALDH10 family are also known as amino-aldehyde dehydrogenases (AMADHs; EC 1.2.1.19) but also 4-aminobutyraldehyde dehydrogenases, 4-guanidinobutyraldehyde dehydrogenases (EC 1.2.1.54) and also as betaine aldehyde dehydrogenases (BADHs; EC 1.2.1.8). This family has been studied in connection to polyamine catabolism i.e. deactivation of reactive and cytotoxic ω -aminoaldehydes, such as 3-aminopropionaldehyde (APAL) or 4-aminobutyraldehyde (ABAL), which appear after oxidation of various polyamines by plant amine oxidases (Sebela et al. 2000). Their oxidation by AMADHs results in a formation of the nontoxic metabolites β -alanine and γ -aminobutyric acid (GABA). In recent years, studying the physiological aspects of plant AMADHs has become attractive for economic reasons as it has been shown that an *AMADH* gene mutation leads to the acetylation of free ABAL (or its cyclic form Δ^1 -pyrroline) and accumulation of 2-acetyl- Δ^1 -pyrroline, a potent flavor component conferring a fragrance to several rice varieties like Jasmine and Basmati or to soybean (Arikiti et al. 2011; Bradbury et al. 2008).

Members of ALDH10 family have also been extensively studied for their role in stress responses and in the production of the osmoprotectant glycine betaine (GB). GB is a major cellular osmolyte and also acts as a molecular chaperone by helping to stabilize protein structure and function (Allakhverdiev et al. 2008). During drought conditions, GB synthesis and accumulation increase to counter the negative consequences of osmotic imbalance. With this respect, higher plants can be divided into GB-accumulating plants and non-accumulating plants (Fitzgerald et al. 2009). Interestingly, both GB-accumulating and non-accumulating plants contain *ALDH10* genes within their genomes, suggesting that the function of these enzymes may be multifaceted. Interestingly, many ALDH10 homologues show preference for aminoaldehyde substrates over betaine aldehyde (BAL). Numerous studies have shown that some ALDH10 isoforms act primarily as AMADHs by metabolizing ABAL, APAL and GBAL (Bradbury et al. 2008; Missihoun et al. 2011; Trossat et al. 1997; Tylichova et al. 2010; Wei et al. 2009). Also data on tomato and maize AMADHs show that all enzymes preferentially oxidize ω -aminoaldehydes rather than BAL (Kopečný et al., unpublished). Nevertheless, there are several ALDH10 members like those from mangrove (*Avicennia marina*), amaranth (*Amaranthus hypochondriacus*) or spinach (*Spinacia oleracea*) showing high activity with BAL and they are referred to BADHs (Hibino et al. 2001; Incharoensakdi et al. 2000; Valenzuela-Soto and Munoz-Clares 1993). From this point of view, members of the ALDH10 family can be categorized into two groups: first one with low BADH activity and second one with high BADH activity. Mammalian genomes do not contain ALDH10 orthologues. In mammals, ALDH7 and ALDH9 are thought to be the primary GB- and γ -butyrobetaine-synthesizing enzymes (Brocker et al. 2010; Vaz et al. 2000).

ALDH10 genes identified within most plant species share a relatively high degree of sequence homology with one another and fall into the *ALDH10A* subfamily (Fig. 4a). The primitive plants, namely algae and mosses, analyzed herein each have a single *ALDH10* gene. The *ALDH10* gene identified in the unicellular algae *C. reinhardtii* is predicted to encode a protein that shares ~72 % sequence identity with the gene product identified in the colonial algae *V. carteri* and ~60 % identity with the other plant sequences, including those from higher plants. Because it shares on average >60 % identity with other plant *ALDH10* family members, the *C. reinhardtii* gene falls into the *ALDH10A* subfamily.

Interestingly, the gene product from *V. carteri* shares significantly less homology with the other plant ALDH10 sequences. This gene therefore represents a new subfamily and was subsequently named *ALDH10C1*. Vascular plants each have two *ALDH10A* genes,

suggesting a duplication event sometime after these groups diverged. Evidence supporting a gene duplication event is noted in *P. trichocarp*. The two *ALDH10A* genes identified in the *P. trichocarp* genome, *ALDH10A8* and *ALDH10A9*, reside on different chromosomes but share common gene architecture within the surrounding chromosomal regions suggestive of a segmental duplication event (data not shown). There is also evidence that some plant species might carry more than two *ALDH10* genes. Namely maize genome reveals presence of three *ALDH10* genes. Two of them, *ALDH10A8* and *ALDH10A9* (also called *AMADH1a* and *AMADH1b*) are paralogs and clearly duplicated.

ALDH11 family

The *ALDH11* gene family represents a group of cytosolic non-phosphorylating glyceraldehyde 3-phosphate dehydrogenases (GAPNs; EC 1.2.1.9). These enzymes catalyze the irreversible NADP⁺-dependent oxidation of GAP to 3-phosphoglycerate and NADPH (Valverde et al. 1999). This reaction is required for the glycolytic ‘bypass’ pathway unique to photosynthetic eukaryotes—including both plants and microalgae—which circumvents the first substrate level phosphorylation step of glycolysis (Plaxton 1996; Valverde et al. 1999). This reaction represents the main source of NADPH utilized for mannitol biosynthesis in many plant species (Gao and Loescher 2000).

Amino-acid sequence comparisons indicate a high degree of sequence homology within the *ALDH11* gene family (Fig. 4b). All members are categorized into the ALDH11A subfamily with the exception of *V. vinifera* ALDH11B1 previously described by Zhang and coworkers. Interestingly, moss genomes have undergone an expansion in the number of *ALDH11* genes. *Physcomitrella patens* and *S. moellendorffii* have five and six *ALDH11* genes, respectively. It remains unclear why these species possess such a high number of these genes or why this would provide a selective advantage.

ALDH12 family

ALDH12 genes encode Δ -1-pyrroline-5-carboxylate dehydrogenases (P5CDH; EC 1.5.1.12). These enzymes play a key role in the degradation of proline (and also arginine) to glutamate. *ALDH12* gene is widely distributed in bacteria, fungi, plants and animals. A thoroughly studied Arabidopsis P5CDH is localized in the mitochondrial matrix and its expression is induced by externally supplied proline but not by the osmotic stress (Deuschle et al. 2001, 2004; Kirch et al. 2005). The *p5cdh* mutants are hypersensitive to proline, arginine and ornithine and it was hypothesized that accumulation of P5C is the causative agent of cell death induced by proline supply. Each plant species analyzed contains a single *ALDH12* gene. The predicted protein sequences exhibited >60 % identity and therefore belong to single subfamily, namely *ALDH12A* (Fig. 5a). Such a high degree of conservation observed between evolutionarily distant species suggests that there must exist strong selective pressure to maintain gene function. For example, the single-celled algae *V. carteri* and the angiosperm *A. thaliana* express putative proteins that share 61 % sequence identity and 74 % sequence similarity.

ALDH18 family

ALDH18 family contains Δ -1-pyrroline-5-carboxylate synthetases (P5CSs; EC 1.2.1.41 and EC 2.7.2.11). *ALDH18* genes represent the most phylogenetically distantly related ALDH family. These enzymes are bi-functional proteins that contain an N-terminal amino-acid kinase domain and a C-terminal aldehyde dehydrogenase domain. In mammals, ALDH18A1 facilitates ATP- and NADPH-dependent conversion of glutamate to pyrroline-5-carboxylate (P5C), which is subsequently converted to ornithine and used for de novo biosynthesis of proline and arginine (Marchitti et al. 2008). In many terrestrial plants, ALDH18 is

significantly up-regulated in response to dehydration (Yoshida et al. 1997). ALDH18 up-regulation is believed to increase proline synthesis and accumulation. Similar to betaine, proline represents one of the most common, compatible osmolytes in plants; proline accumulation also counters osmotic imbalances encountered during water stress.

The *ALDH18* gene is evolutionarily conserved and represented in all green plant species analyzed within this study (Fig. 5b). The *ALDH18* genes identified in *C. reinhardtii* and *V. carteri* encode proteins that share ~80 % amino acid identity with each other but only 40 % sequence identity with other family members indicating a degree of divergence between genes found in algae and terrestrial plants. Furthermore, a gene duplication event appears to have occurred in vascular plant genomes that is not found in the algae or moss species analyzed. A previous study identified a third *ALDH18* gene within the *Z. mays* genome (Jimenez-Lopez et al. 2010). However, reading frame analysis indicates that the gene would translate a truncated 545-amino acid protein that lacks much of the aldehyde dehydrogenase domain and throws any physiological activity into question. Therefore, the gene identified as *ALDH18B3* by Jimenez-Lopez et al. may represent a pseudogene, perhaps generated through partial gene duplication.

ALDH21, 22, 23 and 24 families

The ALDH21, 22, 23 and 24 families represent relatively new additions to the *ALDH* gene superfamily. To date, there is little information related to the physiological actions of these enzymes or to expression patterns and levels in plants.

A single *ALDH21* gene was found within the *S. moellendorffii* and *P. patens* genomes. Another study identified an *ALDH21* homologue in the moss *Tortula ruralis*, suggesting that expression might be restricted to primitive terrestrial plants (Chen et al. 2002). These genes are predicted to encode proteins most closely related to the ALDH11 family members; thus perhaps ALDH11 and ALDH21 might have related origins.

The *ALDH22* gene family was recently identified, and little is known about specific substrate specificities (Fig. 5c). A recent study looking at the gene in *Z. mays* found ALDH22A1 up-regulation in response to a variety of stressors—including dehydration, high salinity and ABA treatment (Huang et al. 2008). They also found that transgenic over-expression of ALDH22A1 elevated stress tolerance. The group also identified the protein is localized to the plastid, whereas the ALDH22A1 protein in *Arabidopsis* is localized in cytosol (Kirch et al. 2005).

The ALDH23 and ALDH24 families are closely related and grouped phylogenetically. The chreALDH24A1 protein sequence shares ~30 % sequence identity with each of the three ALDH23 family members. Further research is needed in order to determine the function of these new families and the roles they play in plant homeostasis.

Discussion and conclusion

Our understanding of how ALDH enzymes participate in plant homeostasis has greatly expanded in recent years. However, there are still areas that are relatively unexplored and warrant further investigation. The observation that many ALDH enzymes appear to play a primary role in abiotic stress-response pathways suggests that these enzymes represent an important target for increasing plant resistance to stressful conditions, such as elevated soil salinity or dehydration; which becomes especially important when developing stress-tolerant crops.

Materials and methods

Amino-acid sequences for plant species were obtained from the Phytozome and NCBI websites. Phytozome.org is maintained by the Department of Energy's Joint Genome Institute and the Center for Integrative Genomics (Goodstein et al. 2012).

Phytozome (v8.0) and NCBI utilized the following sequences for gene prediction analyses. *Volvox carteri* sequences were obtained from the Volvox v2.0 gene set, 8× genome assembly and annotation (Prochnik et al. 2010). *Selaginella moellendorffii* sequences originated from the v1.0 Dec 20, 2007 Filtered Models 3 annotation. Phytozome v8.0 includes the v1.6 gene annotation of *P. patens* (Lang et al. 2005; Rensing et al. 2005). Sequences include the annotation release 10 of the *A. thaliana* genome release 9 from the 'The Arabidopsis Information Resource (TAIR)' (Lamesch et al. 2012). Phytozome v8.0 features the v4.3 release of annotations for the *C. reinhardtii* genome and includes the Augustus update 10.2 (u10.2) annotation of JGI assembly v4.

The MSU Release 7.0 of the genome annotation of the Nipponbare/*japonica* subspecies of *O. sativa* was used for gene prediction (Ouyang et al. 2007). The 5b.60 annotation (unfiltered working set) of the maize "B73" genome v2 produced by the Maize Genome Project was used for *Z. mays* gene prediction. *Sorghum bicolor* gene prediction was based on the v1.0 release that includes the Sbi1 assembly and Sbi1.4 gene set (Paterson et al. 2009). Finally, *V. vinifera* predictions on Phytozome.org were made using the 12× March 2010 release of the draft genome and annotation by the French-Italian Public Consortium for Grapevine Genome Characterization (Jaillon et al. 2007).

Multiple alignments were created using the ClustalW2 program (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). Phylogenetic trees were constructed using TreeIllustrator software (v0.52). Pairwise alignments of amino acid sequences were performed using EMBOSS global alignment software using the Needleman-Wunsch algorithm (<http://www.ebi.ac.uk/Tools/psa/>).

Acknowledgments

We thank our colleagues for critically reviewing this manuscript. This work was supported, in part, by the following National Institutes of Health grants: R01 EY17963 (VV), R21 AA017754 (VV), F31 AA018248 (CB), and P30 ES006096 (DWN). Further support was provided by grant P501/11/1591 from the Czech Science Foundation (DK).

References

- Allakhverdiev SI, Kreslavski VD, Klimov VV, Los DA, Carpentier R, Mohanty P. Heat stress: an overview of molecular responses in photosynthesis. *Photosynth Res.* 2008; 98:541–550. [PubMed: 18649006]
- Arikat S, Yoshihashi T, Wanchana S, Uyen TT, Huong NT, Wongpornchai S, Vanavichit A. Deficiency in the amino aldehyde dehydrogenase encoded by GmAMADH2, the homo-logue of rice Os2AP, enhances 2-acetyl-1-pyrroline biosynthesis in soybeans (*Glycine max* L.). *Plant Biotechnol J.* 2011; 9:75–87. [PubMed: 20497370]
- Aya K, Hiwatashi Y, Kojima M, Sakakibara H, Ueguchi-Tanaka M, Hasebe M, Matsuoka M. The Gibberellin perception system evolved to regulate a pre-existing GAMYB-mediated system during land plant evolution. *Nat Commun.* 2011; 2:544. [PubMed: 22109518]
- Bartels D, Sunkar R. Drought and salt tolerance in plants. *Crit Rev Plant Sci.* 2005; 24:23–58.
- Bouche N, Fait A, Bouchez D, Moller SG, Fromm H. Mitochondrial succinic-semialdehyde dehydrogenase of the gamma-aminobutyrate shunt is required to restrict levels of reactive oxygen intermediates in plants. *Proc Nat Acad Sci USA.* 2003; 100:6843–6848. [PubMed: 12740438]

- Bradbury LM, Gillies SA, Brushett DJ, Waters DL, Henry RJ. Inactivation of an aminoaldehyde dehydrogenase is responsible for fragrance in rice. *Plant Mol Biol*. 2008; 68:439–449. [PubMed: 18704694]
- Brockner C, Lassen N, Estey T, Pappa A, Cantore M, Orlova VV, Chavakis T, Kavanagh KL, Oppermann U, Vasiliou V. Aldehyde dehydrogenase 7A1 (ALDH7A1) is a novel enzyme involved in cellular defense against hyperosmotic stress. *J Biol Chem*. 2010; 285:18452–18463. [PubMed: 20207735]
- Chen X, Zeng Q, Wood AJ. The stress-responsive *Tortula ruralis* gene ALDH21A1 describes a novel eukaryotic aldehyde dehydrogenase protein family. *J Plant Physiol*. 2002; 159:677–684.
- Chugh V, Kaur N, Gupta AK. Role of antioxidant and anaerobic metabolism enzymes in providing tolerance to maize (*Zea mays* L.) seedlings against waterlogging. *Indian J Biochem Biophys*. 2011; 48:346–352. [PubMed: 22165294]
- Cronk QC. Plant evolution and development in a post-genomic context. *Nat Rev Genet*. 2001; 2:607–619. [PubMed: 11483985]
- Cui X, Wise RP, Schnable PS. The rf2 nuclear restorer gene of male-sterile T-cytoplasm maize. *Science*. 1996; 272:1334–1336. [PubMed: 8650543]
- Deuschle K, Funck D, Hellmann H, Daschner K, Binder S, Frommer WB. A nuclear gene encoding mitochondrial Delta-pyrroline-5-carboxylate dehydrogenase and its potential role in protection from proline toxicity. *Plant J Cell Mol Biol*. 2001; 27:345–356.
- Deuschle K, Funck D, Forlani G, Stransky H, Biehl A, Leister D, van der Graaff E, Kunze R, Frommer WB. The role of [Delta]1-pyrroline-5-carboxylate dehydrogenase in proline degradation. *Plant Cell*. 2004; 16:3413–3425. [PubMed: 15548746]
- Fait A, Fromm H, Walter D, Galili G, Fernie AR. Highway or byway: the metabolic role of the GABA shunt in plants. *Trends Plant Sci*. 2008; 13:14–19. [PubMed: 18155636]
- Fitzgerald TL, Waters DL, Henry RJ. Betaine aldehyde dehydrogenase in plants. *Plant Biol*. 2009; 11:119–130. [PubMed: 19228319]
- Flowers JM, Purugganan MD. The evolution of plant genomes: scaling up from a population perspective. *Curr Opin Genet Dev*. 2008; 18:565–570. [PubMed: 19131240]
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*. 1999; 151:1531–1545. [PubMed: 10101175]
- Gao C, Han B. Evolutionary and expression study of the aldehyde dehydrogenase (ALDH) gene superfamily in rice (*Oryza sativa*). *Gene*. 2009; 431:86–94. [PubMed: 19071198]
- Gao Z, Loescher WH. NADPH supply and mannitol biosynthesis. Characterization, cloning, and regulation of the non-reversible glyceraldehyde-3-phosphate dehydrogenase in celery leaves. *Plant Physiol*. 2000; 124:321–330. [PubMed: 10982446]
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012; 40:D1178–D1186. [PubMed: 22110026]
- Grabber JH, Ralph J, Hatfield RD. Cross-linking of maize walls by ferulate dimerization and incorporation into lignin. *J Agric Food Chem*. 2000; 48:6106–6113. [PubMed: 11312783]
- Hibino T, Meng YL, Kawamitsu Y, Uehara N, Matsuda N, Tanaka Y, Ishikawa H, Baba S, Takabe T, Wada K, Ishii T. Molecular cloning and functional characterization of two kinds of betaine-aldehyde dehydrogenase in betaine-accumulating mangrove *Avicennia marina* (Forsk.) Vierh. *Plant Mol Biol*. 2001; 45:353–363. [PubMed: 11292080]
- Huang W, Ma X, Wang Q, Gao Y, Xue Y, Niu X, Yu G, Liu Y. Significant improvement of stress tolerance in tobacco plants by overexpressing a stress-responsive aldehyde dehydrogenase gene from maize (*Zea mays*). *Plant Mol Biol*. 2008; 68:451–463. [PubMed: 18688729]
- Incharoensakdi A, Matsuda N, Hibino T, Meng YL, Ishikawa H, Hara A, Funaguma T, Takabe T. Overproduction of spinach betaine aldehyde dehydrogenase in *Escherichia coli*. Structural and functional properties of wild-type, mutants and E. coli enzymes. *Eur J Biochem/FEBS*. 2000; 267:7015–7023.
- Inostroza-Blancheteau C, Reyes-Diaz M, Aquea F, Nunes-Nesi A, Alberdi M, Arce-Johnson P. Biochemical and molecular changes in response to aluminium-stress in highbush blueberry

(*Vaccinium corymbosum* L.). Plant Physiol Biochem: PPB/ Societe francaise de physiologie vegetale. 2011; 49:1005–1012.

- Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Huguency P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyere C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pe ME, Valle G, Morgante M, Caboche M, Adam-Blondon AF, Weissenbach J, Quetier F, Wincker P, French-Italian Public Consortium for Grapevine Genome C. The grapevine genome sequence suggests ancestral hexaploidization in major angio-sperm phyla. Nature. 2007; 449:463–467. [PubMed: 17721507]
- Jimenez-Lopez JC, Gachomo EW, Seufferheld MJ, Kotchoni SO. The maize ALDH protein superfamily: linking structural features to functional specificities. BMC Struct Biol. 2010; 10:43. [PubMed: 21190582]
- Kirch HH, Nair A, Bartels D. Novel ABA- and dehydration-inducible aldehyde dehydrogenase genes isolated from the resurrection plant *Craterostigma plantagineum* and *Arabidopsis thaliana*. Plant J Cell Mol Biol. 2001; 28:555–567.
- Kirch HH, Bartels D, Wei Y, Schnable PS, Wood AJ. The ALDH gene superfamily of Arabidopsis. Trends Plant Sci. 2004; 9:371–377. [PubMed: 15358267]
- Kirch HH, Schlingensiepen S, Kotchoni S, Sunkar R, Bartels D. Detailed expression analysis of selected genes of the aldehyde dehydrogenase (ALDH) gene superfamily in *Arabidopsis thaliana*. Plant Mol Biol. 2005; 57:315–332. [PubMed: 15830124]
- Kotchoni SO, Kuhns C, Ditzer A, Kirch HH, Bartels D. 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. 2006; 29:1033–1048. [PubMed: 17080931]
- Kotchoni SO, Jimenez-Lopez JC, Gao D, Edwards V, Gachomo EW, Margam VM, Seufferheld MJ. Modeling-dependent protein characterization of the rice aldehyde dehydrogenase (ALDH) superfamily reveals distinct functional and structural features. PLoS ONE. 2010; 5:e11516. [PubMed: 20634950]
- Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, Huala E. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 2012; 40:D1202–D1210. [PubMed: 22140109]
- Lang D, Eisinger J, Reski R, Rensing SA. Representation and high-quality annotation of the *Physcomitrella patens* transcriptome demonstrates a high proportion of proteins involved in metabolism in mosses. Plant Biol. 2005; 7:238–250. [PubMed: 15912443]
- Li Y, Nakazono M, Tsutsumi N, Hirai A. Molecular and cellular characterizations of a cDNA clone encoding a novel isozyme of aldehyde dehydrogenase from rice. Gene. 2000; 249:67–74. [PubMed: 10831839]
- Marchitti SA, Brockner C, Stagos D, Vasilou V. Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. Expert Opin Drug Metab Toxicol. 2008; 4:697–720. [PubMed: 18611112]
- Missihoun TD, Schmitz J, Klug R, Kirch HH, Bartels D. Betaine aldehyde dehydrogenase genes from Arabidopsis with different sub-cellular localization affect stress responses. Planta. 2011; 233:369–382. [PubMed: 21053011]
- Missihoun TD, Kirch HH, Bartels D. T-DNA insertion mutants reveal complex expression patterns of the aldehyde dehydrogenase 3H1 locus in *Arabidopsis thaliana*. J Exp Bot. 2012; 63:3887–3898. [PubMed: 22442412]
- Mitsuya S, Yokota Y, Fujiwara T, Mori N, Takabe T. OsBADH1 is possibly involved in acetaldehyde oxidation in rice plant peroxisomes. FEBS Lett. 2009; 583:3625–3629. [PubMed: 19850038]
- Nair RB, Bastress KL, Ruegger MO, Denault JW, Chapple C. The Arabidopsis thaliana REDUCED EPIDERMAL FLUORES-CENCE1 gene encodes an aldehyde dehydrogenase involved in ferulic acid and sinapic acid biosynthesis. Plant Cell. 2004; 16:544–554. [PubMed: 14729911]

- Oguchi K, Tanaka N, Komatsu S, Akao S. Methylmalonate-semialdehyde dehydrogenase is induced in auxin-stimulated and zinc-stimulated root formation in rice. *Plant Cell Rep.* 2004; 22:848–858. [PubMed: 15045523]
- op den Camp RG, Kuhlemeier C. Aldehyde dehydrogenase in tobacco pollen. *Plant Mol Biol.* 1997; 35:355–365. [PubMed: 9349259]
- Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, Orvis J, Haas B, Wortman J, Buell CR. The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res.* 2007; 35:D883–D887. [PubMed: 17145706]
- Palanivelu R, Brass L, Edlund AF, Preuss D. Pollen tube growth and guidance is regulated by POP2, an Arabidopsis gene that controls GABA levels. *Cell.* 2003; 114:47–59. [PubMed: 12859897]
- Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob ur R, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS. The Sorghum bicolor genome and the diversification of grasses. *Nature.* 2009; 457:551–556. [PubMed: 19189423]
- Plaxton WC. The organization and regulation of plant glycolysis. *Annu Rev Plant Physiol Plant Mol Biol.* 1996; 47:185–214. [PubMed: 15012287]
- Prochnik SE, Umen J, Nedelcu AM, Hallmann A, Miller SM, Nishii I, Ferris P, Kuo A, Mitros T, Fritz-Laylin LK, Hellsten U, Chapman J, Simakov O, Rensing SA, Terry A, Pangilinan J, Kapitonov V, Jurka J, Salamov A, Shapiro H, Schmutz J, Grimwood J, Lindquist E, Lucas S, Grigoriev IV, Schmitt R, Kirk D, Rokhsar DS. Genomic analysis of organismal complexity in the multicellular green alga *Volvox carter*. *Science.* 2010; 329:223–226. [PubMed: 20616280]
- Rensing SA, Fritzowsky D, Lang D, Reski R. Protein encoding genes in an ancient plant: analysis of codon usage, retained genes and splice sites in a moss *Physcomitrella patens*. *BMC Genomics.* 2005; 6:43. [PubMed: 15784153]
- Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujita T, Oishi K, Shin IT, Kuroki Y, Toyoda A, Suzuki Y, Hashimoto S, Yamaguchi K, Sugano S, Kohara Y, Fujiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA, Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu SH, Stueber K, Theodoulou FL, Tu H, Van de Peer Y, Verrier PJ, Waters E, Wood A, Yang L, Cove D, Cumming AC, Hasebe M, Lucas S, Mishler BD, Reski R, Grigoriev IV, Quatrano RS, Boore JL. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science.* 2008; 319:64–69. [PubMed: 18079367]
- Rodrigues SM, Andrade MO, Gomes AP, Damatta FM, Baracat-Pereira MC, Fontes EP. Arabidopsis and tobacco plants ectopically expressing the soybean antiquitin-like ALDH7 gene display enhanced tolerance to drought, salinity, and oxidative stress. *J Exp Bot.* 2006; 57:1909–1918. [PubMed: 16595581]
- Sakthivel K, Sundaram RM, Shobha Rani N, Balachandran SM, Neeraja CN. Genetic and molecular basis of fragrance in rice. *Biotechnol Adv.* 2009; 27:468–473. [PubMed: 19371779]
- Sebela M, Brauner F, Radova A, Jacobsen S, Havlis J, Galuszka P, Pec P. Characterisation of a homogeneous plant amino-aldehyde dehydrogenase. *Biochim Biophys Acta.* 2000; 1480:329–341. [PubMed: 11004571]
- Shen Y, Zhang Y, Yang C, Lan Y, Liu L, Liu S, Chen Z, Ren G, Wan J. Mutation of OsALDH7 causes a yellow-colored endosperm associated with accumulation of oryzamutic acid A in rice. *Planta.* 2012; 235:433–441. [PubMed: 21960163]
- Shin JH, Kim SR, An G. Rice aldehyde dehydrogenase7 is needed for seed maturation and viability. *Plant Physiol.* 2009; 149:905–915. [PubMed: 19052152]
- Skibbe DS, Liu F, Wen TJ, Yandea MD, Cui X, Cao J, Simmons CR, Schnable PS. Characterization of the aldehyde dehydrogenase gene families of *Zea mays* and Arabidopsis. *Plant Mol Biol.* 2002; 48:751–764. [PubMed: 11999848]

- Sophos NA, Vasiliou V. Aldehyde dehydrogenase gene superfamily: the 2002 update. *Chem Biol Interact.* 2003; 143-144:5–22. [PubMed: 12604184]
- Stiti N, Missihoun TD, Kotchoni SO, Kirch HH, Bartels D. Aldehyde dehydrogenases in *Arabidopsis thaliana*: biochemical requirements, metabolic pathways, and functional analysis. *Frontiers Plant Sci.* 2011; 2:65.
- Sunkar R, Bartels D, Kirch HH. Overexpression of a stress-inducible aldehyde dehydrogenase gene from *Arabidopsis thaliana* in transgenic plants improves stress tolerance. *Plant J Cell Molecular Biol.* 2003; 35:452–464.
- Toyokura K, Watanabe K, Oiwa A, Kusano M, Tameshige T, Tatematsu K, Matsumoto N, Tsugeki R, Saito K, Okada K. Succinic semialdehyde dehydrogenase is involved in the robust patterning of *Arabidopsis* leaves along the adaxial-abaxial axis. *Plant Cell Physiol.* 2011; 52:1340–1353. [PubMed: 21690177]
- Trossat C, Rathinasabapathi B, Hanson AD. Transgenically expressed betaine aldehyde dehydrogenase efficiently catalyzes oxidation of dimethylsulfoniopropionaldehyde and [omega]-aminoaldehydes. *Plant Physiol.* 1997; 113:1457–1461. [PubMed: 12223684]
- Tsuji H, Tsutsumi N, Sasaki T, Hirai A, Nakazono M. Organ-specific expressions and chromosomal locations of two mitochondrial aldehyde dehydrogenase genes from rice (*Oryza sativa* L.), ALDH2a and ALDH2b. *Gene.* 2003; 305:195–204. [PubMed: 12609740]
- Tylichova M, Kopečný D, Morera S, Briozzo P, Lenobel R, Snegaroff J, Sebela M. Structural and functional characterization of plant aminoaldehyde dehydrogenase from *Pisum sativum* with a broad specificity for natural and synthetic aminoaldehydes. *J Mol Biol.* 2010; 396:870–882. [PubMed: 20026072]
- Valenzuela-Soto EM, Munoz-Clares RA. Betaine-aldehyde dehydrogenase from leaves of *Amaranthus hypochondriacus* L. exhibits an Iso Ordered Bi Bi steady state mechanism. *J Biol Chem.* 1993; 268:23818–23823. [PubMed: 8226918]
- Valverde F, Losada M, Serrano A. Engineering a central metabolic pathway: glycolysis with no net phosphorylation in an *Escherichia coli* gap mutant complemented with a plant GapN gene. *FEBS Lett.* 1999; 449:153–158. [PubMed: 10338122]
- Vasiliou V, Bairoch A, Tipton KF, Nebert DW. Eukaryotic aldehyde dehydrogenase (ALDH) genes: human polymorphisms, and recommended nomenclature based on divergent evolution and chromosomal mapping. *Pharmacogenetics.* 1999; 9:421–434. [PubMed: 10780262]
- Vaz FM, Fouchier SW, Ofman R, Sommer M, Wanders RJ. Molecular and biochemical characterization of rat gamma-trimethylaminobutyraldehyde dehydrogenase and evidence for the involvement of human aldehyde dehydrogenase 9 in carnitine biosynthesis. *J Biol Chem.* 2000; 275:7390–7394. [PubMed: 10702312]
- Wei Y, Lin M, Oliver DJ, Schnable PS. The roles of aldehyde dehydrogenases (ALDHs) in the PDH bypass of *Arabidopsis*. *BMC Biochem.* 2009; 10:7. [PubMed: 19320993]
- Wen Y, Wang X, Xiao S, Wang Y. Ectopic expression of VpALDH2B4, a novel aldehyde dehydrogenase gene from Chinese wild grapevine (*Vitis pseudoreticulata*), enhances resistance to mildew pathogens and salt stress in *Arabidopsis*. *Planta.* 2012; 236(2):525–539. [PubMed: 22437646]
- Wood AJ, Duff RJ. The aldehyde dehydrogenase (ALDH) gene superfamily of the moss *Physcomitrella patens* and the algae *Chlamydomonas reinhardtii* and *Ostreococcus tauri*. *Bryologist.* 2009; 112:1–11.
- Yang J, An D, Zhang P. Expression profiling of cassava storage roots reveals an active process of glycolysis/gluconeogenesis. *J Integr Plant Biol.* 2011; 53:193–211. [PubMed: 21205184]
- Yoshida Y, Kiyosue T, Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K. Regulation of levels of proline as an osmolyte in plants under water stress. *Plant Cell Physiol.* 1997; 38:1095–1102. [PubMed: 9399433]
- Zhang Y, Mao L, Wang H, Brockner C, Yin X, Vasiliou V, Fei Z, Wang X. Genome-wide identification and analysis of grape aldehyde dehydrogenase (ALDH) gene superfamily. *PLoS ONE.* 2012; 7:e32153. [PubMed: 22355416]

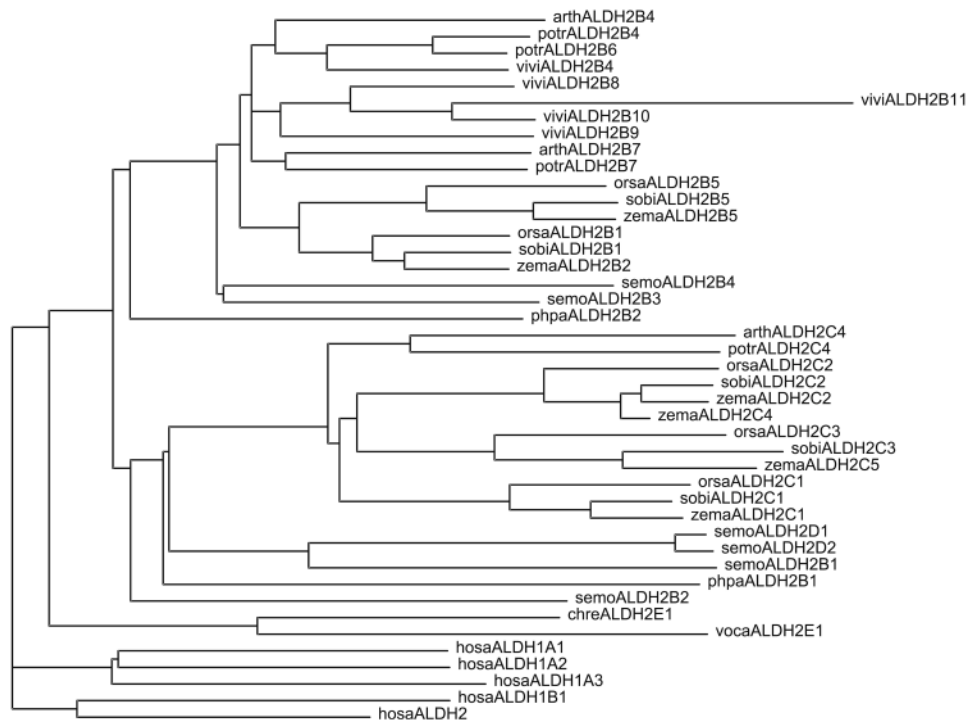


Fig. 1. Phylogenetic analysis of ALDH2 family members. Phylogram created using ALDH2 protein sequences and ClustalW2 multiple sequence alignment software. Tree edited using TreeIllustrator (v0.52) software. The prefixes 'arth' (*Arabidopsis thaliana*), 'chre' (*Chlamydomonas reinhardtii*), 'hosa' (*Homo sapiens*), 'orsa' (*Oryza sativa*), 'phpa' (*Physcomitrella patens*), 'potr' (*Populus trichocarpa*), 'semo' (*Selaginella moellendorffii*), 'sobi' (*Sorghum bicolor*), 'vivi' (*Vitis vinifera*), 'voca' (*Volvox carteri*) and 'zema' (*Zea mays*) are used for clarification when referring to ALDH proteins or genes found within different species



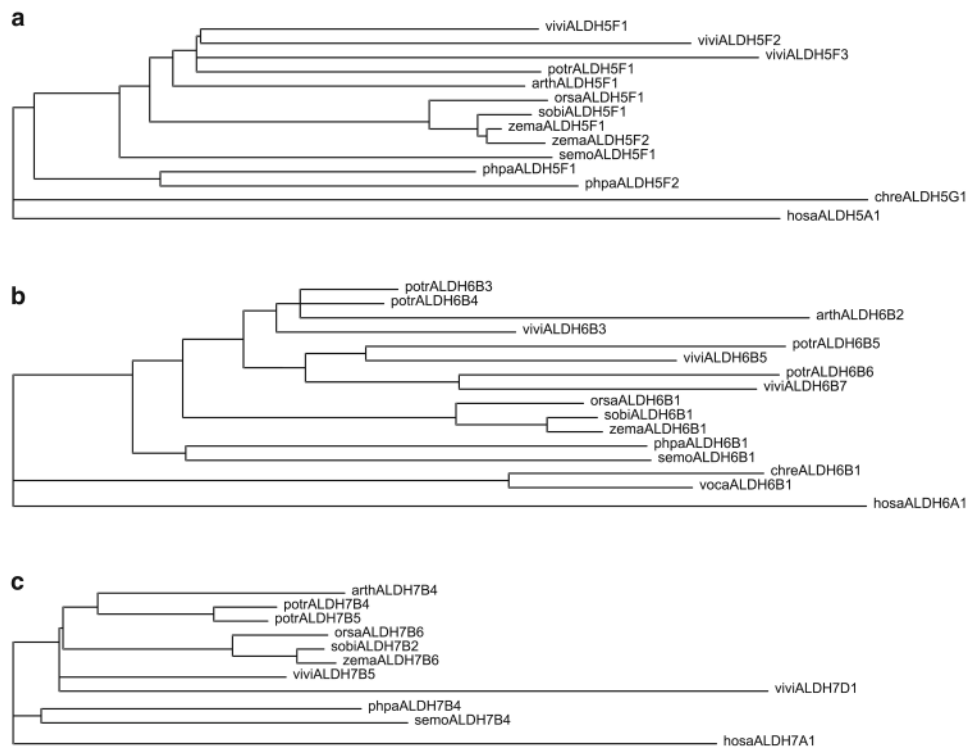


Fig. 3. Phylogenetic analyses of ALDH5 (a), ALDH6 (b) and ALDH7 (c) family members. Creation of the phylogram, use of ClustalW2 multiple sequence alignment software, editing of the tree, and prefix abbreviations of genus and species are identical to that in Fig. 1

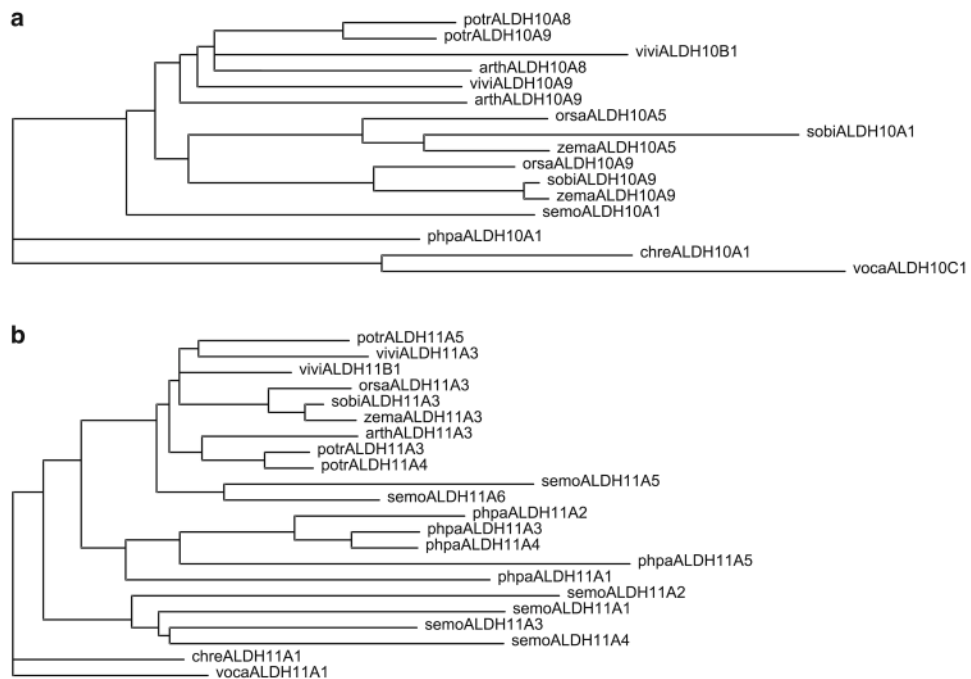
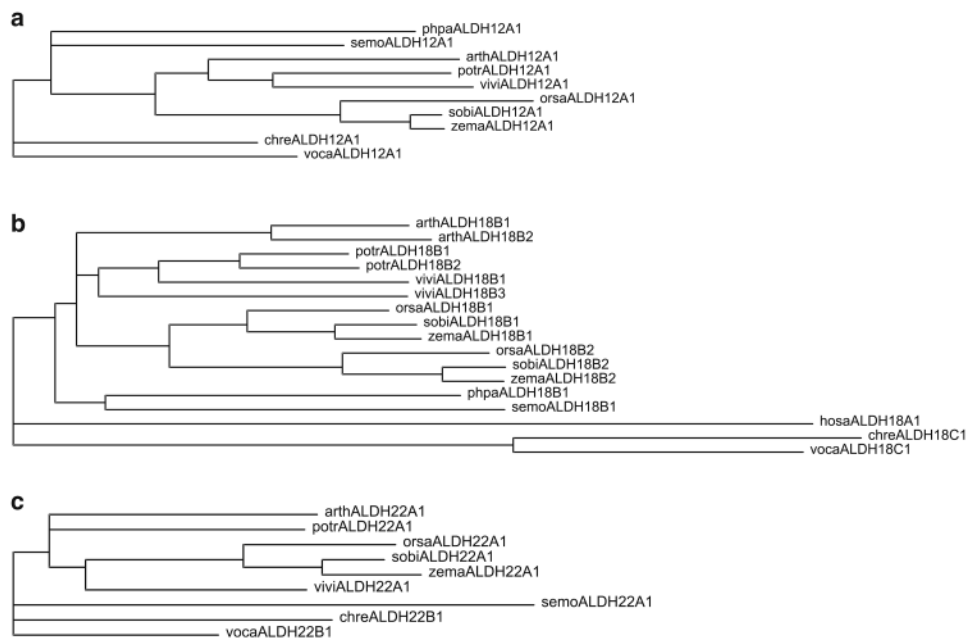


Fig. 4. Phylogenetic analyses of ALDH10 (**a**) and ALDH11 (**b**) family members. Creation of the phylogram, use of ClustalW2 multiple sequence alignment software, editing of the tree, and prefix abbreviations of genus and species are identical to that in Fig. 1

**Fig. 5.**

Phylogenetic analyses of ALDH12 (a), ALDH18 (b) and ALDH22 (c) family members.

Creation of the phylogram, use of ClustalW2 multiple sequence alignment software, editing of the tree, and prefix abbreviations of genus and species are identical to that in Fig. 1

Table 1
Number of ALDH family members identified in analyzed species

Species	ALDH Family																								Total
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24		
<i>A. thaliana</i>	3	3	-	1	1	1	1	-	2	1	1	-	-	-	-	-	2	-	-	-	1	-	-	16	
<i>C. reinhardtii</i>	1	-	-	1	1	1	-	-	1	1	1	-	-	-	-	-	1	-	-	-	1	-	1	9	
<i>O. sativa</i>	5	5	-	1	1	1	1	-	2	1	1	-	-	-	-	-	2	-	-	-	1	-	-	20	
<i>P. patens</i>	2	5	-	2	1	1	1	-	1	5	1	-	-	-	-	-	1	-	-	1	-	1	-	21	
<i>P. trichocarpa</i>	4	6	-	1	4	2	-	-	2	3	1	-	-	-	-	-	2	-	-	-	1	-	-	26	
<i>S. moellendorffii</i>	6	2	-	1	1	1	1	-	1	6	1	-	-	-	-	-	1	-	-	1	1	2	-	24	
<i>S. bicolor</i>	5	4	-	1	1	1	1	-	2	1	1	-	-	-	-	-	2	-	-	-	1	-	-	19	
<i>V. vinifera</i>	5	4	-	3	3	2	-	-	2	2	1	-	-	-	-	-	2	-	-	-	1	-	-	25	
<i>V. carteri</i>	1	-	-	-	1	-	-	-	1	1	1	-	-	-	-	-	1	-	-	-	1	-	-	7	
<i>Z. mays</i>	6	5	-	2	1	1	1	-	2	1	1	-	-	-	-	-	2	-	-	-	1	-	-	22	
<i>H. sapiens</i>	7	4	1	1	1	1	1	1	-	-	-	-	-	-	1	-	1	-	-	-	-	-	-	19	

Table 2
Genome information of analyzed plant species

Species	Common name	Chrm number	Genome size (Mbp)	ALDH number
<i>Arabidopsis thaliana</i>	Thale cress	5	135	16
<i>Chlamydomonas reinhardtii</i>	Unicellular green algae	ND	112	9
<i>Oryza sativa</i>	Asian rice	12	372	20
<i>Physcomitrella patens</i>	Moss	27	480	21
<i>Populus trichocarpa</i>	Black cottonwood	19	403	26
<i>Selaginella moellendorffii</i>	Gemmiferous Spikemoss	27	100	24
<i>Sorghum bicolor</i>	Sorghum	10	660	19
<i>Vitis vinifera</i>	Common grape vine	19	300	25
<i>Volvox carteri</i>	Colonial green algae	ND	138	7
<i>Zea mays</i>	Maize	10	3,000	22
<i>Homo sapiens</i>	Human	23	3,000	19

ND not determined

Table 3
***Arabidopsis thaliana* ALDH superfamily: unified nomenclature and gene information**

Family	Gene name	NCBI gene ID	Other names/aliases	NCBI protein ID	Phytozone ID	Chrm	Exon #	AA #	References
Family 2	<i>ALDH2B4</i>	823955	ALDH2; ALDH2A; ALDH2B4	NP_190383.1	AT3G48000	3	11	538	Sophos and Vasiliou 2003
	<i>ALDH2B7</i>	838991	ALDH2B; F5O8.35; F5O8_35	NP_564204.1	AT1G23800	1	11	534	Sophos and Vasiliou 2003; Kirch et al. 2004
	<i>ALDH2C4</i>	822042	ALDH1A; REF1	NP_566749.1	AT3G24503	3	9	501	Kirch et al. 2004
Family 3	<i>ALDH3FI</i>	829782	F23E13.140; F23E13_140	NP_195348.2	AT4G36250	4	9	484	Sophos and Vasiliou 2003
	<i>ALDH3HI</i>	841020	ALDH4; T7O23.15; T7O23_15	NP_175081.1	AT1G44170	1	10	484	Kirch et al. 2004
	<i>ALDH3II</i>	829573	ALDH3; F10M10.10; F10M10_10	NP_567962.1	AT4G34240	4	11	550	Kirch et al. 2004
Family 5	<i>ALDH5FI</i>	844282	SSADH; SSADH1	NP_178062.1	AT1G79440	1	20	528	Sophos and Vasiliou 2003
Family 6	<i>ALDH6B2</i>	815903	T22C12.10; T22C12_10	NP_179032.1	AT2G14170	2	19	607	Sophos and Vasiliou 2003
Family 7	<i>ALDH7B4</i>	841849	F15I1.19; F15I1_19	NP_175812.1	AT1G54100	1	14	508	Sophos and Vasiliou 2003
Family 10	<i>ALDH10A8</i>	843831	F25A4.11; F25A4_11	NP_001185399.1	AT1G74920	1	15	496	Sophos and Vasiliou 2003
	<i>ALDH10A9</i>	823972	–	NP_190400.1	AT3G48170	3	15	503	Sophos and Vasiliou 2003
Family 11	<i>ALDH11A3</i>	816962	F27D4.18; F27D4_18; NP-GAPDH	NP_001189589.1	AT2G24270	2	9	496	Sophos and Vasiliou 2003; Kirch et al. 2004
Family 12	<i>ALDH12A1</i>	836373	K19B1.14; K19B1_14; P5CDH	NP_568955.1	AT5G62530	5	16	556	Kirch et al. 2004
Family 18	<i>ALDH18B1</i>	818566	ATP5CS; P5CS1; T517.10; T517_10	NP_181510.1	AT2G39800	2	20	717	–
	<i>ALDH18B2</i>	824727	P5CS2	NP_191120.2	AT3G55610	3	20	726	–
Family 22	<i>ALDH22A1</i>	819849	T8E24.4; T8E24_4	NP_974242.1	AT3G66658	3	14	596	Kirch et al. 2004

Exon and amino acid figures obtained from NCBI entries

\$watermark-text

\$watermark-text

\$watermark-text

Chlamydomonas reinhardtii ALDH superfamily: unified nomenclature and gene information

Table 4

Family	Gene name	NCBI gene ID	Other names/ aliases	NCBI protein ID	Phytozome ID	Chrm	Exon #	AA #	References
Family 2	<i>ALDH2E1</i>	159465489	Cr_Aldh2A	XP_001690955.1	Cre12.g500150	12	16	534	Wood and Duff 2009
Family 5	<i>ALDH5G1</i>	159472386	Cr_Aldh5B	XP_001694332.1	Cre08.g381707	8	17 ^a	476 ^a	Wood and Duff 2009
Family 6	<i>ALDH6B1</i>	159475673	Cr_Aldh6A	XP_001695943.1	Cre16.g675650	16	12	553	Wood and Duff 2009
Family 10	<i>ALDH10A1</i>	159482146	Cr_Aldh10A	XP_001699134.1	Cre13.g605650	13	11	504	Wood and Duff 2009
Family 11	<i>ALDH11A1</i>	159472072	Cr_Aldh11A	XP_001694180.1	Cre12.g556600	12	10	498	Wood and Duff 2009
Family 12	<i>ALDH12A1</i>	159477663	Cr_Aldh12A	XP_001696928.1	Cre12.g520350	12	12	548	Wood and Duff 2009
Family 18	<i>ALDH18D1</i>	159475184	–	XP_001695703.1	Cre01.g061800	1	16	664	–
Family 22	<i>ALDH22B1</i>	159463690	–	XP_001690075.1	Cre01.g033350	1	16	629	–
Family 24	<i>ALDH24A1</i>	159481716	Cr_Aldh24A	XP_001698924.1	Cre01.g069750	1	12	490	Wood and Duff 2009

Exon and amino acid figures obtained from NCBI entries unless specified

^aExon and amino acid figures obtained from phytozome entry

\$watermark-text

\$watermark-text

\$watermark-text

Table 5
***Oryza sativa* ALDH superfamily: unified nomenclature and gene information**

Family	Gene name	NCBI gene ID	Other names/aliases	NCBI protein ID	Phytozone ID	Chrm	Exon #	AA #	References
Family 2	<i>ALDH2B1</i>	4340725	Os06g15990; OsALDH2-5; ALDH2b	NP_001057358.1	LOC_Os06g15990	6	11	549	Sophos and Vasilou 2003; Gao and Han 2009
	<i>ALDH2B5</i>	4330612	Os02g49720; OsALDH2-4; ALDH2a	NP_001048010.1	LOC_Os02g49720	2	9	553	Sophos and Vasilou 2003; Gao and Han 2009
	<i>ALDH2C1</i>	4326375	Os01g40860; OsALDH2-1; ALDH1a	NP_001043453.1	LOC_Os01g40860	1	8	502	Sophos and Vasilou 2003; Gao and Han 2009
Family 3	<i>ALDH2C2</i>	4326376	Os01g40870; OsALDH2-2; ALDH1b	NP_001043454.1	LOC_Os01g40870	1	9	507	Gao and Han 2009
	<i>ALDH2C3</i>	No entry	Os06g39230; OsALDH2-3; CU607043	BAD32861.1	LOC_Os06g39230	6	8	423	Gao and Han 2009
	<i>ALDH3E1</i>	4330146	Os02g43194; OsALDH3-1	NP_001047575.1	LOC_Os02g43194	2	10	487	Sophos and Vasilou 2003; Gao and Han 2009
Family 4	<i>ALDH3E2</i>	4336538	Os04g45720; OsALDH3-3	NP_001053442.1	LOC_Os04g45720	4	9	499	Gao and Han 2009
	<i>ALDH3E3</i>	4330149	Os02g43280; OsALDH3-2	NP_001047578.1	LOC_Os02g43280	2	10	491	Gao and Han 2009
	<i>ALDH3H1</i>	No entry	Os12g07810; OsALDH3-5	ABA96616.1	LOC_Os12g07810	12	10	484	Gao and Han 2009
Family 5	<i>ALDH3H2</i>	4349966	Os11g08300; OsALDH3-4	NP_001065921.1	LOC_Os11g08300	11	10	482	Gao and Han 2009
	<i>ALDH5F1</i>	No entry	OsL_06045; Os02g07760; OsALDH5; CU606989	B9F3B6.1	LOC_Os02g07760	20	2	527	Gao and Han 2009
	<i>ALDH6B1</i>	4342610	Os07g09060; OsALDH6	NP_001059082.1	LOC_Os07g09060	7	19	534	Sophos and Vasilou 2003; Gao and Han 2009
Family 7	<i>ALDH7B6</i>	4347172	Os09g26880; OsALDH7	NP_001063281.1	LOC_Os09g26880	9	14	509	Sophos and Vasilou 2003; Gao and Han 2009
Family 10	<i>ALDH10A5</i>	4336081	Os04g39020; OsALDH10-1; BADH	NP_001053016.1	LOC_Os04g39020	4	15	505	Sophos and Vasilou 2003; Gao and Han 2009
Family 11	<i>ALDH10A9</i>	4345606	Os08g32870; OsALDH10-2	NP_001061833.1	LOC_Os08g32870	8	15	503	Gao and Han 2009
	<i>ALDH11A3</i>	4345685	Os08g34210; OsALDH11	NP_001061905.1	LOC_Os08g34210	8	9	499	Gao and Han 2009
Family 12	<i>ALDH12A1</i>	4339448	Os05g45960; OsALDH12	EEE64501.1	No entry	Un	Un	716	Gao and Han 2009
Family 18	<i>ALDH18B1</i>	4338979	Os05g38150; OsALDH18-2; P5CS; BAA19916	NP_001055723.1	LOC_Os05g38150	5	19	716	Sophos and Vasilou 2003; Gao and Han 2009
Family 22	<i>ALDH18B2</i>	4324853	Os01g62900; OsALDH18-1	NP_001044802.1	LOC_Os01g62900	1	20	735	Gao and Han 2009
	<i>ALDH22A1</i>	4344358	Os07g48920; OsALDH22	NP_001060704.1	LOC_Os07g48920	7	14	597	Gao and Han 2009

Exon and amino acid figures obtained from NCBI entries *Un* undetermined

\$watermark-text

\$watermark-text

\$watermark-text

Table 6
***Physcomitrella patens* ALDH superfamily: unified nomenclature and gene information**

Family	Gene name	NCBI gene ID	Other names/ aliases	NCBI protein ID	Phytozome ID	Scaffold	Exon #	AA #	References
Family 2	<i>ALDH2B1</i>	5930606	Pp_Aldh2A	XP_001767457	Pp1s93_81V6	93	12	553	Wood and Duff 2009
	<i>ALDH2B2</i>	5948865	Pp_Aldh2B	XP_001785650	Pp1s496_1V6	496	4	535	Wood and Duff 2009
Family 3	<i>ALDH3H1</i>	5928120	Pp_Aldh3A	XP_001764841	Pp1s71_128V6	71	10	492	Wood and Duff 2009
	<i>ALDH3H2</i>	5943334	Pp_Aldh3E	XP_001780129	Pp1s272_3V6	272	11	583	Wood and Duff 2009
	<i>ALDH3K1</i>	5920386	Pp_Aldh3B	XP_001757163	Pp1s26_253V6	26	11	479	Wood and Duff 2009
	<i>ALDH3K2</i>	5930362	Pp_Aldh3D	XP_001767194	Pp1s90_226V6	90	9	485	Wood and Duff 2009
	<i>ALDH3K3</i>	5933556	Pp_Aldh3C	XP_001770374	Pp1s124_90V6	124	9 ^a	467 ^a	Wood and Duff 2009
Family 5	<i>ALDH5F1</i>	5916037	Pp_Aldh5B	XP_001752876	Pp1s6_180V6	6	19	492	Wood and Duff 2009
	<i>ALDH5F2</i>	5943671	Pp_Aldh5A	XP_001780466	Pp1s279_72V6	279	20	498	Wood and Duff 2009
Family 6	<i>ALDH6B1</i>	5920495	Pp_Aldh6A	XP_001757403	Pp1s27_198V6	27	4	574	Wood and Duff 2009
Family 7	<i>ALDH7B4</i>	5941578	Pp_Aldh7A	XP_001778351	Pp1s237_29V6	237	15	511	Wood and Duff 2009
Family 10	<i>ALDH10A1</i>	5919866	Pp_Aldh10A	XP_001756623	Pp1s23_38V6	23	15	559	Wood and Duff 2009
Family 11	<i>ALDH11A5</i>	5935487	Pp_Aldh11E	XP_001772261	No entry	148	10 ^a	504 ^a	Wood and Duff 2009
	<i>ALDH11A1</i>	5928282	Pp_Aldh11A	XP_001765101	Pp1s73_32V6	73	10	503	Wood and Duff 2009
	<i>ALDH11A2</i>	5918399	Pp_Aldh11B	XP_001755163	Pp1s16_156V6	16	8	496	Wood and Duff 2009
	<i>ALDH11A3</i>	5916920	Pp_Aldh11D	XP_001753784	Pp1s10_228V6	10	9	496	Wood and Duff 2009
	<i>ALDH11A4</i>	5923860	Pp_Aldh11C	XP_001760657	Pp1s44_64V6	44	8	496	Wood and Duff 2009
Family 12	<i>ALDH12A1</i>	5923366	Pp_Aldh12A	XP_001760169	Pp1s41_177V6	41	17	571	Wood and Duff 2009
Family 18	<i>ALDH18B1</i>	5919578	–	XP_001756289	Pp1s22_40V6	22	21	757	–
Family 21	<i>ALDH21A1</i>	5932362	Pp_Aldh21A	XP_001769187	Pp1s11_161V6	111	8	497	Wood and Duff 2009
Family 23	<i>ALDH23A1</i>	5918738	Pp_Aldh23A	XP_001755511	Pp1s18_148V6	18	2	494	Wood and Duff 2009

Exon and amino acid figures obtained from NCBI entries unless specified

^aExon and amino acid figures obtained from Phytozome entry

Watermark-text

Watermark-text

Watermark-text

Table 7
***Populus trichocarpa* ALDH superfamily: unified nomenclature and gene information**

Family	Gene name	NCBI Gene ID	Other names/aliases	NCBI Protein ID	Phytozome ID	Chrm	Exon #	AA #	References
Family 2	<i>ALDH2B4</i>	7483487	Pt-ALDH1.2; POPTRDRAFT_823362	XP_002318034.1	POPTR_0012s08010	12	11	536	–
	<i>ALDH2B6</i>	7453761	Pt-ALDH1.3; POPTRDRAFT_666446	XP_002321583.1	POPTR_0015s08540	15	11	542	–
	<i>ALDH2B7</i>	7481430	Pt-ALDH2.1; POPTRDRAFT_830473	XP_002301540.1	POPTR_0002s19060	2	11	540	–
	<i>ALDH2C4</i>	7458755	Pt-REF1.1; POPTRDRAFT_809215	XP_002323497.1	POPTR_0018s06060	18	9	497	–
Family 3	<i>ALDH3F1</i>	7480157	Pt-ALDH3.2; POPTRDRAFT_832403	XP_002310836.1	POPTR_0007s13650	7	10	480	–
	<i>ALDH3H1</i>	7460961	Pt-ALDH3.3; POPTRDRAFT_831510	XP_002306641.1	POPTR_0005s20150	5	10	488	–
	<i>ALDH3H4</i>	7466457	Pt-ALDH3.1; POPTRDRAFT_754562	XP_002302229.1	POPTR_0002s08230	2	10	488	–
	<i>ALDH3H5</i>	7471422	POPTRDRAFT_928126	XP_002330375.1	POPTR_0001s42150	1	10	458	–
Family 5	<i>ALDH3H6</i>	7479170	POPTRDRAFT_861574	XP_002307060.1	POPTR_0005s07090	5	10	488	–
	<i>ALDH3I1</i>	7480051	POPTRDRAFT_706284	XP_002298405.1	POPTR_0001s26630	1	10	472	–
	<i>ALDH5F1</i>	7481606	Pt-ALDH5.2; POPTRDRAFT_1091519	XP_002315079.1	POPTR_0010s18120	10	20	536	–
	<i>ALDH6B3</i>	7478778	Pt-ALDH6.4; POPTRDRAFT_202375	XP_002313217.1	POPTR_0009s08220	9	17	496	–
Family 6	<i>ALDH6B4</i>	7487827	Pt-ALDH6.5; POPTRDRAFT_846793	XP_002300008.1	POPTR_0001s28990	1	17	497	–
	<i>ALDH6B5</i>	7478777	POPTRDRAFT_1087520	XP_002313216.1	POPTR_0009s08230	9	17	506	–
	<i>ALDH6B7</i>	7476429	–	XP_002307307.1	No entry	5	17	496	–
	<i>ALDH7B4</i>	7477690	Pt-ALDH7.2; POPTRDRAFT_853157	XP_002304124.1	POPTR_0003s06570	3	14	516	–
Family 10	<i>ALDH7B5</i>	7460180	Pt-ALDH7.1; POPTRDRAFT_745820	XP_002331178.1	POPTR_0001s16730	1	14	508	–
	<i>ALDH10A8</i>	7484775	Pt-ALDH10.1; POPTRDRAFT_661953	XP_002318630.1	POPTR_0012s07730	12	15	503	–
	<i>ALDH10A9</i>	7462540	Pt-ALDH10.2; POPTRDRAFT_666405	XP_002322147.1	POPTR_0015s08190	15	15	503	–
	<i>ALDH11A3</i>	7493404	Pt-ALDH11.1; POPTRDRAFT_578767	XP_002325142.1	POPTR_0018s11820	18	9	496	–
Family 11	<i>ALDH11A4</i>	7485231	Pt-ALDH11.2; POPTRDRAFT_717030	XP_002308391.1	POPTR_0006s20090	6	9	496	–
	<i>ALDH11A5</i>	7458414	POPTRDRAFT_912477	XP_002324322.1	POPTR_0018s02400	18	9	498	–
	<i>ALDH12A1</i>	7491541	Pt-FIS1.3; POPTRDRAFT_581353	XP_002330119.1	POPTR_0015s07550	15	16	566	–
	<i>ALDH18B1</i>	7490371	Pt-P5CS.1; POPTRDRAFT_833794	XP_002315202.1	POPTR_0010s20590	10	20	719	–
Family 22	<i>ALDH18B2</i>	No entry	–	No entry	POPTR_0008s06060	8 ^a	19 ^a	691	–
	<i>ALDH22A1</i>	7471107	Pt-ALDH22.1; POPTRDRAFT_820566	XP_002312334.1	POPTR_0008s10550	8	14	594	–

Exon and amino acid figures obtained from NCBI entries unless specified

^aExon and amino acid figures obtained from Phytozome entry

\$watermark-text

\$watermark-text

\$watermark-text

Table 8
***Selaginella moellendorffii* ALDH superfamily: unified nomenclature and gene information**

Family	Gene name	NCBI gene ID	Other names/aliases	NCBI protein ID	Phytozome ID	Scaffold	Exon #	AA #	References
Family 2	<i>ALDH2D1</i>	9657137	–	XP_002977953.1	268124	35	10	491	–
	<i>ALDH2B2</i>	9632783	–	XP_002961274.1	164447	0	11 ^a	500 ^a	–
	<i>ALDH2B4</i>	9633073	–	XP_002961352.1	266598	0	10	491	–
	<i>ALDH2B1</i>	9635245	–	XP_002982106.1	233921	52	10	495	–
	<i>ALDH2B3</i>	9631989	–	XP_002970420.1	231526	14	10	488	–
Family 3	<i>ALDH2D2</i>	9657142	–	XP_002977958.1	233053	35	10	491	–
	<i>ALDH3H1</i>	9652402	–	XP_002966639.1	227647	7	10	479	–
	<i>ALDH3H2</i>	9652527	–	XP_002966702.1	168397	7	10	480	–
	<i>ALDH5F1</i>	9656931	–	XP_002961441.1	230009	1	19	499	–
	<i>ALDH6B1</i>	9636001	–	XP_002965586.1	407088	6	19	557	–
Family 5	<i>ALDH5F1</i>	9656931	–	XP_002961441.1	230009	1	19	499	–
Family 6	<i>ALDH6B1</i>	9636001	–	XP_002965586.1	407088	6	19	557	–
Family 7	<i>ALDH7B4</i>	9642264	–	XP_002961173.1	164276	0	14	511	–
Family 10	<i>ALDH10A1</i>	9658015	–	XP_002974519.1	174224	24	15	503	–
Family 11	<i>ALDH11A1</i>	9643335	–	XP_002962724.1	79174	2	9	540	–
	<i>ALDH11A2</i>	9643415	–	XP_002963205.1	80113	3	9	553	–
	<i>ALDH11A3</i>	9637846	–	XP_002974975.1	232458	26	10 ^a	497 ^a	–
	<i>ALDH11A4</i>	9630484	–	XP_002961562.1	77620	1	10	497	–
	<i>ALDH11A5</i>	9655811	–	XP_002960309.1	75347	0	8	497	–
	<i>ALDH11A6</i>	9655302	–	XP_002963773.1	270375	3	9	497	–
Family 12	<i>ALDH12A1</i>	9650766	–	XP_002968656.1	90262	11	15	526	–
Family 18	<i>ALDH18B1</i>	9646044	–	XP_002970003.1	410685	13	20 ^a	771 ^a	–
Family 21	<i>ALDH21A1</i>	9657643	–	XP_002981136.1	420552	47	8	481	–
Family 22	<i>ALDH22A1</i>	9660395	–	XP_002979353.1	110267	40	14	586	–
Family 23	<i>ALDH23B1</i>	9657190	–	XP_002977683.1	443623	35	16	494	–
	<i>ALDH23B2</i>	9657189	–	XP_002977682.1	417609	35	16	480	–

Exon and amino acid figures obtained from NCBI entries unless specified

^aExon and amino acid figures obtained from Phytozome entry

\$watermark-text

\$watermark-text

\$watermark-text

Table 9
***Sorghum bicolor* ALDH superfamily: unified nomenclature and gene information**

Family	Gene name	NCBI gene ID	Other names/ aliases	NCBI protein ID	Phytozome ID	Chrm	Exon #	AA #	References
Family 2	<i>ALDH2B1</i>	No entry	-	BAB92019.1	Sb10g009790	10	11	556	-
	<i>ALDH2B5</i>	No entry	-	BAB92018.1	Sb04g029040	4	9	551	-
	<i>ALDH2C1</i>	No entry	-	No entry	Sb03g026550	3	7	501	-
	<i>ALDH2C2</i>	8082171	-	XP_0024455869.1	Sb03g026570	3	7	504	-
	<i>ALDH2C3</i>	8065048	-	XP_002437214.1	Sb10g023000	10	6	520	-
Family 3	<i>ALDH3E1</i>	8078880	-	XP_0024452841.1	Sb04g033420	4	9	486	-
	<i>ALDH3E2</i>	8083334	-	XP_002448250.1	Sb06g023975	6	6	319	-
	<i>ALDH3H1</i>	8079901	-	XP_002441918.1	Sb08g004840	8	10	487	-
	<i>ALDH3H2</i>	8079744	-	XP_002449124.1	Sb05g005470	5	10	478	-
	<i>ALDH5F1</i>	No entry	-	No entry	Sb04g004920	4 ^a	20 ^a	527	-
Family 6	<i>ALDH6B1</i>	8075050	-	XP_002459473.1	Sb02g005200	2	19	540	-
Family 7	<i>ALDH7B2</i>	8080995	-	XP_002462451.1	Sb02g025790	2	14	509	Sophos and Vasilou 2003
Family 10	<i>ALDH10A1</i>	8057487	-	AAC49268	Sb06g019200	6	9	494	Sophos and Vasilou 2003
Family 11	<i>ALDH10A9</i>	8075627	-	XP_002444357.1	Sb07g020650	7	15	505	-
	<i>ALDH11A3</i>	8080604	-	XP_002444416.1	Sb07g021630	7	9	498	-
Family 12	<i>ALDH12A1</i>	8068986	-	XP_002441445.1	Sb09g026810	9	15	549	-
Family 18	<i>ALDH18B2</i>	No entry	-	ACU65226.1	Sb03g039820	3	21	729	-
Family 22	<i>ALDH18B1</i>	No entry	-	ACU65227.1	Sb09g022290	9	20	716	-
	<i>ALDH22A1</i>	8077508	-	XP_002463429.1	Sb02g043660	2	14	593	-

Exon and amino acid figures obtained from NCBI entries unless specified

^a Exon and amino acid figures obtained from Phytozome entry

Table 10
***Vitis vinifera* ALDH superfamily: unified nomenclature and gene information**

Family	Gene name	NCBI gene ID	Other names/aliases	NCBI protein ID	Phytozone ID	Chrm	Exon #	AA #	References
Family 2	<i>ALDH2B4</i>	100262043	ALDH2B7_V1, V2, V3	XP_002283132.1	GSVIVT01007784001	17	11	538	Zhang et al. 2012
	<i>ALDH2B8</i>	100267386	VpALDH2_XM_002263443	XP_002263479.1	GSVIVT01020224001	1	11	538	Zhang et al. 2012
	<i>ALDH2B9</i>	100259296	VpALDH2_XM_002274827	XP_002274863.1	GSVIVT01032500001	14	11	535	Zhang et al. 2012
	<i>ALDH2B10</i>	No entry	-	CBI16844.3	GSVIVT01000602001	1	11	388	-
Family 3	<i>ALDH2B11</i>	No entry	-	CBI32134.3	GSVIVT01020227001	1	13	571	-
	<i>ALDH3F1</i>	100247106	VpALDH3F1	XP_002273358.1	GSVIVT01018842001	4	10	485	Zhang et al. 2012
	<i>ALDH3H1</i>	100248915	VpALDH3H1	XP_002285866.1	GSVIVT01008845001	18	10	488	Zhang et al. 2012
	<i>ALDH3H5</i>	100266331	VpALDH3H1	XP_002273730.1	GSVIVT01022356001	7	10	488	Zhang et al. 2012
Family 5	<i>ALDH3J1</i>	100253650	VpALDH3_XM_002285430	XP_002285466.1	GSVIVT01025276001	6	10	485	Zhang et al. 2012
	<i>ALDH5F1</i>	100262240	ALDH5F1_XM_002265478.1	XP_002265514.1	GSVIVT01036719001	19	20	530	Zhang et al. 2012
	<i>ALDH5F2</i>	100251993	ALDH5F1_XM_002265366	XP_002265402.1	GSVIVT01036720001	19	19	493	Zhang et al. 2012
	<i>ALDH5F3</i>	100257122	ALDH5F1_XM_002265318	XP_002265354.1	GSVIVT01036721001	19	21	559	Zhang et al. 2012
Family 6	<i>ALDH6B3</i>	100266351	VpALDH6B2	XP_002266390.1	GSVIVT01000336001	3	19	539	Zhang et al. 2012
	<i>ALDH6B5</i>	100261145	ALDH6B_XM_002266580	XP_002266616.1	GSVIVT01000338001	3	19	571	Zhang et al. 2012
	<i>ALDH6B7</i>	100254417	ALDH6B_XM_002266343	XP_002266379.1	GSVIVT01003951001	Un	19	1031	Zhang et al. 2012
	<i>ALDH7B5</i>	100252622	VpALDH7A1	XP_002278093.1	GSVIVT01015062001	11	14	508	Zhang et al. 2012
Family 7	<i>ALDH7D1</i>	100261802	VpALDH7A1_XM_002272508	XP_002272544.1	GSVIVT01016734001	9	14	530	Zhang et al. 2012
	<i>ALDH10A9</i>	100246770	VpALDH10A9	XP_002283690.1	GSVIVT01007829001	17	15	503	Zhang et al. 2012
	<i>ALDH10B1</i>	100250859	VpALDH10A8	XP_002281984.1	GSVIVT01032588001	14	15	499	Zhang et al. 2012
	<i>ALDH11A3</i>	100249114	VpALDH11_XM_002285250.1	XP_002285286.1	GSVIVT01035891001	4	9	496	Zhang et al. 2012
Family 11	<i>ALDH11B1</i>	100260378	VpALDH11A3	XP_002279374.1	GSVIVT01023590001	11	9	496	Zhang et al. 2012
	<i>ALDH12A1</i>	100251938	VpALDH12A1	XP_002273569.1	GSVIVT01008047001	17	16	555	Zhang et al. 2012
	<i>ALDH18B1</i>	100232881	Vp_ALDH18A_XM_002282319	XP_002282355.1	GSVIVT01016467001	13	20	762	Sophos and Vasilou 2003
	<i>ALDH18B3</i>	100266504	Vp_ALDH18A1_XM_002273220	XP_002273256.1	GSVIVT01034097001	8	20	714	Zhang et al. 2012
Family 22	<i>ALDH22A1</i>	100249882	VpALDH22A	XP_002277743.1	GSVIVT01035003001	5	14	593	Zhang et al. 2012

Exon and amino acid figures obtained from NCBI entries

Table 11
***Volvox carteri* ALDH superfamily: unified nomenclature and gene information**

Family	Gene name	NCBI gene ID	Other names/ aliases	NCBI protein ID	Phytozone ID	Scaffold	Exon #	AA #	References
Family 2	<i>ALDH2E1</i>	9621060	–	XP_002948098.1	73567	6	8	528	–
Family 6	<i>ALDH6B1</i>	9618616	–	XP_002949543.1	59269	18	11	524	–
Family 10	<i>ALDH10C1</i>	9620250	–	XP_002947147.1	73155	4	10	503	–
Family 11	<i>ALDH11A1</i>	9617583	–	XP_002946341.1	56054	1	7	497	–
Family 12	<i>ALDH12A1</i>	9623193	–	XP_002958122.1	69010	37	10	550	–
Family 18	<i>ALDH18D1</i>	9625062	–	XP_002954612.1	65082	14	16	661	–
Family 22	<i>ALDH22B1</i>	9628231	–	XP_002958773.1	84639	1	13	529	–

Exon and amino acid figures obtained from NCBI entries

Table 12
***Zea mays* ALDH superfamily: unified nomenclature and gene information**

Family	Gene name	NCBI gene ID	Other names/aliases	NCBI protein ID	Phytozone ID	Chrm	Exon #	AA #	References
Family 2	<i>ALDH2B2</i>	732806	ALDH2B1 ^b , ZmALDH2B1, rf2a	NP_001105891.1	GRMZM2G058675	9	11	549	Jimenez-Lopez et al. 2010
	<i>ALDH2B5</i>	542567	ALDH2B6 ^b , ZmALDH2B5, rf2b; ac183825.4	NP_001105576.1	GRMZM2G125268	4	9	550	Jimenez-Lopez et al. 2010
	<i>ALDH2C1</i>	541913	ALDH2C2 ^b , ZmALDH2C1a; rf2c	NP_001105046.1	GRMZM2G071021	3	7	503	Jimenez-Lopez et al. 2010
	<i>ALDH2C2</i>	541914	ALDH2C3 ^b , ALDH2C2, rf2d; rf2d1; ALDH5; ZmALDH2C1b	NP_001105047.1	GRMZM2G097706	3	8	511	Jimenez-Lopez et al. 2010
Family 3	<i>ALDH2C4</i>	No entry	rf2e	No entry	GRMZM2G380438G	8 ^a	8 ^a	324 ^c	Jimenez-Lopez et al. 2010
	<i>ALDH2C5</i>	No entry	rf2f	No entry	RMZM2G407949	6 ^a	8 ^a	516	Jimenez-Lopez et al. 2010
	<i>ALDH3E1</i>	100280692	ZmALDH3E1	NP_001147083.1	GRMZM2G169458	5	10	485	Jimenez-Lopez et al. 2010
	<i>ALDH3E2</i>	100284774	ZmALDH3E2	NP_001151141.1	GRMZM2G155502	10	9	489	Jimenez-Lopez et al. 2010
Family 5	<i>ALDH3H1</i>	100285809	ZmALDH3H1	NP_001152171.1	GRMZM2G103546	10	10	491	Jimenez-Lopez et al. 2010
	<i>ALDH3H2</i>	100281700	ZmALDH3H2	NP_001148092.1	GRMZM2G060800	4	10	481	Jimenez-Lopez et al. 2010
	<i>ALDH3H3</i>	100382449	ZmALDH3H3	NP_001168661.1	GRMZM2G118800	2	10	478	Jimenez-Lopez et al. 2010
	<i>ALDH5F1</i>	100280779	ZmALDH5F1	NP_001147173.1	GRMZM2G128114	5	20	493	Jimenez-Lopez et al. 2010
Family 6	<i>ALDH5F2</i>	100284047	ZmALDH5F2	NP_001150417.1	GRMZM2G119482	4	20	527	Jimenez-Lopez et al. 2010
	<i>ALDH6B1</i>	100274311	ZmALDH6B1	NP_001142146.1	GRMZM2G001898	7	13	537	Jimenez-Lopez et al. 2010
	<i>ALDH7B6</i>	100282748	ZmALDH7B6	NP_001149126.1	GRMZM2G130440	2	14	509	Jimenez-Lopez et al. 2010
	<i>ALDH10A5</i>	541949	ZmALDH10A5; AMADH2	ACS74868.1	GRMZM2G135470	10	15	506	Jimenez-Lopez et al. 2010
Family 10	<i>ALDH10A8</i>	100302679	ZmALDH10A8; AMADH1a	ACS74867.1	GRMZM2G013214	4	15	505	Jimenez-Lopez et al. 2010
	<i>ALDH10A9</i>	No entry	ZmALDH10A9; AMADH1b	AEP68091.1	GRMZM2G016189	1	15	506	Jimenez-Lopez et al. 2010
	<i>ALDH11A3</i>	542583	ZmALDH11A3	NP_001105589.1	GRMZM2G035268	4	9	498	Sophos and Vasilou 2003
	<i>ALDH12A1</i>	No entry	ZmALDH12A1	AAL70108.1	GRMZM2G090087	6	15	549	Jimenez-Lopez et al. 2010
Family 18	<i>ALDH18B2</i>	100280719	ZmALDH18B2	NP_001147111.1	GRMZM2G028535	8	21	731	Jimenez-Lopez et al. 2010
	<i>ALDH18B1</i>	No entry	ZmALDH18B1	ACR33941.1	GRMZM2G375504	8	20	717	Jimenez-Lopez et al. 2010
	<i>ALDH22A1</i>	100125658	ZmALDH22A1; AC212124.5	NP_001106059.1	GRMZM2G135341	7	15	593	Jimenez-Lopez et al. 2010
	<i>ALDH22A2</i>	No entry							

Exon and amino acid figures obtained from NCBI entries unless specified

^aExon and amino acid figures obtained from phytozone entry

\$watermark-text

\$watermark-text

\$watermark-text

^bFormally named according to Skibbe et al. 2002

^cThis appears to be incomplete sequence as it can be assembled from two GRMZM numbers