Opinion TRENDS in Plant Science Vol.9 No.8 August 2004

Abiotic stress series:

The ALDH gene superfamily of *Arabidopsis*

Hans-Hubert Kirch[1], Dorothea Bartels[1], Yanling Wei[2], Patrick S. Schnable[2] and Andrew J. Wood[3]

[1]University of Bonn, Institute of Molecular Physiology and Biotechnology of Plants (IMBIO), Kirschallee1, D-53115 Bonn, Germany
[2]Center for Plant Genomics, Roy J. Carver Co-Laboratory, Iowa State University, Ames, IA 50011-3605, USA
[3]Department of Plant Biology, Southern Illinois University – Carbondale, Carbondale, IL 62901-6509, USA

Aldehyde dehydrogenases (ALDHs) represent a protein superfamily of NAD(P)$^+$-dependent enzymes that oxidize a wide range of endogenous and exogenous aliphatic and aromatic aldehydes. The *Arabidopsis* genome contains 14 unique ALDH sequences encoding members of nine ALDH families, including eight known families and one novel family (ALDH22) that is currently known only in plants. Here, we identify members of the ALDH gene superfamily in *Arabidopsis*; provide a revised, unified nomenclature for these ALDH genes; analyze the molecular relationship among *Arabidopsis* ALDH genes and compare them to ALDH genes from other species, including prokaryotes and mammals; and describe the role of ALDHs in cytoplasmic male sterility, plant defense and abiotic stress tolerance.

Many biologically important aldehydes are metabolized by the superfamily of NAD(P)$^+$-dependent aldehyde dehydrogenases [aldehyde:NAD(P)$^+$ oxidoreductases, EC 1.2.1] [1]. Sequence comparisons among ALDH genes from bacteria, animals and plants have identified three diagnostic amino acid motifs: (i) the ALDH glutamic acid active site signature sequence MELGGNA (PROSITE PS00687); (ii) the Rossmann fold GxGxxG coenzyme-binding site; and (iii) the catalytic thiol (PROSITE PS00070) [1–3]. Currently, 555 genes encoding ALDH proteins have been identified throughout all taxa [4]. Aldehydes are intermediates in a range of fundamental biochemical pathways and are generated during the metabolism of carbohydrates, vitamins, steroids, amino acids and lipids [5,6]. Aldehydes can also be generated in response to a suite of environmental stresses that perturb metabolism including salinity, dehydration, desiccation, cold and heat shock [7,8]. Although they are common biochemical intermediates, many aldehydes are chemically reactive and toxic at physiological concentrations [9]. Active ALDH enzymes represent an important mechanism for the detoxification of aldehydes by oxidation to their corresponding carboxylic acids [6–9].

The ALDH Gene Nomenclature Committee (AGNC) has established specific criteria for cataloguing deduced ALDH protein sequences [1]. Protein sequences that are more than 40% identical to other previously identified ALDH sequences compose a family, and sequences more than 60% identical compose a protein subfamily. Protein sequences that are less than 40% identical would describe a new ALDH protein family. Previous classifications of the ALDH gene superfamily in eukaryotes have identified 21 protein families based upon sequence identity [2,3,10–12]. To date, plant enzymes are represented in 11 ALDH families: ALDH2, ALDH3, ALDH5, ALDH6, ALDH7, ALDH10, ALDH11, ALDH12, ALDH18, ALDH19 and ALDH21. Three protein families are unique to plants (ALDH11, ALDH19 and ALDH21) [12] and one of these is apparently unique to mosses (ALDH21) [13].

The *Arabidopsis* genome is the only completely sequenced plant genome, and this allows us for the first time to examine the phylogenetic and molecular relationship of all the ALDH genes in a plant species. Based upon nomenclature developed by the AGNC [1–5], we present here a revised, unified nomenclature for the *Arabidopsis* ALDH genes and describe the most important structural and functional features of the corresponding ALDH protein families. In addition, the unified nomenclature enables a critical view of a functional classification of ALDH enzymatic activities.

Arabidopsis ALDH gene superfamily

Aldehyde dehydrogenase (ALDH) and ALDH-like DNA sequences were retrieved from The *Arabidopsis* Information Resource (TAIR, http://www.arabidopsis.org/), Munich Information center for Protein Sequences (MIPS, http://mips.gsf.de/proj/thal/), the US National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) genome and non-redundant expressed sequence tag (EST) databases using BLASTN, PSI-BLAST and BLASTX [14,15]. The searches were conducted using ALDH3I1 (GenBank Accession number AJ306961), ALDH3H1 (GenBank Accession number AY072122) and the ALDH active site signature sequence (PROSITE PS00687). Annotated *Arabidopsis* ALDH open reading frames (ORFs) were verified by comparison to cDNA and EST sequences. These searches resulted in the identification of 14 *Arabidopsis* sequences that might encode proteins with the diagnostic motifs described above (Table 1). All these ALDH sequences were submitted to

Corresponding author: Andrew J. Wood (wood@plant.siu.edu).
Available online 17 July 2004

 1360-1385/$ - see front matter doi:10.1016/j.tplants.2004.06.004

Table 1. *Arabidopsis thaliana* aldehyde dehydrogenase gene superfamily

Gene Annotation	Previous name	Locus	GenBank Accession number	Protein annotation	Putative function or pathway involved	Refs
ALDH2B4	ALDH2a	At3g48000	AF349447	ALDH2B4	Mitochondrial ALDH	[10]
ALDH2B7	ALDH2b	At1g23800	AF348416	ALDH2B7	Mitochondrial ALDH	[10]
ALDH2C4	ALDH1a	At3g24503	AF349448	ALDH2C4	Cytosolic ALDH, phenyl-propanoid pathway (ferulic acid and sinapic acid biosynthesis)	[10,23]
ALDH3F1	NA	At4g36250	AJ584644	ALDH3F1	Variable substrate ALDH	[a]
ALDH3H1	AthALDH4	At1g44170	AY072122	ALDH3H1	Variable substrate ALDH	[24][a]
ALDH3I1	AthALDH3	At4g34240	AJ306961	ALDH3I1	Variable substrate ALDH (chloroplast), stress-regulated detoxification pathway	[24,25]
ALDH5F1	SSADH1	At1g79440	AF117335	ALDH5F1	SSADH (mitochondria), stress-regulated detoxification pathway	[26,27]
ALDH6B2	NA	At2g14170	NM_126989	ALDH6B2	Mitochondrial MM-ALDH	
ALDH7B4	NA	At1g54100	AJ584645	ALDH7B4	Turgor-responsive	[a]
ALDH10A8	NA	At1g74920	AY093071	ALDH10A8	BADH	
ALDH10A9	NA	At3g48170	AF370333	ALDH10A9	Mitochondrial BADH	
ALDH11A3	NA	At2g24270	AY037205	ALDH11A3	GAPN	
ALDH12A1	AtP5CDH	At5g62530	AY039787	ALDH12A1	Mitochondrial P5CDH, protection from proline toxicity	[34]
ALDH22A1	NA	At3g66658	AJ584646	ALDH22A1	Novel ALDH	[a]

Abbreviations: ALDH, aldehyde dehydrogenase; BADH, betaine aldehyde dehydrogenase; GAPN, non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase; MM-ALDH, methylmalonyl semialdehyde dehydrogenase; NA, not applicable; P5CDH, Δ^1-pyrroline-5-carboxylate dehydrogenase; SSADH, succinic semialdehyde dehydrogenase.
[a]H.H. Kirch *et al.*, unpublished.

the AGNC [1] and the revised nomenclature presented here has been approved.

The ALDH genes contained in the *Arabidopsis* genome encode members of nine ALDH protein families – eight previously identified protein families (ALDH2, ALDH3, ALDH5, ALDH6, ALDH7, ALDH10, ALDH11, ALDH12) and one novel protein family (ALDH22) (Table 1, Figure 1). For the classification of each gene family, the root symbol ALDH is followed by the family designation number (1–22), a subfamily designator if required (A, B etc.) and, finally, the individual gene number. For example, the betaine aldehyde dehydrogenase (BADH) encoding cDNA cloned from *Spinacia oleracea* [16] was renamed using this universal nomenclature as ALDH10A7.

Three of the nine ALDH families in *Arabidopsis* are represented by more than one gene [family 10 (two members) and families 2 and 3 (three members each)], whereas the remaining six families are encoded by single-copy genes. It is therefore interesting that no more than four ALDH genes (*ALDH6B2*, *ALDH22A1* and two

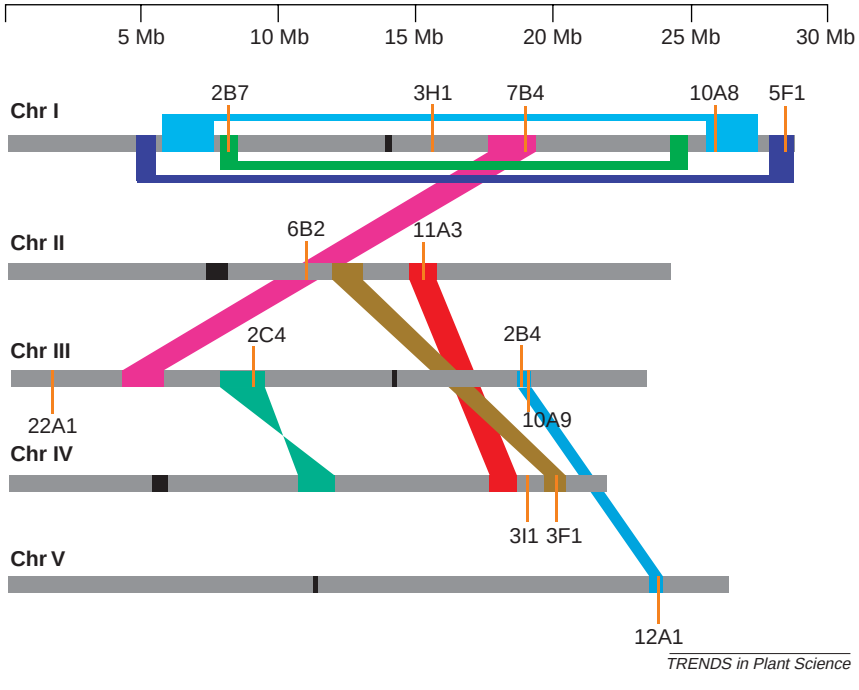

Figure 1. Chromosomal positions of aldehyde dehydrogenase (ALDH) genes in *Arabidopsis*. Chromosomes I–V (Chr I–V) are depicted as horizontal gray bars with centromeres as black bars. ALDH genes are indicated by vertical orange lines and annotated using the proposed unified ALDH nomenclature. Colored bars denote duplicated regions of the *Arabidopsis* genome; the twisted colored bar indicates that the duplicated regions are in reverse orientation. The map was produced using the MATDB Redundancy Viewer (http://mips.gsf.de/proj/thal/db/gv/gv_frame.html). Adapted from Ref. [37].

members of the ALDH3 gene family) are located in non-duplicated regions of the *Arabidopsis* genome (Figure 1). In addition, only three genes (*ALDH2B4*, *ALDH10A9* and *ALDH12A1*) are found within the same duplicated region and thus might represent 'direct' gene duplications, whereas the respective counterparts of the other seven ALDH genes have been lost. Although the three ALDHs show a high protein sequence divergence and even belong to different families, our hypothesis is supported by the fact that the gene contexts around the ALDH genes are highly conserved between the duplicated genomic regions of chromosomes III and V. This implies that functional constraints are responsible for the rapid evolution and sequence divergence of these ALDH genes.

A phylogenetic tree of the *Arabidopsis* ALDH sequences and other putative plant ALDHs available in the TAIR, MIPS, NCBI genome and non-redundant EST databases is depicted in Figure 2. The root of the tree was placed at the phylogenetically most distantly related ALDH12 family, which shows the least sequence conservation to the rest of the plant ALDHs. The phylogenetic analysis demonstrates that the plant ALDHs split up into ten protein families and confirms the assignment of the *Arabidopsis* sequences. Although evolutionary relationships could not be clarified for all the different families, the analysis reveals some interesting observations. Family 21, for example, is represented by just one gene from the moss *Tortula ruralis*, yet it shows a close relationship to the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) family. Similarly, ALDH families 5, 2 and 10 seem to cluster together, and the BADHs (family 10) probably directly diverged from ALDH family 2. Finally, the predicted cytosolic and mitochondrial ALDH forms in family 2 can be clearly separated from each other, which is in accordance with results of a recent characterization of ALDH2 genes from maize and *Arabidopsis* [10].

Family 2 ALDHs

Family 2 ALDHs are mitochondrial or cytosolic homotetrameric enzymes and have been extensively studied in humans and yeast [5,17]. The first ALDH2 gene identified in plants (*rf2*) encodes a nuclear restorer of cytoplasmic male sterility (cms) [18,19], and orthologs have been subsequently characterized in maize (*Zea mays*), rice (*Oryza sativa*) and *Arabidopsis* [10,20,21]. Further studies with one of the mitochondrial ALDHs from rice show that this enzyme might be responsible for efficient detoxification of acetaldehyde during re-aeration after submergence of rice plants [22] and suggest a role during ethanolic fermentation for these ALDHs [20–22]. The *Arabidopsis* genome contains two genes (*ALDH2B7* and *ALDH2B4*) that encode ALDHs predicted to accumulate in the mitochondria [10]. These proteins, ALDH2B7 and ALDH2B4, are 75% identical to each other. *In vitro* activity assays of recombinant proteins indicate that both of these two mitochondrial ALDHs (mtALDHs) can oxidize acetaldehyde and glycolaldehyde, but cannot oxidize L-lactaldehyde, as shown by *E. coli* complementation assays.

Interestingly, all studied plant species, including *Arabidopsis*, contain genes that encode two highly similar mtALDHs. Kinetic studies of the two maize mtALDHs (i.e. RF2A and RF2B) indicate that these two enzymes have overlapping, but non-identical substrates. The RF2A protein has a broad substrate spectrum including aliphatic and aromatic aldehydes, whereas the other mtALDH, RF2B, can oxidize only short-chain aliphatic aldehydes [20]. In addition, these two mtALDHs do not accumulate in the same tissues or at the same times [20]. Hence, it appears that plant mtALDHs have undergone functional specialization.

The *Arabidopsis* genome contains a single gene (*ALDH2C4*) coding for an ALDH that is predicted to accumulate in the cytosol. This protein, ALDH2C4, exhibits 58% and 54% identity to ALDH2B4 and ALDH2B7, respectively, and can oxidize acetaldehyde and glycolaldehyde [10]. In addition, a complementation assay involving an *E. coli* strain that is null for a particular ALDH indicated that this enzyme can also oxidize L-lactaldehyde. More recently, Ramesh B. Nair *et al*. [23] demonstrated that the *reduced epidermal fluorescence1* (*ref1*) mutant of *Arabidopsis* is caused by a mutation in the *ALDH2C4* gene and that ALDH2C4 is involved in the oxidization of sinapaldehyde and coniferaldehyde.

Family 3 ALDHs

In contrast to the mitochondrial and cytosolic ALDH2 family, ALDHs belonging to family 3 are dimeric enzymes located in cytosolic and microsomal fractions. They have been characterized in detail in humans and are associated with carcinogenesis and severe genetic disorders [5]. The first plant ALDH3 gene, *Cp-ALDH*, was isolated from the resurrection plant *Craterostigma plantagineum* in an attempt to identify genes that are crucial for the abscisic acid (ABA) dependent stress response [24]. The *Arabidopsis* genome contains three genes (*ALDH3F1*, *ALDH3I1* and *ALDH3H1*) with 60–76% amino acid identity to *Cp-ALDH*. Sequence analysis of ALDH3I1 predicts a putative chloroplast-targeting peptide of 60 amino acids, whereas ALDH3F1 and ALDH3H1 are probably localized to the cytosol.

ALDH3I1, *ALDH3H1* and *ALDH3F1* differ in their expression patterns. Similar to *Cp-ALDH*, *ALDH3I1* expression is induced after exogenous ABA application, high salinity, dehydration and exposure to heavy metals, H_2O_2 and paraquat (methyl viologen, which binds to thylakoid membranes and causes the formation of superoxide radicals), suggesting a possible role in response to oxidative stress [24,25]. Interestingly, stress-regulated expression of *ALDH3I1* is restricted to leaves and the transcript is almost undetectable in roots. By contrast, *ALDH3H1* is constitutively expressed at a low level in leaves but is activated in response to osmotic stress or after ABA treatment in root tissue, whereas *ALDH3F1* expression is not stress responsive at all (H.H. Kirch *et al.*, unpublished). These results indicate that the *Arabidopsis* ALDH3 gene family members might have evolved as a consequence of functional specialization in different tissues and subcellular compartments.

Kinetic data indicated that Cp-ALDH might be involved in the detoxification of reactive aldehyde species

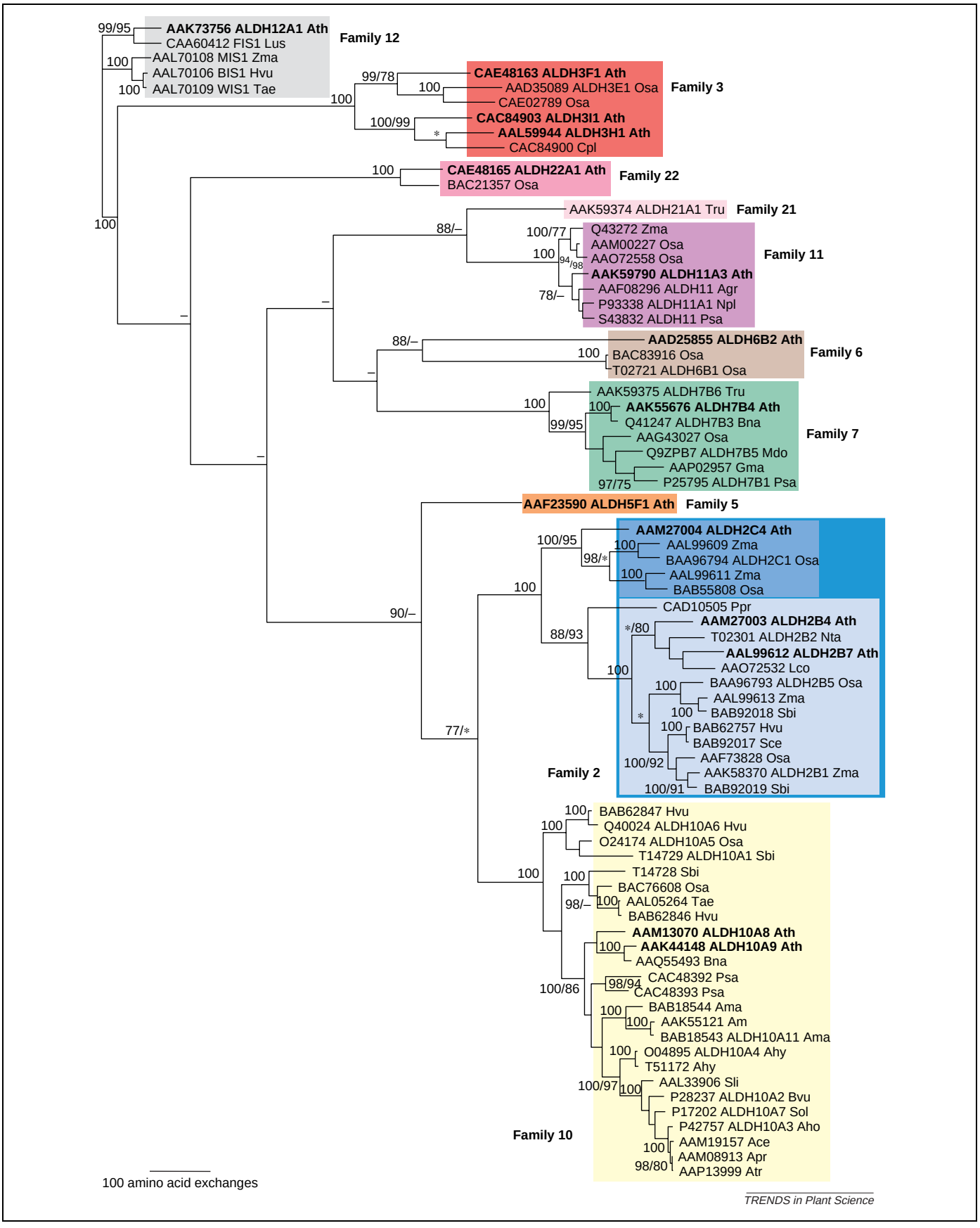

Figure 2. Phylogenetic analysis of the plant aldehyde dehydrogenase (ALDH) superfamily. Deduced protein sequences of 75 plant ALDH sequences were aligned with AlignX™ (Vector NTI™ Suite 5.5, InforMax, North Bethesda, MD, USA) using ClustalW algorithm [38]. Alignments were manually adjusted using GeneDoc [39] and a phylogenetic tree was generated by a neighbor-joining (NJ) method using PAUP* v. 4.0 [40]. Single best trees inferred by maximum parsimony (MP) (7041 steps, not shown) were topologically identical with the NJ tree, with all nodes receiving significant (≥75%) bootstrap support. Bootstrap proportions are expressed as a percentage of 1000 replicates, and values are presented where available at each node [NJ (left) / MP (right)]. Only one value is shown for nodes with identical proportions. Abbreviations: –, bootstrap value not obtainable; *, bootstrap

generated by oxidative-stress-associated lipid peroxidation. A functional analysis of ALDH3I1 suggests that this protein is probably involved in a detoxification pathway in plants that limits aldehyde accumulation and oxidative stress [25]. Transgenic *Arabidopsis* plants ectopically expressing the *ALDH3I1* cDNA exhibit an improved tolerance in response to dehydration, salinity (NaCl), heavy metals (Cu^{2+} and Cd^{2+}), H_2O_2 and paraquat under laboratory conditions. This tolerance is correlated with a decreased accumulation of lipid-peroxidation-derived reactive aldehydes compared with wild-type plants.

Succinic and methylmalonyl semialdehyde dehydrogenases (ALDH families 5 and 6)

The *Arabidopsis* genome contains one succinic semialdehyde dehydrogenase (SSADH; EC 1.2.1.24) gene, *ALDH5F1*, which encodes a protein of 528 amino acids (GenBank Accession number AAF23590; Table 1). To date, *ALDH5F1* is the only identified member of the succinic semialdehyde family in plants (Figure 2). Analysis of the *Arabidopsis* database indicates that a second SSADH 'isoform' might exist, encoding a protein of 509 amino acids (NCBI Accession number AAD30232), with 35 amino acids missing at the N-terminus and an insertion of 16 amino acids at position 86, which could be produced by alternative splicing. However, we have not included this protein in our report here because available cDNA and EST sequences do not support this possibility. SSADHs are involved in γ-aminobutyric acid (GABA) metabolism as part of the 'shunt' from Glu to GABA [26]. The *Arabidopsis* protein is localized to mitochondria and a kinetic analysis showed that the recombinant enzyme was specific for succinic semialdehyde and regulated by adenine nucleotides. T-DNA knockout mutants of *ALDH5F1* result in dwarfed plants with necrotic lesions and are sensitive to both ultraviolet-B light and heat stress [27]. Plants with *ssadh* mutations accumulate elevated levels of H_2O_2, suggesting a role for this gene in plant defense against environmental stress by preventing the accumulation of reactive oxygen species.

The *Arabidopsis* genome contains a single gene, *ALDH6B2*, that encodes a methylmalonyl semialdehyde dehydrogenase (MM-ALDH, EC 1.2.1.27; Table 1). MM-ALDHs are involved in the degradation of valine to propionyl CoA [28]. However, this enzyme has not been extensively studied in plants and the expression of *ALDH6B2* has not been investigated.

Stress-associated ALDHs (ALDH families 7, 10, 11 and 12)

Several osmotic-stress-inducible ALDH genes have been identified in plants and they compose the majority of the protein families ALDH7, ALDH10, ALDH11 and ALDH12 [12]. *Arabidopsis* has a single gene, *ALDH7B4*, that encodes a 'turgor-responsive' [29] or 'stress aldehyde' ALDH [13]. It is postulated that ALDH7 gene products are involved in an unknown adaptive metabolic pathway. In angiosperms, ALDH7B homologs are induced by a range of stresses including dehydration, low temperature, heat shock and high concentrations of ABA [12]. Expression analysis of *ALDH7B4* from *Arabidopsis* also indicates a strong induction by abiotic stress treatments and after application of ABA to whole plants and to roots (H.H. Kirch *et al.*, unpublished). In sharp contrast to this, *ALDH7B6* from the moss *T. ruralis* is constitutively expressed in response to NaCl, ultraviolet-C, ABA and desiccation [13]. In this respect, it is interesting that the moss ALDH is phylogenetically clearly separated from all other plant ALDH7 sequences (Figure 2).

The *Arabidopsis* genome encodes two members of the ALDH10 protein family, ALDH10A8 and ALDH10A9, which are putative dehydration- and salt-inducible BADHs (EC 1.2.1.8) that catalyze the oxidation of betaine aldehyde to the compatible solute glycine betaine [16]. The ability to synthesize and/or accumulate glycine betaine is a ubiquitous adaptation to osmotic stress present in bacteria, animals and plants [30].

The *Arabidopsis* genome contains a single gene, *ALDH11A3*, that encodes a non-phosphorylating GAPDH (GAPN, EC 1.2.1.9). GAPN operates in the cytosol of autotrophic eukaryotes, where it generates NADPH for biosynthetic processes from photosynthetic glyceraldehyde-3-phosphate exported from the chloroplast by the phosphate translocator. ALDH11 catalyzes one of the classic glycolytic 'bypass' reactions unique to plants [31]. Although the role of this enzyme is unclear, Zhifang Gao and Wayne H. Loescher [32] have established that GAPN is the main source of NADPH for mannitol biosynthesis in celery. Expression of *ALDH11A3* from *Arabidopsis* has not been investigated.

A mitochondrial Δ^1-pyrroline-5-carboxylate dehydrogenase (P5CDH, EC 1.5.1.12; Table 1), a key enzyme in the degradation of proline to glutamate [33], is encoded by a single gene, *ALDH12A1*, in *Arabidopsis*. *ALDH12A1* was identified by complementing the yeast Δ*put2* mutant, and its transcript is induced by exogenous proline application and salinity [34].

Novel ALDH (ALDH family 22)

The novel ALDH22 family is so far represented by three plant genes. One of them (*ALDH22A1*) was identified in the *Arabidopsis* genome and encodes a putative aldehyde dehydrogenase of 596 amino acids. Our analyses of rice [35] and maize [36] genome assemblies revealed the

value less than 75%. Members of each ALDH family are highlighted by different colors: gray, family 12 (P5CDH); red, family 3 (variable substrate; class 3 ALDHs); pink, family 22 (novel); purple, family 11 (non-phosphorylating glyceraldehyde-3-phosphate dehydrogenases); brown, family 6 (methylmalonate semialdehyde dehydrogenases); green, family 7 (antiquitin-related; turgor ALDHs); orange, family 5 (succinic semialdehyde dehydrogenases); blue, family 2 (variable substrate, class-1/2 ALDHs; lighter blue, mitochondrial; darker blue, cytosolic); yellow, family 10 (betaine aldehyde dehydrogenases). Each ALDH protein sequence in the tree is identified by its accession number, a gene name (if available) and the respective plant species: Ace, *Atriplex centralasiatica*; Agr, *Apium graveolens*; Aho, *Atriplex hortensis*; Ahy, *Amaranthus hypochondriacus*; Ama, *Avicennia marina*; Apr, *Atriplex prostrata*; Ath, *Arabidopsis*; Atr, *Atriplex triangularis*; Bna, *Brassica napus*; Bvu, *Beta vulgaris*; Cpl, *Craterostigma plantagineum*; Gma, *Glycine max*; Hvu, *Hordeum vulgare*; Lco, *Lotus corniculatus*; Lus, *Linum usitatissimum*; Mdo, *Malus domestica*; Npl, *Nicotiana plumbaginifolia*; Nta, *Nicotiana tabacum*; Osa, *Oryza sativa*; Ppr, *Polytomella*sp. 'Pringsheim 198.80'; Psa, *Pisum sativum*; Sbi, *Sorghum bicolor*; Sce, *Secale cereale*; Sli, *Suaeda liaotungensis*; Sol, *Spinacea oleracea*; Tae, *Triticum aestivum*; Tru, *Tortula ruralis*; Zma, *Zea mays*.

presence of orthologs in both species. The rice ortholog is predicted to encode a closely related protein consisting of 597 amino acids (77% identity, 89% similarity; Figure 2). It is not possible to determine the predicted size of the maize ortholog because only partial genomic and cDNA sequences are available. An initial examination of the *ALDH22A1* transcript accumulation pattern indicates a low constitutive expression, suggesting that the expression of this ALDH gene is not modulated by abiotic stress (H.H. Kirch *et al.*, unpublished).

Conclusions

Under non-stress conditions, ALDHs are involved in fundamental plant metabolism and maintain aldehyde concentrations at nontoxic levels. As abiotic stresses perturb metabolism, ALDHs detoxify stress-generated aldehydes and help maintain the balance of the pool of reducing equivalents. The expression of ALDH-related genes seems to be a common 'stress response' within several divergent plant species from mosses to angiosperms and suggests the existence of common mechanisms for the regulation of their genes, emphasizing the conservation of physiological features in response to osmotic stress.

The *Arabidopsis* genome contains 14 genes belonging to the ALDH gene superfamily, encoding members of nine distinctive protein families, many of which are associated with stress conditions that limit plant growth and development. One research goal will be to establish that ALDH genes encode authentic enzymatic activity against aldehydes and to determine the endogenous substrate(s) and co-enzyme preference (e.g. NAD^+ or $NADP^+$) for each protein family. A second goal will be to use the suite of molecular, genetic and genomic tools available in *Arabidopsis* to identify and characterize the biochemical pathways associated with each protein and/or protein family. Elucidation of the ALDH functions will represent an important step towards understanding basic aspects of osmotic stress responses in plants.

Acknowledgements

Our work was supported in part by grants to A.J.W. (USDA NRI 97–35100), H.H.K. and D.B. (DFG-AFGN #BA 712/3–1), and P.S.S. (USDA NRI 00–0347 and 02–01419). We thank Vasilis Vasiliou (University of Colorado Health Sciences Center, Denver, CO) for helpful discussions.

References

1 Vasiliou, V. *et al*. (1999) Eukaryotic aldehyde dehydrogenase (ALDH) genes: human polymorphisms, and recommended nomenclature based on divergent evolution and chromosomal mapping. *Pharmacogenetics* 9, 421–434
2 Perozich, J. *et al*. (1999) Relationships within the aldehyde dehydrogenase extended family. *Protein Sci.* 8, 137–146
3 Perozich, J. *et al*. (1999) The big book of aldehyde dehydrogenase sequences: an overview of the extended family. In *Advances in Experimental Medicine and Biology* (Vol. 463) (Weiner, H. *et al*.,eds.), pp. 1–7, Kluwer
4 Sophos, N.A. and Vasiliou, V. (2003) Aldehyde dehydrogenase gene superfamily: the 2002 update. *Chem. Biol. Interact.* 143-144, 5–22
5 Yoshida, A. *et al*. (1998) Human aldehyde dehydrogenase gene family. *Eur. J. Biochem.* 251, 549–557
6 Vasiliou, V. *et al*. (2000) Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism. *Chem. Biol. Interact.* 129, 1–19
7 Bartels, D. (2001) Targeting detoxification pathways: an efficient approach to obtain plants with multiple stress tolerance? *Trends Plant Sci.* 6, 284–286
8 Barclay, K.D. and Mckersie, B.D. (1994) Peroxidation reactions in plant membranes – effects of free fatty acids. *Lipids* 29, 877–882
9 Lindahl, R. (1992) Aldehyde dehydrogenases and their role in carcinogenesis. *Crit. Rev. Biochem. Mol. Biol.* 27, 283–335
10 Skibbe, D. *et al*. (2002) Characterization of the aldehyde dehydrogenase gene families of *Zea mays* and *Arabidopsis*. *Plant Mol. Biol.* 48, 751–764
11 Sophos, N.A. *et al*. (2001) Aldehyde dehydrogenase gene superfamily: the 2000 update. *Chem. Biol. Interact.* 130-132, 323–337
12 Wood, A.J. and Krayesky, S. (2002) The role of aldehyde dehydrogenases (ALDHs) in plant stress tolerance. In *Biochemical & Molecular Responses of Plants to the Environment* (Wood, A.J. ed.), pp. 1–14, Research Signpost
13 Chen, X. *et al*. (2002) The stress-responsive gene *ALDH21A1* describes a novel eukaryotic aldehyde dehydrogenase protein family. *J. Plant Physiol.* 159, 677–684
14 Altschul, S.F. *et al*. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410
15 Altschul, S.F. *et al*. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402
16 Weretilnyk, E.A. and Hanson, A.D. (1990) Molecular cloning of a plant betaine aldehyde dehydrogenase, an enzyme implicated in adaptation to salinity and drought. *Proc. Natl. Acad. Sci. U. S. A.* 87, 2745–2749
17 Navarro-Aviño, J.P. *et al*. (1999) A proposal for nomenclature of aldehyde dehydrogenases in *Saccharomyces cerevisiae* and characterization of the stress-inducible *ALD2* and *ALD3* genes. *Yeast* 15, 829–842
18 Ciu, X.Q. *et al*. (1996) The *RF2* nuclear restorer gene of male-sterile, T-cytoplasm maize. *Science* 272, 1334–1336
19 Liu, F. *et al*. (2001) Mitochondrial aldehyde dehydrogenase activity is required for male fertility in maize (*Zea mays* L.). *Plant Cell* 13, 1063–1078
20 Liu, F. and Schnable, P.S. (2002) Functional specialization of maize mitochondrial aldehyde dehydrogenases. *Plant Physiol.* 130, 1657–1674
21 Nakazono, M. *et al*. (2000) Expression of a gene encoding mitochondrial aldehyde dehydrogenase in rice increases under submerged conditions. *Plant Physiol.* 124, 587–598
22 Tsuji, H. *et al*. (2003) Induction of mitochondrial aldehyde dehydrogenase by submergence facilitates oxidation of acetaldehyde during re-aeration in rice. *FEBS Lett.* 546, 369–373
23 Nair, R. *et al*. (2004) The *Arabidopsis thaliana REDUCED EPIDERMAL FLUORESCENCE1* gene encodes an aldehyde dehydrogenase involved in ferulic acid and sinapic acid biosynthesis. *Plant Cell* 16, 544–554
24 Kirch, H.H. *et al*. (2001) Novel ABA- and dehydration-inducible aldehyde dehydrogenase genes isolated from the resurrection plant *Craterostigma plantagineum* and *Arabidopsis thaliana*. *Plant J.* 28, 555–568
25 Sunkar, R. *et al*. (2003) Overexpression of a stress-inducible aldehyde dehydrogenase gene from *Arabidopsis thaliana* in transgenic plants improves stress tolerance. *Plant J.* 35, 452–464
26 Bouché, N. and Fromm, H. (2004) GABA in plants: just a metabolite? *Trends Plant Sci.* 9, 110–115
27 Bouché, N. *et al*. (2003) Mitochondrial succinic-semialdehyde dehydrogenase of the gamma-aminobutyrate shunt is required to restrict levels of reactive oxygen intermediates in plants. *Proc. Natl. Acad. Sci. U. S. A.* 100, 6843–6849
28 Steele, M.I. *et al*. (1992) Characterization of the *mmsAB* operon of *Pseudomonas aeruginosa* PAO encoding methylmalonate-semialdehyde dehydrogenase and 3-hydroxyisobutyrate dehydrogenase. *J. Biol. Chem.* 267, 13585–13592
29 Guerrero, F.D. *et al*. (1990) Turgor-responsive gene transcription and RNA levels increase rapidly when pea shoots are wilted. *Plant Mol. Biol.* 15, 11–26
30 Rhodes, D. and Hanson, A.D. (1993) Quaternary ammonium and tertiary sulphonium compounds in high plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 44, 357–384

31 Plaxton, W. (1996) The organization and regulation of plant glycolysis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47, 185–214
32 Gao, Z. and Loescher, W.H. (2000) NADPH supply and mannitol biosynthesis. Characterization, cloning, and regulation of the non-reversible glyceraldehyde-3-phosphate dehydrogenase in celery leaves. *Plant Physiol.* 124, 321–330
33 Hare, P.D. and Cress, W.A. (1997) Metabolic implications of stress-induced proline accumulation in plants. *Plant Growth Regul.* 21, 79–102
34 Deuschle, K. *et al*. (2001) A nuclear gene encoding mitochondrial delta-pyrroline-5-carboxylate dehydrogenase and its potential role in protection from proline toxicity. *Plant J.* 27, 345–356
35 Yuan, Q.P. *et al*. (2003) The TIGR rice genome annotation resource: annotating the rice genome and creating resources for plant biologists. *Nucleic Acids Res.* 31, 229–233
36 Emrich, S.J. *et al*. (2004) A strategy for assembling the maize (*Zea mays* L.) genome. *Bioinformatics* 20, 140–147
37 The Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
38 Thompson, J.D. *et al*. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680
39 Nicholas, K.B. *et al*. (1997) GeneDoc; analysis and visualization of genetic variation. *EMBnet News* 4, 14 (http://www.hgmp.mrc.ac.uk/embnet.news/)
40 Swofford, D.L. (2002) *PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods) 4.0 Beta*, Sinauer Associates, Sunderland, MA, USA