

Role and structural characterization of plant aldehyde dehydrogenases from family 2 and family 7

Radka Končítíková*†, Armelle Vigouroux‡, Martina Kopečná*, Tomáš Andree†, Jan Bartošš, Marek Šebela*, Solange Moréra‡¹ and David Kopečný*¹

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

†Department of Biochemistry, Faculty of Science, Palacký University, Šlechtitelů 11, Olomouc CZ-783 71, Czech Republic

‡Laboratoire d'Enzymologie et Biochimie Structurales, CNRS, Avenue de la Terrasse, Gif-sur-Yvette 91198, France

§Centre of Plant Structural and Functional Genomics, Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Šlechtitelů 31, Olomouc CZ-78371, Czech Republic

Aldehyde dehydrogenases (ALDHs) are responsible for oxidation of biogenic aldehyde intermediates as well as for cell detoxification of aldehydes generated during lipid peroxidation. So far, 13 ALDH families have been described in plants. In the present study, we provide a detailed biochemical characterization of plant ALDH2 and ALDH7 families by analysing maize and pea ALDH7 (ZmALDH7 and PsALDH7) and four maize cytosolic ALDH(cALDH)2 isoforms RF2C, RF2D, RF2E and RF2F [the first maize ALDH2 was discovered as a fertility restorer (RF2A)]. We report the crystal structures of ZmALDH7, RF2C and RF2F at high resolution. The ZmALDH7 structure shows that the three conserved residues Glu¹²⁰, Arg³⁰⁰ and Thr³⁰² in the ALDH7 family are located in the substrate-binding site and are specific to this family. Our kinetic analysis demonstrates that α -amino adipic semialdehyde, a lysine catabolism intermediate, is the preferred substrate for plant ALDH7. In contrast, aromatic aldehydes including

benzaldehyde, anisaldehyde, cinnamaldehyde, coniferaldehyde and sinapaldehyde are the best substrates for cALDH2. In line with these results, the crystal structures of RF2C and RF2F reveal that their substrate-binding sites are similar and are formed by an aromatic cluster mainly composed of phenylalanine residues and several nonpolar residues. Gene expression studies indicate that the *RF2C* gene, which is strongly expressed in all organs, appears essential, suggesting that the crucial role of the enzyme would certainly be linked to the cell wall formation using aldehydes from phenylpropanoid pathway as substrates. Finally, plant ALDH7 may significantly contribute to osmoprotection because it oxidizes several aminoaldehydes leading to products known as osmolytes.

Key words: aldehyde dehydrogenase 2 (ALDH2), aldehyde dehydrogenase 7 (ALDH7), benzaldehyde, coniferaldehyde, cytokinin, fertility restorer.

INTRODUCTION

Aldehyde dehydrogenases (ALDHs) constitute a superfamily of NAD(P)⁺-dependent enzymes that catalyse irreversible oxidation of aldehydes to the corresponding carboxylic acids. Aldehydes, which are highly reactive molecules, are toxic at high concentrations. Therefore, ALDHs play a crucial role in detoxifying aldehydes produced by various metabolic pathways. They also play a role during adaptation to various stress conditions such as salinity, drought, heat and cold [1]. ALDHs are classified according to their sequence identity. Those sharing more than 40% sequence identity belong to the same family, whereas those with more than 60% form a subfamily

[2]. The superfamily of plant ALDHs currently contains 13 distinct families: ALDH2, ALDH3, ALDH5, ALDH6, ALDH7, ALDH10, ALDH11, ALDH12, ALDH18, ALDH21, ALDH22, ALDH23 and ALDH24 [3]. Only ALDH2, 3, 5, 6, 7 and 18 families possess mammalian orthologues.

The plant ALDH2 family (EC 1.2.1.-) comprises mitochondrial (mtALDH) and cytosolic (cALDH) isoforms split into ALDH2B and ALDH2C subfamilies respectively. They share ~54%–63% amino-acid identity with human ALDH2 (hALDH2), which belongs to the ALDH2A subfamily. hALDH2 plays a role in ethanol metabolism by catalysing the oxidation of ethanol-derived acetaldehyde to acetate [4–6]. Its crystal structure is known (PDB 1CW3) [6]. The first maize (*Zea mays*) *ADLH2*

Abbreviations: ABAL, 4-aminobutyraldehyde; ALDH, aldehyde dehydrogenase; AMADH, aminoaldehyde dehydrogenase; APAL, 3-aminopropionaldehyde; AASAL, α -amino adipate-semialdehyde; BAL, betaine aldehyde; cALDH, cytosolic aldehyde dehydrogenase; DAP, days after pollination; GABA, γ -aminobutyric acid; GBAL, 4-guanidinobutyraldehyde; hALDH2, human ALDH2; LKR, lysine-ketoglutarate reductase; MASAL, adipic semialdehyde methyl ester; mtALDH, mitochondrial aldehyde dehydrogenase; P6C, Δ^1 -piperidine-6-carboxylate; PCAL, pyridine carboxaldehyde; PsALDH, ALDH from *Pisum sativum* (pea); RF, fertility restorer; TMABAL, 4-(trimethylamino)butyraldehyde; SDH, saccharopine dehydrogenase; TMAPAL, 3-(trimethylamino)propionaldehyde; ZmALDH, ALDH from *Zea mays* (maize).

¹ Correspondence may be addressed to either of these authors (email david.kopecnny@upol.cz or morera@lebs.cnrs-gif.fr).

Sequence data can be found in the EMBL/GenBank data libraries under accession numbers KJ004509 for ZmALDH7, KJ004510 for RF2A, KJ004511 for RF2B, KJ004512 for RF2C, KM225857 for RF2D, KM225858 for RF2E and KJ004513 for RF2F.

The structure for ZmALDH2-3 (RF2C) complexed with NAD⁺, ZmALDH2-6 (RF2F) complexed with NAD⁺ and ZmALDH7 complexed with NAD⁺ will appear in the PDB under accession code 4PXL, 4PZZ and 4PXN respectively.

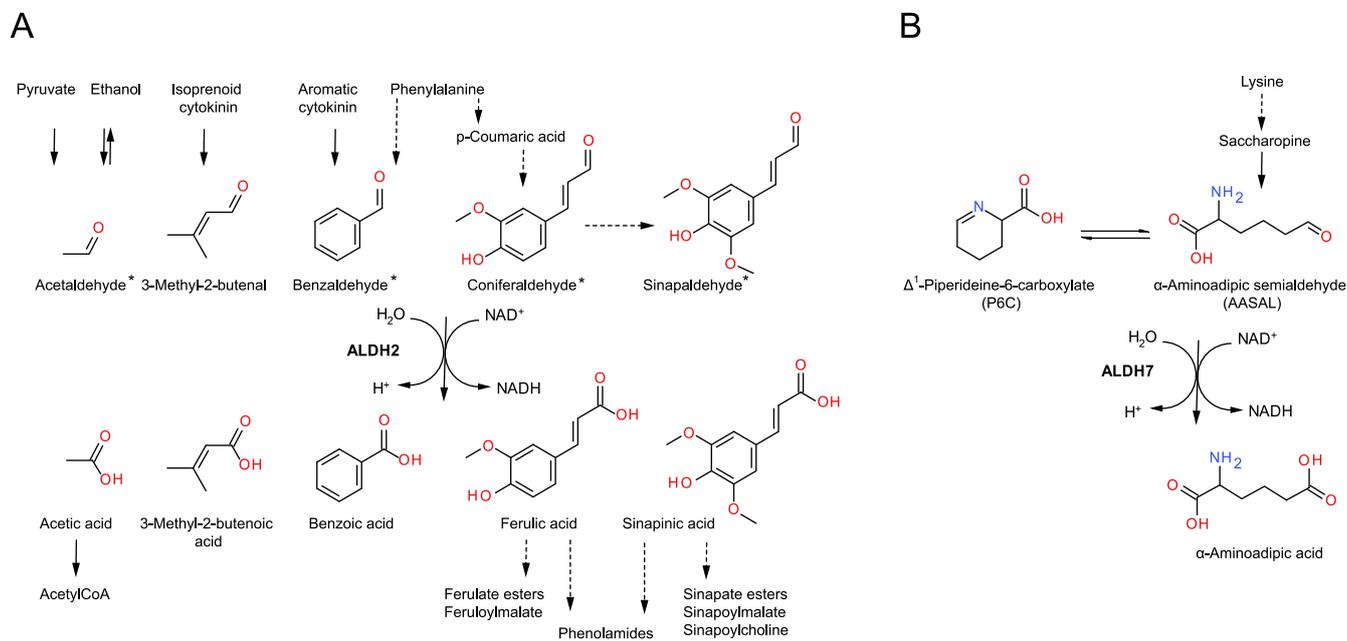


Figure 1 Possible reactions catalysed by plant ALDH2 and ALDH7

(A) Plant ALDH2 members catalyse conversion of acetaldehyde, 3-methyl-2-butenal, benzaldehyde, coniferaldehyde and sinapaldehyde. Confirmed *in vivo* substrates are marked with an asterisk [13–16]. (B) ALDH7 oxidizes AASAL to α -aminoadipic acid as confirmed *in vivo* for hALDH7 [18].

gene was identified as a male fertility restorer (*RF*)2A [7,8] followed by *RF*2B [9]. Both genes code for homotetrameric mtALDHs with an acetaldehyde activity. *RF*2A has a wide substrate specificity covering aliphatic and aromatic aldehydes, whereas *RF*2B oxidizes only short-chain aliphatic aldehydes. *RF*2A, which accumulates in tapetal cells, is required for normal another development. Two more maize genes, *RF*2C and *RF*2D, were identified, but not characterized further [10]. Two rice (*Oryza sativa*) mtALDHs are abundant in panicles [11] and two tobacco mtALDHs are abundant in reproductive tissues and play a role in pollen development [12]. All of the mtALDHs above along with those from *Arabidopsis* [10] produce acetate for acetyl-CoA biosynthesis from acetaldehyde generated via ethanolic fermentation, bypassing pyruvate dehydrogenase [11,13]. The recently characterized snapdragon benzaldehyde dehydrogenase [14] is indeed an mtALDH2. cALDH2 [reduced epidermal fluorescence 1 (REF1), ALDH2C4] isoforms from *Arabidopsis* and oilseed rape [15,16] oxidize sinapaldehyde and coniferaldehyde *in vivo* and play a major role in the formation of soluble and cell wall-linked ferulate and sinapate esters. A reaction scheme with possible *in vivo* substrates is shown in Figure 1A.

ALDH7 (EC 1.2.1.31) is also known as Δ^1 -piperidine-6-carboxylate (P6C) dehydrogenase, α -aminoadipate-semialdehyde (AASAL) dehydrogenase or antiquitin. The *ALDH7* gene is highly conserved across species suggesting a conserved role within the cell using identical physiological substrates. Human ALDH7 (hALDH7 and ALDH7A1) shares ~60% sequence identity with plant orthologues (ALDH7B subfamily) and its crystal structure is known (PDB 2J6L) [17]. The enzyme is primarily involved in the metabolism of lysine and catalyses the conversion of AASAL to α -aminoadipate (Figure 1B) [18]. P6C, which is the cyclic Schiff base of AASAL,

is in equilibrium with AASAL in solution. Plant ALDH7 was first reported from pea (*Pisum sativum*) as a ‘26 g protein’ involved in the regulation of turgor pressure. Its expression is induced by dehydration of leaves and the stem [19]. Later, *ADLH7* genes in oilseed rape (*btg-26*), *Arabidopsis*, rice and foxtail millet have also been reported to be induced by dehydration, high salinity, abscisic acid treatment or other stress conditions [20–23]. Soybean and *Arabidopsis* ALDH7 overexpressers confer tolerance to abiotic and oxidative stresses [24,25]. Stress tolerance is accompanied by a reduction of hydrogen peroxide and malondialdehyde derived from lipid peroxidation. ALDH7 was found to be important for seed viability and maturation in rice. Indeed, *osaladh7* mutants are more sensitive to various stresses and mutant seeds accumulate more malondialdehyde and melanoidin pigment, a product of a Maillard reaction between carbonyl and amino compounds [26].

We have previously kinetically and structurally characterized several ALDH10 [aminoaldehyde dehydrogenase (AMADH)] family members from pea, maize and tomato [27–29] linked to polyamine catabolism, osmoprotection, secondary metabolism (fragrance) and carnitine biosynthesis. In the present study, we investigate the structure–function relationship of plant ALDH2 and ALDH7 families by analysing the maize and pea ALDH7 families represented by a single member per plant species (ZmALDH7 and PsALDH7) as well as the maize ALDH2 family represented by six members with a special focus on four uncharacterized cALDH2 isoforms *RF*2C, *RF*2D, *RF*2E and *RF*2F (Table 1). To better understand substrate specificity, we solved the crystal structures of ZmALDH7 (systematic name ALDH7B6), *RF*2C (ZmALDH2-3, ALDH2C1) and *RF*2F (ZmALDH2-6, ALDH2C5) with NAD^+ at 2.94, 2.25 and 2.40 Å (1 Å = 0.1 nm) resolution respectively, and compared them with the human enzymes. Finally, we analysed the spatial and temporal

Table 1 Enzyme nomenclature and GenBank accession numbers of plant ALDH2 and 7 family members studied in the present work

The alternative and systematic names follow previous publications dealing with nomenclature of plant ALDHs [3,30]. Sequence identities (in %) of maize RF2C and maize ALDH7 with other ALDHs was calculated by Lalign [31].

Name	Alternative/systematic name	GenBank accession number	Maize chromosome	Exon #	AA #	Sequence identity (%)	
						RF2C	ZmALDH7
RF2A	ZmALDH2-1/ALDH2B2	KJ004510	9	11	549	59.1	29.5
RF2B	ZmALDH2-2/ALDH2B5	KJ004511	4	9	550	60.2	28.2
RF2C	ZmALDH2-3/ALDH2C1	KJ004512	3	7	502	–	27.5
RF2D	ZmALDH2-4/ALDH2C2	KM225857	3	8	511	73.8	29.7
RF2E	ZmALDH2-5/ALDH2C4	KM225858	8	10	501	71.1	29.1
RF2F	ZmALDH2-6/ALDH2C5	KJ004513	6	8	519	70.2	29.4
ZmALDH7	ALDH7B6	KJ004509	2	14	509	27.5	–
PsALDH7	ALDH7B1	X54359	–	–	508	28.4	78.2
hALDH7	ALDH7A1	AK312459	–	–	511	30.0	59.3

expression of *ALDH2* and *ALDH7* during the first 2 weeks in developing maize seedlings to assess their roles in early stages of development and also in various organs.

MATERIALS AND METHODS

Cloning, expression and purification of ALDHs from maize and pea

The total RNA from various maize organs (*Z. mays* cv. Cellux, Morseva) and apical meristem of pea seedlings (*P. sativum*) was extracted using the RNAqueous kit and plant RNA isolation aid (Ambion). The RNA was treated twice with the Turbo DNase-free kit (Ambion). The cDNA was synthesized using Superscript II RT (Life Technologies). Six *ZmALDH2* ORFs and three *ALDH7* ORFs, namely *ZmALDH7*, *PsALDH7* and *hALDH7*, were amplified using gene-specific primers and Accuprime Pfx polymerase (Life Technologies). Obtained maize sequences were submitted to GenBank. *ALDHs* were ligated into a pCDFDuet vector (Novagen), and then transformed into T7 expressing *Escherichia coli* cells (NEB). Cells were grown at 37°C in LB media, at $D_{600} = 0.5$, the cultures were supplemented with 0.5 mM isopropyl- β -thiogalactopyranoside for protein expression and incubated at 20°C overnight. Recombinant ALDHs were purified on HisPur Cobalt Spin Columns (Thermo Fisher Scientific) equilibrated with 20 mM Tris/HCl buffer, pH 8.0, containing 100 mM NaCl, 10 mM imidazole and 5% glycerol. Elution was performed using 250 mM imidazole in the same buffer. Enzymes were concentrated using Amicon 30 kDa filters (Merck) and further purified by gel filtration chromatography on a HiLoad 26/60 Superdex 200 column in 50 mM Tris/HCl buffer, pH 8.0, 150 mM NaCl. Protein content was measured using Coomassie plus (Bradford) protein assay kit (Thermo Fisher Scientific) using BSA as a standard and using absorption coefficients calculated from sequences (<http://web.expasy.org/protparam/>). The purity of recombinant ALDHs was confirmed by SDS/PAGE using NuPAGE system (Life Technologies; Supplementary Figure S1). For activity measurements and storage, 5% glycerol was added to the enzyme solution.

Enzyme kinetics

Activity was measured by monitoring the NADH formation ($\epsilon_{340} = 6.62 \text{ mM}^{-1} \cdot \text{cm}^{-1}$) at 37°C on a UV-Vis spectrophotometer 8453 (Agilent) in a cuvette and by monitoring fluorescence emission of NADH at 460 nm upon excitation

at 365 nm on a FluoroLog-3 spectrofluorometer (Horiba). The reaction mixtures contained 150 mM Tris/HCl buffer, pH 8.0 (for RF2C and RF2F) or pH 7.5 (for RF2D and RF2E), 1 mM NAD^+ and ALDH2 or 100 mM sodium pyrophosphate buffer, pH 8.0, 2.5 mM NAD^+ and ALDH7. The enzyme reaction was started by the addition of aldehyde at a final 1 mM concentration. Kinetic constants for selected substrates (up to 6 mM concentration) were determined using GraphPad Prism 5.0 data analysis software (GraphPad Software). Elementary aliphatic and aromatic aldehydes, pyridine carboxaldehydes (PCALs), MASAL (methyl ester of adipic semialdehyde), BAL (betaine aldehyde) chloride together with APAL (3-aminopropionaldehyde) and ABAL (4-aminobutyraldehyde) diethylacetals were purchased from Sigma-Aldrich Chemie. AASAL ethylene acetal was purchased from Chiralix. Diethylacetals of 4-guanidinobutyraldehyde (GBAL), 3-guanidinopropionaldehyde (GPAL), 3-(trimethylamino)propionaldehyde (TMAPAL) and 4-(trimethylamino)butyraldehyde (TMABAL) were synthetic preparations [32,33]. Free aminoaldehydes were prepared by heating their acetals in a plugged test tube with 0.2 M HCl for 10 min [34].

Crystallization and structure determination

Crystallization conditions were screened using Qiagen kits. Crystals were obtained in hanging drops by mixing equal volumes of protein solution and a precipitant solution containing 30% (w/v) PEG 400, 100 mM CaCl_2 and 100 mM sodium acetate, pH 4.5, for RF2C; 35% 2-methyl-2,4-pentanediol (MPD), 100 mM 4-morpholineethanesulfonic acid (MES), pH 6.5, and 15% (w/v) PEG 4000 for RF2F; 40% MPD and 200 mM ammonium phosphate for ZmALDH7. Crystals were directly flash-frozen in liquid nitrogen. Diffraction data were collected at 100 K on the Proxima 1 beamline for RF2C and RF2F and on Proxima 2 beamline for ZmALDH7 at the SOLEIL synchrotron at 2.25, 2.40 and 2.94 Å resolution respectively. Intensities were integrated using the XDS program [35] and data quality was assessed using the correlation coefficient $CC_{1/2}$ [36,37]. The crystal structures were determined by performing molecular replacement with Phaser [38], using the monomers of hALDH2 (PDB codes 1CW3 and 1O05) and of seabream ALDH7 (PDB code 2JG7) as search models [6,39]. Both models were refined with NCS restraints and TLS using Buster 2.10 [40]. Electron density maps were evaluated using COOT [41]. Refinement statistics are presented in Table 2. Models of RF2D

Table 2 Data collection and refinement statistics

	RF2C (ZmALDH2-3)	RF2F (ZmALDH2-6)	ZmALDH7
PDB code	4PXL	4PZ2	4PXN
Space group	P2 ₁ 2 ₁ 2	I4 ₁	I2 ₁ 2 ₁ 2 ₁
Asymmetric unit	1 dimer	1 tetramer	2 monomers
Oligomeric state	Tetrameric	Tetrameric	Tetrameric
Unit cell (Å)			
<i>a</i>	109.9	233.9	79.2
<i>b</i>	126.1	233.9	162.5
<i>c</i>	78.3	82.9	188.6
$\alpha = \beta = \gamma$ (°)	90.0	90.0	90.0
Resolution (Å)	49.1–2.25	50.0–2.40	42.11–2.94
Observed reflections	381340	603883	184717 (28407)
	(59526) [*]	(95778)	
Unique reflections	52406 (8260)	87 611 (13838)	26085 (4014)
Completeness (%)	99.8 (99.0)	99.7 (98.3)	99.3 (95.7)
<i>I</i> / σ (<i>I</i>)	13.62 (2.1)	11.12 (1.2)	9.91 (1.8)
CC _{1/2} [†]	–	99.7 (58.4)	99.6 (58.3)
<i>R</i> _{sym} (%)	10.7 (78.9)	9.7 (155.7)	19.4 (153.3)
<i>R</i> _{cryst} (%)	17.1	19.0	18.8
<i>R</i> _{free} (%)	20.5	21.3	21.7
RMSD bond lengths (Å)	0.010	0.009	0.008
RMSD bond angles (°)	1.11	1.14	1.14
<i>B</i> average value (Å ²)			
Protein	41.9	83.7	80.4
NAD ⁺	55.9	85.2	100.5
Solvent	43.2	72.3	60.1

*Numbers in parentheses represent values in the highest resolution shell: 2.25–2.38 Å (ZmALDH2-3), 2.40–2.54 Å (ZmALDH2-6) and 2.94–3.12 Å (ZmALDH7).

[†]CC_{1/2} stands for a percentage correlation between intensities from random half-datasets [36,37].

and RF2E were constructed using SWISS-MODEL server [42] with RF2C structure as a template. Molecular graphics images were generated using PYMOL (www.pymol.org).

qPCR analysis

The total RNA from 3- to 13-day-old and 3-month-old maize plants was isolated using the RNAqueous kit and plant RNA isolation aid (Ambion) and treated twice with a Turbo DNase-free kit (Ambion). First-strand cDNA was synthesized by RevertAid H Minus reverse transcriptase and oligo(dT) primers (Thermo Fisher Scientific). RNA from four biological replicates was transcribed in two independent reactions and PCR was performed in triplicate. Diluted cDNA samples were used as templates in real-time PCRs containing TaqMan Gene Expression Master Mix (Life Technologies), both primers at 300 nM concentrations and 250 nM TaqMan 6-FAM TAMRA probe on a StepOnePlus Real-Time PCR System. Primers and TaqMan probes were designed using Primer Express 3.0 software (Life Technologies) and are shown in Supplementary Table S1. Plasmid DNA carrying the ORF of the respective maize *ALDH2* or *ALDH7* gene was used as a template for a calibration curve to determine the PCR efficiencies of designed probes and primer pairs as well as to verify their specificity. Cycle threshold values were normalized with respect to elongation factor 1 α and β -actin genes and amplification efficiency. Expression values were determined and statistically evaluated using the DataAssist v3.0 Software package (Life Technologies). Transcript abundance values are shown in Supplementary Table S2.

RESULTS AND DISCUSSION

Gene models and phylogenetic analysis of ALDH2 and ALDH7 families

The maize genome database (<http://ensembl.gramene.org>, AGPv3) indicates the existence of six putative ALDH2 genes and one ALDH7 gene. We cloned the complementary DNAs (cDNAs) of these seven ALDHs using gene-specific primers in order to identify the correct gene models in each case and to obtain the corresponding recombinant proteins. We deposited their sequences in GenBank. The unique ZmALDH7 gene composed of 14 exons lies on chromosome 2 (Table 1; Figure 2A). ZmALDH2 genes are composed of 7–11 exons with the last three conserved, except for *RF2E*, whose last exon is split into two by insertion of a CACTA-like DNA transposon. The first two exons are also conserved among the *ZmALDH2* gene family. Both *RF2A* and *RF2B* genes carry an additional exon at the 5'-end coding for a signal sequence towards the mitochondrion. The high variability in exon number rather than in total length appears in the middle section of the *ALDH2* genes. For example, although two exons (648 and 138 bp long) are found in *RF2C* gene, five exons (154, 230, 90, 174 and 138 bp long) are found in *RF2A* gene (Figure 2A). *RF2D* and *RF2E* are the most homologous, even though the third exon of *RF2D* is split into two in *RF2E* and separated by a long intron. This intron emerges with the insertion of a CACTA-like DNA transposon whose footprint (target site duplication) is unambiguously detectable in the gene/intron sequence. The intron contains at least eight other transposable elements mainly LTR retroelements. Currently, two ORFs (GRMZM2G380438 comprising the first three exons and GRMZM2G407949 comprising the remaining seven exons) represent *RF2E*, but the fact that the complete cDNA covering all 10 exons was cloned (Table 1; Figure 2A) shows that only one ORF exists and the gene is either incorrectly annotated or is different between the maize cultivar we used and the sequenced B73 line. The gene structures are similar to those in rice and *Arabidopsis* [22] with a high variability in exon/intron number in the middle part of the *ALDH2* genes.

Phylogenetic analysis (Figure 2B) shows that both maize *RF2A* and *RF2B* belong to subfamily ALDH2B composed of mtALDH2 members from different plant species including the snapdragon (*Anrrihrinum majus*) benzaldehyde dehydrogenase [14], rice, tobacco and *Aradopsis* mtALDHs [10–12]. The remaining four uncharacterized enzymes *RF2C*, *RF2D*, *RF2E* and *RF2F*, which share about 70% sequence identity (Table 1), belong to subfamily ALDH2C composed of cALDH2 members including *Arabidopsis* ALDH2C4, also named coniferaldehyde or sinapaldehyde dehydrogenase [15]. For comparison, rice, the other monocotyledonous plant, contains two genes, *ALDH2B5* and *ALDH2B1*, coding for mtALDH2 isoforms and three genes *ALDH2C1*, *ALDH2C2* and *ALDH2C3* coding for cALDH2 isoforms [11,22]. ZmALDH7, which shares 78% sequence identity with PsALDH7, appears in the unique ALDH7B subfamily with other plant ALDH7s (Figure 2C). So far, *ALDH7* gene is absent from available algae genomes and exists in bryophytes like *Physcomitrella patens* or *Tortula ruralis* [46]. Angiosperms usually carry one *ALDH7* gene, but some species, for example poplar or soybean, can carry two or more *ALDH7* genes.

Substrate specificity and the role of the maize ALDH2 family

We purified four maize cALDH2 isoforms (ALDH2C subfamily), namely *RF2C*, *RF2D*, *RF2E* and *RF2F* (Supplementary Figure S1) and measured their enzyme kinetics using various

Table 3 Kinetic parameters of four ALDH2 family members from maize for selected substrates

K_m and k_{cat} values are given in μM and s^{-1} respectively. Kinetic constants were measured in 0.15 M Tris/HCl, pH 8.0 (for RF2C and RF2F) and pH 7.5 (for RF2D and RF2E), using 1.0 mM NAD^+ . Specific activity values with 1 mM hexanal were 62 nkat mg^{-1} for RF2C, 12.2 nkat mg^{-1} for RF2D, 14.6 nkat mg^{-1} for RF2E and 8.9 nkat mg^{-1} for RF2F. The K_m value for NAD^+ and NADP^+ was measured at a fixed 1 mM concentration of hexanal. The symbol 'n.d.' stands for not determined.

Ligand	RF2C			RF2D			RF2E			RF2F		
	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\text{M}^{-1}\text{s}^{-1}$)	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\text{M}^{-1}\text{s}^{-1}$)	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\text{M}^{-1}\text{s}^{-1}$)	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\text{M}^{-1}\text{s}^{-1}$)
NAD^+	44 ± 3	3.46 ± 0.04	7.9×10^4	59 ± 7	0.85 ± 0.009	1.5×10^4	45 ± 2	0.86 ± 0.009	1.9×10^4	86 ± 8	0.55 ± 0.004	6.4×10^3
NADP^+	3012 ± 190	0.16 ± 0.01	5.4×10^1	1488 ± 78	0.21 ± 0.005	1.4×10^2	1364 ± 63	0.22 ± 0.004	1.6×10^2	1672 ± 113	0.14 ± 0.005	8.4×10^1
Acetaldehyde	2419 ± 170	0.59 ± 0.02	2.4×10^2	92 ± 9	0.37 ± 0.022	4.1×10^3	90 ± 8	0.59 ± 0.030	6.6×10^3	1090 ± 107	0.05 ± 0.001	4.5×10^1
Propionaldehyde	239 ± 22	2.12 ± 0.07	8.9×10^3	n.d.	n.d.	n.d.	50 ± 4	0.41 ± 0.009	8.1×10^3	944 ± 57	0.10 ± 0.001	1.0×10^2
Butyraldehyde	210 ± 20	3.54 ± 0.11	1.7×10^4	n.d.	n.d.	n.d.	30 ± 2	0.44 ± 0.005	1.5×10^4	291 ± 10	0.13 ± 0.002	4.4×10^2
Valeraldehyde	99 ± 5	3.20 ± 0.07	3.2×10^4	n.d.	n.d.	n.d.	31 ± 1	0.53 ± 0.004	1.7×10^4	142 ± 9	0.22 ± 0.004	1.5×10^3
Isovaleraldehyde	33 ± 3	2.32 ± 0.05	7.0×10^4	n.d.	n.d.	n.d.	52 ± 3	0.12 ± 0.002	2.4×10^3	269 ± 22	0.18 ± 0.004	6.9×10^2
3-Methyl-2-butenal	54 ± 11	0.24 ± 0.06	4.4×10^3	64 ± 3	0.08 ± 0.001	1.2×10^2	63 ± 5	0.06 ± 0.001	9.7×10^2	250 ± 9	0.14 ± 0.002	5.6×10^2
Hexanal	15 ± 1	4.54 ± 0.08	3.0×10^5	48 ± 4	0.91 ± 0.032	1.9×10^4	42 ± 5	0.85 ± 0.031	2.0×10^4	424 ± 21	0.72 ± 0.011	1.7×10^3
<i>l</i> -2-Hexenal	39 ± 5	0.35 ± 0.02	9.0×10^3	31 ± 1	0.08 ± 0.001	2.5×10^3	34 ± 2	0.06 ± 0.001	1.7×10^3	156 ± 11	0.014 ± 0.001	9.0×10^1
Heptanal	172 ± 11	0.77 ± 0.01	4.5×10^3	n.d.	n.d.	n.d.	130 ± 6	0.14 ± 0.002	1.0×10^3	1717 ± 99	0.007 ± 0.001	3.9
Octanal	312 ± 20	1.97 ± 0.04	6.3×10^3	n.d.	n.d.	n.d.	196 ± 30	0.18 ± 0.009	9.4×10^2	1852 ± 96	0.03 ± 0.001	1.5×10^1
Nonanal	51 ± 5	1.63 ± 0.06	3.2×10^4	159 ± 9	0.36 ± 0.009	2.3×10^3	143 ± 13	0.36 ± 0.010	2.5×10^3	181 ± 10	0.41 ± 0.013	2.3×10^3
<i>l</i> -2-Nonenal	19 ± 13	2.54 ± 0.30	1.3×10^5	89 ± 7	0.1 ± 0.004	1.1×10^3	79 ± 4	0.11 ± 0.001	1.4×10^3	177 ± 6	0.03 ± 0.001	1.6×10^2
Benzaldehyde	83 ± 9	1.21 ± 0.06	1.5×10^4	22 ± 2	1.49 ± 0.105	6.8×10^4	12 ± 1	0.99 ± 0.036	8.3×10^4	146 ± 7	0.29 ± 0.003	2.0×10^3
<i>m</i> -Anisaldehyde	203 ± 33	1.25 ± 0.14	6.2×10^3	17 ± 2	0.95 ± 0.037	5.6×10^4	19 ± 3	0.65 ± 0.024	3.4×10^4	260 ± 29	0.35 ± 0.020	1.4×10^3
<i>p</i> -Anisaldehyde	299 ± 7	0.40 ± 0.03	1.3×10^3	6 ± 1	0.93 ± 0.020	1.5×10^5	13 ± 2	1.35 ± 0.078	1.0×10^5	536 ± 31	0.82 ± 0.016	1.6×10^3
Phenylacetaldehyde	480 ± 99	0.13 ± 0.02	2.8×10^2	196 ± 9	0.10 ± 0.002	5.3×10^2	195 ± 11	0.17 ± 0.003	8.9×10^2	1.8 ± 0.23	0.42 ± 0.014	2.3×10^5
Cinnamaldehyde	10 ± 1	2.95 ± 0.17	3.0×10^5	69 ± 7	0.31 ± 0.014	4.5×10^3	52 ± 2	0.65 ± 0.011	1.2×10^4	116 ± 7	0.17 ± 0.005	1.5×10^3
Hydrocinnamaldehyde	22 ± 3	2.41 ± 0.16	1.1×10^5	103 ± 10	0.60 ± 0.031	5.8×10^3	147 ± 35	1.00 ± 0.101	6.8×10^3	203 ± 17	0.17 ± 0.008	8.6×10^2
Coniferaldehyde	19 ± 3	2.65 ± 0.37	1.4×10^5	15 ± 3	0.37 ± 0.055	2.5×10^4	14 ± 3	0.33 ± 0.027	2.4×10^4	10 ± 1	0.08 ± 0.006	7.8×10^4
Sinapaldehyde	14 ± 2	0.24 ± 0.02	2.4×10^4	26 ± 6	0.33 ± 0.088	1.3×10^4	23 ± 4	0.15 ± 0.010	6.6×10^3	7 ± 1	0.06 ± 0.006	8.8×10^4

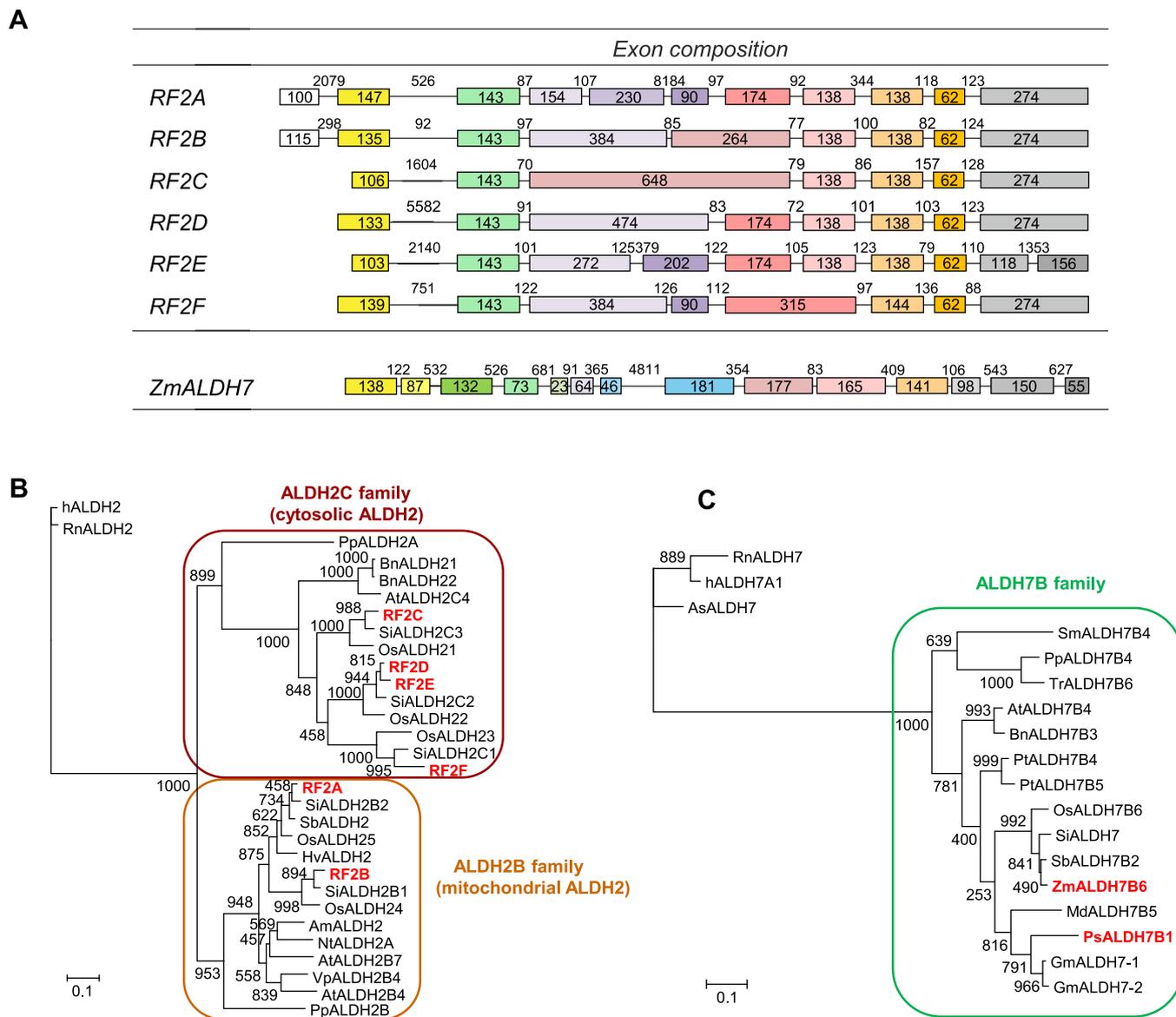


Figure 2 Phylogeny and gene models for studied plant ALDH2 and ALDH7 families

(A) Gene models for *Zea mays* ALDH2 and ALDH7 families. Gene models were constructed using cloned cDNA sequences (present work) and genomic DNA sequences obtained from maize genome (http://ensembl.gamene.org/Zea_mays/Info/Index). (B and C) Phylogenetic trees illustrating the relatedness of maize ALDH2 and ALDH7 to other ALDHs from the same families. Amino acid alignments were performed using MUSCLE v3.8 [43] followed by Gblocks [44]. A maximum likelihood phylogeny with bootstrap analysis was performed with PhyML v3.0 [45] using LG amino acid replacement matrix. Bootstrap values are given for the nodes. Only subset of known ALDHs was chosen, especially those that have been previously studied. ALDH2 sequences include those from *Antirrhinum majus* (FJ151199), *Arabidopsis thaliana* (At3g48000, At1g23800 and At3g24503), *Brassica napus* (FN995990 and FN995991), *Homo sapiens* (AY621070), *Hordeum vulgare* (BAB62757), *Nicotiana tabacum* (CAA71003), *Oryza sativa* (Os06g15990, Os02g49720, Os01g40860, Os01g40870 and Os06g39230), *Physcomitrella patens* (XP_001767457 and XP_001785650), *Rattus norvegicus* (P11884), *Sorghum bicolor* (BAB92019), *Vitis pseudoreticulata* (DQ150256), *Setaria italica* (XM_004953741, XM_004965148, XM_004965940, XM_004968994 and XM_004968990). ALDH7 sequences include those from *Acanthopagrus schlegelii* (AY847462), *Arabidopsis thaliana* (At1g54100), *Brassica napus* (Q41247), *Glycine max* (Glyma.15G178400 and Glyma.09G070300), *Homo sapiens* (AK312459), *Malus domestica* (BAA75633), *Oryza sativa* (Os09g26880), *Physcomitrella patens* (XP_001778351), *Pisum sativum* (X54359), *Populus trichocarpa* (Potri.003G067700 and Potri.001G167100), *Rattus norvegicus* (NP_001258034.1), *Selaginella moellendorffii* (XP_002961173), *Setaria italica* (XM_004956875), *Sorghum bicolor* (XP_002462451.1), *Tortula ruralis* (AY034889) and *Zea mays* (KJ004509).

aldehydes, ω -aminoaldehydes and AASAL at 1 mM concentration (Figure 3A). ALDH2C isoforms display a wide substrate preference for aliphatic and aromatic aldehydes, especially for hexanal, benzaldehyde and *p*-anisaldehyde. Other naturally occurring aldehydes, including AASAL, APAL, ABAL, citral and glycolaldehyde were weak substrates (all below the 10% rate compared with hexanal). Based on the catalytic efficiency values (k_{cat}/K_m ratios; Table 3), hexanal and cinnamaldehyde are

the best substrates for RF2C, whereas RF2D and RF2E prefer *p*-anisaldehyde and benzaldehyde and RF2F preferentially oxidizes phenylacetaldehyde. Saturation curves for the RF2C, RF2E and RF2F show that a strong excess substrate inhibition appears at low concentrations for some aromatic aldehydes, such as for cinnamaldehyde with RF2C or phenylacetaldehyde with RF2F (Figure 3B).

The specific activity of all ZmALDH2 isoforms increases with the chain length of aliphatic aldehydes up to six carbons

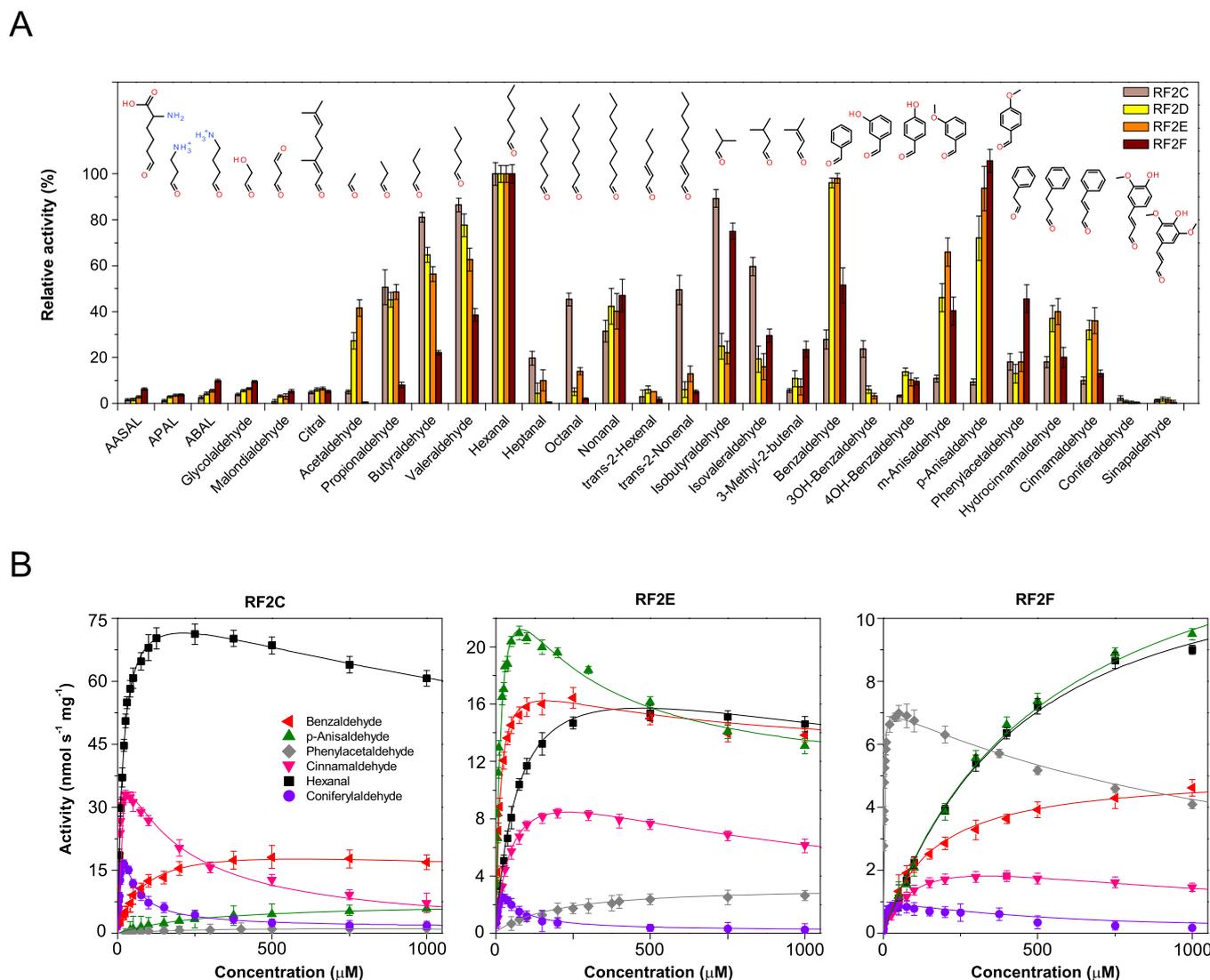


Figure 3 Screening of substrate specificity of maize ALDH2C isoforms

(A) Substrate specificity of four maize ALDH2 isoforms with aliphatic and aromatic aldehydes. Measurements were performed with 1 mM substrate in 0.15 M Tris/HCl buffer, pH 8.0 (for RF2C and RF2F) and pH 7.5 (for RF2D and RF2E), containing 1 mM NAD^+ . Specific activity values with 1 mM hexanal were 62 nkat mg^{-1} for RF2C, $12.2 \text{ nkat mg}^{-1}$ for RF2D, $14.6 \text{ nkat mg}^{-1}$ for RF2E and 8.9 nkat mg^{-1} for RF2F. The activity with hexanal was arbitrarily taken as 100%. Error bars stand for S.D. from four measurements. (B) Saturation curves for activity determination of maize RF2C, RF2E and RF2F. The data were measured as above and are shown for benzaldehyde, *p*-anisaldehyde, phenylacetaldehyde, cinnamaldehyde, hexanal and coniferylaldehyde.

(hexanal). The activity with heptanal and octanal is much lower (except for RF2C and octanal) and increases again with nonanal (nine-carbon chain) up to 50% activity compared with hexanal. The unsaturated C6 and C9 aldehydes, *t*-2-hexenal and *t*-2-nonenal respectively, are weaker substrates than the saturated ones. However, for RF2C isoform, *t*-2-nonenal is among the best substrates. Aliphatic C6 and C9 aldehydes and their unsaturated forms arise from linoleic or linolenic acids via the lipoxygenase pathway [47]. They belong to green leaf volatiles, which are synthesized in response to wounding and also provide fruits and vegetables with their aroma. Branched aliphatic C4 and C5 aldehydes are also good substrates. All isoforms display activity with 3-methyl-2-butenal (isopentenol), which is 8%–25% of that compared with hexanal. Isopentenol is formed by the oxidation of isoprenoid cytokinins by cytokinin oxidase/dehydrogenase [48].

The activity of ZmALDH2 isoforms toward aromatic aldehydes such as benzaldehyde, *m*- and *p*-anisaldehyde,

phenylacetaldehyde, cinnamaldehyde and hydrocinnamaldehyde is variable. However, they are all good substrates. In plants, benzaldehyde originates either from phenylalanine via the non- β -oxidative pathway [49] or from the oxidation of aromatic cytokinins by cytokinin oxidase/dehydrogenase. Phenylacetaldehyde also originates from phenylalanine via several pathways [47]. ZmALDH2 isoforms display an activity with coniferylaldehyde and sinapaldehyde (Table 3), which arise from the phenylpropanoid pathway and have been demonstrated as *in vivo* substrates for ALDH2C from *Arabidopsis* and oilseed rape [15,16].

Substrate specificity and the role of maize and pea ALDH7

A large screening study performed using 1 mM substrates proves that the three ALDH7 isoforms from maize, pea and human

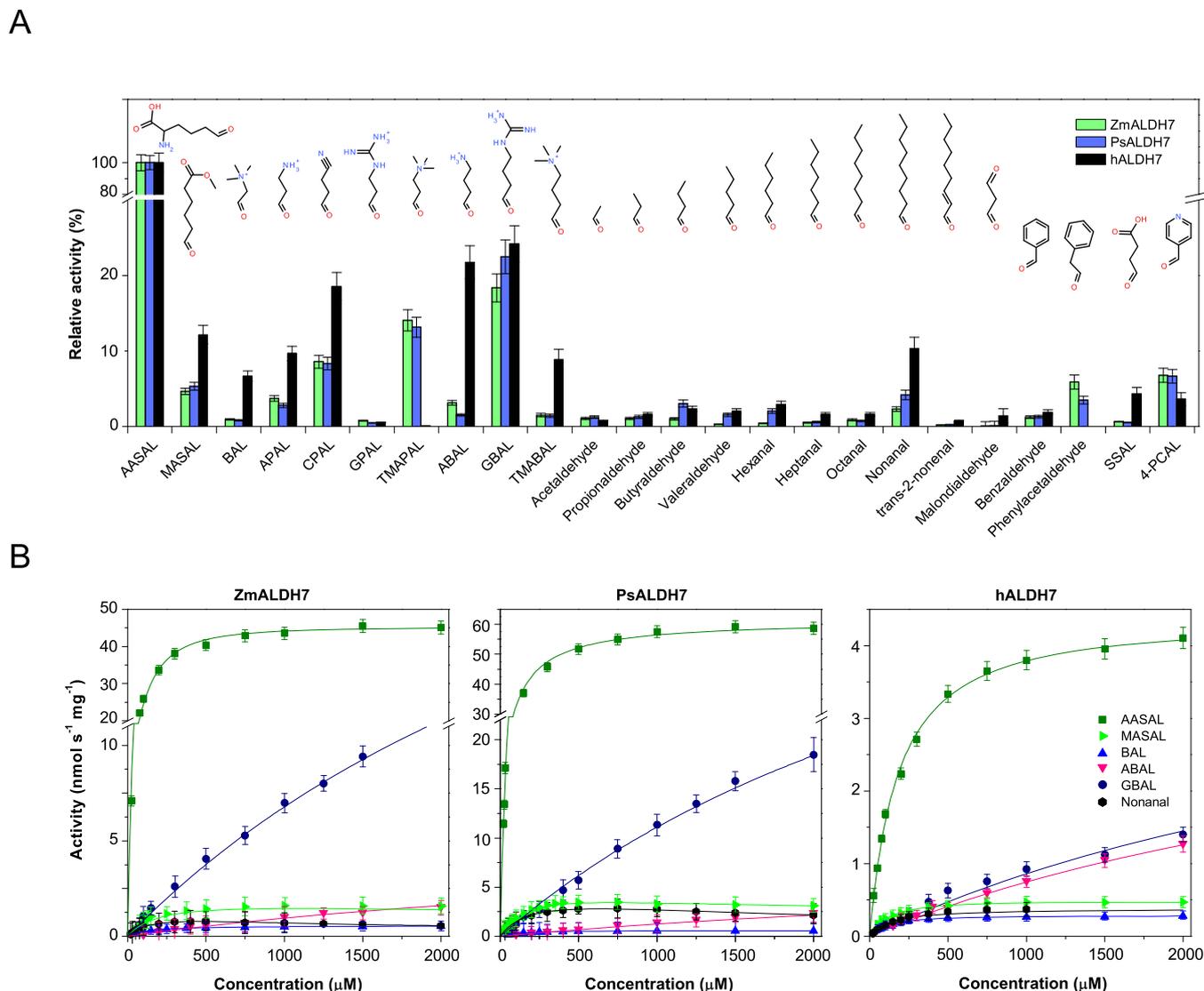


Figure 4 Screening of substrate specificity of ALDH7 isoforms

(A) Substrate specificity of maize, pea and human ALDH7. The measurements were performed with 1 mM substrates in 0.1 M sodium pyrophosphate buffer, pH 8.0, containing 2.5 mM NAD⁺. The activity with AASAL was arbitrarily taken as 100%. Error bars stand for S.D. Specific activity values measured with 1.0 mM AASAL as a substrate were 43.5, 57.5 and 3.8 nkat mg⁻¹ for ZmALDH7, PsALDH7 and hALDH7 respectively. (B) Saturation curves for activity determination of ZmALDH7, PsALDH7 and hALDH7. The data were measured under the same conditions as above and are shown for selected substrates including α -AASAL, MASAL, ABAL, BAL, GBAL and nonanal.

preferentially oxidize AASAL (Figure 4A). The K_m values of 90 μ M for ZmALDH7 and PsALDH7 and 170 μ M for hALDH7 are in agreement with those previously found for seabream ALDH7 (67 μ M) and hALDH7 (169 μ M) [17,39]. MASAL is a much weaker substrate. Interestingly, both plant enzymes have narrower substrate specificity and higher specific activities compared with hALDH7 (45 compared with 4 nkat mg⁻¹). Indeed, natural ω -aminoaldehydes, such as BAL, APAL, ABAL, TMABAL and GBAL, are oxidized by hALDH7 with 10%–25% rates relative to that of AASAL, whereas only GBAL is similarly oxidized by plant enzymes. Aliphatic and aromatic aldehydes, as well as succinic semialdehyde (SSAL), are rather weak substrates. AASAL is clearly the best substrate for all ALDH7 isoforms as shown by the catalytic efficiency values (Table 4; Figure 4B).

Our kinetic analysis shows a very weak activity with lipid peroxidation products *t*-2-hexenal, *t*-2-nonenal and malondialdehyde (Table 4; Figure 4B). Decreased malondialdehyde

levels have been observed in soybean and *Arabidopsis* ALDH7 overexpressors upon a stress treatment [24,25] suggesting a decreased oxidative stress *in vivo*. In line with these observations, *aldh7* mutant rice seeds accumulated more malondialdehyde and a melanoidin pigment in the endosperm during seed storage and desiccation [26]. Later it was shown that the dark yellow pigment corresponds to oryzamutaic acids, which are products of AASAL polymerization [50–52]. Plant ADLH7 from rice and *Arabidopsis* are confined to the cytosol [25,26], in agreement with the cytosolic localization of a bifunctional lysine-ketoglutarate reductase/saccharopine dehydrogenase (LKR/SDH) [53,54], which catalyses the first reaction of the lysine catabolism pathway (saccharopine pathway) leading to AASAL and glutamate. In contrast, hALDH7 was found in the mitochondria, cytoplasm and nucleus [17,55] in line with the existence of two lysine catabolism pathways in humans: the saccharopine pathway localized in mitochondria and the pipecolate pathway predominantly

Table 4 Kinetic parameters of maize and pea ALDH7 for selected substrates

All K_m and k_{cat} values are given in μM and s^{-1} respectively. Kinetic constants were measured in 0.1 M sodium pyrophosphate buffer, pH 8.0, using saturating 2.5 mM NAD^+ concentration. The K_m value for NAD^+ and NADP^+ was measured at a fixed 1.5 mM concentration of AASAL. The symbol 'n.d.' stands for not determined.

Ligand	ZmALDH7			PsALDH7			hALDH7		
	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\text{M}^{-1}\text{s}^{-1}$)	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\text{M}^{-1}\text{s}^{-1}$)	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\text{M}^{-1}\text{s}^{-1}$)
NAD ⁺	358 ± 24	2.68 ± 0.030	7.5 × 10 ³	393 ± 21	3.46 ± 0.046	8.8 × 10 ³	224 ± 28	0.256 ± 0.011	1.2 × 10 ³
NADP ⁺	14017 ± 758	0.27 ± 0.008	2.0 × 10 ¹	12834 ± 988	0.17 ± 0.009	1.3 × 10 ¹	10645 ± 623	0.233 ± 0.009	2.2 × 10 ¹
AASAL	97 ± 7	2.56 ± 0.043	2.6 × 10 ⁴	89 ± 3	3.32 ± 0.027	3.7 × 10 ⁴	170 ± 9	0.242 ± 0.003	1.4 × 10 ³
MASAL	126 ± 21	0.09 ± 0.004	7.1 × 10 ²	147 ± 17	0.26 ± 0.009	1.8 × 10 ³	127 ± 8	0.030 ± 0.001	2.4 × 10 ²
GBAL	2767 ± 146	1.43 ± 0.029	5.2 × 10 ²	2551 ± 159	2.22 ± 0.071	8.7 × 10 ²	3103 ± 119	0.218 ± 0.004	7.0 × 10 ¹
BAL	120 ± 3	0.03 ± 0.001	2.5 × 10 ²	111 ± 7	0.03 ± 0.001	2.9 × 10 ²	105 ± 5	0.016 ± 0.001	1.6 × 10 ²
ABAL	4371 ± 116	0.27 ± 0.005	6.3 × 10 ¹	5167 ± 162	0.46 ± 0.007	9.0 × 10 ¹	4339 ± 331	0.220 ± 0.016	5.0 × 10 ¹
APAL	1280 ± 59	0.15 ± 0.002	1.1 × 10 ²	1698 ± 81	0.34 ± 0.007	2.0 × 10 ²	1007 ± 81	0.038 ± 0.001	3.7 × 10 ¹
Hexanal	1805 ± 186	0.11 ± 0.005	6.1 × 10 ¹	1672 ± 61	0.14 ± 0.002	8.3 × 10 ¹	1477 ± 123	0.013 ± 0.002	8.5
<i>l</i> -2-Hexenal	1652 ± 96	0.09 ± 0.002	5.4 × 10 ¹	1580 ± 57	0.10 ± 0.002	6.6 × 10 ¹	1145 ± 55	0.010 ± 0.002	9.0
Nonanal	174 ± 14	0.07 ± 0.006	4.3 × 10 ²	198 ± 15	0.25 ± 0.010	1.2 × 10 ³	257 ± 12	0.030 ± 0.001	1.2 × 10 ²
<i>l</i> -2-Nonenal	953 ± 89	0.008 ± 0.001	8.4	974 ± 74	0.009 ± 0.001	9.5	993 ± 103	0.003 ± 0.001	3.5
TMAPAL	4820 ± 334	1.54 ± 0.096	3.2 × 10 ²	4656 ± 260	1.57 ± 0.049	3.4 × 10 ²	n.d.	n.d.	n.d.
3-PCAL	437 ± 46	0.14 ± 0.005	3.2 × 10 ²	231 ± 13	0.29 ± 0.004	1.3 × 10 ³	n.d.	n.d.	n.d.
4-PCAL	537 ± 51	0.22 ± 0.007	4.1 × 10 ²	321 ± 10	0.27 ± 0.005	8.3 × 10 ²	n.d.	n.d.	n.d.

peroxisomal/cytosolic [56]. A mutation in *hALDH7* gene results in the accumulation of P6C (cyclic form of AASAL) inactivating pyridoxal 5'-phosphate (PLP) and leading to acute depletion of vitamin B6 [18].

A recent study on hALDH7 pointed out the importance of BAL activity for possible protection from hyperosmotic stress [17]. The oxidation of BAL, which originates from choline, leads to glycine betaine, a well-known osmolyte. In the present study, we demonstrate kinetically for the first time that the plant ALDH7 family possesses a significant AMADH activity as the enzymes oxidize GBAL, APAL and ABAL providing directly (or after a further conversion) the cytosolic osmoprotectants 4-guanidinobutyrate, β -alanine and γ -aminobutyric acid (GABA) respectively. GBAL is an intermediate in arginine catabolism and the released 4-guanidinobutyrate may be further hydrolysed by ureohydrolase to urea and GABA.

Crystal structures of ZmALDH2 and ZmALDH7

To understand the substrate specificity differences within and between ALDH2 and ALDH7 family members, we solved the crystal structures of RF2C (ZmALDH2-3), RF2F (ZmALDH2-6) and ZmALDH7 (Table 2). The asymmetric unit of RF2F crystal contains a tetramer, whereas those of RF2C and ZmALDH7 contain a dimer and two monomers, respectively. However, using the crystallographic symmetry, both RF2C and ZmALDH7 form a tetramer which is the active form in solution in accordance with results from gel permeation chromatography (molecular mass values of 214, 226 and 181 kDa for RF2C, RF2F and ZmALDH7 respectively). Each subunit within the tetramer adopts the characteristic ALDH fold consisting of a coenzyme-binding domain, a catalytic domain and an oligomerization domain. The entrances of the substrate channel and the coenzyme-binding site are located on opposite sides of the monomer. A conserved sodium ion, probably having a structural role, binds in the cavity close to the coenzyme-binding site. Monomers of RF2C and RF2F are very similar to each other (average RMSD of 0.7 Å) and a structural comparison using PDBeFold (<http://www.ebi.ac.uk/msd-srv/ssm/>) shows that they

both resemble the mitochondrial hALDH2 with the lowest RMSD of 0.80–0.85 Å for PDB codes 1O02, 1NZZ, 1NZX and others [57] (Supplementary Figure S1). Both RF2C and RF2F tetramers are very similar to hALDH2 tetramer with average RMSD values of 1.2 and 1.0 Å for 1940 C α atoms respectively. Subtle differences in the oligomerization arrangement can be observed.

A comparison of substrate channel residues in RF2C and RF2F however indicates significant differences as shown in Figures 5A and 5B. The substrate-binding site of RF2C is narrow and formed by an aromatic cluster composed of Tyr¹⁷¹, Phe¹⁷⁸, Phe⁴⁵⁸, Phe⁴⁶⁰ and Phe⁴⁶⁶ and also by several mainly nonpolar residues including Ile³⁰², Val³⁰⁴, Thr¹⁷⁴, Met¹⁷⁵ and Thr²⁹⁷. In contrast, that of RF2F is much wider because of the presence of Val¹⁹² and Met⁴⁷⁷ at the position of the two aromatic residues Phe¹⁷⁸ and Phe⁴⁶⁰ in RF2C. Indeed, Met⁴⁷⁷ side chain, which is shifted outwards by 3 Å, opens the cavity. Two more residues Ala³¹⁶ and Glu⁴⁹⁵ (Ile³⁰² and Met⁴⁷⁸ in RF2C) also contribute to the substrate channel broadening. The cavity width of RF2F clearly correlates with high K_m values for various substrates most probably due to weaker nonpolar interactions except for phenylacetaldehyde. RF2D and RF2E, which share 95 % sequence identity, do not differ in active site residues and thus have similar kinetic properties. A model of RF2E was made using RF2C structure as a template (Figure 5C). The substrate-binding site of RF2E contains Phe⁴⁵⁹ (Phe⁴⁶⁰ in RF2C) making the cavity as narrow as in RF2C. Nevertheless, the cavity is slightly wider in the opposite direction because of the presence of Leu¹⁷⁷ at the position of Phe¹⁷⁸ in RF2C. Finally, only RF2C contains a threonine at position 297, whereas the other three maize isoforms contain a phenylalanine at the equivalent position. This threonine, which occupies less space, might have a role in substrate-binding plasticity and specificity by allowing the side chain of Tyr¹⁷¹ (equivalent to a phenylalanine in other maize isoforms) to move and facilitate substrate accommodation.

A sequence alignment of substrate channel residues in ALDH2C members (Figure 5D) reveals a high variability that allows various isoforms within one species oxidizing several different substrates from multiple pathways. Recently, maize *RF2C* gene was found to be induced in elongating internodes containing cells which exhibit primary cell wall biosynthesis, compared with non-elongating internodes with cells depositing

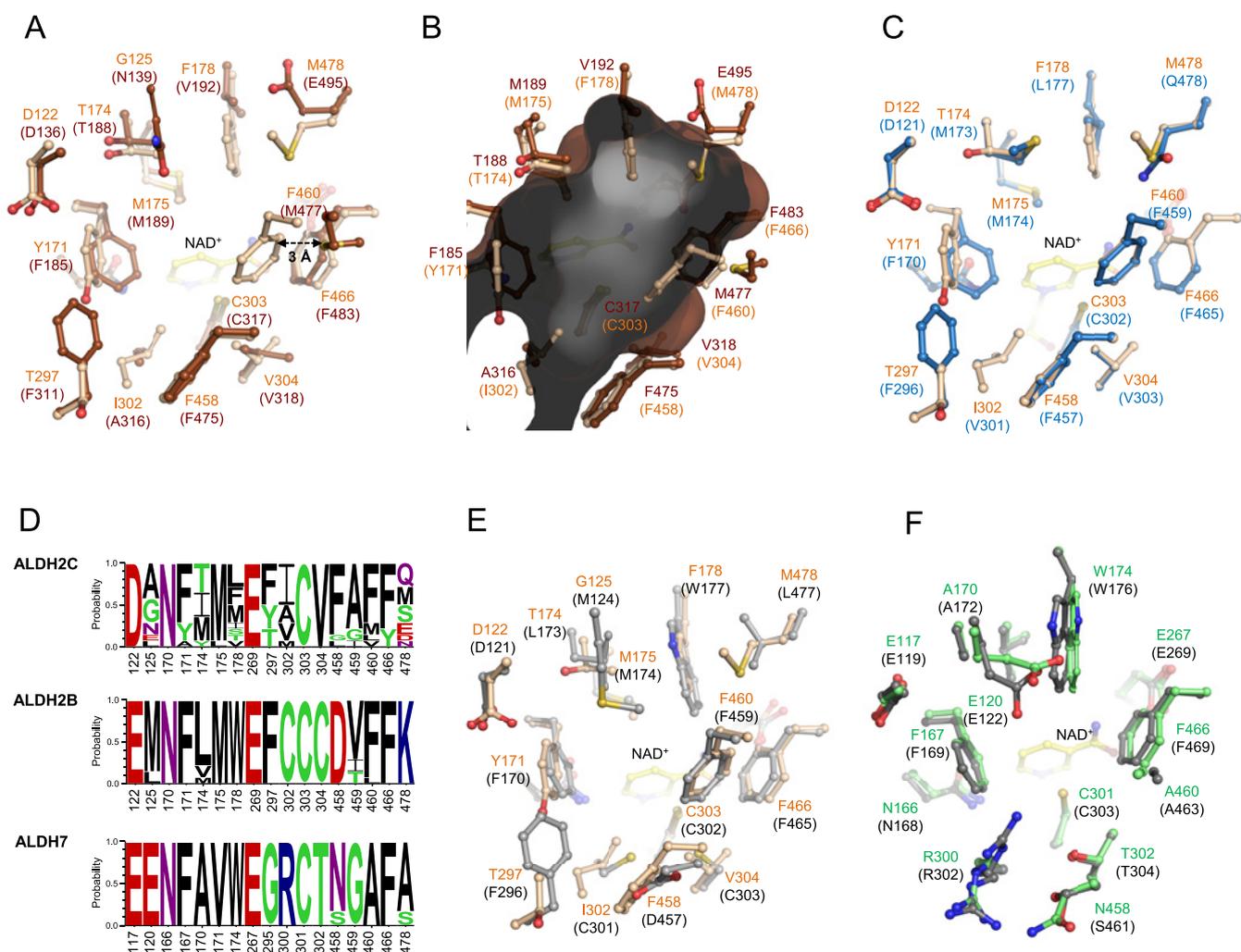


Figure 5 Composition of maize ALDH2 and ALDH7 active sites

(A) A view from the top on substrate channels of maize RF2C (light brown, present work) and RF2F (dark brown, present work). Residues of RF2C labelled in orange and those of RF2F are shown in brackets. (B) A transversal section of the substrate channel of RF2F. The inner surface is grey coloured (foreground is dark grey, background is light grey) and was calculated using Hollow with a 0.7 Å grid spacing and 1.2 Å interior probe. Residues are coloured in dark brown and labelled. Residues of RF2C (in light brown) are shown for comparison to indicate differences affecting diameter of the substrate channel. (C) Composition of the substrate channel in RF2E (blue). The model was made using SWISS-MODEL (<http://swissmodel.expasy.org>) and RF2C as a template. (D) An overview of conservation of amino acid residues forming the substrate channel in ALDH2 and ALDH7 families. Residues are numbered according to maize RF2C for ALDH2C and ALDH2B subfamilies and maize ALDH7 for ALDH7 family. Sequence logos were made using WebLogo (<http://weblogo.threeplusone.com>). (E) Superposition of the substrate channels of maize RF2C (light brown) and hALDH2 (grey, PDB code 1NZX) used as a reference. Residues of hALDH2 are shown in brackets. (F) Superposition of the substrate channels of ZmAMADH7 (green, present work) and hALDH7 (grey, PDB code 2J6L) used as a reference. Residues of hALDH7 are shown in brackets.

secondary cell wall material [58]. In line with our kinetic results, all four maize cALDH2 isoforms seem to be involved in the oxidation of coniferaldehyde and sinapaldehyde and formation of ferulic and sinapinic acids and their cell wall-linked esters similarly to ALDH2C in *Arabidopsis* and oilseed rape [15,16]. However, in the phenylpropanoid pathway, there are several other aldehydes involved including for example cinnamaldehyde, *p*-coumarylaldehyde, caffeylaldehyde or 5-hydroxyconiferaldehyde. Because the last three hydroxy derivatives are not available, they could not be analysed although they probably belong to *in vivo* substrates.

In contrast, plant mtALDH2 isoforms including RF2A and RF2B share a conserved active site and are highly similar to hALDH2 (Figure 5D). Their sequences contain a triple cysteine motif including the catalytic cysteine. Although *in vivo* the preferred substrate of hALDH2 is acetaldehyde, the enzyme has

been shown to oxidize various aliphatic and aromatic aldehydes, especially hexanal, phenylacetaldehyde or benzaldehyde [4]. Likewise, RF2A, but not RF2B, is in addition to acetaldehyde also highly active with benzaldehyde and cinnamaldehyde [9]. MtALDH2 from snapdragon was found to have the highest catalytic efficiency with acetaldehyde but *in vivo* it has been linked to oxidation of benzaldehyde to benzoate in flowers leading to emission of high levels of methylbenzoate and benzylbenzoate in floral scent [14].

ZmALDH7 monomer resembles that of humans and seabream (PDB codes 2J6L and 2JG7) with RMSDs of 0.72 and 0.74 Å respectively (Supplementary Figure S1). The major difference between ZmADLH7 and ZmALDH2 monomers (RMSD of 1.3–1.4 Å for 465 C α atoms) comes from a slight shift in the oligomerization domain and the presence of an N-terminal helix in ZmALDH7 (Supplementary Figure S1). The tetramer

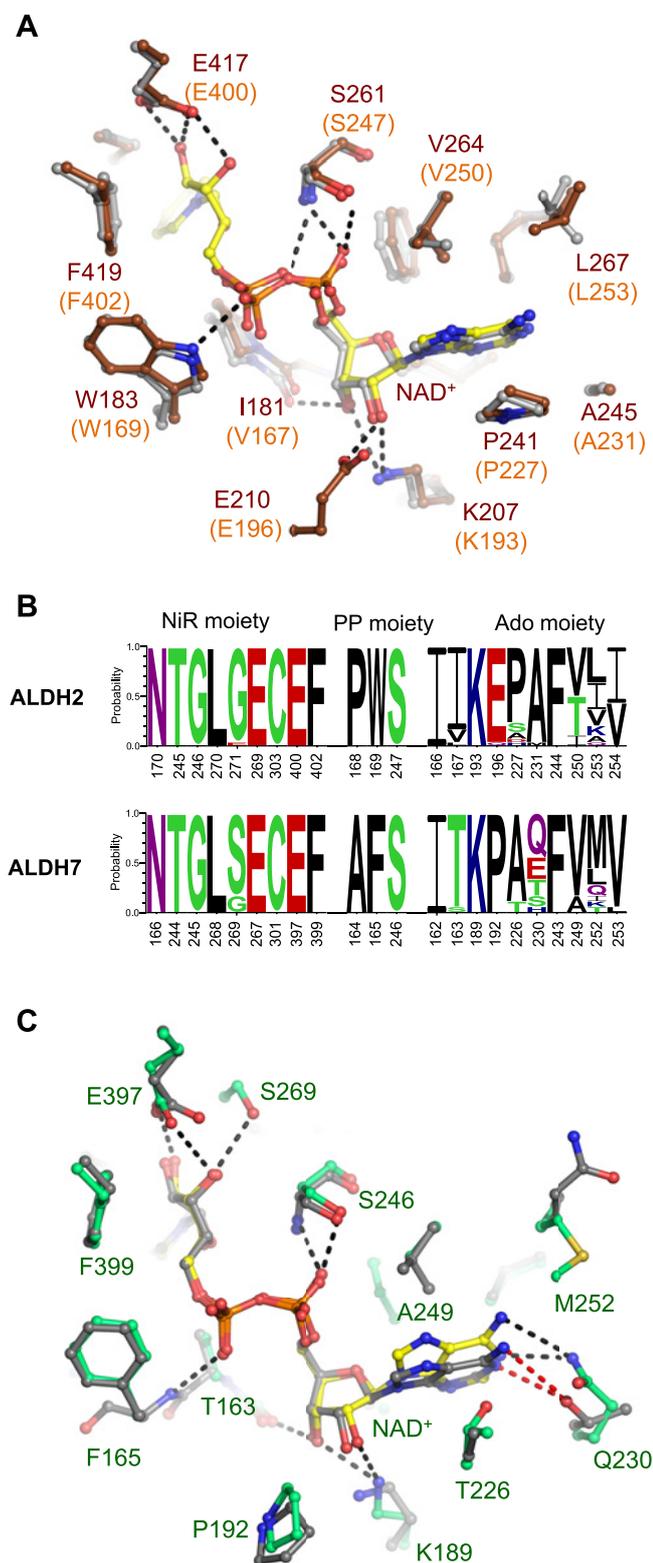


Figure 6 NAD⁺ binding in ALDH2 and ALDH7 family

(A) Superposition of NAD⁺-binding sites of maize RF2F (ZmALDH2-6, dark brown colour, present work) and hALDH2 (grey colour). Residues are labelled and numbered according to RF2F sequence; those of RF2C are shown in orange and brackets. (B) An overview of conservation of amino acid residues relevant for NAD⁺ binding in ALDH2 and ALDH7 families. Residues are split into three clusters for NiR, diphosphate and Ado moiety of NAD⁺ molecule and are numbered according to maize RF2C and ALDH7 sequence respectively. Sequence logos were

of ZmALDH7 slightly differs from both ZmALDH2 tetramers with an average RMSD value of 2.4 Å. Their active site residues are almost identical indicating that these enzymes have the same cellular role in plants, mammals or fish using AASAL as the preferred substrate confirmed by our kinetic data. The three residues Glu¹²⁰, Arg³⁰⁰ and Thr³⁰² located in the substrate-binding site of ZmALDH7 are conserved and specific to the ALDH7 family (Figure 5F). The role of Glu¹²⁰ and Arg³⁰⁰ has been validated by mutagenesis studies on seabream ALDH7 and hALDH7 [39,59] suggesting that Glu¹²⁰ binds the α-amino group of AASAL, whereas Arg³⁰⁰ binds its carboxylate group. The role of Thr³⁰² has not been investigated by mutagenesis yet. However, visual inspection of ZmALDH7 active site together with manual modelling shows that this residue may interact with Arg³⁰⁰ in the binding of AASAL carboxylate side chain. Our kinetic analysis shows that AASAL and MASAL have similar *K_m* values but the *k_{cat}* value for MASAL is about 2-fold lower indicating that protonated amino group of the substrate (and interaction with Glu¹²⁰) is needed to achieve a high reaction rate. High *K_m* values for aminoaldehydes GBAL (~2.5 mM) and ABAL (~4.5 mM) indicate that the carboxy group (and interaction with Arg³⁰⁰ and Thr³⁰²) is necessary to maintain low *K_m* values.

Coenzyme-binding sites in ZmALDH2 and ZmALDH7

In all ZmALDH2 monomers, a NAD⁺ molecule is well defined in the electron density maps (Supplementary Figure S1). In RF2F structure, the coenzyme adopts an extended conformation typically observed for oxidized NAD⁺ and the catalytic cysteine Cys³¹⁷ is oriented towards the substrate channel (attacking conformation). In RF2C structure, it adopts a contracted conformation most probably due to the high concentration of calcium ions present in the crystallization conditions. A calcium ion binds to the pyrophosphate oxygen atoms O1N and O2A of the NAD⁺ molecule in subunit A. An equivalent magnesium ion bound to the contracted coenzyme was previously observed in the hALDH2 structure (PDB code 1NZZ) [57] and it was shown that the presence of magnesium ions may select particular conformations of the coenzyme. The catalytic cysteine Cys³⁰³ is again in the attacking conformation and in subunit A it is oxidized into sulfenic acid [S-(hydroxy)cysteine].

The adenine moiety of NAD⁺ lies in a hydrophobic pocket flanked by two helices with no polar contacts. In contrast, the ribose, both α and β-phosphates and the nicotinamide riboside (NiR) moiety of NAD⁺ make several protein interactions identical to those observed for NAD⁺ binding in hALDH2 (Figure 6A) as well as in plant ALDH10. Therefore, RF2C, RF2D, RF2E, RF2F, hALDH2 and ALDH10 display similar *K_m* values for NAD⁺ (~10⁻⁵ μM) [27,29,60]. These enzymes are NAD⁺ specific because they possess a conserved glutamate (Glu¹⁹⁶ in RF2C, Glu²¹⁰ in RF2F) preventing the binding of the 2'-phosphate group of NADP⁺. Indeed, *K_m* values of 3.0, 1.5, 1.4 and 1.7 mM for NADP⁺ with RF2C, RF2D, RF2E and RF2F respectively are ~20–70 times higher than those for NAD⁺ and the reaction rates appear between 0.5%–9% of those with NAD⁺. Catalytic efficiency values for NADP⁺ are about 2-fold lower (with RF2C more than 3-fold lower) than those for NAD⁺. Thus, NADP⁺

made using WebLogo (<http://weblogo.threeplusone.com>). (C) Superposition of NAD⁺-binding sites of maize ALDH7 (green colour, present work) and hALDH7 (grey colour). Residues in the vicinity of NAD⁺ are labelled and numbered according to ZmALDH7 sequence. Those in human enzymes are grey coloured as a reference. NAD⁺ molecules adopting the extended conformation are shown in yellow and atom-coded colour sticks.

does not function as an effective coenzyme in the plant ALDH2 family. A recent mutagenesis study on this conserved glutamate residue confirmed its importance for NAD⁺ specificity in the plant ALDH3 family [61]. Figure 6B presents an overview of residues forming the NAD⁺-binding site and their frequency. Notably, the residues involved in adenosine moiety (Ado) binding of NAD⁺ are highly variable.

Although the coenzyme-binding site in ZmALDH7 slightly differs in residue composition from that of ALDH2 family and of hALDH7, a NAD⁺ molecule present in each ZmALDH7 monomer adopts the extended conformation with the catalytic Cys³⁰¹ in the attacking conformation and most of the important protein interactions conserved. Indeed, Phe¹⁶⁵ in ZmALDH7 equivalent to Phe¹⁶⁷ in hALDH7 is bound to the β -phosphate oxygen atom via its main chain NH atom restoring the direct interaction between the side chain of the tryptophan residue and the β -phosphate in both ALDH2 and ALDH10 families (Trp¹⁶⁹ and Trp¹⁸³ in RF2C and RF2F; Figure 6C). Ser²⁶⁹ in ZmALDH7 (glycine in ZmALDH2) makes an additional interaction with the O3D atom of the NiR moiety whereas hALDH7 and ALDH2 enzymes possess a glycine at the equivalent position. The conserved glutamate in ALDH2 (Glu¹⁹⁶ in RF2C) and ALDH10, which prevents binding of the 2'-phosphate group of NADP⁺, is replaced by a proline in ALDH7 family (Pro¹⁹² in ZmALDH7). This proline also makes a steric hindrance to 2'-phosphate of NADP⁺ as reflected in the K_m value of 14 mM (40 times higher than that for NAD⁺) and the reaction rate, which is only 0.5% of that measured with NAD⁺ and ZmALDH7. The major difference between ZmALDH7 and hALDH7 or ALDH2 concerns the adenine position of NAD⁺. Indeed, the adenine adopts a different position in ZmALDH7 because of the presence of Thr²²⁶ side chain (alanine in hALDH7 and proline in ALDH2), which pushes the adenine ring up to 1 Å towards Ala²⁴⁹ (Figure 6C). The presence of the couple Thr²²⁶-Ala²⁴⁹ appears only in a small subgroup of monocots including maize, whereas most ALDH7 family members carry a conserved alanine-valine couple (Ala-Val²⁵⁰ in hALDH7) at the corresponding positions. In ALDH7 family, adenine makes one polar interaction between its N1 atom and a glutamate, glutamine, serine, threonine or histidine. All the above differences contribute by one order of magnitude to the higher K_m values for NAD⁺ of ZmALDH7 and PsALDH7 compared with those of ALDH2 and ALDH10 family members. Consequently, a higher saturating concentration of NAD⁺ was used for kinetic measurements.

Expression of ALDH2 and ALDH7 genes in maize

First we analysed the expression pattern of ALDH2 and ALDH7 genes in developing maize seedlings (stem, roots and leaves) during the first two weeks after germination (Figure 7). RF2C is highly expressed compared with the five other ALDH2 genes with a strong and comparable transcript levels in leaves and roots but lower in stem. RF2A was preferentially expressed in leaves, whereas RF2B was preferentially expressed in roots and less in leaves. Transcripts of RF2D and RF2E were the most abundant in roots, but almost absent from the stem, whereas RF2F transcripts appeared especially in the stem of 2-week-old seedlings. The highest levels of ZmALDH7 transcripts were found in leaves.

The pattern of gene expression in older plants, in tassels, silks and kernels was also analysed. Again, RF2C is the most expressed ALDH2 gene in maize. RF2B, RF2C and ZmALDH7 are highly abundant in tassels at 5 days before pollination (DBP) and 0 days after pollination (DAP) suggesting their important role in pollen development. In silks, gene expression was lower compared with that in tassels. Relatively high levels of RF2A,

RF2B, RF2C and RF2F transcripts appear in silks at 6 DAP. In kernel, we detected a strong expression of RF2B and RF2C followed by RF2D and RF2F. RF2A expression was highest in silks at 6 DAP and kernels at 20 DAP, whereas those of RF2B, RF2C and ZmALDH7 were strongest in tassels. The RF2D transcripts accumulated in mature leaves, whereas the RF2E gene was significantly expressed only in the embryo. The RF2F gene was strongly expressed in the stem and main root. The strong expression of ZmALDH7 in kernel, tassels and silks correlates well with a strong expression of LKR/SDH gene observed in maize endosperm [53] and in floral organs of *Arabidopsis* [62] suggesting an active saccharopine pathway in developing seeds and in reproductive organs containing actively dividing cells.

Several expression studies on the plant ALDH gene superfamily, including rice and foxtail millet, show that many ALDH2 genes are induced by abiotic stressors, but some may also be down-regulated [22,23]. Indeed, OsALDH2B5 gene was found up-regulated by submergence, drought and salinity stresses, whereas the OsALDH2C2 gene is down-regulated by drought and salinity stresses [11,22]. Interestingly, a gene coding for mtALDH2 in Chinese wild grapevine *Vitis pseudoreticulata* (VpALDH2B4) [63] is induced by pathogens. Transgenic 35S:VpALDH2B4 plants are more resistant to a downy mildew, powdery mildew and salt stress. This all indicates that in addition to *in vivo* acetaldehyde, benzaldehyde and coniferaldehyde/sinapaldehyde dehydrogenase activities, plant ALDH2 members might be involved in the detoxification of aldehydes produced upon stress like *t*-2-hexenal and *t*-2-nonenal. Indeed, hexenal and *t*-2-nonenal, lipid peroxidation products [64], are among the best substrates for RF2C and when comparing k_{cat}/K_m values, ALDH2C isoforms show much higher catalytic efficiencies with these aldehydes than ALDH7 (Tables 3 and 4).

Phenolamides (hydroxycinnamic acid amides), known to contribute to cell-wall cross-linking, are composed of coumaric, caffeic and ferulic acids fused to either aryl monoamines (tyramine, tryptamine or others) or polyamines (putrescine and spermidine) [65,66]. They accumulate in reproductive organs of many plants, as well as during stress, in order to function as inducible defence against insect herbivory [67]. The accumulation is mainly associated with flower initiation and development and di- and tri-substituted hydroxycinnamoyl conjugates are found in pollen grains in anthers. Male sterility was previously linked to the absence of phenolamides in maize anthers [68]. In *Arabidopsis*, several genes involved in phenolamide biosynthesis are highly expressed in tapetal cells of anthers and major metabolites in pollen coat were identified as N^1, N^5, N^{10} -trihydroxyferuloyl spermidine and N^1, N^5 -di(hydroxyferuloyl)- N^{10} -sinapoyl spermidine [69–71]. The RF2A gene, originally found to restore fertility to Texas cytoplasmic male sterility maize lines, is highly abundant in tapetal cells and important for the anther development [7,8]. We may hypothesize that fertility restoration linked to RF2A is because of the changes in phenolamide biosynthesis. As other cALDH2 isoforms also oxidize coniferaldehyde and sinapaldehyde, they might contribute to the biosynthesis of phenolamides in various plant organs as well as during stress conditions.

CONCLUSIONS

In the present work, we studied maize and pea ALDH7 and compared their kinetics with that of human ALDH7. Based on our kinetic and structural results, we can conclude now that a high conservation of the ALDH7 gene during evolution is linked to the substrate preference of ALDH7 for AASAL in accordance

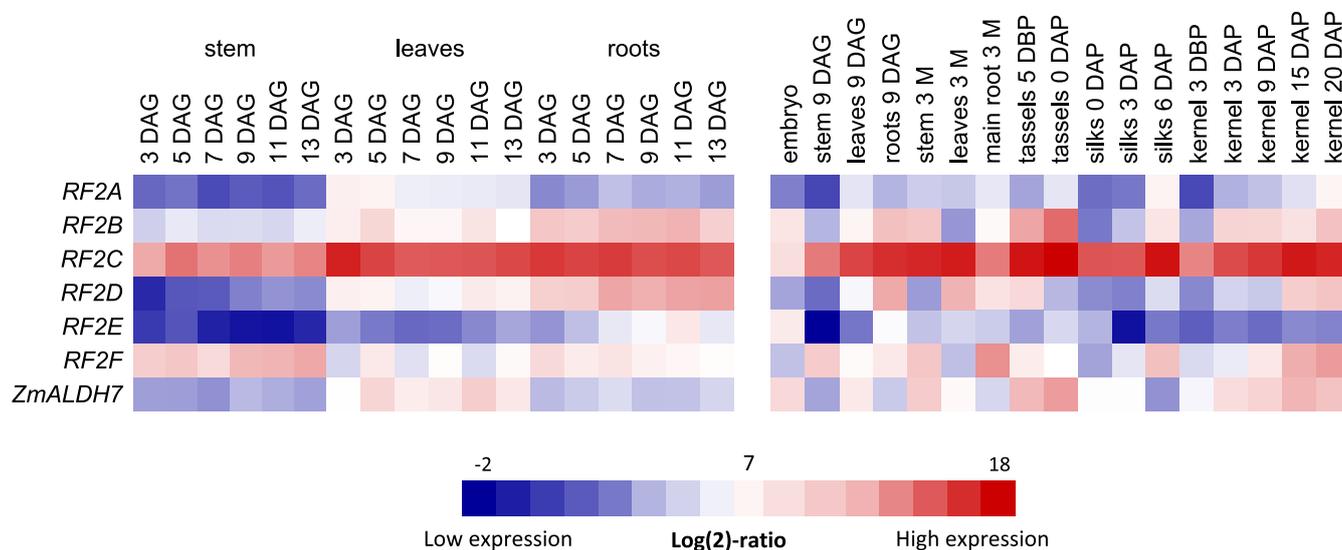


Figure 7 Expression profiles of *ZmALDH2* and *ZmALDH7* genes

Gene expression was studied in stem, leaves and roots during 2 weeks after germination (left side) and in various organs during the lifespan of maize plants (right side). The heat map illustrates transcript levels detected in 1 ng of total RNA. All values are expressed as \log_2 -ratios and are colour-coded from blue to red (the lowest and highest detected transcript numbers, respectively). DAG, days after germination; M, months.

with conservation of the active site of ADLH7 across species. In contrast, maize ALDH2C isoforms show high preferences for benzenoid and phenylpropanoid aldehydes arising via the non- β -oxidative pathway from phenylalanine or via the phenylpropanoid pathway. The latter pathway is very complex and is a starting point for the production of many secondary metabolites, such as flavonoids, coumarins, monolignols (building blocks of lignin), phenolamides and others. Therefore, the ALDH2C subfamily displays a high variability in composition of the active site containing mainly aromatic and nonpolar residues in accordance with differences in phenylpropanoid pathway across species and different hydroxy and methoxy phenylpropanoid aldehydes arising from this pathway. Finally, ALDH2C isoforms may also contribute to the oxidation of 'lipid peroxidation' aldehydes, as well as to oxidation of aliphatic and aromatic aldehydes released after the oxidation of plant hormones cytokinins.

AUTHOR CONTRIBUTION

David Kopečný, Solange Moréra and Marek Šebela designed the research. Radka Končítiková, Tomáš Andree and David Kopečný analysed enzyme kinetics. Armelle Vigouroux and Solange Moréra performed the crystallographic study and contributed material and tools for crystallography. Radka Končítiková, Martina Kopečná and Jan Bartoš analysed gene expression. Solange Moréra and David Kopečný analysed the crystal structures. David Kopečný and Solange Moréra wrote the manuscript.

ACKNOWLEDGEMENTS

We acknowledge SOLEIL for provision of synchrotron radiation facilities (proposal ID 20130869) in using beamlines Proxima 1 and Proxima 2. The present work benefitted from IMAGIF platform facilities (<http://www.imagif.cnrs.fr>) at the Centre de Recherche de Gif-sur-Yvette (FRC3115, CNRS) for crystallization work. We thank William Shepard for a critical reading of the manuscript.

FUNDING

This work was supported by the Czech Science Foundation [grant number 15-22322S]; the National Program of Sustainability I (the Ministry of Education, Youth and Sports, Czech Republic) [grant number LQ1204]; and the CNRS (to S.M. and A.V.).

REFERENCES

- Yoshida, A., Rzhetsky, A., Hsu, L. C. and Chang, C. (1998) Human aldehyde dehydrogenase gene family. *Eur. J. Biochem.* **251**, 549–557 [CrossRef PubMed](#)
- Vasiliou, V., Bairoch, A., Tipton, K. F. and Nebert, D. W. (1999) Eukaryotic aldehyde dehydrogenase (ALDH) genes: human polymorphism, and recommended nomenclature based on divergent evolution and chromosomal mapping. *Pharmacogenetics* **9**, 421–434 [CrossRef PubMed](#)
- Brocker, C., Vasiliou, M., Carpenter, S., Carpenter, C., Zhang, Y., Wang, X., Kotchoni, S. O., Wood, A. J., Kirch, H. H., Kopečný, D. et al. (2013) Aldehyde dehydrogenase (ALDH) superfamily in plants: gene nomenclature and comparative genomics. *Planta* **237**, 189–210 [CrossRef PubMed](#)
- Klyosov, A. A. (1996) Kinetics and specificity of human liver aldehyde dehydrogenases toward aliphatic, aromatic, and fused polycyclic aldehydes. *Biochemistry* **35**, 4457–4467 [CrossRef PubMed](#)
- Steinmetz, C. G., Xie, P., Weiner, H. and Hurley, T. D. (1997) Structure of mitochondrial aldehyde dehydrogenase: the genetic component of ethanol aversion. *Structure* **5**, 701–711 [CrossRef PubMed](#)
- Ni, L., Zhou, J., Hurley, T. D. and Weiner, H. (1999) Human liver mitochondrial aldehyde dehydrogenase: three-dimensional structure and the restoration of solubility and activity of chimeric forms. *Protein Sci.* **8**, 2784–2790 [CrossRef PubMed](#)
- Cui, X., Wise, R. P. and Schnable, P. S. (1996) The *rf2* nuclear restorer gene of male-sterile T-cytoplasm maize. *Science* **272**, 1334–1336 [CrossRef PubMed](#)
- Liu, F., Cui, X., Horner, H. T., Weiner, H. and Schnable, P. S. (2001) Mitochondrial aldehyde dehydrogenase activity is required for male fertility in maize. *Plant Cell* **13**, 1063–1078 [CrossRef PubMed](#)
- Liu, F. and Schnable, P. S. (2002) Functional specialization of maize mitochondrial aldehyde dehydrogenases. *Plant Physiol.* **130**, 1657–1674 [CrossRef PubMed](#)
- Skibbe, D. S., Liu, F., Wen, T. J., Yandea, M. D., Cui, X., Cao, J., Simmons, C. R. and Schnable, P. S. (2002) Characterization of the aldehyde dehydrogenase gene families of *Zea mays* and *Arabidopsis*. *Plant Mol. Biol.* **48**, 751–764 [CrossRef PubMed](#)
- Nakazono, M., Tsuji, H., Li, Y., Saisho, D., Arimura, S., Tsutsumi, N. and Hirai, A. (2000) Expression of a gene encoding mitochondrial aldehyde dehydrogenase in rice increases under submerged conditions. *Plant Physiol.* **124**, 587–598 [CrossRef PubMed](#)
- op den Camp, R. G. and Kuhlemeier, C. (1997) Aldehyde dehydrogenase in tobacco pollen. *Plant Mol. Biol.* **35**, 355–365 [CrossRef PubMed](#)
- Wei, Y., Lin, M., Oliver, D. J. and Schnable, P. S. (2009) The roles of aldehyde dehydrogenases (ALDHs) in the PDH bypass of *Arabidopsis*. *BMC Biochem.* **10**, 7 [CrossRef PubMed](#)

- 14 Long, M. C., Nagegowda, D. A., Kaminaga, Y., Ho, K. K., Kish, C. M., Schnepf, J., Sherman, D., Weiner, H., Rhodes, D. and Dudareva, N. (2009) Involvement of snapdragon benzaldehyde dehydrogenase in benzoic acid biosynthesis. *Plant J.* **59**, 256–265 [CrossRef PubMed](#)
- 15 Nair, R. B., Bastress, K. L., Ruegger, M. O., Denault, J. W. and Chapple, C. (2004) The *Arabidopsis thaliana* REDUCED EPIDERMAL FLUORESCENCE1 gene encodes an aldehyde dehydrogenase involved in ferulic acid and sinapic acid biosynthesis. *Plant Cell* **16**, 544–554 [CrossRef PubMed](#)
- 16 Mittasch, J., Böttcher, C., Frolov, A., Strack, D. and Milkowski, C. (2013) Reprogramming the phenylpropanoid metabolism in seeds of oilseed rape by suppressing the orthologs of reduced epidermal fluorescence1. *Plant Physiol.* **161**, 1656–1669 [CrossRef PubMed](#)
- 17 Brocker, C., Lassen, N., Estey, T., Pappa, A., Cantore, M., Orlova, V., Chavakis, T., Kavanagh, K. L., Oppermann, U. and Vasiliou, V. (2010) Aldehyde dehydrogenase 7A1 (ALDH7A1) is a novel enzyme involved in cellular defense against hyperosmotic stress. *J. Biol. Chem.* **285**, 18452–18463 [PubMed](#)
- 18 Mills, P. B., Struys, E., Jakobs, C., Plecko, B., Baxter, P., Baumgartner, M., Willemsen, M. A., Omran, H., Tacke, U., Uhlenberg, B., Weschke, B. and Clayton, P. T. (2006) Mutations in antiquitin in individuals with pyridoxine-dependent seizures. *Nat. Med.* **12**, 307–309 [CrossRef PubMed](#)
- 19 Guerrero, F. D., Jones, J. T. and Mullet, J. E. (1990) Turgor-responsive gene transcription and RNA levels increase rapidly when pea shoots are wilted. Sequence and expression of three inducible genes. *Plant Mol. Biol.* **15**, 11–26 [CrossRef PubMed](#)
- 20 Stroeder, V. L., Boothe, J. G. and Good, A. G. (1995) Molecular cloning and expression of a turgor-responsive gene in *Brassica napus*. *Plant Mol. Biol.* **27**, 541–551 [CrossRef PubMed](#)
- 21 Kirch, H. H., Schlingensiepen, S., Kotchoni, S., Sunkar, R. and Bartels, D. (2005) Detailed expression analysis of selected genes of the aldehyde dehydrogenase (ALDH) gene superfamily in *Arabidopsis thaliana*. *Plant Mol. Biol.* **57**, 315–332 [CrossRef PubMed](#)
- 22 Gao, C. and Han, B. (2009) Evolutionary and expression study of the aldehyde dehydrogenase (ALDH) gene superfamily in rice (*Oryza sativa*). *Gene* **431**, 86–94 [CrossRef PubMed](#)
- 23 Zhu, C., Ming, C., Zhao-Shi, X., Lian-Cheng, L., Xue-Ping, C. and You-Zhi, M. (2014) Characteristics and expression patterns of the aldehyde dehydrogenase (ALDH) gene superfamily of foxtail millet (*Setaria italica* L.). *PLoS One* **9**, e101136 [CrossRef PubMed](#)
- 24 Rodrigues, S. M., Andrade, M. O., Gomes, A. P., Damatta, F. M., Baracat-Pereira, M. C. and Fontes, E. P. (2006) Arabidopsis and tobacco plants ectopically expressing the soybean antiquitin-like ALDH7 gene display enhanced tolerance to drought, salinity, and oxidative stress. *J. Exp. Bot.* **57**, 1909–1918 [CrossRef PubMed](#)
- 25 Kotchoni, S. O., Kuhns, C., Ditzer, A., Kirch, H. H. and Bartels, D. (2006) Over-expression of different aldehyde dehydrogenase genes in *Arabidopsis thaliana* confers tolerance to abiotic stress and protects plants against lipid peroxidation and oxidative stress. *Plant Cell Environ.* **29**, 1033–1048 [CrossRef PubMed](#)
- 26 Shin, J. H., Kim, S. R. and An, G. (2009) Rice aldehyde dehydrogenase 7 is needed for seed maturation and viability. *Plant Physiol.* **149**, 905–915 [CrossRef PubMed](#)
- 27 Tylichová, M., Kopečný, D., Morera, S., Briozzo, P., Lenobel, R., Sněgaroff, J. and Šebela, M. (2010) Structural and functional characterization of plant aminoaldehyde dehydrogenase from *Pisum sativum* with a broad specificity for natural and synthetic aminoaldehydes. *J. Mol. Biol.* **396**, 870–882 [CrossRef PubMed](#)
- 28 Kopečný, D., Tylichová, M., Sněgaroff, J., Popelková, H. and Šebela, M. (2011) Carboxylate and aromatic active-site residues are determinants of high-affinity binding of ω -aminoaldehydes to plant aminoaldehyde dehydrogenases. *FEBS J.* **278**, 3130–3139 [CrossRef PubMed](#)
- 29 Kopečný, D., Končítiková, R., Tylichová, M., Vigouroux, A., Moskalíková, H., Soural, M., Šebela, M. and Morera, S. (2013) Plant ALDH10 family: identifying critical residues for substrate specificity and trapping a thiohemiacetal intermediate. *J. Biol. Chem.* **288**, 9491–9507 [CrossRef PubMed](#)
- 30 Sophos, N. A., Pappa, A., Ziegler, T. L. and Vasiliou, V. (2001) Aldehyde dehydrogenase gene superfamily: the 2000 update. *Chem.-Biol. Interact.* **130–132**, 323–337 [CrossRef](#)
- 31 Huang, X. and Miller, W. (1991) A time-efficient linear-space local similarity algorithm. *Adv. Appl. Math.* **12**, 337–357 [CrossRef](#)
- 32 Šebela, M., Brauner, F., Radová, A., Jacobsen, S., Havliš, J., Galuszka, P. and Peč, P. (2000) Characterisation of a homogeneous plant aminoaldehyde dehydrogenase. *Biochim. Biophys. Acta* **1480**, 329–341 [PubMed](#)
- 33 Vaz, F. M., Fouchier, S. W., Orfan, R., Sommer, M. and Wanders, R. J. (2000) Molecular and biochemical characterization of rat γ -trimethylaminobutyraldehyde dehydrogenase and evidence for the involvement of human aldehyde dehydrogenase 9 in carnitine biosynthesis. *J. Biol. Chem.* **275**, 7390–7394 [CrossRef PubMed](#)
- 34 Trossat, C., Rathinasabapathi, B. and Hanson, A. D. (1997) Transgenically expressed betaine aldehyde dehydrogenase efficiently catalyzes oxidation of dimethylsulfoniopropionaldehyde and -aminoaldehydes. *Plant Physiol.* **113**, 1457–1461 [PubMed](#)
- 35 Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 [CrossRef](#)
- 36 Karplus, P. A. and Diederichs, K. (2012) Linking crystallographic model and data quality. *Science* **336**, 1030–1033 [CrossRef PubMed](#)
- 37 Diederichs, K. and Karplus, P. A. (2013) Better models by discarding data? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1215–1222 [CrossRef](#)
- 38 Storoni, L. C., McCoy, A. J. and Read, R. J. (2004) Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 432–438 [CrossRef](#)
- 39 Tang, W. K., Wong, K. B., Lam, Y. M., Cha, S. S., Cheng, C. H. and Fong, W. P. (2008) The crystal structure of seabream antiquitin reveals the structural basis of its substrate specificity. *FEBS Lett.* **582**, 3090–3096 [CrossRef PubMed](#)
- 40 Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O. S., Vornrhein, C. and Womack, T. O. (2011), In BUSTER version 2.1.0. Global Phasing Ltd., Cambridge, United Kingdom
- 41 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 [CrossRef PubMed](#)
- 42 Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. (2006) The SWISS-MODEL Workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201 [CrossRef PubMed](#)
- 43 Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 [CrossRef PubMed](#)
- 44 Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 [CrossRef PubMed](#)
- 45 Guindon, S. and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 [CrossRef PubMed](#)
- 46 Chen, X., Zeng, Q. and Wood, A. J. (2002) Aldh7B6 encodes a turgor-responsive aldehyde dehydrogenase homologue that is constitutively expressed in *Tortula ruralis* gametophytes. *Bryologist* **105**, 177–184 [CrossRef](#)
- 47 Dudareva, N., Klempien, A., Muhlemann, J. K. and Kaplan, I. (2013) Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytol.* **198**, 16–32 [CrossRef PubMed](#)
- 48 Kopečný, D., Pethe, C., Šebela, M., Houba-Hérin, N., Madzak, C., Majira, A. and Laloue, M. (2005) High-level expression and characterization of *Zea mays* cytokinin oxidase/dehydrogenase in *Yarrowia lipolytica*. *Biochimie* **87**, 1011–1022 [CrossRef PubMed](#)
- 49 Boatright, J., Negre, F., Chen, X., Kish, C. M., Wood, B., Peel, G., Orlova, I., Gang, D., Rhodes, D. and Dudareva, N. (2004) Understanding *in vivo* benzenoid metabolism in petunia petal tissue. *Plant Physiol.* **135**, 1993–2011 [CrossRef PubMed](#)
- 50 Shen, Y., Zhang, Y., Yang, C., Lan, Y., Liu, L., Liu, S., Chen, Z., Ren, G. and Wan, J. (2012) Mutation of OsALDH7 causes a yellow-colored endosperm associated with accumulation of oryzamutic acid A in rice. *Planta* **235**, 433–441 [CrossRef PubMed](#)
- 51 Nakano, H., Kosemura, S., Suzuki, T., Hirose, K., Kaji, R. and Sakai, M. (2009) Oryzamutic acid A, a novel yellow pigment from an *Oryza sativa* mutant with yellow endosperm. *Tetrahedron Lett.* **50**, 2003–2005 [CrossRef](#)
- 52 Nakano, H., Kosemura, S., Yoshida, M., Suzuki, T., Iwaura, R., Kaji, R., Sakai, M. and Hirose, K. (2010) Oryzamutic acids B-G, new alkaloids from an *Oryza sativa* mutant with yellow endosperm. *Tetrahedron Lett.* **51**, 49–53 [CrossRef](#)
- 53 Kemper, E. L., Neto, G. C., Papes, F., Moraes, K. C., Leite, A. and Arruda, P. (1999) The role of opaque2 in the control of lysine-degrading activities in developing maize endosperm. *Plant Cell* **11**, 1981–1994 [CrossRef PubMed](#)
- 54 Zhu, X., Tang, G. and Galili, G. (2000) Characterization of the two saccharopine dehydrogenase isozymes of lysine catabolism encoded by the single composite AtLKR=AtLKR/SDH locus of *Arabidopsis*. *Plant Physiol.* **124**, 1363–1372 [CrossRef PubMed](#)
- 55 Wong, J. W., Chan, C. L., Tang, W. K., Cheng, C. H. and Fong, W. P. (2010) Is antiquitin a mitochondrial enzyme? *J. Cell. Biochem.* **109**, 74–81 [PubMed](#)
- 56 Hallen, A., Jamie, J. F. and Cooper, A. J. (2013) Lysine metabolism in mammalian brain: an update on the importance of recent discoveries. *Amino Acids.* **45**, 1249–1272 [CrossRef PubMed](#)
- 57 Perez-Miller, S. J. and Hurley, T. D. (2003) Coenzyme isomerization is integral to catalysis in aldehyde dehydrogenase. *Biochemistry* **42**, 7100–7109 [CrossRef PubMed](#)
- 58 Bosch, M., Mayer, C. D., Cookson, A. and Donnison, I. S. (2011) Identification of genes involved in cell wall biogenesis in grasses by differential gene expression profiling of elongating and non-elongating maize internodes. *J. Exp. Bot.* **62**, 3545–3561 [CrossRef PubMed](#)
- 59 Chan, C. L., Wong, J. W., Wong, C. P., Chan, M. K. and Fong, W. P. (2011) Human antiquitin: structural and functional studies. *Chem. Biol. Interact.* **191**, 165–170 [CrossRef PubMed](#)
- 60 Farrés, J., Wang, X., Takahashi, K., Cunningham, S. J., Wang, T. T. and Weiner, H. (1994) Effects of changing glutamate 487 to lysine in rat and human liver mitochondrial aldehyde dehydrogenase. A model to study human (Oriental type) class 2 aldehyde dehydrogenase. *J. Biol. Chem.* **269**, 13854–13860 [PubMed](#)
- 61 Stiti, N., Podgórska, K. and Bartels, D. (2014) Aldehyde dehydrogenase enzyme ALDH3H1 from *Arabidopsis thaliana*: Identification of amino acid residues critical for cofactor specificity. *Biochim. Biophys. Acta* **1844**, 681–693 [CrossRef PubMed](#)

- 62 Tang, G., Miron, D., Zhu-Shimoni, J. X. and Galili, G. (1997) Regulation of lysine catabolism through lysine-ketoglutarate reductase and saccharopine dehydrogenase in *Arabidopsis*. *Plant Cell* **9**, 1305–1316 [PubMed](#)
- 63 Wen, Y., Wang, X., Xiao, S. and Wang, Y. (2012) 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* **236**, 525–539 [CrossRef](#) [PubMed](#)
- 64 Esterbauer, H. and Zollner, H. (1989) Methods for determination of aldehydic lipid peroxidation products. *Free Radic. Biol. Med.* **7**, 197–203 [CrossRef](#) [PubMed](#)
- 65 Bassard, J. E., Ullmann, P., Bernier, F. and Werck-Reichhart, D. (2010) Phenolamides: bridging polyamines to the phenolic metabolism. *Phytochemistry* **71**, 1808–1824 [CrossRef](#) [PubMed](#)
- 66 Gaquerel, E., Gulati, J. and Baldwin, I. T. (2014) Revealing insect herbivory-induced phenolamide metabolism: from single genes to metabolic network plasticity analysis. *Plant J.* **79**, 679–692 [CrossRef](#) [PubMed](#)
- 67 Onkokesung, N., Gaquerel, E., Kotkar, H., Kaur, H., Baldwin, I. T. and Galis, I. (2012) MYB8 controls inducible phenolamide levels by activating three novel hydroxycinnamoyl coenzyme A: polyamine transferases in *Nicotiana attenuata*. *Plant Physiol.* **158**, 389–407 [CrossRef](#) [PubMed](#)
- 68 Martin-Tanguy, J., Perdrizet, E., Prevost, J. and Martin, C. (1982) Hydroxycinnamic acid amides in fertile and cytoplasmic male sterile lines of maize. *Phytochemistry* **21**, 1939–1945 [CrossRef](#)
- 69 Grienemberger, E., Besseau, S., Geoffroy, P., Debayle, D., Heintz, D., Lapierre, C., Pollet, B., Heitz, T. and Legrand, M. (2009) A BAH1 acyltransferase is expressed in the tapetum of *Arabidopsis* anthers and is involved in the synthesis of hydroxycinnamoyl spermidines. *Plant J.* **58**, 246–259 [CrossRef](#) [PubMed](#)
- 70 Fellenberg, C., Milkowski, C., Hause, B., Lange, P. R., Böttcher, C., Schmidt, J. and Vogt, T. (2008) Tapetum-specific location of a cation-dependent O-methyltransferase in *Arabidopsis thaliana*. *Plant J.* **56**, 132–145 [CrossRef](#) [PubMed](#)
- 71 Fellenberg, C., Böttcher, C. and Vogt, T. (2009) Phenylpropanoid polyamine conjugate biosynthesis in *Arabidopsis thaliana* flower buds. *Phytochemistry* **70**, 1392–1400 [CrossRef](#) [PubMed](#)

Received 5 January 2015/23 February 2015; accepted 3 March 2015

Published as BJ Immediate Publication 3 March 2015, doi:10.1042/BJ20150009