Plant Molecular Biology **48:** 751–764, 2002.

Characterization of the aldehyde dehydrogenase gene families of *Zea mays* and *Arabidopsis*

David S. Skibbe[1,2,+], Feng Liu[1,3,+], Tsui-Jung Wen[5,+], Marna D. Yandeau[1,3], Xiangqin Cui[1,3], Jun Cao[1,3], Carl R. Simmons[4] and Patrick S. Schnable[1,2,3,5,6,*]

[1]*Department of Zoology & Genetics, Iowa State University, Ames, IA 50011, USA;* [2]*Molecular, Cellular, and Developmental Biology Program, Iowa State University, Ames, IA 50011, USA;* [3]*Interdepartmental Genetics Program, Iowa State University, Ames, IA 50011, USA;* [4]*Pioneer Hi-Bred Intl. Inc., Johnston, IA 50131, USA;* [5]*Department of Agronomy, Iowa State University, Ames, IA 50011, USA;* [6]*Center for Plant Genomics, Iowa State University, Ames, IA 50011, USA;* [*]*author for correspondence (G405 Agronomy Hall, Iowa State University, Ames, IA 50011, USA; e-mail schnable@iastate.edu);* [+]*these authors contributed equally to this work*

Received 19 February 2001; accepted in revised form 22 July 2001

Key words: cytoplasmic male sterility, functional divergence, gene family evolution, nuclear restorer genes, prediction of functional residues, *rf2*

Abstract

Cytoplasmic male sterility is a maternally transmitted inability to produce viable pollen. Male sterility occurs in Texas (T) cytoplasm maize as a consequence of the premature degeneration of the tapetal cell layer during microspore development. This sterility can be overcome by the combined action of two nuclear restorer genes, *rf1* and *rf2a*. The *rf2a* gene encodes a mitochondrial aldehyde dehydrogenase (ALDH) that is capable of oxidizing a variety of aldehydes. Six additional ALDH genes were cloned from maize and *Arabidopsis*. *In vivo* complementation assays and *in vitro* enzyme analyses demonstrated that all six genes encode functional ALDHs. Some of these ALDHs are predicted to accumulate in the mitochondria, others in the cytosol. The intron/exon boundaries of these genes are highly conserved across maize and *Arabidopsis* and between mitochondrial and cytosolic ALDHs. Although animal, fungal, and plant genomes each encode both mitochondrial and cytosolic ALDHs, it appears that either the gene duplications that generated the mitochondrial and the cytosolic ALDHs occurred independently within each lineage or that homogenizing gene conversion-like events have occurred independently within each lineage. All studied plant genomes contain two confirmed or predicted mitochondrial ALDHs. It appears that these mitochondrial ALDH genes arose via independent duplications after the divergence of monocots and dicots or that independent gene conversion-like events have homogenized the mitochondrial ALDH genes in the monocot and dicot lineages. A computation approach was used to identify amino acid residues likely to be responsible for functional differences between mitochondrial and cytosolic ALDHs.

Abbreviations: ALDH, aldehyde dehydrogenase; *At*, *Arabidopsis thaliana; Bt*, *Bos taurus*; cALDH, cytosolic aldehyde dehydrogenase; cms, cytoplasmic male sterility; cms-T, Texas cytoplasm maize, *Hs*, *Homo sapiens*; mtALDH, mitochondrial aldehyde dehydrogenase; *Nt*, *Nicotiana tabacum*; *Os*, *Oryza sativa*; *Rn*, *Rattus norvegicus*; *Sc*, *Saccharomyces cerevisiae*; T cytoplasm, Texas cytoplasm; *Zm*, *Zea mays*

Introduction

Cytoplasmic male sterility (cms) is a maternally inherited inability to produce viable pollen (reviewed by Duvick, 1965; Laughnan and Gabay-Laughnan, 1983). In Texas cytoplasm maize (cms-T) sterility arises from the premature degeneration of the tapetal

cell layer during microspore development (Warmke and Lee, 1977). The T cytoplasm-specific mitochondrial gene, *T-urf13*, that is responsible for this sterility encodes a 13 kDa protein termed URF13 (Wise *et al*., 1987). The mechanism by which URF13 causes male sterility is not known. However, working in concert dominant alleles of *rf1* and *rf2* (now termed *rf2a*) can overcome URF13-mediated sterility. As such, they are nuclear restorer of fertility genes (reviewed by Schnable and Wise, 1998; Wise *et al*., 1999).

To better understand the molecular mechanism of fertility restoration in cms-T, the *rf2a* gene was cloned (Cui *et al*., 1996). Subcellular localization studies and enzymatic characterizations have demonstrated that *rf2a* is a mitochondrial aldehyde dehydrogenase (mtALDH) (Liu *et al*., 2001). Subsequently, *rf2a* sequences were used to assist with the cloning of mtALDHs from tobacco (op den Camp and Kuhlemeier, 1997), rice (Nakazono *et al*., 2000), and *Arabidopsis* (Li *et al*., 2000).

ALDHs catalyze the nearly irreversible oxidation of aldehydes to the corresponding carboxylic acid. Aldehydes are common by-products of a number of metabolic pathways, including the metabolism of vitamins, amino acids, carbohydrates, and lipids (Schauenstein *et al*., 1977). The resulting aldehydes are highly reactive molecules; because of its electrophilic nature an aldehyde's carbonyl group can attack cellular nucleophiles, including proteins and nucleic acids. The damaging effects of aldehydes have been well studied in man and include cytotoxicity, mutagenicity, and carcinogenicity (Schauenstein *et al*., 1977). Therefore, the removal of aldehydes is essential for cellular survival.

ALDHs have been widely studied in man (for reviews see Lindahl, 1992; Yoshida *et al*., 1998), but relatively few studies have been conducted on the corresponding plant enzymes. Based on the finding that indole-3-acetaldehyde dehydrogenase activity has been detected in cell-free extracts from mung bean seedlings (Wightman and Cohen, 1968), it has been proposed that ALDHs may function in the production of indole-3-acetyl acetate (Marumo, 1986). In another study, mtALDHs isolated from potato tubers and pea epicotyls exhibited kinetic properties consistent with the *in vivo* oxidation of acetaldehyde, glycolaldehyde, and indoleacetaldehyde (Asker and Davies, 1985). This supports the observation that while some classes of ALDHs are highly substrate-specific (e.g. betaine ALDH and glyceraldehyde-3-phosphate dehydrogenase), others, including the *rf2a*-encoded mtALDH (Liu *et al*., 2001), exhibit broad substrate specificities.

*NtAldh2*A mRNA and protein levels are substantially higher in tobacco reproductive tissues than in leaves (op den Camp and Kuhlemeier, 1997). Futhermore, ethanolic fermentation occurs under both aerobic and anaerobic conditions in developing tobacco pollen and produces high levels of acetaldehyde, which can be converted to acetate by ALDH (Tadege *et al*., 1997). Two mtALDH genes have been identified in rice. Northern analyses of these genes have revealed that *OsAldh2a* expression is dramatically increased upon anaerobic treatment, while *OsAldh2b* is not affected by anaerobic treatment (Nakazono *et al*., 2000 and personal communication). Two cALDH genes from rice have been identified, and one, *OsAldh1a*, has been shown by RT-PCR analysis to be expressed highly in roots of rice seedlings (Li *et al*., 2000).

The specific pathway(s) within which plant ALDHs act is an area of considerable interest because *rf2a* is the first, and to date only, nuclear restorer of fertility gene to be cloned from any species. To better understand the roles of ALDHs in plants, six additional ALDH genes from maize and *Arabidopsis* were cloned. All six ALDHs were functionally characterized and display ALDH activity. Phylogenetic and computational analyses were used to reconstruct the origins of gene duplications and to predict those amino acids responsible for functional differences among classes of ALDHs.

Materials and methods

Nomenclature

A nomenclature based on sequence similarity has been developed for eukaryotic ALDH genes (Vasiliou *et al*., 1999; http://www.uchsc.edu/sp/sp/alcdbase/aldhcov.html). In this nomenclature the first digit indicates a 'family' and the first letter a 'subfamily', while the final number identifies an individual gene within a subfamily.

Although the existing trivial names will be used in the body of this report, the official designations are listed here to facilitate comparisons across taxa.

According to the new nomenclature, the maize *rf2a* (Cui *et al*., 1996), *rf2b*, *rf2c* and *rf2d* genes encode the ALDH2B1, ALDH2B6, ALDH2C2, and ALDH2C3 proteins, respectively. The *Arabidopsis* genes previously designated *ALDH1a, ALDH2a* and

ALDH2b by Li *et al*. (2000) (and which are equivalent to *AtALDH2, AtALDH1* and *AtALDH3* in the Schnable laboratory nomenclature) encode proteins designated as ALDH2C4, ALDH2B4 and ALDH2B7, respectively. The tobacco gene *TobAldh2a* (op den Camp and Kuhlemeier, 1997) is equivalent to ALDH2B2. The rice ALDH genes *OsAldh1a*, *OsAldh2a*, and *OsAldh2b* (Li *et al*., 2000; Nakazono *et al*., 2000) are equivalent to *ALDH2C1*, *ALDH2B5*, and *ALDH2B1*, respectively. In this report the prefixes *Zm* (*Zea mays*), *Os* (*Oryza sativa*), *Nt* (*Nicotiana tabacum*) and *At* (*Arabidopsis thaliana*) have been added to gene symbols or protein designations when required for clarity.

Cloning of ALDH genes

The sequences of *rf2a* cDNA (GenBank accession number U43082) and genomic (GenBank accession number AF215823) clones have been described previously (Cui *et al*., 1996, and submitted). A full-length cDNA clone of *rf2b* was obtained from a C131A root plus shoot cDNA library provided by Monica Frey and Alfons Gierl (Technische Universität München, Garching, Germany) by screening with the cDNA insert from prf2a′ (see Results). A full-length clone of *rf2c* was obtained from a W22 immature tassel cDNA library, provided by Alejandro Calderon and Steven Dellaporta (Yale University), by screening with the cDNA insert from p0016.ctsad31r (see Results). A full-length *rf2d* cDNA clone (prf2d-exprA7) was obtained from B73 seedling leaf total RNA using the GeneRacer kit in combination with primers developed from p0104.cabau70r (see Results) (Invitrogen, Carlsbad, CA).

Genomic clones that span the entire coding regions of *rf2b* (*rf2b*λ#25) and *rf2c (rf2c*λ*#5-1)* and a genomic clone that includes most of *rf2d* (*rf2d*λ*#21-1*) were isolated from a λDASHII B73 genomic DNA library obtained from John Tossberg (Pioneer Hi-Bred, Johnston, IA). Overlapping genomic DNA fragments for each λ clone were subcloned into the pCSOS-72 vector and sequenced using the TN1000 system (Gold Biotechnology, St. Louis, MO). The resulting subclones were: for *rf2b* pTN6, pTN20, pTN22, pTN24, pTN33, pTN38, pTNS7, and pTNS36; for *rf2c* pCSOS/cBB5.0, pCSOS/cXX4.8, and pCSOS/cBN4.2; for *rf2d* pCSOS/dP4.8 and pCSOS/dS3.3.

Arabidopsis genomic DNA sequences for *AtALDH1a* (P1 clone MOB24; GenBank accession number AB020746), *AtALDH2a* (BAC clone T17F15; AL049658), and *AtALDH2b* (BAC clone F508; AC005990) were obtained from the *Arabidopsis* genome sequencing project (*Arabidopsis* Genome Initiative, 2000).

Genetic mapping

The mapping of *rf2a* was described previously (Wise and Schnable, 1994). The positions of *rf2b*, *rf2c*, and *rf2d* were defined via RFLP mapping of the Ben Burr recombinant inbred lines (Burr *et al*., 1988). The full-length cDNA insert from pRB73 was used to map the *rf2b* locus, while the *rf2c* and *rf2d* loci were mapped using partial cDNA inserts from Dupont/Pioneer ESTs p0016.ctsad31r and p0104.cabau70r, respectively.

Computational analyses

Figure 1 was prepared by first aligning ALDH amino acid sequences using ClustalX 1.62b (Thompson *et al*., 1997) and then generating a neighbor-joining tree from 465 sites using version 1.00 of MEBoot. Gaps were deleted and distances estimated using the Poisson method. The neighbor-joining tree was then bootstrapped using 1000 replicates.

A multiple alignment of amino acid sequences of the 11 plant ALDHs shown in Figure 1 was created with ClustalX 1.62b. After trimming signal sequences, a neighbor-joining tree was prepared using Paup4b1 (Swofford, 2000); all of the bootstrapping values were larger than 50%. After trimming gaps in the sequences, version 1.01 of the Gu99 software (Gu, 1999) was used to obtain the expected number of amino acid substitutions at each site and to then calculate the coefficient of functional divergence θ, which is the probability that the evolutionary rate at a site is statistically independent between two gene clusters.

Pair-wise alignments of sequences were determined using the Wisconsin GCG software package Version 10.0-UNIX from the Genetics Computer Group.

Primers

All primers were synthesized by the Iowa State University Nucleic Acid Facility (Ames, IA): rf2b1, 5′-ATTGGCCCTGGTTGAAGAAGAC-3′; rf2b5, 5′-AGCTAGCACCGCACCGGCAT-3′; rf2c-expr, 5′-CGAGGCTAGCATGGCGACTGCGAAC-3′; rf2c-rev1, 5′-AATCCCGGCACGACGTTGAG-3′; rf2d-expr, 5′-CGAGGCTAGCATGGCGAGCAACGGC-3′; rf2d-1011L, 5′-TGAAGAAGACGGCAAGCCTC-3′; A117b.U, 5′-CAACATATGACCTCTTCTGCTG-

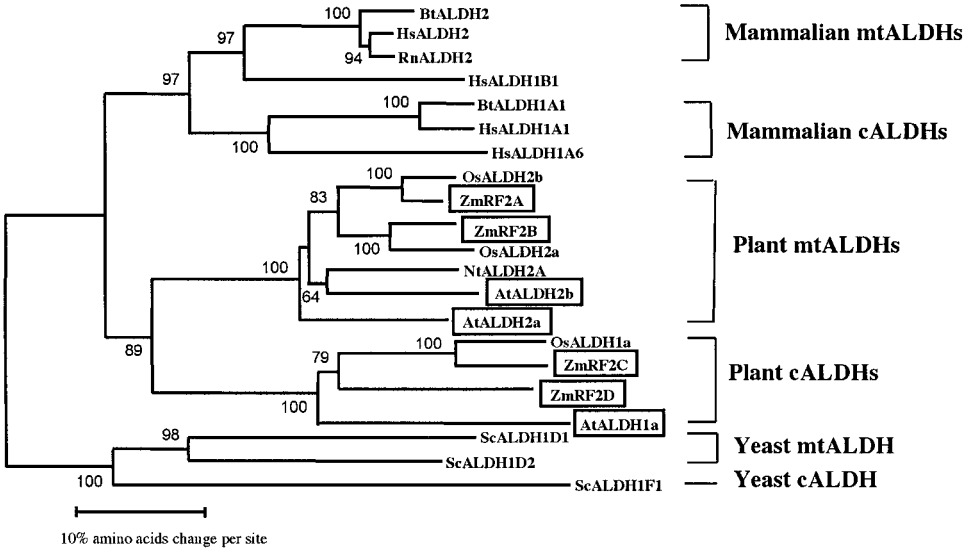

Figure 1. Phylogenetic tree of selected ALDHs. The tree was constructed using the MEBoot program. A neighbor-joining tree was initially used and 1000 bootstrap replicates were conducted using 456 amino acid sites. Bootstrap values above 50% are indicated at each node. At, *Arabidopsis thaliana*; Bt, *Bos taurus*; Hs, *Homo sapiens*; Nt, *Nicotiana tabacum*; Os, *Oryza sativa*; Rn, *Rattus norvegicus*; Sc, *Saccharomyces cerevisiae*; Zm, *Zea mays*; mtALDH, mitochondrial ALDH (actual or predicted); cALDH, cytosolic ALDH (actual or predicted). Boxed ALDHs were characterized in this study. GenBank and SwissProt accessions for ALDHs used in this study: BtALDH2, P20000; HsALDH2, AAA51693; RnALDH2, S03564; HsALDH1B1, XP_005402; BtALDH1A1, AAA74234; HsALDH1A1, AAA51692; HsALDH1A6, AAA79036; OsALDH2b, AB044537; ZmRF2A, U43082; ZmRF2B, AF348417; OsALDH2a, AB030939; NtALDH2a, Y09876; AtALDH2b, AF348416; AtALDH2a, AF349447; OsALDH1a, AB037421; ZmRF2C, AF348413; ZmRF2D, AF348415; AtALDH1a, AF349448; ScALDH1D1, AAB01220; ScALDH1D2, P46367; ScALDH1F, AAB01219.

CTGC-3′; A117b.L, 5′-TCGGAGCTCTTCGCTGTGTTTCT-3′; A217b.u, 5′-ACTCATATGGAGAACGGCAAATGCAA-3′; A2ETL, 5′-TCAGGAATTCCCGCTTCTTTTGAGAG-3′; ALDH3-expr, 5′-CGAGGCTAGCTCGCTGCTGCTGTCGAA-3′; ALDH3-778L, 5′-GCAGTAGCCCCAAATCCAG-3′.

Production of expression constructs

The construction of pMAP11, which expresses *rf2a*, has been described previously (Liu *et al.*, 2001). A similar approach was used to generate expression constructs for *rf2b*, *rf2c*, and *rf2d*. Specifically, an *Nhe*I restriction site was introduced into the appropriate cDNA at either the predicted cleavage site of the predicted mitochondrial targeting sequence (*rf2b*) using the primer rf2b5, or at the start codon (*rf2c* and *rf2d*) using primers rf2c-expr and rf2d-expr, respectively.

To create the *rf2b* expression construct, the PCR product derived from the amplification from the full-length cDNA clone pRB73 using primers rf2b5 and rf2b1 was digested with *Nhe*I and *Hin*dIII and ligated into the *Nhe*I and *Hin*dIII sites of pET17b (Novagen, Madison, WI), thereby generating pNH2. pNH2 was then digested with *Hin*dIII and *Bam*HI and the fragment containing the vector backbone was rescued and ligated with a 1.0 kb pRB73 *Hin*dIII-*Bam*HI fragment, resulting in the expression construct pRB17.

To create the *rf2c* expression construct, the PCR product derived from the amplification from the full-length cDNA clone pBSK+/3-1#3 with primers rf2c-expr and rf2c-rev1 was digested with *Nhe*I and *Bam*HI to release a 0.2 kb fragment. This fragment was ligated into the *Nhe*I-*Eco*RI sites of pET17b along with a 1.5 kb *Bam*HI-*Eco*RI 3′ restriction fragment from pBSK+/3-1#3, thereby creating the expression construct prf2c-expr#22.

To create the *rf2d* expression construct, the PCR product derived from the amplification from clone prf2d-TOPOB6 (a product of 5′ RACE) with primers rf2d-expr and rf2d-1011L was digested with *Nhe*I and *Sac*I to release a 0.8 kb fragment. This fragment was ligated into the *Nhe*I and *Xho*I sites of pET17b along with the 1.0 kb *Sac*I-*Xho*I fragment from the partial cDNA clone p0104.cabau70r, thereby creating the expression construct prf2d-exprA7.

To create the AtALDH1a expression construct, a 0.7 kb PCR product was amplified from the full-length cDNA clone pTJ311 using primers A217b.u

and A2ETL. This PCR product was then purified, digested with *Nde*I and *Eco*RI, and subcloned into the corresponding restriction sites of pET17b. The 0.95 kb fragment obtained by digesting pTJ311 with *Eco*RI and *Not*I was then cloned into the plasmid that contained the 0.7 kb PCR product to create the expression construct pA217b.

To create the AtALDH2a expression construct, a 0.3 kb PCR product was amplified from the full-length cDNA clone pAT6 using primers A117b.U and A117b.L and digested with *Nde*I and *Sac*I. This fragment was cloned into the corresponding sites of pET17b. The 1.3 kb *Sac*I-*Bam*HI fragment obtained by digesting pAT6 with *Sac*I and *Bam*HI was ligated to the pET17b-derived clone that contained the 0.3 kb PCR product to create the expression construct pA117b.

To create the AtALDH2b expression construct, a 0.78 kb PCR product was amplified from the full-length cDNA clone pTJ313 using primers ALDH3-expr and ALDH3-778L. This PCR product was purified and then digested with *Nhe*I and *Bam*HI to release a 0.1 kb *Nhe*I-*Bam*HI fragment. A 1.4 kb *Bam*HI-*Xho*I fragment was generated by digesting pTJ313 with *Bam*HI and *Xho*I. The 0.1 kb *Nhe*I-*Bam*HI and 1.4 kb *Bam*HI-*Xho*I fragments were then subcloned into pET17b that had been digested with *Nhe*I and *Xho*I to create the expression construct pALDH3-expr1.

All of these PCR amplifications were performed using high-fidelity *Taq* polymerase (Life Technologies, Rockville, MD) and the amplified portions of all expression clones were sequenced to identify PCR-induced mutations.

Escherichia coli *complementation studies*

The *E. coli* strain JA111 carries a mutation in an ALDH gene that makes it unable to grow on media in which 1,2-propanediol is the sole carbon source (Hidalgo *et al*., 1991). *E. coli* strain JA111(DE3) contains the recombinant λDE3 phage (Novagen, Inc., Madison, WI) which carries a T7 RNA polymerase gene under the control of the *lacUV5* promoter (Liu *et al*., 2001). Plasmid pALD9 contains an *E. coli* ALDH that is able to complement JA111(DE3) (Hidalgo *et al*., 1991).

Plasmids expressing RF2A, RF2B, RF2C, RF2D, AtALDH1a, AtALDH2a and AtALDH2b were transformed by electroporation into JA111(DE3). The resulting cultures were grown on solid medium containing 100 μg/ml ampicillin. Individual colonies from each transformation experiment were inoculated into liquid dyt medium containing 100 μg/ml ampicillin and grown overnight with shaking. A 10 μl aliquot of each culture was then transferred into a fresh 1 ml dyt with ampicillin and grown at 37 °C for 2 h. A 10 μl portion of each culture was then streaked on a basal medium (Boronat and Aguilar, 1979) supplemented with either 40 mM glucose or 40 mM 1,2-propanediol as the sole carbon source. Plates were scored after one day (glucose medium) or three to four days (1,2-propanediol medium) of growth at 37 °C.

ALDH enzyme assays

Enzyme assays were performed on *E. coli* strain JA111(DE3) harboring each of the ALDH expression constructs as described previously (Liu *et al*., 2001). For each reaction, 48 μg of protein was preincubated in the reaction mix for 40 s with either 1.8 mM acetaldehyde or 0.2 mM glycolaldehyde as substrate. Protein concentrations were measured using the BioRad Protein Reagent concentrated dye solution according to the procedures recommended by the manufacturer (BioRad, Hercules, CA).

Results

Cloning of maize and Arabidopsis *ALDH genes*

The *rf2* gene was cloned using a transposon tagging strategy (Cui *et al*., 1996). During a cDNA library screen for *rf2* (now termed *rf2a*) clones, weakly hybridizing plaques were observed. Subsequent purification and DNA sequencing of one of the corresponding partial cDNA clones (prf2a′) revealed that it was derived from a gene distinct from, but closely related to *rf2a*. This gene was termed *rf2b*.

A search of the DuPont/Pioneer EST database resulted in the identification of two additional classes of cDNA clones with reasonably high sequence similarities to *rf2a* and *rf2b*. These genes have been designated *rf2c* (Pioneer clone p0016.ctsad31r) and *rf2d* (Pioneer clone p0104.cabau70r). Subsequently, full-length cDNA clones were obtained and sequenced for *rf2b* (pRB73; GenBank accession number AF348417), *rf2c* (pBSK+/3-1#3; AF348413), and *rf2d* (prf2d-TOPOB6 and p0104.cabau70r; AF348415) via library screens and 5′ RACE experiments.

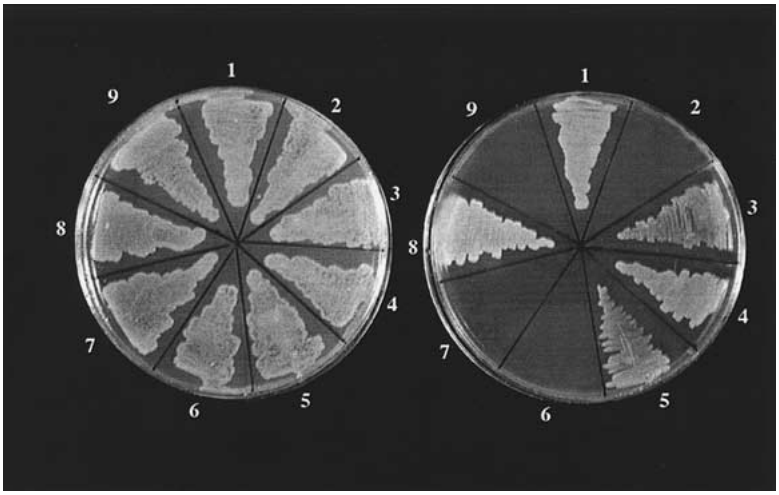

Figure 2. The *rf2a*, *rf2c, rf2d*, and *AtALDH1a* genes can complement an *E. coli* ALDH mutant. Expression constructs containing the *rf2a, rf2b, rf2c, rf2d, AtALDH1a*, *AtALDH2a* and *AtALDH2b* genes were each transformed into JA111(DE3) to assess complementation on basal media containing either glucose (left plate) or 1,2-propanediol (right plate) as the sole carbon source. For both plates: Section 1, JA111(DE3) pMAP11 (*rf2a*); Section 2, JA111(DE3) pRB17 (*rf2b*); Section 3, JA111(DE3) prf2c-expr#22 (*rf2c*); Section 4, JA111(DE3) prf2d-exprA7 (*rf2d*); Section 5, JA111(DE3) pA217b (*AtALDH1a*); Section 6, JA111(DE3) pA117b (*AtALDH2a*); Section 7, JA111(DE3) pALDH3-expr1 (AtALDH2b); Section 8, JA111(DE3) pALD9 (positive control); Section 9, JA111(DE3) pET17b (negative control).

A BLAST search (Altschul *et al*., 1990) of *Arabidopsis* ESTs performed using the RF2A sequence as the query identified *Arabidopsis* ESTs derived from three genes, *AtALDH1a*, *AtALDH2a*, and *AtALDH2b*. A full-length clone of *AtALDH2a* (pAT6; GenBank accession number AF349447) and near-full-length clones of *AtALDH1a* (pAT7, AA395226) and *AtALDH2b* (pAT4, R83958) were obtained from the *Arabidopsis* Biological Resource Center. The 5′ ends of the partial clones were used in cDNA library screens to obtain full-length clones of *AtALDH1a* (pTJ311; AF349448) and *AtALDH2b* (pTJ313; AF348416).

Maize and *Arabidopsis* cDNA clones were also identified that exhibit substantial sequence similarity to fatty aldehyde dehydrogenases; these will be the subject of a subsequent report.

Confirmation of ALDH function

E. coli strain 3 (Caballero *et al*., 1983) can grow on media in which 1,2-propanediol is the sole carbon source. It uses 1,2-propanediol as a carbon source by oxidizing 1,2-propanediol to L-lactaldehyde and subsequently oxidizing lactaldehyde to L-lactate which can be converted to pyruvate and enter central metabolism (Boronat and Aguilar, 1979). *E. coli* strain JA111 differs from strain 3 in that it carries a mutation in a gene that encodes an ALDH capable of oxidizing L-lactaldehyde to L-lactate. Because of this mutation, JA111 can not grow on media in which 1,2-propanediol is the sole carbon source (Hidalgo *et al*., 1991).

To test whether the plant ALDH genes can complement this *E. coli* ALDH mutant, expression constructs containing the coding regions from *rf2a*, *rf2b*, *rf2c*, *rf2d*, *AtALDH1a, AtALDH2a*, and *AtALDH2b* were produced and transformed into JA111(DE3). RF2A-, RF2C-, RF2D-, and AtALDH1a-expressing cultures were able to grow on selective media (Figure 2). RF2B-, AtALDH2a-, and AtALDH2b-expressing cultures were not. Because the *rf2a*, *rf2c*, *rf2d* and *AtALDH1a* genes can complement the ALDH mutation carried by JA111, we conclude that these genes encode functional ALDHs. The finding that the *rf2b*, *AtALDH2a* and *AtALDH2b* genes cannot complement this mutation suggests that either these genes do not encode functional ALDHs or that they are not able to utilize L-lactaldehyde as a substrate when expressed in *E. coli*.

To settle the question of whether the *rf2b*, *AtALDH2a*, and *AtALDH2b* genes encode functional ALDHs and to explore the substrate preferences of this set of proteins, ALDH enzyme assays were performed on crude protein extracts from *E. coli* expressing RF2A, RF2B, RF2C, RF2C, RF2D, AtALDH1a, AtALDH2a, and AtALDH2b. Glycolaldehyde and acetaldehyde were used as substrates in these exper-

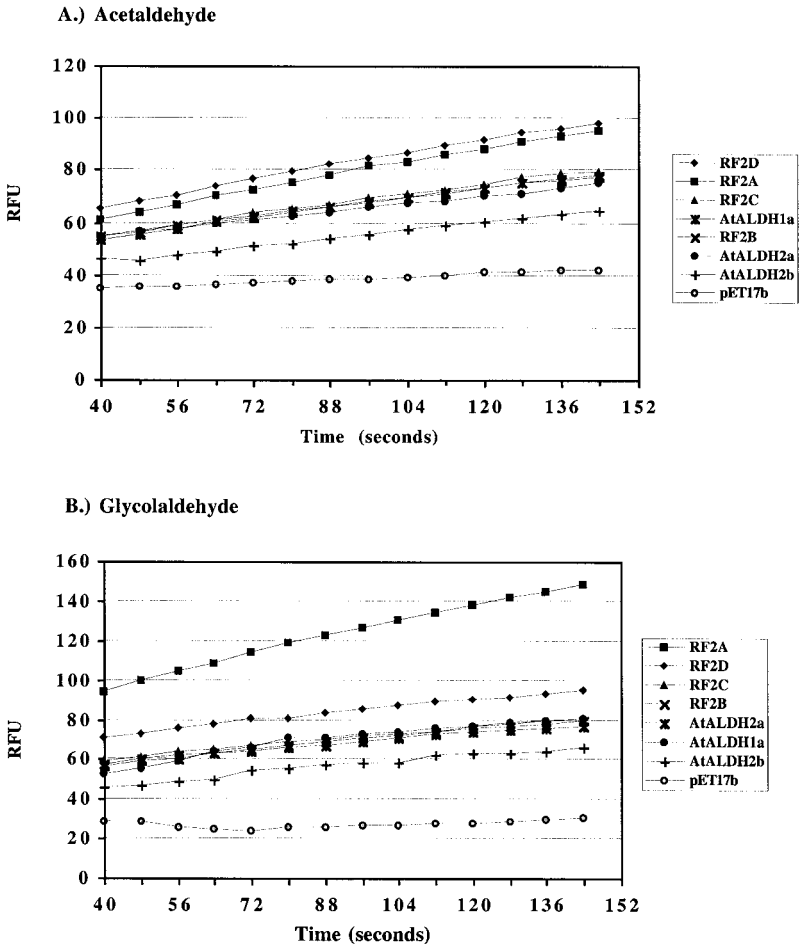

Figure 3. RF2A, RF2B, RF2C, RF2D, AtALDH1a, AtALDH2a, and AtALDH2b all exhibit ALDH activity. Crude extracts of *E. coli* expressing recombinant RF2A, RF2B, RF2C, RF2D, AtALDH1a, AtALDH2a, AtALDH2b protein or containing the empty vector pET17b were assayed for ALDH activity on acetaldehyde (panel A) and glycolaldehyde (panel B). RFU, relative fluorescence units at 460 nm.

iments. All extracts, including RF2B, AtALDH2a and AtALDH2b, exhibited ALDH activity (Figure 3). Hence, it can be concluded that all of these genes encode functional ALDHs.

Phylogenetic relationships among ALDH proteins

The PSORT subcellular localization prediction program (Nakai and Kanehisa, 1992) was used to test whether any of the ALDHs contain subcellular targeting signals. This program predicts with 92% certainty that the RF2A protein localizes to the mitochondrial matrix. Subsequent subcellular fractionation experiments confirmed this prediction (Liu *et al.*, 2001). PSORT also predicts to a high certainty that RF2B, AtALDH2a, and AtALDH2b localize to the mitochondrial matrix. The predicted cleavage sites of the mitochondrial targeting sequences according to Sjöling and Glaser (1998) are after Phe-45 (RF2B), Phe-14 (AtALDH2a), and Tyr-34 (AtALDH2b).

No subcellular localization predictions were obtained via PSORT analyses for RF2C, RF2D, or AtALDH1a. mtALDHs are typically between 520 and 550 amino acids, while cytosolic ALDHs (cALDHs) are typically around 500 amino acids (Perozich *et al.*, 1999). Consistent with the results from the PSORT analysis, RF2C and RF2D are considerably shorter than RF2A and RF2B. RF2C and RF2D consist of 503 and 512 amino acids, respectively, while RF2A and RF2B consist of 549 and 551 amino acids, respectively. In addition, the additional amino acids in RF2A and RF2B relative to RF2C and RF2D are located at the N-termini of the proteins, the location of typical mitochondrial targeting sequences. This result

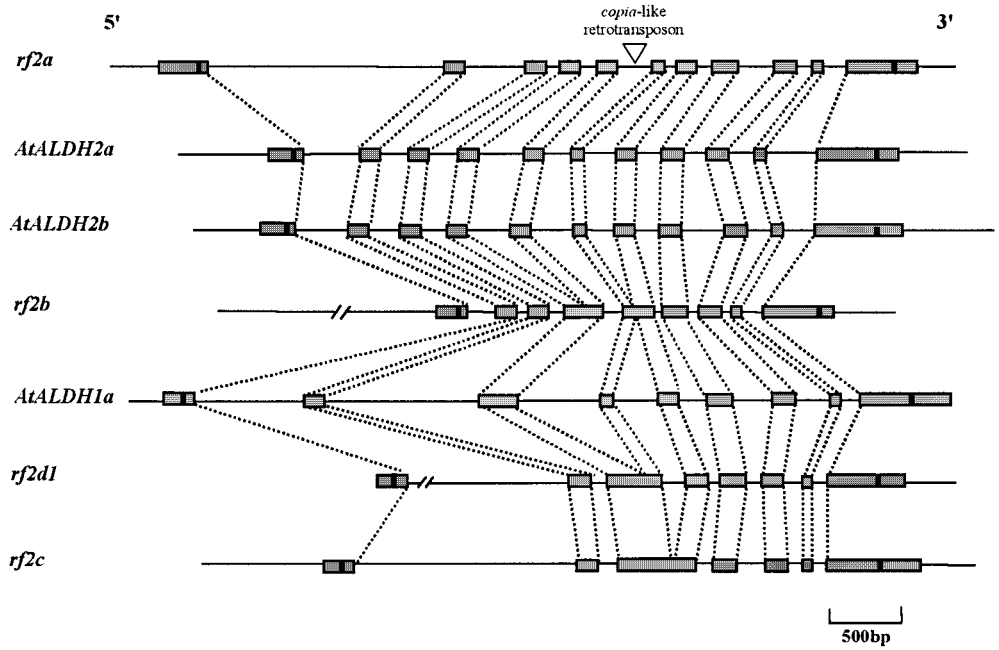

Figure 4. Gene structures of the ALDH genes of maize and *Arabidopsis*. Exons are represented by gray boxes. Horizontal lines represent intronic or non-transcribed sequences. The position of a retrotransposon insertion in *rf2a* is indicated by an open triangle. Start and stop codons are denoted by the vertical bars in exons at the 5′ and 3′ ends of the gene, respectively.

in combination with the failure of PSORT to predict subcellular localizations for RF2C, RF2D, and AtALDH1a have led us to putatively classify these proteins as cALDHs.

Phylogenetic comparison of the maize and *Arabidopsis* ALDH gene families demonstrates that mtALDHs from maize and *Arabidopsis* group together with other predicted plant mtALDHs to the exclusion of the mammalian mtALDHs (Figure 1). Similarly, the predicted cytosolic forms (AtALDH1a, RF2C, and RF2D) cluster together and are most similar to plant mtALDHs. These predicted cytosolic isozymes are equally divergent from both mammalian mtALDHs and mammalian cALDHs.

Conservation of gene structures

The *rf2a* gene structure has been reported previously (Cui *et al*., submitted). Genomic clones that contain the *rf2b* (GenBank accession number AF348418), *rf2c* (AF348412) and *rf2d* (AF348414) loci were isolated from a B73 genomic library and sequenced. The structures of these genes were determined via comparisons to the sequences of the corresponding full-length cDNA clones. The gene structures of the *AtALDH1a*, *AtALDH2a*, and *AtALDH2b* genes were determined via comparisons of the sequences of the full-length cDNAs with the corresponding genomic sequences obtained from the *Arabidopsis* genome project sequencing effort.

Comparisons of the seven resulting gene structures revealed that in addition to being well conserved at the protein level (Table 1), the maize and *Arabidopsis* mtALDH and cALDH genes also exhibit a striking degree of conservation of intron/exon boundaries both within and between species (Figure 4).

Origins of the maize ALDH genes

The maize genome contains at least four ALDH genes. It is possible that some of these genes arose via the segmental allotetraploidization event that restructured the maize genome 20 million years ago (Gaut and Doebley, 1997; Gale and Devos, 1998; Moore, 2000). To test this hypothesis, the genetic map positions of the four genes were determined via RFLP mapping of recombinant inbred lines (Figure 5). If these duplications arose during the segmental allotetraploidization event, then it would be expected that the duplicate genes reside on syntenic regions of the genome. However, in most instances the ALDH genes do not map to syntenic regions.

The single exception concerns the *rf2d* gene. The RFLP mapping experiments revealed that the *rf2d*

Table 1. Comparison of amino acid identity and similarity for the maize and *Arabidopsis* ALDHs. Numbers in shaded and non-shaded boxes represent amino acid similarities and identities, respectively.

	RF2A	RF2B	RF2C	RF2D	AtALDH1a	AtALHD2a	AtALDH2b
RF2A	–	83.4	68.6	69.4	67.8	82.9	83.9
RF2B	78.7	–	68.6	68.4	67.0	80.6	79.4
RF2C	59.0	60.0	–	80.9	76.3	68.3	66.0
RF2D	59.2	58.3	72.2	–	77.0	68.0	63.6
AtALDH1a	56.2	55.9	67.9	68.1	–	67.4	64.3
AtALDH2a	76.4	74.0	57.7	57.9	57.9	–	80.7
AtALDH2b	78.8	73.1	56.6	54.1	54.5	75.1	–

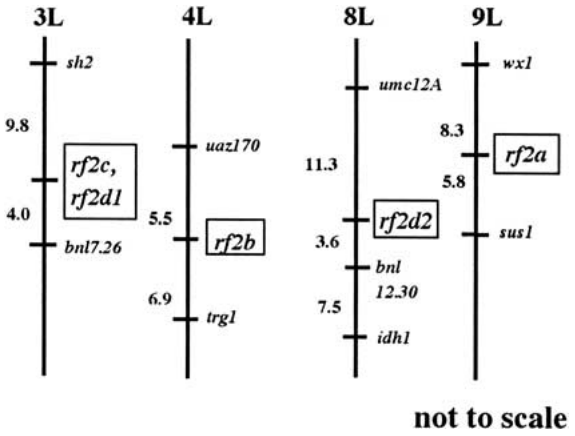

Figure 5. Genetic mapping of maize ALDH genes. *rf2a* was mapped as described previously (Wise and Schnable, 1994). *rf2b*, *rf2c*, and *rf2d* were mapped using the Ben Burr recombinant inbred lines (Burr *et al.*, 1988). Chromosome arms are indicated above each vertical line. Distances between markers are indicated in cM.

probe hybridizes to DNA fragments derived from the long arms of chromosomes 3 (major band) and 8 (minor band). The long arms of chromosomes 3 and 8 are syntenic (Wilson *et al.*, 1999; reviewed by Moore, 2000). The copy of *rf2d* on chromosome 3 (*rf2d1*) is closely linked to the *rf2c* locus. Even though no recombination events were observed between *rf2c* and *rf2d1* in any of the roughly 100 Ben Burr recombinant inbred lines (Burr *et al.*, 1988), *rf2d1* is clearly distinct from *rf2c* because the *rf2c* and *rf2d* probes detect different RFLPs.

Sequence non-conformity in RF2D, AtALDH1a, and AtALDH2b

Protein sequence alignments identified four amino acid residues that are fully conserved across 145 diverse ALDHs (Perozich *et al.*, 1999). These invariant residues are in regions of the ALDH proteins that are predicted to be involved in either the interaction of NAD+ with ALDH or the positioning of the catalytic nucleophile. An additional twelve residues are present in more than 95% of these 145 ALDHs (Perozich *et al.*, 1999). Analyses of the positions of these twelve residues relative to the crystal structures of ALDH proteins (Liu *et al.*, 1997; Steinmetz *et al.*, 1997; Moore *et al.*, 1998) suggest that these residues may be conserved due to their involvement in delineating secondary structure, interacting with NAD+, or catalytic activity. Each of these 16 residues is conserved in RF2A, RF2B, RF2C, and AtALDH2a.

RF2D and AtALDH1a exhibit similar conservation except at the ALDH index position Arg-25 [166]. In this nomenclature the number in brackets designates the index position of a site within the alignment of the 145 ALDHs; Arg-25 is the amino acid residue at that site in PDB ID: 1AD3. This site is equivalent to positions 100 and 89 in RF2D and AtALDH1a, respectively. RF2D and AtALDH1a both contain a Lys at index position 166 (Figure 6). Even though Arg-to-Lys changes are generally considered conservative, the Arg at index position 166 is conserved in all but two of the 145 ALDHs. The two exceptions are an Ω-crystallin from squid and an *E. coli* ALDH which contain at index position 166 Leu and Asp residues, respectively. Because one of these proteins (the Ω-crystallin from squid) is known to lack ALDH activity, RF2D and AtALDH1a represent only the second and third instances of functional ALDHs that contain a non-conserved residue at index position 166.

Position 303 in AtALDH2b contains a Glu residue at ALDH index position Gly-211 [400]. In contrast, all but five of 145 other ALDHs contain Gly at this position (Perozich *et al.*, 1999). Hence, this is a highly conserved residue. This may be because it is located in a structural turn between the catalytic and coenzyme binding domains in the class 3 ALDH

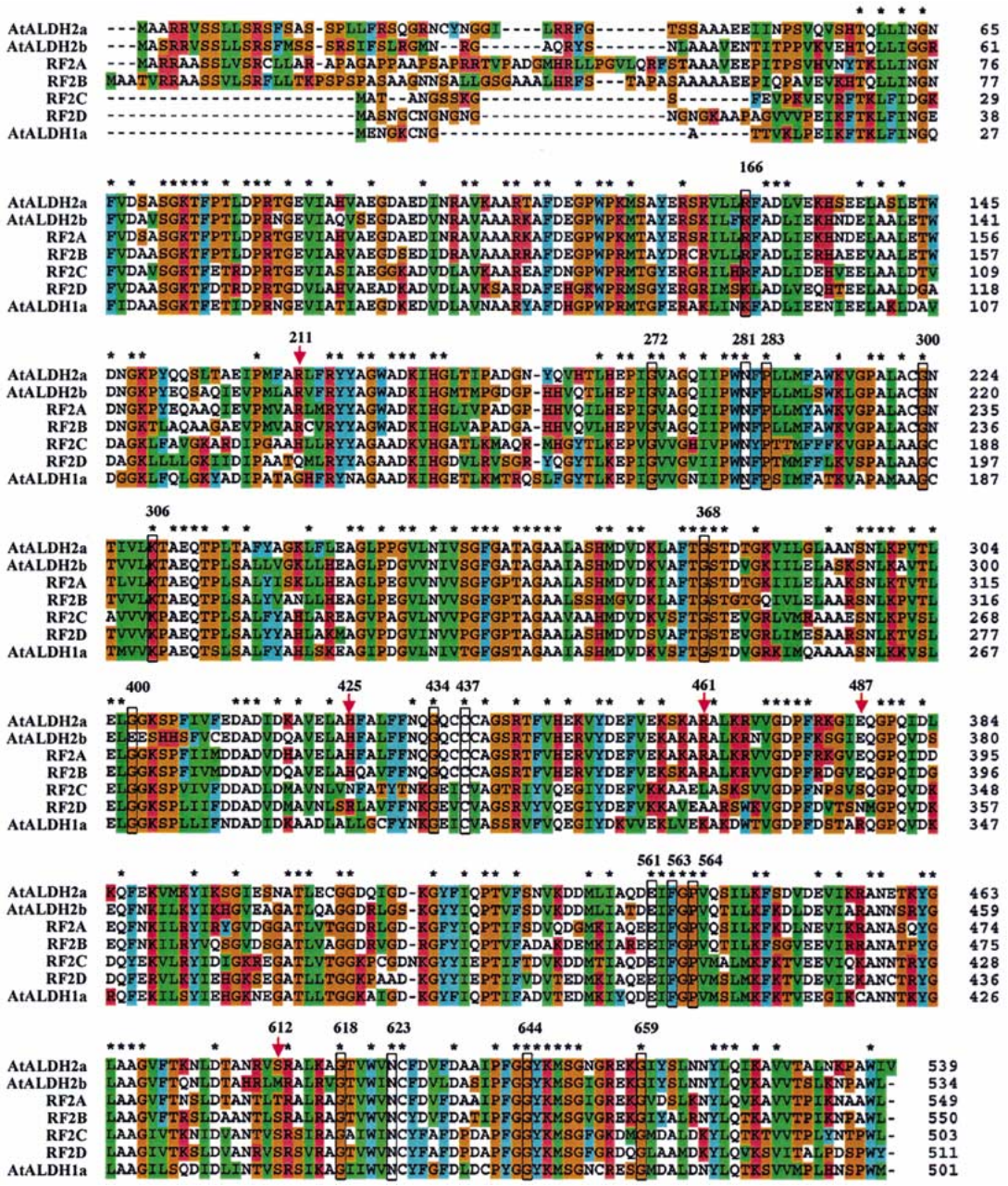

Figure 6. ALDH gene family protein alignment. Asterisks above the alignment represent identical residues for the corresponding ALDH index position. Hyphens designate gaps introduced to optimize alignments. Boxes around the residues represent positions conserved in >95% of analyzed ALDH sequences (Perozich *et al*., 1999). Arrows indicate positions predicted to be responsible for functional differences between plant mtALDHs and cALDHs. Residue colors follow the assignment of the PAM 250 Matrix: orange, hydrophilic (G, P, S, T); red, basic (H, K, R); blue, aromatic (F, W, Y); green, hydrophobic (I, L, M, V); no color, acid-amide and other (A, C, D, E, N, Q).

structure (Hempel *et al*., 1997). None of the five ALDHs that contain a non-Gly residue at index position Gly-211 [400] have been shown unequivocally to be enzymatically active. Therefore, it is possible that a Gly-to-Glu substitution could abolish ALDH function. However, the biochemical characterization of AtALDH2b has revealed that this conserved Gly residue is not essential for ALDH function (Figure 3).

Functional differences between mtALDHs and cALDHs

If mtALDHs and cALDHs have different metabolic functions, then the set of amino acids that vary between the two classes of isozymes would be expected to include those residues that confer functional differences. Inspection of a multiple alignment of the four mtALDHs and three cALDHs boxed in Figure 1 revealed that 173 sites are fully conserved across all members. Thirteen sites are fully conserved within but not between the two clades; at five sites the R-group properties are conserved within but not between the two clades; 55 sites are conserved within one clade, but vary in the other; 15 sites are conserved within one clade and differ from conserved R-group properties in the other clade; and 16 sites exhibit conserved R-group properties within one clade but vary in the other clade.

Sites that are responsible for functional differences between mtALDHs and cALDHs are expected to exhibit rates of evolution that are statistically independent of the rate of evolution between members of the two clades. The probability that the evolutionary rate at a particular site is statistically independent between two gene clades is termed θ. Using an algorithm developed by Gu (1999), the values of θ were determined for each residue in the comparison of plant mtALDHs and cALDHs. These analyses established that of 491 residues analyzed, 315 (64%) had values of θ less than 0.13, 120 (24%) had values between 0.13 and 0.18, and 51 (11%) had values between 0.18 and 0.35. Only five exhibited values of θ greater than 0.4 (Arg-175, His-339, Arg-373, Glu-388, Thr-491). This residue numbering system is based on that of RF2A as shown in Figure 6. These specific residues correspond to index numbers 211, 425, 461, 487 and 612 in the alignment of 145 ALDHs (Perozich *et al*., 1999). Arg-175, His-339, Arg-373, and Glu-388 are all conserved in plant mtALDHs but vary in cALDHs. All the cALDHs have a Ser at the position corresponding to the Thr-491 position of RF2A. However, this site varies among mtALDHs.

Discussion

Six ALDH genes were cloned from maize and *Arabidopsis*. Each of the corresponding proteins has been enzymatically characterized and exhibits ALDH activity. A partial explanation for the presence of so many ALDH genes in plant genomes is the need to provide ALDH activity in various subcellular compartments. This need arises because although some aldehydes (e.g. acetaldehyde) are able to move from one subcellular compartment to another, the molecular sizes of others preclude their passive diffusion across membranes. Therefore, organelles that contain pathways that generate aldehydes but that do not contain an ALDH could experience aldehyde-induced damage.

Phylogenetic analyses indicate that the maize and *Arabidopsis* ALDHs cluster into two clades. One clade contains confirmed and predicted mtALDHs, the other predicted cALDHs. In this respect plants are similar to mammals and fungi. These observations are consistent with an ancient duplication that generated the cALDHs and the mtALDHs. However, plant mtALDHs are more similar to plant cALDHs than to mtALDH from mammals and fungi (Figure 1). Hence, it appears that either the gene duplications that generated the mtALDHs and the cALDH occurred independently within each lineage after the divergence of plants, animals and fungi or that gene conversion-like events that homogenized the mtALDH and cALDH sequences have occurred independently within each lineage.

All studied plant genomes (maize, rice, *Arabidopsis* and tobacco) contain two confirmed or predicted mtALDHs (Figure 1 and data from op den Camp and Kuhlemeier, 1997). If these mtALDHs arose via a single gene duplication that preceded the divergence of monocots and dicots, it would be expected that there would be more mtALDH sequence divergence within species than among species. However, the maize and rice mtALDHs are more similar to each other than to the *Arabidopsis* and tobacco mtALDHs, suggesting that the mtALDH genes arose via independent duplications after the divergence of monocots and dicots or that independent gene conversion-like events have homogenized the mtALDHs in the monocot and dicot lineages. The observation that the maize and rice ALDHs do not cluster by species suggests that this duplication occurred prior to the speciation of maize and rice. This result is consistent with the finding that the two maize mtALDHs do not map to regions of the maize genome that were duplicated during the seg-

mental allotetraploidization event. Because DNA gel blot experiments do not provide any evidence for the presence of additional sequences in the maize genome that have a high degree of sequence similarity to *rf2a* or *rf2b* (data not shown), it can be concluded that either the mtALDH genes did not participate in the segmental allotetraploidization of the maize genome or that any duplications created via this process were subsequently lost.

Maize contains multiple cALDH genes (*rf2c, rf2d1*, and *rf2d2*). The two of these genes for which there is evidence of expression (*rf2c* and *rf2d1*) are very closely linked on the long arm of chromosome 3. This mapping result suggests that the duplication of these cALDH genes occurred via a process unrelated to the segmental allotetraploidization event. Indeed, it appears that this tandem duplication preceded the divergence of maize and rice because two rice ALDH ESTs map to the region of the rice genome (Causse *et al*., 1994) that is syntentic with the long arm of the maize chromosome 3 (Gale and Devos, 1998). One of these ESTs appears to have been derived from the *OsALDH1a* gene described by Li *et al*. (2000); the other appears to be a close relative of *OsALDH1a*. The existence of an *OsALDH1a* relative is consistent with hybridization data from Li *et al*. (2000).

In addition to tandem duplications, there is also evidence that the maize cALDH genes participated in the segmental allotetraploidization event. The second copy of *rf2d* is located on a region of the genome (the long arm of chromosome 8) that is known to be syntenic to the long arm of chromosome 3. As described above, the tandem duplication on the long arm of chromosome 3 appears to have arisen prior to the speciation of maize and rice. After the segmental allotetraploidization event that presumably duplicated both of the closely linked cALDH genes from the long arm of chromosome 3, the duplicate copy of *rf2c* on chromosome 8 was apparently lost from the genome.

Gene duplications are thought to provide the raw material for functional innovations during evolution (Henikoff, 1997). For example, gene duplication permits one copy to maintain its original function, while the other copy is free to accumulate mutations that confer new functions (Ohno, 1970; Li, 1983). Indeed, unless this type of functional divergence occurs it is likely that over time all but one gene copy will be silenced by random mutations. Hence, that this inactivation has not occurred extensively among plant ALDHs provides strong evidence for the existence of functional divergence among members of these gene families.

The nature of this functional divergence is not known. One possibility is that the plant ALDHs encode enzymes with the same biochemical functions, but that accumulate at different times or places within the plant. In the absence of detailed expression data it is not currently possible to address this possibility. One of the ways that we are testing this hypothesis is by determining whether constructs that contain the regulatory regions of *rf2a* in combination with the coding regions of *rf2b*, *rf2c* and *rf2d* can complement the *rf2a* mutation in transgenic maize.

An alternative, but not mutually exclusive, possibility is that the ALDH genes have accumulated polymorphisms in their coding regions that confer functional differences. In this context, functionality can refer to any number of features, including pH optimum, K_m, V_{max}, and K_{cat} that can influence metabolic function.

Mammalian mtALDHs and cALDHs are known to exhibit significant differences in K_m and V_{max} for a variety of substrates (Klyosov, 1996). Inspection of a multiple alignment revealed multiple sites that were better conserved in mtALDHs than in cALDHs, or vice versa, suggesting that plant mtALDHs and cALDHs might also exhibit different properties. Gu (1999) has developed an algorithm that can identify residues likely to be responsible for such functional differences. This algorithm is based on the expectation that residues that confer functional differences will experience different rates of evolution among clades, θ, than those sites that do not.

Comparisons of plant mtALDHs and cALDHs revealed five residues (Arg-175, His-339, Arg-373, Glu-388, and Thr-491) that exhibited high values of θ. Since the crystal structure of a mtALDH has been solved (Steinmetz *et al*., 1997) and used to thread the RF2A protein (Liu *et al*., 2001), the positions of these residues can be established in three-dimensional space. Four of these high-θ residues (Arg-175, His-339, Arg-373, Glu-388) are conserved among plant mtALDHs and are located around the periphery of the active-site pocket on the mtALDH structure (Figure 7). Because these residues are bulky and/or charged, polymorphisms at these sites are likely to affect substrate-binding characteristics. Hence, these data provide computational support for the hypothesis that, like mammalian ALDHs, plant mtALDHs and cALDHs exhibit functional differences. We are

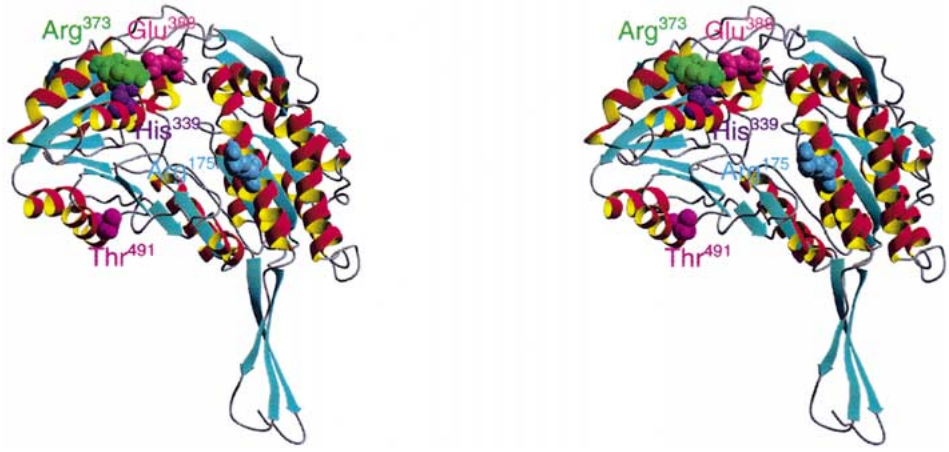

Figure 7. Residues predicted to be responsible for functional differences between mtALDHs and cALDHs. This three-dimensional structure of RF2A was predicted using Swiss-Model (Guex and Peitsch, 1997) with bovine mtALDH (Protein Data Bank number 1AG8) as a template. The stereo images were prepared using MOLMOL (Koradi *et al*., 1996). Residues Arg-175, His-339, Arg-373, Glu-388, and Thr-491 in the RF2A sequence are equivalent to amino acid index numbers 211, 425, 461, 487 and 612 (Perozich *et al*., 1999), respectively. 1AG8 is a homotetramer; only one subunit is illustrated here.

currently testing this hypothesis via biochemical experiments.

Acknowledgments

We thank Tim Fox of Dr Marc Albertson's research group (Pioneer Hi-Bred International, Inc., Johnston, IA) for assistance in screening the Pioneer EST database; Dr Juan Aguilar (Department of Biochemistry, School of Pharmacy, University of Barcelona, Spain) for *E. coli* strain JA111 and pALD9; Drs Alejandro Calderon and Steven Dellaporta (Yale University) for the W22 tassel cDNA library; Drs Monica Frey and Alfons Gierl (Technische Universität München, Garching, Germany) for the C131A 'root plus shoot' cDNA library; Dr Xun Gu (Iowa State University) for assistance with the application of his algorithm; Dr Ben Burr (Brookhaven National Laboratory) for analysis of RI mapping data; and Dr John Tossberg (Pioneer Hi-Bred International) for the B73 genomic DNA library. This work was supported in part by competitive grants from the United States Department of Agriculture National Research Initiative Program (awards 98101805 and 0001478), the Human Frontiers in Science Program (award RG0067), and the Consortium for Plant Biotechnology Research with matching funds from Calgene and Renessen. M.D.Y. was supported by a USDA National Needs Fellowship (award 9706042). This is journal paper 19220 of the Iowa Agriculture and Home Economics Experiment Station (Ames, IA), projects 3390, 3409, and 3554, supported by the Hatch Act and State of Iowa funds.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403–410.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796–815.

Asker, H. and Davies, D.D. 1985. Mitochondrial aldehyde dehydrogenases from higher plants. Phytochemistry 24: 689–693.

Boronat, A. and Aguilar, J. 1979. Rhamnose-induced propanediol oxidoreductase in *Escherichia coli*: purification, properties, and comparison with the fucose-induced enzyme. J. Bact. 140: 320–326.

Burr, B., Burr, F.A., Thompson, K.H., Albertson, M.C. and Stuber, C.W. 1988. Gene mapping with recombinant inbreds in maize. Genetics 118: 519–526.

Caballero, E., Baldoma, L., Ros, J., Boronat, A. and Aguilar, J. 1983. Identification of lactaldehyde dehydrogenase and glycolaldehyde dehydrogenase as functions of the same protein in *Escherichia coli*. J. Biol. Chem. 258: 7788–7792.

Causse, M.A., Fulton, T.M., Cho, Y.G., Ahn, S.N., Chunwongse, J., Wu, K., Xiao, J., Yu, Z., Ronald, P.C., Harrington, S.E., Second, G., McCouch, S.R. and Tanksley, S.D. 1994. Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 138: 1251–1274.

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.

Duvick, D.N. 1965. Cytoplasmic pollen sterility in corn. Adv. Genet. 13: 1–56.

Gale, M.D. and Devos, K.M. 1998. Comparative genetics in the grasses. Proc. Natl. Acad. Sci. USA 95: 1971–1974.

Gaut, B.S. and Doebley, J.F. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. Proc. Natl. Acad. Sci. 94: 6809–6814.

Gu, X. 1999. Statistical methods for testing functional divergence after gene duplication. Mol. Biol. Evol. 16: 1664–1674.

Guex, N. and Peitsch, M.C. 1997. Swiss-Model and the Swiss-PdbViewer: an environment for comparing protein modeling. Electrophoresis 18: 2714–2723.

Hempel, J., Liu, Z.-J., Perozich, J, Rose, J., Lindahl, R. and Wang, B.-C. 1997. Conserved residues in the aldehyde dehydrogenase family: locations in the class 3 tertiary structure. Adv. Exp. Med. Biol. 414: 9–13.

Henikoff, S., Greene, E.A., Pietrokovski, S., Bork, P., Attwood, T.K. and Hood, L. 1997. Gene families: the taxonomy of protein paralogs and chimeras. Science 278: 609–614.

Hidalgo, E., Chen, Y.M., Lin, E.C. and Aguilar, J. 1991. Molecular cloning and DNA sequencing of the *Escherichia coli* K-12 ald gene encoding aldehyde dehydrogenase. J. Bact. 173: 6118–6123.

Klyosov, A.A. 1996. Kinetics and specificity of human liver aldehyde dehydrogenase toward aliphatic, aromatic, and fused polycyclic aldehydes. Biochemistry 35: 4457–4467.

Koradi, R., Billeter, M. and Withrich, K. 1996. MOLMOL: A program for display and analysis of macromolecular structures. J. Mol. Graph. 14: 51–55.

Laughnan, J.R. and Gabay-Laughnan, S. 1983. Cytoplasmic male sterility in maize. Annu. Rev. Genet. 17: 27–48.

Li, W.H. 1983. Evolution of duplicated genes. In: M. Nei and R.K. Koehn (Eds.) Evolution of Genes and Proteins, Sinauer, Sunderland, MA.

Li, Y., Nakazono, M., Tsutsumi, N., and Hirai, A. 2000. Molecular and cellular characterization of a cDNA clone encoding a novel isozyme of aldehyde dehydrogenase from rice. Gene 249: 67–74.

Lindahl, R. 1992. Aldehyde dehydrogenases and their role in carcinogenesis. Crit. Rev. Biochem. Mol. Biol. 27: 283–355.

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.

Liu, Z.-J., Sun, Y.-J., Rose, J., Chung, Y.-J., Hsiao, C.-D., Chang, W.-R., Kuo, I., Perozich, J., Lindahl, R., Hempel, J. and Wang, B.-C. 1997. The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. Nature Struct. Biol. 4: 317–326.

Marumo, S. 1986. Auxins. In: N. Takahashi (Ed.) Chemistry of Plant Hormones, CRC Press, Boca Raton, FL, pp. 9–56.

Moore, G. 2000. Cereal chromosome structure, evolution, and pairing. Annu. Rev. Plant Physiol. Plant Mol. Biol 51: 195–222.

Moore, S.A., Baker, H.M., Blythe, T.J., Kitson, K.E., Kitson, T.M. and Baker, E.N. 1998. Sheep liver cytosolic aldehyde dehydrogenases: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases. Structure 6: 1541–1651.

Nakai, K. and Kanehisa, M. 1992. A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics 14: 897–911.

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.

Ohno, S. 1970. Evolution by Gene Duplication. Springer-Verlag, Berlin.

op den Camp, R.G. and Kuhlemeier, C. 1997. Aldehyde dehydrogenase in tobacco pollen. Plant Mol. Biol. 35: 355–365.

Perozich, J., Nicholas, H., Wang, B.C., Lindahl, R. and Hempel, J. 1999. Relationships within the aldehyde dehydrogenase extended family. Protein Sci. 8: 137–146.

Schauenstein, E., Esterbauer H. and Zollner, H. 1977. Aldehydes in Biological Systems: Their Natural Occurrence and Biological Activities. Pion, London.

Schnable, P.S. and Wise, R.P. 1998. The molecular basis of cytoplasmic male sterility. Trends Plant Sci. 3: 175–180.

Sjöling, O. and Glaser E. 1998. Mitochondrial targeting peptide in plants. Trends Plant Sci. 3: 136–140.

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.

Swofford, D.L. 2000. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, MA.

Tadege, M. and Kuhlemeier C. 1997. Aerobic fermentation during tobacco pollen development. Plant Mol. Biol. 35: 343–354.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl. Acids Res. 24: 4876–4882.

Vasiliou, V., Bairoch, A., Tipton, K.F. and Nebert, D.W. 1999. Eukaryotic aldehyde dehydrogenase (ALDH) genes: human polymophisms, and recommended nomenclaturue based on divergent evolution and chromosomal mapping. Pharmacogenetics 9: 421–434.

Warmke, H.E. and Lee, S.-L.J. 1977. Mitochondrial degeneration in Texas cytoplasmic male-sterile corn anthers. J. Hered. 68: 213–222.

Wightman, F. and Cohen, D. 1968. Intermediary steps in the enzymatic conversion of tryptophan to IAA in cell-free systems from mung bean seedlings. In: F. Wightman and G. Setterfield (Eds.) Biochemistry and Physiology of Plant Growth Substances (Proceedings of the 6th International Conference on Plant Growth Substances), Runge Press, Ottawa, p. 273.

Wilson, W.A., Harrington, S.E., Woodman, W.L., Lee, M., Sorrells, M.E. and McCouch, S.R. 1999. Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated panicoids. Genetics 153: 453–473.

Wise R.P. and Schnable, P.S. 1994. Mapping complementary genes in maize: positioning the *rf1* and *rf2* nuclear restorers relative to visible and RFLP markers. Theor. Appl. Genet. 88: 785–795.

Wise, R.P., Pring D.R. and Gengenbach, B.G. 1987. Mutation to male fertility and toxin insensitivity in T-cytoplasm maize is associated with a frameshift in a mitochondrial open reading frame. Proc. Natl. Acad. Sci. USA 84: 2858–2862.

Wise, R.P., Bronson, C., Schnable, P.S. and Horner, H.T. 1999. T-cytoplasmic male sterility of maize. Adv. Agron. 65: 79–130.

Yoshida, A., Rzhetsky, A., Hsu, L.C. and Chang, C. 1998. Human aldehyde dehydrogenase gene family. Eur. J. Biochem. 251: 549–557.