Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 1408–1412, March 1986
Genetics

Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene

(gene structure/enzyme induction/intron/evolution)

CAREN CHANG AND ELLIOT M. MEYEROWITZ

Division of Biology, California Institute of Technology, Pasadena, CA 91125

Communicated by Ray D. Owen, October 18, 1985

ABSTRACT **_Arabidopsis thaliana_ provides an excellent experimental plant system for molecular genetics because of its remarkably small genome size, near absence of dispersed middle repetitive DNA, and short life cycle. We have cloned and determined the nucleotide sequence of a single-copy gene from _A. thaliana_ likely to be the gene encoding alcohol dehydrogenase (ADH; alcohol:NAD$^+$ oxidoreductase, EC 1.1.1.1). The gene was isolated from a random recombinant library by cross-hybridization with a maize _Adh1_ gene probe. The DNA sequence contains an open reading frame capable of encoding a polypeptide the same length as maize ADH1 and ADH2 (379 amino acids) and having ≈80% homology with both maize enzymes. This open reading frame is interrupted by six introns whose positions are conserved with six of the nine intron positions present in both maize genes. The 5′ and 3′ untranslated regions are, respectively, 58 and 204 base pairs long. Sequences important for eukaryotic gene expression such as the TATA box, polyadenylylation signal, and intron splice-site sequences are found in the expected locations. The gene hybridizes to a specific anaerobically induced RNA in _Arabidopsis_ whose appearance correlates with the anaerobic induction of _Arabidopsis_ ADH protein.**

Alcohol dehydrogenase (ADH; alcohol:NAD$^+$ oxidoreductase, EC 1.1.1.1) is an easily assayed enzyme whose activity has been observed in numerous higher plants including *Arabidopsis*, maize, pearl millet, sunflower, wheat, and pea (1, 2). Most plants have two or three isozymes of ADH, which exist as both hetero- and homodimers in various organs (1). The enzyme is presumably required by plants for NADH metabolism, via reduction of acetaldehyde to ethanol, during periods of anaerobic stress. High levels of ADH activity are found in dry seeds (3, 4) and in anaerobically treated seeds (5, 6), roots (5), and shoots (7).

The most extensive study of a plant ADH system has been in maize (8) from which both *Adh* genes, *Adh1* and *Adh2*, have been cloned and sequenced (9–11). The coding sequences of these genes are 82% homologous, interrupted by nine identically positioned introns that differ in sequence and length. ADH1 and ADH2 belong to a small group of proteins in maize primary root that are selectively translated in response to anaerobiosis (12); the increased levels of ADH are due to induction of Adh mRNA (9–11, 13).

Arabidopsis ADH is similar to the maize ADHs, although genetic experiments indicate only one *Adh* locus in *Arabidopsis* (14). Examination of *Arabidopsis* ADH in crude extracts has shown that the enzyme behaves as a homodimer of M_r 87,000 (14), close to the maize ADH M_r of ≈80,000 (15). ADH is induced anaerobically in *Arabidopsis* (16) as in maize. ADH is also induced in both maize root and *Arabidopsis* callus by the synthetic auxin 2,4-dichlorophenoxyacetic acid (16, 17), and for *Arabidopsis* this has been shown to be the result of *de novo* synthesis of a poly(A) mRNA (16).

Arabidopsis ADH has potential as a biochemical marker for genetic transformation of *Arabidopsis*: null mutations exist, and ADH is easily induced and assayed. We present here the molecular cloning and characterization of an inducible single-copy gene from *Arabidopsis* likely to be the gene encoding ADH.

MATERIALS AND METHODS

***Arabidopsis* Strains.** The Landsberg *erecta* strain of *A. thaliana* was obtained from F. J. Braaksma (Department of Genetics, Biology Centre, Haren, The Netherlands); the Bensheim strain was obtained from A. R. Kranz (Botanisches Institut, J. W. Goethe-Universität, Frankfurt am Main, Federal Republic of Germany).

General Nucleic Acid Methods. *Arabidopsis* DNA was prepared from whole plants as described in ref. 18. Library construction followed the procedure in ref. 19. The clone nomenclature system is described in ref. 20 with the following additions: f for λ EMBL4 (21); b for pMT21, a 1.9-kilobase-pair (kbp) pBR322 derivative (H. V. Huang, personal communication); j and k, respectively, for pSP65 and pSP64 (Promega Biotec, Madison, WI) containing the SP6 RNA polymerase promoter. Radiolabeled DNA probes were produced by nick-translation (22). Hybridizations with the maize probe were in 50% formamide/5× SSPE (5× SSPE is 5 mM Na_2EDTA/40 mM NaOH/50 mM $NaH_2PO_4 \cdot H_2O$/900 mM NaCl) at 37°C with washes in 0.05× SSPE at 37°C. Genome blot hybridizations were in 50% formamide/5× SSPE at 43°C, with washes in 1× SSPE at room temperature. All DNA manipulations were carried out as described in ref. 23. The DNA sequence was determined by the method of Maxam and Gilbert (24).

Anaerobic Treatment. Seeds were treated for 1–2 days at 4°C in Petri dishes containing distilled H_2O-soaked filter paper (Whatman 3). The seeds were germinated on fresh filter paper on 0.7% agar plates at 25°C with constant illumination (7000 lx). After 3–5 days, seedlings were transferred into 1–2 ml of distilled H_2O (60–100 seedlings per ml). Untreated seedlings were left on plates. At various times, seedlings were placed on filter paper in a Büchner funnel with suction to remove excess water. They were then used for either RNA or protein preparations.

RNA Analysis. Seedlings (60–100) were homogenized for 2 min in a 0.5-ml microtube fitted onto a mini-BeadBeater (Biospec Products). Each tube contained 20 μl of 50% phenol/0.2 M Tris, pH 7.7/10 mM NaCl/75 mg of 0.5-mm zirconium oxide beads (Biospec Products). After homogenization, an additional 20 μl of buffer, 10 μl of 10% $NaDodSO_4$, and 2 μl of yeast tRNA (10 mg/ml) were mixed in by

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ADH, alcohol dehydrogenase; kbp, kilobase pair(s).

vortexing. The contents were extracted three times with an equal volume of phenol/chloroform/isoamyl alcohol (50:49.5:0.5; vol/vol), twice with chloroform/isoamyl alcohol (99:1; vol/vol), once with ether, and then precipitated with ethanol.

For blot analysis, the RNA was glyoxylated (25), subjected to electrophoresis in agarose gels and transferred to nitrocellulose (23). Radiolabeled RNA probes were synthesized by using SP6 RNA polymerase (26). RNA induction was quantitated by liquid scintillation counting of bands excised from the blots.

Protein Analysis. Seedlings (60–100 per 0.5-ml microtube) were homogenized for 2 min in the mini-BeadBeater. Each tube contained 5.0 μl of 70 mM Tris·HCl, pH 7.2/25% (vol/vol) glycerol/0.8% (vol/vol) 2-mercaptoethanol/0.25% (wt/vol) bromophenol blue/75 mg of 0.5-mm zirconium oxide beads. A hole was punched in the bottom of each tube with a 26-gauge needle and the homogenates were spun into 1.5-ml microtubes. The resulting supernatants were loaded onto nondenaturing polyacrylamide gels consisting of 7.3% (wt/vol) acrylamide/0.2% (wt/vol) *N,N'*-methylenebisacrylamide/70 mM Tris·HCl, pH 7.2/0.8% (wt/vol) ammonium persulfate/0.08% (vol/vol) *N,N,N',N'*-tetramethylethylenediamine. The gels were run at 10 mA at 22°C with recirculated 8.25 mM Tris/30 mM diethylbarbituric acid electrode buffer (pH 7.4). Gels were stained for ADH activity at 37°C in 0.1 M Tris·HCl, pH 7.6/1.5 mM NAD^+/0.25 mM nitroblue tetrazolium/0.26 mM phenazine methosulfate/0.5% (vol/vol) ethanol. Ethanol was omitted in substrate-dependence controls.

5′ and 3′ RNA Mapping. The 5′ end was mapped by primer extension (27) using induced RNA and a synthetic oligonucleotide primer complementary to the RNA sequence at the 3′ end of the believed first exon. The 3′ end was mapped by ribonuclease protection (28) using a gel-isolated *in vitro*-synthesized complementary strand RNA probe, ≈645 bases long, and having one terminus within the last intron.

RESULTS

Isolation of *Arabidopsis* DNA Clones. Four genome equivalents of an *Arabidopsis* DNA library (described in ref. 18) were screened for cross-hybridization with a maize *Adh1* gene fragment probe (kindly provided by M. Freeling). Four positive clones were detected; restriction digests revealed that all four were identical, each containing the same 4.9-kbp *Eco*RI restriction fragment of *Arabidopsis* DNA (the At3001 fragment; Fig. 1*a*) and a 6.6-kbp *Eco*RI stuffer fragment of the λ vector.

To obtain larger clone segments, a new library was constructed consisting of Landsberg *erecta* strain DNA that had been partially digested with *Mbo* I and ligated into *Bam*HI-digested λEMBL4. After amplification of the library, four genome equivalents were screened with nick-translated bAt3001 (a plasmid containing the At3001 fragment). Three positive clones were detected; two contained 15.6-kbp inserts that were identically oriented and indistinguishable by restriction mapping (λfAt3102), and the third contained a 17.2-kbp insert in the opposite orientation (λfAt3101) (Fig. 1*a*).

Genome blots verified that these clones represent the sequence organization present in genomic DNA. Seven different restriction digests of *Arabidopsis* DNA were probed with nick-translated bAt3001 and jAt3011 (a plasmid containing the 2.5-kbp *Sac* I/*Hin*dIII At3011 fragment of λfAt3102, Fig. 1*b*). The hybridized restriction fragments completely agreed with the map produced by the λ clones, suggesting that the region of DNA spanned by the clones exists in a single copy in the *Arabidopsis* genome (data not shown).

The maize *Adh1* gene fragment cross-hybridized with each of the λ clones as expected; the region of hybridization is

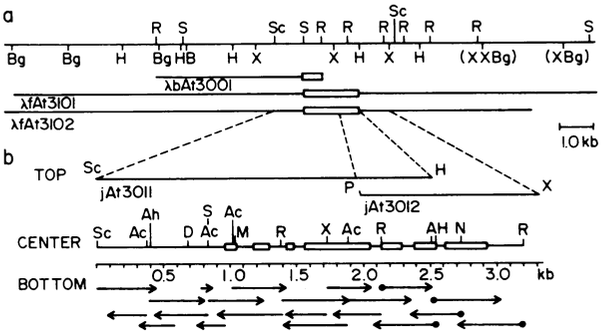

FIG. 1. (*a*) Summary of isolated λ clones. Lines represent the lengths and positions of the three independently cloned inserts relative to the restriction map above. Boxes indicate regions of the clones that hybridize with the maize *Adh1* 2.3-kbp *Hin*dIII gene fragment. (Relative positions have not been verified for sites enclosed in parentheses. Three unmapped *Hin*dIII sites are not shown.) (*b*) DNA sequencing strategy. Top, lines represent the two sequenced subclones, derived from λfAt3102 as indicated by dashed lines. Center, restriction map of sites end-labeled for sequence analysis showing the entire gene structure with boxes representing exons; the gene is shown such that the transcription direction is left to right (determined as in Fig. 2 *a–c*). Bottom, arrows designate the direction and extent of DNA sequence obtained for each fragment. Arrows having a dot denote sequence data from jAt3012; remaining data are from jAt3011. A, *Ava* I; Ac, *Acc* I; Ah, *Aha* III; B, *Bam*HI; Bg, *Bgl* II; D, *Dde* I; H, *Hin*dIII; M, *Msp* I; N, *Nco* I; P, *Pst* I; R, *Eco*RI; S, *Sal* I; Sc, *Sac* I; X, *Xba* I.

shown in Fig. 1*a*. The cloned portion of DNA includes at least 7 kbp on either side of this region.

RNA Induction Correlates with ADH Protein Induction. The fact that ADH activity increases in anaerobically treated *Arabidopsis* plantlets suggested that anaerobically treated plantlets should contain Adh mRNA to which our *Arabidopsis* clones might hybridize. RNA blots were probed with single-stranded RNA derived from SP6 RNA polymerase transcription of *Hin*dIII-linearized jAt3011 and *Sac*

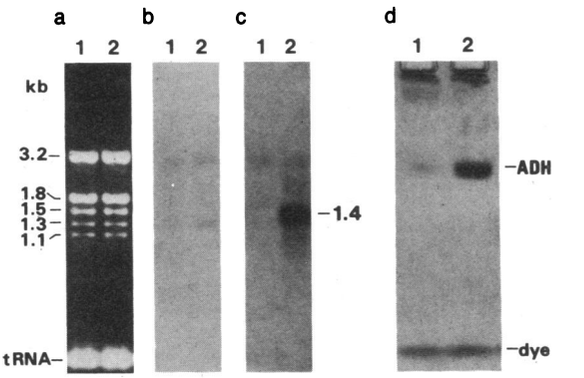

FIG. 2. RNA induction, showing direction of transcription, and ADH protein induction. Lanes 1 and 2, 0 hr and 4 hr of anaerobiosis, respectively. (*a*) Ethidium bromide-stained half of agarose gel containing 10% of total RNA prepared from Bensheim strain plantlets shows similar amounts of RNA for lanes 1 and 2, assessed by the intensity of predominant rRNA bands. (*b*) Autoradiogram of blot of other half of gel, containing 90% of the RNA preparations, probed with a single-stranded RNA probe (described in text) from jAt3011. Little or no specific hybridization is detected. (*c*) Autoradiogram of the same blot hybridized with a single-stranded RNA probe from kAt3011 (after washing off previous probe) shows induction of an RNA in lane 2. An overexposure is shown here to emphasize the low level of hybridization in lane 1. Hybridization conditions were made as close as possible to those of the first hybridization. (*d*) Native protein gel stained for ADH activity shows induction of ADH. Nonspecific protein staining just below the wells indicates that roughly equal amounts of protein were loaded onto both lanes.

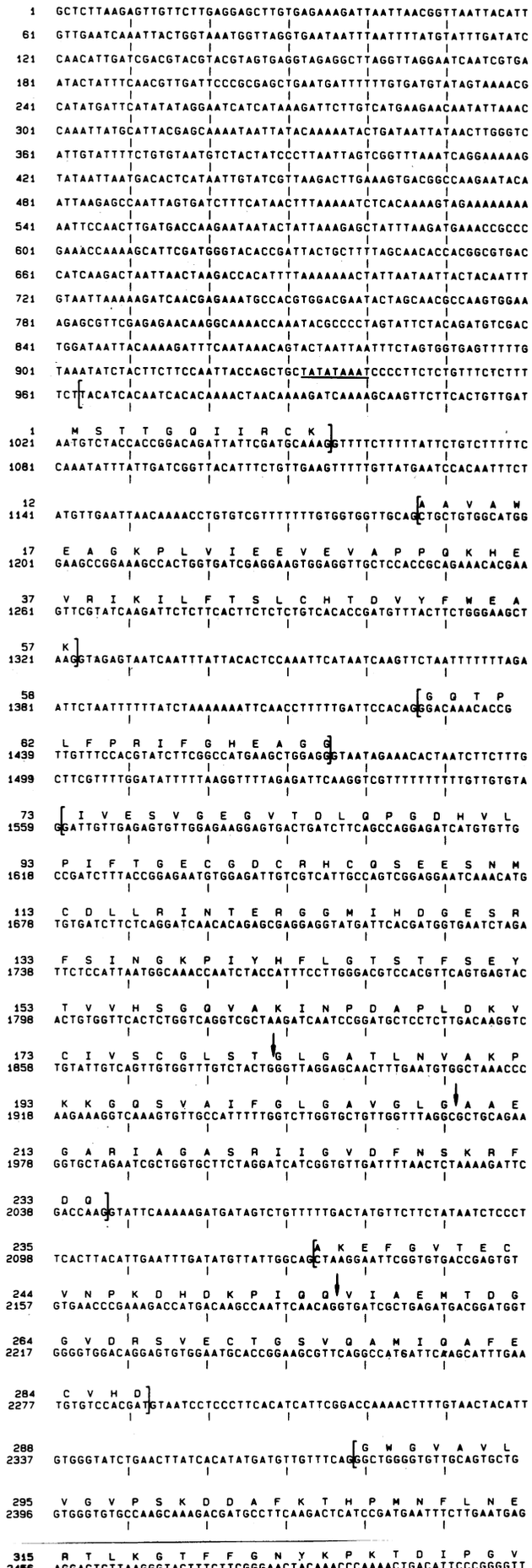

FIG. 3. Complete DNA sequence of the *Adh* gene including introns and flanking regions. The derived amino acid sequence is shown in the one-letter code above the DNA sequence and is numbered separately. Exons are bracketed. Arrows indicate intron positions that are present in the maize *Adh* genes but absent in *Arabidopsis Adh*. The TATA box is noted with a solid underline and the putative polyadenylylation signal is noted with a dotted underline.

I-linearized kAt3011 templates. (kAt3011 is identical to jAt3011 except for the opposite orientation of the At3011 fragment with respect to the SP6 promoter.) These experiments revealed a homologous RNA, ≈1450 nucleotides long, that clearly increased upon anaerobiosis. The induction for Landsberg *erecta* strain plantlets at 2 hr and 4 hr of anaerobiosis was ≈5-fold and 10-fold, respectively. Hybridization to the induced RNA occurred only with the SP6 transcripts from kAt3011 (not from jAt3011), revealing that the *in vivo* transcription direction of the gene, as depicted in Fig. 1, is from left to right (Fig. 2).

Examination of protein from plantlets treated in parallel with those used in the RNA experiments demonstrates that the induction of RNA correlates with induction of ADH activity. Protein extracts were subjected to electrophoresis in nondenaturing polyacrylamide gels and stained for ADH activity. Two bands normally appeared; we identified one of these bands as arising from ADH because of its substrate specificity as well as its shifted migration when *Arabidopsis* ADH electrophoretic variants (14) were used (data not shown). The ADH band showed a marked increase in intensity after anaerobic treatment (Fig. 2*d*).

DNA Sequencing. We determined the DNA sequence of a 3.2-kbp segment (Fig. 1*b*) that contains the region that had hybridized with the induced RNA. The sequence is presented in Fig. 3. Extensive homology was discovered between the *Arabidopsis* gene and the protein coding sequence of both maize *Adh* genes. The homology is interrupted by six probable introns in the *Arabidopsis* gene; although their lengths and sequences differ from the corresponding maize introns, their positions are coincident with six of the nine intron positions present in both of the maize *Adh* genes (Fig. 4). All six begin with the dinucleotide GT and end with the dinucleotide AG, as seen consistently at the intron/exon junctions of eukaryotic genes (29). The average A+T content of these regions is 71%, while the A+T content of the open reading frame is 52%. Translation of the DNA sequence in all possible reading frames produces an open reading frame only when all six of the putative intron sequences are omitted.

The deduced polypeptide encoded by the 1137-nucleotide open reading frame (Fig. 3) shows conservation with both maize ADH sequences: it is of identical length and has 80.5% identity with maize ADH1 and 79% identity with maize ADH2. (The maize ADH proteins are 87% conserved.) The

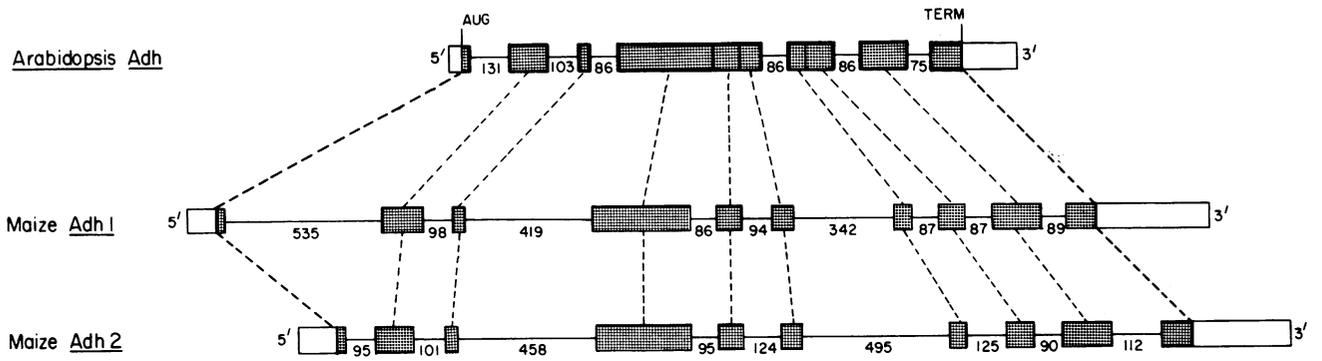

FIG. 4. Comparison of exon/intron structure of *Arabidopsis Adh* and maize *Adh* (11) genes. Boxes represent relative lengths and positions of exons; shaded portions designate protein-coding regions of exons. Solid lines represent relative lengths and positions of introns that are at identical positions in all three genes with respect to the protein sequence; the number below each line gives the length of the intron in nucleotides. Dashed lines connect homologous exons.

nucleic acid sequence coding for this polypeptide has 73% homology with maize *Adh1* and 72% homology with maize *Adh2*. (The maize *Adh* coding regions are 82% conserved.) Of the nucleotide differences with *Adh1*, 53% are silent at the amino acid level and 17% result in conservative amino acid substitutions.

5′ and 3′ RNA Mapping. Primer extension on the RNA resulted in a DNA transcript 58 ± 1 nucleotides beyond the translational initiation codon (data not shown). The position of the initiation codon was determined by homology with the deduced protein sequences of both maize enzymes. Neither the upstream DNA (1021 nucleotides) nor the 5′-untranslated RNA contains an alternative initiation codon that would allow read through. Ribonuclease protection mapping of the 3′ end gave a length of 204 ± 2 bases for the 3′-untranslated RNA using DNA ladder size standards calibrated for RNA (data not shown). The actual length may be as low as 200 nucleotides since the 2–4 terminal adenosine nucleotides of the protected probe are resistant to the ribonucleases used.

DISCUSSION

Sequence analysis indicates that we have cloned the *Arabidopsis Adh* gene. Our clone is single copy in the genome supporting genetic evidence of there being a single *Adh* locus in *Arabidopsis* (14). The calculated molecular weight of the deduced polypeptide is 41,200, which is close to the value 44,000 measured by the migration of *Arabidopsis* ADH monomers in polyacrylamide gels (14). The length of the anaerobically induced RNA is consistent with the predicted exon sizes of the gene, allowing for a poly(A) tail of ≈50 nucleotides.

The gene contains sequences characteristic of expressed eukaryotic genes. A TATA box is located 23 base pairs (bp) upstream of the mapped 5′ end of the RNA, consistent with other genes expressed in plants (30). A possible polyadenylylation signal, AATATAAA, similar to the plant consensus signal (30, 31), is situated 19 bases upstream of the mapped 3′ end. An AGGA box, possibly involved in transcription regulation, is occasionally found 36–59 bp upstream of the TATA box of plant genes (30). Such a sequence for the *Arabidopsis* gene might be TAAACAGTACT, which is similar to the plant consensus sequence $\begin{matrix}C\\T\end{matrix}A_{2-5}\begin{matrix}G\\T\end{matrix}NGA_{2-4}\begin{matrix}CC\\TT\end{matrix}$ and is located 57 bp upstream of the putative TATA box.

Upstream sequences may be required for ADH expression during anaerobiosis or 2,4-dichlorophenoxyacetic acid induction. Although the DNA sequences upstream of the coding sequences have generally diverged, distinct homologous stretches exist 5′ of the translation start sites of the *Arabidopsis Adh* and maize *Adh1* genes (Fig. 5). These sequences are not found in maize *Adh2*, with the exception of the TATA box. The three 8-bp sequences shared by the two maize genes in their 5′-untranslated regions (11) are not seen in the 1021 nucleotides upstream of the initiation codon of the *Arabidopsis* gene.

The six introns of the *Arabidopsis* gene possess the 5′ and 3′ splice site consensus sequences of eukaryotic genes (29) (Fig. 6). Recently, internal signals, such as the conserved TACTAAC sequence at 20–55 nucleotides from the 3′ border of yeast introns (32), have been discovered to be required for gene splicing. Each *Arabidopsis Adh* intron contains the consensus sequence $\begin{matrix}C\\T\end{matrix}T\begin{matrix}A\\G\end{matrix}AT$ at 16–39 nucleotides from the 3′ border, which is homologous with the animal consensus sequences found in identical positions in diverse animal gene introns (33) and homologous with the 5-base-long sequences used *in vitro* for formation of the lariat intermediate in intron excision (34). In addition to this consensus sequence, a second highly conserved sequence was found within the *Arabidopsis* introns (Fig. 6); this second consensus sequence may have some role in *Arabidopsis* gene expression.

The deduced *Arabidopsis* ADH polypeptide is 47% conserved with horse liver ADH. Structurally and functionally important residues defined by the tertiary structure of the horse enzyme (35), such as the seven residues that provide ligands for the catalytic and noncatalytic zinc atoms, are conserved, suggesting that the *Arabidopsis* enzyme has a similar structure.

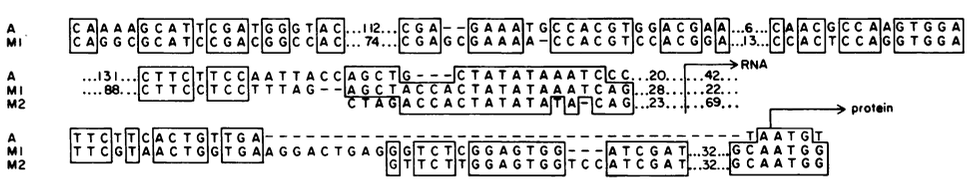

FIG. 5. 5′ sequence homologies for *Arabidopsis Adh* (A), maize *Adh1* (M1), maize *Adh2* (M2). The number of bases between each homologous stretch is given. Contiguous bases, spread apart for alignment purposes, are linked with dashes. The homologies between M1 and M2 are from ref. 11.

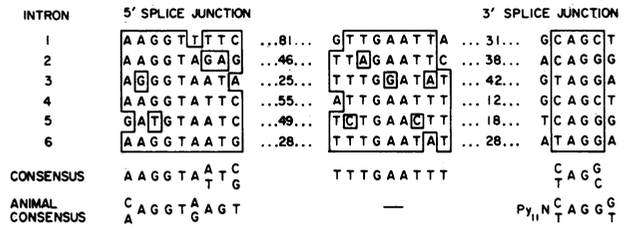

FIG. 6. Intron consensus sequences for the *Arabidopsis Adh* gene. The number of bases between each consensus block is given.

Comparison of the nucleic acid coding sequences and deduced protein sequences of the *Arabidopsis Adh* gene and of the two maize *Adh* genes shows that the two maize genes are more related to each other than either is to the *Arabidopsis* gene. It is thus likely that the maize and *Arabidopsis Adh* genes descended from a single ancestral gene that was subsequently duplicated in the maize but not in the *Arabidopsis* lineage. Further comparison with other angiosperm *Adh* gene sequences, when they become available, should indicate whether the ADH isozymes of other monocots and dicots are likewise descendants of a single gene in their last common ancestor or are the result of more ancient gene duplications, as is generally supposed (1, 11).

The low copy-number sequence content of the maize genome is >20-fold that of the *Arabidopsis* genome (36, 37). This difference is partially reflected in both the size and the copy number of the *Adh* genes of the two species. The *Arabidopsis* gene has three fewer introns than either of the two maize genes; the sum of the intron lengths in the *Arabidopsis* gene is >1000 bases shorter than the comparable sum in the maize *Adh* genes. Other *Arabidopsis* genes have also been seen to be present in fewer copies per haploid genome than the homologous genes in other angiosperms (38).

This work demonstrates the feasibility of isolating a dicot plant gene by homology with a monocot gene. Cross-hybridization is proving to be a useful approach for isolating plant genes. A plant such as *Arabidopsis*, which lends itself well to molecular genetics and cloning, might be exploited for rapid cloning of genes that cross-hybridize with specific genes in other plants (38). Our cloned *Arabidopsis* gene can be tested for the ability to restore ADH function to existing *Arabidopsis* ADH null mutants (14) via Ti plasmid-mediated transformation. Should this approach succeed, we hope to answer questions concerning the regulation of ADH in plants and to develop *Arabidopsis* ADH as a marker in gene fusion experiments.

We are grateful to C. Rice for technical advice and to T. Hunkapiller and C. Martin for DNA sequence analysis programs. We thank M. Garfinkel, P. Mathers, K. Fryxell, and R. Pruitt for comments on the manuscript. This work was supported by Grant PCM-8408504 from the National Science Foundation (E.M.M.). C.C. was supported by National Research Service Award 5 T32GM07616 from the National Institutes of Health.

1. Gottlieb, L. D. (1982) *Science* **216**, 373–380.
2. Leblová, S., Zimáková, I., Barthová, J. & Ehlichová, D. (1971) *Biol. Plant.* **13**, 33–42.
3. Efron, Y. & Schwartz, D. (1973) *Proc. Natl. Acad. Sci. USA* **61**, 586–591.
4. Leblová, S., Zimáková, I., Sofrová, D. & Barthová, J. (1969) *Biol. Plant.* **11**, 417–423.
5. Freeling, M. (1973) *Mol. Gen. Genet.* **127**, 215–227.
6. Banuett-Bourrillon, F. & Hague, D. R. (1979) *Biochem. Genet.* **17**, 537–552.
7. App, A. A. & Meiss, A. N. (1958) *Arch. Biochem. Biophys.* **77**, 181–190.
8. Freeling, M. & Birchler, J. A. (1981) in *Genetic Engineering, Principles and Methods*, eds. Stelow, J. K. & Hollaender, A. (Plenum, New York), Vol. 3, pp. 223–264.
9. Gerlach, W. L., Pryor, A. J., Dennis, E. S., Ferl, R. J., Sachs, M. M. & Peacock, W. J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2981–2985.
10. Dennis, E. S., Gerlach, W. L., Pryor, A. J., Bennetzen, J. L., Inglis, A., Llewellyn, D., Sachs, M. M., Ferl, R. J. & Peacock, W. J. (1984) *Nucleic Acids Res.* **12**, 3983–4000.
11. Dennis, E. S., Sachs, M. M., Gerlach, W. L., Finnegan, E. J. & Peacock, W. J. (1985) *Nucleic Acids Res.* **13**, 727–743.
12. Sachs, M. M., Freeling, M. & Okimoto, R. (1980) *Cell* **20**, 761–767.
13. Ferl, R. J., Brennan, M. D. & Schwartz, D. (1980) *Biochem. Genet.* **18**, 681–691.
14. Dolferus, R. & Jacobs, M. (1984) *Biochem. Genet.* **22**, 817–838.
15. Freeling, M. & Schwartz, D. (1973) *Biochem. Genet.* **8**, 27–36.
16. Dolferus, R., Marbaix, G. & Jacobs, M. (1985) *Mol. Gen. Genet.* **199**, 256–264.
17. Freeling, M. (1973) *Mol. Gen. Genet.* **127**, 215–227.
18. Pruitt, R. E. & Meyerowitz, E. M. (1986) *J. Mol. Biol.*, in press.
19. Meyerowitz, E. M., Guild, G. M., Prestidge, L. S. & Hogness, D. S. (1980) *Gene* **11**, 271–282.
20. Meyerowitz, E. M. & Martin, C. H. (1984) *J. Mol. Evol.* **20**, 251–264.
21. Frischauf, A.-M., Lehrach, R., Poustka, A.-M. & Murray, N. (1983) *J. Mol. Biol.* **170**, 827–842.
22. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
23. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
24. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
25. McMaster, G. K. & Carmichael, G. G. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4835–4838.
26. Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K. & Green, M. R. (1984) *Nucleic Acids Res.* **12**, 7035–7070.
27. Ghosh, P. K., Reddy, V. B., Piatak, M., Lebowitz, P. & Weissman, S. M. (1980) *Methods Enzymol.* **65**, 580–595.
28. Zinn, K., DiMaio, D. & Maniatis, T. (1983) *Cell* **34**, 865–879.
29. Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459–472.
30. Messing, J., Geraghty, D., Heidecker, G., Hu, N.-T., Kridl, J. & Rubenstein, I. (1983) in *Genetic Engineering of Plants*, eds. Kosuge, T., Meredith, C. P. & Hollaender, A. (Plenum, New York), pp. 211–227.
31. Dhaese, P., De Greve, H., Gielen, J., Seurinck, J., Van Montagu, M. & Schell, J. (1983) *EMBO J.* **2**, 419–426.
32. Langford, C. J. & Gallwitz, D. (1983) *Cell* **33**, 519–527.
33. Keller, E. B. & Noon, W. A. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7417–7420.
34. Ruskin, B., Krainer, A. R., Maniatis, T. & Green, M. R. (1984) *Cell* **38**, 317–331.
35. Eklund, H., Nordström, B., Zeppezauer, E., Söderlund, G., Ohlsson, I., Boiwe, T., Söderberg, B.-O., Tapia, O. & Brändén, C.-I. (1976) *J. Mol. Biol.* **102**, 27–59.
36. Hake, S. & Walbot, V. (1980) *Chromosoma* **79**, 251–270.
37. Leutwiler, L. S., Hough-Evans, B. R. & Meyerowitz, E. M. (1984) *Mol. Gen. Genet.* **194**, 15–23.
38. Meyerowitz, E. M. & Pruitt, R. E. (1985) *Science* **229**, 1214–1218.