Volume 12 Number 9 1984 Nucleic Acids Research

Molecular analysis of the alcohol dehydrogenase (*Adh1*) gene of maize

E.S.Dennis, W.L.Gerlach, A.J.Pryor, J.L.Bennetzen, A.Inglis, D.Llewellyn, M.M.Sachs, R.J.Ferl and W.J.Peacock

Division of Plant Industry, CSIRO, Canberra, ACT 2601, Australia

Received 15 February 1984; Accepted 5 April 1984

ABSTRACT

A cDNA clone of maize Adh1 which contains the entire protein coding region of the gene has been constructed. The protein sequence predicted from the nucleotide sequence is in agreement with limited protein sequencing data for the ADH1 enzyme. An 11.5 kb genomic fragment containing the Adh1 gene has been isolated using the cDNA clone as a probe, and the gene region fully sequenced. The gene is interrupted by 9 introns, their junction sequences fitting the animal gene consensus sequence. Within the gene there is a triplication of a segment (104 bp) spanning an intron-exon junction. Presumptive promoter elements have been identified and are similar in nucleotide sequence and location, relative to the start of transcription, to those of other plant and animal genes. No recognizable poly(A^+) addition signal is evident. Comparison of the nucleotide sequences of the cDNA (derived from an Adh1-F allele) and genomic (derived from an Adh1-S allele) clones has identified an amino acid difference consistent with the observed difference in electrophoretic mobility of the two enzymes. The maize ADH1 amino acid sequence is 50% homologous to that of horse liver ADH but is only 20% homologous to yeast ADH.

INTRODUCTION

The alcohol dehydrogenase genes of maize (Adh1 and Adh2) are two of about twenty genes which are co-ordinately induced under anaerobic conditions (1). In the absence of a functional alcohol dehydrogenase (ADH1) activity (EC 1.1.1.1; Alcohol : NAD^+ oxidoreductase) maize seedlings are unable to survive the low oxygen tensions occurring during prolonged flooding (2). This correlation also holds in tissue culture, where maize callus tissues derived from immature embryos require a functional Adh1 gene to survive prolonged anaerobiosis (3). These observations suggest that an isolated Adh1 gene may be useful as a dominant selectable marker for plant cell transformation.

Adh1 is genetically one of the best-characterized loci in higher plants, with many variants and mutants already described (4). Two commonly occurring alleles are Adh1-S and Adh1-F, producing respectively, slow and fast electrophoretic variants of ADH1 which differ by a single charge on

each of the two subunit polypeptides. Previously, we reported a cDNA clone (pZML84) of Adh1-78F corresponding to approximately half of the protein coding region (5). This clone has now been used as a probe to isolate a cDNA clone containing the entire coding region and nearly all of the flanking non-translated sequences, and a genomic segment containing all of the Adh1-1S gene and its flanking regions. Both the cDNA and the genomic segment have been sequenced and the data used to predict the amino acid sequences of the two alleles, to examine the structural features of the gene, and to identify the possible regulatory sequences controlling transcription.

MATERIALS AND METHODS

Maize lines used

The poly(A^+) RNA used for the preparation of a new cDNA clone, pZML793, was isolated from the commercial hybrid line Cromac containing the Adh1-CroF allele. The genomic clones λ1S.1 and λ1S.2 were prepared from DNA isolated from lines having the standard S allele Adh1-1S. Adh1 alleles used were: 1F (Standard F) from Berkeley Fast (BKF); FkF from Funk G4343 hybrid (4); FkF-γ25, a γ radiation induced crm^- null mutant (4); PrF from bz2-m inbred line (6); 78F from a rust resistant line; 33F from Super Gold Pop (4); C^t from teosinte (4); FC^m, an Adh1 duplication (4); 1S5657, 1F63, 1F207 are EMS induced crm^- null mutants (4).

Starch gel electrophoresis

Zymograms were performed as described by Schwartz and Endo (7). Anti-ADH1 anti-serum was a kind gift of M. Freeling.

Southern transfer and hybridization

10 μg of maize DNA was digested with 50 units of restriction endonuclease (N.E.Biolabs), electrophoresed on 0.8% agarose gels, transferred to nitrocellulose (8) with acid treatment reduced to 0.1M HCl for 5 min. The filter was baked, prehybridized and hybridized to 2 X 10^7 cpm of a nick translated cDNA clone (pZML793), radioactively labelled using a nick translation kit (New England Nuclear) to a specific activity of $\sim 10^8$ cpm/μg.

Northern analyses

RNA samples were isolated from anaerobically induced seedlings (1). 40 μg of total RNA or 3 μg poly(A^+) RNA were run in each lane of a methyl mercury hydroxide agarose (1.5%) gel (9), transferred to DPT paper (10), hybridized to pZML793 probe and washed (8).

Cloning

Approximately full-length Adh1-CroF cDNA clones were prepared and isolated as described previously (5) except that a high molecular weight fraction of double stranded cDNA was isolated by A150 chromatography before tailing and cloning. Adh1 cDNA clone pZML84 was used as a colony hybridization probe.

Southern blot analysis indicated that the Adh1-1S gene is located on a single 11.5 kb BamH1 fragment. For genomic cloning DNA was digested completely with BamH1 and a size fraction containing 9-13 kb fragments (isolated from sucrose gradients or by gel electrophoresis) was used for cloning. This DNA was ligated and packaged to produce recombinant clones in the vector λ1059 (11). Recombinant phage were screened by plaque hybridization (method adapted from Benton and Davis, 12) using the insert of pZML84 as probe.

DNA sequencing

pZML793 (cDNA) was sequenced using both the methods of Maxam and Gilbert (13) and of Sanger et al. (14). For Maxam and Gilbert sequencing, 3'-OH ends were labelled with α^{32}P d ATP or α^{32} d CTP by "filling in" recessed 3' ends with the large fragment of DNA polymerase I. For the dideoxy chain termination method (14) the plasmid was digested with Sau3A 1, and all fragments internal to the insert eluted from acrylamide. These were cloned in the BamH1 site of mp7.1 (15) and sequenced. Ten fragments were sequenced and ordered from the Maxam Gilbert sequence.

The genomic clone p1S.1 was sequenced by the dideoxy method (14). Restriction fragments were cloned (Fig. 4) in mp8 and mp9 (15).

S1 mapping

S_1 mapping was essentially as described in Weaver and Weissman (16). The 5' probe was a 220 bp PstI-HinfI fragment 5' end labelled with polynucleotide kinase. 150,000 cpm of probe was precipitated with 40 µg poly(A^+) RNA from BKS seedlings anaerobically treated for 6 hours. This was resuspended in 40 µl formamide, denatured at 68°C and hybridized at 50°C overnight.

Additional confirmation of the transcription start point was obtained using a probe made with the 5' PstI-HindIII fragment subcloned into mp8 (15). A single strand probe complementary to Adh1 mRNA was made by hybridizing the 15 bp universal sequencing primer (NE Biolabs) to the mp8 subclone and incubating with 77 µM each dATP, dGTP, dTTP, and 5 µCi dCTP^{32} (New England Nuclear) in 33 mM Tris-HCl pH7.4, 33 mM $MgCl_2$, 160 mM NaCl with

0.125 units Klenow fragment DNA polymerase I (Boehringer) at room temperature for 15 min.

Protein isolation and sequencing

From purified ADH protein (17), three large fragments were produced by partial proteolytic degradation (18). 3mg of protein were dissolved in Laemmli sample buffer and boiled for 3 minutes. 6 µg of S.aureus protease V8 was added and the samples run on a preparative 12.5% acrylamide gel containing SDS. Electrophoresis was halted for 25 minutes when the sample had migrated through the stacking gel to allow digestion to occur. The major individual partial proteolytic products were visualized by placing the gel in ice-cold .25M KCl. The regions of the gel containing the protein fragments were cut from the gel, the proteins electroeluted into dialysis tubing and precipitated with 1.5 volumes of methanol (containing 4 µl/ml glacial acetic acid) at -20°C. The precipitate was washed with acetone and dried. Other fragments were isolated after cyanogen bromide cleavage. Standard N-terminal sequencing procedures generated a limited amount of amino acid sequence data which was compared to some unpublished sequences (Kelly and Freeling, pers.comm.) and to the sequence inferred from the nucleotide sequence.

RESULTS

Isolation of a cDNA clone containing the complete protein coding region of Adh1-F

The standard electrophoretic slow variant of ADH1, Adh1-1S, has a mRNA 1650 bp long, whereas the standard fast allele, Adh1-1F, produces two mRNAs of lengths 1650 and 1750 bp (5). A complete cDNA clone should be at least as long as the shorter of these lengths. The cDNA clone we described previously, pZML84, was only 980 bp long and contained portion of the coding region (the 168 carboxy-terminal amino acids) and a 364 bp 3' untranslated region. We have now isolated a cDNA clone containing a 1592 bp insert (pZML793) from Cromac, a commercial maize line homozygous for the Adh1-CroF allele. This cDNA lacks 20 bp of the 5' terminus of the mRNA.

The sequence of this clone (Fig. 1) has an open reading frame extending from an ATG triplet 78 bp from the 5' end, which can be aligned with the reading frame already described for pZML84 (5). The protein coding region encompasses 379 amino acids, corresponding to a polypeptide of molecular weight 38 kd and in agreement with experimental estimates of the molecular weight of the ADH1 subunit (19). The predicted amino acid composition is

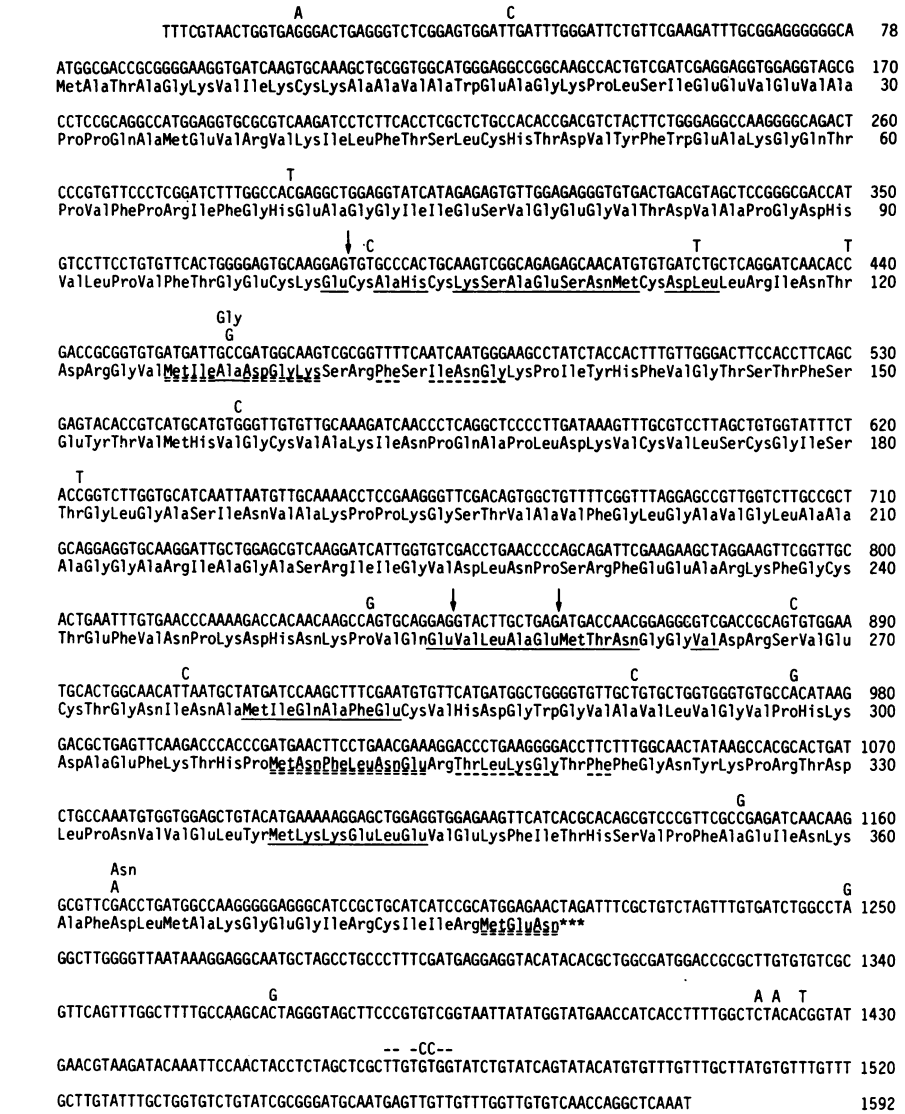

Fig. 1 Adh1-nucleotide and direct amino acid sequences of cDNA pZML793. Sites of cleavage with S.aureas V8 protease are shown ↓. The amino acid sequences we determined are underlined ——— , those supplied by Kelly and Freeling ---- and those determined in both laboratories ====. Differences between the Adh1-CroF and Adh1-1S sequence (see genomic sequence (Fig. 5)) are indicated above the Adh1-F sequence.

similar to that directly determined for the protein (20, and Inglis unpublished). pZML793 encodes the entire ADH1 polypeptide. There are two translation stop codons upstream from the ATG nominated as the initiation methionine.

Amino acid sequence of the ADH1 protein

The N terminus of the ADH1 polypeptide is blocked, so the sequence of this region could not be obtained. A limited amount of amino acid sequence data was obtained from peptides derived by cyanogen bromide digestion or by partial proteolysis with Staphylococcus aureas V8 protease (18). In the S.aureus protease reaction, the conditions were adjusted so that only the most labile glutamic acid-X bonds were cleaved by the protease to produce three major fragments of molecular weights 10, 13 and 15 kd. The 10 kd fragment had a blocked N terminus presumably corresponding to the N terminus of the intact protein. The other fragments each yielded about 15 amino acids of sequence from their N terminus (Fig. 1). The 13 kd fragment consisted of two fragments differing at their N terminal ends, due to the cleavage of two susceptible glutamic acid-X bonds in close proximity (Fig. 1). One was cleaved to a greater extent than the other making it possible to assign the amino acid residues to the two fragments on the basis of the relative yields in the degradation cycle.

Four sequences, consistent with the sequence predicted from the DNA nucleotide sequence, were obtained from cyanogen bromide fragments. The C-terminal cyanogen bromide peptide was confirmed as met-glu-asn. Amino acid sequence data (Fig. 1) are consistent with the sequence predicted from DNA sequencing. We cannot be sure whether the methionine corresponding to the first ATG is the N-terminal amino acid or whether there is post-translational modification of the amino acids at the N terminus of the polypeptide. However, since the products of in vitro translation of Adh1 mRNA co-migrate with the ADH1 polypeptides isolated in vivo, any processing that does occur must affect very few residues.

Expression in E.coli

Extracts from E.coli containing the plasmid pZML793 have a low level of ADH activity of the expected F class of mobility (lane 2, Fig. 2). The origin of the polypeptide from the maize insert of pZML793 and its positive identification as ADH at the immunological level are shown in lanes 4-6 of Fig. 2. The GGAGGG sequence at nucleotide 68 in this plasmid corresponds to the known bacterial ribosomal binding site (Shine-Dalgarno sequence, 21) and could account for the observed bacterial expression. The ADH activity from

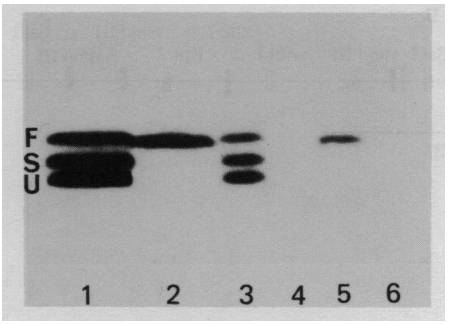

Fig. 2 Expression of Adh1-F cDNA clone pZML793 in E.coli. Starch gel zymograms (7) of alcohol dehydrogenase from maize kernels and E.coli carrying the plasmids pBR322 and pZM793. (1) Standard mixture of scutellar extracts in 10mM Tris HCl pH 8.0 from homozygous lines carrying the Adh1-F Adh1-S and Adh1-U alleles; (2) E.coli + pZML793; (3) Standard mixture; (4) E.coli + pZML793 + ADH anti sera (note loss of activity); (5) E.coli + pZML793; (6) E.coli + pBR322.

the E.coli extracts has a slightly slower electrophoretic migration rate than authentic maize ADH1-F enzyme (Fig. 2), possibly indicating a difference due to post-translational protein modification (e.g. N-terminal amino acid modification) between maize and E.coli. These observations indicate that the assembly of the subunits of maize ADH protein and the insertion of Zn^{++} required for activity can occur in E.coli.

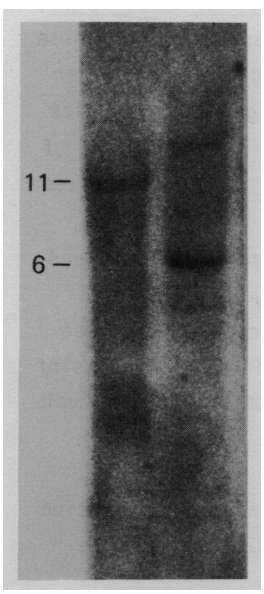

Fig. 3 Genomic organization of Adh1-1S and 1-1F. DNA was isolated from BKS (left) or BKF (right). 10µg of DNA was digested with Bam H1 restriction enzyme, transferred to nitrocellulose and hybridized to radioactively labelled pZML793. The numbers refer to lengths in kb.

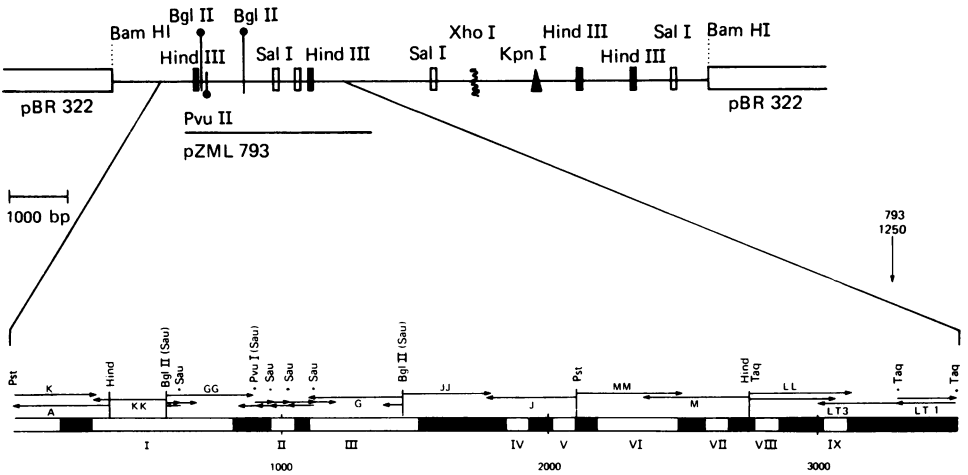

Fig. 4 Map of restriction sites of the Adh1-1S genomic segment cloned in pBR322. The region corresponding to the cDNA clone pZML793 is indicated. The region sequenced is expanded and the sites and subcloned fragments used for sequencing indicated. Exons are filled in, introns are numbered.

Genomic organization of the Adh1 gene

BKS and BKF maize stocks showed a small number (1-3) of fragments, of different sizes in BKS and BKF, when the DNA was restricted with 6 bp recognition site enzymes (BamH1, BglII, HindIII) and hybridized to pZML793 (Fig. 3, and 22). EMS-derived electrophoretic mutants (Adh1-1U and Adh1-1W, 19) exhibited the same patterns as Adh1-1F and Adh1-1S, respectively. These results together with the patterns obtained with double digests and other enzymes indicate that there is a single copy of the Adh1 gene in the maize genome. Single copy reconstruction experiments, in which a known amount of pZML793 was compared to the sequence concentration in the genome, supported this conclusion.

The Adh1-1S gene is located on a single BamH1 fragment of 11.5 kb and the Adh1-1F gene on a BamH1 fragment of 7.0 kb, suggesting that fractionation of these BamH1 fragments would facilitate cloning of these genes. Longer fragments were present to varying extents in BamH1 digests of Adh1-1F in different experiments and probably resulted from incomplete BamH1 digestion as a consequence of modification of some sites.

Cloning and sequencing of the Adh1-S gene

Two independent Adh1-1S clones were identified by plaque hybridization with the cDNA probe. λ1S.1 (originally λAD59) and λ1S.2 (in opposite

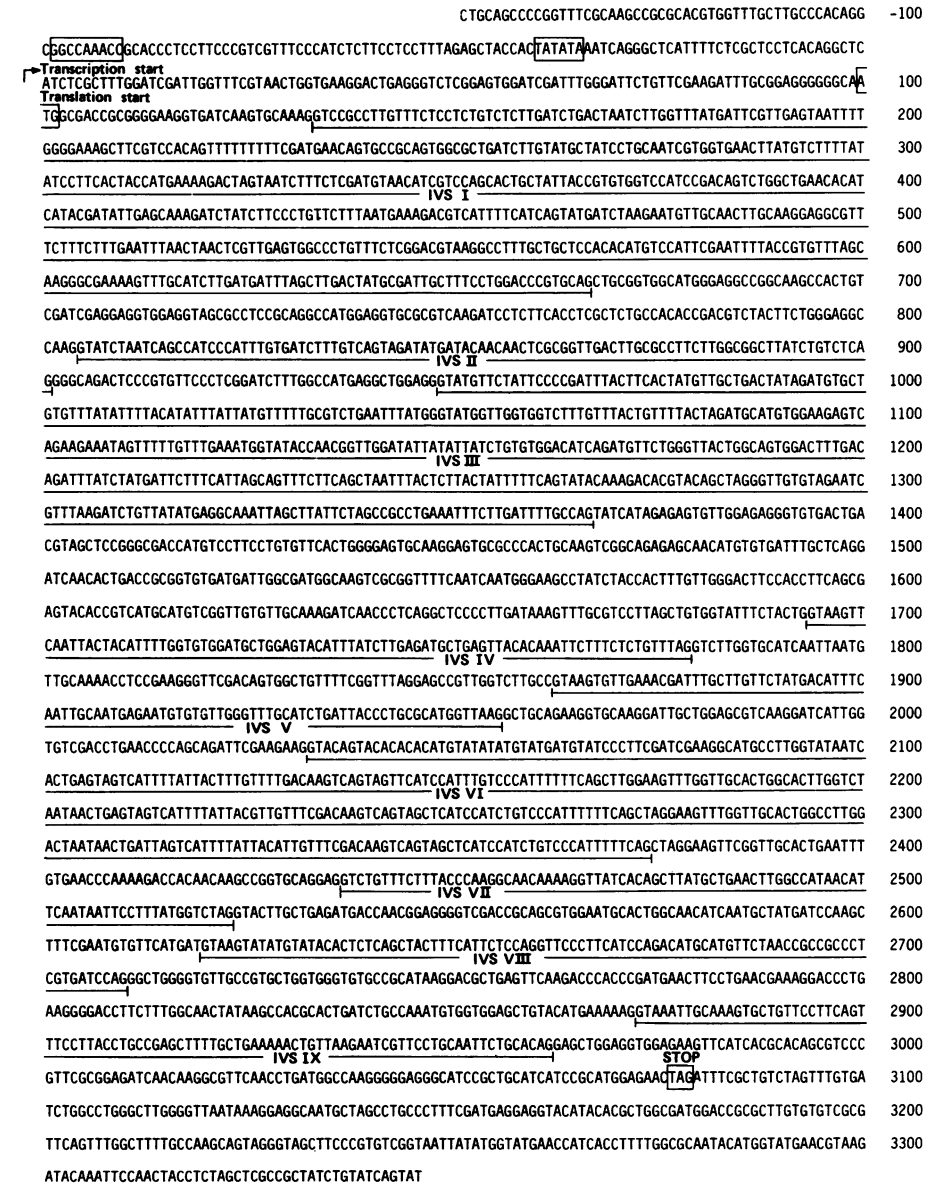

Fig. 5 The nucleotide sequence of the genomic clone λ1S.1. Underlined regions are introns. The presumptive transcription and translation signals are boxed.

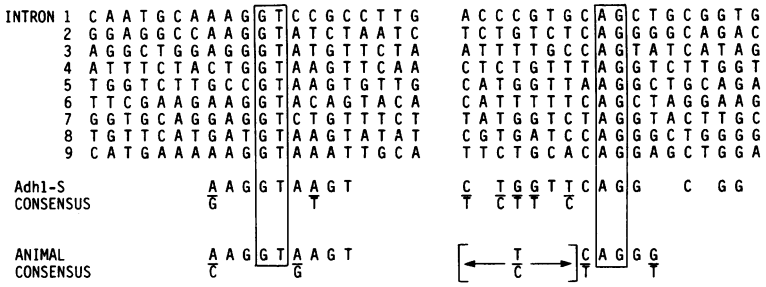

Fig. 6 Nucleotide sequences of intron/exon junctions in the Adh1-S gene. The nucleotide sequence around the intron/exon junction is shown for the 9 Adh1-S intron donor and acceptor sites together with the consensus sequence around the Adh1-S introns and the general consensus sequence for animal genes (23).

orientation to λ1S.1) contain segments of maize DNA which are identical by restriction enzyme mapping and give the map expected from the Southern blotting of genomic DNA. Both inserts were recloned into pBR322; detailed sequencing was carried out on p1S.1, the pBR322 cloned insert from λ1S.1 (Fig. 4). The insert contained about 1.5 kb upstream of the 5' end of the coding region of Adh1 and 6 kb downstream of the 3' end of pZML84. The coding region spans approximately 3 kb and because the mRNA length is 1650 bases, we expected to find intervening sequences in the gene. A number of subclones were prepared in pUC8 and in the M13 vectors mp8 and mp9 (15) (Fig. 4).

The complete sequence of the gene region and some of the adjacent flanking regions was determined (Fig. 5). Comparison of the sequence of the genomic clone with that of the cDNA clone showed nine introns to be present in the coding region of the gene. Each intron is bounded by GT/AG signals which have been found previously at all eukaryote intron/exon junctions (23). The consensus sequence of the Adh1 intron junctions (Fig. 6) is similar to that found in other plants and resembles that described for animal genes. The only differences are some G for C substitutions. All the introns are rich in AT base pairs; their base composition, length and position in the coding sequence are summarized in Table 1.

At the 3' boundary of intervening sequence 6 a region of approximately 100 bp is repeated three times. The first two copies are entirely within intron 6. Both contain in-frame stop codons following the putative splice acceptor sites, and so could only code for a truncated protein sequence even if incorrectly used as the intron exit during splicing of the RNA

Table 1 Characterization of Introns in the Adh1-S gene

Intron	Position in pZML793 (bp #)	Length of intron (bp)	% AT
1	112	535	59
2	249	98	54
3	296	419	66
4	622	86	65
5	705	94	61
6	781	342	63
7	843	87	62
8	939	87	55
9	1101	89	66

transcript. The third copy of the repeat spans the junction of intron 6 and the subsequent exon region. The sequence of the three copies is highly conserved (Fig. 7).

The start of transcription was identified by S_1 mapping using either a 5'-labelled or a uniformly labelled fragment (Fig. 8). Transcription starts at a CAT sequence (see Fig. 5). Upstream from the transcription start point there are putative TATA and CCAAT boxes. The TATA box, which has the sequence TATATAAT, begins at nucleotide -35 (numbering from the transcription start point) and the presumptive CCAAT box (of sequence GGCCAAACC) is at nucleotide -90, their positions being similar to those seen in other eukaryotic genes.

The position of translation initiation, the ATG codon at position 100 in the genomic clone, gives a 100 base 5' leader sequence. The nucleotide sequence around the initiation codon, GCAATGGC, is similar to the sequence defined (24) on the basis of animal gene data as the eukaryotic consensus sequence and strengthens our assignment of this ATG triplet as the

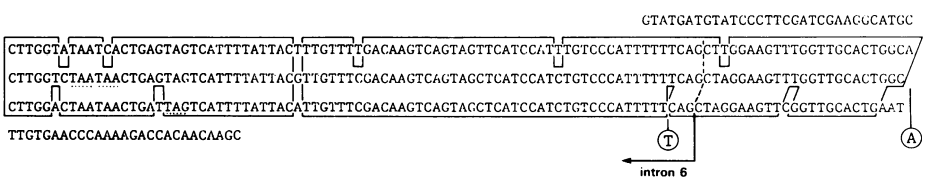

Fig. 7 Nucleotide sequence of the triplicated region of the genomic sequence of Adh1-1S. The region extending from the 3' end of intron 6 and into exon 7 is shown with the three copies of the triplicated region aligned. Changes in nucleotide sequence are indicated as is the intron/exon boundary and the stop codons on the correct reading frame (....).

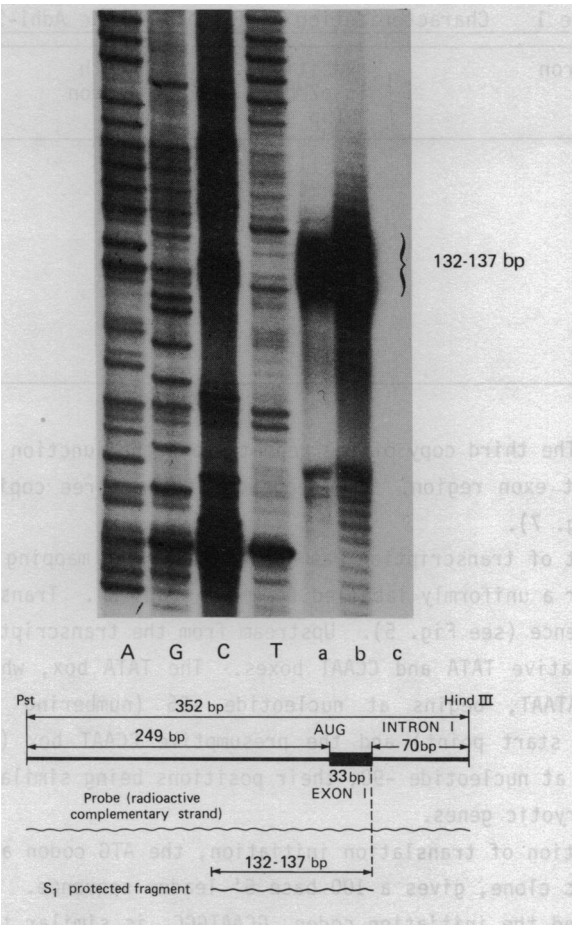

Fig. 8 S_1 mapping of transcription start of Adh1-S. The Pst-HindIII A fragment (Fig. 4) was made uniformly radioactive in the strand complementary to the mRNA. This strand was then hybridized to Adh1-1S mRNA, treated with S_1 nuclease and run on a sequencing gel. AGCT represent dideoxy sequencing tracks of the same fragment. The protected fragments using the following RNAs are shown: a) total BKS RNA, b) poly(A^+) BKS RNA, c) E.coli tRNA.

translation start.

A poly(A^+) addition signal resembling the animal consensus sequence AATAAA is not seen at the 3'-terminus of the cDNA sequence despite the presence of a poly(A) tail on the clone. Some plant genes, for example that of phaseolin (25), do have an AATAAA sequence occurring 10-30 bases (16 bp in phaseolin) upstream of the poly(A) addition site. In the case of Adh1 either the poly(A) addition signal is different from the consensus, or the

cDNA clones we have sequenced have been transcribed from an internal run of A residues in the mRNA and not from the true poly(A) tail (and hence are too short). A third alternative is that the poly(A) addition signal is within an intron. Additional sequencing data for the 3' end of the genomic clone may show the AATAAA further downstream.

Differences between Adh1-S and Adh1-F

The cDNA clone pZML793 was derived from a fast electrophoretic variant, Adh1-CroF, and the genomic clone, p1S.1, from a slow variant, Adh1-1S. There are differences in the nucleotide sequences of these two alleles (Fig. 1). Of the 14 nucleotide substitutions in the protein coding region, 12 occur in the third base of the codon and do not cause any change in the amino acid sequence. There is one C to G change (codon 127) resulting in the substitution of glycine for alanine, and one G to A change (codon 363) causing a substitution of aspartic acid for asparagine (Fig. 1). This latter change, equivalent to unit charge, would account for the difference in electrophoretic mobility observed between the two proteins. This change occurs within 20 amino acids of the C terminus and does not affect the enzymatic activity of the protein since ADH1-F and ADH1-S have the same specific activity (26). Base changes between the two alleles are not random, with transitions accounting for 80% of the substitutions in both translated and non-translated regions; on random expectation transversions should be twice as frequent as transitions.

Although the difference in electrophoretic mobility between F and S can be accounted for, explanation of the other differences which have been observed between the F and S alleles, such as a failure to undergo intragenic recombination (26), different mRNA patterns (5), and differential expression of the two alleles in different organs of the plants (27,28) will have to await further comparison of the nontranslated and flanking sequences of the two alleles.

mRNA of Adh1 variants

Most Adh1 alleles have a single mRNA of $\sim$ 1650 bases, (Fig. 9), however, several F alleles (BKF, CroF, FkF, PrF) show about an equal amount of an additional mRNA of $\sim$ 1750 bases. 78F has the 1750 base class but in a smaller amount relative to the 1650 bases mRNA. One F allele, 33F, has only the 1650b mRNA. The FC^{m} duplication, suggested to be one operon (29), has only one band at 1650 b.

The origin of the longer transcript present in F lines remains obscure but since there is no heterogeneity in the length of the ADH1F polypeptide

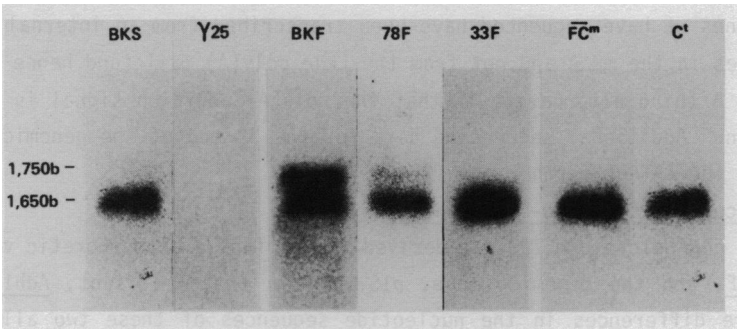

Fig. 9 Northern analysis of mRNA from some Adh1 variants.

the extra 100 nucleotides most likely occur in either the 5' or 3' flanking regions. Too few cDNA clones have been isolated to say whether the 3' non-translated region is longer in half of the Adh1-F mRNAs due to alternative poly(A^+) addition sites. On the other hand there may be alternative transcription start sites in the Adh1-F gene. The situation will not be clarified until the Adh1-F genomic fragment is isolated and subjected to S_1 mapping at both 5' and 3' ends.

Several ADH null mutants have been induced by EMS and one by gamma irradiation. The EMS-induced crm^- nulls S5657, F63 and F207 have the same mRNA pattern as their progenitor alleles (1S and 1F). A gamma ray induced crm^- null Adh1-FkF525 (γ25) shows no detectable mRNA (Fig. 9).

DISCUSSION

Genetic analyses, and in particular mutagenesis and recombination studies, suggested the Adh1 locus is a single copy gene in the maize genome. Also our Southern blot patterns can be explained on the basis of one or few copies of the gene. Restriction site mapping and sequencing of the cloned gene segment confirms that it is a single copy gene.

Sequence analysis of the Adh1 gene shows it to have features typical of eukaryote genes. Translation and transcription signals, and intron-exon junctions are similar to those found in most animal genes. The two putative promoter elements, a TATA box at -35 bp and a CCAAT box at -90 bp, bear strong sequence and position resemblance to the consensus sequences determined for animal genes but only functional analysis will determine if they are truly analogous to the animal promoter elements. We were not able to identify a poly(A^+) addition signal in the gene, the only AATAAA sequence in the 3' untranslated region being more than 150 bp 5' of the poly(A^+)

```
1    .STAGKVIKCKAAVLWEEKKPFSIEEVEVAPPKAHEVRIKMVATGICRSD   49
     |||||||||||||| ||  || |||||||||| | ||| |   |  | |
1    MATAGKVIKCKAAVAWEAGKPLSIEEVEVAPPQAMEVRVKILFTSLCHTD   50

50   DHVVSGTLVTP.LPVIAGHEAAGIVESIGEGVTTVRPGDKVIPIFTPQCG   98
               ||  | | |||| || || ||||| | ||| | | ||  |
51   VYFWEAKGQTPVFPRIFGHEAGGIIESVGEGVTDVAPGDHVLPVFTGECK   100

99   KCRVCKHPEGNFCLKNDLSMPRGTM.QDGTSRFTCRGKPIHHFLGTSTFS   147
     |  ||  | | |        || |  || |||   |||| || ||||||
101  ECAHCKSAESNMCDLLRINTDRGVMIADGKSRFSINGKPIYHFVGTSTFS   150

148  QYTVVDEISVAKIDAASPLEKVCLIGCGFSTGYGSAVKVAKVTQGSTCAV   197
     |||      ||||    || |||   || ||| |    |||    ||| ||
151  EYTVMHVGCVAKINPQAPLDKVCVLSCGISTGLGASINVAKPPKGSTVAV   200

198  FGLGGVGLSVIMGCKAAGAARIIGVDINKDKFAKAKEVGATECVNPQDYK   247
     |||| |||    |  ||| |||||| |  |  |   | || ||| |
201  FGLGAVGLAAAGGARIAGASRIIGVDLNPSRFEEARKFGCTEFVNPKDHN   250

248  KPIQEVLTEMSNGGVDFSFEVIGRLDTMVTALSCCQEAYGVSVIVGVPPD   297
     || |||| || |||||| | |  |    |  | |     || | ||||
251  KPVQEVLAEMTNGGVDRSVECTGNINAMIQAFECVHDGWGVAVLVGVPHK   300

298  SQNLSMNPMLLLSGRTWRGAIFGGFKSKDSVPKLVADFMAKKFALDPLIT   347
           ||  |  ||  |  ||  |      |  |   | |        ||
301  DAEFKTHPMNFLNERTLKGTFFGNYKPRTDLPNVVELYMKKELEVEKFIT   350

348  HVLPFEKINEGFDLLRSGESIRTILTF...                      374
     |  ||  ||  |||   || || |
351  HSVPFAEINKAFDLMAKGEGIRCIIRMEN                       380
```

Fig. 10 Comparison of the amino acid sequence of horse liver alcohol dehydrogenase (top line) with that of maize ADH1 (bottom line). The standard one letter amino acid code is used. Amino acids which are identical in the two enzymes are connected.

addition site. Some other plant genes (soybean leghaemoglobin and maize zein) also do not have AATAAA in the standard location (30,31) and it is possible that there is more than one sequence which can act as the signal for polyadenylation in plant genes.

Although the maize Adh1 intron-exon junctions have the animal consensus sequence the presence of the 104 bp triplication at the junction of intron 6 and exon 7 shows that nucleotide sequence per se is not the only factor determining the functional splicing points within the gene. It is evident that the effective acceptor site is the one in the most 3' copy of the triplicated sequence. This is the only mature mRNA from which ADH polypeptide could be encoded since termination triplets follow the other two acceptor sites. With the most 3' acceptor being used, this triplication provides a natural analogy to the observation that when duplications are artificially introduced into a rabbit β globin the effective donor site is the most 5' copy of the donor sequence and the effective acceptor site is the most 3' copy of the acceptor sequence (32). 30 bp of the 104 bp are in the protein coding region so the triplication must have arisen from the sequence spanning the intron-exon junction. It will be of interest to

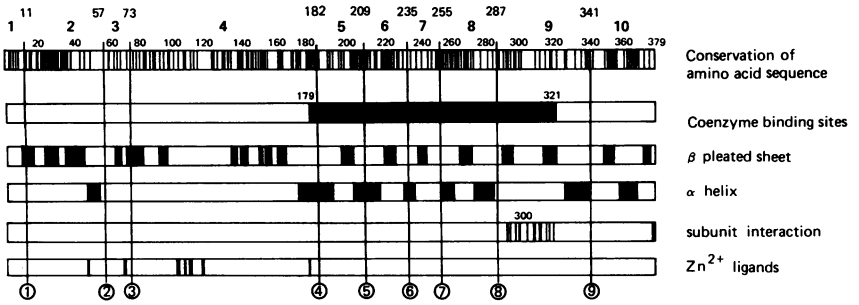

Fig. 11 Diagram of functional regions of the maize ADH1 enzyme as predicted from homology to the horse liver enzyme (32,34). Vertical lines indicate the position of introns in the Adh1-S gene. The top box is a diagrammatic representation of the data in Fig. 10.

determine whether this same region is triplicated in other Adh1 alleles (e.g. Adh1-1F) and in the unlinked Adh2 gene.

Although ADH1 is induced by anaerobic conditions, a nucleotide sequence responsible for the induction has not been identified. In Drosophila melanogaster, sequence and functional analyses have led to the identification of a short nucleotide sequence (CTNGAATNTTCTAGA), in the -60 to -150 region, which is responsible for the property of high temperature induction of a set of heat shock genes (33). A similar control sequence may exist for the anaerobically-induced genes in maize and should be identifiable by a similar functional analysis. Alternatively, it may be possible to identify a sequence common to anaerobically induced loci when their genomic sequences are compared. We already have cDNA clones for a number of them.

The maize ADH1 amino acid sequence has about 20% homology with yeast ADH and 50% with the horse liver enzyme (Fig. 10); 31 of the first 40 amino acids are identical in the maize and horse liver enzymes. The amino acid data strongly suggest a common evolutionary origin of the animal and plant Adh genes and a more distant relationship to the yeast gene. Because of the high degree of homology between the maize and horse liver enzyme the three dimensional structure of the horse liver ADH (34,35) at 2.4Å resolution is probably a reasonable approximation for the structure of the maize ADH protein and can be used to predict the position of β pleated sheet and α helix structures, zinc-binding ligands, regions of subunit interaction and the coenzyme and catalytic domains (Fig. 11). The zinc-binding ligands show complete conservation with those of the horse liver enzyme. However, other

structural domains show little obvious correlation with the position of the 9 introns of the Adh1 gene. The one exception is that the co-enzyme binding domain (amino acids 179-321) is very nearly exactly contained in the 5 exons bounded by introns 4 and 9. When the horse liver gene is cloned and the nucleotide sequence obtained, the question of conservation of both intron position and number and the relationship to functional domains between the maize and horse Adh genes can be properly addressed.

ACKNOWLEDGEMENTS

We are grateful to K. Ferguson, G. Koci, Y. Hort and M. Lee for excellent technical assistance. We thank Janice Kelly and Mike Freeling for supplying unpublished protein sequencing data.

REFERENCES

1. Sachs, M., Freeling, M. and Okimoto, R. (1980) Cell 20, 761-767.
2. Schwartz, D. (1969) Amer.Nat. 103, 479-481.
3. Peacock, W.J., Dennis, E.S., Gerlach, W.L., Llewellyn, D., Lorz, H., Pryor, A.J., Sachs, M.M., Schwartz, D. and Sutton, W.D. (1983) Proc. Miami Winter Symp., Academic Press, N.Y.
4. Freeling, M. and Birchler, J.A. (1981) In "Genetic Engineering Principles and Methods" III pp.223-264 (Setlow, J.K. and Hollaender A., eds) Plenum Press, N.W.
5. Gerlach, W.L., Pryor, A.J., Dennis, E.S., Ferl, R.J., Sachs, M.M. and Peacock, W.J. (1982) Proc.Natl.Acad.Sci.USA 79, 2981-2985.
6. Nuffer, M.G. (1953) Maize Genet. Coop Newsletter 27,68.
7. Schwartz, D. and Endo, T. (1966) Genetics 53, 709-715.
8. Wahl, G.M., Stern, M. and Stark, G.R. (1979) Proc.Natl.Acad.Sci.USA, 76 3683-3687.
9. Chandler, P.M., Rimks, D. and Davidson, N. (1979) Anal.Bioch. 99, 200-206.
10. Seed, B. (1982) Nucleic Acids Res. 10, 1799-1810.
11. Karn, J., Brenner, S., Barnett, L., Cesareni, G. (1980) Proc.Natl. Acad.Sci.USA 77, 5172-5176.
12. Benton, W.D. and Davis, R.W. (1977) Science 196, 180-182.
13. Maxam, A.M. and Gilbert, W. (1980) in "Methods in Enzymology" Vol.65 (L.Grossman and K.Moldave, eds.) Academic Press, New York, 499-560.
14. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc.Natl.Acad.Sci. USA 74, 5463-5476.
15. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
16. Weaver, R.F. and Weissman, C. (1979) Nucl.Acids Res. 7, 1175-1193.
17. Pryor, A.J. and Huppatz, J.L. (1983) Biochem.Internat. 4, 431-438.
18. Clevland, D.W., Fischer, S.G., Kirschner, M.W. and Laemmli, U.K. (1977) J.Biol.Chem. 252, 1102-1106.
19. Ferl, R.J., Dlouhy, S.R. and Schwartz, D. (1979) Molec.gen.genet. 169, 7-12.
20. Kelly, J. and Freeling, M. (1980) Biochem.Biophys.Acta 624, 102-110.
21. Shine, J. and Dalgarno, L. (1975) Nature 254, 34-38.
22. Johns, M.L., Strommer, J.N., Freeling, M. (1983) Genetics 105, in press.
23. Breathnach, R. and Chambon, P. (1981) Ann.Rev.Biochem. 50, 349-383.
24. Kozak, M. (1981) Nucl.Acids Res. 9, 5233-5252.

25. Slightom, J.L. and Hall, T.C. (1983) Proc.Natl.Acad.Sci.USA
26. Freeling, M. (1978) Genetics 89, 211-224.
27. Schwartz, D. (1971) Genetics 67, 411-425.
28. Woodman, J.C. and Freeling, M. (1981) Genetics 98, 354-378.
29. Schwartz, D. Maize Genetics Co-op. Newsletter (1973) 47, 53-55.
30. Jenson, E.O., Palendan, K., Hyldig-Nielson, J.J., Jorgensen, P. and Marcker, K.A. (1981) Nature 291, 677-679.
31. Pedersen, K., Devereux, J., Wilson, D.R., Sheldon, E. and Larkins, B.A. (1982) Cell 29, 1015-1026.
32. Wieringa, B., Meyer, F., Reiser, J. and Weissmann, C. (1983) Nature 301, 38-43.
33. Pelham, H.R.B. (1982) Cell 30, 517-528.
34. Ekland, H., Branden, C-I and Jornuall, H. (1976) J.Mol.Biol. 102, 61-73.
35. Ekland, H., Nordstrom, B., Zeppezauer, E., Soderlund, G., Ohlsson, I., Boiwe, T., Soderberg B-O, Tapia, O, and Branden, C-I. (1976) J.Mol.Biol. 102, 27-59.