THE MOLECULAR CLONING AND CHARACTERIZATION OF THE WHEAT *Adh*-1A GENE

BY

LUKE MITCHELL

A thesis submitted for the degree of Doctor of Philosophy
of The Australian National University
Canberra

No PAGE 18. TEXT O.K.

August, 1987

DECLARATION

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University. To the best of my knowledge, it contains no material previously published, or the result of any work by another person, except where due reference is made in the text.

Luke E Mitchell

Luke Mitchell

~~Chapters 2,3 and 4 of this thesis are presented in manuscript form. Chapters 2 and 3 have been submitted to the Journal of Molecular and General Genetics and Chapter 4 to the Journal of Theoretical and Applied Genetics~~. Some sections of the introduction and results of this thesis form part of a paper titled "Anaerobically regulated gene expression: Molecular adaptations of plants for survival under flooded conditions" which has been accepted for publication in The University of Oxford Surveys of Plant Molecular Biology.

ACKNOWLEDGEMENTS

I would like to thank the Rockefeller Foundation for providing financial support for this project.

I would also like to thank Professor J. Pateman of the Genetics Department for his consent to do this work through the Research School of Biological Sciences, Australian National University.

Throughout the course of this project Dr. W. J. Peacock has been a constant source of inspiration and critical discussion. I would like to thank him for this, as well as for his permission to do this work at CSIRO.

The members of the Division of Plant Industry in particular the biochemistry and genetics sections have provided a very friendly atmosphere in which to work. Dr. W. Bottomley has been especially helpful with his great computer programs, as have Mandy Walker for reading this thesis and making many useful comments, and Rhonda Perriman for her technical assistance.

Finally I would especially like to thank Dr. Elizabeth Dennis for her encouragement, support and enthusiasm over the last 3 1/2 years.

CONTENTS

ABSTRACT

We have cloned and determined the nucleotide sequence of a gene encoding alcohol dehydrogenase (*Adh*) from *Triticum aestivum* cv. Millewa. The cloned gene hybridizes to an anaerobically induced RNA, and sequences important for eucaryotic gene expression such as the TATA box, polyadenylation signal and intron splice sites are found in the expected positions.

Using *T.aestivum* cv Chinese Spring nullisomic-tetrasomic and ditelosomic lines we have mapped the cloned *Adh* gene to a previously unidentified *Adh* locus on the long arm of chromosome 1A. Southern analysis shows that there are three copies of the gene in the *T.aestivum* genome and that the remaining copies map to the long arms of chromosomes 1B and 1D.

The cloned gene has a structure similar to the maize *Adh* genes with an open reading frame encoding a polypeptide of 379 amino acids. The polypeptide is identical in length and ~85% homologous to maize ADH1 and ADH2 at the amino acid level. The open reading frame is interrupted by eight introns which are in identical positions to eight of the nine introns in maize, but their length and sequence differ.

A comparison of the nucleotide sequence of the wheat *Adh*-1A and Barley *Adh3* genes revealed extensive homology in both the coding and non coding regions of these genes. The homology is however discontinuous due to a 1.8 kb insertion (TLM) in the wheat *Adh*-1A gene. This insertion is associated with a direct duplication of 8 bp of flanking genomic DNA, contains several large open reading frames, and the termini of the insertion have an imperfect 14 bp inverted repeat. These structural features are similar to those of eucaryotic transposable elements. Evolutionary studies of ancestral wheats suggest that the insertion is not

CHAPTER 1

INTRODUCTION

1.1 ALCOHOL DEHYDROGENASE AND THE ANAEROBIC RESPONSE.

1.1.1 The anaerobic response.

The sedentary nature of plants has meant that plants do not have the ability to alter their immediate environment by changing their location in the way an animal does. As a result plants are subject to greater extremes of environmental conditions and to different selection pressures. This has resulted in many unique adaptive responses of plants to environmental conditions such as drought, flooding, and temperature extremes.

Insufficient oxygen, resulting from flooding of the root zone, is one particular stress that plants often encounter and that has been studied extensively both at the physiological and molecular level. Flooding induced anaerobiosis has been shown to result in increased ethylene content in the root and shoot of plants (Jackson and Campbell,1975), decreased nitrate uptake and leaf wilting and chlorosis (Drew and Sisworo,1977). In maize, together with physiological changes, there is a dramatic alteration in the pattern of protein synthesized by the primary root. During the first hour of anaerobiosis, aerobic protein synthesis is repressed and there is synthesis of four polypeptides called the transition polypeptides. In the second hour of anaerobiosis the synthesis of another small group of about twenty polypeptides, the anaerobic polypeptides (ANP's), is initiated. After five hours this group of polypeptides accounts for greater than 70% of total protein synthesis and remains unchanged until root death (Sachs et al.1980). A number of the major anaerobic polypeptides have been identified; they include alcohol dehydrogenase 1 and 2, aldolase

(Kelly and Freeling,1984a), sucrose synthase (Werr et al.1985) pyruvate decarboxylase (Wignarajah and Greenway,1976; Lazlo and St Lawrence,1983) and glucose phosphate isomerase (Kelly and Freeling,1984b). Several of these enzymes catalyze reactions involved in glycolysis or, in the case of sucrose synthase, catalyze the reaction providing the substrate for glycolysis. Thus their synthesis ensures that glycolysis will continue when aerobic respiration is not possible. Roberts et al.(1984a) have shown that under anaerobic conditions there is a rapid inhibition of oxidative phosphorylation, due to the absence of O_2 as a terminal electron acceptor, and a resulting increase in NADH levels. To regenerate the NAD^+, necessary for glycolysis under anaerobic conditions, plants progressively change metabolic pathways from oxidative phosphorylation to lactic acid fermentation and finally to alcohol fermentation. The reduction of pyruvate, the end product of glycolysis, to lactic acid by lactate dehydrogenase produces NAD+. However, the lactic acid produced rapidly reduces the cytoplasmic pH resulting in inhibition of lactate dehydrogenase and concurrent activation of pyruvate decarboxylase and alcohol dehydrogenase (Roberts et al.(1984b). Pyruvate decarboxylase converts pyruvate to acetaldehyde which is subsequently oxidized to ethanol by alcohol dehydrogenase. This last reaction regenerates the NAD^+ required for glycolysis.

1.1.2 Transcriptional control.

Several maize cDNA clones corresponding to anaerobically induced polypeptides have been isolated (Gerlach et al.1982; Hake et al.1985) and used in studies on the expression of mRNAs during anaerobic conditions. Hake et al.(1985) have examined the expression of five cDNA clones corresponding to the anaerobically induced mRNAS from; ADH1, sucrose synthase, aldolase and two

unidentified ANPs, ANP31 and ANP40. With the exception of ANP40, mRNAs hybridizing to these cDNA clones begin to accumulate within 30 minutes of anaerobiosis. The ANP40 mRNA does not accumulate until after 3 hours of anaerobic treatment. The expression of these five anaerobic mRNAs could be further distinguished on the basis of the kinetics of their accumulation. It seems likely that these differences are due to both transcriptional and post-transcriptional processes and are related to the different physiological functions of the individual ANPs.

Dennis et al.(1984), have cloned and determined the nucleotide sequence of the anaerobically induced maize *Adh1* gene. This gene has been shown by Schwartz (1969) to be of critical importance in the anaerobic response of maize since seedlings which are homozygous for a non-functional *Adh1* gene die, presumably due to an inability to suppress cytoplasmic acidification as a result of lactic acid fermentation. Walker et al.,(PNAS, in press) have studied the transcriptional induction of the maize *Adh1* gene and have identified within the maize *Adh1* promoter, a 40 base pair element which is essential for anaerobic regulation (ARE). This region can be further divided into two sub regions at position -140, CTGCAGCCCCGGTTTC and position -114, CCGTGGTTTGCTTGCC relative to transcription start. These two regions are separated by ten base pairs, the sequence of which does not appear to be critical for the anaerobic response.

Since induction of ADH activity has been demonstrated for a wide variety of both monocot and dicot plants (McMannon and Crawford,1971; Gottlieb,1982) it might be expected that the maize anaerobic regulatory sequence would be conserved in *Adh* genes from other plant species. Functional analysis of the pea *Adh1* (Llewellyn et al.,1987) indicates that a region between -120 and -

51 is critical for anaerobic expression. Within this region no homology to the maize regulatory sequence is present although a search of the complement of the sequence identified an 7 base pair region CAAACCA/GTTTGGT at position -106 which has homology to each sub region of the maize ARE. The anaerobically induced *Arabidopsis thaliana Adh* gene (Chang and Meyerowitz,1986) also shows a similar inverse complementary sequence CAAAACCAA/GTTTTGGTT at position -166 relative to transcription start.

1.1.3 Translational control.

The translational regulation of the maize anaerobic response is not well understood although Sachs et al.(1980) have shown that in vitro translation of m-RNA's isolated from anaerobically treated tissue results in the synthesis of both the normal aerobic as well as the anaerobic polypeptides. Therefore, the block in expression of the aerobic polypeptides under anaerobic condition is exerted at the translational level. Lin and Key (1967) have also observed that under anaerobic conditions the polyribosomes dissociate. Using cycloheximide they were able to show that this dissociation is not an active process but rather the result of the ribosomes being unable to reinitiate on the m-RNA's. The translational control therefore appears to be exerted at the level of ribosome recognition of the m-RNA species.

1.1.4 ADH amino acid sequences.

Although the biochemical basis for the presence of alcohol dehydrogenase is different in plant and mammalian systems, the reaction catalyzed is nevertheless identical (Branden et al.1975). This functional homology has provided a useful basis for analysis of the secondary and tertiary structure of the plant ADH enzymes.

The most extensively studied ADH enzyme, in terms of structure and catalytic mechanism, is that derived from horse liver (LADH,

EC 1.1.1.1). The horse liver enzyme, like maize, has a molecular weight of 80,000 and is a dimer of two identical subunits each of which binds two zinc ions and has one active site (Eklund et al.,1976). The entire sequence of 374 amino acids has been determined for one subunit of the horse liver ADH enzyme by Jornvall,(1970) while Eklund et al.(1976) have determined the three dimensional crystal structure of the enzyme to a resolution of 2.4A. The horse liver enzyme is highly homologous with ADH enzymes from other animals. In particular the mouse ADH (Edenberg et al.1985) and human β and τ chain ADH enzymes (Hempel et al.,1984; Buhler et al.,1984) are all greater than 85% homologous, at the amino acid level, with the horse liver enzyme. This extensive conservation of amino acid sequence is also found to extend to ADH enzymes from plants, with the amino acid sequences derived from the maize *Adh1* (Dennis et al.,1984) and maize *Adh2* (Dennis et al.,1985) genes being about 50% homologous to the horse liver ADH enzyme (Fig 1.1).

Within each subunit of the horse liver ADH enzyme there is a coenzyme binding domain and a catalytic domain which are separated by a deep hydrophobic cleft. The coenzyme binding domain comprises residues 176-318 and is highly ordered with only 10% of residues having no regular secondary structure. The core structure of this domain is a six stranded pleated sheet similar to that of lactate dehydrogenase (Adams et al.,1970) and malic dehydrogenase (Hill et al.,1972). Binding of the coenzyme analogue ADP-ribose has identified a hydrophobic pocket involved in interactions with adenine moiety of NADH as well as a number of residues involved in hydrogen bonding and steric interactions with the coenzyme. The catalytic domain is derived from the amino and carboxyl ends of the polypeptide and comprises residues 1-175 and 319-374. One

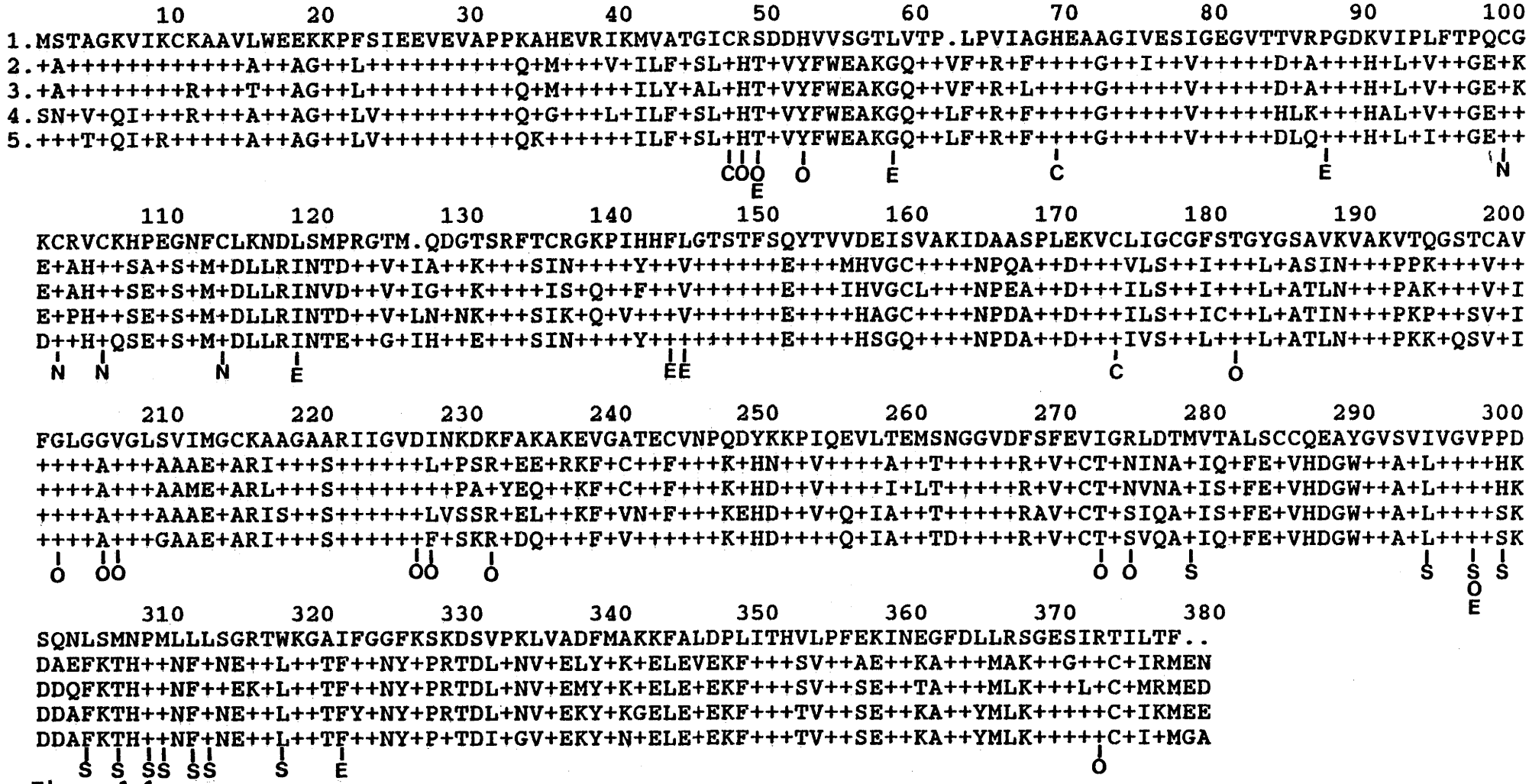


Figure 1.1

Comparison of the amino acid sequences of (1).Horse liver ADH, (2).Maize *Adh1*, (3).Maize *Adh2*, (4).Arabidopsis *Adh1*, (5).pea *Adh1*. The first amino acid of the pea *Adh* sequence was omitted to enable the sequences to be aligned. Amino acids identical to horse liver ADH are indicated by a (+), non identical amino acids are given. Residues presumed to be involved in the catalytic cleft (E), cofactor binding (O), catalytic (C) and non-catalytic (N) zinc binding ligands and subunit interactions (S) are indicated.

histidine and two cysteine residues within this domain provide ligands to the catalytic zinc atom which is located at the bottom of the hydrophobic cleft which separates the two domains. The catalytic zinc atom is bound (in the absence of substrate) in a distorted tetrahedral structure, by coordination of a water molecule or hydroxyl ion depending on pH. This cleft also binds the substrate and the nicotinamide moiety of the coenzyme and constitutes the active site of the enzyme. The second zinc atom is bound in a distorted tetrahedral structure by four cysteine residues and is located on one side of a second cleft in the enzyme surface. It has been suggested that this zinc ion may be involved in stabilization of the protein structure (Drum et al.,1969) or that it may represent a second active site which may no longer be functional.

The extensive amino acid sequence homology between the maize ADH enzymes and horse liver ADH has enabled three dimensional models for the maize polypeptides to be constructed (Eklund and Dennis per. comm.). Only minor structural differences exist between the maize and horse liver ADH's with all the maize side chain substitutions being compatible with the same folding of the main carbon chain as the horse liver ADH enzyme. The maize ADH sequences are five amino acids longer at the carboxyl end relative to horse liver ADH. This region however, is at the surface of the molecule and these additional residues are not thought to significantly alter the three dimensional structure of the enzyme.

This tertiary structure analysis, together with direct amino acid sequence comparisons has allowed the identification of residues in the maize ADH proteins with functional properties equivalent to those from horse liver ADH. These include residues involved in the active site, coenzyme binding, subunit

interactions and those providing ligands for the catalytic and non-catalytic zinc atoms (Fig 1.1).

The coding region of the pea *Adh1* and *Arabidopsis Adh* genes have been determined both by sequence comparison with the maize *Adh1* gene and by cDNA cloning. The *Arabidopsis Adh* gene is expected to encode a polypeptide of 379 amino acids with a subunit molecular weight of approximately 40,000. The pea *Adh1* gene has an additional amino acid in the first exon and is therefore expected to encode a polypeptide of 380 amino acids. Alignment of the predicted amino acid sequence of these genes with the sequence of the maize *Adh1* enzyme is obvious throughout and shows extensive conservation of those residues predicted by Eklund et al.(1976) to be of functional importance (Fig 1.1).

1.1.5 *Adh* nucleotide coding sequences.

The high degree of ADH amino acid conservation between plant species (Fig 1.1) is a result of conservation of the nucleotide coding sequence. This conservation is particularly evident in the first and second positions of each codon. The exceptionally high percentage of guanine nucleotides in the first position of each codon (approximately 40%) results in the high content of hydrophobic amino acids especially valine, alanine and glycine found in the ADH polypeptides. Guanine however is the least frequently used nucleotide in the second position of each codon. This results in relatively high levels of the polar amino acids asparagine and glutamine. Due to the degeneracy of the genetic code essentially random fluctuations are expected in the third position of each codon. This periodical codon pattern (G,non-G,N) found in the *Adh* coding region has also been recognized in a large number of m-RNA transcripts from animals, plants, bacteria and viruses (Trifonov,1987). Numerous suggestion have been made to

account for this non uniform distribution of bases in the codons including; non random mutation (Nassar & Cook,1976), primeval coding pattern (Shepard,1981) and constraints on protein structure and composition (Grantham et al,1980; Bibb et al,1984). Trifonov (1987) has proposed that this pattern may be responsible for monitoring the correct reading frame during translation. This suggestion is supported by the observation that there are several regions within the E.coli 16S rRNA which have complementary C-periodical structure and are known to be involved in rRNA-mRNA interactions.

1.1.6 *Adh* intervening sequences.

Gilbert (1978) has suggested that exons may correspond to units of protein function and that enzymes are built up by random association of these functional units. Introns are therefore presumed to be non-functional sequences which separate the exons. This theory is consistent with the increasing rate of evolution (Gilbert,1978) and can be used to explain why alcohol dehydrogenase, glyceraldehyde-3-phosphate (Buehner et al.,1973) and lactate dehydrogenase (Adams et al.,1970) all have similar coenzyme binding domains yet their catalytic domains are quite different.

Nucleotide sequence analysis of genomic and cDNA clones from the maize *Adh1*, maize *Adh2* and pea *Adh1* genes shows that these genes are all interrupted by nine introns. The positions of these nine introns are identical in each gene although their length and sequence differ. Chang and Meyerowitz (1986) have shown that the *Arabidopsis Adh* gene has only 6 introns whose positions correspond to six of the nine introns in maize and pea *Adh*.

Naora and Deacon (1982) have attempted to correlate the relationship between the total size of exons and introns in

protein-coding genes of higher eucaryotes. Using the correlation developed in this study it would be concluded that the maize *Adh* genes are intron deficient and that perhaps introns were lost prior to the divergence of the monocot and dicot lineages. If this is the case then it is difficult to assess, for the plant *Adh* genes, Gilbert's hypothesis of exons encoding discrete functional units. Nevertheless, Branden et al.(1984) have attempted to correlate the exons of maize *Adh1* with functional domains determined from the 3D crystal structure of the horse liver ADH enzyme. Introns 1,2,3 and 9 are all found within the catalytic domain while intron 7 divides the coenzyme domain into two $(\alpha\beta)_3$ supersecondary structural units. One of these units is further subdivided by introns 5 and 6 into three $(\alpha\beta)_1$ structural units. Intron 4 separates the two domains of the enzyme. If Gilbert's hypothesis is correct the low intron/exon size ratio of the maize *Adh* genes is a result of loss of introns which once subdivided these supersecondary structures.

The research described in chapter 2 of this thesis is directed toward an understanding of the molecular structure of an inducible *Adh* gene from wheat. The agronomic importance of wheat makes such an analysis worthwhile in terms of providing a potentially species specific anaerobic inducible promoter and an easily assayable genetic marker for the future transformation of wheat. Subclones of this gene would also be useful in RFLP mapping of wheat and in determining the genetic relationships between wheat and other members of the Gramineae family.

In chapter 2 we describe the molecular cloning and characterization of a wheat *Adh* gene. However, during our analysis of the wheat gene we became aware of extensive, although discontinuous structural homology between this gene and an *Adh*

gene from barley which Martin Trick from the Genetics Department at Cambridge University had been studying. Subsequent analysis revealed that the wheat and barley genes are evolutionarily closely related and that the structural discontinuity was a consequence of an insertion in the wheat *Adh* gene. This insertion was found to have structural features very similar to the class of transposable elements having short inverted repeats found in other plant species. The potential use of such an element in gene tagging or as an integration module in wheat transformation resulted in a change in direction of the project and a focussing on the structure of the insertion element, its distribution within the wheat genome and experiments designed to induce genetic instability at the *Adh* locus. This work is described in chapter 3 of this thesis. To place this research in perspective I have included a brief synopsis of known plant transposable elements and insertion sequences.

1.2 PLANT TRANSPOSABLE ELEMENTS.

In Mendelian genetics the genome is envisaged as a stable ordered array of loci. However, the discovery of controlling elements with their ability to excise and transpose, causing altered gene expression and chromosomal rearrangements, has radically altered our perceptions from one of a fixed to a more dynamic genome.

1.2.1 The maize controlling elements.

Controlling elements were first described in maize (McClintock,1948,1949) and have subsequently been found in a wide variety of both procaryotes and eucaryotes (see review, Shapiro, 1983). The most extensively studied plant controlling element family both in terms of genetics and molecular structure is Ac-Ds

of maize. *Activator* (Ac) is the autonomous member of the Ac-Ds family and is an inherently unstable allele capable of self excision and transposition (McClintock,1950a). *Ds* (*Dissociation*) is dependent on the presence of *Ac* for activity (McClintock,1951a). The *Ac* element was first recognized by its ability to cause chromosome breakage at a *Ds* locus (McClintock,1945,1946). *Ac* induced chromosome breakage at *Ds* was identified by the uncovering of recessive alleles as a result of partial chromosome loss. Cytogenetic studies have shown that breakage at *Ds* induces a series of breakage-fusion-bridge cycles (McClintock,1942). The acentric chromosome fragments are lost and the dicentric chromatids generated are subsequently broken at mitosis. This breakage resulted in the sequential loss of dominant alleles and the uncovering of recessive alleles on the unbroken homologous chromosome.

Ac can also mediate the transposition of *Ds* resulting in insertions of *Ds* either within or near a locus. Depending on the site of insertion in the host gene a change in phenotype may occur (McClintock, 1951a,b). *Ac* mediated somatic transposition of *Ds* can in some cases also be accompanied by chromosomal rearrangements McClintock (1950b). *Ac* elements have also been shown to give rise to *Ds* like derivatives (McClintock,1955,1956) suggesting that *Ac* and *Ds* elements are structurally related.

Pohlman et al. (1984) have cloned and determined the nucleotide sequence of an *Ac* element from the *wx-m9* allele. The *Ac* element was shown to be 4.5 kb long and to have an imperfect 11bp terminal inverted repeat. The insertion event appears to have resulted in an 8 bp direct duplication of the host DNA. It was speculated (Pohlman et al., 1984) that two large open reading frames in the *Ac* element may encode two polypeptides similar to

the transposase and resolvase enzymes of the bacterial transposon Tn3 (see Review Shapiro,1983). However, Kunze et al.(1987), by cDNA cloning of transcripts derived from Ac, have shown that the element encodes a single polypeptide of 807 amino acids containing four introns. By analogy with the P element of *Drosophila* this polypeptide is very likely to be the transposase enzyme which can act in trans on *Ds* elements and cause their transposition.

Since most mutations caused by transposition are probably deleterious to the plant there is likely to be selection pressure for a low level of transposition and hence a low number of autonomous elements. Southern analysis indicates that there are usually very few active copies of *Ac* and 40 to several hundred copies of the *Ds* elements in the maize genome. The relatively large number of copies of *Ds* is thought to result from *Ds* transposition during chromosome replication (Greenblatt, 1974). Replicative transposition as found in procaryotes is not known to occur in maize.

A number of *Ds* elements have been cloned and shown to have inverted repeats similar to *Ac* and to be flanked by direct duplications of 8 bp of host DNA. Sequence comparisons with *Ac* indicates that many *Ds* elements including Ds9 (Pohlman et al.,1984), Ds, *sh-m5933* (DÖring et al.,1984) and the *Ds* element from the *sh-m6233* (Weck et al.,1984) allele are all derived from *Ac* by internal deletions. In the case of Ds9 the deletion of only 194 bp is sufficient to render the element non-autonomous. The Ds1 element cloned by Peacock et al.(1983), and a very similar *Ds* element from the *wx-m1* allele (Wessler et al.,1986) may be an exception to this rule with only the 11 bp inverted repeat being homologous to Ac. The observation of *Ac* mediated transposition of both Ds1 and *Ds wx-m1* indicates that the 11 bp inverted repeat is

essential for recognition and transposition.

A number of revertants arising from excision of *Ac* and *Ds* elements have been analyzed at the molecular level and are described in Table 1.1. The excision events are most often imprecise and in only one instance have resulted in restoration of the progenitor sequence. Depending on the site of insertion of the element these imprecise excisions can result in revertants having phenotypes that differ from wild type (Wessler et al.,1986;Dennis et al.,1986).

A number of other maize transposable elements are now known to have structural features similar to Ac-Ds. The *Spm* (*Suppressor-mutator* or *En*) controlling elements were first recognized in maize by Peterson (1953). As in Ac-Ds, the *Spm* family also includes autonomous and non-autonomous or "defective" elements (see review Fedoroff,1983). The autonomous elements, like Ac, can mediate both chromosome breakage and somatic excision of defective *Spm* elements. Pereira et al.(1985) have cloned and determined the nucleotide sequence of an unstable mutation caused by the insertion of an autonomous *Spm* element (*En-1*) in the *wx* gene of *Zea mays*. Molecular analysis indicates that *En-1* is 8.4 kb long, has a 13 bp inverted terminal repeat and generates a 3 bp target site duplication. Several internal inverted terminal repeats were also found in the element suggesting extensive secondary structure which may be involved in the recognition and transposition of this element. Comparative sequence analysis of *En-1* and a defective *Spm* element (*Spm-18*) from the *wx-m8* allele (Schwartz-Sommer et al.,1984) indicates that the defective element, like Ds, is derived by internal deletion of an autonomous element. Schwarz-Sommer et al.(1984) have used the *Spm-18* element as a probe in Southern blot analysis of maize DNA and have shown that also like

TABLE 1.1

Physical characteristics of plant transposable elements.

Element	Size (kb)	Locus	Wild type Sequence	Terminal region of insertion element (Direct genomic duplications are in bold lettering)	Revertants		References
Ac9	4.563	wx-m9	-CATGGAGA-	-**CATGGAGA**TAGGGATGAAA----------TTTCATCCCTG**CATGGAGA**-	-CATGGAGATGGAGA-		Pohlman et al
Ac	4.563	wx-m7	-GGTCAGGC-	-**GGTCAGGC**TAGGGATGAAA----------TTTCATCCCTG**GGTCACGC**-			Müller-Neuman et al 1984
Ds9	4.369	wx-m9	-CATGGAGA-	-**CATGGAGA**TAGGGATGAAA----------TTTCATCCCTG**CATGGAGA**-			Pohlman et al 1984
Ds1	0.45	adh1-Fm335	-GGGACTGA-	-**GGGACTGA**TAGGGATGAAA----------TTTCATCCCTA**GGGACTGA**-			Sutton et al 1984
					-GGGACTGTCGGACTGA-		Sachs et al 1983
					-GGGACTGTCCGACTGA-		" " " "
					-GGGACTGTCACTGA-		" " " "
					-GGGACTGGGACTGA-		" " " "
					-GGGACTGGGACTGA-		Peacock et al 1984
					-GGGACTGGACTGA-		" " " "
					-GGGACTGTCGGACTGA-		" " " "
					-GGGACTGA-		" " " "
Ds	0.409	wx-m1	-GCATCACC-	-**GCATCACC**TAGGGATGAAA----------TTTCATCCCTA**GCATCACC**-	-GCATCACGGCCATCACC -GCATCACCATCACC-		Wessler et al 1986
Ds	2.0[1]	sh-m5933	-CGAAGTGG-	-**CGAAGTGG**TAGGGATGAAA----------TTTCATCCCTA**CGAAGTGG**-			Döring et al 1984
Ds	2.0[1]	sh-m6233	-CTTGTCCC-	-**CTTGTCCC**TAGGGATGAAA----------TTTCATCCCTA**CTTGTCCC**-	-CTTGTCCTTGTCCC-		Weck et al 1984
En-1	8.4	wx-844	-ATA-	-**ATA**CACTACAAGAAAA----------TTTTCTTGTAGTG**ATA**-			Pereira et al 1985
Spm-18	2.0	wx-m8	-GTT-	-**GTT**CACTATTAGAAAA----------TTTTGTAATAGTG**CTA**-			
Mu1	0.1367	adh1-S3034	-TTTTGGGGA-	-**TTTTGGGGA**GAGATAATTGCCATT--215bp[4]-AATGGCAATTATCTC**TTTTGGGGA**-			Barker et al 1984
Cin1-001	0.691	nfl[3]	[2]	-**ACTCGT**GTTGG----------CCAACA**TTCTCG**-			Shepard et al 1984
Cin1-002	0.690	LC102[3]	"	-**ATAATT**GTTGG----------CCAACA**ATAAT**-			" " " "
Cin1-003	0.520	LC103[3]	"	-**GTTAGT**GTTGG----------CCAATA**GTTAG**-			" " " "
Tan1	17.0	niv53	-ATA-	-**ATA**CACTACAACAAAA----------TTTTGTTGTAGTG**ATA**-			Bonas et al 1984
Tan2	5.0	niv44	-GTA-	-**GTA**CTACACTACAACAAAAA----------TTTTTGTTGTAGTG**GTA**-			Upadhyaya et al 1985
Tan3	3.5	niv98	-ATCTCAGC-	-**ATCTCAGC**TAAAAGATGTGAA----------TTCACATCTTTTA**ATCTCAGC**-	-ATCTCAGGCTACCATACTCTCAGC-	Sonner et al 1985	
Tan3	3.5	Pal[rec]	-TACCC-	-**TACCC**TAAAGATGTGAA----------TTCACATCTTTTA**TACCC**-	-TACCC-	" "	
					-TAACCC-	" "	
					-CC-	" "	

[1].The DNA insert is considerably larger, but the size of the Ds element is approx. 2kb.

[2].Wild type sequence was inferred by homology with similar regions form other maize lines LC1 and TG2.

[3].This is the particular maize line containing the insertion, the actual locus is unknown.

[4].The inverted repeat extends for 215bp.

Ac-Ds there are multiple copies of *Spm* related sequences in the maize genome.

Robertsons Mutator (Mu) is another system in maize characterized by frequent mutations, many of which are unstable (Robertson,1978; Robertson and Mascia,1981). A *Mu1* related insertion (S3034) in the maize *Adh1* allele has been identified and cloned by Barker et al.(1984). The element is 1367 bp long, is flanked by a 9 bp direct target site duplication and has two large terminal inverted repeats of 215 bp. These inverted repeats are perfectly matched over the terminal 28 bp. Copy number estimates for this element range from 10 to 50 (Bennetzen et al.,1984).

1.2.2 Maize insertion sequences.

Three other putative maize transposable elements *Cin1*, *Tz86* and *Bs1* have also been identified and analyzed at the molecular level. These elements have structural features in common with other transposable elements and are present in multiple copies in the maize genome. However, they are not characterized by unstable mutations and have generally been identified by sequence comparison with closely related alleles. For this reason these sequences are more correctly termed "insertion elements".

Cin1-001 (Shepard et al.,1982) was identified as a 700bp insertion in a clone from line LC1 of *Zea mays* that was not present in a Northern flint line (nfl). This insertion has a 6 bp terminal inverted repeat and is flanked by 5 bp direct repeats. Shepard et al. (1984) have cloned and sequenced two related *Cin1* elements, *Cin1*-102 and *Cin1*-103. Both elements have similar inverted repeats to the *Cin*-001 element and are each flanked by a 5 bp direct duplication.

Tz86 (Dellaporta et al.,1986) and *Bs1* (Mottinger et al.,1984) were both identified as insertion elements following infection of

maize with barley streak mosaic virus (BSMV) (Sprague et al.,1963). Dellaporta et al (1986) have suggested that the apparent genetic instability resulting from BSMV infection may be a result of genomic shock (McClintock,1978) in which the genome is rapidly restructured by the mobilization of transposable elements. The *Tz86* element was identified as a 3.6 kb insertion in the 3' region of the *Sh1-B* allele. The insertion has produced a 10 bp direct duplication but unlike other transposable elements has no terminal inverted repeat. Numerous minor internal direct and inverted repeats have been found in the element. Similarly, *Bs1* was identified as a 3.3 kb insertion in the maize *Adh1* gene following BSMV infection. The insertion is flanked by a 6 bp target site duplication and has terminal direct repeats of 304 bp (Johns et al.,1985). This structure although unlike plant transposable elements, is similar to Ty1 in yeast and copia like elements from *Drosophila*.

1.2.3 Transposable elements in other plant species.

Transposable elements have also been recognized in other plants including snapdragon and soybean. *Tam1* and *Tam2* are related but independent insertions in the *nivea* locus of *Antirrhinum majus* which encodes chalcone synthase, a key enzyme in anthocyanin biosynthesis. *Tam1* is a 17 kb insertion in the promoter region of the chalcone synthase gene and has a 13 bp terminal inverted repeat and induces a 3 bp duplication of the target site upon integration (Bonas et al.,1984a). Somatic instability, leading to variegated flower phenotype is due to frequent excision of this element (Bonas et al.,1984b). *Tam2*, which appears to be a deletion derivative of *Tam1*, is a 5 kb insertion element which has an identical inverted repeat to *Tam1* except that it is one nucleotide longer (Upadhyana et al.,1985). This element is located in the

first intron/exon boundary and is also flanked by a 3 bp target site duplication. There is no evidence for genetic instability resulting from excision of this element. Both the *Tam1* and *Tam2* elements have extensive structural homology with the *En-1* and *Spm-18* elements of *Zea mays*. All four elements generate a 3 bp target site duplication and have considerable secondary structure resulting from numerous internal inverted repeats. Furthermore apart from one mismatch the inverted repeats of all these elements are identical. This extensive structural conservation of transposable elements from different plant species might indicate that all these elements originated from a common progenitor or that horizontal spread of the element occurred. Alternatively, convergent evolution of these elements may have resulted in the selection of a particular structure.

Two independent *Tam3* elements have been cloned and the nucleotide sequence determined. Both elements are integrated in the promoter regions of genes involved in flower pigmentation and are 3.5 kb long. The *Tam3* insertion in the *nivea* locus (Sommer et al.,1985) which encodes chalcone synthase has a 12 bp terminal inverted repeat and like Ac-Ds generates an 8 bp duplication of target site sequences upon integration. The unstable *Tam3* element from the *pallida* locus (Coen et al.,1986) has an identical 12 bp inverted repeat to the *nivea* element but results in only a 5 bp target site duplication. Imprecise excision of this element has been correlated with altered spatial patterns and intensities of flower pigmentation.

Goldberg et al.(1983) have identified a naturally occurring null mutant of the soybean Le1 gene which encodes seed lectin. Sequence comparison of the mutant and functional genes (Vodkin et al.,1983) has identified a 3.4 kb insertion (*Tgm1*) in the coding

region. This insertion has a 13 bp inverted repeat and is flanked by 3 bp direct duplications. The nucleotide sequence of the inverted repeat is very similar to the *Spm-18* and *En-1* elements of *Zea mays* as well as the *Tam1* and *Tam2* elements from *Antirrhinum majus*.

1.2.4 Related transposable elements from animals.

Transposable elements from other eucaryotes, including the P elements of *Drosophila melanogaster* (which are involved in hybrid dysgenesis), and the *Tc1* element from *Caenorhabditis elegans* have also been found to have short inverted repeat structures similar to plant transposable elements.

From a structural point of view the P element has three large open reading frames, a 31 bp inverted repeat and is flanked by an 8 bp direct duplication of the target site DNA (O'Hare and Rubin,1983). An unusual feature of the P elements is that they appear to have some target site specificity (O'Hare and Rubin,1983). Rio et al.(1986) have shown that the three large open reading frames of the P element encode a biologically active transposase.

The *Tc1* element from *Caenorhabditis elegans* (Rosenzweig et al.,1983a) is a 1610 bp insertion element which is present in multiple copies in the genome and has a 54 bp inverted repeat and is flanked by a 2 bp direct duplication. There is, however, some evidence to suggest that this direct duplication may form part of the inverted repeat structure (Rosenzweig et al.,1983b) as the direct repeats are also complementary to each other.

Transposable elements with structures different to those found for the plant transposable element have been identified in a number of other eucaryotes including *Drosophila*, mice, butterflies and yeast. However, a detailed discussion of these elements is beyond the scope of this introduction.

1.3 WHEAT EVOLUTION.

The presence of the insertion element in the cloned wheat *Adh* gene described in chapter 2 and its corresponding absence in a related gene from barley raises the question: when did the insertion event occur? This inevitably leads into a study of the evolution of wheat. During the course of this study we found that the cloned *Adh* gene, together with the insertion element as a molecular tag, provided useful probes not only for analyzing the timing of the insertion event but also for analyzing the evolution of the diploid and polyploid wheats. These results are described in chapter 4 of this thesis.

1.3.1 Classification.

The cultivated wheats belong to the *Triticum* genus, one of seven genera in the Triticinae subtribe, the wild diploid relatives of wheat were classified in a different genus; *Aegilops*. Subsequently, Stebbins (1956) and Bowden (1959) proposed that the *Aegilops* genus be reclassified and combined with the *Triticum* genus. This reclassification has become widely accepted and for the purpose of clarity will be used throughout this thesis.

The taxonomic classification of the *Triticum* species is difficult due to introgression (Vardi, 1973), polyploid hybridization (Zohary and Feldman, 1962) and the suppression of characters of generic importance in wide crosses involving polyploids (Kimber, 1983). As a result of these limitations karyotype analysis has been used as a means of determining evolutionary relationships in the *Triticum* species. Numerous techniques and criteria have been used in the genomic classification of the diploid wheat species. These include centromeric position, number and size of satellited chromosomes (Waines and Kimber, 1972), giemsa banding patterns, *in-situ*

hybridization and protein electrophoresis (Jaaska, 1976; Johnson, 1975). Probably the most successful method for recognition of genomic similarities has been by the "analyzer method" (see review Lilienfeld,1951). In this analysis a hybrid is made and chromosome pairing at meiosis is taken as an indication of genomic similarity. Numerical methods (Kimber et al.,1981) have also been developed that allow quantitation of both the amount and pattern of genomic affinities in hybrids. Using these methods eleven different diploid genomes have been recognized (Table 1.2).

1.3.2 Polyploidy in the wheats. AND IT'S RELATIVES.

~~The wild and cultivated species of~~ Wheat IS known to exist at three levels of ploidy; diploid, tetraploid and hexaploid. However, interspecific hybridization within diploid species of the *Triticum* genus is very rare. The only species in which interspecific hybrids have been well documented are *T. longissimum* and *T. sharonensis* (Ankori and Zohary,1962). This virtual absence of natural hybridization is a result of evolutionary divergence of the eleven diploid species and subsequently reduced pairing affinities at meiosis resulting in sterile hybrids. This suggests that the polyploid wheat species are unlikely to have arisen by interspecific hybridization followed by chromosome doubling as is commonly thought. An alternative mechanism to account for polyploidy has been proposed by deWet (1980). In this scheme, failure of the first meiotic division results in rare, unreduced diploid gametes. Such gametes, because of their diploid character, are capable of interspecific hybridization with a normal haploid gamete to form a triploid sporophyte (2n+n). The diploid nature of the unreduced gamete is thought to act as a buffer ensuring some fertility in the resulting hybrid. Subsequent meiotic failure producing triploid gametes followed by fertilization by haploid

TABLE 1.2 *

GENOMIC CONSTITUTION OF DIPLOID AND POLYPLOID WHEATS

	Species	Genomic Formula
Diploids	*T. monococcum boeoticum*	AA
	T. monococcum urartu	AA
	T. dichasians	CC
	T. tauschii	DD
	T. comosum	MM
	T. tripsacoides	MtMt
	T. uniaristatum	MuMu
	T. speltoides	SS
	T. bicorne	$S^b S^b$
	T. sharonensis	$S^1 S^1$
	T. longissimum	$S^1 S^1$
	T. searsii	$S^s S^s$
	T. umbellulatum	UU
Polyploids sharing A genome	*T. turgidum dicoccoides*	AABB
	T. timopheevi araraticum	AAGG
	T. aestivum	AABBDD
Polyploids sharing D genome	*T. cylindricum*	DDCC
	T. crassum	$DDM^{c\,r}M^{c\,r}$
	T. crassum	$DDD2D_2\,M^{c\,r}M^{c\,r}$
	T. syriacum	$DDM^{c\,r}M^{c\,r}S^1S^1$
	T. juvenale	$DDM^{c\,r}M^{c\,r}UU$
	T. ventricosum	$DDMM^vM^v$
Polyploids sharing U genome	*T. triunciale*	UUCC
	T. macrochaetum	UUM^bM^b
	T. columnare	UUM^cM^c
	T. triaristatum	UUM^tM^t
	T. triarestatum	$UUM^fM^fM^{f\,2}M^{f\,2}$
	T. ovatum	UUM^oM^o
	T. variabile	UUS^vS^v
	T. kotschyi	UUS^vS^v

* This table was adapted from Feldman and Sears, 1981.

gametes of the diploid parent, results in tetraploid progeny. Similar, meiotic failure in the tetraploids would eventually give rise to the hexaploid species.

1.3.3 The pivotal genome hypothesis.

The pivotal-differential hypothesis was first proposed by Zohary and Feldman (1962) to account for the extensive interspecific hybridization found to occur between polyploid wheats and, to a lesser extent, polyploid and diploid wheats. In this theory, which is analogous to the meiotic failure hypothesis of deWet, interspecific hybridization is thought to be facilitated by the shared genome acting as a buffering agent ensuring at least some fertility in the resulting hybrids. There is therefore selection pressure to maintain an essentially unaltered genome while the other genome is free to recombine.

Cytogenetic evidence has been used to classify the genome constitution of the polyploid wheats. Using the analyzer method three groups have been recognized within the polyploid wheat species. The species within each group share one common genome (the pivotal genome(s)) while the other genome(s) is different (differential genome). One of the species clusters is based on the U genome of *T. umbellulatum*, one on the D genome of *T. tauschii* and a third on the A genome of *T. monococcum* (see Table 1.2).

The pivotal differential pattern of evolution has been well documented for the U genome cluster of polyploid species (Zohary and Feldman,1962; Feldman, 1965). Zhao and Kimber (1984) however, have shown that the D genome cluster can be further subdivided into three groups based on how closely the common genome is related to the D genome of the diploid progenitor species *T. tauschii*. *T. cylindricum* and *T. ventricosum* appear to be very similar to the D diploid progenitor while the D genome of both the

tetraploid and hexaploid species of *T. crassum* and *T. juvenale* and *T. syriacum* are progressively more divergent. Hence the D genome cluster of polyploid species only loosely conforms to the pivotal-differential hypothesis. The A genome cluster of tetraploid and hexaploid species, which includes all the cultivated forms of wheat, are thought to have an A genome essentially unaltered from the diploid progenitor species *T. monococcum*.

Two wild tetraploid species in the A genome polyploid group have been identified; *T. timopheevi var. araraticum*, having the genome constitution AAGG, and *T. turgidum var. dicoccoides* with the genome constitution AABB. In the absence of an obvious diploid progenitor for either the B or G genomes a number of models have been proposed to account for the origin of the B and G genomes. These include monophyletic, diphyletic and introgressive models and are discussed in more detail in Chapter 4.

From the above it can be seen that the pivotal-differential genome hypothesis occupies a central position in our understanding of the evolution of the polyploid wheats. At the molecular level this concept, of an unaltered genome and a differential genome within the one species, would predict a lower frequency of polymorphic changes for pivotal genome derived loci relative to similar loci from the differential genome(s). In Chapter 4 this hypothesis is tested for the A genome polyploid cluster using the cloned wheat *Adh* gene described in chapter 2. In addition we have used the RFLP data from the above analysis to provide supporting evidence for a monophyletic/introgressive origin of the B and G genomes of the tetraploid wheats. The insertion element has also been used as a molecular probe to study the evolution of the wheat *Adh* locus to more accurately define when the insertion event occurred.

1.4 References

Adams, M.J., Ford, G.C., Koekok, R., Lentz, P.J. Jr., McPherson, A. Jr., Rossmann, M.G., Smiley, I.E., Schevitz, R.W. and Woncott, A.J. (1970). Structure of lactate dehydrogenase at 2.8A resolution. Nature. 227, 1098-1103.

Ankori, H. and Zohary, D. (1962). Natural hybridization between *Aegilops sharonensis* and *Ae. longissimum*. Cytologia 27, 314-324.

Barker, R.F., Thompson, D.V., Talbot, D.R., Swanson, J. and Bennetzen, J.L. (1984). Nucleotide sequence of the maize transposable element Mul. Nucl. Acids. Res. 12, 5955-5967.

Bennetzen, J.L., Swanson, J., Taylor, W.C. and Freeling, M. (1984). DNA insertion in the first intron of maize *Adh*, affects message levels - cloning of progenitor and mutant *Adh* alleles. Proc. Natl. Acad. Sci. USA 81, 4125-4128.

Bib, M.J., Finlay, P.R. and Johnson, M.W. (1984). The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein coding sequences. Gene, 30, 157-166.

Bonas, V., Sommer, H. and Saedler, H. (1984a). The 17kb *Tam1* element of *Antirrhinum majus* induces a 3-bp duplication upon integration into the chalcone synthase gene. Embo J. 3, 1015-1019

Bonas, V., Sommer, H., Harrison, B.J. and Saedler, H. (1984b). The transposable element *Tam1* of *Antirrhinum majus* is 17kb long. Mol. Gen. Genet. 194, 138-143.

Bowden, W.M. (1959). The taxonomy and nomenclature of the wheats, barleys, and ryes and their wild relatives. Can. J. Bot., 37, 130-136.

Branden, C.I., Jornvall, H., Eklund, H. and Furugren, B. (1975) Alcohol dehydrogenase in The Enzymes P.B. Boyer, ed., Academic Press 104-190.

Branden, C-I., Eklund, H., Cambillau, C and Pryor, A.J. (1984). Correlation of exons with structural domains in alcohol dehydrogenase. EMBO J. 3, 1307-1310.

Buehner, M., Ford, G.C., Moras, D., Olsen, K.W. and Rossmann, M.G. (1973). D-glyceraldehyde-3-phosphate dehydrogenase: three dimensional structure and evolutionary significance. Proc. Natl. Acad. Sci. U.S.A. 70, 3052-3054.

Buhler, R., Hempel, J., Kaiser, R., de Zalenski, C., von Wartburg, J-P and Jörnvall, H. (1984). Human liver alcohol dehydrogenase 2. The primary structure of the τ protein chain. Eur. J. Biochem. 145, 447-453.

Chang, C. and Meyerowitz, E.M. (1986). Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. Proc. Natl. Acad. Sci. U.S.A., 83, 1408-1412.

Coen, E.S., Carpenter, R., Martin, C. (1986) Transposable elements generate novel spatial patterns of gene expression in *Antirrhinum majus*. Cell 47, 285-296.

Dellaporta, S.L., Chomet, P.S., Mottinger, J.P., Wood, J.A., Yu, S.M. and Hicks, J.B. (1986) Endogenous transposable elements associated with virus infection in maize. Cold Spring Harbour Symp. Quant. Biol. XLIX, 321-328.

Dennis, E.S., Gerlach, W.L., Pryor, A.J., Bennetzen, J.L., Inglis, A., Llewellyn, D., Sachs, M.M., Ferl, R.J. and Peacock, W.J. (1984). Molecular analysis of the alcohol dehydrogenase (*Adh1*) gene of maize. Nucl. Acids Res. 12, 3983-4000.

Dennis, E.S., Sachs, M.M., Gerlach, W.L., Finnegan, E.J. and Peacock, W.J. (1985). Molecular analysis of the alcohol dehydrogenase 2 (*Adh2*) gene of maize. Nucl. Acids Res. 13, 727-743.

Dennis, E.S., Gerlach, W.L., Peacock, W.J. and Schwartz, D. (1986). Excision of the *Ds* controlling element from the *Adh* gene of maize. Maydica XXXI 47-57.

deWet, J.M.J. (1980). Origins of polyploids. In: Polyploidy biological relevance. Ed. Lewis, W.H., Plenum, New York.

Döring, H.-P., Tillman,E., Starlinger, P. (1984). DNA sequence of the maize transposable element *Dissociation*. Nature 307, 127-130.

Drew, M.C. and Sisworo, E.J. (1977). Early effects of flooding on nitrogen deficiency and leaf chlorosis in barley. New Phytol. 79, 567-571.

Drum, D.E., Li, T-K. and Vallee, B.L. (1969). Zinc isotope exchange in horse liver alcohol dehydrogenase. Biochemistry 8, 3792-3797.

Edenberg, H.M., Zhang, K., Fong, K., Bosron, W.F. and Li, T-K (1985). Cloning and sequencing of a cDNA encoding the complete mouse liver alcohol dehydrogenase. Proc. Natl. Acad. Sci. U.S.A. 82, 2262-2266.

Eklund, H., Nordström, B., Zeppezauer, E., Söderlund, G., Ohlsson, I., Boiwe, T., Söderberg, B-O., Tapia, O., Bränden, C-I., Akeson, A. (1976). Three dimensional structure of horse liver alcohol dehydrogenase at 2.4A resolution. J. Mol. Biol. 102, 27-59.

Fedoroff, N.V. (1983). Controlling elements in maize. In: Mobile Genetic Elements (Shapiro, J.A.,ed) Academic, New York.

Feldman, M. (1965). Further evidence for natural hybridization between tetraploid species of *Aegilops* section *Pleionathera*. Evolution 19, 162-174.

Feldman, M. and Sears, E.R. (1981). The wild gene resources of wheat. Scientific American June, 98-108.

Gerlach, W.L., Pryor, A.J., Dennis, E.S., Ferl., R.J., Sachs, M.M. and Peacock, W.J. (1982). cDNA cloning and induction of the alcohol dehydrogenase gene (*Adh1*) of maize. Proc. Natl. Acad. Sci. USA 79, 2981-2985.

Gilbert, W. (1978). Why genes in pieces? Nature 271, 501.

Goldberg, R.B., Hoschek, G. and Vodkin, L.O. (1983). An insertion sequence blocks the expression of a soybean lectin gene. Cell. 33, 465-475.

Gottlieb, L.D. (1982). Conservation and duplication of isozymes in plants. Science 216, 373-380.

Grantham, R., Gautier, C., Govy, M., Mercier, R. and Pave, A. (1980). Codon catalog usage and the genome hypothesis. Nucl. Acids. Res. 8, r49-r62.

Greenblatt, I.M. (1974). Movement of modulator in maize: a test of an hypothesis. Genetics 77, 671-678.

Hake, S., Kelley, P., Taylor, W.C. and Freeling, M. (1985) Coordinate induction of alcohol dehydrogenase 1, aldolase, and other anaerobic RNAs in maize. J. Biol. Chem. 260, 5050-5054.

Hempel, J., Buhler, R., Kaiser, R., Holmquist, B., de Zalenski, C., von Wartburg, J-P., Vallee, B. and Jörnvall, H. (1984). Human liver alcohol dehydrogenase 1. The primary structure of the β.β. isoenzyme. Eur. J. Biochem. 145, 437-445.

Hill, E., Tsermoglou, D., Webb, L. and Banaszak, L.J. (1972). Polypeptide conformation of cytoplasmic malate dehydrogenase from an electron density map at 3.0A resolution. J. Mol. Biol. 72, 577-591.

Jaaska, J.A. (1976). Aspartate aminotransferase isoenzymes in the polyploid wheats and their diploid relatives. On the origin of tetraploid wheats. Biochem. Physiol. Pflanzen. 170, 159-171.

Jackson, M.B. and Campbell, D.J. (1975). Movement of ethylene from roots to shoot a factor in the response of tomato plants to waterlogged conditions. New Phytol. 74, 397-406.

Johns, M.A., Mottinger, J. and Freeling, M. (1985). A low copy number, copia-like transposon in maize. EMBO J. 4, 1093-1102.

Johnson, B.L. (1975). Identification of the apparent B genome donor of wheat. Can. J. Genet. Cytol. 17, 21-39.

Jörnvall, H. (1970). Horse liver alcohol dehydrogenase. The primary structure of the protein chain of the ethanol active isozyme. Eur. J. Biochem. 16, 25-40.

Kimber, G. (1983). The suppression of characters of generic importance in hybrids and amphiploids in the Triticeae. Cereal Res. Comm. 11, 9-14.

Kimber, G., Alonso, L.C., and Sallee,P.J. (1981). The analysis of meiosis in hybrids. 1. Aneuploid hybrids. Can. J. Genet. Cytol. 23, 209-219.

Kelley, P.M. and Freeling, M. (1984a). Anaerobic expression of maize fructose-1,6-diphosphate aldolase. J. Biol. Chem. 259, 673-677.

Kelley, P.M. and Freeling, M. (1984b). Anaerobic expression of maize glucose phosphate isomerase. I. J. Biol. Chem. 259, 673-677.

Kunze, R., Stochaj, U., Laufs, J. and Starlinger, P. (1987). Transcription of transposable element *activator* (Ac) of *Zea mays* L. EMBO J. 6, 1555-1563.

Lazlo, A. and St. Lawrence, P. (1983) Parallel induction and synthesis of PDC and ADH in anoxic maize roots. Mol. Gen. Genet. 192, 110-117

Lilienfeld, F.A. (1951). H. Kihara: Genome-analysis in *Triticum* and *Aegilops*. Concluding review. Cytologia 16, 101-123.

Lin, C.Y. and Key, J.L. (1967). Dissociation and reassembly of polyribosomes in relation to protein synthesis in the soybean root. J. Mol. Biol. 26, 237-247.

Llewellyn, D.J., Finnegan, E.J., Ellis, J.G., Dennis, E.S. and Peacock, W.J. (1987). Structure and expression of an alcohol dehydrogenase 1 gene from *um sativum* (cv. Greenfeast). J. Mol. Biol. 195, 115-123.

McClintock, B. (1942). The fusion of broken ends of chromosomes following nuclear fusion. Proc. Natl. Acad. Sci. U.S.A. 28, 458-463.

McClintock, B.(1945). Cytogenetic studies of maize and rospora. Carnegie Inst. Washington Year Book 44, 108-112.

McClintock, B. (1946). Maize genetics. Carnegie Inst. Washington Year Book 45, 176-186.

McClintock, B. (1948). Mutable loci in maize. Carnegie Inst. Washington Year Book 47, 155-169.

McClintock, B. (1949). Mutable loci in maize. Carnegie Inst. Washington Year Book 48, 142-154.

McClintock, B. (1950a). The origin and behaviour of mutable loci in maize. Proc. Natl. Acad. Sci. U.S.A. 36, 344-355.

McClintock, B. (1950b). Mutable loci in maize. Carnegie Inst. Washington Year Book 49, 157-167.

McClintock, B. (1951a). Chromosome organization and genetic expression. Cold Spring Harbour symp. Quant. Biol. 16, 13-47.

McClintock, B. (1951b). Mutable loci in maize. Carnegie Inst. Washington Year Book, 50, 174-181.

McClintock, B. (1955). Controlled mutation in maize Carnegie Inst. Washington Year Book 54, 245-255.

McClintock, B. (1956). Controlling elements and the gene. Cold Spring Harbour Symp. Quant. Biol. 21, 197-216.

McClintock, B. (1978). Mechanisms that rapidly reorganise the genome. Stadler Genet. Symp. 10, 25-48.

McMannon, M. and Crawford R.M.M. (1971) Metabolic theory of flooding tolerance: The significance of enzyme distribution and behaviour. New Phytol.70, 299-306.

Mottinger, J.P., Johns, M.A. and Freeling, M. (1984). Mutations of the *Adh1* gene in maize following infection with barley stripe mosaic virus. Mol. Gen. Genet. 195, 367-369.

Nassar, R.F. and Cook, R.D. (1976). Non-randomness of nucleotide bases in mRNA codons. Genet. Res. 27, 353-362.

Naora, H. and Deacon, N.J. (1982). Relationship between the total size of exons and introns in protein coding genes of higher eucaryotes. Proc. Natl. Acad. Sci. U.S.A. 79, 6196-6200.

O'Hare K., and Rubin, G.M.(1983). Structures of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell 34, 25-35.

Peacock, W.J., Dennis, E.S., Gerlach, W.L., Llewellyn, D., Lörz, H. (1983). Gene transfer in maize. Controlling elements and the alcohol dehydrogenase genes. Proc. 15th Miami Winter Symp. ed. K. Downry, R.w. Vollmy, F. Ahmad, J. Schultz, 20, 311-326. New York:Academic.

Pereira, A., Schwarz-Sommer, Z., Gierl, A., Bertram, I., Peterson, P.A. and Saedler, H. (1985). Genetic and molecular analysis of the enhancer (En) transposable element system of *Zea mays*. EMBO. J. 4, 17-23.

Peterson, P.A. (1953). A mutable pale green locus in maize. Genetics 38, 682-683.

Pohlman, R.F., Fedoroff, N.V. and Messing, J. (1984). The nucleotide sequence of the maize controlling element *Activator*. Cell 37, 635-643.

Rio, D.C., Laski, F.A. and Rubin, G.M. (1986). Identification and immunochemical analysis of biologically active P element transposase. Cell 44, 21-32.

Roberts, J.K., Callis, J., Wemmer, D., Walbot, V. and Jardetzky, O. (1984a) Mechanisms of cytoplasmic pH regulation in hypoxic maize root tips and its role in survival under hypoxia. Proc. Natl. Acad. Sci. USA. 81, 3379-3383.

Roberts, J.K., Callis, J., Jardetzky, O., Walbot, V. and Freeling, M. (1984b) Cytoplasmic acidosis as a determinant of flooding intolerance in plants. Proc. Natl. Acad. Sci. USA. 81, 6029-6033.

Robertson, D.S. (1978) Characterization of a mutation system in maize. Mutat. Res. 51, 21-28.

Robertson, D.S. and Mascia, P.N. (1981). Tests of four controlling-element systems of maize for mutator activity and their interaction with Mu-mutator. Mutat. Res. 84, 283-289.

Rozenzweig, B., Liao, L.W. and Hirsch, D. (1983a). Sequence of the *C. elegans* transposable element *Tc1*. Nucl. Acids Res. 11, 4201-4209.

Rozenzweig, B., Liao, L.W. and Hirsch, D. (1983b). Target sequences for the C. elegans transposable element *Tc1*. Nucl. Acids Res. 11, 3437-3440.

Sachs, M.M., Freeling, M. and Okimoto, R. (1980) The anaerobic proteins of maize. Cell 20, 761-767.

Schwartz, D. (1969). An example of gene fixation resulting from selective advantage in sub-optimal conditions. Amer. Nat. 103, 479-481.

Schwarz-Sommer, Zs., Gierl, A., Klosgen, R.B., Wienand, U., Peterson, P.A. and Saedler, H. (1984). The *Spm* (En) transposable element controls the excision of a 2kb DNA insert at the *wx*$^{m-8}$ allele of *Zea mays*. EMBO J. 3, 1021-1028.

Shapiro, J.A. (1983). (ed.) Mobile Genetic Elements. New York, Academic.

Shepard, J.C.W. (1981). Method to determine the reading frame of a protein from the purine/pyrimidine genome sequence and its possible evolutionary justification. Proc. Natl. Acad. Sci. U.S.A. 78, 1596-1600.

Shepard, N.S., Schwarz-Sommer, Z., Wienand, U., Sommer, H., Deumling, B., Peterson, P.A. and Saedler, H. (1982). Cloning of a genomic fragment carrying the insertion element *Cin 1* of *Zea mays*. Mol. Gen. Genet. 188, 266-271.

Shepard, N.S., Schwarz-Sommer, Z., Blumberg, J., Gupta, M., Wienand, U. and Saedler, H. (1984). Similarity of the *Cin 1* repetitive family of *Zea mays* to eucaryotic transposable elements. Nature. 307, 185-187.

Sommer, H., Carpenter, R., Harrison, B.J. and Saedler, H. (1985). The transposable element *Tam3* of *Antirrhinum majus* generates a novel type of sequence alteration upon excision. Mol. Gen. Genet. 199, 225-231.

Sprague, G.F., McKinney, H.H., Greeley, L. (1963). Virus as a mutagenic agent in maize. Science, 14, 1052-1053.

Stebbins, G.L. (1956). Taxonomy and the evolution of genera, with special reference to the family Gramineae. Evolution 10, 235-245.

Trifonov, E.N. (1987). Translation framing code and frame monitoring mechanism as suggested by the analysis of mRNA and 16s RNA nucleotide sequences. J. Mol. Biol. 194, 643-652.

Upadhyana, K.C., Sommer, H., Krebbers, E. and Saedler, H.(1985). The paramutagenic line niv-44 has a 5kb insert, *Tam2* in the chalcone synthase gene of *Antirrhinum majus*. Mol. Gen. Genet. 199, 201-207.

Vardi, A. (1973). Introgression between different ploidy levels in the wheat group. Proc. 4th Int. Wheat Genet. Symp. Columbia, Mo. 131-141.

Vodkin, L.O., Rhodes, P.R. and Goldberg, R.B. (1983). A lectin gene insertion has the structural features of a transposable element. Cell 34, 1023-1031.

Waines, J.G. and Kimber, G. (1972). Satellite number and size in *T. monococcum* and the evolution of polyploid wheats. Can. J. Genet. Cytol. 15, 117-122.

Walker J.C., Howard E.A., Dennis E.S., Peacock W.J. (1987). DNA sequences required for anaerobic expression of the maize alcohol dehydrogenase-1 gene. Proc. Natl. Acad. Sci. USA (in press).

Weck, C., Courage, U., Döring, H.-P., Fedoroff, N. and Starlinger, P. (1984). Analysis of *sh-m6233*, a mutation induced by the transposable element *Ds* in the sucrose synthase gene of *Zea mays*. EMBO J. 3, 1713-1716.

Werr, W., Frommer, W.-B., Maas, C. and Starlinger, P. (1985) Structure of the sucrose synthase gene on chromosome 9 of *Zea mays* L. EMBO J. 4, 1373-1380.

Wessler, S.R., Baran, G., Varagona, M. and Dellaporta, S.L. (1986). Excision of *Ds* produces waxy proteins with a range of enzymatic activities. EMBO. J. 5, 2427-2432.

Wignarajah, K., Greenway, H.(1976). Effect of anaerobiosis on activities of alcohol dehydrogenase and pyruvate decarboxylase in roots of *Zea mays*. New Phytol. 77, 575-584.

Zhao, Y.H. and Kimber, G. (1984). New hybrids with D genome wheat relatives. Genetics 106, 509-515.

Zohary, D. and Feldman, M. (1962). Hybridisation between amphidiploids and the evolution of polyploids in the wheat (*Aegilops - Triticum*) group. Evolution 16, 44-61.

CHAPTER 2

Molecular analysis of an Alcohol Dehydrogenase (*Adh*) gene from Chromosome 1 of wheat.

L. E. Mitchell*, E. S. Dennis, W. J.Peacock

Division of Plant Industry
CSIRO
G.P.O. Box 1600
Canberra, A.C.T 2601, Australia

* Also from Research School of Biological Sciences
Australian National University
Canberra

Abbreviations: ADH alcohol dehydrogenase,
NT nullisomic-tetrasomic
IEF isoelectric focussing

2.1 Introduction

Alcohol dehydrogenase (ADH; alcohol: NAD^+ oxidoreductase, EC1.1.1.1) activity has been found in numerous higher plants including wheat (Hart,1970), barley (Hanson and Brown,1984), maize (Schwartz,1971), pea (McMannon and Crawford,1971) and *Arabidopsis* (see review Gottlieb,1982).

The enzyme is expressed under hypoxic conditions such as are found in dry seeds or after prolonged flooding of seedlings. One function of the enzyme is to convert acetaldehyde to ethanol thereby regenerating NAD^+ which is necessary for glycolytic ATP production during anaerobiosis (Freeling and Birchler,1982).

The most extensive study of plant *Adh* genes has been in maize where both *Adh1* and *Adh2* have been cloned and sequenced (Dennis et al.,1984,1985). Both maize ADHs are NAD^+ dependent dimeric enzymes belonging to a small group of about 20 proteins that are selectively expressed under anaerobic conditions (Sachs et al. 1980). Several of the wheat ADH's are similar to the maize ADH's in that they are also dimeric enzymes with a MW of approximately 80,000 and in some cases are anaerobically inducible (Langston et al.,1979,1980; Hart,1980). However, the cultivated wheats are hexaploid and consequently most of the gene loci including *Adh* are triplicated in these species.

To date three NAD^+ dependent alcohol dehydrogenases as defined by their different substrate specificities, have been identified in wheat and mapped to specific chromosomes using the aneuploid lines of *T.aestivum* cv Chinese Spring developed by Sears (1966). Alcohol dehydrogenase 1 (ADH-1) is produced in the seed and is active on aliphatic alcohols (Hart,1969). Aneuploid studies indicate that there is a triplicate set of *Adh1* genes located on chromosome arms 4Ap, 4Bp and 4Dp (Hart,1970). A second triplicate

set of *Adh* genes has been found closely linked to *Adh1* but is induced in seedlings only under anaerobic conditions (Hart,1980). Alcohol dehydrogenase 3 (ADH-3) is encoded by a triplicate set of genes located in the 6q arms and utilizes cinnamyl alcohol as a substrate. Jaaska (1978) has identified an NADP dependent ADH which can utilize cinnamyl alcohol as a substrate and has mapped the three molecular forms of the enzyme to the 5q arms of Chinese Spring. Heterodimers of the enzyme are not found suggesting that the enzyme is monomeric (See review Hart,1983).

Northern blot analysis of wheat RNA showed sufficient homology to a maize *Adh1* probe to permit isolation of the wheat *Adh* genes. A cloned wheat *Adh* gene may have potential as a genetic marker for transformation of wheat as well as providing a useful probe for RFLP mapping of ancestral wheats and other members of the Gramineae family. A cloned wheat *Adh* gene would also be useful as a molecular probe to analyze particular alien chromosome substitutions or additions in wheat backgrounds.

We present here the molecular cloning and characterization of an inducible *Adh* gene from a previously unknown series of *Adh* genes found on chromosome 1 of hexaploid wheat.

2.2 Materials and methods

(a) Plant material: Bulk seed of hexaploid wheat (*Triticum aestivum*) cv. Millewa, cv. Chinese Spring and Chinese Spring compensating nullisomic-tetrasomic lines were obtained from P. Larkin (CSIRO). Chinese Spring ditelosomic lines $1A_L$, $1B_S$, $1B_L$ and $1D_L$ were obtained from P. Anderson (CSIRO).
Seedlings were germinated in moist vermiculite in a 25°c controlled growth room and harvested after 6-7 days.

(b) Genomic cloning: A partial Sau3A1 digest of Millewa DNA was sedimented through a 10-40% glycerol gradient (in 10mM Tris pH7.5, 0.5mM EDTA pH8.5) and the 12-20kb fraction isolated and ligated into the BamH1 site of Charon 35 (Wilhelmine et al. 1983). The library was screened by plaque hybridization modified from Benton and Davis (1977) using a mixed probe of the inserts of the maize *Adh1* (pZML793) and maize *Adh2* (pZML841) cDNAs (Dennis et al. 1984,1985). Four positive clones λTA 1-4 were isolated. Two fragments of λTA 1 were subcloned into pUC13: pTA1, a 5.7kbp Xba1/Hpa1 fragment containing the entire wheat *Adh* gene and p3NTR a 0.4kb Spe1/Hpa1 fragment from the 3' untranslated region (Fig 2.1).

(c) Southern analysis: Genomic DNA was prepared by the method of Dellaporta et al.(Dellaporta, S.L., Wood, J., Hicic, J.D. Maize Genetics Co-operation Newsletter (1983) 57: 26-29). 10ug of genomic DNA was digested with 25 units of restriction endonuclease and electrophoresed on a 0.8% agarose gel. The gel was acid washed (0.1M HCl 5 min), denatured (0.5M NaOH, 1.5M NaCl, 30 min.), neutralised (1M NH_4CH_3COO pH 8.3, 30 min.) and transferred to nitrocellulose (Southern 1975). The filter was baked, prehybridized and hybridized using the method of Dennis et al.(1984) with the exception that $3x10^7$ cpm of an oligolabelled probe prepared using an oligolabelling kit (BRESA) to a specific activity of 10^8 cpm/ug was used.

(d) Northern analysis: Total RNA was isolated from both aerobically and anaerobically grown seedlings by the method of Sachs et al. (1980). Anaerobic induction of *Adh* mRNA was achieved by immersion of seedlings for 6 hrs in $5x10^{-4}$M $ZnSO_4$ at 25°C. RNAs were analyzed by electrophoresis in 1% agarose gels containing 2.2M formaldehyde and transferred directly to nitrocellulose

(Palukaitis et al.1983). Specific *Adh* mRNAs were detected by hybridization to either primer extended or oligolabelled p3NTR.

(e) DNA sequencing. pTA1 was sequenced using the dideoxy chain termination method of Sanger et al.(1980). A series of overlapping subclones in M13 phage vectors mp18 and mp19 (Norrander et al. 1983) was generated for the first 2.5 kb of pTA1 using the method of Dale et al.(1986). The remainder of pTA1 was sequenced using overlapping restriction fragments subcloned into the above phage vectors.

Restriction enzyme mapping, subcloning and other standard recombinant DNA techniques were essentially those compiled by Maniatis et al.(1982).

2.3 Results

Isolation of *T. aestivum* cv. Millewa *Adh* genomic clones.

Approximately $1x10^6$ recombinant plaques from a Charon 35 library of Sau3A1 partially digested Millewa DNA were screened using a mixed probe of the maize *Adh1* and maize *Adh2* cDNAs. Four positive clones were identified, λTA 1-4.

DNA was isolated from each of the four positive clones which were then further screened by Southern hybridization to determine if they contained an entire *Adh* gene. Two subclones of the maize *Adh1* gene were used as probes: DS2A, a 743 bp Pst1/HindIII fragment from the maize *Adh1*-Fm335 allele extending from -148 to +595 which contains the first exon of maize *Adh1* (Sutton et al. 1984) and 1F1D , a 950bp HindIII/Pst1 fragment from the maize *Adh1*-1F allele containing exons 9 and 10 of the maize *Adh1* gene (Sachs et al.1986). λTA 2, 3 and 4 hybridized only to the 1F1D probe and not to the DS2A probe and therefore do not contain an entire *Adh* gene. λTA1 however, hybridized to both the 1F1D and

DS2A probes and was presumed to contain an intact *Adh* gene. The restriction map and hybridization data for λTA1 is shown in Fig 2.1a.

To verify that the restriction map of pTA 1 represented the sequence organization found in the genomic DNA we used a subclone from the 3' untranslated region of the wheat *Adh* gene (p3NTR) as a probe in genomic blots of Millewa DNA. It was expected that this probe would be specific for the cloned *Adh* gene based on observations by Dennis et al.(1985) that even though the coding regions of the maize *Adh1* and *Adh2* genes are highly conserved the 3' untranslated regions are completely divergent. In all digests a restriction fragment of the size expected from the restriction map of pTA1 was found to hybridize to p3NTR (Fig 2.2). However, two additional bands were also found. Subsequently we have shown that these two bands correspond to other copies of the cloned *Adh* gene on homeologous chromosomes. The p3NTR probe is therefore not specific to a particular *Adh* gene but hybridizes to a series of homologous *Adh* genes. A number of restriction digests showed only one additional band (Fig 2.2,lanes 3,4). This is presumably due to the third fragment comigrating with the observed bands or alternatively, the fragment may be either too small or too large to be efficiently transferred to the nitrocellulose membrane.

Chromosomal location of the *Adh* genes.

The known wheat *Adh* genes have been mapped by the zymogram technique to chromosomes 4, 5 and 6 (see introduction). It was expected therefore that λTA1 would map to one of these chromosomes. DNA was prepared from *T.aestivum* Chinese Spring nullisomic-tetrasomic stocks of chromosomes 4, 5 and 6. 10 ug of each DNA sample was restricted with HpaI/BglII and probed with

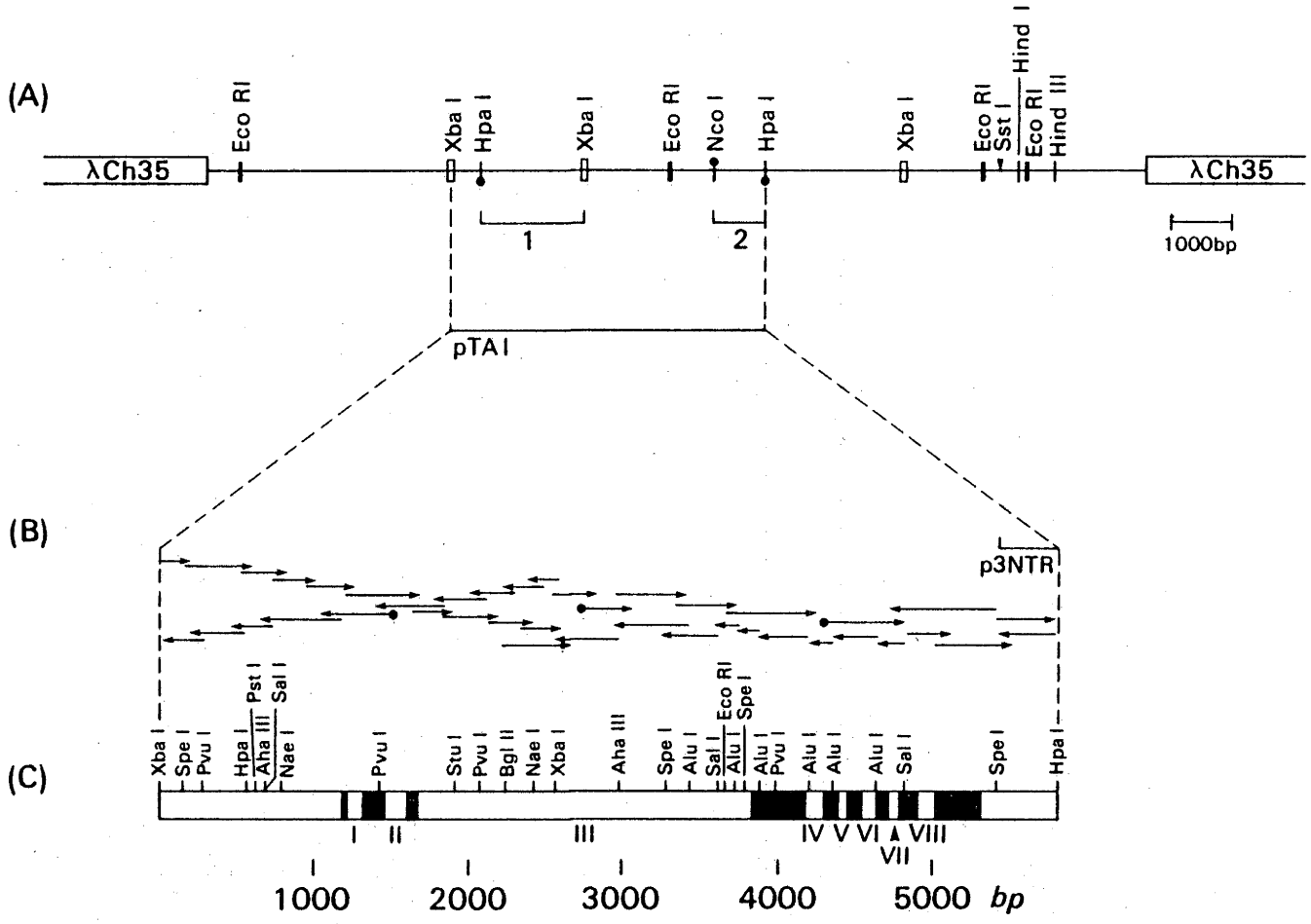


Figure 2.1
(a) Restriction map of λTA 1. Bracketed regions 1 and 2 represent restriction fragments which hybridized to DS2A and 1F1D probes respectively. Subclones pTA1 and p3NTR are represented as solid lines
(b) DNA sequencing strategy. Arrows represent direction and length of DNA sequence obtained for each fragment. Arrows having a dot indicate priming using a synthetic oligonucleotide.
(c) Detailed restriction map of pTA1 showing gene structure with exons denoted by solid boxes and introns numbered.

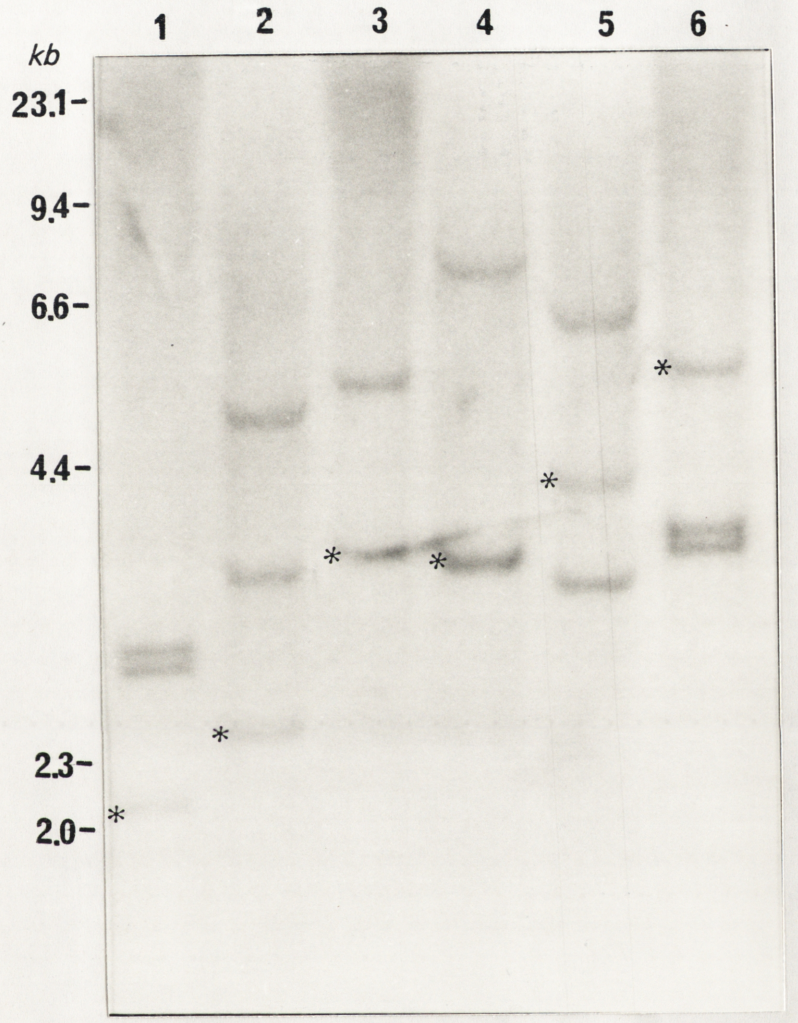


Figure 2.2
Southern of *T.aestivum* cv.Millewa DNA probed with oligolabelled p3NTR.Asterisk denotes expected restriction fragment sizes as derived from the sequence of pTA1. (1) Hpa1/Pvu1, (2) Hpa1/EcoR1, (3) Hpa1/Nae1, (4) Hpa1/Xba1,(5) Hpa1/Stu1, (6) Hpa1/Pst.

oligolabelled p3NTR. All the digests showed three hybridizing bands exactly as found in the Millewa control indicating that the TA1 *Adh* gene was not from any of the previously identified *Adh* loci (data not shown).

Further Southern blots using p3NTR as probe on Chinese Spring nulli-tetra stocks of other chromosome groups indicated that the cloned wheat *Adh* gene mapped to chromosome 1A (Fig 2.3). As expected for hexaploid wheat there were three copies of the gene resulting in the three bands seen in genomic Southerns. The additional copies of the gene were mapped to chromosomes 1B and 1D. Using Chinese Spring group 1 ditelosomic lines we have shown that all three copies of the *Adh* gene are located on the long arm of their respective chromosomes (Fig 2.3).

The nucleotide sequence of the wheat *Adh* gene.

The nucleotide sequence of a 5.7kb fragment (Fig 2.1b) which spans the region hybridized by the DS2A and 1F1D probes was determined (Fig 2.4). Sequence homology with the maize *Adh* cDNA clone pZML793 (Dennis et al.1984) has shown that this region, as expected from the hybridization data, contains an entire *Adh* gene. Dot matrix analysis of the wheat *Adh* genomic clone and pZML793 allowed identification of the wheat *Adh1* coding region. The predicted coding region has an open reading frame of 1137 nucleotides which is exactly the same length as the maize *Adh1* and *Adh2* coding regions. Within the coding region there is 82% homology with the maize *Adh1* gene and 86% homology with the maize *Adh2* gene at the nucleotide level. At the amino acid level there is 86% and 90% homology respectively.

Unlike the maize *Adh1* and *Adh2* genes which have nine introns the wheat *Adh* coding region is interrupted by only eight introns.

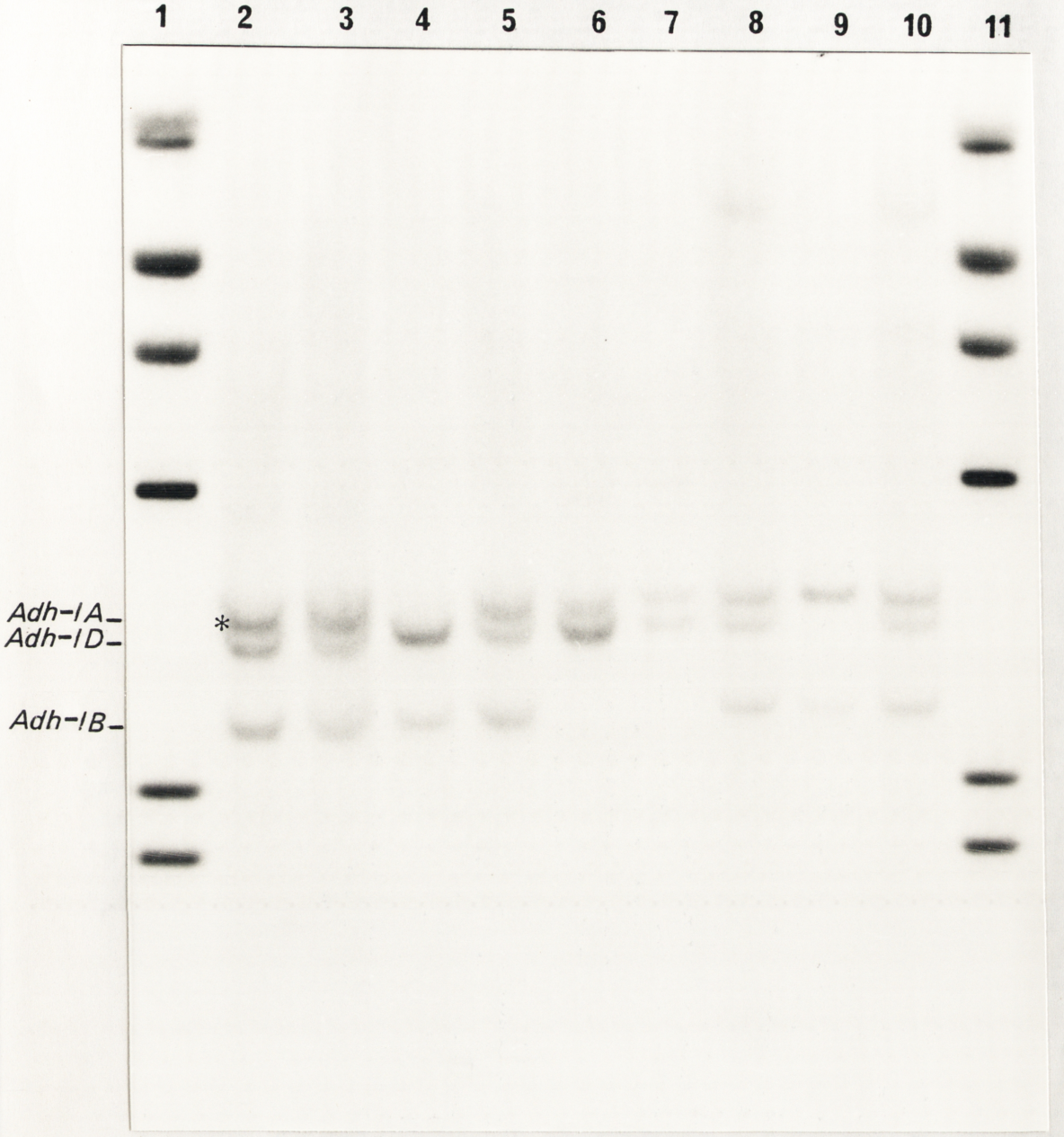


Figure 2.3
Genomic Southern showing mapping of the cloned wheat *Adh* gene to the long arm of chromosome 1A with additional copies on the long arms of chromosomes 1B and 1D. All genomic DNA samples were digested with Hpa1/BglII. Restriction fragments correspond to the Adh-1A, Adh-1B and Adh-1D loci are labelled. Asterisk denotes restriction fragment corresponding in size to the cloned wheat *Adh* gene. Lanes (1) and (11) are end labelled MW markers 23.1kb, 9.4kb, 6.6kb, 4.4kb, 2.3kb, 2.0kb; (2) *T.aestivum* cv. Millewa; (3) *T.aestivum* cv. Chinese Spring; (4) Chinese Spring NT 1A-1D; (5) Chinese Spring ditelo.1Al; (6) Chinese Spring NT 1B-1D; (7) Chinese Spring ditelo 1Bs; (8) Chinese Spring ditelo.1Bl; (9) Chinese Spring NT 1D-1A; (10) Chinese Spring

Figure 2.4

1
TCTAGACATGAACGGTTGCAGTTTGCCGAAAAAGAGGACTTGTGTAATGAAAATACGATTGTTGTTCTTATTGGGCCGCAGCCGATGTCGGGCCGCCTGT 100

AGGCGAGCGGAAGGCGATACATACTACTCCATTTCGATTCCCTAAAAAAAACTAGTCCATTTCGCTGAATTTTGTACTTGTGCCGCTGGGCTGAGACAGT 200

CAACTTTATGTGCTGAGGTCAAATAAATTTTTTATTACTTCCTACCTAACAGTGTTTGCTGAAGTTACGAATTACGATCGGCCTTTTCAACTGAGCTTCG 300

TCCGGCAGTGCCTAATCGGCGCTTCAGGCGCCATTACTAAACCTGCCGCTTGCGTGGACTGGCCAAACAAGCGCACGGATGACTCGCTCACGCATGCAAC 400

CGATTTCGTTCATTCGCCGTTTTTTTTCCGCTCCATTAACTGTTTGTCGATCAACTTTGACCAATTTAACCAGTTTGACCATTGACTTAAAACAACATGG 500

ATCTAAAAAAGTTGATAAATTTGAAAATAGTTTATGTAATTCAAAATTCCGTAATTTGAGAAAAGTTAACGAGTACGAAGAATGTCTGCAGATTTTAAAA 600

CGTTTCAGTAAAACGAGGGATTGGCGCCTGGTCCCGTCGACGACGATGCGCCAGGTCGTGGGCTCGTGGACCGTTGCAAAACAGGGTACCGAGCTGTTGT 700

TTTCGCAGTTGGTTTGCT**CGGTCGCCCCGGTTTC**GGCCACCAAATGCAAGTCAGCCCTCTCCGGCCGGCCTCAGCGATGGTTTTTTACTCTTTCCCAGCT 800

ACCGGTTGACCTGTAGGCTGTAGCCCCCTGATCACGTCACAACGTTTTCCAGATACCACACGGGCCCAATTTTGAAACCAGCGGACTCGCGCGCAGAGCC 900

GCAGGGAGACACCCCTGTTTCAGCCGCCTTCTTCTTTCCCCCCCTTGCACTCCAAGCCACCCTTAATTACTCGCCGATGACTCCAGCGCCCGATAATGCT 1000

CTATCTACGAAAAGGTCTACGTAAGGTAACGTAACCTTCCTAA**CAAT**AATTCTCTTCCCCCCCCTGATTTTCAGTGGGGGTGGGCCCC**CAAT**CCCCTCTA 1100

ATC**CAAT**CCTGATTGTCCATGTTGC**ATATTA**CGTAAAACACGTACGTTAGTGTACGTAGGTGTAGCAACGAACTGCAAAAGCAGAACCAGCGACAGAGAT 1200

1
CCATCCTTCTCTTCTTCTAGCGACAAGG**ATGGCGACCGCCGGGAAGGTGATCGAGTGCAAAG**GTGTGTTCCTCAGATGGTAGTCAGTTTCATCTGATTTT 1300

2
GGTTGATGTTTTGAGGACGGGTGGAGATCTGAATTGTGCGGTTCGTTGGTTGCAG**CGGCGGTGGCATGGGAGGCCGGGAAGCCGCTGTCGATCGAGGAGG** 1400

TGGAGGTGGCGCCGCCGCACGCCATGGAGGTGCGCGTCAAGATCCTCTACACCGCCCTCTGCCACACCGACGTCTACTTCTGGGAAGCCAAGGTATCCAC 1500

TCAAATAATTGTATATGTATCTCAAAAGCGTTTGTGTTTTCCTATTGGGTAAATTGCAGTGATGCATCAGTGGCTAATGAAATTTGTAATGATGTTCTGC 1600

3
TTGGATTTGTGCAG**GGTCAAACTCCTGTTTTCCCTAGGATCTTAGGCCATGAAGCTGGAGG**GTAAATATGCATCCTCTCTGTCTCTGTCTCTGTCTTATT 1700

Figure 2.4 (Ctd)

1800

CTGTTTCAGCTGTTGCTTGGAGGAGGAATCTTGAATCCTAAATCATAGTTTATTCTCTCTGGCTTCTAATTCTAGTCATCCTGAGTACTGATCTCTGTCA

1900

ATGGTATTAGATTCGTATACCCTACATAAAGCAAGTGTCCATATTTTATTGCCAATGGGATAAAACATACTATGAAGCACTGGCGGAGGCCTCCAGTGGG

2000

CAGGGGCAGGGTATGGCAGCCGCCATACCTTGACTTGGAGCAAAAAAATTGTATACGTATGTATGTTACGGCTGGCTAGTTTTGATCGCTGGCAGTTTTA

2100

CGCTAAAAGGATAGATGGGCCGACGCATCCAGCTAGTTGGGCTAGAATCGATCGAGCGCCCATGCACGAAGCAGATGAACCGGTCGAGGAGAGCAGAAAT

2200

TTGGAGCCGGAATTTTTTGCACGGACGCTGGTGATTATGTGGGCAGTTTGGTGGGCAAGGAGAAGGGCGATCCATGAGGAGGAATTTCAAGCTCCACTGT

2300

CCACGCATGCATTTGTCACAAGGTTTATGGCTGACGTAGTAGATGCACAAATAAGAAACCGTCAGTTGAAAAAGGCCGGAGGAAGAACCCCTGTTCCATC

2400

AGTTTGTTGGACACCTCCTGCGACGGGCATAGCCAAGATACATGCCGATGGTGGATTATCTAAAGATCTGCAAGTGGGAGCAGCAGCAGCTGTATGCCGG

2500

CAACGACCAAGGCATTTTTCAAGGTGCTTCGGCAGCTGTGCAAGCTGGAATGACTGATCCTCAAACCCTCGAAGCACAAGCGATATGTGAAGCTATGGCT

2600

TTGGCTGCTGACCTACAGGTTTCTAGAGTGCTTATAATATCAGACTGCTTAAGTGTCATCAAGAACATTAACTCAGGCAACAGAATGACAGCTTATGGTG

2700

CTATCCTGAAGGAAGTAGATAGGTGGAAAGATAGCTTTCAGATGGTCATCTTCGACCACGAACGGTGAGAGTCAAATAGGGACGCTCACGCTCTAGCTAA

2800

ATATATGCAACCTCCCTGCGTGGGGGTAGGCATGTCTGGTTCACAGGGCCTCCAGACCCGTTTTGTATCCCTATAAACATTAATATGCAATAAAGCTCTC

2900

CCGTTCCCTCAAAAAAATATATAACCAGGTCGAGGTGTCGAGCCAGGTCGAGAACACATGACCCAATCGAGACACACACACACGGGAAAATACCAATTGT

3000

TCCTTATTCCTTGGGGTGCTCCCCATGCTCCAATTTGAAAACATCAATTTTAAATGTCTCAAAATTTTCAGAAAAAAATGTGAATGCTCACAAGGAATAT

3100

GAATACAACCCCTAAAAATTTCAGATCCAAACTCGAAATTTACTTTGAGAAACAAAAAAGACAAATTCAGATATGGATAGTGTCTAATGTTGCTTTTGTC

3200

TTTTATGACACTATTTATGCTAGATTTCACATTTTCGATTCTCAATTTTCATTTAGAGCACCCCAAGGATGGGAGCACCTGATATTTCCCCCCACACGCA

3300

CGTCTCCACGACTCCACCACTTCTTACTCCACCCGGCGGCCGACGCAGCGAACCCTAACTAGTGCGGCGGCGCCGCTAGCCAGGTCGCCAGGACAAGGCG

3400

GGTGACCGCACCGGGGCAGCCAACGGCAGTTGGCCAGGTCGCTAGGACGAGGCAGGCGACCGCGACGGGGCAGCCAACGACGGCAGGCTGGGGGCCTAGA

Figure 2.4 (Ctd)

3500
GGGCAGCGACAACAACGCCATTGGCAGCTCCAGCCAGGACCTAGGCGCCTAGCGGGTAGGGCACTGCCATTGCCTGGGTCTTCTAGCCATTGGAAGATTC
3600
ATTATCAATATGTTTTAGGTTGTTTTCCTATTGAAACATGCCTACATTTTTATTGTATTTGATTTTTTTGCTGTAATGTGTGTCGACTCTTTCGATTTGC
3700
ATCAAAAGAATTCTTTCCTTAAGACTTTGCCAGACCTTAATTTCTTTCCGTCCTCCGCCTCTGCTATGAAGAATCATGGGAAAACGTCATCCATGACACA
3800
AGCTCACATCACAAACTGCTCTGTACAATCATACAGTTTTGGACTAGTTAATCTCCTCAGCATGTTTGCTCTTCATAGTTTATCTTTTCTGGCATCTCTT
4 3900
TACACCAAATAGCTTCTTTGTCGCAG**CATCGTCGAGAGCGTCGGAGAAGGTGTGACTGAGCTTGTGCCGGGTGACCATGTCCTCCCGGTGTTCACCGGCG**
4000
AGTGCAAGGACTGTGCCCACTGCAAGTCAGAGGAGAGCAACCTTTGTGATCTCCTCAGGATCAACGTCGATCGTGGCGTGATGATCGGCGATGGGCAGTC
4100
TCGCTTCACCATCAATGGGAAGCCGATTTTCCACTTCGTCGGGACCTCCACCTTCAGCGAGTACACCGTCATCCATGTCGGTTGCCTCGCAAAGATCAAC
4200
CCCGAGGCGCCCCTTGACAAAGTTTGTGTTCTCAGCTGTGGCATCTCAACTGGTAAGACTGATATAATAGCTCGTCACATGCCATTGTTTCCAAATGAAG
5 4300
CACAATGGTTTAGGATGCCTTGATTAGTTGCTGAGTTGTCATTTTTCTTCAG**GACTTGGTGCTACGCTCAATGTCGCAAAACCAAAAAAGGGTTCCACGG**
4400
TGGCAATTTTCGGTCTTGGAGCTGTAGGACTTGCTGTGAGTAACATTCCAAGATGTTAATCTTCCACGACATATTTTTTACAGGCTGAAGAGAATTATGC
6 4500
GCTATAGCCATGAATAATGACTGTATGCTTTTCACATGAGCAG**GCCATGGAAGGGGCCAGGATGGCTGGGGCATCAAGGATCATTGGTGTGGATTTGAAC**
4600
CCTGCAAAATACGAACAAGGTACAATAATCTTAAAACAAGGGTGTCTTGGCAAATGTTTTCGGATTGATAGAACACTGAACCTAACAACATGTACTACAT
7 4700
TATTTTTCAG**CTAAGAAATTTGGATGCACAGACTTTGTGAACCCAAAGGACCACACTAAGCCCGTGCAAGAG**GTCTGTTGTCCATGTCCCTCAGTAAACA
8 4800
TTTTCCCCTTTATTCAATCTTCATGACATGAACAATGATGCTCTCCAG**GTGCTTGTCGAGATGACCAATGGCGGAGTCGACCGTGCGGTTGAGTGCACTG**
4900
GCCACATCGATGCAATGATCGCCGCCTTCGAATGTGTCCACGATGTACGTAGCTCGCACTATACCTACTATAAATAACATATCTGTACTCCAACGATAAT
9 5000
TAATCACACCCTCTGCTTTAGAAGAGTTTAGACTGAATGTTACAACCTGGTGGATGCAG**GGGTGGGGCGTGGCCGTGCTGGTTGGCGTGCCGCACAAGGA**
5100
GGCGGTGTTCAAGACCTACCCAATGAACTTCCTCAACGAGAGGACCCTGAAAGGCACCTTCTTCGGCAACTACAAGCCACGCACCGACCTCCCGGAAGTC

Figure 2.4 (Ctd)

5200

GTCGAGATGTACATGAGGAAGGAGCTGGAGCTGGAGAAGTTCATCACGCATAGCGTGCCCTTCTCGCAGATCAACACGGCGTTCGACCTCATGCTCAAGG

5300

GGGAGGGCCTGCGCTGCATCATGAGGATGGACCAGTAGATCACCGCCTCCCATCATTCATCAATGTGTCTCGCTGTTTTCAGATTTGGCGCTTGAGAAGC

5400

AAAAGGCAAAAAAACAAAAAGTTAGAAATCTCCCCTTGGATAATAATGTACTAGTCGCGAGATGTGTTCATCTCTATGGTCTCAGATAGAACGAATGGGT

5500

CTTTGTTTTGAGTGTGTCGTCGGCAGCTTTCAGTTAGCATGTAAATTACTATCATTGTGCACAATGGTAATTAATATCTTGCAGCATCTGTCTTAATGTT

5600

GGGAGCGTCTATGACAGGTGCCTGGACTTGTTGTTCCAGTCACCCTTTTTTTGTGCAGAACGATTTGCTTTCGACGGCAACACCTAGATCAAGTAGAGAA

5700

AACGTGACTGGTTCTTTTTTGCTGTTTAGACCGATTTTCACTTACCTGGCAAAGCAACAGCCCAAACGGAGAGAAGGCCAGCCAGGCTCACCTATCTATC

AGAAACTGTCCACATAGAGGATGGA**AATAAG**CTATTGGTAAACGGTTAACAGTTCTTGCTTGTCCTGTCTGCGTAGAGA

Figure 2.4
The nucleotide sequence of pTA1. Exons are numbered, printed in bold lettering and underlined. Presumptive polyadenylation signal, TATA, CAAT boxes and anaerobic regulatory element (see text) are also in bold lettering and are underlined.

The eight introns are in the same positions as the corresponding maize *Adh* introns although their length and sequence differ. The consensus sequence for the 5' and 3' splice sites are as expected for eucaryotic genes (Fig 2.5).

Transcription & expression of the *Adh* genes on chromosome 1.

Alcohol dehydrogenase has been shown in many plants to be induced under anaerobic conditions. In maize this control of ADH expression is due to transcriptional regulation of the *Adh* genes (Rowland and Strommer 1986). We have isolated total RNA from wheat seedlings grown aerobically and after six hours following transfer to anaerobic conditions. Hybridization of the probe (p3NTR) to an anaerobically induced message was found (Fig 2.6). Northern analysis indicates that the length of the induced transcript is approximately the same size as the maize *Adh* mRNAs (Fig 2.6).

Hart (personal communication) has identified by IEF an isozyme which stains for ADH activity and utilizes ethanol as a substrate. He has correlated the disappearance of this isozyme band with Chinese Spring lines which are nullisomic for chromosome 1D. The absence of ADH isozyme bands corresponding to the expression of *Adh* genes on chromosomes 1A and 1B may be due either to their comigration with other ADH isozymes or alternatively, the 1A and 1B *Adh* genes may not be expressed or have different kinetics.

2.4 Discussion

Genetic and isozyme studies have shown that there are a number of *Adh* loci in wheat. The different types of alcohol dehydrogenases can be classified by their different substrate specificities and co-factor requirements. Recent experiments indicate that these differences may be due to relatively minor

			INTRON			
		5'			3,	
1.	GAGTGCAAAG	GT	GTGTTCCT	TTGGTTGC	AG	CGGCGGTGGC
2.	GGAAGCCAAG	GT	ATCCACTC	ATTTGTGC	AG	GGTCAAACTC
3.	AAGCTGGAGG	GT	AAATATGC	TTTGTCGC	AG	CATCGTCGAG
4.	ATCTCAACTG	GT	AAGACTGA	TTTTCTTC	AG	GACTTGGTGC
5.	AGGACTTGCT	GT	GAGTAACA	ACATGAGC	AG	GCCATGGAAG
6.	TACGAACAAG	GT	ACAATAAT	TATTTTTC	AG	CTAAGAAATT
7.	CGTGCAAGAG	GT	CTGTTGTC	TGCTCTCC	AG	GTGCTTGTCG
8.	TGTCCACGAT	GT	ACGTAGCT	GTGGATGC	AG	GGGTGGGGCG
Wheat Adh-1A	AAG	GT	ATGT	T_TT_TGC	AG	G__CGGG__G
Consensus	G		GAAA	A GG T		C TTAA C
Animal	AAG	GT	AAGT	T C	AG	GG
Consensus	C		G	C T		T

Figure 2.5
Nucleotide sequence of intron/exon splice junctions for the cloned wheat *Adh*-1A gene. The derived consensus sequence forthe wheat *Adh*-1A gene and the general consensus sequence for animal genes are given.

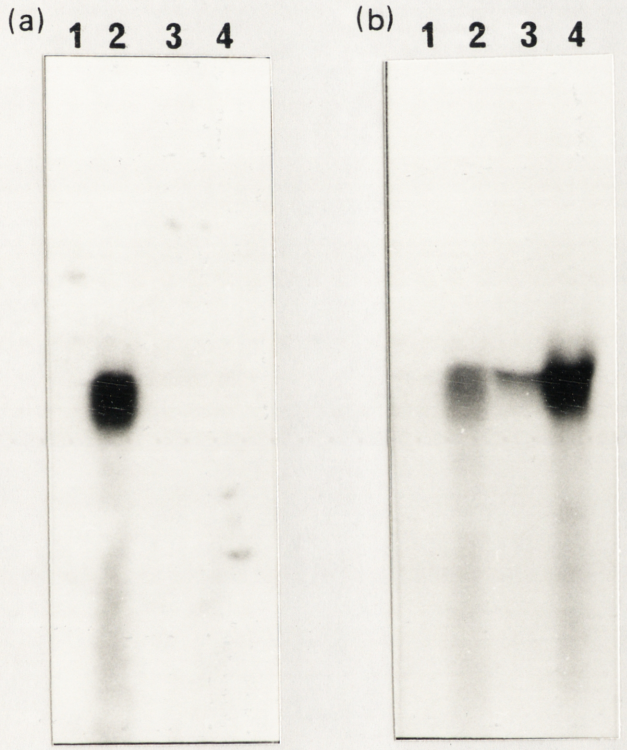


Figure 2.6
(a) Northern of uninduced and ANAEROBICALLY induced total RNA (50ug) from *T.aestivum* cv.Millewa (Lanes 1,2) and maize (Lanes 3,4) probed with p3NTR.
(b) Rehybridization of the Northern in (a) probed with the maize *Adh1* cDNA pZML793.

structural changes which can cause significant alterations in the function of an enzyme (Murali and Creaser 1986).

Sequence analysis indicates that we have cloned a wheat *Adh* gene. The gene contains sequences characteristic of eucaryotic genes including transcription and translation start signals and intron/exon junctions similar to those found in most animal genes (Fig 2.4).

In vivo functional analysis of the maize *Adh1* promoter by Walker et al. (PNAS in press) has identified two regions of approximately 15bp in length which are required for anaerobic regulation of the maize *Adh1* gene. These two regions are immediately 5` of sequences which become hypersensitive to DNAse 1 digestion under anaerobic conditions (Paul et al.1986), and may represent binding sites for transacting factors required for anaerobic expression of the maize *Adh1* gene. Region 1 of the maize *Adh1* gene which extends downstream from position -140 (CTGCAGCCCCGGTTTC) shares 81% homology with an 16bp sequence (CGGTCGCCCCGGTTTC) which is located approximately 500bp 5` of the start of translation of the wheat *Adh*-1A gene (Fig 2.4). Region 2 of the maize *Adh1* gene (CCGTGGTTTGCTGCC) is located 10bp downstream of region 1. The central 9bp of this element are identical to sequences immediately 5' of the putative region 1 of the wheat *Adh*-1A gene (TGGTTTGCT). Either both or one of these regions may be involved in the anaerobic response of this gene although functional studies would be necessary to confirm this.

Chromosome mapping studies indicate that the cloned *Adh* gene is from the long arm of chromosome 1A. Due to the polyploidy of *Triticum aestivum* additional copies of the *Adh* gene are located on the long arms of chromosomes 1B and 1D. Dennis et al.,(1985) have suggested that maize *Adh1* and *Adh2*, because of their extensive

sequence homology and because they have the same position and number of introns, arose by duplication of an ancestral gene. The time of the duplication is such that the 5' and 3' flanking regions of these two genes, which must once have been identical, are now extensively diverged. In wheat however, Southern analysis under stringent hybridization conditions (50% formamide, 45^{c}) indicates that the 3' non-coding regions of the *Adh* 1A, 1B and 1D genes are at ~~least 90%~~ HIGHLY homologous. This suggests that the lineage of the ancestral wheat *Adh* gene predates the diploid and polyploid wheats and that speciation of the diploid wheats occurred relatively recently.

Llewellyn et al.(1987) have shown that the pea *Adh1* gene has the same number and position of introns as the maize *Adh1* and *Adh2* genes. This conservation of intron position and number between *Adh* genes of monocot and dicot species suggest that both the pea and maize *Adh* genes arose from the same ancestral gene. Chang and Meyerowitz (1986) have shown that the *Arabidopsis Adh* gene has only six introns whose positions correspond to that of six of the nine introns in maize and pea. Similarly the wheat *Adh* gene has eight introns whose positions are the same as eight of the nine introns in maize and pea *Adh*. This suggests that during evolution there are processes resulting in the loss of introns. The alternative would require the ancestral gene to have six introns ,similar to *Arabidopsis*, and for both pea and maize *Adh* to have acquired three introns in identical locations and by independent events.

Nucleotide sequence comparison of the cloned wheat *Adh* gene with *Adh* genes from other species including maize *Adh1* and *Adh2*, pea, *Arabidopsis* and barley *Adh1* and *Adh2* (M. Trick Personal communication) indicate extensive homology in the coding region

but complete sequence divergence in the introns and 5` and 3` non coding regions. Barley *Adh3* however, was 95.6% homologous to the wheat *Adh* 1A gene in the coding region and greater than 80% homologous in the non coding regions. Dot matrix analysis of these two genes indicates the presence of an insertion of 1794bp into intron 3 of the wheat *Adh* gene. This insertion is flanked by eight base pair direct repeats and has inverted terminal repeats, structures normally associated with transposable elements. The insertion accounts for the unusually large size of intron 3 compared to other *Adh* genes. Detailed analysis of the insertion and its evolution will be discussed elsewhere.

The inferred amino acid sequence of the wheat ADH polypeptide is 50% conserved relative to horse liver ADH. Eklund et al.(1976) have determined the three dimensional crystal structure for the horse liver enzyme and have identified structurally important amino acid residues including the seven residues that provide ligands for the catalytic and structural zinc atoms. These seven residues as well as many of the residues implicated in coenzyme binding are conserved in the wheat *Adh* gene. Regions of high variability in the sequence tend to be associated with residues that form no obvious secondary structure but are involved in the joining of secondary and super secondary structures.

Acknowledgements

We would like to thank Gary Hart, Danny Llewellyn and Martin Trick for sharing data prior to publication. This work was supported by a grant from the Rockefeller Foundation.

2.5 References

Benton, W.D., Davis, R.W. (1977). Screening λgt Recombinant clones by hybridization to single plaques insitu. Science 196:180-182

Chang, C. and Meyerowitz, E.M. (1986). Molecular Cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. Proc. Natl. Acad. Sci. USA 83:1408-1412

Dale, R.M.K., McClure, B.A., Houchins, J.P. (1986). A rapid single stranded colony strategy for producing a sequential series of overlapping clones for use in DNA sequencing. Plasmid 13:31-41

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). Molecular analysis of the alcohol dehydrogenase (*Adh1*) gene of maize. Nucl. Acids Res. 12:3983-3999

Dennis, E.S., Sachs, M.M., Gerlach, W.L., Finnegan, E.J., Peacock, W.J. (1985). Molecular analysis of the alcohol dehydrogenase 2 (*Adh2*) gene of maize. Nucl. Acids Res. 13:727-743

Eklund, H., Nordstrom, B., Zeppezaur, E., Soderlund, G., Ohlsson, I., Boiwe, T., Soderberg, B., Tapia, O., Branden, C,, Akeson. A. (1976). Three Dimensional structure of Horse liver alcohol dehydrogenase at 2.4A resolution. J. Mol. Biol. 102:27-59

Freeling, M. and Birchler, J.A. (1982). In: Setlow JK, Hollaender A (eds) Genetic Engineering Principles and Methods III. Plenum Press, pp 223-264

Gottlieb, L.D. (1982). Conservation and duplication of Isozymes in Plants. Science 216:373-380

Hanson, A.D., Brown, A.H.D. (1984). Three alcohol dehydrogenase genes in wild and cultivated barley: Characterization of the product of variant alleles. J Biochem. Genet. 22:495-515

Hart, G.E. (1969). Genetic control of alcohol dehydrogenase isozymes in *Triticum dicoccum*. Biochem. Genet. 3: 617-625

Hart, G.E. (1970). Evidence for triplicate genes for alcohol dehydrogenase in hexaploid wheat. Proc. Natl. Acad. Sci. USA 66:1136-1141

Hart, G.E. (1980). Evidence for a triplicate set of alcohol dehydrogenase structural genes in hexaploid wheat. Genetics 94: 108-113.

Hart, G.E. (1983). Hexaploid wheat. In: Tanksley SD, Orton TJ (eds) Isozymes in plant genetics and breeding, Part B. Elsevier Science Publishers B.V., Amsterdam

Jaaska, V. (1978). NADP dependant aromatic alcohol dehydrogenase in polyploid wheats and their diploid relatives. On the origin and phylogeny of polyploid wheats. Theor. Appl. Genet. 53:209-217

Langston, P.J., Hart, G.E., Pace, C.N. (1979). Purification and partial characterization of alcohol dehydrogenase from wheat. Arch. Biochem. Biophys. 196:611-618

Langston, P.J., Pace, C.N., Hart, G.E. (1980). Wheat alcohol dehydrogenase isozymes: Purification, characterization and gene expression. Plant Phys 65:518-522

Llewellyn, D.J., Finnegan, E.J., Ellis, J.G., Dennis, E.S., Peacock, W.J. (1987). Structure and Expression of an Alcohol Dehydrogenase 1 gene from *Pisum sativum* (cv.Greenfeast). J. Mol. Biol. 195:115-123.

Maniatis, T., Fritch, E.F., Sambrook. J. (1982). Molecular Cloning. Cold Spring Harbour Press, Cold Spring Harbour

McMannon. M., Crawford, R.M.M. (1971). Metabolic Theory of Flooding Tolerance:The significance of enzyme distribution and behaviour. New Phytol. 70:299-306

Murali, C., Creaser, E.H. (1986). Protein engineering of alcohol dehydrogenase 1. Effects of two amino acid changes in the active site of yeast ADH1. Prot. Eng. 1:55-57

Norrander, J., Kempe, T., Messing, J. (1983). Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26:101-106

Palukaitis, P., Garcia-Arenal, F., Sulzinski, M.A., Zaitlin, M. (1983). Replication of Tobacco Mosaic Virus. Virol. 131:533-545

Paul, A.L., Ashraf, M., Bollinger, S., Ferl, R.J. (1986). Architecture of the 5' Flanking Regions of the Maize *Adh* genes. In: First International Symposium on chromatin.Structure of plant genes. Frankfur FRG

Rowland, L.J., Strommer, J.N. (1986). Anaerobic Treatment of Maize Roots Affects Transcription of *Adh1* and transcript stability. Mol. and Cell Biol. 6:3368-3372

Sachs, M.M., Freeling, M., Okimoto, R. (1980). The Anaerobic Proteins in Maize. Cell 20:761-766

Sachs, M.M., Dennis, E.S., Gerlach, W.L., Peacock, W.J. (1986). Two alleles of maize alcohol dehydrogenase 1 have 3` structural and poly(A) addition polymorphisms. Genetics 113: 449-467

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H., Roe, B.A. (1980). Cloning in single stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. 143:161-178

Schwartz, D. (1971). Genetic Control of Alcohol Dehydrogenase-A competition model for regulation of gene action. Genetics 67:411-425

Sears, E.R. (1966). Nullisomic-tetrasomic combinations in hexaploid wheat. In: Riley R, Lewis K.R (eds) Chromosome Manipulations and Plant Genetics. Oliver and Boyd, London, pp29-45

Southern, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-505

Sutton, W.D., Gerlach, W.L., Schwartz, D., Peacock, W.J. (1984). Molecular nature of *Ds* controlling element mutations at the *Adh1* locus of maize. Science 223:1265-1268

Walker, J.C., Howard, E.A., Dennis, E.S., Peacock, W.J. (1987). DNA sequences required for anaerobic expression of the maize alcohol dehydrogenase-1 gene. Proc. Natl. Acad. Sci. USA (in press).

Wilhelmine, A.M., Blattner, F.R. and L. (1983) Lambda Charon vectors (Ch 32, 33, 34, 35) adapted for DNA cloning in recombination-deficient hosts. Gene 26:171-179

CHAPTER 3

Evidence for transposable elements in wheat.

L. E. Mitchell*, E. S. Dennis, W. J.Peacock

Division of Plant Industry
CSIRO
G.P.O. Box 1600
Canberra, A.C.T 2601, Australia

* Also from Research School of Biological Sciences
Australian National University
Canberra

Abbreviations: *Adh*; alcohol dehydrogenase, CS; *T. aestivum* cv.Chinese Spring, Chris; *T.aestivum* cv.Chris, Selkirk; *T.aestivum* cv.Selkirk
MW; Molecular weight

3.1 Introduction

Since the discovery of the Ac-Ds transposable element system (McClintock,1948,1949) there has been a gradual accumulation of evidence for a variety of other transposable elements in maize. A number of these elements have been studied at the molecular level and include the *Spm-18* (Schwartz-Sommer et al.,1984), *Mu1* (Robertson.,1978), *Cin 1* (Shepard et al.,1984), *Bs1* (Johns et al.,1985) and *Tz86* (Dellaporta et al.,1986) elements. Several transposable elements have also been recognized in other plant species including the *Tam1* (Bonas et al.,1984), *Tam2* (Upadhyana et al.,1985) and *Tam3* (Sommer et al.,1985) elements from *Antirrhinum majus* and the *Tgm1* element (Vodkin et al.,1983) from soybean.

The *Tz86*, *Bs1*, *Tam2* and *Tgm1* elements were identified as insertions which inactivate gene function. These insertions have structural features similar to other transposable elements but genetic instability of these element has not been observed. Thus these sequences may represent non-autonomous elements which have lost the ability to transpose or alternatively, they may be "cryptic" elements and transpose only under specific enviromental conditions.

The origin and function of transposable elements is still disputed (McClintock,1951; Orgel and Crick,1980) but it is well documented that the insertion and/or subsequent excision of such elements can create genetic diversity (Wessler et al.,1986: Dennis et al.,1986). McClintock (1951) was the first to realize that following chromosome breakage transposable elements which were normally inactive could be induced to transpose. This activation of cryptic transposons has also been recognized following virus infection of plants (Mottinger et al.,1984) and the tissue culture regeneration of plants (Peschke et al.,1986). McClintock (1978)

has suggested that this activation of transposable elements may be as a result of environmental stress and that such a system allows a rapid reorganisation of the plant genome providing the genetic diversity required for an eventual increase in fitness.

The potential of transposable elements in gene tagging experiments and as integration modules for the transformation of plants (see review,Freeling,1984) makes their isolation from agriculturally important crops particularly relevant. In this paper we report the identification of an insertion in the wheat *Adh*-1A gene which has structural features similar to those of a transposable element. We also report the results of our attempts to induce genetic instability of the insertion at the *Adh* locus.

3.2 Materials and methods

Plant Material.

(a).Cultivars.

The distribution of the TLM insertion sequence in some members of the Gramineae family was determined using the following cultivars; *T.aestivum* cv. Millewa, *S.cereale* cv. Imperial, *H.vulgare* cv. Proctor and the Berkley Fast line of *Zea mays*. The wheat-barley addition lines were provided by Dr. A. Brown C.S.I.R.O. as was *H.vulgare* cv. Betzes and *T.aestivum* cv. Chinese Spring.

(b).Tissue culture regenerants of *T.aestivum* cv. Millewa.

Callus cultures were initiated from 46 immature embryos of *T.aestivum* cv. Millewa. Following cell proliferation, shoot regeneration was initiated by transferring the cultures to a low auxin medium (MS9) containing 0.5mg/l IAA and 1mg/l BAP. Ninety eight primary regenerants (SC_1) were removed and transferred to MS media (Murashige and Skoog, 1962) for induction of root formation (Larkin et al.,1984). Selfing of these primary regenerants gave

rise to the subsequent generations; SC_2, SC_3 etc. DNA was prepared from four seedlings of each of 90 SC_2, 4 SC_3 and 4 SC_4 lines and analyzed for tissue culture induced transposition by the method of Southern (1975) using oligolabelled pTLM (Fig 3.1b) as probe.

We are grateful to Dr. P. Larkin (CSIRO) for providing all the above somaclonal regenerants.

(c). Wheat lines carrying gametocidal genes from diploid *Triticum* species.

Three cultivars of common wheat, *Triticum aestivum*; cv. Chinese Spring, cv. Chris and cv. Selkirk, each carrying gametocidal chromosomes from *T.longissim*um or *T.speltoides*, in either a hetero- or hemizygous state were used in this investigation (Maan,1976,1980; Endo,1978,1985). A fourth line of wheat, R8, derived from [*T.speltoides* x *T.aestivum* cv.Chinese Spring] and back crossed several times to *T.aestivum* cv. Chris, is isogenic to *T.aestivum* and homozygous for the Gc1 gametocidal gene from *T.speltoides* (Tsujimoto and Tsunewaki,1984). R8 was crossed reciprocally with *T.aestivum* cv. Chris to produce progeny heterozygous for the Gc1 gene. The characteristics caused by these gametocidal chromosomes differ depending on the species from which the gametocidal agent was derived (Tsujimoto and Tsunewaki,1985). The gametocidal gene Gc1 from *T.speltoides* causes sterility, temperature dependent seed shrivelling and greatly increased mutation rate in the F1 generation. Gametocidal chromosomes, $4S^{sh}$ of *T.sharonensis* and $4S^{l}$ of T. longissimum, cause high rates of mutation and chromosome aberration in addition to sterility.

Tsujimoto and Tsunewaki (1985) have reported that the frequency of mutants in the F1 generation between wheat cultivars and *T.aestivum* lines CS-Gc1Gc1, CS+$4S^{sh}4S^{sh}$ and CS+$4S^{l}$ ranged from 7.1 to 19.8% and suggest that this high frequency of mutants may

be the result of a process similar to hybrid dysgenesis in *Drosophila*.

Table 3.1 describes the lines used and the number of seed assayed by Southern hybridization using pTLM as probe. Although the sample size analyzed for each gametocidal line (Table 3.1) was relatively small, it is likely that many mutations, which cause genetic rearrangements, result in no observable change in phenotype. The mutation frequency estimated by Tsujimoto and Tsunewaki (1985) is therefore probably considerably lower than the true mutation rate.

We are very grateful to Dr. S. S. Maan (North Dakota State University) for providing all of the above lines.

Sequence analysis.

Comparative nucleotide sequence analysis of the wheat *Adh*-1A and barley *Adh3* genes was done using the DIAGON program written by Staden (1982). The computer programs for DNA sequence analysis, developed by Dr. W. Bottomley, C.S.I.R.O., were used to analyze the TLM sequence for open reading frames and stem-loop structures.

Southern analysis.

Genomic DNA was prepared by the method of Dellaporta et al.(personal communication). 10ug of genomic DNA was digested with 25 units of restriction endonuclease and electrophoresed on a 0.8% agarose gel. The gel was acid washed (0.1M HCl 5 min), denatured (0.5M NaOH, 1.5M NaCl, 30 min.), neutralized (1M NH_4CH_3COO pH 8.3, 30 min.) and transferred to nitrocellulose (Southern 1975). The filter was baked, prehybridized and hybridized using the method of Dennis et al. (1984) with the exception that $3x10^7$ cpm of oligolabelled TLM probe, prepared using a oligolabelling kit (BRESA) to a specific activity of 10^8 cpm/ug was used.

Table 3.1

ANALYSIS OF GAMETOCIDAL LINES OF *T. AESTIVUM*.

T. aestivum lines containing gametocidal chromosomes $4S^l$ from *T. longissimum* OR THE GENE R8 FROM *T. SHARONENSIS*. $4S^{sh}$ from *T. sharonensis*.	No. of seed
1. Chinese Spring x (Chris + $4S^l$)	20
2. Chris x (Chris + $4S^l$)	20
3. Chris x (Chris + $4S^{sh}$)	15
4. (Chris + $4S^{sh}$) x Chris	20
5. Chris(R8) x Chris	19
6. Chris x Chris(R8)	45

3.3 Results.

We have cloned and determined the nucleotide sequence of an *Adh* gene from *T.aestivum* cv. Millewa (Mitchell et al.,1987). The gene has been mapped to the long arm of chromosome 1A and has been designated *Adh*-1A. Southern analysis indicates extensive sequence homology between the 3' untranslated region of the cloned wheat *Adh*-1A gene (p3NTR) and the homeologous *Adh* genes on the long arms of chromosome 1B and 1D of *T.aestivum*. This extensive sequence and position homology between *Adh* genes, each derived from a different genome, indicates that these genes probably arose from a common ancestor. Southern analysis of DNA from the related genus *H.vulgare* cv.Betzes, using p3NTR as probe, resulted in hybridization of the probe under stringent conditions to a single 2.9 kb BamH1 fragment. This result suggested that there may be an *Adh* gene in barley which is very closely related to the wheat *Adh*-1A gene. The *T.aestivum* cv.Chinese Spring - *H.vulgare* cv. Betzes addition lines developed by Islam et al.,(1981) were used to determine the chromosomal location of this fragment in barley. DNA was prepared from each of the six available addition lines containing barley chromosomes 1,2,3,4,6 or 7, restricted with BamH1 and probed with p3NTR. For each addition line only three cross hybridizing bands were observed which corresponded to the *Adh*-1A, 1B and 1D genes present in the *T.aestivum* cv.Chinese Spring background (data not shown). This result indicated, by difference, that the cross hybridizing band observed in *H.vulgare* cv. Betzes ~~must~~ PROBABLY originates from chromosome 5. This is consistent with extensive chromosomal mapping of isozyme loci in barley and wheat which has identified chromosome 5 of barley as being homeologous to chromosome 1 of wheat (See summary, Brown,1983).

M.Trick (per.comm) has recently cloned and determined the

nucleotide sequence of the three *Adh* genes from *H.vulgare* cv.Proctor. These genes have been tentatively assigned as *Adh1*, *Adh2*, and *Adh3*. Restriction mapping of Proctor, using p3NTR as probe, indicated that the wheat *Adh*-1A gene was homologous to barley *Adh3*. The probe did not hybridize to either barley *Adh1* or *Adh2*.

Nucleotide sequence comparison revealed extensive homology in the coding region of the wheat *Adh*-1A genes and all three of the barley genes. This is probably due to selection pressure to maintain enzyme function and has been observed for many other *Adh* genes (Chang and Meyerowitz,1986; Dennis et al.,1985). No significant homology between the untranslated sequences of wheat *Adh*-1A and barley *Adh1* was found. Barley *Adh2* showed slight homology to intron sequences of wheat but only in stretches of 8 to 10 bases, usually confined to sequences at or near the intron splice junctions. Barley *Adh3* however, showed extensive but discontinuous homology to both the coding and noncoding regions of the wheat *Adh*-1A gene (Fig 3.1a).

The discontinuity of sequence between the wheat *Adh*-1A and barley *Adh3* genes is a result of an insertion (TLM) of 1794 bp in intron 3 of the wheat gene. The restriction map of this insertion is shown in Fig 3.1b. Closer inspection of this region indicates that there has been a direct duplication in wheat of an 8bp sequence (CTATGAAG) common to both the wheat and barley *Adh* genes. Immediately internal to these repeats is found an extensive GC rich, 14 bp imperfect inverted repeat (Fig 3.2). These structural features are characteristic of the short terminal inverted repeat class of transposable elements and indicates that the discontinuous homology between the wheat *Adh*-1A and barley *Adh3* genes is very likely the result of an insertion due to a

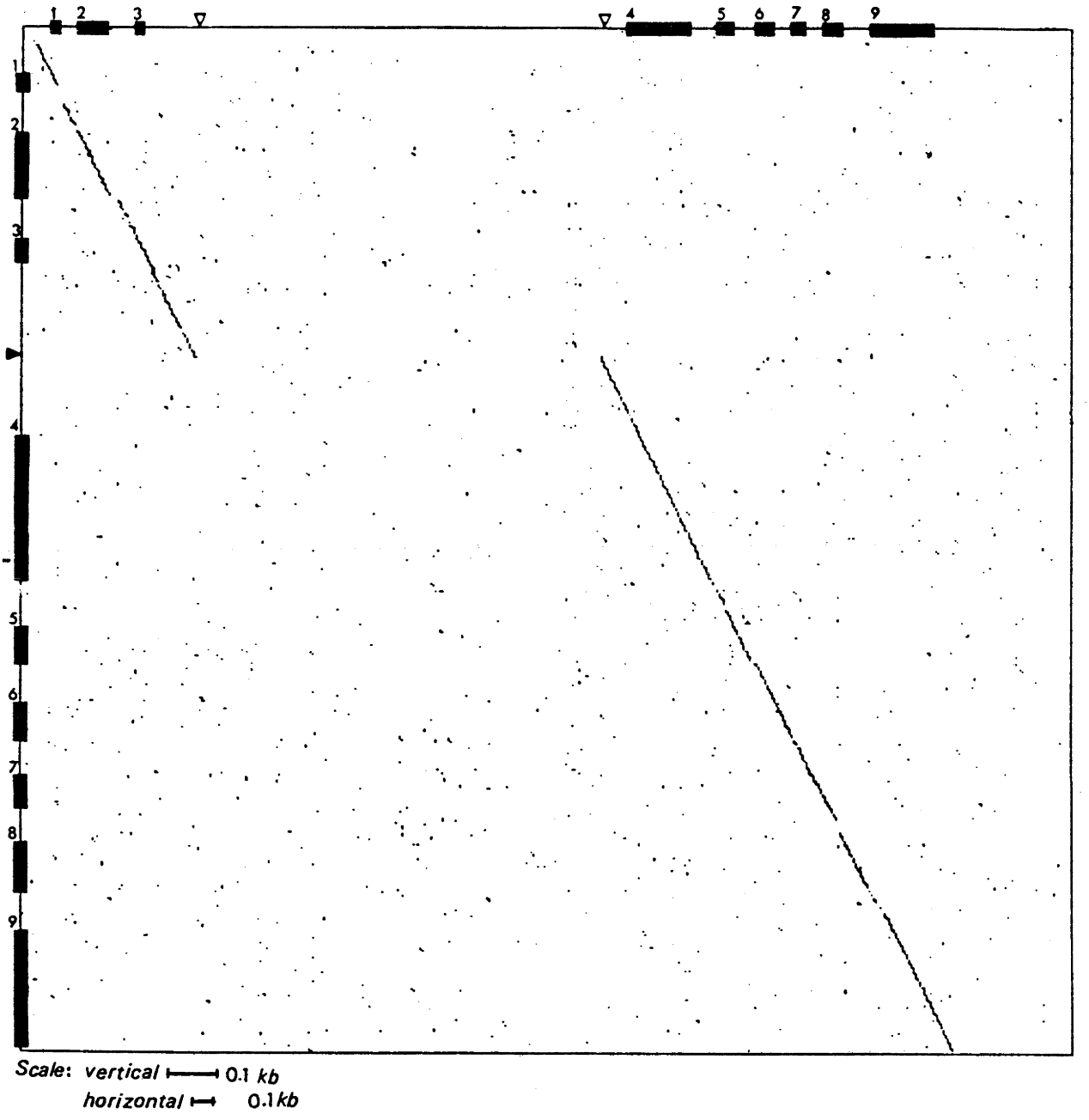


Figure 3.1(a)
Diagonal comparison of the nucleotide sequences of the wheat *Adh*-1A (horizontal) and barley Adh3 (vertical) genes. Exons are denoted by solid boxes and are numbered. For technical reasons the scales on the X and Y axis are different. The breakpoint in homology between the wheat and barley genes is indicated by a solid triangle. Open triangles represent the terminal regions of the insertion (TLM) in the wheat gene.

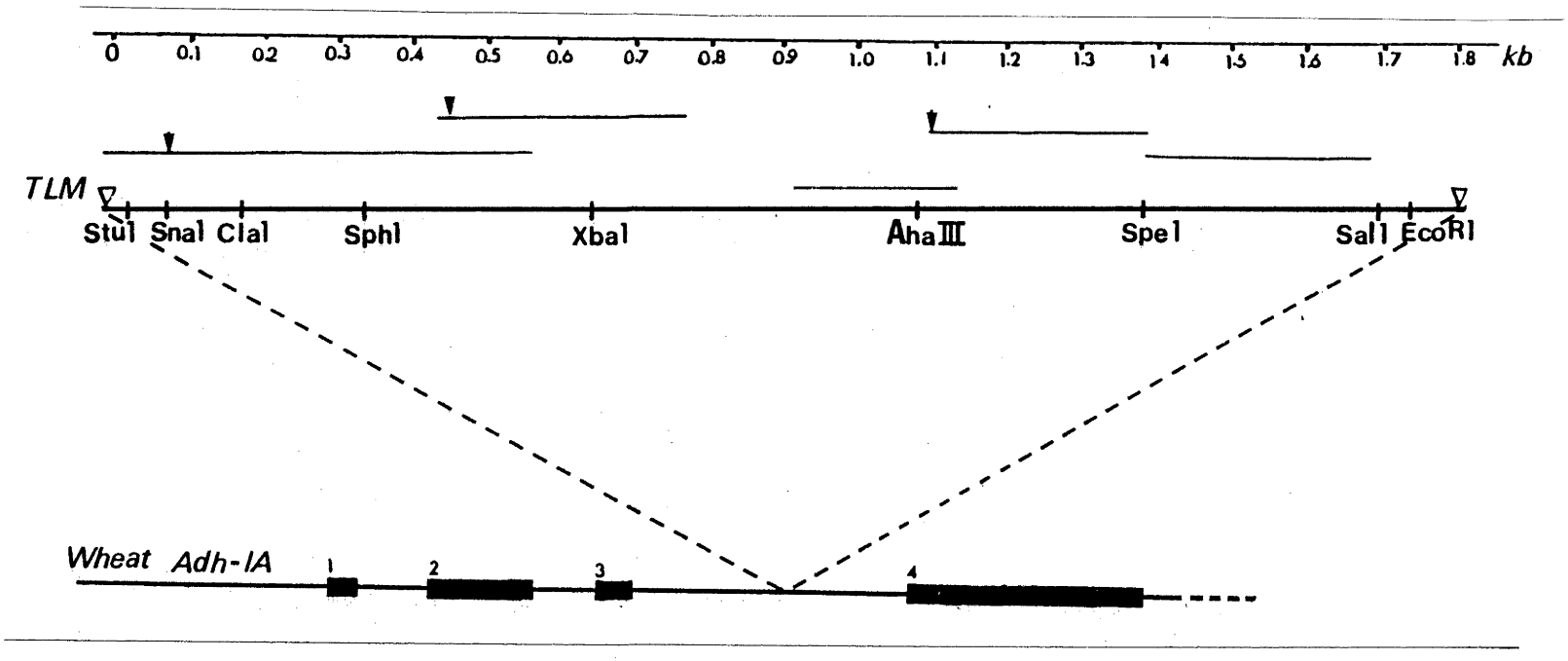
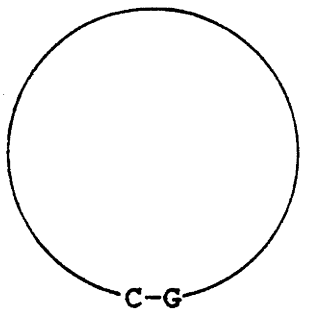


Figure 3.1(b)
Restriction map of the TLM insertion showing the site of insertion in the wheat *Adh*-1A gene. Exons are denoted by solid boxes and are numbered. Open triangles represent the terminal regions of the insertion. Open reading frames are indicated above the restriction map with vertical arrows indicating the position of the first methionine. The subclone (pTLM) used as probe in Southern analysis extends from the Stu1 site to the EcoR1 site.



C-G
C T
G-C
G-C
A-T
G-C
G-C
C-G
G-C
G-C
T T
C C
A-T
C-G

Wheat *Adh*-1A
AATGGGATAAAACATA**CTATGAAG** **CTATGAAG**AATCATGGGAAAA

AATGGGATAAAACATA**CTATGAAG**AATCATGGGAAAA
Barley *Adh3*

Figure 3.2
Comparison of the nucleotide sequence of the barley *Adh3* and wheat *Adh*-1A genes at the site of insertion. The direct 8 bp genomic duplication is shown in bold lettering. The 5' and 3'termini of the TLM insertion are drawn in a stem-loop structure.

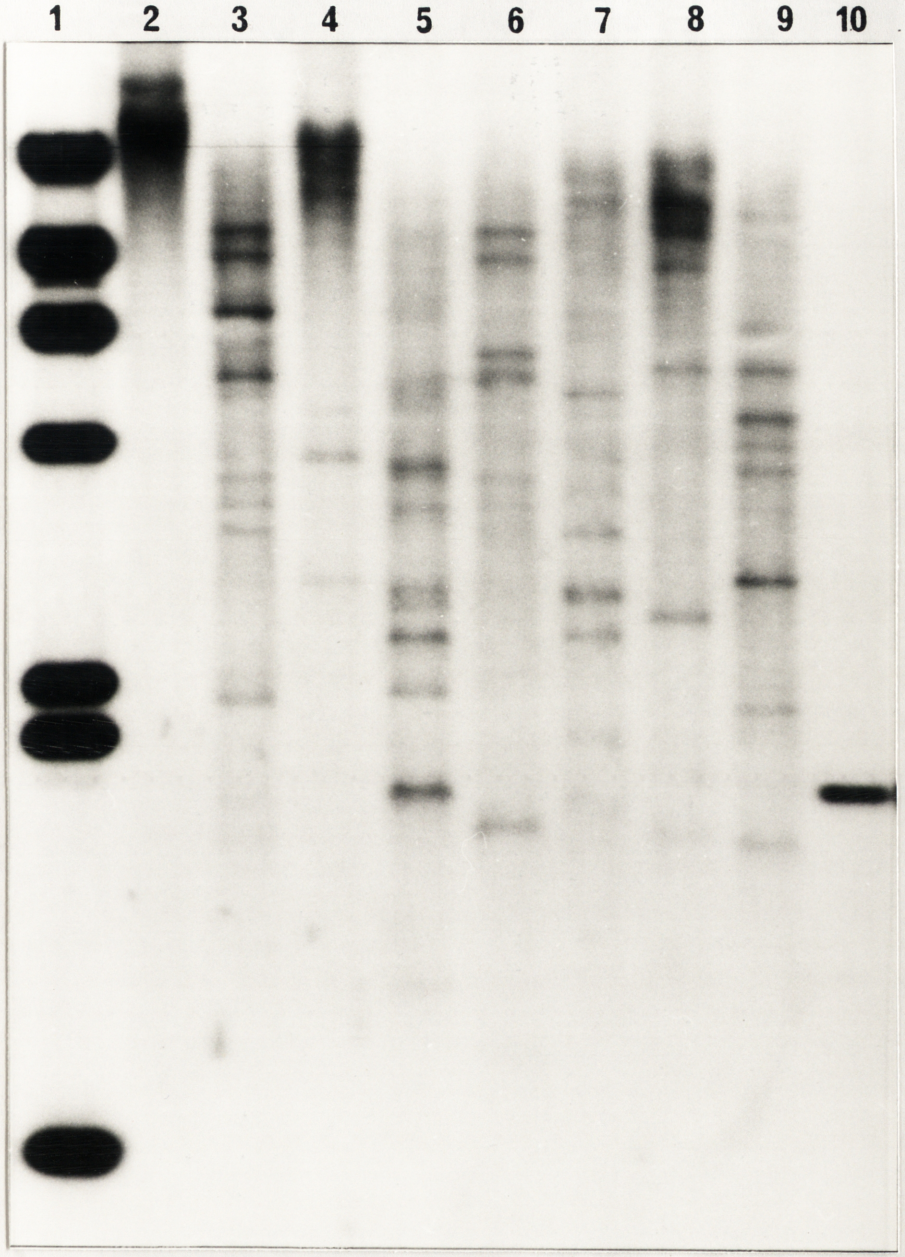


Figure 3.3
Southern analysis of *T.aestivum* cv Millewa probed with pTLM. Lane 1. end labelled MW marker 23.1kb, 9.4kb, 6.6kb, 4.4kb, 2.3kb, 2.0kb, 0.56kb; 2.uncut *T.aestivum* cv Millewa; 3.EcoR1; 4.sal1; 5.Stu1/EcoR1; 6.EcoR1/Cla1 7.Sal1/Stu1; 8.Sal1/Cla1; 9.Nco1; 10.pTLM (control).

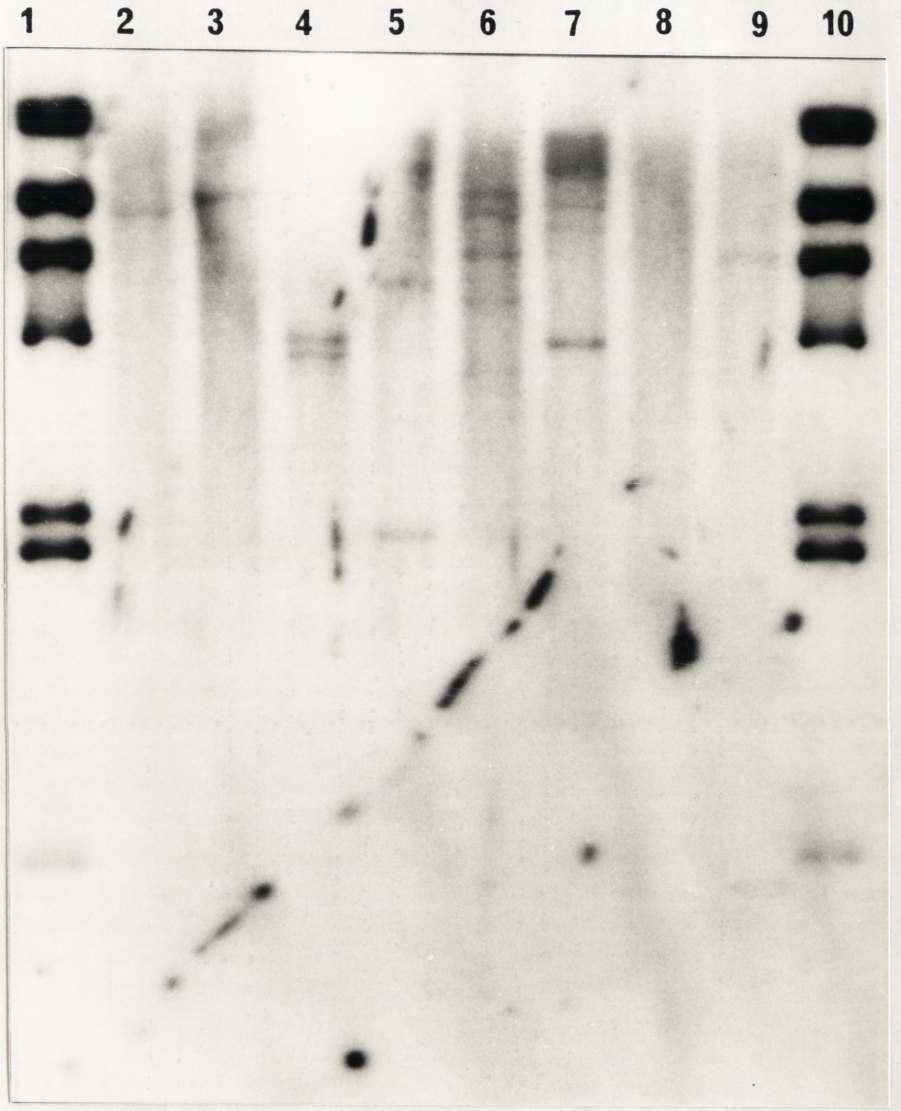


Figure 3.4
Southern analysis of A, B and D diploid wheats, barley, maize and rye digested with EcoR1 and probed with pTLM. Lane 1,10 end labelled MW marker 23.1kb, 9.4kb, 6.6kb, 4.4kb, 2.3kb, 2.0kb, 0.56kb; 2.*T.monococcum* (AA); 3.*T.speltoides* (BB); 4.*T.sharonensis* (BB); 5.*T.tauschii* (DD); 6.*T.aestivum* cv Millewa; 7.*H. vulgare* cv Proctor; 8.*Z.mays* (BKF); 9.*S.cereale* cv Imperial.

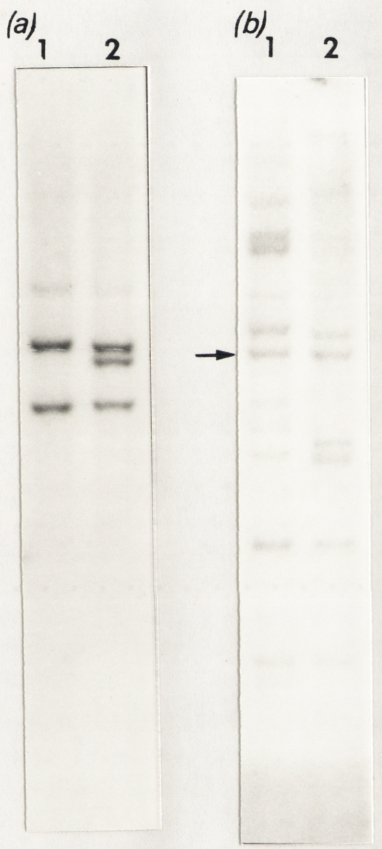


Figure 3.5
Southern analysis of *T.dicoccoides* (AABB,2n=28) digested with Hpa1/Bgl11 and probed with (a)p3NTR and (b)reprobed with pTLM. Lane 1 *T.dicoccoides* TA1058-1; 2.*T.aestivum* cv Millewa. The p3NTR probe is specific for the wheat *Adh* genes on chromosomes 1A, 1B and 1D. The pTLM probe reveals that the insertion element (TLM) is also present in the *Adh*-1A gene from tetraploid wheat as indicated by the arrow.

similar to hybrid dysgenesis in *Drosophila*, which is known to be associated with P element transposition (Rubin et al.,1982), suggests that induction of transposition may be occurring in these lines (see materials and methods). Southern hybridization was used to determine if wheat lines carrying these gametocidal genes were activating transposition of sequences related to the TLM insertion. DNA was prepared from individual seedlings of each line described in Table 3.1 and probed with oligolabelled TLM. In no case was there any change in either band number or position indicating that no transposition of the TLM insertion or cross hybridizing sequences had occurred.

3.4 Discussion.

Transposition was first recognised by McClintock (1948,1949) and since then has been well documented in both procaryotes and eucaryotes (see review Shapiro, 1983). Molecular analysis of the elements involved in this transposition has revealed that they can be classified on the basis of their terminal structures and the rearrangements of genomic sequences caused by integration of the element. All the known plant controlling elements, with the exception of *Bs1*, belong to the short inverted terminal repeat class of transposable elements which cause short, direct duplications of flanking genomic DNA. The 8 bp direct duplication and 14 bp imperfect inverted terminal repeat structures of the integrated TLM insertion are typical of this class of insertion.

The presence of sequences related to the TLM insertion in diploid *Triticum* genomes A, B, D, and the absence of a restriction fragment "ladder" in Southerns of genomic DNA probed with the TLM insertion, indicates that these related sequences are not in tandem arrays but are dispersed throughout the genome. The

dispersed repetitive sequences, in *Drosophila melanogaster*, have been shown to have termini that are either direct or inverted repeats and have been found to cause duplication of short stretches of genomic DNA upon insertion (Calos and Miller,1980). This would appear to indicate that nearly all of the middle repetitive fraction of *Drosophila* DNA is, or once was transposable. In the case of wheat 30 to 40% of the genome is present as middle repetitive DNA (Flavell et al., 1981). An extensive study of the nucleotide sequence organization in the chromosomes of wheat, rye, barley and oats, which all evolved from a common ancestor, has provided evidence for sequence translocations in these species during evolution (Flavell et al.,1981). These results would seem to suggest that part of the genome of cereals has been amplified and dispersed by DNA transposition. Due to natural selection, and as a result of agricultural practices which select for uniformity and constancy of phenotype, it is unlikely that this dispersed repetitive DNA presently consists of autonomously transposing sequences. It follows then, that the results of transposition events are more likely to be observed as stable integrated insertions similar to the TLM insertion.

There is however, some evidence to suggest that stable insertions can, under certain conditions, be induced to transpose. In maize, induction of two transposable element systems, *Ac* and Dt has been observed as the result of repeated chromosome breakage (McClintock,1951; DÖerschug,1973), systemic infection with barley stripe mosaic virus (Sprague et al.,1963) was correlated with the presence of unrelated insertion mutants in the *Adh* and *Sh* loci (Dellaporta et al.,1986; Mottinger et al.,1984) and transposition of cryptic *Ac* elements is reported to have been induced in plants

regenerated from tissue culture (Peschke et al.,1986). If this transposition is enviromentally induced as suggested by McClintocks "genome stress" hypothesis (McClintock,1978) it is likely that different controlling elements are induced by different enviromental conditions. Similarly, observations that wheat lines with different gametocidal genes result in different characteristics (Tsujimoto and Tsunewaki,1985) may indicate the induction of different controlling elements by each gametocidal agent. Thus the absence of induced transposition of the TLM insertion or related sequences in wheat plants regenerated from tissue culture or wheat lines carrying gametocidal genes is not entirely unexpected.

There are a number of large open reading frames in the cloned TLM insertion which may be involved in the encoding of a transposase. Translations of these ORF's show no amino acid sequence sequence homology to the transposase from either the P element of *Drosophila* (Rio et al.,1986) or the putative transposase from the maize *Ac* element (Kunze et al.,1987). There are no striking nucleotide sequence homologies to other insertion or transposable elements, although the four bases at the 5' terminus of the insertion are identical to those found for the *Tam* elements of *Antirrhinum majus* and the *Spm-18* element of maize.

Acknowledgements

We would like to thank Rhonda Perriman for her excellent technical assistance related to this work. She is responsible for the DNA preparations and Southern analysis of the plants regenerated from tissue culture described on page 51 paragraph 4 of this thesis.This research was supported by the Rockefeller Foundation.

3.5 References

Bonas, V., Sommer, H. and Saedler, H. (1984). The 17kb *Tam1* element of *Antirrhinum majus* induces a 3-bp duplication upon integration into the chalcone synthase gene. Embo J. 3, 1015-1019

Brown, A.H.D. (1983). Hexaploid wheat. in Isozymes in Plant Genetics and Breeding, Part B (Tanksley, S.D. and Orton, J.J. (ed)) Elseiver Science Publishers, B.V.,Amsterdam.

Calos, M. and Miller, J.H. (1980). Transposable elements. Cell 20, 579-595.

Chang, C. and Meyerowitz, E.M.(1986). Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. Proc. Natl. Acad. Sci. U.S. A. 83, 1408-1412.

Dellaporta, S.L. Chomet, P.S., Mottinger, J.P., Wood, J.A., Yu, S.M. and Hicks, J.B. (1986) Endogenous transposable elements associated with virus infection in maize. Cold Spring Harbour Symp. Quant. Biol. XLIX, 321-328.

Dennis, E.S., Gerlach, W.L., Pryor,A.J., Bennetzen,J.L., Inglis,A., Llewellyn,D., Sachs,M.M., Ferl,R.J. and Peacock,W.J. (1984). Molecular analysis of the alcohol dehydrogenase (*Adh1*) gene of maize. Nucl. Acids Res. 12, 3983-3999.

Dennis, E.S., Sachs, M.M., Gerlach, W.L., Finnegan, E.J. and Peacock, W.J. (1985). Molecular analysis of the alcohol dehydrogenase 2 (*Adh2*) gene of maize. Nucl. Acids Res. 13, 727-743.

Dennis, E.S., Gerlach, W.L., Peacock, W.J. and Schwartz, D. (1986). Excision of the *Ds* controlling element from the *Adh* gene of maize. Maydica XXXI 47-57.

Doerschug, E.B. (1973). Studies of *Dotted*, a regulatory element in maize. Theor. Appl. Genet. 43, 182-189.

Endo, T.R. (1978). On the Triticum chromosomes having gametocidal action on common wheat. Proc.5th Int. Wheat Genetics Symposium pp. 306-314.

Endo, T.R. (1985) Two types of gametocidal chromosome of *Triticum sharonensis* and *Ae. longissimum*. Jpn. J. Genet. 60, 125-135.

Flavell, R.B., O'Dell, M., Hutchinson, J. (1981). Nucleotide sequence organization in plant chromosomes and evidence for sequence translocation during evolution. Cold Spring Harbour Symp. Quant. Biol. 45, 501-508.

Freeling, M. (1984). Plant transposable elements and insertion sequences. Ann. Rev. Plant Physiol. 35, 277-298.

Helbaek, H. (1959). Domestication of food plants in the old world. Science 130, 365-372.

Islam, A.K.M.R., Shepherd,K.W. and Sparrow, D.H.B. (1981). Isolation and characterization of euplasmic wheat-barley chromosome addition lines. Heredity 46, 161-174.

Johns, M.A., Mottinger, J. and Freeling, M. (1985). A low copy number, copia-like transposon in maize. EMBO J. 4, 1093-1102.

Kunze, R., Stochaj, U., Laufs, J. and Starlinger, P. (1987). Transcription of transposable element *activator* (Ac) of *Zea mays* L. EMBO J. 6, 1555-1563.

Larkin, P.J., Ryan, S.A., Brettell, R.I.S. and Scowcroft, W.R. (1984). Heritable somaclonal variation in wheat. Theor. Appl. Genet. Theor. Appl. Genet. 67, 443-455.

Maan, S.S. (1975). Exclusive preferential transmission of an alien chromosome in common wheat. Crop Science 15, 287-292.

Maan, S.S. (1976). Alien chromosome controlling sporophytic sterility in common wheat. Crop Science 16, 580-583.

Maan, S.S. (1980). Alteration of sporophytic sterility mechanism in wheat. Heredity 71, 75-82.

McClintock, B. (1948). Mutable loci in maize. Carnegie Inst. Washington Year Book 47, 155-169.

McClintock, B. (1949). Mutable loci in maize. Carnegie Inst. Washington Year Book 48, 142-154.

McClintock, B. (1951). Chromosome organization and genetic expression. Cold Spring Harbour symp. Quant. Biol. 16, 13-47.

McClintock, B. (1978). Mechanisms that rapidly reorganise the genome. Stadler Genet. Symp. 10, 25-48.

Mitchell, L.E., Dennis, E.S. and Peacock, W.J. (1987). Molecular analysis of an alcohol dehydrogenase gene from chromosome 1 of wheat (in preparation).

Mottinger, J.P., Johns, M.A. and Freeling, M.(1984). Mutations of the *Adh1* gene in maize following infection with barley stripe mosaic virus. Mol. Gen. Genet. 195, 367-369.

Murashige, T. and Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol. Plant 15, 473-497.

Orgel, L.E. and Crick, F.H.C. (1980) Selfish DNA: The ultimate parasite. Nature 284, 604-607.

Peschke, V.M., Phillips, R.L. and Gengenbach, B.G. (1986). Transposable element activity in progeny of regenerated maize plants. 6th international congress of plant tissue and cell culture.

Rio, D.C., Laski, F.A. and Rubin, G.M. (1986). Identification and immunochemical analysis of biologically active P element transposase. Cell 44, 21-32.

Robertson, D.S. (1978) Characterization of a mutation system in maize. Mutat. Res. 51, 21-28.

Rubin, G.M., Kidwell, M.G. and Bingham, P.M. (1982). The molecular basis of P-M Hybrid dysgenesis: The nature of induced mutations. Cell 29, 987-994.

Schwarz-Sommer, Zs., Gierl, A., Klosgen, R.B., Wienand, U., Peterson, P.A. and Saedler, H. (1984). The *Spm* (En) transposable element controls the excision of a 2kb DNA insert at the *wx*$^{m-8}$ allele of *Zea mays*. EMBO J. 3, 1021-1028.

Shapiro, J.A. (1983). (ed.) Mobile Genetic Elements. New York, Academic.

Shepard, N.S., Schwarz-Sommer, Z., Blumberg, J., Gupta, M., Wienand, U. and Saedler, H. (1984). Similarity of the Cin 1 repetitive family of *Zea mays* to eucaryotic transposable elements. Nature. 307, 185-187.

Sommer, H., Carpenter, R., Harrison, B.J. and Saedler, H. (1985). The transposable element *Tam3* of *Antirrhinum majus* generates a novel type of sequence alteration upon excision. Mol. Gen. Genet. 199, 225-231.

Southern, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503-507.

Sprague, G.F., McKinney, H.H., Greeley, L. (1963). Virus as a mutagenic agent in maize. Science 14, 1052-1053.

Staden, R.(1982). An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. Nucl. Acids Res. 10, 2951-2961.

Tsujimoto, H. and Tsunewaki, K. (1985). Hybrid dysgenesis in common wheat caused by gametocidal genes. Jpn. J. Genet. 60, 565-578.

Tsujimoto, H. and Tsunewaki, K. (1984). Gametocidal genes in wheat and its relatives. I.Genetic analyses in common wheat of a gametocidal gene derived from *Triticum speltoides*. Can. J. Genet. Cytol. 26,78-84.

Upadhyana, K.C., Sommer, H., Krebbers, E. and Saedler, H.(1985). The paramutagenic line niv-44 has a 5kb insert, *Tam2* in the chalcone synthase gene of *Antirrhinum majus*. Mol. Gen. Genet. 199, 201-207.

Vodkin, L.O., Rhodes, P.R. and Goldberg, R.B. (1983). A lectin gene insertion has the structural features of a transposable element. Cell 34, 1023-1031.

Wessler, S.R., Baran, G., Varagona, M. and Dellaporta, S.L. (1986). Excision of *Ds* produces waxy proteins with a range of enzymatic activities. EMBO. J. 5, 2427-2432.

CHAPTER 4

The Wheat *Adh*-1A gene as a molecular probe for analysis of the evolution of diploid and polyploid wheats.

L.E.Mitchell*, E.S.Dennis, W.J.Peacock.

Division of Plant Industry,
CSIRO
GPO Box 1600
Canberra ACT 2601 Australia

Abbreviations; RFLP restriction fragment linked polymorphism

* Also from Research School of Biological Sciences
Australian National University

4.1 Introduction

Polyploidy has played a major role in the evolution of wheat with the species of wheat existing at three main ploidy levels: diploid, tetraploid and hexaploid, all established on a basic chromosome number of seven.

The diploid wheat species of which there are eleven ~~known~~ RELATED members, appear to have diverged from a common ancestor and have developed characteristic morphological and genomic differences. Phylogenetic relationships indicate that the cultivated tetraploid species *T.turgidum ssp. dicoccoides* (AABB,2n=28) and a further wild tetraploid *T. timopheevi ssp. araraticum* (AAGG,2n=28) have a common A genome possibly derived from the diploid *T. monococcum* L. (AA,2n=14) (Kihara 1954). This feature is characteristic of the polyploid wheats and has led Zohary and Feldman (1962) to describe the common genome as the pivotal genome. Although interspecific hybridization is rare in the diploid wheats it is a very common occurrence in the polyploid wheats (Feldman,1965). This may be due to the pivotal genome acting as a buffer in the polyploid wheats ensuring at least some fertility in resulting hybrids. If this is the case then it might be expected that the pivotal genome would be essentially unaltered from the diploid progenitor.

A number of theories have been put forward to account for the existence of the B and G genomes of T.*turgidum ssp. dicoccoides* and *T. timopheevi ssp. araraticum* as there is no obvious diploid progenitor for either of these genomes. Tsunewaki and Ogihara (1983) proposed, because of the cytoplasmic differences between *T.dicoccoides* and *T. araraticum*, a diphyletic origin for the tetraploid wheats. This theory requires two independent hybridization events with *T.longissimum* contributing both the B genome and cytoplasm to *T.dicoccoides*, and *T.speltoides*

contributing the G genome and cytoplasm to *T. araraticum*.

Several workers (Sarkar and Stebbins 1956, Johnson 1975, Wagenaar 1961) have suggested a monophyletic origin for the two tetraploid species. This is based on cytogenetic evidence which indicates that the B genome of *T.dicoccoides* and the G genome of *T. araraticum* show more affinity to each other than to the genome of any other putative diploid progenitor. Such an origin implies the existence of a mechanism to account for the differentiation of the B and G genomes. Zohary and Feldman (1962) suggested that this differentiation may arise by the common A genome acting as a buffer during hybridization while the B and G genomes recombine. Others have proposed that the acquisition of extensive amounts of chromatin (Dhaliwal and Johnson 1978), differential accumulation of major structural changes and rapid nucleotide divergence (Dvorjak and Appels 1982) as alternative mechanisms which may have resulted in the differentiation of the B and G genomes.

Gill and Chen (1987) have proposed a modification of the monophyletic model in which *T. araraticum* arose prior to *T.dicoccoides*. They suggest that *T.araraticum* arose by hybridization of *T.monococcum* L. and *T. speltoides* (Maan 1973) and that subsequent introgressive hybridization of *T.araraticum* with an unknown *Triticum* species resulted in both the distinctive cytoplasm and B genome of *T. dicoccoides*.

Although the relationships among the tetraploid species are still uncertain it is generally agreed that the A genome derived hexaploid wheats arose by subsequent hybridization of the tetraploid species with a further diploid (McFadden and Sears 1944). *T.araraticum*, by hybridization with *T.monococcum* L. gave rise to the hexaploid *T.zhukovskyi* (AAAAGG, 2n = 42) and *T.dicoccoides* hybridized with *T.tauschii* to give the common bread

wheat *T.aestivum* (AABBDD, 2n = 42)

The evolutionary relationships between the diploid and polyploid wheats have been determined mainly by cytology, chromosome pairing studies and more recently using molecular probes to highly repeated DNA sequences. In this paper we report the use of a subclone of the wheat *Adh*-1A gene as a probe for the RFLP mapping of the alcohol dehydrogenase loci on chromosome 1. This analysis has confirmed the common origin of the A genome in all A genome derived polyploid wheats and identified particular accessions of *T.dicoccoides* and *T.tauschii* which are likely to be the direct ancestors of *T.aestivum*. We have also used this technique to distinguish between the diphyletic, monophyletic and introgressive models for the origin of the *T.dicoccoides* and *T.araraticum* tetraploid wheats.

4.2 Materials and methods

a) Plant material: *T.aestivum* cv Millewa, cv Chinese Spring were obtained from P.Larkin (CSIRO). *T.dicoccoides* accessions AUS20303, CPI99970, AUS17480, AUS17980 and *T.tauschii* accessions 80 and 82 were provided by Dr. O. Frankel (CSIRO). *T.tauschii* accessions D1, D2 and D3 were supplied by Dr. R. Appels (CSIRO). All remaining lines were obtained from Dr. B Gill, Dept. of Plant Pathology, Kansas State University.

b) Southern analysis: Genomic DNA was prepared by the method of Dellaporta et al. (personal communication). 10ug of genomic DNA was digested with 25 units of restriction endonuclease and electrophoresed on 0.8% agarose gel. The gel was acid washed (0.1M HCl, 5 min.), denatured (0.5M NaOH, 1.5M NaCl, 30 min.), neutralised (1M NH_4CH_3COO pH 8.3, 30 min.) and transferred to nitrocellulose (Southern 1975). The filter was baked,

prehybridized and hybridized as described by Dennis et al. (1984) with the exception that $3x10^7$ cpm of an oligolabelled probe prepared using a oligolabelling kit (BRESA) to a specific activity of 10^8 cpm/ug was used.

(c) Description of probes: Mitchell et al.(1987a) have cloned and determined the nucleotide sequence of an *Adh* gene from *T.aestivum* cv Millewa that maps to the long arm of chromosome 1A. Additional copies of the gene are present on chromosomes 1B and 1D. Two subclones of the wheat *Adh*-1A gene were used as molecular probes for analyzing the evolution of wheat: p3NTR, a 0.4kb Spe1/Hpa1 fragment from the 3' untranslated region and pTLM, a 1.8kb Stu 1/EcoR1 fragment from intron 3 (Fig 4.1).

The p3NTR subclone when used as a probe in genomic Southerns of hexaploid wheat hybridizes to three identical bands in both cv. Millewa and cv. Chinese Spring corresponding to the *Adh* genes on chromosomes 1A, 1B and 1D. The designation of each gene was determined using Chinese Spring nullisomic-tetrasomic lines (Mitchell et al.,1987a). Nucleotide sequence analysis of the wheat *Adh*-1A and a closely related *Adh* gene from barley (M.Trick per comm) has identified pTLM as being an insertion into intron 3 of the wheat *Adh*-1A gene (Mitchell et al.,1987b). Southern blots of hexaploid wheat DNA (*T. aestivum* cv Millewa, cv Chinese Spring) using oligolabelled pTLM as probe indicate that there are approximately twelve copies of this insertion in the hexaploid wheat genome and that neither the 1B nor 1D *Adh* genes contain the pTLM insertion.

4.3 Results

One function of alcohol dehydrogenase (ADH; alcohol: NAD oxidoreductase EC1.1.1.1) is to catalyze the reduction of

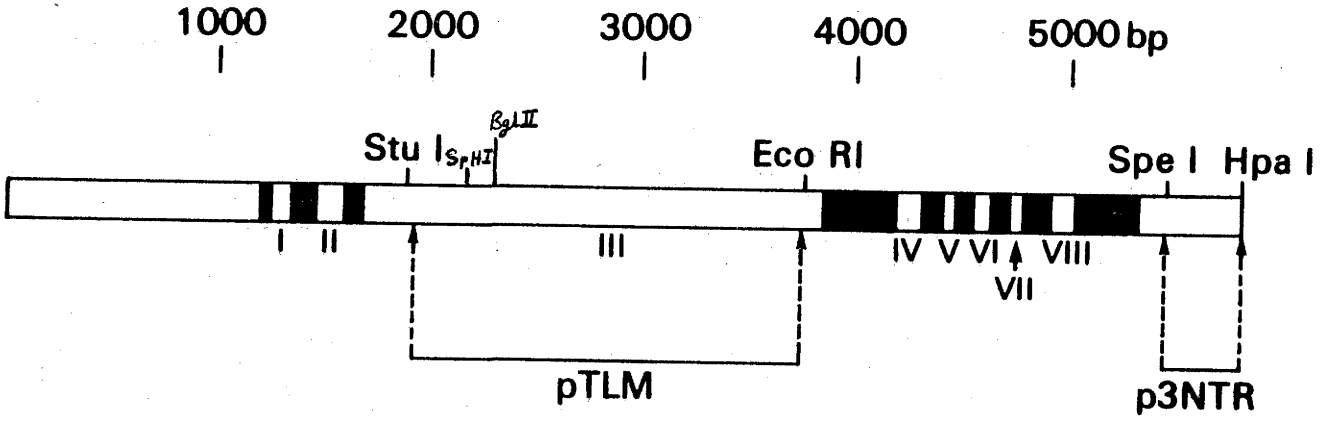


Figure 4.1
Restriction map of the wheat *Adh*-1A gene. Bracketed regions indicate the extent of the pTLM and p3NTR subclones. Exons are denoted by solid bars. *Introns* are numbered.

acetaldehyde to ethanol thereby regenerating NAD^+ which is necessary for glycolytic ATP production during hypoxic stress.

Numerous alcohol dehydrogenases have been identified in wheat (see review, Hart 1983) and mapped to specific chromosomes using the aneuploid lines of *Triticum aestivum* cv. Chinese Spring developed by Sears (1966). In all cases aneuploid genetic studies of hexaploid wheat indicate that there is a triplication of each *Adh* locus.

The p3NTR probe (Fig 4.1) used in this work hybridizes equally to the *Adh* genes from chromosome 1A, 1B and 1D and as such is a useful probe for analyzing simultaneously three closely related genes each from a different genome. The pTLM probe (Fig 4.2) is specific for the *Adh*-1A locus (see materials and methods).

Analysis of *Adh* loci in *T.dicoccoides* (AABB,2n=28).

DNA was isolated from nine accessions of *T.dicoccoides* (AABB, 2n=28). Each DNA sample was restricted with Hpa1/BglII and Hpa1/Sph1, electrophoresed on 0.8% agarose and probed with oligolabelled p3NTR (Figs 4.2a, 4.3a).

In every accession of *T. dicoccoides*, and for both Hpa1/BglII and Hpa1/Sph1 digests a restriction fragment of a size similar to that for the *Adh*-1A gene from hexaploid wheat hybridized to the p3NTR probe (Figs 4.2a, 4.3a). To confirm that these restriction fragments were directly related to the *Adh*-1A gene from hexaploid wheat the p3NTR probe was melted off and the filters reprobed with oligolabelled pTLM (Figs 4.2b, 4.3b). Superposition of the autoradiograms indicated that the pTLM probe hybridized to the putative *Adh*-1A gene from each tetraploid accession. This confirmed that the tetraploid *Adh* locus also had the pTLM insertion and justified its assignment as *Adh*-1A.

The *Adh*-1B gene of *T.dicoccoides* appeared to be considerably

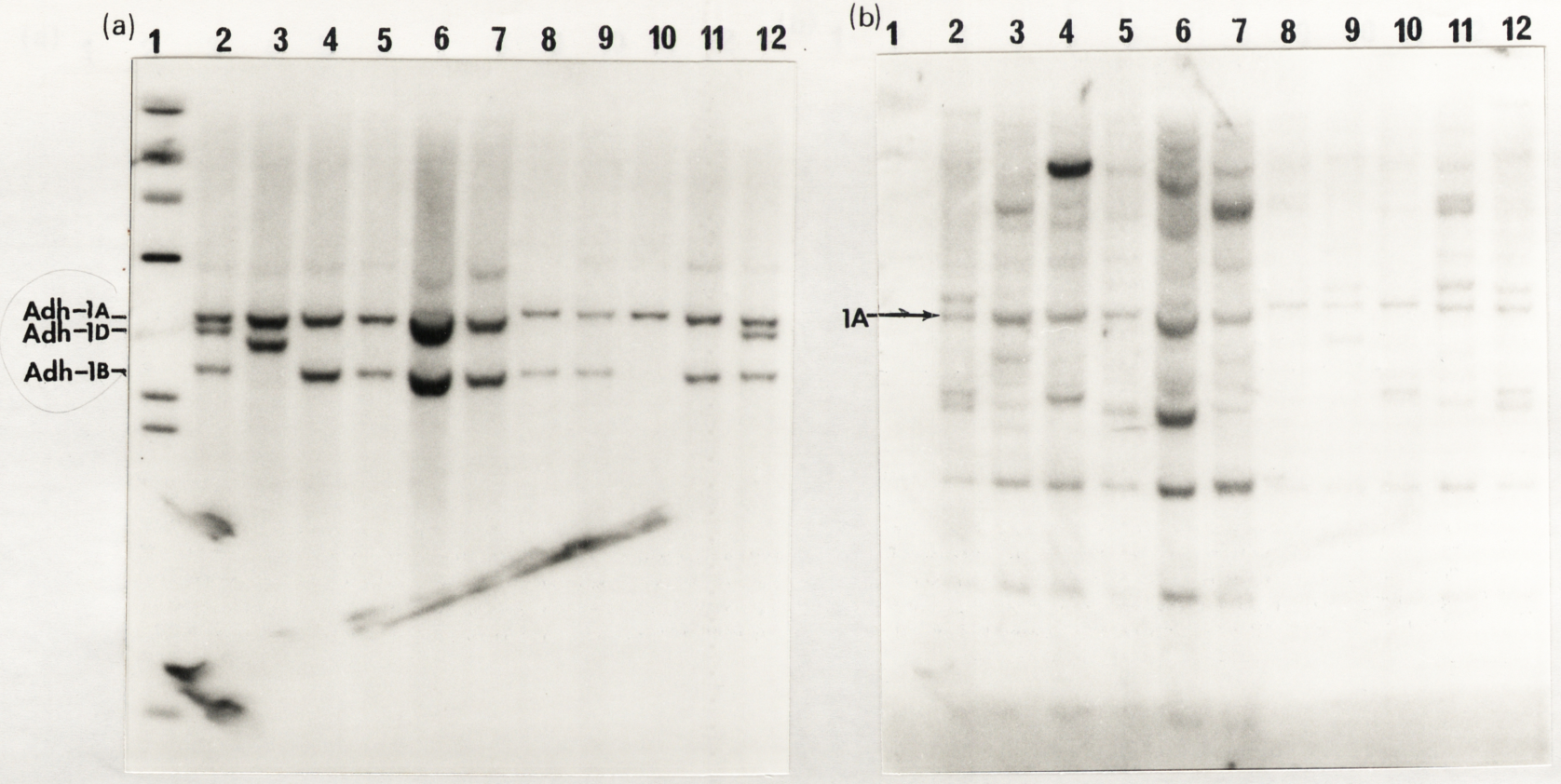


Figure 4.2
Southern analysis of *T.dicoccoides* accessions digested with HpaI/BglII and probed with: (a) p3NTR and (b) reprobed with pTLM. Lane 1. end labelled MW markers 23.1kb, 9.4kb, 6.6kb, 4.4kb, 2.3kb, 2.0kb, 0.56kb; Lanes 2. and 12. *T.aestivum* cv. Millewa; 3. *T.dicoccoides* accession AUS20303; 4.CPI99970; 5. AUS17480; 6. AUS17980; 7. G278; 8. TA54; 9. TA71; 10. TA84; 11. TA1058-1. The wheat *Adh*-1A, 1B and 1D genes are labelled. Arrow indicates the position of the wheat *Adh*-1A gene which contains the pTLM insertion.

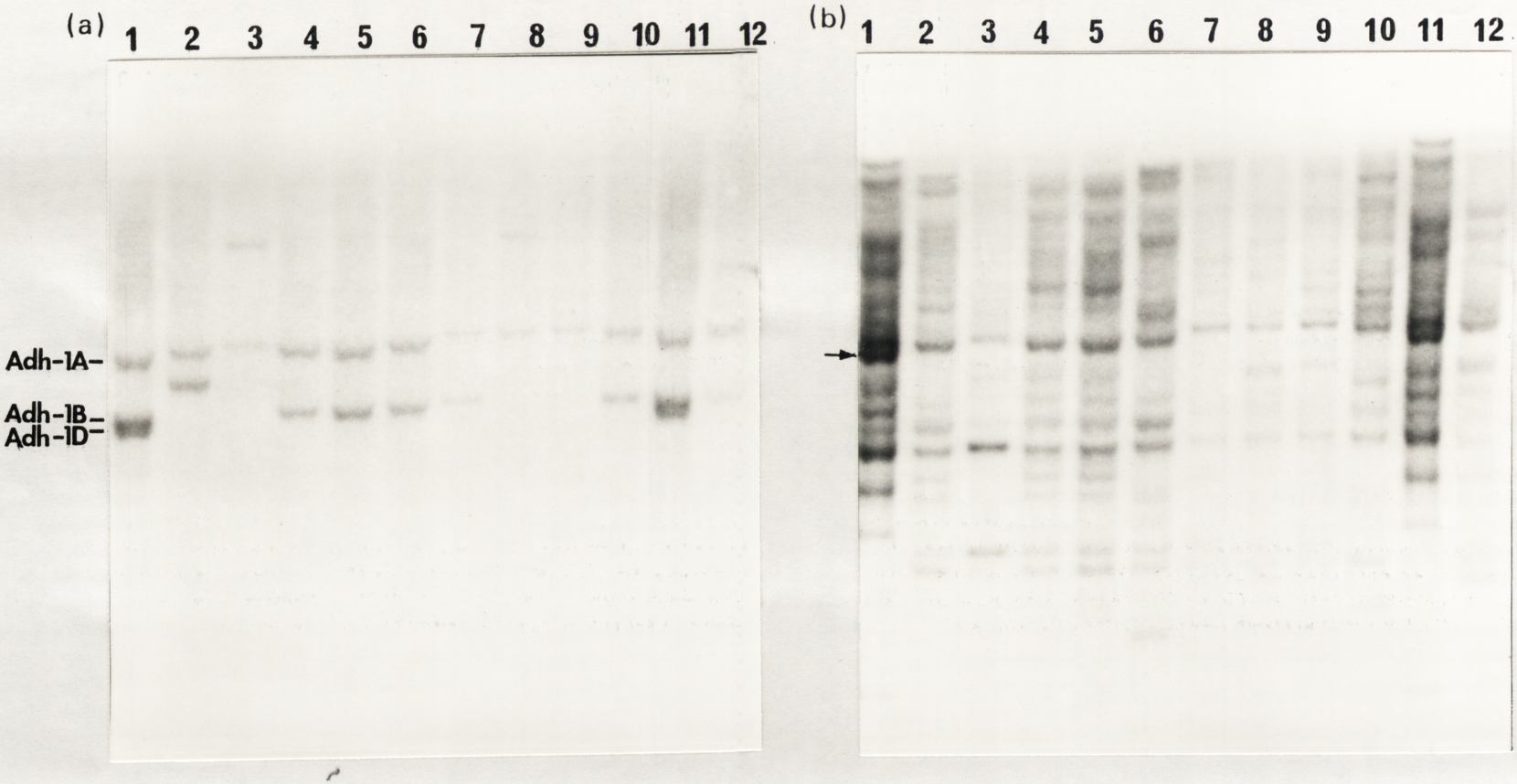


Figure 4.3
Southern analysis of *T.dicoccoides* accessions digested with HpaI/SphI and probed with; (a) p3NTR and (b) reprobed with pTLM. Lanes 1. and 11. are *T.aestivum* cv. Millewa; 2. *T.dicoccoides* accession AUS20303; 3. CPI99970; 4. AUS17480; 5. AUS17980; 6. G278; 7. TA54; 8. TA71; 9. TA84; 10. TA1058-1; 12. *T.aestivum* cv. Chinese Spring Nullisomic 1D-tetrasomic 1A. The wheat *Adh*-1A, 1B and 1D genes are labelled. Arrow indicates the position of the wheat *Adh*-1A gene which contains the pTLM insertion.

more polymorphic than the *Adh*-1A gene, with three of the nine accessions examined being polymorphic for either one or both of the Hpa1/BglII and Hpa1/Sph1 digests. In addition accession TA84 appeared to be null for the *Adh*-1B gene.

The five *T.dicoccoides* accessions which appeared to have similar *Adh*-1A and *Adh*-1B genes to those of hexaploid wheat were analyzed further by Southern hybridization. Using p3NTR as probe and different combinations of restriction enzymes: Hpa1/Stu1, Hpa1/Pvu1 and Hpa1/Pst1 failed to distinguish any of these tetraploids as being polymorphic at either *Adh* loci relative to hexaploid wheat.

Analysis of accessions of T.araraticum (AAGG, 2n=28)

Two accessions of *T.araraticum* were analyzed by Southern hybridization using p3NTR and pTLM as probes. Both *T.araraticum* accessions 35 and 37 when digested with either Hpa1/BglII (Fig 4.4) or Hpa1/Sph1 (Fig 4.5)and probed with oligolabelled p3NTR gave restriction fragments comparable in size to those expected for the *Adh*-1A gene of hexaploid wheat. The *Adh*-1A assignment was confirmed by rehybridization of the filter with oligolabelled pTLM as described above (data not shown).

In both *T.araraticum* accessions the *Adh*-1G gene was conserved for the Hpa1/Bgl11 restriction fragment but polymorphic for Hpa1/Sph1 fragment relative to the *Adh*-1B gene of hexaploid wheat. Both *T. araraticum* accessions appeared to have the same Hpa1/Sph1 polymorphism.

Analysis of putative A diploid progenitors.

A characteristic of the polyploid wheats is that they all have a common A genome possibly derived from *T.monococcum* L. (Kihara 1954). The A genome is considered to be largely unaltered from the

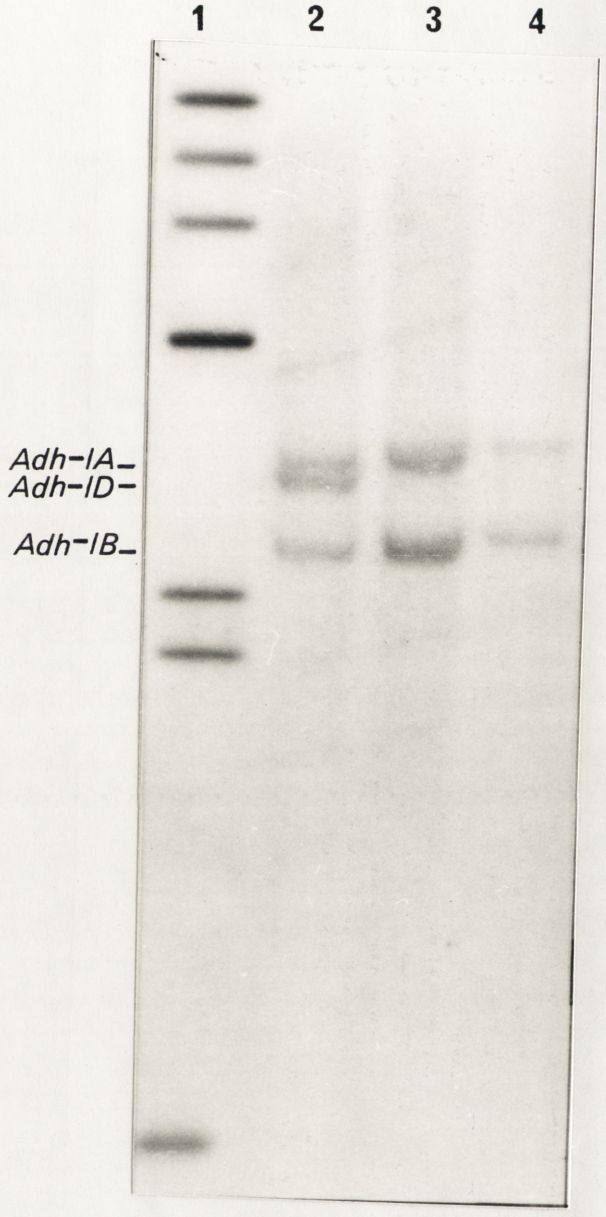


Figure 4.4
Southern analysis of *T.araraticum* accessions digested with HpaI/BglII and probed with p3NTR. 1. end labelled MW marker 23.1kb, 9.4kb, 6.6kb, 4.4kb, 2.3kb, 2.0kb, 0.56kb; 2. *T.aestivum* cv. Millewa; 3. *T.araraticum* accession 35; 4. *T.araraticum* accession 37. The wheat *Adh*-1A, 1B and 1D genes are labelled.

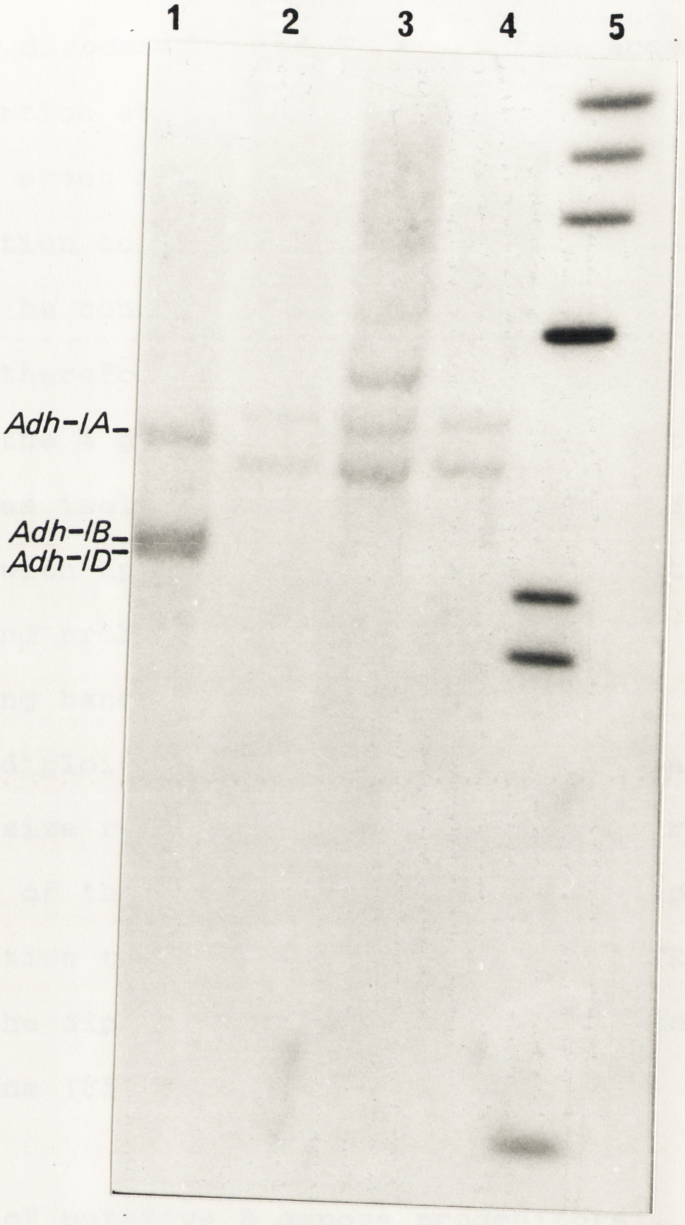


Figure 4.5
Southern analysis of *T.araraticum* accessions digested with HpaI/SphI and probed with p3NTR. 1. *T.aestivum* cv. Millewa; 2. *T.araraticum* accession 35; 3. *T.araraticum* accession 34; 4. *T.araraticum* accession 37; 5. end labelled MW marker 23.1kb, 9.4kb, 6.6kb, 4.4kb, 2.3kb, 2.0kb, 0.56kb. The wheat *Adh*-1A, 1B and 1D genes are labelled.

diploid progenitor (although chromosome 4A may have been derived from some currently unidentified species (Miller et al.,1981)) and is described as the pivotal genome (see introduction).

Molecular analysis of the tetraploid species indicates that all the *T.dicoccoides* and *T. araraticum* accessions tested have the pTLM insertion at the *Adh*-1A locus. This indicates that the insertion event may have occurred in the A diploid prior to hybridization to form the tetraploid species. The pTLM probe can therefore be considered as a molecular tag to trace the *Adh*-1A gene and therefore the particular A diploid accession involved in donating the A genome to the polyploid wheats.

DNA was isolated from five A diploid *Triticum* species, digested with Hpa1/BglII or Hpa1/Sph1, electrophoresed on 0.8% agarose and probed with oligolabelled p3NTR (Fig 4.6). A single hybridizing band was found for each diploid species. However, none of the A diploids tested showed conservation of restriction fragment size relative to the *Adh*-1A gene from hexaploid wheat. Reprobing of the filter with oligolabelled pTLM showed no hybridization to bands detected by the p3NTR probe indicating that none of the diploid A species tested had the pTLM insertion in the *Adh*-1A gene (Fig 4.6).

Analysis of putative B genome progenitors .

Tsunewaki and Ogihara (1983) proposed a diphyletic origin for the tetraploid wheats and suggested *T.longissimum* as the donor of both the B genome and cytoplasm to *T.dicoccoides*. However, the monophyletic and introgressive models (see introduction) for the evolution of wheat suggest that *T.speltoides* was the B/G genome donor either directly or indirectly to *T.dicoccoides*.

We have analyzed by RFLP mapping five *Triticum* species using

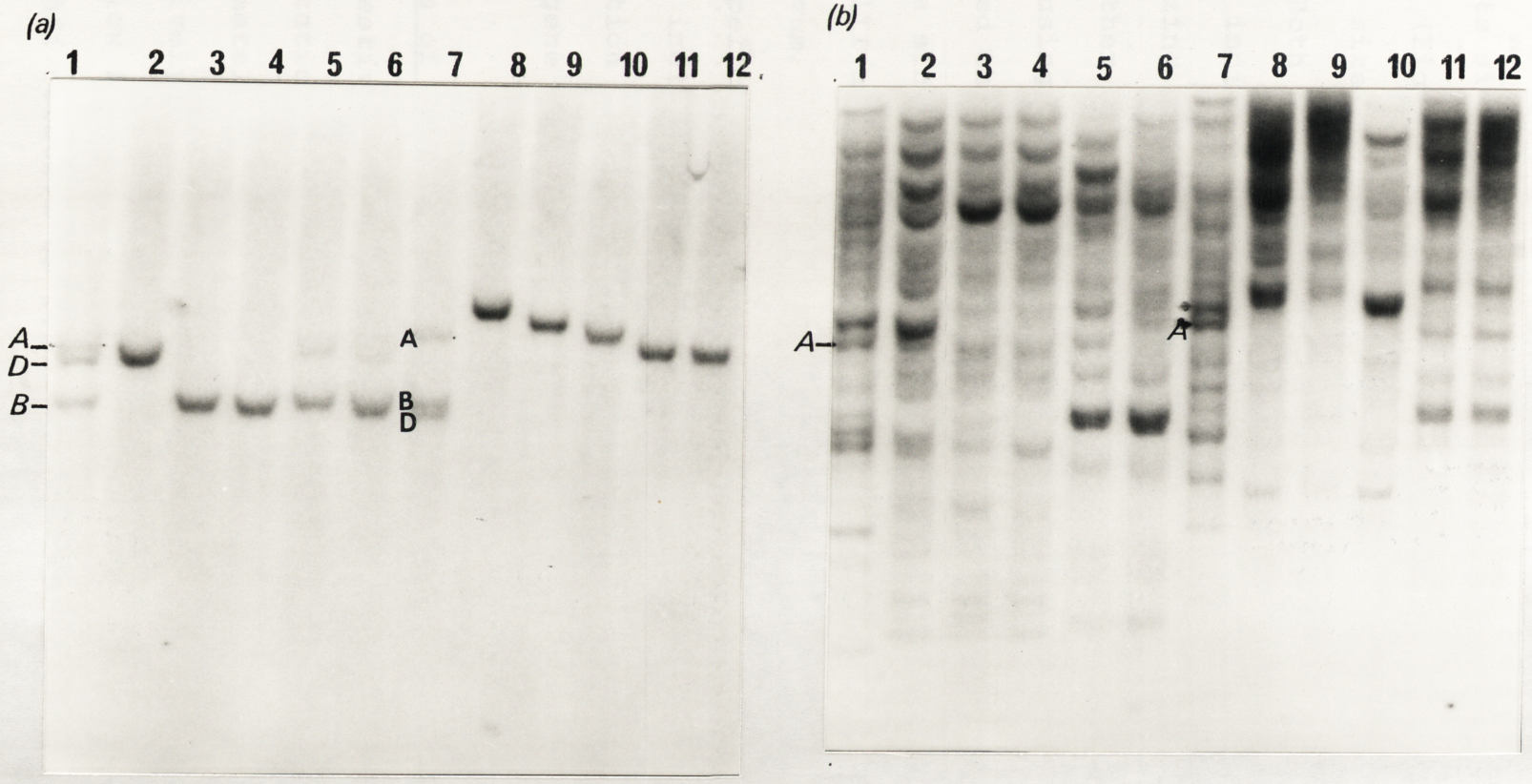


Figure 4.6
Southern analysis of A diploid *Triticum* species probed with (a) p3NTR or (b) reprobed with pTLM. Lanes 1-6 are digested with HpaI/BglII, lanes 7-12 are duplicates digested with HpaI/SphI. Lane 1. *T.aestivum* cv. Millewa; 2. *T.monococcum*; 3. *T.boeticum* Acc1; 4. *T.boeticum* Acc2; 5. *T.urartu* Acc1; 6. *T.urartu* Acc2. The wheat *Adh*-1A, 1B and 1D genes are labelled A, B, D.

p3NTR as probe to determine their evolutionary relationship with the B genome of hexaploid wheat. All three accessions of *T.speltoides* when digested with Hpa1/BglII gave restriction fragments similar in size to the *Adh*-1B gene in *T.aestivum* cv Millewa (Fig 4.7). *T.searsii* 76 and *T.sharonensis* 88 also gave similar sized Hpa1/BglII restriction fragments when probed with p3NTR. Both of these accessions appeared to have more than one *Adh* gene as indicated by the additional high molecular weight hybridizing bands.

Further Southern blot analysis of the putative B diploid donors using the restriction endonucleases Hpa1/Sph1 (Fig 4.7) indicated that only *T.speltoides* 71 fulfilled the criteria of having a single copy *Adh* gene with conservation of both the Hpa1/BglII and Hpa1/Sph1 sites relative to the *Adh*-1B gene from *T.aestivum*.

T.speltoides 71 was analyzed further using other restriction enzymes including Hpa1/Pst1, Hpa1/Stu1 and Hpa1/EcoR1. All these restriction sites were found to be polymorphic relative to the *Adh*-1B gene of *T.aestivum* (data not shown).

Analysis of D genome donor.

T.aestivum (AABBDD, 2n = 42) is thought to have arisen by hybridization of *T.dicoccoides* and *T.tauschii* (DD, 2n = 14) approximately 10,000 years ago (see review, Helbaek 1959). This is a relatively recent event from an evolutionary viewpoint and as such a low frequency of restriction fragment polymorphisms of *T.tauschii* relative to *T.aestivum* is expected.

Southern analysis of five accessions of *T.tauschii* using p3NTR as probe has identified three accessions, which by RFLP mapping appear to have an *Adh* gene identical to the *Adh*-1D gene from

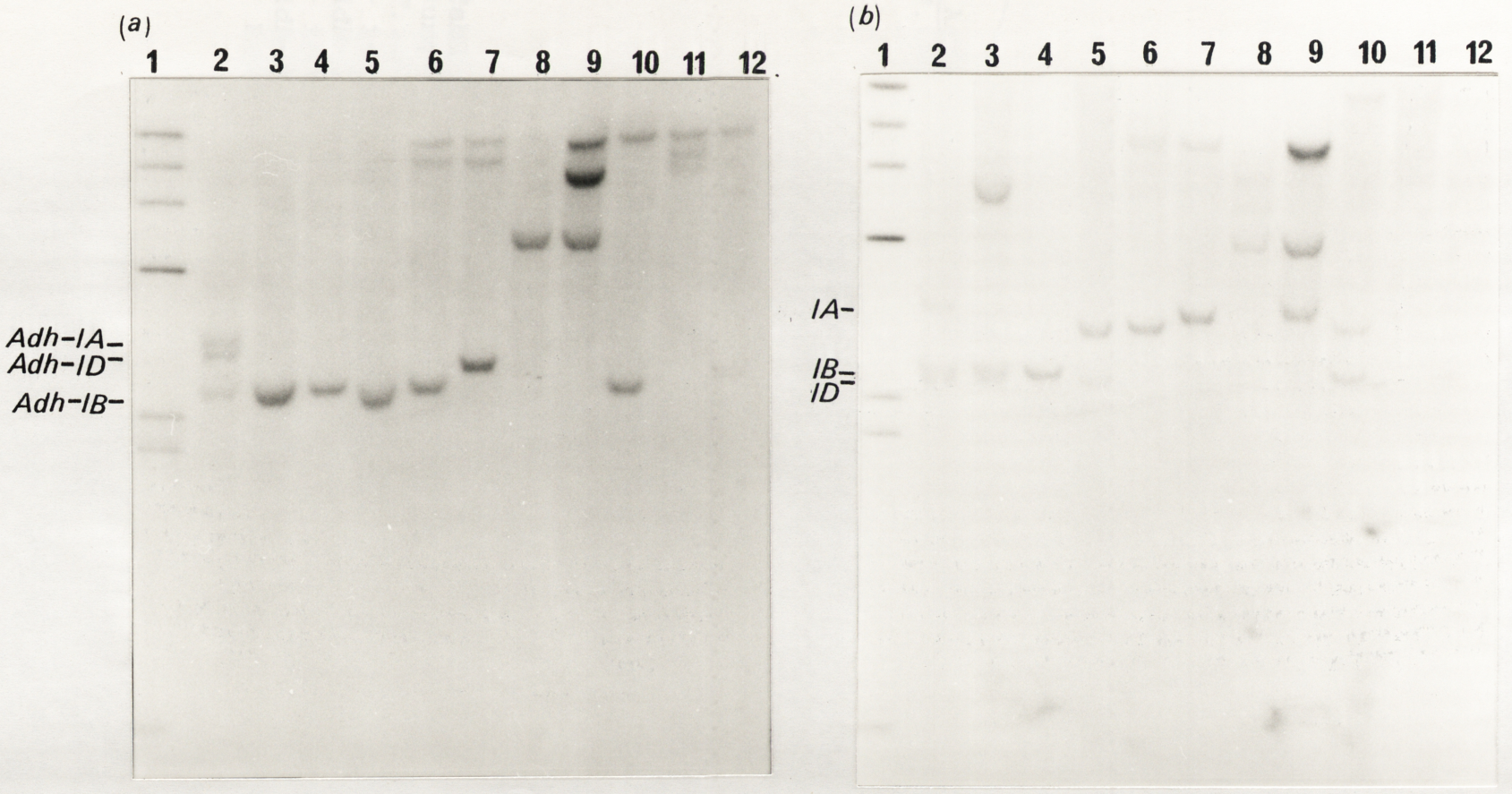


Figure 4.7
Southern analysis of putative B genome donors digested with; (a) HpaI/BglII or (b) HpaI/SphI and probed with p3NTR. 1. end labelled MW markers 23.1kb, 9.4kb, 6.6kb, 4.4kb, 2.3kb, 2.0kb, 0.56kb; 2. *T.aestivum* cv. Millewa; 3. *T.speltoides* 69; 4. *T.speltoides* 71; 5. *T.speltoides* 93; 6. *T.searsii* 76; 7. *T.searsii* 62; 8. *T.bicornis* 79; 9. *T.bicornis* 60; 10. *T.sharonensis* 83; 11. *T.sharonensis* 83; 12. *T.longissimum* 57. The wheat *Adh*-1A, 1B and 1D genes are labelled.

Accession \ Digest	Hpa1/ Bgl11	Hpa1/ Sph1	Hpa1/ EcoR1	Hpa1/ Pvu1	Hpa1/ Pst1	Hpa1/ Stu1
T.tauschii 80	+	?	+	+	+	?
" " 82	-	-	-	-	-	-
" " D1	+	+	+	+	+	+
" " D2	-	+	+	+	+	-
" " D3	+	+	+	+	+	+

Table 4.1
Summary of results of Southern analysis of five *T.tauschii* accessions probed with p3NTR.
+ indicates restriction fragment equivalent in size to *Adh*-1D of *T.aestivum*.
- indicates restriction fragment different in size to *Adh*-1D of *T.aestivum*.
? Result not clear.

T.aestivum. The results are summarized in Table 4.1.

4.4 Discussion

Numerous different techniques, all of which have inherent advantages and disadvantages, have been used to analyze the genomic constitution of the polyploid wheats and hence the evolutionary relationships between the wheat species. These techniques all, either directly or indirectly, measure the relative similarity of the DNA from different species. The fundamental difference between the techniques is the quantity of DNA compared. In the "analyzer" technique developed by Kihara (See review Lilienfeld 1951), which has been the most consistent and reliable method of determining evolutionary relationships, chromosome pairing at meiosis in artificial hybrids is taken as an indication of genomic similarity. This method has the advantage of comparing large amounts of DNA and can hence determine the relative homology between distantly related species. The disadvantage lies in the resolution of the technique and its inability to quantitatively differentiate between different cultivars or very closely related species. For such detailed evolutionary studies much smaller regions of DNA must be compared and in greater detail. The *Adh* genes of wheat represent just such a region.

The RFLP mapping of the *Adh* genes from nine different accessions of *T.dicoccoides* and two accessions of *T.araraticum* using p3NTR as probe indicates that both tetraploid species appear to have a common *Adh*-1A gene and that this gene is closely related to the *Adh*-1A gene of hexaploid wheat. This is in agreement with phylogenetic relationships which show that all the polyploid wheats share a common A genome.

All of the nine accessions of *T.dicoccoides* (AABB, 2n=28) examined had similar restriction fragment sizes to the *Adh*-1A gene of hexaploid wheat. Five of these accessions also had *Adh*-1B genes that were identical to the corresponding *Adh*-1B gene from hexaploid wheat. These particular accessions are likely to be directly related to the *T.dicoccoides* accessions involved in hybridization with *T.tauschii* (DD,2n=14) to form *T.aestivum* (AABBDD,2n=28). Of the remaining four accessions of *T.dicoccoides* three appeared to be polymorphic and one null for the *Adh*-1B gene. This apparent lower frequency of polymorphic change for the A genome locus relative to the B genome locus implies the existence of selection pressure to maintain an unaltered A genome. This is consistent with the pivotal genome hypothesis proposed by Zohary and Feldman,(1962).

Southern analysis using pTLM as probe, indicates that the *Adh*-1A gene from all the tetraploid accessions examined also contain the pTLM insertion. This places the insertion event of pTLM into the wheat *Adh*-1A gene sometime after the divergence of the wheat and barley lineages and prior to the formation of the polyploid wheat species. This means that the insertion of pTLM into the polyploid wheat *Adh*-1A gene may have taken place either in the A diploid species or during hybridization to form the tetraploid species assuming a monophyletic origin. If the insertion occurred very early in the divergence of the A genome then we might expect the *Adh* gene from each diploid A accession to have the pTLM insertion. Alternatively, if the insertion event occurred immediately prior to the formation of the tetraploid wheats then we might expect to be able to identify the particular A diploid accession involved in the hybridization to form the polyploid wheats. Of five diploid A species examined none of the *Adh* loci

detected by the p3NTR probe contained the pTLM insertion. This suggests that the insertion event may have occurred late in the A diploid species and that an exhaustive study of other A accessions would be necessary to identify *Adh* genes carrying the pTLM insertion. The possibility also exists that the particular A cultivar involved in formation of the polyploid wheats is no longer available.

Southern analysis of two accessions of *T.araraticum* (AAGG, 2n=28) using p3NTR as probe shows that the *Adh*-1G gene gives rise to a similar sized Hpa1/BglII restriction fragment as the *Adh*-1B gene from *T.aestivum* yet the Hpa1/Sph1 restriction fragment is not conserved between *T.araraticum* and *T.aestivum*. This partial conservation of restriction fragment sizes between the *Adh*-1B and *Adh*-1G genes is unlikely to be consistent with the diphyletic model for the evolution of the tetraploid wheats as this model predicts that the 1B and 1G genes were derived from two different species. The results are however consistent with both the monophyletic and introgressive models for the origin of the tetraploid wheats as both of these models predict the 1B and 1G *Adh* genes to be derived from a common ancestor. From the limited amount of data it would seem that the *Adh*-1G gene of *T. araraticum* is more polymorphic than the *Adh*-1B gene of *T. dicoccoides*. This would be consistent with the introgressive model which predicts that *T. araraticum* predates *T. dicoccoides*.

RFLP analysis of putative B/G genome donors using p3NTR as probe indicates that the *Adh* gene from *T.speltoides* may be more closely related than the corresponding genes from either *T.longissimum*, *T.searsii* or *T.sharonensis* to the *Adh*-1B gene from the polyploid wheats.

Hybridization of *T.dicoccoides* with *T.tauschii* almost

certainly gave rise to *T.aestivum* (AABBDD,2n=42). RFLP analysis using p3NTR as probe of four accessions of *T.tauschii* indicates that the *Adh* gene from *T.tauschii* is closely related to the *Adh*-1D gene from hexaploid wheat although certain accessions appear to have a more highly conserved *Adh* gene relative to the *Adh*-1D gene from hexaploid wheat than others.

In conclusion, analysis of the *Adh* loci on chromosome 1 using the p3NTR and pTLM probes has identified particular accessions of *T.dicoccoides* and *T.tauschii* that appear to be closely related to *T.aestivum*.

RFLP analysis of the *Adh*-1B gene from *T.dicoccoides* and the *Adh*-1G gene from *T.araraticum* suggests that these genes are somewhat related and are probably derived from a common ancestor thereby providing support for the monophyletic/introgressive models for the origins of tetraploid wheats.

Analysis of the *Adh* genes from putative B genome donors has not conclusively identified the source of the B genome of the polyploid wheats but provides supporting evidence that *T.speltoides* may have been involved in the hybridization.

Finally we have been able to show at a molecular level that the *Adh*-1A genes from the polyploid wheats are all derived from a common A diploid accession. Although we have not been able to identify this particular A diploid accession, it would be expected to have an *Adh* gene carrying the pTLM insertion and this criteria will be used for screening further A diploid accessions as they become available.

Acknowledgements.

This research was supported by the Rockefeller Foundation.

4.5 References

Dhaliwal, R.S. and Johnson, B.L. (1978). Diploidization and chromosome pairing affinities in the tetraploid wheats and their putative amphiploid progenitors. Theor. Appl. Genet. 61:117-123.

Dvorjak, J. and Appels, R. (1982). Chromosome and nucleotide sequence differentiation in genomes of polyploid Triticum species. Theor. Appl. Genet. 63:349-360.

Feldman, M. (1965). Further evidence for natural hybridization between tetraploid species of Aegilops section Pleionathera. Evolution 19:162-174.

Gill, B.S. and Chen, P.D. (1987). Cytogenetic evidence on introgressive hybridization during the evolution of the polyploid wheat species. Proc. Natl. Acad. Sci. USA

Helbaek, H. (1959). Domestication of food plants in the old world. Science 130, 365-372.

Johnson, B.L. (1975). Identification of an apparent B genome donor of wheat. Can. J. Genet. Cytol. 17:21-39.

Kihara, H. (1954). Considerations on the evolution and distributions of Aegilops species based on the analyzer method. Cytologia 19:336-357.

Lilienfeld, F.A. (1951). H. Kihara: Genome-analysis in *Triticum* and *Aegilops*. Concluding review. Cytologia 16, 101-123.

Maan, S.S. (1973). Cytoplasmic and cytogenetic relationships among tetraploid Triticum species. Euphytica 22:287-300.

Miller, T.E., Shepherd, K.W. and Riley, R. (1981).The relationship of chromosome 4A of diploid wheat to that of hexaploid wheat: a clarification of an earlier study. Cereal Res. Comm. 9, 327-329.

Mitchell, L.E., Dennis, E.S. and Peacock, W.J. (1987a). Molecular analysis of an alcohol dehydrogenase gene from chromosome 1 of wheat (in preparation).

Mitchell, L.E., Dennis, E.S. and Peacock, W.J. (1987b). Evidence for transposable elements in wheat.(in preparation).

McFadden, E.S. and Sears, E.R. (1944). The artificial synthesis of T. spelta Rec. Genet. Soc. Am. 13:26-27.

Sarkar, P. and Stebbins, G.L. (1956). Morphological evidence concerning the origin of the B genome in wheat. Amer. J. Bot. 43:297-304.

Tsunewaki, K. and Ogihara, Y. (1983). The molecular basis of genetic diversity among cytoplasms of Triticum and Aegilops species. II. Genetics 104:155-171.

Wagenaar, E.B. (1961). Studies on the genome of Triticum timopheevi Zhuk. I. Evidence for genetic control of meiotic irregularities in tetraploid hybrids. Can. J. Genet. Cytol. 3:47-60.

Zohary, D. and Feldman, M. (1962). Hybridization between amphidiploids and evolution of polyploids in the wheat (Aegilops-Triticum) group. Evol. 16:44-61

CHAPTER 5

EPILOGUE

Molecular and cytogenetic analysis indicates that we have cloned an alcohol dehydrogenase gene from chromosome 1A of wheat. However, preliminary S1 mapping experiments have failed to detect a transcript from the *Adh*-1A locus. As the gene appears to have all the signals required for its expression the absence of a transcript may be a consequence of the TLM insertion in intron 3. Insertions in non-coding regions resulting in inactivation or reduced expression of a locus have been observed in other species (Barker et al.,1984; Schwarz-Sommer et al.,1985). If the *Adh*-1A gene has been inactivated as a result of the TLM insertion it must, from an evolutionary point of view, have been a relatively recent event as no detectable mutations to non-functional sequences have occurred in the coding region or splice junctions of the gene. Although the *Adh*-1A locus may not be active, northern analysis has revealed that at least one of the group 1 *Adh* loci is expressed and induced under anaerobic conditions. Taking into account these two considerations, (a)that the *Adh*-1A gene may not be expressed but retains a functional nucleotide sequence and (b)that closely related *Adh* genes are anaerobically induced, it is valid to analyze the promoter region of the *Adh*-1A gene for the presence of anaerobic regulatory elements (ARE). The wheat *Adh*-1A gene shows extensive homology to the ARE of maize *Adh1* although the relative orientation of the two sub-regions is reversed (Fig 5.1). Anaerobically induced genes from other species also have homology to the central portion of the sub-regions of the maize *Adh1* ARE (Fig 5.1) (See review Walker et al.,1987). The observation that both the *Arabidopsis Adh* and pea *Adh1* sequences

Fig. 5.1

Organism	Gene		Comments
Maize	Adh1-1S (Dennis et al., 1984)	-130 -120 -110 -100 CTGCAGCCCCGGTTTCGCAAGCCGCGCCGTGGTTTGCTTGC	Identified as functionally important by linker scanner analysis (Walker et al., 1987) and by DMS protection (Ferl & Nick 1987)
	Ferl & Nick (1987)#	-185 -195 GGACGTGGTTTTCGCTCGT	Identified by DMS protection (Ferl & Nick, 1987)
	Adh2-N (Dennis et al., 1985)	-140 -130 CTCCCTGGTTTCTAACCGCG	Sequence homology to Adh1-1S
	Sucrose synthase (Werr et al., 1985)	+431 +441 CGTGGTTTGCTTG -170 -160 AATTCTGGTTTTGAGTAA	Sequence homology to ARE of Adh1-1S
	Aldolase (Dennis et al., 1987)	-70 -60 TTTCGCTGGTTTCTTTCCCCTT	Lies in functionally important region (-60 to -110)
	#	-180 -190 CCTTTCTGGTTTTTTTGTTTT	Sequence homology, inverted orientation
Wheat	Adh-1A (Mitchell et al., 1987)	-525 -490 CAGTTGGTTTGCTCGGTCGCCCCGGTTTCGGCCAC	Sequence homology to ARE of Adh1-1S. Numbering is from translation start not transcription start.
Pea	Adh (Llewellyn et al., 1987)#	-100 -110 CCGCTTTTGGTTTGTTT	Lies in functionally important region (-51 to -129)
Arabidopsis	Adh (Chang & Meyerowitz, 1986)#	-150 -160 GTATTTGGTTTTGCCTT	Sequence homology
CONSENSUS SEQUENCE		TC TC CGTGGTTTGT GT CG	

#Occurs in inverse orientation

Fig. 5.1 Putative anaerobic regulatory elements in the upstream region of anaerobically induced genes.

have homology to the reverse complement of this central region suggests that the element may work in either orientation similar to many enhancer sequences.

Both the INFERRED amino acid composition (Fig 5.2) and the inferred primary amino acid sequence of the wheat *Adh*-1A gene are very similar to other plant *Adh* genes (TABLE 5.3). There is also extensive conservation of both chemically and structurally important amino acid residues (TABLE 5.3) (see introduction). This conservation of primary sequence implies a similar tertiary structure to other ADH enzymes although it must be remembered that minor changes in amino acid sequence can result in dramatic changes in enzyme kinetics (Murali and Creaser 1986).

The nucleotide coding sequence of the wheat *Adh*-1A gene is also similar to that of other plant *Adh* genes and displays the typical codon usage pattern (Fig 5.4a) as well as the (G,non G,N) pattern described earlier (Fig 5.4b). It is interesting to note that the eight introns which interrupt the coding sequence of the wheat *Adh*-1A gene are in identical positions to eight of the nine introns in maize *Adh*. Comparison of the maize and *Arabidopsis Adh* sequences established that intron loss occurs in the *Arabidopsis Adh* gene (see discussion chapter 2). It seems likely that the wheat *Adh*-1A gene has lost intron 9 resulting in the in-frame fusions of exons 9 and 10. It appears that intron loss does not occur with intron 1 as intron 1 of maize *Adh1* appears to serve a functional role in high level gene expression (for review Freeling,1984) and is conserved in all plant species so far examined. The suggestion of a possible functional role for intron 1 of maize *Adh1* was based on the observation that three mutants with reduced ADH1 activity were found to have insertions in intron 1 (Freeling,1984). Callis (1986) has shown that transient

TABLE 5.2

THE AMINO ACID COMPOSITION OF THE POLYPEPTIDE DERIVED FROM THE WHEAT *Adh*-1A GENE. (Molecular Weight = 41119.55)

Amino Acid		No.	Weight Percent	Moles percent
Ala	A	32	5.53	8.42
Cys	C	13	3.26	3.42
Asp	D	16	4.48	4.21
Glu	E	30	9.42	7.89
Phe	F	18	6.44	4.74
Gly	G	37	5.13	9.74
His	H	12	4.00	3.16
Ile	I	20	5.50	5.26
Lys	K	25	7.79	6.58
Leu	L	27	7.43	7.11
Met	M	13	4.15	3.42
Asn	N	12	3.33	3.16
Pro	P	19	4.49	5.00
Gln	Q	6	1.87	1.58
Arg	R	13	4.94	3.42
Ser	S	13	2.75	3.42
Thr	T	23	5.66	6.05
Val	V	40	9.64	10.53
Trp	W	3	1.36	0.79
Tyr	Y	7	2.78	1.84

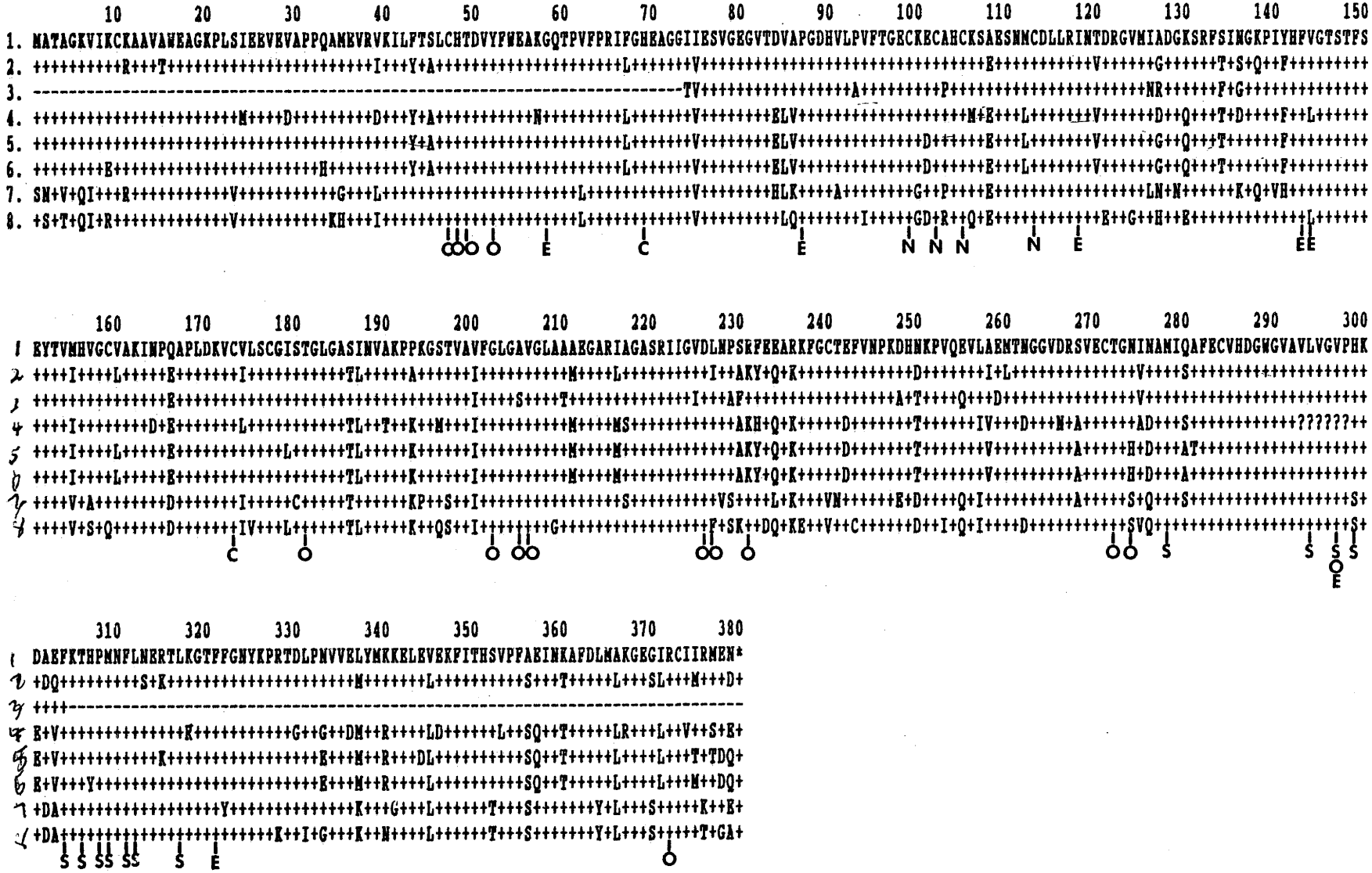


TABLE 5.3

Derived amino acid sequence of the maize *Adh1* gene (top line, 1) aligned with maize *Adh2* (2), barley *Adh1* (3), barley *Adh2* (3), barley *Adh3* (4), wheat *Adh*-1A (5), pea *Adh1* (7) and Arabidopsis *Adh* (8). Conserved residues are indicated by a (+). (-) indicates sequence data not available and (?) indicates amino acid deletions. The first residue (methionine) of the pea ADH polypeptide has been omitted to allow maximum sequence alignment. Residues presumed to be involved in the catalytic cleft (E), cofactor binding (o), catalytic (C) and non-catalytic (N) zinc binding ligands and subunit interactions (S) are indicated.

TABLE 5.4

CODON (a) AND BASE (b) USAGE FOR THE WHEAT *Adh*-1A GENE.

(a)

TTT(Phe) = 2	TCT(Ser) = 1	TAT(Tyr) = 0	TGT(Cys) = 5
TTC(Phe) = 16	TCC(Ser) = 2	TAC(Tyr) = 7	TGC(Cys) = 8
TTA(Leu) = 1	TCA(Ser) = 3	TAA(***) = 0	TGA(***) = 0
TTG(Leu) = 1	TCG(Ser) = 2	TAG(***) = 1	TGG(Trp) = 3
CTT(Leu) = 7	CCT(Pro) = 3	CAT(His) = 4	CGT(Arg) = 2
CTC(Leu) = 12	CCC(Pro) = 4	CAC(His) = 8	CGC(Arg) = 4
CTA(Leu) = 0	CCA(Pro) = 4	CAA(Gln) = 3	CGA(Arg) = 0
CTG(Leu) = 6	CCG(Pro) = 8	CAG(Gln) = 3	CGG(Arg) = 0
ATT(Ile) = 3	ACT(Thr) = 5	AAT(Asn) = 3	AGT(Ser) = 0
ATC(Ile) = 17	ACC(Thr) = 13	AAC(Asn) = 9	AGC(Ser) = 5
ATA(Ile) = 0	ACA(Thr) = 1	AAA(Lys) = 7	AGA(Arg) = 0
ATG(Met) = 13	ACG(Thr) = 4	AAG(Lys) = 18	AGG(Arg) = 7
GTT(Val) = 5	GCT(Ala) = 6	GAT(Asp) = 6	GGT(Gly) = 8
GTC(Val) = 15	GCC(Ala) = 11	GAC(Asp) = 10	GGC(Gly) = 13
GTA(Val) = 1	GCA(Ala) = 7	GAA(Glu) = 7	GGA(Gly) = 7
GTG(Val) = 19	GCG(Ala) = 8	GAG(Glu) = 23	GGG(Gly) = 9

(b)

BASE	POSITION		
	1	2	3
T	13.7	31.1	15.8
C	17.9	21.6	40.5
A	27.6	28.7	10.8
G	40.8	18.7	32.9

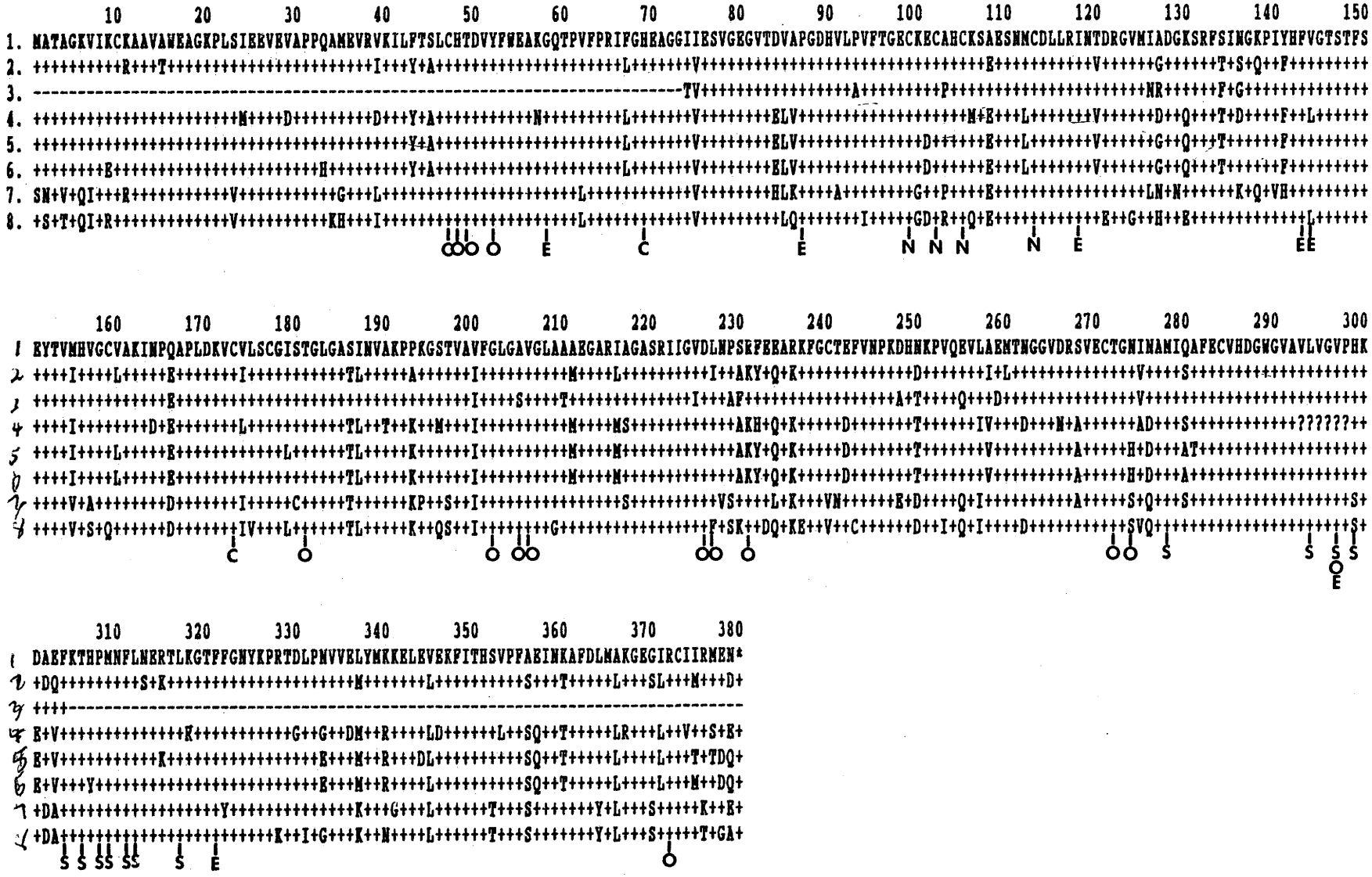


TABLE 5.3

Derived amino acid sequence of the maize *Adh1* gene (top line, 1) aligned with maize *Adh2* (2), barley *Adh1* (3), barley *Adh2* (3), barley *Adh3* (4), wheat *Adh*-1A (5), pea *Adh1* (7) and Arabidopsis *Adh* (8). Conserved residues are indicated by a (+). (-) indicates sequence data not available and (?) indicates amino acid deletions. The first residue (methionine) of the pea ADH polypeptide has been omitted to allow maximum sequence alignment. Residues presumed to be involved in the catalytic cleft (E), cofactor binding (o), catalytic (C) and non-catalytic (N) zinc binding ligands and subunit interactions (S) are indicated.

TABLE 5.4

CODON (a) AND BASE (b) USAGE FOR THE WHEAT *Adh*-1A GENE.

(a)

TTT(Phe) = 2	TCT(Ser) = 1	TAT(Tyr) = 0	TGT(Cys) = 5
TTC(Phe) = 16	TCC(Ser) = 2	TAC(Tyr) = 7	TGC(Cys) = 8
TTA(Leu) = 1	TCA(Ser) = 3	TAA(***) = 0	TGA(***) = 0
TTG(Leu) = 1	TCG(Ser) = 2	TAG(***) = 1	TGG(Trp) = 3
CTT(Leu) = 7	CCT(Pro) = 3	CAT(His) = 4	CGT(Arg) = 2
CTC(Leu) = 12	CCC(Pro) = 4	CAC(His) = 8	CGC(Arg) = 4
CTA(Leu) = 0	CCA(Pro) = 4	CAA(Gln) = 3	CGA(Arg) = 0
CTG(Leu) = 6	CCG(Pro) = 8	CAG(Gln) = 3	CGG(Arg) = 0
ATT(Ile) = 3	ACT(Thr) = 5	AAT(Asn) = 3	AGT(Ser) = 0
ATC(Ile) = 17	ACC(Thr) = 13	AAC(Asn) = 9	AGC(Ser) = 5
ATA(Ile) = 0	ACA(Thr) = 1	AAA(Lys) = 7	AGA(Arg) = 0
ATG(Met) = 13	ACG(Thr) = 4	AAG(Lys) = 18	AGG(Arg) = 7
GTT(Val) = 5	GCT(Ala) = 6	GAT(Asp) = 6	GGT(Gly) = 8
GTC(Val) = 15	GCC(Ala) = 11	GAC(Asp) = 10	GGC(Gly) = 13
GTA(Val) = 1	GCA(Ala) = 7	GAA(Glu) = 7	GGA(Gly) = 7
GTG(Val) = 19	GCG(Ala) = 8	GAG(Glu) = 23	GGG(Gly) = 9

(b)

BASE	POSITION		
	1	2	3
T	13.7	31.1	15.8
C	17.9	21.6	40.5
A	27.6	28.7	10.8
G	40.8	18.7	32.9

expression of the maize *Adh*-1S all ele in maize protoplasts is 50-100 fold higher than when the coding region of the *Adh*-1S allele is replaced by its cDNA equivalent. Callis (1986) has also shown that addition of the first intron to the equivalent maize *Adh*-1S cDNA allele was sufficient to restore ADH-1S activity nearly equivalent to that of the genomic *Adh*-1S allele. Introns 8 and 9 of maize were also found to increase the level of *Adh* expression but not to the same extent as intron 1. It is unlikely that the absence of intron 9 in the wheat *Adh*-1A (and presumably *Adh*-1B and *Adh*-1D genes) causes the low level of expression of the enzyme and accounts for the inability to detect enzymatic activity in isozyme studies. Marked effects of introns on expression and regulation have also been observed for the β globin genes (Hamer and Leder 1979) and for the immunoglobulin genes (Grosschedl and Baltimore 1985).

We were surprised to have identified previously unknown *Adh* loci on the group 1 chromosomes of wheat considering the very extensive ADH isozyme studies that have been done by Hart and Jaaska in that species (see chapter 2). The apparent absence of isozymes corresponding to expression from the group 1 *Adh* loci may be a result of the assay procedure used or alternatively may be related to the structure of the individual loci as discussed above.

The *Adh* isozyme assay relies on the ability of enzymes to catalyze the dehydrogenase reaction (ie. ethanol to acetaldehyde). In this assay, following electrophoresis, ethanol and NAD+ are used as substrates and the acetaldehyde produced in the presence of ADH reacts with Nitro-Blue Tetrazolium to form an insoluble blue precipitate. There are a number of reasons why isozyme studies may have failed to detect the chromosome 1 *Adh* loci. These

include comigration with other ADH isozymes, high Michaelis constant for ethanol resulting in poor staining for ADH activity in isozyme assays, altered substrate specificity, low level of expression, and possibly altered catalytic direction. Since alcohol dehydrogenase may also play a role in the degradation of cytotoxic alcohol upon restoration of aerobic conditions it might be expected that ADH enzymes exist with activities favouring the catalysis of opposing reactions ie. with Km's which favour either ethanol or acetaldehyde as substrate. However, the standard isozyme assay is more effective in detecting those enzymes involved in oxidation of ethanol to acetaldehyde, the reverse reaction to that occurring under anaerobic stress. Perhaps the ADH enzymes encoded by the chromosome 1 loci catalyze the conversion of acetaldehyde to ethanol. A similar type of catalysis of opposing reactions has been observed for the cytosolic yeast *Adh1* and *Adh2* isozymes (Bennetzen and Hall.,1982).

The use of subclones of the wheat *Adh*-1A and TLM insertion as probes for analyzing the evolution of wheat demonstrates the ability of molecular probes to resolve minor genetic differences between cultivars. Although the practical application of this technique is limited it nevertheless provides us with a means of determining some of the molecular processes that occur during evolution. Transposition of controlling elements is one mechanism by which a whole range of genetic rearrangements (see introduction) can occur. Thus transposable elements provide one means by which genome modification can occur generating the sequence diversity needed in evolution. The presence of the TLM insertion in the wheat *Adh*-1A gene and its absence in the closely related barley *Adh3* gene indicates that insertion elements have played a role in the evolution of wheat. As discussed in chapter 3

it is unlikely that the TLM element or any cross hybridizing sequences are active as autonomous transposable elements. Although this does not preclude the possibility of inducing the transposition of the TLM element in wheat and its use as a molecular tag for isolation of genes it seems more likely that the eventual transformation of wheat with well characterized transposable elements from other species such as *Ac* from maize will allow such tagging experiments to be achieved in wheat. Interspecific activity of *Ac* has already been reported in tobacco (Baker et al.,1986).

5.1 References

Baker, B., Schell, J., Lorz, H. and Fedoroff, N. (1986). Transposition of the maize controlling element "*Activator*" in tobacco. Proc.Natl.Acad. USA 83, 4833-4848.

Barker, R.F., Thompson, D.V., Talbot, D.R., Swanson, J. and Bennetzen, J.L. (1984). Nucleotide sequence of the maize transposable element Mul. Nucl. Acids. Res. 12, 5955-5967.

Bennetzen , J.L., Hall, B.D. (1982). The primary structure of the Saccharomyces cerevisiae gene for alcohol dehydrogenase 1. J.. Biol. Chem. 257, 3018-3025.

Callis, J. (1986). PhD thesis Stanford University. Studies on maize alcohol dehydrogenase.

Freeling, M. (1984). Plant transposable elements and insertion sequences. Ann. Rev. Plant Physiol. 35, 277-298.

Grosschedl, R., Baltimore, D. (1985). Cell type specificity of immunoglobulin gene expression is regulated by at least three DNA sequence elements. Cell 41, 885-897.

Hamer, D.H. and Leder, P. (1979). Splicing and the formation of stable mRNA. Cell 18, 1299-1302.

Murali, C. and Creaser,E.H. (1986). Protein engineering of alcohol dehydrogenase-1. Effects of two amino acid changes in the active site of yeast ADH-1. Prot. Eng. 1, 55-57.

Schwarz-Sommer, Z., Gierl, A., Cuypers, H., Peterson, P.A. and Saedler, H. (1985). Plant transposable elements generate the sequence diversity needed in evolution. EMBO J. 4, 591-597.

Walker J.C., Howard E.A., Dennis E.S., Peacock W.J. (1987). DNA sequences required for anaerobic expression of the maize alcohol dehydrogenase-1 gene. Proc. Natl. Acad. Sci. USA (in press).

ADDENDUM

1. NUCLEOTIDE SEQUENCE ANALYSIS AND NOMENCLATURE

It is not possible, because of the differing size and sequence divergence of introns and untranslated regions to produce sequence alignments as has been done for the protein coding regions of the *Adh* genes (Fig 5.3). A thorough analysis of the intron and untranslated regions of plant *Adh* genes therefore requires a detailed dot matrix analysis of all combinations of *Adh* genes. This analysis has been done and because of the visual nature of the data require hundreds of pages of printing and therefore was not included in this thesis. As stated in the text, weak homology between *Adh* genes was found in and around intron splice sites and, apart from the particular case of the wheat *Adh*-1A and barley *Adh* 3/*Adh*2 genes, no other signifigant homology was observed in the introns and untranslated regions. This is not to say that controlling regions do not exist in these genes but rather that in the absence of functional analysis it is impossible to asses the relevance of weak homologies in thes non-coding regions. This is emphasised for example by the fact that multiple branch sites for lariet structures involved in intron excission were identified in every wheat *Adh*-1A intron. (The putative wheat *Adh*-1A anaerobi regulatory element (Fig 5.1) was identified on the basis of functional studies which had been done for maize *Adh*1.)

Although it is possible to condense the dot matrix information by an averaging procedure this is only vali for genes which are very closely related as shown in Fig 3.1a for the wheat *Adh*-1A and barley *Adh* 3 genes which are > 95% related at the nucleotide level. If this averaging procedure were used in the analysis of intron an untranslated regions all information would be lost as putative control sequences are likely to be only weakly related.

It is possible however to align the nucleotide sequence of the coding regions of the plant *Adh* genes . This data is summarized in Table 1A From this analysis it can be seen that the pea and *Arabidopsis Adh* genes are more closely related to each other than to the monocot genes as well as being more closely related to maize *Adh* 1 than maize *Adh* 2. Similarly, barley *Adh*1 is more closely related to maize *Adh*1 while barley *Adh* 2 and *Adh*3 ar more like maize *Adh*2. This implies that after divergence of wheat and barley the barley *Adh* 2 gene was duplicated to produce barley *Adh* 3 which was subsequetly translocated to barley chromosome 5. This thesis describes the cloning and characterization of the wheat *Adh* gene equivalent to barley *Adh* 3. The extensive homology between the two genes implies that the duplication event in barley probably occurred shortly before the divergence of the wheat and barley lineages. As a result of this the homologous group 1 *Adh* gene loci which throughout the text has been designated *Adh*-1A, *Adh*-1B, *Adh*-1D, should be assigned the new and more consistent designations *Adh*-A2, *Adh*-B2 and *Adh*-D2 ie. considered to compose the *Adh*-2 set of gene loci.

TABLE 1A

Percent nucleotide (and amino acid) sequence homologies among plant *Adh* genes. Square parentheses indicate incomplete sequence information

	Maize *Adh*1	Maize *Adh*2	Barley *Adh*1	Barley *Adh*2	Barley *Adh*3	Wheat *Adh*-1A	Pea *Adh*	*Arabidopsis Adh*
Maize *Adh*1	100	82.2(87.6)	[84.4(91.3)]	79.1(81.8)	81.3(85.8)	82.4(86.3)	74.4(83.6)	73.2(81.3)
Maize *Adh*2		100	[77.8(83.5)]	85.3(85.5)	85.7(90.0)	85.8(90.2)	71.1(81.8)	72.0(79.4)
Barley *Adh*1			100	[72.3(75.3)]	[77.1(80.1)	[77.2(81.0)]	[73.8(79.2)]	[71.9(76.6)]
Barley *Adh*2				100	88.6(90.1)	88.5(90.3)	69.6(76.7)	71.0(75.6)
Barley *Adh*3					100	95.6(97.9)	72.1(78.6)	71.8(78.1)
Wheat *Adh*-1A						100	73.1(79.2)	71.8(78.4)
Pea *Adh*							100	76.5(85.0)
Arabidopsis Adh								100

2. INSERTION ELEMENT STRUCTURE

It should be noted that the 8 bp genomic duplication produced by the TLM insertion, as described in Fig 3.2, is the maximum possible direct duplication and that both 4 and 6 bp duplications are also possible.

```
    C-G                 C-G
    C T                 C T
    G-C                 G-C
    G-C                 G-C
    A-T                 A-T
    G-C                 G-C
    G-C                 G-C
    C-G                 C-G
    G-C                 G-C
    G-C                 G-C
    T T                 T T
    C C                 C C
    A-T                 A-T
    C-G                 C-G
    G-C                 G-C
    A-T
ATGA  ATGA          TATGAA  TATGAA
```

3. EVOLUTION OF THE INSERTION ELEMENT

Southern analysis of *T. aestivum cv.* Millewa probed with pTLM (Fig 3.3) indicates that under stringent hybridization conditions (50% formamide, 42°C) there are approximatly 10-15 related copies of the insertion element in the hexaploid genome. A combination of restriction enzymes which were known to cut the cloned TLM sequence internally were chosen for the Southern analysis and a wide range of restriction fragment sizes and hybridization intensities were observed. This range of hybridization intensities may be due to differing degrees of homology with the probe or alternatively to several hybridizing fragments having the same restriction fragment length. This situation is similar to that found for Ds elements in maize and probably implies that many (if not all) of the cross hybridizing sequences are inactive transposons which have diverged in sequence.

Southern analysis of A, B and D diploid wheats using pTLM as probe indicates that the insertion is present in all three genomic backgrounds and implies that it originated from a more ancient species. If this is the case then it would not be surprising that many of the TLM related sequences are inactive. No discrete hybridizing fragments were found in maize. However, this evidence is not sufficient to say that sequences related to TLM are not present in this species.

4. PUTATIVE BB AND DD GENOME DONORS

In chapter 4 the p3NTR probe has been used to detect RFLPs in wheat and its relatives with the view to establishing the course of evolution of the polyploids. It should be noted that this type of analysis is by no means rigourous and that RFLPs can arise from single base changes in many different locations in the adjacent DNA. As a result of this the analysis of putative B (p.65) and D (p.66) genome donors and the discussion on the diphyletic monophyletic and introgressive models for the origin of tetraploid wheat are inconclusive. However, if a survey of hexaploid wheat cultivars revealed a characteristic pattern for the species with little or no variation between cultivars it might be possible to make inferences about possible progenitors if in a wide survey of candidate species only one species matched the appropriate pattern in wheat for that particular genome.

5. CORRECTIONS

In the introduction to chapter 2 it should be noted that the cloning of the wheat *Adh* gene will also allow comparison of the molecular structure and organization of equivalent genes in polyploid and diploid wheat species where the selection pressures are presumed to be very different

On page 12 line 16 it is stated that " Replicative transposition as found in procaryotes is not known to occur in maize". However, Bennetzen, J.L. et al (1987) Mol. Gen. Genet. 208, 57-62 have recently reported evidence that Mu1 of maize probably transposes by a replicative mechanism.

It should be noted that there is no wild form of hexaploid wheat contrary to what is implied on Page 20 lines 11-12.

6. MATERIALS AND METHODS

Genomic Southerns

DNA samples were digested with a 2 fold excess of restriction enzyme for seven to eight hours. The digested samples were run on a 0.7-0.8% agarose gel containing RNAase A (0.1mg/ml) overnight. The following day the gel was washed with the following solutions; 0.1M HCl 5min, water 5min, 0.5M NaOH/1.5M NaCl 30 min, water 5 min, 1M ammonium acetate pH8.3 30 min.
The DNA was transferred from the gel to nitrocellulose overnight using the wick method(Maniatis et al., 1982) and 1M ammonium acetate pH8.3 as transfer buffer. The nitrocellulose filter was washed briefly in 2x SSC and baked in a vacuum oven 80, 40 min. For a 15 x 20 cm filter approximatly 50 ml of prehybridization solution was used.

Prehybridization solution

formamide	25ml
100 x denhardts	2.5ml
1M Phosphate buffer pH 6.5	2.5ml
50% dextran sulphate	10ml
20 x SSC	10ml
10% SDS	0.625ml
Denatured salmon sperm DNA (6mg/ml)	8.75ml

Prehybridizations were carried out at 42°C for a minimum of 4hrs, with continuous shaking.

Radioactive probes were prepared by the method of oligolabelling as described in the text. The probed was denatured by boiling in the presence of 5ml salmon sperm DNA (6mg/ml) prior to adding to the hybridization solution.

Hybridization conditions were carried out at 42°C overnight in the following solution;

formamide	25ml
100 x denhardts	1.0ml
1M phosphate buffer pH6.5	0.9ml
20 X SSC	12.5ml
50% Dextran sulphate	10ml
10% SDS	0.75ml

Washings were all done at room temperature using the following solutions;

2 x SSC, 0.1% SDS	15 min
2 x SSC, 0.1% SDS	15 min
0.1 x SSC 0.1% SDS	15 min
0.1 x SSC 0.1% SDS	15 min

DNA sequencing

All DNA sequencing was done using the dideoxy chain termination method as described by Sanger et. al.,(1980) J. Mol. Biol. 143, 161-178. This method uses a synthetic primer, which after annealing to a specific region of the M13 viral positive strand 3' to the cloned DNA is extended through the insert using the large fragment of DNA Pol 1 and deoxyribonucleotides. The incorporation of a dideoxy nucleotide into a newly synthesised DNA strand results in termination at that residue as the absence of a 3' hydroxyl group on the ribose moiety of the dideoxynucleotide prevents further polymerization. Thus the dideoxy nucleotides act as specific chain terminators and can be used to determine the nucleotide sequence of the inserted DNA 3' to the annealed

primer. All the sequencing reported in this thesis was carried out using a kit for dideoxy sequencing supplied by BRESA Biotechnology Research Enterprises S.A. Pty. Ltd.
Figure 2.1 indicates that a number of primers were synthesised to aid in sequencing the wheat Adh-1A gene. These primers were all 17 mers and were taken from;

Intron 2	GATGCATCAGTGGCTAA
Intron 3	GACGCTCACGCTCTAGC
Intron 4	TAGGATGCCTTGATTAG

Genomic cloning

The Charon 35 lambda vector was developed in 1983 by Wilhelmine et al., (see reference in text) and is similar to the earlier Charon series of vectors except that a deletion (W113) was made in the gam gene.

Charon 35 phage DNA was prepared by the liquid lysis method using the host strain ED8767 as described by Maniatis et al. and referenced earlier in this text. The vector DNA was digested with Bam H1 and the arms purified from the stuffer fragment by centrifugation on a 10-40% glycerol gradient. Millewa DNA was partially digested with Sau 3A to give fragments in the range 12-20kb and subsequently ligated into the lambda vector and transformed into E. Coli C600 cells for screening with the maize *Adh*1 probe as described in the text.