Molecular Analysis of Genes of White Clover (*Trifolium repens* L.)

A thesis presented in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy in Biotechnology
at Massey University, Palmerston North,
New Zealand

Nicholas Ellison

1990

Dedicated to
Kay, Jessica, Anna and Catherine

Abstract

The expression of the genes for the small subunit (SSU) of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), alcohol dehydrogenase (Adh) and lectin in the white clover plant was investigated by Northern analysis, using heterologous plant probes. SSU was shown to be expressed in the leaves/stems, Adh in the roots and lectin in both the leaves/stems and the roots of the mature plant.

A series of independent, white clover, leaf/stem and root cDNA libraries was constructed in the lambda vector λgt10 from polyadenylated messenger RNA isolated from mature plants. A number of SSU and Adh cDNA clones was isolated from these libraries and the inserts from these clones were characterized by restriction enzyme mapping and DNA sequence analysis. These clones included a partial SSU cDNA clone from a leaf/stem library and a full length Adh clone from a root library. Two uncharacterized lectin cDNA clones were also isolated from each of the leaf/stem and root cDNA libraries.

A fully-representative genomic library was constructed in the lambda vector λEMBL3 from total white clover DNA. This library was screened with the previously isolated white clover SSU and Adh cDNA clones. One SSU and three Adh genomic clones were isolated and the inserts from these clones were characterized by restriction enzyme mapping and Southern blotting. Restriction fragments, to which the corresponding cDNA probe hybridized, were subcloned and characterized by additional restriction enzyme mapping and DNA sequence analysis.

The one SSU genomic clone represented a functional white clover *rbcS* gene, corresponding to the white clover SSU cDNA clone, and was complete with 5′ and 3′ non-transcribed regions. Conserved sequences were identified in this gene that have been implicated in the regulation of plant gene expression in general and the regulation of *rbcS* genes in

particular. The three Adh genomic clones represented different, non-functional, white clover *Adh* pseudogenes, each with regions of strong homology to only limited regions of the white clover Adh cDNA clone.

List of Publications

Data from this work have been published in part in the following papers:

Ellison, N.W., Yu, P.L., and White, D.W.R. (1990a). Nucleotide sequence of a white clover ribulose bisphosphate carboxylase small subunit gene. Nucl. Acids Res. 18, 4914.

Ellison, N.W., Yu, P.L., and White, D.W.R. (1990b). Nucleotide sequence of a white clover alcohol dehydrogenase cDNA. Nucl. Acids Res. 18, 4913.

Acknowledgements

I wish to thank my supervisors, Dr P.L. Yu and Dr D.W.R. White, for their continued support throughout the duration of this study.

I am indebted to Dr W.D. Sutton for my introduction to plant molecular biology and to Dr J.P. Kerr, Director, Plant Physiology Division, and Mr J.A. Lancashire, Director, Grasslands Division, DSIR, for allowing that brief introduction to develop.

Thanks are also due to my fellow Molecular Biologists in the Palmerston North research community, and in particular to Dr G. Limsowtin, Dairy Research Institue, and Dr B. Mansfield, Massey University, for their helpful suggestions and technical advice on many aspects of molecular biology throughout the course of this study.

The assistance of Mr P. Spring (DSIR) with the photography, Mr H. Coenders (DSIR) with the growth of white clover plants in the glasshouse, and Mr C Tunnicliffe (DSIR) with keeping various items of equipment going, is gratefully acknowledged.

Table of Contents

List of Figures

List of Tables

Abbreviations

The following abbreviations have been used in this thesis:

A	absorbance
Adh	alcohol dehydrogenase
ASCII	American Standard Code for Information Interchange
bp	base pair
BSA	bovine serum albumin
cDNA	complementary DNA
ddNTP	dideoxynucleoside triphosphate
dNTP	deoxynucleoside triphosphate
DTT	dithiothreitol
kb	kilobases
LB	Luria-Bertani medium
mw	molecular weight
PEG	polyethylene glycol
pfu	plaque forming units
rpm	revolutions per minute
Rubisco	ribulose-1,5-bisphosphate carboxylase/oxygenase
SSU	small subunit of Rubisco
U	units

Abreviations not included in this list are either defined where appropriate in the text or are in accordance with "Instructions to Authors", Biochem. J. 241, 1-24, (1987).

Chapter 1

Introduction

1.1 White clover

White clover, *Trifolium repens* L., is the most agronomically important of the 250 - 300 species in the genus *Trifolium*. This genus, together with *Medicago* and *Melilotus*, is usually classified taxonomically in tribe Trifolieae of subfamily Papilionoideae of the family Leguminosae (also referred to as family Fabaceae) (Corby *et al.*, 1983; Williams, 1987).

In addition to *Trifolium repens*, the genus *Trifolium* contains more than 20 other clover species which are cultivated as forages. Among the wild species, about 54 occur only in North America and about 33 occur only in Africa south of the Sahara. The remainder are distributed mainly in Europe, northern and south-west Asia, and north Africa, with the heaviest concentration around the Mediterranean. *T. repens* itself is native throughout Europe and is now widely distributed throughout the world (Williams, 1987).

White clover has a chromosome number 2n = 32 (Britten, 1963; Ellison *et al.*, 1986). The basic chromosome number in the whole of the tribe Trifolieae, including *Trifolium*, is 8 (Senn, 1938), as it is for all closely related species in Section *Amoria* which have been counted so far (Britten, 1963; Pritchard, 1962; Pritchard, 1969; Chen and Gibson, 1971). White clover is therefore considered to be a tetraploid species. The origins of white clover and its relationships with other species of *Trifolium* remain unclear but the techniques of molecular genetics could well clarify these issues in the future.

1.2 Plant genome organization

The nuclear genome of eukaryotes contains most of the genes expressed by a species. In addition to coding for RNA and protein components of the cell, the nuclear genome also contains sufficient informational DNA to regulate the expression of these gene products and structural DNA required for the maintenance of chromosomal morphology and function.

From studies of the DNA renaturation kinetics of fragmented plant DNA, it has been established that plant genomes consist of a variable but often large proportion of repetitive DNA, which reassociates relatively rapidly, and a smaller proportion of single copy DNA, which reassociates as if present only once per haploid genome (Walbot and Goldberg, 1979; Flavell, 1980). In the case of the pea genome, approx. 85% of the DNA is composed of repetitive sequences while only 15% behaves as if it is present in one or a few copies (Murray *et al.*, 1978). The repetitive DNA fraction of plant genomes has been further subdivided into three kinds of DNA sequence organization pattern (Flavell, 1980; Flavell *et al.*, 1981). The first kind consists of repetitive DNA segments, usually shorter than 600 bp, interspersed with single- or low-copy DNA segments that are usually shorter than 1200 bp. The second kind consists of long, tandem arrays of the same repeating unit, and the third consists of stretches of repeated DNA in which different short sequences are interspersed with one another in complex permutations. Long tandem arrays of repeating units have been isolated as "satellites" by centrifugation of plant nuclear DNAs in cesium salt gradients (Ingle *et al.*, 1973), although this pattern of repetitive DNA is not limited to satellites. The lengths of the repeating units in satellite DNA arrays range from a few base pairs to thousands of base pairs, and the number of such repeating units in the chromosome ranges from a few hundred to over a million (Flavell, 1980). There is no experimental evidence regarding the function of satellite DNA. However in

in situ hybridization experiments, this DNA has been found associated with heterochromatic regions around the centromere and the telomere of the chromosome (Bedbrook *et al.*, 1980a; Peacock *et al.*, 1981; Jones and Flavell, 1982; Appels, 1983; Zhu *et al.*, in preparation). These observations suggest that satellite DNA may be important for chromosome structure and replication.

Together, interspersed and tandemly arranged, highly repetitive DNA can constitute up to 80% of plant nuclear DNA (Walbot and Goldberg, 1979). The haploid genome size of various plants and the proportion of repetitive DNA in each genome is given in Table 1.1. As can be seen from this table, a larger genome size tends to correlate with a higher proportion of repetitive DNA in the genome. The *Arabidopsis* genome, at 70,000 kb, is approx. 20 times the size of the *E. coli* genome and only five times larger than that of the yeast genome (Leutwiler, 1984).

Of the single copy DNA within a plant genome, not all is actually transcribed into RNA. Thus in the tobacco plant, approx. 45% of the single copy DNA is represented in the nuclear RNA of all vegetative and floral organ systems. The majority of these sequences are developmentally regulated with 18% of the single copy DNA being represented in the nuclear RNA of each organ (Kamalay and Goldberg, 1984). As a result of post-transcriptional selection mechanisms, only 11% of the single copy DNA, equivalent to 60,000 diverse structural genes, is actually expressed in the form of mRNA in the entire tobacco plant during the dominant phase of its life cycle. This amount of genetic information constitutes just 5% of the tobacco genome.

Table 1.1. Haploid genome size in various plants and the proportion of repetitve DNA in each genome. The data have been taken from Hake and Walbot (1980), Leutwiler (1984), Sorensen (1984) and Meyerowitz and Pruitt (1985).

Plant	Haploid genome size (bp × 10^9)	Proportion of repetitive DNA (%)
Arabidopsis	0.07	26
Millet	0.2	69
Mung bean	0.47	34
Cotton	0.78	40
Spinach	0.8	74
Tobacco	1.5	55
Soybean	1.8	61
Parsley	3.0	70
Pea	4.5	70
Barley	5.0	70
Wheat	5.9	75
Maize	6.3	77
Rye	7	75
Oats	13	70
Broadbean	44	80

1.3 Structure and expression of plant genes

The structure of a typical plant gene, the pea *rbcS-3A* gene, is shown in in Figure 1.1 (based on Fluhr *et al.*, 1986; Giuliano *et al.*, 1988; Kuhlemeier *et al.*, 1988). For descriptions of the main structural features, see below.

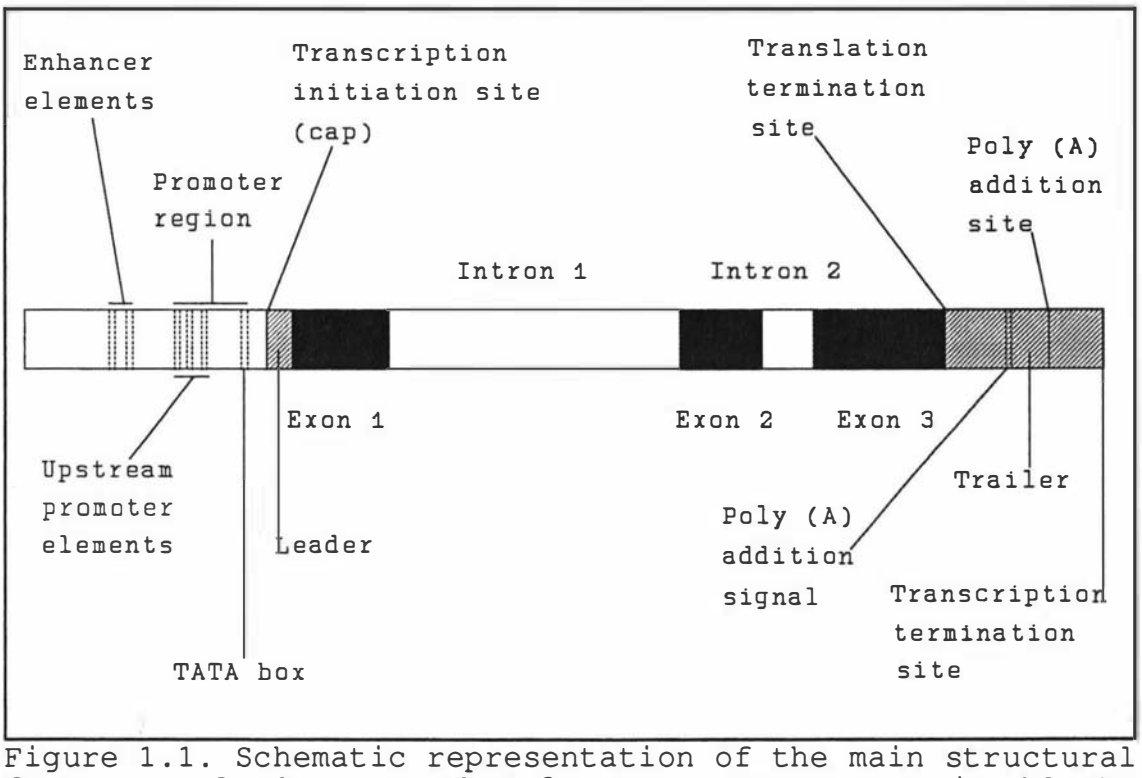


Figure 1.1. Schematic representation of the main structural features of the pea *rbcS-3A* gene. Exons are in black; untranslated regions of the corresponding mRNA are cross-hatched.

1.3.1 RNA polymerases

Plant nuclear DNA sequences are transcribed by three separate forms of RNA polymerase, I, II and III. The enzymes are structurally and functionally distinct and closely resemble the RNA polymerases found in animal cell nuclei. RNA polymerase I is located in the nucleolus where it transcribes the genes for rRNA. RNA polymerase II is located in the nucleoplasm and is important for transcribing the genes coding for proteins into mRNA. RNA polymerase III is believed to be responsible for the transcription of low molecular weight sequences in the nucleoplasm, including the genes for 5S and tRNA (Grierson and Covey, 1984).

1.3.2 Transcription initiation

Transcription initiation by RNA polymerase II is controlled by two types of regulatory sequences, namely promoters and

enhancers. Both types of sequences are typically composed of a variety of short sequence motifs (*cis* elements) to which different transcription factors bind. Such factors can be either general or gene specific. The promoter usually consists of a sequence around the transcription initiation site (cap site), an AT-rich sequence (TATA box) and one or more upstream elements of about 10 bp (see below). The upstream promoter elements and their associated transcription factors control the frequency of transcription initiation and therefore the level of transcripts of the corresponding gene. Enhancers, on the other hand, increase the rate of transcription initiation and can be located either upstream or downstream of the initiation site, close to the promoter or thousands of base pairs away.

The transcription factors involved in the initiation of gene transcription consist of at least three different domains which specify DNA binding, translocate the protein to the nucleus, and interact with a component, or components, of the general transcription apparatus. A transcription factor can also contain an additional domain for interaction with regulatory proteins that can up- or down-regulate the transcriptional activity of the factor.

While considerable progress has been made in understanding the structure and function of transcription factors that bind to promoters and enhancers, there is still little known about the way these factors interact with the general transcription apparatus. Two principal models, the looping and the scanning model, have been proposed for this process (Müller and Schaffner, 1990). In the looping model, initiation of transcription is stimulated by the interaction of enhancer elements with promoter elements through proteins bound to the DNA. The DNA between the enhancer and promoter elements is thereby looped out. In the scanning model, enhancers are recognized by RNA polymerase II (or a transcription factor), which binds and then slides in either direction along the DNA

until it reaches the promoter elements where it facilitates the formation of a transcription initiation complex. Of these two models, the available evidence tends to support the looping model (Schleif, 1988; Müller and Schaffner, 1990).

1.3.3 Gene regulatory sequences

A number of relatively short nucleotide sequences has been identified that are considered to be important for controlling gene transcription (see above) as well as for controlling mRNA processing and translation. Some of these sequences, or variations on them, are found in all eukaryotic genes while others are specific for particular genes.

One of the putative transcriptional controls of most protein-encoding genes is the TATA box promoter element. The binding of transcription factor IID (TFIID) to this element is the first step in transcription from these genes (Schmidt *et al.*, 1989). This is followed by the consecutive binding of TFIIA, TFIIB, RNA polymerase II and TFIIE to form a transcription complex (Buratowski *et al.*, 1989). In plants, the TATA box promoter element is located approx. 32 nucleotides upstream from the transcription initiation site (Joshi, 1987a). The sequence of the TATA box varies slightly in different plant genes but a consensus sequence of this element and flanking regions, TCACTATATATAG, has been identified, where the region underlined is the actual TATA box (Joshi, 1987a). The TATA box of animal genes, with a consensus sequence of TATA(A/T)A(A/T), is also located in a similar position relative to the transcription initiation site (Breathnach and Chambon, 1981).

It was originally believed that the TATA box itself determined the actual site of transcription initiation (Grierson and Covey, 1984). However it is now evident that an independent initiator element (Inr), containing within itself the transcription initiation site, is involved in transcription

from this initiation site. When present, the TATA box activates transcription initiation from the Inr but this is not essential (Smale and Baltimore, 1989). A consensus sequence, 5'-PyAPyPyPyPyPy-3'has been identified around the transcription iniation site of a number of animal genes, with transcription beginning at the A residue (Breathnach and Chambon, 1981). The equivalent consensus sequence for plant genes is 5'-CTCATCA-3', with transcription beginning at the underlined A, although different groups of genes exhibit varied consensus sequences (Joshi, 1987a).

A second possible transcription regulation sequence, the CAAT box is located approx. 70 nucleotides upstream from the transcription initiation site. In some plant genes, the CAAT box is replaced by an AGGA box (Heidecker and Messing, 1986). Evidence that this element is important for gene regulation comes from studies of animal gene expression which indicate that it may regulate, along with additional enhancer elements, the extent of transcription (Grierson and Covey, 1984).

1.3.4 Gene introns

A typical eukaryotic gene consists of short coding sequences (exons) interrupted by relatively long non-coding sequences (introns). Plant introns are usually between 70 and 1000 nucleotides in length whereas vertebrate introns range from 70 to thousands of nucleotides, fungal introns from 40 to 100 nucleotides and insect introns from 50 to 100 nucleotides (Hawkins, 1988). A minimum functional length for a plant intron has been determined to be between 70 and 73 nucleotides (Goodall and Filipowicz, 1990).

The introns in a primary RNA (pre-mRNA) from a transcribed gene are removed by the splicing apparatus in the nucleus, after which the spliced RNA (mRNA) is transported to the cytoplasm where it is translated. The splicing process is

regulated by several sequences at the exon/intron junction, as well as by sequences within the intron itself. The splice sites on each side of an intron have been found to possess consensus sequences that are probably recognized by the splicing apparatus. In eukaryotic genes in general, the 5′ splice junction of the intron consists of the 8-nucleotide conserved sequence AG:GT(A/G)AGT while the 3′ splice junction consists of a pyrimidine-rich region of about 11 nucleotides followed by (C/T)AG:G (Shapiro and Senapathy, 1987), where (:) indicates the intron/exon boundary.

The splice junction sequences of plant genes are very similar to the animal consensus sequences (Brown, 1986; Hanley and Schuler, 1988). However significant differences between monocot and dicot genes have been found in both the 5′ and 3′ junction regions which may account for the low efficiency of expression of monocot genes in dicot plants (Hanley and Schuler, 1988). On the other hand, a dicot gene can be correctly spliced in a monocot plant (Peterhans *et al.*, 1990). Plant introns in general have a high AT content relative to the introns of vertebrate genes (Goodall and Filipowicz, 1989) and tend to lack the pyrimidine-rich region at the 3′ end of the intron (Weibauer *et al.*, 1988). It has been postulated that these two features of plant introns can account for the inability of plant cells to process most of the animal introns that have been tested (Goodall and Filipowicz, 1989).

Little is known about the actual function of introns in the expression of genes *in vivo*. A large number of genes are known to undergo alternative patterns of pre-mRNA splicing, giving rise to protein isoforms sharing extensive regions of identity and varying only in specific domains (Smith *et al.*, 1989). This feature thus permits an increase in the diversity of gene products from a single locus. In maize cells, introns of the maize *Adh1* gene have been shown to increase the expression of both *Adh1* and a chimeric gene when located near

the 5′ end of the gene (Callis *et al.*, 1987). This is in contrast to the finding that intronless bean phaseolin genes transferred to tobacco are capable of full expression of phaseolin protein in callus tissue (Chee *et al.*, 1986). In animal cells, both negative and positive gene regulatory elements have been located within an intron (Bruhat *et al.*, 1990).

1.4 Structure of plant mRNAs

Nuclear genes coding for proteins are transcribed by RNA polymerase II into mRNAs. The mRNAs represent 2 - 4% of the cellular RNA and range in size from a few hundred to several thousand nucleotides long. In addition to the coding region, the mRNAs contain 5′ and 3′ sequences of untranslated nucleotides of variable length which have particular features.

Chemical analysis of some plant mRNAs has shown that they contain a 5′ cap structure related to that found in animal mRNAs (Grierson and Covey, 1984). In this cap structure, a terminal 7-methylguanosine residue and the penultimate nucleotide are joined by their 5′-hydroxyl groups through a triphosphate bridge (Shatkin, 1976). This 5′-5′ linkage is inverted, relative to the normal 3′-5′ phophodiester bonds in the remainder of the polynucleotide chain. The cap structure is added soon after the start of mRNA synthesis and before mRNA chain completion (Salditt-Georgieff *et al.*, 1980). The cap is not essential for mRNA translation *in vitro* but is thought to function by binding a specific protein which may regulate initiation of protein synthesis (Grierson and Covey, 1984). It may also function to stabilize the pre-mRNA in the nucleus, since mammalian extracts appear to contain a 5′ to 3′ exonuclease activity which is blocked by the cap structure (Green, 1986).

The length of plant mRNA leader sequences (the distance from the cap site to the translation initiation codon) varies from 9 - 193 nucleotides, with a predominant length of 40 - 80 nucleotides (Joshi, 1987b). Most of the leader sequences analyzed were found to be AT rich.

The 3′ ends of nearly all mRNAs in eukaryotic cells terminate in a homopolymer of 20 - 250 adenosine nucleotides. In general, most RNA polymerase II transcripts are polyadenylated, with mRNAs for histones in some organisms being the only known exceptions (Hunt, 1988). Polyadenylation occurs in the nucleus and immediately follows a cleavage reaction at the polyadenylation site in which a phosphodiester bond in the mRNA precursor is cleaved. Adenosine nucleotides are then added to the new 3′ hydroxyl group one at a time. This poly(A) 'tail' is usually shortened once the mRNA enters the cytoplasm. The poly(A) tail both stimulates translation and stabilizes the mRNA (Callis *et al.*, 1987; Wickens, 1990).

1.5 Regulatory signals on plant mRNAs

In addition to the regulatory sequences associated with plant genes that control aspects of gene transcription (see above), other regulatory sequences associated with the primary transcripts of these genes control post-transcriptional processes. In the interests of simplicity in describing these sequences both here and in the Discussion (Chapter 3), the sequence of the DNA coding strand is used instead of the mRNA sequence (i.e. T in place of U).

1.5.1 Translation initiation

In the majority of plant mRNAs that have been analyzed, the initiation of translation occurs at the first ATG triplet from the 5′ end (Heidecker and Messing, 1986; Joshi, 1987). This is also a feature of eukaryotic mRNAs in general (Kozak,

1984). Since additional ATGs do occur in some mRNAs upstream of the functional ATG, the sequence context within which the functional ATG occurs is also considered to be important for translation initiation (Kozak, 1984; Heidecker and Messing, 1986; Joshi, 1987). This sequence context has been referred to as the Kozak sequence. In animal mRNAs, the consensus for this sequence is CACCATG and in plant mRNAs it is AACAATGGC (Lütcke *et al.*, 1987; Joshi, 1987), where the underlined ATG is the functional initiation codon.

The modified scanning ribosome model (Kozak, 1984) has been proposed to explain how the correct ATG translation initiation codon is selected. In this model, 40S ribosomal subunits scan along from the 5′ end of the mRNA in a linear fashion until the first ATG is located. If the sequence preceding the first ATG conforms closely to the consensus sequence, the search for an ATG stops and translation initiation occurs.

1.5.2 Polyadenylation

In animal cells, the polyadenylation of mRNAs (see Section 1.4) requires the specific hexanucleotide sequence AATAAA. This hexanucleotide is typically located 15 - 25 nucleotides upstream of the actual site of polyadenylation and is highly conserved. Every base in this sequence is required, although the stringency of the requirement varies with the position of the base substituted (Sheets *et al.*, 1990). The only other requirement for polyadenylation besides AATAAA is a minimum distance between this sequence and the 3′ end of the mRNA. However the actual identity of the downstream sequence does not appear to be important for polyadenylation.

The cleavage reaction that normally precedes polyadenylation requires both AATAAA and additional sequences located downstream of the poly(A) addition site. These downstream elements are generally T- or TG-rich but are less highly conserved than AATAAA.

The sequence requirements for the polyadenylation of mRNAs in plants have not been studied to the same extent as they have in animals. Sequences similar to AATAAA occur in a large number of plant genes at a distance from the poly(A) addition site comparable to that found in animal genes but appear to be entirely lacking from many other plant genes (Hunt *et al.*, 1987; Joshi, 1987b). However, when the search for this sequence was extended to 200 nucleotides upstream of the poly(A) addition site, nearly all plant genes were found to contain at least one sequence related to AATAAA (Hunt and McDonald, 1989). In addition to the position of AATAAA relative to the polyadenylation site, the degree of variation in this sequence is much greater in plant genes than it is in animal genes (Dean *et al.*, 1986; Hunt *et al.*, 1987; Joshi, 1987b). In many cases, multiple polyadenylation signals occur in plant genes and these give rise to a heterogeneous population of mRNAs from the one transcription unit (Dean *et al.*, 1986; Hunt, 1988). The total lack of recognition (Barta *et al.*, 1986), or inefficient and incorrect utilization (Hunt *et al.*, 1987) of animal polyadenylation signals in plant cells suggests that there are also possible functional differences between the poly(A) signals of plant and animal genes.

1.6 Codon usage in plant genes

Codon usage is the selective and nonrandom use of synonymous codons by an organism to encode the amino acids in the genes for its proteins. Analysis of a large group of plant gene sequences indicates that the patterns of codon usage in plants are distinct from those of *E. coli*, yeast and man (Murray *et al.*, 1989). A clear distinction between dicot and monocot codon usage is also evident. Dicot genes use a set of 44 preferred codons with a slight preference for codons ending in A or T while monocot codon usage is more restricted with an average of 38 codons preferred which are predominantly those ending in C or G (Murray *et al.*, 1989; Campbell and

Gowri, 1990). These differences could well influence the ability of one group of plants (monocot or dicot) to translate genes derived the other group.

1.7 Ribulose-1,5-bisphosphate carboxylase

1.7.1 Role of ribulose-1,5-bisphosphate carboxylase

The enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) catalyses the initial steps of two opposing metabolic pathways, carboxylation and oxygenation of ribulose-1,5-bisphosphate. The carboxylation reaction is the first step in the photosynthetic fixation of CO_2. The reaction yields two molecules of phosphoglycerate, which are partly recycled in the Calvin cycle to regenerate ribulose-1,5-bisphosphate, and partly converted to starch, the main storage form of photosynthetic chemical energy. The oxygenation reaction yields one molecule each of phosphoglycerate and phosphoglycolate. Phosphoglycolate is metabolized in the photorespiratory pathway where reduced carbon is oxidized to CO_2.

1.7.2 Synthesis and structure of Rubisco

Rubisco is a chloroplast enzyme consisting of 16 subunits, eight large subunits (LSU) of molecular weight approx. 55,000, and eight small subunits (SSU) of molecular weight approx. 14,000. The large subunit of the enzyme contains the catalytic site, is encoded by a single gene (*rbcL*) on the chloroplast genome, and is synthesized on chloroplast ribosomes. The small subunit (*rbcS*) is encoded by the nuclear genome as a small multigene family of 5 to 15 members in most higher plants (Section 1.7.3).

The SSU protein is synthesized on cytoplasmic ribosomes as a large precursor protein with an amino-terminal extension, the

transit peptide, which mediates the transport of the polypeptide into the chloroplast and is cleaved off during or after transport. The mature SSU assembles in the chloroplast stroma with the LSU to form the holoenzyme (for reviews, see Miziorko and Lorimer, 1983, and Manzara and Gruissem, 1988). The contribution of the SSUs to the function of the holoenzyme is largely unresolved but may include modulation of substrate binding and possibly the carboxylation/oxygenation ratio (Schneider *et al.*, 1990). The SSUs are not directly involved in the catalytic chemistry but appear to have an effect by way of conformational alterations to the LSUs on assembly of the holoenzyme (Andrews, 1988).

1.7.3 Organization and structure of *rbcS* genes

For recent reviews on the organization and structure of *rbcS* genes, see Manzara and Gruissem (1988) and Dean *et al.* (1989a). Of the large number of *rbcS* genes that have been characterized, most of these have been derived from the gene families of the dicots petunia, tomato, potato, tobacco, pea, soybean and *Arabidopsis* and the monocots *Lemna gibba* (duckweed), wheat and maize (see below).

1.7.3.1 Petunia

Petunia has eight *rbcS* genes, all of which have been cloned and sequenced (Tumer *et al.*, 1986; Dean *et al.*, 1987). The genes are divided into three subfamilies based on nucleotide sequence homology. One subfamily contains six genes, five of which are closely linked in the petunia genome, while the other two contain single genes. There is 10.2% nucleotide sequence divergence in the mature SSU coding regions among the genes of the three different subfamilies and 0% to 3% within each subfamily. In general, the nucleotide sequence encoding the transit peptide shows a higher degree of divergence among subfamilies than the sequence encoding the mature SSU.

1.7.3.2 Tomato

Tomato has five *rbcS* genes, all of which have been cloned and sequenced (Sugita *et al.*, 1987). Two of the genes are present as single genes at individual loci and the three others are organized in a tandem array within 10 kb at a third independent locus. The coding sequences of the linked genes diverge by up to 4.7% in the transit peptide region and 1.6% in the mature SSU. The coding sequences of the two unlinked genes diverge by 14% from each other and by 10.4% and 13.3% from one of the linked genes (Sugita *et al.*, 1987).

1.7.3.3 Potato

Potato has at least five *rbcS* genes which have been cloned, four as genomic clones and one as a cDNA clone (Wolter *et al.*, 1988). Three of the genes are organized in a tandem array and one of the others occurs at a single locus. No sequence data are available for these genes.

1.7.3.4 Tobacco

The *rbcS* genes have been characterized from three *Nicotiana* species: *N. tabacum* (Mazur and Chui, 1985; O'Neal *et al.*, 1987), *N. sylvestris* (Pinck *et al.*, 1986), and *N. plumbaginifolia* (Poulsen *et al.*, 1986). Five unlinked *N. tabacum rbcS* genes have been isolated and, of these, three have been completely sequenced (Mazur and Chui, 1985; O'Neal *et al.*, 1987). One of the *N. tabacum* genes has been shown to be a transcriptionally active pseudogene (O'Neal *et al.*, 1987). In addition, there could be up to three other *rbcS* genes in this species (O'Neal *et al.*, 1987). In *N. sylvestris* there are at least seven *rbcS* genes as determined by Southern analysis (Pinck *et al.*, 1986). One *N. sylvestris* genomic clone, encoding two genes separated by a 200 bp intergenic region, has been isolated and the genes have been sequenced (Jamet *et al.*, 1990). One of these genes has a high degree of

homology to the transcriptionally active *N. tabacum* pseudogene (Jamet *et al.*, 1990). The one *N. plumbaginifolia* gene that has been sequenced (Poulsen *et al.*, 1986) has a high degree of homology to the sequence of another *N. tabacum* gene sequence.

1.7.3.5 Pea

Pea has at least five *rbcS* genes, which are clustered on chromosome 5 (Polans *et al.*, 1985). All of these genes have been cloned and sequenced (Coruzzi *et al.*, 1984; Timko *et al.*, 1985; Fluhr *et al.*, 1986a). There is little nucleotide sequence divergence between the coding regions of the five genes and all encode an identical mature SSU.

1.7.3.6 Soybean

Soybean has 6 - 10 *rbcS* genes and two of these have been cloned and sequenced (Berry-Lowe *et al.*, 1982; Grandbastien *et al.*, 1986). There is little nucleotide sequence divergence between the coding regions of the two genes, and the proteins for which they code differ at only one residue.

1.7.3.7 *Arabidopsis*

Arabidopsis has four *rbcS* genes, all of which have been cloned and sequenced, and these can be divided into two families based on linkage and nucleotide sequence similarities (Krebbers *et al.*, 1988). Three of the genes, sharing greater than 95% similarity in nucleotide sequence, are organized in a tandem array and the remaining gene occurs at a separate locus.

1.7.3.8 *Lemna*

Lemna gibba has approx. twelve *rbcS* genes, six of which have been cloned and five of which have been sequenced (Wimpee *et*

al., 1983; Silverthorne *et al.*, 1990). Two pairs of the genes that have been characterized show linkage.

1.7.3.9 Wheat

The hexaploid genome of *Triticum aestivum* contains at least ten *rbcS* genes (Broglie *et al.*, 1983; Smith *et al.*, 1983). Of these, one gene has been characterized as a genomic clone and four have been characterized as different cDNA clones (Smith *et al.*, 1983; Broglie *et al.*, 1983).

1.7.3.10 Maize

One genomic *rbcS* clone (Lebrun *et al.*, 1987) and four different *rbcS* cDNA clones (Sheen and Bogorad, 1986; Matsuoka *et al.*, 1987) have been isolated from *Zea mays*, although the nucleotide sequence of one of the cDNA clones is identical to the transcribed region of the genomic clone with the sequence of the intron removed.

1.7.3.11 Intron positions

All *rbcS* genes from dicots sequenced so far have at least two introns that are always found in the same positions (Manzara and Gruissem, 1988; Dean *et al.*, 1989). In the monocot species, the *rbcS* genes of *Lemna gibba* lack the first of these two introns (Silverthorne *et al.*, 1990) while the single wheat, maize and rice genes characterized lack the second of these introns (Xie and Wu, 1988; Dean *et al.*, 1989). One *rbcS* gene in each of the species petunia, tomato, potato, and tobacco, all members of the family *Solanaceae*, contains an additional intron downstream from the second intron (Manzara and Gruissem, 1988; Dean *et al.*, 1989a).

1.7.4 Regulation of Rubisco gene expression

The expression of the *rbcS* gene is highly regulated, with transcripts being most abundant in leaves and less abundant or undetectable in other organs (Coruzzi *et al.*, 1984; Fluhr *et al.*, 1986a). In addition, the expression is at least 20-fold induced by light. The effective wavelength of the light for induction depends on the developmental stage of the plant. In immature pea plants, the photoreceptor phytochrome mediates the induction of one of the *rbcS* genes, since transcription is turned on by a pulse of red light, whereas the effect can be reversed by far-red light (Fluhr *et al.*, 1986b). In mature plants, however, there is evidence for an additional blue photoreceptor acting in concert with phytochrome (Fluhr and Chua, 1986). While regulation of *rbcS* gene expression is primarily at the level of transcription, differential mRNA stabilities also appear to be involved since soybean *rbcS* RNA is degraded more rapidly in light than in darkness (Shirley and Meagher, 1990).

Nagy *et al.* (1988) propose a model for phytochrome-regulated gene expression. The first step in this process involves the conversion of the red light absorbing form of phytochrome to the far-red light absorbing form in response to illumination. Changes in the structural features of phytochrome as a result of this conversion in some way trigger the chain of events involved in the signal transduction pathway itself. This leads to the replacement of a repressor, bound to multiple copies of redundant light-regulatory elements, by an activator and this in turn leads to the induction of gene transcription.

1.7.5 *RbcS* upstream regulatory elements

A number of conserved sequence elements involved in the regulation of *rbcS* gene expression has been identified within the upstream sequence of *rbcS* genes from a variety of plants

(Green *et al.*, 1987; Giuliano *et al.*, 1988; Manzara and Gruissem, 1988). These are in addition to the TATA and the CAAT boxes (Section 1.3.3), although sequences in the vicinity of the TATA box itself have also been implicated in light-dependent expression of the pea *rbcS* genes (Morelli *et al.*, 1985; Kuhlemeier *et al.*, 1989). In some cases, multiple copies of these *cis*-acting elements occur within the upstream sequence of a particular gene. For the location of these conserved elements in the pea *rbcS-3A* gene, see Figure 1.2 (based on Green *et al.*, 1987; Giuliano *et al.*, 1988; Manzara and Gruissem, 1988). Additional elements, which are only conserved among the *rbcS* genes of a specific *rbcS* gene family or among *rbcS* genes of closely related species, have been identified in some *rbcS* genes (Fluhr *et al.*, 1986; Manzara and Gruissem, 1988). In petunia, promoter sequences that contribute to quantitative differences in expression between the members of the *rbcS* gene family have also been identified (Dean *et al.*, 1989b). It should be noted that regulatory elements are not necessarily restricted to the upstream region of the *rbcS* gene. Sequences downstream of the initiation codon, and possibly downstream of the coding region itself, have been shown to contribute to quantitative differences in expression between two *rbcS* genes of petunia (Dean *et al.*, 1989c).

1.7.5.1 Box I

The 'box I' sequence 5'-TTCAA-3' has been identified in all of the pea *rbcS* genes analyzed (Fluhr *et al.*, 1986; Manzara and Gruissem, 1988). This element may be a component of a positive regulatory factor in the pea *rbcS* genes (Green *et al.*, 1987).

1.7.5.2 Box II

The sequence 5'-GTGTGGTTAATATG-3', termed 'box II', is conserved among all *rbcS* genes analyzed (Fluhr *et al.*, 1986;

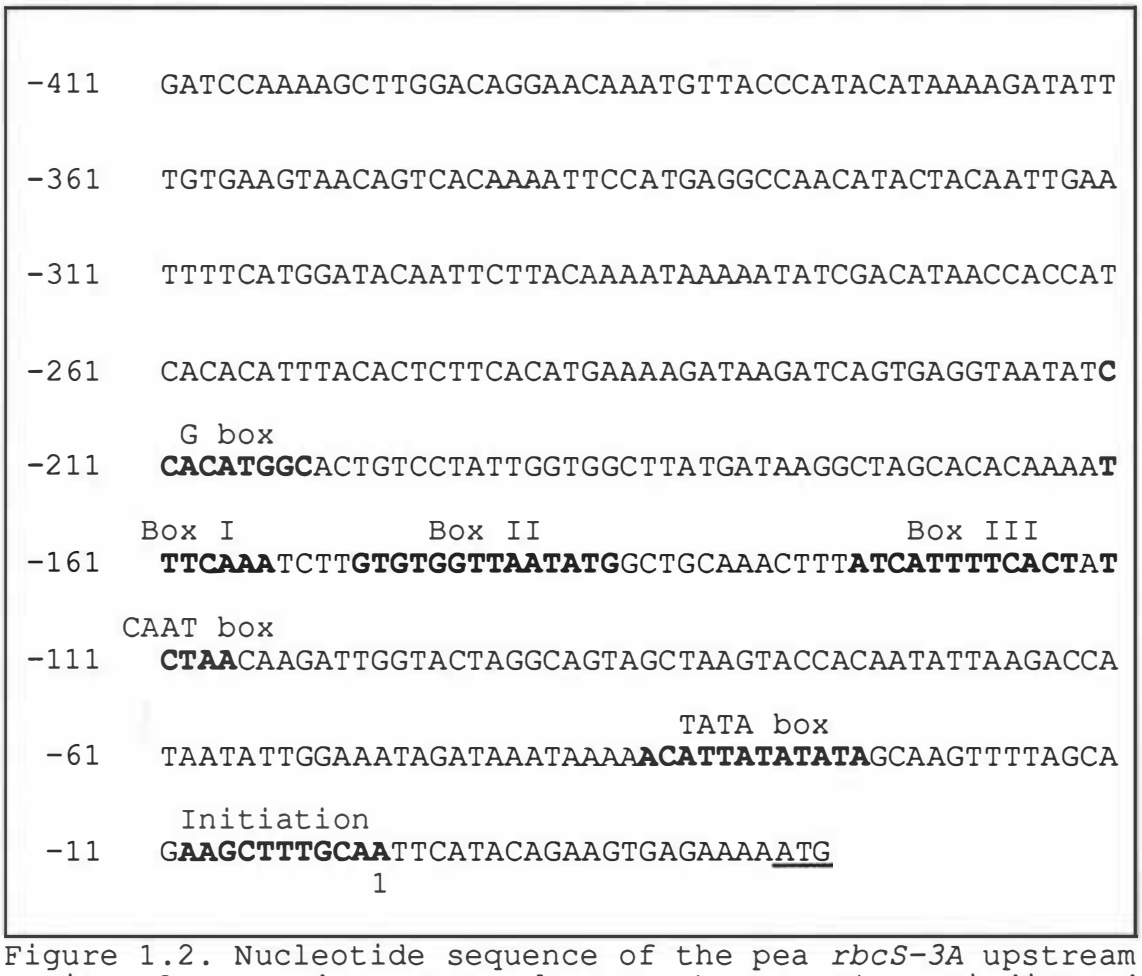


Figure 1.2. Nucleotide sequence of the pea *rbcS-3A* upstream region. Conserved sequence elements (see text) are indicated by bold type; position 1 indicates the start of transcription; the ATG start codon of the gene is underlined.

Manzara and Gruissem, 1988). This element has been shown to function as a *cis*-acting element (Kuhlemeier *et al.*, 1987a, 1988) as well a silencer in the dark for the pea small subunit genes (Fluhr *et al.*, 1986; Kuhlemeier *et al.*, 1987a; Manzara and Gruissem, 1988). The box II element has significant homology with the SV40 core enhancer, a constitutive mammalian enhancer element (Coruzzi *et al.*, 1984) In pea, a nuclear protein factor that specifically interacts with the box II sequence, designated GT-1, has been identified in nuclear extracts by gel retardation assay (Green *et al.*, 1987).

1.7.5.3 Box III

The sequence 5'-ATCATTTTCACT-3', termed 'box III', is conserved among all of the pea and soybean *rbcS* genes analyzed (Fluhr *et al.*, 1986; Manzara and Gruissem, 1988). This element appears to be analogous to the box II element in pea (Section 1.7.5.2) and is also bound by the factor GT-1 (Green *et al.*, 1987).

1.7.5.4 G box and I box

Further upstream from the box I element, 'G box' (5'-C/ACACGTGGC-3') and 'I box' (5'-GGATGAGATAAGAYTA-3') sequences have been identified (Giuliano *et al.*, 1988; Manzara and Gruissem, 1988), which have also been implicated in *rbcS* gene regulation (Ueda *et al.*, 1989). The G box has been identified in all of the *rbcS* genes analyzed whereas the complete I box has been identified only in tomato, tobacco and soybean *rbcS* genes (Giuliano *et al.*, 1988; Manzara and Gruissem, 1988). In both tomato and *Arabidopsis thaliana* nuclear extracts, a protein factor, designated GBF, that specifically binds to the G box of *rbcS* genes of these plants, as well as to the G boxes of *rbcS* genes of heterologous plants, has been identified (Giuliano *et al.*, 1988). By site-specific mutagenesis, a requirement for conserved G-box and I-box elements for expression from the *Arabidopsis rbcS-1A* promoter has been established (Donald and Cashmore, 1990).

1.7.6 Conservation of *rbcS* regulatory elements

Of the above sequence elements found upstream of the white clover *rbcS* ATG initiation codon, the TATA box, the CAAT box, the box II and the G box elements have been shown to be conserved among all *rbcS* genes of dicotyledenous plants examined (Giuliano *et al.*, 1988; Manzara and Gruissem, 1988; Hutchison *et al.*, 1990). The G-box, an octamer with perfect

dyad symmetry (5'-CCACGTGG-3'), has also been implicated in the transcriptional regulation of the *Arabidopsis thaliana Adh* gene (DeLisle and Ferl, 1990; McKendree *et al.*, 1990) (see Section 1.8.2) and the parsley chalcone synthase gene (Schulze-Lefert *et al.*, 1990). It is interesting to note that the 5'-GTGG-3' motif of the G-box, without the dyad symmetry, occurs in the separate box II element. It is likely that the highly conserved elements within the promoter of the *rbcS* gene play an essential role in the regulation, as well as in the light-inducible and/or organ-specific expression, of this gene. It is also likely that additional, as yet unidentified, elements will be found that contribute to the regulation of these genes. The ability of several of the *rbcS* upstream sequences to confer regulated expression in heterologous plant species (Broglie *et al.*, 1984; Nagy *et al.*, 1985; Fluhr and Chua, 1986) suggests that the mechanisms that mediate light-responsive transcription are conserved to some degree. However, not all of these mechanisms are conserved among all species, as indicated by the observation that a wheat *rbcS* gene under the control of its own promoter is inactive in transgenic tobacco plants (Keith and Chua, 1986).

1.7.7 Organ-specific expression of *rbcS*

As well as their light-responsive nature, *rbcS* genes appear to be expressed only in chloroplast-containing cells (Kuhlemeier *et al.*, 1987; Aoyagi *et al.*, 1988). This organ specificity is controlled by *cis*-acting elements that are independent of the light-regulatory elements and are located in the enhancer region of the gene rather than the promoter region (Kuhlemeier *et al.*, 1989).

1.8 Alcohol dehydrogenase in plants

Alcohol dehydrogenase (ADH) catalyses the reversible inter-conversion of ethanol and acetaldehyde with the cofactors NAD and NADH respectively. The plant enzymes are dimeric with

subunits of approx. 40 kD size (Freeling and Bennett, 1985) and can utilize many primary and secondary alcohols as substrates. Most plants contain two or three isozymes, all located in the cytosol (Gottlieb, 1982). ADH activity has been found in the roots (Sachs and Freeling, 1978; Kimmerer, 1987; Xie and Wu, 1989), seeds (Yamashita *et al.*, 1976; Yamashita *et al.*, 1978; Woodman and Freeling, 1981; Dolferus and Jacobs, 1984; Dolferus *et al.*, 1985), germinated seeds (Dolferus *et al.*, 1985; Ricard *et al.*, 1986), leaves (Tanksley and Jones, 1981; Kimmerer, 1987; Xie and Wu, 1989), pollen (Freeling, 1976; Dolferus and Jacobs, 1984; Xie and Wu, 1989) and fruit (Bicsak *et al.*, 1982) of monocot and dicot species. ADH activity can be induced by anaerobiosis (Freeling, 1973; Tanksley and Jones, 1981) and 2,4-dichlorophenoxyacetic acid (2,4-D) (Freeling, 1973). *Adh* genes from *Arabidopsis* (Chang and Meyerowitz, 1986), barley (Trick *et al.*, 1988), maize (Dennis *et al.*, 1984; Dennis *et al.*, 1985; Sachs *et al.*, 1986), pea (Llewellyn *et al.*, 1987), rice (Xie and Wu, 1990) and strawberry (Wolyn and Jelenkovic, 1990) have been cloned and sequenced.

1.8.1 Regulation of plant *Adh* gene expression

The most extensively studied plant ADH system has been that of maize. The *Adh* genes of this plant (*Adh1* and *Adh2*) are two of approx. 20 genes which are coordinately induced in primary roots under anaerobic conditions (Sachs *et al.*, 1980). In the absence of ADH1 activity, maize seeds or seedlings are unable to survive the anaerobic conditions induced by flooding (Schwartz, 1969). The effect of ADH2 activity on survival of anaerobiosis is minimal and only apparent in plants devoid of ADH1 activity (Freeling and Bennett, 1985). The two *Adh* loci are unlinked, with *Adh1* located on chromosome 1 (Schwartz, 1971) and *Adh2* located on chromosome 4 (Dlouhy, 1980, cited in Trick *et al.*, 1988).

In the maize plant, *Adh1* expression is controlled by both developmental and environmental signals. ADH1 activity is constitutively expressed in pollen aleurone and scutellum, but expression is at either low or undetectable levels in roots, mesocotyl and epicotyl (Freeling and Bennett, 1985). Under anaerobic conditions, *Adh1* expression is induced in a variety of organs, including roots, endosperm and anther wall, but not in mature leaves (Okimoto *et al.*, 1980). Nuclear run-on transcription experiments reveal that transcriptional induction of *Adh1* can be detected within one hour of anaerobic treatment and is maximal after five hours (Rowland and Strommer, 1986). When anaerobically-stressed seedlings are returned to an aerobic environment, transcription of *Adh1* falls to almost undetectable levels within one hour, but *Adh1* mRNA can be detected for at least 26 hours (Rowland and Strommer, 1986).

In tomato, only ADH1 is synthesized during embryogenesis, while only ADH2 is synthesized after flooding (Tanksley and Jones, 1981). In *Lupinus angustifolius*, the *Adh1* gene is expressed both during embryogenesis and flooding, while the *Adh2* gene is expressed only in roots of young seedlings (Marshall *et al.*, 1974). In sunflower (*Helianthus annuus*), both *Adh1* and *Adh2* genes are expressed during embryogenesis, seed dormancy, and early germination while only *Adh2* is expressed in flooded roots and in pollen (Torres *et al.*, 1977). In soybean, only one ADH is synthesized in roots, the activity of which is high in the embryo and radicle but decreases rapidly during germination and is not induced by anaerobiosis (Brzezinski *et al.*, 1986). The single *Adh* gene of *Arabidopsis thaliana* is expressed in pollen, seeds, and roots of young seedlings and is induced by flooding (Dolferus and Jacobs, 1984).

1.8.2 *Adh* upstream regulatory elements

In the maize *Adh1* gene, an anaerobic regulatory element (ARE) has been located within the promoter region and this element has been functionally defined by a deletion analysis of constructs expressed in a transient expression system (Howard *et al.*, 1987; Walker *et al.*, 1987) (Figure 1.3). This 40 bp element is composed of two essential regions (region I and region II) of approx. 15 bp each, separated by a 10 bp region that does not appear to be important for anaerobic expression (Walker *et al.*, 1987). A segment of the maize promoter containing this element is sufficient for anaerobic regulation in transgenic tobacco when it is combined with upstream promoter elements from the octopine synthase gene or the cauliflower mosaic virus 35S promoter (Ellis *et al.*, 1987). Homologous sequences also occur within the promotor region of the pea *Adh1* gene (Llewellyn *et al.*, 1987) and the *Arabidopsis Adh* gene (Ferl and Laughner, 1989). *In vivo* foot-printing experiments indicate that *trans*-acting regulatory factors bind to a DNA sequence within this element in the maize gene (Ferl and Nick, 1987) as well as to an homologous sequence within the promoter region of the *Arabidopsis Adh* gene (Ferl and Laughner, 1989).

In addition to the ARE, several factor binding sites occur within the promoters of both the maize *Adh1* and the *Arabidopsis Adh* genes (Ferl and Nick, 1987; Ferl and Laughner, 1989). The sequence motif 5′-GTGG-3′ is usually associated with these binding sites (Ferl and Laughner, 1989). At one position in the *Arabidopsis Adh* gene, this motif is present in both strands as the perfect dyad 5′-CCACGTGG-3′, but this does not occur in the maize *Adh1* gene (Ferl and Laughner, 1989). The sequence of this dyad is identical to the sequence of the G-box which occurs in the 5′-flanking sequence of a number of *rbcS* genes as well as several other genes (see Sections 1.7.5.4 and 1.7.6). In general, the factors which bind to the GTGG sites of the

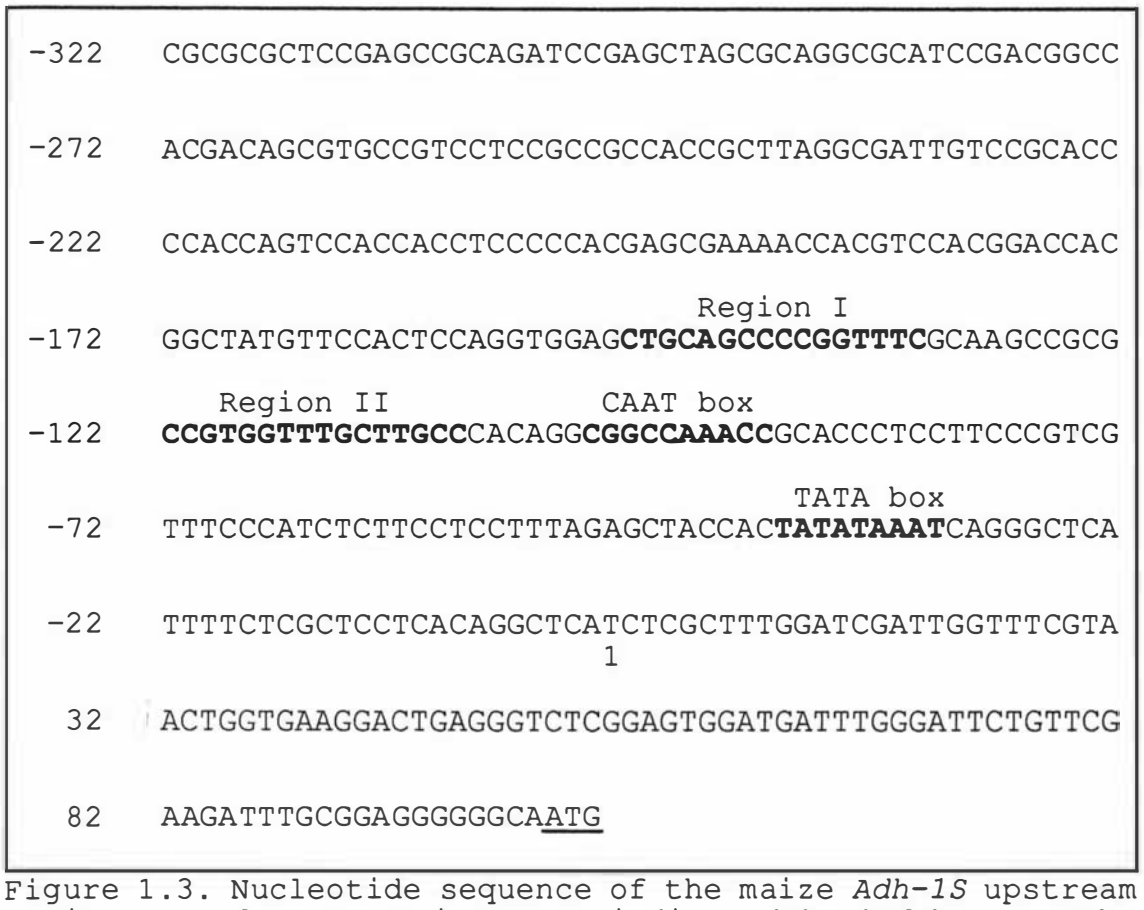


Figure 1.3. Nucleotide sequence of the maize *Adh-1S* upstream region. Regulatory regions are indicated by bold type; the sequence is numbered from the transcription start site (position 1); the ATG start codon of the gene is underlined.

maize *Adh* promoter appear to be involved in regulating the rate of transcription from the gene but not in anaerobic regulation (Ferl and Laughner, 1989). However, the G-box binding factor of the *Arabidopsis Adh* gene is separate and distinct from the GTGG binding factors (McKendree *et al.*, 1990).

1.9 Aim of this study

No work has been reported on the molecular structure or organization of genes from white clover. This study was aimed at determining the structure and the patterns of expression of certain genes from this plant. Of primary interest were genes that were expressed in the white clover plant in an organ-specific manner, such as the *rbcS* and *Adh* genes.

Chapter 2

Materials and Methods

2.1 Germination and growth of white clover

2.1.1 Germination of white clover seeds

Seeds of white clover (*Trifolium repens* cv. Huia) were immersed in drum ethanol for 30 min and then the ethanol was poured off. The seeds were covered with acidified, 0.2% (w/v) $HgCl_2$ (5 ml of conc. HCl per litre of $HgCl_2$ solution), left for 6 min and then rinsed 5 times with sterile H_2O.

2.1.2 Growth of white clover

Germinated seeds were planted out in trays containing potting mix consisting of either 100% sand or equal parts of sand and peat. The plants were grown in a glasshouse and watered three times per week with Modified Hoaglands Solution - ½ Strength.

Requirements

- Modified Hoaglands Solution - ½ Strength (10 litres): mix 67 ml Solution A, 67 ml Solution B and 67 ml of concentrated HCl. Make up to 10 litres with H_2O. This solution should have a pH of 6.5 - 7.5.
- Solution A (10 litres): dissolve 1,480 g $Ca(NO_3)_2.4H_2O$, 52 g Sequestrene 330 Fe in H_2O. Make up to 10 litres with H_2O.
- Solution B (10 litres): dissolve 170 g KH_2PO_4, 632 g KNO_3, 616 g $MgSO_4.7H_2O$, 3.58 g boric acid, 2.26 g $MnCl_2.4H_2O$, 0.28 g $ZnSO_4.7H_2O$, 0.10 g $CuSO_4.5H_2O$, 0.034 g $Na_2MoO_4.2H_2O$, 7.88 g KCl in H_2O. Make up to 10 litres with H_2O.

2.1.3 Harvesting of white clover plants

After eight weeks of growth, leaves and stems of plants growing in each tray were trimmed off at the surface of the potting mix, weighed into 20 g lots and stored at -70°C. Roots were recovered by washing off the potting mix with a stream of water (this was successful for those plants grown in 100% sand only), weighed into 20 g lots, and stored at -70°C.

2.2 Isolation of total RNA from white clover tissues

2.2.1 Phenol extraction procedure

This method was based on that of Lichtenstein and Draper (1985). Frozen plant material (100 g) was ground to a fine powder in a 900 ml Waring blender container, prechilled to -20°C, with a small amount of dry ice. Homogenization buffer (300 ml; see below) was added and the plant material was homogenized in the blender as the mixture thawed. This mixture was transferred to sterile 250 ml centrifuge bottles (Nalgene) and, in some cases, was further homogenized briefly with a Polytron (Brinkmann Instruments) set at either 5 or at maximum speed. Yields of RNA were improved up to ten-fold with this additional homogenization (see Section 3.1). The homogenate was then centrifuged in a Sorvall GSA rotor at 5,000 rpm and 4°C for 10 min. The supernatant was filtered through four layers of autoclaved cheesecloth into a 1 litre conical flask and SDS was added from a 10% stock (Section 2.22.6) to give a final concentration of 0.5%. Equal volumes of water saturated phenol and chloroform (Section 2.22.3) were then added and mixed with the filtrate for 20 min on a magnetic stirrer, at room temperature, at a speed just sufficient to keep the phases emulsified. The mixture was transferred to 250 ml polypropylene centrifuge bottles (Nalgene) and the phases were separated by centrifugation at

8,000 rpm and 4°C for 10 min. The aqueous phase was removed to a clean 1 litre flask, an equal volume of chloroform was added and this was mixed for 10 min on a magnetic stirrer as above. The centrifugation step was repeated and the aqueous phase was removed to 50 ml centrifuge tubes (Oak Ridge, Nalgene). Sodium acetate/acetic acid, pH 6.0 (3 M, 1/20 volume) and 2 vol. of absolute ethanol were added to each tube and the tubes were incubated at -20°C overnight to precipitate the nucleic acids. The tubes were then centrifuged in a Sorvall SS-34 rotor at 10,000 rpm and 4°C for 20 min and the pellets were washed twice with 5 ml of ice-cold 3 M sodium acetate/acetic acid, pH 6.0 to remove the ethanol and the DNA. The pellets (RNA) were washed twice with 80% ethanol, 0.1 M potassium acetate and stored under this solution at -20°C until required.

Requirements

- Homogenization buffer: 0.2 M Tris-HCl pH 8.5, 0.2 M sucrose (RNase free), 30 mM magnesium acetate, 60 mM KCl. Autoclave and either use fresh or store frozen at -20°C. Before use, add polyvinylpolypyrolidone (Sigma) to 1% (w/v) and 2-mercaptoethanol to 0.31% (v/v).
- Water saturated phenol: add H_2O to a 500 g bottle of commercial phenol and leave overnight to liquefy. Store in aliquots at -20°C.

2.2.2 Guanidine hydrochloride procedure

This method was based on that of Logemann *et al.* (1987). Frozen plant material (100 g) was ground to a fine powder, as described above for the phenol extraction procedure, and then 200 ml of guanidine buffer was added. The mixture was homogenized in the Waring blender and further homogenized with a Polytron (see Section 2.2.1). The homogenate was transferred to a 1 litre conical flask and 150 ml each of water-saturated phenol (Section 2.2.1) and chloroform

(Section 2.22.3) were added. The homogenate was mixed on a magnetic stirrer for 20 min as above. The mixture was transferred to 250 ml polypropylene centrifuge bottles and the phases were separated by centrifugation in a Sorvall GSA rotor at 8,000 rpm and 4°C for 60 min. The RNA-containing aqueous phase was removed to clean 250 ml centrifuge bottles and mixed with 0.7 volume of cold ethanol (-20°C) and 0.2 volume of 1 M acetic acid. The RNA was precipitated by incubation at -20°C overnight and then pelleted by centrifugation in a GSA rotor at 8,000 rpm and 4°C for 15 min, leaving DNA and residual proteins in the supernatant. The RNA pellets were washed with 10 ml of 3 M sodium acetate/acetic acid, pH 5.2 at room temperature, combined in a 50 ml centrifuge tube (Oak Ridge, Nalgene) and centrifuged in a Sorvall SS-34 rotor at 10,000 rpm and 4°C for 10 min. The supernatant was discarded and the pellet (RNA) was washed with 20 ml of 70% ethanol and recentrifuged to remove residual salt. The RNA pellet was subsequently dried for 5 min under vacuum, redissolved in sterile H_2O, and run on an oligo(dT) cellulose column as described in Section 2.3.1.

Requirements

- Guanidine buffer: 8 M guanidine hydrochloride, 20 mM Mes, 20 mM EDTA, 50 mM 2-mercaptoethanol, pH 7.0.

2.3 Isolation of Poly(A)$^+$ RNA

2.3.1 Oligo(dT) cellulose chromatography

The method used for separating poly (A)$^+$ RNA (mRNA) from other species of RNA by affinity chromatography on an oligo(dT) cellulose column was based on that of Lichtenstein and Draper (1985). Oligo(dT) cellulose (0.5 g; Boehringer Mannheim) was preswollen overnight in 5 ml of binding buffer. The swollen oligo(dT) cellulose was then packed into a 4.0 cm × 0.7 cm (internal diameter) column (Bio-Rad, Econo-Column)

such that the glass barrel was almost full. This gave a column volume of approx. 1 ml. The column was washed through sequentially with 5 ml of sterile distilled water, 0.3 M NaOH until the eluent was alkaline, and then 10 ml of binding buffer. The ethanol precipitated total RNA (Sections 2.2.1 and 2.2.2) was centrifuged for 10 min in 1.5 ml microfuge tubes. The supernatants were removed and the pellets were dried for 5 min under vacuum and then dissolved in a total volume of 0.5 ml of sterile, distilled H_2O, with heating to 65°C for 4 min. An equal volume of 2× binding buffer was added and the solution was allowed to cool to room temperature. The RNA solution was then slowly added to the oligo-(dT) column, the column was washed through with 10 ml of binding buffer and 1 ml fractions were collected in sterile microfuge tubes, until the A_{260} approached zero. The column was then washed through with 5 ml of washing buffer and the poly(A) mRNA was then eluted with 5 ml of eluting buffer. Fractions (0.5 ml) were collected in 1.5 ml microfuge tubes, the A_{260} of each fraction was determined to locate the poly(A) mRNA-containing fractions and these fractions were combined in a siliconized 15 ml glass centrifuge tube (Corex). Sodium acetate/acetic acid, pH 5.2 (3 M, 0.1 volume) and 2 vol. of absolute ethanol were added to precipitate the mRNA. The mRNA was stored, as a precipitate, at -20°C until required.

Requirements

- Binding buffer: 20 mM Tris-HCl, pH 7.5, 0.5 M NaCl, 1.0 mM EDTA, 0.05% SDS.
- Washing buffer: 20 mM Tris-HCl, pH 7.5, 0.1 M NaCl, 1.0 mM EDTA, 0.05% SDS.
- Eluting buffer: 20 mM Tris-HCl pH 7.5, 1.0 mM EDTA, 0.02% SDS.
- Autoclave the buffers before adding SDS from a 10% stock solution (Section 2.22.6).

2.3.2 Regeneration of oligo(dT) cellulose

The oligo(dT) cellulose column was washed through sequentially with 2× 5 ml of binding buffer (Section 2.3.1), 10 ml of sterile H_2O, 5 ml of 0.3 M NaOH, 30 ml of sterile H_2O, and 30 ml of 95% ethanol. The washed oligo(dT) cellulose was dried in the column overnight under vacuum and then removed from the column and stored at -20°C.

2.4 Synthesis of double-stranded cDNA

The synthesis of double-stranded cDNA from mRNA was based on the methods of Gubler and Hoffmann (1985) and Watson and Jackson (1985). The procedure relies on the activity of reverse transcriptase to synthesize a complementary DNA strand from an oligo(dT)-primed mRNA, and the combined activities of *E. coli* RNase H and DNA polymerase I to synthesize a second complementary cDNA strand from the mRNA-cDNA hybrid by RNA-primed nick translation. As originally described, the reaction products of the first strand synthesis need to be purified before being used for the second strand synthesis reaction. A one-tube protocol based on the work of Lapeyre and Amalric (1985), D'Alessio *et al.* (1987) and Gerard and D'Alessio (1987), whereby the first strand reaction was diluted to give a solution of the correct composition for the second strand reaction, was developed and used in this study. Subsequent steps involved making the double-stranded products of the second strand reaction blunt-ended, attaching *Eco*RI linkers, and cloning the linkered cDNAs into the bacteriophage lambda vector λgt10.

In order to optimize the first and second strand synthesis reactions, trial reactions were carried out with synthetic RNA consisting of a series of six poly(A)$^+$ RNAs (0.24 - 9.5 kb RNA ladder, BRL). These RNAs, once labelled, were also used to provide size markers for subsequent analysis of reaction products on alkaline agarose gels.

2.4.1 First Strand cDNA Synthesis

A volume of white clover poly(A)$^+$ RNA (Section 2.3.1) equivalent to 10 µg was centrifuged in a sterile 1.5 ml microfuge tube in a Sorvall SS-34 rotor with an appropriate adaptor (Sorvall, No. 00381) at 15,000 rpm and 4°C for 30 min. The RNA pellet was rinsed with 70% ethanol, dried for 5 min under vacuum and redissolved in 19 µl of sterile H_2O. For trial reactions and reactions to provide size markers for agarose gel analysis of reaction products, 5 µg of synthetic RNA, as supplied by the manufacturer (see above), was used for the reaction and made up to 19 µl with sterile H_2O. Immediately before use, the RNA was heated at 65°C for 5 min and then chilled on ice. The first strand cDNA synthesis reaction was set up in a 1.5 ml microfuge tube as follows:

Poly(A)$^+$ RNA	19 µl
5× RT buffer	10 µl
5× d(ACGT)TPs	5 µl
200µg/ml oligo $(dT)_{15}$	5 µl
Reverse transcriptase	11 µl
TOTAL	50 µl

A 10 µl aliquot of the reaction was transferred to a fresh tube containing 1 µl of [α-^{32}P] dCTP so that the efficiency of the reaction could be monitored. Both reactions were incubated at 42°C for 2 h.

After 2 h, the 10 µl reaction was extracted with an equal volume of phenol/chloroform (Section 2.22.2). The first strand product from this reaction was precipitated by adding 0.5 volume of 7.5 M ammonium acetate and 2 vol. of absolute ethanol and the tube was centrifuged for 20 min in a microfuge at room temperature (see Section 2.21). The pellet was redissolved in 10 µl of sterile H_2O and the ethanol precipitation from ammonium acetate was repeated. The pellet was then redissolved in 25 µl of TE (Section 2.22.4) for

subsequent analysis of the first strand products by gel electrophoresis (Section 2.11.1). The 40 µl reaction was used for the second strand cDNA synthesis reaction, as described in Section 2.4.2.

Requirements

- 5× RT buffer: 250 mM Tris-HCl, pH 8.3 at 42°C, 40 mM $MgCl_2$, 250 mM KCl.
- 5× d(ACGT)TP: dissolve each deoxynucleoside triphosphate (Boehringer Mannheim) at a concentration of 5 mM; store at -70°C.
- Reverse transcriptase (avian myeloblastosis virus) at a concentration of 9 U/µl was obtained from Promega.
- [α-^{32}P] dCTP: see Section 2.14.1.

2.4.2 Second Strand cDNA Synthesis

The second strand cDNA synthesis reaction was set up as follows:

1st strand reaction	40 µl
5× 2nd strand buffer	64 µl
E. coli DNA polymerase 1	8 µl
RNase H	1 µl
5 mM d(ACGT)TPs	12 µl
[α-^{32}P] dCTP	2 µl
H_2O	193 µl
TOTAL	320 µl

The reaction was incubated at 16°C for 2 h and then 10 µl of 0.5 M EDTA was added and the reaction was extracted twice with an equal volume of phenol/chloroform. The aqueous phase was distributed into two 1.5 ml microfuge tubes and 0.5 volume of 7.5 M ammonium acetate and 2 vol. of absolute ethanol were added to each tube to precipitate the DNA (see Section 2.21). The tubes were centrifuged for 20 min and then

the supernatants were removed. The DNA pellets were re-dissolved in a total volume of 100 µl of H_2O and combined into one tube. The ethanol precipitation from ammonium acetate was repeated and the cDNA pellet was finally re-dissolved in 25 µl of H_2O. As an alternative to ethanol precipitation of the double stranded cDNA, the DNA was purified with Geneclean (BIO 101) as described in Section 2.11.8 and eluted from the powdered glass with two aliquots of 12.5 µl of H_2O. The DNA was counted by Cerenkov counting in a liquid scintillation counter.

Requirements

- 5× 2nd strand buffer: 94 mM Tris-HCl, pH 8.3, 470 mM KCl, 20 mM $MgCl_2$, 20 mM DTT (from 1 M stock; see Section 2.14.1).
- 5× d(ACGT)TP: see Section 2.4.1.
- *E. coli* DNA polymerase 1 at a concentration of 10 U/µl was obtained from New England Biolabs.
- RNase H at a concentration of 2.3 U/µl was obtained from Boehringer Mannheim.
- [α-^{32}P] dCTP: see Section 2.14.1.

2.4.3 Filling in reaction

The filling in reaction to make the double-stranded cDNAs blunt-ended was set up as follows:

cDNA (from Section 2.4.2)	25 µl
5× d(ACGT)TPs	1 µl
10× T4 polymerase buffer	3 µl
T4 DNA polymerase	1 µl
TOTAL	30 µl

The reaction was incubated at 37°C for 30 min and then it was extracted with 60 µl of phenol/chloroform (Section 2.22.2). The phenol phase was re-extracted with 30 µl TE (Section

2.22.4) and the aqueous phases were combined in a 0.6 ml microfuge tube. The blunt-ended cDNAs were purified with Geneclean (Section 2.11.8) and eluted from the powdered glass with two 5 µl aliquots of H_2O.

Requirements

- 5× d(ACGT)TPs: Section 2.4.1.
- 10× T4 DNA polymerase buffer: 330 mM Tris-acetate, pH 7.9, 660 mM potassium acetate, 100 mM magnesium acetate, 1 mg/ml BSA, 5 mM DTT (from 1 M stock; see Section 2.14.1).
- T4 DNA polymerase at a concentration of 10 U/µl was obtained from Promega.
- BSA (nuclease free) at a concentration of 50 mg/ml was obtained from BRL.

2.4.4 Phosphorylation of *Eco*RI linkers

*Eco*RI linkers are short, self-complementary, chemically synthesized oligomers which form blunt end duplexes containing an *Eco*RI recognition sequence. Before use, the linkers were phosphorylated as follows:

Dephosphorylated *Eco*RI linkers	5 µl
10× kinase buffer	5 µl
T4 polynucleotide kinase	3 µl
H_2O	37 µl
TOTAL	50 µl

The reaction was incubated at 37°C for 1 h and then 5 µl aliquots were frozen at -70°C.

Requirements

- Dephosphorylated *Eco*RI linkers (New England Biolabs; No. 1004) were dissolved in sterile H_2O at a concentration of 1 µg/µl.

- 10× kinase buffer: 500 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$, 100 mM DTT (from 1 M stock; see Section 2.14.1), 10 mM ATP (from 0.1 M stock).
- 0.1 M ATP: dissolve 60 mg of ATP in 0.8 ml of H_2O. Adjust the pH to 7.0 with 0.1 M NaOH. Adjust the volume to 1.0 ml with H_2O. Dispense the solution into 200 µl aliquots and store at -70°C.

2.4.5 Methylation of *Eco*RI sites

*Eco*RI sites within the double-stranded and end-filled cDNAs (section 2.4.3) were methylated to protect them from subsequent *Eco*RI digestion as follows:

Double-stranded cDNA	10.0 µl
800 µM SAM	2.0 µl
5× methylase buffer	4.0 µl
*Eco*RI methylase	0.5 µl
H_2O	3.5 µl
TOTAL	20.0 µl

The reaction was incubated at 37°C for 30 min and then it was extracted with 40 µl of phenol/chloroform (Section 2.22.2). The phenol phase was re-extracted with 20 µl of TE (Section 2.22.4) and the aqueous phases were combined in a 0.6 ml microfuge tube. The methylated cDNAs were purified with Geneclean as described in Section 2.11.8 and eluted from the powdered glass with 5 µl of phosphorylated linkers (Section 2.4.4) and then with 5 µl of H_2O.

Requirements

- SAM: 30 mM S-Adenosylmethionine (New England Biolabs) in 5 mM sulphuric acid : ethanol (9:1 v:v) was stored at -20°C and diluted to 800 µM in sterile H_2O immediately before use.

- 5× methylase buffer: 500 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM EDTA, 1 mg/ml BSA (for 50 mg/ml BSA, see Section 2.4.3).
- *Eco*RI methylase (New England Biolabs) was at a concentration of 20 U/µl.

2.4.6 Ligation of linkers

The phosphorylated *Eco*RI linkers (Section 2.4.4) were ligated onto the double-stranded cDNAs (Section 2.4.5) as follows:

Methylated cDNAs plus linkers	10.0 µl
5× ligase buffer	2.0 µl
T4 DNA ligase	1.2 µl
TOTAL	13.2 µl

The reaction was incubated at 14°C overnight and then heated at 65°C for 10 min.

Requirements

- 5× ligase buffer: see section 2.15.2.
- T4 DNA ligase at a concentration of 0.72 U/µl (Weiss units) was obtained from BRL.

2.4.7 Digestion with *Eco*RI

The linker ligation reaction (Section 2.4.6) resulted in the covalent attachment of oligomers of the *Eco*RI linkers to the ends of the cDNAs. Excess linkers were removed by digestion with *Eco*RI, leaving a single linker at each end, as follows:

Ligated cDNA/linkers	13.2 µl
0.5 M NaCl	1.2 µl
High concentration *Eco*RI	1.6 µl
TOTAL	16.0 µl

The reaction was incubated at 37°C for 2 h. TE (30 µl; Section 2.22.4) was then added and the reaction was extracted with 100 µl of phenol/chloroform (Section 2.22.2). The phenol phase was re-extracted twice with 50 µl of TE and the aqueous phases were combined and distributed into two 1.5 ml microfuge tubes. *Eco*RI-digested λgt10 DNA (10 µg; Section 2.4.8) was added to each tube and the DNAs were precipitated by the addition of 0.1 volume of 3 M sodium acetate/acetic acid, pH 5.2 and 2 vol. of absolute ethanol. The tubes were centrifuged in a Sorvall SS-34 rotor (with a microfuge tube adaptor, Sorvall, No. 00381), at 15,000 rpm for 20 min. The pellets were rinsed with 70% ethanol, dried for 5 min under vacuum and redissolved in 96 µl of sterile H_2O.

Requirements

- High concentration *Eco*RI, at a concentration of 90 U/µl, was obtained from Boehringer Mannheim.

2.4.8 *Eco*RI digestion of λgt10 DNA

The λgt10 vector DNA (Section 2.5.4) was digested with *Eco*RI in preparation for cDNA cloning as follows:

λgt10 DNA (0.83 µg/µl)	60 µl
10× M buffer	30 µl
*Eco*RI	20 µl
H_2O	190 µl
TOTAL	300 µl

The reaction was incubated at 37°C for 3 h. An aliquot was analyzed by agarose gel electrophoresis (Section 2.11.3) to confirm that the *Eco*RI digestion was complete. EDTA (0.5 M; 1.5 µl) was then added and the digest was heated at 65°C for 10 min and extracted three times with an equal volume of phenol/chloroform (Section 2.22.2). The DNA was precipitated by the addition of 0.1 volume of 3 M sodium acetate/acetic

acid, pH 5.2 and 2 vol. of ethanol and pelleted by centrifugation for 20 min (see Section 2.21). The DNA pellet was rinsed sequentially with 1 ml of 70% ethanol and 1 ml of ether and was then redissolved in sterile H_2O at a final concentration of 1 µg/µl. The DNA solution was stored at -20°C until required.

Requirements

- 10× M: see Section 2.18.
- *Eco*RI at a concentration of 20 U/µl was obtained from New England Biolabs.

2.4.9 Removal of excess linkers

The efficiency of cloning cDNAs into the λgt10 vector is improved by the removal of competing linker fragments. This is usually achieved by passage over an agarose-based resin, Bio-Gel A-50-m (Bio-Rad). However it has been reported that this resin contains a ligation inhibitor which itself reduces cloning efficiency (Pape and Kim, 1987). Therefore for this study, NACS PREPAC mini-columns (BRL; cartridge style) contaning an ion exchange resin were used to remove excess linkers following the method described in the NACS PREPAC manual.

NaCl (5 M; 4 µl) was added to the redissolved cDNAs plus the *Eco*RI-digested λgt10 DNA (Section 2.4.7) to give a NaCl concentration (0.2 M) equivalent to that of buffer A. Using a 5 ml disposable syringe, the NACS PREPAC column was hydrated with 3 ml of buffer D and then equilibrated with 3 ml of buffer A. Each sample was loaded onto a column and the column was washed with 3 ml of buffer A to separate the bound cDNAs from the unbound linkers. The cDNAs were then eluted from the column with 2× 100 µl of buffer B followed by 2× 100 µl of buffer D. The cDNA effluents from each column were combined in a 1.5 ml microfuge tube. Ethanol (600 µl) was

added to each tube and the tubes were centrifuged in a Sorvall SS-34 rotor at 15,000 rpm (with a microfuge tube adaptor, Sorvall, No. 00381) for 20 min. The pellets were then dried for 5 min under vacuum and redissolved in 7 µl of 10 mM Tris HCl, pH 7.4, 10 mM $MgCl_2$.

Requirements

- Buffer A: 0.2 M NaCl, 10 mM Tris-HCl, pH 7.2, 1 mM EDTA.
- Buffer B: 1.0 M NaCl, 10 mM Tris-HCl, pH 7.2, 1 mM EDTA.
- Buffer D: 2.0 M NaCl, 10 mM Tris-HCl, pH 7.2, 1 mM EDTA.

2.4.10 Ligation of cDNAs with vector DNA

The cDNA plus the *Eco*RI digested λgt10 DNA (Section 2.4.9) were incubated at 42°C for 15 min to allow the λgt10 cohesive ends to anneal and then 2 µl of 10 mM ATP, 0.1 M DTT and 1 µl of T4 DNA ligase were added. The ligation reactions were incubated overnight at 14°C. The ligated DNAs were then packaged into phage particles *in vitro* as described in Section 2.4.11.

Requirements

- 10 mM ATP, 0.1 M DTT: prepare from 0.1 M ATP (Section 2.4.4) and 1 M DTT (Section 2.14.1) stocks.
- T4 DNA ligase at a concentration of 0.72 U/µl (Weiss units) was obtained from BRL.

2.4.11 *In vitro* packaging of λgt10/cDNA hybrids

In vitro packaging refers to the use of a cell extract derived from phage-infected *E. coli* to supply the mixture of proteins and precursors required for encapsidating lambda DNA. For this study, the Packagene extract system (Promega) was used following the protocol supplied with the system.

Two packaging extracts were used for each ligation. The extracts were thawed on ice and then the ligated λgt10/cDNAs were added. The contents were mixed by gently tapping the bottom of each tube and the reactions were incubated at 22°C for 2 h. PSB (0.5 ml) was then added to each tube and the tubes were gently vortexed. Chloroform (25 µl; Section 2.22.3) was added and the tubes were gently vortexed again. The packaged phage (unamplified cDNA library) were stored at 4°C over chloroform.

Requirements

- PSB (phage suspension buffer): 10 mM Tris-HCl, pH 7.4, 100 mM NaCl, 10 mM $MgSO_4$, 0.05% gelatin. The gelatin is dissolved separately in H_2O with heating and then added to the remaining components. Autoclave to sterilize.

2.4.12 Assay of cDNA libraries

Appropriate dilutions of the packaged λgt10/cDNA were made in PSB (see Section 2.4.11). Duplicate 10 µl aliquots of each dilution were plated out with 100 µl of *E. coli* C600 and 100 µl of *E. coli* C600ΔHfl plating bacteria (Sections 2.5 and 2.9.1) as described in Section 2.4.14.

2.4.13 Amplification of cDNA libraries

The libraries of recombinant λgt10/cDNA phage (Section 2.4.12) were amplified by preparing plate stocks directly from the packaging mixtures. Plating out of phage was as described in Section 2.4.14 except that each packaging mixture was mixed with 125 µl of *E. coli* C600ΔHfl (Sections 2.5 and 2.9.1) and plated out on 150 mm diameter plates with 5 ml of lambda top agar. The plates were incubated, right side up, overnight at 37°C. The plates were then overlaid with 10 ml of PSB (Section 2.4.11) and incubated overnight at 4°C with gentle shaking. The overlay solution was transferred

to 50 ml centrifuge tubes (Oak Ridge, Nalgene) and 200 µl of chloroform was added to each tube. The tubes were centrifuged in a Sorvall SS-34 rotor at 7,000 rpm and 4°C for 10 min. The supernatants were then transferred to sterile, screw-capped, 10 ml tubes (Kimax). Chloroform was added to 0.3% and the tubes were stored at 4°C. For long term storage, dimethyl-sulphoxide was added to the aqueous phase to a final concentration of 7% (v/v) and the phage were stored at -70°C. The amplified libraries were assayed as described in Section 2.4.12.

2.4.14 Plating out λ phage

For most purposes, 85 mm diameter plates were used for plating out λ phage. The volumes of *E. coli* plating bacteria and top agar required for larger plates are indicated in the relevant section, as is the type of container required for the adsorption of the phage to the *E. coli*.

An aliquot of the λ suspension (typically 10 µl) was mixed with 100 µl of *E. coli* plating bacteria (Section 2.9.1) in sterile, capped, 16 mm (external diameter), glass tubes. The tubes were incubated at 37°C in a constant temperature block heater for 20 min to allow the phage to adsorb to the *E. coli* cells. Lambda top agar (3.5 ml), pre-equilibrated at 55°C, was then added to each tube. The contents were quickly vortexed and transferred to an LB plate pre-equilibrated at 37°C. The plate was gently swirled so that the top agar was spread evenly over the plate and then left at room temperature until the top agar had hardened. The plate was inverted and incubated overnight at 37°C to allow plaques to develop.

Requirements

- Lambda top agar (200 ml): 2 g Bacto-tryptone, 1 g yeast extract, 2 g NaCl, 1.3 g agar. Adjust the pH of the

solution to 7.5 with NaOH and then autoclave for 25 min. Cool to 60°C and add 2 ml of sterile 1 M $MgSO_4$. Store at room temperature. Before use, melt the agar in a microwave oven (loosen top of bottle) and then cool to 55°C.

- LB plates (per litre): 10 g Bacto-tryptone, 5 g yeast extract, 10 g NaCl. Adjust pH to 7.5 with NaOH and autoclave for 25 min. Cool to 60°C and pour into plates in a laminar flow cabinet. Once the agar has hardened, the lids are removed and the plates are allowed to dry for 20 min in the laminar flow cabinet. Dried plates are sealed in a plastic bag and stored at room temperature.

2.5 Growth of λgt10

For background on the phage lambda vector, λgt10, as well as on the methods used for constructing cDNA libraries in this vector, see Huynh *et al.*, (1985). λgt10 contains a single *Eco*RI cleavage site within the phage repressor gene (*c*I). Insertion of a DNA fragment into this gene generates a cI^- phage which forms a plaque with a clear centre, whereas the cI^+ λgt10 parental phage forms a plaque with a turbid centre. This allows recombinant cI^- phage to be distinguished from the parental phage. Spontaneous clear-plaque phage do occur in the cI^+ λgt10 population and therefore a stock containing the lowest proportion of clear-plaque phage is selected from a number of prepared stocks. Stocks of λgt10, *E. coli* C600, and *E. coli* C600ΔHfl were obtained from Dr B. Mansfield, Massey University.

Requirements

- *E. coli* C600 (Jendrisak et al., 1987): F^-, *thi*-1, *leu*B6, *lac*Y1, *ton*A21, *sup*E44, λ^-.
- *E. coli* C600ΔHfl (Jendrisak et al., 1987): F^-, *thi*-1, *leu*B6, *lac*Y1, *ton*A21, *sup*E44, λ^-, *hfl*A150, [*chr*.Tn10].

2.5.1 Preparation of λgt10 plate stock

Methods for the preparation of λgt10 plate stocks were based on those of Maniatis *et al.* (1982). The original stock of λgt10 was assayed on *E. coli* C600 as described in Section 2.4.12. Five plugs of top agar, each with a single, turbid plaque, were excised from the assay plate with a small spatula and transferred to individual 1.5 ml microfuge tubes containing 100 µl of PSB (Section 2.4.11). One drop of chloroform was added to each tube and the tubes were stored overnight at 4°C. A 20 µl aliquot of each plaque isolate was then plated out with 100 µl of *E. coli* C600 (Section 2.9.1) as described in Section 2.4.14. After overnight incubation at 37°C, when confluent lysis was evident, 5 ml of PSB was added to each plate. The plates were incubated for 5 h at 4°C with intermittent shaking to allow the phage to diffuse out of the top agar and into the PSB. The PSB was then removed from each plate and transferred to sterile, 10 ml, screw-capped tubes (Kimax). Each of these plate stocks was assayed on *E. coli* C600 as above. Duplicate aliquots of appropriate dilutions of each plate stock were then plated out on *E. coli* C600 to give approx. 5×10^3 pfu per plate. After overnight incubation at 37°C, the plates were scored for the presence of turbid plaques and the stock with the lowest number of turbid plaques was selected for subsequent use.

2.5.2 Preparation of λgt10

Methods for the large-scale growth of λgt10 plate lysates were based on those of Huynh *et al.* (1985) and Maniatis *et al.* (1982). Aliquots of the λgt10 plate stock, containing 2.5×10^6 pfu (Section 2.5.1), were plated out with 200 µl of *E. coli* C600 (Section 2.9.1) and 5 ml of lambda top agar on ten 150 mm LB glucose plates as described in Section 2.4.14. The plates were inverted and incubated at 37°C for 6 h. The plates were then cooled to 4°C and 10 ml of lambda diluent was added to each plate. The plates were incubated at 4°C

overnight to allow the phage to diffuse out of the top agar and into the lambda diluent. The overlay solution was then carefully removed from each plate and combined in a sterile, 250 ml centrifuge bottle (Nalgene). Bacterial debris was pelleted by centrifugation in a Sorvall GSA rotor at 8,000 rpm and 4°C for 15 min. The supernatant was transferred to two 38.5 ml cellulose nitrate tubes (Beckman, No. 302237) and the phage were pelleted by centrifugation in a Beckman SW27 rotor at 23,000 rpm and 4°C for 90 min. The phage pellets were then resuspended in a total volume of 2 ml of lambda diluent and purified on a CsCl block gradient as described in Section 2.5.3.

2.5.3 Purification of λgt10

A CsCl block gradient was made up in an 11 ml polypropylene tube (Beckman, No. 331372) by adding 3 ml of 1.6 g/ml CsCl/PSB to the tube and then overlaying this consecutively with 3 ml of 1.5 and 1.4 g/ml CsCl/PSB without mixing the layers. The resuspended λgt10 (Section 2.5.2) was carefully layered onto the gradient and the gradient was centrifuged in a Beckman SW41 rotor at 22,000 rpm and 4°C for 2 h.

Following centrifugation, the phage particles, which were visible as a bluish-white band in the gradient, were recovered by piercing the bottom of the tube and collecting 0.5 ml fractions in 1.5 ml microfuge tubes. Fractions containing the phage were combined in an 11 ml Beckman tube and made up to 3 ml with 1.6 g/ml CsCl/PSB. The refractive index of this solution was then adjusted to 1.3995 (equivalent to 1.7 g/ml CsCl/PSB) by the addition of solid CsCl. The phage solution was overlaid consecutively with 2.5 ml of 1.6, 1.5 and 1.4 g/ml CsCl/PSB without allowing the layers to mix. This second gradient was then centrifuged in a Beckman SW41 rotor and the phage band recovered as above. Fractions containing the phage band were combined, transferred to a dialysis sac and dialysed at room temperature

against 1 litre of 10 mM NaCl, 50 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$ for 2× 1 h.

Requirements

- CsCl solutions: make up in 50 ml of PSB (Section 2.4.11) according to Table 2.1. Check the refractive index of each solution in a refractometer.
- Dialysis sac: prepare by rinsing a piece of dialysis membrane internally and externally with deionized H_2O.

Table 2.1. Cesium chloride solutions for λ step gradient.

Density (g/ml)	Refractive index	CsCl (g/50 ml PSB)
1.4	1.3722	28.5
1.5	1.3815	37.2
1.6	1.3905	47.2

2.5.4 Extraction of λgt10 DNA

Methods for the extraction of λgt10 DNA were based on those of Maniatis *et al.* (1982). The dialysed phage (Section 2.5.3) were transferred to a 15 ml, glass, centrifuge tube (Corex), the volume was determined and 0.5 M EDTA, 10% SDS (Section 2.22.6) and 20 mg/ml proteinase K were added to final concentrations of 20 mM, 0.5% and 50 µg/µl respectively. The solution was mixed gently by inversion and incubated at 65°C for 1 h. An equal volume of phenol (Section 2.22.1) was then added and the tube was mixed by inversion and centrifuged in a Sorvall SS-34 rotor at 4,500 rpm and room temperature for 5 min. The aqueous phase was re-extracted sequentially with an equal volume of phenol/chloroform (Section 2.22.2) and with an equal volume of chloroform (Section 2.22.3), with centrifugation as above. The phenol, phenol/chloroform and

chloroform phases were back-extracted sequentially with an equal volume of TE (Section 2.22.4). The aqueous phases were combined, transferred to a dialysis sac (Section 2.5.3) and dialysed over a period of 48 h at 4°C against 1 litre of TE (three changes). The DNA solution was then transferred to a 15 ml, glass, centrifuge tube (Corex) and the DNA was precipitated by the addition of 0.05 volume of 5 M NaCl and 2 vol. of absolute ethanol. The tube was centrifuged in a Sorvall SS-34 rotor at 10,000 rpm and 4°C for 15 min. The DNA pellet was then redissolved in 1 ml of TE and the concentration of DNA in solution was determined spectrophotometrically (Section 2.19).

2.6 Screening of white clover cDNA libraries

The screening of a library of recombinant phage involves a series of steps: (i) growth of recombinant phage on plates to produce plaques; (ii) production of filter replicas of the surfaces of the plates; (iii) denaturation and immobilization of phage DNA onto the filters; (iv) hybridization of the filters with radioactively labelled DNA probes; (v) identification and purification of positive recombinants. The methods used for screening the libraries were based on those of Benton and Davis (1977), Kaiser and Murray (1985), Vogeli and Kaytes (1987) and Wahl and Berger (1987).

2.6.1 Plating out cDNA libraries

An aliquot of each amplified cDNA library, containing approx. 1×10^4 pfu, was plated out with 1 ml of 2× concentrated *E. coli* C600ΔHfl plating bacteria (Sections 2.5 and 2.9.1) as described in Section 2.4.14. Adsorption was carried out in a sterile 50 ml centrifuge tube (Oak Ridge, Nalgene) and 20 ml of lambda top agarose (Section 2.4.12), pre-warmed to 55°C, was added to the tube. The phage were plated out on a 22.5 × 22.5 cm LB plate, pre-warmed to 45°C.

Requirements

- 22.5 x 22.5 cm LB plates: pour into 22.5 x 22.5 cm plates (Bio-Assay Dish, Nunc) as described in Section 2.4.14. Approx. 250 ml of medium is required per plate. Dry plates with lids removed for 1 h in a laminar flow cabinet.

2.6.2 Primary screening of cDNA library

A replica nitrocellulose filter was taken off the cDNA library plate (Section 2.6.1) as described in Section 2.6.8. The primary plaque hybridization was carried out as described in Section 2.6.9, using 50 ml of prehybridization solution per bag, with each bag containing one filter.

2.6.3 Secondary screening of positive cDNA isolates

The method used to screen positive cDNA/λgt10 isolates from the primary screen (Section 2.6.2) was based on that of Porteous (1986). A number of plaques in the vicinity of each positively hybridizing plaque (typically 4 - 6) was picked with a yellow pipette tip (200 µl) and the pipette tip was transferred to 100 µl of SM in a microtitre plate screening apparatus. A fresh lawn of *E. coli* C600 (Section 2.5) was prepared by mixing 100 µl of 2x concentrated cells (Section 2.9.1) with 3 ml of lambda top agarose and plating this on an 85 mm LB plate as described in Section 2.4.14. The array of tips from the screening apparatus was touched onto the surface of the plate and the plate was then inverted and incubated at 37°C overnight. A nitrocellulose filter replica was then taken off the plate as described in Section 2.6.8. Pairs of filters were prehybridized as described in Section 2.6.9, using 9 ml of prehybridization solution per bag. The prehybridization solution was then discarded and replaced with 9 ml of recycled hybridization solution from the primary screen (see Section 2.6.9). The filters were hybridized at

37°C overnight. Post-hybridization rinsing, autoradiography of the filters and identification of positive plaques were as described in Section 2.6.9. The microtitre plate containing the primary isolates was sealed and stored at 4°C.

Requirements

- Microtitre plate screening apparatus: this apparatus consists of one microtitre plate with the base of each well drilled out which is fixed above a second microtitre plate, each well of which contains 100 µl of PSB (Section 2.4.11). The distance between the two plates is arranged such that, when a yellow pipette tip (200 µl) is introduced through a hole in the top plate, the tip locates in the corresponding well in the bottom plate and is suspended in the PSB. For the subsequent plating onto 85 mm diameter plates, the outline of such a plate is marked on the top microtitre plate.

2.6.4 Plaque purification of isolates

The method used to purify positive cDNA/λgt10 isolates was based on the methods of Porteous (1986) and Carlock (1986). The isolates that showed up as positive at the secondary screening (Section 2.6.3) were transferred from the microtitre plate wells to individual 0.6 ml microfuge tubes. One drop of chloroform was added to each tube and the tubes were stored at 4°C until required.

A 10 µl aliquot of each positive isolate was transferred to 100 µl of PSB (Section 2.4.11) on a sheet of parafilm. Successive 10 µl dilutions were then made to four adjacent 100 µl aliquots of PSB. A 10 µl aliquot of each dilution was transferred to a fresh lawn of *E. coli* C600 (see Section 2.6.3). The plate was allowed to dry at room temperature and was then incubated overnight at 37°C. By that time, individual, well-isolated plaques were usually visible for

one of the dilutions. A nitrocellulose filter replica was taken off the plate as described in Section 2.6.8. Filters were prehybridized as described in Section 2.6.9, using 9 ml of prehybridization solution per bag. The prehybridization solution was then discarded and replaced with 9 ml of recycled hybridization solution from the primary screen (see Section 2.6.9). The filters were hybridized at 37°C overnight. Post-hybridization rinsing, autoradiography of the filters and identification of isolated, positive plaques were as described in Section 2.6.9.

2.6.5 Amplification of positive isolates

A plug of top agar containing a single, well-isolated, positive plaque was excised from the plaque purification plate (Section 2.6.4) and transferred to 100 µl of PSB (Section 2.4.11) in a 600 µl microfuge tube. One drop of chloroform was added and the tube was stored at 4°C. A 1 µl aliquot of the isolate was spotted, in duplicate, onto a fresh lawn of *E. coli* C600 (see Section 2.6.3). After overnight growth at 37°C, the plate was sealed and stored at 4°C.

2.6.6 Small-scale liquid lysates of positive isolates

A plug of top agar containing one of the amplified plaques was excised from the amplification plate (Section 2.6.5) and transferred to a 1.5 ml microfuge tube. A 100 µl aliquot of 2× concentrated *E. coli* C600 cells (Sections 2.5 and 2.9.1) was added and the tube was incubated at 37°C for 20 min. The adsorbed phage were then inoculated into 30 ml of LAM medium, in a 150 ml flask, and incubated overnight at 37°C with vigorous agitation.

Requirements

- LAM medium (per litre): 10 g Bacto-tryptone, 5 g yeast extract, 10 g NaCl, 1.2 g $MgSO_4$.

2.6.7 Purification of λ DNA from liquid lysates

The method used to purify cDNA/λgt10 DNA from liquid lysates was based on that of Kaslov (1986). To the overnight liquid lysate (Section 2.6.6), chloroform was added to 2%, DNase I and RNase A to 1 µg/ml, and solid NaCl to 1 M final concentration. The culture was incubated on a shaker at 37°C for 30 min. The aqueous phase was then transferred to a 50 ml centrifuge tube (Oak Ridge, Nalgene) and centrifuged in a Sorvall SS-34 rotor at 10,000 rpm and 4°C for 10 min. The supernatant was transferred to a 150 ml flask and solid PEG 6000 was added to 10% (w/v). The culture was gently shaken at room temperature until the PEG had completely dissolved and was then incubated on ice for at least 1 h. The culture was transferred to another 50 ml centrifuge tube and the tube was centrifuged as above for 20 min. The supernatant was then discarded and the tube was completely drained. The phage pellet was resuspended in 1 ml of PSB (Section 2.4.11), transferred to a 1.5 ml microfuge tube and centrifuged for 5 min. The supernatant was transferred to another tube, 0.5 ml of 30% (w/v) PEG, 2.5 M NaCl was added and the tube was centrifuged for 5 min. The phage pellet was resuspended in 200 µl of PSB and 10 µl of 10% SDS (Section 2.22.6), 8 µl of 0.5 M EDTA and 20 µl of 1 mg/ml proteinase K were added. The tube was incubated at 68°C for 30 min to remove the phage coat proteins and the phage DNA was then extracted sequentially with an equal volume of phenol (Section 2.22.1) and an equal volume of phenol/chloroform (Section 2.22.2). The DNA was precipitated by adding 100 µl of 7.5 M ammonium acetate and 750 µl of absolute ethanol and the tube was centrifuged for 15 min (see Section 2.21). The DNA pellet was rinsed sequentially with 1 ml of 70% ethanol and 1 ml of ether, with

a 1 min centrifugation at each rinse. The pellet was then air dried and redissolved in 50 µl of TE (Section 2.22.4). An aliquot (typically 2 µl) of the DNA solution was analyzed by agarose gel electrophoresis (Section 2.11.3).

2.6.8 Replica filters for plaque hybridization

Plates from which replica filters were to be taken were cooled to 4°C for at least 1 h to minimize the possibility of the top agar or agarose tearing. All manipulations of filters were with gloved hands or with a pair of non-serrated forceps. Nitrocellulose filters, either pre-cut sheets (Hybond-C, 0.45 µm, Amersham) or cut from a roll of nitrocellulose (BA 85, 0.45 µm, Schleicher and Schuell), were labelled with a ball-point pen. Large filters (22 × 22 cm) were completely wetted in distilled H_2O and then transferred to paper towels to remove excess H_2O. Small filters were used dry. Each filter was gently folded and laid onto a plate such that the surface of the plate was first touched in the centre of the plate with the filter along the fold and then the edges of the filter were also carefully lowered onto the plate surface. Any air bubbles between the filter and the plate surface were removed by gently rubbing the filter with a gloved finger. A series of holes was made asymmetrically around the perimeter of the filter with a syringe needle, through both the filter and the underlying agar, to enable the subsequent realignment of the filter with the plate. After 5 min, one edge of the filter was carefully lifted from the plate with a pair of forceps and then the filter was completely removed. The filter was placed plaque-side up on a sheet of Whatman 3MM paper, slightly larger then the nitrocellulose filter, saturated (but not overly wet) with denaturation solution. After 5 min, the filter was transferred to a sheet of 3MM paper saturated with neutralization solution for 5 min. The neutralization step was repeated and then the filter was rinsed briefly in 2× SSC, dried on a sheet of 3MM paper at room temperature and baked for 4 h at 68°C between

two sheets of 3MM paper. Filters were sealed, singly or in pairs with plaque-sides facing outwards, in a plastic bag with an impulse heat sealer, and stored at 4°C until required. The plates were sealed and stored at 4°C.

Requirements

- Denaturation solution: 0.5 M NaOH, 1.5 M NaCl.
- Neutralization solution: 0.5 M Tris-HCl, pH 8.0, 1.5 M NaCl.
- SSC (1×): 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0; see Section 2.22.5.

2.6.9 Plaque hybridization

Background radioactivity in plaque hybridization experiments was controlled by including heparin in the prehybridization and hybridization solutions as described by Singh and Jones (1984). One corner of the bag containing replica nitrocellulose filters was cut off and prehybridization solution was added to the bag, using a 10 ml pipette attached to a Pi-Pump for small volumes and a plastic funnel for large volumes. The bag was resealed and the filters were prehybridized for at least 2 h at 37°C with gentle shaking in an incubator. Another corner was then cut off the bag, the radioactively labelled, denatured, DNA probe (see Section 2.14.2) was introduced using an adjustable pipette and the bag was resealed. Alternatively, the prehybridization solution was replaced with recycled hybridization solution from a previous round of plaque hybridization. The filters were hybridized overnight as for the prehybridization step.

Following hybridization, the hybridization solution was removed from the bag and either discarded or transferred to a heat resistant, glass bottle (Schott) for storage at -20°C. The bag was cut along three sides, one side of the bag was peeled back with forceps and the filter(s) was lifted off and

transferred to a tray containing 2× SSC (500 ml for 22 × 22 cm filters, 250 ml for 150 mm diameter filters and 150 ml for 85 mm diameter filters). The filters were then rinsed sequentially in the same volume of 2× SSC, 0.1% SDS for 2× 15 min at 37°C and 0.1× SSC, 0.1% SDS for 2× 15 min at 37°C. This was followed by an optional rinse in 0.5× SSC, 1% SDS for 30 min at 65°C, depending on background levels, as indicated by a hand-held radiation monitor, as well as on the required stringency. The filters were then placed on paper towels to remove excess rinse solution and wrapped, while still damp, in polyethylene film. Each wrapped filter was taped to a piece of paper along its edges and transferred to an auto-radiography cassette. Working in a darkroom with appropriate safelighting, a sheet of X-ray film (Fuji RX or Kodak XAR-5) was cut to a size slightly larger than the filter and the film was then fixed over the filter, onto the underlying paper, with pieces of tape. The outline of the film was marked onto the underlying paper with a pen to allow the subsequent realignment of the film with the filter. An intensifying screen (Cronex Lightening Plus, DuPont) was placed over the film and the cassette was closed. The filters were autoradiographed at -70°C for 2 - 7 days and then the autoradiograph was developed (Section 2.6.10). The dried, developed film was aligned with the filter and the original plate/filter alignment marks were transferred to the film. The film was then aligned with the corresponding plate and any positively hybridizing plaques on the film were marked on the underside of the plate.

Requirements

- Prehybridization solution: 40% or 50% (v/v) formamide (depending on the hybridization stringency required), 5× SSC, 0.5% SDS, 100 µg/ml heparin. Filter the solution through a 0.45 µm cellulose acetate filter and store at -20°C. Warm to 37°C before use.

- Recycled hybridization solution: thaw the hybridization solution from a previous round of plaque hybridization at 37°C. Immediately before use, heat an appropriate volume in a heat resistant glass bottle at 80°C for 5 min and then chill in iced water.

2.6.10 Development of autoradiographs

The autoradiography cassette was removed from the -70°C freezer and allowed to warm to room temperature if necessary and then transferred to a darkroom. Using an appropriate safelight, the X-ray film was removed from the cassette and transferred sequentially to a tray of developer solution (G150, Agfa) for 3 min, stop solution (2% v/v acetic acid) for 30 sec and fixer solution (G334, Agfa), 2.5% (v/v) hardener (Hypan Hardener, Ilford) for 6 min. The film was then rinsed extensively with running water and dried. The developer and fixer solutions were made up according to the manufacturer's instructions.

2.7 Isolation of DNA from white clover

During the course of this study, a variety of methods for isolating total DNA from white clover tissues was tried (Dellaporta *et al.*, 1983a, 1983b and 1985; Hattori *et al.*, 1987; Lichtenstein and Draper, 1985; Murray and Thompson, 1980; Sutton, 1974). The Dellaporta *et al.* (1983a; 1983b; 1985) and the Hattori *et al.* (1987) methods consistently produced DNA that could be easily digested with restriction enzymes, as well as being of a size suitable for the construction of a genomic DNA library. Modifications of these two methods, both involving the isolation of DNA from freeze dried seedlings, are outlined below. The former method (Section 2.7.2) involves the use of potassium acetate to precipitate proteins and polysaccharides from lysed cells and the subsequent precipitation of DNA with ethanol. The latter method (Section 2.7.3) involves the use of phenol to denature

the proteins of lysed cells, insoluble polyvinylpolypyrrolidone (PVP) to remove phenolic compounds and the subsequent precipitation of DNA with polyethylene glycol (PEG). For both methods, the DNA was further purified by two rounds of CsCl/ethidium bromide gradient centrifugation.

2.7.1 Freeze drying white clover seedlings

White clover seeds (10 g) were germinated in a sterile 250 ml glass bottle as described in Section 2.1.1. The seedlings were grown for 7 days with occasional shaking to prevent clumping. They were then transferred to a sterile 250 ml freeze drier container (Virtis). The seedlings were freeze dried for approx. 48 h.

2.7.2 DNA isolation - Method 1

This method was based on the plant DNA minipreparation procedure of Dellaporta *et al.* (1983a; 1983b; 1985) with modifications to make it suitable for the isolation of larger amounts of DNA. Freeze dried material (1 g; Section 2.7.1) was mixed with 4 g of alumina (Sigma, Type A5) and ground to a fine powder with a mortar and pestle. The mixture was transferred to a 125 ml conical flask and 60 ml of Dellaporta extraction buffer was added. The material was mixed gently on a magnetic stirrer until it was completely resuspended and it was then transferred to four 50 ml centrifuge tubes (Nalgene, Oak Ridge). SDS (2ml of a 10% stock; see Section 2.22.6) was added to each tube, the contents were mixed by inversion and the tubes were incubated at 65°C for 10 min. Potassium acetate (5 M, 5 ml) was then added, the contents were mixed again and the tubes were incubated on ice for 20 min. Most proteins and polysaccharides were precipitated as a complex with the insoluble potassium dodecyl sulphate at this stage and were removed by centrifugation in a Sorvall SS-34 rotor at 16,000 rpm and 4°C for 20 min. The supernatants were gently poured off and the pellets were dried by inverting the

tubes on paper towels for 10 min. Each pellet was then redissolved in 700 µl of 50 mM Tris-HCl, pH 8.0, 10 mM EDTA, transferred to a 1.5 ml microfuge tube and the tube was centrifuged for 10 min to pellet insoluble material. The supernatants were combined in a siliconized, 15 ml, glass, centrifuge tube (Corex) and 300 µl of 3 M sodium acetate-/acetic acid, pH 5.2 and 2 ml of isopropanol were added. The precipitated DNA was pelleted by centrifugation at 10,000 rpm for 5 min. The supernatant was then discarded, 5 ml of 70% ethanol was added to the tube and the tube was recentrifuged for 5 min. The supernatant was again discarded and the DNA pellet was dried for 5 min under vacuum and then redissolved in 4 ml of 4.5 M CsCl solution. This solution was transferred to a 5 ml polyallomer centrifuge tube (Quick-Seal, Beckman, No. 342412) and the tube was sealed. The DNA was further purified on CsCl gradients as described in Section 2.7.4.

Requirements

- Dellaporta extraction buffer: 100 mM Tris-HCl, pH 8.0, 50 mM EDTA, 500 mM NaCl. Autoclave for 20 min and then add 2-mercaptoethanol to 10 mM.
- 4.5 M CsCl: 4.5 M CsCl, 50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 200 µg/ml ethidium bromide. Check the refractive index in a refractometer and adjust to 1.390.

2.7.3 DNA isolation - Method 2

This method was based on that of Hattori *et al.* (1987). Freeze dried plant material (3 g; Section 2.6.1) was mixed with 10 g of alumina (Sigma, Type A5) and ground to a fine powder with a mortar and pestle. Hattori extraction buffer was then slowly added to a final volume of 200 ml with constant, gentle mixing with a glass rod. The mixture was then transferred to two 250 ml polypropylene centrifuge bottles (Nalgene). Phenol (40 ml; Section 2.22.1) was added to each centrifuge bottle and the contents were mixed gently

by inversion. The centrifuge bottles were incubated on ice for 10 min with intermittent mixing and then centrifuged in a Sorvall GSA rotor at 7,000 rpm and 4°C for 10 min to separate the phases. The aqueous phase from each bottle was transferred to another centrifuge bottle, using a 10 ml glass pipette inverted in a 10 ml Pi-Pump. Each aqueous phase was then extracted sequentially with 40 ml of phenol (Section 2.22.1), 40 ml of phenol/chloroform (Section 2.22.2) and 40 ml of chloroform (Section 2.22.3), with centrifugation at each step as above. The phenol, phenol/chloroform and chloroform phases were sequentially back-extracted with 20 ml of Hattori extraction buffer. All of the aqueous phases were combined in a 1 litre flask and mixed with PVP (slurry from 2 g dry weight; see below) for 5 min at room temperature. The mixture was then transferred to 250 ml centrifuge bottles and centrifuged in a Sorvall GSA rotor at 7,000 rpm and 4°C for 10 min to pellet the PVP. The supernatants were filtered through Whatman number 1 filter paper and combined in one 250 ml centrifuge bottle. NaCl and PEG 6000 were added to final concentrations of 0.5 M and 10% (w/v) respectively and allowed to dissolve with gentle shaking.

Following overnight incubation on ice to allow precipitation of DNA to occur, the DNA was pelleted by centrifugation as above but at 8,000 rpm. The supernatant was discarded, the bottle was recentrifuged briefly, and residual supernatant was removed. The DNA pellet was redissolved in 16 ml of TE (Section 2.22.4) with gentle shaking at 37°C for 1 h followed by a 5 min incubation at 65°C. The DNA solution was then transferred to a 50 ml centrifuge tube (Oak Ridge, Nalgene) and centrifuged in a Sorvall SS-34 rotor at 10,000 rpm and 4°C for 10 min to pellet insoluble material. The supernatant was transferred to a clean 50 ml centrifuge tube and 15 g of CsCl and 340 µl of 10 mg/ml ethidium bromide were added. The refractive index of the solution was checked in a refractometer and adjusted to a refractive index of 1.390. This produced a cloudy solution which was centrifuged at

10,000 rpm and 4°C for 10 min. The clear supernatant was transferred to 5 ml polyallomer centrifuge tubes (Quick-Seal, Beckman, No. 342412) and the tubes were sealed. The DNA was further purified on CsCl gradients as described in Section 2.7.4.

Requirements

- Hattori extraction buffer: 0.1 M Tris-HCl, pH 8.0, 0.05 M EDTA. This solution was autoclaved for 20 min and then SDS was added to 1% from a 10% stock (Section 2.22.6).
- PVP (Polyvinylpolypyrrolidone; Sigma): this was washed with 0.1 M HCl, neutralized with 50 mM Tris-HCl pH 8.0, and filtered through Whatman number 1 filter paper to obtain a slurry.

2.7.4 Purification of plant DNA on CsCl gradients

Sealed tubes containing DNA/CsCl/ethidium bromide solution (Sections 2.7.2 and 2.7.3) were centrifuged in a VTi65 rotor at 55,000 rpm and 20°C for 16 h. Using proper eye protection, the tubes were then examined in a darkroom under long wavelength u.v. illumination (366 nm) from a hand-held u.v. light source (Model UVL-56, UVP Inc.). The top of each tube was punctured with a syringe needle and the DNA band in each tube was removed by side puncture of the tube using a 2 ml syringe fitted with an 18 gauge needle. The DNA was transferred to two new centrifuge tubes, the tubes were filled with 4.5 M CsCl (Section 2.7.2), sealed and centrifuged as above. The DNA bands were then removed from the tubes as above and combined in a 15 ml, glass, centrifuge tube (Corex). The ethidium bromide was removed by partitioning five times against NaCl-saturated isopropanol. The DNA solution was then transferred to a dialysis sac (see Section 2.5.3) and the CsCl was removed by dialysis against two, four litre, volumes of TE (Section 2.22.4) at 4°C over a period of 24 h. The

volume of the dialysed DNA was determined and the DNA concentration was determined spectrophotometrically (Section 2.20). The DNA solution was then distributed in 400 µl aliquots into 1.5 ml microfuge tubes and 40 µl of 3 M sodium acetate/acetic acid, pH 5.2 and 900 µl of absolute ethanol were added to each tube. The tubes were stored at -20°C. Before use, the DNA was pelleted by centrifugation in a microfuge for 10 min at room temperature and redissolved in sterile H_2O at a concentration of 1 µg/µl.

2.8 Construction of white clover genomic library

The methods involved in the construction of a white clover genomic library in the bacteriophage lambda replacement vector, λEMBL3, were based on the methods of Frischauf *et al.*, (1983), Kaiser and Murray (1985) and Promega Technical Bulletin Number 037. The procedure involves the fractionation of white clover DNA by partial digestion with *Sau*3A and the subsequent cloning of this fractionated DNA into λEMBL3 arms generated by *Bam*HI digestion.

2.8.1 Optimization of *Sau*3A partial digestion

The optimum size range for cloning into λEMBL3 is 18 - 22 kb. The optimum concentration of *Sau3*A restriction enzyme required to generate white clover DNA fragments within this size range was first determined in a set of small-scale *Sau3*A digestions.

A 10 µl aliquot of white clover DNA (10 µg; Section 2.6.4) was mixed with 10 µl of 10× H buffer (Section 2.19) and adjusted to 100 µl with sterile, distilled H_2O. A 20 µl aliquot of this solution was transferred to a 1.5 ml microfuge tube (tube 1) and 10 µl aliquots were transferred to additional tubes (tubes 2 - 9). All tubes were incubated on ice. Immediately before use, the *Sau*3A enzyme was diluted

to 1 U/µl in enzyme dilution buffer. A 1 µl aliquot of the diluted enzyme was added to tube 1 (0.5 U/µg of DNA) and the contents were mixed. A 10 µl aliquot was then transferred from tube 1 to tube 2. The serial dilution was continued to tube 8, with the contents of each tube being mixed thoroughly at each dilution. Tube 9 was used as an undigested control. All tubes were then incubated in a 37°C water bath for 1 h. Each digestion was stopped by adding EDTA to 20 mM from a 0.5 M stock. All the samples were cooled on ice and then analyzed by electrophoresis on a 0.3% agarose gel (Section 2.11.3). The agarose gel was poured in a cold room and, once the gel was hardened, it was overlaid with electrophoresis buffer (Section 2.11.3) and the comb was removed.

Requirements

- Enzyme dilution buffer: 50% (v/v) glycerol, 500 µg/ml BSA (for 50 mg/ml stock, see Section 2.4.3), 10 mM Tris-HCl, pH 7.4, 0.1 mM EDTA, 50 mM KCl.
- *Sau3*A restriction enzyme at a concentration of 5 U/µl was obtained from New England Biolabs.

2.8.2 Large-scale partial *Sau*3A digestion

The redissolved white clover DNA (Section 2.7.4) was distributed into three 1.5 ml microfuge tubes such that each tube contained 100 µg (100 µl) of DNA (tubes 1, 2 and 3). A 100 µl aliquot of 10× H (Section 2.19) and 800 µl of sterile distilled H_2O were added to each tube and the tubes were incubated at 37°C. Dilutions of *Sau3*A were made in enzyme dilution buffer (Section 2.8.1) to give enzyme concentrations of 0.64, 0.32 and 0.16 U/µl. A 5µl aliquot of 0.64 U/µl *Sau3*A was added to tube 1, 5 µl of 0.32 U/µl *Sau*3A was added to tube 2, and 5 µl of 0.12 U/µl *Sau*3A was added to tube 3. The tubes were then incubated at 37°C for 1 h. The digestions were stopped by adding EDTA to 20 mM. A 10 µl aliquot of each digest was analyzed by electrophoresis on a 0.3% agarose gel

(see Section 2.8.1) to check that the correct DNA size distribution had been obtained. The digestions were then extracted twice with equal volumes of phenol/chloroform (Section 2.22.2) and the aqueous phases were combined in an acid-washed, siliconized, 15 ml, glass, centrifuge tube (Corex). The digested DNA was precipitated by the addition of 0.1 volume of 3 M sodium acetate/acetic acid, pH 5.2 and 2 vol. of ethanol. The tube was incubated overnight at -20°C and then centrifuged in a Sorvall HB4 rotor at 10,000 rpm and 4°C for 20 min. The DNA pellet was rinsed with 70% ethanol and the tube was recentrifuged for 5 min. The DNA was then redissolved in 400 µl of sterile TE (Section 2.22.4) with heating at 65°C for 10 min to ensure that the DNA was completely dissolved. The DNA solution was transferred to a 1.5 ml microfuge tube and stored at 4°C.

2.8.3 Size fractionation of DNA in a velocity gradient

Two 5 - 25% (w/v) NaCl gradients were poured in 12.5 ml centrifuge tubes (Beckman, No. 331372), using a standard gradient former and pouring at approx. 1 ml per minute. A 200 µl aliquot of *Sau*3A digested white clover DNA (Section 2.8.2) was then layered onto the top of each gradient. The tubes were centrifuged in a Beckman SW41 rotor at 37,000 rpm and room temperature for 4.5 h. The bottom of each tube was then pierced and 0.25 ml fractions were collected in 1.5 ml microfuge tubes. A 25 µl aliquot was taken from each alternate fraction and mixed with 75 µl of TE (Section 2.22.4) and 10 µl of 10× sample loading buffer (Section 2.11.3). These samples were analyzed by electrophoresis on a 0.3% agarose gel (see Section 2.8.1) to determine which fractions contained DNA of the required size-range. DNA was precipitated from these fractions by adding 250 ml of H_2O and 1 ml of ethanol to each tube and incubating the tubes at -20°C overnight. The tubes were then centrifuged in a Sorvall SS-34 rotor (with microfuge tube adaptor, Sorvall, No. 00381) at

15,000 rpm and 4°C for 20 min. The DNA pellets were rinsed twice with 70% ethanol and dried for 5 min under vacuum. Each DNA pellet was redissolved in 20 µl of sterile H_2O and the DNA solutions derived from equivalent fractions of the two gradients were combined. The concentration of DNA in each solution was determined by electrophoresis of a 2 µl aliquot on an agarose gel (Section 2.11.3) with known amounts of undigested λEMBL3 DNA.

2.8.4 Ligation of size fractionated DNA with λEMBL3 arms

The ligation of size fractionated white clover DNA with λEMBL3 *Bam*HI arms was carried out as follows (8 reactions):

λEMBL3 *Bam*HI arms	1.2 µl
18 - 22 kb white clover DNA	2.3 µl
10× ligase buffer	0.5 µl
T4 DNA ligase	1.0 µl
TOTAL	5.0 µl

The ligation reactions were incubated at room temperature for 3 h and then transferred to 4°C overnight.

Requirements

- λEMBL3 *Bam*HI arms at a concentration of 0.5 µg/µl were obtained from Promega.
- 18 - 22 kb white clover DNA: fractions spanning the optimum insert length were combined (see Section 2.8.3); the DNA concentration was 0.2 µg/µl.
- 10× ligase buffer: 400 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$, 100 mM DTT (from 1 M stock; Section 2.14.1), 10 mM ATP (from 0.1 M stock; Section 2.4.4), 500 µg/ml BSA. The buffer was freshly made.

- T4 DNA ligase at a concentration of 400 U/µl (cohesive end ligation unit; one cohesive end ligation unit equals 0.0025 Weiss unit) was obtained from New England Biolabs.
- BSA: Section 2.4.3.

2.8.5 *In vitro* packaging of ligated DNA

The methods involved in packaging the ligated DNA (Section 2.8.4) are described in Section 2.4.11. One packaging extract (Packagene extract system, Promega) was used for each ligation. The packaged phage (white clover, unamplified gene library) were combined and stored at 4°C over chloroform.

2.8.6 Assay of gene library

Appropriate dilutions were made in PSB (Section 2.4.11) and 10 µl of each dilution was plated out with 100 µl of *E. coli* MB406 plating bacteria (Section 2.9.2) as described in Section 2.4.14.

Requirements

- *E. coli* MB406 (Promega Notes No. 11): *sup*E, *rec*B21, *rec*C22, *sbc*B15, *hfl*A, *hfl*B, *hsd*R$^-$.

2.8.7 Amplification of gene library

The libraries of recombinant phage (Section 2.8.5) were amplified by preparing plate stocks directly from the packaging mixtures. The phage were plated out as described in Section 2.4.14 except that 1 ml aliquots of the combined packaging mixture were adsorbed with 8 ml of *E. coli* MB406 plating bacteria (Section 2.9.2) in 50 ml centrifuge tubes (Oak Ridge, Nalgene) and plated out with 40 ml of lambda top agar on 22.5 × 22.5 cm plates (see Section 2.6.2), pre-warmed to 45°C. The plates were incubated, right side up, overnight at 37°C. The plates were then overlaid with 50 ml of PSB

(Section 2.4.11) and incubated overnight at 4°C with gentle shaking. The overlay solution was transferred to a sterile 250 ml bottle (Schott), chloroform was added to 5% and the bottle was stored at 4°C. For long term storage, dimethylsulphoxide was added to the aqueous phase to a final concentration of 7% (v/v) and the phage were stored at -70°C. The amplified library was assayed as described in Section 2.4.12 for cDNA libraries.

2.9 Preparation of plating bacteria

Two methods were used for preparing *E. coli* cells for plating out with λ bacteriophage. Either a 2× concentrated cell suspension (Davis *et al.* 1980) or a non-concentrated cell suspension (Promega Technical Bulletin number 037) was used.

2.9.1 2× concentrated plating bacteria

A single colony of the required *E. coli* strain, grown on a fresh LB plate (Section 2.4.14), was used to inoculate 20 ml of maltose supplemented LB in a 125 ml flask. Following overnight incubation at 37°C with vigorous shaking, the culture was transferred to a 50 ml centrifuge tube (Oak Ridge, Nalgene). The cells were pelleted by centrifugation in a Sorvall SS-34 rotor at 8,000 rpm and 4°C for 5 min and then resuspended in 10 ml of 0.01 M $MgSO_4$. The resuspended cells were stored at 4°C for up to 4 weeks but fresh cells were used when optimum phage plating efficiencies were required.

Requirements

- Maltose supplemented LB (per litre): 10 g bacto-tryptone, 5 g yeast extract, 10 g NaCl. Adjust pH to 7.5 with NaOH. Autoclave for 25 min, cool and then add maltose to 0.2% from a sterile 20% stock (autoclave).

2.9.2 Non-concentrated plating bacteria

Maltose supplemented LB (10 ml) was inoculated as described in Section 2.9.1, in a loosely capped 50 ml centrifuge tube (Oak Ridge, Nalgene). The culture was grown overnight at 37°C with vigorous shaking and then transferred directly to 4°C. The cell suspension was used within one week.

2.10 Screening of white clover gene library

For background relating to the screening of recombinant λ libraries, see Section 2.6.

2.10.1 Plating out the gene library

Aliquots of the amplified gene library were plated out on six 22.5 × 22.5 cm plates to give approx. 2×10^5 plaques per plate, (see Section 2.8.7) as described in Sections 2.4.14 and 2.8.7 except that 3 ml of a fresh overnight culture of *E. coli* MB406 (Sections 2.8.6 and 2.9.2) and 30 ml of TB top agarose (Section 2.10.4) were used. The plates were inverted and incubated at 37°C for approx. 11 h, by which time confluent lysis was apparent, and then transferred to 4°C.

2.10.3 Primary plaque hybridization

A replica nitrocellulose filter was taken off each of the gene library plates (Section 2.10.1) as described in Section 2.6.8. The primary plaque hybridization was carried out as described in Section 2.6.9.

2.10.4 Plaque purification of positive isolates

A plug of top agarose, approx. 0.5 cm in diameter, was excised from the primary library plate at each mark, corresponding to a positive hybridization signal, (Section 2.10.3)

and transferred to a 1.5 ml microfuge tube containing 500 µl of PSB (Section 2.4.11). Chloroform (20 µl) was added to each tube and the tubes were stored at 4°C. The isolates were assayed, as described in Section 2.8.6, by making appropriate dilutions of each isolate (typically 10^{-4} or 10^{-5} dilutions) and plating out 10 µl of each dilution with 100 µl of *E. coli* MB406 (Sections 2.8.6 and 2.9.2) on TB plates. An aliquot of each isolate containing approx. 5000 pfu was then plated out with 300 µl of *E. coli* MB406 and 11 ml of TB top agarose on 150 mm diameter TB plates as described in Section 2.4.14. Filter replicas were made of each plate as described in Section 2.6.8. Prehybridization and hybridization were as described in Section 2.6.9 except that 20 ml of prehybridization solution was used per bag and this solution was replaced with recycled hybridization solution from the primary screen (Section 2.10.3). The filters were hybridized at 37°C overnight. Post-hybridization rinsing, autoradiography of the filters and identification of positive plaques were as described in Sections 2.6.9 and 2.6.10.

One further round of plaque purification was carried out as above with minor modifications. A volume of an appropriate dilution of each secondary isolate was plated out with 100 µl of *E. coli* MB406 (Sections 2.8.6 and 2.9.2) and 3 ml of TB top agarose on 85 mm TB plates to give approx. 150 pfu per plate (typically 5 - 20 µl of a 10^{-4} dilution). Replica nitrocellulose filters (BA 85, 0.45 µm, Schleicher and Schuell) were prehybridized in pairs with 10 ml of prehybridization solution per bag and then hybridized with 10 ml of recycled hybridization solution (see Section 2.6.9). At a density of 150 pfu per 85 mm plate, it was possible to excise individual, well-isolated, positive plaques. Each plaque was transferred to a microfuge tube containing 100 µl PSB (Section 2.4.11), one drop of chloroform was added and the isolates were stored at 4°C.

Requirements

- TB plates (per litre): 10 g Bacto-tryptone, 5 g of NaCl, 15 g of agar. Autoclave for 25 min, cool to 60°C and then add 10 ml of sterile 1 M $MgSO_4$. Pour as described in Section 2.4.14 for LB plates.
- TB top agarose (200 ml): 2.0 g bacto-tryptone, 1.0 g NaCl, 1.6 g agarose (Ultra Pure, BRL). Autoclave in a 200 ml bottle (Schott) for 25 min, cool to 60°C and then add 2 ml of sterile 1 M $MgSO_4$. Store at room temperature. Before use, melt agarose in a microwave oven (loosen top) and cool to 50°C in a waterbath.

2.10.5 Plate stocks of positive isolates

The method used to prepare phage stocks from positive isolates was based on that of Miller (1987). A 10 µl aliquot of each isolate was plated out with 100 µl of *E. coli* MB406 (Sections 2.8.6 and 2.9.2) as described in Section 2.4.14 except that 2.5 ml of TB medium and 2.5 ml of TB top agar, both equilibrated at 45°C, were added to the tube and the phage were plated out on TB plates (Section 2.10.4). The plates were incubated, without being inverted, at 37°C for approx. 4 - 5 h until lysis was first apparent. The top agar layer was then scraped off the plate, with a flat spatula, into a 15 ml glass centrifuge tube (Corex). Two drops of chloroform were added to complete the lysis of the host *E. coli* cells. The tube was briefly vortexed and then centrifuged in a Sorvall SS-34 rotor at 10,000 rpm and 4°C for 10 min. The supernatant containing the phage was transferred to a 10 ml screw-capped tube (Kimax) and stored at 4°C. The plate stocks were assayed (Section 2.4.14) by plating out 10 µl of appropriate dilutions (typically 10^{-6} and 10^{-7} dilutions) with 100 µl of *E. coli* LE392 (Section 2.9.2) and 3ml of TB top agar on TB plates (Section 2.10.4).

Requirements

- TB medium (per litre): 10 g Bacto-tryptone, 5 g NaCl. Autoclave for 25 min and then add 10 ml of sterile 1 M $MgSO_4$.
- TB top agar (200 ml): as for TB top agarose (Section 2.10.4) but with 1.6 g of agar instead of agarose.
- *E. coli* LE392 (Murray *et al.*, 1977): F^-, *hsd*R574 (r_K^-, m_K^+), *sup*E44, *sup*F58, *lac*Y1, or Δ(*lac*1ZY)6, *gal*K2, *gal*T22, *met*B1, *trp*R55, λ^-.

2.10.6 Large-scale liquid lysates of positive isolates

Large-scale liquid lysates of positive isolates were prepared according to the method of Miller (1987) with minor modifications. A single colony of *E. coli* MB406 (Section 2.8.6) was inoculated into 250 ml of maltose supplemented LB (Section 2.9.1), in a 1 litre flask, and incubated overnight at 37°C with vigorous shaking. The A_{550} of the overnight culture was measured and an appropriate volume of the culture was inoculated into 500 ml of LB medium (Section 2.15.8), in a 2 litre flask, such that the initial A_{550} was less than 0.2. The bacteria were grown until an A_{550} of 0.2 was reached and then an aliquot of the phage isolate equivalent to 5×10^5 pfu, based on the assay of each stock (Section 2.10.5), was added to the flask. The flask was incubated overnight at 37°C with vigorous shaking. Chloroform (3 ml) was then added to the flask, whether or not the culture was already lysed. The flask was stored at 4°C until the culture could be further processed (Section 2.10.7).

2.10.7 Isolation of phage DNA from large-scale liquid lysate

The methods used to purify phage from a large-scale liquid lysate and to isolate DNA from the purified phage were based

on the methods of Ziai *et al.* (1988). The phage liquid lysate (Section 2.10.6) was centrifuged in a Sorvall GSA rotor at 8,000 rpm and 4°C for 10 min to pellet bacterial debris. The supernatant was then transferred to a 1 liter flask. NaCl to a concentration of 0.8 M and solid PEG 6000 to 10% (w/v) were added and the flask was shaken at room temperature to dissolve the PEG. The lysate was left at 4°C overnight and then centrifuged at 5,000 rpm and 4°C for 15 min. The pellets were resuspended in a total volume of 30 ml of phage buffer, transferred to 50 ml centrifuge tubes (Oak Ridge, Nalgene) and an equal volume of saturated ammonium acetate was added to each tube. The tubes were incubated on ice for 1 h and then centrifuged in a Sorvall SS-34 rotor at 15,000 rpm and 4°C for 20 min. The supernatants were discarded and the phage pellets were resuspended in a total volume of 5 ml of TE (Section 2.22.4). The phage suspension was transferred to a 15 ml glass centrifuge tube (Corex) and SDS was added to a concentration of 0.1%. The suspension was then incubated sequentially with RNase A at a final concentration of 20 µg/ml at room temperature for 30 min and proteinase K at a final concentration of 200 µg/ml at 60°C for 20 min to lyse the phage. The phage lysate was then transferred to ice for 5 min, mixed with 500 µl of 1 M NaOH and incubated on ice for a further 10 min. The phage lysate was then neutralized by adding 2 ml of 10 M ammonium acetate/acetic acid, pH 6.0, and centrifuged in a Sorvall SS-34 rotor at 10,000 rpm and 4°C for 10 min. The supernatant was extracted twice with an equal volume of phenol (Section 2.22.1) and once with an equal volume of chloroform (Section 2.22.3). The final aqueous phase was transferred to a 15ml glass centrifuge tube (Corex) and mixed with an equal volume of isopropanol.

Following overnight incubation at -20°C to precipitate the phage DNA, the tube was centrifuged in a Sorvall SS-34 rotor at 10,000 rpm and 4°C for 20 min. The DNA pellet was then rinsed twice with cold 70% ethanol, dried under vacuum for 5

min, and redissolved in 500 µl of H_2O. The DNA solution was stored at -20°C until required.

Requirements

- Phage buffer: 20 mM Tris-HCl, pH 7.4, 100 mM NaCl, 10 mM $MgSO_4$.

2.10.8 Small-scale liquid lysates of positive isolates

A small-scale culture of each positive library isolate (Section 2.10.5) was grown by inoculating 20 ml of TB medium (Section 2.10.5), in a 150 ml flask, with 1 ml of *E. coli* MB406 (Sections 2.8.6 and 2.9.2) which had been preadsorbed with 10 µl of each isolate (see Section 2.4.14). The flask was incubated overnight at 37°C with vigorous shaking. Two drops of chloroform were then added to the culture and the flask was stored at 4°C until the culture could be further processed (Section 2.10.9).

2.10.9 Isolation of phage DNA from small-scale liquid lysate

The methods used to purify phage from a small-scale liquid lysate of a λEMBL3 recombinant (Section 2.10.8) with DEAE-cellulose and to isolate DNA from the purified phage were based on the methods of Pohl (1988) and Ziai *et al.* (1988). The lysate was transferred to a 50 ml centrifuge tube (Oak Ridge, Nalgene), an equal volume of resuspended DEAE cellulose was added and the tube was inverted several times to mix the contents. The tube was then centrifuged in a Sorvall SS-34 rotor at 15,000 rpm and 4°C for 10 min to pellet bacterial debris as well as the bacterial DNA and RNA bound to the DEAE cellulose. The supernatant, containing the phage particles, was transferred to a fresh tube and the phage were precipitated by adding NaCl to 1 M and PEG 6000 to 10%. The

tube was incubated on ice for 1 h and then centrifuged in a Sorvall SS-34 rotor at 10,000 rpm and 4°C for 10 min. The phage pellet was subsequently resuspended in 200 µl of TE (Section 2.22.4) and transferred to a 1.5 ml microfuge tube. SDS and proteinase K were added to 0.1% and 200 µg/ml respectively and the tube was incubated at 60°C for 20 min and then transferred to ice. Ammonium acetate/acetic acid, pH 6.0 (7.5 M, 120 µl), was added and the tube was centrifuged for 10 min. The supernatant was subsequently extracted once with an equal volume of phenol (Section 2.22.1) and once with an equal volume of chloroform (Section 2.22.3). The aqueous phase was then mixed with 2 vol. of absolute ethanol and the DNA was pelleted by centrifugation for 15 min. The DNA pellet was subsequently rinsed with 1 ml of cold 70% ethanol with centrifugation for 5 min. The DNA was finally redissolved in 50 µl H_2O. A 2 µl aliquot of the DNA preparation was sufficient for restriction enzyme digestion and analysis by agarose gel electrophoresis (Section 2.19).

Requirements

- Resuspended DEAE-cellulose: suspend 10 g DEAE-cellulose (Whatman, DE52, pre-swollen) in 200 ml of 0.05 M HCl. Neutralize the suspension with 400 µl of 10 M NaOH and allow the DEAE-cellulose particles to settle. Decant and discard the supernatant. Wash the DEAE-cellulose 3 times with 1 litre of saline and resuspend so that the final suspension consists of 75% (v/v) DEAE-cellulose and 25% saline. Add ampicillin to 200 µg/ml and store at 4°C.
- Saline: 0.9% (w/v) NaCl.

2.11 Agarose gel electrophoresis

Agarose gel electrophoresis was initially carried out in an apparatus purpose-built to accomodate the lid of a microtitre plate (127 mm × 85 mm × 4 mm), in either orientation. Subsequently, a commercial unit giving a gel size of 10 mm

(length) × 15 mm (width) was used (Wide Mini-Sub Cell, Bio-Rad). The comb size used for each gel depended upon the sample volume to be subsequently loaded on the gel and the number of samples to be electrophoresed.

2.11.1 Alkaline agarose gel electrophoresis

Alkaline agarose gels for the analysis of the first and second strand cDNA synthesis reaction products (Sections 2.4.1 and 2.4.2) were prepared and electrophoresed as per Maniatis *et al.* (1982). Immediately before use, 0.2 vol. of 5× alkaline loading buffer was added to each DNA sample. Equal counts of each sample, up to a maximum of 20 µl per well, were loaded on a 1.4% alkaline agarose gel. The gel was electrophoresed in alkaline electrophoresis buffer, at approx. 2.5 V/cm for 5 h or at approx. 1 V/cm overnight, with constant buffer recirculation.

Following electrophoresis, the gel was soaked in 200 ml of 7% trichloroacetic acid in H_2O for 30 min at room temperature. The gel was then dried in a vacuum drier (Bio-Rad) at 60°C for 30 min, wrapped with polyethylene film, and autoradiographed at either room temperature or -70°C in an autoradiography cassette. Typically, autoradiography was carried out for from 2 h (room temperature) to overnight (-70°C). The autoradiograph was developed as described in Section 2.6.10.

Requirements

- 5× alkaline loading buffer: 250 mM NaOH, 5 mM EDTA, 25% Ficoll (type 400), 0.125% bromocresol green.
- Alkaline electrophoresis buffer: 30 mM NaOH, 1 mM EDTA.
- 1.4% alkaline agarose gel: add 0.98 g agarose to 70 ml of 50 mM NaCl, 1 mM EDTA and melt in a microwave oven (see Requirements, Section 2.11.3). Cool to approx. 60°C and pour into a gel tray with a comb in place. Soak the

hardened gel in alkaline electrophoresis buffer for at least 30 min before use.

2.11.2 Gel electrophoresis of RNA

Poly(A)$^+$ selected RNA species (Section 2.3.1) were separated by electrophoresis on an agarose gel with 2.2 M formaldehyde in the presence of 50% formamide. The method was based on that of Gerard and Miller (1986).

Aliquots containing approx. 2 µg of poly(A)$^+$-enriched white clover RNA (Section 2.3.1) were centrifuged in 1.5 ml microfuge tubes for 15 min. The supernatants were then discarded and the RNA pellets were rinsed twice with 70% ethanol, dried under vacuum for 5 min and redissolved in 2.2 µl of Buffer A. Formaldehyde/formamide (4.8 µl; see below) was added to each sample and the samples were heated at 70°C for 10 min. The samples were then transferred to ice and 1.5 µl of Gel Loading Buffer and 1 µl of 1 mg/ml ethidium bromide (see Section 2.11.3 for 10 mg/ml; dilute 1/10 with H_2O) were added. The samples were electrophoresed on a 1.4% agarose gel in Electrophoresis Buffer at 60 V until the leading dye had migrated approx. one third of the length of the gel. Halfway through the electrophoresis run, the buffer in the two chambers of the gel tank was mixed together. At the end of the electrophoresis run the gel was photographed (Section 2.11.4).

Requirements

- Electrophoresis Buffer: 1× Mops/EDTA Buffer.
- 10× Mops/EDTA Buffer: 0.5 M Mops (pH 7.0), 0.01 M EDTA (pH 7.5).
- Buffer A: mix 294 µl of 10× Mops/EDTA Buffer with 706 µl of H_2O.
- Formaldehyde/formamide: mix 89 µl of 37% formaldehyde (12.3 M) with 25 µl of deionized formamide.

- Deionized formamide: add 5 g of Amberlite MB1 mixed bed resin to 100 ml of formamide and mix gently on a magnetic stirrer for 30 min. Filter through Whatman number 1 filter paper to remove the resin. Store at -20°C in 10 ml aliquots.
- Gel Loading Buffer: add 5 mg of xylene cyanol, 5 mg of bromophenol blue and 400 mg of sucrose (Molecular Biology Reagent, Sigma) to 322 µl of Buffer A and mix. Add 178 µl of 37% formaldehyde and 500 µl of deionized formamide.
- 1.4% agarose gel: add 0.98 g agarose (Molecular Biology Reagent, Sigma) to 70 ml of 1× Mops/EDTA Buffer and melt in a microwave oven. Cool to approx. 60°C and pour into a gel tray with a comb in place. Allow the gel to harden and carefully remove the comb. Transfer the gel to the electrophoresis tank and pour in Electrophoresis Buffer to completely cover the gel.

2.11.3 Analytical gel electrophoresis of DNA

DNA samples to be electrophoresed were mixed with either 0.2 volume of 5× sample loading buffer or 0.1 volume of 10× sample loading buffer. If the sample consisted of λ DNA, it was heated at 65°C for 5 min to denature the cohesive ends and then cooled on ice. An appropriate aliquot of each sample, depending on the size of the wells, was loaded into a well of the gel. For most purposes, a 0.7% agarose gel was made with, and electrophoresed in, 1× Tris-acetate (TAE) electrophoresis buffer. However, for specific applications, gels of 0.3 - 3%, in either TAE electrophoresis buffer or Tris-borate (TBE) electrophoresis buffer, were also used depending on the size of the DNA fragments to be resolved and the degree of resolution required. The gel was electrophoresed at a voltage gradient of 3 - 4 V/cm for 2 - 4 h, depending on the degree of DNA fragment separation required. Alternatively, the gel was electrophoresed overnight at approx. 1 V/cm. At the end of the electrophoresis run the gel was photographed (Section 2.11.4).

When a specific DNA fragment from an analytical gel was to be purified and subcloned, the gel was made up with and electrophoresed in TAE electrophoresis buffer. The gel was then examined in a darkroom under long wavelength u.v. light (Section 2.7.4), before being photographed, and a gel slice containing the band of interest was excised from the gel with a scalpel and transferred to a 1.5 ml microfuge tube. The gel slice was stored at 4°C until it could be further processed. The DNA was purified from the agarose with Geneclean as described in Section 2.11.8.

Requirements

- 5× sample loading buffer: 50 mM Tris-HCl, pH 8.0, 100 mM EDTA, 1% (w/v) SDS, 7.5% Ficoll 400, 0.05% bromophenol blue, 0.05% xylene cyanol. This loading buffer leads to unsatisfactory banding of DNA fragments when used with TBE electrophoresis buffer.
- 10× sample loading buffer: 125 mM EDTA, 0.1% (w/v) SDS, 50% glycerol, 0.05% bromophenol blue.
- 50× TAE Buffer (500 ml): dissolve 121 g Tris base, either 41 g of sodium acetate mw 82 or 68 g of sodium acetate mw 136, and 18.6 g EDTA (di-sodium salt) in approx. 400 ml of H_2O. Adjust pH to 8.0 with glacial acetic acid and make up to 500 ml with H_2O. Store at room temperature.
- TAE electrophoresis buffer (1 litre): add 20 ml of 50× TAE buffer and 50 µl of 10 mg/ml ethidium bromide to 980 ml of H_2O. This gives a working solution of 40 mM Tris-acetate pH 8.0, 20 mM sodium acetate, 2 mM EDTA, 0.5 µg/ml ethidium bromide.
- 10× TBE electrophoresis buffer (1 litre): dissolve 108 g Tris base, 55 g boric acid, and 7.4 g EDTA (di-sodium salt) in H_2O. Make up to a final volume of 1 litre with H_2O. Store at room temperature.
- 1× TBE Buffer (1 litre): add 100 ml of 10 × TBE Buffer and 50 µl of 10 mg/ml ethidium bromide to 900 ml of H_2O. This

gives a working solution of 89 mM Tris-borate, 2 mM EDTA, 0.5 µg/ml ethidium bromide.

- 10 mg/ml ethidium bromide: add 1 g of ethidium bromide (a mutagen which requires the avoidance of skin contact) to 100 ml of H_2O and stir for several hours on a magnetic stirrer to dissolve. Store at 4°C in a dark bottle.
- Agarose gel: measure out the required amount of electrophoresis buffer (1× TAE or 1× TBE) into a 250 ml conical flask. The volume used depends on the tray type and the depth of well required (70 ml for a microtitre plate lid and 50 - 100 ml for a 10 mm × 15 mm Bio-Rad tray). Weigh out sufficient agarose (Ultra Pure, BRL) to give the required gel concentration and add to the buffer. Invert a small plastic beaker over the top of the flask and weigh the flask. Place the flask in a microwave oven and bring the contents to the boil by heating on high power, mix the contents by swirling the flask and then heat for an additional 3 - 4 min, depending on volume, at approx. 30% power (700 watt oven). Re-weigh the flask and add H_2O to the original weight, thereby compensating for losses caused by evaporation. As an alternative to melting in a microwave oven, add a magnetic stirring bar to the agarose and melt the agarose on a magnetic stirrer/hot plate. Cool to approx. 60°C and pour onto a gel tray with the comb set so that the bottom of the teeth are approx. 1 mm above the tray. The ends of the Bio-Rad tray are first sealed with a piece of autoclave tape. Allow the gel to harden and then remove the comb and transfer the gel to the electrophoresis tank. Gels of less than 0.5% need to be poured and allowed to set in a coldroom.

2.11.4 Photography of agarose gels

Gels were transferred to a darkroom and, using proper eye protection, illuminated with u.v. light on a transilluminator (254 nm). Gels were photographed through a red filter onto Polaroid Type 667 film using a Polaroid MP4 camera. Gels

poured and electrophoresed in microtitre plate lids were inverted onto the transilluminator stage while gels poured and run in Bio-Rad trays were transferred right-side up onto the stage. A typical exposure setting was 1 second at f5.6. The Polaroid film was processed for 30 seconds.

2.11.5 Preparative electrophoresis of DNA

Preparative agarose gel electrophoresis was used when significant amounts of a specific DNA fragment were required, e.g. for labelling to produce probes. Low melting temperature agarose gels of 0.5 - 1.5%, depending on the size of the DNA fragment of interest, were made up with 1× TAE using a preparative comb. The DNA sample was mixed with 0.1 vol. of 10× Sample Loading Buffer (Section 2.11.2), heated at 65°C for 5 min and chilled on ice. Aliquots were then loaded into preparative wells in the gel (typically 150 µl per well) and the gel was electrophoresed as described in Section 2.11.3. At the end of the run, the gel was examined in a darkroom under long wavelength u.v. light (Section 2.7.4). The band of interest was excised from the gel with a scalpel and transferred to a 1.5 ml microfuge tube. The tube was weighed to determine the volume of the gel fragment and then stored at 4°C until it could be further processed (Section 2.11.7). The remainder of the gel was photographed (Section 2.11.4) to confirm that the correct fragment had been excised.

Requirements

- Low melting temperature agarose gels are made in the same way as standard agarose gels (Section 2.11.3) except that low melting temperature agarose (SeaPlaque, FMC Corp.) is used and the gels are poured in a coldroom.

2.11.6 Determination of fragment sizes

A set of standard fragments was run in the gel with each set of samples (Sections 2.11.2 and 2.11.3). For DNA gels, a *Hind*III digest of λ DNA was used for fragments larger than approx. 1 kb while an *Alu*I digest of pBR322 DNA was used for fragments smaller than approx. 1 kb. For RNA gels, a 0.24 - 9.5 kb RNA ladder (BRL) was used. Following electrophoresis, the gel was photographed onto Polaroid film (Section 2.11.4). On the photograph, the distance each fragment had migrated from the origin (a line drawn through the wells) was measured. A standard curve of the distances migrated by the standard fragments as a function of the logarithm of their sizes was plotted on semi-log graph paper. The sizes of unknown fragments were determined from the distances each had migrated, with reference to the standard curve.

2.11.7 Purification of DNA from low melting temperature agarose

The excised gel fragment containing the DNA of interest (Section 2.11.5) was heated at 65°C for 5 - 10 min, or until the agarose had completely melted, and then incubated at 37°C. An equal volume of phenol (Section 2.22.1) was added, the tube was vortexed for 30 seconds, and then centrifuged for 5 min. The aqueous phase and the cloudy interface were transferred to another tube and the phenol extraction was repeated. The aqueous phase was then extracted with an equal volume of chloroform (Section 2.22.3). The DNA was precipitated by the addition 0.1 volume of 3 M sodium acetate/acetic acid, pH 5.2 and 2 vol. of ethanol and pelleted by centrifugation for 20 min. The DNA pellet was rinsed with 70% ethanol and redissolved in an appropriate volume of H_2O.

2.11.8 Purification of DNA with Geneclean

The Geneclean kit (BIO 101) was used to purify DNA from a standard agarose gel slice (Section 2.11.3) as well as from solution as an alternative to ethanol precipitation. The purification procedure followed those of the kit instructions and Vogelstein and Gillespie (1979). The DNA, either in agarose or in solution and contained in a 1.5 ml microfuge tube, was mixed with 2 - 3 vol. of 6 M NaI solution (supplied with kit). Agarose, when present, was then dissolved by incubating the tube at 45 - 55°C for 5 min. A 5 µl aliquot of a suspension of powdered glass (supplied with kit) was added to the tube, the contents were briefly vortexed and the tube was incubated on ice for 5 min to allow the DNA to bind to the glass powder. The tube was then centrifuged in a microfuge for the length of time taken for the microfuge to reach full speed plus an additional 5 s. The supernatant was removed by aspiration through a fine pipette tip (Costar, No. 4853) connected to a water-jet aspirator pump. The DNA-/powdered glass pellet was resuspended in 5 - 10 vol. of wash solution by pipetting up and down in a 1 ml pipette tip and the tube was recentrifuged as above. The washing procedure was repeated twice more and then the pellet was resuspended in an appropriate volume of H_2O or low-salt buffer (typically 5 µl) and incubated at 45 - 55°C for 2 - 3 min to elute the DNA from the powdered glass. The tube was centrifuged for 1 min and the supernatant containing the eluted DNA was transferred to another tube. The elution step was repeated when maximum recovery of DNA was required. The DNA was then ready for further processing.

Requirements

- Wash solution (supplied with Geneclean kit or made up): 10 mM Tris-HCl, pH 7.5, 0.1 M NaCl, in 50% ethanol.

2.12 Transfer of RNA and DNA to solid supports

RNA and DNA were transferred from agarose gels by capillary transfer (Northern and Southern blotting respectively) onto either nitrocellulose or nylon membranes.

2.12.1 Northern blotting

RNA samples were electrophoresed on a formaldehyde agarose gel as described in Section 2.11.2. The RNA was then transferred from the gel to a sheet of Zeta-Probe nylon membrane (Bio-Rad) as described in Section 2.12.4, using 3 mM NaOH as the transfer solution. Following transfer, the membrane was lightly blotted, sealed in a plastic bag and stored at 4°C.

2.12.2 Southern blotting onto nitrocellulose

DNA samples were electrophoresed on an agarose gel in TAE electrophoresis buffer as described in Section 2.11.3. When digests of white clover DNA had been electrophoresed on a gel, the gel was immersed in 2 vol. of 0.25 M HCl for 2× 5 min with gentle agitation to partially depurinate the DNA. The gel was then rinsed briefly in distilled H_2O and soaked in 2 vol. of 0.5 M NaOH, 1.5 M NaCl for 2× 15 min to denature the DNA. Finally, the gel was neutralized by soaking in 2 vol. of 0.5 M Tris pH 7.4, 3 M NaCl for 2× 15 min. When digests of recombinant λ DNA were run on a gel, no pre-treatment of the gel was required. The DNA was transferred to a nitrocellulose membrane (Schleicher and Schuell, BA 85, 0.45 μm) as described in Section 2.12.4, using 20× SSC (Section 2.22.5) as the transfer solution. Following transfer, the membrane was baked for 2 h at 80°C under vacuum to fix the denatured DNA onto the nitrocellulose. The baked filter was sealed in a plastic bag with an impulse heat sealer and either used immediately for hybridization (Section 2.13.2) or stored at -20°C until required. The gel was

restained in TAE electrophoresis buffer (Section 2.11.3) with gentle agitation for 2 h and then rephotographed to confirm that efficient transfer of DNA had occurred.

2.12.3 Alkaline Southern blotting

DNA samples were electrophoresed on an agarose gel as described in Section 2.11.3. Gels on which clover DNA digests had been electrophoresed were immersed in 0.25 M HCl and rinsed in distilled H_2O as described in Section 2.12.2; gels on which digests of recombinant λ DNA had been electrophoresed were not pretreated. The DNA was transferred to either Zeta-Probe (Bio-Rad) or BioTrace RP (Gelman), both positively charged nylon membranes, as described in Section 2.12.4, using 0.4 M NaOH as the transfer solution. Transfer of DNA to the nylon membrane was carried out for 4 - 16 h. Following transfer, the membrane was air dried, sealed in a plastic bag and either used immediately for hybridization (Section 2.13.2) or stored at -20°C.

2.12.4 Capillary transfer

The pretreated agarose gel (Sections 2.12.1, 2.12.2, and 2.12.3) was placed on a piece of Whatman 3MM filter paper, larger than the gel, which was elevated above a tray of transfer solution with its ends immersed in the buffer (Fig. 2.1). A piece of nitrocellulose or nylon membrane, cut to the same size as the gel, was completely wetted in deionized H_2O and transferred to a tray containing transfer solution. The membrane was then carefully overlaid onto the gel. Two pieces of Whatman 3MM paper, also cut to the size of the gel and wetted in transfer solution, were laid onto the membrane. Pieces of Parafilm were inserted between the membrane and the gel, around the edges of the gel, to prevent the upper layers from contacting the transfer solution directly. The assembly was placed onto a light box to visualize any trapped air bubbles and these bubbles were removed by gently rolling a

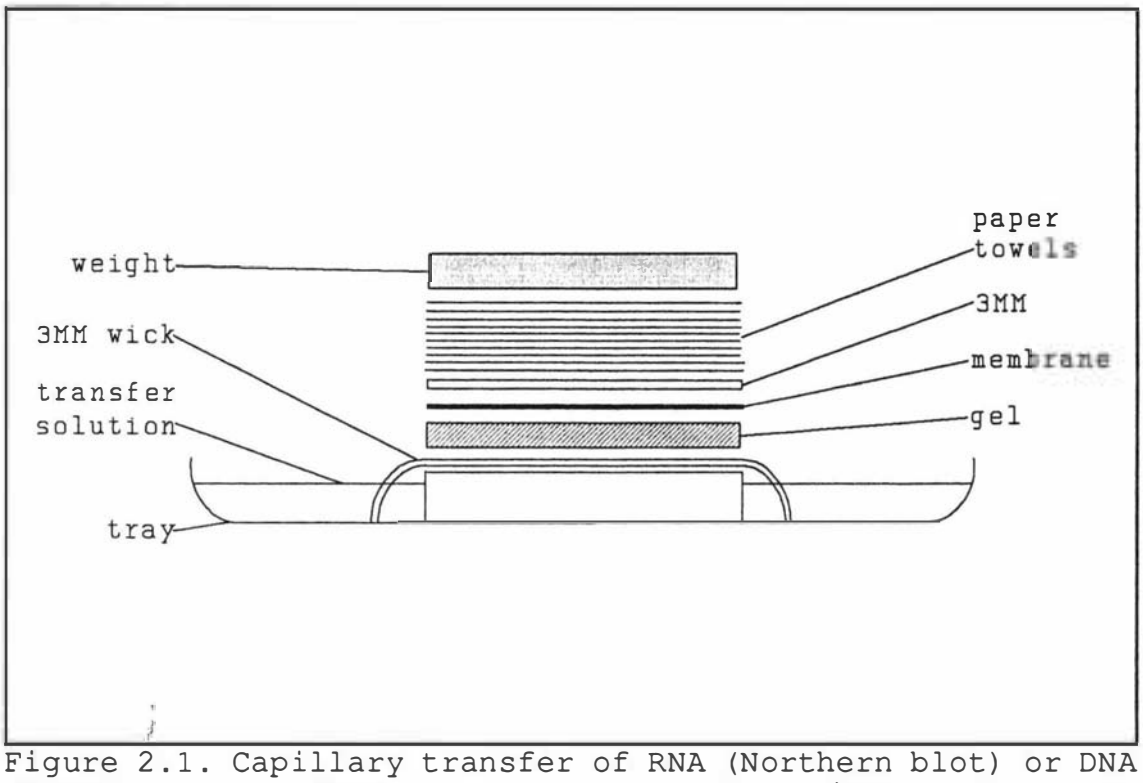


Figure 2.1. Capillary transfer of RNA (Northern blot) or DNA (Southern blot) from an agarose gel to a nitrocellulose or nylon membrane.

glass pipette over the filter paper. A wad of paper towels, cut to a size slighty larger than the gel and approx. 10 cm thick, was placed onto the 3MM filter paper. A plastic tray with a weight of approx. 250 g was placed onto the wad of paper towels. The whole assembly was left overnight for the transfer of nucleic acids from the gel to the membrane to occur. The upper layers were then removed and the membrane was rinsed for 5 min in 2× SSC with gentle agitation and air dried.

2.12.5 Bidirectional Southern blotting

Bidirectional Southern blotting depends solely upon the fluid within the gel itself to transfer DNA and results in the production of two replicate blots from the one gel (Fig. 2.2). Thus duplicate hybridization experiments can be performed from the same gel transfer. This procedure was used, as an alternative to conventional Southern blotting (Sections 2.12.2 and 2.12.3), for gels on which digests of

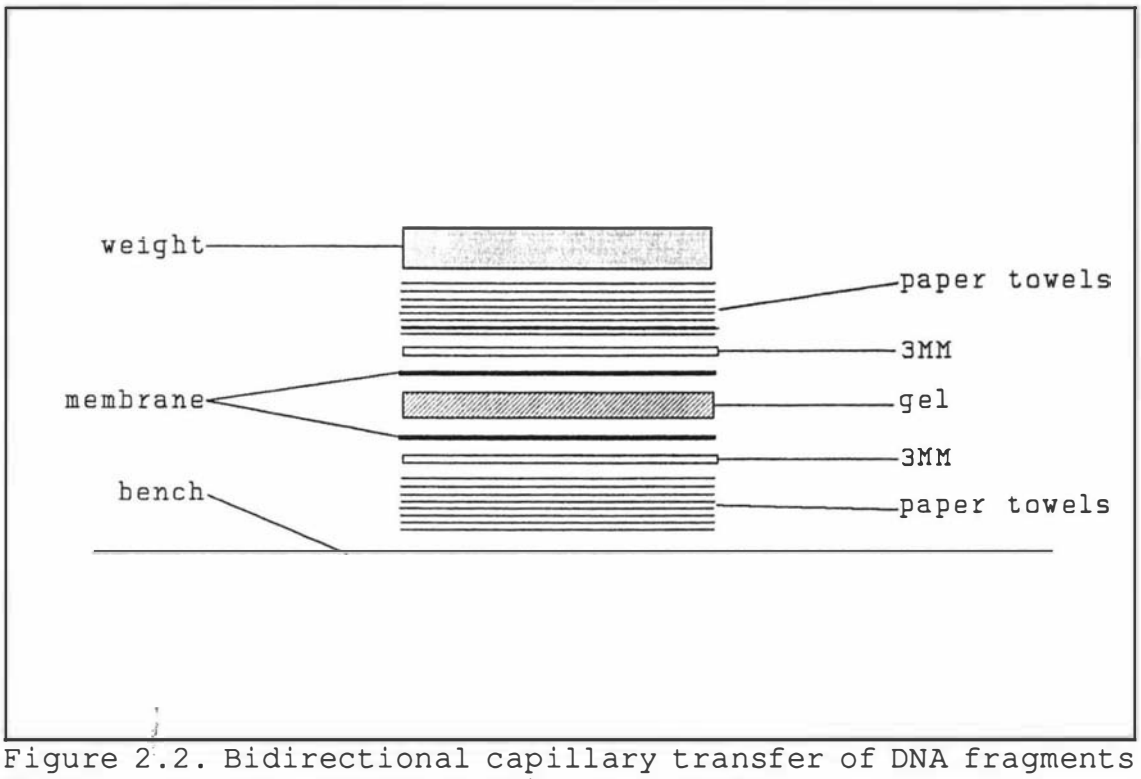


Figure 2.2. Bidirectional capillary transfer of DNA fragments from an agarose gel to duplicate membranes.

recombinant λ DNAs were run. Both nitrocellulose (see Section 2.12.2) and Zeta-Probe (see Section 2.12.3) membranes were used. For transfer to a nitrocellulose membrane, the gel was pretreated with gentle shaking in 2 vol. of 1.5 M NaCl, 0.5 M NaOH for 2× 15 min to denature the DNA followed by 2 vol. of 1 M ammonium acetate, 0.02 M NaOH for 2× 15 min to neutralize the gel. No pretreatment was required for transfer to a Zeta-Probe membrane.

A wad of paper towels, cut to a size slightly larger than the gel and approx. 5 cm thick, was placed directly on the laboratory bench. A piece of Whatman 3MM paper was soaked in either 1 M ammonium acetate, 0.02 M NaOH for transfer to nitrocellulose, or distilled H_2O for transfer to Zeta-Probe. The 3MM paper was placed onto the paper towels and a membrane, wetted in the same solution, was placed onto the 3MM paper. The gel was carefully placed onto the membrane to avoid trapping air bubbles and another membrane and a piece of 3MM paper, wetted as before, were sequentially placed onto

the gel. Another 5 cm thick wad of paper towels was placed on top of the gel sandwich and a plastic tray with a weight of approx. 500 g was placed on top of the wad of paper towel. Transfer of DNA was carried out for 1 - 2 h and then the assembly was dismantled and the membranes were removed. Following transfer, nitrocellulose membranes were rinsed for 5 min in 2× SSC, air dried and then baked for 2 h at 80°C under vacuum; Zeta-Probe membranes were transferred to sheets of Whatman 3MM paper saturated with 0.4 M NaOH for 10 min, rinsed in 2× SSC for 5 min, and then air dried. Membranes were sealed in plastic bags and stored at -20°C.

2.13 Northern and Southern blot hybridizations

For comprehensive reviews of the detection of immobilized nucleic acids, see Meinkoth and Wahl (1984) and Wahl *et al.* (1987).

2.13.1 Northern blot hybridization

Northern blots onto nylon membranes were prepared as described in Section 2.12.1. Background radioactivity in the Northern blot hybridization was controlled by including heparin in the prehybridization and hybridization solutions (see Section 2.6.9). Prehybridization solution was added to the bag containing the membrane (10 ml per 100 cm^2 of membrane) and the bag was resealed. The membrane was pre-hybridized for 2 - 4 h in a 37°C incubator with gentle shaking. The denatured probe (Section 2.14.2) was then added to the bag and the membrane was hybridized overnight as for the prehybridization step.

Following hybridization, the hybridization solution was removed from the bag and the membrane was transferred to a tray containing 150 ml of 2× SSC. The membrane was rinsed, with gentle shaking, sequentially with 150 ml of 2× SSC, 0.1% SDS at 65°C for 2 × 30 min and 0.1× SSC, 0.1% SDS at 37°C for

30 min and then blotted lightly on a piece of paper towel. The damp membrane was wrapped in polyethylene film, attached to a sheet of paper and transferred to an autoradiography cassette. Working in a darkroom with appropriate safe-lighting, a sheet of X-ray film (Kodak XAR-5) was placed over the membrane, an intensifying screen (DuPont, Cronex Lightening Plus) was placed over the film and the cassette was closed. The membrane was autoradiographed at -70°C for 16 - 48 h. Following autoradiography, the cassette was warmed to room temperature and the film was developed as described in Section 2.6.10.

Requirements

- Prehybridization solution: see Section 2.6.9.
- SSC (1×): 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0; see Section 2.22.5 for 20× stock.

2.13.2 Southern blot hybridization

Southern blots onto nitrocellulose or nylon membranes were prepared as described in Sections 2.12.2, 2.12.3 and 2.12.4. Prehybridization solution was added to the bag containing the membrane (10 ml per 100 cm^2 of membrane) and the bag was sealed. The membrane was prehybridized at 42°C in an incubator with gentle shaking for 2 - 4 h. The denatured probe (Section 2.14.2) was then added to the bag and the membrane was hybridized overnight as above.

Following hybridization, the hybridization solution was removed from the bag and the membrane was transferred to a tray containing 150 ml of 2× SSC. The membrane was then washed, with gentle shaking, sequentially in 150 ml of 2× SSC, 0.1% SDS at room temperature for 2× 15 min and 0.1× SSC, 0.1% SDS at 42°C for 2× 15 min and then blotted lightly on a piece of paper towel (see Section 2.22.5 for 20× SSC stock and Section 2.22.6 for 10% SDS stock). When a higher strin-

gency was required, the membrane was washed at 55 - 65°C instead of 42°C. The damp membrane was wrapped in polyethylene film, attached to a sheet of paper and autoradiographed as described in Section 2.13.1. Typically, autoradiography was carried out for 2 - 7 days. The autoradiograph was subsequently developed as described in Section 2.6.10.

Requirements

- Prehybridization solution (nitrocellulose membrane): as for prehybridization solution used in plaque hybridization (Section 2.6.9) except that 25 - 50% formamide was used, depending on the hybridization stringency required.
- Prehybridization solution (nylon membrane): 25 - 50% formamide (depending on the hybridization stringency required), 5× SSC (Section 2.22.5), 0.5% SDS (Section 2.22.6), 5× Denhardt's solution, 100 µg/ml sonicated, denatured herring sperm DNA.
- 100× Denhardt's solution: 2% Ficoll 400, 2% polyvinylpyrrolidone and 2% bovine serum albumin (Type III sodium salt, Sigma). Make up solutions of individual components separately in sterile H_2O at 6% concentration and then mix together equal volumes of each. Keep working aliquots at 4°C and freeze remainder at -20°C. These solutions are too viscous to be sterilized by filtration and BSA cannot be autoclaved without causing it to denature, so freezing is necessary to prevent microbial growth.
- 100 µg/ml sonicated, denatured herring sperm DNA: dissolve the DNA (Boehringer, Cat. No. 223 646) in sterile H_2O at a concentration of 10 mg/ml. If necessary, stir the solution on a magnetic stirrer for 2 - 4 h at room temperature to help the DNA to dissolve. Sonicate to reduce the molecular size to 1 - 5 kb; the viscosity of the solution should decrease noticeably. Store in 1 ml aliquots at -20°C.

2.13.3 Removal of probe from blots

Hybridized probe was removed from both Northern and Southern blots (nitrocellulose and nylon membranes), allowing the same membranes to be reprobed with alternative probes. The membrane was removed from the polyethylene film used for autoradiography and placed in a plastic container. Boiling 0.1× SSC, 0.1% SDS (200 ml; see Section 2.22.5 for 20× SSC stock and Section 2.22.6 for 10% SDS stock) was poured onto the membrane and the container was agitated gently for 15 min, allowing the solution to cool to room temperature. The solution was then removed and the treatment was repeated. Following this treatment, the membrane was rinsed briefly with 0.1× SSC at room temperature, lightly blotted on a piece of paper towel, sealed in a plastic bag and then stored at -20°C.

2.14 Labelling of DNA

2.14.1 Nick translation

The labelling of double-stranded DNA by nick translation depends on the combined abilities of *E. coli* DNA polymerase I (Pol I) to add nucleotide residues to the 3′-hydroxyl terminus of a nick, generated by DNase I, and to remove nucleotides from the adjacent 5′ phosphoryl terminus. The method used to make DNA probes by nick translation was based on the methods of Rigby *et al.* (1977) and K. Reed (unpublished). The nick translation reaction was set up in a 1.5 ml microfuge tube as follows:

DNA (0.2 µg) + deionized H_2O	18.0 µl
10× NTB	3.5 µl
0.1 M DTT	2.0 µl
1 mg/ml BSA	3.5 µl
0.5 mM dNTP mix	1.0 µl
[α-^{32}P] dCTP	2.0 µl
200 ng/ml DNase I	5.0 µl
TOTAL	35.0 µl

The tube was vortexed, centrifuged briefly and incubated at 14°C for 15 min. A 0.5 µl aliquot was removed during this incubation for subsequent analysis by PEI-cellulose chromatography (Section 2.14.3), when required. A 1 µl aliquot of Pol I was then added and the tube was vortexed briefly and incubated at 14°C for 15 min. The reaction was stopped by adding 4 µl of 10× stop solution and a second 0.5 µl aliquot was removed for PEI-cellulose chromatography. The labelled DNA was precipitated by adding 20 µl of 7.5 M ammonium acetate and 120 µl of ethanol and pelleted by centrifugation for 20 min (see Section 2.21). The supernatant was then discarded and the pellet was rinsed with 80% ethanol and recentrifuged for 5 min. The rinse was repeated and the pellet was finally redissolved in 50 µl TE (Section 2.22.4). The tube containing the probe was placed inside a liquid scintillation vial and counted (Cerenkov) in a liquid scintillation counter. The probe was then stored at -20°C until required.

Requirements

- 10× NTB: 0.5 M Tris-HCl, pH 7.5, 0.075 M magnesium acetate. Sterilize by filtration (0.2 µm, TCM-200, Gelman) and store in 20 µl aliquots at -20°C.
- 1 M DTT: dissolve 3.09 g of DTT in 20 ml of 0.01 M sodium acetate/acetic acid, pH 5.2. Sterilize by filtration (0.2 µm, TCM-200, Gelman) and store in 200 µl aliquots at -20°C. Dilute 1/10 with H_2O for 0.1 M stock.

- 1 mg/ml BSA: dilute 20 mg/ml stock of nuclease-free BSA (Boehringer) in sterile, deionized H_2O and store in 50 µl aliquots at -20°C.
- 0.5 mM dNTP mix: mix 5 µl of 20 mM dATP, 5 µl of 20 mM dGTP, 5 µl of 20 mM dTTP and 185 µl of TE (Section 2.22.4). Store in 20 µl aliquots at -70°C, with one working aliquot at -20°C.
- [α-^{32}P] dCTP (aqueous solution, Amersham) at 10 mCi/ml and a specific activity of approx. 3000 Ci/mmol was purchased monthly and stored at -20°C.
- DNase I: dissolve 5 mg of deoxyribonuclease I (D4527, Sigma) in 2.5 ml of 0.15 M NaCl, 50% glycerol. Store in 50 µl aliquots at -70°C, with one working aliquot at -20°C. Prepare a solution of 200 ng/ml by two successive dilutions of 1 : 100 into deionized H_2O immediately before use.
- Pol I: *E. coli* DNA polymerase I at a concentration of 5 U/µl was obtained from Amersham.
- 10× stop solution: 125 mM EDTA, 5% (w/v) SDS, from 0.5 M EDTA and 10% SDS (Section 2.22.6) stocks.

2.14.2 Denaturation of probe

Labelled DNA (Section 2.14.1) was made single-stranded immediately before being used in a hybridization experiment. An equal volume of formamide was added to the probe and the probe was heated at 70°C for 5 min and then quenched on ice. Alternatively, 0.1 volume of 1 M NaOH was added to the probe and the probe was incubated at 37°C for 5 min.

2.14.3 PEI-cellulose chromatography

The nick translation reaction was optimized and the quality of different batches of [α-^{32}P] dCTP were checked by analyzing aliquots from the reaction (Section 2.14.1) with PEI-cellulose chromatography. The two samples from each nick translation reaction (Section 2.14.1) were spotted onto

origin marks on a sheet of PEI-cellulose (polyethyleneimine-cellulose coated plastic sheets for thin layer chromatography; Schleicher and Schuell, F 1440 PEI). and air-dried. The samples were chromatographed in a foil-covered beaker containing 0.75 M KH_2PO_4, adjusted to pH 3.5 with orthophosphoric acid, to a depth of approx. 2 mm. The sheet was removed from the beaker when the solvent front was near the top and wrapped in polyethylene film. The sheet was then transferred to an autoradiography cassette and autoradiographed onto X-ray film, without an intensifying screen, for 10 - 30 min at room temperature. The autoradiograph was developed as described in Section 2.6.10.

2.15 Subcloning DNA fragments

DNA fragments were routinely subcloned into the plasmid vectors pGEM-3Z (Promega), pSP72 (Promega) or pUBS (pUC19 with the original polylinker replaced by the KS polylinker from Bluescript (Stratagene); obtained from R. Forster, DSIR). Both pGEM-3Z and pUBS encode the alpha fragment of beta-galactosidase when expressed in a strain encoding the complementing beta fragment, such as *E. coli* JM109, *E. coli* DH5α, or *E. coli* MC1022. The alpha fragment coding sequence, under control of the *lac* promoter, is interrupted by the polylinker sequence that maintains the proper open reading frame for production of functional alpha fragment. Insertion of DNA fragments into the polylinker disrupts this open reading frame, resulting in an inability to produce functional beta-galactosidase; thus clones containing insertions can be identified on appropriate indicator plates.

Requirements

- *E. coli* JM109 (Yanisch-Perron *et al.*, 1985): *end*A1, *rec*A1, *syr*A96, *thi*, *hsd*R17(r_K^-, m_K^+), *rel*A1, *sup*E44, λ^-, Δ(*lac-pro*AB), [F′, *tra*D36, *pro*AB, *lac*$I^qZ\Delta M15$].

- *E. coli* DH5α (BRL): F^-, ϕ80d, *lac*ZΔM15, *end*A1, *rec*A1, *hsd*R17(r_K^-, m_K^+), *sup*E44, *thi*-1, d^-, *gyr*A96, Δ(*lac*ZYA-*arg*F), U169.
- *E. coli* MC1022 (Stanford *et al.*, 1989): *ara*D139, Δ(*ara*, *leu*)7697, *lac*ZΔM15, *gal*U, *gal*K, *str*A.

2.15.1 Digestion of plasmid vector DNA

Approx. 0.5 µg of vector DNA was digested with the appropriate restriction enzyme(s) at 37°C (Section 2.19). When the digest was with a single restriction enzyme, 0.5 µl of phosphatase was added and the digest was incubated for an additional 30 min at 37°C. The digested DNA was then purified with Geneclean and used in a ligation reaction (Section 2.15.2). When problems were encountered with incomplete digestion of vector DNA, the digest was electrophoresed on an agarose gel, without additional purification, and the band consisting of linearized vector DNA was excised from the gel (see Section 2.11.3). The DNA fragment was then purified with Geneclean and used in a ligation reaction (Section 2.15.2).

Requirements

- Phosphatase: alkaline phosphatase from calf intestine, molecular biology grade, at a concentration of 22 U/µl was obtained from Boehringer.

2.15.2 Ligation of cloned DNA with vector DNA

DNA samples, digested with the appropriate restriction enzyme(s) (Section 2.15.1), were electrophoresed on an agarose gel (0.5 - 3% depending on the size of fragment to be subcloned). The fragment of interest was excised from the gel and transferred to a 1.5 ml microfuge tube (see Section 2.11.3). The digested vector DNA (Section 2.15.1) was added to the tube and the combined DNAs were purified with Geneclean and eluted with 15 µl of H_2O as described in

Section 2.11.8. The eluted DNA was mixed with 4 µl of 5× ligase buffer and 1 µl of T4 DNA ligase. Ligations involving vector and fragment DNAs with sticky ends (complementary 3′ or 5′ overhangs) were incubated at room temperature for 2 - 16 h; ligations involving vector and insert DNAs with blunt ends were incubated at 14°C for at least 16 h. The ligation was then either used immediately to transform competent *E. coli* cells (Section 2.15.4) or stored at -20°C until required.

Requirements

- 5× ligase buffer (King and Blakesley, 1986): 50 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 5% PEG 8000, 1 mM ATP (from 0.1 M stock; Section 2.4.4), 1 mM DTT (from 1 M stock; Section 2.14.1).
- T4 DNA ligase at a concentration of approx. 1 U/µl (Weiss unit) was obtained from BRL.

2.15.3 Preparation of competent *E. coli* cells

Competent *E. coli* cells were prepared by $CaCl_2$ treatment following the method of Miller (1987). Initially, *E. coli* JM109 was used but this strain was subsequently replaced by *E. coli* strains DH5α and MC1022 (Section 2.15), both of which gave transformation efficiences of approx. 100× that of *E. coli* JM109 when using $CaCl_2$-treated, competent cells.

A 10 ml overnight culture in LB medium (Section 2.15.8) was grown up overnight at 37°C with shaking. The overnight culture was then diluted 40-fold into 250 ml of LB medium in a 1 litre flask and the flask was incubated at 37°C with vigorous agitation until an A_{550} of 0.4 - 0.5 was reached. The culture was chilled by swirling in an ice-water bath and then transferred to two 250 ml centrifuge bottles (Nalgene) and centrifuged in a Sorvall GSA rotor at 5,000 rpm and 4°C for 10 min. The supernatants were poured off and each pellet was

resuspended in 60 ml of ice-cold 100 mM $CaCl_2$ by sucking up and down with a 10 ml pipette attached to a Pi-pump. The resuspended cells were combined into one centrifuge bottle and incubated on ice for 30 min with occasional agitation. The cells were then centrifuged as above and resuspended in 20 ml of ice-cold 100 mM $CaCl_2$, 15% glycerol. Aliquots of 200 µl were distributed into sterile 1.5 ml microfuge tubes and the tubes were incubated on ice for approx. 24 h. The competent cells were then frozen in an ethanol-dry ice bath and stored at -70°C.

2.15.4 Transformation of *E. coli*

A tube of frozen competent cells (Section 2.15.3) was thawed on ice. A 10 µl aliquot of the ligation reaction (Section 2.15.2) was gently mixed with the cells, and the cells were incubated on ice for at least 30 min. The cells were then heat shocked for 2 - 3 min in a 43°C water bath and transferred to ice for 1 min. A 400 µl aliquot of SOC medium was mixed with the cells and the tube was incubated for 1 h at 37°C without shaking. An aliquot of transformed cells (typically 100 - 200 µl) was then spread on an LB Amp plate and the plate was incubated overnight at 37°C. Plate cultures were stored at 4°C for up to 1 month.

Requirements

- LB Amp plates (per litre): 10 g Bacto-tryptone, 5 g yeast extract, 10 g NaCl 15 g agar. Adjust pH to 7.5 with NaOH and autoclave for 25 min. Cool to 60°C. In a laminar flow cabinet, add 50 mg of ampicillin (sodium salt, Sigma). Swirl to dissolve the ampicillin and pour into plates. Allow the plates to harden and then dry for 20 min in the laminar flow cabinet with the lids of the plates removed. Seal plates in a plastic bag and store at room temperature for up to 1 month.

- SOC medium: LB medium (Section 2.15.8) supplemented with 10 mM $MgSO_4$, 10 mM $MgCl_2$, and 20 mM glucose.
- 1 M $MgSO_4$, 1 M $MgCl_2$: sterilise by filtration through a 0.22 µm filter (TCM-200, Gelman).
- 2 M glucose: filter sterilise as for 1 M $MgSO_4$.

2.15.5 Selection of transformants

Cloned inserts in the plasmids pGEM-3Z and pUBS were selected by blue/white colour screening (see Section 2.15). An LB Amp plate (Section 2.15.4) was spread with 30 µl each of X-Gal and IPTG and the plate was allowed to dry for approx. 30 min. An aliquot of transformed cells (Section 2.15.4) was then spread on the plate. After overnight growth, cells that had been transformed with a plasmid containing an insert produced white colonies while cells transformed with a plasmid lacking an insert produced smaller, blue colonies. When necessary, the blue colour was made more intense by incubating the plate at 4°C for 1 - 2 days.

Requirements

- X-Gal: dissolve 20 mg of X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) in 1 ml of N,N'dimethylformamide and store at -20°C.
- IPTG: dissolve 20 mg of IPTG (isopropyl-β-D-thio-galactopyranoside) in 1 ml of sterile H_2O and store frozen at -20°C.

2.15.6 Characterization of transformants

A number of colonies (typically 6 white colonies, Section 2.14.5 or 12 non-colour selected colonies) was patched out onto an LB Amp plate (Section 2.15.4; see Davis *et al.*, (1980) for templates) and the plate was incubated overnight at 37°C. Plasmid DNAs were prepared from each patch culture (Section 2.15.7) and an aliquot of each DNA preparation was

characterized by restriction enzyme digestion with the appropriate enzyme(s) and agarose gel electrophoresis (Section 2.19).

2.15.7 LiCl-boiling method for plasmid mini-preps

The method used for preparing mini-preps of plasmid DNAs was based on that of Wilimzig (1985). Cells, either scraped off an overnight plate or pelleted from an overnight culture by centrifugation for 5 min, were resuspended in 100 µl of TELT buffer in a 1.5 ml microfuge tube and 10 µl of a freshly prepared aqueous solution of lysozyme (10 mg/ml) was added. The tube was vortexed and placed in a boiling water bath for 1 min. The tube was then cooled on ice for 5 min and centrifuged for 10 min at room temperature to pellet cell debris along with LiCl-precipitated rRNA and protein aggregates. The pellet was removed with a 200 µl pipette tip and 250 µl of ethanol was added to the supernatant. The tube was then centrifuged for 20 min to pellet the plasmid DNA. The supernatant was removed and the pellet was rinsed with 500 µl of 80% ethanol. The tube was recentrifuged for 5 min and the supernatant was again removed. The plasmid DNA pellet was redissolved in 30 µl of H_2O with heating at 65°C for 10 min. An aliquot of each DNA preparation (typically 3 µl) was characterized by restriction enzyme digestion and agarose gel electrophoresis (Section 2.19).

Requirements

- TELT buffer: 50 mM Tris-HCl, 62.5 mM EDTA, 0.4% Triton X-100, 2.5 M LiCl, pH 7.5.

2.15.8 Storage of recombinant clones

The method used for the storage of *E. coli* recombinant clones was based on the method of G. Limsowtin (personal commun-

ication). Each positive isolate was inoculated into 5 ml of LB Amp medium in a Universal bottle and incubated overnight at 37°C with shaking at approx. 200 rpm. Two 500 µl aliquots were then transferred to sterile 1.5 ml microfuge tubes and the tubes were centrifuged for 2 min. The supernatants were removed and each cell pellet was resuspended in 250 µl of Revco medium. The tubes were then placed directly into a -70°C freezer for storage.

Requirements

- LB Amp medium: LB medium supplemented with ampicillin from a 100 mg/ml stock to 100 µg/ml.
- LB medium (per litre): 10 g Bacto-tryptone, 5 g yeast extract, 10 g NaCl. Adjust pH to 7.5 with NaOH. Autoclave for 25 min.
- 100 mg/ml ampicillin: dissolve 1 g of ampicillin (sodium salt, Sigma) in 10 ml of H_2O. Sterilize by filtration through a 0.22 µm filter (TCM-200, Gelman) and store in 200 µl aliquots at -20°C.
- Revco medium: mix 10 ml of 10× Revco salts with 30 ml of glycerol and 60 ml of H_2O and autoclave.
- 10× Revco salts: dissolve 7 g K_2HPO_4, 3 g KH_2PO_4, 0.5 g sodium citrate, and 0.1 g $MgSO_4.7H_2O$ in H_2O and make up to 100 ml.

2.16 Large-scale plasmid DNA preparation

2.16.1 Plasmid DNA isolation by alkaline lysis - Method 1

This method was based on that described in the Promega Technical Bulletin, Number 009. A single colony was inoculated into 5 ml of LB Amp medium (Section 2.15.7) and the culture was incubated overnight at 37°C with shaking. A 0.5 ml aliquot of this culture was then inoculated into 250 ml of the same medium in a 1 liter flask and the flask was

incubated overnight at 37°C with vigorous shaking. The cells were transferred to a 250 ml centrifuge bottle (Nalgene) and harvested by centrifugation in a Sorvall GSA rotor at 5,000 rpm and 4°C for 15 min. The cell pellet was resuspended in 6 ml of freshly prepared lysis buffer by pipetting up and down with a 10 ml pipette attached to a Pi-pump. The resuspended cells were transferred to a 50 ml centrifuge tube (Oak Ridge, Nalgene) and incubated in ice water for 20 min. Freshly prepared 0.2 M NaOH, 1% SDS (6 ml) was added, the tube was mixed by inversion and incubated in ice water for 10 mins. Sodium acetate/acetic acid, pH 4.6 (3 M with respect to sodium; 7.5 ml) was added, the tube was again mixed by inversion and incubated for 20 min in ice water. The tube was then centrifuged in a Sorvall SS-34 rotor at 15,000 rpm and 4°C for 15 min. The supernatant was transferred to another tube, mixed with 50 µl of RNase A (1 mg/ml) and incubated at 37°C for 20 min. The tube was then almost filled with phenol/chloroform (Section 2.22.2), vortexed for 30 seconds and centrifuged as above for 5 min. The phenol/chloroform extraction was repeated and the aqueous phase was distributed into 2 tubes. Ethanol (2 vol.) was added to each tube, the tubes were centrifuged at 10,000 rpm and 4°C for 10 min and the supernatants were removed. The DNA pellets were dissolved, with gentle agitation at 37°C, in a total volume of either 1.6 ml of H_2O for further purification by PEG precipitation (Section 2.16.3) or 4 ml TE (Section 2.22.4) for further purification by CsCl gradient centrifugation (Section 2.16.4).

Requirements

- Lysis buffer: 25 mM Tris-HCl, pH 7.5, 10 mM EDTA, 15% (w/v) sucrose, 2 mg/ml lysozyme (Sigma, No. L6876).

2.16.2 Plasmid DNA isolation by alkaline lysis - Method 2

This method was based on that described by Grinsted and Bennett (1988). The yield of DNA obtained using this method was lower than that of Method 1 (Section 2.16.1) but the DNA was found to be more suitable for digestion with *E. coli* exonuclease III (Section 2.17.1), probably because of less nicking of the plasmid DNA from endogenous DNases during the initial cell lysis step.

A 250 ml overnight culture was harvested as described in Section 2.16.1. and the cell pellet was resuspended in 13 ml of 5 mM EDTA. The resuspended cells were transferred to a 50 ml centrifuge tube (Oak Ridge, Nalgene) and mixed with 1 ml of 10% SDS (Section 2.22.6) and 1 ml of 2 M NaOH. After a 5 min incubation at room temperature, 7.5 ml of potassium acetate solution was added and the tube was mixed by inversion and incubated at room temperature for an additional 5 min. The tube was then centrifuged in a Sorvall SS-34 rotor at 15,000 rpm and 4°C for 5 min. The supernatant was transferred to another tube and mixed with 0.4 ml of 4 M sodium acetate/acetic acid, pH 6 and 12 ml of isopropanol. The tube was centrifuged as above and then the supernatant was removed. The tube was filled with ethanol, centrifuged as above for 1 min and the supernatant was again removed. The pellet was dried for 5 min under vacuum and redissolved in 4 ml of TE (Section 2.22.4) with gentle shaking at 37°C. The DNA was further purified by CsCl gradient centrifugation (Section 2.16.4).

Requirements

- Potassium acetate solution: add 11.5 ml of glacial acetic acid and 28.5 ml of H_2O to 60 ml of 5 M potassium acetate. The solution is 3 M with respect to potassium and 5 M with respect to acetate.

2.16.3 Plasmid DNA purification by PEG precipitation

This method was based on that described in the Promega Technical Bulletin, Number 009. The DNA solution in a total volume of 1.6 ml (Section 2.16.1) was transferred to a 15 ml glass centrifuge tube (Corex) and mixed with 0.4 ml of 4 M NaCl and 2 ml of 13% (w/v) polyethylene glycol 6000 PEG. The tube was incubated in ice water for at least 1 h and then centrifuged in a Sorvall SS-34 rotor at 10,000 rpm and 4°C for 10 min. The supernatant was removed and the pellet was rinsed with 5 ml of cold (-20°C) 70% ethanol and redissolved in 500 µl of H_2O. An estimate of the DNA concentration was made by diluting the DNA solution 1/5 with H_2O and electrophoresing 1 µl of the diluted DNA on an agarose gel (Section 2.11.3). The DNA solution was stored at -20°C.

2.16.4 Plasmid DNA purification by CsCl gradient centrifugation

The DNA solution, in a total volume of 4 ml (Sections 2.16.1 and 2.16.2), was mixed with 400 µl of 10 mg/ml ethidium bromide solution (Section 2.11.3) and 4.4 g of CsCl. The CsCl was dissolved by gentle agitation. The density of the solution was determined by weighing a 1 ml aliquot and adjusted, if necessary, to 1.55 - 1.59 g/ml by adding TE (Section 2.22.4) or additional CsCl. The solution was transferred to a 5 ml polyallomer centrifuge tube (Beckman, No. 342412) and the tube was filled, if necessary, with a CsCl, ethidium bromide solution of the same density (see above) and sealed. The tube was centrifuged in a Beckman VTi65 rotor at 48,000 rpm and 14°C for 16 h or at 55,000 rpm and 14°C for 5 h. The tube was then removed from the rotor and examined in a darkroom under long wavelength u.v. illumination (Section 2.7.4). Two bands of DNA were usually visible - a lower plasmid band and an upper, fainter, chromosomal band. The tube was punctured at the top with a syringe needle and an 18

gauge needle, attached to a 5 ml syringe, was inserted into the tube just below the plasmid band. The plasmid band was removed and transferred to a second 5 ml polyallomer centrifuge tube. The second tube was filled with CsCl, ethidium bromide solution (see above), sealed and centrifuged as above.

Following the second centrifugation, the plasmid band was removed from the tube as above. The plasmid DNA was purified from CsCl and ethidium bromide according to the method of Grinsted and Bennett (1988) with minor modifications. The DNA was transferred to 1.5 ml microfuge tubes such that each tube contained 270 µl of the DNA solution. To each tube was added 530 µl of 10 mM Tris-HCl, pH 7.5, 70 µl of 4 M sodium acetate/acetic acid, pH 6, and 0.5 ml of isopropanol. The tubes were mixed by inversion and centrifuged for 5 min. The supernatants were then removed and the pellets were rinsed sequentially with 1 ml of 40% (v/v) isopropanol, 0.2 M sodium acetate/acetic acid, pH 6 and 1 ml of ethanol, with a 1 min centrifugation at each rinse. The pellets were dissolved and combined in a total volume of 0.8 ml of 10 mM Tris-HCl, pH 7.5 in one of the tubes. An equal volume of phenol (Section 2.22.1) was then added, the tube was vortexed for 30 sec and centrifuged for 1 min. The aqueous phase was transferred to another tube and mixed with 70 µl of 4 M sodium acetate-/acetic acid, pH 6, and 0.5 ml of isopropanol. The tube was mixed by inversion and centrifuged for 1 min. The supernatant was removed and the pellet was rinsed sequentially with 1 ml of ethanol and 1 ml of ether, with a 1 min centrifugation at each rinse. The pellet was then dried for 5 - 10 min at 37°C and redissolved in 500 µl of H_2O. The DNA concentration was determined spectrophotometrically (Section 2.20). The DNA solution was stored at -20°C.

2.17 Generation of deletion derivatives for sequencing

Ordered sets of deletions of a cloned plasmid insert were generated by methods involving either *E. coli* exonuclease III (Exo III) or *Bal*31 exonuclease digestion of the double-stranded DNA. Exo III catalyses the stepwise removal in a 3′ - 5′ direction of nucleotides from the 3′ termini of blunt or 5′ protruding ends but not from 3′ protruding ends. In contrast, *Bal*31 catalyses the stepwise removal of nucleotides from both 5′ and 3′ termini. Both methods can generate a set of deletions originating close to a sequencing primer and extending various lengths along the insert DNA. Each successively longer deletion enables a new region of the insert to be sequenced from the same primer site. Thus the entire insert can be sequenced in a series of separate reactions and the complete sequence can be assembled from the sequence of each deletion derivative. The DNA fragment from which deletion derivatives were to be subsequently generated was cloned into one of the vectors pGEM-3Z, pUBS or pSP72 (Section 2.15).

2.17.1 Unidirectional digestion with Exo III

The method used to generate deletion derivatives by digestion with Exo III was based on the methods of Henikoff (1984; 1987) and the instructions supplied with a commercial kit (Erase-a-Base, Promega). Reagents were either prepared independently or used as components of the Promega kit.

An aliquot of CsCl gradient-purified plasmid DNA (Section 2.16.4), equivalent to 5 - 10µg, was transferred to a 1.5 ml microfuge tube and the DNA was digested with two restriction enzymes - one which generated a 4-base 3′ protrusion to protect the primer binding site and the other which generated a 5′ protrusion or a blunt end adjacent to the insert. Digestion was according to the restriction enzyme manufac-

turer's instructions. Typically, 100 U of restriction enzyme was used in a total volume of 100 µl in a 2 h digestion. If salt conditions did not allow simultaneous digestion with two different restriction enzymes, the first digestion was carried out in low salt and then additional NaCl and the second enzyme were added. In cases where a restriction enzyme site generating a 3′ protrusion did not occur adjacent to the primer binding site, 3′ recessed ends were filled-in with α-phosphorothioate deoxynucleotides and Klenow DNA polymerase as described in Section 2.17.2.

Following restriction enzyme digestion, an equal volume of phenol/chloroform (Section 2.22.2) was added to the tube, the tube was vortexed for 30 sec and then centrifuged for 5 min. The aqueous phase was transferred to another tube and the DNA was precipitated by adding 0.1 volume of 3 M sodium acetate-/acetic acid, pH 5.2 and 2 vol. of ethanol. The tube was centrifuged for 20 min and the pellet was rinsed with 500 µl of 80% ethanol and dried for 5 min under vacuum. The pellet was redissolved in 54 µl of H_2O and then 6 µl of 10× Exo III buffer was added.

The doubly digested, or singly digested and α-phosphoro-thioate protected, plasmid DNA was warmed to 37°C. A series of 0.6 ml microfuge tubes, the number depending on the number of time points to be taken (typically 25), was placed on ice and 7.5 µl of S1 mix was added to each. A 2.5 µl aliquot of plasmid DNA was removed into the first of the time point tubes (time zero). Approx. 500 U of Exo III was then added to the tube containing the plasmid DNA. The tube was vortexed and returned to 37°C. At 30 sec intervals, 2.5 µl aliquots were removed to the series of time point tubes and mixed by briefly pipetting up and down. After all the samples had been taken, the time point tubes were incubated at room temp-erature for 30 min so that the 5′ protrusions resulting from the Exo III digestion could be removed by the S1 nuclease. A 1 µl aliquot of S1 stop buffer was then added to each tube

and the tubes were incubated at 70°C for 10 min to inactivate the S1 nuclease. The tubes were transferred to ice and a 2 µl aliquot of each was electrophoresed on a 1% agarose gel (Section 2.11.3) to determine the extent of Exo III digestion of the insert. The tubes were transferred to 37°C, 1 µl of Klenow mix was added to each and the tubes were incubated for 3 min. A 1 µl aliquot of dNTP mix was then added to each and the tubes were incubated for an additional 5 min at 37°C. The tubes were then transferred to room temperature and 40 µl of ligase mix was added to each tube. The tubes were incubated at room temperature for 1 h. Alternatively, analysis of each time point on an agarose gel was omitted and the whole of each sample was run on a low melting point agarose gel as described in Section 2.16.3.

Requirements

- 10× Exo III buffer: 660 mM Tris-HCl, pH 8.0, 6.6 mM $MgCl_2$.
- 7.4× S1 buffer: 0.3 M potassium acetate/acetic acid, pH 4.6, 2.5 M NaCl, 10 mM $ZnSO_4$, 50% glycerol.
- S1 mix (for 25 time points): mix 172 µl deionized H_2O, 27 µl of 7.4× S1 buffer and 60 U of S1 nuclease. Make fresh for each experiment.
- S1 stop buffer: 0.3 M Tris base, 0.05 M EDTA.
- 1× Klenow buffer: 20 mM Tris-HCl, pH 8.0, 100 mM $MgCl_2$.
- Klenow mix: mix 30 µl of 1× Klenow buffer and 3 - 5 U of Klenow DNA polymerase. Make fresh for each experiment.
- dNTP mix (supplied with the Promega kit or made up): 0.125 mM each of dATP, dCTP, dGTP, and dTTP (see Section 2.4.1 for 5 mM stocks).
- 10× ligase buffer: 500 mM Tris-HCl, pH 7.6, 100 mM $MgCl_2$, 10 mM ATP (for 0.1 M ATP, see Section 2.15.2).
- Ligase mix: mix 790 µl deionized H_2O, 100 µl of 10× ligase buffer, 100 µl of 50% (w/v) PEG, 10 µl 100 mM DTT (for 1 M DTT, see Section 2.14.1), and 5 U T4 DNA ligase. Make fresh for each experiment.

- Exo III was obtained from New England Biolabs at a concentration of 100 U/µl or was supplied as a component of the Promega kit.
- S1 nuclease was obtained from Boehringer at a concentration of 400 U/µl or was supplied as a component of the Promega kit.
- Klenow DNA polymerase was obtained from Boehringer at a concentration of 5 U/µl or was supplied as a component of the Promega kit.
- T4 DNA ligase was obtained from BRL at a concentration of 0.72 U/µl (Weiss units) or was supplied as a component of the Promega kit.

2.17.2 Filling-in with α-phosphorothioate deoxynucleotides

As an alternative to using 3′ protrusions to block Exo III digestion of vector sequences (Section 2.17.1), 3′ recessed ends were filled-in with α-phosphorothioate deoxynucleotides and Klenow DNA polymerase. The plasmid DNA was digested with a restriction enzyme which generated a 5′ protrusion at a site between the primer binding site and the insert as described in Section 2.17.1. The digested DNA was subsequently purified as described in Section 2.17.1. The DNA was redissolved in 50 µl of 1× Klenow buffer (see Section 2.17.1) and a mixture of all four α-phosphorothioate deoxynucleoside triphosphates was added to a final concentration of 40 µM each. DTT (see Section 2.14.1 for 1 M stock) and Klenow DNA polymerase (see Section 2.17.1), to 1 mM and 50 U/ml respectively, were added and the sample was incubated at 37°C for 10 min. The sample was then heated at 70°C for 10 min to inactivate the Klenow. If the second restriction enzyme was active in 1× Klenow buffer, the enzyme was added to the sample and the DNA was digested as above. Alternatively, the DNA was purified by phenol/chloroform extraction and ethanol precipitation (Section 2.17.1), redissolved in the appropriate buffer and the second restriction enzyme was then

added. Subsequent procedures were as described in Section 2.17.1.

Requirements

- A mixture of all four α-phosphorothioate deoxynucleoside triphosphates at a concentration of 0.4 mM each was obtained from Promega.

2.17.3 Fractionation of Exo III-digested DNAs

As an alternative to direct ligation of Exo III-digested DNAs (Section 2.17.1), the DNA from each time point was filled-in with Klenow and dNTPs as described in Section 2.17.1 and then the whole of each sample was run on a 1% low melting point agarose gel (see Section 2.17.5), as suggested by Nakayama and Nakauchi (1989). The major band in each lane was then excised from the gel under long wavelength u.v. light (see Section 2.11.5) and transferred to a 1.5 ml microfuge tube. The DNA was eluted from each gel slice as described in Section 2.17.6 and redissolved in 16 µl of H_2O. A 4 µl aliquot of 5× ligase buffer and 0.25 µl of ligase (Section 2.15.2) were added to each sample and the samples were incubated at room temperature for 4 h.

2.17.4 Transformation of *E. coli* with Exo III plasmid derivatives

A 10 µl aliquot of each directly ligated plasmid DNA (Section 2.17.2) was mixed in a 1.5 ml microfuge tube with 50 µl of competent *E. coli* cells (Section 2.15.3). Alternatively, 2 µl of each sample of agarose gel-fractionated and religated plasmid DNA (Section 2.17.3) was mixed with 20 µl of competent cells. The sample was mixed gently and incubated on ice for 30 min. The tube was then heated to 42°C for 2 min and placed on ice for 2 min. A 200 µl aliquot of SOC medium (Section 2.15.4) was added and the tube was incubated at 37°C

for 1 h. The entire mixture was then plated out on an LB Amp plate with X-gal and IPTG as described in Section 2.15.5.

2.17.5 Digestion with *Bal*31

The method used to generate deletion derivatives by digestion with *Bal*31 was based on the method of Nixon, (1989). An aliquot of CsCl-purified plasmid DNA (Section 2.16.4), equivalent to 30 µg, was transferred to a 1.5 ml microfuge tube and the DNA was digested with a restriction enzyme with a site near one end of the insert (see Section 2.19). Digestion with a particular restriction enzyme, in a total volume of 200 µl and using 6 - 7 U/µg of DNA, was according to the restriction enzyme manufacturer's instructions. A 1 µl aliquot of the digest was analyzed by agarose gel electrophoresis (Section 2.11.3) to confirm that the digestion had gone to completion. The digest was then extracted with an equal volume of phenol/chloroform (Section 2.22.2) and the DNA was precipitated from the aqueous phase by adding 0.1 volume of 3 M sodium acetate/acetic acid, pH 5.2 and 2 vol. of absolute ethanol. The tube was centrifuged for 20 min and then the pellet was rinsed with 500 µl of 80% ethanol. The tube was centrifuged for 5 min and the pellet was redissolved in H_2O at a concentration of 1 µg/µl.

The sensitivity of the linearized plasmid DNA to digestion with *Bal*31 was determined before proceeding with the actual *Bal*31 digestion. A 2 µl aliquot of the linearized plasmid DNA (2 µg; see above) was mixed with 38 µl of H_2O and 5 µl of 10× *Bal*31 buffer. A 9 µl aliquot was transferred to each of nine 1.5 ml microfuge tubes and 1 µl of 1/10, 1/20, 1/40, or 1/80 *Bal*31 nuclease, diluted in *Bal*31 buffer, was added to each tube. The tubes were incubated at 37°C for 30 min and then the reactions were stopped by adding 1 µl of 200 mM EGTA to each tube and heating the tubes at 65°C for 5 min. The samples were analyzed by agarose gel electrophoresis (Section

2.11.3) to determine the dilution of *Bal*31 nuclease which completely digested the DNA in the 30 min incubation period.

The linearized plasmid DNA was digested with *Bal*31 nuclease, using the amount of enzyme determined above for each 2 µg of DNA. A 26 µl aliquot of the linearized DNA was mixed with 559 µl of H_2O and 65 µl of 10× *Bal*31 buffer and the tube was incubated at 37°C for 10 min. A 5 µl aliquot of 200 mM EGTA was added to a series of thirteen 1.5 ml microfuge tubes. The determined amount of *Bal*31 nuclease was added to the tube of linearized DNA at 37°C, the tube was vortexed gently and returned to 37°C. At 1 min intervals, a 45 µl aliquot of the digestion was transferred to the appropriate time point tube containing 200 mM EGTA at room temperature. When all the time points had been taken, the tubes were incubated at 65°C for 5 min. The DNA in each tube was then precipitated by adding 5 µl of 3 M sodium acetate/acetic acid, pH 5.2 and 110 µl of ethanol. The tube was centrifuged for 20 min and the DNA pellet was rinsed with 80% ethanol and redissolved in 17 µl of H_2O. A 2 µl aliquot of the appropriate 10× restriction enzyme buffer (Section 2.19) and 1 µl of a restriction enzyme having a site at the end of the insert opposite to that digested with *Bal*31 were added. The digests were incubated at 37°C (see Section 2.19) and then electrophoresed on a 0.7% low melting temperature agarose gel. The major band in each lane was individually excised from the gel under long wavelength u.v. light (see Section 2.11.5) and transferred to a 1.5 ml microfuge tube. The DNA was eluted from each gel slice and redissolved in 30 µl of H_2O as described in Section 2.17.6

Requirements

- 10× *Bal*31 buffer: 0.3 M Tris-HCl, pH 8.0, 50 mM $MgCl_2$, 50 mM $CaCl_2$, 3 M NaCl.
- *Bal*31 nuclease at a concentration of 2.5 U/µl was obtained from New England Biolabs.

- Low melting temperature agarose gel: this was poured as described in Section 2.11.5 except that a standard comb was used.

2.17.6 Elution of DNA fragments from agarose

The method used to elute Exo III- and *Bal*31-digested DNA fragments from gel slices (Sections 2.17.3 and 2.17.5) was similar to that described in Section 2.11.7 for the routine purification of DNA fragments from low melting temperature agarose but on a smaller scale. Each tube containing a low melting temperature agarose gel slice was weighed to determine the volume of the gel slice and then TE buffer (Section 2.22.4) was added to a total volume of 400 µl. The tube was incubated at 65°C until the agarose had completely melted. TE-saturated phenol (1 ml; Section 2.22.1) was then added and the tube was vortexed for 30 sec and centrifuged for 5 min. The aqueous phase, combined with the milky interface, was transferred to another tube and the phenol extraction was repeated. The DNA was precipitated from the aqueous phase by adding 0.1 volume of 3 M sodium acetate/acetic acid, pH 5.2 and 2 vol. of absolute ethanol and the tube was centrifuged for 20 min. The pellet was rinsed with 80% ethanol, redissolved in H_2O and stored at 4°C.

2.17.7 Preparation of plasmid vector for cloning *Bal*31-digested fragments

An aliquot of PEG-purified pGEM-3Z DNA (Section 2.16.3) containing 10 µg of DNA was mixed with 5 µl of 10× *Sma*I buffer, 48 U of *Sma*I restriction enzyme and H_2O to a final volume of 50 µl. The reaction was incubated at 37°C for 1 h and was then extracted with phenol/chloroform (Section 2.22.2). The DNA was precipitated with ethanol (Section 2.21) and the pellet was rinsed with 80% ethanol and redissolved in 40 µl of H_2O. The linearized vector was then digested, at 37°C in a total volume of 50µl, with the same enzyme used to cut

out the *Bal*31-digested inserts (Section 2.17.5). A 0.5 µl aliquot of phosphatase (see Section 2.15.1) was added for the last 20 min of the digestion. The reaction was then extracted with an equal volume of phenol/chloroform and the DNA was again precipitated with ethanol. The DNA pellet was rinsed with 80% ethanol and redissolved in 50 µl of TE (Section 2.22.4). The solution of doubly digested DNA was stored at -20°C until required.

Requirements

- 10× *Sma*I buffer: 330 mM Tris-acetate, pH 7.9, 100 mM magnesium acetate, 660 mM potassium acetate, 5 mM DTT (from 1 M stock, Section 2.14.1).
- *Sma*I restriction enzyme at a concentration of 12 U/µl was obtained from Boehringer.

2.17.8 Ligation of *Bal*31-digested insert fragments with plasmid vector

A 7 µl aliquot of each eluted *Bal*31-digested fragment (Section 2.17.6) was mixed with 1 µl of prepared vector DNA (Section 2.17.7), 2 µl of 5x ligase buffer and 1 µl of ligase (Section 2.15.2). The ligations were incubated at 14°C overnight.

2.17.9 Transformation of *E. coli* with *Bal*31 plasmid derivatives

The entire ligation reaction derived from each *Bal*31 time point (Section 2.17.8) was used to transform 50 µl of competent *E. coli* DH5α (Section 2.15) as described in Section 2.17.4.

2.17.10 Characterization of Exo III and *Bal*31 deletion derivatives

A number of white colonies (typically 12; more for non-fractionated Exo III derivatives) resulting from each Exo III or *Bal*31 digestion time point (Sections 2.17.4, 2.17.5, and 2.17.9) were patched out on an LB Amp plate (see Sections 2.15.4 and 2.15.6) and the plate was incubated overnight at 37°C. Positive transformants were characterized as described in Section 2.15.6. Glycerol stocks of positive isolates were prepared from 500 µl of an overnight culture as described in Section 2.15.8. Plasmid DNA for sequencing was prepared from the remainder of the overnight culture as described in Section 2.18.1.

2.18 DNA sequencing of plasmid DNA

Modifications of the original enzymatic chain termination method (Sanger *et al.*, 1977) were used to sequence cloned DNA. For routine sequencing, T7 DNA polymerase was used as the sequencing enzyme but *Taq* DNA polymerase was used to sequence long poly(A) tracts. Recombinant plasmid DNA, either CsCl purified (Section 2.16.2) or from a rapid plasmid preparation (see below), was used directly as a template for all DNA sequencing.

2.18.1 Preparation of template DNA by alkaline lysis

Most sequence data were obtained from plasmid DNA prepared using a method based on the methods of Hattori and Sakaki (1986) and Mierendorf and Pfeffer (1987). A positive transformant (Sections 2.15.6 and 2.17.10) was inoculated from the original patch culture, or from a single colony derived from that patch culture, into 5 ml of LB Amp medium (Section 2.15.8) in a universal bottle.

Following overnight incubation at 37°C with shaking at approx. 200 rpm, the universal bottle was centrifuged in a Sorvall SS-34 rotor, fitted with a rubber cushion, at 4,000 rpm and 4°C for 5 min. The supernatant was removed and the cell pellet was resuspended in 100 µl of ice-cold lysis buffer and transferred to a 1.5 ml microfuge tube. The tube was incubated for 5 min at room temperature and then 200 µl of a freshly prepared solution of 0.2 M NaOH, 1% SDS was added (see Section 2.22.6 for 10% SDS stock). The tube was mixed by inversion, incubated on ice for 5 min and then 150 µl of ice-cold potassium acetate solution (Section 2.16.2) was added. The tube was again mixed by inversion, incubated on ice for 5 min and centrifuged for 10 min. The supernatant was transferred to a fresh tube, RNaseA was added to a final concentration of 100 µg/ml and the tube was incubated at 37°C for 20 min. An equal volume of phenol/chloroform (Section 2.22.2) was then added, the tube was vortexed for 30 sec and centrifuged for 5 min. The aqueous phase was transferred to a fresh 1.5 ml tube and 2.5 vol. of of absolute ethanol were added. The tube was vortexed briefly, centrifuged for 20 min and the supernatant was removed. The DNA pellet was rinsed with 1 ml of cold, 80% ethanol, the tube was centrifuged for 5 min and the supernatant was again removed. The DNA pellet was redissolved in 48 µl of H_2O with heating at 65°C for 10 min. The tube was then placed on ice and 12 µl of 4 M NaCl and 60 µl of 13% PEG were added. The tube was vortexed and incubated on ice for at least 20 min. The tube was then centrifuged for 15 min, the supernatant was removed and the DNA pellet was rinsed with 80% ethanol as above. The DNA pellet was dried for 5 min under vacuum and redissolved in 20 µl H_2O. The amount of DNA present was estimated by agarose gel electrophoresis (Section 2.11.3) of a 0.5 µl aliquot. The DNA sample was stored at -20°C until required.

Requirements

- Lysis buffer: 50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl, pH 8.0. Immediately before use, add lysozyme to 2 mg/ml to an aliquot of buffer sufficient for the number of cultures to be processed.

2.18.2 Preparation of template DNA by boiled lysis

Towards the end of this study and based on the observations of Toneguzzo *et al.* (1988), some sequence data were obtained from template DNA prepared using the LiCl-boiling method (Section 2.15.7). DNA was prepared from a 5 ml overnight culture as described (Section 2.15.7) and redissolved in 50 µl of H_2O. The amount of DNA present was estimated by agarose gel electrophoresis (Section 2.11.3) of a 1 µl aliquot. The DNA sample was stored at -20°C until required.

2.18.3 Alkali denaturation of template DNA

Template DNA was made single-stranded by treatment with alkali, primarily according to the method of Mierendorf and Pfeffer (1987). An aliquot of template DNA (Sections 2.16.2, 2.18.1, 2.18.2), equivalent to 2 - 4 µg, was transferred to a 0.6 ml microfuge tube. H_2O was added to bring the volume up to 20 µl and then 2 µl of 2 M NaOH, 2 mM EDTA was added and the tube was incubated for 5 min at room temperature. The solution was neutralized by adding 3 µl of 3 M sodium acetate/acetic acid, pH 5.2, and then 7 µl of H_2O was added and the contents were mixed. Absolute ethanol (75 µl) was added to precipitate the DNA and the tube was vortexed briefly and centrifuged for 20 min. The supernatant was removed and the pellet was rinsed with 200 µl of cold 80% ethanol. The tube was centrifuged for 5 min, the supernatant was removed and the pellet was dried for 5 min under vacuum. The dried DNA was either immediately redissolved in H_2O and

annealed with a sequencing primer (Section 2.18.4) or stored at -20°C until required.

During the course of this study, several modifications were made to the above protocol. Based on the method of Zhang *et al.* (1988), the denatured DNA was neutralized by adding 2 µl of 2 M ammonium acetate/acetic acid, pH 4.6, and immediately precipitated by adding 60 µl of absolute ethanol. Template DNA prepared by using the LiCl-boiling method (Section 2.18.2) was alkali denatured with an incubation at 85°C for 5 min according to the method of Toneguzzo *et al.* (1988).

2.18.4 Annealing of sequencing primer and denatured template

The denatured DNA (Section 2.18.3) was dissolved in 7 µl (for T7 DNA polymerase sequencing; Section 2.18.5) or 10 µl (for Taq DNA polmerase sequencing; Section 2.18.6) of H_2O and 1 µl of sequencing primer was added to the tube. The tube was vortexed and incubated at 65°C for 5 min. The tube was then centrifuged briefly and 2 µl of 5× T7 sequencing buffer or 2 µl of Taq DNA polmerase reaction buffer (Section 2.18.6) was added. The tube was vortexed again and incubated at 37°C for 10 - 15 min. The annealed template was then used immediately for a set of sequencing reactions.

Requirements

- Sequencing primer: SP6 (19mer), T3 (20mer) and T7 (20mer) promoter primers were from Promega or New England Biolabs at a concentration of 10 ng/µl.
- 5× T7 sequencing buffer: 200 mM Tris-HCl, pH 7.5, 50 mM $MgCl_2$, 250 mM NaCl.

2.18.5 Sequencing with T7 DNA polymerase

Sequencing with T7 DNA polymerase was carried out with either the Sequenase kit (USB) or the T7 DNA Polymerase Sequencing System (Promega). The procedure used followed the instructions supplied with each kit and the method of Tabor and Richardson (1987).

For each set of four sequencing reactions (A, C, G, and T), a sequencing mix was prepared on ice as follows:

100 mM DTT	1.0 µl
diluted labelling mix	2.0 µl
[α-^{35}S] dATP	0.5 µl
diluted T7 DNA polymerase	2.0 µl

An aliquot of 5.5 µl of the sequencing mix was mixed with each annealed template-primer (Section 2.18.4) by pipetting the solution up and down. The tube was centrifuged briefly and incubated at room temperature for 5 - 10 min (labelling reaction).

Four 1.5 ml microfuge tubes were labelled A, C, G, and T and 2.5 µl of the appropriate nucleotide mix was added to each of the four tubes. These tubes were pre-warmed at 37°C for at least 1 min and then 3.5 µl of the labelling reaction was added to each. The tubes were vortexed briefly and incubated at 37°C for 5 - 10 min (termination reactions). A 4 µl aliquot of stop solution was then added to each of the termination reactions. The tubes were vortexed and then stored either on ice (hours) or at -20°C (days) until they were ready to be loaded on a sequencing gel (Section 2.18.8).

Requirements

- 100 mM DTT: supplied with kit.

- Diluted labelling mix: dilute labelling mix (7.5 µM dCTP, 7.5 µM dGTP, 7.5 µM dTTP; supplied with kit) 5-fold with H_2O.
- [α-^{35}S] dATP: 10 mCi/ml, 1000 Ci/mmol from Amersham or New England Nuclear was stored at -20°C and used within 4 - 6 weeks.
- Diluted T7 DNA polymerase: dilute Sequenase (USB) 1:8 or T7 DNA polymerase (Promega) to 1.5 U/µl in cold TE (Section 2.22.4) immediately before use.
- Nucleotide mix (supplied with kit): A mix is 80 µM each of the four dNTPs, 8 µM (USB) or 6 µM (Promega) ddATP; C mix is 80 µM each dNTP, 8 µM ddCTP; G mix is 80 µM each dNTP, 8 µM ddGTP; T mix is 80 µM each dNTP, 8 µM (USB) or 9 µM (Promega) ddTTP (dNTP: deoxynucleoside triphosphate; ddNTP: dideoxynucleoside triphosphate).
- Stop solution (supplied with kit): 95% formamide, 0.05% (USB) or 0.01% (Promega) bromophenol blue, 0.05% (USB) or 0.01% (Promega) xylene cyanol, 20 mM EDTA.

2.18.6 Sequencing with *Taq* DNA polymerase

Sequencing with *Taq* DNA polymerase was carried out using the TAQuence DNA sequencing kit (USB) following the instructions supplied with the kit. The following were added to the annealed template-primer (Section 2.18.4) on ice:

Template-primer	13.0µl
Labelling mix	2.0µl
[α-^{35}S] dATP	0.5µl
Diluted *Taq* DNA polymerase	2.0µl

The tube was vortexed briefly and incubated at 45°C for 5 min (labelling reaction).

Four 1.5 ml microfuge tubes were labelled A, C, G, and T and a 4 µl aliquot of the appropriate nucleotide mix was added to each of the four tubes. A 4 µl aliquot of the labelling

reaction was then added to each of the four tubes and the tubes were vortexed briefly and incubated at 70°C for 5 min (termination reactions). The reactions were cooled to room temperature and 4 µl of stop solution was added to each. The tubes were vortexed and stored on ice until they were ready to be loaded on a sequencing gel (Section 2.18.8).

Requirements

- Labelling mix (supplied with kit): 1.5 µM dCTP, 1.5 µM dGTP, 1.5 µM dTTP.
- [α-^{35}S] dATP: see Section 2.17.5.
- Diluted *Taq* DNA polymerase: dilute *Taq* DNA polymerase (supplied with kit) 1:8 in ice-cold enzyme dilution buffer (supplied with kit).
- Nucleotide mix: supplied with kit.
- Stop solution (supplied with kit): 95% formamide, 0.05% bromophenol blue, 0.05% xylene cyanol, 20 mM EDTA.

2.18.7 Preparation of sequencing gels

A pair of glass plates (for Model S2, BRL) was cleaned with ethanol and allowed to dry. One surface of the shorter (back) plate was wiped over with a piece of paper towel soaked with a 2% (v/v) solution of dimethyl-dichlorosilane in chloroform, in a fume cupboard and wearing gloves. This plate was allowed to dry for 10 min and then it was wiped over with a piece of paper towel soaked with deionized H_2O. Both plates were wiped over with a dry, lint-free cloth and then taped together, separated by 0.4 mm (standard) or 0.4 - 1.2 mm (wedge) side spacers, with Sleek tape. The assembly was placed in a 40 × 50 cm tray. The gel was poured by tilting the assembly at approx. 30° from the horizontal and sloped slightly to one side and pouring the sequencing gel mix down the lower side. The plates were completely filled and were then laid down in the tray so that the top end was slightly elevated. Two sharkstooth combs (standard or doublefine, BRL) were inserted

between the plates, flat edge first, to a depth of approx. 3 mm. The top corners of the plates were clamped, near the combs, with 2 fold-back spring clips and the gel was allowed to set for at least one hour. The gel was then left at room temperature until required (Section 2.18.7). If the gel was to be left overnight, the top was wrapped in polyethylene film. Before use, the tape was removed from around the glass plates and the combs were removed and rinsed with deionized H_2O.

Between runs, glass plates, combs and side spacers were soaked in a solution of laboratory cleaner (Pyroneg). The glass plates were rubbed with a synthetic pot cleaner and were then rinsed sequentially with warm H_2O and distilled H_2O and allowed to dry. The combs and side spacers were also rinsed with H_2O and allowed to dry.

Requirements

- Weigh out acrylamide in a fume cupboard wearing gloves; wear gloves whenever handling acrylamide solutions.
- 40% acrylamide stock (500 ml): dissolve 190 g acyrlamide (Electran, BDH) and 10 g NN'-methylenebisacrylamide (GPR, BDH) in H_2O with stirring. Adjust to a final volume of 500 ml with H_2O. Filter solution through a 0.45 µm filter (cellulose nitrate, Sartorius) under vacuum and store at 4°C.
- 6% sequencing gel mix - standard gel (31 × 38.5 cm; 0.4 mm spacers): mix 15 ml of 40% acrylamide stock, 50 g urea (analytical reagent), 10 ml 10× TBE electrophoresis buffer (Section 2.17.7), and 35 ml of H_2O on a shaker at 37°C until the urea has dissolved. Filter the solution through a 0.45 µm filter (cellulose nitrate, Sartorius) under vacuum. Add 1 ml of 10% ammonium persulphate solution and 20 µl of TEMED (NNN'N'-tetramethylethylenediamine; GPR, BDH) immediately before use and mix briefly on a magnetic stirrer.

- 6% sequencing gel mix - wedge-shaped gel (0.4 - 1.2 mm spacers): as for standard gel but use 22 ml of 40% acrylamide stock, 75 g urea, 15 ml of 10× TBE electrophoresis buffer and 149 ml of H_2O. Add 1.5 ml of TEMED and 3 ml of 10% ammonium persulphate solution.
- 10% ammonium persulphate solution: dissolve 0.5 g ammonium persulphate (AnaLar, BDH) in 5 ml of H_2O. Prepare fresh for each gel, or set of gels, to be poured.

2.18.8 Denaturing gel electrophoresis

A prepared 6% sequencing gel (Section 2.18.7) was assembled in the sequencing apparatus (Model S2, BRL) and 500 ml of TBE electrophoresis buffer was added to both the top and the bottom buffer chambers. The top of the gel was rinsed with buffer and then the sharkstooth combs were inserted between the glass plates so that the teeth of each comb made slight indentations in the gel. The gel was pre-run for 15 - 60 min at a constant power of 65 watts. The sequencing reactions (Sections 2.18.5 and 2.18.6) were heated at 80°C for 4 min immediately before use and then placed on ice. Before loading each set of four reactions onto the gel, four wells were rinsed with buffer. An aliquot of either 3 µl (for standard sharkstooth comb; BRL) or 1.5 µl (for doublefine sharkstooth comb; BRL) from each reaction was loaded onto the gel using a sequencing pipette tip (Costar, No. 4854) attached to a P20 Pipetman (Gilson) or a Finnpipette Digital, 0.5 - 10 µl (preferably the latter) in the order A, C, G, T. The gel was run at a constant power of 65 watts.

To maximize the amount of data generated from each set of sequencing reactions, duplicate loadings of each reaction were electrophoresed on either the same gel or on separate gels. If electrophoresed on the same gel, the first loading was electrophoresed for 3 h, the reactions were reheated at 80°C for 4 min, placed on ice and a second aliquot was loaded onto the gel. The gel was then electrophoresed for an

additional 2 h or until the blue dye from the second loading had reached the bottom of the gel. If electrophoresed on separate gels, duplicate loadings were made at the same time and one gel was electrophoresed for 2 h, or until the blue dye had reached the bottom of the gel, and the other for 5 h.

As an alternative to duplicate loadings, wedge-shaped gels were sometimes used (Section 2.18.7) but these did not yield as much sequence data. In addition to the usual duplicate electrophoresis runs, extended electrophoresis runs of up to 10 h were sometimes required. The sequencing reactions were stored at -20°C after use in case additional electrophoresis runs were subsequently required.

Requirements

- TBE electrophoresis buffer: mix 100 ml of 10× TBE buffer with 900 ml of H_2O.
- 10× TBE electrophoresis buffer: 121 g Tris base, 55 g boric acid, 7.4 g EDTA (di-sodium salt). Dissolve in H_2O and adjust final volume to 1 litre. Store at room temperature.

2.18.9 Fixing, drying, and autoradiography of sequencing gels

The sequencing gel (Section 2.18.8) was removed from the sequencing apparatus and laid, back plate uppermost, on the bench. The combs and the side spacers were removed and the glass plates were gently prised apart with a small spatula. The top plate was carefully lifted off the gel and the bottom plate with the gel attached was placed into a tray containing 3 litres of 5% methanol, 5% acetic acid fixing solution. The gel was fixed for at least 15 min for a standard gel and 60 min for a wedge-shaped gel (gels were left overnight in fixing solution if necessary). The glass plate with the gel on top of it, but not necessarily still attached, was care-

fully lifted out of the tray and placed on the bench. Two pieces of Whatman 3MM paper were cut to the size of the glass plate and one of them was wetted in the fixing solution and laid onto the gel. A wad of paper towels was used to soak up excess solution from the 3MM paper on the gel. A corner of the 3MM paper was then lifted and, if the gel was stuck to the paper, the paper and the gel were completely removed from the glass plate. If the gel was not stuck to the 3MM paper, another corner was tried. The gel was transferred, paper-side down, onto the second piece of 3MM paper and the gel was covered with polyethylene film. The edges of the paper and the polyethylene film were trimmed off and the gel was dried in a vacuum drier (Bio-Rad) at 80°C for at least 1 h or until no cold spots could be felt on the neoprene gasket. The polyethylene film was then removed and the dried gel was transferred to an autoradiography cassette. A sheet of X-ray film (Fuji RX) was placed over the gel in a darkroom (no intensifying screen) and the cassette was closed. The gel was autoradiographed at room temperature for 48 - 62 h and then the X-ray film was developed as described in Section 2.6.10.

2.18.10 Resolution of band compressions

Some DNA templates containing strong secondary structure were not fully denatured during electrophoresis (Section 2.18.8). This resulted in abnormal migration of some DNA fragments, evidenced as band compressions on the subsequent autoradiograph. One method used to eliminate sequence compressions involved the substitution of the nucleotide analogue 7-deaza dGTP for dGTP in the T7 DNA polymerase labelling and nucleotide mixes (Section 2.18.5) The deaza dGTP solutions were obtained from Promega (Deaza Reagent Kit).

An alternative method to eliminate sequence compressions involved the use of a 7 M urea, 40% formamide sequencing gel in place of the standard 8 M urea sequencing gel (Section 2.18.7) as described by Martin (1987). Such a gel required

the use of 80 µl of TEMED to achieve the normal rate of polymerization but running conditions and subsequent treatment were the same as a standard gel (Sections 2.18.8 and 2.18.9).

2.18.11 Handling of sequence data

Autoradiograms of sequencing gels (Section 2.18.9) were placed on a sheet of white paper or on a light box and the DNA sequence was read off. The sequence was then entered, as an individual file, into an IBM PC compatible computer in ASCII format using a text editor (Sidekick Notepad, Borland International, or PC-Write, Version 3.03, Quicksoft). Sequence overlaps between adjoining sequences were determined visually or with sequence alignment software (Nucaln, National Institutes of Health, or Align, Scientific and Educational Software). Files containing adjoining sequences were combined and the final sequence was assembled using PC-Write. Sequences of complementary strands were compared using the sequence alignment software. Restriction enzyme sites were determined and restriction enzyme maps were produced with mapping software (Clone 3, Scientific and Educational Software).

2.18.12 Sequence alignment

Pairs of sequences were aligned and the percent homology between the two sequences was calculated with the program Align (Scientific and Educational Software). Multiple sequences were aligned and dendrograms were constructed with the Clustal suite of programs (Higgins and Sharp, 1988; Higgins and Sharp, 1989). All of these programs run on an IBM PC compatible computer; the use of at least an AT class computer is desirable for the Clustal programs though not essential.

2.19 Restriction enzyme digestion of DNA

DNA samples were digested with restriction enzymes, using either the buffer supplied by the manufacturer with each enzyme or an alternative buffer (L, M, or H) giving as close to 100% enzyme activity as possible. When the DNA was to be digested with more than one enzyme, the digest requiring least salt was carried out first and then additional salt and the second enzyme were added. For multiple digests where this was not possible, the DNA was purified with Geneclean (Section 2.11.8) after the first digest and then digested with the second enzyme. Digests were carried out in 1.5 ml microfuge tubes, either in a constant temperature block heater for digests of 1 - 2 h or in a laboratory incubator for digests of a longer duration. Digests were stopped by adding sample loading buffer (Section 2.11.3) and were then electrophoresed, either on a standard agarose gel as described in Sections 2.11.3, or on a preparative agarose gel as described in Section 2.11.5. Digests of λ DNA were usually heated at 65°C for 5 min to denature the λ cohesive ends before being loaded on the gel.

Requirements

- 10× L Buffer: 100 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$.
- 10× M Buffer: 500 mM NaCl, 100 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$.
- 10× H Buffer: 1 M Nacl, 500 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$.

2.20 Spectrophotometric quantitation of DNA and RNA

Solutions of DNA or RNA were quantitated by measuring the A_{260} of each solution. The concentration of nucleic acid in the sample was calculated on the basis that 1 A_{260} unit of double-stranded DNA corresponds to approx. 50 µg/ml while 1 A_{260} unit

of single-stranded RNA corresponds to approx. 40 µg/ml. DNA solutions were usually diluted 1/20 in H_2O for A_{260} measurements.

2.21 Ethanol precipitation of DNA

In general, DNA was precipitated from solution with ethanol as the sodium salt. This was achieved by adding either 1/20 volume of 5 M NaCl or 1/10 volume of 3 M sodium acetate-/acetic acid, pH 5.2 and 2 vol. of absolute ethanol to the DNA sample. Alternatively, the DNA was precipitated as the ammonium salt by adding 1/2 volume of 7.5 M ammonium acetate and 2 vol. (DNA sample plus ammonium acetate) of absolute ethanol (Crouse and Amorese, 1987). The sample was then immediately centrifuged at 12,000 *g* for 20 min at room temperature. When the highest yield of precipitated DNA was required, samples were centrifuged in 1.5 ml microfuge tubes in a Sorvall SS-34 rotor at 15,000 rpm for 20 min at 4°C, using a microfuge tube adaptor (Sorvall, No. 00381). For samples centrifuged in microfuge tubes, the supernatant was removed by aspiration through a fine pipette tip (Costar, No. 4853) connected to a water-jet aspirator pump. For larger volumes, a Pasteur pipette was used in place of the pipette tip. The DNA pellet was then rinsed with a volume of 70% or 80% ethanol, pre-chilled to -20°C, approx. equal to the volume of the original supernatant. The tube was then centrifuged for 5 min and the supernatant was removed as before. The tube was then centrifuged briefly and any residual supernatant was also removed. The DNA pellet was either redissolved directly or was dried for 3 - 5 min under vacuum and then redissolved in H_2O or in an appropriate buffer.

2.22 Buffers and solutions

Buffers and solutions for specific applications are listed under Requirements at the end of each relevant section.

2.22.1 Buffer saturated phenol

The preparation of buffer saturated phenol was based on the method of Grinsted and Bennett, (1988). Add 7.5 ml of 2 M NaOH, 130 ml of H_2O, 6 ml of 1 M Tris-HCl, pH 7.5 to a 500 g bottle of commercial phenol and leave overnight to liquefy. This gives a solution of pH 7.5 and 10 mM Tris. When working with small volumes of sample, it is important to use phenol that is completely saturated with buffer. This can be achieved by overlaying the phenol with TE (Section 2.22.4).

2.22.2 Phenol/chloroform

Mix equal volumes of phenol and chloroform and overlay with TE (Section 2.22.4).

2.22.3 Chloroform

Add octanol to chloroform to a concentration of 4% (v/v).

2.22.4 TE

TE is 10 mM Tris-HCl, pH 8.0, 1 mM EDTA. Autoclave to sterilize.

2.22.5 20× SSC

Dissolve 175.3 g of NaCl and 88.2 g of sodium citrate in 800 ml of H_2O. Adjust pH to 7.0 with NaOH. Adjust volume to 1 litre. This gives a solution of 3 M NaCl, 0.3 M sodium citrate. Autoclave to sterilize.

2.22.6 10% SDS

Dissolve 25 g of sodium dodecyl sulphate (SDS) in 175 ml of H_2O with heating to 65°C. Adjust pH to 7.2 with HCl. Adjust

volume to 250 ml. For RNA work, incubate at 65°C for 1 - 2 hr to inactivate RNases.

Chapter 3

Results and Discussion: RNA Analysis and cDNA Cloning

3.1 Isolation of white clover RNA

Initially, RNA was isolated from both leaf/stem and root material using a standard phenol extraction procedure (Section 2.2.1). Yields from leaf/stem material were subsequently improved by including an additional homogenization step using a Polytron, with maximum yields achieved with the Polytron set at maximum speed (Table 3.1). However, leaf/stem RNA preparations using the phenol extraction procedure tended to be contaminated with a gelatinous material which subsequently led to difficulties in separating poly $(A)^+$ RNA from the other species of RNA on oligo (dT) cellulose columns (Section 2.3.1). This was overcome by using a guanidine hydrochloride procedure (Section 2.2.2) to isolate RNA from leaf/stem material. The guanidine hydrochloride procedure resulted in a yield of RNA similar to that obtained using the phenol extraction procedure when additional homogenization using a Polytron was included (Table 3.1).

As can be seen from Table 3.1, optimum RNA yields were achieved using either the phenol extraction procedure or the guanidine extraction procedure with a brief homogenization with a Polytron set at max. speed.

Table 3.1. Yields of RNA from white clover.

Clover material	Isolation procedure	RNA yield (µg/g material)
leaf/stem #1	standard phenol	16.4
leaf/stem #2	standard phenol	19.5
root #1	standard phenol	19.5
root #2	standard phenol	17.4
leaf/stem #3	phenol + Polytron (set at 5)	112
leaf/stem #4	phenol + Polytron (set at max)	198
leaf/stem #5	guanidine + Polytron (set at max)	176

3.2 Northern blot analysis of leaf/stem and root RNAs

Since one of the objectives of this study was the isolation of white clover cDNA clones using heterologous probes, it was first necessary to determine whether or not these probes would hybridize with white clover mRNA species. Samples of mRNA-enriched leaf/stem and root RNA were electrophoresed on a formaldehyde agarose gel (Section 2.11.2) and a Northern blot was made from this gel (Section 2.12.1). The blot was probed sequentially with a petunia Rubisco SSU cDNA clone (pSSU117; Dunsmuir *et al.*, 1983), a pea cDNA clone of Adh1 (pPsR546; Llewellyn *et al.*, 1987) and a pea seed lectin cDNA clone (pPS15-104; Higgins *et al.*, 1983).

The petunia SSU cDNA probe hybridized with a single mRNA species of approx. 800 bp in the leaf/stem poly(A)$^+$ RNA lane but no hybridization was detected in the root poly(A)$^+$ RNA

lane (Figure 3.1). The pea Adh1 cDNA probe hybridized with a single mRNA species of approx. 1400 bp in the root poly(A)$^+$ RNA lane (Figure 3.2). The background hybridization apparent in the leaf/stem poly(A)$^+$ RNA lane probed with the pea Adh cDNA (Figure 3.2) probably represents residual SSU probe which had not been completely removed prior to the pPsR546 hybridization experiment. The pea lectin probe hybridized with two mRNA species of approx. 1600 and 1400 bp in the root poly(A)$^+$ RNA lane and to three mRNA species of approx. 3000, 1600, and 640 bp in the leaf/stem poly(A)$^+$ RNA lane (Figure 3.3). The low level of smearing below the predominant bands of hybridization in all three Northern hybridization experiments indicated that minimal degradation of poly(A)$^+$ RNA had occurred during RNA isolation.

3.3 cDNA synthesis

Preliminary experiments were carried out to check the efficiency of the first and second strand cDNA synthesis reactions using a commercial preparation of a mixture of poly(A)$^+$ RNAs (Section 2.4). The standard cDNA synthesis procedure involved purification of the first strand synthesis reaction products by extraction with phenol/chloroform and precipitation with ethanol before proceeding with the second strand synthesis reaction. The first and second strand synthesis reaction products were analyzed on a gel and the gel was subsequently autoradiographed (Figure 3.4). It was evident from Figure 3.4 that the first strand synthesis reaction products consisted of a set of discrete fragments of the expected size but the second strand synthesis reaction products did not consist of the corresponding set of discrete fragments and instead resulted in the smear of bands evident in this autoradiograph. It was concluded from this experiment that the DNA polymerase in the second strand synthesis reaction was being inhibited by residual phenol, and as a result, a range of incomplete second strands was synthesized.

Figure 3.1. Hybridization of a petunia SSU cDNA probe (pSSU117) to a Northern blot of white clover mRNA. Lane 1: root mRNA; lane 2: leaf/stem mRNA.

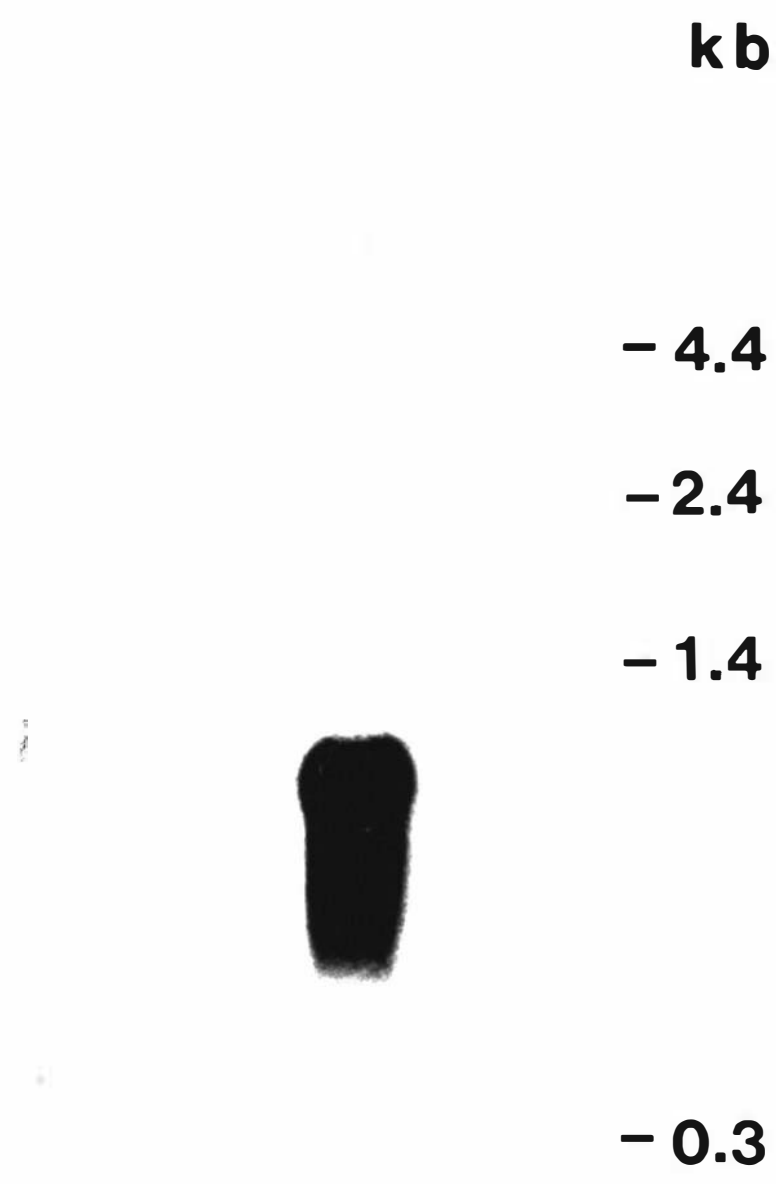

kb
– 4.4
– 2.4
– 1.4
– 0.3
1
2

Figure 3.2. Hybridization of a pea Adh1 cDNA probe (pPsR546) to a Northern blot of white clover mRNA. Lane 1: root mRNA; lane 2: leaf/stem mRNA.

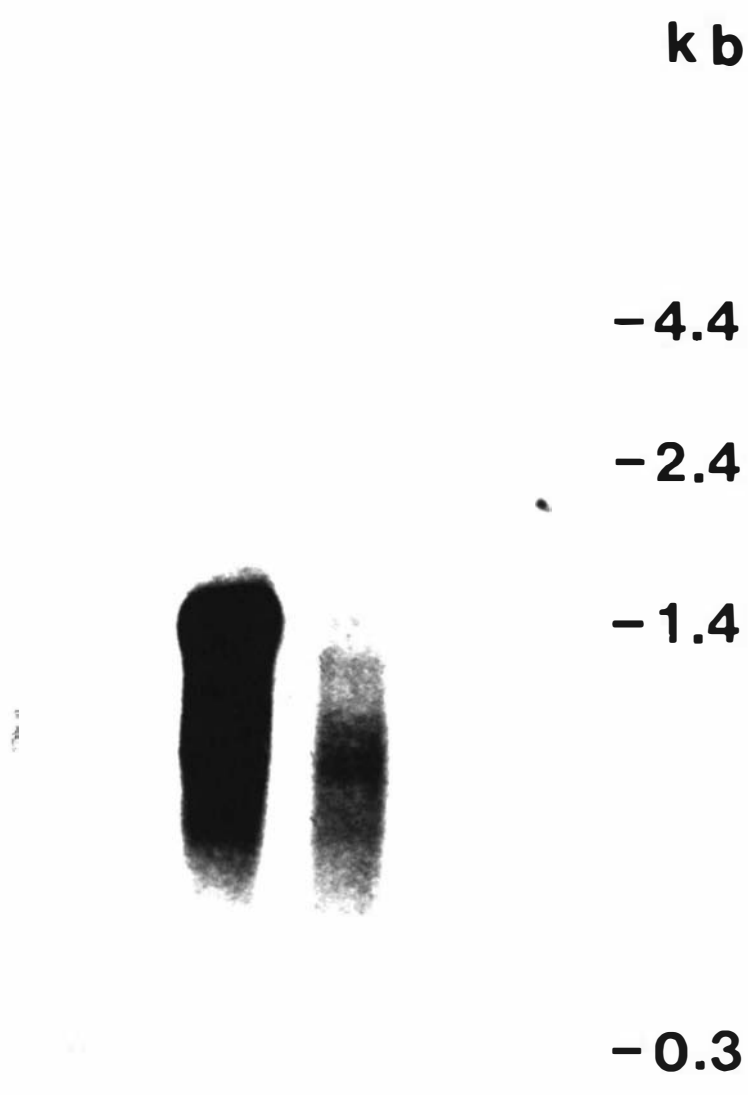

kb
4.4
2.4
1.4
0.3
1
2

Figure 3.3. Hybridization of a pea lectin cDNA probe (pPS15-104) to a Northern blot of white clover mRNA. Lane 1: root mRNA; lane 2: leaf/stem mRNA.

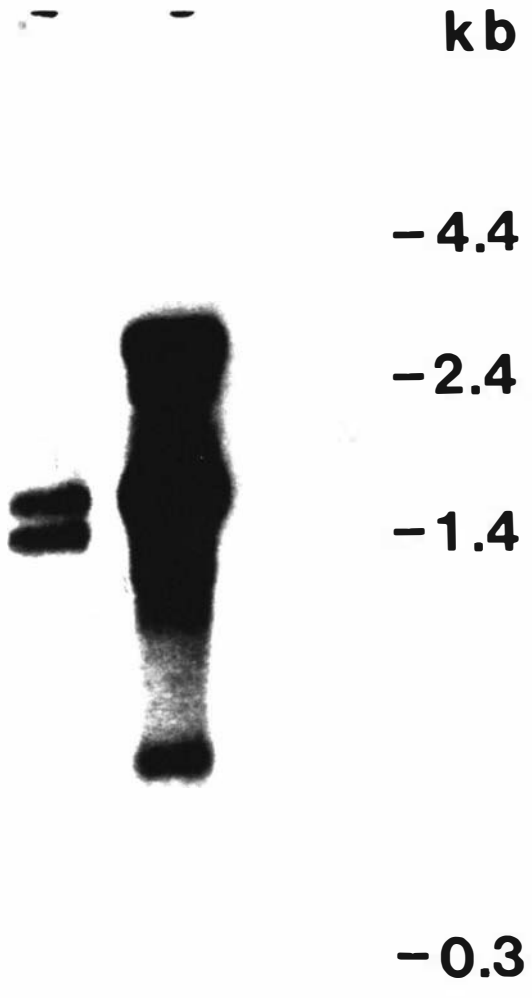
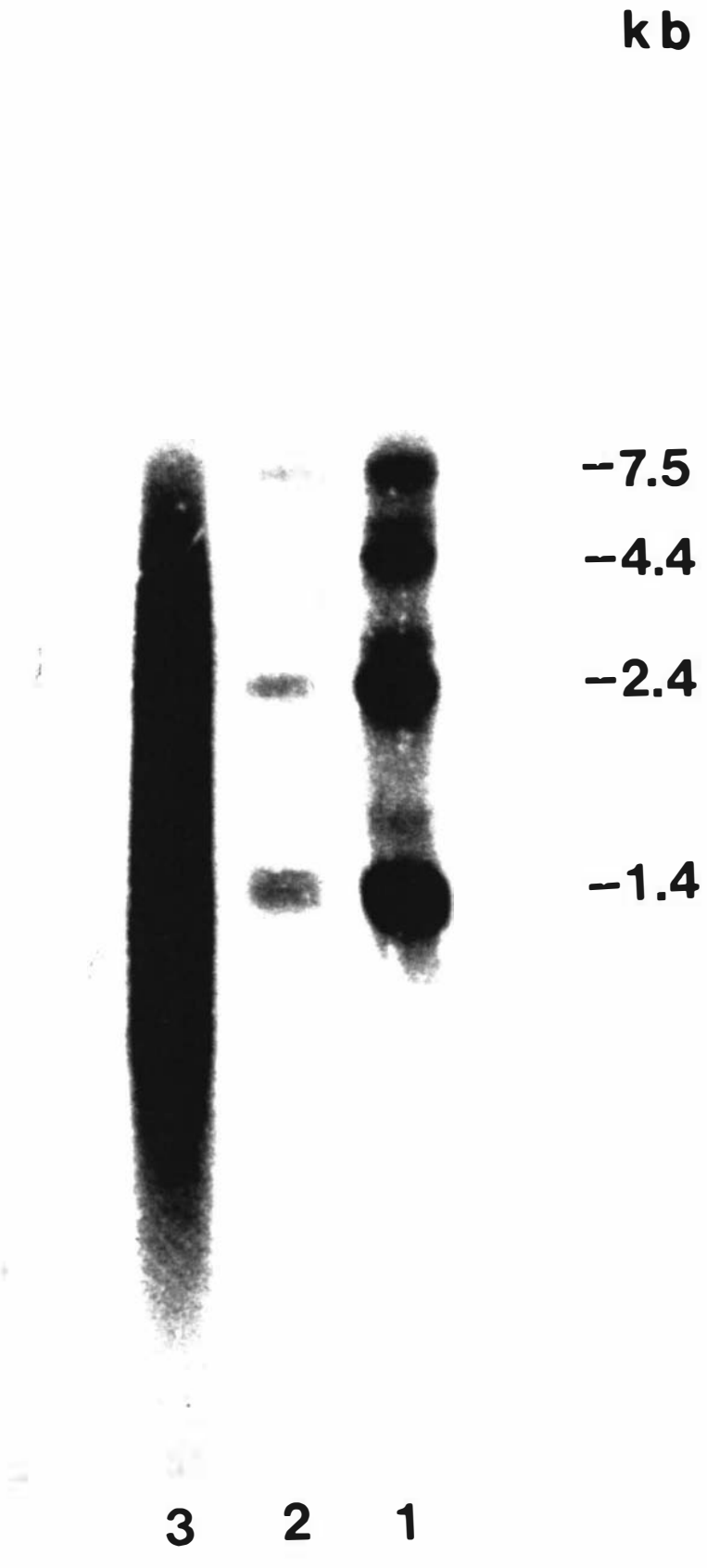

kb
4.4
2.4
1.4
0.3
1
2

Figure 3.4. Trial first and second strand cDNA synthesis reactions - two-step protocol. First strand reaction products were extracted with phenol and ethanol precipitated before being used for the second strand reaction. Lanes 1 and 2: first strand reaction reaction products from different batches of reverse transcriptase; lane 3: second strand reaction products.


kb
7.5
4.4
2.4
1.4
3
2
1

Since the two-step first and second strand cDNA reactions resulted in unsatisfactory second strand reaction products, a one-tube protocol which did not require the intermediate purification of the first strand synthesis reaction products was developed (Section 2.4). Trial reactions using the one-tube protocol produced discrete bands of the correct sizes for both the first and the second strand synthesis reactions (Figure 3.5). This figure clearly illustrates that reactions based on the one-tube protocol can effectively synthesise double-stranded cDNAs of at least 7.5 kb. However, the synthesis of the 0.3 kb fragment in the second strand reaction does appear to be impaired.

The one-tube protocol was subsequently used for a series of cDNA synthesis reactions with white clover leaf/stem- and root-derived poly(A^+) RNAs. A typical result of first and second strand synthesis reactions is presented in Figure 3.6. A feature of the white clover cDNA synthesis reactions (Figure 3.6) that is not apparent in the trial one-tube protocol cDNA synthesis reactions (Figure 3.5), is the doubling in size of the second strand reaction products (in Figure 3.6, compare the top band in the first strand leaf/stem reaction at approx. 1.5 kb with the equivalent band in the second strand reaction at approx. 3.0 kb). This increase in size was more pronounced in some reactions than in others and probably results from the priming of the second strand synthesis reaction from hairpin loop structures at the 3′ termini of the single-stranded cDNAs. The absence of this phenomenon in the trial reactions possibly reflects structural differences between the 5′ termini of the commercially produced and the white clover-derived poly(A^+) RNAs. Self-priming from hairpin loop structures can be utilized for the synthesis of double-stranded cDNA (Maniatis *et al.*, 1982), but this procedure requires the subsequent digestion of the double-stranded cDNAs with nuclease S1 to remove the hairpin loops. Nuclease S1 digestion can in turn lead to the loss of sequences corresponding to the 5′ termini of the mRNAs.

Figure 3.5. Trial first and second strand cDNA synthesis reactions - one-tube protocol. Lane 1: first strand reaction products; lane 2: second strand reaction products.

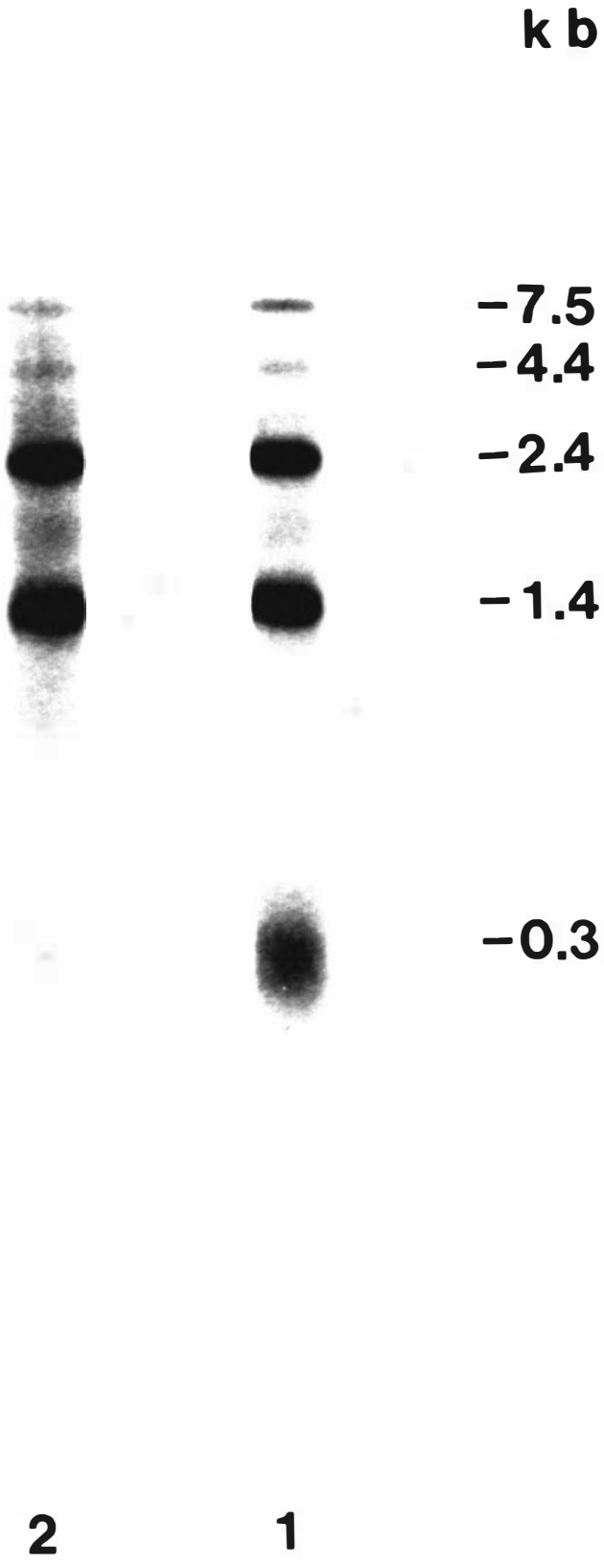
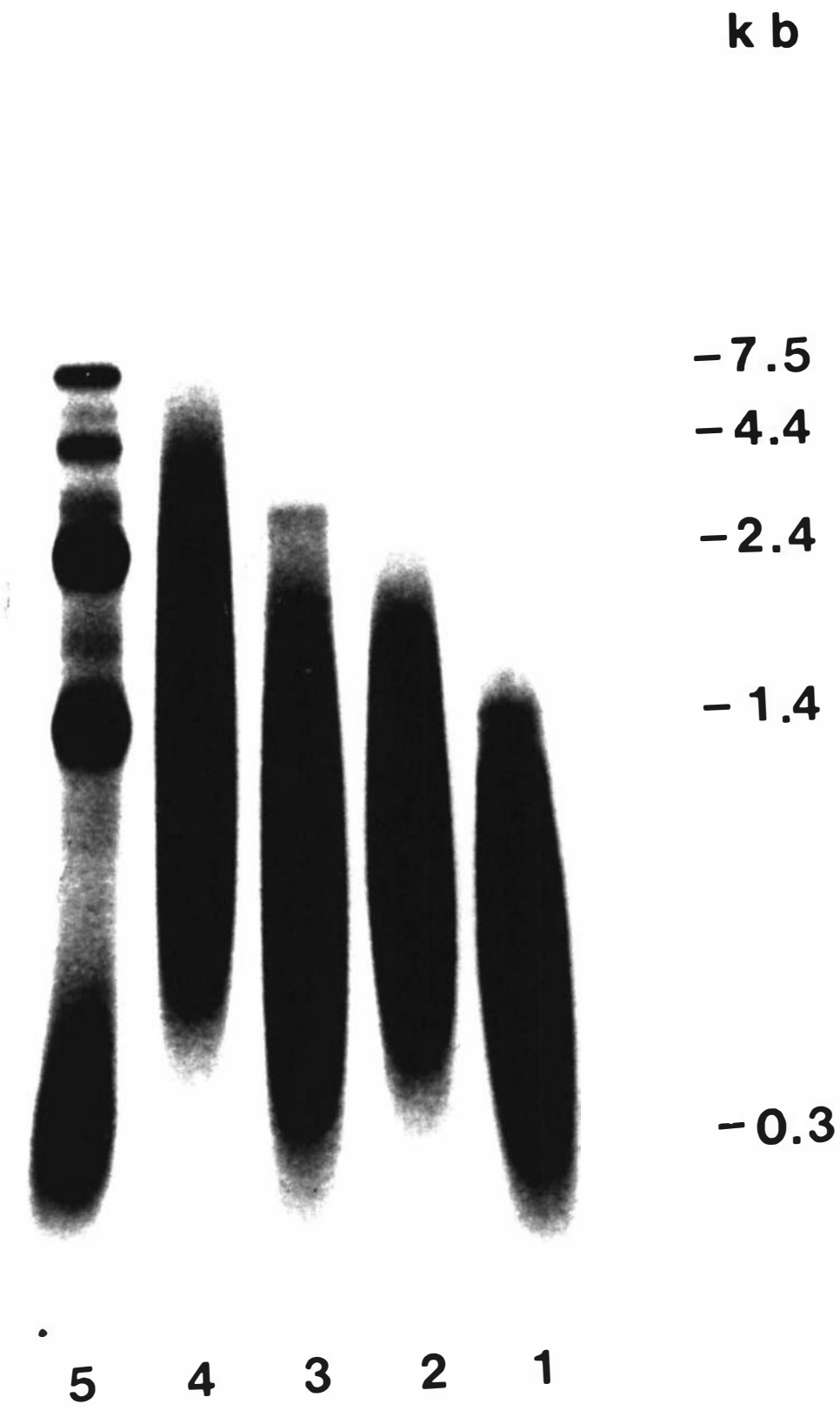

k b
-7.5
-4.4
-2.4
-1.4
-0.3
2
1

Figure 3.6. First and second strand cDNA reactions from white clover poly(A)$^+$ RNAs using the one-tube protocol. Lane 1: leaf/stem poly(A)$^+$ first strand reaction; lane 2: root poly(A)$^+$ first strand reaction; lane 3: leaf/stem poly(A)$^+$ second strand reaction; lane 4: root poly(A)$^+$ second strand reaction; lane 5: poly(A)$^+$ size standards.


k b
-7.5
-4.4
-2.4
-1.4
-0.3
5
4
3
2
1

The inclusion of a nuclease S1 digestion step was considered for the cDNA synthesis protocol in this study. However this digestion was not included since it was not clear what effect a proportion of double-stranded cDNAs with hairpin loop structures would have on the size of the subsequent cDNA library and how the size distribution of cloned cDNAs would be affected by attempting to remove these structures.

3.4 Construction of cDNA libraries in λgt10

A series of independent cDNA synthesis reactions was carried out with white clover leaf/stem- and root-derived mRNAs and the resultant double-stranded cDNAs were cloned into λgt10 (Section 2.4). The libraries were assayed on either *E. coli* C600 or *E. coli* C600ΔHfl (Section 2.5) before and after amplification (Table 3.2).

Table 3.2. Assay of white clover cDNA libraries before and after amplification.

Library #	Source of mRNA	Total recombinant pfu (unamplified)	Recombinant pfu/ml (amplified)
1	root	2.7×10^5	3.0×10^9
1	leaf/stem	8.0×10^4	3.6×10^8
2	root	5.8×10^5	1.8×10^{10}
2	leaf/stem	1.4×10^5	-
3	leaf/stem	2.8×10^4	8.0×10^{10}
4	leaf/stem	2.1×10^4	6.6×10^{10}

The number 2 leaf/stem library failed to form plaques at the amplification stage. This was probably due to failure of the host *E. coli* to grow, on account of the components of the lambda *in vitro* packaging mix being at too high a concen-

tration in the lambda/*E. coli* plating mix for this particular experiment. All other libraries gave a satisfactory level of amplification. Both of the root libraries are of such a size that the probability of isolating clones representative of low-abundance root mRNA sequences should be high (Maniatis *et al.*, 1982). However the probability of isolating clones representative of low-abundance leaf/stem mRNA sequences from the leaf/stem libraries should be somewhat lower.

As well as considering the population size of the cDNA libraries, the quality of some of the libraries was assessed by isolating clones at random, preparing DNA from these clones and determining the insert size of each clone. These results are summarized in Table 3.3.

Table 3.3. Insert sizes of random cDNA isolates from root (R) and leaf/stem (LS) cDNA libraries.

Isolate	Library #	Source	Insert size (kb)
R1	1	root	1.1
R2	1	root	1.5
R3	1	root	0.72
R4	2	root	0.40
R5	2	root	1.3
R6	2	root	0.78
R7	2	root	0.60
R8	2	root	0.67
R9	2	root	1.5
LS1	1	leaf/stem	0.43
LS2	1	leaf/stem	0.66

Taken together, these results indicate that the average insert size is 0.88 kb for clones derived from the number 1 and number 2 cDNA libraries (predominantly root). Since the average size of a mRNA isolated from mammalian cells is approx. 1.8 kb (Sambrook *et al.*, 1989), it is likely that the white clover cDNA libraries consist of clones with inserts that are not full length relative to the corresponding population of white clover mRNAs. This suggests that the inclusion of a double-stranded cDNA size selection step before cloning into λgt10 would have been desirable to optimize the probability of subsequently isolating near full length, white clover, cDNA clones.

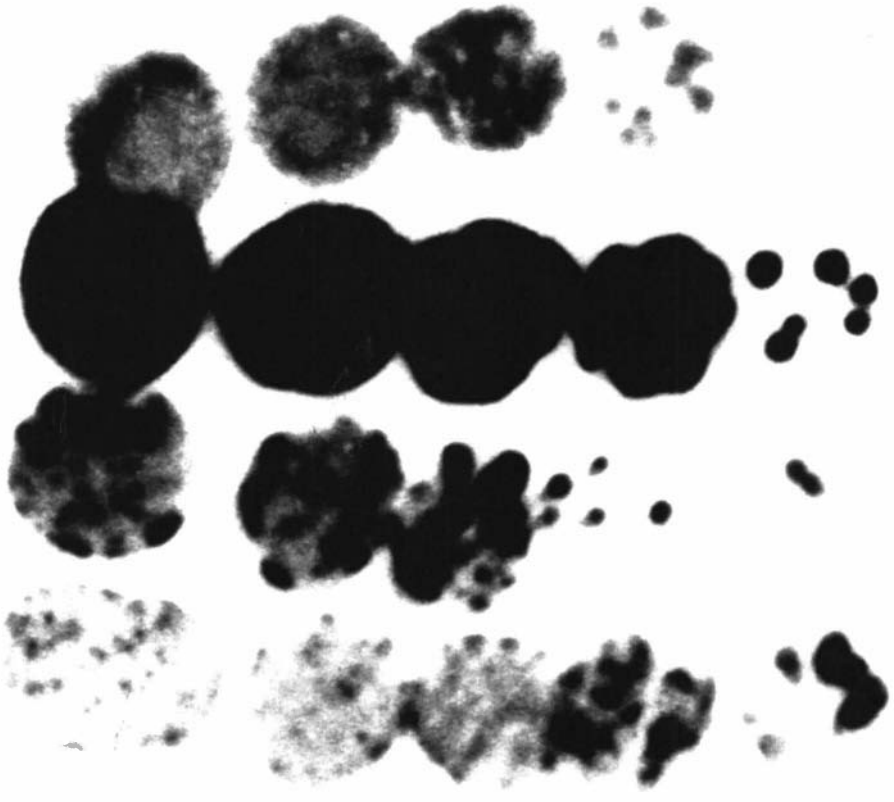
3.5 Isolation of white clover SSU cDNA clones

The Northern hybridization results (Section 3.2) indicated that the petunia Rubisco SSU cDNA (pSSU117; Dunsmuir *et al.*, 1983) would be useful for isolating white clover SSU cDNA clones from the leaf/stem cDNA libraries. Approx. 1×10^4 pfu of the amplified number 1 leaf/stem cDNA library (Table 3.3) were plated out on a 22.5 × 22.5 cm plate and transferred to a nitrocellulose filter (Section 2.6). The filter was probed with the labelled, purified insert from pSSU117, using hybridization solution containing 40% formamide and an incubation temperature of 37°C, and then autoradiographed (Figure 3.7). Four positively hybridizing cDNA clones were identified and subjected to a second cycle of plaque hybridization (Section 2.6.3) which subsequently confirmed the isolation of all four clones (Figure 3.8). A further cycle of plaque hybridization (Section 2.6.4) was carried out to identify individual plaques of each of the positive isolates (Figure 3.9). One of the isolates hybridized only weakly at this stage and was not processed further. DNA samples were prepared from the remaining three isolates and digested with *Eco*RI. The digests, when analyzed on an agarose gel, indicated that the three isolates all had inserts of identical size of approx. 550 bp (data not shown).

Figure 3.7. Primary screening of a white clover leaf/stem cDNA library with a petunia SSU cDNA probe (pSSU117). Examples of putative, positively hybridizing plaques are indicated by arrows.

Figure 3.8. Secondary screening of putative white clover cDNA phage isolates.

Figure 3.9. Plaque purification of putative white clover cDNA isolates. Serial dilutions of each of the original isolates are indicated.

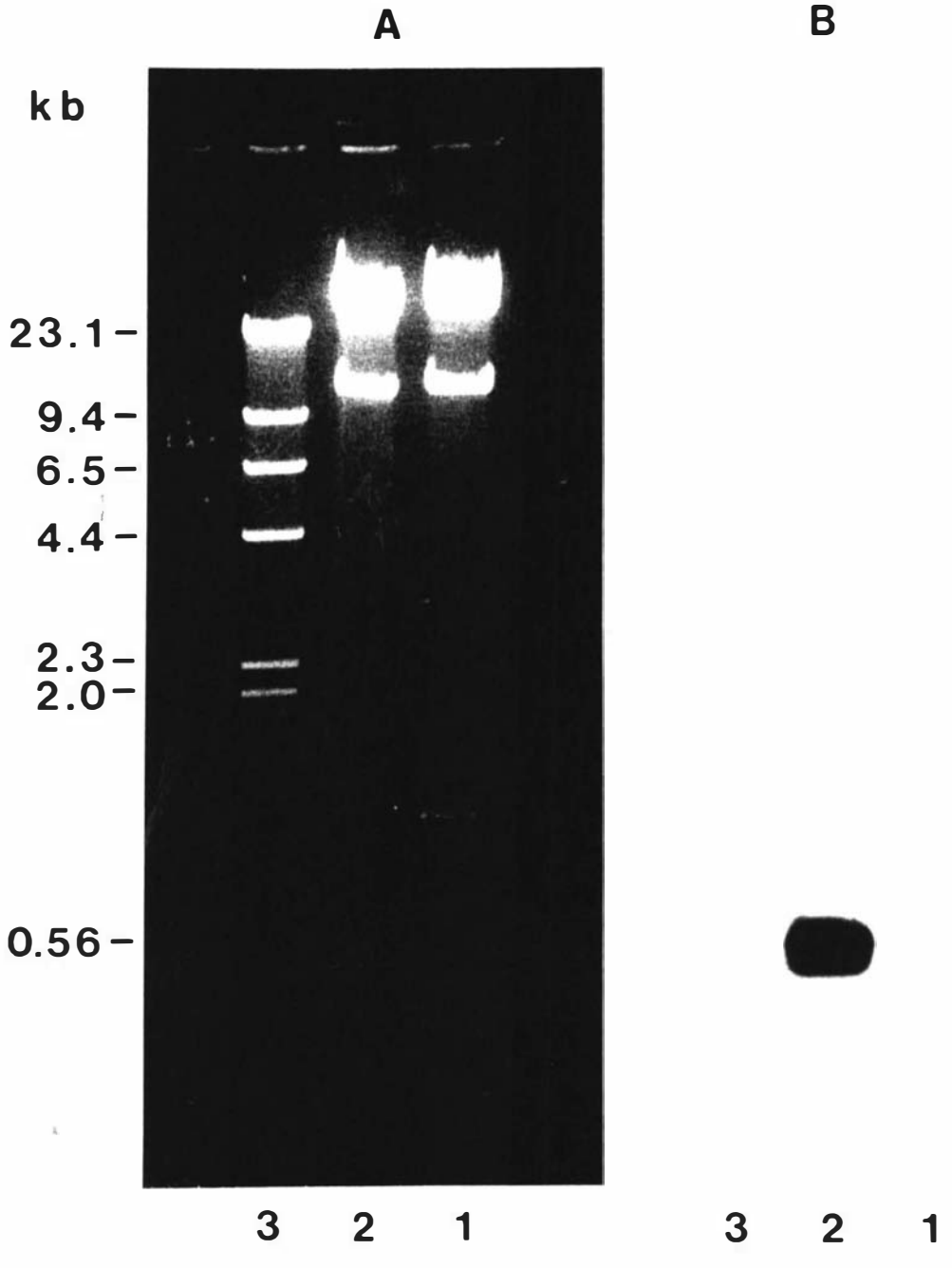


10^{-2} 10^{-3} 10^{-4} 10^{-5} 10^{-6}

The white clover SSU isolates were therefore considered to be independent isolates of the same clone. These independent, identical, isolates probably arose through amplification of a single SSU clone in the original, unamplified, cDNA library. A Southern blot of one of the digests probed with the petunia SSU cDNA confirmed that this particular white clover cDNA isolate (λTrS2) was an SSU clone with an insert size of approx. 550 bp (Figure 3.10; the 550 bp fragment does not show up in this reproduction of the original gel photograph). A restriction enzyme map of this clone, based on this and subsequent digests, is presented in Figure 3.11.

The isolation of just one positive white clover SSU cDNA clone from a primary screening of approx. 1×10^4 recombinants derived from white clover leaf/stem mRNA indicates that this particular library is not as representative as it should be. In pea, mRNA for the SSU has been shown to represent approx. 2% of the total polyadenylated RNA in illuminated plants (Bedbrook *et al.*, 1980b). In a soybean leaf cDNA library, SSU sequences represented approx. 1.5% of the recombinant phage (Shirley *et al.*, 1990). On the basis of these results, screening 1×10^4 white clover leaf/stem cDNA clones with a SSU probe should result in the detection of approx. 200 white clover SSU clones. The reasons for the underrepresentation of SSU clones in this particular white clover leaf/stem cDNA library remain unclear.

Figure 3.10. Gel electrophoresis of digests on a 0.7% gel (A) and Southern blot (B) of a white clover SSU cDNA isolate. Lane 1: *Eco*RI digest of λTrA2 (Section 3.6); lane 2: *Eco*RI digest of λTrS2; lane 3: *Hind*III digest of λ DNA


A
B
kb
23.1
9.4
6.5
4.4
2.3
2.0
0.56
3
2
1
3
2
1

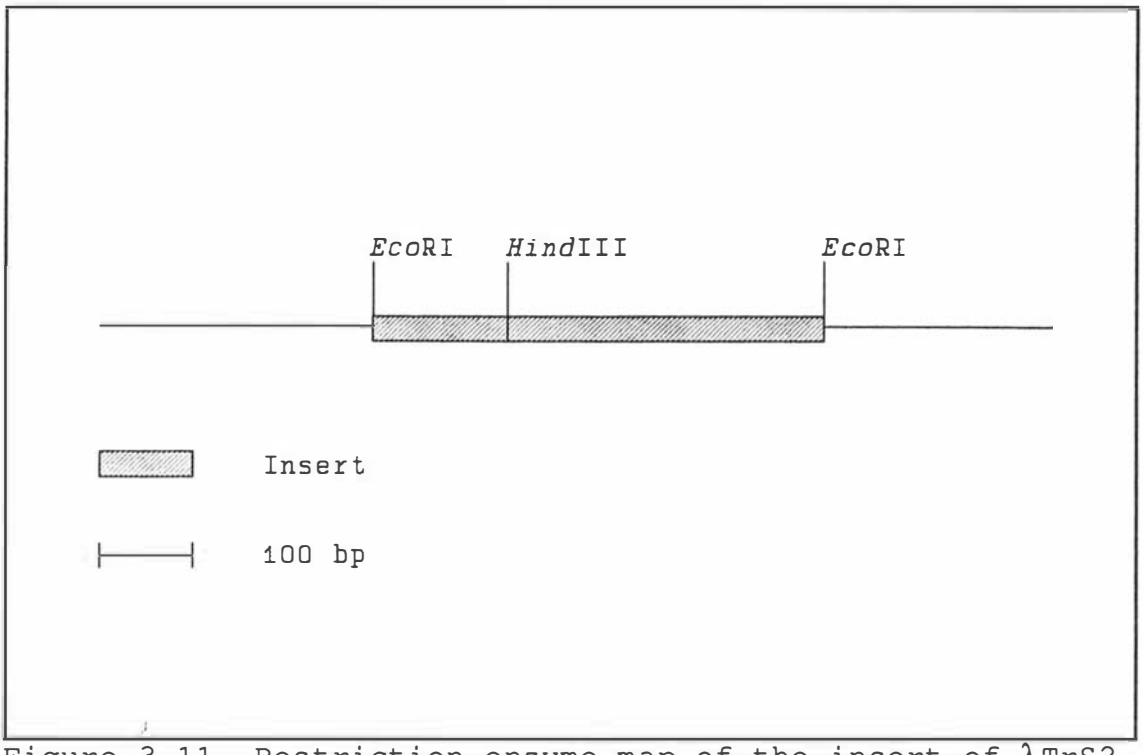


Figure 3.11. Restriction enzyme map of the insert of λTrS2, a white clover/λgt10 Rubisco SSU cDNA clone.

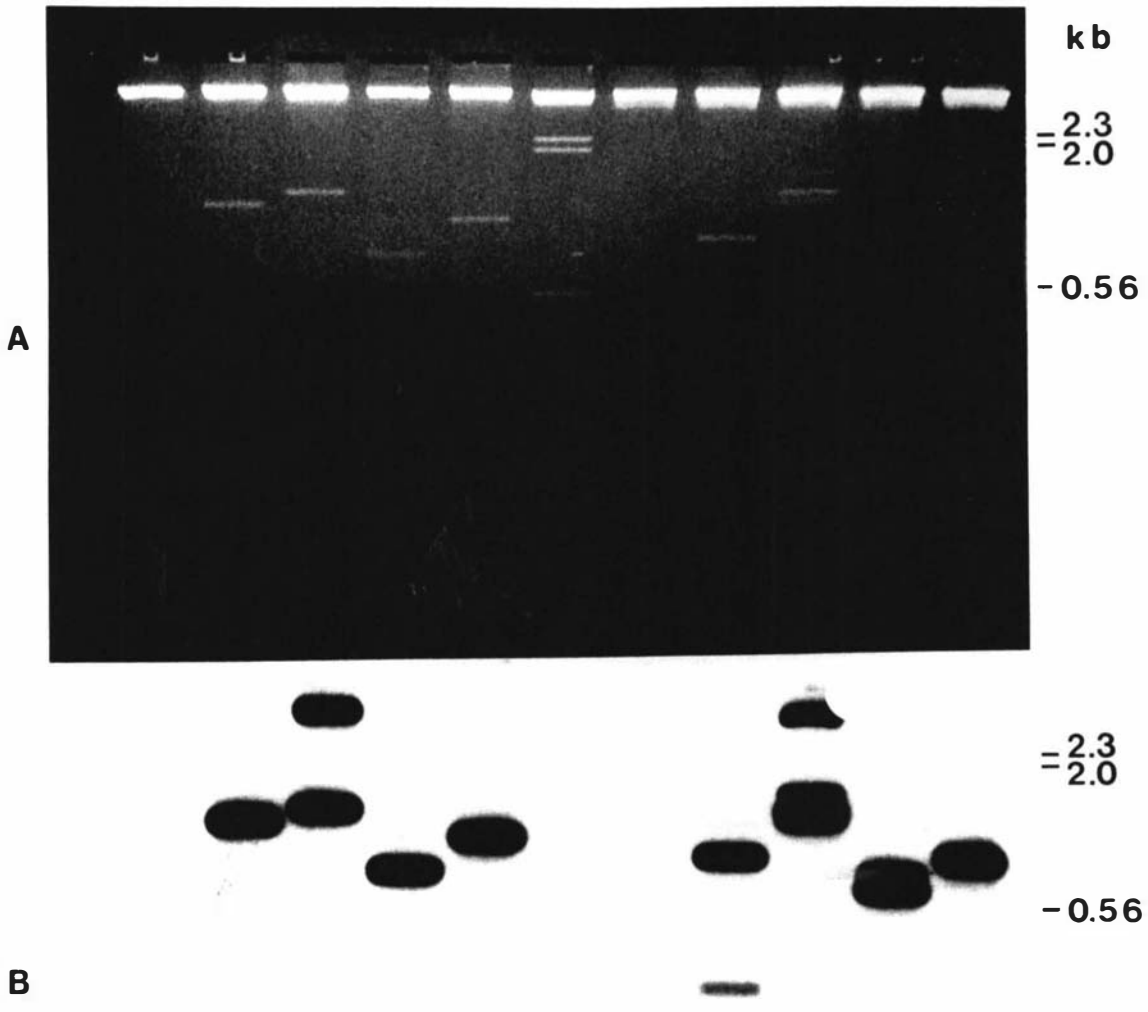
3.6 Isolation of white clover Adh cDNA clones

The Northern hybridization results (Section 3.2) indicated that the pea Adh1 cDNA (pPsR546; Llewellyn *et al.*, 1987) would be useful for isolating white clover Adh cDNA clones from the root cDNA libraries. Approx. 1×10^4 pfu of the amplified number 1, and 5×10^3 pfu of the amplified number 2 root cDNA libraries (Table 3.3) were plated out on 22.5 × 22.5 cm plates and transferred to nitrocellulose filters (Section 2.6). The filters were probed with the labelled, purified insert from pPsR546, using hybridization solution containing 40% formamide and an incubation temperature of 37°C, and then autoradiographed (for typical primary screening result, see Figure 3.7). Ten positively hybridizing clones were identified from the number 1 library and five from the number 2 library. These clones were purified as outlined in Section 3.5 for the isolation of white clover SSU

cDNA clones. Restriction enzyme digests of DNA isolated from each of these fifteen clones, when analyzed on an agarose gel, indicated that the isolates derived from the number 1 library all had inserts of the same size whereas the isolates derived from the number 2 library had inserts of different sizes (data not shown). The isolates from the number 1 library thus appeared to be independent isolates of the same clone, resulting from amplification of a single clone in the original cDNA library (see Section 3.5). In contrast, the isolates from the number 2 library were different.

Southern blots of digests of some of the positive isolates probed with the pea Adh1 cDNA confirmed that these white clover cDNA isolates were Adh clones (Figure 3.12). Insert sizes of approx. 1000, 800, 1300, 1200, and 200 bp were calculated for λTrA2, λTrA3, λTrA4, λTrA5, and λTrA6 respectively. The additional hybridizing bands above the λTrA4 1300 bp insert (Figure 3.12B, lanes 3 and 9) represent incomplete digestion of this particular DNA sample. Additional restriction enzyme mapping (data not shown) indicated that all of the clones shared a common *Hind*III site (Figure 3.13), suggesting that these clones were closely related. Since the insert of λTrA6 was so small compared with the inserts of the other clones and did not represent a different portion of the corresponding mRNA, no further characterization of this insert was carried out.

Figure 3.12. Gel electrophoresis of digests on a 2% gel (A) and Southern blot (B) of white clover Adh cDNA isolates. Lane 1: *Eco*RI plus *Hind*III digest of λTrA2 DNA; lane 2: *Eco*RI plus *Hind*III digest of λTrA3 DNA; lane 3: *Eco*RI plus *Hind*III digest of λTrA4 DNA; lane 4: *Eco*RI plus *Hind*III digest of λTrA5 DNA; lane 5: *Eco*RI plus *Hind*III digest of λTrA6 DNA; lane 6: *Hind*III digest of λ DNA; lane 7: *Eco*RI digest of λTrA2 DNA; lane 8: *Eco*RI digest of λTrA3 DNA; lane 9: *Eco*RI digest of λTrA4 DNA; lane 10: *Eco*RI digest of λTrA5 DNA; lane 11: *Eco*RI digest of λTrA6 DNA.


kb
=2.3
=2.0
-0.56
A
=2.3
=2.0
-0.56
B
11 10 9 8 7 6 5 4 3 2 1

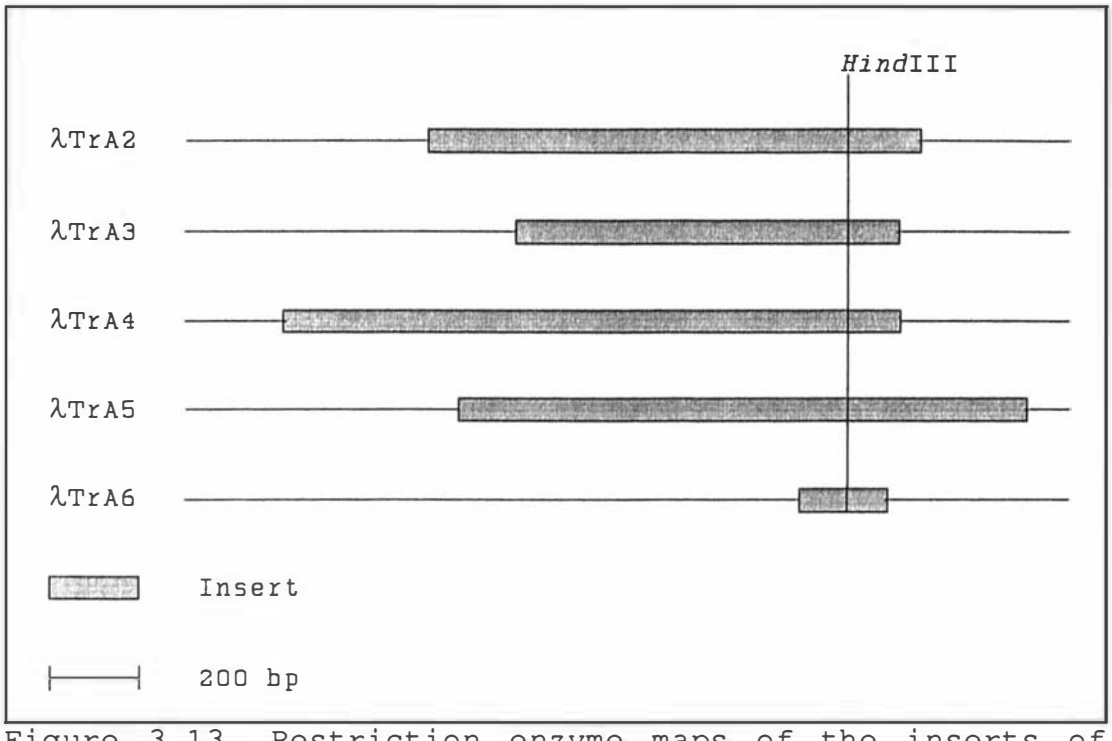


Figure 3.13. Restriction enzyme maps of the inserts of white clover/λgt10 Adh cDNA clones. All inserts are bounded by *Eco*RI sites as a consequence of linker addition during cloning.

3.7 Isolation of white clover lectin cDNA clones

The Northern hybridization results (Section 3.2) indicated that the pea seed lectin, cDNA clone (pPS15-104; Higgins *et al.*, 1983) would be useful for isolating white clover lectin cDNA clones from both the leaf/stem and the root cDNA libraries. Approx. 1×10^4 pfu of each of the amplified root number 2 and unamplified leaf/stem number 4 libraries (Table 3.3) were probed with the labelled, purified insert from the pea seed lectin cDNA clone, using hybridization solution containing 25% formamide and an incubation temperature of 37°C. Positive clones were identified as outlined in Section 3.5 (for typical primary screening result, see Figure 3.6). Two positively hybridizing clones were purified from each of the libraries, as outlined in Section 3.5 for the isolation of white clover SSU cDNA clones. A Southern blot of digests

of DNA samples prepared from each of these clones, probed with the pea seed lectin cDNA as above, is presented in Figure 3.14. Hybridization can be seen to each of the inserts of the lectin clones but not to the insert of λTrS2 (Section 3.5). Hybridization to the upper bands of the λTrLR1 and λTrLR2 digests (Figure 3.14, lanes 1 and 2) represents incomplete digestion. Insert sizes of 0.30, 3.6, 0.66 and 2.6 kb were determined for the λTrLR1, λTrLR2, λTrLL1 and λTrLL2 inserts respectively (the λTrLR1 and λTrLL1 inserts do not show up in the reproduction of the original gel photograph for Figure 3.14A). Additional restriction enzyme mapping failed to reveal any relationships between these isolates. Whether all these isolates represent genuine white clover lectin cDNA clones remains uncertain, since the hybridization stringency used to isolate and characterize the clones was relatively low. The inserts from these lectin/λgt10 cDNA isolates were subcloned into plasmids (Section 3.8) and stored.

3.8 Subcloning λgt10/cDNA inserts into plasmid vectors

Inserts were excised from the original λgt10/SSU (Section 3.5), λgt10/Adh (Section 3.6) and λgt10/lectin (Section 3.7) clones by digestion with *Eco*RI and cloned into the plasmid vector pGEM-3Z (Section 2.5). The identities of the original λgt10 clones and the corresponding plasmid clones are presented in Table 3.4; where an insert occurs in the plasmid in both orientations, two identities are listed for that particular insert.

Figure 3.14. Gel electrophoresis of digests on a 0.7% gel (A) and Southern blot (B) of white clover lectin cDNA isolates from leaf/stem (λTrLL) and root (λTrLR) cDNA libraries. Lane 1: *Hind*III digest of λ DNA; lane 2: *Eco*RI digest of λTrLR1 DNA; lane 3: *Eco*RI digest of λTrLR2 DNA; lane 4: *Eco*RI digest of λTrLL1 DNA; lane 5: *Eco*RI digest of λTrLL2 DNA; lane 6: *Eco*RI digest of λTrS2 DNA.

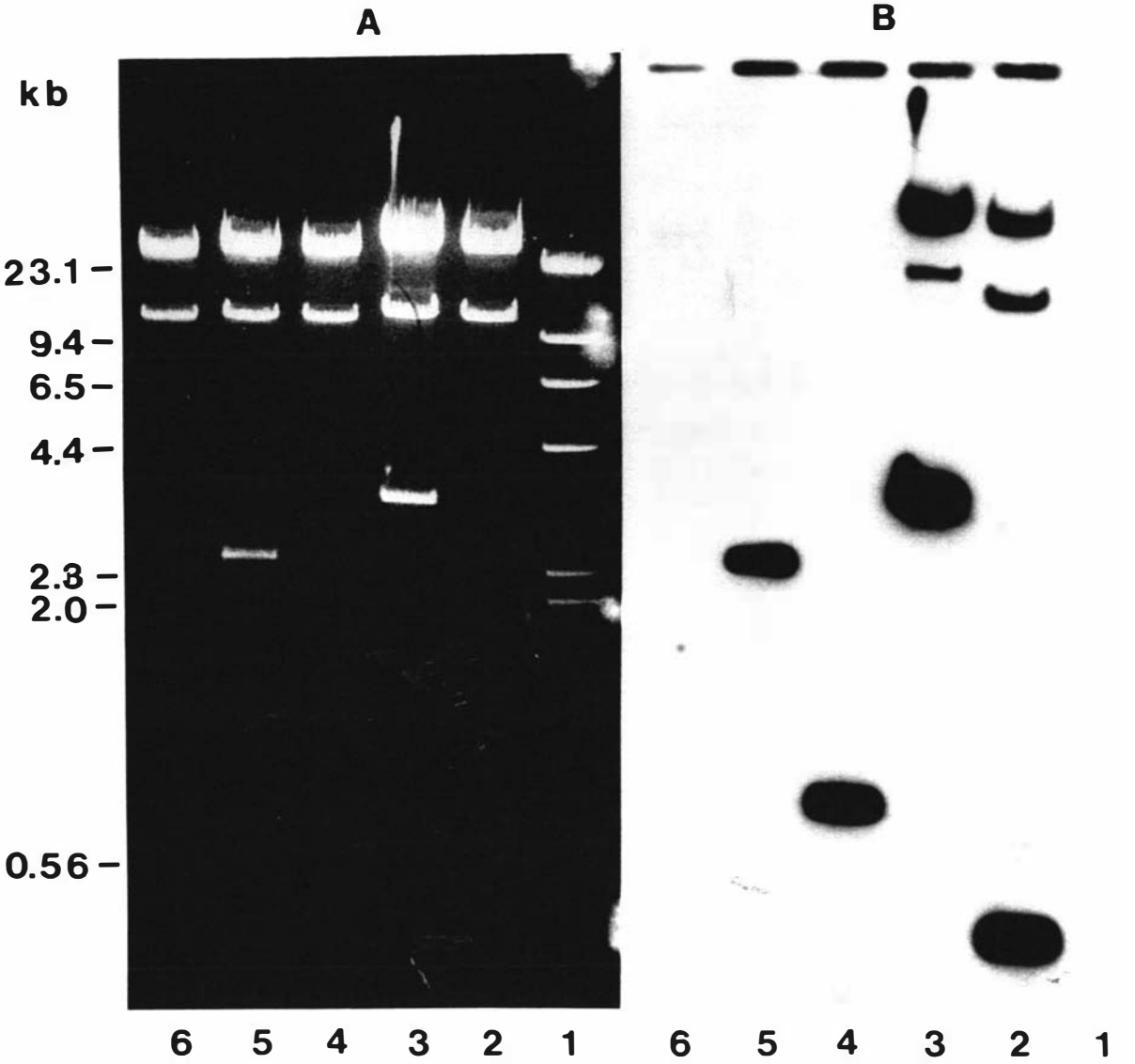

A
B
kb
23.1
9.4
6.5
4.4
2.3
2.0
0.56
6 5 4 3 2 1
6 5 4 3 2 1

Table 3.4. Identities of white clover cDNA clones subcloned into the plasmid pGEM-3Z.

λgt10 clone	Putative insert	pGEM-3Z clone
λTrS2	SSU	pTrS15/pTrS20
λTrA2	Adh	pTrA21
λTrA3	Adh	pTrA36/pTrA38
λTrA4	Adh	pTrA45/pTrA46
λTrA5	Adh	pTrA51/pTrA58
λTrLR1	lectin	pTrLR1
λTrLR2	lectin	pTrLR2
λTrLL1	lectin	pTrLL1
λTrLL2	lectin	pTrLL2

3.9 Restriction enzyme mapping of Adh cDNA clones pTrA45 and pTrA58

Restriction enzyme mapping of the five original λgt10/Adh cDNA clones indicated that these clones were closely related (Section 3.6). Additional mapping of the two plasmid sub-clones with the largest inserts (pTrA45 and pTrA58) was carried out to more accurately define the relationship between them. From the mapping results (Figure 3.15), it is clear that these two clones are very closely related and could be different sized clones derived from the same, original, mRNA species.

3.10 Sequence analysis of the white clover SSU cDNA

The strategy adopted for sequencing pTrS20, the putative white clover SSU cDNA clone, is outlined in Figure 3.16. Some

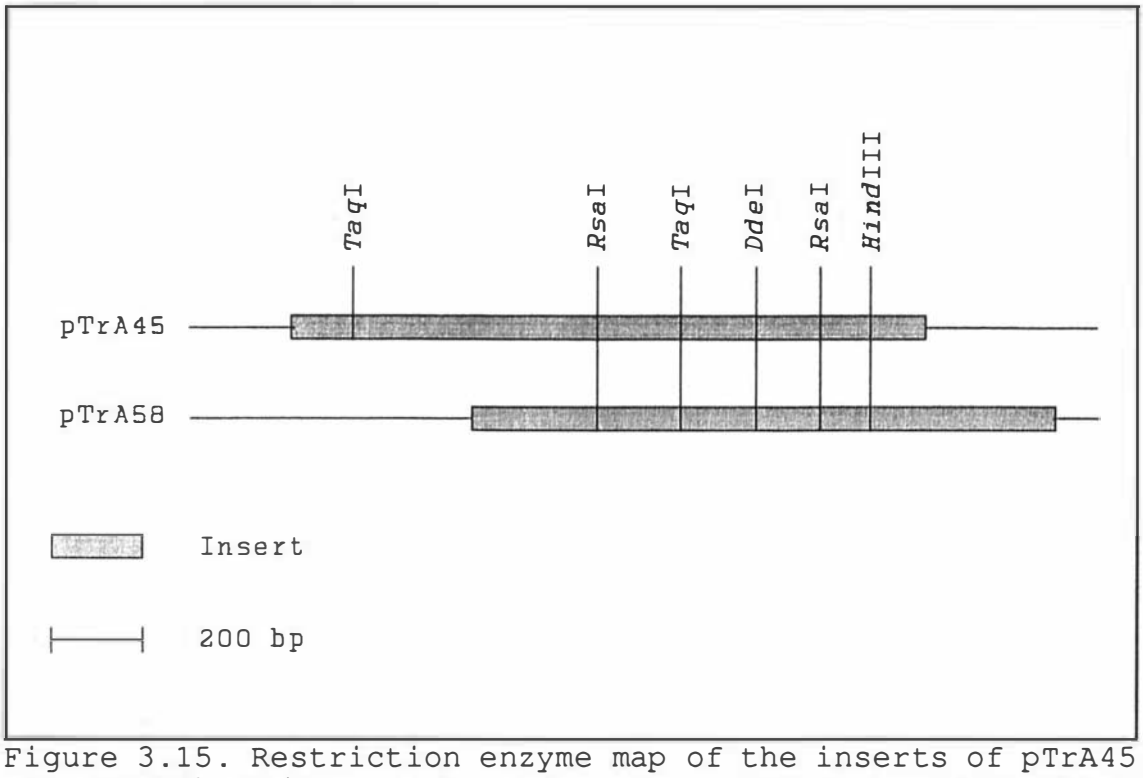


Figure 3.15. Restriction enzyme map of the inserts of pTrA45 and pTrA58 white clover Adh cDNA subclones. Only those restriction enzyme sites that could be located are indicated.

difficulty was encountered sequencing from the T7 promoter primer due to the presence of a poly(A) tract at that end of the insert and hence a portion of the insert was sequenced in one direction only.

The complete nucleotide sequence of the cDNA insert of the pTrS20 clone, with the sequence of the *Eco*RI linker at each end deleted, is shown in Figure 3.17. The insert, which includes a 50 bp poly(A) tract, is 440 bp long. This clone is an incomplete SSU cDNA clone lacking the sequences encoding the transit peptide and the first 30 amino acids of the mature protein (see below). The sequence can be translated into a polypeptide in reading frame two with a stop codon located at position 229. This sequence has been entered into the EMBL database under accession number X53954.

A comparison of the sequence of pTrS20 with that of pSSU117, the partial, petunia SSU, cDNA clone (Dunsmuir *et al.*, 1983) used to isolate the original white clover cDNA clone, is

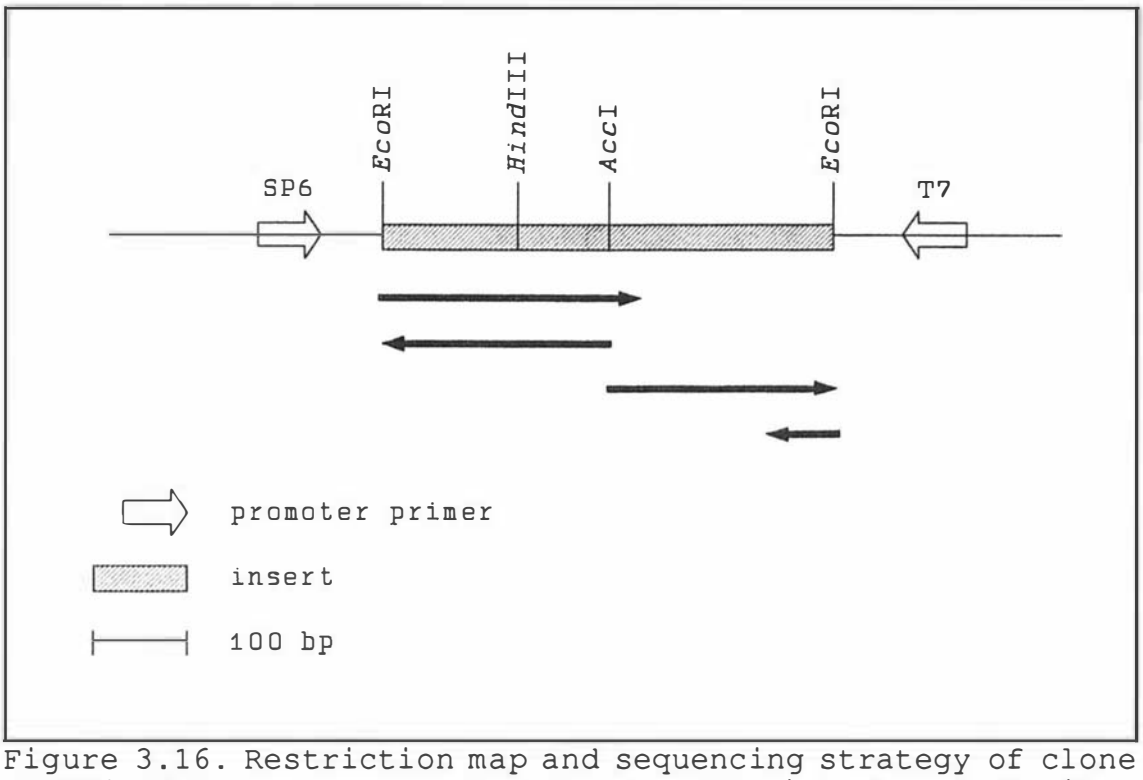


Figure 3.16. Restriction map and sequencing strategy of clone pTrS20. Subclones were generated from the internal *Acc*I site. The arrows below the restriction map indicate the direction and approx. length of the DNA strands sequenced.

```
  1   AGGATTTGTC CACCGTCAGT ACAACAGTTC ACCAGGATAC TATGATGGAC
 51   GTTACTGGAC AATGTGGAGG TTGCCATTGT TTGGAACCAC TGATGCTGCT
101   CAGGTGTTGA AGGAAGTTGC TGAATGTAAA GCAGAATACC CAGAAGCTTT
151   CATCCGTATC ATCGGATTTG ACAACGTTCG TCAAGTGCAA TGCATTAGTT
201   TCATTGCAAG CACACCCAAA GTCTACTAAA TTTGATATTT GCACCACCCT
251   TTATTACTAC TTTGTTTGTA CTTCACCATT GTAAGAACTA TATTTCCCAT
301   TTGTTTTATG TTTTTTAATA TTTCATCATC AATATAGTAT CATCCTGTTG
351   TATTTTTGGT TATTATGTAT TCGGATTTCC ATTGGAAATT AAAAAAAAAA
401   AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
```

Figure 3.17. DNA sequence of white clover SSU cDNA clone pTrS20. The stop codon (TAA) is underlined.

presented in Figure 3.18. There is 74% homology at the nucleotide level between the sequences of the coding regions of these two clones.

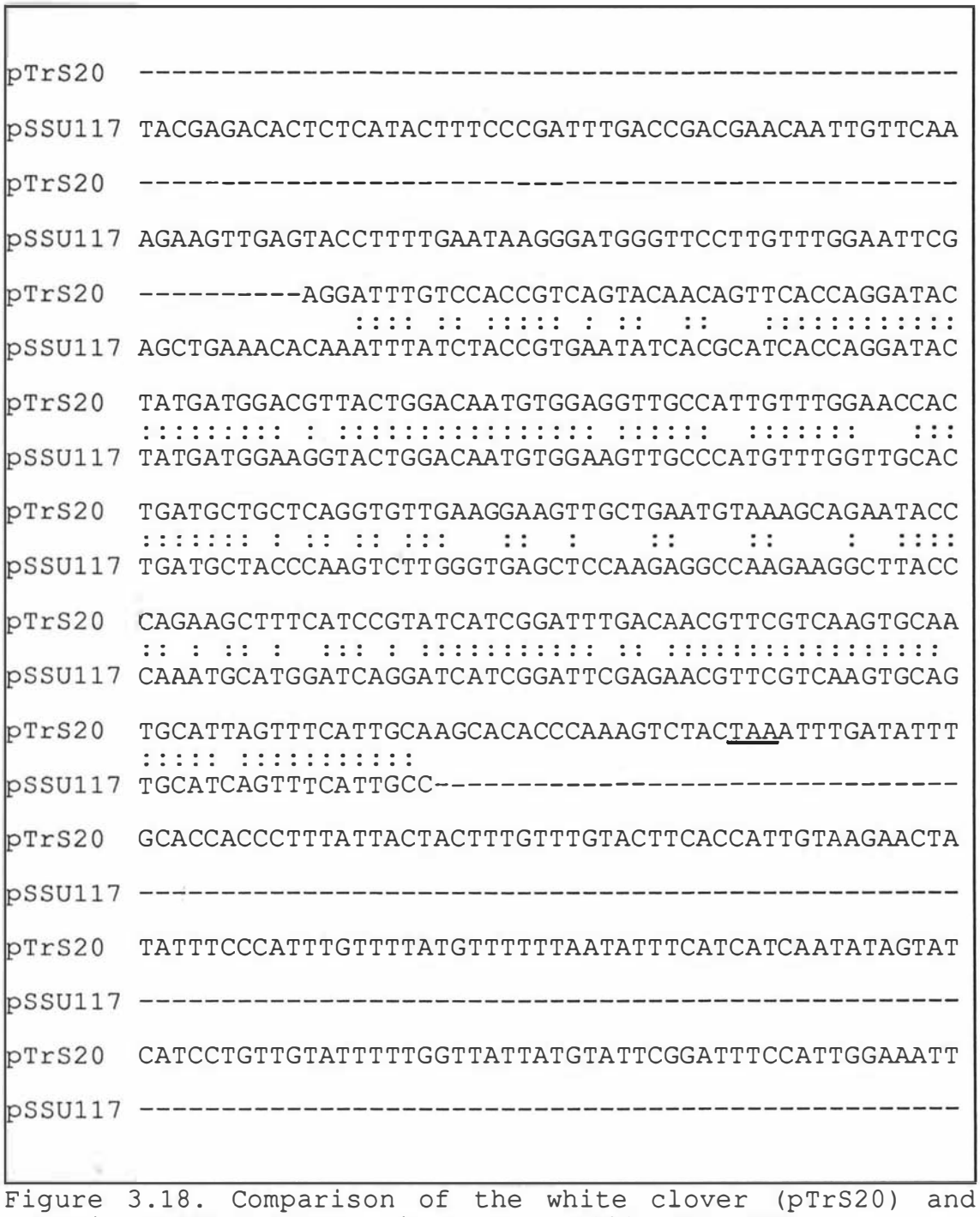


Figure 3.18. Comparison of the white clover (pTrS20) and petunia (pSSU117; Dunsmuir *et al.*, 1983) SSU cDNA sequences. The translation stop codon (TAA) of the white clover sequence is underlined.

A comparison of the sequence of pTrS20 with that of pSSU1, a near full length, pea SSU, cDNA clone (Bedbrook *et al.*, 1980b), is presented in Figure 3.19.

```
pTrS20  --------------------------------------------------
pSSU1   AACACTGACATTACAAGCAATGGTGAAAGAGTAAAGTGCATGCAGGTGTG
pTrS20  --------------------------------------------------
pSSU1   GCCTCCAATTGGAAAGAAGAAGTTTGAGACTCTTTCCTATTTGCCACCAT
pTrS20  --------------------------------------------------
pSSU1   TGACGAGAGATCAATTGTTGAAAGAAGTTGAATACCTTCTGAGGAAGGGA
pTrS20  -------------------------------AGGATTTGTCCACCGTCA
                                        :::::::::  :: :: :
pSSU1   TGGGTTCCATGCTTGGAATTTGAGTTGCTCAAAGGATTTGTGTACGGTGA
pTrS20  GTACAACAGTTCACCAGGATACTATGATGGACGTTACTGGACAATGTGGA
        : ::::::  :::::: :::::::::::::: : ::::::::::::::::
pSSU1   GCACAACAAGTCACCAAGATACTATGATGGAAGATACTGGACAATGTGGA
pTrS20  GGTTGCCATTGTTTGGAACCACTGATGCTGCTCAGGTGTTGAAGGAAGTT
         : : ::  ::::::: ::::::::: ::::::: ::  :::::::: :::
pSSU1   AGCTTCCTATGTTTGGCACCACTGATCCTGCTCAAGTCGTGAAGGAGGTT
pTrS20  GCTGAATGTAAAGCAGAATACCCAGAAGCTTTCATCCGTATCATCGGATT
        : :::: :  :: : ::::: ::::::::: : ::: ::::::: ::
pSSU1   GATGAAGTTGTTGCCGCTTACCCCGAAGCTTTCGTTCGTGTCATCGGTTT
pTrS20  TGACAACGTTCGTCAAGTGCAATGCATTAGTTTCATTGCAAGCACACCCA
          :::::::::::::::: :::::::: ::::::::::::  ::::::
pSSU1   CAACAACGTTCGTCAAGTTCAATGCATCAGTTTCATTGCACACACACCAG
pTrS20  AAGTCTACTAAATTTGATATTTGCACCACCCTTTATTACTACTTTGTTTG
        ::  ::::::: ::   :   :       ::: : :   :  : :: ::
pSSU1   AATCCTACTAAGTTCACTGCATTGGAGTTCCTATTTATATGTTATGCTTT
pTrS20  TACTTCACCATTGTAAGAACTATATTTCCCATTTGTTTTATGTTTTTTAA
        ::  :  :  ::::     : : : :    : ::::::  : :::  ::
pSSU1   TAAGTTCCTTTTGTTGTGTATTTTTATAATTTCTGTTTTTGGATTTCCAA
pTrS20  TATTTCATCATCAATATAGTATCATCCTGTTGTATTTTTGGTTATTATGT
          :   :    :    :::  :    :      : ::::::   :
pSSU1   ATTGCAAATGGGATGTGTGTAAGAGTTAATGAATGATATGGTTAACTTTA
pTrS20  ATTCGGATTTCCATTGGAAATT----------------------------
         : :  : ::   ::::   :
pSSU1   TTCCCAAGTTTACTTGGCGGTTTGTACTGTGTGGCTTTCGTTGTTCAGTG
```

Figure 3.19. Comparison of the white clover (pTrS20) and pea (pSSU1; Bedbrook *et al.*, 1980b) SSU cDNA sequences. The processing site (ATG) between the signal peptide sequence and the mature SSU polypeptide sequence of pSSU1, and the translation stop codons (TAA) of both sequences are underlined.

There is 80% homology at the nucleotide level between the sequences of the coding regions of these two clones. However, the 3′ untranslated regions of the two sequences are quite divergent. Since the *rbcS* genes of tomato, petunia, tobacco and pea all encode polypeptides of 180 or 181 amino acids (Manzara and Gruissem, 1988), it can be concluded from Figure 3.19 that the sequence of pTrS20 encodes 61% of the mature, white clover, SSU polypeptide.

3.11 Sequence analysis of white clover Adh cDNA clones

Deletion derivative subclones of the insert of the white clover Adh cDNA clone, pTrA45, were generated by digestion with exonuclease III in both directions (Section 2.17). The DNA sequence of each of the derivatives was determined and the complete DNA sequence in both directions was assembled from the sequences of overlapping subclones. The complete sequence (Figure 3.20) includes 85 bp of 5′ non-translated sequence, 1140 bp of translated sequence and 43 bp of 3′ non-translated sequence. The translated sequence (85 - 1225) codes for a polypeptide of 380 amino acids. This is the same number of amino acids as in the Adh of pea (Llewellyn *et al*, 1987) and one more than that in the Adh of maize (Dennis *et al*, 1984). The nucleotide sequence of pTrA45 has been entered into the EMBL database under accession number X14826.

A comparison of the sequence of pTrA45 with the sequence of pPsR546, the partial pea Adh1 cDNA clone (pPsR546 represents approx. 50% of the full length pea Adh1 transcript) used to isolate the white clover Adh cDNA clones (Figure 3.21), indicates that there is 92% homology at the nucleotide level between these two sequences. It is interesting to note that the two sequences shar a high degree of homology downstream of the translation s codons. Comparison of the coding sequence of pTrA45 wi the coding sequence of pZmL793, a full length maize Adh cDNA clone (Dennis *et al.*, 1984),

```
   1 CACAGAGTTA ACAAAACAAA GCAAAGCAAA GAAAAAGAAA AACAAAGCAA
  51 ACAAAGAACT CAGTCATCAT CATCATCTGA TAATCATGTC GAACACTGCT
 101 GGTCAGGTCA TCAAGTGCAG AGCTGCGGTT GCATGGGAGG CAGGGAAGCC
 151 ACTGGTGATT GAAGAAGTAG AGGTGGCGCC ACCACAGGCC GGTGAAGTCC
 201 GTCTTAAGAT ACTCTTCACC TCCCTTTGCC ACACTGATGT TTACTTCTGG
 251 GAAGCTAAGG GTCAGACTCC ATTGTTTCCT CGTATATTTG GTCATGAAGC
 301 TGGAGGGATT GTGGAGAGCG TAGGCGAAGG TGTGACTCAT CTGAAACCAG
 351 GAGACCATGC CCTGCCTGTA TTCACAGGCG AGTGTGGGGA ATGCCCACAT
 401 TGTAAGTCAG AGGAGAGTAA CATGTGTAAT CTTCTTAGGA TTAACACCGA
 451 CAGAGGTGTC ATGATCAATG ACAACAAGTC AAGATTCTCT ATTAAGGGAC
 501 AACCTGTACA CCATTTTGTC GGTACCTCTA CATTCAGCGA GTACACTGTC
 551 GTTCATGCGG GATGTGTTGC AAAGATCAAC CCTGATGCAC CACTTGACAA
 601 AGTTTGTATT CTCAGCTGTG GAATATGCAC AGGTCTTGGT GCTACTGTCA
 651 ATGTTGCAAA ACCGAAACCC GGTTCTTCTG TTGCAATCTT TGGACTTGGA
 701 GCTGTTGGCC TTGCTGCTGC TGAAGGGGCA AGGATGTCTG GTGCATCTCG
 751 AATCATTGGA GTTGATTTAG TTTCTAGCCG ATTTGAATTA GCTAAGAAGT
 801 TTGGGGTAAA TGAGTTCGTC AACCCGAAAG ATCATGACAA ACCTGTTCAA
 851 CAGGTAATTG CTGAAATGAC TGATGGAGGT GTGGATCGTG CTGTTGAATG
 901 TACCGGGAGC ATCCAGGCCA TGATCTCAGC ATTCGAATGT GTCCATGACG
 951 GATGGGGTGT TGCTGTACTT GTTGGAGTTC CAAAAAAAGA TGATGCATTC
1001 AAAACTCATC CTATGAATTT CTTGAATGAG AGGACTCTTA AGGGTACTTT
1051 CTATGGTAAC TACAAGCCTC GAACCGATCT TCCTAATGTT GTAGAGCAAT
1101 ACATGAAAGG GGAGCTGGAA CTTGAAAAAT TCATCACTCA CTCAATCCCA
1151 TTTTCAGAGA TTAACAAAGC TTTTGATTAC ATGCTGAAAG GGGAGTCCAT
1201 CAGATGTATC ATCCGAATGG AGGAGTAAAA CTGTAAAACA ATGATGAAAT
1251 AGTCCCTACT GGGGGACT
```

Figure 3.20. DNA sequence of the white clover Adh cDNA clone pTrA45. The translation start (ATG) and stop codons (TAA) are underlined.

indicates that there is 73% homology at the nucleotide level between these two sequences (Figure 3.22).

On the basis of the above results, it is apparent that pTrA45 is a near full length clone of a white clover Adh cDNA, lacking a small portion of the 3', non-coding sequence. Since clone pTrA58 appeared to be closely related to clone pTrA45,

with additional sequence at the 3′ end (Section 3.9), the 300 bp *Hind*III - *Eco*RI fragment was subcloned from pTrA58 (to give subclone pTrA52) and the insert of this subclone was sequenced in both directions. A comparison of the sequence of pTrA52 with the sequence of the equivalent *Hind*III - *Eco*RI fragment from pTrA45 (Figure 3.23) indicates that minor sequence differences occur between them. Minor differences have also been found between the sequences of two independent potato Adh cDNA clones (Matton and Brisson, 1990a). It has been suggested that, since potato has a tetraploid genome, several copies of the Adh gene are probably transcribed in potato (Matton *et al.*, 1990b). It is likely that a similar situation exists in white clover, which is also tetraploid, thus accounting for the observed sequence variation between the two Adh cDNA clones.

Figure 3.21. Comparison of the white clover (pTrA45) and pea (pPsR546; Llewellyn *et al.*, 1987) Adh cDNA sequences. The translation start codon (ATG) of the white clover sequence and the stop codons (TAA) of both of the sequences are underlined.

```
pTrA45   CACAGAGTTAACAAAACAAAGCAAAGCAAAGAAAAAGAAAAACAAAGCAA
pPsR546  --------------------------------------------------
pTrA45   ACAAAGAACTCAGTCATCATCATCATCTGATAATCATGTCGAACACTGCT
pPsR546  --------------------------------------------------
pTrA45   GGTCAGGTCATCAAGTGCAGAGCTGCGGTTGCATGGGAGGCAGGGAAGCC
pPsR546  --------------------------------------------------
pTrA45   ACTGGTGATTGAAGAAGTAGAGGTGGCGCCACCACAGGCCGGTGAAGTCC
pPsR546  --------------------------------------------------
pTrA45   GTCTTAAGATACTCTTCACCTCCCTTTGCCACACTGATGTTTACTTCTGG
pPsR546  --------------------------------------------------
pTrA45   GAAGCTAAGGGTCAGACTCCATTGTTTCCTCGTATATTTGGTCATGAAGC
pPsR546  --------------------------------------------------
pTrA45   TGGAGGGATTGTGGAGAGCGTAGGCGAAGGTGTGACTCATCTGAAACCAG
pPsR546  --------------------------------------------------
pTrA45   GAGACCATGCCCTGCCTGTATTCACAGGCGAGTGTGGGGAATGCCCACAT
pPsR546  --------------------------------------------------
pTrA45   TGTAAGTCAGAGGAGAGTAACATGTGTAATCTTCTTAGGATTAACACCGA
pPsR546  --------------------------------------------------
pTrA45   CAGAGGTGTCATGATCAATGACAACAAGTCAAGATTCTCTATTAAGGGAC
pPsR546  --------------------------------------------------
pTrA45   AACCTGTACACCATTTTGTCGGTACCTCTACATTCAGCGAGTACACTGTC
pPsR546  --------------------------------------------------
pTrA45   GTTCATGCGGGATGTGTTGCAAAGATCAACCCTGATGCACCACTTGACAA
pPsR546  --------------------------------------------------
```

Figure 3.21. (Continued)

```
pTrA45    AGTTTGTATTCTCAGCTGTGGAATATGCACAGGTCTTGGTGCTACTGTCA

pPsR546   ----------------------------------------------------

pTrA45    ATGTTGCAAAACCGAAACCCGGTTCTTCTGTTGCAATCTTTGGACTTGGA

pPsR546   ----------------------------------------------------

pTrA45    GCTGTTGGCCTTGCTGCTGCTGAAGGGGCAAGGATGTCTGGTGCATCTCG

pPsR546   ----------------------------------------------------

pTrA45    AATCATTGGAGTTGATTTAGTTTCTAGCCGATTTGAATTAGCTAAGAAGT
                                :: :::::::::::::::::::::::::
pPsR546   ----------------------TCCAGCCGATTTGAATTAGCTAAGAAGT

pTrA45    TTGGGGTAAATGAGTTCGTCAACCCGAAAGATCATGACAAACCTGTTCAA
          :::::::::::::::::::: ::::: ::::: :: :::::::::::: :::
pPsR546   TTGGGGTAAATGAGTTCGTAAACCCAAAAGAGCACGACAAACCTGTGCAA

pTrA45    CAGGTAATTGCTGAAATGACTGATGGAGGTGTGGATCGTGCTGTTGAATG
          ::::::::::::::::::::  :::::::::: ::::: ::::::::::
pPsR546   CAGGTAATTGCTGAAATGACGAATGGAGGTGTAGATCGAGCTGTTGAATG

pTrA45    TACCGGGAGCATCCAGGCCATGATCTCAGCATTCGAATGTGTCCATGACG
          :::::: :::::::::::::::::::::::::: :::::::::::::: :
pPsR546   TACCGGTAGCATCCAGGCCATGATCTCAGCATTTGAATGTGTCCATGATG

pTrA45    GATGGGGTGTTGCTGTACTTGTTGGAGTTCCAAAAAAAGATGATGCATTC
          : ::::::::::::::::::::::::: ::::  :::::::::: :::
pPsR546   GTTGGGGTGTTGCTGTACTTGTTGGAGTGCCAAGCAAAGATGATGCCTTC

pTrA45    AAAACTCATCCTATGAATTTCTTGAATGAGAGGACTCTTAAGGGTACTTT
          ::::::::::::::::: :::::::::::::::::::::::::::::: ::
pPsR546   AAAACTCATCCTATGAACTTCTTGAATGAGAGGACTCTTAAGGGTACCTT

pTrA45    CTATGGTAACTACAAGCCTCGAACCGATCTTCCTAATGTTGTAGAGCAAT
          ::: :: :::::::: :: :: :: :::::::::::::::::::: : :
pPsR546   CTACGGAAACTACAAACCCCGCACTGATCTTCCTAATGTTGTAGAGAAGT

pTrA45    ACATGAAAGGGGAGCTGGAACTTGAAAAATTCATCACTCACTCAATCCCA
          ::::::::::::::::::::::: :: :::::::: :: ::: :::: :::
pPsR546   ACATGAAAGGGGAGCTGGAACTGGAGAAATTCATTACACACACAATACCA

pTrA45    TTTTCAGAGATTAACAAAGCTTTTGATTACATGCTGAAAGGGGAGTCCAT
          ::::::::::::::::::::::::::::::::::::::::::::::::::
pPsR546   TTTTCAGAGATTAACAAAGCTTTTGATTACATGCTGAAAGGGGAGTCCAT

pTrA45    CAGATGTATCATCCGAATGGAGGAGTAAAACTGTAAAACAATGATGAAAT
          ::::::::::::  :::::::::::::: :: ::::::  ::::: ::
pPsR546   CAGATGTATCATCAAAATGGAGGAGTAAATCTCTAAAACTGTGATGTAAC
```

Figure 3.21 (Continued)

```
pTrA45    AGTCCCTACTGGGGGACT----------------------------
          ::::::::::::::::::
pPsR546   AGTCCCTACTGGGGGACTGTGGTCACTCTCTAGTTTTGCAAAAATAAATT

pTrA45    ----------------------------------------------

pPsR546   TCTTAAATAATCCCTTTTGTGTTGTTGAGTTTGTGAAATTTTTATGTAAA

pTrA45    ----------------------------------------------

pPsR546   GAATTGTTGTCGCTGTATGGTAATTTAATAATATGTTGAAGGAATAAAAA
```

Figure 3.22. Comparison of the white clover (pTrA45) and maize (pZmL793; Dennis *et al.*, 1984) Adh cDNA coding sequences.

```
pTrA45    ATGTCGAACACTGCTGGTCAGGTCATCAAGTGCAGAGCTGCGGTTGCATG
          ::: ::   :: :: ::  :::: :::::::::: ::::::::: :::::
pZmL793   ATGGCG---ACCGCGGGGAAGGTGATCAAGTGCAAAGCTGCGGTGGCATG

pTrA45    GGAGGCAGGGAAGCCACTGGTGATTGAAGAAGTAGAGGTGGCGCCACCAC
          :::::: :: ::::::::: ::: :: :: :: ::::: ::::: :: :
pZmL793   GGAGGCCGGCAAGCCACTGTCGATCGAGGAGGTGGAGGTAGCGCCTCCGC

pTrA45    AGGCCGGTGAAGTCCGTCTTAAGATACTCTTCACCTCCCTTTGCCACACT
          :::::   :: :: ::  : ::::: :::::::::: :: ::::::::
pZmL793   AGGCCATGGAGGTGCGCGTCAAGATCCTCTTCACCTCGCTCTGCCACACC

pTrA45    GATGTTTACTTCTGGGAAGCTAAGGGTCAGACTCCATTGTTTCCTCGTAT
          :: :: :::::::::: :: ::::: :::::::  :::: ::::: ::
pZmL793   GACGTCTACTTCTGGGAGGCCAAGGGGCAGACTCCCGTGTTCCCTCGGAT

pTrA45    ATTTGGTCATGAAGCTGGAGGGATTGTGGAGAGCGTAGGCGAAGGTGTGA
           ::::: :: :: ::::::: ::  : ::::: :: :: :: :::::::
pZmL793   CTTTGGCCACGAGGCTGGAGGTATCATAGAGAGTGTTGGAGAGGGTGTGA

pTrA45    CTCATCTGAAACCAGGAGACCATGCCCTGCCTGTATTCACAGGCGAGTGT
          :: :  :    :: :: ::::::: ::: ::::: ::::: :: :::::
pZmL793   CTGACGTAGCTCCGGGCGACCATGTCCTTCCTGTGTTCACTGGGGAGTGC

pTrA45    GGGGAATGCCCACATTGTAAGTCAGAGGAGAGTAACATGTGTAATCTTCT
            ::: ::  : :: :: ::::: :  ::::: ::::::::: :::: ::
pZmL793   AAGGAGTGTGCCCACTGCAAGTCGGCAGAGAGCAACATGTGTGATCTGCT

pTrA45    TAGGATTAACACCGACAGAGGTGTCATGATCAATGACAACAAGTCAAGAT
           ::::: ::::::::: : ::::: :::::    ::   ::::::  : :
pZmL793   CAGGATCAACACCGACCGCGGTGTGATGATTGCCGATGGCAAGTCGCGGT

pTrA45    TCTCTATTAAGGGACAACCTGTACACCATTTTGTCGGTACCTCTACATTC
          : :: :: :: ::  : ::: :  :::: ::::: :: :: :: :: :::
pZmL793   TTTCAATCAATGGGAAGCCTATCTACCACTTTGTTGGGACTTCCACCTTC

pTrA45    AGCGAGTACACTGTCGTTCATGCGGGATGTGTTGCAAAGATCAACCCTGA
          :::::::::: ::: : :::: ::: ::::::::::::::::::::: :
pZmL793   AGCGAGTACACCGTCATGCATGTGGGTTGTGTTGCAAAGATCAACCCTCA

pTrA45    TGCACCACTTGACAAAGTTTGTATTCTCAGCTGTGGAATATGCACAGGTC
           :: :: ::::: ::::::::  : :: :::::::: :: :  :: ::::
pZmL793   GGCTCCCCTTGATAAAGTTTGCGTCCTTAGCTGTGGTATTTCTACCGGTC

pTrA45    TTGGTGCTACTGTCAATGTTGCAAAACCGAAACCCGGTTCTTCTGTTGCA
          :::::::  :  : ::::::::::::::       :::::  : :: ::
pZmL793   TTGGTGCATCAATTAATGTTGCAAAACCTCCGAAGGGTTCGACAGTGGCT

pTrA45    ATCTTTGGACTTGGAGCTGTTGGCCTTGCTGCTGCTGAAGGGGCAAGGAT
           : :: ::  : ::::: ::::: ::::: ::::: : ::: ::::::::
pZmL793   GTTTTCGGTTTAGGAGCCGTTGGTCTTGCCGCTGCAGGAGGTGCAAGGAT
```

Figure 3.22. (Continued)

```
pTrA45   GTCTGGTGCATCTCGAATCATTGGAGTTGATTTAGTTTCTAGCCGATTTG
           :::: :: ::  : :::::::: :: ::  :     : ::: :::: :
pZmL793  TGCTGGAGCGTCAAGGATCATTGGTGTCGACCTGAACCCCAGCAGATTCG

pTrA45   AATTAGCTAAGAAGTTTGGGGTAAATGAGTTCGTCAACCCGAAAGATCAT
         ::  ::::: :::::: ::    : ::: :: :: ::::: ::::: ::
pZmL793  AAGAAGCTAGGAAGTTCGGTTGCACTGAATTTGTGAACCCAAAAGACCAC

pTrA45   GACAAACCTGTTCAACAGGTAATTGCTGAAATGACTGATGGAGGTGTGGA
          :::: :: :: ::  ::::: ::::::: :::::  : ::::: :: ::
pZmL793  AACAAGCCAGTGCAGGAGGTACTTGCTGAGATGACCAACGGAGGCGTCGA

pTrA45   TCGTGCTGTTGAATGTACCGGGAGCATCCAGGCCATGATCTCAGCATTCG
          ::   ::: ::::: :: :: : :::  : :: ::::::  ::: ::::
pZmL793  CCGCAGTGTGGAATGCACTGGCAACATTAATGCTATGATCCAAGCTTTCG

pTrA45   AATGTGTCCATGACGGATGGGGTGTTGCTGTACTTGTTGGAGTTCCAAAA
         ::::::: ::::: :: :::::::::::::: :: :: :: :: ::: :
pZmL793  AATGTGTTCATGATGGCTGGGGTGTTGCTGTGCTGGTGGGTGTGCCACAT

pTrA45   AAAGATGATGCATTCAAAACTCATCCTATGAATTTCTTGAATGAGAGGAC
         :: :: : ::  ::::: :: :: :: ::::: ::: :::: :: :::::
pZmL793  AAGGACGCTGAGTTCAAGACCCACCCGATGAACTTCCTGAACGAAAGGAC

pTrA45   TCTTAAGGGTACTTTCTATGGTAACTACAAGCCTCGAACCGATCTTCCTA
          :: ::::: :: :::: ::: ::::: ::::: :: :: ::::: :: :
pZmL793  CCTGAAGGGGACCTTCTTTGGCAACTATAAGCCACGCACTGATCTGCCAA

pTrA45   ATGTTGTAGAGCAATACATGAAAGGGGAGCTGGAACTTGAAAAATTCATC
         :::: :: ::::  :::::::::  :::::::::  : :: :: ::::::
pZmL793  ATGTGGTGGAGCTGTACATGAAAAAGGAGCTGGAGGTGGAGAAGTTCATC

pTrA45   ACTCACTCAATCCCATTTTCAGAGATTAACAAAGCTTTTGATTACATGCT
         :: :::    :::: ::  : ::::: ::::: :: :: ::    :::
pZmL793  ACGCACAGCGTCCCGTTCGCCGAGATCAACAAGGCGTTCGACCTGATGGC

pTrA45   GAAAGGGGAGTCCATCAGATGTATCATCCGAATGGAGGAG
          :: :::::::  :::: : :: :::::::: :::::: :
pZmL793  CAAGGGGGAGGGCATCCGCTGCATCATCCGCATGGAGAAC
```

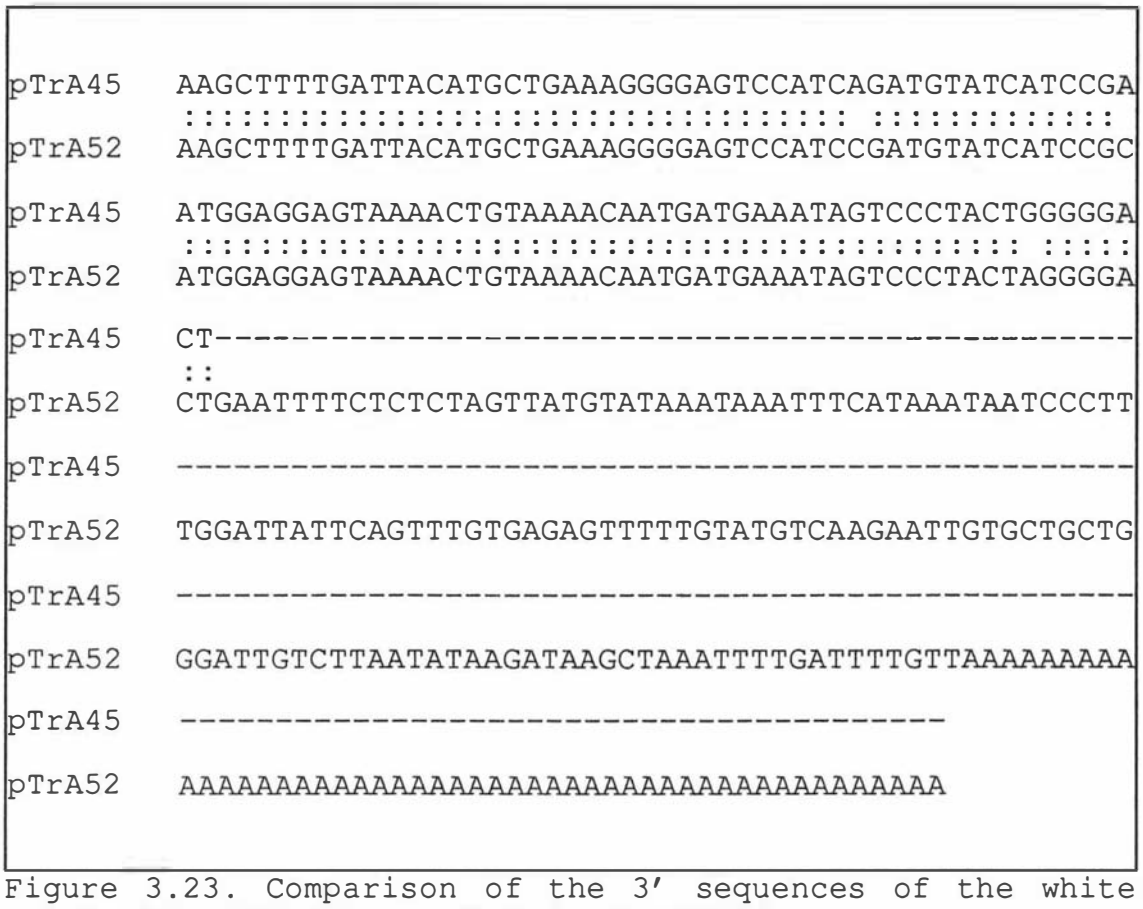


Figure 3.23. Comparison of the 3′ sequences of the white clover Adh cDNA clones pTrA45 and pTrA52.

Chapter 4

Results and Discussion: DNA Analysis and Gene Cloning

4.1 Isolation of white clover DNA

A variety of methods was tried for isolating total DNA from white clover tissues and of these, two methods consistently produced good quality DNA (Section 2.7). Method 2, based on the method of Hattori *et al.* (1987), produced DNA of a slightly greater size than that of Method 1, based on the method of Dellaporta *et al.* (1983a, 1983b, 1985) (data not shown). For both methods, the DNA was further purified by two rounds of CsCl gradient centrifugation. For subsequent use in constructing white clover genomic libraries, it was considered that the DNA should be of such a size as to comigrate with intact lambda DNA (approx. 50 kb) when analyzed on an agarose gel. DNA smaller than this would give a lower yield of useful fragments for genomic cloning after partial restriction enzyme digestion since a higher proportion of the fragments would have one sheared end and one restriction enzyme-generated end. Such fragments could not be successfully ligated with restriction enzyme digested, vector DNA.

DNA isolations were made from seeds and from germinated seedlings. The yields of DNA from seeds were typically 2 - 4 µg DNA/g seeds. However the large amount of insoluble polysaccharide in the preparations made such DNA isolations difficult. The use of freeze dried tissue offers certain advantages in the isolation of plant DNA (Murray and Thompson, 1980). Firstly, dry tissue can be efficiently disrupted while the DNA is not hydrated and the DNA is thus

less susceptible to shear. Secondly, nucleolytic degradation of the DNA can be minimized since the DNA is hydrated directly in the presence of EDTA and SDS. Freeze dried seedlings proved to be a convenient source of white clover DNA. When used in conjunction with the Method 2 DNA isolation procedure, a DNA yield of approx. 160 µg/g dry weight of seedling material was achieved.

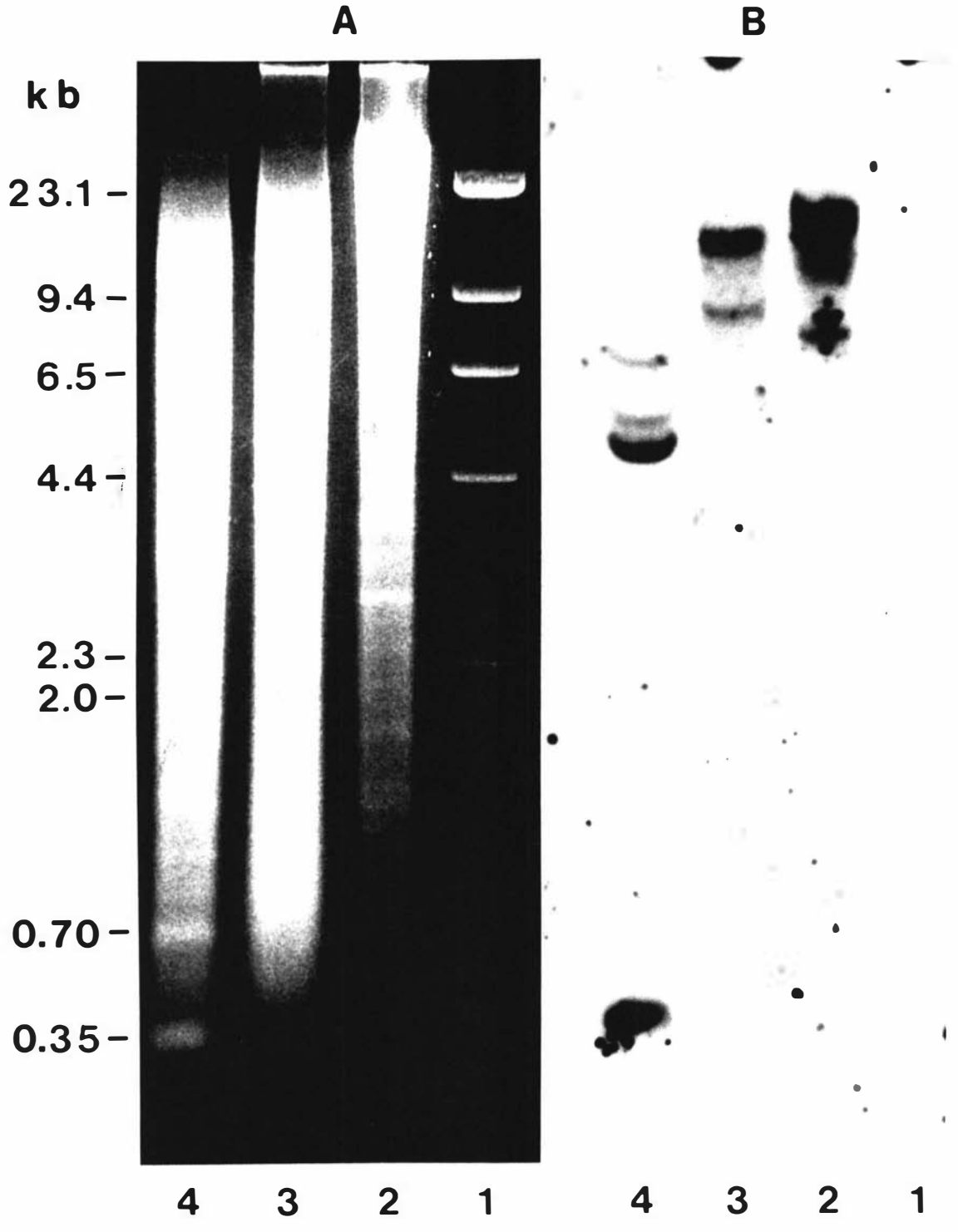
4.2 Restriction enzyme digestion of white clover DNA

The initial steps in the analysis of the organization of particular DNA sequences within the white clover genome involved digesting the DNA with various restriction enzymes and running the digested DNA on an agarose gel. Typically, 10 µg of CsCl-purified DNA was digested with 50 units of restriction enzyme at 37°C for 3 - 5 hours. After separation by agarose gel electrophoresis, each digest shows a number of characteristic discrete fragments superimposed on a background of heterogeneous restriction fragments (Figure 4.1A). The discrete fragments that are visible probably represent a combination of repeated nuclear DNA sequences and DNA sequences of chloroplast origin. It has been shown that chloroplast DNA can account for approx. 11% of the total DNA prepared from mung bean leaves (Murray *et al.*, 1979). However, by cloning and sequencing representative members from the 350 bp and 700 bp fragments that are clearly visible in the *Hind*III digest, these fragments have been shown to be tandemly arranged repeat elements in the white clover nuclear genome (Ellison *et al.*, 1986).

4.3 Southern blot analysis of white clover DNA

A Southern blot analysis was performed on white clover DNA digested with *Bam*HI, *Eco*RI and *Hind*III restriction enzymes using the labelled insert from pTrS20, the white clover SSU cDNA clone (Section 3.8), as a probe (Figure 4.1B).

Figure 4.1. Gel electrophoresis on a 0.7% agarose gel (A) and Southern blot of white clover DNA digests probed with the insert from the white clover SSU cDNA clone, pTrS20 (B). Lane 1: *Hind*III digest of λ DNA; lane 2: *Bam*HI digest of white clover DNA; lane 3: *Eco*RI digest of white clover DNA; lane 4: *Hind*III digest of white clover DNA.


A
B
kb
23.1
9.4
6.5
4.4
2.3
2.0
0.70
0.35
4 3 2 1
4 3 2 1

In the Southern blot (Figure 4.1B), four distinct bands of 9.0, 6.7, 5.6 and 0.4 kb were detected in the white clover DNA *Hind*III digest. The lower two *Hind*III bands were significantly more intense than the upper two bands. Two distinct bands were detected in each of the white clover DNA *Bam*HI and *Eco*RI digests and for these two enzymes the upper band was significantly more intense than the lower band. The upper, diffuse, band in the *Bam*HI digest probably represented incomplete digestion of the DNA. *Bam*HI is subject to limited methylation inhibition (Watson and Thompson, 1986) and this could account for the incomplete digestion in this case. The lower band in the *Bam*HI digest was found to have a size of 11 kb while the lower band in the *Eco*RI digest was found to have a size of 13 kb. These results suggested that white clover has at least two *rbcS* genes with one gene represented by the more intense bands and the others represented by the less intense bands in all three digests.

Studies in other higher plants have revealed that the SSU is encoded as a multigene family containing from five to twelve genes (Section 1.7.3). In both petunia and tomato, the *rbcS* genes are encoded at three genetic loci (Dean *et al.*, 1985a; Sugita *et al.*, 1987) whereas in pea the *rbcS* gene family is encoded at a single genetic locus (Polans *et al.*, 1985).

4.4 Preparation of white clover DNA for cloning

A series of trial *Sau*3A digests of white clover DNA was carried out to determine the concentration of enzyme that would generate DNA fragments within the correct size range for cloning into the vector, λEMBL3 (Section 2.8.1). The digests were subsequently analyzed on an agarose gel (Figure 4.2). From Figure 4.2, it was determined that *Sau*3A at a concentration of 0.031 U/μg DNA produced the maximum fluorescence in the 18 - 22 kb size range. The intensity of fluorescence is related to the mass distribution of the DNA. Optimal representation in the subsequent library depends upon

maximizing the number of molecules in the optimum size range and this is achieved by selecting conditions which give half the extent of digestion necessary for maximum fluorescence (Seed *et al.*, 1982). Therefore, in the scaled-up *Sau*3A digestion of white clover DNA (Section 2.8.2), the enzyme was used at a concentration of 0.016 units/µg DNA.

The *Sau*3A-digested white clover DNA was size fractionated on a NaCl gradient and each fraction from the gradient was analyzed on an agarose gel (Figure 4.3). From Figure 4.3, it was determined that fraction numbers 4 - 6 spanned the required size range and thus these fractions were used for the subsequent construction of a white clover genomic library.

4.5 Construction of a white clover genomic library

White clover DNA in the 18 - 22 kb size range was ligated to EMBL3 arms at a mass ratio of 1.5:1 and the ligated DNAs were packaged using an *in vitro* packaging system. The resulting genomic library was assayed on *E. coli* strain MB406 (Section 2.8.6). The total number of recombinants in the white clover genomic library was determined to be 9.2×10^5.

Figure 4.2. Trial digests of white clover DNA with varying concentrations of *Sau*3A electrophoresed on a 0.3% agarose gel. Lanes 1 and 10: λgt10 DNA digested with *Bam*HI plus *Bgl*II; lane 2: 0.0039 U/µl *Sau*3A; lane 3: 0.078 U/µl *Sau*3A; lane 4: 0.016 U/µl *Sau*3A; lane 5: 0.031 U/µl *Sau*3A; lane 6: 0.065 U/µl *Sau*3A; lane 7: 0.125 U/µl *Sau*3A; lane 8: 0.25 U/µl *Sau*3A; lane 9: 0.50 U/µl *Sau*3A; lane 11: λ DNA digested with *Hind*III.

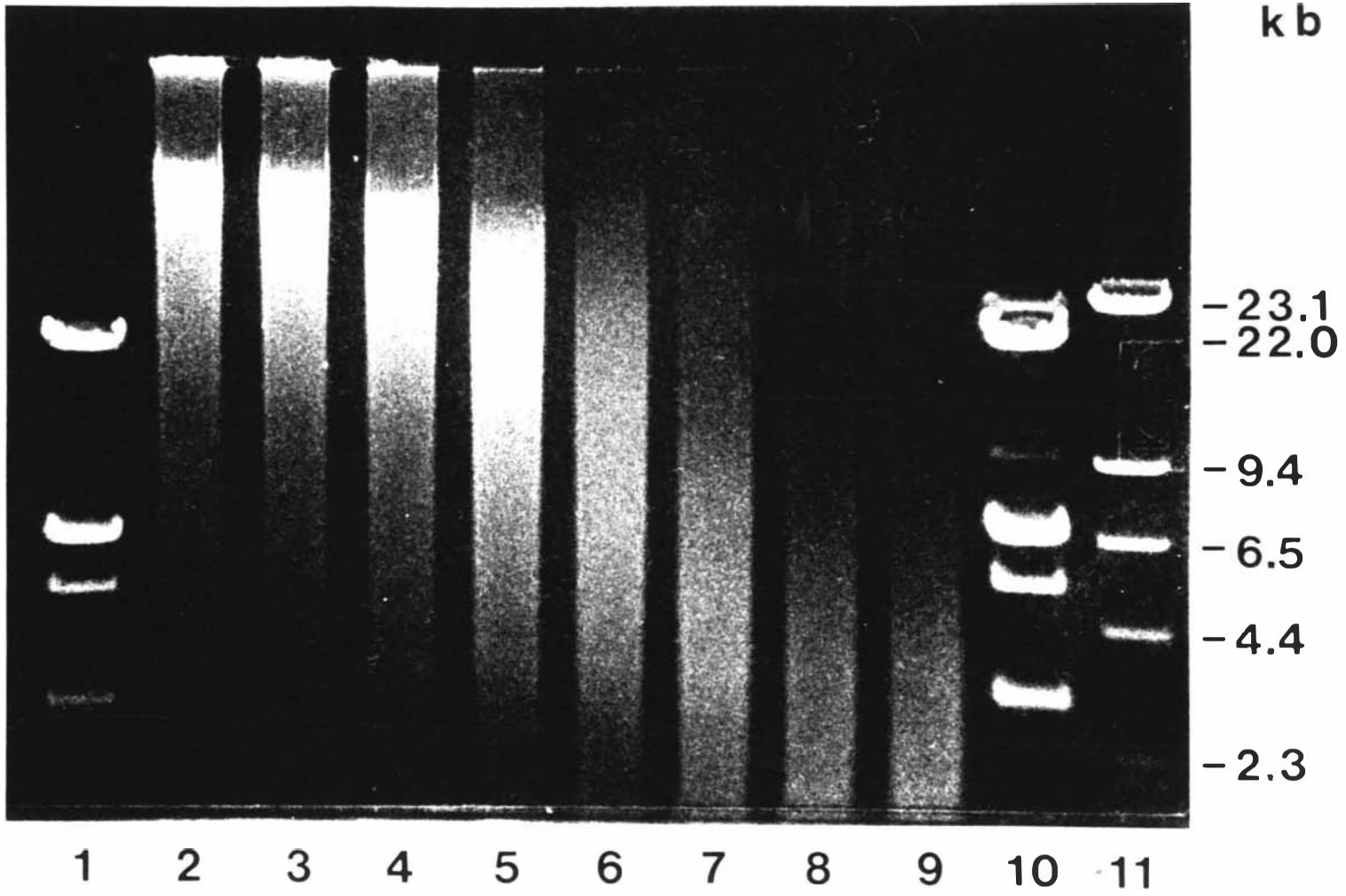
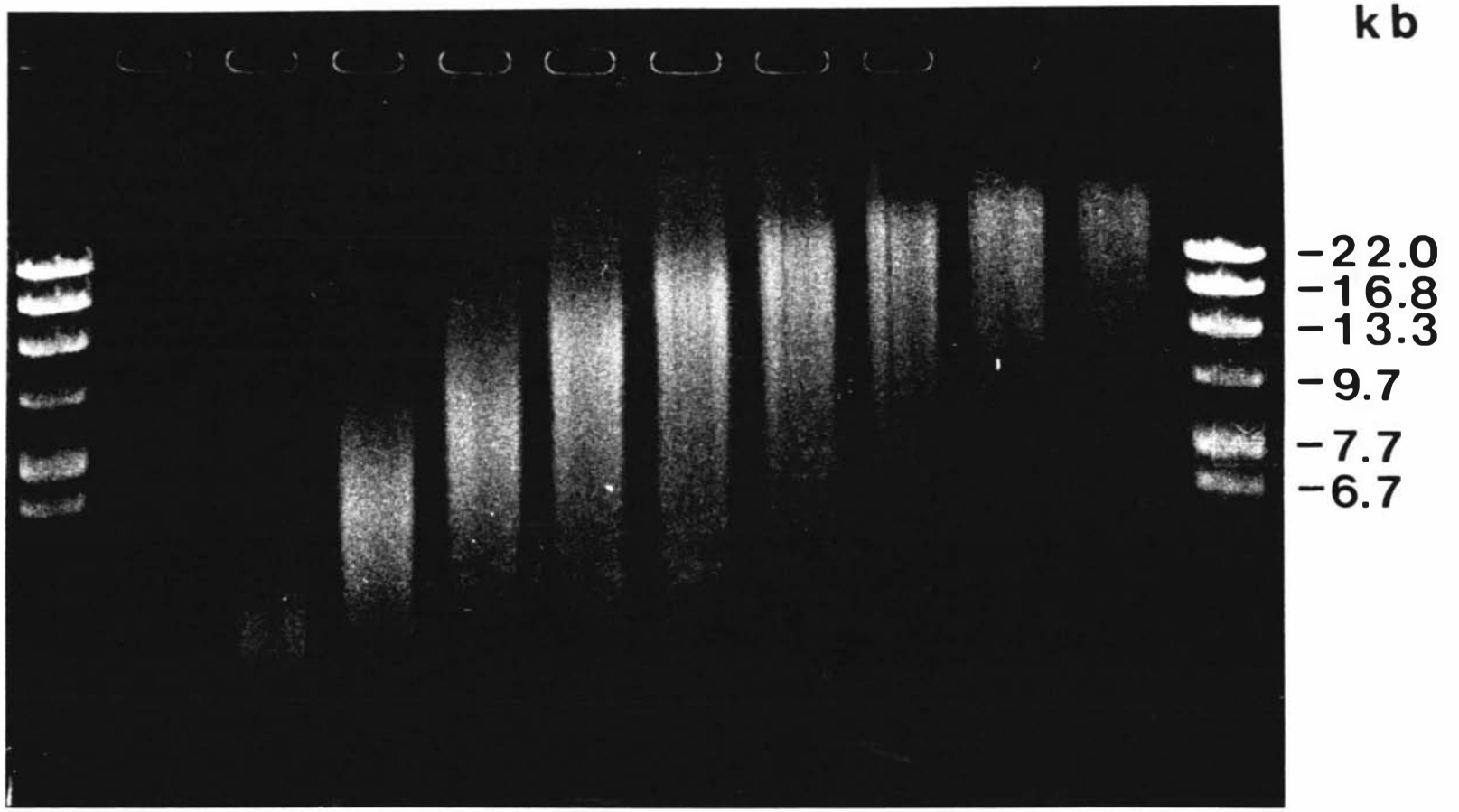

kb
23.1
22.0
9.4
6.5
4.4
2.3
1
2
3
4
5
6
7
8
9
10
11

Figure 4.3. Gel electrophoresis on a 0.3% agarose gel of white clover DNA partially digested with *Sau*3A and fractionated by velocity gradient sedimentation. Lanes 1 and 12: a mixture of λ DNA digested with *Bam*HI and λ DNA digested with *Bgl*II; lanes 2 - 11: *Sau*3A digested white clover DNA fractions.


kb
22.0
16.8
13.3
9.7
7.7
6.7
12
11
10
9
8
7
6
5
4
3
2
1

The *E. coli* strain MB406 is *rec*B⁻ *rec*C⁻ *sbc*B⁻. The use of a *rec*B⁻ *rec*C⁻ *sbc*B⁻ *E. coli* host for amplifying and screening genomic libraries is necessary for equalizing the representation of genomic sequences containing inverted and direct repetitions (Wyman and Wertman, 1987). It has been shown that the majority of recombinant phages in a library of genomic *Physarum* DNA could form plaques only on a *rec*B⁻ *rec*C⁻ *sbc*B⁻ *E. coli* host (Nader *et al.*, 1985).

Since the genome of a particular plant species can consist of up to 80% repetitive DNA sequences (see Table 1.1), the optimization of sequence representation is clearly an important feature of plant genomic library construction.

4.6 Amplification and analysis of a white clover genomic library

The white clover genomic library was amplified on *E. coli* MB406 (see Section 4.5 for discussion). The amplified library was subsequently assayed on the same host and the concentration of recombinant phage was determined to be 6.2×10^9 pfu/ml.

Eight plaques were isolated at random from the amplified library assay plate and DNA preparations were made from each of these isolates. Each DNA sample was digested with *Bam*HI and *Sal*I and the insert size of each isolate was determined by agarose gel electrophoresis with appropriate size standards (Table 4.1)

Table 4.1. Insert sizes of random isolates from the white clover genomic library.

isolate	1	2	3	4	5	6	7	8
size (kb)	16.1	16.5	15.7	12.5	17.2	11.9	11.2	17.0

Based on the insert sizes of these random isolates, an average insert size of 15 kb was calculated. Since the *Sau*3A-digested white clover DNA was size selected in the 18 - 22 kb size range, the lower value for the average insert size suggests that the size selection procedure was not as effective as it should have been.

The size of a genomic library that will have an arbitrary probability of including a particular DNA sequence can be calculated from the formula:

$$N = \frac{\ln(1-P)}{\ln\left(1-\frac{x}{y}\right)}$$

where x is the average insert size, y is the size of the haploid genome in the same units, N is the number of clones in the library and P is the probability of the library containing a particular DNA sequence (Clarke and Carbon, 1976). Therefore, the probability of having a given sequence represented in the white clover genomic library reported above, based on a genome size (y) for white clover of 3 pg (D. White, personal communication), or approx. 3×10^9 bp, can be calculated:

$$9.2x10^5 = \frac{\ln(1-P)}{\ln\left(1-\left[\frac{1.5x10^4}{3.0x10^9}\right]\right)}$$

$$\therefore P = 99\%$$

It was therefore concluded that the white clover genomic library was suitable for the isolation of specific white clover genomic sequences.

4.7 Isolation of a white clover SSU genomic clone

A total of approx. 1.2 x 10^6 pfu of the amplified white clover gene library was plated out on six 22.5 x 22.5 cm plates. After overnight growth of the phage, near confluent lysis on each plate was evident. The plaques were transferred from each plate to nitrocellulose filters and the filters were probed with the labelled insert from pTrS20, the white clover SSU cDNA clone (Section 3.8). After autoradiography, a total of seven, putative, positively hybridizing plaques was identified (for typical result see Figure 4.4). Each positively hybridizing plaque was isolated and subjected to a second round of plaque hybridization using the pTrS20 insert probe (for typical result see Figure 4.5). At this stage, only two of the initial isolates continued to show up as positive.

After an additional round of plaque hybridization, just one of these isolates (λTrS64) could still be identified as being definitely positive. DNA was prepared from λTrS64 and analyzed by restriction enzyme digestion and agarose gel electrophoresis (Figure 4.6A). A Southern blot prepared from this gel was hybridized with the white clover SSU cDNA (pTrS20) insert probe (Figure 4.6B). Based on these results as well as on additional restriction enzyme mapping (data not shown), a restriction enzyme map and hybridization profile of λTrS64 was constructed (Figure 4.7). The 3.8 kb *Bam*HI - *Eco*RI fragment was shown to encompass the entire region to which the cDNA hybridized (Figure 4.7). The 5.6 and 0.4 kb *Hind*III fragments to which the cDNA probe hybridizes (Figure 4.6B) correspond with the 5.6 and 0.4 kb *Hind*III fragments in the Southern blot of *Hind*III digested white clover genomic DNA (Figure 4.1).

Figure 4.4. Typical primary screening of the white clover genomic library with a white clover cDNA probe. Examples of putative, positively hybridizing plaques are indicated by arrows.

Figure 4.5. Typical secondary screening of a primary white clover genomic library isolate.

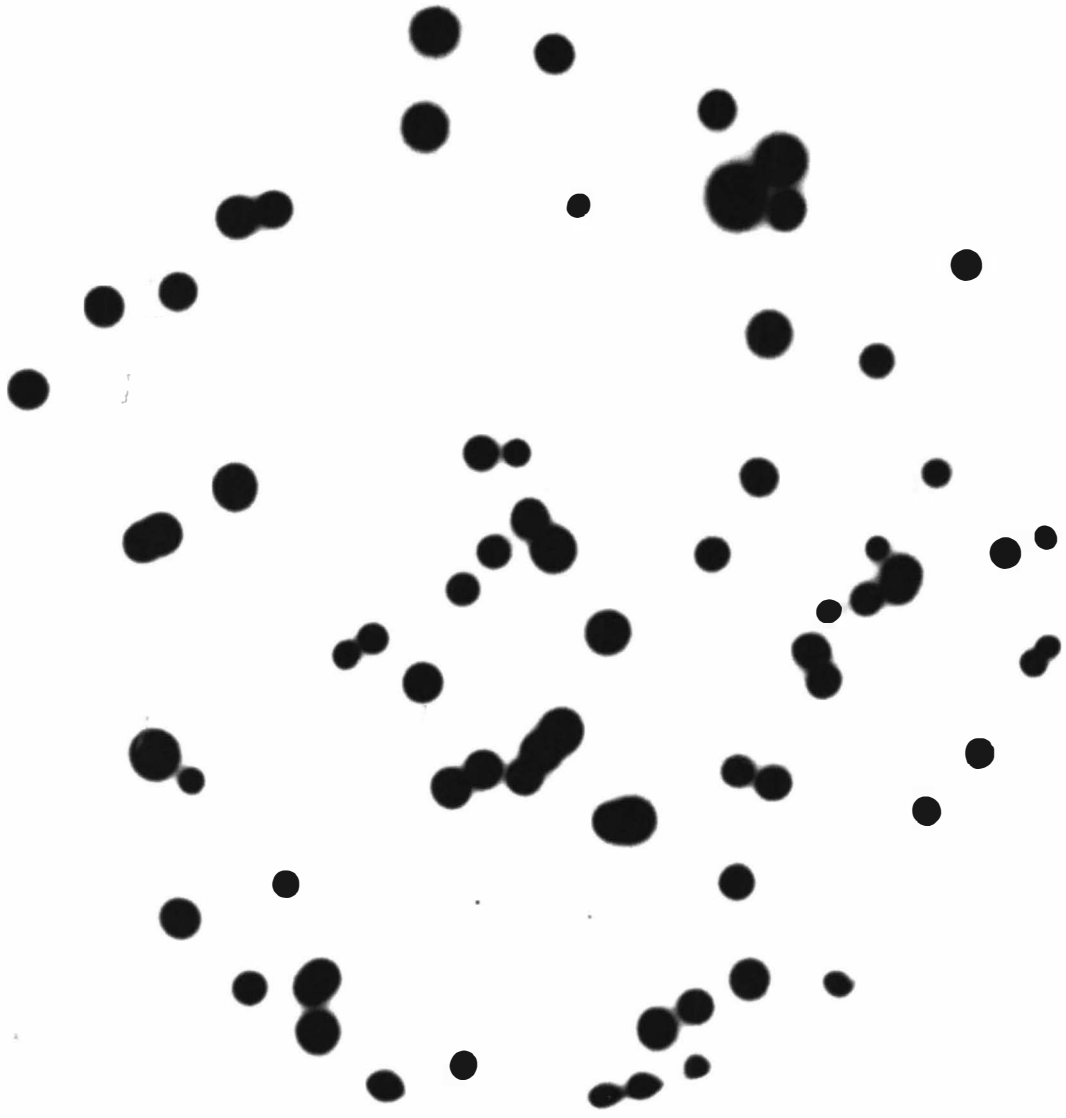
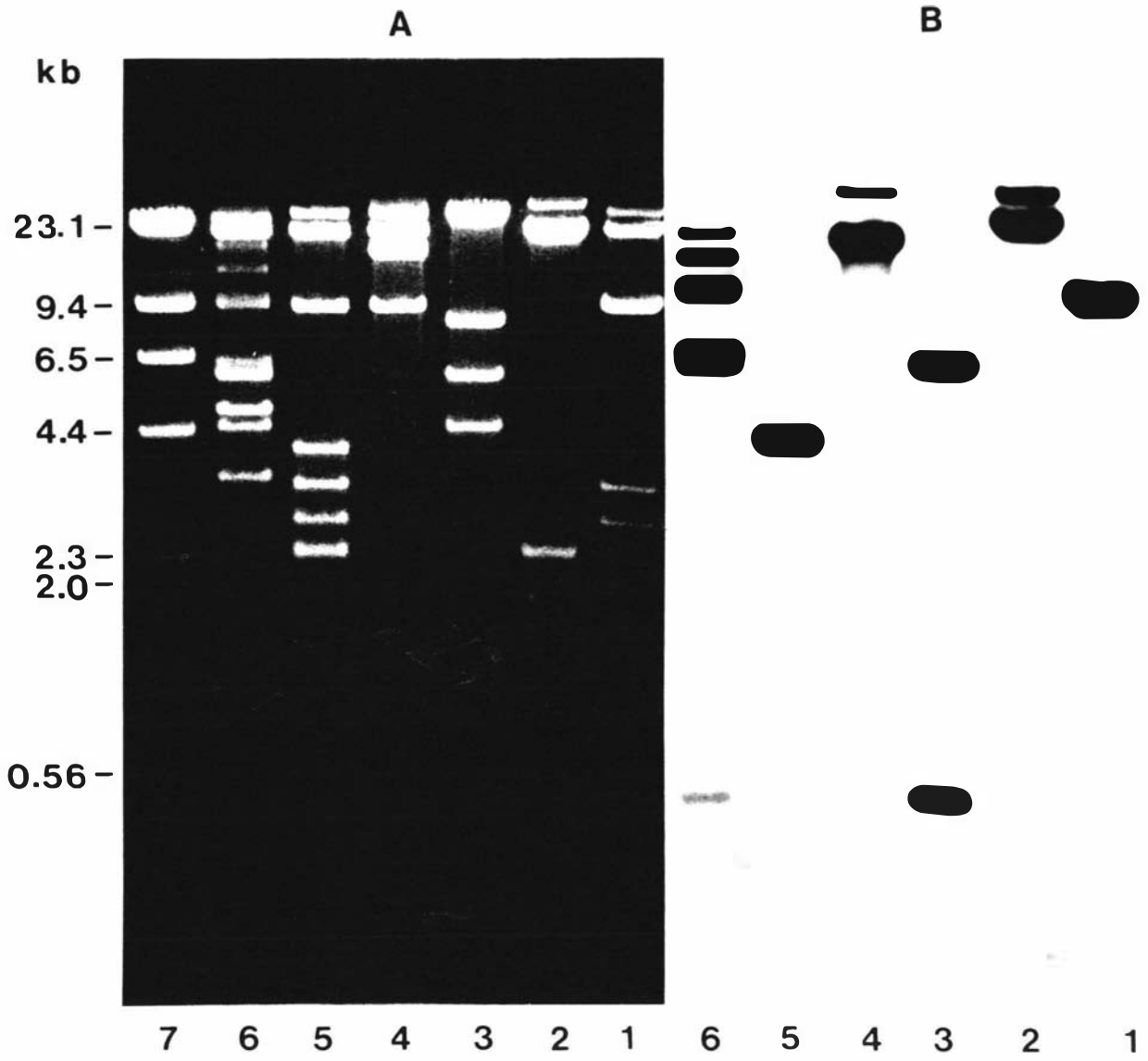


Figure 4.6. Gel electrophoresis on a 0.7% agarose gel (A) and Southern blot (B) of digests of the white clover SSU genomic clone, λTrS64. Lane 1: *Bam*HI; lane 2: *Eco*RI; lane 3: *Hind*III; lane 4: *Sal*I; lane 5: *Bam*HI plus *Eco*RI; lane 6: *Hind*III plus *Sal*I; lane 7: *Hind*III digest of λ DNA.


A
B
kb
23.1
9.4
6.5
4.4
2.3
2.0
0.56
7 6 5 4 3 2 1 6 5 4 3 2 1

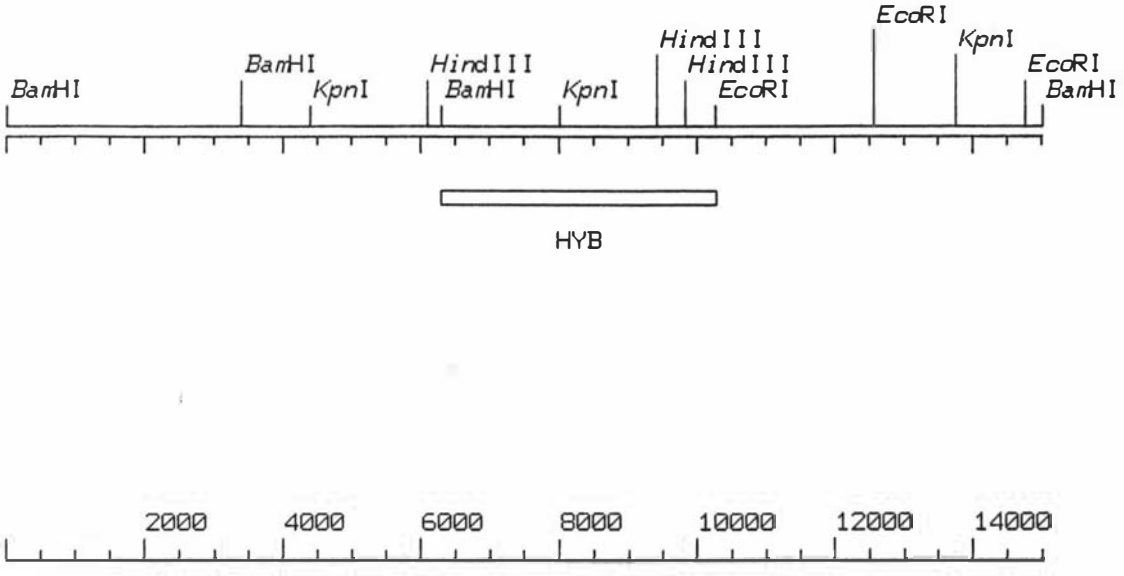


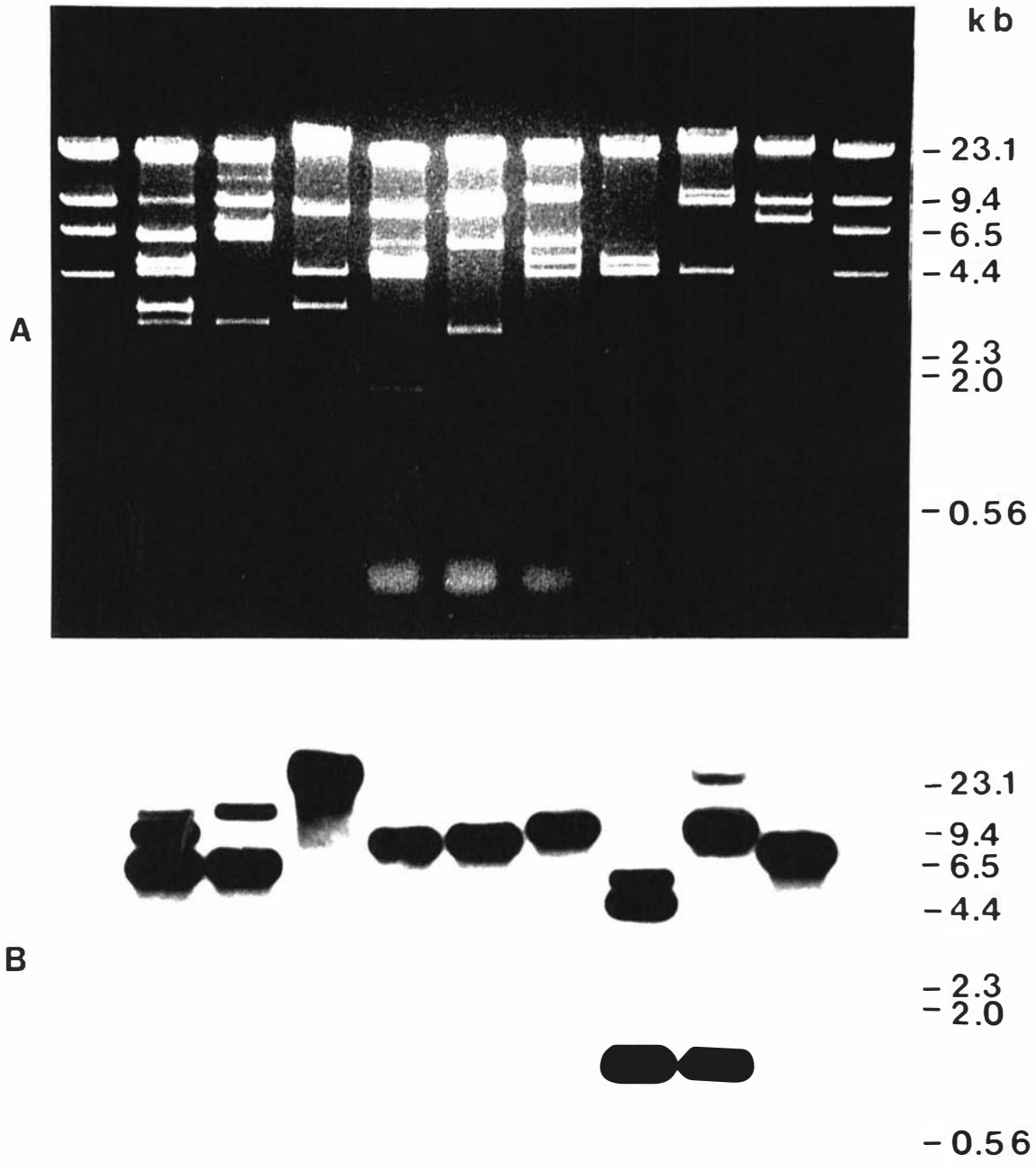
Figure 4.7. Restriction enzyme map of the insert from the white clover SSU genomic clone λTrS64. The restriction fragment to which the white clover SSU cDNA probe hybridizes (HYB) is indicated beneath the map. The scale is in base pairs. This insert is in the opposite orientation with respect to the left and right arms of λEMBL3.

4.8 Isolation of white clover Adh genomic clones

The set of genomic library plaque hybridization filters prepared for screening with the white clover SSU cDNA probe (Section 4.7) was reused for the Adh screening following the removal of the previous probe. The filters were reprobed with the labelled insert from pTrA21, one of the white clover Adh cDNA clones (Section 3.8). After autoradiography, a total of seventeen, putative, positively hybridizing plaques was identified (for typical result see Figure 4.4). Each positively hybridizing plaque was isolated and subjected to a second round of plaque hybridization using the pTrA21 insert probe (for typical result see Figure 4.5). At this stage, nine of the initial isolates were still positive. The five isolates that gave the strongest hybridizing signals at the second round (λTrA11, λTrA12, λTrA15, λTrA21 and λTrA22) were subjected to an additional round of plaque hybridization to further purify the isolates.

DNA was prepared from each of the purified white clover Adh genomic isolates. Restriction enzyme digestion of each DNA revealed that λTrA11 and λTrA15 were identical, as were λTrA12 and λTrA21 (data not shown). DNA samples prepared from λTrA11, λTrA12 and λTrA22 were analyzed further by restriction enzyme digestion and agarose gel electrophoresis (Figure 4.8A). A Southern blot prepared from this gel was hybridized with the pTrA21 insert probe (Figure 4.8B). Less intense hybridization to larger fragments in lanes 3, 4, 9, and 10 (Figure 4.8B) represents incomplete digestion with restriction enzymes. Based on these results as well as on additional restriction enzyme mapping (data not shown), restriction enzyme maps and hybridization profiles of λTrA11, λTrA12 and λTrA22 were constructed (Figures 4.9, 4.10 and 4.11).

Figure 4.8. Gel electrophoresis on a 0.7% agarose gel (A) and Southern blot (B) of digests of the white clover Adh genomic clones. Lanes 1 and 11: *Hind*III digest of λ DNA; lane 2: *Bam*HI digest of λTrA11; lane 3: *Hind*III digest of λTrA11; lane 4: *Bam*HI plus *Hind*III digest of λTrA11; lane 5: *Hind*III digest of λTrA12; lane 6: *Sal*I digest of λTrA12; lane 7: *Hind*III plus *Sal*I digest of λTrA12; lane 8: *Hind*III digest of λTrA22; lane 9: *Sal*I digest of λTrA22; lane 10: *Hind*III plus *Sal*I digest of λTrA22.


kb
A
23.1
9.4
6.5
4.4
2.3
2.0
0.56
B
23.1
9.4
6.5
4.4
2.3
2.0
0.56
11 10 9 8 7 6 5 4 3 2 1

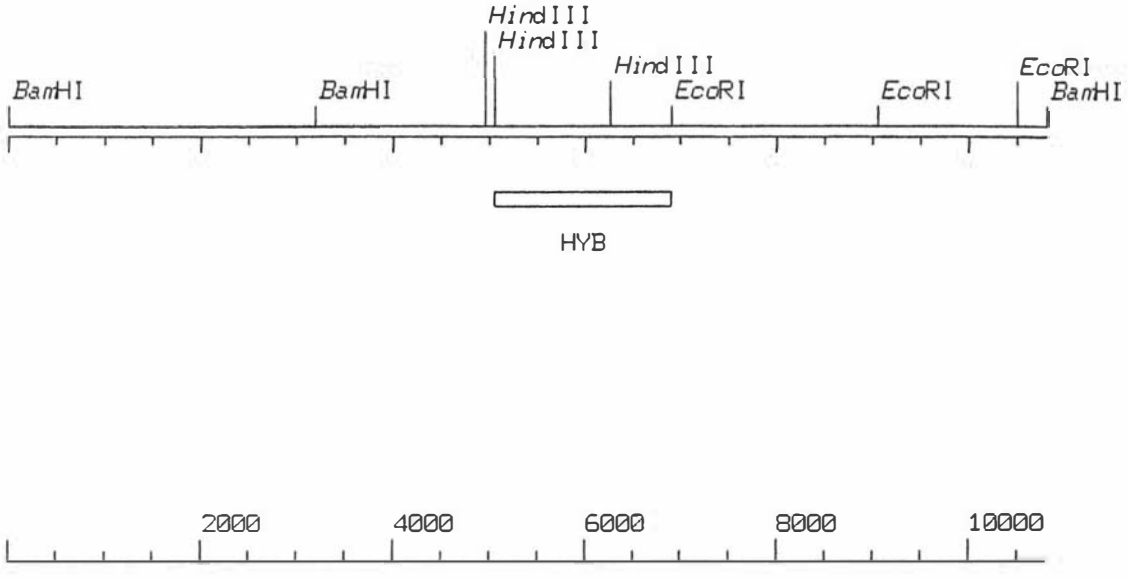


Figure 4.9. Restriction enzyme map of the insert from the white clover Adh genomic clone λTrA11. The restriction fragment to which the white clover Adh cDNA probe hybridizes (HYB) is indicated beneath the map. The scale is in base pairs. This insert is in the same orientation with respect to the left and right arms of λEMBL3.

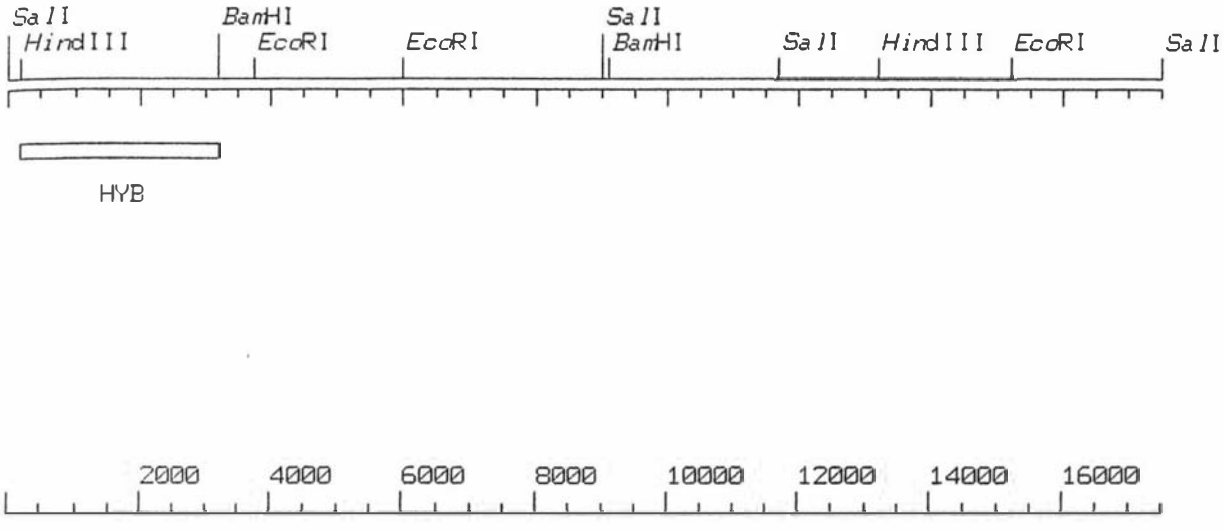


Figure 4.10. Restriction enzyme map of the insert from the white clover Adh genomic clone λTrA12. The restriction fragment to which the white clover SSU cDNA probe hybridizes (HYB) is indicated beneath the map. The scale is in base pairs. This insert is in the opposite orientation with respect to the left and right arms of λEMBL3.

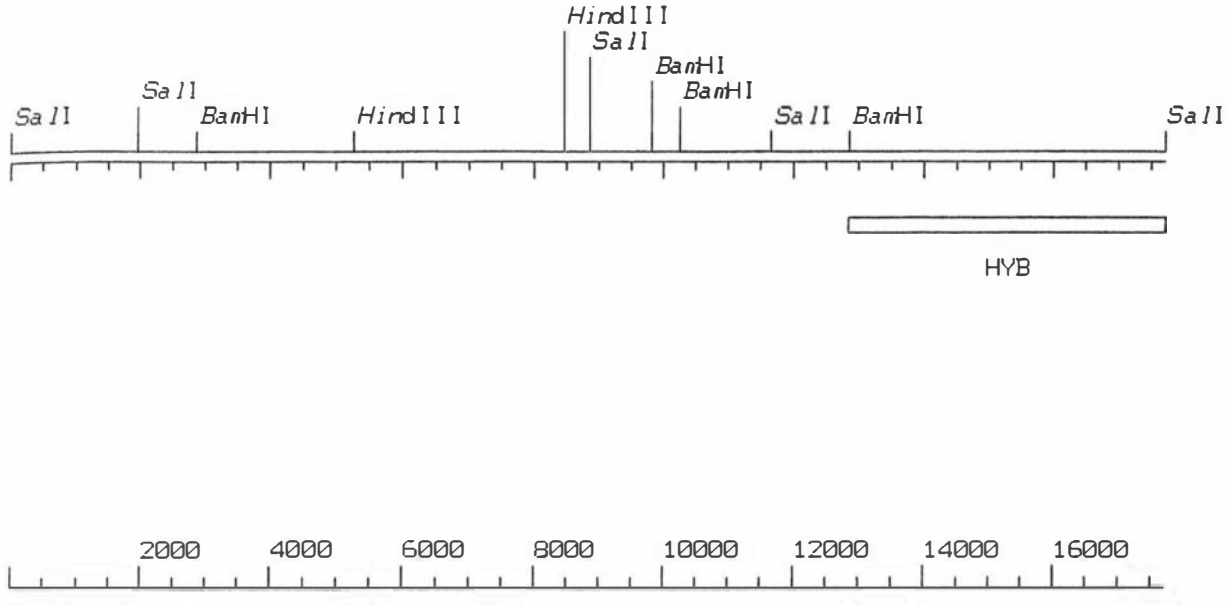


Figure 4.11. Restriction enzyme map of the insert from the white clover Adh genomic clone λTrA22. The restriction fragment to which the white clover Adh cDNA probe hybridizes (HYB) is indicated beneath the map. The scale is in base pairs. This insert is in the opposite orientation with respect to the left and right arms of λEMBL3.

4.9 Subcloning fragments from λEMBL3/genomic clones into plasmid vectors

For each of the white clover SSU and Adh genomic clones isolated (Sections 4.7 and 4.8), fragments spanning the region to which the corresponding white clover cDNA probe hybridized were subcloned into pGEM-3Z and/or pUBS plasmid vectors (see Section 2.15). For λTrS64 (Figure 4.7), the 3.1 kb *Bam*HI - *Eco*RI fragment was subcloned to give pTrS640; for λTrA11 (Figure 4.9), the 7.6 kb *Bam*HI - *Bam*HI fragment was subcloned to give pTrA110; for λTrA12 (Figure 4.10), the 3.7 kb *Sal*I - *Eco*RI fragment was subcloned to give pTrA131; for λTrA22 (Figure 4.11), the 4.8 kb *Bam*HI - *Sal*I fragment was subcloned to give pTrA230.

4.10 Sequence analysis of the white clover small subunit genomic clone

Deletion derivative subclones of the white clover SSU genomic clone pTrS640 (Section 4.9) were generated by a combination of *Bal*31 and exonuclease III digestion of the insert in both directions. The DNA sequence of each of the derivatives was determined and the complete DNA sequence in both directions was assembled from the sequences of overlapping subclones (Figure 4.12). When the sequence of pTrS640 was compared with that of pTrS20, the white clover SSU cDNA clone (Section 3.10), the sequence of pTrS20 was found to be identical to the corresponding region from pTrS640 (Figure 4.13). It was therefore concluded that pTrS640 represented a transcribed white clover *rbcS* gene and that pTrS20 represented a partial cDNA clone of the transcript derived from this gene. The nucleotide sequence of the insert of pTrS640 has been entered into the EMBL nucleotide sequence database under accession number X52293.

Figure 4.12. Nucleotide sequence of the insert from the white clover *rbcS* clone, pTrS640. The three *rbcS* gene exons are underlined.

```
   1  GGATCCGGCC GTCATCTGGC CGTACATCAT CAGACCGGCT TTATCCAGCT
  51  CGTGGAAGTG GTCCCAGGTG GCCCAGTGCG GCACCAGGTT AGAGTTGGCA
 101  ATCAGTACGC GTGGTGCATC GGCATGGGTA CGGAATACGC CCACCGGCTT
 151  ACCGGACTGG ACCAGCAGCG TCTCTTCCGG CTGCAGCGCC TGCAGTGAAC
 201  GCAGGATCTG CTCGAAGCAC GCCCAGTTAC GGGCTGCTTT ACCAATGCCA
 251  CCGTATACCA CCAAATCCTC CGGGCGTTCG GCCACATCCG GGTCGAGATT
 301  GTTCTGGATC ATGCGGTAAG CGGCTTCGAT CAGCCAGTTG GCACAGTGCA
 351  GTTCACTGCC GTGCGGGGCA CGGATCGTGC GGGCGACGGC TTTGCTTACG
 401  GAATCAGTCA TTATTTTTTC CTTTTCAAAC AGGCTTAGCG CGCGCGTGCA
 451  GGCTGCAGGG CGGCGATCGT TTCTGGCCAG TGGTCTGATG TCGCCCACAG
 501  GCGCATCTGC TCAATATCCG GGGCCATCAG GCGATCTTTT TCCAGATAAG
 551  CCACACGTTC GCGGACGCCG GCCAGCACCT GCTCCAGCTG TGGAGAGCTT
 601  TTCAGCGGAC GGATAAAATC AATACCCTGT GCGGCGGCCA TCGCCTCAAT
 651  CCCCACCACG GCAGCCGTGT TAAAGCACAT GGCCCCCAGG CGACGGGCCG
 701  CATAGGTGGC CATCGAAACG TGATCTTCCT GATTGGCGGA GGTAGGCAGG
 751  CTGTCCACGC TCCCCGGGTG GGCGAGGGAT TTGTTTTCAG AGGCCAGCGC
 801  CGCGGCGGTC ACCTGGGCTA TCATAAAACC GGAGTTCACC CCGCCATCGT
 851  TGACCAGGAA GGGCGGCAGG CCGGAAAGGC CGGTATCCAG CAGCAGGGCA
 901  AGCCGGCGCT CGGAAATCGC CCCGATTTCC GCCACGGCCA GGGCGATAAT
 951  ATCGGCGGCA AAGGCAACCG GCTCGGCATG GAAGTTCCCA CCGGAGATCT
1001  TATGTTTTCA ATATTGTATG TTTGGCTACT AATCAAACAA GGTTGATAAG
1051  CCCCATCTTT CTCTCTACCA CCCCTTATCT TATGTACACA CTTTGCTTAT
1101  CTTCAATTTC CCCATTGTAA ATTCTAAACA AAATATAGTA ACAAAAAAAA
1151  ATGTTAATGA AAATTATTGT TGTAATATTA AAAAGCATTG GTCCCATGCA
1201  ATGGTGAAGA TACAAACCTA TTTTCAACTA TCCAACAAAA GGGATAAGAG
1251  TGTCAGTTGA AAGATTTTCT TTGAAGTTCT ATTCCTTCCC CAAGAAATTA
1301  ACATTTACTA TCTAACTTTC AAATTTGATA TTTGATACAT TGACTAAATT
1351  ACATCAACTA AATTAACAAT TAAATTTGAT ATCTAAGTTA ATTTTATTAT
1401  TGTGTAATTT AATTATTTTA ATCTCAAATA AATAGTTAAG ATAATTTGTG
1451  TTTTTTACGG TATGTAGTTT GAAATCCTAA ATTTCACTCC TACACTATAA
1501  AACATAATTG TAGCAACAAA GGTACAAATG CTATTAGTAA TTCACTCAAT
1551  ATAGACATAA GAAATTATGG TCACAATTTC ACAAATATGT TAAGAAGTAA
1601  AATAGTGTGT GAATGTGAAA TAAAAAGATA TCAACTTCAC CACAATCACA
1651  CATTTTATGT ACTTCCAAAG AAGAGATAAG ATTATGGAGC CAAATCCACG
1701  TGGCATTACC ATAGTGGTAC CTAACGATAA GGCTACCATT TCAAAACATT
1751  ACATTCTCTT GTGGCCAATA TCATCAAATA TTCAATCCAA CGGATGACAA
1801  TTTTCAGCCA CCAATTTTTT TCAATCCATA CCATTAGATT AGTGACATCA
1851  AAATTTCTCC ATTATATATA GCAAGTTTGA GTTGAACTAT AAGCAGAAGC
1901  AAAAAGAAGA AGTACTTGGT GAACTAAGAA GGAAAAGAAG AAAAATGGCT
1951  TTGATTTCCT CCGCCGCAGT CACCACCATT AACCGCGCTC CGGTACAAGC
2001  CAACTTGGCT ACTCCATTCA CCGGTCTCAA ATCCTCAGCT GGATTCCCAG
2051  TCACCAAGAA GAACAATGAC ATTACCTCCA TCACAAGCAA TGGTTCAAGA
2101  GTTAACTGCA TGCAGGTAAC ATACCTAATC CATATACCCC ATCAAAAATT
2151  AATTCAATTT CATGTTATGT TACATATATT TATGTACCTT AATGTTGCTT
2201  AATTTTTGGT TTTCGTAACC AATATCCGAC CCACTGAACC GTTTAATCTG
2251  TTTCGGTGGT CAGTTCTGAT TCAAAAAAAT GTTACAAAAA ATAAAATAAT
2301  TGATTTAATT GAAATTTTAT CTACTACAAT GTATGTATAC CAAAATTGAT
2351  ACAAGTAATT AAAATTGTTA AATTTTGCTA ACCCTTTTAC TTAATTTTAT
2401  GTATAATATT TATAGGTGTG GCCACCAGTT GGCAAGAAGA AGTTTGAGAC
2451  CCTTTCATAC CTTCCACCCC TCACTGATGA GCAATTGCTT AAGGAAGTAG
2501  AGTATCTTCT AAGGAAGGGA TGGGTTCCAT GTGTTGAATT TGAGTTGGAG
```

Figure 4.12. (Continued)

```
2551  GTTAATTTTA TTTACTCTTG CAACCCTTTT TTAAAATTTA TTAGTTTACG
2601  AGTTTAATTA CTATACACTA CCAATAAAAA ACTTTTCTAT AGATTGAATA
2651  TATCAACAAA TTATCTCTGT GGCAAAATAT GATTGAATGA AATATTTTAC
2701  ATAGTCAGTA CATATAGGTC TTTTGAAAAT TATATTTTTA TATTGGTTAG
2751  AGAAATTAAG ATTGTCTAAA GTCGATAATC CTTTCATCCG CGCAGCTAGC
2801  ACATCAGTGA CTGTTCTATT GAGATCAGTC CGAACAGTAC GATCTCAATC
2851  GAACAGTGTT CACAGATGTA CTGACTGCGT GAATACTGAA GATTTCTGAC
2901  TGCATATGAC AATATAAATT TGTGTTACGA GTTACAAATA TTTTAGGGTT
2951  GATGTATTTG TTTTGTATTT TATAGAAAGG ATTTGTCCAC CGTCAGTACA
3001  ACAGTTCACC AGGATACTAT GATGGACGTT ACTGGACAAT GTGGAGGTTG
3051  CCATTGTTTG GAACCACTGA TGCTGCTCAG GTGTTGAAGG AAGTTGCTGA
3101  ATGTAAAGCA GAATACCCAG AAGCTTTCAT CCGTATCATC GGATTTGACA
3151  ACGTTCGTCA AGTGCAATGC ATTAGTTTCA TTGCAAGCAC ACCCAAAGTC
3201  TACTAAATTT GATATTTGCA CCACCCTTTA TTACTACTTT GTTTGTACTT
3251  CACCATTGTA AGAACTATAT TTCCCATTTG TTTTATGTTT TTTAATATTT
3301  CATCATCAAT ATAGTATCAT CCTGTTGTAT TTTTGGTTAT TATGTATTCG
3351  GATTTCCATT GGAAATTATG AATGGATGAG AACTATCAAT AATAATAATA
3401  ATATGTTGTT TCTTTGTTCC CAAATTATAT TGCTTCAAAG TAGAGGGTGT
3451  AGGTAGTAAT TTATATATTT GTTGAAACTA CCTATTTCAA ACATTCATTA
3501  CTTACTTACT ATCTTTTGAT TTCCCTCTAG TGGGTTGAAA GCTTCACTAA
3551  ATCAAAGTTT TTATCAAGAT TTCAAATTGA TTTTTGTTGC AATTTTCAAT
3601  ATTATGAAAG TTTGCGGTTA ACATCAAAAA CATTGATGTT ACAATTGTAT
3651  TAGTTGATGT AACATCAACT AATACAATTG TGACTTAGTC ACGAAGACAT
3701  CAACAACATT GATGTCGCGG TCACAATTAT AGTCACATTT AAGAACTTAG
3751  TTATCATTCT TGTGTTAGTT ACTAAATGAG TGTTAACAAA GCAACTTTTA
3801  TGCATAGATG TTCTTATATA TACTTTGGTT ATTCATAAGC GAAATGAAGT
3851  ACATAGTATA ATTTGTGCTG AGCAAATATG ACTAGGGTGC ACACATTTTC
3901  TTGTCTATTT AAGTAATATC TTGCATTCAT GTTATGAGGA AGCTACTAAA
3951  CATCCCAAAC CAAATTGTGA ATTC
```

Figure 4.13. Sequence comparison of part of the SSU genomic clone, pTrS640, with the complete sequence of the SSU cDNA clone, pTrS20. The numbering of the pTrS640 sequence corresponds with the numbering in Figure 4.12. The poly(A) tail of the pTrS20 sequence (Figure 3.17) has been deleted.

```
                2960       2970       2980       2990       3000
                   *          *          *          *          *
pTrS640   GATGTATTTGTTTTGTATTTTATAGAAAGGATTTGTCCACCGTCAGTACA
                                        :::::::::::::::::::::::
pTrS20    -------------------------AGGATTTGTCCACCGTCAGTACA

                3010       3020       3030       3040       3050
                   *          *          *          *          *
pTrS640   ACAGTTCACCAGGATACTATGATGGACGTTACTGGACAATGTGGAGGTTG
          :::::::::::::::::::::::::::::::::::::::::::::::::::
pTrS20    ACAGTTCACCAGGATACTATGATGGACGTTACTGGACAATGTGGAGGTTG

                3060       3070       3080       3090       3100
                   *          *          *          *          *
pTrS640   CCATTGTTTGGAACCACTGATGCTGCTCAGGTGTTGAAGGAAGTTGCTGA
          :::::::::::::::::::::::::::::::::::::::::::::::::::
pTrS20    CCATTGTTTGGAACCACTGATGCTGCTCAGGTGTTGAAGGAAGTTGCTGA

                3110       3120       3130       3140       3150
                   *          *          *          *          *
pTrS640   ATGTAAAGCAGAATACCCAGAAGCTTTCATCCGTATCATCGGATTTGACA
          :::::::::::::::::::::::::::::::::::::::::::::::::::
pTrS20    ATGTAAAGCAGAATACCCAGAAGCTTTCATCCGTATCATCGGATTTGACA

                3160       3170       3180       3190       3200
                   *          *          *          *          *
pTrS640   ACGTTCGTCAAGTGCAATGCATTAGTTTCATTGCAAGCACACCCAAAGTC
          :::::::::::::::::::::::::::::::::::::::::::::::::::
pTrS20    ACGTTCGTCAAGTGCAATGCATTAGTTTCATTGCAAGCACACCCAAAGTC

                3210       3220       3230       3240       3250
                   *          *          *          *          *
pTrS640   TACTAAATTTGATATTTGCACCACCCTTTATTACTACTTTGTTTGTACTT
          :::::::::::::::::::::::::::::::::::::::::::::::::::
pTrS20    TACTAAATTTGATATTTGCACCACCCTTTATTACTACTTTGTTTGTACTT

                3260       3270       3280       3290       3300
                   *          *          *          *          *
pTrS640   CACCATTGTAAGAACTATATTTCCCATTTGTTTTATGTTTTTTAATATTT
          :::::::::::::::::::::::::::::::::::::::::::::::::::
pTrS20    CACCATTGTAAGAACTATATTTCCCATTTGTTTTATGTTTTTTAATATTT

                3310       3320       3330       3340       3350
                   *          *          *          *          *
pTrS640   CATCATCAATATAGTATCATCCTGTTGTATTTTTGGTTATTATGTATTCG
          :::::::::::::::::::::::::::::::::::::::::::::::::::
pTrS20    CATCATCAATATAGTATCATCCTGTTGTATTTTTGGTTATTATGTATTCG
```

Figure 4.13. (Continued)

```
                3360       3370       3380       3390       3400
                  *          *          *          *          *
pTrS640  GATTTCCATTGGAAATTATGAATGGATGAGAACTATCAATAATAATAATA
         ::::::::::::::::::
pTrS20   GATTTCCATTGGAAATTA----------------------------------
```

Based on the comparison between the white clover genomic and cDNA SSU sequences (Figure 4.13), the stop codon and the polyadenylation site have been located within the *rbcS* gene sequence at positions 3204 and 3367 respectively. A putative polyadenylation signal, AATATA, has been identified between positions 3308 and 3313. The sequence of the white clover *rbcS* polyadenylation signal is in good agreement with the proposed consensus plant polyadenylation signal, AATAAA (Joshi, 1987b). Additional putative polyadenylation signals, AATAAT, are present between positions 3388 and 3402. These polyadenylation signals are thus downstream of the polyadenylation site used for the transcript corresponding to the pTrS20 cDNA clone. Multiple polyadenylation sites which give rise to transcripts differing at their 3′ ends have been shown to occur in a number of plant genes (Dean *et al.*, 1986; Hunt and MacDonald, 1989).

It is not yet clear whether transcripts of the *rbcS* gene which differ at their 3′ ends, reflecting the multiple, putative polyadenylation sites, do occur in white clover. The nucleotide in the white clover *rbcS* sequence which corresponds with the first A in the poly(A) tail of the cDNA sequence is an A residue (at position 3368), as is the case for all of the petunia small subunit genes analyzed (Dean *et al.*, 1986).

4.11 Comparison of the white clover *rbcS* sequence with other *rbcS* sequences

A comparison of the sequence of pTrS640 with the published sequences of the coding regions of *rbcS* genes from a variety of plants (Figure 4.14; the identities of the *rbcS* sequences used are listed in Table 4.2) confirmed the identity of the white clover sequence as a *rbcS* gene.

Figure 4.14. Comparison of *rbcS* gene coding sequences from various higher plants. The identity of each sequence is listed in Table 4.2. The processing site (ATG) between the transit peptide and the mature SSU polypeptide for each sequence is indicated in bold type. The white clover sequence (TR) corresponds to positions 1945 - 3206 in Figure 4.12 with the intron sequences (2116 - 2415 and 2551 - 2975) removed. Nucleotides which are conserved across all sequences are indicated (*).

```
AT1       ATGGCTTCCTCTATGCTC------TCTTCCGCTACTATGGT---------
AT2       ATGGCTTCCTCTATGCTC------TCCTCTGCCGCTGTGGT---------
AT3       ATGGCTTCCTCTATGCTC------TCCTCCACCGCTGTGGT---------
AT4       ATGGCTTCCTCTATGCTC------TCCTCCGCCGCTGTGGT---------
GM1       ATGGCTTCCTCAATGATC------TCCTCCCCAGCTGTTAC--CACCGTC
GM2       ATGGCTTCCTCAATGATC------TCTTCCCCAGCTGTTAC--CACTGTC
PS1       ATGGCTTCT---ATGATA------TCCTCTTCAGCTGTGAC--TACAGTC
PS2       ATGGCTTCT---ATGATA------TCCTCTTCCGCTGTGAC--AACAGTC
TR        ATGGCTT------TGATT------TCCTCCGCCGCAGTCAC--CACCATT
HA        ATGGCTTC------GATC------TCCTCCTCAGTCGCG-----ACCGTT
LE1       ATGGCTTCCTCAATTGTC------TCATCGGCAGCCGCTGC--TACCCGT
LE2       ATGGCTTCCTCTGTCATT------TCTTCAGCAGCTGTTGC--CACACGC
LE3       ATGGCTTCTTCAGTAATG------TCCTCAGCAGCTGTTGC--CACCCGC
LE4       ATGGCTTCTTCAGTAATG------TCCTCAGCAGCTGTTGC--CACCCGC
LE5       ATGGCTTCCTCTATAGTT------TCTTCAGCTGCTGTTGC--CACCCGC
NT        ATGGCTTCCTCAGTTCTT------TCCTCTGCAGCAGTTGC--CACCCGC
PE1       ATGGCTTCCTCTGTGATT------TCCTCTGCAGCTGTTGC--TACTCGC
PE2       ATGGCTTCCTCAGTGATG------TCCTCAGCTGCAGTTGC--CACAAGC
LL        ATGGCTTCCTCTATCATGGCTCTGTCCTCCACAGCTGCAGTGGCAGCGGT
OS        ATGGCCCCCTCCGTGATGGC---G---TCGTCGGC--------------
ZM        ATGGCGCCCACCGTGATGAT---GGCCTCGTCGGC--------------
          *****           *          **  *

AT1       -----TGCCTCTCCGGCTCAGGC-----CACTATGGTCGCTCCTTTCAAC
AT2       -----TACCTCCCCGGCTCAAGC-----CACCATGGTCGCTCCATTCACT
AT3       -----TACCTCCCCGGCTCAAGC-----CACCATGGTCGCTCCATTCACC
AT4       -----TACATCCCCGGCTCAGGC-----CACCATGGTCGCTCCATTCACC
GM1       AACCGTGCC-----GGTGC---------CGGCATGGTTGCTCCATTCACC
GM2       AACCGTGCC-----GGTGC---------CGGCATGGTTGCTCCATTCACT
PS1       AGCCGTGCTTCTACGGTGCAATC-----GGCCGCGGTGGCTCCATTCGGC
PS2       AGCCGTGCCTCTAGGGGGCAATC-----CGCCGCAGTGGCTCCATTCGGC
TR        AACCGCGCTCC---GGTACAAGC-----CAACTTGGCTACTCCATTCACC
HA        AGCCGGACCGCCCCTGCTCAGGC-----CAACATGGTGGCTCCGTTCACC
LE1       AGCAATGTT------GCTCAAGC-----TAGCATGGTCGCACCCTTCACC
LE2       AGCAATGTT------ACACAAGC-----TAGCATGGTTGCACCTTTCACT
LE3       GGCAATGGT------GCACAAGC-----TAGCATGGTTGCACCCTTCACT
LE4       GGCAATGGT------GCACAAGC-----TAGCATGGTTGCACCCTTCACT
LE5       GGCAATGGT------GCACAAGC-----TAGCATGGTTGCACCCTTCACT
NT        AGCAATGTT------GCTCAAGC-----TAACATGGTTGCACCTTTCACT
PE1       ACTAATGTG------GCTCAAGC-----TAGCATGGTTGCACCTTTTAAT
PE2       ACCAATGCT------GCTCAAGC-----CAGCATGGTTGCACCCTTCACT
LL        AGCCGCGCC-CTCCAAGACAGGCAACAGCAATGTGGTGTCGGCGTTCACG
OS        --------------------------CACCACCGTCGCTCCCTTCCAG
ZM        --------------------------CACCGCCGTCGCTCCGTTCCAG
                                                *   *  * **
```

Figure 4.14. (Continued)

```
AT1       GGACTTAAGTCCTCCGCTGCCTTCCCAGCCACCCGCAAGGCTAACAA---
AT2       GGTTTGAAGTCATCCGCTTCTTTCCCGGTTACCCGCAAGGCCAACAA---
AT3       GGCTTGAAGTCATCCGCTTCTTTCCCGGTCACCCGCAAGGCCAACAA---
AT4       GGCTTGAAGTCATCCGCTGCATTCCCGGTCACCCGCAAGACCAACAA---
GM1       GGCCTCAAATCCATGGCTGGCTTCCCCACGA---GGAAGACCAACAA---
GM2       GGCCTCAAGTCCATGGCTGGCCTCCCCACCA---GGAAGACCAACAA---
PS1       GGCCTCAAATCCATGACTGGATTCCCAGTTA---AGAAGGTCAACAC---
PS2       GGCCTCAAATCCATGACTGGATTCCCAGTGA---AGAAGGTCAACAC---
TR        GGTCTCAAATCCTCAGCTGGATTCCCAGTCA---CCAAGAAGAACAA---
HA        GGCCTTAAGTCCAACGCCGCCTTCCCCACCACCAAGAAGGCTAAC-----
LE1       GGACTCAAATCCGCCGCTTCTTTTCCCGTTACCAAGAAGAACAACAACGT
LE2       GGTCTCAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG--CAA-AACCT
LE3       GGACTCAAGTCCACCGCTTCTTTCCCTGTTTCAAGGAAG--CAA-AACCT
LE4       GGACTCAAGTCCACCGCTTCTTTCCCTGTTTCAAGGAAG--CAA-AACCT
LE5       GGACTCAAGTCCACTGCTTCTTTCCCTGTTTCAAGGAAG--CAA-AACCT
NT        GGCCTTAAGTCAGCTGCCTCATTCCCTGTTTCAAGGAAG--CAA-AACCT
PE1       GGTCTTAAGTCTGCTGTCTCCTTCCCAGTTTCAAGCAA--GCAA-AACCT
PE2       GGCCTCAAGTCTGCAGCCTCCTTCCCTGTTTCCAGGAA--ACAG-AACCT
LL        GGGCTCAAGTCCATGGCTCAATTCCCTTCCAGCAAGACGATGAGCAACGC
OS        GG-CTCAAGTCCACCGCCGGCATGCC-GTCGCCCG---------CCGTCC
ZM        GGGCTCAAGTCCACCGCCAGCCTCCCCGTCGCCCG---------CCGCTC
          **  * ** **           * **

AT1       CGACATTACTTCC---------ATCACAAGCAACGGCGGAAGAGTTAACT
AT2       CGACATTACTTCC---------ATCACAAGCAATGGGGGAAGAGTTAGCT
AT3       CGACATTACTTCC---------ATCACAAGCAACGGAGGAAGAGTTAGCT
AT4       GGACATCACTTCC---------ATCACAAGCAACGGGGGAAGAGTTAGCT
GM1       TGACATTACCTCC---------ATTGCTAGCAACGGTGGAAGAGTACAAT
GM2       TGACATTACCTCC---------ATTGCTAGCAACGGTGGAAGAGTGCAAT
PS1       TGACATTACTTCC---------ATTACAAGCAATGGTGGAAGAGTAAAGT
PS2       TGACATTACTTCC---------ATTACAAGCAATGGTGGAAGAGTAAAGT
TR        TGACATTACCTCC---------ATCACAAGCAATGGTTCAAGAGTTAACT
HA        -GACTTCTCCACC---------CTTCCCAGCAACGGTGGAAGAGTTCAAT
LE1       TGACATTACCTCC---------CTTGCTAGCAATGGTGGACGCGTTAGAT
LE2       TGACATCACTTCC---------ATTGCTAGCAATGGTGGAAGAGTTAGCT
LE3       TGACATTACCTCC---------ATTGCTAGCAACGGTGGAAGAGTCAGTT
LE4       TGACATTACCTCC---------ATTGCTAGCAACGGTGGAAGAGTCAGTT
LE5       TGACATTACCTCC---------ATTGCTAGCAACGGTGGAAGAGTCAGTT
NT        TGACATCACTTCC---------ATTGCCAGCAACGGCGGAAGAGTGCAAT
PE1       TGACATCACTTCC---------ATTGCTAGCAATGGTGGAAGAGTCCAAT
PE2       TGACATTACTTCC---------ATTGCTAGCAATGGTGGAAGAGTTCAAT
LL        TGGCGCTGAATGGGAGCAAAAGACAACGAGCAACGGCTCCCGCGTACGAT
OS        GAACTCCAGCTTCG-GCAA-----CGTCAGCA-TGGCGGCAGGATCAGGT
ZM        CTCCAGAAGCCTCG-GCAA-----CGTCAGCAACGGCGGAAGGATCCGGT
             *                          ****  **     *  *    *
```

Figure 4.14. (Continued)

```
AT1       GCATGCAGGTGTGGCCTCCGATTGGAAAGAAGAAGTTTGAGACTCTCTCT
AT2       GCATGAAGGTGTGGCCACCAATCGGAAAGAAGAAGTTTGAGACTCTATCT
AT3       GCATGAAGGTGTGGCCACCAATCGGAAAGAAGAAGTTTGAGACTCTATCT
AT4       GCATGAAGGTGTGGCCACCAATTGGAAAGAAGAAGTTTGAGACTCTATCT
GM1       GCATGCAGGTGTGGCCACCAATTGGCAAGAAGAAGTTCGAGACTCTTTCC
GM2       GCATGCAGGTGTGGCCACCAGTTGGCAAGAAGAAGTTTGAGACTCTTTCC
PS1       GCATGCAGGTGTGGCCTCCAATTGGAAAGAAGAAGTTTGAGACTCTTTCC
PS2       GCATGCAGGTGTGGCCTCCAATTGGAAAGAAGAAGTTTGAGACTCTTTCC
TR        GCATGCAGGTGTGGCCACCAGTTGGCAAGAAGAAGTTTGAGACCCTTTCA
HA        GCATGAAGGTGTGGCCACCACTTGGATTGAAGAAGTACGAGACTCTCTCA
LE1       GCATGCAGGTGTGGCCACCAATCAACATGAAGAAATACGAGACATTGTCA
LE2       GCATGCAGGTGTGGCCACCAATTAACATGAAGAAGTACGAGACACTCTCA
LE3       GCATGCAGGTGTGGCCACCAATTAACATGAAGAAGTACGAGACTCTGTCG
LE4       GCATGCAGGTGTGGCCACCAATTAACATGAAGAAGTACGAGACTCTGTCG
LE5       GCATGCAGGTGTGGCCACCAATTAACATGAAGAAGTACGAGACTCTGTCG
NT        GCATGCAGGTGTGGCCACCAATTAACAAGAAGAAGTACGAGACTCTCTCA
PE1       GCATGCAGGTGTGGCCCCCATATGGCAAGAAGAAGTACGAGACTCTCTCA
PE2       GCATGCAGGTGTGGCCACCATACGGCAAGAAGAAGTACGAAACTCTCTCA
LL        GCATGCAGGTGTGGCCTCCTTACGCGAATAAAAAGTTTGAGACTCTGTCG
OS        GCATGCAGGTGTGGCCGATTGAGGGCATCAAGAAGTTCGAGACCCTCTCC
ZM        GCATGCAGGTGTGGCCGGCCTACGGCAACAAGAAGTTCGAGACGCTGTCG
          ***** **********             ** ** *  ** **  * **

AT1       TACCTTCCTGACCTTACCGATTCCGAATTGGCTAAGGAAGTTGACTACCT
AT2       TACCTCCCTGACCTTACTGACGTCGAATTGGCTAAGGAAGTTGACTACCT
AT3       TACCTCCCTGACCTTAGTGACGTTGAATTGGCTAAGGAAGTTGACTACCT
AT4       TACCTCCCTGACCTTAGTGACGTCGAATTGGCTAAGGAAGTTGACTACCT
GM1       TACTTGCCAGACCTCGATGATGCCCAATTGGCCAAGGAAGTCGAATACCT
GM2       TACCTGCCAGACCTTGATGATGCACAATTGGCAAAGGAAGTAGAATACCT
PS1       TATTTGCCACCATTGACCAGAGATCAGTTGTTGAAAGAAGTTGAATACCT
PS2       TATTTGCCACCATTGACGAGAGATCAATTGTTGAAAGAAGTTGAATACCT
TR        TACCTTCCACCCCTCACTGATGAGCAATTGCTTAAGGAAGTAGAGTATCT
HA        TACTTACCACCACTAACTGAAACTCAGTTGGCTAAGGAAGTCGACTACTT
LE1       TACCTTCCTGACTTGTCCGATGAGCAATTGCTTAGCGAAATTGAGTATCT
LE2       TACCTTCCTGATTTGTCTGACGAGCAATTGCTTAGTGAAATTGAGTACCT
LE3       TACCTTCCTGATTTGTCCGACGAGCAATTGCTCAGCGAAATTGAGTACCT
LE4       TACCTTCCTGATTTGTCCGACGAGCAATTGCTCAGCGAAATTGAGTACCT
LE5       TACCTTCCTGATTTGTCCGACGAGCAATTGCTCAGCGAAATTGAGTACCT
NT        TACCTTCCTGATTTGAGCCAGGAGCAATTGCTTAGTGAAGTTGAGTACCT
PE1       TACCTTCCTGATTTAACCGACGAGCAATTGCTCAAGGAGATTGAGTACCT
PE2       TACCTTCCTGATTTGACTGACGAGCAGCTCCTCAAGGAAATTGAGTACCT
LL        TATCTCCCTCGCTTGACCCCGGAGCAACTGGTGAAGGAGGTGGAGTACCT
OS        TACCTGCCACCGCTCACCGTGGAGGACCTCCTGAAGCAGATCGAGTACCT
ZM        TACCTGCCGCCGCTGTCGACGGACGACCTGCTGAAGCAGGTGGACTACCT
          **  * **     *           *  *    *   *  * ** **  *
```

Figure 4.14. (Continued)

```
AT1       T-ATCCGCAACAAGTGGATTCCTTGTGTTGAATTCGAGTTGGAGCACGGA
AT2       T-CTCCGCAACAAGTGGATTCCTTGTGTTGAATTCGAGTTGGAGCACGGA
AT3       T-CTCCGCAACAAGTGGATTCCTTGTGTTGAATTCGAGTTGGAGCACGGA
AT4       T-CTCCGCAACAAGTGGATTCCTTGTGTTGAATTCGAGTTAGAGCACGGA
GM1       T-CTAAGGAAAGGATGGATTCCTTGCTTGGAATTCGAGTTGGAGCACGGT
GM2       T-CTTAGGAAGGGATGGATTCCTTGCTTGGAATTCGAGTTGGAGCACGGT
PS1       T-CTCAGGAAGGGATGGGTTCCTTGCTTGGAATTTGAGTTGGAGAAAGGA
PS2       T-CTGAGGAAGGGATGGGTTCCATGCTTGGAATTTGAGTTGGAGAAAGGA
TR        T-CTAAGGAAGGGATGGGTTCCATGTGTTGAATTTGAGTTGGAGAAAGGA
HA        G-CTCCGCAAAAAATGGGTTCCTTGTTTGGAATTCGAGTTGGAGCACGGT
LE1       T-TTGAAAAATGGATGGGTTCCTTGCTTGGAATTCGAGACTGAGCGCGGA
LE2       T-TTGAAAAATGGATGGGTTCCTTGCTTGGAATTTGAGACTGAGCACGGA
LE3       A-TTGAAAAATGGATGGGTTCCTTGCTTGGAATTCGAGACTGAGCACGGA
LE4       A-TTGAAAAATGGATGGGTTCCTTGCTTGGAATTCGAGACTGAGCACGGA
LE5       A-TTGAAAAATGGATGGGTTCCTTGCTTGGAATTCGAGACTGAGCACGGA
NT        T-TTGAAAAATGGATGGGTTCCTTGCTTGGAATTCGAGACTGAGCACGGA
PE1       T-TTGAACAAGGGATGGGTTCCTTGCTTGGAATTTGAGACTGAGCACGGA
PE2       T-TTGAACAAGGGATGGGTTCCTTGCTTGGAATTCGAGACTGAGCACGGA
LL        G-CTGAAGAACAAGTGGGTGCCCTGCCTGGAATTCGAG---GAGGATGGT
OS        AGCTCCGTTCCAAGTGG-TGCCCTGCCTCGAGTTCAG---CAAGGTCGGA
ZM        G----CGCAACGGCTGGATACCCTGCCTCGAGTTCAG---CAAGGTCGGC
                       *** * ** **  * ** **        **   **

AT1       TTTGTGTACCGTGAGCACGGTAACTCACCCGGATACTATGATGGACGGTA
AT2       TTTGTGTACCGTGAGCACGGAAACACTCCCGGATACTACGATGGACGGTA
AT3       TTTGTGTACCGTGAGCACGGAAACACTCCCGGATACTATGATGGACGATA
AT4       TTTGTGTACCGTGAGCACGGAAACACTCCCGGATACTACGATGGACGGTA
GM1       TTCGTGTACCGTGAGCACAACAGGTCACCTxGATACTATGATGGACGCTA
GM2       TTCGTGTACCGTGAGCACAACAGGTCACTAGGATACTACGATGGACGCTA
PS1       TTTGTGTACCGTGAGCACAACAAGTCACCAGGATACTATGATGGAAGATA
PS2       TTTGTGTACCGTGAGCACAACAAGTCACCAGGATACTATGATGGAAGATA
TR        TTTGTCCACCGTCAGTACAACAGTTCACCAGGATACTATGATGGACGTTA
HA        TTTGTCTACCGTGAGAACGCCAGATCCCCCGGATACTATGACGGAAGATA
LE1       TTTGTGTACCGAGAGAACAACAGTTCCCCTGGATACTACGATGGTAGATA
LE2       TTTGTCTACCGTGAGAACAACAAGTCACCAGGATACTATGATGGAAGGTA
LE3       TTTGTGTACCGTGAGAACCATAAGTCACCAGGATACTACGATGGCAGATA
LE4       TTTGTGTACCGTGAGAACCATAAGTCACCAGGATACTACGATGGCAGATA
LE5       TTTGTGTACCGTGAGAACCATAAGTCACCAGGATACTACGATGGCAGATA
NT        TTTGTCTACCGTGAAAACAACAAGTCACCAGGATACTATGATGGCAGATA
PE1       TTTGTCTACCGTGAATACCACGCCTCACCTAGATACTATGATGGAAGGTA
PE2       TTCGTCTACCGTGAGTACCATGCATCTCCAAGGTACTATGATGGCAGGTA
LL        GAAATAAAGAGAGTGTATGGGAATAGCCCAGGGTACTACGACGGGAGATA
OS        TTTGTCTACCGTGAGAACCACAAGTCCCCTGGATACTACGACGGCAGGTA
ZM        TTCGTGTACCGCGAGAACTCCACCTCCCCGTGCTACTACGACGGCCGCTA
              *  *  *     *           *   * ***** ** **  * **
```

Figure 4.14. (Continued)

```
AT1       CTGGACAATGTGGAAGCTTCCCTTGTTCGGTTGCACCGACTCCGCTCAAG
AT2       CTGGACAATGTGGAAGCTTCCATTGTTCGGATGCACCGACTCTGCTCAAG
AT3       CTGGACAATGTGGAAGCTTCCATTGTTCGGATGCACCGACTCCGCTCAAG
AT4       CTGGACAATGTGGAAGCTTCCATTGTTCGGATGCACCGACTCCGCTCAAG
GM1       CTGGACCATGTGGAAGCTGCCTATGTTTGGCTGCACTGATGCTTCTCAGG
GM2       CTGGACCATGTGGAAGCTGCCTATGTTTGGTTGCACTGATGCTTCTCAGG
PS1       CTGGACAATGTGGAAGCTTCCTATGTTTGGTACCACTGATGCTTCTCAAG
PS2       CTGGACAATGTGGAAGCTTCCTATGTTTGGTACCACTGATGCTTCTCAAG
TR        CTGGACAATGTGGAGGTTGCCATTGTTTGGAACCACTGATGCTGCTCAGG
HA        CTGGACAATGTGGAAATTGCCTATGTTCGGTTGCACCGACTCAGCCCAAG
LE1       CTGGACCATGTGGAAGTTACCTATGTTTGGGTGCACTGATGCAACACAGG
LE2       CTGGACCATGTGGAAGTTGCCTATGTTTGGGTGCACTGATGCAACCCAAG
LE3       CTGGACCATGTGGAAGTTGCCCATGTTCGGGTGCACTGATGCAACCCAGG
LE4       CTGGACCATGTGGAAGTTGCCCATGTTCGGGTGCACTGATGCAACCCAGG
LE5       CTGGACCATGTGGAAGTTGCCCATGTTTGGGTGCACTGATGCAACCCAGG
NT        CTGGACCATGTGGAAGCTACCTATGTTCGGATGCACTGATGCCACCCAAG
PE1       CTGGACCATGTGGAAGTTGCCCATGTTTGGGTGCACTGATGCAACTCAGG
PE2       CTGGACCATGTGGAAGCTGCCCATGTTCGGGTGCACCGATGCCACCCAAG
LL        CTGGGTGATGTGGAAGCTGCCTATGTTCGGATGCACAGAGGCATCGCAGG
OS        CTGGACCATGTGGAAGCTGCCCATGTTCGGGTGCACCGACGCCACCCAGG
ZM        CTGGACCATGTGGAAGCTGCCCATGTTCGGCTGCAACGACGCCACCCAGG
          ****   *******   * **  **** **   **  **  *  * ** *

AT1       TGTTGAAGGAAGTGGAAGAGTGCAAGAAGGAGTACCCCAATGCCTTCATT
AT2       TATTGAAGGAAGTTGAAGAATGCAAGAAGGAGTACCCGGGCGCCTTCATT
AT3       TGTTGAAGGAAGTTGAAGAATGCAAGAAGGAGTACCCTGGCGCCTTCATT
AT4       TGTTGAAGGAGGTTGAAGAATGCAAGAAGGAGTACCCGGGCGCCTTCATT
GM1       TGTTGAAGGAGCTTCAAGAGGCTAAGACTGCATACCCCAATGGCTTCATC
GM2       TGTTGAAGGAGCTTCAAGAGGCTAAGACTGCATACCCCAACGGCTTCATC
PS1       TCTTGAAGGAGCTTGATGAAGTTGTTGCCGCTTACCCCCAAGCTTTCGTC
PS2       TCTTGAAGGAGCTTGATGAAGTTGTTGCCGCTTACCCTCAAGCTTTCGTT
TR        TGTTGAAGGAAGTTGCTGAATGTAAAGCAGAATACCCAGAAGCTTTCATC
HA        TGATGAAGGAGCTTGCTGAATGCAAGAAGGAGTACCCCCAGGCCTGGATC
LE1       TGTTGGCTGAGGTTCAAGAGGCGAAGAAGGCGTACCCACAAGCCTGGGTT
LE2       TGTTGGCTGAGGTTCAAGAGGCTAAAAAGGCATACCCACAAGCATGGGTC
LE3       TCTTGGCTGAGGTGCAGGAGGCAAAGAAGGCTTACCCACAGGCATGGGTC
LE4       TCTTGGCTGAGGTGCAGGAGGCAAAGAAGGCTTACCCACAGGCATGGGTC
LE5       TCTTGGCTGAGGTGCAGGAGGCAAAGAAGGCTTACCCTCAGGCATGGGTC
NT        TGTTGGCTGAGGTGGAAGAGGCGAAGAAGGCATACCCACAGGCCTGGATC
PE1       TGTTGGGTGAGCTCCAAGAGGCCAAGAAGGCTTACCCTAATGCATGGATC
PE2       TCTTGGGTGAGCTCCAAGAGGCCAAGAAGGCTTACCCCAATGCCTGGATC
LL        TGTTGAACGAGGTGAACGAGTGTGCGAAGGCATACCCCAACGCCTTCATC
OS        TCGTCAAGGAGCTCGAGGAGGCCAAGAAGGCGTACCCTGATGCATTCGTC
ZM        TGTACAAGGAGCTGCAGGAGGCCATCAAATCCTACCCGGACGCCTTCCAC
          *       **  *    **             *****    *  *
```

Figure 4.14. (Continued)

```
AT1       AGGATCATCGGATTCGACAACACCCGTCAAGTCCAGTGCATCAGTTTCGT
AT2       AGGATCATCGGATTCGACAACACCCGTCAAGTCCAGTGCATCAGTTTCAT
AT3       AGGATCATCGGATTCGACAACACCCGTCAAGTCCAATGCATCAGTTTCAT
AT4       AGGATCATCGGATTCGACAACACCCGTCAAGTCCAATGCATCAGTTTCAT
GM1       CGTATCATTGGATTTGACAACGTTCGCCAAGTGCAGTGCATCAGCTTCAT
GM2       CGTATCATCGGATTCGACAACGTTCGCCAAGTGCAGTGCATCAGCTTCAT
PS1       CGTATCATCGGTTTCGACAACGTTCGTCAAGTTCAATGCATCAGTTTCAT
PS2       CGTATCATCGGTTTCGACAACGTTCGTCAAGTTCAATGCATCAGTTTCAT
TR        CGTATCATCGGATTTGACAACGTTCGTCAAGTGCAATGCATTAGTTTCAT
HA        CGTATCATCGGATTTGACAATGTTCGTCAAGTTCAATGTATCATGTTCAT
LE1       CGTATTATCGGATTCGACAACGTTCGTCAAGTCCAGTGCATCAGCTTCAT
LE2       AGAATCATTGGATTCGACAATGTGCGTCAAGTGCAGTGTATCAGTTTCAT
LE3       CGTATCATCGGATTCGACAATGTTCGTCAAGTGCAGTGCATCAGTTTCAT
LE4       CGTATCATCGGATTCGACAATGTTCGTCAAGTGCAGTGCATCAGTTTCAT
LE5       CGTATCATCGGATTCGACAATGTTCGTCAAGTGCAGTGCATCAGTTTCAT
NT        CGTATCATTGGATTCGACAACGTGCGTCAAGTGCAGTGCATCAGTTTCAT
PE1       AGAATCATCGGATTCGACAACGTCCGTCAAGTGCAATGCATCAGTTTCAT
PE2       AGAATCATTGGATTCGACAACGTGCGTCAAGTGCAATGCATCAGTTTCAT
LL        CGCGTCATCGGATTCGACAACGTCCGCCAAGTGCAGTGCATCTCCTTCAT
OS        CGTATCATCGGCTTCGACAACGTTAGGCAGGTGCAGCTCATCAGCTTCAT
ZM        CGCGTCATCGGCTTCGACAACATCAAGCAGACGCAGTGCGTCAGCTTCAT
           *  * ** ** ** *****       **    **     *    *** *

AT1       TGCCTACAAGCC------ACCAAGCTTCACCGG---TTAA
AT2       TGCCTACAAGCC------CCCAAGCTTCACTGATGCTTAA
AT3       TGCCTACAAGCC------CCCAAGCTTCACCGAAGCTTAA
AT4       TGCCTACAAGCC------CCCAAGCTTCACCGAAGCTTAA
GM1       CGCCTACAAGCC------CCCAGGCTTC---------TAA
GM2       CGCCTACAAGCC------CCCAAGCTTC---------TAA
PS1       TGCCCACACACC------AGAATCCTAC---------TAA
PS2       TGCACACACACC------AGAATCCTAC---------TAA
TR        TGCAAGCACACC------CAAAGTCTAC----------TAA
HA        TGCTTCCAGGCC------AGATGGTTAC---------TAA
LE1       TGCCTACAAGCC------AGAAGGATTC---------TAA
LE2       TGCCTACAAGCC------AGAAGGCTAC---------TAA
LE3       CGCTTACAAGCC------CGAAGGATAC---------TAA
LE4       CGCTTACAAGCC------CGAAGGATAC---------TAA
LE4       CGCTTACAAGCC------CGAAGGATAC---------TAA
NT        TGCCTACAAGCC------AGAAGGCTAC---------TAA
PE1       TGCCTACAAGCC------ACCAGGCTTC---------TAA
PE2       TGCCTACAAGCC------CCCAGGCTAC---------TAG
LL        CGTCCACAAGCC------TGAA---TACAA------TTAA
OS        CGCCTACAACCCGGGCTGCGAGGAGTCTGGTGGCAACTAA
ZM        CGCCTACAA----GCCCCCG--------GGCAGCGACTAG
           *    **                          **
```

Table 4.2. Identities of plant *rbcS* gene sequences used for the multiple alignment presented in Figure 4.14 and the dendrogram presented in Figure 4.15.

Sequence	Plant	EMBL/Genbank Acession No.
AT1	*Arabidopsis*	X13611
AT2	*Arabidopsis*	X13610
AT3	*Arabidopsis*	X13610
AT4	*Arabidopsis*	X13610
GM1	soybean	V00458
GM2	soybean	M16889
PS1	pea	X04333
PS2	pea	X04334
TR	white clover	X52293
HA	sunflower	Y00431
LE1	tomato	X05982
LE2	tomato	X05983
LE3	tomato	X05984
LE4	tomato	M13544
LE5	tomato	X05985
NT	tobacco	X02353
PE1	petunia	X03820
PE2	petunia	X03821
LL	larch	X16039
OS	rice	X07515
ZM	maize	Y00322

This sequence comparison also allowed the identification of the complete coding region of the white clover *rbcS* gene (1945 - 3206 in Figure 4.12) in the absence of a complete cDNA sequence, as well as the assignment of intron-exon boundaries within the gene. The translation start codon (ATG) of the white clover gene is located at position 1945 and the processing site between the transit peptide and the mature small subunit polypeptide (ATG) is located at position 2110 (Figure 4.12).

As can be seen from the multiple alignment of *rbcS* sequences (Figure 4.14), there is a high degree of sequence conservation at the processing site between the transit peptide and the mature SSU polypeptide as well as to either side of this site. Within the sequence for the transit peptide itself, there are regions of homology located in the amino-terminal, the central and the carboxy-terminal portions of the peptide. By deletion analysis of the transit peptide sequence, it has been shown that the amino-terminal region of the peptide is important for transport into chloroplasts while the carboxy-terminal portion is important for both transport and for correct processing of the SSU precursor into the mature form of the SSU (Reiss *et al.*, 1987). In contrast, the middle portion of the transit peptide is dispensable for transport (Reiss *et al.*, 1987). Within the sequence for the mature SSU polypeptide, approx. 37% of the nucleotides are conserved across all of the *rbcS* sequences.

The coding region of the white clover *rbcS* gene includes two introns and both of these introns have the GT-AG dinucleotide border elements (positions 2116/2415 and 2551/2975 in Figure 4.12) which are almost invariably found in plant introns (Brown, 1986; Hanley and Schuler, 1988) (Section 1.3.4). Two introns occur within the coding regions of most of the *rbcS* genes of dicotyledenous plants that have been sequenced, with the exception of one of the five *rbcS* genes of tomato (Sugita *et al.*, 1987), one of the eight *rbcS* genes of petunia (Dean

et al., 1985; Tumer *et al.*, 1986) and the one tobacco *rbcS* gene that has been sequenced (Mazur and Chui, 1985) (see Section 1.7.3.11). It is probably significant from an evolutionary point of view that all of the three intron *rbcS* genes have been isolated from plants belonging to one family, the *Solanaceae*. In contrast, *rbcS* genes containing one intron have been reported in the monocotyledenous plants rice (Xie and Wu, 1988), maize (Lebrun *et al.*, 1987) and wheat (Broglie *et al.*, 1983).

The position of introns in *rbcS* genes has been found to be highly conserved within higher plants (Wolter *et al.*, 1988). The intron positions of the *Solanaceae* three intron *rbcS* genes are identical with respect to the coding sequence of the mature polypeptide. This is also the case with all of the two intron *rbcS* genes, where the intron positions correspond to the positions of intron 1 and 2 of the three intron genes. The positions of the single intron in the monocot *rbcS* genes correspond to the position of intron 1 of the dicot genes. In all cases, the first exon of the gene codes for the transit peptide and the first codon of the mature SSU polypeptide (Figure 4.14).

The relationships between the *rbcS* sequences used in the multiple alignment of *rbcS* sequences (Figure 4.14) are presented in the form of a dendrogram in Figure 4.15 (the identities of the *rbcS* sequences used are listed in Table 4.2). In the dendrogram, the white clover sequence is clearly grouped with the *rbcS* sequences of other legumes. Within the group of legume *rbcS* sequences, the white clover sequence is most closely related to the *rbcS-3A* (PS1) and *rbcS-3C* (PS2) genes of pea (Fluhr *et al.*, 1986a).

The complete coding region of the white clover gene (1945 - 3206 in Figure 4.12) exhibits 79% identity to the pea *rbcS-3A* and *rbcS-3C* genes (Fluhr *et al.*, 1986) although the mature polypeptide coding region (2110 - 3206 in Figure 4.12)

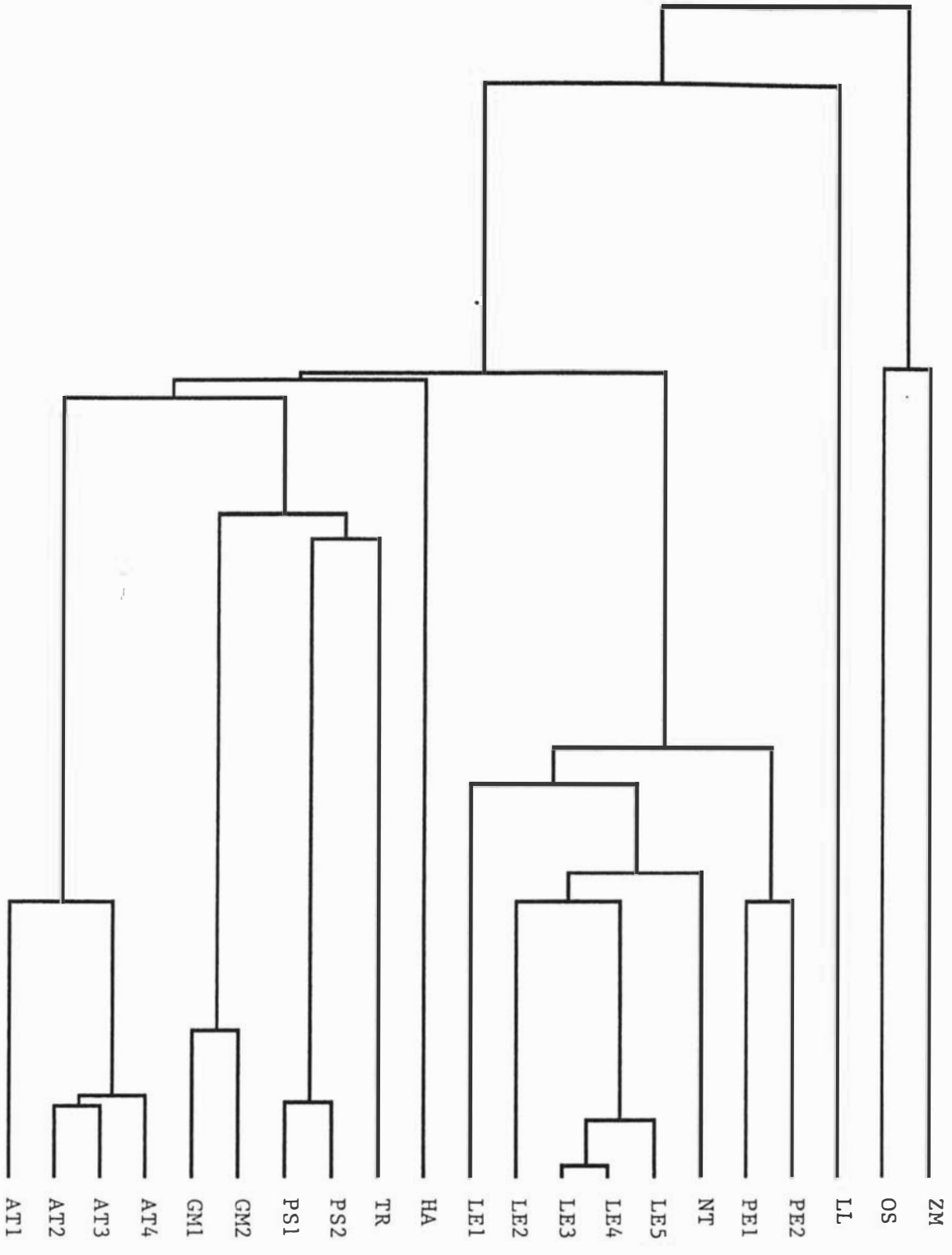


Figure 4.15. Dendrogram of the *rbcS* sequences used in the multiple sequence alignment presented in Figure 4.14. The identity of each sequence is listed in Table 4.1.

exhibits a slightly higher identity to the *rbcS-3C* gene (82%) than it does to the *rbcS-3A* gene (81%). The transit peptide coding region of the white clover gene (1945 - 2110 in Figure 4.12) exhibits 74% and 73% identity to the pea *rbcS-3A* and *rbcS-3C* genes respectively, reflecting the higher divergence of transit peptide sequences relative to those of the mature polypeptide (Karlin-Neumann and Tobin, 1986; Wolter *et al.*, 1988). No significant homology was detected between the introns of the white clover gene and the corresponding introns of the pea genes.

4.12 Analysis of white clover *rbcS* 5′ sequences

A number of conserved sequence elements involved in the regulation of gene expression has been identified within the upstream sequence of *rbcS* genes from a variety of plants (Green *et al.*, 1987; Manzara and Gruissem, 1988; Giuliano *et al.*, 1988) (see Section 1.8.4). Analysis of the upstream sequence of the white clover *rbcS* gene revealed a number of regions that exhibited strong homology with some of these conserved elements (Figure 4.16).

The white clover *rbcS* sequence elements include the eukaryotic TATA box, a sequence similar to the eukaryotic CAAT box, the box II sequence GTGTGGTTAATATG, the box I sequence, the G box sequence, and the I box sequence (see also Figure 1.2). It is likely that these conserved sequences in the white clover *rbcS* gene have functions similar to those determined for conserved elements in *rbcS* genes of other plant species (Section 1.8.4).

```
1401   TGTGTAATTT AATTATTTTA ATCTCAAATA AATAGTTAAG ATAATTTGTG

1451   TTTTTTACGG TATGTAGTTT GAAATCCTAA ATTTCACTCC TACACTATAA

1501   AACATAATTG TAGCAACAAA GGTACAAATG CTATTAGTAA TTCACTCAAT

1551   ATAGACATAA GAAATTATGG TCACAATTTC ACAAATATGT TAAGAAGTAA

1601   AATAGTGTGT GAATGTGAAA TAAAAAGATA TCAACTTCAC CACAATCACA

                                  I box                      G
1651   CATTTTATGT ACTTCCAAAG AAGAGATAAG ATTATGGAGC CAAATCCACG

       box                                                Box I
1701   TGGCATTACC ATAGTGGTAC CTAACGATAA GGCTACCATT TCAAAACATT

                       Box II
1751   ACATTCTCTT GTGGCCAATA TCATCAAATA TTCAATCCAA CGGATGACAA

                                                         CAAT
1801   TTTTCAGCCA CCAATTTTTT TCAATCCATA CCATTAGATT AGTGACATCA

       box          TATA box
1851   AAATTTCTCC ATTATATATA GCAAGTTTGA GTTGAACTAT AAGCAGAAGC

1901   AAAAAGAAGA AGTACTTGGT GAACTAAGAA GGAAAAGAAG AAAAATGGCT

1951   TTGATTTCCT CCGCCGCAGT CACCACCATT AACCGCGCTC CGGTACAAGC
```

Figure 4.16. Conserved sequence elements identified upstream of the white clover *rbcS* gene. The conserved elements are underlined; the start codon (ATG) is indicated by bold type; the numbering corresponds with that in Figure 4.12.

4.13 Sequence analysis of the white clover Adh genomic clones

Deletion derivative subclones of the white clover Adh genomic clones pTrA110, pTrA131 and pTrA230 (Section 4.9) were generated by exonuclease III digestion of the insert and the sequence of each derivative was determined. Where exonuclease III deletion derivatives did not span the entire region of interest, additional derivative subclones were generated by subcloning appropriate restriction enzyme-generated fragments. Where it was not possible to resolve sequence ambiguities due to sequence compressions, the complementary strands of particular subclones were also sequenced. The DNA

sequence of each of the inserts from the original plasmid subclones, pTrA110, pTrA131 and pTrA230, was assembled from overlapping subclones as well as from the sequence of complementary strands where necessary (Figures 4.17, 4.18 and 4.19). For each of the pTrA110 and pTrA230 inserts (see Section 4.9), the sequence of the entire insert was not determined.

A comparison of the sequence of pTrA110 with that of pTrA45, the white clover Adh cDNA clone (Section 3.11) revealed that there was strong homology between these two sequences within only a limited region of the genomic sequence (Figure 4.20). The gaps introduced by the alignment program in the pTrA45 sequence between positions 634 and 635 and between positions 716 and 717 suggested the presence of introns in the pTrA110 sequence in these regions. Similarly, comparisons of the pTrA45 sequence with that of pTrA131 (Figure 4.21) and with that of pTrA230 (Figure 4.22) also revealed regions of strong homology. In the case of the pTrA131 sequence, no putative intron sequences were apparent but the gap introduced in the pTrA45 sequence between positions 717 and 718 suggested the presence of one putative intron in the pTrA230 sequence. No additional regions of homology were detected outside the region of each of the genomic clone sequences selected for the above alignments.

Clearly the three white clover Adh genomic clones represent different portions of the white clover genome but do not represent functional Adh genes. No open reading frames of significant length were detected in any of the sequences. The strong, though limited, sequence homology that each of these clones exhibits with the white clover Adh cDNA sequence can account for the initial isolation of these genomic clones with the white clover Adh cDNA probe.

Since all three of the white clover Adh genomic clones did have some sequence homology with the white clover cDNA

sequence and two out of the three appeared to have at least one intron, each sequence was compared with the sequence of a known Adh gene. The gene chosen was the pea *Adh1* gene, since the white clover Adh cDNA sequence had already been shown to exhibit strong homology with the pea Adh1 cDNA sequence (Section 3.11). The coding region of the pea *Adh1* gene spans 2106 bp and includes 9 introns (Llewellyn *et al.*, 1987). A comparison of the sequences (Figure 4.23) indicated that the pTrA131 sequence exhibited strong homology with all of exon 3 and with approx. 40% of exon 4 of the pea gene. In contrast, the pTrA110 sequence exhibited strong homology with all of exon 4, exon 5 and exon 6 of the pea gene while pTrA230 exhibited strong homology with all of exon 5, exon 6 and exon 7 of the pea gene. No additional regions of homology with the pea gene sequences were detected for any of the white clover sequences. Where putative introns occur within the white clover sequences, the GT-AG dinucleotide border elements which are almost invariably found in plant introns (Brown, 1986; Hanley and Schuler, 1988) (Section 1.3.4) are preserved.

```
   1  TTTTTTAATC TTTGTCATTC ACCATCAAGA AATAAGCATT TGCGACATTT
  51  CATTTGCGAC CCTTAAGTCA AAACTGTCGC AATGTCTTTT TTTTCAAAAT
 101  TTGGCGACAC TGTATGCGAC AGTTTTGGCC ACCGTTGCAA CTTTTAGATC
 151  TAAAACCAAA TTTTAAGCTA GCGCCAAGTG AATATTTTTG TGAAATCTCT
 201  CACTTAGGCA AACTGAAATC CCTCTTCACT TTTTGTTACA GTATAAACCA
 251  AAAAAATCTC AAATCCCTAA ATTCCTTTTC CAAATCTCTT GAATCAACAA
 301  ACCAAGCTTC TCCTATATGA TAAATCACTC TCCAAATCGT GCTCCCCTCA
 351  TCGTTCCAAT CCCTACAAAT CCCTAAATCC CCTTCATCGT TCCCATCAAC
 401  GAAGCTTAAT CGTCTTCATC GTTTTGTGTC TCTTCTTGTT TCGTTATCTC
 451  TCATTATGTC TCCGATAACC AGGCTTCACT TTCACTACTC GGGTTCCCGC
 501  CGCACACCCG CCTGTCACGT CCATCTCCGC CGATACACCC TCCGCAGTTT
 551  ATACGTTCAG GTTCAGGTTT GTTTTCCCTT CTTCCTCCTT CCCTTTCATC
 601  GTTCTACTTA GTTTTCCATT GGGTTTGAGA ATCACGATTA TGGATCTGAG
 651  GAATAGTGTT TGGTGGTCAG AGGGATGTGA GTTGAGTTTG AGAATCGCGA
 701  TTATGGGTAT GAGGAATGGG TTTGAGAATT ACAGAATCAC TTGATGTTTA
 751  TTAGAAACGC ATAACCTCGT TTGTTTACCA GAATCAGTTG ATGTTTTCAT
 801  TTTATTCATT TTTTTCTTTT ATCTGTTATG TAGGTTTGGG TCCTCAATTT
 851  GATAATGACC CTTTAATTTG CTTCAGTTCC ATTTGAGCTC TGATTTCTTT
 901  TATGTGTTAA AGTAGGTAAT AATAATGAAG TTTAGATTGT CCATTGATAA
 951  ATTTCTAAAA TGCATGCTAT ATAATGTTTT TTCTTCATTG AACCCCTATA
1001  TTTGGTTGTA TTGTATTTGA TTTATGTTGT GTTTCTTTAT TAGGTACATA
1051  GTGAAAGAAG TAGCATTATC GGTTCCATTA GTGTTAGTCA CTTTCACTAC
1101  TAGAAATAAG ACATTTGGTT TGAATGAATG TTATTGATAA TTGTTGCCAT
1151  GATATTGATC AGGATTGTGG AGAGCATAGG CGAGGGTGTG ACTCATCTGA
1201  AACCAGGGGA CACGCCCTGC CAGTATTCAC AGGGGAGTGT GGGGAATGTC
1251  CACATTGTTA AGTCAGAGGA GAGTAGCATG TGTGATCTTC TTAGAATCAA
1301  CACAGACAGA GGTGTCATGA TCAATGACAA CCAGTCTAGA TTCTCTATTA
1351  AGGGACAACT CGTACACCAT TTTGTCGGTA CTTCTACATT CAGCAAGTAC
1401  ACTGTGGTCC ATGCTGGTTG TGTTGCAAAG ATCAACCCTG ATGCACCACT
1451  TGACAAAGTT TGTATTCTCA GTTACGGATT ATGCACAGGT ATGAGCATGT
1501  TTAGATTGAT TTATTTGAGT TTTATTGTTA GCATACATAT TTGTGAGACT
1551  GTTTGATAGA ATTTATGAAA CAACTTATGA CATGTTTTTT GAGCTTATTA
1601  TTATAAGCTT TCTTAAATAG CTTATGAAAA TAGTTTATAT CTTACATAAA
1651  AACAATTTGA CTTTATTTTA TCCTTAGTGA TAAAAATAGC TTATACATAA
1701  AGACTTATAT GATAAGGGTT TATTCCAAAC ATTTTAAAAG ATCCCTATCT
1751  TAAAATATTA ATATGTAACT AGGTTGAGTT CATGCTTGTA GGTCTTGGTG
1801  CTACTGTCAA TGTTGCAAAA CCGAAACCCG GTTTATCCGC TTCTATCTTT
1851  GGACTAAGAG CCGCTGGCCT TGCTGTATGT TATTCTCATT CGCTTTAATA
1901  CAAATTTCAG GCTGCTGAAG GGGCAAGGAT ATCTGGTGCA TTGCAAATCA
1951  TTGGAGTTGA TTTAGTTTCC AGCCGATTTG AATTAGGTAC AATGGCACTC
2001  AATTTTAATC CAATGGCACT CAATTTTAAT CCAAAATCAA AATCAAATTA
2051  ACTCAAACAC TAAACATGCT AAAACAACAC AACCACAATT TCAACTCAAA
2101  AATCAAAACA ACTAAGAAAA TAGAAAAATA TCAACAAGTC CATTAATTGA
2151  TTTAATCTTA AAAAGTAGTA GCTACACTAT CTATAGTTCT TTTGAGTTCT
2201  CAATACTTGT TGTACTAAAC ATGACCATTA CATTTGTAGT AATGTGCTCC
2251  ATGAATTCTA GTTTTTTGCA AAATTAGTTG CTATGTGAGT TATATATATA
2301  TATCACTGTT GGATTTTCTT AGAGGTTAAT CAAATGAAAC AAGTTTTACT
2351  TTCATTTTGA ATGTAGCTTC GTTGTTAATT GGATGGTAAT GTCTTATATC
2401  ACTACTCAAG CAGATGCAGA GACGTTAGCG CAATTAGTAT GTTTGACTGC
2451  AACAAGTTGA TAAGATGATG AAAAAGAAAA TGAGAAGATG ATGAAAATGC
2501  AAGCTGAAAC ATTTCACCGA AGGATAAGGC TTCACGGGTT GATATCTTGA
2551   AGGAGCAAGT TGCATTCTTA ATACAATGCA AATT
```

Figure 4.17. Nucleotide sequence of the insert from the white clover *Adh* genomic clone, pTrA110.

Figure 4.18. Nucleotide sequence of the insert from the white clover *Adh* genomic clone, pTrA131.

```
   1  GTCGACCTGC AGGTCAACGG ATCATGATGA TGATGATGAT TTTGTAGTGT
  51  GAGTTTCTCT CTCTCTCTCT ATGTCTTGTT CTACCACGTT CTTTGTTCTG
 101  TTAAACAACG ATTGGTATTG GTTTTGTGTT CAAAGCTTCA AAGCTGGTCG
 151  TTGATGTTCT GTTTGGAAGA GTACAAATTT TGGTCCCTTG TTATCGAATT
 201  AATTTTTATT TTTTTTCACC GCCGGTTTAA TCTAATTCGG AAATTAGCAA
 251  GTGGTTTCAG CTTTCTCTTA ATCATAATTA TGCTTATCAA ACTAATTTTT
 301  AATACATCAA AATTACAATT TCTAATAGTA TTTTTTATAG TCGAATGGCT
 351  AGAAATTTCA TCATTAAATG TGAATTAAAT TTATAATACT CTCTAATATT
 401  TAATTTAATA GCAAGTCATT GAAAAATTGG AAACACACTT TTTTTTCCTT
 451  TTAATATCTT AGTGAGTGAC AAAGCAATGG AGTAGTTAAA TTTTAATTAA
 501  AAATATAAGA GTTTTTTTTT TGTTGATTTG GGAAACAGAG GATTCTGTGT
 551  CTACTAAAAC AGAATTTGGC ATAAAGGAAA TTTGGCATAT ACTAATCCTG
 601  CCATAATATG AAATCAATTT TGCACCTATT AAGTGAAATG GTTGACAGTA
 651  AATTATGTAT GAGAAGAAGT TGCTAACAAC TGTCTTAATT TCCATTTATA
 701  AACCAACATT AACTAATCTA AAAAATGGAC ATTGATTATT TTTCTCAAGA
 751  TAAGGTAGAA TTATTTTAAA AAATTGAACA TTAAGAGCAT TTTTTATGAA
 801  AGTACCTAAG GAAATTTTAT AGGCTAATTA TTTTATCATT GAAACATTAT
 851  ATGACTCAAG TACGTTATTT AAGGTGTCAT ATTATCAATT AACAATTTAA
 901  TATCGACTTA TTTAGTGAAT CATATTAAAT AATATATATA AGTTACATAA
 951  TTTTGAGCAA AATATGTATG AGTTATAAAA ATATTTTATG ACATTTAGAA
1001  ACCACGTATT AATAATGTTA CAAATAAAAG ATGAAGAATC AAAATGTTAC
1051  AAAAAAAATA AGATAAAGAA CTCCTGATGC TATTTTGCCA AATTGTTTTT
1101  ACATCTTATG AAAATTATTT TAAGAAAAAA ATACTACAAA TTATTTTATT
1151  TTTTATTTTT TTTCTATAAA TGTGATTTTT TTTTTCCTTA TAATTTGGGA
1201  AGAAAGTATT TTTATAAACT ACATTGAAAA CTTATAAAAA TAAGTTGATA
1251  AAAACTTATA AAAATTGTCA TAAACTATTT CCTCCACAAA CAATTTCACA
1301  AGACAAAACT TATGTTAATC GATAAACTTA AATAAATTAA TCGAAACATA
1351  AACCTAAATA GTGAAAGGTA AAAAATCATA CCTTTTAGAA AAGAAAACAA
1401  CGTTTAAAGG ATGAAATTGA CCAAGACTCT GTTTGGTAAA CCAAATAAAT
1451  TAGCTTATAG CTTTTAGCTT ATAGCTAAAA GCTCTGTTTG GTAACAGTTT
1501  TTGAAAAAGA GCTTATAGCT TATTTTTACT AGCTTATAGC TTATTTTCCA
1551  AATGCTATTT CAAGTAGCTT ATAGCTTATT ATCTTTTTTT CCAATTTTAC
1601  CCCTATCATC TTACTTGAAA AGAATTAAAT ATTAATTAAA CATATCATTT
1651  TTATGTCATT TCATACTTAT AGTTAACTCA ACCGCTAATT TTACCAAACA
1701  CTACACATTC AATCAGCTAG CTTATTCGCT ATAAGCTAAA AGCTAGCTTA
1751  TAGCTATCCG CTATAAGCTA GCTTATCAGC TATCCGCTAT TTTTTACCAA
1801  ACAGATCCTG TATGCAAAGC ATTAACCGTA ACGTACACAA CAATATGCTT
1851  AAGTATAAGG AAACATGCAT GAAATATCTA ATTCCTTTTG GATGTTTTTG
1901  CACTGGTAAA CCAGTTTTGG ATATCGATGT TTTTGGAAAA TACCAAGATT
1951  AAGTAAACAA GGGGCTTAAA ATTTTACAAA ACCATGAATA TGTATATAAA
2001  ATCCAAAAAA GAAGTAAATG AATGTAAACT ATAGGGAAAT TGCATAGTTA
2051  AATAGAGAAA TATAATTGAG TAGAGGACAA CGTGAAGGGA AATGGTGGTC
2101  CAAAAAATAT TATATCTACG TCGCTGATCA AGAAGAATAA TTTATTTTTT
2151  TCTAGTTGGT GTTTACACCC TTTTGTCTTG TCTTCAAAGT ACTTAATTAT
2201  GTGTTTATGG ATTGAGTACA AGTACTCCCT TATAATATAT TTTTCTTATA
2251  AAAATAATAA TTATATATAT TATAATTAAA AAGATGTAAA TTACTAAATT
2301  AGTAATGTGC ATTTTTTTAA TTGGATTAAA ATATCCCTTT AACAATTTTA
2351  AATTAAAGGA TAATTTTGTA AAATTAAATT AATATTAAAA CATAAATCAT
2401  ACGAACATTT TTTAATTGAT ATTAATTTTT GTTTCCAATT TATTTTTTGA
2451  AGTTGTTTAT TTTTTATTTT TAAAATTGTC ATGTGTTTAT ATATATATAT
2501  ATATATATAT ATATATATGA TTGTTGGAGG TGATTGCAAT ATAAGTTCCA
```

Figure 4.18. (Continued)

```
2551  CATTGCCAAA CATAAAGTTT CACAAATTTT GATATTAAGG ACCACCTAAT
2601  ATTAACCTTC TGGTTTTCAA GAAAGGGGCT CTGGGTCAAA CTCCGCTATT
2651  TCCTCGTATA TTTGGTCATG AAGCTGGAGG GAATGTGGAG AGCGTAGGCG
2701  AGGGAGTGAC TCATCTGAAA CCAGTGGACA AGGCCCTGCC TGTATTCACA
2751  AGGGAGTGTG GGGAATGTCC ACATTGTAAG TCAGAGGAGA GTAATTAACA
2801  TGTGTAATCA TCTTAGGATC TGCAACAGTG CCAGAGCGAC CACGCGGCCC
2851  AGCCAGAGAA GTGCTTCATC TGTGCCACCA CGATGGTGTA AACAAACAAC
2901  CCCACCACAA TCGACGGCGC AGACAGCAGG ATGTCGTTAA TAAAGCGGGT
2951  GACTTCGGCC AGCCAGGATT TACGGCCATA TTCGGCCAGA TAGATGCCCG
3001  GATAATGCCC AGTGGCGTAC CGACGGCCGT TGACCAGAGA TCAGCAGCCC
3051  GCTGCCCGCC AGTTCGCCAG ACCACCGCCT GCCGTATTCG GTGGCGGGGT
3101  ATTTTCGGTG AACAGCGACA TCGTCATGCC ATCAATACCG CGTGTGACGG
3151  TTGAAAACAG GATCCATACC AGCCAGAACA AACCAAATGC CATGGTCAGT
3201  AACGACAGGG TCAGGGCAAT CCGGTTTTTC ATCCGGCGCC ACGCCTGCAT
3251  TTTGCGGCGT GAGGCTAACA GTTCGTCACG CGACTGTAAT TCCATCGTGG
3301  TCATGAACGT GCTCCTTCAC TTTTCGCCAG GCGCAGGATC ATCAGTTTCG
3351  AACAGGCCAG CACGATAAAG GTAATCACAA ACAGGATCAG CCCCAGCTCC
3401  ATCAGCGCAG CGTATGCACA CCCGACTCGG CTTCAGCAAA TTCGTTGGCC
3451  AGCGCAGAGG TAATACTGTT GCCCGGCATA AACAGCGATG CGCTGTCGAG
3501  CTGGTAGGTG TTACCGATAA TAAAGGTCAC CGCCATCGTT TCGCCCAGCG
3551  CACGCCCAGA CCCAGCATCA CGCCGCCAAT CACCCCATTT TTGGTGAATG
3601  GCAGTACCAC ACGCCACATC ACTTCCCAGG TGGTACAGCC GATACCGTAA
3651  GCGGACTCCT TCATCATGAC CGGCGTCTGC TCGAATACGT CACGCATCAC
3701  TGAGGCGATG TAAGGGATAA TCATAATCGC GAGGATAATC CCGGCGGCCA
3751  GAATTC
```

Figure 4.19. Nucleotide sequence of the insert from the white clover *Adh* genomic clone, pTrA230.

```
   1  GGATCCTGAG ATGATCAGGC CAGGTCACCG GCCTGTGAAA AGATGTTAGT
  51  CAGCGCGCAC GTCCACTAAC ACCGGACAGC TGGCACGTTC AATCACTGCC
 101  GCACTGCACG AACCTTTCAG CAGGCGGTTA AAAGAGGAGA GATGGCGACG
 151  CCCATGATAA TCATCGCCGC CTGAAGGCTG TTGATTCGGC CACGATGGTT
 201  TCTGCCGATT CCCCGGCGAC CACTCTGCCC CGCGCCACCA CGCCGGCCCG
 251  GTGCAGTTCC GCCAGCGCCT GGCGTACCAC CGCTTCGGCA GTGTTTTGCT
 301  CGTCCAGCGC CAGGCCAAAA TCACCAGGGT CTTCACCCGC CTCAATATCA
 351  AACGGGCCCG AGTCTGCGTA GTGGCATCCA CACAGCACAC CACCACGACT
 401  TCAGCCTGTA AGCGCAGCGC CTGTTCGCTC GCCAGTGAAA CAACTTTGCG
 451  GGCGATGGCT GAATTATCTA CAGCCACTAA AAGCGTTTTC ATCGTCTTCT
 501  CCTGTCAGTC CCGTAAGCAG CTTTACCAAT TTCATCGATT AATGCCTCTT
 551  TACACCGGAG ATCGCCTCCC GGTTACGCGC CTCATGCGCT GAAAGCGGGC
 601  GCCAGGAATT AACAAGACTG ATGGAAAAGC GTCCCAGAAT AGTGTCGATC
 651  GCCACCGCCA TGGTGGAGAT CCCTTCCAGT GTTTCTCCCC GGTCATAAGC
 701  CAGACCGGTA CGGCGAATTT CAGTCAGGTC GCCAGCAGCG CAGGGAAGTC
 751  GCGCACGGTC AGTTCAGTGA TGGCCTGAAA CGCCTCCCCG ACCCGCTCTC
 801  TGGCTGTCTC ATCGCTGTCC ACGCCAGCAG CGCGCGGCCA CCGGACGTGC
 851  TCTAAAGCGG CAGGTTCATG CCGATGCGCG GCACAACCCG CAGTTCCCGG
 901  TCTGCCACCA CATAGTGAAC AATGGCCAGC TGCAGGCCGC TGGGCGCGCT
 951  AGGGAAACTG TTTCTCCCGT CTGGTCGCAC AGTGCCTGAA GCCAGGGCTG
1001  GGCGATCGCC ACCACGTCAG TATGAACATT GGCGATCAGT CGCAGCAGAG
1051  CTGGCCCGAG GCGCACGCTG CTGCCGCACC GTGGCTGCGT ACCAGTTGAA
1101  CCGAGTCAAG CGCACGCAAT TCGCTGTACG GTGGAACGCG GCAGATCGAC
1151  CGCCTGGGCA ATTTCGCCCA GGTCATCCCG TCCGGTTTAG CCCCCAGCGC
1201  GTCAAGAATT TTCGCCGCGC GGGTGATCAC CTGAATGCCG CCGGAACGCT
1251  CGTCATCGCG GCACGCTGAA GACTCAAACA TAGGGATTCT TTTCCTTGTT
1301  ACCGGCCATT ACTGTAGCAC TGAATGTCTC ACTGGATACA GTGTACCGCA
1351  TTGCAATACA TCGCATCACC CTGTAATATT CCGCACAATG TATCACGATG
1401  CAATACAGTG CATCGATAAT AAAAATAGCG TTGGAGTCCC TGCAATGAAT
1451  GGCCAGCCTG CGGCAATGCC GCTCCGCACC CTTTCACCCT GCGCCTGGCG
1501  TGGGTCTGGT GGGTGTGTTG ATTGCCGCAC TCAGCTCCGG CCTCAACGAT
1551  CGGGTGACGG ACCTCGCCCT GGCCGATATC CGTGCCGCGC TGGGCATCGA
1601  TTTTGATACG GGCAGTTGGA TCACCTCCGC CTATCAGGCA GCGGAAGTGG
1651  CTGCCATGAT GGTGGCCCCT GGTTTGCGGT CACTTTTTCC CTGCGTCGTT
1701  TTACGCTGGC TGCCACGCTG GGTTTCTGCC TGATGGCGGC GATCCCTGTC
1751  TTAAAATATT AATATATAAC TAGGTTGAGT TCATGCTTGT AGGTCTTGGT
1801  GCTACTGTCA ATGTTGCAAA ACCGAAACCC GGTTCATCCG TTTCTATCTT
1851  TGGACTCAGA GTTGTTGGCC TTGCTGTATA TTATTCTCAT TCGCTTTAAT
1901  ACAAATTTCT GGTTGCTGAA GGGACAAGGA TATCTGGTGC ATCGCGAATC
1951  ATTGGAGTTG ATTTAGTTTC CAGCCGATTT GAATTAGGTA CAATGCTCTT
2001  TTATTTCATC AAATGCTTTA AGATTGAGGT CAACATAAAT TATATGACTA
2051  AATTTTTGTA CTGTTCTTGT TCTCAGCTAA GAAGTTTGGT GTGATTGAGT
2101  TTGTGAACCC AAAAGATCAT GACAAACCTG TACAACAGGT TAGTTTAGCT
2151  CAGCAATAAT TCTTATTTAT TTATTTATTT ATTTATTAAT TACATTTTTG
2201  TTATCTCTAA ACTTATTACA TTTTCAACTT CCTTTGATTG AAATTTGAAA
2251  CCTTATTGAT ATTTGCTTTC CATGTTGCCT ATGGTATCAT CTGCTGTCCT
2301  TTGCTTTTAA GGTATATGTG CATCATTAAG ATGCTAAGAT TTAATTTGTA
2351  ATGATACTAG GTATTAATTT TACAATTAAA TTGATATTCT GTAACCCCTT
2401  TTTAACCTCT ATGTGGTAGT AATTGAAGGT ATATGTGGTA GTAATGATAC
2451  TAGGTATATG TGGTAGTAAT TGATATTCTT TAAGGTATTC ATTTGCAATA
2501  CTTGTTTGTT ATTGTCTAAT AGCCTCTCTA TTCTTTAAGG TAGTATCGAG
```

Figure 4.19. (Continued)

```
2551  TTGTGTTAAA TGTGAGTGTG TGTATGATAG TAAGATCAGT AGATAAAATT
2601  TTAATTAGGT ATTAAATGCA TTATAGGAAC TTCCTAGCAC TATTTAGTTT
2651  AGGATTGCCT TGTTTTTTTA ATGTATTTAT TCGATGTCAT AATCACAAAA
2701  CAAGACCAAC CAGGGGCAGG CACCTATGGG TTGGGAAGAA GAAATTTAAC
2751  AAGACTTTTA TTCAGCTTAG TGCACTAGGT TCTCTATTTC TAGAGTGAAC
2801  CAATTATACA AAGACATAAT TGAATTGTAT ATACNCTATA TATATATATA
2851  TGCAGGACCA AACTAAATCT CAGTGCTAGT TCTACTTCAT GGGTTGCAAT
2901  GATTTTCCAG CACCCATTGT TTCAGAAGCA AATTCTAGTG GCTGGATTCG
2951  ATAACACATT TAATGATTTA GTAACACAAA AATTAGATGA GCATTGAAAT
3001  TAGATGATTT ACTTTTCTAA TTAATTATAT ACCAATCAAG ATTGTTTTAA
3051  TATGTTTTTT GAGATGCTAT TGTGTTTAGG TCCTTGTAAT TTGGAATATA
3101  TCTATAGTGC TTAGAAAAAA GATGTCATTA CGTTATGTTT AAATGAACAT
3151  ATTATTTTGT GGAATTATAT GCCATATGGA CTCACTGTCT TCTTTAGGAA
3201  TGGTCGTTGG AACTTCTCTT GTGTATGGCT AGATAATTGT TCTCGTTATT
3251  TAAGTGTCTT TTTTTGGAGG CATTTAAGTA TCTAATCGAT ACTTATATTA
3301   TTTAGGAACT GCATTAGAAG AGTCAATAAA TGAAA
```

Figure 4.20. Sequence comparison of selected regions of each of the inserts from the white clover Adh cDNA clone, pTrA45, and genomic clone, pTrA110. The numbering of the pTrA45 sequence corresponds with the numbering in Figure 3.20. The beginning of the pTrA110 sequence corresponds with position 1162 in Figure 4.17.

```
                310        320        330        340        350
                  *          *          *          *          *
pTrA45    TGGAGGGATTGTGGAGAGCGTAGGCGAAGGTGTGACTCATCTGAAACCAG
               :::::::::::::: ::::::: ::::::::::::::::::::::::
pTrA110   -----GGATTGTGGAGAGCATAGGCGAGGGTGTGACTCATCTGAAACCAG

                360        370        380        390        400
                  *          *          *          *          *
pTrA45    GAGACCATGCCCTGCCTGTATTCACAGGCGAGTGTGGGGAATGCCCACAT
          : :: :: :::::::: :::::::::: :::::::::::::: ::::::
pTrA110   GGGA-CACGCCCTGCCAGTATTCACAGGGGAGTGTGGGGAATGTCCACAT

                410        420        430        440
                  *          *          *          *
pTrA45    TGT-AAGTCAGAGGAGAGTAACATGTGTAATCTTCTTAGGATTAACACCG
          ::: ::::::::::::::::: ::::::: :::::::::: :: ::::: :
pTrA110   TGTTAAGTCAGAGGAGAGTAGCATGTGTGATCTTCTTAGAATCAACACAG

       450        460        470        480        490
         *          *          *          *          *
pTrA45    ACAGAGGTGTCATGATCAATGACAACAAGTCAAGATTCTCTATTAAGGGA
          :::::::::::::::::::::::::: :::: ::::::::::::::::::
pTrA110   ACAGAGGTGTCATGATCAATGACAACCAGTCTAGATTCTCTATTAAGGGA

       500        510        520        530        540
         *          *          *          *          *
pTrA45    CAACCTGTACACCATTTTGTCGGTACCTCTACATTCAGCGAGTACACTGT
          ::::  :::::::::::::::::::: :::::::::::: ::::::::::
pTrA110   CAACTCGTACACCATTTTGTCGGTACTTCTACATTCAGCAAGTACACTGT

       550        560        570        580        590
         *          *          *          *          *
pTrA45    CGTTCATGCGGGATGTGTTGCAAAGATCAACCCTGATGCACCACTTGACA
           :: ::::: :: ::::::::::::::::::::::::::::::::::::::
pTrA110   GGTCCATGCTGGTTGTGTTGCAAAGATCAACCCTGATGCACCACTTGACA

       600        610        620        630
         *          *          *          *
pTrA45    AAGTTTGTATTCTCAGCTGTGGAATATGCACAG----------------
          ::::::::::::::::: :  ::: :::::::::
pTrA110   AAGTTTGTATTCTCAGTTACGGATTATGCACAGGTATGAGCATGTTTAGA


pTrA45    ------------------------------------------------

pTrA110   TTGATTTATTTGAGTTTTATTGTTAGCATACATATTTGTGAGACTGTTTG
```

Figure 4.20. (Continued)

```
pTrA45   --------------------------------------------------
pTrA110  ATAGAATTTATGAAACAACTTATGACATGTTTTTTGAGCTTATTATTATA

pTrA45   --------------------------------------------------
pTrA110  AGCTTTCTTAAATAGCTTATGAAAATAGTTTATATCTTACATAAAAACAA

pTrA45   --------------------------------------------------
pTrA110  TTTGACTTTATTTTATCCTTAGTGATAAAAATAGCTTATACATAAAGACT

pTrA45   --------------------------------------------------
pTrA110  TATATGATAAGGGTTTATTCCAAACATTTTAAAAGATCCCTATCTTAAAA

                                                      640
                                                       *
pTrA45   ------------------------------------GTCTTGGTGCTACT
                                              ::::::::::::::
pTrA110  TATTAATATGTAACTAGGTTGAGTTCATGCTTGTAGGTCTTGGTGCTACT

          650        660        670        680        690
           *          *          *          *          *
pTrA45   GTCAATGTTGCAAAACCGAAACCCGGTTCTTCTGTTGCAATCTTTGGACT
         :::::::::::::::::::::::::::::  :: : : : ::::::::::
pTrA110  GTCAATGTTGCAAAACCGAAACCCGGTTTATCCGCTTCTATCTTTGGACT

          700        710
           *          *
pTrA45   TGGAGCTGTTGGCCTTGCT----------------------------
           :::: : ::::::::::
pTrA110  AAGAGCCGCTGGCCTTGCTGTATGTTATTCTCATTCGCTTTAATACAAAT

                 720        730        740        750        760
                  *          *          *          *          *
pTrA45   -----GCTGCTGAAGGGGCAAGGATGTCTGGTGCATCTCGAATCATTGGA
               :::::::::::::::::::: :::::::::: : ::::::::::
pTrA110  TTCAGGCTGCTGAAGGGGCAAGGATATCTGGTGCATTGCAAATCATTGGA

                 770        780        790                  800
                  *          *          *                    *
pTrA45   GTTGATTTAGTTTCTAGCCGATTTGAATTAGCTAAGAAG--------TTT
         :::::::::::::: ::::::::::::::::: ::  : :        :::
pTrA110  GTTGATTTAGTTTCCAGCCGATTTGAATTAGGTACAATGGCACTCAATTT
```

Figure 4.21. Sequence comparison of selected regions of each of the inserts from the white clover Adh cDNA clone, pTrA45, and genomic clone, pTrA131. The numbering of the pTrA45 sequence corresponds with the numbering in Figure 3.20. The beginning of the pTrA131 sequence corresponds with position 2615 in Figure 4.18.

```
                160        170        180        190        200
                  *          *          *          *          *
pTrA45    ACTGGTGATTGAAGAAGTAGAGGTGGCGCCACCACAGGCCGGTGAAGTCC

pTrA131   ----------------------------------------------------

                210        220        230        240        250
                  *          *          *          *          *
pTrA45    GTCTTAAGATACTCTTCACCTCCCTTTGCCACACTGATGTTTACTTCTGG
           : : :::: :
pTrA131   -TTTCAAGAAA-----------------------------------------

                260        270        280        290        300
                  *          *          *          *          *
pTrA45    GAAGCTAAGGGTCAGACTCCATTGTTTCCTCGTATATTTGGTCATGAAGC
          :  :::  :::::: :::::  : ::::::::::::::::::::::::::
pTrA131   GGGGCTCTGGGTCAAACTCCGCTATTTCCTCGTATATTTGGTCATGAAGC

                310        320        330        340        350
                  *          *          *          *          *
pTrA45    TGGAGGGATTGTGGAGAGCGTAGGCGAAGGTGTGACTCATCTGAAACCAG
          :::::::: :::::::::::::::::: :: ::::::::::::::::::
pTrA131   TGGAGGGAATGTGGAGAGCGTAGGCGAGGGAGTGACTCATCTGAAACCAG

                360        370        380        390        400
                  *          *          *          *          *
pTrA45    GAGACCATGCCCTGCCTGTATTCACAGGCGAGTGTGGGGAATGCCCACAT
            ::: : :::::::::::::::::: : :::::::::::::: ::::::
pTrA131   TGGACAAGGCCCTGCCTGTATTCACAAGGGAGTGTGGGGAATGTCCACAT

                410        420            430        440
                  *          *              *          *
pTrA45    TGTAAGTCAGAGGAGAGTAA----CATGTGTAATCTTCTTAGGATTAACA
          ::::::::::::::::::::    :::::::::: :::::::::
pTrA131   TGTAAGTCAGAGGAGAGTAATTAACATGTGTAATCATCTTAGGAT-----

           450        460        470        480        490
             *          *          *          *          *
pTrA45    CCGACAGAGGTGTCATGATCAATGACAACAAGTCAAGATTCTCTATTAAG

pTrA131   ----------------------------------------------------

           500        510        520        530        540
             *          *          *          *          *
pTrA45    GGACAACCTGTACACCATTTTGTCGGTACCTCTACATTCAGCGAGTACAC

pTrA131   ----------------------------------------------------
```

Figure 4.22. Sequence comparison of selected regions of each of the inserts from the white clover Adh cDNA clone, pTrA45, and genomic clone, pTrA230. The numbering of the pTrA45 sequence corresponds with the numbering in Figure 3.20. The beginning of the pTrA230 sequence corresponds with position 1779 in Figure 4.19.

```
                510        520        530        540        550
                  *          *          *          *          *
pTrA45    AACCTGTACACCATTTTGTCGGTACCTCTACATTCAGCGAGTACACTGTC

pTrA230   -------------------------------------------------

                560        570        580        590        600
                  *          *          *          *          *
pTrA45    GTTCATGCGGGATGTGTTGCAAAGATCAACCCTGATGCACCACTTGACAA
          ::::::::
pTrA230   GTTCATGC---------------------------------------

                610        620        630        640        650
                  *          *          *          *          *
pTrA45    AGTTTGTATTCTCAGCTGTGGAATATGCACAGGTCTTGGTGCTACTGTCA
             ::::                        ::::::::::::::::::::
pTrA230   ---TTGT----------------------AGGTCTTGGTGCTACTGTCA

                660        670        680        690        700
                  *          *          *          *          *
pTrA45    ATGTTGCAAAACCGAAACCCGGTTCTTCTGTTGCAATCTTTGGACTTGGA
          :::::::::::::::::::::::::: :: ::: : ::::::::::::  ::
pTrA230   ATGTTGCAAAACCGAAACCCGGTTCATCCGTTTCTATCTTTGGACTCAGA

                710
                  *
pTrA45    GCTGTTGGCCTTGCT-----------------------------------
          : ::::::::::::
pTrA230   GTTGTTGGCCTTGCTGTATATTATTCTCATTCGCTTTAATACAAATTTCT

             720        730        740        750        760
               *          *          *          *          *
pTrA45    -GCTGCTGAAGGGGCAAGGATGTCTGGTGCATCTCGAATCATTGGAGTTG
           : :::::::::: ::::::: :::::::::::: ::::::::::::::::
pTrA230   GGTTGCTGAAGGGACAAGGATATCTGGTGCATCGCGAATCATTGGAGTTG

             770        780        790        800        810
               *          *          *          *          *
pTrA45    ATTTAGTTTCTAGCCGATTTGAATTAGCTAAGAAGTTTGGGGTAAATGAG
          :::::::::: ::::::::::::::::: ::  :       : :   : :
pTrA230   ATTTAGTTTCCAGCCGATTTGAATTAGGTACAA------TGCTCTTTTAT

             820         830        840        850        860
               *          *          *          *          *
pTrA45    TTCGTCAAC--CCGAAAGATCATGACAAACCTGTTCAACAGGTAATTGCT
          ::: ::::   :   :::::  :::       : :::::     ::   :
pTrA230   TTCATCAAATGCTTTAAGAT--TGA-------GGTCAAC-----ATAAAT
```

Figure 4.23. Comparison of the sequences of exons 3, 4, 5, 6 and 7 of the pea *Adh1* gene (Llewellyn *et al.*, 1987; here referred to as PsADH1) with homologous regions of the three white clover Adh genomic clones pTrA131 (TrA131), pTrA110 (TrA110) and pTrA230 (TrA230). Only sequence differences are shown; (.) indicates sequence identity with the pea *Adh1* sequence. This Figure was constructed with a text editor from independent alignments (Align) of each white clover sequence with the pea sequence.

```
                1110      1120      1130      1140      1150
                   *         *         *         *         *
PsADH1   GCAATAGCTTAAAATGGTTTTCTTTTTTCCCCTTGTTTCAGGGTCAGACT
TrA131   ----------------------------------------.....A...
TrA110   --------------------------------------------------
TrA230   --------------------------------------------------

                1160      1170      1180      1190      1200
                   *         *         *         *         *
PsADH1   CCATTGTTTCCTCGTATATTTGGTCATGAAGCTGGAGGGTATGCTTCTCT
TrA131   ..GC.A.................................-----------
TrA110   --------------------------------------------------
TrA230   --------------------------------------------------

                1210      1220      1230      1240      1250
                   *         *         *         *         *
PsADH1   TCTCAACACTGAAAGTTAAATCACACATATTATATACTTTGTCTTATATT
TrA131   --------------------------------------------------
TrA110   --------------------------------------------------
TrA230   --------------------------------------------------

                1260      1270      1280      1290      1300
                   *         *         *         *         *
PsADH1   ATTATTGATTGAAGCATTTGATTATCTTAGTAATTTAGTATGTAATGTGA
TrA131   --------------------------------------------------
TrA110   --------------------------------------------------
TrA230   --------------------------------------------------

                1310      1320      1330      1340      1350
                   *         *         *         *         *
PsADH1   ATGAATGTTATTGATAATTGTTGGGATAATGTTGATCAGGATTGTGGAGA
TrA131   ----------------------------------------..A.......
TrA110   ----------------------------------------..........
TrA230   --------------------------------------------------

                1360      1370      1380      1390      1400
                   *         *         *         *         *
PsADH1   GTGTAGGTGAGGGTGTGACACATCTTAAACCAGGGGATCATGCTCTGCCG
TrA131   .C.....C.....A.....T.....G.......T...CA.G..C.....T
TrA110   .CA....C..........T.....G..........-..C..C.....A
TrA230   --------------------------------------------------
```

Figure 4.23. (Continued)

```
                1410        1420        1430        1440
                  *           *           *           *
PsADH1   GTATTCACGGGGGAATGTGGGGAATGTCCACATTGT-AAGTCAGAGGAAA
TrA131   ........AA....G.....................-..........G.
TrA110   ........A.....G.....................T..........G.
TrA230   ----------------------------------------------

       1450          1460        1470        1480        1490
         *             *           *           *           *
PsADH1   GCAA----CATGTGTGATCTTCTTAGGATCAACACAGATAGAGGTGTAAT
TrA131   .T..TTAA.......A...A..........TG..ACAG.GCCA.A.CG.C
TrA110   .T.G----.................A...........C........C..
TrA230   ----------------------------------------------

           1500        1510        1520        1530        1540
             *           *           *           *           *
PsADH1   GCTCAATGACAACAAGTCCAGATTCTCTATTAAGGGACAACCTGTACACC
TrA131   CACGCGGCC..G.C..AGA..TGCT..ATC.GT.CC.CC..GATGGTGTA
TrA110   .A...........C....T.....................TC.......
TrA230   ----------------------------------------------

           1550        1560        1570        1580        1590
             *           *           *           *           *
PsADH1   ATTTCGTCGGCACCTCCACGTTCAGCGAATACACTGTGGTTCATGCTGGT
TrA131   .ACAAACAAC.CC.C.ACAA.CG.CG.CGC.G..A.CA.GATG.CG.TAA
TrA110   ....T.....T..T..T..A......A.G...........C.........
TrA230   ----------------------------------------------

           1600        1610        1620        1630        1640
             *           *           *           *           *
PsADH1   TGTGTTGCAAAGATCAATGCTGATGCACCGCTTGACAAAGTTTGTATTCT
TrA131   .AAAGC.GGTGAC.TCGGC.A.CCAGGATTTAC.G.C.TA..C.GCCAGA
TrA110   .................CC..........A....................
TrA230   ----------------------------------------------

           1650        1660        1670        1680        1690
             *           *           *           *           *
PsADH1   CAGCTGTGGAATATGCACTGGTATATATATACAACATCTCTAACTTCAGA
TrA131   T..A..CCCGGATAATG.CC----------------------------
TrA110   ...T.AC...T.......AG----------------------------
TrA230   ----------------------------------------------

           1700        1710        1720        1730        1740
             *           *           *           *           *
PsADH1   CATTAATATATAACTAGGCTGAATCTATGTTTTCCTAATCCAATTTGCAT
TrA131   ----------------------------------------------
TrA110   ----------------------------------------------
TrA230   ----------------------------------------------
```

Figure 4.23. (Continued)

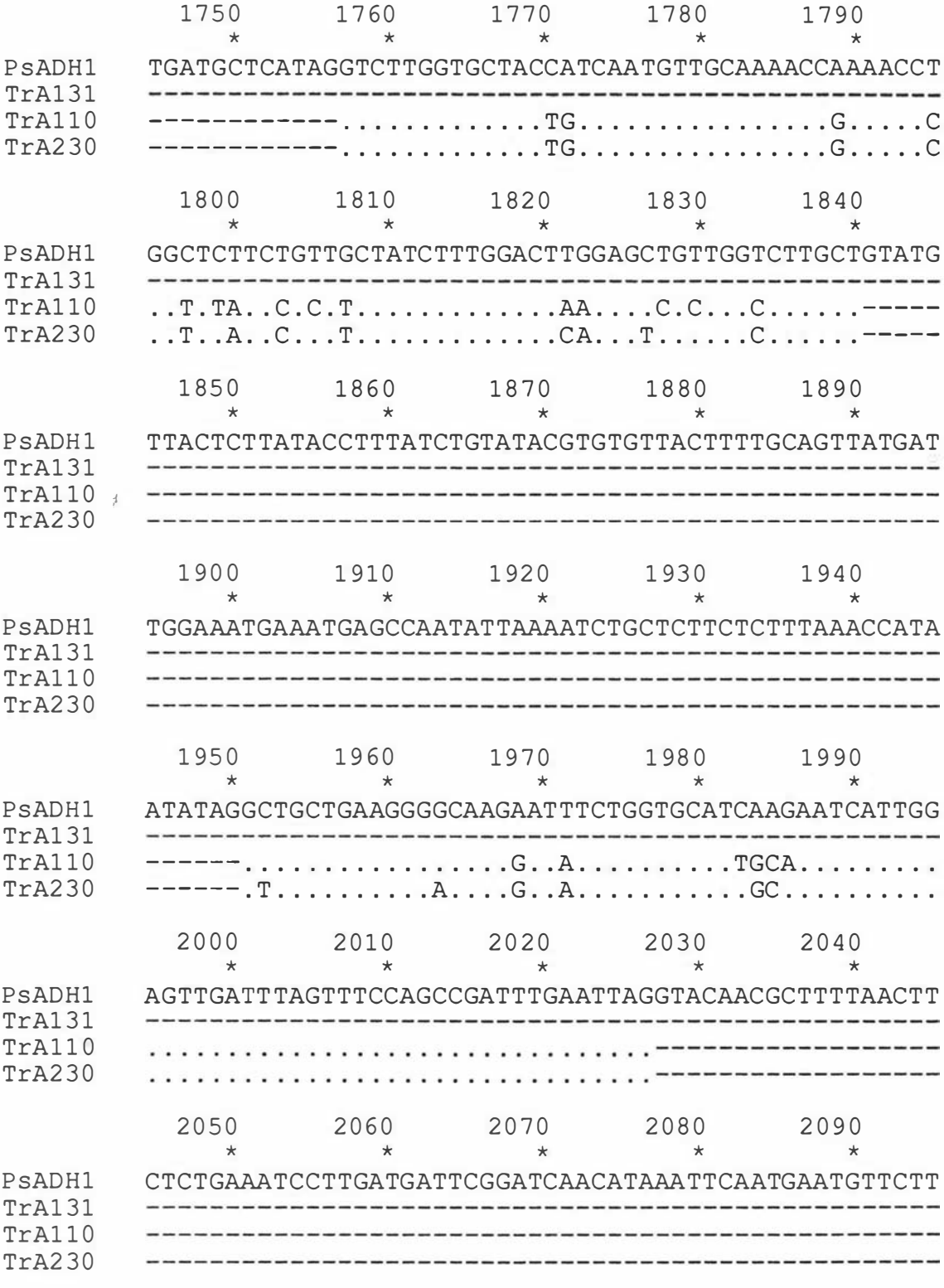


Figure 4.23. (Continued)

```
            2100       2110       2120       2130       2140
               *          *          *          *          *
PsADH1   TTGTTGACGGCAATTCCTATTCTTTTCATGTTGGCAGCTAAGAAGTTTGG
TrA131   ----------------------------------------------------
TrA110   ----------------------------------------------------
TrA230   ----------------------------------.....------------

            2150       2160       2170       2180       2190
               *          *          *          *          *
PsADH1   GGTAAATGAGTTCGTAAACCCAAAAGAGCACGACAAACCTGTGCAACAGG
TrA131   ----------------------------------------------------
TrA110   ----------------------------------------------------
TrA230   T..G.T......T..G...........T..T...........A......-

            2200       2210       2220       2230       2240
               *          *          *          *          *
PsADH1   TTAGTTCATTGATAATTTATTATATTTCTCTTTTCTATTATAATTTTTGA
TrA131   ----------------------------------------------------
TrA110   ----------------------------------------------------
TrA230   ----------------------------------------------------
```

Homology searches of the Genbank and EMBL nucleotide sequence databases (Genbank release number 63.0; EMBL release number 23.0) were carried out with each of the pTrA131, pTrA110 and pTrA230 sequences using the FASTA program (Pearson and Lipman, 1988) and the complete sequence determined for each insert. Alcohol dehydrogenase sequences accounted for the top 21 scores in the pTrA110 search, the top 16 scores in the pTrA230 search and the top 2 scores in the pTrA131 search (data not shown). In none of the searches were sequences other than Adh sequences found which exhibited significant identity to the sequences of the white clover clones.

Clearly, the three putative white clover Adh clones isolated from the genomic library do not represent functional *Adh* genes. Furthermore, based on the lack of open reading frames in the region of each clone sequenced as well as on the lack of sequence homology with sequences other than Adh sequences in the nucleotide sequence databases, it must be concluded that none of these clones represents a functional white clover gene of any type and that the regions sequenced represent white clover *Adh* pseudogenes.

A search of the the Genbank nucleotide sequence database (Genbank release number 64.0) for plant pseudogenes using the Information Retrieval Experimental Workbench (IRX) program (Genbank Online Services) located 42 entries. Plant pseudogenes that have been reported include genes that have become inactivated through a single base pair deletion or addition leading to a frame-shift mutation and hence the inability to code for a full-length protein (Pikaard *et al.*, 1986; Slightom *et al.*, 1985; Voelker *et al.*, 1986). A mutation has also been reported to occur in an upstream sequence important for transcription of the gene (Wiborg *et al.*, 1983) and in the coding region of the gene to produce a premature stop codon (O'Neal *et al.*, 1987). Other plant pseudogenes include genes that have become inactivated through large deletions (Heim *et al.*, 1989; Lee and Verma, 1984; Walling *et al.*,

1988). Pseudogenes lacking introns can result from integration into the genome of reverse transcripts of cytoplasmic poly(A)$^+$ messenger RNA of genes that are expressed in the germ line (Drouin and Dover, 1987).

Of the classes of plant pseudogenes that have been reported, the three white clover *Adh* pseudogenes appear to belong to the class resulting from large deletions. It is apparent that, for each of the white clover *Adh* sequences, a large deletion has occurred in both the 5′ and the 3′ regions of the gene and these deletions have caused the inactivation of the respective *Adh* gene. The precise relationship between these *Adh* sequences and the sequence of the functional white clover *Adh* gene must await the isolation and characterization of the functional gene.

Chapter 5

General Discussion

This study is the first to examine gene structure and expression in white clover. The study has involved the development of a number of techniques necessary for the analysis of gene expression, the construction of cDNA and genomic libraries, isolation of recombinant clones and the characterization of isolates by DNA sequence analysis.

A series of white clover leaf/stem- and root-specific cDNA libraries has been constructed in the lambda vector λgt10. One Rubisco SSU, five Adh and four lectin cDNA clones were successfully isolated from these cDNA libraries using heterologous plant probes after first establishing that these probes hybridized to white clover mRNA transcripts in Northern blot analyses. Inserts from each of the positive λgt10 isolates were subcloned into plasmids for restriction enzyme mapping and DNA sequencing.

The SSU cDNA clone, pTrS20, subcloned from an isolate from one of the leaf/stem cDNA libraries, represents approx. 50% of the full length white clover SSU transcript, based on a transcript size of approx. 800 bp as determined by Northern blot analysis (Section 3.2). As discussed above (Section 3.5), the isolation of a single SSU clone from one of the leaf/stem libraries was unexpected. A number of factors could account for the low representation of SSU clones in the leaf/stem library including instability of the transcript, low affinity of SSU poly(A)$^+$ transcripts for oligo(dT) cellulose, inefficient first and/or second strand cDNA synthesis for this particular leaf/stem library, inefficient first and/or second strand cDNA synthesis from leaf/stem-

derived poly(A)$^+$ RNA in general, or poor amplification of SSU cDNA clones during library amplification. Insufficient evidence is available to clearly distinguish between these possibilities but some conclusions can be drawn. Since two putative lectin clones were isolated from a different, unamplified, leaf/stem cDNA library (Section 3.7), it would appear that the poor representation of SSU clones is associated with either the cDNA synthesis reactions themselves or poor amplification of SSU clones within the particular cDNA library that was screened with the petunia SSU cDNA clone. This is based on the assumption that SSU transcripts are more abundant than lectin transcripts in white clover leaf/stem material. The small insert sizes of the SSU cDNA and the random cDNA clones isolated from the leaf/stem number 1 cDNA library (Section 3.4) suggests that the problem of poor representation is associated with the efficiencies of the first or second strand cDNA synthesis reactions for this particular library rather than differential growth of recombinants during library amplification. This could be confirmed by screening one of the other leaf-/stem cDNA libraries with a SSU probe. Of the remaining possibilities, it has been suggested that subtle conformational differences between the RNAs derived from two different soybean *rbcS* genes leading to differential retention on oligo(dT) cellulose could account for the low frequency of one class of cDNA clone relative to the other (Shirley *et al.*, 1990).

In the case of the white clover root cDNA libraries probed with a pea Adh cDNA, ten identical Adh cDNA isolates were obtained after screening 10,000 recombinants from one amplified library (i.e. one positive per 1000 recombinants screened) and five independent Adh cDNA isolates were obtained after screening 5,000 recombinants from a second, amplified library. Taken together, these results indicate that Adh cDNA clones represent 0.1% of the total root cDNA clones. The actual level of expression of Adh in white clover

has not been determined but the value for the proportion of Adh clones in the root cDNA libraries compares favourably with the 0.07% of Adh1 clones in pea anaerobic root cDNA libraries (Llewellyn *et al.*, 1987). In pea the *Adh1* gene is induced by anaerobic conditions, *Adh1* transcripts being barely detectable in aerobic roots (Llewellyn *et al.*, 1987). However for this white clover study, no attempt was made to induce *Adh* gene expression by subjecting the white clover roots to anaerobic conditions. It is therefore concluded that either the white clover Adh cDNA clones represent a gene that is expressed constitutively in the roots of white clover or the watering regime during the growth of the white clover plants (Section 2.1.2) was such that anaerobiosis was inadvertantly induced. There is no evidence to support either of these alternatives but it is possible that the functional white clover *Adh* gene corresponding to the cDNA isolate is expressed under aerobic conditions in the root and does not require anaerobic induction, as is the case in particular organs of some other plants (see Section 1.8.2). Clearly, the regulation of *Adh* gene expression in the roots of white clover, whether it is constitutive or anaerobically induced, requires further investigation.

Amongst the five Adh cDNA clones isolated from the white clover root cDNA libraries there is considerable variation in both their 3′ and 5′ termini (Figure 3.11). From the sequence data available for the subclones from clones λTrA4 and λTrA5 (Section 3.13), it is apparent that only the λTrA5 insert includes sequence corresponding to the poly(A) tail of the mRNA and that only the λTrA4 insert includes sequence corresponding to the 5′ region of the mRNA. However λTrA4 does represent an almost full length *Adh* gene transcript and includes the entire coding region corresponding to the gene. Since the first strand cDNA synthesis is initiated from oligo (dT)-primed mRNA (Section 2.4), it is likely that the variability at the 5′ termini represents either incomplete first strand synthesis by reverse transcriptase or incomplete

second strand synthesis from the mRNA-cDNA hybrid, whereas the variability at the 3′ termini represents incomplete second strand synthesis from the mRNA-cDNA hybrid.

The problem of incomplete and variable-sized inserts is common to all of the white clover SSU and Adh cDNA clones that have been isolated and characterized during the course of this study (see above). It is probable that this problem is associated in some way with the apparent self-priming from hairpin loop structures during the second strand cDNA synthesis reaction (Section 3.3), although the mechanism whereby variable-sized cDNAs are produced remains unclear. In the method used in this study, the second strand cDNA synthesis from the mRNA-cDNA hybrid is catalysed by the combined activities of *E. coli* RNase H and DNA polymerase I by RNA-primed nick translation (Section 2.4). For this method to be successful, second-strand synthesis by DNA polymerase I must prime from the 5′ (capped) end of the mRNA. Failure to prime from this oligonucleotide can lead to hairpin-primed synthesis of second strand cDNA (D'Alessio and Gerard, 1988), which could give rise to variable-sized second strand reaction products.

In addition to the white clover leaf/stem- and root-specific cDNA libraries, a fully representative white clover genomic library has been constructed in the lambda vector λEMBL3. The white clover SSU and Adh cDNA clones were used to isolate one SSU and three Adh genomic clones from this library. Restriction enzyme fragments from each of these clones, identified by Southern blot analysis to contain sequences homologous to the cDNA clones, were subcloned and sequenced.

A screening of the white clover genomic library with the white clover SSU cDNA resulted in the isolation of one 15 kb SSU genomic clone. The sequence of a 4 kb fragment subcloned from this isolate (pTrS640) consists of a 1259 bp white clover *rbcS* gene coding region, 1944 bp of sequence upstream

from the translation initiation codon and 771 bp of sequence downstream from the translation termination codon. The coding region of the white clover *rbcS* gene includes two introns.

In petunia, pea and tomato, most of the *rbcS* genes of the multigene family of each plant have been sequenced and the organization and/or the chromosomal localization has been determined (Dean *et al.*, 1985; Polans *et al.*, 1985; Vellegos *et al.*, 1986; Sugita *et al.*, 1987) (Section 1.7.3). In tomato, the five *rbcS* genes are encoded at three separate genetic loci located on chromosome 2 and on chromosome 3 (Sugita *et al.*, 1987). The eight *rbcS* genes of petunia are also encoded at three genetic loci (Dean *et al.*, 1985b). In contrast, the five member pea *rbcS* gene family maps to a single locus on chromosome 5 (Polans *et al.*, 1985). According to the Southern blot analysis of white clover DNA (Section 4.3), white clover appears to have at least two *rbcS* genes which are encoded at two genetic loci. However, in a white clover *in situ* hybridization experiment using the white clover SSU cDNA probe, hybridization was detected on chromosomes 6, 11 and 12 (J.M. Zhu, personal communication), suggesting three *rbcS* genetic loci for white clover. The reasons for this discrepency remain unclear but could reflect differences in the hybridization stringencies used for the filter hybridization and the *in situ* hybridization experiments.

The three Adh clones isolated from the white clover genomic library (λTrA11, λTrA12 and λTrA22) all contain regions which exhibit strong homology with the white clover Adh cDNA probe used to isolate them. However none of these isolates represents a functional white clover *Adh* gene.

Based on the results presented in this study, a number of avenues for future work can be identified. One of a family of white clover *rbcS* genes has been characterized and it would be of interest to isolate the remaining members of this gene

family, along with a cDNA clone of each of their transcripts, so that the relative levels of *rbcS* gene expression in white clover could be determined. It would also be of interest to characterize promoter elements associated with the *rbcS* genes so that those elements responsible for conferring differential gene expression could be identified. Since highly conserved regions in *rbcS* genes from a variety of plants have been identified, it should be possible to use polymerase chain reaction (PCR) techniques with primers to these conserved regions to isolate the full complement of white clover SSU cDNA clones directly from RNA. These clones could in turn be used to screen the genomic library directly. Alternatively, primers specific for each cDNA could be synthesized and used with PCR techniques to isolate each of the genes directly from white clover DNA.

In the case of the white clover *Adh* gene, it would be of interest to analyse gene regulation and to isolate a functional gene so that promoter elements responsible for conferring root-specific expression of the gene could be identified. Since the region of homology with the white clover cDNA clone for each of the non-functional genomic clones that have been isolated corresponds to the central portion of the cDNA, it should be possible to choose a subclone from the cDNA clone for use as a probe which would preclude the re-isolation of these clones in any future screening of the genomic library. Alternatively, primers specific for each end of the Adh cDNA clone could be synthesized and used with PCR techniques to isolate the gene directly from white clover DNA.

Chapter 6

Bibliography

Andrews, T.J. (1988). Catalysis by cyanobacterial ribulose-bisphosphate carboxylase large subunits in the complete absence of small subunits. J. Biol. Chem. 263, 12213-12219.

Aoyagi, K., Kuhlemeier, C., and Chua, N.-H. (1988). The pea *rbcS-3A* enhancer-like element directs cell-specific expression in transgenic tobacco. Mol. Gen. Genet. 213, 179-185.

Appels, R. (1983). Chromosome structure in cereals: the analysis of regions containing repeated sequence DNA and its application to the detection of alien chromosomes introduced into wheat. In: Genetic Engineering of Plants, (T. Kosuge, C.P. Meredith, and A. Hollaender eds.), pp. 229-256. Plenum Press, New York.

Barta, A., Sommergruber, K., Thompson, D., Hartmuth, K., Matzke, M.A., and Matzke, A.J.M. (1986). The expression of a nopaline synthase - human growth hormone chimaeric gene in transformed tobacco and sunflower callus tissue. Plant Mol. Biol. 6, 347-357.

Bedbrook, J.R., Jones, J., O'Dell, M., Thompson, R.D., and Flavell, R.B. (1980a). A molecular description of telomeric heterochromatin in Secale species. Cell 19, 545-560.

Bedbrook, J.R., Smith, S.M., and Ellis, R.J. (1980b). Molecular cloning and sequencing of cDNA encoding the precursor to the small subunit of chloroplast ribulose-1,5-bisphosphate carboxylase. Nature 287, 692-697.

Benton, W.D., and Davis, R.W. (1977). Screening λgt recombinant clones by hybridization to single plaques in situ. Science 196, 180-182.

Berry-Lowe, S.L., McKnight, T.D., Shah, D.M., and Meagher, R.B. (1982). The nucleotide sequence, expression, and evolution of one member of a multigene family encoding the small subunit of ribulose-1,5-bisphosphate carboxylase in soybean. J. Mol. Appl. Genet. 1, 483-498.

Bicsak, T.A., Kann, L.R., Reiter, A., and Chase, T.Jr. (1982). Tomato alcohol dehydrogenase: purification and substrate specificity. Arch. Biochem. Biophys. 216, 605-615.

Breathnach, R., and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Ann. Rev. Biochem. 50, 349-383.

Britten, E.J. (1963). Chromosome numbers in the genus *Trifolium*. Cytologia 28, 428-449.

Broglie, R., Coruzzi, G., Lamppa, G., Keith, B., and Chua, N.-H. (1983). Structural analysis of nuclear genes coding for the precursor to the small subunit of wheat ribulose-1,5-bisphosphate carboxylase. Biotechnology 1, 55-61.

Brown, J.W.S. (1986). A catalogue of splice junction and putative branch point sequences from plant introns. Nucl. Acids Res. 14, 9549-9559.

Bruhat, A., Tourmente, S., Chapel, S., Sobrier, M.L., Couderc, J.L., and Dastugue, B. (1990). Regulatory elements in the first intron contribute to transcriptional regulation of the β_3 tubulin gene by 20-hydroxyecdysone in Drosophila Kc cells. Nucl. Acids Res. 18, 2861-2867.

Brzezinski, R., Talbot, B.G., Brown, D., Klimuszko, D., Blakeley, S.D., and Thirion, J.-P. (1986). Characterization of alcohol dehydrogenase in young soybean seedlings. Biochem. Genet. 24, 643-656.

Buratowski, S., Hahn, S., Guarente, L., and Sharp, P.A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. Cell 56, 549-561.

Callis, J., Fromm, M., and Walbot, V. (1987). Introns increase gene expression in cultured maize cells. Genes Dev. 1, 1183-1200.

Campbell, W.H., and Gowri, G. (1990). Codon usage in higher plants, green algae, and cyanobacteria. Plant Physiol. 92, 1-11.

Carlock, L.R. (1986). Analyzing lambda libraries. Focus 8:2, 6-8.

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.

Chee, P.P., Klassey, R.C., and Slightom, J.L. (1986). Expression of a bean storage protein 'phaseolin minigene' in foreign plant tissues. Gene 41, 47-57.

Chen, C.-C., and Gibson, P.B. (1971). Karyotypes of fifteen *Trifolium* species in section *Amoria*. Crop Sci. 11, 441-445.

Clarke, L., and Carbon, J. (1976). A colony bank containing synthetic Col El hybrid plasmids representative of the entire *E. coli* genome. Cell 9, 91-99.

Corby, H.D.L., Polhill, R.M., and Sprent, J.I. (1983). Taxonomy. In: Nitrogen Fixation, Volume 3, Legumes (W.J. Broughton ed.), pp. 1-35. Clarendon Press, Oxford.

Coruzzi, G., Broglie, R., Cashmore, A., and Chua, N.-H. (1983). Nucleotide sequences of two pea cDNA clones encoding the small subunit of ribulose 1,5-bisphosphate carboxylase and the major chlorophyll a/b-binding thylakoid polypeptide. J. Biol. Chem. 258, 1399-1402.

Coruzzi, G., Broglie, R., Edwards, C., and Chua, N.-H. (1984). Tissue-specific and light-regulated expression of a pea nuclear gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase. EMBO J. 3, 1671-1679.

Crouse, J., and Amorese, D. (1987). Ethanol precipitation: ammonium acetate as an alternative to sodium acetate. Focus 9(2), 3-5.

D'Alessio, J.M., Noon, M.C., Ley, H.L.III, and Gerard, G.F. (1987). One-tube double-stranded cDNA synthesis using cloned M-MLV reverse transcriptase. Focus 9(1), 1-4.

D'Alessio, J.M., and Gerard, G.F. (1988). Second-strand cDNA synthesis with *E. coli* DNA polymerase I and RNase H: the fate of information at the mRNA 5' terminus and the effect of *E. coli* DNA ligase. Nucl. Acids Res. 16, 1999-2014.

Davis, R.W., Botstein, D., and Roth, J.R. (1980). Advanced Bacterial Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Dean, C., van den Elzen, P., Tamaki, S., Dunsmuir, P., and Bedbrook, J. (1985a). Differential expression of the eight genes of the petunia ribulose bisphosphate carboxylase small subunit multi-gene family. EMBO J. 4, 3055-3061.

Dean, C., van den Elzen, P., Tamaki, S., Dunsmuir, P., and Bedbrook, J. (1985b). Linkage and homology analysis divides the eight genes for the small subunit of petunia ribulose 1,5-bisphosphate carboxylase into three gene families. Proc. Natl. Acad. Sci. USA 82, 4964-4968.

Dean, C., Tamaki, S., Dunsmuir, P., Favreau, M., Katayama, C., Dooner, H., and Bedbrook, J. (1986). mRNA transcripts of several plant genes are polyadenylated at multiple sites *in vivo*. Nucl. Acids Res. 14, 2229-2240.

Dean, C., van den Elzen, P., Tamaki, S., Black, M., Dunsmuir, P., and Bedbrook, J. (1987). Molecular characterization of the *rbcS* multi-gene family of *Petunia* (Mitchell). Mol. Gen. Genet. 206, 465-474.

Dean, C., Pichersky, E., and Dunsmuir, P. (1989a). Structure, evolution, and regulation of *RbcS* genes in higher plants. Ann. Rev. Plant Physiol. Plant Mol. Biol. 40, 415-439.

Dean, C., Favreau, M., Bedbrook, J., and Dunsmuir, P. (1989b). Sequences 5' to translation start regulate expression of petunia *rbcS* genes. Plant Cell 1, 209-215.

Dean, C., Favreau, M., Bond-Nutter, D., Bedbrook, J., and Dunsmuir, P. (1989c). Sequences downstream of translation start regulate quantitative expression of two petunia *rbcS* genes. Plant Cell 1, 201-208.

DeLisle, A.J., and Ferl, R.J. (1990). Characterization of the *Arabidopsis Adh* G-box binding factor. Plant Cell 2, 547-557.

Dellaporta, S.L., Wood, S.L., and Hicks, J.B. (1983a). Maize DNA minipreps. Maize Genetics Cooperation Newsletter 57, 26-29.

Dellaporta, S.L., Wood, J., and Hicks, J.B. (1983b). A plant DNA minipreparation: version II. Plant Mol. Biol. Rep. 4, 19-21.

Dellaporta, S.L., Wood, J., and Hicks, J.B. (1985). In: Molecular Biology of Plants - A Laboratory Course Manual, pp. 36-37. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

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.

Dolferus, R., and Jacobs, M. (1984). Polymorphism of alcohol dehydrogenase in *Arabidopsis thaliana* (L.) Heynh.: genetical and biochemical characterization. Biochem. Genet. 22, 817-838.

Dolferus, R., Marbaix, G., and Jacobs, M. (1985). Alcohol dehydrogenase in *Arabidopsis:* analysis of the induction phenomenon in plantlets and tissue cultures. Mol. Gen. Genet. 199, 256-264.

Donald, R.G.K., and Cashmore, A.R. (1990). Mutation of either G box or I box sequences profoundly affects expression from the *Arabidopsis rbcS-1A* promoter. EMBO J. 9, 1717-1726.

Drouin, G., and Dover, G.A. (1987). A plant processed pseudogene. Nature 328, 557-558.

Ellis, J.G., Llewellyn, D.J., Dennis, E.S., and Peacock, W.J. (1987). Maize *Adh-1* promoter sequences control anaerobic regulation: addition of upstream promoter elements from constitutive genes is necessary for expression in tobacco. EMBO J. 6, 11-16.

Ellison, N., Zhu, J-M, and Yu, P-L. (1986). Highly repeated DNA of *Trifolium repens*: sequence analysis and *in situ* hybridization. In: Programme and Abstracts - Genetic Manipulation, Its Application and Potential. Hanmer Springs, New Zealand.

Ferl, R.J., and Nick, H.S. (1987). *In vivo* detection of regulatory factor binding sites in the 5′ flanking region of maize *Adh1*. J. Biol. Chem. 17, 7947-7950.

Ferl, R.J., and Laughner, B.H. (1989). *In vivo* detection of regulatory factor binding sites of *Arabidopsis thaliana Adh*. Plant Mol. Biol. 12, 357-366.

Flavell, R. (1980). The molecular characterization and organization of plant chromosomal DNA sequences. Ann. Rev. Plant Physiol. 31, 569-596.

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

Fluhr, R., and Chua, N.-H. (1986). Developmental regulation of two genes encoding ribulose-bisphosphate carboxylase small subunit in pea and transgenic petunia plants: phytochrome responses and blue-light induction. Proc. Natl. Acad. Sci. USA 83, 2358-2362.

Fluhr, R., Moses, P., Morelli, G., Coruzzi, G., and Chua, N.-H. (1986a). Expression dynamics of the pea *rbcS* multigene family and organ distribution of the transcripts. EMBO J. 5, 2063-2071.

Fluhr, R., Kuhlemeier, C., Nagy, F., and Chua, N.-H. (1986b). Organ-specific and light-induced expression of plant genes. Science 232, 1106-1112.

Freeling, M. (1973). Simultaneous induction by anaerobiosis or 2,4-D of multiple enzymes specified by two unlinked

genes: differential *Adh1-Adh2* expression in maize. Mol. Gen. Genet. 127, 215-227.

Freeling, M. (1976). Intragenic recombination in maize: pollen analysis methods and the effect of parental *Adh1*$^{+}$ isoalleles. Genetics 83, 701-717.

Freeling, M., and Bennett, D.C. (1985). Maize *Adh1*. Ann. Rev. Genet. 19, 297-323.

Frischauf, A.-M., Lehrach, H., Poustka, A., and Murray, N. (1983). Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. 170, 827-842.

Gerard, G.F., and Miller, K. (1986). Comparison of glyoxal and formaldehyde gels for sizing rRNAs. Focus 8(3), 5-6.

Gerard, G.F., and D'Alessio, J.M. (1987). Comparison of reverse transcriptase and *E. coli* DNA polymerase I for synthesis of second strand cDNA. Focus 9(4), 4-7.

Giuliano, G., Pichersky, E., Malik, V.S., Timko, M.P., Scolnik, P.A., and Cashmore, A.R. (1988). An evolutionary conserved protein binding sequence upstream of a plant light-regulated gene. Proc. Natl. Acad. Sci. USA 85, 7089-7093.

Goodall, G.J., and Filipowicz, W. (1989). The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. Cell 58, 473-483.

Goodall, G.J., and Filipowicz, W. (1990). The minimum functional length of pre-mRNA introns in monocots and dicots. Plant Mol. Biol. 14, 727-733.

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

Grandbastien, M.A., Berry-Lowe, S., Shirley, B.W., and Meagher, R.B. (1986). Two soybean ribulose-1,5-bisphoshate carboxylase small subunit genes share extensive homology even in distant flanking sequences. Plant Mol. Biol. 7, 451-465.

Green, M.R. (1986). Pre-mRNA splicing. Ann. Rev. Genet. 20, 671-708.

Green, P.J., Kay, S.A., and Chua, N.-H. (1987). Sequence-specific interactions of a pea nuclear factor

with light-responsive elements upstream of the rbcS-3A gene. EMBO J. 6, 2543-2549.

Grierson, D., and Covey, S.N. (1984). Plant Molecular Biology. Blackie, Glasgow and London.

Grinsted, J., and Bennett, P.M. (1988). Preparation and electrophoresis of plasmid DNA. Methods Microbiol. 21, 129-142.

Gubler, U., and Hoffman, B.J. (1983). A simple and very efficient method for generating cDNA libraries. Gene 25, 263-269.

Hake, S., and Walbot, V. (1980). The genome of *Zea mays*, its organization and homology to related grasses. Chromosoma 79, 251-270.

Hanley, B.A., and Schuler, M.A. (1988). Plant intron sequences: evidence for distinct groups of introns. Nucl. Acids Res. 14, 7159-7176.

Hattori, M., and Sakaki, Y. (1986). Dideoxy sequencing method using denatured plasmid templates. Anal. Biochem. 152, 232-238.

Hattori, J., Gottlob-McHugh, S.G., and Johnson, D.A. (1987). The isolation of high-molecular-weight DNA from plants. Anal. Biochem. 165, 70-74.

Hawkins, J.D. (1988). A survey of intron and exon lengths. Nucl. Acids Res. 16, 9893-9908.

Heidecker, G., and Messing, J. (1986). Structural analysis of plant genes. Annu. Rev. Plant Physiol. 37, 439-466.

Heim, U., Schubert, R., Bäumlein, H., and Wobus, U. (1989). The legumin gene family: structure and evolutionary implications of *Vicia faba* genes and pseudogenes. Plant Mol. Biol. 13, 653-663.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28, 351-359.

Henikoff, S. (1987). Exonuclease III generated deletions for DNA sequence analysis. Promega Notes 8, 1-3.

Higgins, D.G., and Sharp, P.M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73, 237-244.

Higgins, D.G., and Sharp, P.M. (1989). Fast and sensitive multiple sequence alignments on a microcomputer. Comp. Applic. Biosci. 5, 151-153.

Higgins, T.J.V., Chandler, P.M., Zurawski, G., Button, S.C., and Spencer, D. (1983). The biosynthesis and primary structure of pea seed lectin. J. Biol. Chem. 258, 9544-9549.

Howard, E.A., Walker, J.C., Dennis, E.S., and Peacock, W.J. (1987). Regulated expression of an alcohol dehydrogenase 1 chimeric gene introduced into maize protoplasts. Planta 170, 535-540.

Hunt, A.G., Chu, N.M., Odell, J.T., Nagy, F., and Chua, N.-H. (1987). Plant cells do not properly recognize animal gene polyadenylation signals. Plant Mol. Biol. 8, 23-35.

Hunt, A.G. (1988). Identification and characterization of cryptic polyadenylation sites in the 3′ region of a pea ribulose-1,5-bisphosphate carboxylase small subunit gene. DNA 7, 329-336.

Hunt, A.G., and MacDonald, M.H. (1989). Deletion analysis of the polyadenylation signal of a pea ribulose-1,5-bisphosphate carboxylase small-subunit gene. Plant Mol. Biol. 13, 125-138.

Hutchison, K.W., Harvie, P.D., Singer, P.B., Brunner, A.F., and Greenwood, M.S. (1990). Nucleotide sequence of the small subunit of ribulose-1,5-bisphosphate carboxylase from the conifer *Larix laricina*. Plant Mol. Biol. 14, 281-284.

Huynh, T.V., Young, R.A., and Davis, R.W. (1985). Constructing and screening cDNA libraries in λgt10 and λgt11. In: DNA Cloning (D.M. Glover ed.), Vol. I, pp. 49-78. IRL Press, Oxford and Washington.

Ingle, J., Pearson, G.G., and Sinclair, J. (1973). Species distribution and properties of nuclear satellite DNA in higher plants. Nature New Biol. 242, 193-197.

Jamet, E., Fargeas, C., Durr, A., and Fleck, J. (1990). Nucleotide sequences of two genes encoding the small subunit of RUBISCO in *Nicotiana sylvestris*. Nucl. Acids Res. 18, 4589.

Jendrisak, J., Young, R.A., and Engel, J. (1987). In: Guide to Molecular Cloning Techniques (S. Berger and A. Kimmel eds.), pp. 359-371. Academic Press.

Jones, J.D.G., and Flavell, R.B. (1982). The mapping of highly-repeated DNA families and their relationship to C-bands in chromosomes of *Secale cereale*. Chromosoma 86, 595-612.

Joshi, C.P. (1987a). An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucl. Acids Res. 15, 6643-6653.

Joshi, C.P. (1987b). Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis. Nucl. Acids Res. 15, 9627-9640.

Kaiser, K., and Murray, N.E. (1985). The use of phage lambda replacement vectors in the construction of representative genomic DNA libraries. In: DNA Cloning (D.M. Glover ed.), Vol. I, pp. 1-47. IRL Press, Oxford and Washington.

Kamalay, J.C., and Goldberg, R.B. (1984). Organ-specific nuclear RNAs in tobacco. Proc. Natl. Acad. Sci. USA 81, 2801-2805.

Karlin-Neumann, G.A., and Tobin, E.M. (1986). Transit peptides of nuclear-encoded chloroplast proteins share a common amino acid framework. EMBO J. 5, 9-13.

Kaslov, D.C. (1986). A rapid biochemical method for purifying lambda DNA from phage lysates. Nucl. Acids Res. 14, 6767.

Keith, B., and Chua, N.-H. (1986). Monocot and dicot pre-mRNAs are processed with different efficiencies in transgenic tobacco. EMBO J. 5, 2419-2425.

Kimmerer, T.W. (1987). Alcohol dehydrogenase and pyruvate decarboxylase activities in leaves and roots of eastern cottonwood (*Populus deltoides* Bartr.) and soybean (*Glycine max* L.). Plant Physiol. 84, 1210-1213.

King, P.V., and Blakesley, R.W. (1986). Optimizing DNA ligations for transformation. Focus 8(1), 1-3.

Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucl. Acids Res. 12, 857-872.

Krebbers, E., Seurinck, J., Herdies, L., Cashmore, A.R., and Timko, M.P. (1988). Four genes in two diverged subfamilies encode the ribulose-1,5-bisphosphate carboxylase small subunit polypeptides of *Arabidopsis thaliana*. Plant Mol. Biol. 11, 745-759.

Kuhlemeier, C., Fluhr, R., Green, P.J., and Chua, N.-H. (1987a). Sequences in the pea *rbcS-3A* gene have homology to constitutive mammalian enhancers but function as negative regulatory elements. Genes Dev. 1, 247-255.

Kuhlemeier, C., Green, P.J., and Chua, N.-H. (1987b). Regulation of gene expression in higher plants. Ann. Rev. Plant Physiol. 38, 221-257.

Kuhlemeier, C., Cuozzo, M., Green, P.J., Goyvaerts, E., Ward, K., and Chua, N.-H. (1988). Localization and conditional redundancy of regulatory elements in *rbcS-3A*, a pea gene encoding the small subunit of ribulose-bisphosphate carboxylase. Proc. Natl. Acad. Sci. USA 85, 4662-4666.

Kuhlemeier, C., Strittmatter, G., Ward, K., and Chua, N.-H. (1989). The pea *rbcS-3A* promoter mediates light responsiveness but not organ specificity. Plant Cell 1, 471-478.

Lapeyre, B., and Amalric, F. (1985). A powerful method for the preparation of cDNA libraries: isolation of cDNA encoding a 100-kDal nucleolar protein. Gene 37, 215-220.

Lebrun,M., Waksman,G. and Freyssinet,G. (1987). Nucleotide sequence of a gene encoding corn ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit (rbcs). Nucl. Acids Res. 15, 4360.

Lee, J.S., and Verma, D.P.S. (1984). Structure and chromosomal arrangement of leghemoglobin genes in kidney bean suggest divergence in soybean leghemoglobin gene loci following tetraploidization. EMBO J. 3, 2745-2752.

Leutwiler, L.S., Hough-Evans, B.R., and Meyerowitz, E.M. (1984). The DNA of *Arabidopsis thaliana*. Mol. Gen. Genet. 194, 15-23.

Lichtenstein, C., and Draper, J. (1985). Genetic Engineering of Plants. In: DNA Cloning (D.M. Glover ed.), Vol. II, pp. 67-119. IRL Press, Oxford and Washington.

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 *Pisum sativum* (cv. "Greenfeast"). J. Mol. Biol. 195, 115-123.

Logemann, J., Schell, J., and Willmitzer, L. (1987). Improved method for the isolation of RNA from plant tissues. Anal. Biochem. 163, 16-20.

Lütcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F., and Scheele, G.A. (1987). Selection of AUG initiation codons differs in plants and animals. EMBO J. 6, 43-48.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular Cloning - A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Manzara, T., and Gruissem, W. (1988). Organization and expression of the genes encoding ribulose-1,5-bisphosphate carboxylase in higher plants. Photosynth. Res. 16, 117-139.

Marshall, D.R., Broué and Oram, R.N. (1974). Genetic control of alcohol dehydrogenase isozymes in narrow-leafed lupins. J. Hered. 65, 198-203.

Martin, R. (1987). Overcoming DNA sequencing artifacts: stops and compressions. Focus 9(1), 8-9.

Matsuoka, M., Kano-Murakami, Y., Tanaka, Y., Ozeki, Y., and Yamamoto, N. (1987). Nucleotide sequence of cDNA encoding the small subunit of ribulose-1,5-bisphosphate carboxylase from maize. J. Biochem. 102, 673-676.

Matton, D.P., and Brissom, N. (1990a). Nucleotide sequence of two potato alcohol dehydrogenase cDNAs. Nucl. Acids Res. 18, 3070.

Matton, D.P., and Brisson, N. (1990b). Alcohol dehydrogenase gene expression in potato following elicitor and stress treatment. Plant Mol. Biol. 14, 775-783.

Mazur, B.J., and Chui, C.-F. (1985). Sequence of a genomic DNA clone for the small subunit of ribulose bis-phosphate carboxylase-oxygenase from tobacco. Nucl. Acids Res. 13, 2373-2386.

McKendree, W.L., Paul, A.-L., DeLisle, A.J., and Ferl, R.J. (1990). In vivo and in vitro characterization of protein interactions with the dyad G-box of the *Arabidopsis Adh* gene. Plant Cell 2, 207-214.

Meinkoth, J., and Wahl, G. (1984). Hybridization of nucleic acids immobilized on solid supports. Anal. Biochem. 138, 267-284.

Meyerowitz, E.M., and Pruitt, R.E. (1985). *Arabidopsis thaliana* and plant molecular genetics. Science 229, 1214-1218.

Mierendorf, R.C., and Pfeffer, D. (1987). Direct sequencing of denatured plasmid DNA. Methods Enzymol. 152, 556-562.

Miller, H. (1987). Practical aspects of preparing phage and plasmid DNA: growth, maintenance, and storage of bacteria and bacteriophage. Methods Enzymol. 152, 145-170.

Miziorko, H.M., and Lorimer, G.H. (1983). Ribulose-1,5-bisphosphate carboxylase-oxygenase. Ann. Rev. Biochem. 52, 507-535.

Morelli, G., Nagy, F., Fraley, R.T., Rogers, S.G., and Chua, N.-H. (1985). A short conserved sequence is involved in the light-inducibility of a gene encoding ribulose 1,5-bisphosphate carboxylase small subunit of pea. Nature 315, 200-204.

Müller, H.-P., and Schaffner, W. (1990). Transcriptional enhancers can act *in trans*. Trends Genet. 6, 300-304.

Murray, E.E., Lotzer, J., and Eberle, M. (1989). Codon usage in plant genes. Nucl. Acids Res. 17, 477-498.

Murray, M.G., Cuellar, R.E., and Thompson, W.F. (1978). DNA sequence organization in the pea genome. Biochemistry 17, 5781-5790.

Murray, M.G., Palmer, J.D., Cuellar, R.E., and Thompson, W.F. (1979). Deoxyribonucleic acid sequence organization in the mung bean genome. Biochemistry 18, 5259-5266.

Murray, M.G., and Thompson, W.F. (1980). Rapid isolation of high molecular weight plant DNA. Nucl. Acids Res. 8, 4321-4325.

Nader, W.F., Edlind, T.D., Huettermann, A., and Sauer, H.W. (1985). Cloning of *Physarum* actin sequences in an exonuclease-deficient bacterial host. Proc. Natl. Acad. Sci. USA 82, 2698-2702.

Nagy, F., Morelli, G., Fraley, R.T., Rogers, S.G., and Chua, N.-H. (1985). Photoregulated expression of a pea *rbcS* gene in leaves of transgenic plants. EMBO J. 4, 3063-3068.

Nagy, F., Kay, S.A., and Chua, N.-H. (1988). Gene regulation by phytochrome. Trends Genet. 4, 37-42.

Nakayama, K., and Nakauchi, H. (1989). An improved method to make sequential deletion mutants for DNA sequencing. Trends Genet. 5, 325.

Nixon, B.T. (1989). Construction of nested deletions for DNA sequencing. In: Current Protocols in Molecular Biology (F.M. Ausubel, R. Brent, R.E. Kingston, D.D.Moore, J.G. Seidman, J.A. Smith, and K. Struhl eds.), Vol. I, pp. 7.3.9-7.3.21. John Wiley & Sons, New York.

O'Neal, J.K., Pokalsky, A.R., Kiehne, K.L., and Shewmaker, C.K. (1987). Isolation of tobacco SSU genes: characterization of a transcriptionally active pseudogene. Nucl. Acids Res. 15, 8661-8677.

Okimoto, R., Sachs, M.M., Porter, E.K., and Freeling, M. (1980). Patterns of polypeptide synthesis in various maize organs under anaerobiosis. Planta 150, 89-94.

Pape, M.E., and Kim, K.-H. (1987). Improvements in production of cDNA libraries in λgt11: detection of ligation inhibitor and separation of cDNA from free linkers. Gene Anal. Techn. 4, 14-16.

Peacock, W.J., Dennis, E.S., Rhoades, M.M., and Pryor, A.J. (1981). Highly repeated DNA sequence limited to knob

heterochromatin in maize. Proc. Natl. Acad. Sci. USA 78, 4490-4494.

Pearson, W.R., and Lipman, D.J. (1988). Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA 85, 2444-2448.

Peterhans, A., Datta, S.K., Datta, K., Goodall, G.J., Potrykus, I., and Paszkowski, J. (1990). Recognition efficiency of *Dicotyledoneae*-specific promoter and RNA processing signals in rice. Mol. Gen. Genet. 222, 361-368.

Pikaard, C.S., Mignery, G.A., Ma, D.P., Stark, V.J., and Park, W.D. (1986). Sequence of two apparent pseudogenes of the major potato tuber protein, patatin. Nucl. Acids Res. 14, 5564-5566.

Pinck, M., Dore, J.-M., Guilley, E., Durr, A., Pinck, L., Hirth, L., and Fleck, J. (1986). A simple gene-expression system for the small subunit of ribulose bisphosphate carboxylase in leaves of *Nicotiana sylvestris*. Plant Mol. Biol. 7, 301-309.

Pohl, I. (1988). Quick lambda DNA preps. Colloquium (Boehringer) 1/1988, 73.

Polans, N.O., Weeden, N.F., and Thompson, W.F. (1985). Inheritance, organization, and mapping of *rbcS* and *cab* multigene families in pea. Proc. Natl. Acad. Sci. USA 82, 5083-5087.

Porteous, D.J. (1986). Rapid isolation and characterization of hybridization selected recombinants from lambda genomic libraries. Anal. Biochem. 159, 17-23.

Poulsen, C., Fluhr, R., Kauffman, J.M., Boutry, M., and Chua, N.-H. (1986). Characterization of an *rbc*S gene from *Nicotiana plumbaginifolia* and expression of an *rbc*S-CAT chimeric gene in homologous and heterologous nuclear background. Mol. Gen. Genet. 205, 193-200.

Pritchard, A.J. (1962). Number and morphology of chromosomes in African species in the genus *Trifolium* L. Aust. J. Agric. Res. 13, 1023-1029.

Pritchard, A.J. (1969). Chromosome numbers in some species of *Trifolium*. Aust. J. Agric. Res. 20, 883-887.

Reiss, B., Wasmann, C.C., and Bohnert, H.J. (1987). Regions in the transit peptide of SSU essential for transport into chloroplasts. Mol. Gen. Genet. 209, 116-121.

Ricard, B., Mocquot, B., Fournier, A., Delseny, M., and Pradet, A. (1986). Expression of alcohol dehydrogenase in rice embryos under anoxia. Plant Mol. Biol. 7, 321-329.

Rigby, P.W.J., Dieckmann, M., Rhodes, C., and Berg, P. (1977). Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. 113, 237-251.

Rowland, L.J., and Strommer, J.N. (1986). Anaerobic treatment of maize roots affects transcription of *Adh1* and transcript stability. Mol. Cell. Biol. 6, 3368-3372.

Sachs, M.M., and Freeling, M. (1978). Selective synthesis of alcohol dehydrogenase during anaerobic treatment of maize. Mol. Gen. Genet. 161, 111-115.

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

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

Salditt-Georgieff, M., Harpold, M., Chen-Kiang, S., and Darnell, J.E.Jr. (1980). The addition of 5′ cap structures occurs early in hnRNA synthesis and prematurely terminated molecules are capped. Cell 19, 69-78.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning - A Laboratory Manual, Second Edition. Cold Spring Harbor Laboratory Press, New York.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Schleif, R. (1988). DNA looping. Science 240, 127-128.

Schmidt, M.C., Kao, C.C., Pei, R., and Berk, A.J. (1989). Yeast TATA-box transcription factor gene. Proc. Natl. Acad. Sci. USA 86, 7785-7789.

Schneider, G., Knight, S., Andersson, I., Brändén, C.-I., Lindqvist, Y., and Lindqvist, T. (1990). Comparison of the crystal structures of L_2 and L_8S_8 Rubisco suggests a functional role for the small subunit. EMBO J. 9, 2045-2050.

Schulze-Lefert, P., Dangl, J.L., Becker-André, M., Hahlbrock, K., and Schulz, W. (1989). Inducible *in vivo* DNA footprints define sequences necessary for UV light activation of the parsley chalcone synthase gene. EMBO J. 8, 651-656.

Schwartz, D. (1969). Alcohol dehydrogenase in maize: genetic basis for multiple isozymes. Science 164, 585-586.

Schwartz, D. (1971). Genetic control of alcohol dehydrogenase - a competition model for regulation of gene action. Genetics 67, 411-425.

Seed, B., Parker, R.C., and Davidson, N. (1982). Representation of DNA sequences in recombinant DNA libraries prepared by restriction enzyme partial digestion. Gene 19, 201-209.

Senn, H.A. (1938). Chromosome number relationships in Leguminosae. Bibliographia Genetica 12, 175-336.

Shapiro, M.B., and Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and funtional implications in gene expression. Nucl. Acids Res. 15, 7155-7174.

Shatkin, A.J. (1976). Capping of eukaryotic mRNAs. Cell 9, 645-653.

Sheen, J.-Y., and Bogorad, L. (1986). Expression of the ribulose-1,5-bisphosphate carboxylase large subunit gene and three small subunit genes in two cell types of maize leaves. EMBO J. 5, 3417-3422.

Sheets, M.D., Ogg, S.C., and Wickens, M.P. (1990). Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation *in vitro*. Nucl. Acids Res. 18, 5799-5805.

Shirley, B.W., Ham, D.P., Senecoff, J.F., Berry-Lowe, S.L., Zurfluh, L.L., Shah, D.M., and Meagher, R.B. (1990).

Comparison of the expression of two highly homologous members of the soybean ribulose-1,5-bisphosphate carboxylase small subunit gene family. Plant Mol. Biol. 14, 909-925.

Shirley, B.W., and Meagher, R.B. (1990). A potential role for RNA turnover in the light regulation of plant gene expression: ribulose-1,5-bisphosphate carboxylase small subunit in soybean. Nucl. Acids Res. 18, 3377-3385.

Silverthorne, J., Wimpee, C.F., Yamada, T., Rolfe, S.A., and Tobin, E.M. (1990). Differential expression of individual genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase in *Lemna gibba*. Plant Mol. Biol. 15, 49-58.

Singh, I., and Jones, K.W. (1984). The use of heparin as a simple cost-effective means of controlling background in nucleic acid hybridization procedures. Nucl. Acids Res. 12, 5627-5638.

Slightom, J.L., Drong, R.F., Klassy, C., and Hoffman, L.M. (1985). Nucleotide sequences from phaseolin cDNA clones: the major storage proteins from *Phaseolus vulgaris* are encoded by two unique gene families. Nucl. Acids Res. 13, 6483-6498.

Smale, S.T., and Baltimore, D. (1989). The "initiator" as a transcription control element. Cell 57, 103-113.

Smith, S.M., and Speirs, J. (1983). Characterization of three cDNA clones encoding different mRNAs for the precursor to the small subunit of wheat ribulosebisphosphate carboxylase. Nucl. Acids Res. 11, 8719-8734.

Smith, C.W., Patton, J.G., and Nadal-Ginard, B. (1989). Alternative splicing in the control of gene expression. Ann. Rev. Genet. 23, 527-577.

Sorenson, J.C. (1984). The structure and expression of nuclear genes in higher plants. Adv. Genet. 22, 109-144.

Stanford, A., Bevan, M., and Northcote, D. (1989). Differential expression within a family of novel wound-induced genes in potato. Mol. Gen. Genet. 215, 200-208.

Stiekema, W.J., Wimpee, C.F., and Tobin, E.M. (1983). Nucleotide sequence encoding the precursor of the small subunit of ribulose 1,5-bisphosphate carboxylase from *Lemna gibba* L.G-3. Nucl. Acids Res. 11, 8051-8061.

Sugita, M., Manzara, T., Pichersky, E., Cashmore, A., and Gruissem, W. (1987). Genomic organization, sequence analysis and expression of all five genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from tomato. Mol. Gen. Genet. 209, 247-256.

Sutton, W.D. (1974). Some features of the DNA of *Rhizobium* bacteroids and bacteria. Biochim. Biophys. Acta 366, 1-10.

Tabor, S., and Richardson, C.C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA 84, 4767-4771.

Tanksley, S.D., and Jones, R.A. (1981). Effects of O_2 stress on tomato alcohol dehydrogenase activity: description of a second ADH coding gene. Biochem. Genet. 19, 397-409.

Timko, M.P., Kausch, A.P., Castresana, C., Fassler, J., Herrera-Estrella, L., Van den Broeck, G., Van Montagu, M., Schell, J., and Cashmore, A.R. (1985). Light regulation of plant gene expression by an upstream enhancer-like element. Nature 318, 579-582.

Toneguzzo, F., Glynn, S., Levi, E., Mjolsness, S., and Hayday, A. (1988). Use of a chemically modified T7 DNA polymerase for manual and automated sequencing of supercoiled DNA. BioTechniques 6, 460-469.

Torres, A.M., Diedenhofen, U., and Johnstone, I.M. (1977). The early allele of alcohol dehydrogenase in sunflower populations. J. Hered. 68, 11-16.

Trick, M., Dennis, E.S., Edwards, K.J.R., and Peacock, W.J. (1988). Molecular analysis of the alcohol dehydrogenase gene family of barley. Plant Mol. Biol. 11, 147-160.

Tumer, N.E., Clark, W.G., Tabor, G.J., Hironaka, C.M., Fraley, R.T., and Shah, D.M. (1986). The genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase

are expressed differentially in petunia leaves. Nucl. Acids Res. 14, 3325-3342.

Ueda, T., Pichersky, E., Malik, V.S., and Cashmore, A.R. (1989). Level of expression of the tomato *rbcS-3A* gene is modulated by a far upstream promoter element in a developmentally regulated manner. Plant Cell 1, 217-227.

Vallejos, C.E., Tanksley, S.D., and Bernatzky, R. (1986). Localization in the tomato genome of DNA restriction fragments containing sequences homologous to the rRNA (45s), the major chlorophyll a/b binding polypeptide and the ribulose bisphosphate carboxylase genes. Genetics 112, 93-105.

Voelker, T.A., Staswick, P., and Chrispeels, M.J. (1986). Molecular analysis of two phytohemagglutinin genes and their expression in *Phaseolus vulgaris* cv. Pinto, a lectin-deficient cultivar of the bean. EMBO J. 5, 3075-3082.

Vogeli, G., and Kaytes, P.S. (1987). Amplification, storage, and replication of libraries. Methods Enzymol. 152, 407-415.

Vogelstein, B., and Gillespie, D. (1979). Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. USA 76, 615-619.

Wahl, G.M., and Berger, S.L. (1987). Screening colonies or plaques with radioactive nucleic acid probes. Methods Enzymol. 152, 415-423.

Wahl, G.M., Berger, S.L., and Kimmel, A.R. (1987). Molecular hybridization of immobilized nucleic acids: theoretical concepts and practical considerations. Methods Enzymol. 152, 399-407.

Waksman, G., Lebrun, M., and Freyssinet, G. (1987). Nucleotide sequence of a gene encoding sunflower ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit (rbcs). Nucl. Acids Res. 15, 7181.

Walbot, V., and Goldberg, R. (1979). Plant genome organization and its relationship to classical plant

genetics. In: Nucleic Acids in Plants (T.C. Hall and J.W. Davies eds.), Vol I, pp. 3-40. CRC Press, Boca Raton.

Walker, J.C., Howard, E.A., Dennis, E.S., and Peacock, W.J. (1987). DNA sequences required for anaerobic expression of the maize alcohol dehydrogenase 1 gene. Proc. Natl. Acad. Sci. USA 84, 6624-6628.

Walling, L.L., Chang, Y.C., Demmin, D.S., and Holzer, F.M. (1988). Isolation, characterization and evolutionary relatedness of three members from the soybean multigene family encoding chlorophyll a/b binding proteins. Nucl. Acids Res. 16, 10477-10492.

Watson, C.J., and Jackson, J.F. (1985). An alternative procedure for the synthesis of double-stranded cDNA for cloning in phage and plasmid vectors. In: DNA Cloning (D.M. Glover ed.), Vol. I, pp. 79-88. IRL Press, Oxford and Washington.

Watson, J.C., and Thompson, W.F. (1986). Purification and restriction endonuclease analysis of plant nuclear DNA. Methods Enzymol. 118, 57-75.

Wiborg, O., Hyldig-Nielsen, J.J., Jensen, E.Ø., Paludan, K., and Marcker, K.A. (1983). The structure of an unusual leghemoglobin gene from soybean. EMBO J. 2, 449-452.

Wickens, M. (1990). How the messenger got its tail: addition of poly(A) in the nucleus. Trends Biochem. Sci. 15, 277-281.

Wiebauer, K., Herrero, J.-J., and Filipowicz, W. (1988). Nuclear pre-mRNA processing in plants: distinct modes of 3'-splice-site selection in plants and animals. Mol. Cell. Biol. 8, 2042-2051.

Wilimzig, M. (1985). LiCl-boiling method for plasmid mini-preps. Trends Genet. 1, 6.

Williams, W.M. (1987). White clover taxonomy and biosystematics. In: White Clover (M.J. Baker and W.M. Williams eds.), pp. 323-342. C.A.B International, Wallingford.

Wolter, F.P., Fritz, C.C., Willmitzer, L., Schell, J., and Schreier, P.H. (1988). *rbcS* genes in *Solanum tuberosum*:

conservation of transit peptide and exon shuffling during evolution. Proc. Natl. Acad. Sci. USA 85, 846-850.

Wolyn, D.J., and Jelenkovic, G. (1990). Nucleotide sequence of an alcohol dehydrogenase gene in octoploid strawberry (*Fragaria* x *Ananassa* Duch.). Plant Mol. Biol. 14, 855.

Woodman, J.C., and Freeling, M. (1981). Identification of a genetic element that controls the organ-specific expression of *Adh1* in maize. Genetics 98, 357-378.

Wyman, A.R., and Wertman, K.F. (1987). Host strains that alleviate underrepresentation of specific sequences: overview. Methods Enzymol. 152, 173-180.

Xie, Y., and Wu, R. (1988). Nucleotide sequence of a ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit gene (rbcS) in rice. Nucl. Acids Res. 16, 7749.

Xie, Y., and Wu, R. (1990). Molecular analysis of an alcohol dehydrogenase-encoding genomic clone (*adh2*) from rice. Gene 87, 185-191.

Yamashita, I., Nemoto, Y., and Yoshikawa, S. (1976). NAD-dependent alcohol dehydrogenase and NADP-dependent alcohol dehydrogenase from strawberry seeds. Agric. Biol. Chem. 40, 2231-2235.

Yamashita, I., Iino, K., and Yoshikawa, S. (1978). Alcohol dehydrogenases from strawberry seeds. Agric. Biol. Chem. 42, 1125-1132.

Yanisch-Perron, C., Viera, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC vectors. Gene 33, 103-119.

Zhang, H., Scholl, R., Browse, J., and Somerville, C. (1988). Double stranded DNA sequencing as a choice for DNA sequencing. Nucl. Acids Res. 16, 1220.

Ziai, M.R., Giordano, A., Armandola, E., and Ferrone, S. (1988). Purification by ammonium sulfate precipitation of bacteriophage λgt11 DNA for restriction analysis of cloned cDNA inserts. Anal. Biochem. 171, 192-196.