Construction and screening of plant genomic libraries

Yaish, Sami Abdul-Rahman

Construction and screening of plant genomic libraries

by

Sami Abdul-Rahman Yaish
B.Sc. Nablus (Palestine), M.Sc. Glasgow (UK)

A thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy in the University of Durham

July, 1990

To my parents, my wife and my kids

✳❑ ❍● ❒❁❏❉■▼▲✌ ❍● ◗❋❊❉ ❁■❉ ❍● ❋❋❊▲

ABSTRACT

A library of pea (*Pisum sativum* L) genomic DNA in bacteriophage EMBL3 was screened for seed storage protein genes of the legumin and vicilin families. Three genomic clones were isolated. One of the clones was found to contain a gene in the *Leg A* sub-family which was designated *Leg E*. The nucleotide and predicted amino acid sequence of *Leg E* were compared to those of *Leg A*. The coding sequences of both genes are strongly homologous with only 9 bases difference out of 1560 bases.

A second genomic clone contained two genes from the *Leg J* subfamily, *Leg J* and *Leg K*. The clone was shown to overlap with a genomic clone isolated previously, JC5 (Gatehouse *et al*. 1988). Strong homology was found between the *Leg K* and *Leg J* sequences. The *Leg K* gene is predicted to be pseudogene, due to the conversion of the ATG methionine start codon to a GTG valine codon and the presence of a stop codon in the 5' end of the coding sequence in the reading frame predicted by the first subsequent start codon.

A genomic library was constructed for *Arabidopsis thaliana*, using EMBL3 as a vector to sub-clone Sau3AI partially digested *Arabidopsis* genomic DNA. About 8 x 10^4 random clones were obtained when the ligated vector DNA and insert were *in vitro* packaged.

The *Arabidopsis* gene library was screened for clones containing sequences encoding the cell wall protein extensin, using a rape (*Brassica napus* L extensin cDNA as a probe. Six clones were isolated, two of which were restriction mapped. One of them was partially sequenced. This clone did not contain an extensin gene homologous to the probe sequence, and only contained a short extensin-like sequence which was responsible for the observed hybridisation. The putative gene may represent another type of protein, since it was expressed in the root of *Arabidopsis* and *Brassica napus* L, as shown by "Northern" blots which were probed with labelled DNA from the clone.

CHAPTER THREE - RESULTS

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr J.A. Gatehouse for his encouragement, support and advice throughout this work and also Professor D. Boulter for his support whenever necessary.

I would also like to thank Dr R.R.D. Croy and Dr A. Shirsat for providing the pea genomic library and for the stimulating discussions. I am extremely grateful to Miss J. Bryden and Mr D. Bown for their help in producing some of the sequences appearing in this thesis.

Special thanks go to Mr Hugo Minney for producing some of the figures in this thesis on the computer, and to Dr M. Evans for making the rape cDNA extensin probe available for my use. Dr Philip Taylor for providing the *in vitro* packaging extracts.

Other members of the Department of Biological Sciences, especially Mr A. Fordham-Skelton for critical reading of parts of this manuscript and Mrs A. Richardson for typing.

Finally, I should like to thank the British Council for providing me with financial support to carry out this study.

ABBREVIATIONS

bp	= base pairs
kb	= kilobase pairs
DNAseI	= deoxyribonuclease
RNAse	= ribonuclease
EtBr	= ethidium bromide
SDS	= sodium dodecyl sulphate
SSC	= saline sodium citrate
PEG	= polyethylene glycol
A_{260}	= absorbance at 260
X-gal	= 5 dibromo-4-chloro-3-indoylgalactoside
IPTG	= isopropylthiogalactoside
5'	= 5' terminal phosphate of a DNA or RNA molecule
3'	= 3' terminal hydroxyl of a DNA or RNA molecule
cDNA	= complementary DNA
BSA	= bovine serum albumin
pfu	= plaque forming unit
daf	= days after flowering
poly(A^+)RNA	= polyadenylated RNA
DS	= Denharts solution
EDTA	= ethylenediaminetetra acetic acid (disodium salt)
EtOH	= ethanol
Ap	= ampicilin
mA	= milliamps
^oC	= degree centigrade
cpm	= counts per minute
hr	= hour
g	= gram
mg	= milligram
ug	= microgram

ng	=	nanogram
l	=	litre
ml	=	millilitre
ul	=	microlitre
u	=	molar
mM	=	millimolar
cm	=	centimetre
mm	=	millimetre
min	=	minute
sec	=	second
v	=	volts
v/v	=	volume per volume
w/v	=	weight per volume
Fig	=	Figure
Var	=	variety
LMP	=	low melting point
Mes	=	(2[N-Morpholino] Ethane Suffonic Acid)

FIGURES

TABLES:

CHAPTER ONE
INTRODUCTION

1.1 GENERAL INTRODUCTION

Legume and cereal seeds comprise about 70% of mankind's food. Animals account for the remaining 30%. The major source of protein in the diet of human beings and animals is the protein present in seeds, with cereals forming the major source of plant protein followed by legumes. However the percentage of protein in legume seeds is 20-50% of the seed dry weight while in cereals it is only 8-15% (Danielsson 1949).

Storage proteins are those proteins which are found in seeds and used as a nitrogen source to establish the seedling upon germination (Basha and Beavers 1975; Thompson *et al.* 1978) and constitute 50-90% of the protein (Shewry *et al.* 1981). On the basis of their solubility storage proteins have been divided into four classes (Osborn 1924); albumin which is soluble in water, globulin which is soluble in dilute salt solutions at neutral pH, prolamin is an alcohol-soluble class, while glutelin is alkali-soluble. Legume seed storage proteins fall into the globulin class, i.e. those soluble in aqueous salt solution. Legume globulins are of two main types, legumin and vicilin. The variation in the relative amounts of these types within the leguminosae is very wide. For example, *Pisum sativum* contains about equal amounts of both (Schroeder 1983), *Phaseolus vulgaris* contains very little legumin (Derbyshire *et al.* 1976), while *Vicia faba* contains an excess of legumin over vicilin (Gatehouse *et al.* 1980). The presence of similar globulin storage proteins in other legumes e.g. *Glycine max, Vigna unguiculata, vigna radiata*, has been reported by Derbyshire *et al.* 1976. An overall similarity of the legumin and vicilin proteins between closely related species of leguminosae has been demonstrated (Croy *et al* 1979).

1.2 STORAGE PROTEIN SYNTHESIS IN PEA

During seed development, the synthesis of legumin starts from the initial stages of embryo development (Domoney *et al.* 1980). It is probable that vicilin is also similarly synthesised (Boulter 1981). However, storage proteins are only present in small amounts until the cell division phase of embryo development has ceased. Cell expansion in embryo development then results in a massive increase in storage protein synthesis until by day 20 of seed development, legumin and vicilin comprise about 60% of the total protein of the mature seed. This is accompanied by both an increase in the amount of rough endoplasmic reticulum (Boulter *et al.* 1972) and an increase in the ratio of polysomes to monosomes in cotyledon cells (Beevers and Poulson 1972).

Storage proteins are laid down in large amounts during the cell expansion phase of seed development so as to be used as a nitrogen supply on germination. Because of the specificity of the synthesis of these storage proteins to the seed (Gatehouse *et al.* 1982), the developing seed provides a good model system for studying developmentally regulated gene expression.

The investigation of mRNAs responsible for the production of legumin has been used to investigate the mode of synthesis of the protein. The sequence complexity of pea seed mRNA has been found to decrease sharply when the synthesis of storage protein is at its highest level (Morton *et al.* 1983). This coincidence between the abundance of certain mRNA classes and the accumulation of legumin makes it likely that the increased rate of legumin synthesis is a direct result of increasing amounts of legumin mRNA. It has been shown that the levels of legumin and vicilin mRNAs during seed development increased and decreased in agreement with estimated rates of synthesis of the respective polypeptide.

Storage protein synthesis in seeds is under strict developmental control. In pea, as well as in other legumes, there is a rapid accumulation of protein during the cell expansion phase of seed development (Boulter, 1981). During this cell expansion phase, the genomes of the cotyledon parenchyma cells undergo extensive endoreplication resulting in increases in nuclear DNA content (Millerd and Spencer, 1974; Smith, 1973).

The seed storage protein precursors of pea are synthesised by polysomes on the endoplasmic reticulum (Chrispeels *et al*, 1982) and are transported to the golgi apparatus and then to the protein bodies via electron-dense vesicles

(Chrispeels, 1984). The polypeptides are assembled into oligomers in the lumen of the endoplasmic reticulum. The glycosylation occurs here (Chrispeels *et al*, 1982). In the protein bodies the post-translational proteolysis of the precursors takes place.

1.3 PEA SEED STORAGE PROTEINS: THEIR STRUCTURE AND GENES

1.3.1 LEGUMIN POLYPEPTIDES AND THEIR ENCODING GENES

The 11S storage protein fraction of peas and broad beans is given the name legumin. Other legumin-type proteins occur in other legumes, e.g. glycinin in soyabean (Neilson, 1984) and in many other plants, including cereals, e.g. glutelin of rice (Zhao *et al.*, 1983). In some cases sequence homologies between the proteins have been demonstrated.

Legumin is a hexameric molecule of Mr approximately 390,000 (Casey, 1979 ; Croy, *et al.*, 1979). Each monomer consists of an acidic subunit (Mr 40,000) covalently linked by disulphide bonds to a basic subunit (Mr 20,000) (Wright and Boulter, 1976, Gilroy *et al.*, 1979). In pea, both the acidic and basic subunits show heterogeneity in charge and molecular weight (Gatehouse *et al.*, 1980). The various subunit pairs have been divided into "major" and "minor" legumin species (Casey *et al.*, 1981, Matta *et al.*, 1981). The authors showed that the Mr 54,000 disulphide bonded subunit pairs constituted the major part in total legumin. However subunit pairs with Mr varying from 35,000 to 58,000 were also present. The subunit pair associated in various ways to give at least three distinct molecular forms of legumin, one of which contained only "major" subunit pairs. The assembly of subunit pairs into hexameric molecular species was not random. Thus, heterogeneity of legumin, at the level of native protein size and net surface charge is a consequence of heterogeneity in its constituent subunits and their polypeptides.

The existence of polypeptide heterogeneity in *Pisum* legumin was firmly established (Thomson *et al.*, 1978), and it was estimated that *Pisum* legumin comprises at least 22 different α and 11 different β-polypeptides. This is almost certainly a consequence of sequence heterogeneity within a small gene family and not due to the presence of covalently bound carbohydrate on some polypeptides (Casey 1979; Gatehouse *et al.*, 1980).

Sequencing of two classes of pea legumin gene (leg A and leg J; see later) has indicated that pea legumins fall into two groups with about 50% sequence homology. Similar studies (Otto *et al.*, 1984) have shown that *Vicia* legumins also fall into two groups (legumins A and B) and that are about 50% homologous. Comparison of these legumin sequences with homologous regions of glycinin has shown that pea leg A and *Vicia* legumin A sequences resemble group I glycinins, whereas Pea leg J and *Vicia* legumin B are more similar to group II glycinins which suggests that genes for the two classes of 11S proteins diverged before speciation of *Pisum, Vicia* and *Glycine*.

The pea genome has been reported to contain several 11S protein loci, and a closely related group of genes is contained in each locus. DNA sequences encoding legumin polypeptides have been isolated from two loci. Leg A type gene has been identified with the locus near r whereas the locus near a contains leg J type genes. From the locus which is near r (Lycett *et al.*, 1984) λleg 1 was isolated, and λJC5 was isolated from one of the two legumin loci that map near a (Domoney *et al.*, 1986). Four EcoRI fragments homologous to legumin cDNA are found in the legumin locus near r. One of these fragments (i.e. the fragment in λLeg 1) contained two sequences designated leg A and leg D (Lycett *et al.*, 1984a; Bown *et al.*, 1985). Similarly λJC5 is believed to contain a pair of directly repeated legumin genes (Domoney *et al.*, 1986) It has been concluded from the evidence available that the legumin locus near r contains four genes and one pseudogene while the other (one of two near a) contains at least three genes.

The r linked genes in pea are: leg A, B, C and D. Sequence analysis of leg A has shown that it contains three introns in analogous positions to those of class I and class II glycinin (Wobus *et al.*, 1985). In leg J, however the sequence contains two introns only (Gatehouse *et al.*, 1988). Near the 5' end of both leg A TATA, AGGA and CAAT box sequences can be found

(TATA at a position -66 and AGGA/CAAT at position -126 of leg A) (Lycett *et al.*, 1984). In the region of the AGGA/CAAT box there is a region of 30 bp which is strongly homologous to the corresponding region of leg J. This region was referred to as "legumin box" which can be also found in the corresponding region of leg D (Bown *et al.*, 1985) leg D differs from leg A by a series of deletion and frameshift errors which destroy intron/exon boundaries, as well as two in-frame stop codons, which suggested that leg D was a pseudogene (Bown *et al.*, 1985).

Sequence analysis of the 5'-flanking regions of leg A, leg B and leg C (Lycett *et al.*, 1985) has revealed that the three sequences contain TATA and AGGA/CAAT boxes, a sequence which is 90% homologous to the "legumin box". Upstream of position 322 the leg A gene begins to diverge from leg B and leg C; both leg B and leg C have a second set of TATA and CAAT boxes which is absent from leg A. However, no direct evidence has suggested leg B and leg C are functional, while leg A has been shown to be functional (Evans *et al.*, 1985).

At the 3' end of leg A and leg J sequences multiple poly (A) addition signal can be found. It has not been determined whether these sequences function as poly (A) addition signals, but it has been suggested that the putative mRNA secondary structure in the region of the second poly (A) addition signal of leg A has a stem-loop structure (Lycett *et al.*, 1984a)

Two genes encoding minor legumins in pea were isolated and characterised (Gatehouse *et al.*, 1988). The two genes which are designated leg J and leg K are arranged in tandem, in the same orientation with leg K 5' to leg J. The whole leg J gene and part of gene leg K were sequenced (Gatehouse *et al.*, 1988).

In addition to 1742 bases of coding sequence, the leg J sequence contains introns, 616 bases of 5' flanking sequence and 611 bases of 3' flanking sequence. The leg K sequence runs from the end of the genomic clone, at a position corresponding to base 503 of the leg J sequence, to a point 3' to the end of the coding sequence and thus includes 273 bases of 3' flanking sequence.

The high homology (97%) between the coding regions of the two genes (i.e. leg J and leg K) strongly suggests that they are of the same gene subfamily (Gatehouse *et al.*, 1988).

A polypeptide precursor of Mr 57024 and a final legumin subunit pair with an α-subunit of Mr 34485 and a β-subunit of Mr 20300 has been predicted by the leg J coding sequence.

It has been confirmed that the leg J is an expressed gene and that the gene subfamily including leg J and leg K encodes polypeptides of the minor legumin type.

Sequence comparison between leg J and leg A has shown a 48% homology between the coding sequences of both genes and a significant homology in amino-acid sequence.

1.3.2 VICILIN POLYPEPTIDES AND THEIR ENCODING GENES

It has been found that the vicilin (7S) fraction of pea seed storage protein is heterogenous (Thompson *et al.*, 1978; Gatehouse *et al.*, 1981). Polysomal RNA isolated from developing pea cotyledons when translated *in vitro* gives only two vicilin polypeptides of 50,000 Mr and 47,000 Mr (Croy *et al.*, 1980). The absence of the 47,000 Mr polypeptide in vicilin isolated from mature seeds has suggested that the 47,000 Mr polypeptide is the precursor of the smaller vicilin polypeptides (Gatehouse *et al.*, 1981) which are produced from this subunit as a result of post-translational proteolysis. Vicilin has been suggested to be a protein which has a trimeric structure consisting of subunits of 50,000 Mr (Gatehouse *et al.*, 1981). A similar trimeric structure has been described for the related 7S protein of *Phaseolus vulgaris*, phaseolin (Puztai and Stewart, 1980).

Vicilin is glycosylated (Derbyshire *et al.*, 1976), unlike legumin which does not contain carbohydrate (Gatehouse *et al.*, 1980).

Two classes of cDNAs which encode vicilin have been isolated, a class of those encoding a 47,000 Mr vicilin precursor (Croy *et al.*, 1982; Lycett *et al.*, 1983), and those encoding 50,000 Mr precursors (Croy *et al.*, 1982; Delauney, 1984).

A complete sequence for a vicilin precursor has been compiled (Lycett *et al.*, 1983). The predicted amino acid sequence shows that there are 15 residues upstream from the N-terminus of the mature protein. The sequence is rich in hydrophobic residues and is thought to be a signal sequence (Lycett *et al*, 1983). The point of initiation of translation (Lycett *et al.*, 1983) is thought to be methionine codon (AUG) at the start of the signal sequence.

A polypeptide of Mr 71,000 which is immuno-precipitable by antivicilin antibodies was reported in addition to the vicilin polypeptides described above (Gatehouse *et al.*, 1981) this polypeptide, unlike vicilin is not glycosylated, it is called convicilin.

Convicilin accumulates later during seed development than vicilin (Croy *et al.*, 1980), Gatehouse *et al.*, 1984). Analysis of the nucleotide sequence of a convicilin cDNA and its predicted amino acid sequence shows it to be partly homologous to vicilin, phaseolin and conglycinin sequences (Casey *et al.*, 1984).

Five vicilin loci, named Vc-1 to Vc-5, and a single convicilin locus, *Cvc*, have been shown by genetic mapping to be found in pea.

It has been estimated that there are 5-7 and 4-6 vicilin genes encoding the 47 and 50 kDa precursors, respectively (Domoney and Casey, 1985). Only Vc-1, Vc-2 and Vc-3 of the five vicilin loci have been mapped (Ellis *et al.*, 1986); they all map close to the r locus on chromosome 7. Ellis *et al.*, (1986) isolated cDNA and genomic clones representing four of the five loci. From locus VC-4, two genes of the class encoding a 50 kDa precursor have been cloned and sequenced. The first, *vic B*, contains an insertion 12 amino acids into exon 6, and is probably not expressed (Levasseur, 1988). The faithful expression of the second gene in transgenic tobacco plants has demonstrated that it is fully functional (Higgins *et al.*, 1988). A *vic C* gene of a class encoding a 50 kDa precursor, but only approximately 85% homologous to *vicB*, and lying at locus Vc-5, has been partially sequenced, and thought to be non functional due to the presence of an in-frame stop codon after only 28 residues (J. Gatehouse, unpublished). Genes that encode the 47 kDa are thought to be contained in locus Vc-2. The *vic J* gene from this locus has been sequenced but it may not be expressed because of an insertion, similar to that of *vic B* near its 3' end (D. Bown and J. Gatehouse, unpublished). A subfamily of at least 4 genes of greater than 95% homology (Higgins *et al.*, 1988) is estimated to be found on the Vc-4 locus. A gene designated *vic A* (Sawyer, 1986), the locus of which is not determined, was isolated and sequenced. This gene was suggested to be functional due to the strong homology with pDUB9 and the absence of deletions, frame-shift error or stop codon within the coding sequence.

The convicilin locus cvc, is thought to contain two genes, in both pea cultivars "Feltham First" and "Dark Skinned Perfection", one of which cvcA from "Dark Skinned Perfection", has been cloned and sequenced (Bown *et al.*, 1988).

1.4 GENOMIC DNA LIBRARIES

The perfect genomic library of a certain organism is a collection of recombinant clones which together contain all (or nearly all) the DNA sequences of the entire genome. Assuming complete representation of the entire genomic sequences, it should be possible to find any desired sequence within the genome. It is important to know how to calculate the number of genomic clones needed to cover the entire genome; this depends on the size of the genome and the size of the insert in the clone. A formula has been derived which allows the calculation of the number of independent recombinant clones required to construct libraries of any level of representation (Clarke and Carbone, 1976).

If x is the insert size and y is the size of genome

$$N = \frac{\ln(1 - p)}{\ln(1 - \frac{x}{y})}$$

clones will have a probability P of containing any particular sequence, and when assuming P = 0.99, the N will be the number of clones needed to give 99% representation.

A recombinant clone is an insert cloned in a vector, which if introduced into a host cell, is capable of replicating to produce genetically identical copies. When needed the insert can be separated from the vector by restriction endonuclease digestion.

The vectors used in genomic library construction have different properties, genotypes and phenotypes, therefore, when a vector is chosen to construct a certain genomic library it should have the properties needed to get best results from the library.

Three types of vectors are used in the construction of genomic libraries; plasmids, cosmids or bacteriophage. Plasmids are extrachromosomal DNA molecules which are capable of autonomous replication and in most cases are stably inherited from one generation to another (only the naturally occurring plasmid like ColE1). Plasmids constructed for cloning purposes have unique restriction enzyme cleavage sites. Plasmid vectors are useful for cloning small fragments of DNA and for preparing genomic libraries from small sized genomes such as *E. coli*. It has been estimated that 1400 recombinants of 13 kb long fragments (Clarke and Carbon, 1976) cloned in to ColE1 is sufficient to create a 99% complete genomic library (Hershfield *et al.* 1974).

Recombinant plasmids of about 15 kb can be introduced (transformed) into host cell at a very high frequency (10^5 per μ g) (Sherratt, 1979). Assuming a transformation frequency of 10^5 per μg of DNA, 1.7×10^6 individual recombinant clones would be required to construct a 99% complete library (Bolivar *et al.*, 1977) of pea in pBR322. Also it has been reported (Graf,

1979) under optimum conditions that only 20-35% of the cloned covalent circular molecules formed by ligating a mixture of plasmid and insert would be the desired hybrid. The rest would consist of recircularized plasmids and inserts and very long plasmids resulting from multiple ligations, therefore about 56 g of DNA would be required to construct a pea genomic library. Because large DNA molecules transform at a lower frequency than small ones that would bring some doubt about the accuracy of such calculations.

The variation in size of different recombinant plasmids could lead to variation in copy number per cell between different recombinants since larger plasmids tend to have lower copy number than smaller plasmids. The growth rates of cells harbouring different recombinants may be affected by the nature and number of genes expressed in the cell. So recombinants with lower copy number and those with fewer foreign genes may quickly take over and become predominant in culture when a genomic library is amplified.

Plasmid vectors like pBR 322, pAT 153, pUC 18 etc, although existing in cells at very high copy number are not as stable as ColEl (the plasmid they were derived from) and plasmid-free cells tend to appear in cultures of cells containing them (Summers *et al.* 1985). All those disadvantages make plasmid vectors less popular and useful in constructing genomic libraries than cosmids and bacteriophage.

As we have seen, the efficiency of transformation of recombinants is not high enough to enable the use of small amounts of DNA to get enough transformants for a representative library of a genome like pea genome. In the case of introducing restricted phage DNA into the host bacterium by transfection of competent bacteria the efficiency is relatively low (10^5 plaques per μg of DNA) and in gene manipulation experiments in which the vector DNA is restricted and then ligated with foreign DNA the frequency is reduced to about 10^4-10^3 plaques per microgram of DNA. In some contexts where 10^6 or more plaques are required this transfection would not be useful to achieve such a figure, but packaging *in vitro* yields about 10^6 plaques per microgram of vector after the ligation reaction.

The principle of packaging *in vitro* is to supply the ligated recombinant DNA with high concentrations of phage head precursor, packaging proteins and phage tails. Practically, this is most efficiently performed in a very concentrated mixed lysate of two induced lysogens, one of which is blocked at the prehead stage by an amber mutation in gene D and therefore accumulates the precursor while the other is prevented from forming any head structure by an amber mutation in gene E (Hohn and Murray, 1977). In the mixed lysate, genetic complementation occurs and exogenous DNA is packaged. Although concatemeric DNA is the substrate for *in vitro* packaging added monomeric DNA could be packaged.

The fact that concatemers of DNA molecules can be efficiently packaged *in vitro* if the cos sites are 37-52 kb apart (Hohn, 1975) has lead to the construction of plasmids with contain a fragment of DNA including the cos site (Collins & Bruning, 1978; Collins & Hohn, 1979). These plasmids have been termed cosmids and can be used in conjunction with the *in vitro* packaging systems to construct genomic libraries.

Cosmid recombinants of size between 37-52 kb are the most likely to be packaged, therefore the insertion 32-47 kb foreign DNA in a cosmid vector of 5 kb would be demanded. This provides an efficient means of cloning large pieces of foreign DNA.

After *in vitro* packaging, the particles are used to infect host cells, DNA is injected and circularizes like phage DNA but replicates as a normal plasmid without the expression of any phage functions. Transformed cells are selected on the basis of a vector drug resistance marker. Because of their capacity for large fragments of DNA, cosmids are particularly attractive vectors for constructing libraries of eukaryotic genomes.

In place of cosmid vectors phage lambda vectors may be chosen. Phage lambda is a genetically complex but very extensively studied virus of *E. coli*.

The DNA of phage lambda is a linear duplex molecule of about 49 kb pairs. Genes concerned with recombination are located in the central region of the genome. Functionally related genes are found in the region left to the central region while genes concerned with regulation and prophage immunity to superinfection, DNA synthesis, late function regulation and host cell lysis are located in the region to the right of the central region.

Wild-type lambda DNA is not suitable as a vector due to the presence of several sites of the commonly used restriction endonucleases.

Therefore, derivatives of the wild-type phage were constructed, some have a single site at which foreign DNA is inserted (insertional vectors), or have a pair of sites defining a fragment which can be removed and replaced by foreign DNA (replacement vectors). Many vector derivatives of both the insertional and replacement type, have been constructed (Thomas *et al.* 1974; Murray & Murray, 1975; Blattner *et al.* 1977; Leder *et al.* 1977). Here we shall be concerned only with the class of vectors known as replacement vectors.

Different replacement vectors have the same characteristics and yet differ in some others. In the replacement vector, λgt.λc (Thomas *et al.* 1974), the replacement fragment lies between two *Eco*RI sites. It contains *att, int* and *xis* and gives the phage the capability of forming stable lysogens. When the fragment is replaced by foreign DNA, however, the phage becomes integration defective. The replacement vector NM 781, has a replaceable *Eco*RI fragment which carries the gene, supE, for a mutant tRNA of *E. coli*. It suppresses an amber mutation in the *lac*Z gene of the bacterial host which enables the phage to produce red plaques on lactose/MacConkey agar or blue plaques on agar containing x-gal. Recombinant phage give colourless plaques on both indicators. In NM 762 a *sup*F gene is carried on a Hind III fragment.

Another important class of vectors developed by Karn *et al.* (1980), and Loenen & Brammar (1980), employ the spi$^-$ phenotype as a selection process.

Wild-type lambda cannot grow on *E. coli* strains lysogenic for phage P2, they display the spi$^+$ phenotype-sensitivity to P2 inhibition (Lindahl *et al.* 1970). This sensitivity is due in part to the products of the Lambda genes exo, bet and gam which are situated within the replaceable region of the vector. Replacement of those genes by foreign DNA thus allows the growth of the recombinant in a P2 lysogen.

Whether using bacteriophage or cosmid vector, random cleavage of the eukaryotic DNA must be achieved so as not to exclude any particular class of fragment which would then be under represented in the library. A restriction endonuclease such as Sau3A which recognises a tetra nucleotide sequence (Roberts, 1981) is a suitable choice, since this sequence should be present, assuming a random distribution, every 256 nucleotides. This enzyme also has the advantage that its recognition sequence is contained within the BamHI recognition sequence.

The replacement vector EMBL3 (Frischauf *et al.* 1983) which has a convenient polylinker flanking the replaceable fragment and confers the spi$^-$ phenotype has been used by the author to construct a genomic library of *Arabidopsis thaliana*.

1.5 *ARABIDOPSIS THALIANA*

1.5.1 SPECIAL FEATURES

Arabidopsis is a flowering small weed in the mustard family which has been the subject for study in classical genetics for over 40 years. *Arabidopsis* has been reported in many different regions and climates. Although it has no exact geographic origin it is believed to be native to the old world.

The plant is a harmless weed of no food or economic value which has its own special features. It has a generation time of only 5 weeks and can produce more than 10,000 seeds per plant, and it is so small in size that tens of thousands can be grown in a small space. A rapid growth of the plant can be achieved under continuous light, at 25^{o}C with good nutrition. *Arabidopsis* grows well in soil, and can also be grown in sterile nutrient agar (Langridge, 1957), or floating in liquid medium (Redei & Perry, 1971). All these properties along with the ability to self or cross fertilization make the plant a convenient subject for studies in classical genetics.

1.5.2 THE GENOME AND ORGANISATION

The haploid chromosome number was established as 5 (Laibach, 1907). Recent reports have indicated a haploid nuclear genome size of 70,000 kb. These reports are consistent with previous ones which suggested that *Arabidopsis* has the smallest genome among flowering plants (Sparrow *et al.* 1972), which is only 15 times the size of *Escherichia coli* genome (Daniels & Blattner, 1987).

About 1,000,000 random lambda clones of 20 kb average insert size must be screened to have a 99% chance of selecting a desired fragment of pea genome from a pea genomic library, and 370,000 clones from tobacco genomic library, while only 16,000 clones from *Arabidopsis* genomic library would have to be screened which indicates that screening and construction of *Arabidopsis* genomic libraries is relatively simple and economical due to its small genome.

Furthermore, *Arabidopsis* has a unique genomic organisation. Its nuclear genome has a very low content of repeated sequences (Pruitt & Meyerowitz, 1986), where the repeated or foldback DNA comprise about 10-15% of the genome, 7.5% is composed of tandem repeats of ribosomal RNA coding sequences, and as little as 1% may be dispersed repeats. Some of the characterized highly repeated DNA consists of 4,000-6,000 repeats of a 180 base pair sequence in one or more long tandem arrays (Martinez-Zapater & Estelle, 1986).

Arabidopsis genome is convenient to use to perform chromosome-walking experiments owing to the fact that it is nearly devoid of dispersed repeats and those elements that do exist are usually very far from each other, which makes *Arabidopsis* quite different from other angiosperms for which similar information is available.

These peculiar characteristics make *Arabidopsis* genome a very useful tool in molecular genetics such as the study of multigene families and the DNA sequence analysis of individual genes. By taking advantage of wide cross-hybridization of plant genes, genes cloned from the small *Arabidopsis* genome might be used as probes for the isolation of homologous sequences from plants of economic importance.

1.6 CELL WALL EXTENSINS

1.6.1 PLANT CELL WALL

Cell wall structure plays an important role in plant morphogenesis. The cell wall provides a protective barrier for plant cells; it also performs an effective strategy for disease resistance in plants.

The established theory that cell walls were composed entirely of carbohydrate, was contradicted by the observation that cell walls contain most of the hydroxyproline in plant cells. Recently, the view that cell walls contain important structural proteoglycins became generally accepted by the advances in characterisation of one class of cell wall hydroxyproline-rich polypeptides, the extensins. At least three classes of hydroxyproline-rich glycoproteins exist in plants (McNeil *et al.*, 1984). These are the extensins, the arabinogalactins and the solanaceae lectins. Other classes of polypeptides have been described in plant cell walls. A large number of enzymes have been reported to be present in plant cell walls (Lamport and Caat, 1981). Recently, a novel class of glycine-rich structural cell wall proteins has been reported (Condit *et al.*, 1986; Keller *et al.*, 1988; Keller *et al.*, 1989).

1.6.2 EXTENSINS

Lamport named the hydroxyproline-rich cell wall glycoprotein extensin in 1963 and began to characterise fragments of this insoluble wall polymer from suspension culture of cell walls. It was clearly indicated from biochemical analysis that extensin was an unusual glycopolypeptide with an extremely biased amino acid composition (Hyp, Ser, Tyr, Lys and Val) and a simple pattern of glycosylation (Ser-0-Gal1-3 and Hyp-0-Ara3-4). Five hydroxyproline-rich peptides from tomato which account for about 35% of the cell wall hydroxyproline were solubilized and characterised by mild acid hydrolysis, followed by trypsin digestion (Lamport, 1977). It was indicated that extensin has a repeating polypeptide structure because all five peptides contained copies of the pentapeptides Ser-Hyp-Hyp-Hyp-Hyp. It was found that two of the peptides also contained diphenyl ether-linked isodityrosine (Fry, 1982), as an intra-molecular crosslink (Epstein and Lamport, 1984).

A sophisticated picture of the structure of this insoluble cell wall polymer system has recently been provided. This picture has shown that the polymer was a covalently cross-linked three-dimensional network formed by the inter- and intra-molecular cross-linking of a family of soluble extensin monomers. This view was based on the aberration that several soluble hydroxyproline-rich glycoproteins from several plant tissues were found to resemble the insoluble tomato glycopeptides, the soluble glycoproteins were slowly insolubilised following secretion into the cell wall. Also, monomeric and polymeric extensins were directly visualised by electron microscope.

Large amounts of hydroxyproline are deposited in the cell wall of carrot roots following slicing and aeration (Chrispeels, 1969). Such observation has helped in the elucidation of the structure and biosynthesis of extensin. In wounded carrot root cells, extensin was found to be synthesised by the sequential translation of extensin mRNA on rough endoplasmic reticulum, hydroxylation of peptidylproline by a propylhydroxylase, glycosylation of Hyp by oligo-arabinosides and Ser by galactose in the Golgi apparatus, and secretion into the cell wall (Chrispeels, 1974). Hydroxyproline-rich glycoprotein from wounded carrot root was subsequently purified and characterised (Stuart and Varner, 1980; Van Holst and Varner, 1984). Most of the hydroxyproline was discovered to accumulate in a single soluble 80 kilodalton glycoprotein containing about one-third protein and two-thirds carbohydrate. The 33 kilodalton hydroxyproline-rich polypeptide has a biased amino acid composition (Hyp, Ser, Tyr, Lys, His and Val), and is heavily glycosylated with short tri- and tetra-arabinosides 0-linked to hydroxyproline, and small amounts of galactose.

Circular dichroism used in studying the secondary structure of this soluble carrot extensin has revealed that 100% of the peptide bonds are in the polyproline II helical conformation. It was indicated from electron micrographs of the native glycoprotein that it appears as an 80 nm long rod with kinks along its length (Van Holst and Varner, 1984; Staftrom and Stachelin, 1986). The hydrogen bonding is thought to stabilise the single polypeptide proline II helix (Van Holst and Varner, 1984).

It has been reported that soluble extensins have been isolated from a number of plant species and tissues, e.g. potato tubers (Leach *et al.*, 1982) and tobacco callus (Mellon and Helgeson, 1982). These extensins were called bacterial agglutinins and have been indicated to be identical in composition to carrot extensin. Two different types of soluble extensins have been isolated and characterised from tomato (Smith *et al.*, 1984, 1986).

Soluble extensins in both carrot and tomato have been reported to be slowly insolubilised following secretion into cell walls, having soluble half lives of about 12 hours (Cooper and Varner, 1983; Smith *et al.*, 1984). It is thought that isodityrosine, synthesised in the cell wall during insolubilisation (Cooper and Varner, 1983) forms both intra- and inter-molecular cross-links which serve to entangle the extensin monomers with cell wall polysaccharides and covalently cross-link the monomers to each other to form a three-dimensional glycoprotein network. It has been demonstrated that soluble extensin dimers, trimers, tetramers and higher oligomers exist and form an open network structure by forming intermolecular crosslinks primarily between the ends of glycoprotein rods (Staftsrom and Staehelin, 1986). The observation that the two different tomato extensins are both incorporated into the insoluble network with comparable kinetics has suggested that plant cells can construct different types of extensin networks, depending upon which types of extensin monomers are secreted into the wall.

1.6.3 EXTENSIN GENES

Early attempts to identify extensin mRNA in wounded carrot root tissues utilized strategies based on the extensins repeating Ser, Hyp, Hyp, Hyp, Hyp peptide sequence. An attempt to immunoprecipitate an extensin precursor from the *in vitro* translation products of wounded carrot root mRNA by raising polyclonal antibodies to the synthetic peptide Ser-Pro-Pro-Pro-Pro was unsuccessful (Smith, 1981). This was explained by the fact that the proline codon is CCX and the extensin mRNA should contain repeats of CCX CCX CCX CCX and thus might hybridise to an oligo-dG cellulose column (Stuart *et al.*, 1982). A hybrid-selected mRNA was obtained by this method which translated into a putative extension precursor, a 55 kilodalton proline-rich, leucine-poor peptide. Regretably, the yield of putative extensin mRNA obtained by this approach was too low to permit cDNA cloning.

It has been reported that an extensin cDNA clone has been isolated (Chen and Varner, 1985a) from a cDNA library prepared from size fractionated, wounded carrot mRNA, enriched in the mRNA encoding a putative extensin polypeptide precursor and the clone was designated pDC5. It has subsequently hybrid-selected a mRNA encoding a 33 kilodalton proline-rich, leucine-poor polypeptide. An open reading frame of 462 nucleotides encoding a polypeptide with two distinct domains was revealed from the DNA sequence of the 370 bp cDNA insert of pDC5. The C-terminal domain contained four repeats of the canonical extension pentapeptide Ser-Pro-Pro-Pro-Pro, while the N-terminal portion of the polypeptide contained nine repeats of a different proline-rich pentapeptide, His-Lys-Pro-Pro-Val/Ile (Chen and Varner, 1985a). Several clones were isolated using pDC5 as a screening probe and subsequently characterised by sequence analysis. The clones varied in the extent and location of homologous sequence to the pDC5 sequence. Two of the clones were homologous only to the 5' half of pDC5 and one clone was homologous only to the 3' half of the clone. A genomic clone, pDC5A1, which was isolated by hybridisation to pDC5, has shown perfect homology with the 3' region of pDC5, except for a putative intron in the 200 bp 3' nontranslated region, and a complete diversion from the sequence of the 5' region of pDC5 (Chen and Varner, 1985b).

A tomato genomic library was screened for extensin genes by using pDC5 as a probe, and a genomic clone pTom5 was isolated. 36 Ser-Pro-Pro-Pro-Pro repeats were found to be contained within 1,100 bp open reading frame in the clone (Showalter *et al.*, 1985).

A family of cross-hybridising sequences isolated from a cDNA library of *Brassica napus* L. roots (Evans *et al.*, 1990), has been shown to encode proteins homologous to carrot (Chen and Varner, 1985a) and tomato extensins (Lamport, 1977). Subsequently, a sub-clone designated pRR_t566 containing sequence from the family was used to screen an oilseed rape (*Brassica napus* L.) genomic library and a gene designated extA was isolated (Gatehouse *et al.*, 1990), and found to encode a protein homologous to carrot and tomato extensins.

Sequences homologous to an extensin genomic clone pDC5A1 have been reported to be isolated from a number of plant species including, potato, tobacco, petunia, melon bean, pea and sunflower.

CHAPTER TWO

MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 GLASSWARE AND PLASTICWARE

All glassware and plasticware was autoclaved and siliconised before being used for handling nucleic acid samples.

2.1.2 CHEMICAL AND REAGENTS

Reagents, unless otherwise indicated, were obtained from BDH Chemicals Ltd., Poole, Dorset, UK and were of analytical grade of the best available. The following materials were purchased from the designated sources.

Pronase, protease K, lysozyme, ampicillin, RNase A, EtBr, and ATP were from Sigma Chemical Co., Poole, Dorset, UK.

3MM were from Whatman Ltd., Maidstone, Kent, UK. Nitrocellulose filters were from Schliecher and Schuell GmbH., Dassel, W. Germany.

Disposable nappies were from Boots (UK) PLC, Nottingham, UK.

7X detergent was from Flow Laboratories, Rickmansworth, Herts. UK.

Sephadex G-50 and Ficoll-400 were from Pharmacia Fine Chemicals, Uppsala, Sweden. Agarose, low melting point agarose and high gelling temp agarose were from BRL Ltd., Uxbridge, Middlesex, UK.

Triton X-100 was from Koch-Light Ltd., Colnbrook, Berks, UK.

Bacto agar and bacto tryptone were from Difco Laboratories, Detroit, Michigan, USA. Yeast extract was from Biolife S.r.l., Milan, Italy.

Radiolabelled chemical and nick translation kit were from Amersham International Plc, Amersham, Oxon, UK. Restriction endonucleases and DNA modifying enzymes were from Boehringer Manheim GmbH, Manheim, W. Germany, New England Biolabs, Beverly, M.A., USA and Northumberland Biochemicals Ltd, Cramlington, UK. C5C1, X-gal and IPTG were also from Boehringer Manheim. Fuji R x 100 x-ray film was from Fuji Phot Co. Japan.

2.1.3 BACTERIAL STRAINS, PLASMIDS AND BACTERIOPHAGE VECTORS

Bacterial strains were derivations of *E. coli* K12. Table 1 lists these strains, plasmids and bacteriophage used as vectors or probes and the source or references for each.

TABLE 1

Bacterial strain	Genotype	Reference or source
K803	rk^{-} mk^{-} gal^{-} Mel^{-}	Wood, W.B. (1966)
Q358	rk^{-} mk^{+} 80^{R} $SuII^{+}$	Karn *et al.* (1980)
Q359	rk^{-} $^{m}k^{+}$ 80^{R} $SuII^{+}$	Karn *et al.* (1980)
DH5	F^{-}, endA1, sup E44, Thi-1	
	$^{-}$, recA1, gyrA96, relA1	
	080 dlac Z M15	
Plasmid		
pUC18	Ap^{R} Lac 2	
pBR322	Ap^{R} Tc^{R}	Bolivar *et al.* (1977)
Bacteriophages		
EMBL3	spi^{-} trpE	Frischauf *et al.* (1983)

2.1.4 BACTERIAL CULTURE MEDIA AND ANTIBIOTICS

The following media were used for the growth of bacterial cultures:

L-broth: 10g/l bactotryptone, 5g/l yeast extract, 5g/l NaCl, pH to 7.0 with NaOh, 10 ml 20% sterile glucose solution added after autoclaving (Miller, 1972).

2 x YT 6g/l bactotryptone, 10g/l yeast extract, 5g/l NaCl (Miller 1972).

For solid media 15g/l agar was added before autoclaving. After autoclaving, (121°C, 15 psi) and the media cooled to 55°C before antibiotic (ampicillin at 50 µg/ml). For the detection of functional -galactosidase in *E. coli*, agar was supplemented with 40 µg/ml x-gal and 0.1 mM IPTG.

2.2 METHODS

2.2.1 BIOCHEMICAL TECHNIQUES

2.2.1.1 STORAGE OF BACTERIAL CELLS

Bacterial cultures were stored at 4^{o} for up to 6 weeks on inverted agar plates sealed with Nescofilm. For long term storage, bacterial lawn from single colonies on selective agar plates were transferred to sterile 2ml aliquots of a solution containing 60% L broth and 40% glycerol, mixed thoroughly by vortexing and stored at $-80^{o}C$.

2.2.1.2 ALCOHOL PRECIPITATION OF DNA

0.1 volume of 3M sodium acetate or potassium acetate pH 5.2 (only when needed) and 2.0 volumes of ethanol were added to DNA solution and kept at -70 for 30-60 min or at -20 for 1 hr - overnight. The precipated DNA was pelleted by centrifugation at 12,000g for 10 min for small samples, or at 25,000mg for 10 min for large samples. The pellet was washed with 70% v/v ethanol, dried briefly under vacuum and resuspended in sterile water of TE buffer.

2.2.1.3 DEPROTEINISATION OF DNA SAMPLES

Solutions of DNA were deproteinised by two extractions with phenol, where an equal volume of phenol was added the DNA sample and mixed by vortexing. The aqueous phase and organic phase were separated by a brief centrifugation in a microfuge. Phenol extractions were followed by two extractions with an equal volume of chloroform - isoamyl alcohol (24:1, v/v) to remove the remaining traces of phenol. After deproteinisation, DNA was recovered by alcohol precipitation.

2.2.1.4 SPECTROPHOTOMETRIC ANALYSIS OF NUCLEIC ACID SOLUTIONS

The optical densities (OD) of nucleic acid solution in 1 cm quartz cells were recorded from 320 to 230 nm in a Pye Unicam SP8-150 UV/VIS spectrophotometer operated in the scanning mode.

An OD260 of 0.02 corresponds to a DNA concentration of 1 μg/ml. An OD260 of 0.024 corresponds to an RNA concentration of 1 μg/ml.

2.2.2 NUCLEIC ACID ISOLATION

2.2.2.1 LARGE-SCALE PREPARATION OF BACTERIOPHAGE DNA

Bacterial cells (e.g. P2392) were grown in 10 ml L Broth with 0.5% maltose, at 37°C with shaking for 4 hrs. Host cells (0.2 ml) were mixed with 10^6-10^7 phage particles and incubated at room temperature for 20 min. The mixture was then added to 400 ml L Broth (no glucose or maltose) containing 10 mM MgCl2 in a baffled flask and slowly shaken at 37°C overnight. Cells and debris were consequently pelleted by centrifugation (15,000g, 10 min at 4°C), and the supernatant was removed to a fresh flask. DNase was added to 5 µg/ml and the mixture was incubated at 37°C for 30 min. Phage particles were precipitated by adding and slowly dissolving PEG 6000 to 10% w/v, followed by incubation on ice for 2 hrs. The precipitate was collected by centrifugation, and, after discarding the supernatant, the phage pellet was resuspended in 6 ml SM buffer in a 15 ml glass centrifuge tube. An equal volume of chloroform was added and the mixtures was vortex mixed for 30 sec. The aqueous phase was subjected to phenol, phenol-chloroform and chloroform extractions. The phage DNA was then ethanol precipitated, collected by centrifugation (12,000g, 10 min) and resuspended in 200-300 µl of TE buffer containing RNAase at 50 µg/ml.

2.2.2.2 SMALL-SCALE PREPARATION OF PLASMID DNA

This was essentially the method of Brinboim and Doly (1979). 5 ml aliquots of YT containing 50 µg/ml ampicillin were inoculated with a single bacterial colony and grown with shaking overnight at 37°C. 1.5 ml of cells were centrifuged in the microfuge for 30 sec, the supernatant removed and tube drained onto absorbent tissue. 150 µl of solution I (50 mM glucose, 10 mM EDTA, 25 mM Tris/HCl pH 8.0, 4 mg/ml lysozyme) was added, tubes vortexed and kept on ice for 20 min. 200 µl of solution II (0.2 N NaOH, 1% SDS) was added, mixed gently and kept on ice for 5 min. 150 µl of 5M sodium acetate pH 4.8 was added, tubes inverted to mix and left on ice for 5 min. Bacterial DNA and debris were pelleted by centrifugation for 5 min. Supernatant was removed into a new tube and extracted with phenol or acid-phenol for DNAs intended for sequencing followed by two chloroform extractions, then DNA was precipitated with ethanol, left at -20°C for 30 min, DNA was harvested by centrifugation for 10 min on microfuge, washed once or twice with cold 70% ethanol, harvested and pellet was dried briefly under vacuum, then resuspended in 25 µl of TE buffer.

2.2.2.3 LARGE-SCALE PREPARATION OF PLASMID DNA

A single colony was inoculated into 100 ml of L broth with the appropriate antibiotic, incubated with shaking overnight at 37°C. Cells were harvested by centrifugation at 12,000g, 4°C for 10 min. The cell pellet was resuspended in 3.3 ml cold 25% sucrose, 0.25M Tris/HCl, pH 8.0. 0.67 ml of fresh lysozyme solution, made up at 10 mg/ml in 0.25 M Tris/HCl, pH 8.0, was added and the cell suspension swirled frequently, on ice, for 10 min. 1.3 ml of 0.25 M EDTA, pH 8.0 was added and swirled again, while on ice, for 5 min. 5.3 ml lytic mixture (2% Triton X-100, 0.05 M Tris pH 8.0, 0.05 M EDTA pH 8.0) was added gently to the cell suspension, the lysate was swirled gently to ensure thorough mixing. The cells were judged to have lysed when the mixture was clear and viscous, this usually took about 5 min. The lysate was centrifuged at 43,000 g, 4°C for 20 min to pellet the chromosomal DNA and membranous material. The supernatant, containing mainly plasmid DNA, was carefully decanted and phenol extracted twice or three times, chloroform extracted twice and traces of phenol and chloroform were removed by ether extraction. The clean aqueous phase was topped up by two volumes of ethanol and placed at -20°C for 30 min, DNA was pelleted by centrifugation at 12,000g, 4°C for 15 min, washed twice with cold 70% ethanol. DNA pellet was dried under vacuum and resuspended in 500 µl of TE buffer with 50 µg/ml RNAase.

2.2.2.4 SINGLE COLONY CLEARED LYSATE

This technique provides a quick method for analysing the total DNA content of a particular clone. The single colonies (isolated from transformation), were patched onto selective plates. Clumps of cells were resuspended in 100 µl SCFSB (2.5% Ficoll (w/v), 0.5% SDS (w/v), 0.06% bromophenol blue (w/v), 0.06 orange G (w/v), made with buffer E). These suspensions were left at 42°C for 15 min. The lysates were spun in a microfuge for 15 min: 50 µl of supernatant was loaded directly onto agarose gels and electrophoresed at 30V overnight.

2.2.2.5 ISOLATION OF DNA FROM *ARABIDOPSIS THALIANA* TISSUES

This was a modified method based on that of Ryan (Pers. comm.) to allow the isolation of reasonable amounts of DNA from a small amount of tissue and at the same time clean enough to be used in manipulation techniques.

About 10g of leaf tissue was ground to a fine powder under liquid nitrogen, ground tissue was dissolved in 60 ml of extraction buffer (100 mM Tris HCl pH 8.0, 50 mM EDTA pH 8.0, 500 mM NaCl, 10 mM 2- Mercaptoethanol), 4 ml of 20% SDS was added, the mixture was incubated at 65°C for 30 min. Then it was incubated for 20 min on ice after the addition of 20 ml 5M/3M KAoC. Debris was removed by centrifugation at 10,000g for 10 min, supernatant was deproteinised by phenol extraction and traces of phenol were removed by chloroform extraction. The aqueous phase was filtered through a miracloth filter into 55 ml of isopropanol, mixed and incubated at -20 for 1 h. Crude DNA was harvested by centrifugation at 10,000g for 10 min. The tube was drained onto an absorbent tissue. Pellet was dissolved in 5 ml of TE buffer, RNAase was added at a final concentration of 100 µg/ml and incubated at 37°C for 30 min. Equal volume of phenol was added and vortex mixed for 30 sec, aqueous phase was removed into a fresh tube after centrifugation at 10,000g for 5 min, this was followed by 2 chloroform extractions and ether extraction to remove traces of phenol and chloroform, DNA was ethanol precipitated and dried briefly under vacuum and finally resuspended in 500 µl of TE buffer.

2.2.2.6 TOTAL RNA EXTRACTION FROM *ARABIDOPSIS THALIANA* TISSUES

Five grams of tissue was ground to a fine powder under liquid nitrogen and transferred to a precooled centrifuge tube, 10 ml of extraction butter was added and mixed at room temperature, the mixture was centrifuged at 10,000g,

4^oC for 10 min. Supernatant was transferred to a fresh tube, equal volume of phenol/chloroform/isoamyl alcohol (24:24:1 v/v was added and mixed by gentle inversion, the aqueous phase was separated from the organic phase by centrifugation at 10,000g for 45 min, and then transferred to a fresh tube 0.7 vol. EtOH and 0.2 vol. 1M AcOH were added and mixed, the mixture was stored at -20^oC overnight. The RNA was recovered by centrifugation at 10,000g for 10 min, the RNA pellet was washed twice with 3M AcONa pH 5.2 centrifuging between and after the washes at 10,000g for 5 min. A final wash with 70% EtOH was carried out and RNA precipitate was dried down and dissolved in sterile redistilled water.

Extraction buffer: 8M guanidine hydrochloride, 20 mM MES, 20 mM EDTA, 50 mM 2-Mercaptoethanol pH 7.0.

2.2.3 ENZYMATIC REACTIONS USED ROUTINELY IN DNA MANIPULATION

2.2.3.1 RESTRICTION WITH ENDONUCLEOYTIC ENZYMES

DNA molecules were digested with restriction endonucleases, each in the buffer supplied with the enzyme by the manufacturers. Only when no buffer was supplied one of the three buffers recommended by Maniatis *et al.* (1982) was used. The buffers are shown in table 2.

Table 2 Endonuclease Digestion Buffers

Buffer	Components (mM)			
	Tris/HCl pH 7.5	Mg Cl2	DTT	NaCl
low salt	10	10	1.0	-
medium salt	10	10	1.0	50
high salt	50	10	1.0	100

Generally the enzymes were used at a concentration of 2-5 U/µg DNA and incubated at the temperature recommended by the manufacturers for 1-3 h. Many of the enzymes have been shown to work adequately at different NaCl concentrations and hence multiple digestions could usually be performed simultaneously in the same buffer.

2.2.3.2 5' DEPHOSPHORYLATION OF DNA USING ALKALINE PHOSPHATASE

The 5' phosphate groups of DNA molecules were removed by treatment with calf intestine alkaline phosphatase in 50 mM Tris-HCl pH 9.0, 1 mM MgCl2, 0.1 mM Zn Cl2 and 1 mM spermidine (Maniatis *et al.* 1982). For fragments with protruding 5' ends, the reaction mixture was incubated for 1 hr at 37°C with 0.2 U enzyme/µg of DNA. To dephosphorylate blunt ended molecules, the reaction was incubated for 15 min periods first at 37°C, then 65°C. A second aliquot of phosphatase was then added and the incubation at both temperatures repeated.

2.2.3.3 DNA LIGATION

DNA molecules with compatible, protruding ends or blunt ends were covalently joined by treatment with T4 DNA ligase in a minimal volume of ligase buffer (20 mM Tris HCl pH 7.6, 10 mM Mg Cl2 10 mM DTT, 0.6 mM ATP). Cohesive termini were ligated at 4°C for 12-16 h using 1U enzyme/µg DNA. Blunt ended molecules were ligated at 4°C or room temperature for 12-16 h using 2U enzyme/µg DNA.

2.2.4 PREPARATION OF COMPETENT CELLS

2.2.4.1 T.S.S. METHOD

Cells were grown in LB broth to the early exponential phase (OD600 0.3-0.4) and pelleted by centrifugation at 1000g for 5 min at room temperature and resuspended in one-tenth of the original volume in ice-cold transformation and storage solution [T.S.S. which consists of LB broth with 20% (wt/vol) PEG 6000, 5% (vol/vol DU50 (DMSO), and 20-50 mM Mg^{2+} (MgSO4 or MgCl2), at a final pH of 6.5]. A 0.2 ml aliquot was used per transformation reaction (Chung et al., 1989).

2.2.4.2 CaCl2 METHOD

100 ml of YT broth was inoculated with 1 ml of overnight culture of desired cells and shaken at 37°C until an OD600 of 0.3-0.4 was reached. Cells were centrifuged at 4,000g for 4 min at 4°C and the pellets were resuspended in 0.5 volumes of 0.05M ice-cold CaCl2. Tubes were left on ice

for 10 min and centrifuged at 4,000g for 20 min at 4^{o}C. Pellets were resuspended in 1/15 of the original culture volume of ice cold 0.05 U $CaCl_2$ and left on ice for at least 30 min prior to use.

2.2.5 INTRODUCTION OF DNA INTO BACTERIAL CELLS

0.25 of the volume of the ligation mix was added to 200 µl of competent cells and incubated on ice for 30 min. Cells were heat shocked at 37^{o}C for 45 sec. When using competent cells prepared by T.S.S. method no heat shock was needed. 0.5 ml of YT broth was added and incubated with shaking at 37^{o}C for 1 h. 100 µl of cells were plated out onto YT X-gal/amp plates.

2.2.6 GEL ELECTROPHORESIS

2.2.6.1 AGAROSE GEL ELECTROPHORESIS OF DNA

The correct amount of agarose depending on the concentration of the gel was added to Alec's gel buffer (40 mM Tris-acetate pH 7.7, 2 mM EDTA) and boiled until the agarose dissolved. The solution was cooled to 50-60^{o}C and EtBr added to a final concentration of 0.5 µg/ml. The solution was allowed to set in a Perspex mould (190 x 150 mm) adhered to a glass plate using silicone grease, and containing a suitable well forming comb. The formed gel was transferred to an electrophoresis tank and Alec's gel buffer containing 0.5 µg/ml EtBr to a level 1-2 ml above the surface of the gel, the DNA samples, containing 20% loading buffer (0.25 w/v bromophenol blue, 0.25% w/v xylene cynol, 30% v/v glycerol, 10 mM EDTA) were loaded into the wells and electrophoresis performed usually at 50 mA (30 V) overnight. The DNA was visualised under 330 mm UV light and photographed with a polaroid MP-4 land camera through a Kodak 23A Wratlan filter, using Polaroid type 667 film.

Minigels were used to estimate DNA concentration and to monitor progress of reactions. These were essentially the same as the gels already described except 100 x 80 x 5 mm gel moulds and Tris-borate electrophoresis buffer (0.89 M Tris borate, 0.089 M boric Acid, 2 mM EDTA).

2.2.6.2 AGAROSE GEL ELECTROPHORESIS OF RNA

The method was that of Miller (1987). Electrophoresis was performed on a 1%, 180 x 150 mm agarose gel prepared by mixing 0.9g high gelling temperature agarose with 9 ml of 10 x MOPS EDTA buffer (0.5M MOPS pH 7.0, 0.01M EDTA Ph 7.5 and 65 ml H20. After dissolving the agarose, the solution was allowed to cool to 60°C and 17 ml of 37% formaldehyde was added, mixed, poured and left to set at least for 1h at room temperature. The gel was pre-electrophoresed at 60 V for 30 min. RNA samples and DNA size marker were prepared by adding 2.2 µl buffer A (294 µl 10 x MOPS EDTA buffer 706 µl H20) to dry samples. Then 4.8 µl formaldehyde/formamide (final concentration of 2.2 M formaldehyde and 50% formamide). Samples were heated at 70°C for 10 min and then placed on ice for 5 min. After the addition of 1.5 µl gel loading buffer (322 µl buffer A, 5 mg xylene cyanol, 5 mg bromocreosol green, 400 mg sucrose, 17.8 µl 37% formaldehyde, 500 µl formamide), samples were loaded and electrophoresed at 100 V for 2-3h.

2.2.7 RECOVERY OF DNA FROM AGAROSE GELS BY ELECTROELUTION

The DNA was digested with the appropriate restriction endonucleases, electrophoresed through an agarose gel and the required fragment cut out from the gel, trimming off excess agarose. The gel slice was placed in a piece of dialysis tubing secured at one end, 0.5 ml of Tris/acetate electrophoresis buffer was added, and the open end of the tubing closed, excluding air bubbles. The tubing was placed in an electrophoresis tank containing Tris/acetate buffer and electrophoresed at 60 V for about 30 min until the DNA was visible under UV illumination, as a thin line on the tubing. The current was reversed for 30 sec and the buffer removed from the tubing and placed in a 1.5 ml eppendorf. The buffer was phenol, chloroform extracted and the volume reduced by precipitation with ethanol.

2.2.8 RECOVERY OF DNA FROM LOW MELTING POINT AGAROSE GELS

A slice containing the required fragment of DNA was cut out and placed in a 1.5 ml eppendorf. The tube was incubated in a water-bath at 65°C for 5 min. Two volumes of L.M.P. buffer (5 mM Tris HCl pH 8.0, 0.5 mM EDTA) were added, thoroughly mixed, and the mixture was incubated for a further 5 min at 65°C, at the end of the incubation an equal volume of phenol was added at room temperature mixed and the aqueous phase was recovered by centrifugation in a microfuge for 5 min, the phenol extraction was repeated and followed by two chloroform extractions. DNA was ethanol precipitated and then resuspended either in TE buffer or in sterile distilled water.

2.2.9 SOUTHERN TRANSFER OF DNA TO NITROCELLULOSE FILTERS

After photography, the DNA was depurinated by agitating the gel for 10 min in 1% HCl. The gel was transferred to denaturation solution (1.5 M NaOH, 1 mM EDTA) and agitated for 45 min. The gel was then agitated for 45 min in neutralisation buffer (3.0 M NaCl, 0.5 M Tris-HCl pH 7.0 1 mM EDTA) with one change of buffer and rinsed in 10 x S.S.C. (3.0 M NaCl, 0.3 Na-citrate pH 7.0) Southern (1975). The gel was transferred to a capillary blotting apparatus consisting of a tray containing a glass plate overlaid with 3 layers of 3 MM paper and filled to a level a few cm below the top of the glass plate with 10 x S.S.C.. Nitrocellulose paper (0.45 nm pore size, pre-soaked in 10 x S.S.C.) was placed over the gel. The nitrocellulose filter was overlaid with 3 layers of 3 MM paper followed by 4 layers of disposable nappies and a 1 kg weight was placed on top of the nappies. Transfer of DNA was continued for 6-20 h at 4°C or room temperature. After transfer, the position of the gel loading wells were marked on the filter and the filter air dried and baked at 80°C under vacuum for 1h.

2.2.10 TRANSFER OF RNA TO NITROCELLULOSE FILTER BY NORTHERN BLOTTING

The method used was that of Thomas (1980). RNA separated on 1.5% formaldehyde agarose gel was transferred directly to nitrocellulose filters on a blotting apparatus and using the same buffer as described for Southern transfer of DNA. After transfer filters were baked at 80°C under vacuum for 1h.

2.2.11 IN VITRO ^{32}P-LABELLING OF DNA BY NICK TRANSLATION

In vitro labelling of DNA was performed using the Amersham nick translation kit. The method used was as described in the instructions. The reaction contained 0.5 µg DNA 10 µl nucleotide/buffer solution (100 µM each of dATP, dTTP and dGTP containing Tris-HCl pH 7.8, MgCl2 and -mercaptoethanol in unspecified concentrations), 5 µl (50 Ci; µ125 pmol) of [α ^{32}P]-dCTP, water to 45 µl and 5 µl of enzyme solution (2.5 units DNA polymerase I and 50 µg DNAse I in a buffer containing Tris-HCl, MgCl2, glycerol and BSA in unspecified concentrations). The mixture was incubated at 14°C for 2h and the reaction terminated.

Labelled DNA was separated from unincorporated radionucleotide by passage through a 5 cm column of Sephadex G50, equilibrated in 150 mM NaCl, 10 mM EDTA, 5 mM Tris-HCl pH 7.5 and 10% SDS. 0.5 ml fractions were collected and 1 µl aliquots counted in a β-scintillation counter. Fractions corresponding to the first peak of radio activity contained labelled DNA, and were pooled for use as a hybridisation probe.

2.2.12 HYBRIDISATION OF ^{32}P-LABELLED PROBES TO FILTER-BOUND DNA

Filters were equilibrated in prehybridisation solution containing 5 x SSC (0.75 M NaCl M Na-citrate pH 7.0, 5 x Denhardt's solution (0.1% (w/v) each of Ficoll 400, BSA and PVP) and 100 µg/ml sheared and denatured herring sperm DNA, at 65°C with shaking for 1-3h in a heat sealed plastic bag. At the end of prehybridisation, ^{32}P-labelled DNA previously denatured by boiling for 5 min, was added and incubation continued for 8-20h. After incubation, the hybridisation solution was removed and the filter washed for 30 min at 65°C in solutions containing 0.1% SDS 3 x SSC, 1 x SSC and 0.1 x SSC. Filters were then air dried and autoradiographed.

2.2.13 HYBRIDISATION OF ^{32}P-LABELLED PROBES TO FILTER-BOUND RNA

The Filters were prehybridised in a solution containing 50% de-ionised formamide, 5 x Denhardt's, 5 x SSC and 100 μg/ml sheared and denatured herring sperm DNA for 2h at 42^{o}C with shaking. Hybridisation was carried out under similar conditions, except 2 x Denhardt's solution and 220 μg/ml herring sperm DNA was used. ^{32}P-labelled DNA probe was denatured by placing in a water bath at 95^{o}C for 5-10 min before addition. Hybridisation was performed at 42^{o}C for 48h. Filters were washed with two washes of 10 min each in 3 x SSC, 1 x SSC and 0.1 x SSC containing 0.1% SDS. The Filters were air dried and autoradiographed.

2.2.14 AUTORADIOGRAPHY

Autoradiography for ^{32}P was performed using a preflashed film (Fuji RX, Fujimex, Swindon, Wilts, UK) and an intensifying screen (Dupont, Wilmington, Delaware, USA) exposing at -80 for the appropriate time (Thomas, 1980).

2.2.15 TRANSFER OF PLAQUES TO NITROCELLULOSE FILTERS

Nitrocellulose filters, marked with a hole punch to aid in orientation, were placed on the surface of the agar plates for 5 min. Holes were punched through into the agar via the holes in the filters. Filters were then removed, placed plaque side up on blotting paper soaked with solution II (0.25 Tris-HCl pH 7.5, 1.5 M NaCl) for 10 min and then to blotting paper soaked with 2 x SSPE (0.36 M NaCl, 0.02 M NaH2P40 pH 7.7, 0.002 M Na2 EDTA) for a further 10 min. Finally filters were baked between two layers of 3 MM paper at 80^{o}C for 1h.

2.2.16 REMOVAL OF PROBE AND RE-USE OF DNA BLOTS

Nitrocellulose filters were submerged in a solution of 0.1% SDS. Brought to boil and left for 30-60 min and allowed to cool to room temperature.

After probe removal filters were autoradiographed to ensure removal of probe, filters were prehybridised and hybridised to the new probe as described in 2.2.12.

For Hybond-N filters an alternative method was used for removal of probe. Filters were incubated at 45°C in 0.4 M NaOH for 30 min, transferred to a solution containing 0.1 x SSC, 0.1% SDS and 0.2 M Tris-HCl pH 7.5 and incubated at 45°C for a further 30 min.

2.2.17 REMOVAL OF PROBE AND RE-USE OF RNA BLOTS

Probes were removed from nitrocellulose and Hybond-N filters by washing blots at 65°C for 1-2h in a solution containing 0.005 M Tris-HCl, pH 8.0, 0.002 M Na2 EDTA and 0.1 x Denhardt's solution. Filters were prehybridised and hybridised as described in section 2.2.13.

2.2.18 TITRATION OF PHAGE STOCK USING POUR PLATE METHOD

This was carried out using a microtitre plate. 180 µl of S M buffer was dispensed into the wells of the microtitre plate, 10 wells were filled. 20 µl of phage stock was dispensed into the left hand well of the row. Mixed gently by drawing the liquid into the pipette tip 3 or 4 times, along with stirring using the end of the pipette tip. 20 µl from the first well were transferred into the second well. 20 µl from the second well to the third well etc until well 10. 100 µl aliquots of each dilution was taken and added to 100 µl plating cells. The mixture was incubated for 20 min at 37°C to allow phage to absorb to the plating cells. About 3 ml of top layer agar was added, mixed gently and poured onto a plate. This was repeated with all the wells. After all the 10 plates were poured and closed they were left to stand at room temperature for 15 min until the top layer had set. The plates were incubated at 37°C overnight.

The titre of the stock was estimated as follows. If a dilution of 10^{-6} produced 10 plaques when 100 µl were plated out, the number of plaques at 10^{0} dilution was 10 x 10^{6} in 50 µl of original stock. Therefore in 1.0 ml of the stock there were 10 x 10^{6} x 10 phage = 10^{8} pfu/ml.

2.2.19 SCREENING OF GENOMIC LIBRARIES

Screening of genomic libraries was carried out using megaplates. Aliquots of phage stock or packaging extract containing 2 x 10^{4} bacteriophage resuspended in 3 ml SM buffer were mixed with 3 ml of plating cells and incubated at 37°C for 15 min. The transfected cells were mixed with 40 ml of molten top layer agarose at 50°C and carefully poured over the surface of the plate when the top layer agarose had set, the plate was incubated wt 37°C.

2.2.20 PREPARATION OF PLATING CELLS FOR TITRATION

In a McCartney bottle containing 10 ml of L-broth + 0.4% maltose a single colony was inoculated and incubated at 37°C overnight on a rotary shaker. The cells were harvested by centrifugation at 4,000g for 10 min, and resuspended in 3.3 ml of SM buffer. The cells were stored at 4°C until required.

2.2.21 RESTRICTION MAPPING OF GENOMIC CLONES

The DNA obtained from large-scale preparations was digested with a number of restriction endonucleases in single and multiple digests. Digests were electrophoresed through an agarose gel with a variety of size markers. From the restriction sites of the enzymes used in the vector DNA and the bands seen of EtBr stained gels, the site for those enzymes in the DNA of the insert were located and a partial restriction map of the clone was constructed. In order to locate the position of a certain part of the clone, usually gene sequence, the gels were blotted onto nitrocellulose filters and hybridised to the appropriate probe.

2.2.22 GENOMIC LIBRARY CONSTRUCTION

2.2.22.1 PREPARATION OF EMBL3, BamHI ARMS

Cohesive ends of EMBL3 were annealed by the addition of MgCl2 to 0.01 M and incubation at 42°C for 1h. A 0.2 µg aliquot was checked on a 0.7% agarose gel. Annealed EMBL3 DNA was digested at a concentration of 0.2 µg/µl with threefold excess of BamH1 at 37°C for 3h. Digested EMBL3 DNA was extracted twice with phenol and chloroform, precipitated with ethanol and resuspended in TE buffer. DNA was fractionated on a 25 ml sucrose gradient (20-40%) centrifuged at 26,000 rpm for 18h at 15°C. 0.6 ml fractions were collected and 15 µl of every third fraction was diluted with 35 µl H20 and analysed on a 0.5% agarose gel. Fractions which contained predominantly EMBL3 BamH1 arms were pooled and dialysed extensively against a large volume of TE at 4°C for 12-16h, ethanol precipitated and resuspended in TE buffer.

2.2.22.2 PREPARATION OF 10-20 KB FRAGMENTS OF Sau3A CLEAVED *ARABIDOPSIS* DNA

Restriction enzyme conditions were established which yielded the maximum amount of DNA in the required size range. Large scale digestions were then performed, using 50 µg of *Arabidopsis* DNA per digestion. After cleavage, the DNA was ethanol precipitated and then resuspended in TE buffer pH 7.6 and pooled together in one tube and loaded on a 0.5% agarose gel, electrophoresed at 30 V overnight. A slice of agarose in the area containing the DNA in the required size range (i.e. 10-20 kb) was cut out. DNA was electroeluted and purified by phenol, chloroform and ether extractions, then ethanol precipitated and resuspended in TE buffer pH 7.6.

2.2.23 PREPARATION OF *IN VITRO* PACKAGING EXTRACTS

2.2.23.1 PREPARATION OF SONICATED EXTRACT

Sufficient cells from a 100 ml overnight culture of *E. coli* BHB2690 were inoculated into 500 ml of NZM broth prewarmed to 32°C, to give an initial OD600 ð 0.1. Culture was incubated with aeration at 32°C until an OD600 = 0.3 was reached. Lysogens were induced by placing the flask into a water bath at 45°C. Culture was swirled continuously for 15 min. Induced cells were incubated at 38-39°C for 2-3h with vigorous aeration. Successful induction was checked by adding a drop of chloroform to a small sample culture, which cleaved within a few minutes Cells were harvested by centrifugation at 4,000g for 10 min at 4°C. Liquid was drained onto absorbent tissue, and remaining medium was removed with a Pasteur pipette and Q-tips. Walls of centrifuge bottles were dried with tissue paper. 3.6 ml of freshly prepared sonication buffer (20 mM Tris-HCl (pH 8.0) 1 mM EDTA, 5 mM -mercaptoethanol) were added, pellet was resuspended thoroughly and transferred to a small, clear plastic tube. The homogeneous suspension was sonicated in short bursts (10 seconds) at maximum power using a microtip probe. The tube was immersed in ice-water and the temperature was not allowed to exceed 4°C. The suspension was allowed to cool for 20-30 seconds between each burst of sonication. Sonication continued until solution cleaved and its viscosity decreased. Sonicated sample was transferred to a centrifuge tube and debris was removed by centrifugation at 12,000g for 10 min at 4°C. The the supernatant an equal volume of cold sonication buffer and 1/6 volume of freshly prepared packaging buffer (6 mM Tris HCl pH 8.0, 50 mM spermidine, 50 mM putresceine, 20 mM MgCl2, 30 mM ATP, 30 mM -mercaptoethanol) were added. Aliquots of 15 µl were dispensed into precooled (4°C), 1.5 ml eppendorf tubes. Caps of the tubes were immediately closed, immersed briefly in liquid nitrogen and transferred to -80°C for long-term storage.

2.2.23.2 PREPARATION OF FREEZE-THAW LYSATE

E. coli BHB2688 cells were used to prepare the freeze-thaw lysate, they were cultured, induced, checked for successful induction and recovered exactly as those used to prepare sonicated extract. They were resuspended in 3 ml ice-cold sucrose solution (10% sucrose in 0.25 M Tris HCl pH 8.0) 0.5 ml of the suspension was distributed in each of six precooled ($4^{O}C$) eppendorf tubes. To each tube 25 µl of fresh, ice-cold lysozyme solution (2 mg/ml lysozyme in 0.25 M Tris HCl pH 8.0) were added. Tubes were immersed in liquid nitrogen. Extracts were thawed in ice, and 25 µl of freshly prepared packaging buffer were added to each tube and mixed. Thawed extracts were combined in a centrifuge tube and centrifugated at 48,000g for 1h at $4^{O}C$. Supernatant was dispensed into precooled eppendorf tubes, 10 µl each. Tubes were immersed in liquid nitrogen until all the aliquot had frozen, then the tubes were transferred to -$80^{O}C$ for long-term storage.

2.2.23.3 *IN VITRO* PACKAGING OF EMBL3 DNA

Packaging extracts were removed from storage at -80^{O} and allowed to thaw on ice. The freeze-thaw lysate (usually thaws first) was transferred to the still frozen, sonicated extract. The two extracts were mixed gently, when they were almost totally thawed. The DNA to be packaged (up to 1 µg dissolved in 5 µl of 10 mM Tris HCl pH 7.9 and 10 mM MgCl2). The DNA was mixed very gently with packaging extracts and incubated for 1h at room temperature. 0.5 ml of 5M buffer and a drop of chloroform were added and mixed. Debris was removed by centrifugation in a microfuge for 30 seconds and titre of the viable EMBL3 particles was measured as described in 2.2.18.

CHAPTER THREE

RESULTS

3. RESULTS

3.1 SCREENING PEA GENOMIC LIBRARY FOR *LEG A* TYPE, *LEG J* TYPE AND VICILIN GENES

In order to detect DNA sequences representing seed protein gene families, the following DNA clones were used.

The DNA clone pDUB6 [originally referred to as pAD 4.4, Delauney (1984)] contains a 1.1 kb cDNA derived from the coding and 3' flanking region of the *Leg A* gene, starting 850 bp downstream of the transcription start. This clone was used to detect DNA sequences of *Leg A* subfamily ("Major" legumin) which contains five genes; *Leg A, B, C, D* and *E*.

Leg J subfamily ("Minor", legumin), which consists of three genes, of which *Leg J* and *Leg K* have been characterised (Gatehouse *et al.*, 1988). A DNA clone designated pJC 5.2, which is an EcoRI genomic clone containing 1.42 kb of the transcriptional unit of *Leg J*.

Vicilin family genes; the cDNA clone pAD 2.1 (pDUB2) was used as a probe. This contains a complete mature vicilin polypeptide coding sequence.

A pea genomic library, prepared from DNA extracted from var. "Feltham First" that had been prepared previously by Shirsat (1984). The library was screened for *Leg A* type, *Leg J* type and vicilin genes, initially using a combination of all three probes mentioned above. The plaques which hybridised to the mixed probe were picked off, replated and characterised by hybridisation to individual probes. Positive plaques were further purified by repeated cycles of plating out and hybridisation until greater than 95% of the well-separated plaques on a plate were seen to hybridise to the probe (Figure 1). A single plaque which strongly and clearly hybridised was picked up and amplified to make a stock of the genomic clone. This was successfully achieved with the three targeted DNA

FIGURE 1 Purification of genomic clones.

These are autoradiographs of the final stage in plaque lifting of:

A - λ*Leg E1*
B - λ*Leg J2*
C - λ*Vic 3*

Plaques on A were hybridised to pDUB6
Plaques on B were hybridised to pJC 5.2
Plaques on C were hybridised to pDUB 2

All filters were washed to high stringency (0.1 x SSC 65°C for 30 min)

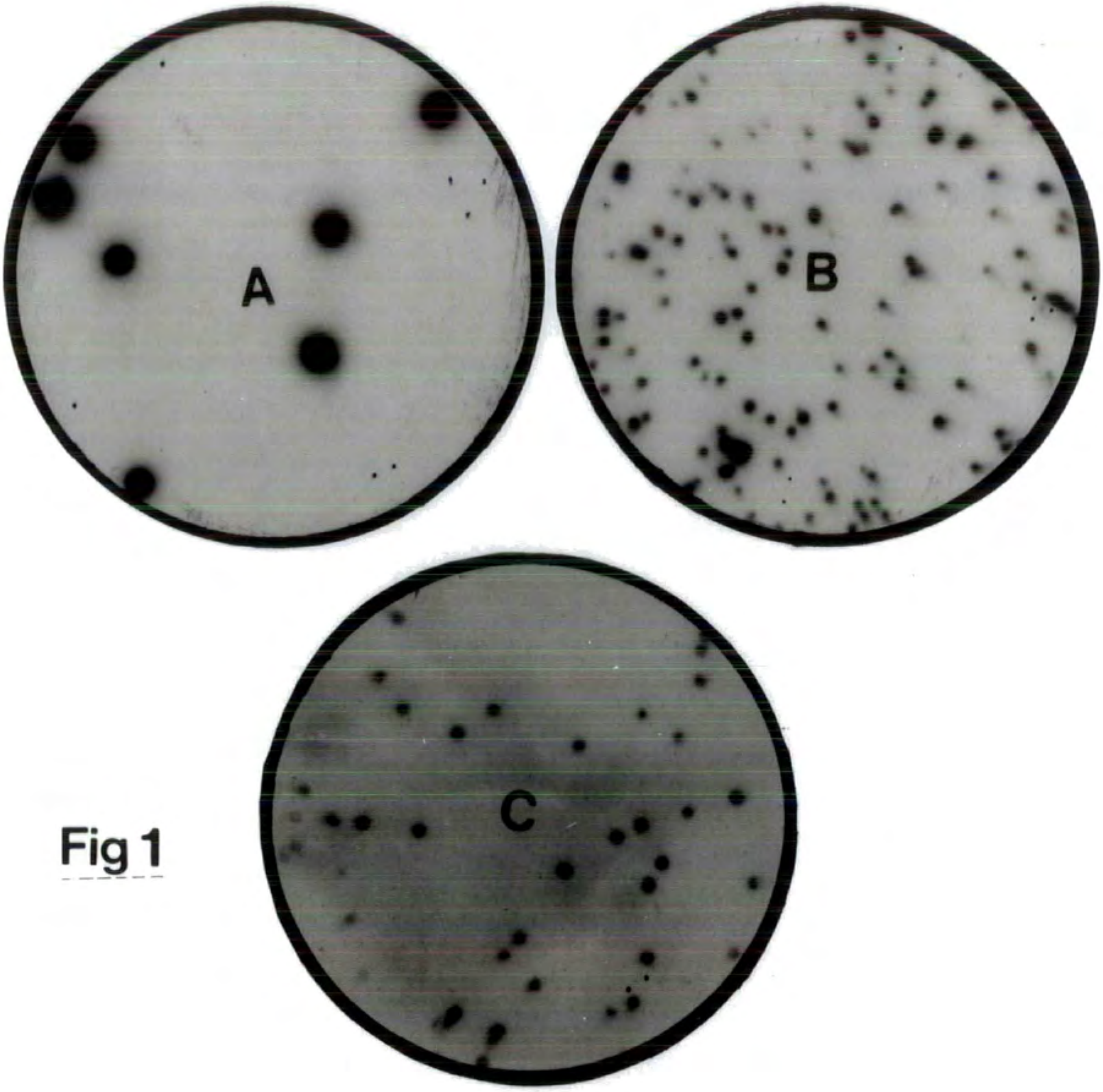

Fig 1

sequences (i.e. *Leg A* type, *Leg J* type and vicilin). The three genomic clones isolated were designated λ*Leg E1*, λ*Leg J2* and λ*Vic 3*.

3.2 CHARACTERISATION OF THE GENOMIC CLONES

3.2.1. RESTRICTION MAPPING OF λ*LEG E1* GENOMIC CLONE

After the stock of λ*Leg E1* was titrated, a large scale DNA preparation was made to be used in characterising the clone by restriction mapping. Aliquots of 1 µg DNA from λ*leg E1* were digested with restriction enzymes in single and double digests. A typical gel showing restriction digests of λ*Leg E1* is shown in Figure 2 . These gels were blotted and hybridised to the *Leg A* probe (Figure 3), in order to locate the legumin gene on this genomic clone, referred to as *Leg E* hereafter.

The sizes of the fragments obtained from the restriction data show that the genomic DNA insert contained in λ*Leg E1* was approx. 13.5 kb. Since none of the restriction fragments which hybridised to the probe was large enough to be either of the vector arms plus part of the insert, the *Leg E* gene was in the middle of the insert. Five fragments were obtained from the EcoRI digestion, one of which was the left arm plus approximately 1.2 kb of insert sequence, another fragment was the right arm plus approximately 3.4 kb of insert sequence; neither of these fragments hybridised to the probe (Figure 4).

The restriction map (Figure 4) obtained from the data was not entirely accurate because of the failure to detect very small fragments on agarose gels and due to the difficulty of estimating exact sizes of the fragments, especially the very large ones.

The sizes of the fragments hybridising to the probe are given below

FIGURE 2 Restriction enzyme analysis of λ*LegE1*

A) EMBL3 DNA cleaved with SalI
B) λ*Leg E1* DNA cleaved with SalI
C) λ*Leg E1* DNA cleaved with SalI and EcoRI
D) λ*Leg E1* DNA cleaved with EcoRI
E) EMBL3 DNA cleaved with EcoRI
F) λ*Leg E1* DNA cleaved with BamHI
G) λ*Leg E1* DNA cleaved with BamHI and EcoRI
H) λ*Leg E1* DNA cleaved with AvaI
I) λ*Leg E1* DNA cleaved with HindIII
J) λ*Leg E1* DNA cleaved with HindIII and EcoRI
K) λ*Leg E1* DNA cleaved with Bgl2 and EcoRI
L) λ*Leg E1* DNA cleaved with Bgl2

FIGURE 3 Autoradiograph of the gel from Figure 2, after Southern blotting and hybridisation to ^{32}P labelled insert from pDUB6 and washing to high strigency (0.1 x SSC, 65^{o}C for 30 min)

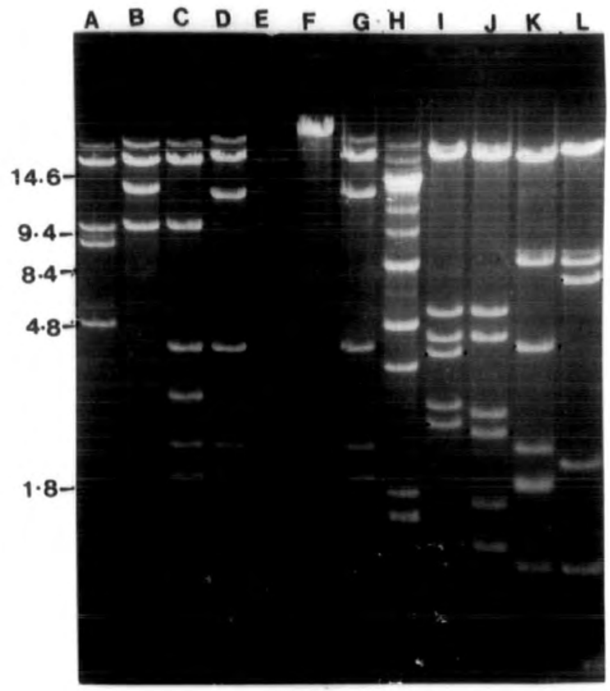

Fig 2

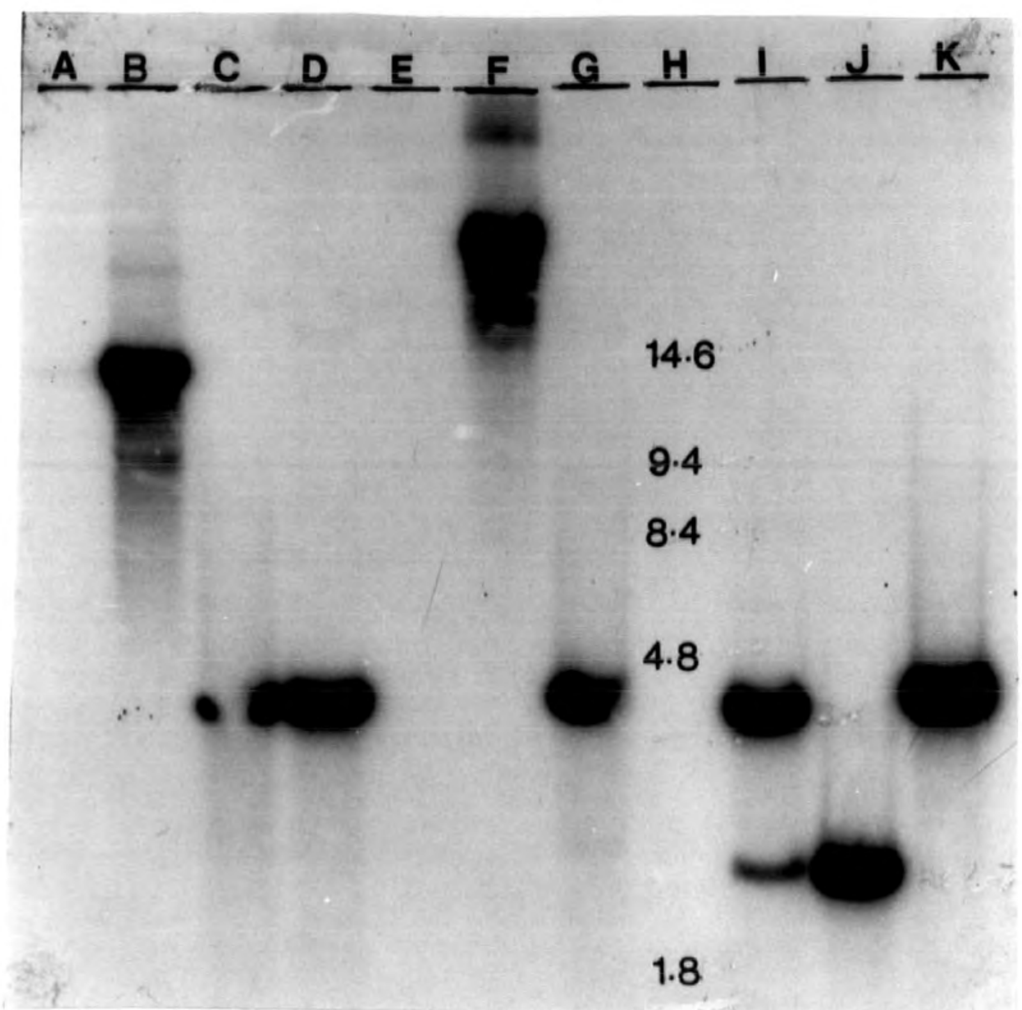

Fig 3

Fig4

λ Leg E1 Restriction Map

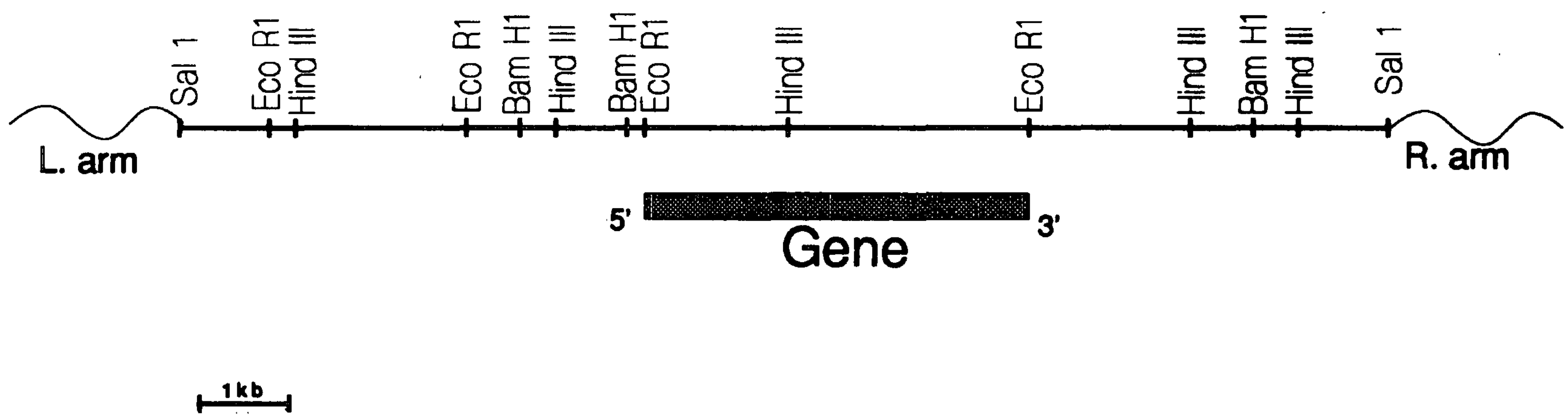

FIGURE 4 Partial restriction map of λ *Leg E1*. Bar indicates fragment hybridised to cDNA probe pDUB6

SalI	13.5 kb
EcoRI	4.2 kb
HindIII	3.7 kb
EcoRI-HindIII	6.5 kb

Before any further characterisation was carried out, the λ*Leg E1* restriction map was compared to restriction maps of other genomic clones containing *Leg A*-type genes which had already been characterised, such as *Leg 1* and *Leg 2* (Croy *et al.*, 1984), *Leg 3* (Shirsat, 1984) and *Leg D* (Bown *et al.*, 1985), and no homology was found. This suggested that λ*Leg E1* was not the same as any of them and did not overlap with any of the other genomic clones. Furthermore, λ*Leg E1* was checked for any cross-hybridisation to pBR322 sequence, and the results were negative. Some cross-hybridisation was seen with pDUB2 (vicilin) probe but this was removed completely after washing the filter in 1 x SSC at 64^{o}C; similarly, when probed with pJC 5.2 probe, cross-hybridisation was observed but was removed after washing the filter in 0.1 x SSC at 65^{o}C.

3.2.2. RESTRICTION MAPPING OF λ*Leg J2*

The λ*Leg J2* DNA was restriction digested with a selection of restriction enzymes in single and double digests.

The size of λ*Leg J2* was calculated from the SalI (Figure 5) digest, where two fragments in addition to the vector arms were obtained and were approximately 2.0 and 13 kb in size, which suggested that the size of the fragment was about 15 kb. However, this digest failed to provide information about the location of the gene(s) because both fragments hybridised to the probe. Accurate information regarding the gene(s) location was obtained from the analysis of data from other digestions, especially the SalI/EcoRI double digestion (Figure 5), which revealed the existence of two genes within the insert. This was confirmed when the

FIGURE 5 Restriction enzyme analysis of λ*LegJ2*

A) λ*Leg J2* DNA cleaved with SalI
B) λ*Leg J2* DNA cleaved with SalI and EcoRI
C) λ*Leg J2* DNA cleaved with EcoRI
D) λ*Leg J2* DNA cleaved with AvaI and BamHI
E) λ*Leg J2* DNA cleaved with HindIII
F) λ*Leg J2* DNA cleaved with HindIII and EcoRI
G) λ*Leg J2* DNA cleaved with Bgl2 and EcoRI
H) λ*Leg J2* DNA cleaved with Bgl2

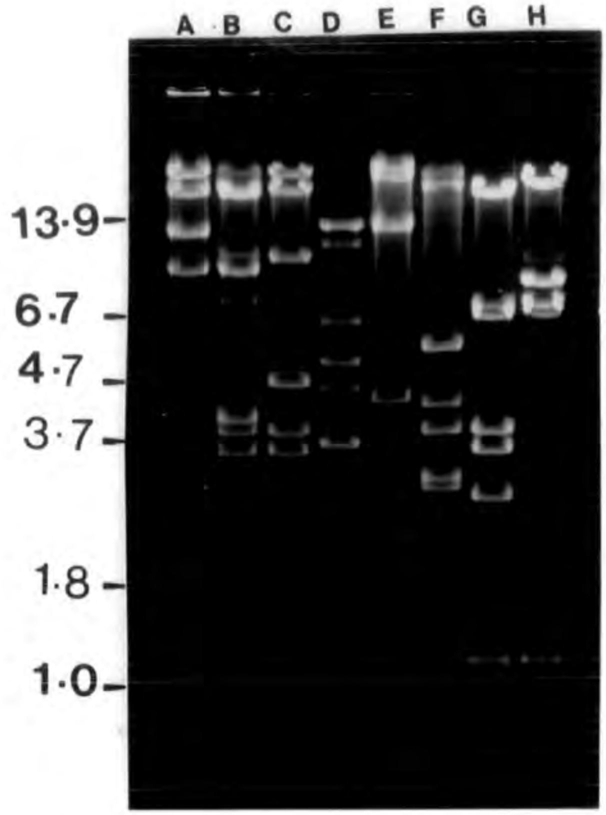

Fig5

Fig6

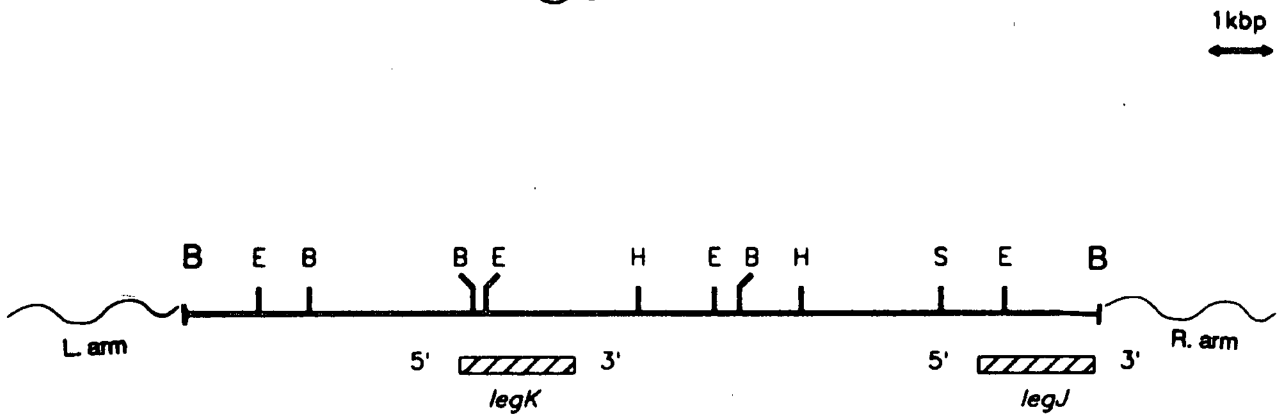

FIGURE 6 Partial restriction map of λ*Leg J2*. Bars indicate fragments hybridised to cDNA probe pJC5.2.

restriction sites for all the enzymes used were located and the restriction map constructed (Figure 6).

The sizes of the hybridising fragment are shown below

SalI	2 and 13 kb
BamHI	4.2 + 24 kb
EcoRI	3.5 + 21 kb

The presence of two genes in the insert provided a strong possibility that the clone could be similar to λJC5 genomic clone previously characterised (Gatehouse *et al.*, 1988). When the restriction maps of λ*Leg* J2 and λJC5 (Figure 7) were compared they were found to be identical in the region extending from the EcoRI site close to the 3' end of *Leg J* gene to the EcoRI site near the 5' end of *Leg K* gene in λJC5 (Figure 7), which is the region where they overlap. The restriction map shows that the 5' flanking sequence of the *Leg K* gene, which was missing in λJC5, is present in λ*Leg J2*.

3.2.3 RESTRICTION MAPPING OF *lVic 3*

Like λ*Leg El* and λ*Leg J2*, λ*Vic3* was digested with selected restriction enzymes in single and multiple digestions (Figure 8). The size of the insert was calculated at about 16 kb, and a restriction map of λ*Vic3* (Figure 10) was constructed. Data obtained from hybridization of gel blots to the probe (Figure 9) indicated that the vicilin gene was close to the left arm, because the fragments which hybridised to the probe in the case of single digests (e.g. HindIII, EcoRI, BamHI) were large (>20 kb) extending from the left arm into the insert. When λ*Vic 3* was double digested with those restriction enzymes and SalI, which has a site in the polylinker at either end of the insert, hybridization to the large fragments was lost, confirming that the gene was located close to the left arm.

Fig 7

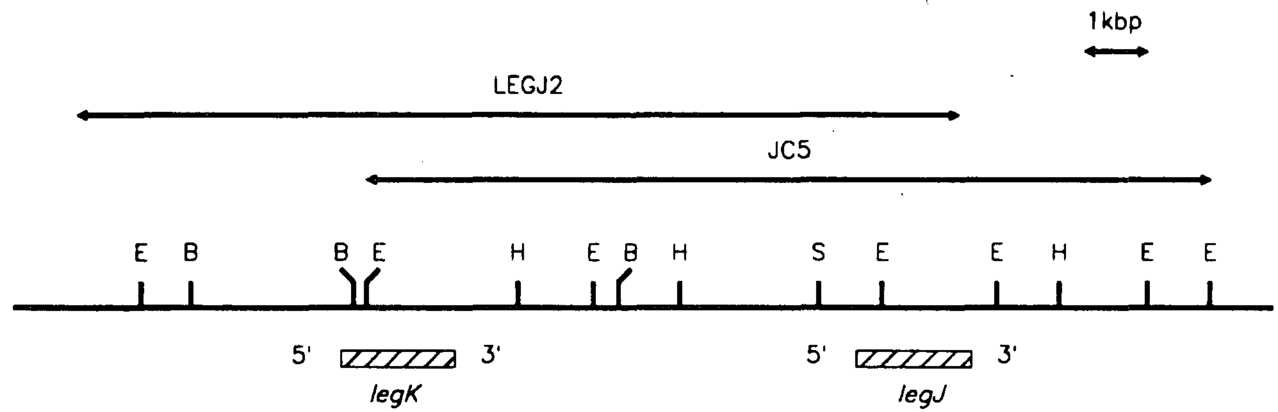

FIGURE 7 Comparison between λ*Leg J2* and λJC5 clones showing the 17 kb fragment covered by the clones. Arrows indicate the fragment covered by each clone. Boxes indicate the position of the genes.

E - EcoRI, B - BamHI, H - HindIII, S - SauI

FIGURE 8 Restriction enzyme analysis of λ*Vic3*

A) λ*Vic 3* DNA cleaved with SalI
B) λ*Vic 3* DNA cleaved with SalI and EcoRI
C) λ*Vic 3* DNA cleaved with EcoRI
D) λ*Vic 3* DNA cleaved with EcoRI and BamHI
E) λDNA cleaved with AvaI and BamHI
F) λ*Vic 3* DNA cleaved with BamHI
G) λ*Vic 3* DNA cleaved with HindIII and EcoRI
H) λ*Vic 3* DNA cleaved with HindIII
I) λ*Vic 3* DNA cleaved with Bgl2

All sizes are in kb

FIGURE 9 Autoradiograph of the gel from Figure 8 after hybridisation to the ^{32}P labelled insert from pDUB2 and washed to high stringency (0.1 x SSC, 65^{o}C for 30 min)

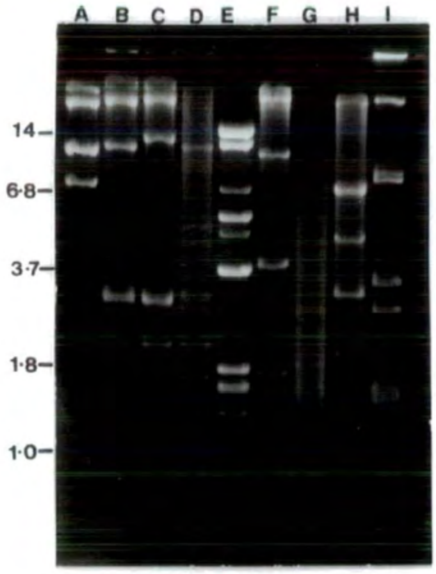

Fig 8

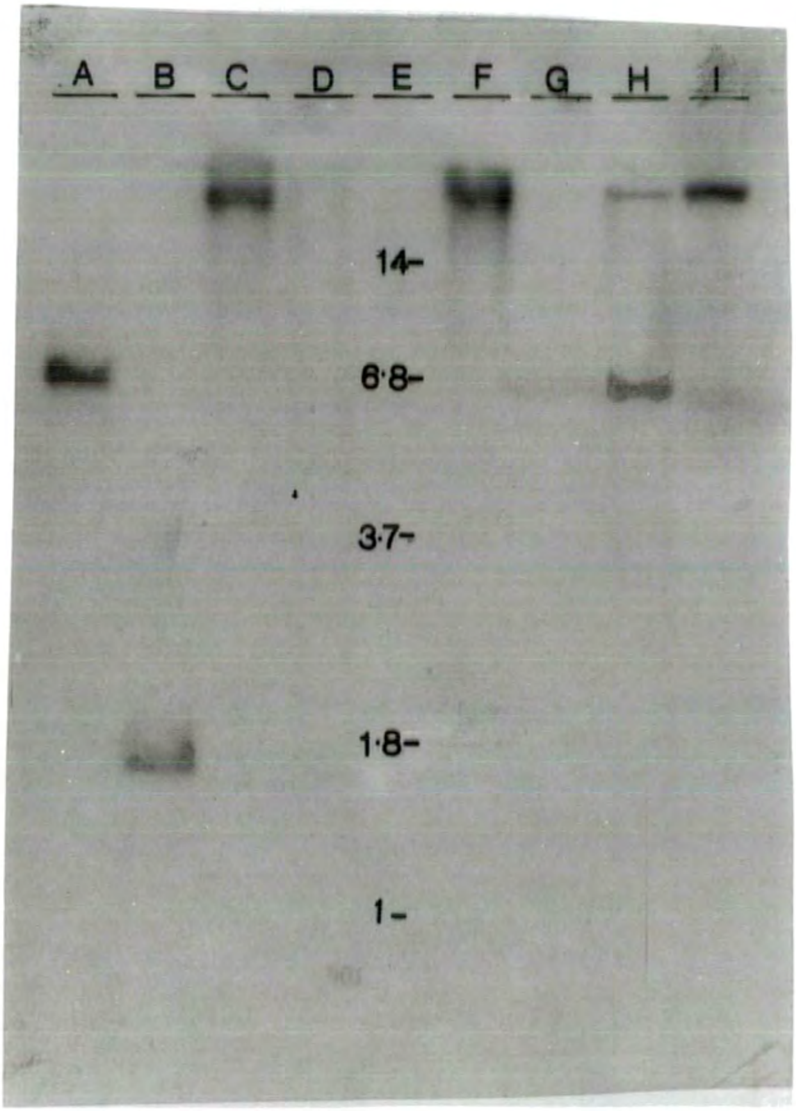

Fig 9

Fig10

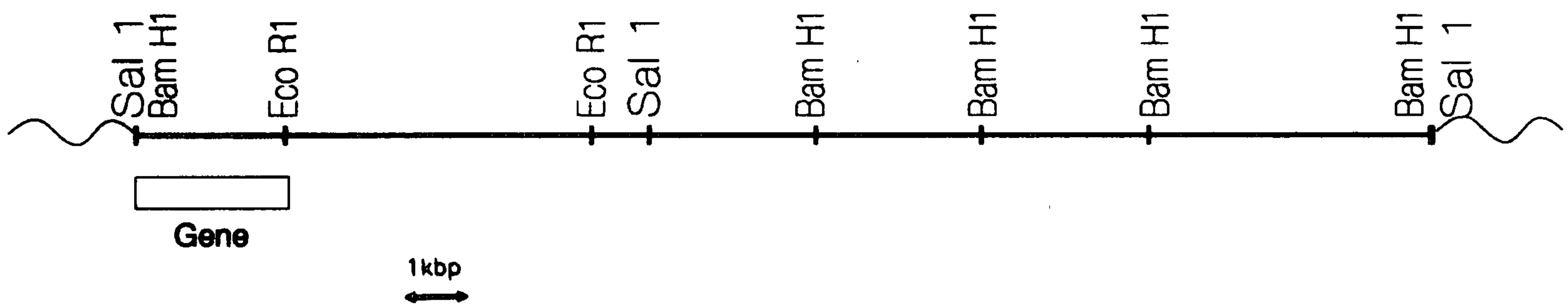

FIGURE 10 Partial restriction map of λ*Vic 3*. Bar indicates the fragment that hybridises to cDNA probe pDUB2.

When the restriction map of λ*Vic 3* was compared to that of λ*Vic 1* (Sawyer 1986) it was found that λ*Vic 3* and λ*Vic 1* were exactly the same, therefore any further characterisation or analysis of λ*Vic 3* was not necessary.

3.2.4 SEQUENCE DETERMINATION OF *LEG E*

Three EcoRI fragments from λ*Leg E1*, which are 4.3, 2.4 and 2 kb in size (Figure 4), were subcloned into pUC18 and designated pSY1, SY2 and pSY3 respectively. The recombinant plasmid pSY1 was further restriction mapped in order to localise restriction sites within the insert which could be used in sub-cloning smaller fragments into M13 in order to be sequenced. The restriction map of the insert is shown in Figure 11. Twenty-three different fragments indicated by arrows were sub-cloned in M13. Those fragments cover the whole of the insert and overlap in many areas as indicated by the arrows in Figure 11. The sequence of the gene *Leg E* and its 5' and 3' flanking regions was determined by conventional techniques using single stranded templates produced from the M13 subclones. Both strands of the DNA were sequenced. The sequence determined is shown in Figure 12.

3.2.5 CHECKING PEA GENOMIC DNA FOR *Leg A* TYPE SEQUENCES

Ten μg of pea genomic DNA was restriction digested to completion with EcoRI and was run on an agarose gel which also contained λ*Leg E1* DNA cleaved with the same enzyme (Figure 13). After being transferred to a nitrocellulose filter the DNA fragments were hybridised against the insert from pDUB6. Four different fragments in the pea genomic DNA hybridised to the probe (Figure 14). The smallest fragment corresponded to the 4.3 kb EcoRI fragment from λ*Leg E1*.

Fig11

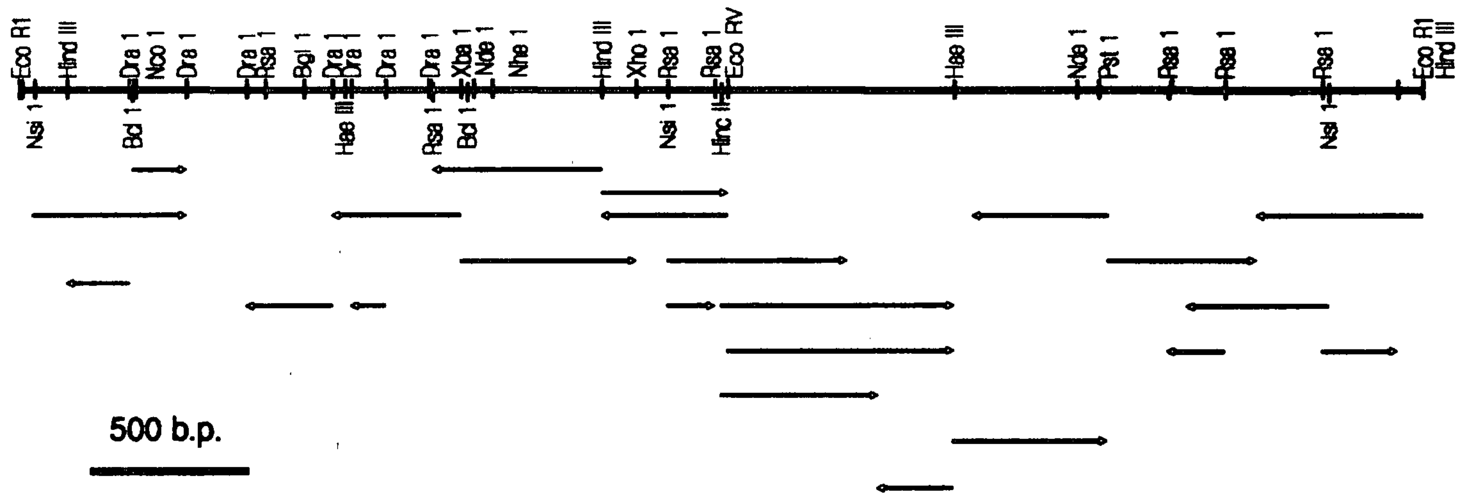

FIGURE 11 Restriction and sequencing map of pSY1 showing position of restriction sites. Sequencing runs are indicated by arrows.

Fig12

```
legE .................................................................TTTAAAGAAAAATAACTCATATTAACAAATAGGTTTTGACTCTCCTTA -961

legE ATTCAACTAAGTGTCTGAGCGTTGTACGATGATATGTAAACACAATAACTTTGGAACTAGTTAGATAAGAGTAAAAATGTCCCAATACAAATATTATGACGAGTAATTCCTTCTTAATGG -841

legE TAGTCTAGTTTACTTTTCATTCAAGATCTGTTCACATAATATATAGTATGGATTTTATTTTATTTTATTGACTTCAAGTTCGTAATCACTAGCATTGCAAATCAGTTTAAAAAGTAATGA -721

legE TCAATTAAGGATTATGATTACGGAAGCGGGCCCAAATCAATTTAAAAAGAAAATAGTCAAGTGAGGGTTATGATTACGAATGTGATGGCTAACTTTTCTTAAATAAGTATTAAAGAACTT -601

legE AAACAATCATCATCAAATTTAAAGCAAATCAATTGTTATTGATCCAAATAATTTGTGTGATTTTATATAACATATATGACTATATTGTAAATTGTAATTCAAGCACTTTAATTTGAAGTT -481

legE GTTATCGTCCACCTATATTCAACTACACAAAATTGTGCTTTACCATTAAACTTTAAAAATTTGTACAGACCATGAACTAAATCCTATCCAACATACAATATACAGACATTAACTAGCTTG -361

legE AAAGTGAATCAGGTTCATATATCTAGATATTACAAGACAGTAATGATCAAACTCACGTACATATGTAAAAAGAGAAATCAACTATATACTATACATGGTCCCCAACACCACCGATCTCAG -241

legE CTAGCTATCTAATTACTCAACTCTCACTTGAAGCCACCTCTGCTGTAATGAGACATTTTGATTTTATGAGGTGTAACACACAAGGCTTCCATAACCATGCAAGATGAAGAATGTCGAATG -121
     .............................<ENHANCER>.............................................<      LEGUMIN BOX      >......

legE GTCAGTAACTAATGTATCTTCTTGTAGACGATGTGTCCCTCCTATACTTTCCTATGTTCACTATAAATCCCTATGCCAGATTAAGGTTCTTCGCGTCACAAACATATATTCTATCCAACT   -1
     ....................................................<TATA BOX>...............................................

legE ATGGCTACTAAGCTTCTTGCACTTTCTCTTTCATTCTGTTTTCTACTTTTGGGGGGCTGTTTTGCTTTGAGAGAACAGCCAGAGCAAAATGAGTGCCAGCTAGAACGCCTCAATGCCCTC  120
     <M  A  T  K  L  L  A  L  S  L  S  L  C  F  L  L  L  G  G  C  F  A :L  R  E  Q  P  E  Q  N  E  C  Q  L  E  R  L  N  A  L   18

legE GAGCCTGATAACCGTATAGAATCGGAAGGTGGGCTCATTGAGACTTGGAATCCCAACAACAAGCAATTCCGATGTGCTGGTGTGGCCCTCTCTCGTGCTACCCTTCAACGTAACGCCCTT  240
      E  P  D  N  R  I  E  S  E  G  G  L  I  E  T  W  N  P  N  N  K  Q  F  R  C  A  G  V  A  L  S  R  A  T  L  Q  R  N  A  L   58

legE CGCAGACCTTACTACTCCAATGCTCCCCAAGAAATTTTCATCCAACAAGGTTACTTATTTTAATCTTATACCAACTTCTTTACGTACATTACATGCATATTAGCATACTAATTAGTGTTC  360
      R  R  P  Y  Y  S  N  A  P  Q  E  I  F  I  Q  Q  <------------------------------------------IVS-1------------

legE TACTATACCAATTACAGGTAATGGATATTTTGGGATGGTATTCCCCGGTTGTCCTGAGACCTTTGAAGAGCCACAAGAATCTGAACAAGGAGAGGGACGCAGGTACAGAGACAGACATCA  480
     --------------->G  N  G  Y  F  G  M  V  F  P  G  C  P  E  T  F  E  E  P  Q  E  S  E  Q  G  E  G  R  R  Y  R  D  R  H  Q  109

legE AAAGGTTAACCGATTCAGAGAGGGTGATATCATTGCAGTTCCTACTGGTATTGTATTTTGGATGTACAACGACCAAGACACTCCAGTTATTGCCGTCTCTCTTACTGACATTAGAAGCTC  600
       K  V  N  R  F  R  E  G  D  I  I  A  V  P  T  G  I  V  F  W  M  Y  N  D  Q  D  T  P  V  I  A  V  S  L  T  D  I  R  S  S  149

legE CAATAACCAGCTTGATCAGATGCCTAGGGTGAGCACTGAGCATAATTAAACTTCCCATATAAGATAATATGTTGTCCAAAACAGTAACATAGATTCTATCTATCTATGTTTGACAGAGAT  720
       N  N  Q  L  D  Q  M  P  R <-----------------------------------IVS-2----------------------------------> R   159

legE TCTATCTTGCTGGGAACCACGAGCAAGAGTTTCTACGATACCAGCATCAACAAGGAGGAAAGCAAGAACAAGAAAATGAAGGCAACAACATTTTCAGTGGCTTCAAGAGGGATTTCTTGG  840
      F  Y  L  A  G  N  H  E  Q  E  F  L  R  Y  Q  H  Q  Q  G  G  K  Q  E  Q  E  N  E  G  N  N  I  F  S  G  F  K  R  D  F  L   199

legE AAGATGCTTTCAACGTGAACAGGCATATAGTAGACAGACTTCAAGGCAGGAATGAAGACGAAGAGAAGGGAGCCATTGTCAAAGTGAAAGGTGGACTCAGCATCATAAGCCCACCCGAGA  960
      E  D  A  F  N  V  N  R  H  I  V  D  R  L  Q  G  R  N  E  D  E  E  K  G  A  I  V  K  V  K  G  G  L  S  I  I  S  P  P  E   239

legE AGCAAGCGCGCCACCAGAGAGGCAGCAGACAAGAGGAAGATGAAGATGAAGATGAAGAGAGGCAGCCGCGCCACCAGAGAGGCAGCAGACAAGAGGAAGAGGAAGATGAAGATGAAGAGA 1080
      K  Q  A  R  H  Q  R  G  S  R  Q  E  E  D  E  D  E  D  E  E  R  Q  P  R  H  Q  R  G  S  R  Q  E  E  E  E  D  E  D  E  E   279

legE GGCAGCCGCGTCATCAAAGGAGAAGAGGAGAGGAGGAAGAAGAAGACAAGAAAGAGCGCCGCGGCAGCCAAAAAGGCAAAAGCAGAAGGCAAGGAGACAATGGGCTTGAGGAAACAGTTT 1200
      R  Q  P  R  H  Q  R  R  R  G  E  E  E  E  E  D  K  K  E  R  R  G  S  Q  K  G  K  S  R  R  Q  G  D  N  G  L  E  E  T  V   319

legE GCACTGCTAAACTTCGATTGAACATTGGCCCGTCTTCATCACCAGACATCTACAACCCTGAAGCTGGTAGAATCAAAACTGTTACCAGCCTGGACCTCCCAGTTCTCAGGTGGCTCAAAC 1320
      C  T  A  K  L  R  L  N  I  G  P  S  S  S  P  D  I  Y  N  P  E  A  G  R  I  K  T  V  T  S  L  D  L  P  V  L  R  W  L  K   359

legE TAAGTGCTGAGCATGGATCTCTCCACAAAGTATGTTTTTTCATATTTTAATTTGTTTTTCCATGAATCAATTTCATGTCGAACTATGTGCTAACTCATTACAATCTTCATACAGAATGCT 1440
      L  S  A  E  H  G  S  L  H  K <------------------------------IVS-3-------------------------------------> N  A   371
```

```
legE ATGTTTGTGCCTCACTACAACCTGAATGCAAACAGTATAATATACGCATTGAAGGGACGTGCAAGGCTACAAGTAGTGAACTGCAATGGCAACACCGTGTTTGATGGAGAGCTAGAAGCC 1560
      M  F  V  P  H  Y  N  L  N  A  N  S  I  I  Y  A  L  K  G  R  A  R  L  Q  V  V  N  C  N  G  N  T  V  F  D  G  E  L  E  A  411

legE GGACGTGCATTGACAGTGCCACAAAACTATGCTGTGGCTGCAAAGTCACTAAGCGACAGGTTCTCATATGTAGCATTCAAGACCAATGATAGAGCTGGTATTGCAAGACTTGCAGGGACA 1680
      G  R  A  L  T  V  P  Q  N  Y  A  V  A  A  K  S  L  S  D  R  F  S  Y  V  A  F  K  T  N  D  R  A  G  I  A  R  L  A  G  T  451

legE TCATCAGTTATAAATAATCTGCCGTTGGATGTTGTGGCAGCTACATTCAACCTGCAGAGGAATGAGGCAAGGCAGCTCAAGTCCAACAATCCCTTCAAATTTCTAGTTCCAGCTCGTCAG 1800
      S  S  V  I  N  N  L  P  L  D  V  V  A  A  T  F  N  L  Q  R  N  E  A  R  Q  L  K  S  N  N  P  F  K  F  L  V  P  A  R  Q  491

legE TCTGAGAACAGAGCTTCGGCTTAGATTTCGCACCAAATCAATGAAAGTAATAATGAAAAGTCTGAATAAGAATACTTAGGCTTAGATGCCTTTGTTACTTGTGTAAAATAACTTGAGTCA 1920
      S  E  N  R  A  S  A  *>..........................................................................................

legE TGTACCTTTGGCGGAAACAGAATAAATAAAAGGTGAAATTCCAATGCTCTATGTATAAGTTAGTAATACTTAATGTGTTCTACGGTTGTTTCAATATCATCAAACTCTAATTGAAACTTT 2040
     ..................( polyA+ )...........................................................................

legE AGAACCACAAATCTCAATCTTTTCTTAATGAAATGAAAAATCTTAATTGTACCATGTTTATGTTAAACACCTTACAATTGGTTGGAGAGGAGGACCAACCGATGGGACAACATTGGGAGA 2160

legE AAGAGATTCAATGGAGATTTGGATAGGAGAACAACATTCTTTTTCACTTCAATACAAGATGAGTGCAACACTAAGGATATGTATGAGACTTTCAGAAGCTACGACAACATAGATGAGTGA 2280

legE GGTGGTGATTCCTAGCAAGAAAGACATTAGAGGAAGCCAAAATCGAACAAGGAAGACATCAAGGGCAAGAGACAGGACCATCCATCTCAGGAAAAGGAGCTTTGGGATAGTCCGAGAAGT 2400

legE TGTACAAGAAATTTTTTGGAGGGTGAGTGATGCATTGCTGGTGACTTTAACTCAATCAAAATTGAGAAAGAAAGAAAAGGGAGGGGGCTCACATGTGAATAGAAGGGAAACGGGAGAATT 2520

legE TTACAGTTTTGATCTAATGGGCATCCCAGCTAGTGGTAACATATTCACCATGTTTAACCTTCACGGAAGTGTGATAGGTAGGTTAGACAAATTTCTGTTGGATGGAACACAAGGAATTC. 2639
```

FIGURE 12 Sequence of *Leg E* and flanking regions. The complete amino acid encoded by the gene is given. IVS = intron.

Fig12a

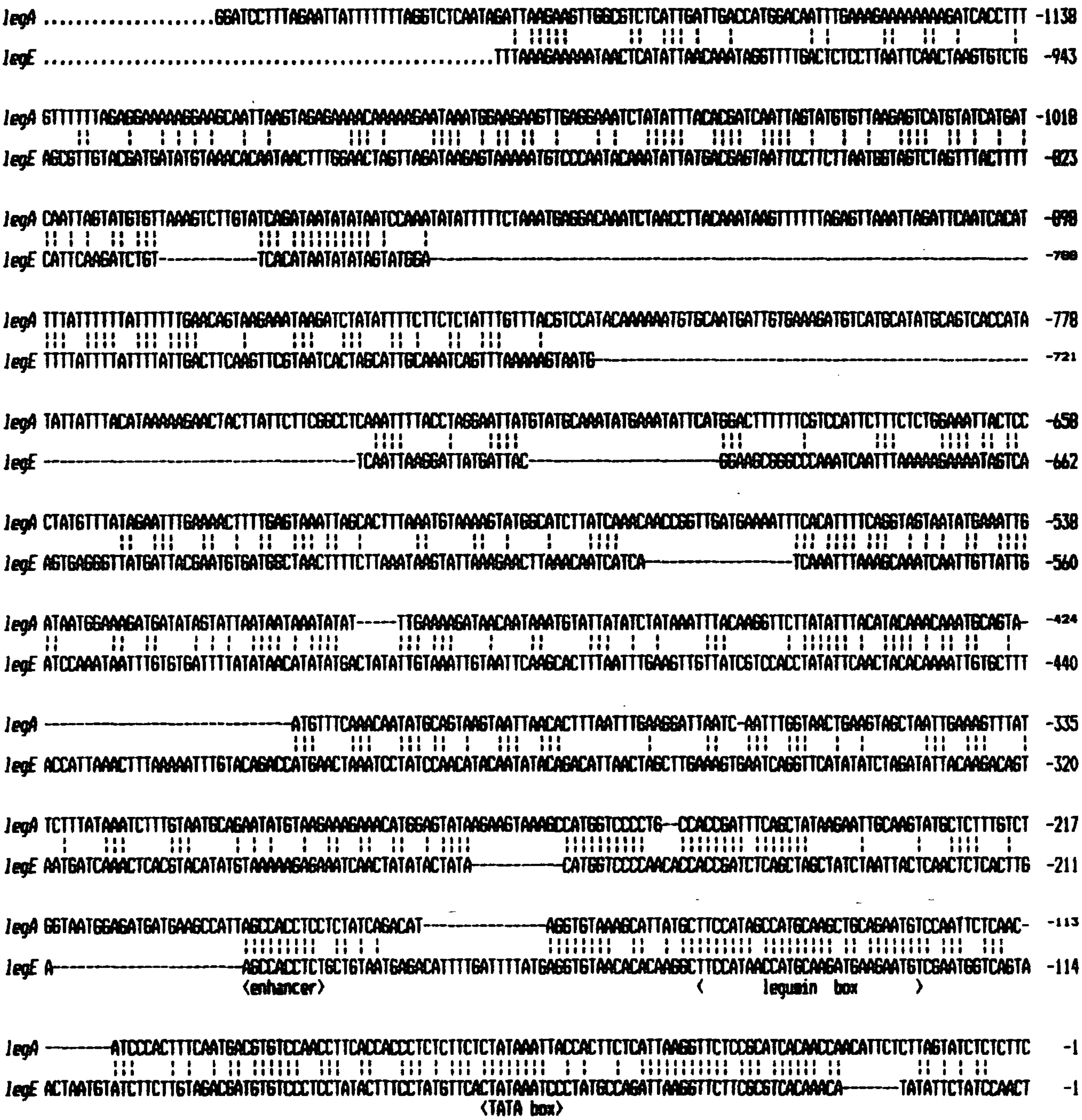

FIGURE 12a Comparison between the 5′ flanking regions of *Leg E* and *Leg A* genes.

: = regions where the two genes are homologous

Fig 12b

```
legA tctgagaacagagcttcggcttagATTTGGCACCAAATCAATGAAAGTAAT------------GAATAAGAAAACTAAGGCTTAGATGCCTTTGTTACTTGTGTAAAATAACTCGAGTCA
legE tctgagaacagagcttcggcttagATTTCGCACCAAATCAATGAAAGTAATAATGAAAAGTCTGAATAAGAATACTTAGGCTTAGATGCCTTTGTTACTTGTGTAAAATAACTTGAGTCA 1920

legA TGTACCTTTTAGCGGAAACAGAATAAATAAAAGGTAAAATTTCAGTGCTCTATGCTTTTCTACTCCAAGTTATAACCAGATGATATATATAACAATCACAATAAATAAATGTGAGTAAAA
legE TGTACCTTTG-GCGGAAACAGAATAAATAAAAGGTGAAATTCCAATGCTCTATGTATAAGTTAGTAATACTTAATGTGTTCTACGGTTGTTTCAATATCATCAAACTCTAATTGAAACT- 2038
                              ( polyA+ )

legA AAATATTGAAGAAAAATGATGTATTGAAATTATAACTAGCCGGATTGGATTTAAGGAGTTACAATGAAATTTTTTGAGATAGCTTGGAGGATGTTTGATTCTAGTAGATTTTTCCACACT
legE --------------------------------TTAGAACCACAAATCTCAATCTTTTCTTAATGAAATGAAAAATCTTAATTGTACCATGTTTATGTTAAACACCTTACAATTGGTTGGAGAG 2129

legA GTTATGGGTATTTT----------------------AAATATAAATAATGTGGGTAATAATAGGGATGCATATTAGGGTTGGAGTCATAAATGTAACTGGGGATCCATGAAGGAATCATAAAAAAA
legE GAGGACCAACCGATGGGACAACATTGGGAGAAAGAGATTCAATGGAGATTTGGATAGGAGAACAACATTCTTTTTCACTTCAATACAAGATGAGTGCAACACTAAGGATATGTATGAGAC 2249

legA GAACAGCATCATCATCAGTGGCATAGTGAGTGAC----ATTGGAAGCAAATGAAGTGACATTATCATTGTGATACTTATAAGGAAAAATATCCGCTTAAAACACCATGTCTAGAAATGAA
legE TTTCAGAAGCTACGACAACATAGATGAGTGAGGTGGTGATTCCTAGCAAGAAAGACATTAGAGGAAGCCAAAATCGAACAAGGAAGACATCAAGGGCAAGAGACAGGACCATCCATCTCA 2369

legA TGTGCTATCAAGATACATGTCATAAAGATCATAAACCTCAGTTTTACATCAGACAAAAATTCACTTCCTAACTCTAGGGTCAAGCTT
legE GGAAAAGGAGCTTTGGGATAGTCCGAGAAGTTGTACAAGAAATTTTTTGGAGGGTGAGTGATGCATTGCTGGTGACTTTAACTCAATCAAAATTGAGAAAGAAAGAAAAGGGAGGGGGCT 2489

legE CACATGTGAATAGAAGGGAAACGGGAGAATTTTACAGTTTTGATCTAATGGGCATCCCAGCTAGTGGTAACATATTCACCATGTTTAACCTTCACGGAAGTGTGATAGGTAGGTTAGACA 2609

legE AATTTCTGTTGGATGGAACACAAGGAATTC
```

__FIGURE 12b__ Comparison between the 3′ flanking regions of *Leg E* and *Leg A* genes.

: = regions where the two genes are homologous

FIGURE 13 λ*Leg E1* and Pea DNA were cleaved with EcoRI to completion on a 0.5% agarose gel

A) λDNA cleaved with AvaI and BamHI
B) λ*Leg E1* DNA cleaved with EcoRI
C) Pea DNA cleaved with EcoRI

All sizes are in kb

FIGURE 14 Autoradiograph of the gel from Figure 13 after Southern blotting and hybridisation to the ^{32}P labelled insert from pDUB6. The nitrocellulose filter was washed to high stringency (0.1 x SSC, 65°C for 30 min)

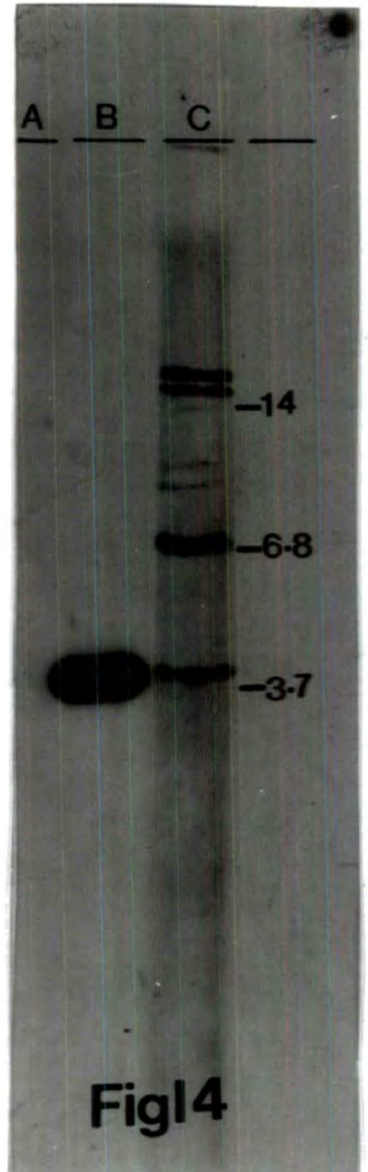

Fig14

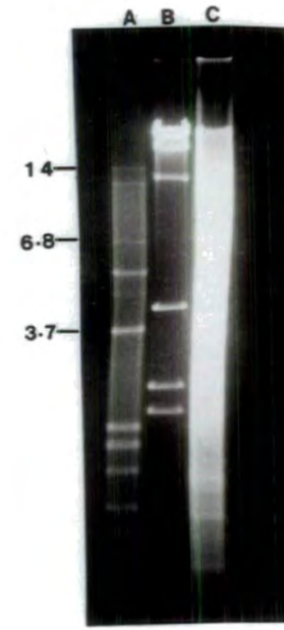

Fig13

3.2.6 DETERMINATION OF *LEG K* SEQUENCE

To produce single stranded templates for sequencing, restriction fragments from *Leg K* were subcloned in M13mp18 and 19. Twenty-six different fragments were sub-cloned, two of which were relatively large, a 1.6 kb EcoRV and 2.56 kb Sphl fragment, the others were much smaller (Figure 15). The determined sequence of *Leg K* gene and its 3' and 5' non-coding flanking sequences is presented in Figure 16. As shown in Figure 15 the fragments sub-cloned cover the whole region needed to be sequenced and overlap to allow the correct order to be established. Both strands of the DNA were fully sequenced.

3.3 CONSTRUCTION OF AN *ARABIDOPSIS THALIANA* GENOMIC LIBRARY

Due to the small amounts of material available, conventional methods for extraction of genomic DNA from *Arabidopsis* tissues did not prove successful. DNA was extracted from *Arabidopsis* using a modified method (see Methods section), which combined features of several methods usually used to isolate genomic DNA from plant tissues. Approximately 1 mg of DNA was obtained from 10 g of leaf tissues. The yield was not high but the DNA obtained was very clean. When the DNA was checked on agarose gel it ran slower than lambda DNA which indicated that it was very large in size and neither degraded nor sheared (Figure 17).

The EMBL3-BamHl arms, free of stuffer fragment were prepared by restriction of the vector followed by sucrose density gradient centrifugation (Maniatis *et al.*, 1982).

Sau3AI was chosen to partially digest *Arabidopsis genomic* DNA to yield a population of fragments that is close to random and yet can be cloned directly into BamHl EMBL3 arms.

Fig15

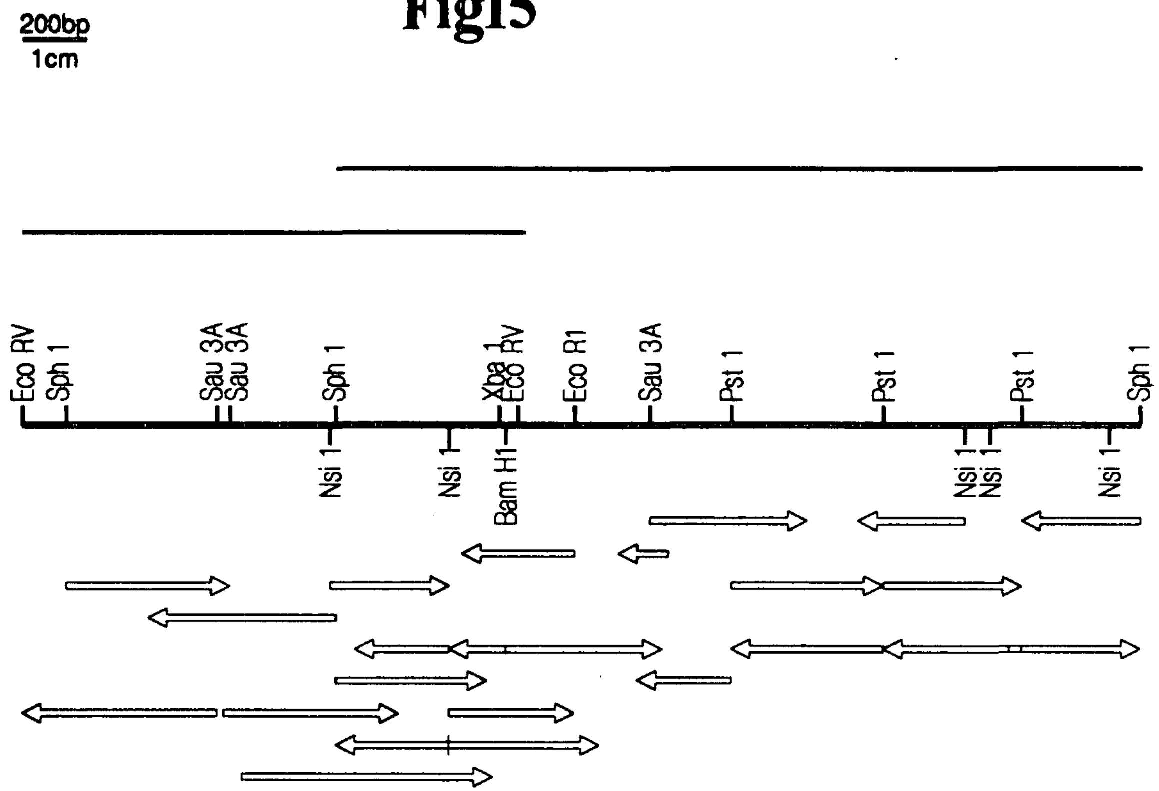

FIGURE 15 Sequencing map of *Leg K*, showing positions of restriction sites. Sequencing runs are indicated by arrows.

FIGURE 16 Sequence of *Leg K* and flanking regions. The sequence of *Leg J* is also given over those regions where the genes are homologous. The complete amino acid sequence encoded by *Leg J* is given; residues differing in *Leg K* are given above the sequence.

Fig 16

LegK ..........................GCATGCAAGTAAGTAATTAAAAAACCTTTATAGTAAAATATTATTATGCTTGATGTAAGCGTGTAGTTTATATCAAACTTCAATTTATAAATTT

LegK ACATAAATACTAAAAAGGTGTGAGGAATTTGAACACAAAATTAAGTGCACTAATACACACTAAAATTTACTTGTGCAGTCATCATTACGTCGTCTTCATCTTCTTATTTGCAGTATCCCC
LegJ ------------------------------GTTAACACAAGCTAAAATTTATTTGTGCAATCATCATCATGTCATCTTCATCTTCTAATTTG---------

LegK GACACAAGATGTCACACTCACTGAGAAATGAAAATTGAGCAAAATACATCACCAGCCAATCTTGAATTTACCTAAAAGAAAGACAACTCTATCTAAATTATTAAGT------------
LegJ ----------------------AAATGAAAATTTAGCAAAATACATAACCAGTCAATCTAGAATTTACCTAAAAGAGAGACAACTGTATCTATATTATATCAGGGAGTAATACACCA -406

LegK --AGTCGAGTGAAATATATGCTAGAGACAGTAAATTAATAATTGAATTAAAGAGATAAAT
LegJ GCAGTACATTTTGAGTGGAGGAGGCCAATTATTAAAGTTTATAAAGTAGTAAAACATGCAAGAGTCGAATGAAATATATGCTCTAGACAGTAAATTAATAGTTGAGTTAAAGAGATAAAT -286

LegK GCTATAAATAACCAGAAATAAGATTAAGTGTATATTGTGTACAAGGTAATAATGTTGTATTATACAGAGATTTATTTTTAATGGTCGGATCACTAAGTCCACCATGTCTGTTACTGATCT
LegJ --

LegK AACGACGAGATTTACTAATCCAATGGTAGAAACAAATATTTAGAGGTGAGACTTTAAAATTAAATTATGAATAGAAGTAATACAGTATTAGTATTAGTACATATTCAGTATTAGTATTAT
LegJ --

LegK GAATATAACTAATAATCACTGAATTGAAGAGATACGCAGAGAGTCGATGCAGAGAAGAGAAGTAGAGAATTGTATGGGACCATCTCCTGCAACATATAAGAATAGCAACAAATATTCATT
LegJ ------------------------------GCATAGAGTCGACGCAGAGAAAAGAACTAGAGAAGTGAAGGGGACCATCC------ACATATAAGAATACCAACAAATATTCATT -207

LegK CTGTTTCTCTGTGGTAATATGGATATATACTAATCATCATCTATCTGTGAAGAATGAATGAAGCGGTCACCTACGTCGCGTCTTACATATGATGTGTACCATATTAAGATTCCATAGCCA
LegJ --GTCTCTTTGTGGT-ATTTGGATATATACTAATT---ATCAATCTGTGAAGAATGAATGAAGCGGCTACTTGCGCTGCGTCCCACATATGATGTGTATCAATTTAGGACTCCATAGCCA -93
....(.........

LegK TGCATGCTCAACAATGTCACACACATTCTGTCACACGTGTTCCTCTCTCACTCTTCCCCTCTTCCTATAAATCACCACAACACAGCTTCTCCAATTCACCACTTCACTCATCAATCTCTC
LegJ TGCATGCTGAACAATGTCATACACATTCTGTCACACGTGTTCCTATCTCACCCTTCCCCTCTTCCTATAAATCACCACAACACAGCTTCTCCACTTCACCACTTCACTCACCAATCTCTC 28
"Legumin" BOX..)...(TATA BOX)..................^...........................

A.A. (V S F R
LegK ATTAGTATTAGTAGTGTATCATCACACTCACAGTGTCCAAACCTTCTCTATCTTTGTTTTCACTTTCCTTGCTACTCTTTGCAAGCGCATGTTTAGCAACTCGCTCTGAGTTTGACAGAC
LegJ CTTAGT......AGTTTATGATCAGAGTCACAATGTCCAAACCTTTTCTATCTTTGCTTTCACTTTCCTTGCTACTCTTTGCAAGCGCATGTTTAGCAACTAGCTCTGAGTTTGACAGAC 142
A.A. ....(M S K P F L S L L S L S L L L F A S A C L A :T S S E F D R 7

A.A. N N
LegK TCAACCAATGCCAACTAGACAACATCAATGCATTGGAACCTGACCACCGTGTTGAGTCCGAAGCCGGTCTCACTGAGACATGGAATCCAAATAACCCTGAGCTAAAATGCGCTGGTGTGT
LegJ TTAACCAATGCCAGCTAGACAGTATCAATGCATTGGAACCAGACCACCGTGTTGAGTCCGAAGCTGGTCTCACTGAGACATGGAATCCAAATCACCCTGAGCTAAAATGCGCCGGTGTGT 262
A.A. L N Q C Q L D S I N A L E P D H R V E S E A G L T E T W N P N H P E L K C A G V 47

A.A. L
LegK CACTTATCAGACGCACCATTGACCCTAATGGACTCCACTTGCCATCTTTCTCACCCTCTCCACAGTTGATTTTCATCATCCAAGGAAAGGGTGTTCTTGGACTTTCACTTCCTGGTTGTC
LegJ CACTTATTAGACGCACCATCGACCCTAATGGACTCCACTTGCCATCTTTCTCCCCCTCTCCACAGTTGATTTTCATCATCCAAGGAAAGGGTGTTCTTGGACTTTCATTTCCTGGCTGTC 382
A.A. S L I R R T I D P N G L H L P S F S P S P Q L I F I I Q G K G V L G L S F P G C 87

A.A. G I
LegK CCGAGACTTATGAAGAGCCACGTTCATCACAATCTAGACAAGGATCCAGGCAGCAACAAGGTGACAGTCACCAGAAGATTCGTCGATTCAGAAAAGGTGATATCATTGCCATTCCATCGG
LegJ CTGAGACTTATGAAGAGCCTCGTTCATCACAATCTAGACAAGAATCCAGGCAGCAACAAGGTGACAGTCACCAGAAGGTTCGTCGATTCAGAAAAGGTGATATCATTGCCATTCCATCGG 502
A.A. P E T Y E E P R S S Q S R Q E S R Q Q Q G D S H Q K V R R F R K G D I I A I P S 127

LegK GAATTCCTTATTGGACATATAACCATGGCGATGAACCTCTTGTTGCCATTAGCCTTCTTGACACTTCCAACATTGCAAACCAGCTCGATTCAACCCCAAGAGTAAGTGATAGTGTATCCA
LegJ GAATTCCTTATTGGACATATAACCATGGCGATGAACCTCTTGTTGCCATTAGTCTTCTTGACACTTCCAACATTGCAAACCAGCTCGATTCAACCCCAAGAGTAAGTAATAGTGTATCCA 622
A.A. G I P Y W T Y N H G D E P L V A I S L L D T S N I A N Q L D S T P R (.................

LegK TTCAT------------------------ACAGTATGCTCTTTCGATTATAACTTAAAAAGTTTCTAAT-----------------------------GTAAATATGTGTATGCAGG
LegJ TACATTACATTATCTCTTATAAATTGTTCATACAGCATGCTCATTCGATTATAACTTTAAAAGTTTCTAATGTATAATTTGTTATACTAAATCAATCACACGTAAATATGTGTATGCAGG 742
A.A. .... Intron-1 ....)

LegK TATTTTACCTTGGTGGAAACCCAGAAACAGAGTTCCCCGAAACACAGGAGGAACAACAAGGAAGGCATCGGCAAAAGCATAGTTACCCTGTTGGACGTAGGAGTGGACATCACCAACAAG
LegJ TATTTTACCTTGGTGGAAACCCAGAAACAGAGTTCCCCGAAACACAGGAGGAACAACAAGGAAGGCATCGGCAAAAGCATAGTTACCCTGTTGGACGTAGGAGTGGACATCACCAACAAG 862
A.A. V F Y L G G N P E T E F P E T Q E E Q Q G R H R Q K H S Y P V G R R S G H H Q Q 201

```
A.A.                                   V
LegK AAGAGGAATCCGAAGAACAAAACGAAGGTAACAGCGTGCTGAGTGGCGTCAGCTCAGAGTTTTTAGCACAAACGTTCAACACTGAAGAGGATACAGCGAAGAGACTTCGATCTCCACGAG
LegJ AAGAGGAATCTGAAGAACAAAACGAAGGTAACAGCGTGCTGAGTGGCTTCAGCTCAGAGTTTTTAGCACAAACGTTCAACACTGAAGAGGATACAGCGAAGAGACTCCGATCTCCACGAG 982
A.A. E  E  E  S  E  E  Q  N  E  G  N  S  V  L  S  G  F  S  S  E  F  L  A  Q  T  F  N  T  E  E  D  T  A  K  R  L  R  S  P  R  241

A.A.                                             N        E                             E
LegK ACGAAAGGAGTCAAATTGTGCGAGTTGAGGGAGGTCTCCGCATTATCAACCCCAAGGGGAAGGAAGAAGAAGAAGAAAAAGAACAGAGTCATTCTCACTCTCACAGAGAGGAGGAGGAAG
LegJ ACGAAAGGAGTCAAATTGTGCGAGTTGAGGGAGGTCTCCGCATTATCAAACCCAAGGGGAAGGAA---GAAGAAGAAAAAGAACAAAGTCATTCTCACTCTCACAGAGAGGAGAAGGAAG 1099
A.A. D  E  R  S  Q  I  V  R  V  E  G  G  L  R  I  I  K  P  K  G  K  E  -  E  E  E  K  E  Q  S  H  S  H  S  H  R  E  E  K  E  280

A.A.          -  -  -       -                                                                                      G
LegK AAGAAGAAG---------AAGATGAGGAG---AAACAAAGAAGTGAGGAAAGAAAGAATGGTTTGGAAGAAACTATCTGTAGTGCCAAAATTCGAGAGAACATTGCGGACGCTGCAGGTG
LegJ AAGAAGAAGAAGAAGAAGAAGATGAGGAGGAGAAACAAAGAAGTGAGGAAAGAAAGAATGGTTTGGAAGAAACTATCTGTAGTGCCAAAATTCGAGAGAACATTGCGGACGCTGCACGTG 1219
A.A. E  E  E  E  E  E  E  D  E  E  E  K  Q  R  S  E  E  R  K  N :G  L  E  E  T  I  C  S  A  K  I  R  E  N  I  A  D  A  A  R  320

A.A.                             R
LegK CCGACCTCTATAACCCACGTGCTGGTCGTATCAGAACTGCAAACAGTTTAACTCTCCCAGTCCTCCGCTATTTACGCCTCAGCGCTGAGTATGTTCGTCTCTACAGGGTGTGTATAGTAC
LegJ CCGACCTTTATAACCCACGTGCTGGTCGTATCAGCACTGCAAACAGTTTAACTCTCCCAGTCCTCCGCTATTTACGCCTCAGTGCTGAGTATGTTCGTCTCTACAGGGTAACTATTAAAC 1339
A.A. A  D  L  Y  N  P  R  A  G  R  I  S  T  A  N  S  L  T  L  P  V  L  R  Y  L  R  L  S  A  E  Y  V  R  L  Y  R <...........

LegK TAACTATTTAATCAATATATTTCCAATTGATGATT-GTTGAAAAAAATGAAA-TTTAATGAGCTAATTAATAACATGTATATATGTATATGCAGAATGGTATATATGCTCCACACTGGAA
LegJ --------TAATGTGTATATTTCCATGATATGATTAGTTACAT-AAATGATTTTTTAATAAACTAATCAATAACGTGTATGTATGTATATGCAGAATGGTATATATGCTCCACACTGGAA 1450
A.A. ......................... Intron-2 ......................................................> N  G  I  Y  A  P  H  W  N  368

A.A.                                                                          F        D  A
LegK CATAAACGCCAACAGTCTGCTGTACGTGATTAGAGGAGAAGGAAGAGTTAGGATTGTGAACTTCCAAGGAGACGCAGTGTTCGACAACAAGGTCAGAAAGGGACAGTTGGTGGTGGTACC
LegJ CATAAACGCCAACAGTCTGCTGTACGTGATAAGGGGAGAAGGAAGAGTTAGGATTGTGAATTGCCAAGGAAACACGGTGTTCGACAACAAGGTGAGAAAGGGACAGTTGGTGGTGGTACC 1570
A.A.  I  N  A  N  S  L  L  Y  V  I  R  G  E  G  R  V  R  I  V  N  C  Q  G  N  T  V  F  D  N  K  V  R  K  G  Q  L  V  V  V  P  408

A.A.                                                                                                       L           A
LegK ACAAAACTTTGTGGTGGCGGAACAAGCTGGGGAGGAAGAAGGATTAGAGTATGTGGTGTTCAAGACAAATGACAGAGCTGCGGTTAGCCACGTACAACAGGTGCTTAGGGCCACTCCTGC
LegJ GCAAAACTTTGTGGTGGCGGAACAAGCTGGGGAGGAAGAAGGATTAGAGTATGTTGTGTTCAAGACAAATGACAGAGCTGCTGTTAGCCACGTACAACAGGTGTTTAGGGCCACTCCTTC 1690
A.A.  Q  N  F  V  V  A  E  Q  A  G  E  E  E  G  L  E  Y  V  V  F  K  T  N  D  R  A  A  V  S  H  V  Q  Q  V  F  R  A  T  P  S  445

A.A.                                                                                            Q
LegK AGAGGTTCTTGCAAATGCTTTTGGTCTTCGTCAACGCCAAGTCACGGAGTTAAAGCTCAGTGGAAACCGTGGCCCTCTGGTTCACCCTCAGTCGCAATCTCAATCTCATTGAGATGATGC
LegJ AGAGGTTCTTGCAAATGCTTTTGGTCTTCGTCAACGCCAAGTCACGGAGTTAAAGCTCAGTGGAAACCGTGGCCCGCTGGTTCACCCTCGGTCTCAATCTCAATCTCATTGAGATGATGC 1810
A.A.  E  V  L  A  N  A  F  G  L  R  Q  R  Q  V  T  E  L  K  L  S  G  N  R  G  P  L  V  H  P  R  S  Q  S  Q  S  H  *>.......

LegK TAT------GGAGTATAATAATGAGATGGCCATCTTATCTT-----------AAATAATAAATTTTGAATGTACTGTAGAGAAGAATTTCAGTTCCGATAATAAAACAATAAAGTATGGC
LegJ TATGATAATGCAATACAATAACAAGATGGCCATCTTGTCTTGAATAATAATGATAAAATAAATTTTGAATGAACTGTAGTAAAAAATTTCACTTCCTATAATAAAACAATAAAGTATCGC 1930
     ...................................................(PolyA+).................................(PolyA+)(PolyA+)......

LegK CTTAAAATCCCAATCTTAATCTAAATTTGTATGCATCTATA-AGGGGCGAATAACACTAGTTTTGTT-CACCTTGCAATTGCCATAATAAAATG-CATA-CACTTTTTACTATTGCTTAT
LegJ CCTACTACCCTGATCTTACTCTGAATTTGTATGCATGTAAAGAGGGGTGAATAACAATTGGTTTTGTACACCTTCCAATTGCCATAATAAAATGGCATATCACTTTTTAAAAAATTCTCC 2050
     ...............................................................................(PolyA+)...........................

LegK ATGTTTGTTTGAATCATATAAAAAACACAACTACAAATCTGCATTTTTCTTCGGCATTTGATTATATATCTGCAGGAAACCGATTTCATGACGCAAACCCTGATGTTTTTCATTCCATGC
LegJ TTCATTATCTTCTTCATTAGTACTAATGATGAATTGTCTCAATAATAATATCAGCTTTTTGAATACAGCAACGAGACAGCAAACTTTAACAATCACAATTATAAGGTAGTGTTATTAATT 2170

LegK AAATACCTTCAATGAACCGAGTTGAAAAACTCCAAGCGGGAAGTTATATTCTACAAGAGAATATGATGGCTGTATAGAGGAACAAAGTAAGAACCAGCAAATGAAATACATAAATTGGTA
LegJ GTTTTCTAAAAATCTTTTCTGTGTGACGGAAGGAATCATTCATTTCCTTATCAATATAATTGTCAAACTATAAAGGATAAAAATAAATACTTGAGAGGCAAGAGGTAAGAAGATGTGGTT 2290

LegK TCTGAAATGCTTCAAACACAAACATGACTCCTAGTAAATATGACCTGCTAACTAACTTATATTCTGAAAGCCCTTCTTTTTTCTTCAACCAGTTATTAACTCTAATAAAATGATTCCAAA
LegJ AATCCCACTACTCGCTAGAGAATGAACACACTTGTCATGTTGATAAAAATGAGAACCGAAAAATCTACAAATAAATACTAGTTGATTAAATAAAGAATTTGTATTGTTGATAAATACTGA 2410

LegK TGCATATACATCTACCTTACACTTCCAGCATCAAAATACTCCGTTGCAATGTCCGTGACATGTATCTTGAGTAATAAGCAGCCCTAATTTGCATGC
LegJ ATTC....(2414)
```

FIGURE 17 Checking *Arabidopsis* genomic DNA

1.5 μg samples of lambda DNA and *Arabidopsis* genomic DNA were analysed on a 0.5% agarose gel

A - Lambda DNA
B - *Arabidopsis* genomic DNA

Size is in kb

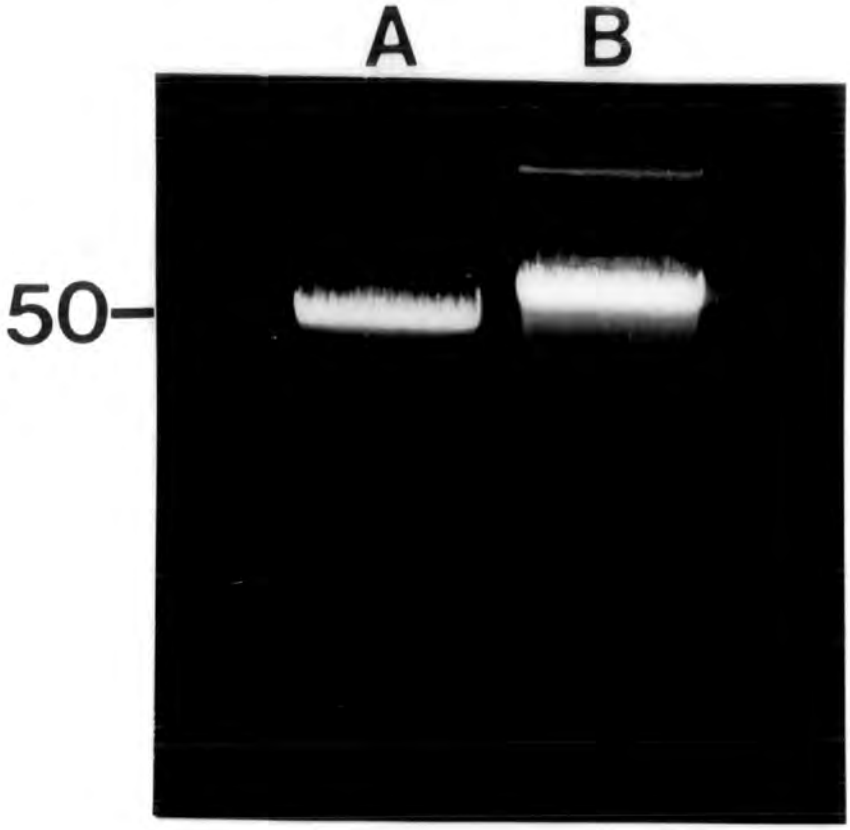

Fig17

To establish the conditions for partial digestion of genomic DNA to produce the maximum intensity of fluorescence in the 10-20 kb region when run on an agarose gel, a range of digestions were achieved using the same amount of DNA and the same incubation period, but with different amounts of enzyme ranging from 0.1 to 0.003 U of enzyme per µg of *Arabidopsis* DNA (Figure 18). The maximum amount of DNA in the correct size range appeared to be in track D. Therefore the digestion shown in that track, which used 0.025 U of Sau3A I per µg of DNA, was chosen as the optimal digestion. On the basis of this, 3 large scale digestions were performed using 0.5 µl and 2 times the amount of Sau3AI used in the optimal digestion. At the end of the incubation, DNA from all 3 tubes was pooled together, ethanol precipitated and resuspended in 60 µl TE buffer. Along side a size marker the DNA was loaded onto a 0.4% agarose gel and run overnight (Figure 19). The agarose containing the DNA in the region 10-20 kb was cut out of the gel, electroeluted, carefully cleaned and resuspended in 10 ul of TE buffer pH 7.6.

Pilot ligation mixtures containing different ratios of genomic DNA to vector DNA were set up and subsequently 1 µl aliquots were checked on a gel (Figure 20) and 3 µl aliquots were *in vitro* packaged to establish the conditions for the large-scale ligation which gives optimum number of plaques upon *in vitro* packaging.

The number of plaques obtained from various ligation mixtures after *in vitro* packaging and plating on 2 different bacterial strains (Q359 and K803) is shown in Table 3.

FIGURE 18 Agarose gel analysis of partial digests of *Arabidopsis* genomic DNA. All digestions were at 37°C for 1 hr. 3 ug of DNA were used in all digestions

A) *Arabidopsis* DNA uncleaved
B) *Arabidopsis* DNA cleaved with 1.0 U Sau3AI
C) *Arabidopsis* DNA cleaved with 0.5 U Sau3AI
D) *Arabidopsis* DNA cleaved with 0.25 U Sau3AI
E) λDNA cleaved with AvaI BamHI
F) *Arabidopsis* DNA cleaved with 0.125 U Sau3AI
G) *Arabidopsis* DNA cleaved with 0.0625 U Sau3AI
H) *Arabidopsis* DNA cleaved with 0.03125 U Sau3AI

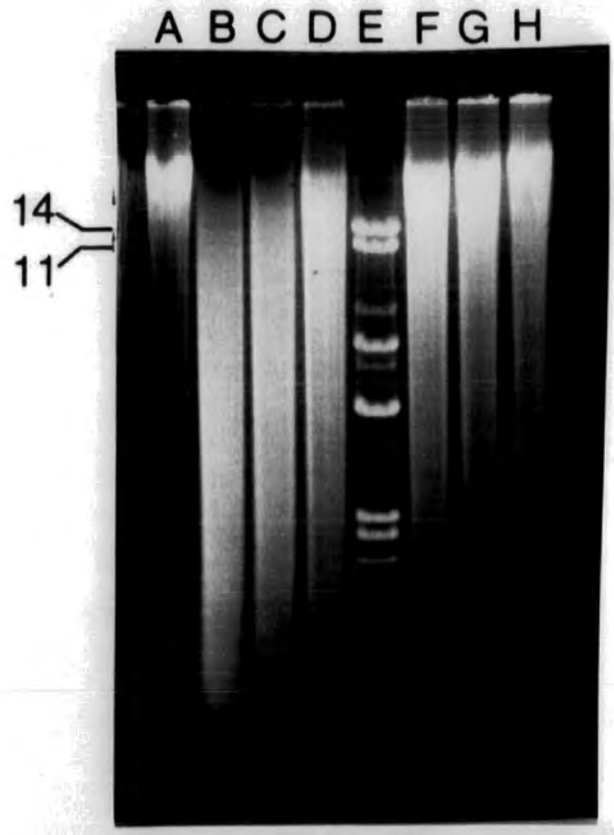

Fig18

FIGURE 19 Agarose gel analysis of the large-scale partial Sau3AI restriction digestion of *Arabidopsis* DNA

A) λDNA cleaved with AvaI and BamHI
B) *Arabidopsis* DNA partially cleaved with Sau3AI

All sizes are in kb

FIGURE 20 Agarose gel analysis of the ligated and unligated mixtures of EMBL3 arms and *Arabidopsis* DNA

A) 1 µl of mixture 1 unligated
B) 1 µl of mixture 2 unligated
C) 1 µl of mixture 3 unligated
D) 1 µl of mixture 4 unligated
E) 0.5 µg of alkaline phosphatase insert DNA ligated
F) λDNA
G) 1 µl of mixture 1 ligated
H) 1 µl of mixture 2 ligated
I) 1 µl of mixture 3 ligated

Fig19

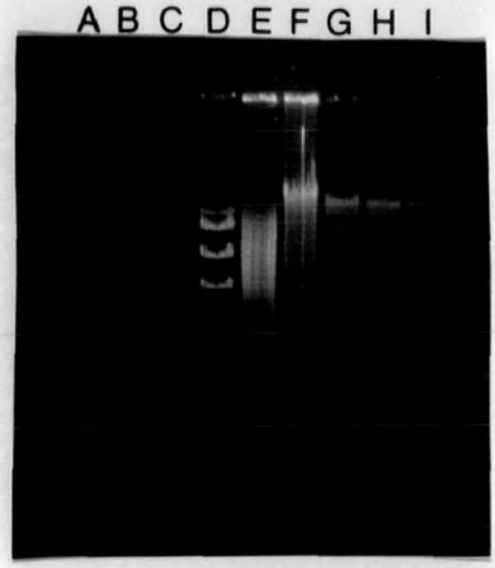

Fig20

Table 3

Ligation Mix	Arms:Insert	Host	No. of Plaques
1	1:1	K803	1.1×10^4
		Q359	8.5×10^3
2	2:1	K803	6.3×10^3
		Q359	5.0×10^3
3	1:2	K803	4.8×10^3
		Q359	3.6 x 10+3
4	Arms only	K803	Nil
		Q359	Nil

3 µl of each ligation mixture was *in vitro* packaged as described, using packaging extract, and aliquots were plated on bacterial hosts as stated.

In order to check the host strains used (i.e. K803 and Q359), cells were transfected with EMBL3 vector DNA. A large number of plaques was obtained with K803 while no plaques were obtained from Q359 cells. This shows that the selection system in Q359 is viable.

The optimum ratio of arms to insert and other optimised conditions (in terms of enzyme, ATP and incubation temperature and time) were applied to set up a large-scale ligation mixture. The ligated DNA was *in vitro* packaged in 3 tubes, the contents of two tubes were mixed with 1.2 ml of K803 cells and plated out on 4 megaplates (20 x 20 cm). Approximately 8×10^4 plaques were obtained. Phages were extracted overnight in 0.01 M MgSO4. When the phage stock was titrated using Q359 and P2392 as host cells, it was observed that the number of plaques obtained when Q359 cells were used as host cells was about 3-4% less than the number obtained when P2392 cells were used. The decrease was consistent in all the dilutions used (10^{-2}, 10^{-3}, 10^{-4}, 10^{-5} and 10^{-6}), and suggests that the background of EMBL3 vector was $<$ 5%. The titre of the stock was calculated at about 1.8×10^8 in Q359.

The content of the third tube of the *in vitro* packaged ligated DNA mixture was reserved to be used to screen the library for cell wall extensin sequence-containing genomic clones.

3.4 SCREENING *ARABIDOPSIS THALIANA* GENOMIC DNA FOR CELL WALL EXTENSIN SEQUENCES

Before the work on screening *Arabidopsis* genomic library for extensin sequences was started, it was intended to screen the genomic DNA of *Arabidopsis* for those sequences.

Arabidopsis genomic DNA was restriction digested with EcoRI, BglII, BamHl and HindIII and run on a 0.4% agarose gel (Figure 21), blotted on nitrocellulose filter and probed with pRR_t566 probe which contains 700 bp of cDNA of the coding region of extensin from oilseed rape (*Brassica napus* L.) (Evans *et al.*, 1990). Obviously the number and size of bands representing fragments hybridising to the probe varied from one digest to another. In the track containing the EcoRI digest there were 4 or 5 bands (Figure 22) ranging in size from approximately 5.5 to 10 kb. With BglII, 7bands of 3.6 to 10 kb in size were evident, as were 3 bands ranging in size from 9 to 15 kb when BamHI was used. The track contained the HindIII digest 3 bands which are 3.4, 6.6 and 9 kb in size. This experiment shows that the pRR_t366 probe detects homologous sequences in *Arabidopsis* genomic DNA.

3.5 SCREENING *ARABIDOPSIS THALIANA* GENOMIC LIBRARY FOR CELL WALL EXTENSIN SEQUENCES

To avoid any error in the representation of the sequences in the library the content of the reserved tube of the treated arms and insert DNA

FIGURE 21 *Arabidopsis* DNA was restriction digested to completion and analysed by electrophoresis on an agarose gel

A) *Arabidopsis* DNA cleaved with EcoRI
B) *Arabidopsis* DNA cleaved with Bgl2
C) *Arabidopsis* DNA cleaved with BamHI
D) λDNA cleaved with AvaI and BamHI
E) *Arabidopsis* DNA cleaved with HindIII
F) *Arabidopsis* DNA uncleaved

All sizes are in kb

FIGURE 22 Autoradiograph of the gel from Figure 21 after hybridisation to the ^{32}P labelled insert from pRR_t566

All sizes in kb

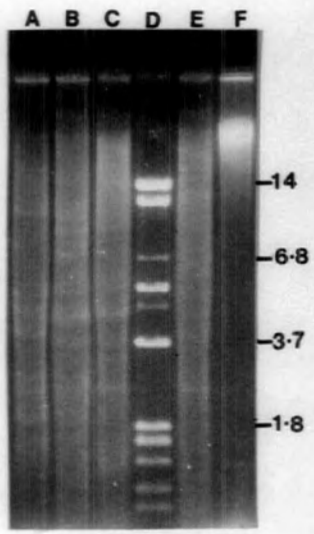

Fig 21

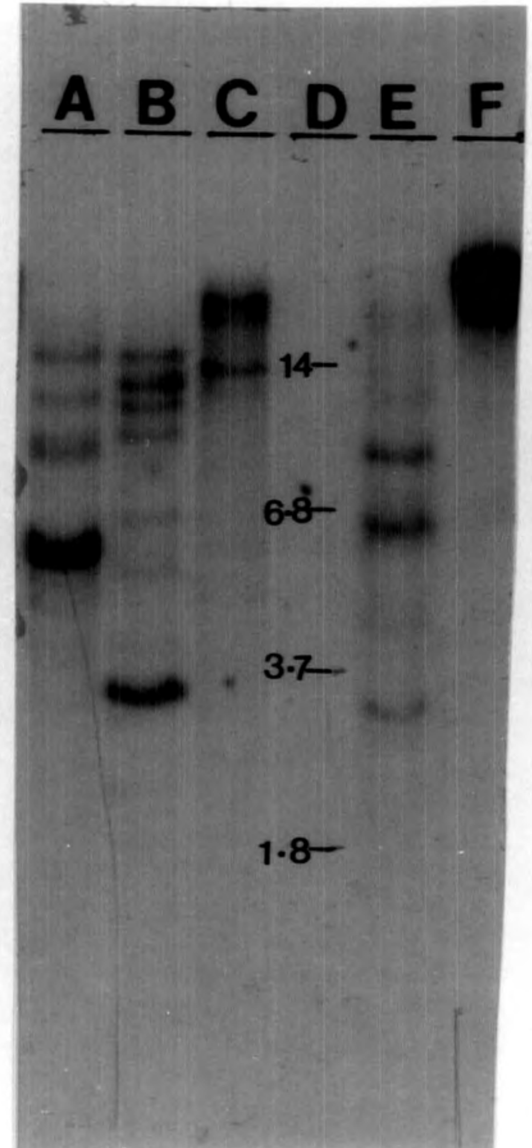

Fig 22

was *in vitro* packaged, plated out on megaplates, and incubated at 37°C for about 10 hr until the phage plaques had a diameter of 1 mm, and were just beginning to come into contact with each other. Plaques were transferred onto nitrocellulose filters and probed with pRR_t566. 6 positive plaques were picked from the first screen and all six were further screened and purified until more than 95% of the plaques were positive (Figure 23). A single well isolated positive plaque was picked from each plate of the six representing the six independent positives and stocks were made for all of them.

3.6 RESTRICTION MAPPING OF λ*ExtA*

One of the genomic clones isolated was designated λ*ExtA*. To restriction map the DNA, a series of single and double restriction digestions was carried out (Figure 24). EcoRI, BamHI, HindIII, BglII, SalI were among the restriction enzymes chosen to map the clone. The SalI digestion has shown one band in addition to the two arms of the vector; the size of the fragment was approximately 11.5 kb which gave an indication about the size of the insert. This was confirmed when the size of the whole insert was calculated from the two fragments which were seen in addition to the arms in the BamHI digest; they were 2.8 and 8.7 kb in size which when added together give the total of 11.5 kb.

The hybridisation data have indicated that the targetted sequence was located in the middle of the insert, and the smallest single fragment to hybridise to the probe was a 3.2 kb EcoRI-BamHI fragment (Figure 25). A partial restriction map of λ*ExtA* is shown in Figure 26.

3.7 SUB-CLONING OF λ*ExtA*

The 3.2 kb EcoRI-BamHI fragment which hybridised to the probe and the 4.5 kb EcoRI fragment were sub-cloned in pUC18 and the recombinant plasmids were called pL01 and pL02 respectively.

<u>FIGURE 23</u> Purification of extensin clones from *Arabidopsis thaliana* genomic library.

These are autoradiographs of the final stage of plaque lifting of:

A - λ*ExtA*
B - λ*ExtB*
C - λ*ExtC*
D - λ*ExtD*
E - λ*ExtE*
F - λ*ExtF*

Filters were washed at 0.3 x SSC, 65°C for 30 min

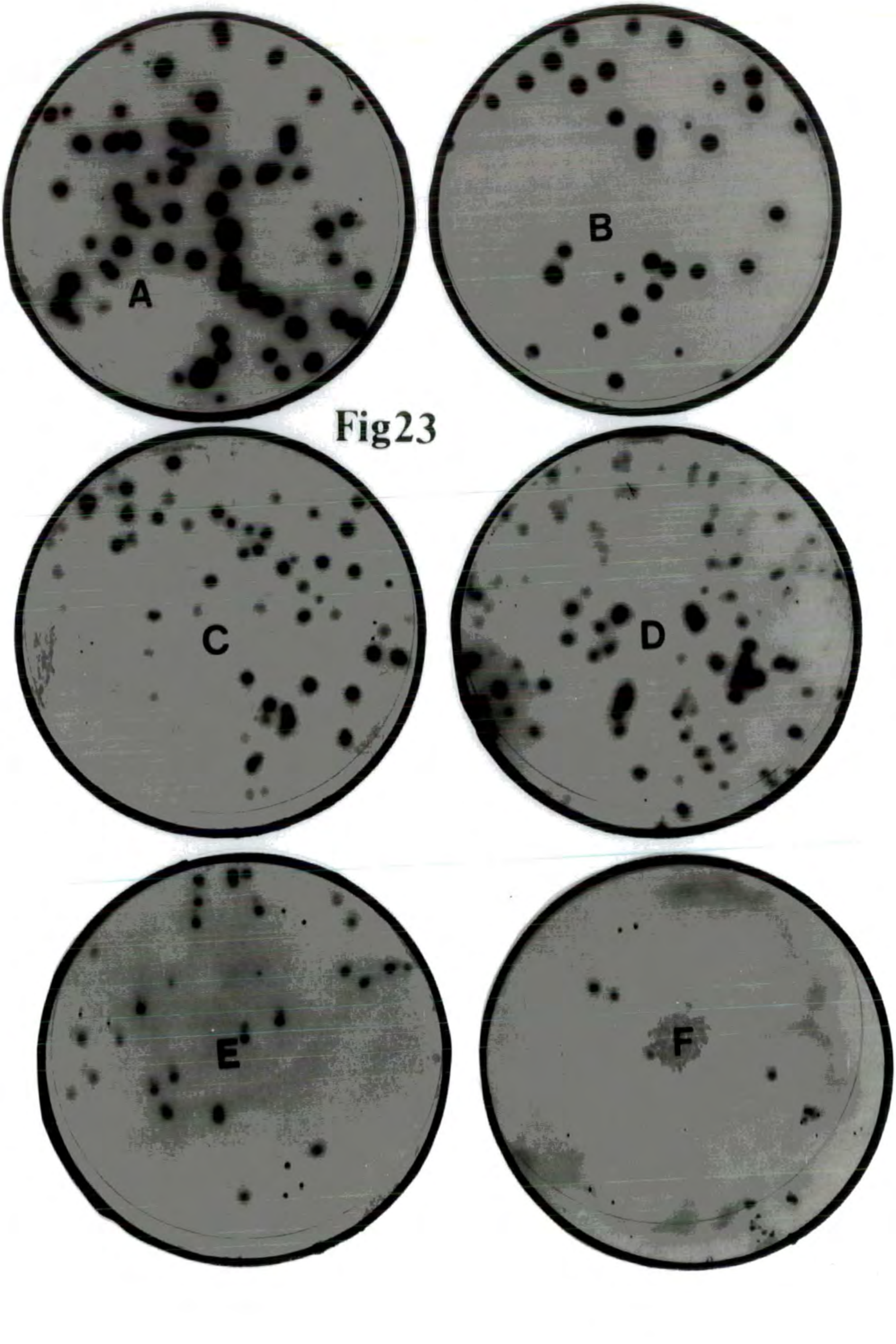

Fig23

FIGURE 24 Restriction enzyme analysis of λ*ExtA*

A) λDNA cleaved with AvaI and BamHI
B) λ*ExtA* DNA cleaved with Bgl2
C) λ*ExtA* DNA cleaved with Bgl2 and SalI
D) λ*ExtA* DNA cleaved with EcoRI
E) λ*ExtA* DNA cleaved with EcoRI and SalI
F) λ*ExtA* DNA cleaved with SalI
G) λ*ExtA* DNA cleaved with SalI and BamHI
H) λ*ExtA* DNA cleaved with BamHI

FIGURE 25 Autoradiograph of the gel from Figure 24, after Southern blotting and hybridisation to the ^{32}P labelled insert from pRR_t566

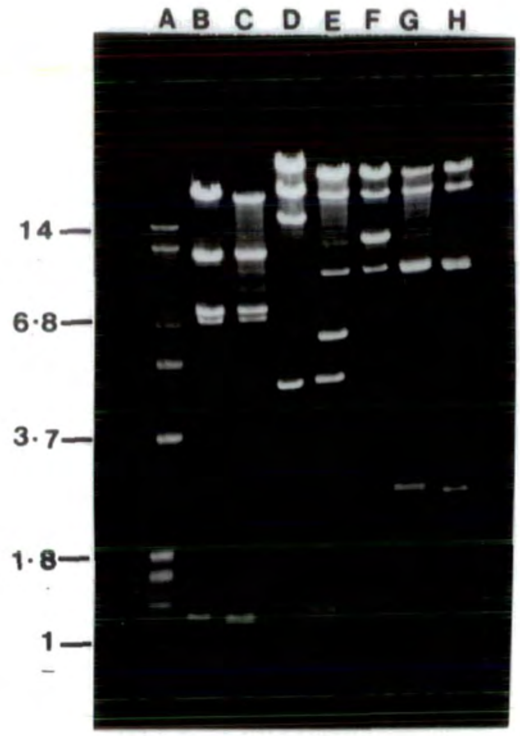

Fig24

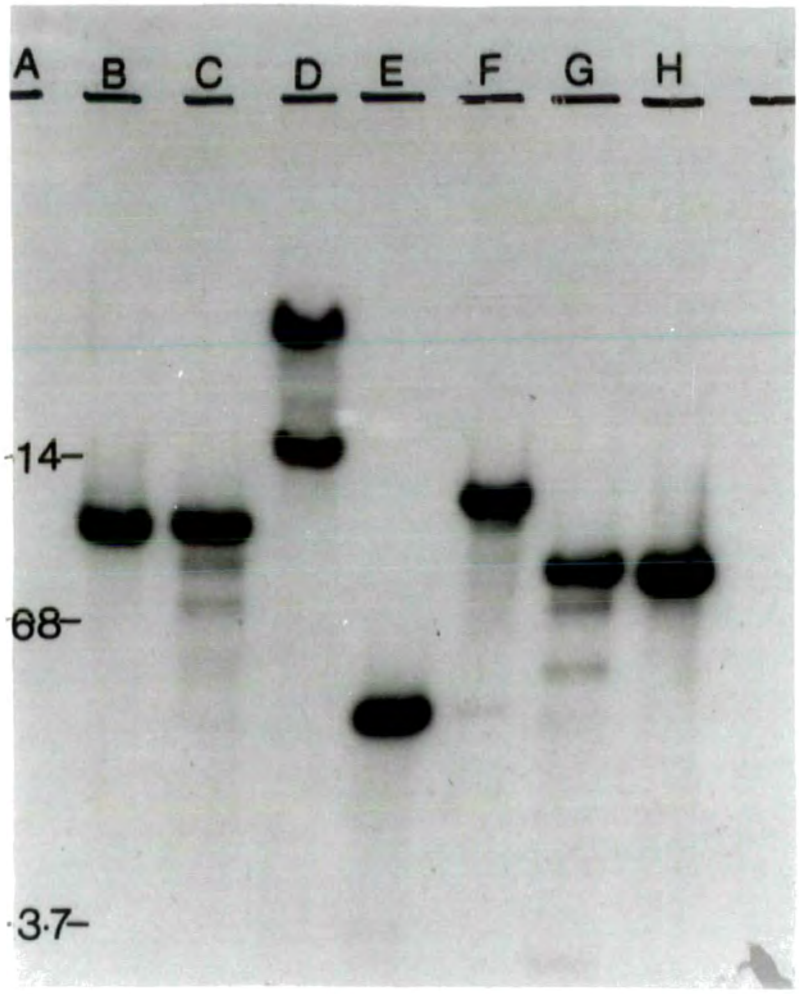

Fig25

Fig26

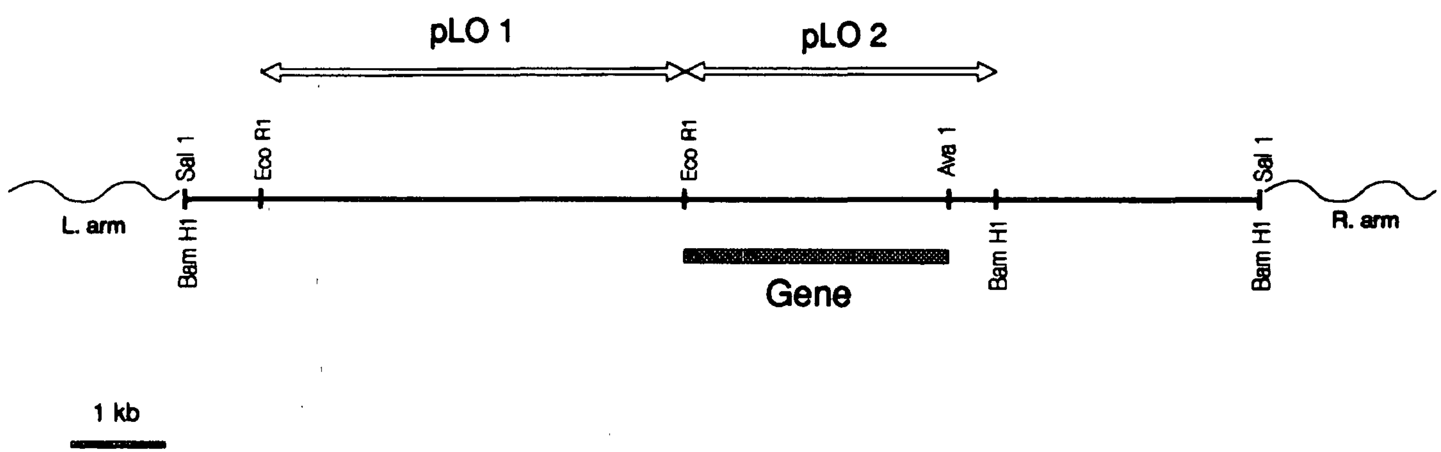

__FIGURE 26__ Partial restriction map of λ*ExtA*. Bar indicates the fragment that hybridises to cDNA probe pRR_t566. Arrows indicate the fragments sub-cloned in pUC18.

3.8 IDENTIFICATION OF A RESTRICTION FRAGMENT IN GENOMIC DNA CORRESPONDING TO λ*ExtA*

In order to check the occurrence in *Arabidopsis* genomic DNA of a sequence equivalent to the 3.2 kb EcoRI-BamHI fragments, λ*ExtA* and *Arabidopsis* genomic DNA were restriction digested to completion with EcoRI-BamHI and run on 0.5% agarose gel (Figure 27), transferred onto nitrocellulose filter and probed with the insert from pRR_t566. The autoradiograph showing the bands hybridising to the probe is shown in Figure 28. Obviously the 3.2 kb EcoRI fragment from λ*ExtA* corresponded to one of at least 4 fragments from *Arabidopsis* genomic DNA.

3.9 RESTRICTION MAPPING OF λ*ExtB* GENOMIC CLONE

Like λ*ExtA*, λ*ExtB* was restriction digested with an assortment of restriction enzymes in single and double digests. The size of the insert was calculated from the size of the single band obtained from the BamHI digest, which was about 12.5 kb in size. Also, when digested with SalI, two bands were obtained which were 7.2 kb and 5.2 kb in size which when added together give the sum of 12.4 kb. The data obtained from hybridisation to pRR_t566 had shown that the region homologous to the probe was located within the 5.6 kb EcoRI fragment. One of the mapping gels used for this analysis is shown in Figure 29. Southern blotting of these fragments and subsequent hybridisation against the insert of pRR_t566 gave the autoradiograph shown in Figure 30. The partial restriction map of λ*ExtB* is shown in Figure 31.

3.10 SUB-CLONING OF λ*ExtB*

In order to sub-clone fragments that cover most of the insert in λ*ExtB*, the 2 and 5.6 kb EcoRI fragments were inserted into pUC18 and designated pL05 and pL06 respectively.

FIGURE 27 *Arabidopsis* DNA was restriction digested with EcoRI and BamHI to completion and run a 0.5% agarose gel against pL02 and λ*ExtA* cleaved with the same enzymes.

A - pL02 DNA cleaved with EcoRI and BamHI
B - λ*ExtA* DNA cleaved with EcoRI and BamHI
C - *Arabidopsis* DNA cleaved with EcoRI and BamHI

FIGURE 28 Autoradiograph of the gel from Figure 27, after Southern blotting and hybridisation to pRR_t566. The filter was washed at high stringency (0.1 x SSC, 65°C for 30 min)

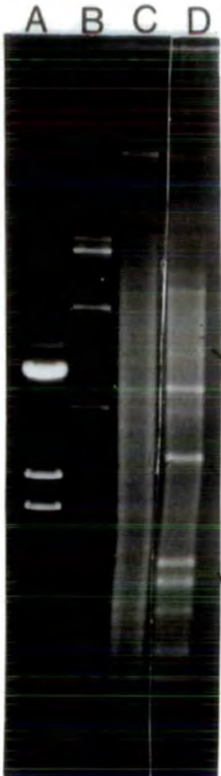

Fig27

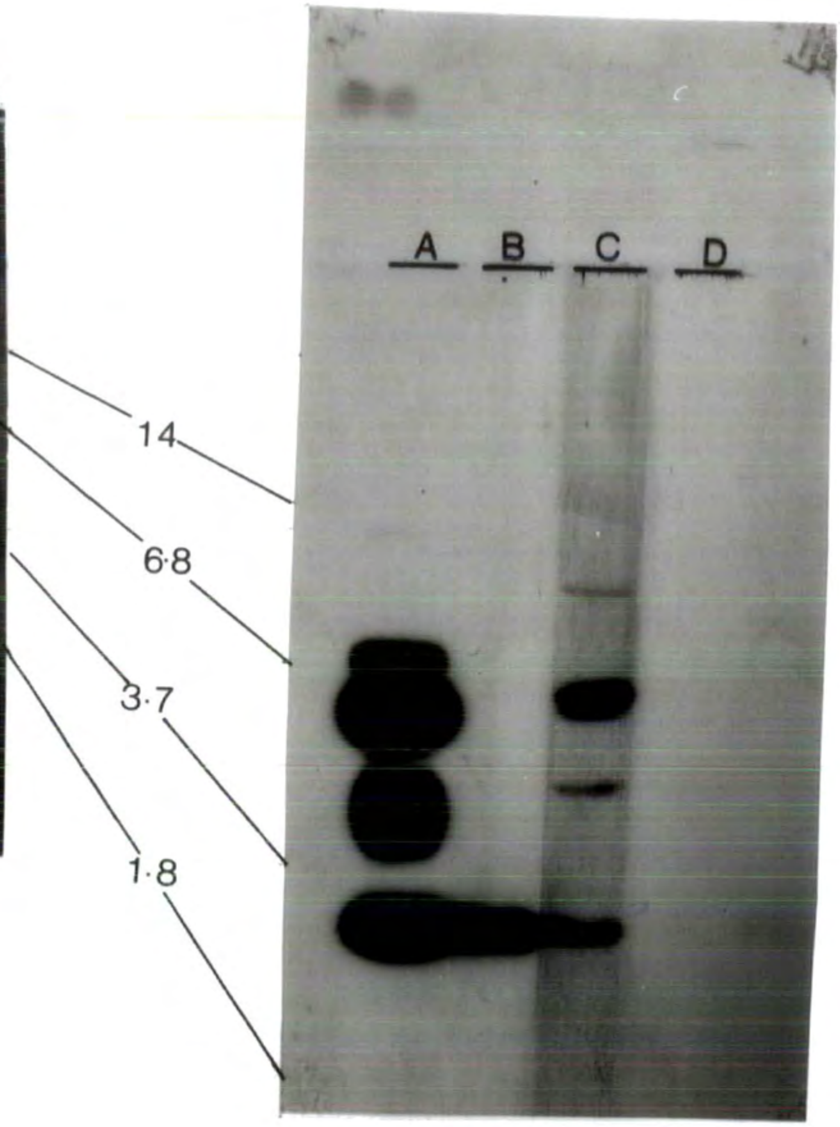

Fig28

FIGURE 29 Restriction enzyme analysis of λ*ExtB*

A) λ*ExtB* DNA cleaved with Bgl2
B) λ*ExtB* DNA cleaved with Bgl2 and BamHI
C) λ*ExtB* DNA cleaved with EcoRI
D) λ*ExtB* DNA cleaved with BamHI and EcoRI
E) λ*ExtB* DNA cleaved with BamHI
F) λ*ExtB* DNA cleaved with BamHI and SalI
G) λ*ExtB* DNA cleaved with SalI
H) λDNA cleaved with AbaI and BamHI

FIGURE 30 Autoradiograph of the gel from Figure 29 after Southern blotting and hybridisation with the ^{32}P labelled insert from pRR_t566. Hybridisation was at high stringency (0.1 x SSC, 65°C for 30 min)

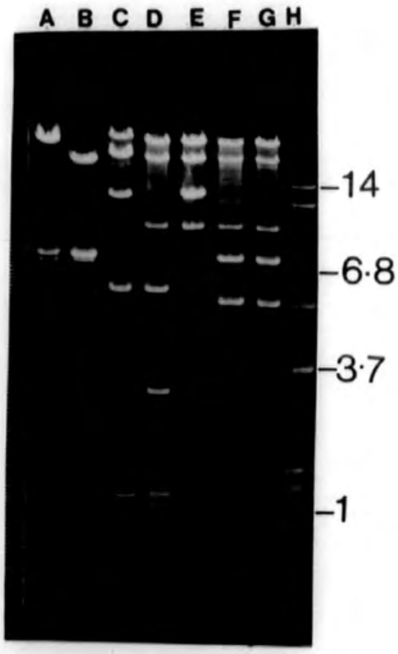

Fig29

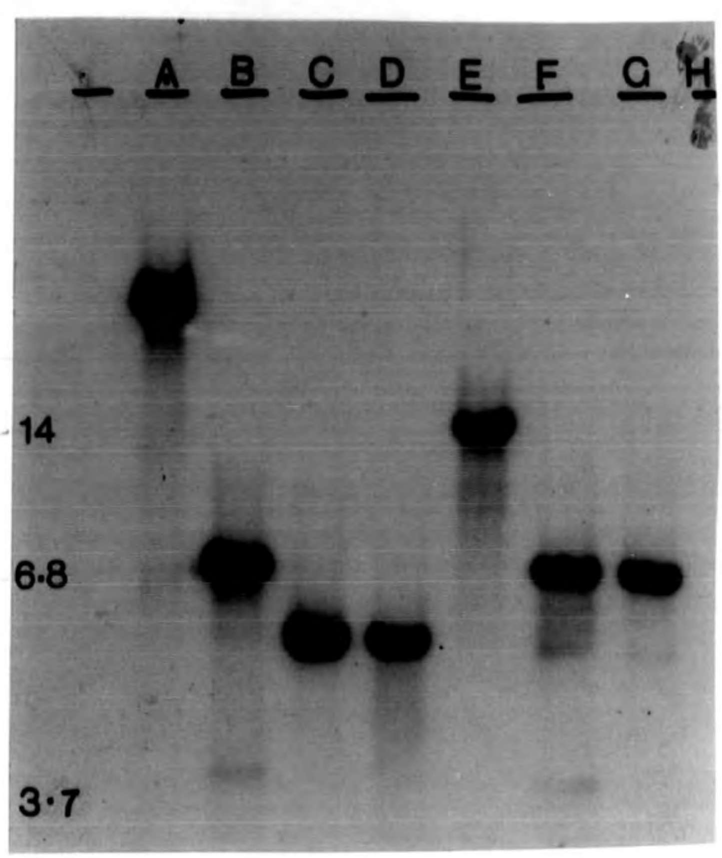

Fig3O

Fig31

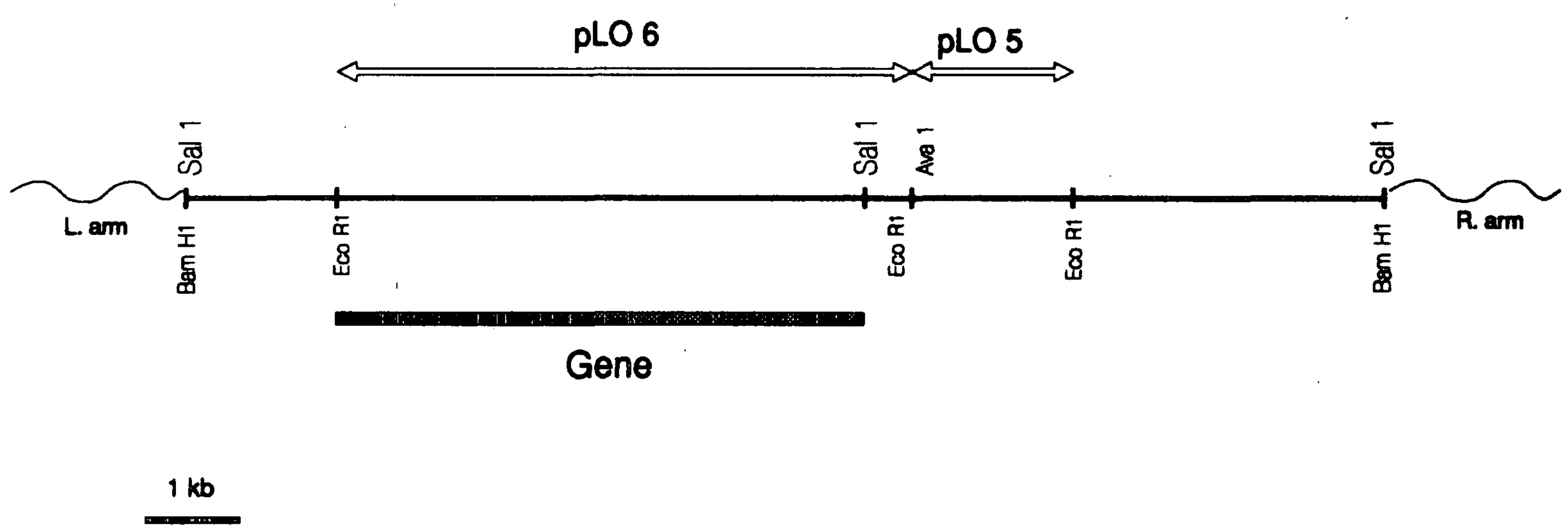

FIGURE 31 Partial restriction map of λ*ExtB*. Bar indicates the fragment hybridises to cDNA probe pRR_t566. Arrows indicate fragments sub-cloned in pUC18.

3.11 SEQUENCE DETERMINATION OF *ExtA*

In order to determine the sequence of *ExtA*, the pL02 sub-clone was mapped with respect to further restriction enzyme cleavable sites by a combination of hybridisation analysis and an examination of the sizes of fragments produced. pL02 was also restriction digested with several enzymes in single and double digestions. Restriction enzymes used were HincII, HindIII, EcoRV, EcoRI-HindIII and EcoRI-HincII (Figure 32). pUC18 was included as an aid to mapping. Subsequently these fragments were southern blotted and hybridised against the insert from pRR_t566, the results of this are shown in Figure 33. The sizes of the hybridising fragments are given below:

RsaI	1.5 kb
EcoRV	0.95 kb
HincII	2.2 kb
HindIII	1.6 kb

In order to sub-clone fragments from the insert in pL02 four different strategies were used:

3.11.1 Sau3AI SHOTGUN OF THE WHOLE INSERT

To achieve this, pL02 was restriction digested with EcoRI-BamHI. The 3.2 kb insert, was purified out of agarose gel and then restriction digested with Sau3AI. Subsequently the restricted DNA was shotgunned into M13mp18 restricted with BamHI. This resulted in the sub-cloning of several fragments which were subsequently sequenced.

3.11.2 SUB-CLONING OF DIRECTED FRAGMENTS

The useful sites (RsaI, EcoRV, HincII and HindIII) were used to sub-clone fragments in pUC18 and 19 vectors (Figure 34) and were sequenced directly from pUC vectors.

FIGURE 32 Restriction enzyme analysis of pLO2

A) pLO2 DNA cleaved with RsaI
B) pUC18 DNA cleaved with EcoRI
C) pUC18 DNA cleaved with RsaI
D) pLO2 DNA cleaved with HincII
E) λDNA cleaved with AvaI and BamHI
F) pLO2 DNA cleaved with HindIII
G) pLO2 DNA cleaved with PUUI
H) pL02 DNA cleaved with EcorV
I) pLO2 DNA cleaved with Bgl2

Figure 33 Autoradiograph of the gel from Figure 32 after Southern blotting and hybridisation with ^{32}P labelled insert from pRR_t566. Hybridisation was at high stringency (0.1 x SSC, 65°C for 30 min)

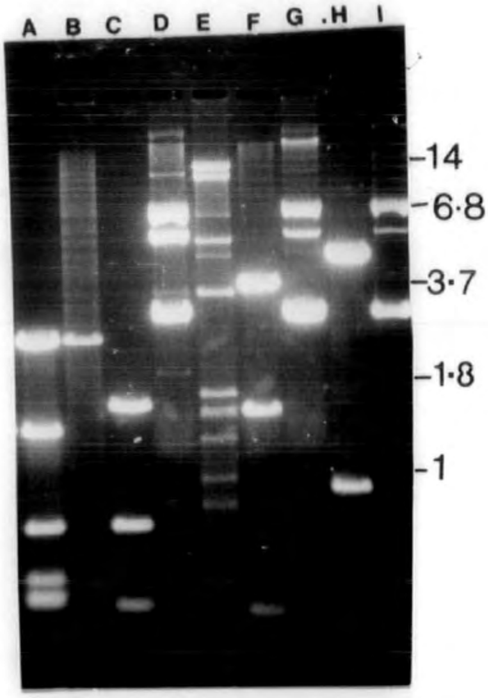

Fig 32

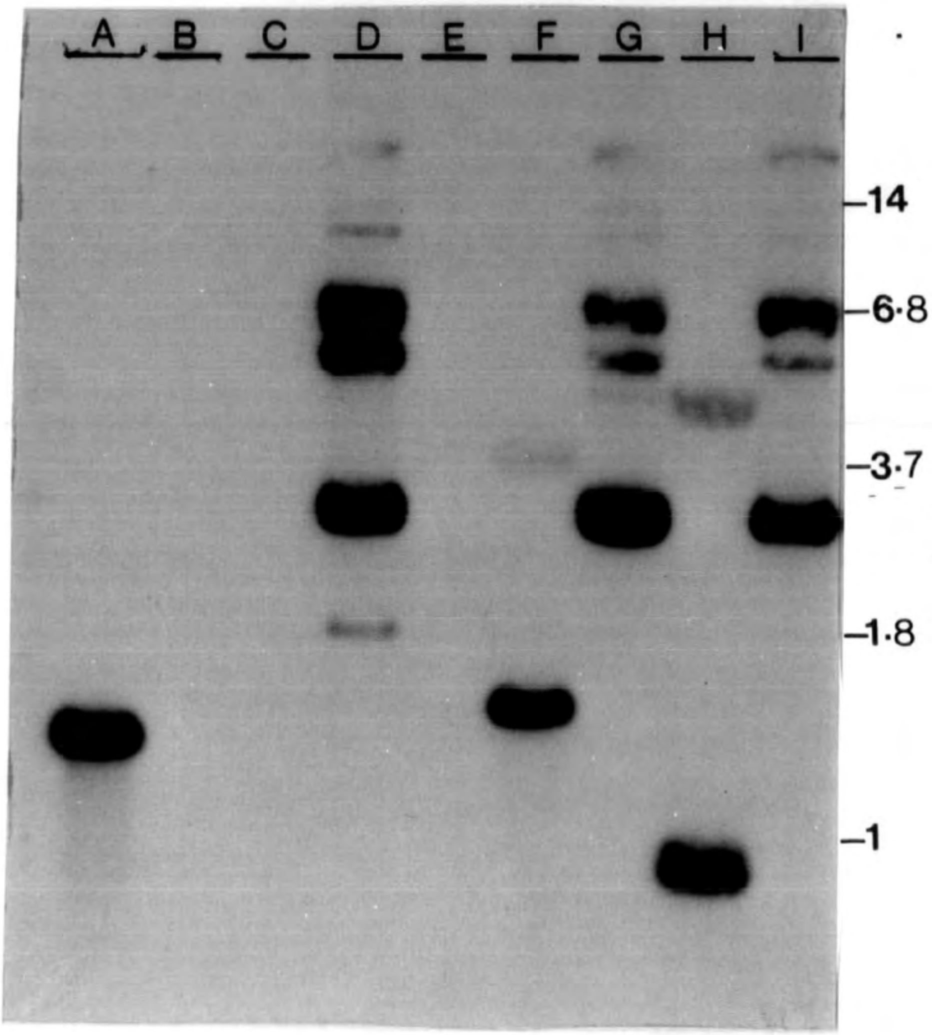

Fig 33

Fig34

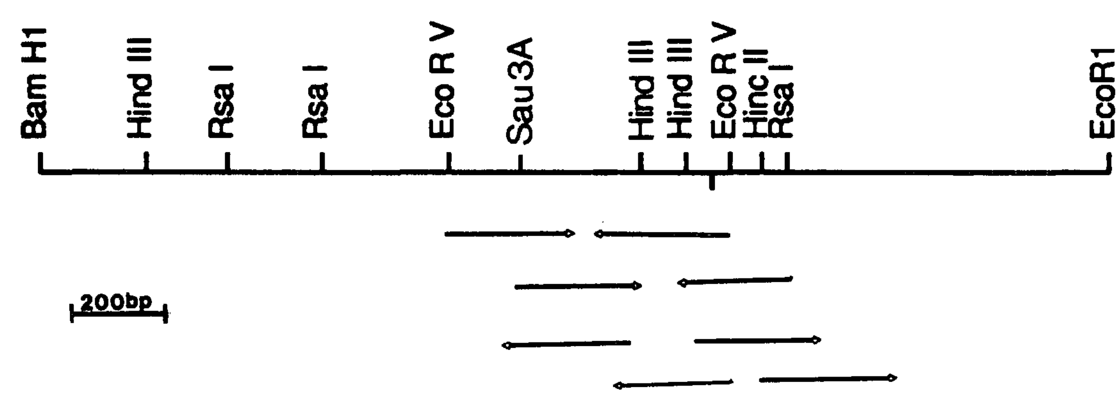

FIGURE 34 Restriction and sequencing map of pLO2. Sequencing runs are indicated by arrows.

3.11.3 DELETION OF FRAGMENTS

In the cases where certain restriction enzymes had one site within the insert and another on the polylinker of the vector, the fragment between those two sites was deleted and the remainder of recombinant plasmid was religated. This allowed the sequencing of the remainder of the insert directly from the pUC vectors.

3.11.4 TRANSFERRING FRAGMENTS FROM pUC VECTORS TO M13 VECTORS

The 950 bp EcoRV fragment, which was the smallest fragment to hybridise to the insert from pRR_t566, was sub-cloned into the SmaI site in pUC18 to allow the insert to be sequenced in both orientations. However, because the reverse primer reaction proved unsuccessful, the fragment was transferred to M13mp18 and M13mp19 vectors in order to provide single stranded templates for sequencing.

The arrows in Figure 34 indicate the parts of *ExtA* which were sequenced. All the sequences obtained are revealed in Figure 35.

3.12 CHECKING *ARABIDOPSIS* mRNA FOR MESSAGES HOMOLOGOUS TO *ExtA*

Total RNA was made from root, stem, leaf and whole plant tissues of *Arabidopsis* and run on a gel (Figure 36). The gel was Northern blotted and subsequently hybridised against the ^{32}P labelled DNA of the 950 bp EcoRV fragment from λ*ExtA*. RNA from *Brassica napus* root was used as a control. The probe hybridised to two bands of sizes 1.3 and 1.6 kb (track A, Figure 37) in the root RNA from *Brassica napus*. No hybridisation was observed to leaf and stem RNA from *Arabidopsis* (tracks C and D, Figure 37). In the tracks containing RNA from roots or whole plants of *Arabidopsis* (tracks B and E, Figure 37) one band corresponding to the 1.3 kb band from *Brassica napus* RNA was seen; no hybridisation to a sequence corresponding to the 1.6 kb band was observed.

FIGURE 35 Partial sequence of λ*Ext A*. The sequence obtained by sequencing fragments shown in Figure 34 and its complementary strand is given. The amino acid sequence encoded by the nucleotide sequence is given.

Fig35

Sequenced strand

GATCACTTAACCAACAATCATTACTCATTAGCAGCCACCCCACAACATCATCGTTTCAGTTTCACTGACAAACAAATTCAACCGTCTCACATAAATTCGTCTCGCACAATATAAATAAAG
D H L T N N H Y S L A A T P Q H H R F S F T D K Q I Q P S H I N S S R T I * I K
I T * P T I I T H * Q P P H N I I V S V S L T N K F N R L T * I R L A Q Y K * S
S L N Q Q S L L I S S H P T T S S F Q F H * Q T N S T V S H K F V S H N I N K Y

TTTTTTGAATAAAAACTAAAAAGTATGTTTTAAATAATATATACTAGCTATAGTCACAAATCACAACATTCCACAACCTTCTAAAATAGCAAAATAATTTTATGGAAAACTTTATAACCT
F F E * K L K S M F * I I Y T S Y S H K S Q H S T T F * N S K I I L W K T L * P
F L N K N * K V C F K * Y I L A I V T N H N I P Q P S K I A K * F Y G K L Y N L
F * I K T K K Y V L N N I Y * L * S Q I T T F H N L L K * Q N N F M E N F I T *

AATCAAATTAATTAAAAATCGAAAATATTAATAATCAAGTTTATAGTCACAAAGTAGTATAAATAAATAAAAAATGAAAAAAACAGGCGGAGTATATATCTAAAGCAACTTCAGTTTAGCG
N Q I N * K S K I L I I K F I V T K * Y K * I K N E K T G G V Y I * S N F S L A
I K L I K N R K Y * * S S L * S Q S S I N K * K M K K Q A E Y I S K A T S V * R
S N * L K I E N I N N Q V Y S H K V V * I N K K * K N R R S I Y L K Q L Q F S E

AAATAATTTATTGTAATTCATACAAAGTACTCATAATTAATAGTGTTATTTATTATTTCCTCGAAAATACATATCGAACTTTTATTTATGATTAAAGGTTAACTACATTTAACAATTAAC
K * F I V I H T K Y S * L I V L F I I S S K I H I E L L F M I K G * L H L T I N
N N L L * F I Q S T H N * * C Y L L F P R K Y I S N F Y L * L K V N Y I * Q L T
I I Y C N S Y K V L I I N S V I Y Y F L E N T Y R T F I Y D * R L T T F N N * H

ATATCGGTATACACTGAATTAATTTGTGTAACCACCATTTGTAAATTATAGTATCAACTTTAAGATTTATCATCCGATATCTTTAATCAAAAAAAATAATAAACGTTATCACAGGTTACT
I S V Y T E L I C V T T I C K L * Y Q L * D L S S D I F N Q K K * * T L S Q V T
Y R Y T L N * F V * P P F V N Y S I N F K I Y H P I S L I K K N N K R Y H R L L
I G I H * I N L C N H H L * I I V S T L R F I I R Y L * S K K I I N V I T G Y S

CAAATCACATGGACGTTCCAAACGAACACATGGGCTTTGCATATTACCACAAAGAAGAAAAGAACAAAAATCGAATTGCATTACACGTATAGTCCAAAAATGCATTATTTGCTTACGGTT
Q I T W T F Q T N T W A L H I T T K K K R T K I E L H Y T Y S P K M H Y L L T V
K S H G R S K R T H G L C I L P Q R R K E Q K S N C I T R I V Q K C I I C L R F
N H M D V P N E H M G F A Y Y H K E E K N K N R I A L H V * S K N A L F A Y G L

TAAATTAAGTTAGGGATAAGCTTGTCTTTCTCCACTTAATTAATATTTTTTGCATGTTGTGCAAACTGATGTTTCTTAACAGGTTGTAAGATCAATCAATACATAATAATTATAATAATA
* I K L G I S L S F S T * L I F F A C C A N * C F L T G C K I N Q Y I I I I I I
K L S * G * A C L S P L N * Y F L H V V Q T D V S * Q V V R S I N T * * L * * *
N * V R D K L V F L H L I N I F C M L C K L M F L N R L * D Q S I H N N Y N N N

ATCATTCATCATGAATTATTTCATCTAATATTTAATTCATTTGGTTGCCTCTCGTCAGTTTGTAATGAAAACATACAACACCAATATCACAATCTTTGCCCTATAAAAACCCTAAGACCA
I I H H E L F H L I F N S F G C L S S V C N E N I Q H Q Y H N L C P I K T L R P
S F I M N Y F I * Y L I H L V A S R Q F V M K T Y N T N I T I F A L * K P * D H
H S S * I I S S N I * F I W L P L V S L * * K H T T P I S Q S L P Y K N P K T I

TCTCCACATTCTTCACACAACACACCTCAAGCTTTATAACTTCTAAAACAAACTCTTATAACTTTTTACTTCCTCCTCCATGGCTTATTCTAAGATTGCTCTTCTCCTCATCTTCAATGT
S P H S S H N T P Q A L * L L K Q T L I T F Y F L L H G L F * D C S S P H L Q C
L H I L H T T H L K L Y N F * N K L L * L F T S S S M A Y S K I A L L L I F N V
S T F F T Q H T S S F I T S K T N S Y N F L L P P P W L I L R L L F S S S S M S

CATCTTCTTCACTTTAGTCAGCTCGACTTCAGTCCCTTGTCCACCACCACCGCCCAAGAGCCACCACAAGAAACCCGCAACGCCATCTCCTAAACCCACTTGTAAAGACGCTCTTAAACT
H L L H F S Q L D F S P L S T T T A Q E P P Q E T R N A I S * T H L * R R S * T
I F F T L V S S T S V P C P P P P P K S H H K K P A T P S P K P T C K D A L K L
S S S L * S A R L Q S L V H H H R P R A T T R N P Q R H L L N P L V K T L L N L

TAAGGTATGTGCTAACGTGTTGGATTTGGTTAAGGTTTCTCTGCCACCAACGTCCAACTGTTGCGCTCTTATCAAAGGTCTAGTTGATCTTGAAGCCGCGGTCTGTCTTTGCACTGCCCT
* G M C * R V G F G * G F S A T N V Q L L R S Y Q R S S * S * S R G L S L H C P
K V C A N V L D L V K V S L P P T S N C C A L I K G L V D L E A A V C L C T A L
R Y V L T C W I W L R F L C H Q R P T V A L L S K V * L I L K P R S V F A L P *

AAAGGCTAATGTTCTTGGTATCAACCTTAATGTTCCCATTTCTTTGAACGTTGTCCTAAACCATTGTGGTAAGAAGGTTCCATCTGGTTTCAAATGTGCCTAGAGATTACACATTTTAAA
K G * C S W Y Q P * C S H F F E R C P K P L W * E G S I W F Q M C L E I T H F K
K A N V L G I N L N V P I S L N V V L N H C G K K V P S G F K C A * R L H I L N
R L M F L V S T L M F P F L * T L S * T I V V R R F H L V S N V P R D Y T F * I

TTGAGATTTCTATATCTCACATTTCATTCCTGTTTTTAAAATTCCTCGTTTGATGATATC
L R F L Y L T F H S C F * N S S F D D I
* D F Y I S H F I P V F K I P R L M I
E I S I S H I S F L F L K F L V * * Y

Complementary strand

GATATCATCAAACGAGGAATTTTAAAAACAGGAATGAAATGTGAGATATAGAAATCTCAATTTAAAATGTGTAATCTCTAGGCACATTTGAAACCAGATGGAACCTTCTTACCACAATGG
D I I K R G I L K T G M K C E I * K S Q F K M C N L * A H L K P D G T F L P Q W
I S S N E E F * K Q E * N V R Y R N L N L K C V I S R H I * N Q M E P S Y H N G
Y H Q T R N F K N R N E M * D I E I S I * N V * S L G T F E T R W N L L T T M V

TTTAGGACAACGTTCAAAGAAATGGGAACATTAAGGTTGATACCAAGAACATTAGCCTTTAGGGCAGTGCAAAGACAGACCGCGGCTTCAAGATCAACTAGACCTTTGATAAGAGCGCAA
F R T T F K E M G T L R L I P R T L A F R A V Q R Q T A A S R S T R P L I R A Q
L G Q R S K K W E H * G * Y Q E H * P L G Q C K D R P R L Q D Q L D L * * E R N
* D N V Q R N G N I K V D T K N I S L * G S A K T D R G F K I N * T F D K S A T

CAGTTGGACGTTGGTGGCAGAGAAACCTTAACCAAATCCAACACGTTAGCACATACCTTAAGTTTAAGAGCGTCTTTACAAGTGGGTTTAGGAGATGGCGTTGCGGGTTTCTTGTGGTGG
Q L D V G G R E T L T K S N T L A H T L S L R A S L Q V G L G D G V A G F L W W
S W T L V A E K P * P N P T R * H I P * V * E R L Y K W V * E M A L R V S C G G
V G R W W Q R N L N Q I Q H V S T Y L K F K S V F T S G F R R W R C G F L V V A

CTCTTGGGCGGTGGTGGTGGACAAGGGACTGAAGTCGAGCTGACTAAAGTGAAGAAGATGACATTGAAGATGAGGAGAAGAGCAATCTTAGAATAAGCCATGGAGGAGGAAGTAAAAAGT
L L G G G G G Q G T E V E L T K V K K M T L K M R R R A I L E * A M E E E V K S
S W A V V V D K G L K S S * L K * R R * H * R * G E E Q S * N K P W R R K * K V
L G R W W W T R D * S R A D * S E E D D I E D E E K S N L R I S H G G G S K K L

TATAAGAGTTTGTTTTAGAAGTTATAAAGCTTGAGGTGTGTTGTGTGAAGAATGTGGAGATGGTCTTAGGGTTTTTATAGGGCAAAGATTGTGATATTGGTGTTGTATGTTTTCATTACA
Y K S L F * K L * S L R C V V * R M W R W S * G F Y R A K I V I L V L Y V F I T
I R V C F R S Y K A * G V L C E E C G D G L R V F I G Q R L * Y W C C M F S L Q
* E F V L E V I K L E V C C V K N V E M V L G F L * G K D C D I G V V C F H Y K

AACTGACGAGAGGCAACCAAATGAATTTAAATATTAGATGAAATAATTCATGATGAATGATTATTATTATAATTATTATGTATTGATTGATCTTACAACCTGTTAAGAAACATCAGTTTGC
N * R E A T K * I K Y * M K * F M M N D Y Y Y N Y Y V L I D L T T C * E T S V C
T D E R Q P N E L N I R * N N S * * M I I I I I I M Y * L I L Q P V K K H Q F A
L T R G N Q M N * I L D E I I H D E * L L L * L L C I D * S Y N L L R N I S L H

ACAACATGCAAAAAATATTAATTAAGTGGAGAAAGACAAGCTTATCCCTAACTTAATTTAAACCGTAAGCAAATAATGCATTTTTGGACTATACGTGTAATGCAATTCGATTTTTGTTCT
T T C K K Y * L S G E R Q A Y P * L N L N R K Q I M H F W T I R V M Q F D F C S
Q H A K N I N * V E K D K L I P N L I * T V S K * C I F G L Y V * C N S I F V L
N M Q K I L I K W R K T S L S L T * F K P * A N N A F L D Y T C N A I R F L F F

TTTCTTCTTTGTGGTAATATGCAAAGCCCATGTGTTCGTTTGGAACGTCCATGTGATTTGAGTAACCTGTGATAACGTTTATTATTTTTTTTGATTAAAGATATCGGATGATAAATCTTA
F L L C G N M Q S P C V R L E R P C D L S N L * * R L L F F L I K D I G * * I L
F F F V V I C K A H V F V W N V H V I * V T C D N V Y Y F F * L K I S D D K S *
S S L W * Y A K P M C S F G T S M * F E * P V I T F I I F F D * R Y R M I N L K

AAGTTGATACTATAATTTACAAATGGTGGTTACACAAATTAATTCAGTGTATACCGATATGTTAATTGTTAAATGTAGTTAACCTTTAATCATAAATAAAAGTTCGATATGTATTTTCGA
K L I L * F T N G G Y T N * F S V Y R Y V N C * M * L T F N H K * K F D M Y F R
S * Y Y N L Q M V V T Q I N S V Y T D M L I V K C S * P L I I N K S S I C I F E
V D T I I Y K W W L H K L I Q C I P I C * L L N V V N L * S * I K V R Y V F S R

GGAAATAATAAATAACACTATTAATTATGAGTACTTTGTATGAATTACAATAAATTATTTCGCTAAACTGAAGTTGCTTTAGATATATACTCCGCCTGTTTTTTCATTTTTTATTTATTT
G N N K * H Y * L * V L C M N Y N K L F R * T E V A L D I Y S A C F F I F Y L F
E I I N N T I N Y E Y F V * I T I N Y F A K L K L L * I Y T P P V F S F F I Y L
K * * I T L L I M S T L Y E L Q * I L S L N * S C F R Y I L R L F F H F L F I Y

ATACTACTTTGTGACTATAAACTTGATTATTAATATTTTCGATTTTTAATTAATTTGATTAGGTTATAAAGTTTTCCATAAAATTATTTTGCTATTTTAGAAGGTTGTGGAATGTTGTGA
I L L C D Y K L D Y * Y F R F L I N L I R L * S F P * N Y F A I L E G C G M L *
Y Y F V T I N L I I N I F D F * L I * L G Y K V F H K I I L L F * K V V E C C D
T T L * L * T * L L I F S I F N * F D * V I K F S I K L F C Y F R R L W N V V I

TTTGTGACTATAGCTAGTATATATTATTTAAAACATACTTTTTAGTTTTTATTCAAAAAACTTTATTTATATTGTGCGAGACGAATTTATGTGAGACGGTTGAATTTGTTTGTCAGTGAA
F V T I A S I Y Y L K H T F * F L F K K L Y L Y C A R R I Y V R R L N L F V S E
L * L * L V Y I I * N I L F S F Y S K N F I Y I V R D E F M * D G * I C L S V K
C D Y S * Y I L F K T Y F L V F I Q K T L F I L C E T N L C E T V E F V C Q * N

ACTGAAACGATGATGTTGTGGGGTGGCTGCTAATGAGTAATGATTGTTGGTTAAGTGATC
T E T M M L W G G C * * V M I V G * V I
L K R * C C G V A A N E * * L L V K *
* N D D V V G W L L M S N D C W L S D

FIGURE 36 *Arabidopsis* RNA ANALYSIS

Approx 10 µg samples of RNA were loaded on a 1% agarose (with formaldehyde) gel and run at 100 V for 3 h.

A RNA from *Brassica napus* roots
B RNA from *Arabidopsis* root tissue
C RNA from *Arabidopsis* leaf tissue
D RNA from *Arabidopsis* stem tissue
E RNA from *Arabidopsis* whole plant

FIGURE 37 An autoradiograph of the gel in Figure 36.

The filter was washed at high stringency (0.1 x SSC, 65^{o}C for 30 min)

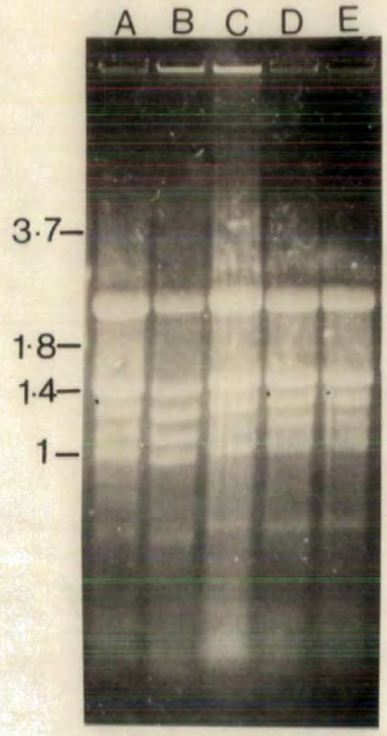

Fig36

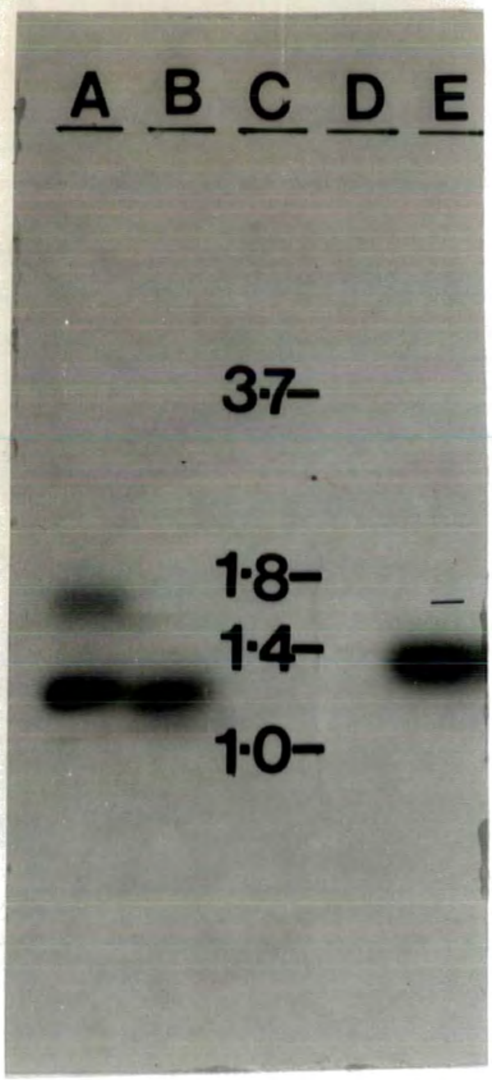

Fig37

CHAPTER FOUR

DISCUSSION

DISCUSSION

Pea has a haploid genome size of 4.8 x 10^9 base pairs. Therefore in a genomic library such as the one screened in this thesis, which has an average cloned insert size of 16 kb, only one in 3 x 10^5 independent recombinant phage will carry a particular single copy sequence. In the case of genes in the *Leg A* subfamily, the analysis of pea genomic DNA (Figure 14) has shown that four fragments on the EcoRI digest hybridised to the insert from pDUB6. Isolation of one of these sequences in every 7.5 x 10^4 independent recombinant lambda clones would therefore be expected and a similar value has been obtained in practice (Shirsat, 1984). The *Leg A* gene subfamily input contains five members, since a psuedo legumin gene, *Leg D*, was found to be located approximately 1.3 kb 3' of *Leg A* (Bown *et al.* 1985) on the 9.6 kb EcoRI fragment in the genomic clone λ*Leg 1*. Of the five *Leg A* type legumin genes in the pea genome, four have already been isolated and characterised (Lycett *et al.* 1984, Lycett *et al.* 1986, Shirsat 1984). The remaining gene could be characterised from Southern blots of pea genomic DNA as lying on a 4.2 kb EcoRI fragment of the genomic DNA. This fragment corresponded to the EcoRI fragment from λ*Leg E1* and thus the gene on this genomic clone is the remaining uncharacterised gene from the *Leg A* subfamily in pea (*Pisum sativum* L).

When the insert from the genomic clone was sub-cloned in pUC18, three EcoRI fragments, of sizes of 4.2, 2.4 and 2 kb, were inserted in pUC18 and denoted pSY1, pSY2 and pSY3 respectively. These three fragments cover most of the insert. The pSY1 sub-clone contained the fragment which hybridised strongly to the insert from pDUB6 and was assumed to contain *Leg A*-type sequence (*Leg E*). Therefore it was chosen for further analysis which started with restriction mapping the 42 kb EcoRI fragment in it.

The sub-cloning of fragments from the insert into M13 mp18 and mp19 provided the single stranded templates for sequencing. The fragments sequenced covered the 4.2 kb EcoRI fragment, and the overlapping between sequences in both the same or the opposite orientation provided confirmation of the whole sequence. It was possible to construct the sequence of the whole fragment from those pieces of sequence obtained from different sequencing runs.

The complete sequence of the coding region including the introns and both the 3' and the 5' flanking regions was obtained (Figure 12).

The *Leg E* sequence contains 1560 bases of coding sequence plus introns, 1009 bases of 5' flanking sequence, and 815 bases of 3' flanking sequence.

When the *Leg E* sequence was compared to the *Leg A* sequence, the 5' flanking regions contained some homology (460/1009) (Figure 12a). The strongest homology between the two regions was in the 160 bases nearest the coding sequence, this strong homology starts at position -186 (*Leg A* numbering) and includes perfect homology (9/9 bases) in the "enhancer" region. The homology stops after the "enhancer" region due to a deletion in the *Leg A* sequence at position -165 (*Leg A* numbering), but it becomes very strong starting from the first base after the deletion and continues to the end of the sequence. The "legumin box" is located between positions -154 and -129 (*Leg E* numbering) and shares a 23/26 homology with *Leg A*. 7/8 homology is observed between the "TATA box" of *Leg E* and that of *Leg A* (Figure 12a). The 5' flanking sequence ends at position -1 (Figure 12) where the ATG methionine translation start codon occurs. There are two other ATG methionine codons occur in this region of 5' flanking sequence of *Leg E*, at positions -108 and -88 (*Leg E* numbering), but both of them occur before the "TATA box" and are not in agreement with the consensus sequence (AACAATGGC) for plant gene translation starts (Lutke *et al*. 1987). The designated translation start (Figure 12) is therefore correct for this gene.

A very high degree of homology is observed between coding sequences of *Leg E* and *Leg A*. A six base deletion in *Leg A* at position 1014 (*Leg E* numbering) and 9 base changes (six active, three silent) are the only observed differences between the 1560 bases of coding sequence of *Leg E* and 1554 bases of coding sequence of *Leg A*, which means that the homology between the coding sequences of the two genes and the predicted amino acid sequence encoded by both genes is more than 99%.

High levels of homology are observed between the introns of both genes. In both genes the size of IVS-1 is the same (88 bases), and the sequences are almost identical with only one base change out of 88 bases. IVS-2 in *Leg E* is identical to IVS-2 in *Leg A* both in size and sequence. IVS-3 in *Leg E* is 13 bases shorter than that of *Leg A*, yet the 84 bases of IVS-3 in *Leg E* share a homology of 82/84 bases with only 2 base differences at positions 1364 and 1365 (*Leg E* numbering).

A homology of 141/150 is observed between the first 150 bases of the 3' flanking sequence of both genes. This region contains the polyadenylation addition signal sequence ((poly A^{+}) site) within it, and there is complete homology over this site between *Leg A* and *Leg E*. Homology decreases in a 3' direction after this region, and there is no significant homology further than 150 bp beyond the stop codon.

Since the *Leg A* gene is functional (Lycett *et al*. 1984) and *Leg E* resembles and is homologous with *Leg A* in the regions which affect the function of plant genes (i.e. "TATA box", "Legumin box", start codon, coding sequence etc), it can be concluded that *Leg E* is a functional and encodes a "Major" legumin polypeptide with a very similar amino acid composition to that encoded by *Leg A*. The *Leg J* subfamily is one of three distinct subfamilies encoding "minor" legumin polypeptides (Thompson, 1989). One of these three subfamilies contains three genes; *Leg J* was isolated and fully characterised by Gatehouse *et al.* 1988 as well as part of another gene *Leg K*, which was found on the same genomic clone. The last gene in the subfamily, designated *Leg L* was identified by Southern blots of pea genomic DNA (Domoney and Casey, 1985). The three genes were mapped to a locus *Lg-2*, near *a* on chromosome 1 of the pea genome (Domoney and Casey 1986).

After λ*Leg J2* was isolated and partially restriction mapped, it was compared to the restriction map of λJC5 (Gatehouse *et al.* 1988). This comparison (Figure 7) has shown that the two clones overlap to cover a region of 17 kb. More importantly the comparison also indicates that the λ*Leg J2* contains the entire sequence of gene *Leg K*, of which only the 3' flanking region and most of coding region was contained on λJC5.

The fact that the whole sequence of *Leg K* gene is found in λ*Leg J2*, prompted the sequencing of the entire *Leg K* gene.

The fragments sub-cloned in M13 mp18 and mp19 in order to provide template for sequencing (Figure 16) have covered the whole sequence and also provided regions of overlapping between fragments in order to confirm the sequence produced.

The determined sequence of *Leg K* has revealed that it contains the entire sequence of *Leg K* gene, this new complete sequence has shown no differences with the previously characterised sequence of *Leg K*, although JC5 was isolated from pea line Dark Skinned Perfection and l*Leg J2* from cv. Feltham First.

A high degree of homology (97%) was observed between *Leg K* and *Leg J* with one addition at base 1048 (*Leg J* numbering) and four deletions three of which at base 1108 and one at base 1128 (*Leg J* numbering). In addition to these deletions and addition, 46 base changes were observed, 20 of which were "active" base changes and resulted in 20 amino acid substitutions while the other 26 were silent mutations with no effect on the amino acid composition. Only one of the amino acid substitutions seems to have an important impact on the gene as a whole, the first base of the start codon in *Leg J* is changed from A to G which has resulted in mutating the start codon from ATG to GTG, a valine codon. The first subsequent ATG in the coding sequence of *Leg K* is at base 117 (*Leg J* numbering) which is out of frame relative to *Leg J*, and would give an open reading frame of only 6 amino acids due to the presence of a stop codon TGA at base 135. Furthermore the sequence around this ATG codon is not in agreement with the consensus sequence (AACAATGGC) for plant gene translation starts (Lutke, *et al.*, 1987), 3/9 bases agree compared with 7/9 for the first ATG. It is therefore unlikely that this is used as a start codon. This evidence, and the presence of small insertion in the *Leg K* sequence in the region where the ribosome would be expected to bind (relative to *Leg J*) provide the circumstances under which an mRNA produced by transcription of *Leg K* will be transcriptionally ineffective, have a very short half-life, and a low steady state level. This explains the failure to observe expression from *Leg K* (Thompson, 1989).

An 89% homology is maintained between the 5' flanking sequences of *Leg J* and *Leg K* if deletions and insertions are ignored, which supports the assumptions from studies on other similarly "damaged" genes that the promoter sequence sequence of *Leg K* is active and therefore the gene is expressed but no product accumulates. The fact that *Leg K* cDNA pCD40 (Thompson, 1989) was produced from pea line "Birte", strongly suggests that *Leg K* is expressed in other pea lines. This indicates that the mutation preventing the gene product from accumulating in pea line Feltham First may be absent in "Birte", hinting at a possible mechanism for some of the line-line variation seen in minor legumin polypeptides.

Both *Leg J* and *Leg K* genes have two introns, the position of introns in *Leg J* is confirmed by comparison with sequences of homologous cDNA clones (Gatehouse *et al.* 1988). In gene *Leg K* intron 1 is 81 bases long and intron 2 is 105 bases while in gene *Leg J* they are 138 and 98 bases long respectively. In both genes the 3' ends of both introns show strong homology to the extent of 56% and 71% (or 96% and 74% if deletions are ignored). This suggests a relatively recent sequence divergence of the two genes. The introns are A + T rich like those in many other plant genes.

Apart from the last 60 bases of the *Leg K* 3' flanking sequence the 3' flanking sequences of both genes show significant homology. At least 4 polyadenylation sites are present in both genes within 220 bases of the stop codon, in gene *Leg K* the first site is of the multiple overlapping (AATAATAAA) type. Since a homologous cDNA has a poly(A) tail at a point corresponding to base 1935 in the *Leg J* gene, the second or third sites are the most likely to be used.

The 5' flanking sequences of both genes contain clearly defined 'TATA' boxes, and both show the same sequence (CCTATAAATT) which is in reasonable agreement with the consensus sequence for this promoter element [T(C/G)TATA(T/A)ATA] (Messing *et al.* 1983). Like many other plant genes there is no "CAAT" box.

Since the construction of the first genomic library was reported (Maniatis *et al.* 1976), there has been great progress in developing new vectors. Replacement vectors which rely on the replacement of the central region of phage lambda derivatives have been described (Karn and Brenner 1980, Loenen and Brammar 1980, Mizusawa and Ward 1982, Rimm *et al.* 1989, Frischnauf *et al.* 1983). Recombinants produced by replacing the central region by genomic DNA in these vectors are *gam*$^-$, *red*$^-$, which allows them to grow in bacterial hosts lysogenic for the phage P2. Such recombinants are termped *Spi*$^-$, while non-recombinants are termed *Spi*$^+$. This provides a selection system. These replacement vectors also allow the cloning of Sau3A cleaved genomic DNA into BamHI sites. For these reasons EMBL3 (Frischnauf *et al.* 1983) was chosen as a vector to construct the *Arabidopsis thaliana* genomic library.

In the construction of genomic libraries it is desirable to reduce the number of non-recombinant transformants in order to reduce the number of clones which have to be screened for the sequence of interest. In this work, this was achieved by isolating the annealed arms of the vector after cleavage with BamHI and EcoRI to create BamHI ends for the arms and EcoRI ends for the replaceable central fragment, the arms were separated from the central fragment by centrifugation through a sucrose gradient.

The unligated vector arms should not be packaged, as they are only 29 kb in length while DNA molecules of the length between 38 and 53 kb will be packaged (Feiss *et al.* 1977). Concatenates of arms could be formed as a result of their ligation at the *cos* ends. These concatenates could be a substrate for *in vitro* packaging, but for the fact that cleavage at the *cos* sites of these molecules will produce forms too small to be packaged. Hence only recombinant molecules, in which DNA has been inserted at the BamHI, will be packaged.

In fact, the development of *in vitro* packaging system is an essential factor in the construction of gene libraries in lambda vectors, as it allows efficiencies of at least two orders of the magnitude above those obtained by $CaCl_2$ transfection (Thomas *et al.* 1974).

The production of an efficient *in vitro* packaging system and the preparation of the annealed arms of EMBL3 meant that only one additional component, the partially digested *Arabidopsis* genomic DNA, was necessary for construction of the genomic library. DNA used in library construction must be of high molecular weight, free of nuclease contamination, and with few or no single strand breaks. Suitable DNA can be prepared from whole tissues such as leaf, embryo and root, or from isolated cellular organelles such as nuclei, mitochondria, or chloroplasts. Since the aim of this work was to prepare a total genomic library, in which all DNA sequences would be represented, leaf tissue was chosen as an abundant and convenient source of material. As judged by agarose gel analysis, the majority of the DNA was of a size greater than 50 kb (Figure 17). Accurate estimation of DNA size was prevented due to the poor resolution of high molecular weight DNA achieved in agarose gels. The absence of endogenous nuclease activity was judged by incubating a sample of the DNA in restriction enzyme buffer at 37^{o}C for 1 hr after which the DNA showed no appreciable degradation (Figure 18). The DNA was readily digested by restriction enzymes, indicating that no inhibiting materials were present in the preparation.

Conditions for the partial digestion of *Arabidopsis* genomic DNA were established (Figure 18). The maximum amount of 10-20 kb fragments appeared to be in track D as judged by the intensity of fluorescence. On the basis of this result, three large-scale digestions of *Arabidopsis* DNA were performed and pooled, using 0.5, 1 and 2 times the amount of Sau3AI which yielded the maximum amount of 10-20 kb fragments. This was done in order to randomise the 10-20 kb size fractionated DNA population, by ensuring that molecules which varied in their degree of susceptibility to cleavage by the enzyme were included in the preparation for cloning. Relatively small amounts of 10-20 kb DNA fragments were needed to construct the *Arabidopsis* gene library. Therefore the DNA from the 3 large-scale digestions was pooled and loaded on one track (Figure 19) and a 10-20 kb size fractionated DNA population was purified from the agarose. To prevent self-ligation the purified DNA was alkaline phosphatased; 1 µg of this DNA was incubated with ligase and checked by agarose gel electrophoresis (Figure 20) track E. This has shown that the DNA was evenly distributed in the region of 10-20 kb on the gel, indicating that the purified DNA was of the correct desired size and efficiently phosphatased.

When ligating the BamHI vector arms and the Sau3AI insert, in theory a 2:1 molar ratio of arms to insert should give an equimolar amount of the two molecules. However this assumes an ideal situation in which all the molecules in the ligation reaction have cohesive termini. Since some molecules will have lost their cohesive ends in extraction and purification steps, this ratio will alter, and a series of test ligations with varying arms to insert ratio was set up to monitor this (Table 3). In order to show that all possible inhibitors of ligation had been removed, and that the cohesive termini generated on the vector and *Arabidopsis* DNA were capable of ligation, small aliquots of all mixtures of the vector DNA and *Arabidopsis* DNA were ligated and the products were analysed by agarose gel electrophoresis (Figure 20). All mixtures were seen to form molecules of λ size or larger, indicating that concatenates had been formed.

The 1:1 ratio of arms to insert was found to produce the greatest number of recombinants as shown in Table 3. As shown in Table 3 the number of plaques obtained from Q359 was about 20% less than the number obtained from K803 which is due to the failure of clones containing methylated genomic DNA to grow on Q359. Therefore the *in vitro* packaged DNA was grown on K803 host cells first and then on Q359. When the library phage stock (prepared by eluting plaques obtained from K803 host cells) was titrated using Q359 and P2392 as host cells, it was observed that the number of plaques obtained with Q359 host cells was about 3-4% less than the number obtained with P2392 cells. This difference is attributed to the background of non-recombinants, therefore Q359 host cells were used when screening was carried out.

As the purpose of constructing gene libraries is to make the isolation of a particular sequence possible, it was intended to try to isolate sequences encoding extensin proteins from the *Arabidopsis* genomic library. Logically, before screening the gene library, *Arabidopsis* genomic DNA was checked for sequences homologous to the insert from pRR_t566 which is actually an extensin cDNA from *Brassica napus* L. The HindIII digest in track E has shown that at least three fragments had hybridised to the probe (Figure 22), indicating that sequences homologous to the extensin cDNA are present in the *Arabidopsis* genomic DNA.

The *Arabidopsis* genomic library was screened for clones containing extensin encoding sequences. As stated six independent positives were isolated, the degree of homology to the probe was variable as judged by the autoradiographs shown in Figure 23. The filters of the plaque lifts were washed in 0.3 x SSC because it was impossible to predict the extent of homology between the probe and the extensin sequences in *Arabidopsis*. Actually the two genomic clones (λ*Ext A* and λ*Ext B*) which showed strong homology with the probe were chosen to be studied. Only phage stocks were made from the other four clones (i.e. λ*ExtC,* λ*Ext D,* λ*Ext E,* and λ*Ext F*).

The restriction enzyme analysis of both λ*Ext A* and λ*Ext B* indicated that they were different clones. The southern blotting and the subsequent hybridisation of the blots has revealed that the probe had hybridised selectively to specific fragments in both clones, although the filters were washed at high stringency (0.1 SSC, 65°C, 30 min).

Although the two genomic clones hybridised to the probe the intensity of bands seen on autoradiographs indicated that λ*Ext A* was more homologous to the probe. Therefore it was chosen for further characterisation.

The 3.2 kb EcoRI-BamHI fragment (Figure 34) of λ*Ext A* consistently hybridised to the extensin probe. Further restriction enzyme mapping of the fragment and the subsequent southern blotting and hybridisation has revealed that the 950 bp EcoRV fragment (Figure 34) was the smallest single fragment to hyridise to the probe, indicating that the coding sequence or at least most of it should be found within that fragment.

The information obtained from the sequence produced (Figure 35), which does not cover the whole 3.2 kb fragment but nevertheless covers the 950 bp EcoRV fragment, has revealed that the sequence does not represent an extensin gene. However, the sequence may represent a gene of some kind, since the ATG codon at base 1077 could be a translational start. As stated above, the gene could not be an extensin gene, although the sequence between codon GTC at position 1113 and the AAG codon at position 1138 looks like an extensin proline rich pentapeptide it is not repeated anywhere else in the sequence obtained. Nevertheless it provides the explanation for the hybridsation between the extensin probe and λ*Ext A*.

These results have suggested the idea of investigating the possibility that λ*Ext A* harbours a gene. Therefore mRNA from different tissues of *Arabidopsis* was probed with ^{32}P labelled DNA from the 950 bp EcoRV fragment in λ*Ext A*. The result has shown that the probe had hybridised to a 1.3 kb long message in mRNA prepared from root and whole plant and had also hybridised to two messages in mRNA prepared from *Brassica napus* root (Figure 37). This has suggested that the sequence contained in λ*Ext A* may represent a functional gene in *Arabidopsis*. The expression of this gene seems to be tissue specific since no hybridisation was observed to mRNA prepared from leaf and stem (Figure 37). It is also possible that the probe is detecting homologous sequences in the mRNA derived from true extensin genes in *Arabidopsis* and *Brassica*, in the same way that the *Brassica* cDNA hybridised strongly to the *Arabidopsis* genomic clone λ*Ext A*.

The data available is not sufficient to predict any further details about the nature of the gene, until further studies are carried out.

A second genomic clone λ*Ext B* was restriction mapped and the partial restriction map produced (Figure 31), which revealed that the sequence homologous to the probe was located within the 5.6 kb EcoRI fragment. Therefore that fragment, along with the 2 kb EcoRI fragment were sub-cloned in pUC18 and called pL06 and pL05 respectively (Figure 37). This will make the task of further characterising the clone reasonably easy.

One of the aims of isolating those two genomic clones was to check the *Arabidopsis* genomic library constructed in this work. Obviously the information obtained from characterising the two clones has revealed the inserts in both clones have BamHI restriction site on their ends and they were 11.5 kb and 12.8 kb in size indicating that the quality of the gene library was high.

REFERENCES

REFERENCES

Basha, S.M.M. and Beevers, L. (1975) The development of proteolytic activity and protein degradation during germination of *Pisum sativum* L. Planta, 124, 77-87.

Beevers, L. and Poulson, R. (1972) Protein synthesis in cotyledons of *Pisum sativum* L. Plant physiol. 49, 476-481.

Benoist, C. and Chambon, P. (1981) *In vivo* sequence requirements of SV40 early promotor region. Nature, 290 , 304-310.

Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) The ovalbumin gene-sequence of putative control regions. Nucleic Acid Res. 8, 127-142.

Buinboim, H.C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. 7, 1513-1523.

Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L.A., Grunland, D.J., Kiefer, D.O., Moore, D.O., Schum, J.W., Sheldon, E.L. and Smithies, O. (1977) Charon phages safer derivatives of bacteriophage lambda for DNA cloning. Science 196, 161-169.

Blobel, G. and Dobberstein, B. (1975) Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobin. Light chains on membrane-bound ribosomes of murine myeloma. J. Cell. Biol. 67, 835-851.

Bolivar, F., Rodriguex, R.L., Gween, P.J., Bettach, M.C., Heyencker, H.L., Boyer, H.W., Crosa, J.H. and Falkow, S. (1977) Construction and characterisation of new cloning vehicles. II. A multi-purpose cloning system. Gene, 2, 95-113.

Boulter, D. (1981) Biochemistry of storage protein synthesis and deposition in the developing legumin seed. Adv. Bot. Res. 9, 1-31.

Boulter, D., Ellis, R.J. and Yarwood, A. (1972) Biochemistry of storage protein in plants. Biol. Rev. 47, 113-175.

Bown, D., Ellis, T.H.N. and Gatehouse, J.A. (1988) The sequence of a gene encoding convicilin from pea (*Pisum sativum* L) shows that convicilin differs from vicilin by an insertion near the N-terminus. Biochem. J. Biochem. 251, 717-726.

Bown, D., Levasseur, M., Croy, R.R.D., Boulter, D. and Gatehouse, J.A. (1985) Sequence of pseudogene in the legumin gene family of pea (*Pisum sativum* L). Nucleic Acids Res. 13, 4527-4538.

Breathnach, R. and Chambon, P. (1981) Organisation and expression of eukaryotic split genes coding for protein. Ann. Rev. Biochem. 50, 349-383.

Casey, R. (1979) Genetic variability in the structure of the -subunits of legumin from *Pisum* - a two dimensional gel electrophoresis study. Heredity 43, 265-272.

Casey, R. and Domoney, C. (1984) The genetics of legume storage proteins. Phil. Trans. Roy. Soc. Lon. B., 304, 349-358.

Casey, R., Domoney, C. and Ellis, T.H.N. (1986) Legume storage proteins and their genes. Oxford Surv. Plant Mol. Cell Biol. 3, 1-95.

Casey, R., Domoney, C. and Stanley, J. (1984) Convicilin mRNA from pea (*Pisum sativum* L) has sequence homology with other legume 7s storage protein mRNA species. Biochem. J. 224, 2661-2666.

Casey,R., March, J.F., Sharman, J.E. and Short, M.N. (1981) The purification, N-terminal amino acid sequence and some other properties of an -subunit of legumin from the pea (*Pisum sativum* L). Biochem. Biophys. Acta, 670, 428-432.

Chen, J. and Varner, J.E., (1985a) An extracellular matrix protein in plants: characterisation of a genomic clone for carrot extensin. EMBO J. 4, 2145-2151.

Chen, J., Varner, J.E. (1985b) Isolation and characterisation of cDNA clones for carrot extensin and a proline-rich 33-kDa protein. Proc. Natl. Acad. Sci. 82, 4399-4403.

Chung, C.T., Niemela, S.L. and Miller, R.H. (1989) One step preparation of competent *Escherichia coli*: Transformation and storage of bacterial cells in same solution. Proc. Natl. Acad. Sci.

Chrispeels, M.J. (1969) Synthesis and secretion of hydroxyprotein containing macromolecules in carrots. I. Kinetic analysis. Plant Physiol. 44, 1187-1193.

Chrispeels, M.J. (1984) Biosynthesis, processing and transport of storage proteins and lectins in cotyledons of developing legume seeds. Phil. Trans. Roy. Soc. Lon. B. 304, 309-322.

Chrispeels, M.J., Higgins, T.J.V., Craig, S. and Spencer, D. (1982) Role of the endoplasmic reticulum in the synthesis of reverse proteins and the kinetics of their transport to protein bodies in developing pea cotyledons. J. Cell Biol. 93, 5-14.

Chrispeels, M.J., Sadava, D. and Cho, Y.P. (1974) Enhancement of extensin biosynthesis in ageing carrot storage tissue. J. Exp. Bot. 25, 1157-1166.

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

Collins, J. and Bruning, H.J. (1978) Plasmids useable as gene cloning vectors in an *in vitro* packaging by coliphage : 'cosmids'. Gene 4, 85-107.

Collins, J. and Hohn, B. (1979) Cosmids: a type of plasmid gene cloning vector that is packageable *in vitro* in bacteriophage heads. Proc. Nat. Acad. Sci. USA 5, 4242-4246.

Condit, C.M., and Meagher, R.B. (1986) A gene encoding a novel glycine-rich structural protein of petunia. Nature 323, 178-181.

Cooper, J.B. and Varner, J.E. (1983) Insolubilisation of hydroxyproline-rich cell wall glycoprotein in aerated carrot root slices. Biochem. Biophys. Res. Comm. 112, 161-167.

Croy, R.R.D., Derbyshire, E., Krishna, T.G. and Boulter, D. (1979) Legumin of *Pisum sativum* and *Vicia faba*. New Phytol., 83, 29-35.

Croy, R.R.D., Gatehouse, J.A., Evans, I.M. and Boulter, D. (1980) Characterisation of the storage protein subunits synthesised *in vitro* by polyribosomes and RNA from developing pea *Pisum sativum* L). I. Legumin. Planta, 148, 49-56.

Croy, R.R.D., Gatehouse, J.A., Evans, I.M. and Boulter, D. (1980b) Characterisation of storage protein subunits synthesised *in vitro* by polyribosome and RNA from developing pea (*Pisum sativum* L). II. Vicilin. Planta 148, 5763.

Croy, R.R.D., Hoque, M.S., Gatehouse, J.A. and Boulter, D. (1984) The major albumin proteins from pea (*Pisum sativum* L). Purification and some properties. Biochem. J. 218, 795-803.

Croy, R.R.D., Lycett, G.W., Gatehouse, J.A., Yarwood, J.N. and Boulter, D. (1982) Cloning and analysis of cDNAs encoding plant storage protein precursors. Nature 295, 76-79.

Danielsson, C.E. (1949) Seed globulins of the Graminae and Leguminosae. Biochem. J. 44, 387-400.

Daniels, D.L. and Blattner, F.R. (1987) Mapping using gene encyclopedias. Nature 325, 831.

Delauney, A.J. (1984) Cloning and characterisation of cDNAs encoding the major pea storage proteins and expression of vicilin in *E. coli*. Ph.D. thesis, University of Durham, U.K.

Derbyshire, E., Wright, D.J. and Boulter, D. (1976) Legumin and vicilin, storage proteins of legume seeds. Phytochemistry 15, 3-24.

Domoney, C., Barker, D. and Casey, R. (1986) The complete deduced amino acid sequence of legumin beta-polypeptides from different genetic loci in *Pisum*. Plant Mol. Biol. 7, 467-474.

Domoney, C. and Casey, R. (1985) Measurement of gene number for seed storage protein in *Pisum*. Nucleic acids Res. 13, 687-699.

Domoney, C., Davies, D.R. and Casey, R. (1980) The initiation of legumin synthesis in immature embryos of *Pisum sativum* L. grown *in vivo* and *in vitro*. Planta 149, 454-460.

Ellis, T.H.N., Domoney, C., Castleton, J., Clearly, W. and Davies D.R. (1986) Vicilin genes of *Pisum*. Mol. Gen. Genetics 205, 164-169.

Epstein, L. and Lamport, D.T.A. (1984) An intracellular linkage involving isodityrosine in extensin. Phytochem. 23, 1241-1246.

Estelle, M.A. and Somerville, C.R. (1986) The mutants of *Arabidopsis*. TIG, 89-93.

Evans, I.M., Bown, D., Lycett, G.W., Croy, R.R.D., Boulter, D. and Gatehouse, J.A. (1985) Transcription of legumin gene from pea (*Pisum sativum* L) *in vitro*. Planta 165, 554-566.

Evans, M., Gatehouse, L., Gatehouse, J.A., Boulter, D. and Croy, R.R.D. (1990) Abundant cDNA sequences from roots of oilseed rape (*Brassica napus* L) encode proteins homologous to extensin. Mol. Gen. Genet. In press.

Feiss, M., Fischer, R.A. Crayton, M.A. and Egner, C. (1977) Packaging of the bacteriophage lambda chromosome: effect of chromosome length. Virology 77, 281-293.

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

Fry, S.C. (1982) Isodityrosine, a new crosslinking amino acid from plant cell wall glycoprotein. Biochem. J. 204, 449-455.

Gatehouse, J.A., Bown, D., Gilroy, J., Levasseur, M., Castleton, J. and Ellis, T.H.N. (1988) Two genes encoding "Minor" legumin polypeptides in pea (*Pisum sativum* L). Biochem. J. 250, 15-24.

Gatehouse, J.A., Croy, R.R.D. and Boulter, D. (1980) Isoelectric focussing properties and carbohydrate content of pea *Pisum sativum* L. legumin. Biochem. J. 185, 497-503.

Gatehouse, J.A., Croy, R.R.D. and Boulter, D. (1984) The synthesis and structure of pea storage proteins. CRC Crit. Rev. Plant Sci. 1, 287-314.

Gatehouse, J.A., Croy, R.R.D., Morton, H., Tyler, M. and Boulter, D. (1981) Characterisation and subunit structures of the vicilin storage proteins of pea (*Pisum saticum* L). Eur. J. Biochem. 118, 627-633.

Gatehouse, J.A., Evans, I.M., Bown, D., Croy, R.R.D. and Boulter, D. (1982) Control of storage protein synthesis during seed development in pea (*Pisum sativum* L). Biochem. J. 208, 119-127.

Gatehouse, L., Evans, M., Gatehouse, J.A., Yarwood, J. and Croy, R.R.D. (1990) Isolation and characterisation of a gene encoding extensin from oilseed rape (*Brassica napus* L). Mol. Gen. Genet. In press.

Gilroy, J., Wright, D.J. and Boulter, D. (1979) Homology of basic subunits of legumin from *Glycin max.* and *Vicia faba*. Phytochem. 18, 315-316.

Graf, H. (1979) Optimising of conditions for the *in vitro* formation of hybrid DNA molecules by DNA ligase. Biochem. Biophys. Acta. 564, 225-234.

Gray, R.E. and Cashmore, A.R. (1976) RNA synthesis in plant leaf tissue, the characterisation of messenger RNA species lacking and containing polyadenylic acid. J. Mol. Biol. 108, 595-808.

Higgins, T.J.V., Newbiggin, E.J., Spencer, D., Llewellyn, D.J. and Craigh, S. (1988) The sequence of a pea vicilin gene and its expression in transgenic plants. Plant Mol. Biol. 11, 683-696.

Hershfield, V., Boyer, H.W., Yanofsky, C., Lovatt, N. and Helinski, D.R. (1974) Plasmid ColE1 as a molecular vehicle for cloning and amplification of DNA. Proc. Nat. Acad. Sci. U.S.A. 71, 3455-3559.

Hohn, B. (1975) DNA as substrate for packaging into bacteriophage lambda *in vitro*. J. Mol. Biol. 98, 93-106.

Hohn, B. and Murray, K. (1977) Packaging recombinant DNA molecules into bacteriophage particles *in vitro*. Proc. Nat. Acad. Sci. U.S.A. 74, 3259-3263.

Karn, J., Brenner, S.and Barnett Cesareni, G. (1980) Novel bacteriophage cloning vector. Proc. Nat. Acad. Sci. U.S.A. 77, 5172-5176.

Keller, B., Sauer, N. and Lamb, C.J. (1988) EMBO J. 7, 3625-3633.

Keller, B., Templeton, M.D. and Lamb, C.J. (1989) Proc. Nat. Acad. Sci. U.S.A. 86, 1525-1533.

Laibach, F. (1907) Zur Frage nach der Individualitat der chromosomen in Pflanzenreich. Beih. Bot. Zentral bl.1 Abt 22, 191-210.

Lamport, D.T.A. (1977) Structure biosynthesis and significance of cell wall glycoproteins. Recent Adv. Phytochem. 11, 79-111.

Lamport, D.T.A. and Caat, J.W. (1981) Glycoproteins and enzymes of the cell wall. Encyclopedia Plant Physiol. 13B, 133-165.

Langridge, J. (1957) The asceptic culture of *Arabidopsis thaliana* L. Hegnh. Aust. J. Biol. Sci. 10, 243-252.

Leach, J.E., Cantrell, M.A. and Sequiera, L. (1982) A hydroxyproline-rich bacterial agglutinin from potato, extraction, purification and characterisation. Plant Physiol. 70, 1353-1358.

Leder, P., Tiemier, D. and Enquist, L. (1977) EK2 derivatives of bacteriophage lambda useful in the cloning of DNA from higher organisms: the gt WES system. Science 196, 175-177.

Levasseur, M.D. (1988) Comparative studies on the nucleotide sequence of pea seed storage protein genes. Ph.D. Thesis, University of Durham.

Lindhal, G., Sironi, G., Bialy, H., and Calendar, R. (1970) Bacteriophage lambda abortive infection of bacteria lysogenic for phage P2. Proc. Nat. Acad. Sci. 66, 592-596.

Lloyd, A.M. *et al.* (1986) Transformation of *Arabidopsis thaliana* with *Agrobacterium tumefaciens*. Science 234, 464-466.

Loenen, W.A.M. and Brammar, W.J. (1980) A bacteriophage lambda vector for cloning large DNA fragments made with several restriction enzymes. Gene 20, 249-259.

Lutke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F. and Scheele, G.A. (1987). EMBO J. 4, 883-889.

Lycett, G.W., Croy, R.R.D., Shirsat, A.H. and Boulter, D. (1984) The complete nucleotide sequence of a legumin gene from pea (*Pisum sativum* L). Nucl. Acids Res. 12, 4493-4506.

Lycett, G.W., Croy, R.R.D., Shirsat, A.H., Richards, D.M. and Boulter, D. (1985) The 5' flanking regions of three pea legumin genes; comparison of the DNA sequences. Nucl. Acids Res. 13, 6733-6743.

Lycett, G.W., Delauney, A.J., Gatehouse, J.A., Gilroy, J., Croy, R.R.D. and Boulter, D. (1983) The vicilin gene family of pea (*Pisum sativum* L): a complete cDNA sequence for preprovicilin. Nucleic Acids Res. 11, 2367-380.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, New York.

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connel, C. and Quon, D. (1978) The isolation of structural genes from libraries of eukaryotic DNA. Cell 15, 687-701.

Martinez-Zapater, J.M., Estelle, M. and Somerville C.R. (1986) A highly repeated DNA sequence in *Arabidopsis thaliana*. Mol. Gen. Genet. 204, 417-423.

Matta, N.K., Gatehouse, J.A. and Boulter, D. (1981) Molecular and subunit heterogeneity of legumin of *Pisum sativum* L (garden pea): a multidimensional gel electrophoretic study. J. Exp. Bot. 32, 1295-1307.

McNeil, M., Darvill, A.G., Fry, S.C. and Albersheim, P. (1984) Ann. Rev. Biochem. 53, 625-663.

Mellon, J.E., Helgeson, J.P. (1982) Interaction of hydroxyproline-rich glycoprotein from tobacco callus with potential pathogens. Plant Physiol. 70, 401-405.

Messing, J., Geraphty, D., Heidecker, G., Hu, N.T., Kridl, J. and Rubenstein, I. (1983) Plant gene structure. I. "Genetic engineering of plants" ed. Meredith, C.P., Kosuge, T., Hollaender, A.

Meyerowitz, E.M. (1987) *Arabidopsis thaliana*. Ann. Rev. Genet. 1987 21, 93-111.

Meyerwitz, E.M. and Pruitt, R.E. (1985) *Arabidopsis thaliana* and plant molecular genetics.

Miller, J.H. (1972) Experiments in molecular genetics. Cold Spring Harbor Lab., New York.

Miller, K. (1987) Gel electrophoresis of RNA. Focus (Bethesda Research Labs) 9:3, 14-15.

Millerd, A. and Spencer, D. (1974) Changes in RNA synthesising activity and template activity in nuclei from cotyledons of developing pea seeds. Aust. J. Plant Physiol. 1, 331-341.

Mizusawa, S. and Ward, D.F. (1982) A bacteriophage vector for cloning with BamHI and Sau3A. Gene 20, 317-322.

Morton, H., Evans, I.M., Gatehouse, J.A. and Boulter, D. (1983) Cell free translation systems from new sources: pea cotyledons and rice germ. Phytochem. 22, 25-26.

Murray, K. and Murray, N.E. (1975) Phage lambda receptor chromosomes for DNA fragments made with restriction endonuclease 3 of *Haemophilus influenza* and restriction endonuclease 1 of *Escherichia coli.*

Neilsen, N.C. (1984) The chemistry of legume storage proteins. Phil Trans. R. Soc. Lond. B. 304, 287-296.

Osborne, T.B. (1924) 'The vegetable proteins'. Longmans, Green and Company, London.

Otto, A., Kraft, R. and Etzold, G. (1984) N-terminal sequence analysis of basic subunits of legumin from *Vicia faba* by solid-phase sequencing. Kulturpflanze 32, S219-S221.

Pruitt, R.E. and Meyerowitz, E.M. (1986) Characterisation of the genome of *Arabidopsis thaliana*. J. Mol. Biol. 187, 169-83.

Pusztai, A. and Stewart, J.C. (1980) Molecular size, subunit structure and microheterogeneity of glycoprotein II from the seeds of kidney bean (*Phaseolus vulgaris* L). Biochem. Biophys. Acta. 623, 418-428.

Redei, G.P. (1975) *Arabidopsis* as a genetic tool. Ann. Rev. Genet. 9, 111-127.

Redei, G.P. and Perry, C.M. (1971) Submerged aseptic culture of intact plants in liquid medium. Arab Inf. Serv. 8, 34.

Rimm, D.L., Horness, D., Kucera, J. and Blattner, F.R. (1980) Construction of coliphage lambda charon vectors with BamHI cloning sites. Gene 12, 301-309.

Roberts, R.J. (1981) Restriction and modification enzymes and their recognition sequences. Nucl. Acids Res. 9, 75-96.

Schroeder, H.E. (1983) Quantitative studies on the cotyledonary proteins in the genus *Pisum*. J. Sci. Food Agric. 33, 623-633.

Seed, B., Parker, R. and Davidson, N. (1982) Representation of DNA sequences in recombinant DNA libraries prepared by restriction enzyme partial digestion. Gene 19, 201-209.

Sawyer, R.M. (1986), Isolation of a vicilin gene from pea (*Pisum sativum* L.), and nuclease sensitivity of seed storage protein genes in pea chromatin. Ph.D. thesis, University of Durham.

Shewry, P.R., Miflin, B.J., Forde, B.G. and Bright, S.W.J. (1981) Conventional and novel approaches to the improvement of nutritional quality of cereal and legume seeds. Sci. Prog. Oxf. 67, 575-600.

Sherratt, D.J. (1979) Plasmid vectors for genetic manipulation *in vitro*. Biochem. Soc. Symp. (1979), 44, 29-36.

Shirsat, A.H. (1984) A gene for legumin: a major storage protein of *Pisum sativum* L. Ph.D. thesis, University of Durham.

Showalter, A.M., Bell, J.N., Cramer, C.L., Bailey, J.A., Varner, J.E. and Lamb, C.J. (1985) Accumulation of hydroxyprotein-rich glycoprotein mRNAs in response to fungal elicitor and infection. Proc. Nat. Acad. Sci. 82, 6551-6555.

Smith, D.L. (1973) Nucleic acid protein and starch synthesis in developing cotyledons of *Pisum arvense* L. Ann. Bot. 37, 795-804.

Smith, M. (1981) Characterisation of carrot cell wall protein. II. Immunological study of cell wall protein. Plant Physiol.

Smith, J.J., Muldoon, E.P. and Lamport, D.T.A. (1984) Isolation of extensin precursors by direct elution of tomato cell suspension cultures. Phytochem., 23, 1233-1239.

Smith, J.J., Muldoon, E.P., Willard, J.J. and Lamport, D.T.A. (1986) Tomato extensin precursors P1 and P2 are highly periodic structures. Phytochem. 25, 1021-1030.

Southern E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503-517.

Sparrow, A.J., Price, H.J. and Underbrink, A.G. (1972) A survey of DNA content per cell and per chromosome of prokaryotic and eukaryotic organisms: some evolutionary considerations. Brookhaven Symp. Biol. 23, 451-494.

Stafstrom, J.P. and Staehelin, L.A. (1986) Crosslinking patterns in salt-extractable extensin from carrot cell walls. Plant Physiol. 81, 234-241.

Stafstrom, J.P. and Staehelin, L.A. (1988) The role of carbohydrate in maintaining extensin in an extended conformation. Plant Physiol. 81, 242-246.

Stuart, D.A., Mozer, T.J. and Varner, J.E. (1982) Cytosine-rich messenger RNA from carrot root disc. Biochem. Biophys. Res. Comm. 105, 582-588.

Stuart, D.A. and Varner, J.E. (1980) Purification and characterisation of a salt-extractable hydroxyproline-rich glycoprotein from aerated carrot discs. Plant Physiol. 66, 787-792.

Summers, D., Yaish, S., Archer, J. and Sherrat, D. (1985) Multimer resolution system of ColE1 and ColK; localisation of the crossover site. Mol. Gen. Genet. 201, 334-338.

Thomas, M., Cameron, J.R. and Davis, R.W. (1974) Viable molecular hybrids of bacteriophage lambda and eukaryotic DNA. Proc. Natl. Acad. Sci. U.S.A. 71, 4579-4583.

Thomas, P.S. (1980) Hybridisation of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. U.S.A. 77, 5202-5205.

Thompson, A.J. (1989) Regulation of gene expression in developing pea seeds. Ph.D. thesis, University of Durham.

Thompson, J.A., Millerd, A. and Schroeder, H.E. (1979) Seed protein improvement in cereals and grain legumes. Proc. Symp. Neuberberg. 1, 231-2140. 1, A.E.A. Vienna.

Thompson, J.A., Shroeder, H.E. and Dudman, W.F. (1978) Cotyledonary storage protein in *Pisum sativum*. I. Molecular heterogeneity. Aust. J. Pl. Physiol. 5, 263-279.

Van Holst, G.J. and Varner, J.E. (1984) Reinforced polyproline II conformation in a hydroxyproline-rich cell wall glycoprotein from carrot root. Plant. Physiol. 74, 247-251.

Vieira, J. and Messing, J. (1982) The pUC plasmids on M13 mp7 derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene. 19, 259-268.

Wobus, U., Baumlein, H., Bassuner, R., Grafe, R., Jung, R., Nuntz, K., Saalbach, G. and Wescheke, W. (1985) Cloning and characterisation of *Vicia faba* seed storage protein genes. Biol. Zentralblatt.

Wood, W.B. (1966) Host specificity of DNA produced by *Escherichia coli* bacterial nations affecting the restriction and modification of DNA. J. Mol. Biol. 16, 118-133.

Wright, D.J. and Boulter, D. (1974) Purification and subunit structure of legumin of *Vicia faba* L (broad bean). Biochem. J. 141, 413-418.

Zhao, W-M., Gatehouse, J.A. and Boulter, D. (1983) The purification and partial amino acid sequence of a polypeptide from the glutelin fraction of rice grains; homology to pea legumin. FEBS Letts. 162, 96-102.